

# Integrated Species Distribution Modelling System - a user friendly front end to the GARP modelling toolkit

(BY) Timothy P. Sutton

Thesis presented in partial fulfilment of the requirements  
for the degree of Master of Arts at the  
University of Stellenbosch.

April 2004

*Pectora laborant cultus recti*

## **DECLARATION**

" I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any University for a degree.

Signature:

Date:

"

---

## ABSTRACT

At a social, ecological and biological level it is important that we gain a better understanding of species distribution and the constraints to species distribution. Various modelling tools and approaches are available to provide this type of functionality. The GARP (Genetic Algorithm for Rule set Production) Modelling System (GMS) was selected because of its strong predictive modelling abilities and its ability to represent the results of model iterations in both a tabular and cartographic manner. A shortcoming in this system was identified in that it requires strong information technology skills in order to carry out the modelling process. This can be attributed to the lack of a user-friendly interface to the system. In order to address this a loosely coupled system was developed that provides an easy to use web-based front end to the GMS. This Integrated Modelling System extends the core functionality of the GMS by providing a system that provides detailed history for each analysis, allows fine tuning of the modelling process, integrates directly with a biodiversity database containing specimen observations, and provides a simple 'wizard' interface to the modelling process.

---

## OPSOMMING

Van 'n sosiale, ekologiese en biologiese standpunt is dit belangrik dat ons spesies verspreiding en die beperkings daarvan verstaan. 'n Verskeidenheid sagteware pakkette en metodologieë is beskikbaar om spesies verspreiding te modelleer. Die GARP (Genetic Algorithm for Rule set Production) sagteware was gebruik vir sy sterk voorspellingsvermoë, en sy kapasiteit vir kartografiese en tubulêre tentoonstelling van model resultate. 'n Tekortkoming met hierdie stelsel was geïdentifiseer - dit is nie gebruikersvriendelik nie en gebruikers het sterk informasie tegnologie vermoëns nodig. Om hierdie tekortkominge aan te spreek was 'n sagteware program ontwerp wat van GARP gebruik maak deur middel van 'n webblaaier.

Hierdie geïntegreerde stelsel bou op die basiese funksionaliteit van GARP om 'n werk omgewing te skep wat 'n gedetailleerde geskiedenis van elke model stoor, fyn beheer oor die model toelaat, direk met 'n biodiversiteits databasis koppel, en van 'n eenvoudige 'wizard' stelsel gebruik maak om gebruikers opsies te bepaal.

## ACKNOWLEDGEMENTS

I would like to thank the following people for their invaluable assistance. Andrew Turner (WCNCB) and Dr Ernst Baard (WCNCB) for the useful input and access to the biodiversity database they provided. Dr David Stockwell for making the GARP software available to me. I would also like to thank Marcelle Sutton for being such a patient wife while I worked on this thesis.

Many thanks to Dr Edwin Hees for proofreading this document.

I would like to dedicate this paper to Chris Burgers (1952 - 2002) whose enthusiasm and in-depth knowledge of the Cape Floristic Kingdom were a source of great inspiration.

# CONTENTS

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	SPECIES DISTRIBUTION MODELLING IN CONTEXT . . . . .	1
1.2	BACKGROUND TO GARP . . . . .	3
1.3	THE GARP MODELLING PROCESS - A BRIEF OVERVIEW . . . . .	6
1.4	RESEARCH GOALS . . . . .	8
<b>2</b>	<b>SYSTEM ARCHITECTURE &amp; DESIGN METHODOLOGY</b>	<b>10</b>
2.1	HARDWARE COMPONENTS . . . . .	11
2.2	SOFTWARE COMPONENTS . . . . .	11
2.2.1	Database services . . . . .	12
2.2.2	Application services . . . . .	13
2.2.3	Client services . . . . .	15
2.2.4	System integration and detailed system design . . . . .	16
<b>3</b>	<b>ISMS ADMINISTRATION</b>	<b>21</b>
3.1	USER MANAGEMENT . . . . .	22
3.2	MANAGING PROJECTIONS . . . . .	23

3.2.1	Projection definitions . . . . .	24
3.2.2	Ellipsoid definitions . . . . .	24
3.2.3	User projection definitions . . . . .	26
3.3	EXTENT DEFINITIONS . . . . .	26
3.3.1	Independent data layers . . . . .	27
3.4	MODEL ENVIRONMENT CONFIGURATION (PARAMETER SETS) . . . . .	29
3.5	PREPARATION OF LAYER SETS . . . . .	32
<b>4</b>	<b>ISMS USAGE</b>	<b>35</b>
4.1	RUNNING A MODEL ITERATION . . . . .	35
4.2	SELECTION OF A PARAMETER SET . . . . .	37
4.3	SELECTION OF A LAYERSET . . . . .	37
4.4	SELECTION OF A TARGET TAXON . . . . .	38
4.5	OPTIONAL SELECTION OF A SURROGATE ABSENCE TAXON . . . . .	39
4.6	SELECTION OF DATA-QUALITY CRITERIA . . . . .	41
4.7	FEEDBACK FROM THE RASTERISATION PROCESS . . . . .	41
4.8	HISTORY BROWSER . . . . .	45
4.9	SPATIAL MANIPULATION AND VISUALISATION OF RESULTS . . . . .	45
4.9.1	Importing the dataset into Grass 5.10 . . . . .	47
4.9.2	Viewing model results in 3D . . . . .	47
<b>5</b>	<b>CONCLUSION</b>	<b>49</b>
5.1	REALISATION OF DESIGN GOALS . . . . .	50

5.2	ADVANTAGES OF THE SYSTEM . . . . .	51
5.3	SHORTCOMINGS OF THE SYSTEM . . . . .	52
5.4	FUTURE RESEARCH DIRECTIONS . . . . .	53
	<b>REFERENCES</b>	<b>55</b>
	<b>A APPLICATION FLOW CHART</b>	<b>60</b>
	<b>B CLASS DIAGRAM FOR THE ISMS APPLICATION</b>	<b>62</b>
	<b>C ISMS PROJECTION DATABASE</b>	<b>64</b>
	<b>D ISMS ELLIPSOID DATABASE</b>	<b>70</b>
	<b>E ISMS INSTALLATION</b>	<b>72</b>
E.1	INSTALLING DEPENDENT SOFTWARE . . . . .	72
E.1.1	mySQL installation . . . . .	72
E.1.2	PostgreSQL installation . . . . .	72
E.1.3	LaTeX installation . . . . .	73
E.1.4	Installation of Proj4 . . . . .	73
E.1.5	Installation of PHP with mySQL drivers . . . . .	73
E.1.6	Optional installation of the Grass GIS . . . . .	75
E.1.7	Start system services . . . . .	76
E.2	POSTGRESQL CONFIGURATION . . . . .	76
E.2.1	Create a non-privileged user . . . . .	76



E.2.2	Create a privileged user . . . . .	76
E.2.3	Allow TCP/IP access to PostgreSQL . . . . .	77
E.2.4	Install sample specimen database . . . . .	77
E.3	MYSQL CONFIGURATION . . . . .	78
E.3.1	Creating a mySQL user . . . . .	78
E.3.2	Loading the ISMS database . . . . .	79
E.4	ISMS APPLICATION DEPLOYMENT . . . . .	79
E.4.1	Allowing Apache access to the 'AT' command scheduler . . . . .	79
<b>F</b>	<b>INDEPENDENT VARIABLE PREPARATION - A DETAILED AC- COUNT</b>	<b>80</b>
F.1	PREPARING THE MASK LAYER . . . . .	81
F.1.1	Obtaining the file properties of an ASCII grid file . . . . .	82
F.1.2	A sample 'paramete' listing for the mask file . . . . .	82
F.1.3	Carrying out the rasterisation process . . . . .	82
F.2	RASTERISING CATEGORICAL DATA . . . . .	83
<b>G</b>	<b>SAMPLE OUTPUT FROM AN ISMS MODEL ITERATION</b>	<b>85</b>
G.1	MODEL ENVIRONMENT . . . . .	86
G.1.1	Extents of the modelled area . . . . .	86
G.1.2	Parameters used . . . . .	86
G.1.3	Layers used . . . . .	87
G.2	MODEL RESULTS . . . . .	89

---

G.2.1	Cartographic outputs . . . . .	89
G.2.2	Initial outputs . . . . .	93
G.2.3	Presample outputs . . . . .	93
G.2.4	Verify outputs . . . . .	93
G.2.5	Explain outputs . . . . .	94
G.2.6	Translate outputs . . . . .	94
<b>H</b>	<b>GRASS PROJECT SETUP</b>	<b>140</b>

# FIGURES

1.1	Using interpolation, locality data (represented as red dots) can be used to infer an organism’s distribution range (represented by area in green). . . . .	2
1.2	A more sophisticated approach to interpolation of locality data is to develop a habitat preference envelope. . . . .	3
1.3	Sample output from a GARP model predicting <i>Cordylus cordylus</i> distribution in the Western Cape Province, South Africa. Areas shaded green represent presence prediction. . . . .	8
2.1	The ISMS hardware and networking infrastructure is composed of two servers connected to each other over a local Ethernet network. . . . .	12
2.2	The ISMS Software architecture. . . . .	15
2.3	High-level user interface flow chart for the execution of a single GARP model iteration. . . . .	18
2.4	A simplified database schema for the ISMS back-end (where application-level data are stored). . . . .	19
3.1	ISMS User Administration Screen. New user accounts are created here and existing user accounts can be modified using this interface. . . . .	23
3.2	The ISMS Ellipsoid database is generated using the PROJ4 projection utility. On completion of the import process, the complete ellipsoid listing is displayed to the administrator. . . . .	25

3.3	ISMS User Projections are defined by combining specific parameters with standard projection and ellipsoid definitions. . . . .	27
3.4	Each independent variable layer that is available to end users is referenced in the Layers table. . . . .	29
3.5	The ISMS Layer management module is used to define the attributes and file system location of data layers stored in the GARP binary format. . . . .	30
3.6	Parameter storage in the ISMS database. . . . .	30
3.7	A sample parameter set object record. . . . .	31
3.8	The parameter set definition form is dynamically rendered based in Control Type field entries in the parameter set Object table. . . . .	32
3.9	The ISMS layer set definition process. . . . .	33
3.10	Associating layers with a layer set. . . . .	34
4.1	Diagrammatic representation of the 'wizard' workflow for carrying out a GARP model iteration using ISMS. . . . .	36
4.2	The modelling process is commenced with a small information screen. . . . .	36
4.3	During the parameter set selection phase, the user is prompted to select one of the predefined parameter sets they have created. . . . .	37
4.4	During the layer set selection stage, a predefined set is chosen from a list of the layer associations compiled by the user. . . . .	38
4.5	The presence taxon search form. Wildcard searches are implemented here, so partial matches to the search string will be found. . . . .	39
4.6	The presence taxon search results form. The number of records for each matching taxon is shown in parentheses after the taxon name. . . . .	40
4.7	An optional surrogate absence taxon may be selected in a similar manner to the selection of the target (presence) taxon. . . . .	41

4.8	The number of records for each matching surrogate absence taxon is indicated in brackets after the taxon name. . . . .	42
4.9	Metadata describing the temporal, spatial and taxonomic accuracy of the data are used to filter the independent variable dataset used. . . . .	43
4.10	Feedback from the rasterisation process allows users to decide if they wish to continue with the model iteration. . . . .	44
4.11	The ISMS model history browser allows individual users to manage their personal collection of model results. . . . .	46
4.12	A GARP model displayed in 3D relief using the GRASS GIS NVIZ module. . . . .	48
A.1	The complete flow chart for the ISMS application. . . . .	61
B.1	Class diagram for the ISMS application. . . . .	63
G.1	predI (Predicted presence / absence) . . . . .	89
G.2	specI (Rule specialisation) . . . . .	90
G.3	probi (Probability of occurrence) . . . . .	91
G.4	stocI - This is an experimental undocumented output from the GMS. . . . .	92
H.1	Setting up a new location in Grass. . . . .	140
H.2	New location confirmation in Grass. . . . .	140
H.3	Coordinate system selection for the new project. . . . .	141
H.4	Setting the location description, projection and ellipsoid in Grass. . . . .	141
H.5	Defining the region of interest for the new location. . . . .	142

# TABLES

1.1	A sample 'paramete' file used by the GMS system to control the model process.	7
3.1	The parameter objects table. . . . .	31
4.1	Sample listing from the biodiversity database illustrating data stored in decimal degrees (first record) and quarter-degree Square (second record). . . . .	48

## CHAPTER 1: INTRODUCTION

### 1.1 SPECIES DISTRIBUTION MODELLING IN CONTEXT

We live in a rapidly changing society, with a growing population and the accompanying pressure this places on our natural resources. Land-use change is considered an important determinant of changes in biodiversity (Sala, Chapin III, Armesto, Berlow, Bloomfield, Dirzo, Huber-Sanwald, Huenneke, Jackson, Kinzig, Leemans, Lodge, Mooney, Oesterheld, LeRoy Poff, Sykes, Walker, Walker & Wall 2000). Increasingly biologists and conservationists are required to use their specialised knowledge to guide development in order to pre-empt its potentially adverse impacts. In addition to local impacts from development, the projected climate changes because of global warming are expected to have a significant impact on climate and consequently faunal and floral distribution patterns (Rutherford, Midgely, Bond, Powrie, Roberts & Allsopp 1999; Van Jaarsveld, Chown, Erasmus, Kshatriya & Wessels 2000; Erasmus, Van Jaarsveld, Chown, Kshatriya & Wessels 2002). The sheer volume of taxa and their varied spatial distribution present problems to biologists and planners - often the biologists simply do not know where species occur (Berendsohn 2000)!

One approach that can improve understanding of species distribution is to use interpolation techniques to generalise sparse point locality data, usually collected at a fine temporal and spatial resolution (Hollander, Davis & Stoms 1993; Delaney 2001). If appropriate techniques can be found, distribution modelling offers an inviting alternative to time-consuming and expensive field-collection exercises (Austin 1991).

Traditional GIS interpolation algorithms (see Figure 1.1) such as nearest neighbour, spline curve and inverse distance-weighted interpolation are of limited use for distribution modelling. This is because the actual observation sites will in many cases not be suitably spatially disaggregated to allow valid distribution deductions to be made, and because species distribution usually occurs in response to complex habitat requirements. GIS packages have been applied to the species-distribution problem through the use of simple overlay techniques (Delaney 2001; Nieminen 2002; Brown, Swetnam, Manchester, Mountford, Caldow & Treweek 1996) and as a useful tool for the collation and visualisation of biological data (Cocks & Baird 1991).



These tools require a high degree of manual interaction and do not employ sophisticated statistical techniques to develop a range map prediction. An additional problem with the overlay techniques as used by Delaney (2001) and Nieminen (2002) is that they require *a priori* understanding of the habitat requirements for the taxon under study.

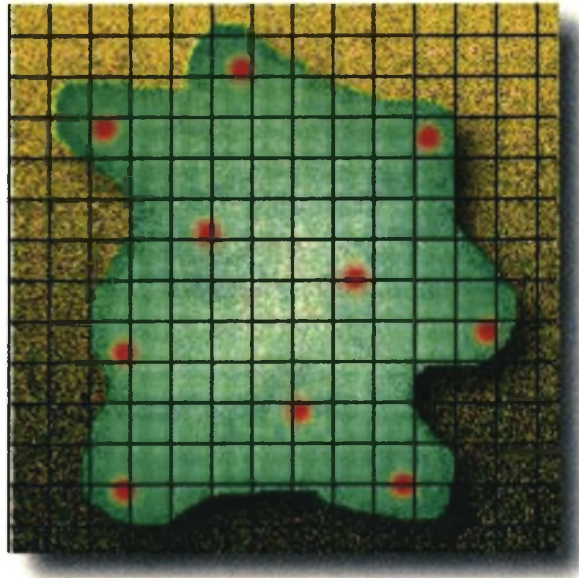


Figure 1.1: Using interpolation, locality data (represented as red dots) can be used to infer an organism's distribution range (represented by area in green).

To address these problems statistical models that automate the development of habitat preference envelopes have been developed. These preference envelopes correlate specimen localities with environmental data to determine the parameters of acceptable habitat conditions for an organism - see Figure 1.2. This is then used as a basis for modelling the predicted distribution of the species across the landscape. Tokumine, McClean & Lovett (2002) organise the statistical models used to predict species distribution data into two categories: generalised linear models and bioclimatic envelope models. Generalised linear models use largely complete dependent variable datasets (i.e. presence and absence data). Bioclimatic envelope models BIOCLIM (Busby 1991) type models are typically derived from incomplete dependent variable datasets (i.e. presence only data). Collecting presence/absence data presents considerable difficulty (particularly with animal taxa) because numerous site visits are required to ascertain that an organism is indeed absent (Hirzel, Hausser, Chessel & Perrin 2002). It is the spatially disaggregated, incomplete nature of these datasets that has upped the ante in the drive to produce more robust prediction techniques. To this end various approaches have been developed to predict distribution including (but not limited to!) Bioclimatic Envelope Modelling (Nix 1986), GAP Analysis (Davis, Stoms, Hollander, Thomas, Stine, Odion, Borchert, Thorne, Gray, Walker, Warner & Graae 1998), logistical regression models (Tokumine *et al.* 2002),



multivariate regression trees (De' ath 2001), fuzzy k-means modelling (Burrough & McDonnel 2000), principal component analysis (Caithness 1995), Bayesian statistics (Pereira & Itami 1991) and Genetic Algorithms (Stockwell 1999; Tokumine *et al.* 2002).

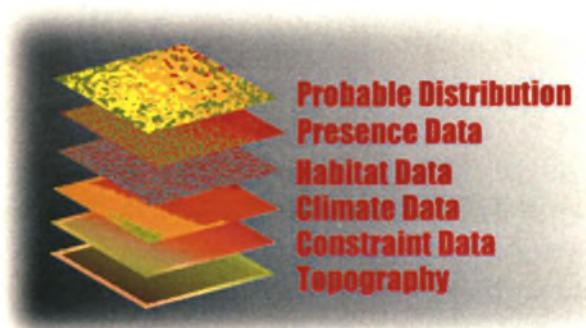


Figure 1.2: A more sophisticated approach to interpolation of locality data is to develop a habitat preference envelope.

The range maps produced from these tools can provide useful insight into present and potential future distribution of organisms, and for steering planning and development. However there is a need to provide access to distribution modelling applications to biologists in an easy to use format which allows the user to quickly generate *ad hoc* predicted distribution maps for a given taxon. The value of such tools is increased when they can provide explanatory rules that explain the distribution in terms of the independent variables used to drive the model. After preliminary testing (Sutton 1999) the GARP Modelling System (GMS) was selected for use as the modelling engine for this purpose. A brief outline of the GMS is provided in the following section.

## 1.2 BACKGROUND TO GARP

The goal of predictive distribution mapping is to establish the ecological niche that is occupied by an organism and to identify all parts of the study area that match the identified niche. The niche of an organism in this context is expressed in terms of the environmental factors such as elevation, soil, aspect and climate (Franklin 1995).

Machine learning software the results of past iterations to improve model results. The GARP Modelling System (GMS) is a machine learning approach to the problem of predicting the distribution of species through the production of bioclimatic envelopes. The GMS utilises genetic algorithms for the development of explanatory rules. This form of species distribution

modelling has been found to produce better results than expert systems, in terms of both speed of model execution and predictive ability (Stockwell, Davey, Davis & Noble 1990).

Ideally, production of bioclimatic envelopes should be treated as an NP-complete problem (Garey & Johnson 1979), where every permutation of species observation site habitat data is explored in order to ascertain optimal model fit. However this 'brute force' approach to bioclimatic envelope development is computationally intensive and leads to extremely long computation times (Tokumine *et al.* 2002). The Genetic Algorithms utilised in the GMS bypass the need for a 'brute force' approach by employing computational heuristics using a natural selection process. In this way the GMS can develop a near optimal bioclimatic envelope, with the benefit of much reduced computational times. In the GMS an organism's response to the environment is explained through the development of rule sets. Each rule in the rule set consists of a precondition and a conclusion (Stockwell & Noble 1992) that takes the form of an 'if...then...' statement, for example:

```
IF  maxtemp=20.06838  AND  mintemp=17.5  AND  pptseas=231.61661
THEN Taxon=PRESENT
```

Four types of rules are produced by the GMS:

1. Atomic rules (the simplest rule that uses single values as illustrated in the example above);
2. Logit rules (where the output of each rule is expressed as a probability based on a logistical regression analysis);
3. Bioclimatic envelope (a 95 percentile range statistical envelope in which an organism is known to occur);
4. Range rules (expressed as the upper and lower bounds of acceptability for each climatic factor).

The result sets produced by the model are suitable for being used to predict the distribution of an organism (Stockwell & Gitay 1992). The GMS addresses a number of problems inherent to modelling biological data, including (Stockwell & Peters 1999):

- The 'missing value' problem, where, in general, biological surveys do not record where organisms are absent, only where they are present;

- The ‘variable prior proportions’ problem where, recordings of presence and absence vary with both species abundance and the amount of effort spent in favourable habitat;
- The ‘large/small dataset’ problem, where, in the former instance, overly large datasets lead to long computation times with little gain in model performance, and in the latter instance small datasets limit the range of sample points from which the model can be developed.

The GARP application is implemented as a suite of programs that can be executed within a UNIX or GNU/Linux<sup>1</sup> operating system environment. In addition to tabular rule sets, the GMS is also capable of producing cartographic output representing the predicted distribution of the taxon under study. This output can be produced in ImageMagick ‘.miff’ format, or Arc/Info ‘ASCII Grid’ format.

Validating the results of species distribution tools is problematic, as the impetus for using such tools is usually that the distribution of the taxon is poorly understood, leaving little against which to measure the results. One approach to model validation is to count the errors of commission (false positives) and omission (false negatives) (Fielding & Bell 1997). In the former instance, areas where the taxon is known not to occur are predicted as ‘presence areas’. In the latter instance, areas where the taxon is known to occur are not included in the range map produced. The GMS provides a simple mechanism for description of the model validation process through the use of a ‘confusion matrix’ (Stockwell 1999). The ‘confusion matrix’ provides a summary of the number of omissions. More problematic, however, is the problem of commission - given that the goal of the analysis is to assert the presence of a taxon in areas which have not been surveyed. Errors of commission are evidenced by an organism being predicted as present in an area where it is in fact known to be absent (Anderson, Lew & Townsend Peterson 2003). During testing of the GMS (Sutton 1999), validating model outputs was carried out through the use of expert (human) assessment. Fielding & Bell (1997) offer further methodologies for validating model results such as prediction error cost criteria whereby (for example) false positives are rated as ‘more costly’ than false negatives.

The GARP modelling process is discussed in more detail in the following section.

---

<sup>1</sup>The GNU/Linux operating system is an amalgam of the Linux kernel and the GNU (Gnu’s Not Unix) utilities which provide common Unix-like software packages such as the Bash shell. The popular press generally refers to the GNU/Linux as simply ‘Linux’ - this is in fact a misnomer (Stallman 2002).



### 1.3 THE GARP MODELLING PROCESS - A BRIEF OVERVIEW

The execution of the GMS requires the user to complete a series of activities. The first task is to produce raster format independent variable datasets. To this end the RASTERIZ application is used to convert either tabular data or data in Arc/Info ASCII grid format into the GARP binary raster format. The independent variable data can be categorical data (e.g. vegetation types) or continuous range data (e.g. average precipitation in millimetres). Independent data will typically be prepared first in a desktop GIS application so that they all provide a continuous coverage over the same geographical area. The GARP binary format normalises the data to a range of 0-255 (i.e. one byte per cell), allowing data of differing types and numerical scales to be compared. Once created, the datasets can be used for carrying out one or more model iterations.

Dependent variable data (i.e. distribution data) are rasterised in a similar manner using the RASTERIZ program. In many cases these data will first be extracted from a GIS or relational database using a SQL query and stored in a 'flat' text file in a format suitable for the RASTERIZ application. An additional step at this stage may involve altering the geographical projection prior to the RASTERIZ process so that the dependent data projection matches that of the independent variable datasets. Typically this process will be carried out using a GIS application such as ArcView.

From the dependent variable dataset a training and a test dataset are produced (using the EXPLAIN application). The training dataset is used to build the model. The test dataset is used to provide an independent assessment of model performance. The VERIFY program is then used to generate the GARP rule set. Following this, the PREDICT application is used to apply the model to predict the probability of occurrence at each cell in the raster matrix. Typically, the results of the PREDICT application are 'piped' from the standard output stream into the IMAGE application, which will then produce a cartographic output of the model results, either in ImageMagick '.miff' format, or in 'ASCII/grid' format. A final application - TRANSLATE - will produce a readable printout of the final rule set in tabular format. For a more detailed explanation of the GARP modelling process, the reader is referred to Stockwell (1999) and Payne & Stockwell (1999).

The user interface to the GMS uses the command line of a UNIX shell (e.g. BASH). Typically the GARP programs are invoked using UNIX pipes and redirects to channel user data and application outputs between the various constituent GARP modules outlined above. The GMS model behaviour is governed by 'paramete' files - parameter files stored in ASCII format. The paramete file includes information about which independent variable datasets should be used,

the type of values these independent variable datasets contain, the geographical extent of the study area, as well as information about the dependent variable data layer. In addition to the data layer definitions, a range of parameters can be used to govern the model behaviour. For example, specifying:

Bioclimflag 1

would induce the GMS to use bioclimatic envelope rules during rule set production. A sample paramete file is provided as Table 1.1. Having prepared an appropriate paramete file, the various GARP applications are launched in turn at the command line. If the user wishes to repeat the model with different parameters, the paramete file is edited and the process is repeated. On invocation of the model, tabular model results are sent to stdout (which by default is the console display). Cartographic output can be produced using a combination of the GARP IMAGE application and either the ImageMagick 'convert' application (see Figure 1.3), or imported into a desktop GIS package such as ArcView or GRASS. This latter process is demonstrated in Chapter 4.9.

With an understanding of the options for carrying out species distribution modelling, and the challenges facing non-technical users when confronted with the non-linear command line interface of the GMS, it was possible to establish clear goals for this project, which are outlined in the following section.

Table 1.1: A sample 'paramete' file used by the GMS system to control the model process.

Edgeflag							1
Overlayflag							1
Bioclimflag							1
Maskvalues							255
Bioclimonlyflag							1
Columns		-330101.00000		328899.00000			500.00000
Rows		-4061519.00000		-3552019.00000			500.00000
Variable	0	Taxon	0	1	s	species	%2.0f
Variable	1	wcdem	0.00	2247.00	r	m	%2.0f
Variable	2	geology	1.00	12.00	c	catno	%2.0f
Variable	3	avgrain	6666.13	3264.52	r	mm	%2.0f
Variable	4	rivdist	0.00	41067.02	r	m	%2.0f



Figure 1.3: Sample output from a GARP model predicting *Cordylus cordylus* distribution in the Western Cape Province, South Africa. Areas shaded green represent presence prediction.

## 1.4 RESEARCH GOALS

From the brief explanation of the GARP modelling process provided in the previous section, it becomes apparent that the operation of the GMS requires a relatively high level of technical expertise, which can be very time consuming to acquire. Skills such as performing SQL queries, altering the geographic projection of data and carrying out console-based operations on a GNU/Linux server are not readily available amongst the biologists interested in the outputs of such a system. Consequently the need for a better environment for interacting with the GMS was identified and the following goals were identified as priorities for the Integrated Species Modelling System:

1. To provide seamless access to the Western Cape Nature Conservation Board's (WCNCB) specimen database, which includes a detailed taxonomic database for vertebrates and vascular plants for the Western Cape Province, South Africa;
2. To automate the projection of dependent variable data into the same projection system as that of the independent variable data used in a GARP model iteration;
3. To provide a graphical user interface (GUI) environment for carrying out model iterations and visualising the results of model iterations. The primary goal here is to allow biologists to explore interactively the effect of different model parameters and independent data layers for a given dependent variable. An important aspect of the proposed user environment was the abstraction of all the GMS applications into one integrated modelling environment that does not require specific knowledge of the operation of the GMS on the user's part;



4. To provide a model output management environment so that users can browse the collection of model outputs that are built up during GMS iterations;
5. To make the GMS available over an Intranet to multiple simultaneous users using a variety of operating systems and hardware resources;
6. To explore the concept of a loosely coupled system in order to produce a functional application with a unified user interface. The project also seeks to explore techniques for the integration of geographical data, tabular data and 'user space' applications across multiple physical devices and operating systems.

It should be noted that, while the proposed system aims to simplify the operation of the GMS, users of the system would still need to have a basic understanding of the concepts behind the GMS (as described in Payne & Stockwell (1999)). The reason for this is that the GMS should not be regarded as a 'black box' approach to modelling species distribution - there is a wide range of parameter permutations possible which can result in highly disparate model outputs (range predictions) for a given taxon. In a similar vein, it can be noted from the goals mentioned above that little attention has been placed on fidelity of model output to the 'on the ground' situation. The choice of modelling application to be used for this application should not be construed as an endorsement of GARP over other modelling approaches, and this thesis does not attempt to evaluate the relative strengths and weaknesses of the various distribution modelling tools available. For an assessment of the advantages and disadvantages of distribution modelling approaches, the reader is directed to Caithness (1995) and Fielding & Bell (1997). This thesis rather seeks to explore the feasibility of deploying sophisticated spatial analysis tools using Internet technologies and the loose coupling of software and hardware infrastructure. GARP is used as an example modelling tool to illustrate how such a system can be assembled.

In the following chapter (Chapter 2), the system design technical approach used to address the above-mentioned goals is described. In Chapter 3 the administrative requirements and user interface for the ISMS system are discussed. In Chapter 4 a description is given of how the ISMS system is used. In the concluding chapter, a discussion is provided of the successes and shortcomings of the final system.

## CHAPTER 2: SYSTEM ARCHITECTURE & DESIGN METHODOLOGY

The research goals described in the previous chapter aim to produce a single integrated environment for modelling the potential distribution of species. Thus the system was named the Integrated Species Modelling System (ISMS) and is referred to by this acronym for the remainder of this paper. The system design for the ISMS application was based on the loose coupling of a variety of software applications across various hardware devices. The concept of ‘coupling’ is used by software engineers to indicate the level of interconnection between various modules in a software architecture. It provides a qualitative measure of the interdependence between these modules. Software architectures with low levels of coupling are more easily understood and less prone to error propagation through the system (Pressman 2000). The \*NIX<sup>1</sup> environment is well suited to this development model - traditionally \*NIX applications are small, single-purpose programs that communicate with each other using pipes and redirection (Moritsugu, Edwards, S, Horvath, Marler, Pedersen, Pitts & Wilson 1998). A three-tiered approach to the development of this application was adopted (Williams & Lane 2002). The three-tiered application consists of three conceptual ‘layers’ providing:

1. The **client tier**, which allows the user to interact with the application;
2. The **middle tier**, implements application logic and communicates data between other tiers;
3. The **database tier**, stores data for query, update, insert and delete operations.

The three-tiered system design provides the advantage of separating the data, application logic and user interface to provide a more flexible growth path for the system. For example, new data sources can be added without requiring any changes to the application logic, different user interfaces can be developed using the same application logic layer, etc.

The three-tier system architecture will be used as a basis for the discussion of the ISMS architecture provided in the following sections.

---

<sup>1</sup>\*NIX has been used throughout this paper to refer to UNIX-like operating systems in a generic manner. It should be assumed by the reader that GNU/Linux is included within this umbrella term.



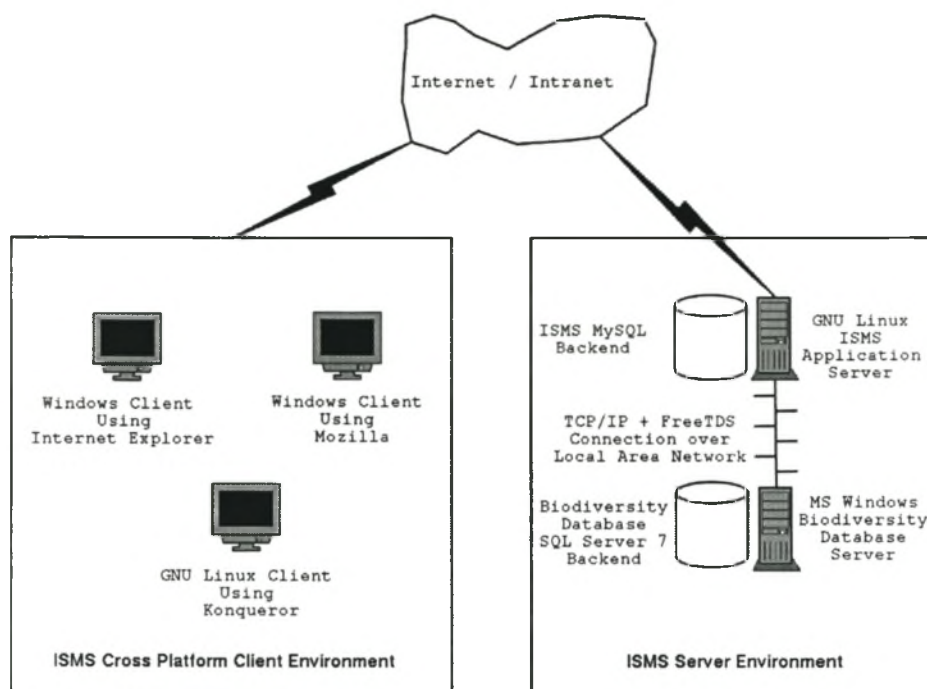
## 2.1 HARDWARE COMPONENTS

The database and application tiers were implemented using servers running Windows 2000 and GNU/Linux respectively. These were comprised of 'commodity off-the-shelf' (COTS) computer hardware (i.e. typical desktop machines). The use of COTS hardware for 'back-end' servers is becoming increasingly prevalent as COTS devices provide increasing levels of processing power at significantly lower prices than high-end servers such as offerings from 'brand-name vendors'. The database tier applications existed before the ISMS project's inception, resulting in a design constraint - the ISMS server would need to be able to communicate with the biodiversity database running on a different physical server on a different operating system. A similar design constraint was imposed by the GARP modelling software, which is written to run in a \*NIX environment. GNU/Linux was selected as a hosting environment for the GMS, because it is free and (important from the hardware perspective) runs on COTS i386 architecture hardware. The concept of 'free' as applied to GNU/Linux refers both to price (GNU/Linux is available for free) and to the freedom to modify any component of the system as the complete source code for the system is made available to the user (GNU/Linux can be freely modified). The two servers were installed in a dedicated server room on the same physical local area network (LAN). Because of the platform-independent nature of the ISMS application, the only restrictions on the client hardware platform were that ISMS clients needed to be able to run a web browser and have connectivity to the Intranet or Internet (depending on how the ISMS application was configured). For testing and development purposes, both Intranet and Internet configurations were used. A typical hardware layout for the system is illustrated in Figure 2.1.

Client machines connect to the ISMS application server via an Intranet/Internet link. With sufficient hardware infrastructure in place, the software components could be implemented. The rationale behind, and design of, the software architecture is outlined in the following section.

## 2.2 SOFTWARE COMPONENTS

The selection of software components for the system was governed by institutional issues (such as the need to integrate with existing systems), financial constraints (there was no funding available for capital expenses such as software or new computer hardware) and design requirements (the desire to develop a modular, extensible system). These constraints are reflected in the software architecture outlined in the following sections.



**ISMS Networking Architecture**

Figure 2.1: The ISMS hardware and networking infrastructure is composed of two servers connected to each other over a local Ethernet network.

### 2.2.1 Database services

Two database systems were involved in the development of the ISMS, the Western Cape Nature Conservation Board (WCNCB) Biodiversity Database (hereafter referred to as ‘the biodiversity database’), and a database for internal use of the ISMS (hereafter referred to as ‘the ISMS database’). The biodiversity database is a Microsoft SQL Server 7 relational database and forms the basis of the first-tier data-storage system within the three-tier application model. This database functions as a repository for all taxon observation data gathered in the field during WCNCB field observations and includes collection data from a number of external institutions. In addition, the biodiversity database houses a taxonomic database of vertebrate fauna and vascular plants for the Western Cape Province. The biodiversity database was developed independently of the ISMS project. Because of this, various minor adaptations needed to be made in order to incorporate the biodiversity database into the ISMS design. These modifications consisted mainly of creating additional views (predefined database queries) for obtaining species lists, species observation data and metadata related to the spatial, temporal and taxonomic accuracy of observation data. Another important modification that needed to be made to the database tier was to enable Tabular Data Stream (TDS) connections to the server

using the TCP/IP protocol. This was implemented in order to allow the ISMS application layer to connect directly to the database tier.

A separate database engine was implemented on the ISMS server using the MySQL database engine. This ISMS Database was used for the storage of ISMS administrative data such as user lists, projection lists, tabular model outputs, etc. This approach allowed for the separation of ISMS application-specific data from the biodiversity database back-end. A secondary motivation behind this design decision was to minimize the load being placed on the biodiversity database. This is advantageous because the biodiversity database is heavily used independently of the ISMS.

### 2.2.2 Application services

GNU/Linux was selected as the platform for the development of the ISMS application. This freely available, open-source operating system provides a plethora of software development tools and server applications ideally suited for delivering applications across a client-server architecture such as that provided in an Internet/Intranet environment. Another important consideration in the selection of Gnu/Linux was the fact that it is well supported with an active user community - making problem solving more streamlined.

The role of the GNU/Linux server is to act as a gateway between the database tier and the client tier. All the application logic is encapsulated within the applications on this machine, which functions as the middle tier in the three-tier model. In addition to providing the platform on which application logic is carried out, the GNU/Linux server also provides part of the first-tier database management services - all model-specific information in the ISMS Database is stored on this machine. GNU/Linux provides a robust, reliable operating system. The flexible 'mix and match' architecture ensures the server environment can be customised as needed, whilst still providing reliable services under heavy workloads.

Low-level functionality was implemented using the Apache web server, the PHP4 web application language and the Sendmail electronic mail (email) transport agent. The aforementioned applications are crucial in making the application available over the Hypertext Transport Protocol, and providing user feedback in the form of providing email notification on model completion. The Apache web server and PHP web application language formed the foundation for the development of the ISMS application.

"PHP is a server-side HTML-embedded scripting language that allows developers to build web applications" (Knudsen 1999). PHP was selected as the back-end application language



because it provides a flexible application development environment, can be programmed in an object-orientated manner (Knudsen 1999), and because the author was already familiar with this programming language. Object orientation was an important consideration in the selection of the application-layer programming language. Development using object-orientated programming (OOP) techniques reduces application complexity and allows the program to be divided into coherent parts that closely model the problem domain. The OOP paradigm also allows a complex system to be assembled from simple components in an evolutionary manner (Graham 1991). The OOP implementation in PHP is not as comprehensive as more 'pure' OOP languages such as JAVA or Smalltalk - PHP is weakly typed and was primarily developed for procedural programming. Notable omissions from the OOP 'feature set' are the lack of encapsulation (e.g. declaring class members private and forcing access to these members through accessor and mutator member functions) (Allen & Hornberger 2002) and the lack of other advanced features such as multiple inheritance and class polymorphism (Ullman 2001). However, the OOP capabilities in PHP are sufficient to support a 'clean' compartmentalised application design and to support code reuse through class inheritance - useful for reducing programmer effort (Allen & Hornberger 2002).

Additional application logic was implemented using the JavaScript programming language, which allows client-side functionality (such as form validation) to be implemented (Flanagan 2002). Several other applications form part of the loosely coupled system architecture of the ISMS application, as illustrated in Figure 2.2. In this architecture, low-level applications provide foundation services such as networking operations to higher-level applications. The system is constructed as a loosely coupled system, where discrete programs are coordinated by the ISMS software to form a single cohesive system.

Console-based applications under the \*NIX operating system make it possible to carry out the development of the ISMS application directly on the server from any client computer capable of running a remote terminal application. To support the development process and manage the source code repository, the concurrent version system (CVS) was installed on the server. CVS provides an advanced software development environment, where each revision of the application source code is logged and poor design decisions can be 'rolled back' with minimal effort. Higher-level applications such as the GARP applications, PROJ4, ImageMagick and a range of Gnu Utilities (including 'awk', 'ln', 'tail', 'at' and others) were installed on the application server to provide services for the ISMS application. Typically these applications were invoked from the ISMS application using PHP4 system() and 'back tick' system calls. The \*NIX *stdout* and *stderr* mechanisms could then be used to evaluate the successful execution of these applications from within the ISMS application.

The ISMS was designed to collate user choices as they progress through the model preparation

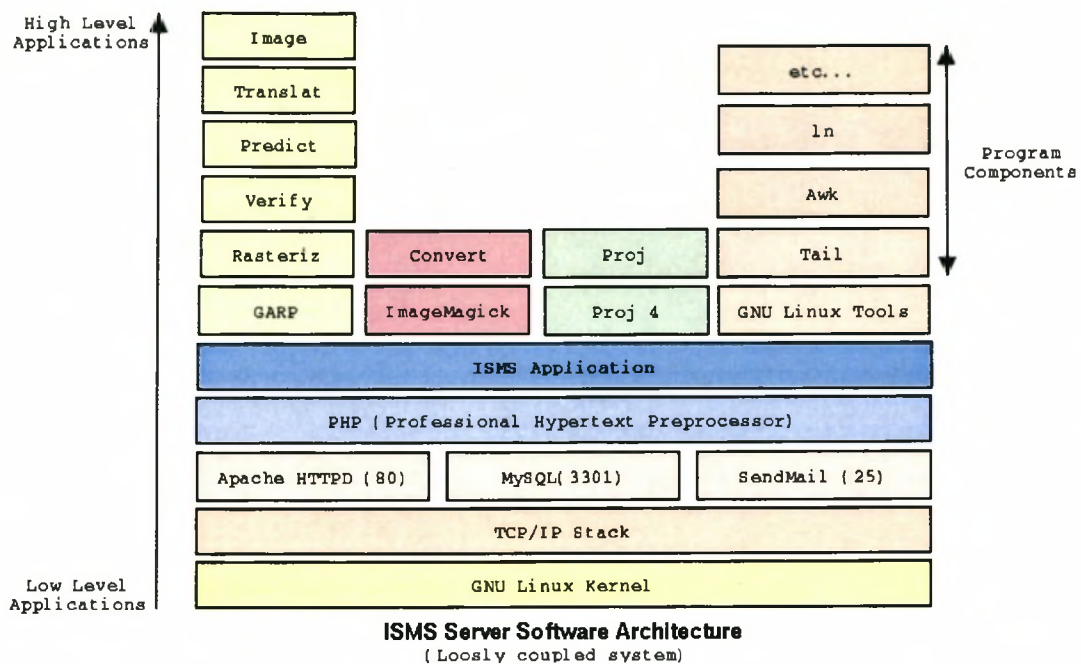


Figure 2.2: The ISMS Software architecture.

'wizard' and assemble a Bash batch file. The 'at' command was then used to queue the model run. When the model iteration reaches completion, the batch file invokes additional applications to post-process the data. For example, the ImageMagick 'convert' program generates standard-sized image outputs and image thumbnails of the model results.

Additional post-processing was implemented for the history browser functionality, including the use of the 'tex2pdf' application to dynamically produce a 'printer friendly' portable document format ('PDF') file from the model outputs which are rendered as latex documents by the ISMS reporting module (see Appendix G for an example model outputs).

### 2.2.3 Client services

At the third tier the client workstations allow the ISMS user interface to be exposed to the user. The architecture of the ISMS allows for multiple concurrent users of the system, thus this tier can be represented by  $n$  workstations. Each concurrent use of the system is allocated a session working directory on the ISMS server, which is created for that iteration. One of the goals of the ISMS application was to keep the hardware requirements for end users to a minimum. This was achieved by executing all of the GARP code on the ISMS server. Client processing is limited to displaying web pages in a browser, executing some minor JavaScript code and communicating with the ISMS server via the TCP/IP networking protocol suite. No direct interaction between

the client workstations and the biodiversity database server occurs. This was an important design decision prompted by concerns on the part of WCNCB that users of the ISMS do not have access to the raw dependent variable data stored in the biodiversity database server. There were two main reasons for this requirement. Firstly, many of the taxa represented in the database are Red Data Book listed (Branch 1988) or have otherwise been classified as important taxa for conservation. There was a strong concern that locality data could easily be misused (e.g. by illegal collectors) should they be made available. Secondly, the biodiversity database contains accession data from a wide range of organisations, often under intellectual property rights (IPR) agreements governing the use and distribution of data. Another important design consideration was to ensure that the Hypertext Mark-up Language (HTML) rendered by the ISMS application does not include any vendor-specific implementations of HTML and associated technologies (e.g. Microsoft vbscript). Producing standards-based HTML output would ensure that the ISMS client application would be accessible from a wide range of web browsers and would be useable within a heterogeneous software and hardware environment.

Being a stateless protocol, the Hypertext Transfer Protocol (HTTP) presents certain problems when producing linear applications. Statelessness in the context of HTTP refers to the fact that each request/response transaction with the HTTP server (Apache) is considered to be a new transaction and does not in itself carry over any state information established in prior transactions. There are various mechanisms to deal with the stateless nature of HTTP. In the case of the ISMS application, server side 'cookies' were used to store user information such as the currently logged-in user name. In addition, hidden HTML fields were used as a simple mechanism to allow for persistent storage of certain variables. Lastly, the ISMS database was used to store user choices in the Iteration table. At the point of model execution, the iteration table could simply be queried and all relevant parameters for the model iteration easily retrieved.

#### 2.2.4 System integration and detailed system design

The individual software components (as enumerated in the preceding sections) were first identified and tested manually, usually by invoking the application from the Bash shell command line. The use of command line interface (CLI) applications was preferred to GUI-only applications. The latter were avoided to ensure that each application could be invoked by a PHP batch process. At the culmination of this initial testing phase, it was verified that a complete GARP iteration could be completed using a purely CLI process (Sutton 1999). The challenge of this project was thus to integrate the array of applications (and operating systems) used to carry out a GARP iteration in a way that provides a seamless, consistent user interface for end users who may be using any recent web browser on any operating system.



---

To this end the system model was refined into a high-level process chart (see Figure 2.3) and a detailed process chart (see Appendix A) that would map out the further development of the system. The high-level process chart defines in general terms the steps a user needs to carry out in order to prepare and execute a GARP model using the ISMS front end, which can then be implemented as individual user interface screens. The detailed process chart breaks down the interaction the user will have into atomic units, which can then be implemented as discrete functions within the application codebase. Using this prototyping approach allowed potential flaws in the application design to be identified at an early stage before any code was actually written. Input forms that have dependencies on information that should have been previously provided by the user could easily be identified in this way. For example, layer set management can only take place once a valid extent has been selected (see section 3.3.1 for more details). The process chart provided a useful tool to identify these dependencies and rearrange the application workflow accordingly.

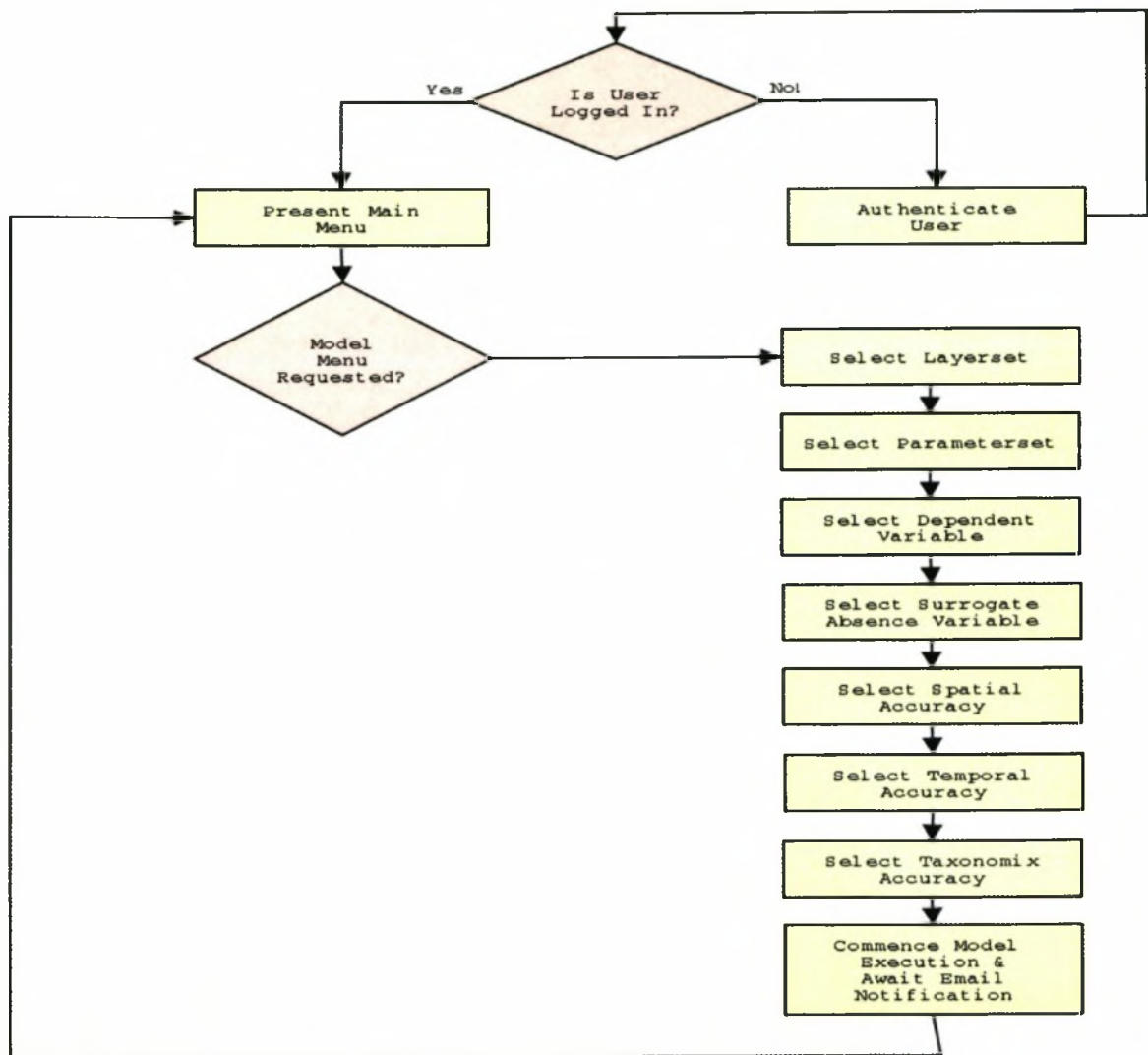


Figure 2.3: High-level user interface flow chart for the execution of a single GARP model iteration.



Once the program flow was conceptualised, a detailed database entity relationship schema was developed to map the various data-centric parts of the application to tables and relational joins between the tables. The database schema for the ISMS application internal database is presented as Figure 2.4. The ISMS database stores information about users who may log on to the system and records the models executed by the user in the iterations table. The independent variable datasets are also recorded on this database, as well as various 'lookup tables' required at various points in the modelling process. Additionally, the database provides the facilities for each user to construct a personalised working environment within the ISMS application framework. The MySQL database engine has certain limitations; in particular, it is not a true relational database and does not support concepts such as foreign key constraints. Other common database features are also absent, including triggers, stored procedures and views. These limitations did not seriously limit the development of the ISMS application owing to the relatively simple structure of the ISMS back-end database and the ability to implement these types of functionality within the PHP application space.

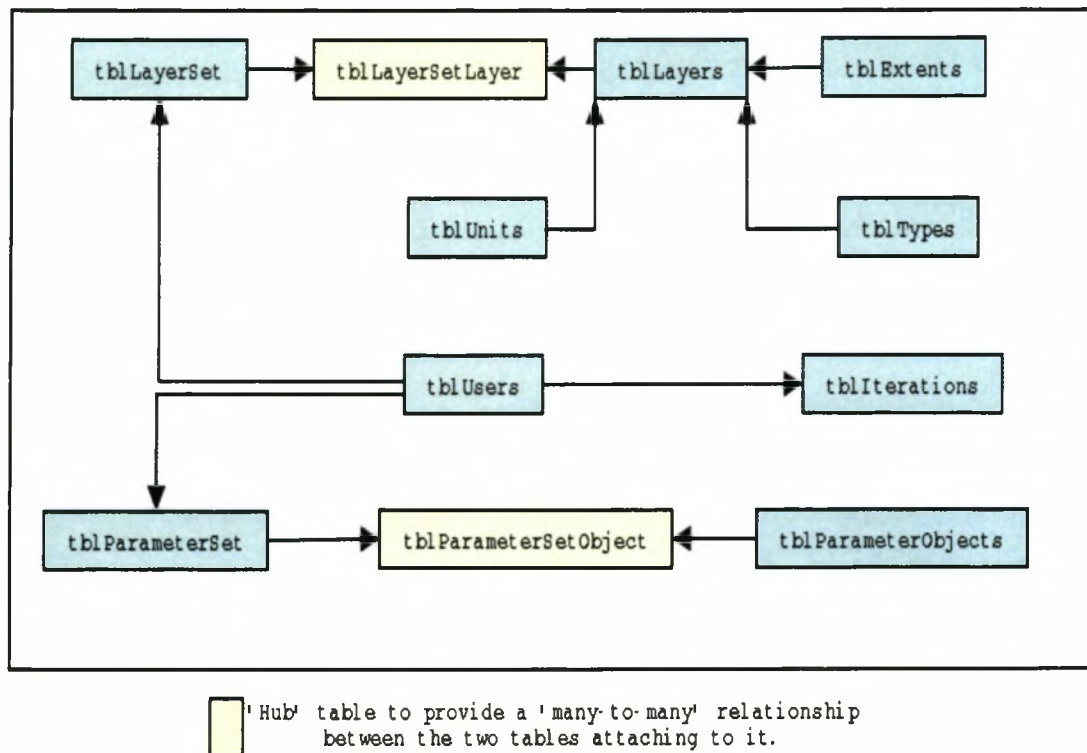


Figure 2.4: A simplified database schema for the ISMS back-end (where application-level data are stored).

Once the database design process had been completed, it was possible to carry out the object orientated design aspects of the application suitable for implementation as classes in the PHP

programming environment. A design for the application (based on the manual testing phase) was prototyped using a Unified Modelling Language (UML) class model. A 'single rooted' class hierarchy (Liberty 2002) was developed with all core ISMS classes extending (or inheriting from) the base ISMSObj class (see Appendix B for the complete class diagram). This base class provides standard functionality such as database abstraction, GUI element rendering (in HTML) and debugging. The object-orientated features of the PHP language allowed for the development of a clear, modular application framework. For example, the entire model execution process is implemented through the instantiation of an Iteration object. The Iteration class provides the various methods to generate the web GUI interface for each step of the data entry process. Thus, the PHP class hierarchy defines and implements the functionality required to carry out the various parts of the modelling process. With the system process flow, database schema and UML object diagram in place, it was possible to implement the ISMS Application following a well-defined roadmap. The aforementioned design documents were treated as 'live' documents and as such were adapted as needed through the design process.

With the software design strategy in place, the various software components were implemented and tested. This resulted in a web application that provided two key areas of functionality - an administrative environment and a model execution environment. In the following chapter, the administrative environment for the ISMS application is examined in more detail.

## CHAPTER 3: ISMS ADMINISTRATION

The ISMS provides a flexible administrative interface. With the exception of creating parameter sets and creating layer sets, the functionality described in this chapter is only available to administrators of the ISMS. The purpose of the administrative interface is to maintain components of the ISMS that have global application to all users. Because of their global nature, only users that have been assigned administrative rights can perform tasks such as adding layers to an extent, defining new user projections, updating the projection and datum tables, adding new extents and layer sets, etc. Administrators can also add new users to the system.

Two distinct database back-ends are used by the ISMS; the ISMS database back-end and the biodiversity database back-end. Administrative settings are stored in the ISMS database. The biodiversity database is only used to store dependent variable data - access to the biodiversity database from the ISMS application is read only. This approach provides good separation between application-specific data and non-application-specific data, and allows for future support of dependent data stored on other types of database back-ends with minimal effort. This latter capability was tested by implementing a biodiversity database 'clone' in PostgreSQL, which required only minor changes to the codebase.

A small bootstrapping configuration file (ISMConfig.php) stores the required information for connecting to the two back-end databases and is stored in the ISMS 'inc/' directory. A simple text file parser was written to scan this configuration file for database host names and login details. When implementing the ISMS application on a new server, a one-time edit of this file is made to update the connection details (user name, password, etc.) for the ISMS and biodiversity database back-ends. Programmatically, the administration tasks are carried out by instantiating objects from the various ISMS application PHP classes. Specifically the following PHP classes were created to carry out these tasks:

- Admin Menu class
- User class
- Projection class

- Ellipsoid class
- Extent class
- Layer class
- User Projection class
- Parameter Set
- Layer Set class

Each of these classes has methods to produce HTML data input forms and to display reports based on simple database queries. The following sections describe the user interface functionality related to these classes in more detail.

### 3.1 USER MANAGEMENT

Two types of user accounts can be created using the ISMS administration interface: administrators and normal users. While both types of user can carry out GARP model iterations, 'normal' users are limited in the range of functions they can carry out in the administration interface (as outlined in the previous section). This hierarchical user access level was implemented in order to prevent the inadvertent corruption of the application environment by inexperienced users. As illustrated in Figure 3.1, a variety of data are stored for each user, including each user's email address. The latter is important because the email address is used to email the user on completion of each model iteration.

A log-on screen is used to authenticate users of the ISMS against the users table on the ISMS database back-end. When the user successfully logs into the ISMS environment, a 'cookie' is stored on the client computer containing the session identifier for their session. Server-side 'cookies' preserve session information so that the user does not need to authenticate with each page reload. The session identifier is used by PHP to automatically retrieve state information from a complementary session file that is stored on the ISMS server at log-in time. The session information is also used to determine which administrative tools can be used by the currently logged-in user.



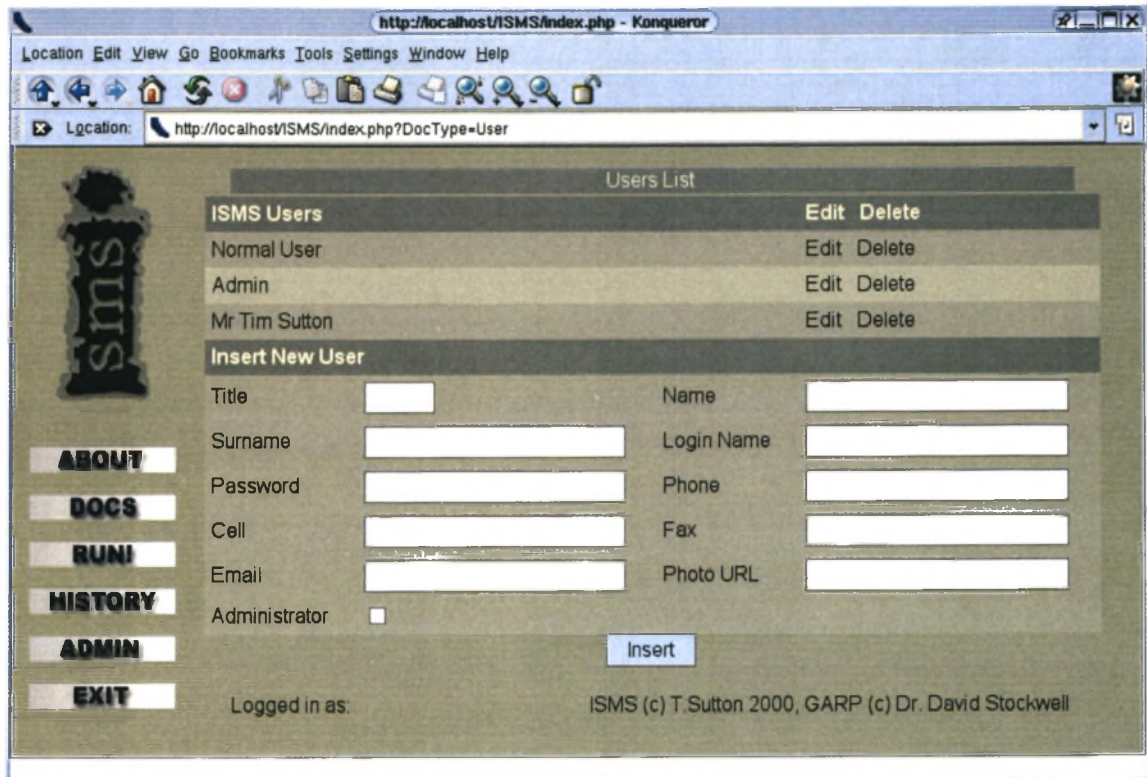


Figure 3.1: ISMS User Administration Screen. New user accounts are created here and existing user accounts can be modified using this interface.

## 3.2 MANAGING PROJECTIONS

“[A projection is a] geometric transformation that converts latitude and longitude coordinates to planar coordinates. Projections can be based on a developable surface (such as a plane, cylinder or cone) or a mathematical function” (Chrisman 2002).

“A projection describes a geometric transformation that converts spatial data from one coordinate system to another” (Kang-tsung 2002).

The GMS requires that data used in the modelling process are already stored in the correct geographical projection. As a further requisite the independent data layers must match the projection of the dependent variable layer that is brought into the modelling environment from the biodiversity database. Thus, the user is normally responsible for handling the task of data projection conversion before commencing with a model iteration. This was seen by the author as a major limiting factor in the use of the GMS. To address this, the ISMS uses the Open Source projection engine ‘PROJ4’ (Evenden 1995) to convert independent variable data (stored

in the biodiversity database as longitude/latitude coordinates) into a user-specified projection. There are three ISMS administrative modules for managing the ISMS projection engine: the projection definition module, the ellipsoid definition module and the user projection definition module. These are discussed in more detail in the following subsections.

### 3.2.1 Projection definitions

PROJ4 is capable of projecting latitude/longitude coordinates into planar coordinate systems (Evenden 1995). Within the ISMS architecture, each projection definition is stored in the ISMS MySQL back-end for easy access from other program modules. The projection database created by ISMS is shown in Appendix C. Three attributes are stored in the projection table for each projection definition: projection name, notes about the projection and the expected parameters for the projection.

'proj' (the PROJ4 command line application) is able to produce a listing of supported projections by issuing the command:

```
proj -lP
```

In order to automate the process of populating the Projections table and to cater for new projections being supported in future versions of PROJ4, a PHP script was written to invoke the above-mentioned command using a named pipe. Data obtained over the named pipe were then parsed and valid SQL statements were created for insertion of these records into the ISMS database back-end. Using this approach allows the ISMS projection database to be dynamically maintained and paves the way for effortless addition of new projections into ISMS as they are implemented in new releases of the PROJ4 application. It should be noted that in certain instances the expected parameters obtained using the above procedure did not contain sufficient information to properly define the projection. In these cases the Proj User's Manual (Evenden 1995) was consulted.

### 3.2.2 Ellipsoid definitions

"[An ellipsoid is a] three dimensional object formed by rotating an ellipse around its minor axis, an oblate ellipsoid approximates the shape of the earth (geoid), computed by the best fit to geodetic observations" (Chrisman 2002).

Associated with each projection is an ellipsoid definition. In keeping with the projection database, the ISMS application contains a parser that allows 'proj' to be queried for a list of all supported ellipsoids, with the resulting list being stored in a MySQL table. Three items of information are stored for each ellipsoid definition: the ellipsoid name, radius and a third parameter, usually (but not always) the flattening on the semi-major axis of the ellipsoid (Evenden 1995).

A user with administrative privileges is able to rebuild the ellipsoid database by using the web-based ISMS administration menu. In a similar manner the Ellipsoid database can be queried from within the ISMS administration menu (as illustrated in Figure 3.2). Because of the integration of ISMS with PROJ4, only ellipsoids directly supported by PROJ4 can be used and the user may not interactively describe new ellipsoid definitions.

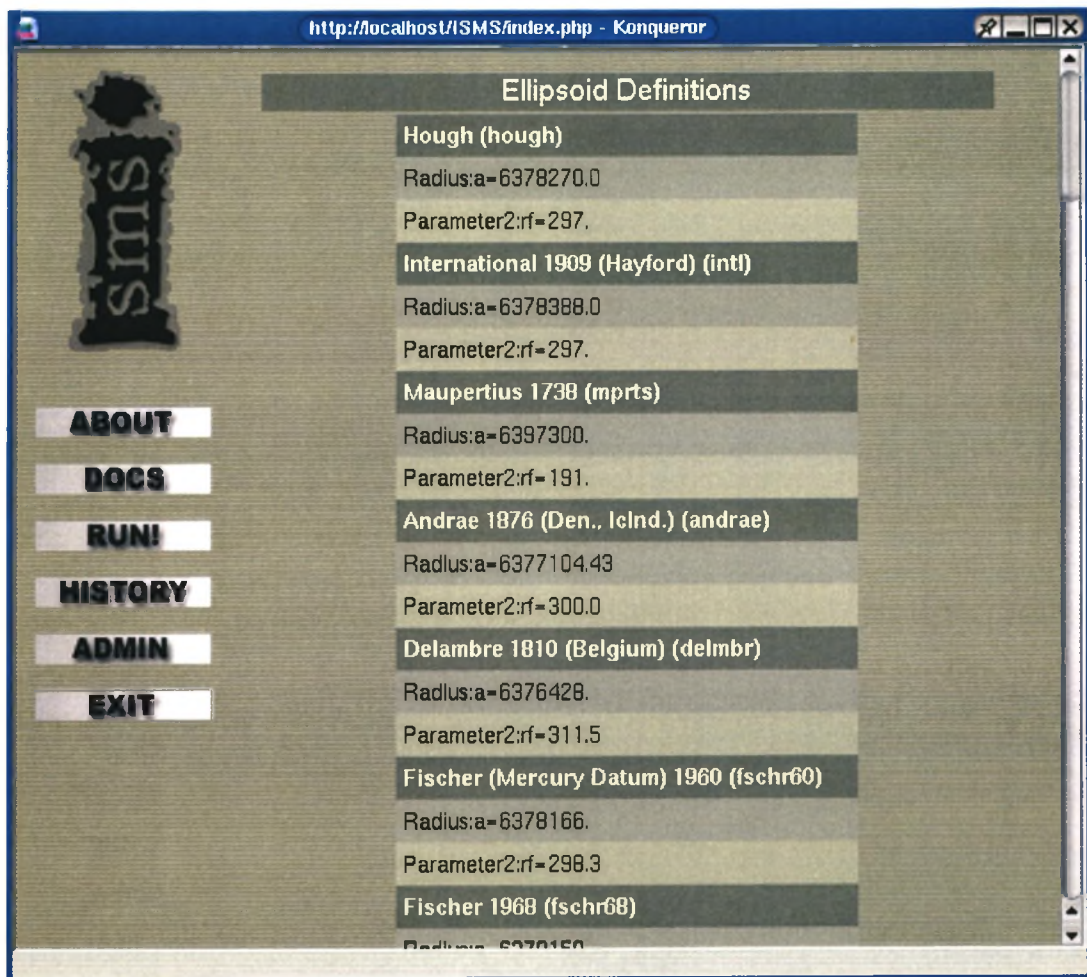


Figure 3.2: The ISMS Ellipsoid database is generated using the PROJ4 projection utility. On completion of the import process, the complete ellipsoid listing is displayed to the administrator.



### 3.2.3 User projection definitions

User-defined projections consist of a PROJ4 projection name, ellipsoid and one or more user-specified parameters - as illustrated in Figure 3.3. This approach allows a large amount of flexibility in the independent data layers that can be used, as the ISMS administrator simply defines a user projection through which dependent variable data will be filtered. Each user-defined projection is given a descriptive name by the ISMS administrator. These user-defined projections are then used during the extents definition process (see the following section for more details on this). For example, a user projection might be defined by selecting 'Lambert Conic Conformal' from the projection list and choosing the 'modified Clarke 1880' ellipsoid from the ellipsoids list. Projection specific parameters are then entered to facilitate the projection of independent data between the 32°S and 34°S parallels on a central meridian of 21°E. The user projection might then be named 'Clarke 1880 32/34/21' by the administrator defining this user projection. This user projection would then be available to be used in the creation of a new extent definition as outlined in the following section.

## 3.3 EXTENT DEFINITIONS

Once a user projection is defined, the ISMS administrator can be used to describe extents that belong to that projection. An extent is an ISMS structure that defines the spatial range of dependent variable data layers ([minx,miny] to [maxx,maxy]), expressed in Cartesian coordinates. The ISMS does not support geographic 'unprojected' data. This is a side effect of the independent layer generation process where the Spatial Analyst extension does not support the creation of grid layers in a non-Cartesian coordinate space (ESRI 1996). In addition to the minimum and maximum spatial extremities, the extents definition also stores information about the cell size of elements in the independent variable raster layer. Thus three elements determine the uniqueness of an extent: projection, spatial extent, cell size. Stated differently, more than one extent can be defined to cover the same geographical area, provided that the cell size, and/or projection, for each extent differs.

Extent definitions are stored in the ISMS database Extents table. Each independent variable layer has an associated record in the Layers table in the ISMS database. A third table stores projection parameters. Each projection can have one or more related extent record, as each extent can have many layers associated with it. The relationships between these tables are enforced by constraints hard coded into the ISMS application. The concept of extents used by the ISMS provides a degree of data normalisation (Pratt & Adamski 1994), reducing the



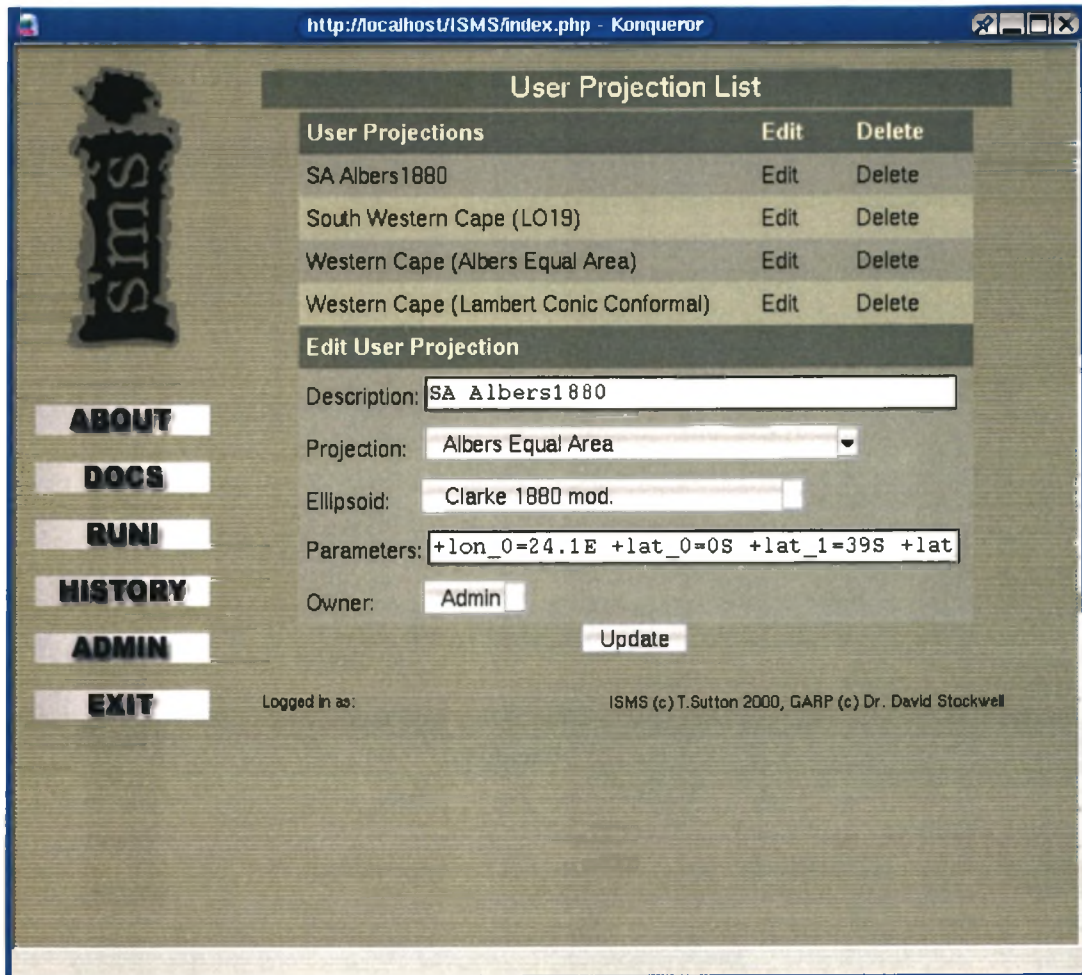


Figure 3.3: ISMS User Projections are defined by combining specific parameters with standard projection and ellipsoid definitions.

need to store the same information on a per layer basis, and allowing the user to quickly obtain metadata for a given independent variable data layer. The system administrator defines extents using the graphical user interface (GUI) based extents administration module. During model execution, the extent definition information is appended to the 'paramete' file as previously illustrated in the 'Columns' and 'Rows' directives of Table 1.1.

### 3.3.1 Independent data layers

In its current implementation the ISMS does not handle the process of actually rasterising independent variable data layers into the GARP binary format. One of the goals of the ISMS application is to shield the end user from the complexities of data preparation and

model parameter settings. These complexities are largely dealt with internally by the ISMS application, and partly conferred to the ISMS administrator. In order to effectively administer the ISMS, it is necessary to understand the way in which ISMS deals with independent variable data. The first step in the generation of an independent data layer for use in GMS/ISMS will typically involve data manipulation in a GIS package. For example, during the testing process (Sutton 1999) it was decided to use elevation as a model parameter following Payne & Stockwell (1999) as a precedent. A seamless vector dataset consisting of contour lines and spot heights was obtained in ESRI shapefile format. In order for this data to be useable in GARP, it was first converted into a Digital Elevation Model (DEM). A commonly used approach to this conversion process is to interpolate contour data and spot heights (Burrough & McDonnell 2000; Hutchinson 1989). The vector data were first reprojected to the same projection system to be used during the GARP modelling process. The contour lines were converted into a series of equally spaced points along the length of each contour and the resulting point dataset merged with the spot height data. Each point in the new dataset was assigned the Z-axis value of its underlying contour. Using the ArcView Spatial Analyst spline curve interpolator, the data were converted into a raster elevation model and exported to an Arc/Info ASCII grid file. The GARP toolkit includes an application called 'rasteriz', which was used to convert the ASCII grid file into the GMS native binary data file format. A similar process was repeated for each independent variable dataset introduced into the system.

This independent variable preparation process (as briefly illustrated above) requires specialist skills in the use of GIS software, as well as access to relevant datasets. Typically, however the process only needs to be carried out once for a given extent (by a system administrator), after which point the independent variable data become globally available to all users of the system. A more detailed guide to independent variable dataset preparation is available as Appendix F.

In order to facilitate multiple concurrent user modelling sessions that access the same independent data, and to minimize hard-drive storage requirements, a compartmentalised environment was needed for each model iteration. In order to achieve this, an inventory of available independent data was stored in the ISMS database. In addition, symbolic links (symlinks) are created during model invocation. These symlinks point to a central pool of independent data layers. In order to prevent users attempting to utilise independent data sets that do not cover the same aerial extent, or are not stored in the same geographic projection system, independent data layers are stored in conceptual aggregations called Extents (see section 3.3). An extent is represented in the file system as a directory, with all independent variable data layers that belong to a given extent being stored in the same directory. Once the various independent data layers have been generated for a given extent, they are stored in the ISMS directory tree under:

ISMS/data/independent/[extentName]/

To keep track of layers within the ISMS application, each independent data layer has an accompanying layer definition record created in the ISMS database layers table. This is illustrated in Figure 3.4.

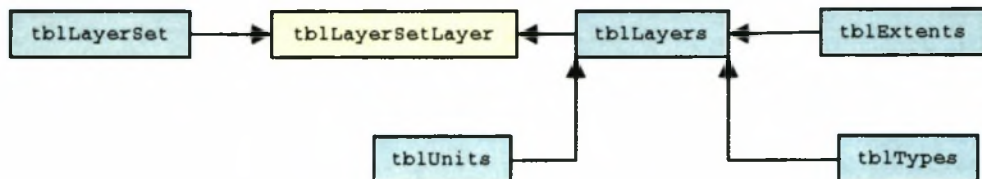


Figure 3.4: Each independent variable layer that is available to end users is referenced in the Layers table.

Each layer needs to be assigned to a valid extent. A layer, once referenced on the Layers table, can then become part of one or more layer sets. The process of populating the Extents table and associating layers stored in the file system with extents (as defined in the ISMS database) is facilitated using the layers management administrative interface (Figure 3.5). The range of data values for each layer, the type of data and a brief description are also recorded using this form. These metadata are used when dynamically constructing 'parameter' files during the model execution process.

### 3.4 MODEL ENVIRONMENT CONFIGURATION (PARAMETER SETS)

Model parameters are grouped together in aggregations called parameter sets. Parameter sets were implemented in order to prevent the situation where a user must repeatedly define parameters for each model iteration. Parameter sets are defined on a per user basis and are Extent independent. The potential parameter types that can be used to carry out a model iteration are conceptualised as parameter objects. A parameter set describes a many-to-many entity relationship. This relationship expresses the range of possible parameters that can be used for any given model iteration. To achieve this an intermediate joining table is used. Parameter set objects 'bind' a particular parameter object to a parameter set (following the methodology described by Pratt & Adamski (1994)). This is illustrated in Figure 3.6.

Parameter sets are 'named' parameter collections that facilitate the development of a range of different parameter profiles. A parameter object is an atomic unit consisting of one potential



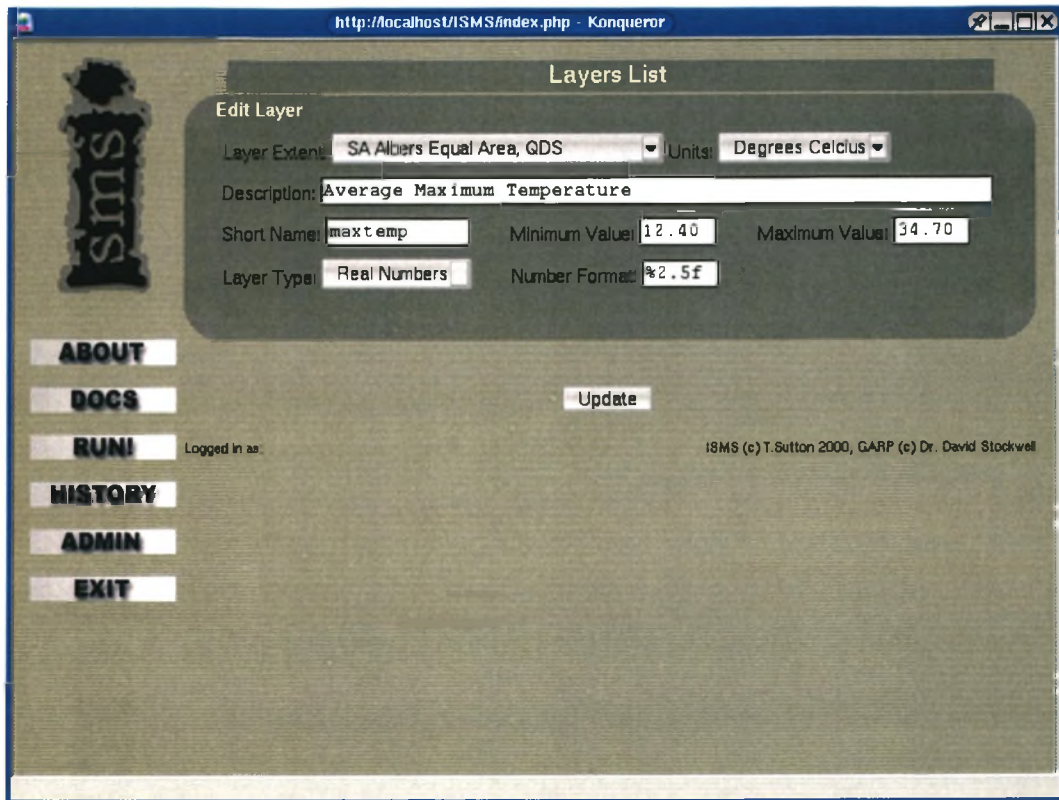


Figure 3.5: The ISMS Layer management module is used to define the attributes and file system location of data layers stored in the GARP binary format.

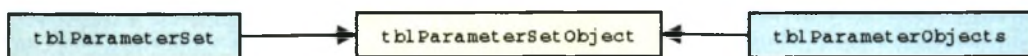


Figure 3.6: Parameter storage in the ISMS database.

parameter that can be passed to the GMS. The actual value of the parameter is stored in the parameter set object table. It is the aforementioned table that also allows a 'many to many' relationship to exist between parameter objects and parameter sets. The instance properties of a given parameter object in a parameter set are also stored in the parameter set object 'join' table. For example, if a user selects the model flag 'atomic' and assigns it a value of true (1), a new record would be created on the parameter set object table, populated with data. This is illustrated in Figure 3.7. In this diagram, the ParameterSetObjectID is a unique identifier for the parameter set object, ParameterSetID is the foreign key to the parameter set to which the parameter object belongs, ParameterObjectID is the foreign key for the parameter being used (in this case 'atomic') and Value is the instance value for that parameter (in this case a Boolean value).

The 'Value' field in this table thus stores the state of the parameter object as it will be actually

ParameterSetObjectID	ParameterSetID	ParameterObjectID	Value
1	38	1	1

Figure 3.7: A sample parameter set object record.

be used when the parameter file is generated. Each parameter object record (see Table 3.1 for an example subset of the available parameters.) contains information in the ControlType field used to govern the presentation of that parameter object on ISMS GUI input forms. For example, a ControlType value of 'Checkbox' will place a Boolean checkbox selector on the parameter set form, which the user can then tick indicating whether the value for this parameter should be true or false. Other possible ControlType values are 'Integer' and 'Text'. Using this simple mechanism, the construction of the parameter set definition GUI is completely soft coded - the ISMS constructs the GUI input form 'on-the-fly' based on the values found in the ParameterSetObjects and ParameterObjects tables. An example of such a dynamically generated parameter form is illustrated in Figure 3.8. The master list of potential parameters that can be used to control the GARP modelling process (held in the ParameterSetObjects table) was populated based on current available documentation for GARP (Payne & Stockwell 1999).

Table 3.1: The parameter objects table.

ID	Name	Control Types	Def. Value	Option Type	Notes
1	Edgeflag	Checkbox	0	Image	Perform edge detection on output image.
2	Overlayflag	Checkbox	1	Image	Overlay dependent variable layer on specI.
3	Atomicflag	Checkbox	1	Explain	Check this to use atomic rules.
4	Logitflag	Checkbox	1	Explain	Check this to use rules produced by logistic regression model.
5	Bioclimflag	Checkbox	1	Explain	Check this to use bioclimatic envelope rules.
6	Garpflag	Checkbox	1	Explain	Check this to use rules produced by the GARP modelling system.
7	Specialiseflag	Checkbox	1	Explain	Promotes rules that are more specialized.

There are two phases to the creation of a parameter set. During the initial phase the parameter set is described using a short text-based 'plain English' description. Following this, member

ParameterSet :GARP, BIOCLIM, ATOMIC  
Description: Basic model with most parameters using GARP defaults.

Parameter	Value		
Atomicflag (Check this to use atomic rules.)	<input checked="" type="checkbox"/>	Update	Delete
Bioclimflag (Check this to use bioclimatic envelope rules.)	<input checked="" type="checkbox"/>	Update	Delete
Garpflag (Check this to use rules produced by the Garp modelling system.)	<input checked="" type="checkbox"/>	Update	Delete
Maskvalues (Enter the cell value for areas to be masked.)	255	Update	Delete
Totalexperiments ()	5	Update	Delete

Add a parameter  
Accuracylimit ▼ Insert

ABOUT  
DOCS  
RUNI  
HISTORY  
ADMIN  
EXIT

Figure 3.8: The parameter set definition form is dynamically rendered based in Control Type field entries in the parameter set Object table.

parameters are selected. During this phase the user selects which parameters will belong to the parameter set being defined. At this point the user must also set the value for each parameter object. Reasonable default values are supplied at this point to speed up the process and to guide users. Administrator status is not required for the creation of parameter sets.

### 3.5 PREPARATION OF LAYER SETS

Layer sets are user-defined collections of independent variable data layers that belong to a common extent. The user may select all possible layers in an extent, or a given subset of the layers that belong to a given extent. As in the case of parameter set definitions, administrator privileges are not required for the creation of layer sets. The Layer Set administration form allows the user to compile these named lists of frequently used layers. Using this form allows the user to create, edit and delete layer sets, and assign one or more layers to a Layer Set collection. Figure 3.9 shows the list of named layer sets for the current user and the form used to create a new (blank) named layer set. Each user-defined layer set is named and given a brief



description. The definition of at least one Layer Set is required before the user can commence with a model iteration. The name and description given to a layer set may be changed at a later stage. The user may add and remove layers to the layer set using any available layer data in the same extent as the layer set (as illustrated in Figure 3.10). During the GARP modelling process the 'paramete' file contains the information about each independent layer that will be used in a model iteration. Typically, when GARP is used from the command line interface (e.g. a Bash shell), the user would use a text editor such as 'vim' to add appropriate independent variable references to the 'paramete' file. When a GARP model is initiated using the ISMS front-end, the need for low-level file editing is removed and the 'paramete' file is constructed automatically. The ISMS database layers table stores the information required to generate the layers section of the 'paramete' file.

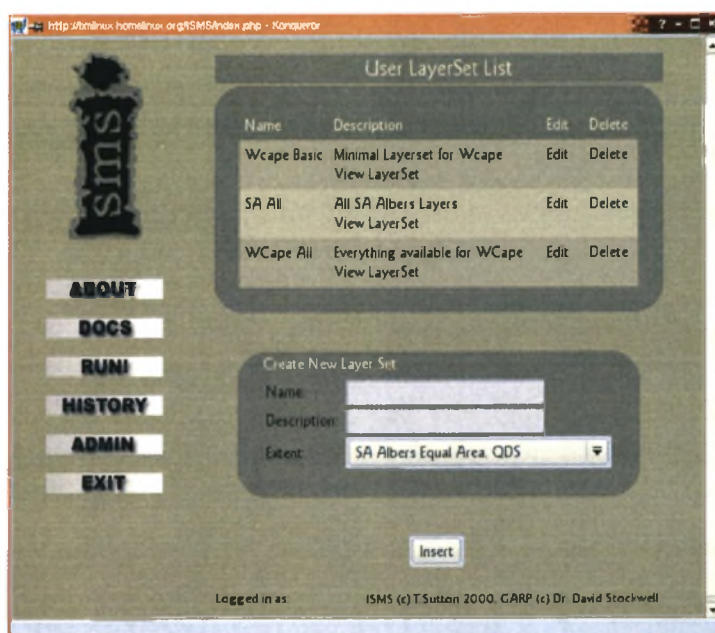


Figure 3.9: The ISMS layer set definition process.



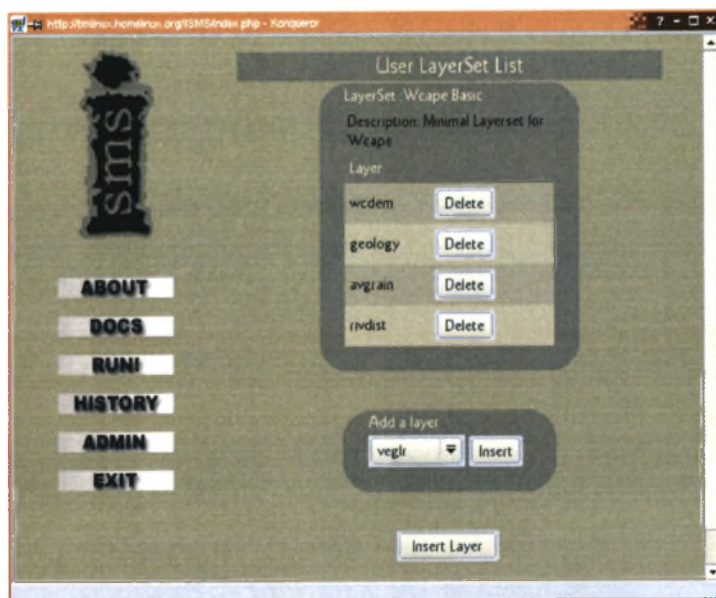


Figure 3.10: Associating layers with a layer set.

## CHAPTER 4: ISMS USAGE

### 4.1 RUNNING A MODEL ITERATION

For the user interface component of the ISMS application that enables GARP models to be executed three main areas of functionality were required:

1. Automation of the generation of the 'paramete' file which controls the GARP modelling process;
2. Creation of a multi-user environment for the GMS so that more than one user could execute a model at any given time;
3. Seamless orchestration of the extraction of dependent data from the biodiversity database and preparation of this data (through means of projection and rasterisation processes) for use in a given model iteration.

A series of input forms was implemented to guide the user through the model preparation process. Each form prompts for specific information accompanied by a brief description of the nature of the information required. The resulting interface provides a 'wizard-like' interface with a predefined workflow (see Figure 4.1). Before the user can carry out a model iteration, at least one Layer Set and one Parameter Set must have been defined by the user using the administrative tools discussed in Chapter 3. The individual forms used to carry out an iteration are described in the following sections.

Allowing concurrent application usage was achieved by dynamically creating \*NIX filesystem symlinks to the GARP binary files and the specific independent variable layers to be used for a given model iteration. When the user commences a new model iteration by clicking the 'Run!' icon, a new record is added to the ISMS Iterations table. The ISMS application layer also prepares for the new model by creating a working directory on the ISMS server file system. The GARP binary applications are also symlinked into this directory, so that when executed they will find the iteration-specific independent data layers in the working directory. Finally,

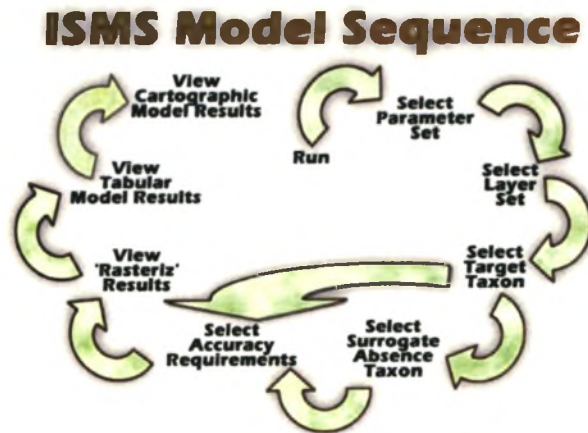


Figure 4.1: Diagrammatic representation of the 'wizard' workflow for carrying out a GARP model iteration using ISMS.

an information screen is presented to the user (Figure 4.2). In the background the system prepares a directory in which the model results will be stored. Symbolic links are also made in this directory that point to the GARP binary files.

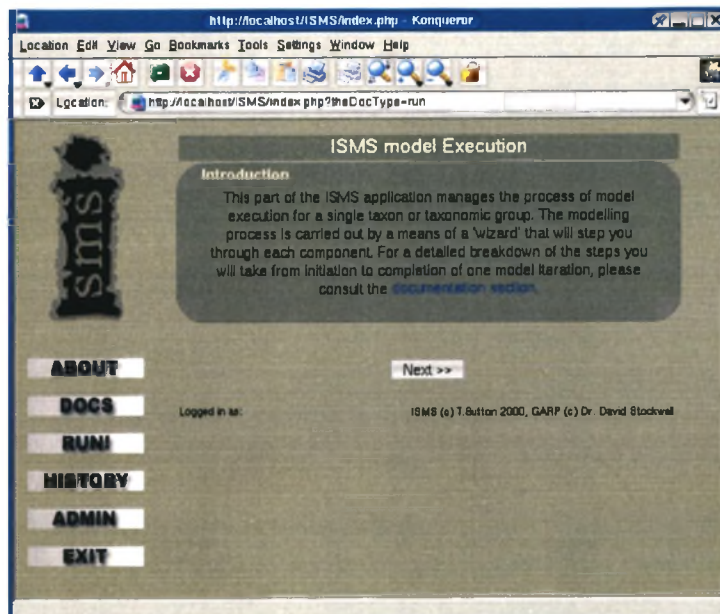


Figure 4.2: The modelling process is commenced with a small information screen.

## 4.2 SELECTION OF A PARAMETER SET

Once the preparation of the model directory is complete (as outlined in the previous section), The ISMS database is queried to determine the parameter sets which have been defined for the active user. This list is then presented to the user in the form of an HTML select box (as illustrated in Figure 4.3). The selection made from this list is stored in the Iterations table when the 'Next' button is pressed. When the execution of the model commences, each of the Parameter Set objects will be written to the 'paramete' file as GARP parameters.

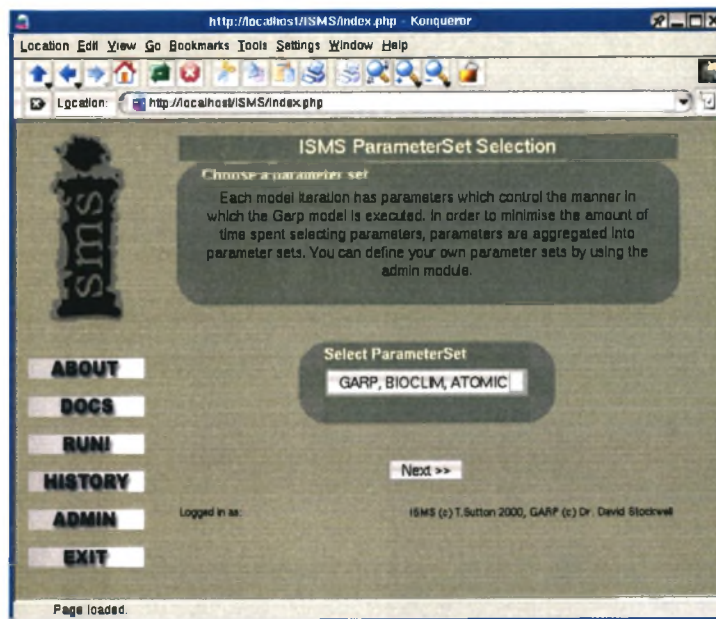


Figure 4.3: During the parameter set selection phase, the user is prompted to select one of the predefined parameter sets they have created.

## 4.3 SELECTION OF A LAYERSET

Having selected an appropriate parameter set, the user is presented with a list of Layer Sets that are associated with their user profile (see Figure 4.4). The selection of a layer set during this stage will cause a number of background processes to be carried out when the form is submitted. Firstly the layer set table is queried and each of the layers associated with that layer set is identified. Following this, each layer in the layer set is symlinked into the model iteration directory. Additionally, the layer set record primary key is stored in the iteration record.



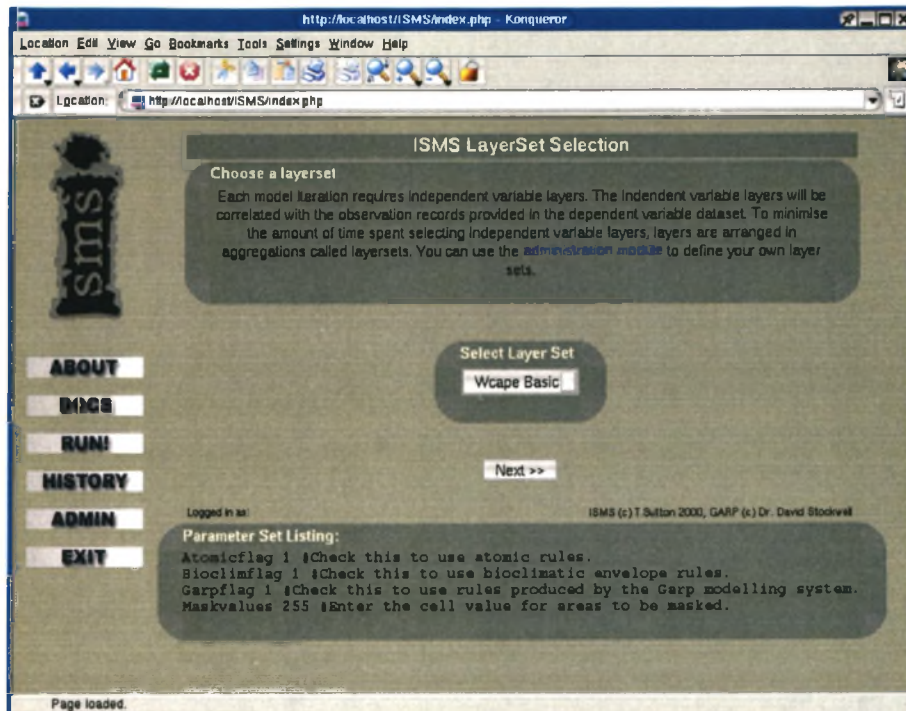


Figure 4.4: During the layer set selection stage, a predefined set is chosen from a list of the layer associations compiled by the user.

#### 4.4 SELECTION OF A TARGET TAXON

In the context of the GMS, dependent variable data consist of localities where a taxon has been observed. The ISMS differentiates between a ‘target’ or ‘presence’ taxon and an optional ‘surrogate absence’ taxon. Surrogate absence is described in the following section (Section 4.5). Together, these data comprise the dependent variable inputs into the GARP model. It is at the point of target taxon search that the user will first begin to interact with the biodiversity database. The search form (see Figure 4.5) allows the freeform entry of a search string by the user. This was implemented because it is more efficient than a hierarchical selection process whereby the user would select the Kingdom, Phylum, etc. from a series of checklists, eventually ‘drilling down’ to the taxon of interest. Scientific names were used as the basis for searching. The search phrase entered by the user is internally wrapped in SQL wildcard delimiters (‘%’) and an appropriate SQL query string is constructed. For example, if a user enters ‘Bufo’ in the search box, the following search string will be constructed:

```
select * from vwTempTaxon where TaxonName like '%Bufo%';
```



Using this simple mechanism, a measure of user uncertainty with regards to a particular taxon name (or its spelling) is catered for. Once the user has entered a search string, the form that follows presents the user with the narrowed down list of taxa that meet the original search criteria. In order to assist the user in ascertaining how many records are available for each of the narrowed down list of taxa, a record count is performed during the database query process and the result is displayed in brackets, as illustrated in Figure 4.6. During the model preparation phase, the ISMS application will query the biodiversity database for presence data and insert the result set into the localities file for the presence taxon. Data appended to this file will have a one (1) inserted after the locality information - which GARP uses to identify presence records.

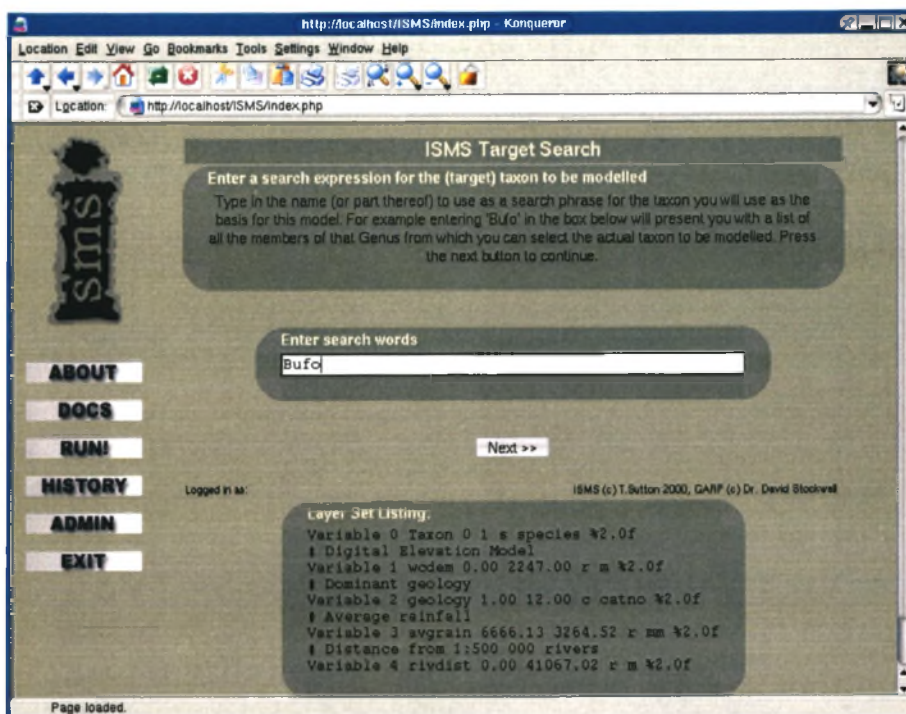


Figure 4.5: The presence taxon search form. Wildcard searches are implemented here, so partial matches to the search string will be found.

## 4.5 OPTIONAL SELECTION OF A SURROGATE ABSENCE TAXON

The biodiversity database, in common with most other field observation databases, does not contain difficult to determine absence records for specimens (Townsend Peterson & Cohoon 1999). An experimental approach was adopted here to facilitate the use of surrogate

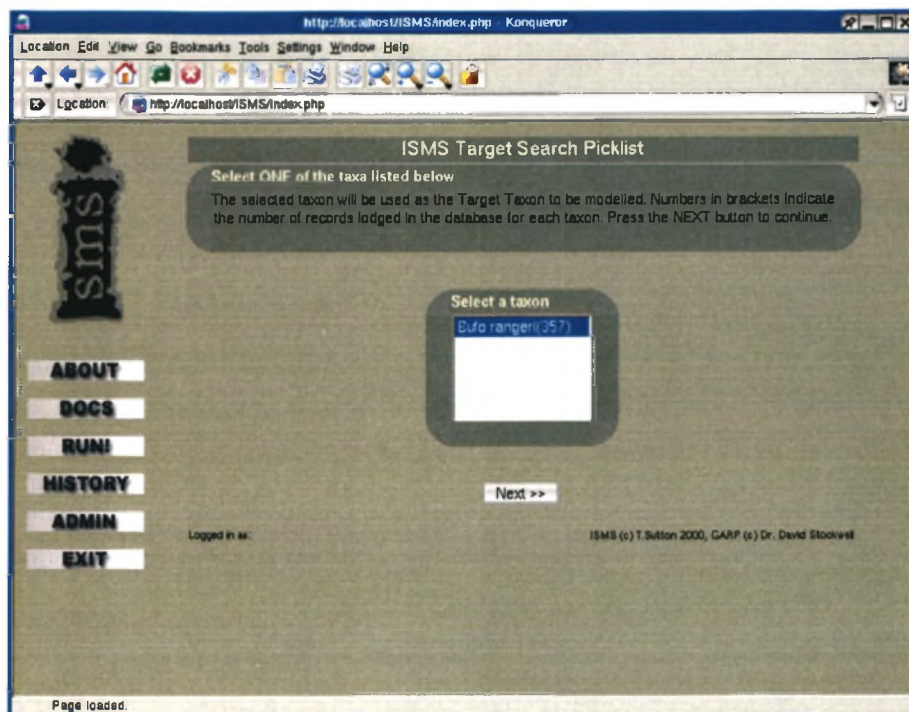


Figure 4.6: The presence taxon search results form. The number of records for each matching taxon is shown in parentheses after the taxon name.

absence data (where the presence of one taxon indicates the absence of another) in modelling distribution. The premise behind the use of a surrogate absence taxon is that certain organisms have mutually exclusive distribution ranges, so the presence of one taxon may be an indication that another taxon is absent from that same area. The ISMS facilitates the optional selection of a surrogate absence taxon, which is selected using a similar mechanism (see Figures 4.7 and 4.8) to that used in selecting the target (presence) taxon. During the model preparation phase, the ISMS application will query the biodiversity database for surrogate absence data and append it to the localities file for the presence taxon. Surrogate absence data appended to this file will have a zero (0) inserted after the locality information - which GARP uses to identify absence records. In keeping with the target taxon search process, the number of locality records per matching taxon is displayed in parentheses after each taxon name (as illustrated in Figure 4.8). This allows the user to decide whether an appropriate quantity of data exists to merit the use of a given taxon.

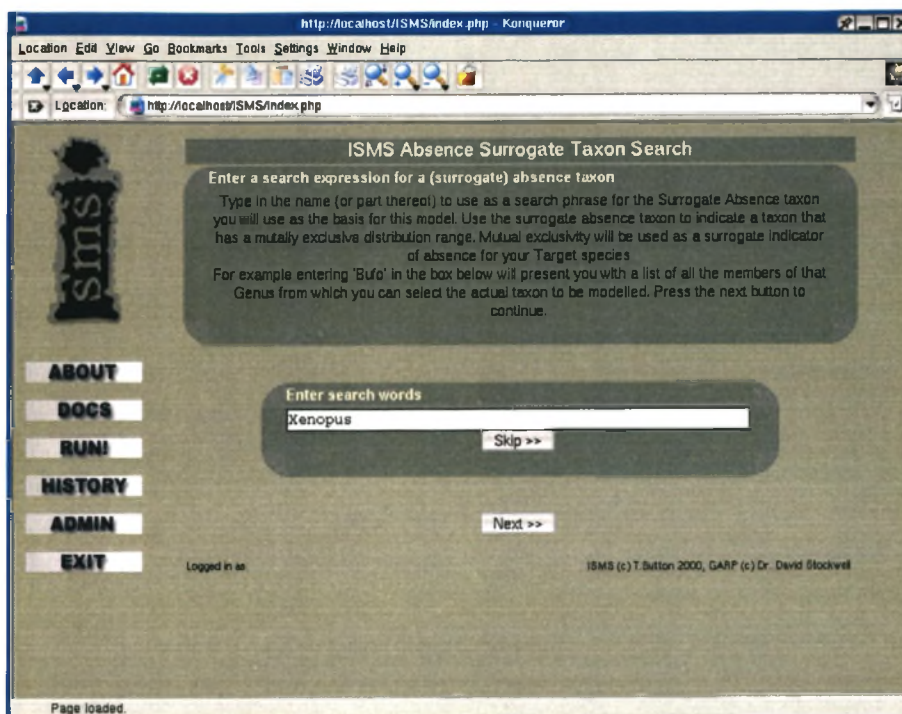


Figure 4.7: An optional surrogate absence taxon may be selected in a similar manner to the selection of the target (presence) taxon.

## 4.6 SELECTION OF DATA-QUALITY CRITERIA

The biodiversity database stores three important items of metadata for each specimen record: spatial accuracy of the locality data, taxonomic precision and temporal precision. This information is utilised within the model preparation process to filter the independent variable dataset to be used in the GARP model iteration. Selection of data quality criteria is made using an input form that uses three pick lists, as illustrated in Figure 4.9. This option is useful when working on a small scale, where it is desirable to omit specimen records that have not been resolved to the nearest second of a degree. Similarly, records where the taxonomy or collection date is uncertain can easily be omitted. Choosing the 'unknown' option in each case will include all records regardless of data quality.

## 4.7 FEEDBACK FROM THE RASTERISATION PROCESS

The goal of the dependent variable selection phase is to produce a GARP binary file called 'Layer00' from which the GARP software will obtain the dependent variable data for a model



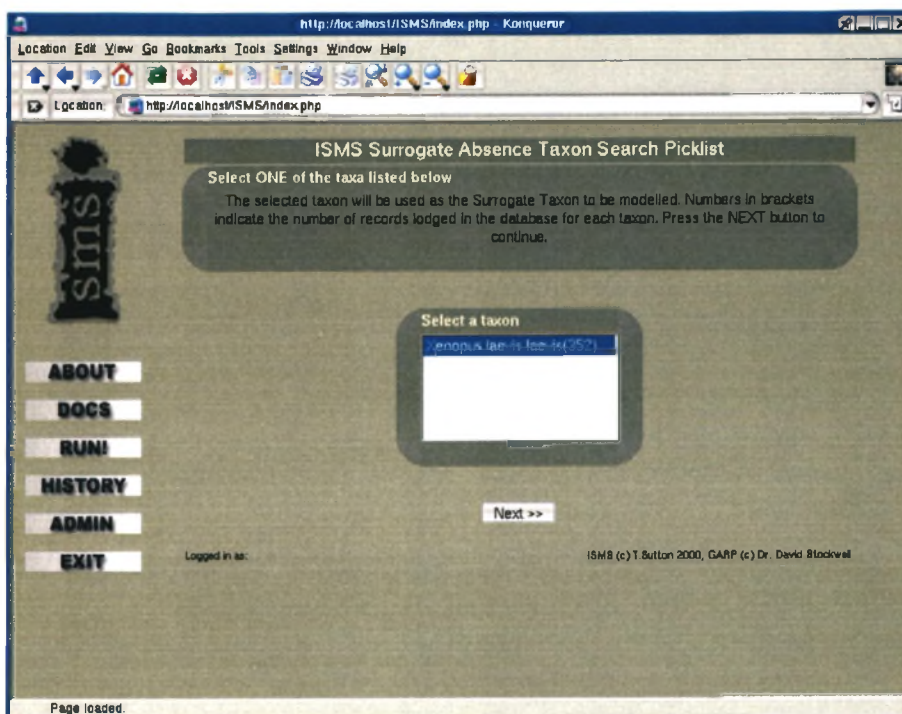


Figure 4.8: The number of records for each matching surrogate absence taxon is indicated in brackets after the taxon name.

iteration. In addition to the actual identification of the target taxon (and optionally a surrogate absence taxon), a number of steps are performed in the transition of data lodged in the biodiversity database into a GARP binary file. Data in the biodiversity database are stored in either decimal degrees or as a quarter-degree square (QDS) reference. In both cases it is necessary to transform the locality data into the projection system that matches the projection of the layer set selected by the user for the model iteration. In order to facilitate this, the PROJ4 (Evenden 1995) utility is used. PROJ4 provides functions to move locality data from one coordinate system to another. Before PROJ4 can be used for this purpose, it is necessary to format input data into degrees, minutes, seconds (DMS) format. The biodiversity database uses a dual system to store positional information for specimen observations. In the first instance observations are stored in decimal degree notation, with two floating point fields in the database storing the latitude and longitude. In other cases only the QDS in which the specimen was observed is known and this is stored in a text field called 'grid' (see Table 4.1 for a sample listing).

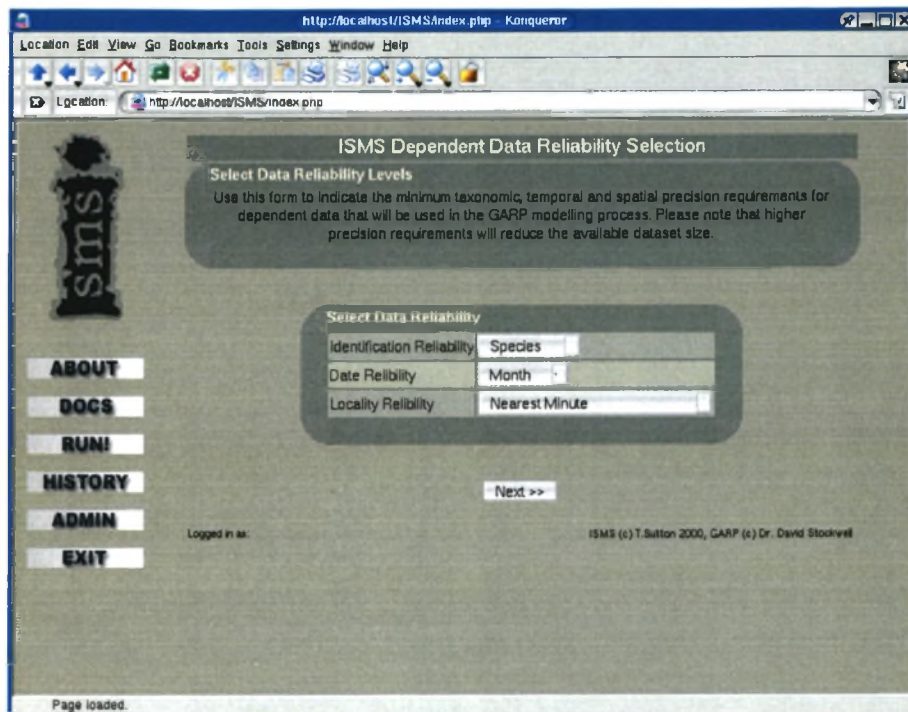


Figure 4.9: Metadata describing the temporal, spatial and taxonomic accuracy of the data are used to filter the independent variable dataset used.

Thus two conversion methods were implemented in ISMS. Type 1 conversions are used to convert directly from decimal degrees to a DMS format suitable for the PROJ4 application. Type 2 conversion is a two-step process whereby the centroid of the QDS is calculated and the resulting coordinate (in decimal degree format) is then converted to PROJ4 DMS format. One potential difficulty with the latter approach is the illusion that site data have been recorded at a finer resolution than is actually the case. To overcome this ISMS provides a mechanism for the user to select independent data that are appropriate to the selected independent variable data layers (see Section 4.3 for details on independent variable layer selection). This selection process is facilitated by meta data stored in the biodiversity database for each locality record. The user's required level of spatial accuracy is cross matched to the meta data for each locality record. If there is a conflict (for example where a user has specified that spatial quality should be 'nearest second' or better), locality records that do not match this criteria will be excluded (for example where the meta data indicates that spatial precision is only to the nearest quarter degree). Conversely if the user has chosen a low level of spatial precision such as 'quarter degree', both records that were originally point locality records, and centroid records will be used.

Once the necessary data conversion has taken place, the ISMS application invokes the PROJ4 application to re-project the data into the same projection system as the independent variable



layers. The information required to do this is retrieved from the ISMS Database Projection table. The resulting text file containing projected specimen localities is then passed to the GARP 'rasteriz' program using a UNIX pipe. During this rasterisation stage the sample size may be further reduced if data are not within the spatial extents of the study area, are within a masked area, or if duplicate entries occur for the same grid cell. Feedback from the 'rasteriz' process (as illustrated in Figure 4.10) is provided to users at this stage, who may choose to abort the process should they be unhappy with the number of samples that are being made available to the GARP engine. If the user elects to continue at this point, the GARP model will commence using the various data assembled during the model preparation phase.

It was found that in many cases long execution times would cause the client web browser to time out before the model had reached completion. To circumvent this problem, it was decided to use the Gnu 'at' command to queue ISMS jobs. Additional application logic was written to automatically email the job owner once the job reaches completion, notifying the ISMS user of the successfully completed job. Email details associated with each ISMS user account are used for this purpose. In addition the ISMS Database Iteration table is populated by the ISMS model execution batch script with the results of each stage of the GARP model after it has executed.

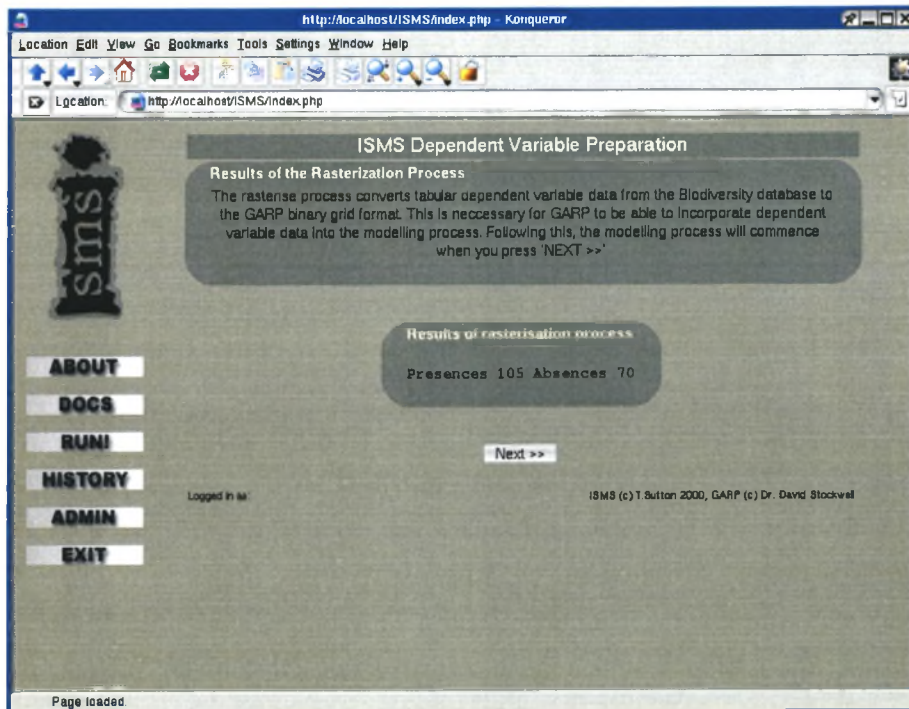


Figure 4.10: Feedback from the rasterisation process allows users to decide if they wish to continue with the model iteration.

## 4.8 HISTORY BROWSER

Once the user receives email notification of the completion of their model iteration, the history browser can be used to access the model outputs. Model outputs are grouped by extent, requiring the user to first select the extent (geographical area) in which the model was carried out. Once an extent has been selected, a complete listing of GARP model iterations is provided. These history listings are linked to the user currently logged in - only iterations 'owned' by that user are displayed. Facilities provided in the model history listing include the ability to delete model iterations, view model outputs as an html page, or view model outputs as a Portable Document Format (PDF) document. Figure 4.11 shows the history browser form.

In order to accommodate deleting a model result from the system, an icon is placed next to each iteration listed in the ISMS model history browser. By clicking on this icon, the iteration record in the Iteration table is deleted and associated hard-drive space is made available by deleting the working directory for that model. The functionality provided by the PDF and html formatted reports are fundamentally different. The PDF format is intended to be used to transfer models reports from the ISMS server to a client machine. The HTML format report is intended to be viewed online and includes hyperlinks to downloadable Arc/Info files containing model results. The latter are generated by the ISMS using the GARP image program.

The production of PDF output is achieved using the LATEX typesetting language. LATEX is a typesetting command language used to produce complex documents (Kopka & Daly 1999). The PHP ISMSHistory class executes a series of database queries to obtain all relevant model metadata and outputs and a LATEX document is dynamically generated when the user presses the 'printable view' icon on the history browser page. A PHP system call then invokes the 'texi2pdf' command to convert the generated LATEX file into a PDF document. This approach provides a highly flexible method of producing a visually rich, cross-platform, 'printer friendly' output from the ISMS. An example of the content of the PDF output for a completed model run is provided as Appendix G.

## 4.9 SPATIAL MANIPULATION AND VISUALISATION OF RESULTS

One of the key factors in selecting the GMS as the modelling engine for the ISMS was its ability to produce model outputs in both a tabular and spatial format. Tabular model outputs are difficult to visualise in a spatial context and the production of a range map at the end of the



Figure 4.11: The ISMS model history browser allows individual users to manage their personal collection of model results.

modelling process allows the end user to make a rapid assessment of the model results. This chapter briefly describes the test procedure used to ensure that ISMS outputs can be usefully employed within a GIS environment.

The ISMS uses the ImageMagick ‘convert’ program suite to manipulate the model outputs produced by the GARP IMAGE program. These manipulations include resampling the image to a size more suited to the bandwidth and display medium constraints of a web browser interface, and production of a thumbnail strip of the various visual outputs. It was however realised that users may wish to combine the outputs of GARP with desktop GIS applications. To this end, the facility was added to the ISMS application to allow users to download the model outputs as Arc/Info ASCII Grid files (as outlined in Section 4.8). In order to test that these output files would work successfully in a GIS desktop application, the GRASS GIS package was selected. GRASS is an Open Source (Neteler & Mitasova 2002) application that runs on the GNU/Linux platform. The following sections describe the procedure followed in order to demonstrate the visualisation of outputs generated from the GMS/ISMS in a GIS software package.



### 4.9.1 Importing the dataset into Grass 5.10

In order to import ASCII/Grid data into the GRASS GIS, a project was first created with extent parameters matching the study area. Figures H.1 through H.5 in Appendix H illustrate this process. The model outputs from any given model iteration could then be imported into GRASS by decompressing the model outputs provided by the history browser and then using the Arc/Info ASCII grid import tool.

### 4.9.2 Viewing model results in 3D

In cases where abiotic data for the study include a digital elevation model (DEM), the GARP model results can be visualised in three dimensions (3D)<sup>1</sup> using the GRASS NVIZ module. A DEM is used by GIS software to represent the variations in altitude across a landform in three dimensions (Delaney 2001).

'NVIZ' is a three-dimensional (3D) spatial data visualisation tool for the GRASS GIS. NVIZ uses the 'Z' attribute of each cell in a matrix to create an elevation model. Additional functionality allows other datasets (including vector data) to be 'draped' over the 3D surface, and to fine tune the lighting and shading effects of the 3D scene produced.

Visualising model outputs in a 3D format allows biologists to better visualise the dependent variable data in relation to predicted distribution range. Once GARP model outputs had been imported into the GRASS environment, a Digital Elevation Model that forms part of the independent datasets was used to produce a 3D scene. The outputs from the GARP model were then draped over the 3D surface. In addition the dependent variable data were imported into the system and overlaid over the model outputs. This proved a useful way to study the model outputs in relation to the actual data that were used to derive the model. As discussed in Chapter 1, errors of commission and omission are a problem with distribution modelling exercises (Csuti & Crist 1998). Using overlay techniques, biologists can make an informed interpretation of how successful the model iteration was and quickly spot errors of omission and errors of commission (as illustrated in Figure 4.12).

---

<sup>1</sup>Three-dimensional visualisation in GIS applications of the nature described here is more accurately described as 2.5D as certain aspects of the landform (e.g. caves and overhangs) are not accurately represented by the digital elevation model (Weibel & Heller 1991).



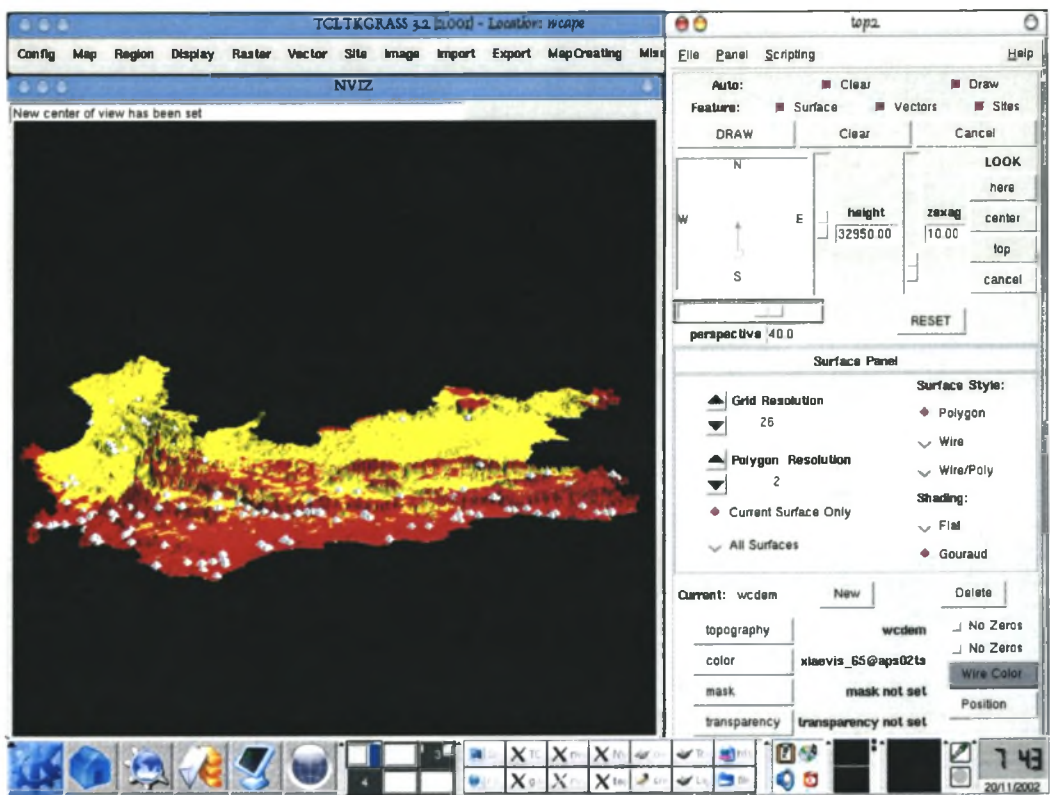


Figure 4.12: A GARP model displayed in 3D relief using the GRASS GIS NVIZ module.

Table 4.1: Sample listing from the biodiversity database illustrating data stored in decimal degrees (first record) and quarter-degree Square (second record).

Name	ID	Dec. De- grees East	Dec. Degrees South	Qrt. Degree Grid
Xenopus laevis laevis	1184	27.0825	-31.31749	3127AC
Xenopus laevis laevis	1184			2823BA

## CHAPTER 5: CONCLUSION

The GARP Modelling System was identified as a useful tool for modelling the distribution of plant and animal taxa (Sutton 1999). The user interface to the GMS is not, however, Graphical User Interface (GUI) based and requires a fairly substantial time investment to learn. The Integrated Species Modelling System was conceived as a way to address the problems experienced by the author in usability of the GMS. A requirement that the system should be delivered using Internet technologies was made in order to ensure the widespread accessibility of the proposed system.

A detailed system design process was carried out that included the establishment of hardware and software system requirements. It was established that a loosely coupled software architecture would provide a robust and extensible environment for the development of the ISMS. Also it was decided to use a three-tier (Williams & Lane 2002) application architecture for the system.

Detailed workflow diagrams were developed to 'map out' the various interactions that need to be carried out by users in order to facilitate administration, model execution and visualisation of model results. The workflow diagrams produced were used to prototype a UML class model and a database entity relationship diagram for the system. Various minor modifications were made to the existing biodiversity database in use by the Western Cape Nature Conservation Board, and a technique for querying this database from the GNU/Linux platform was identified. The biodiversity database forms the basis for the database tier in the three-tier software model.

A series of PHP classes were implemented using the UML prototypes developed during the design phase. These PHP classes formed the application layer in the three tier hierarchy, implementing all of the core application logic for the system. The PHP classes also carry out all of the coupling work - invoking external programs at different stages of the model preparation, execution and results display process. GUI rendering routines were also encapsulated within these classes, allowing for input and output to be implemented using the Hypertext Mark-up Language (HTML).

When the software implementation was complete, users were able to interact with the system using a web browser such as Internet Explorer, Mozilla or Konqueror. The web browser installed

on end users' computers thus formed the client tier in the three-tier development model. User response was provided informally after implementation of the system. The user experience was positive as the removal of the difficult to use command line interface barrier enabled relatively effortless access to the functionality offered by the GARP Modelling System.

An additional function was implemented which allows end users to download the model outputs as ASCII Grid files - a generic format useable by a variety of Geographical Information System software packages, including ArcView 3.1 (with the Spatial Analyst extension) and the GRASS Open Source GIS package. The possibilities for incorporation of model results into standard GIS package functions, such as overlay analysis and 3D visualisation of model results, were then explored.

## 5.1 REALISATION OF DESIGN GOALS

Using a structured development model allowed each of the primary design goals to be realised. Seamless access to the WCNCB biodiversity database was achieved using the FreeTDS driver. Program code was implemented to automate the projection of dependent variable data into the same projection system as that of the independent variable data used in a GARP model iteration. An easy to use GUI environment was developed for carrying out model iterations and visualising the results of model iterations. The system allows biologists to interactively explore the effect of different model parameters and independent data layers for a given dependent variable. Abstraction of all the GMS applications into one integrated modelling environment was achieved allowing users to focus on the problem domain (distribution prediction) rather than trying to understand the operation of the various GARP applications. The ISMS history browsing module provides model output management services, allowing users to browse the collection of outputs for models they have initiated. Concurrent access to the system is facilitated using \*NIX symlinks, command scheduling and the multitasking facilities of the underlying server operating system. The system was implemented in a heterogeneous environment consisting of various operating systems and hardware devices. The system was implemented as a loosely coupled system that provides a unified user interface to a variety of discrete software applications.



## 5.2 ADVANTAGES OF THE SYSTEM

The ISMS provides numerous advantages over the command line interface approach taken by the original GARP application. These advantages are discussed briefly below:

**Concurrent model execution:** The GARP application is not well suited to concurrent use. If more than one user wants to use GARP at the same time (or one user wishes to execute more than one model at the same time), the GARP binary files need to be copied for each instance and a separate working directory created with copies of any common independent and dependent variable data. This results in a high administrative overhead and high consumption of hard-disk space. The ISMS addresses this by automating the creation of multiple working directories and by using \*NIX symlinks to allow for efficient utilisation of disk space. As a result the number of concurrent GARP models that can be carried out is now limited by processor and memory resources, rather than administrative load. Any limitations on concurrent model execution can be addressed relatively easily and cost effectively by upgrading the hardware platform on which the ISMS is installed.

**Complete annotated history for each model iteration:** The Portable Document Format (PDF) was used to create a detailed report for each completed model iteration. This report includes the non-spatial outputs from the GARP model, a detailed listing of all parameters used to drive the model and all spatial outputs from the model - provided as images within the summary document. This report can then be displayed as an embedded document within the user's web browser or downloaded for printing or offline perusal.

**Seamless integration with the biodiversity database:** The biodiversity database contains dependent variable data. The administrative overhead in preparing dependent data for model execution is substantially reduced through the use of automated procedures to extract data from the biodiversity database, project the spatial component of this data into a Cartesian coordinate space, and finally to prepare the GARP binary file from this data ready for use in a model iteration.

**Remote, platform-independent model execution using Internet technology:** By adopting a 'thin client' approach to the application architecture and by standardising on the use of a web browser to provide an interface to the system, the ISMS is usable even on fairly low-end personal computers. As a result the minimum requirements for the client machine are



a suitable network connection and a web browser. In addition, the use of Internet technology has allowed the system to be developed in such a way that it will run on virtually any operating system platform that meets the aforementioned requirements.

**Simple, easy to use interface to GARP:** The development of an easy to use GUI interface makes the GARP modelling process accessible to non-technical users. The user interface provided by the ISMS removes many of the barriers that may prevent biologists from carrying out experiments using the GMS. Although a small amount of training is still required, carrying out GARP iterations is far easier and less confusing than using the Command Line Interface to GARP (based on the author's personal experience).

### 5.3 SHORTCOMINGS OF THE SYSTEM

The setup and installation of the ISMS remains a non-trivial task. It is assumed that an experienced system administrator would be available to carry out the installation process. This is perhaps a limitation in single-user environments where desktop applications such as the recently released Desktop GARP application (Townsend Peterson, Sanchez-Cordero, Beard & Ramsey 2002) may prove to be a better option. The aforementioned option does not allow transparent access to dependent variable data residing in a biodiversity database or detailed parameter specification, so the gains achieved from ease of installation may be offset by data preparation overheads in the modelling process.

Another shortcoming that has been identified is that the system has been 'hard-coded' to use a Microsoft SQL Server database for the independent data store. It would certainly be of benefit to refactor the ISMS application so that dependent variable data are extracted from the database tier via a database abstraction layer, so that multiple database platforms could be supported. This would also make the technologies implemented by the GMS and ISMS applications more accessible to poorly funded organisations, as the entire implementation could then be implemented using free, Open Source software. For example, a PostgreSQL database could be used in place of MS SQL Server.

Lastly, it should be realised that, although the ISMS makes the execution of GARP models a relatively straightforward process for end-users, users of the system should nonetheless have a thorough understanding of the theory (Stockwell 1999) and limitations of the GARP modelling system.

## 5.4 FUTURE RESEARCH DIRECTIONS

The application of Internet technologies to biological and spatial problems presents exciting possibilities for the future. Certainly an area ripe for further development is the spatial representation and exploration of the results produced by the GMS and ISMS. In its current implementation the ISMS presents model outputs as static JPEG images, or as ASCII grid export files. The former output is limited by bandwidth constraints - only relatively small images of spatial data can be presented over the Internet in this manner. The latter output presents its own problems as the users are required to have their own desktop GIS package such as ArcView or GRASS in order to interpret the results, as well as sufficient skill to use these desktop applications effectively.

These limitations certainly pave the way for integrating a web mapping platform such as MapServer (Vatsavai, Burk, Wilson & Shekhar 2000) into the ISMS application suite. The latter could also be used effectively to allow the user to visualise independent variable data as a 'spatial preview' in order to determine if this data is of a satisfactory quality. Such a tool could allow the user to use the same working environment (i.e. a web browser) to perform more detailed visual analysis of the GARP model outputs.

The field of bioinformatics is developing rapidly, with major developments taking place in the development of inter-organisational knowledge, data and software collaboration 'federations' (e.g. Stockwell, Arzberger, Fountain & Helly (2000)). Examples of these federations already in operation include the Species2000 project (Bisby 2000), Species Analyst (Vieglas, Stockwell, Cundari, Beach, Peterson & Krishtalka 1998) and the Global Biodiversity Information Framework (GBIF) project (Edwards, Lane & Nielsen 2000). These advances in interoperability will make modelling at a global scale using distributed data from around the world feasible. The development of formalised ontologies for GIS data (Egenhofer 2002) and potentially for biological data will make the 'harvesting' of data across the Internet a more viable alternative, reducing the possibility of data incompatibility through the development of a consistent semantic model within the distribution modelling problem domain. The next generation GARP toolkit, called 'WhyWhere' is already pioneering global species distribution modelling by providing a set of useful independent data on a global scale (Stockwell 2001). However the latter suffers from many of the same limitations as the current GARP system (users don't have ability to dynamically use their own dependent and independent variable data holdings and no model history system is provided). Thus a new area of potential research (for the ISMS) relates to the issue of biodiversity data exchange. There are various examples of the use of technologies such as XML, the eXtensible Mark-up Language (Bisby 2000) and Z39.50 (Vieglas *et al.* 1998). To provide a truly ubiquitous modelling environment, it would be

---

useful to find ways to integrate the ability to incorporate these data exchange protocols into the ISMS.

As briefly pointed out in the opening chapter of this paper, it should also be noted that other alternatives to GARP exist for modelling species distribution. A more ideal modelling environment would encompass the ability to select from a variety of distribution modelling algorithms, so that the most applicable algorithm could be used for any model iteration, using data from a number of federated species distribution databases connected to the Internet.

## REFERENCES

- Allen J & Hornberger C, 2002. *Mastering PHP 4*. Sybex, London.
- Anderson RP, Lew D & Townsend Peterson A, 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, 162(2003):211–232.
- Austin MP, 1991. Vegetation: Data Collection and Analysis. In Cargules CR & Austin MP (eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. Australia: CSIRO.
- Berendsohn WA (ed.), 2000. *Resource Identification for a Biological Collection Information Service in Europe (BioCISE)*. Berlin: Botanical Garden and Botanical Museum Berlin-Dahlem.
- Bisby FA, 2000. The Quiet Revolution: Biodiversity Informatics and the Internet. *Science*, 289:2309–2312.
- Branch WR, 1988. South African Red Data Book - Reptiles and Amphibians. South African National Scientific Programmes Report. 151, CSIR, Pretoria.
- Brown NJ, Swetnam RD, Manchester SJ, Mountford JO, Caldow RWG & Treweek JR, 1996. A wetlands GIS: an ecologist's tool for wet grassland management. Tech. rep., ITE Annual Report, Institute of Terrestrial Ecology, Huntingdon.
- Burrough PA & McDonnell RA, 2000. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, UK.
- Busby JR, 1991. *BIOCLIM - A Bioclimatic Analysis and Prediction System*. In: Margules, CR & Austin, MP (eds.) *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Canberra.
- Caithness N, 1995. *Pattern, process and the evolution of the African antelope (Mammalia Bovidae)*. PhD Thesis, University of the Witwatersrand. Johannesburg.
- Chrisman N, 2002. *Exploring Geographical Information Systems*. Second Edition. John Wiley and Sons, Inc., USA.



- Cocks KD & Baird IA, 1991. *Margules, CR & Austin, MP (eds) In: Nature conservation: cost effective biological surveys and data analysis*, chap. The role of geographic information systems in the collection, extrapolation and use of survey data, pp. 74–78. Melbourne: CSIRO.
- Csuti B & Crist P, 1998. Methods for Assessing Accuracy of Animal Distribution Maps for GAP Analysis. Internet : <http://www.gap.uidaho.edu/gap/handbook/VDA/index.htm> [6.6.1999].
- Davis FW, Stoms AD, Hollander KA, Thomas PA, Stine D, Odion MI, Borchert JH, Thorne MV, Gray RE, Walker K, Warner & Graae J, 1998. *The California Gap Analysis Project-Final Report*. University of California, Santa Barbara, CA. Available: [http://www.biogeog.ucsb.edu/projects/gap/gap\\_rep.html](http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html) [7.6.1999].
- De' ath G, 2001. Multivariate Regression Trees: A New Technique For Modelling Species-Environment Relationships. *Ecology*, 83(4):1105–1117.
- Delaney J, 2001. *Geographical Information Systems - An Introduction*. Oxford University Press, Australia.
- Edwards J, Lane MA & Nielsen ES, 2000. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*, 289:2312–2314.
- Egenhofer MJ, 2002. *Toward the Semantic Geospatial Web*. Proceedings of the tenth ACM international symposium on Advances in geographic information systems. ACM Press New York, NY, USA.
- Erasmus BFN, Van Jaarsveld AS, Chown SL, Kshatriya M & Wessels KJ, 2002. Vulnerability of South Arican taxa to climate change. *Global Change Biology*, 8:679–693.
- ESRI, 1996. *ArcView Spatial Analyst user's manual*. Environmental Systems Research Institute, Inc. USA.
- Evenden GI, 1995. *Cartographic Projection Procedures for the UNIX Environment - A User's Manual*. United States Department of the Interior Geological Survey, Open File Report 90-284.
- Fielding AH & Bell JF, 1997. A review of methods for assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1):38–49.
- Flanagan D, 2002. *Javascript, the Definitive Guide (Fourth Edition)*. O'Reilly & Associates, California.
- Franklin J, 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19:474–499.

- Garey MR & Johnson DS (eds.), 1979. *A Guide to the Theory of Np-Completeness*. New York: W H Freeman Co.
- Graham I, 1991. *Object Orientated Methods*. Addison-Wesley Publishers Ltd, UK.
- Hirzel AH, Hausser J, Chessel D & Perrin N, 2002. Ecological-Niche Factor Analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83(7):2027–2036.
- Hollander AD, Davis FW & Stoms DM, 1993. *Hierarchical representations of species distributions using maps, images and sighting data*. In Miller RI (ed.) *Mapping the Diversity of Nature*. Chapman & Hall, Oxford, Great Britain.
- Hutchinson MF, 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, 106:211–232.
- Kang-tsung C, 2002. *Introduction to Geographic Information Systems*. McGraw-Hill Higher Education, New York.
- Knudsen C, 1999. PHP version 4. Internet : <http://noframes.linuxjournal.com/lj-issues/issue67/3649.html>.
- Kopka H & Daly PW, 1999. *A Guide to Latex*. London: Addison-Wesley, 3rd edn.
- Liberty J, 2002. *Teach Yourself C++ in 10 Minutes*. Indianapolis, Indiana, USA: Sams Publishing, 2nd edn.
- Moritsugu S, Edwards J, S G, Horvath D, Marler G, Pedersen J, Pitts D & Wilson D, 1998. *Using Unix, Second Edition*. Que, USA.
- Neteler M & Mitasova H, 2002. *Open Source GIS: A GRASS GIS Approach*. Kluwer Academic Publishers, Boston, Dordrecht.
- Nieminen J, 2002. In search of Habitats testing a GNU approach. In *Proceedings of the Open Source GIS - GRASS users conference 2002*. Trento, Italy.
- Nix HA, 1986. *A biogeographic analysis of Australian Elapid snakes in Longmore, R. (ed.) Atlas of Australian Elapid Snakes*. 8. Australian Flora and Fauna Series.
- Payne K & Stockwell DRB, 1999. GARP Modelling System user's guide and technical reference. Internet : <http://biodi.sdsc.edu/Doc/GARP/Manual/manual.htm>.
- Pereira JMC & Itami RM, 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing*, 57:1475–1486.

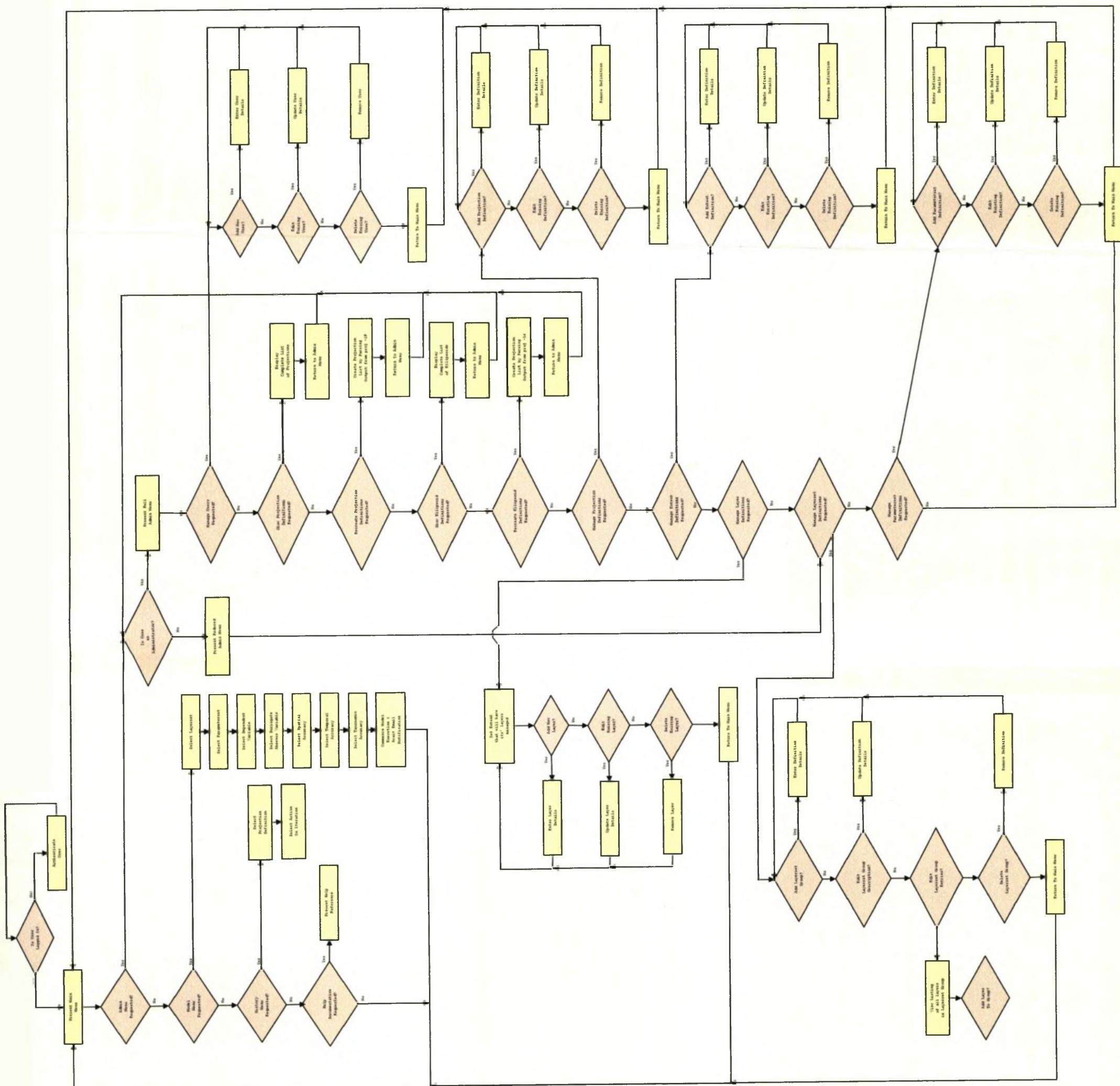
- Pratt PJ & Adamski J, 1994. *Database Systems Management and Design*. USA: CTI Publishers, 3rd edn.
- Pressman RS, 2000. *Software Engineering, A Practitioner's Approach - European Adaptation, Fifth Edition*. McGraw-Hill Publishing, UK.
- Rutherford MC, Midgely GF, Bond WJ, Powrie LW, Roberts R & Allsopp J, 1999. *Plant Biodiversity : Vulnerability and Adaptation Assessment*. South African Country Study on Climate Change, Dept of Environment Affairs and Tourism. Pretoria.
- Sala OE, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, LeRoy Poff N, Sykes M, Walker B, Walker M & Wall D, 2000. Global Biodiversity Scenarios for the Year 2100. *Science*, 287:1770–1774.
- Stallman R, 2002. Linux and the GNU Project. Internet : <http://www.gnu.org/gnu/linux-and-gnu.html> [09.09.2002].
- Stockwell D, 2001. Introduction to WhyWhere. Internet : <http://biodi.sdsc.edu/Doc/WhyWhere/tutorial.html> [01.12.2002].
- Stockwell D, Arzberger P, Fountain T & Helly J, 2000. An Interface between Computing, Ecology and Biodiversity: Environmental Informatics. *Korean Journal of Ecology*, 23(2):101–106.
- Stockwell D & Peters D, 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science.*, 13(2):143–158.
- Stockwell DRB, 1999. The role of biological and environmental data in modelling landscape patterns and processes. Internet : <http://biodi.sdsc.edu/Doc/GARP/Manual/process.html> [3.6.1999].
- Stockwell DRB, Davey SM, Davis JR & Noble IR, 1990. Using Induction of Decision Trees to Predict Greater Glider Density. *AI Applications*, 4(4):33–43.
- Stockwell DRB & Gitay H, 1992. *Using rule-sets to explain animal response to the environment*. In Machine Learning and the Problem of Prediction and Explanation in Ecological Modelling. Thesis submitted for the degree of Doctor of Philosophy to the Australian National University in May , 1992. pp175:206.
- Stockwell DRB & Noble IR, 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, 33(1992):385–390.



- Sutton TP, 1999. *A review of the techniques available for modelling the probable distribution of lower vertebrates in the Western Cape Province of South Africa*. Literature review presented in partial fulfilment of the requirements for the Masters Degree (in Geographical and Environmental Studies) at the University of Stellenbosch.
- Tokumine ST, McClean C & Lovett JC, 2002. *Genetic Algorithm based bioclimatic modelling of continental scale sub-Saharan African plant distributions in a GIS*. Unpublished.
- Townsend Peterson A & Cohoon KP, 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling*, 117:159–164.
- Townsend Peterson A, Sanchez-Cordero V, Beard C & Ramsey J, 2002. Ecological Niche Modeling and Potential Reservoirs for Chagas Disease, Mexico. *Emerging Infectious Diseases*, 8(7):662–667.
- Ullman LE, 2001. *PHP Advanced for the World Wide Web*. Berkeley, CA, USA: Peach Pit Press.
- Van Jaarsveld AS, Chown SL, Erasmus BFN, Kshatriya M & Wessels KJ, 2000. Vulnerability and adaptation assessment of South African animal taxa to climate change. Tech. rep., Centre for Environmental Studies, University of Pretoria.
- Vatsavai RR, Burk TE, Wilson TB & Shekhar S, 2000. A Web-based browsing and spatial analysis system for regional natural resource analysis and mapping. In *Proceedings of the Eighth ACM International Symposium on Advances in Geographic Information Systems*. ACMGIS.
- Viegas DA, Stockwell DRB, Cundari CM, Beach J, Peterson AT & Krishtalka, 1998. *The Species Analyst: Tools Enabling a Comprehensive Distributed Biodiversity Network*. Biodiversity, Biotechnology & Biobusiness Conference. Perth.
- Weibel R & Heller M, 1991. Digital Terrain Modelling. In J MD, Goodchild MF & Rhind DW (eds.), *Geographical Information Systems - Principles and Applications*, vol. 1, pp. 269–297. Avon, UK: Longman Scientific and Technical.
- Williams HE & Lane D, 2002. *Web Database Applications with PHP and MySQL*. O' Reilly & Associates, California.

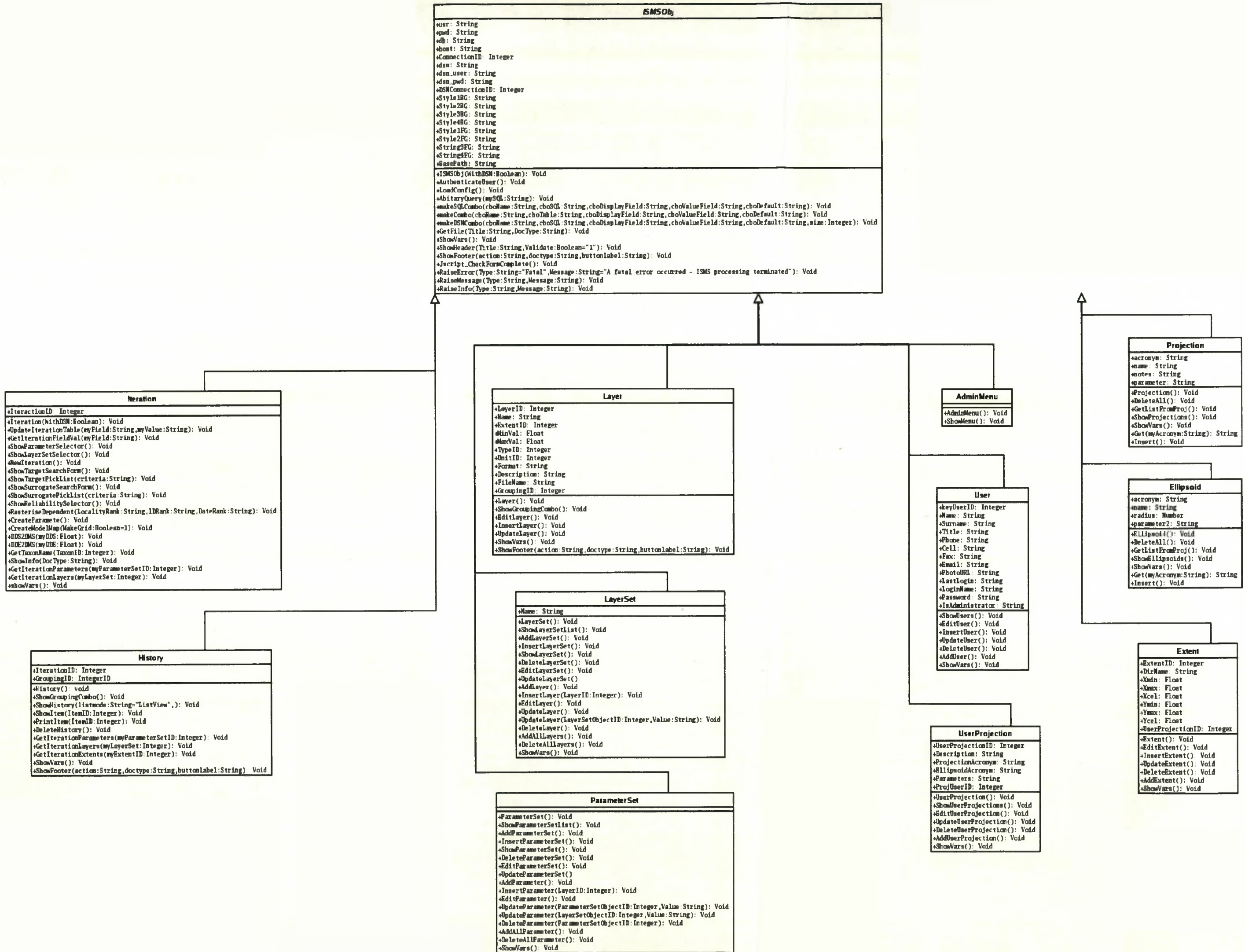


## **APPENDIX A: APPLICATION FLOW CHART**



## **APPENDIX B: CLASS DIAGRAM FOR THE ISMS APPLICATION**







## APPENDIX C: ISMS PROJECTION DATABASE

The projection database consists of named projection definitions (compliant with the PROJ4 internal naming schema), along with the parameters expected by the PROJ4 application in order to carry out a projection operation.

Acronym	Name	Description	Parameters
aea	Albers Equal Area	Conic Sph&Ell	lat_1= lat_2=
aeqd	Azimuthal Equidistant	Azi, Sph&Ell	lat_0 guam
airy	Airy	Misc Sph, no inv.	no_cut lat_b=
aitoff	Aitoff	Misc Sph	
alsk	Mod. Stererographics of Alaska	Azi(mod)	
apian	Apian Globular I	Misc Sph, no inv.	
august	August Epicycloidal	Misc Sph, no inv.	
bacon	Bacon Globular	Misc Sph, no inv.	
bipc	Bipolar conic of western hemi- sphere	Conic Sph.	
boggs	Boggs Eumorphic	PCyl., no inv., Sph.	
bonne	Bonne (Werner lat_1=90)	Conic Sph&Ell	lat_1=
cass	Cassini	Cyl, Sph&Ell	
cc	Central Cylindrical	Cyl, Sph	
cea	Equal Area Cylindrical	Cyl, Sph&Ell	lat_ts=
chamb	Chamberlin Trimetric	Misc Sph, no inv.	lat_1= lon_1= lat_2= lon_2= lat_3= lon_3=

Acronym	Name	Description	Parameters
<b>collg</b>	Collignon	PCyl, Sph.	
<b>crast</b>	Craster Parabolic (Putnins P4)	PCyl., Sph.	
<b>denoy</b>	Denoyer Semi-Elliptical	PCyl., no inv., Sph.	
<b>eck1</b>	Eckert I	PCyl., Sph.	
<b>eck2</b>	Eckert II	PCyl. Sph.	
<b>eck3</b>	Eckert III	PCyl, Sph.	
<b>eck4</b>	Eckert IV	PCyl, Sph.	
<b>eck5</b>	Eckert V	PCyl, Sph.	
<b>eck6</b>	Eckert VI	PCyl, Sph.	
<b>eqc</b>	Equidistant Cylindrical (Plate Caree)	Cyl, Sph	lat_ts=
<b>eqdc</b>	Equidistant Conic	Conic, Sph&Ell	lat_1= lat_2=
<b>euler</b>	Euler	Conic, Sph	lat_1= and lat_2=
<b>fahey</b>	Fahey	Pcyl, Sph.	
<b>fouc</b>	Foucaut	PCyl., Sph.	
<b>fouc_s</b>	Foucaut Sinusoidal	PCyl., Sph.	
<b>gall</b>	Gall (Gall Stereographic)	Cyl, Sph	
<b>gins8</b>	Ginsburg VIII (TsNIIGAiK)	PCyl, Sph., no inv.	
<b>gnom</b>	Gnomonic	Azi, Sph.	
<b>gn_sinu</b>	General Sinusoidal Series	PCyl, Sph.	m= n=
<b>goode</b>	Goode Homolosine	PCyl, Sph.	
<b>gs48</b>	Mod. Stererographics of 48 U.S.	Azi(mod)	
<b>gs50</b>	Mod. Stererographics of 50 U.S.	Azi(mod)	
<b>hammer</b>	Hammer & Eckert-Greifendorff	Misc Sph, no inv.	W= M=
<b>hatano</b>	Hatano Asymmetrical Equal Area	PCyl, Sph.	
<b>imw_p</b>	International Map of the World Polyconic	Mod. Polyconic, Ell	lat_1= and lat_2= [lon_1=]

Acronym	Name	Description	Parameters
kav5	Kavraisky V	PCyl., Sph.	
kav7	Kavraisky VII	PCyl, Sph.	
labrd	Laborde	Cyl, Sph	Special for Madagascar
laea	Lambert Azimuthal Equal Area	Azi, Sph&Ell	
lagrng	Lagrange	Misc Sph, no inv.	W=
larr	Larrivee	Misc Sph, no inv.	
lask	Laskowski	Misc Sph, no inv.	
latlong	Lat/long (Geodetic)		
lcc	Lambert Conformal Conic	Conic, Sph&Ell	lat.1= and lat.2= or lat.0
leac	Lambert Equal Area Conic	Conic, Sph&Ell	lat.1= south
lee_os	Lee Oblated Stereographic	Azi(mod)	
longlat	Lat/long (Geodetic)		
loxim	Loximuthal	PCyl Sph	
lsat	Space oblique for LANDSAT	Cyl, Sph&Ell	lsat= path=
mbtfpp	McBride-Thomas Flat-Polar Parabolic	Cyl., Sph.	
mbtfpq	McBryde-Thomas Flat-Polar Quartic	Cyl., Sph.	
mbtfps	McBryde-Thomas Flat-Polar Sinusoidal	PCyl, Sph.	
mbt_fps	McBryde-Thomas Flat-Pole Sine (No. 2)	Cyl., Sph.	
mbt_s	McBryde-Thomas Flat-Polar Sine (No. 1)	PCyl., Sph.	
merc	Mercator	Cyl, Sph&Ell	lat.ts=
mill	Miller Cylindrical	Cyl, Sph	
mil_os	Miller Oblated Stereographic	Azi(mod)	
moll	Mollweide	PCyl., Sph.	

Acronym	Name	Description	Parameters
mpoly	Modified Polyconic	Polyconic, Sph	lat_1= and lat_2= lotsa
murd1	Murdoch I	Conic, Sph	lat_1= and lat_2=
murd2	Murdoch II	Conic, Sph	lat_1= and lat_2=
murd3	Murdoch III	Conic, Sph	lat_1= and lat_2=
nell_h	Nell-Hammer	PCyl., Sph.	
nell	Nell	PCyl., Sph.	
nicol	Nicolosi Globular	Misc Sph, no inv.	
nsper	Near-sided perspective	Azi, Sph	h=
nzm	New Zealand Map Grid	fixed Earth	
ob_tran	General Oblique Transformation	Misc Sph	o.proj= plus parameters for projection
oce	Oblique Cylindrical Equal Area	Cyl, Sph lonc= alpha= or	lat_1= lat_2= lon_1= lon_2=
oea	Oblated Equal Area	Misc Sph	n= m= theta=
omerc	Oblique Mercator	Cyl, Sph&Ell	no_rot rot_conv no_uoff and
ortel	Ortelius Oval	Misc Sph, no inv.	
ortho	Orthographic	Azi, Sph.	
pconic	Perspective Conic	Conic, Sph	lat_1= and lat_2=
poly	Polyconic (American)	Conic, Sph&Ell	
putp1	Putnins P1	PCyl, Sph.	
putp2	Putnins P2	PCyl., Sph.	
putp3p	Putnins P3	PCyl., no inv., Sph.	
putp3	Putnins P3	PCyl., Sph.	
putp4p	Putnins P4	PCyl., Sph.	
putp5p	Putnins P5	PCyl., Sph.	
putp5	Putnins P5	PCyl., Sph.	
putp6p	Putnins P6	PCyl., Sph.	
putp6	Putnins P6	PCyl., Sph.	



Acronym	Name	Description	Parameters
qua_aut	Quartic Authalic	PCyl., Sph.	
robin	Robinson	PCyl., Sph.	
rpoly	Rectangular Polyconic	Conic, Sph., no inv.	lat_ts=
sinu	Sinusoidal (Sanson-Flamsteed)	PCyl, Sph&Ell	
somerc	Swiss. Obl. Mercator	Cyl, Ell	For CH1903
stere	Stereographic	Azi, Sph&Ell	lat_ts=
tcc	Transverse Central Cylindrical	Cyl, Sph, no inv.	
tcea	Transverse Cylindrical Equal Area	Cyl, Sph	
tissot	Tissot	Conic, Sph	lat_1= and lat_2=
tmerc	Transverse Mercator	Cyl, Sph&Ell	
tpeqd	Two Point Equidistant	Misc Sph	lat_1= lon_1= lat_2= lon_2=
tpers	Tilted perspective	Azi, Sph	tilt= azi= h=
ups	Universal Polar Stereographic	Azi, Sph&Ell	south
urm5	Urmaev V	PCyl., Sph.	n= q= alpha=
urmfps	Urmaev Flat-Polar Sinusoidal	PCyl, Sph.	n=
utm	Universal Transverse Mercator (UTM)	Cyl, Sph	zone= south
vandg2	van der Grinten II	Misc Sph, no inv.	
vandg3	van der Grinten III	Misc Sph, no inv.	
vandg4	van der Grinten IV	Misc Sph, no inv.	
vandg	van der Grinten (I)	Misc Sph	
vitk1	Vitkovsky I	Conic, Sph	lat_1= and lat_2=
wag1	Wagner I (Kavraisky VI)	PCyl, Sph.	
wag2	Wagner II	PCyl., Sph.	
wag3	Wagner III	PCyl., Sph.	
wag4	Wagner IV	PCyl., Sph.	
wag5	Wagner V	PCyl., Sph.	

Acronym	Name	Description	Parameters
wag6	Wagner VI	PCyl, Sph.	
wag7	Wagner VII	Misc Sph, no inv.	
weren	Werenskiold I	PCyl., Sph.	
wink1	Winkel I	PCyl., Sph.	lat_ts=
wink2	Winkel II	PCyl., Sph., no inv.	lat_1=

## APPENDIX D: ISMS ELLIPSOID DATABASE

Acronym	Name	Radius	Parameter2
MERIT	MERIT 1983	a=6378137.0	rf=298.257
SGS85	Soviet Geodetic System 85	a=6378136.0	rf=298.257
GRS80	GRS 1980(IUGG, 1980)	a=6378137.0	rf=298.257222101
IAU76	IAU 1976	a=6378140.0	rf=298.257
airy	Airy 1830	a=6377563.396	b=6356256.910
APL4.9	Appl. Physics. 1965	a=6378137.0.	rf=298.25
NWL9D	Naval Weapons Lab., 1965	a=6378145.0.	rf=298.25
mod_airy	Modified Airy	a=6377340.189	b=6356034.446
andrae	Andrae 1876 (Den., Iclnd.)	a=6377104.43	rf=300.0
aust_SA	Australian Natl & S. Amer. 1969	a=6378160.0	rf=298.25
GRS67	GRS 67(IUGG 1967)	a=6378160.0	rf=298.2471674270
bessel	Bessel 1841	a=6377397.155	rf=299.1528128
bess_nam	Bessel 1841 (Namibia)	a=6377483.865	rf=299.1528128
clrk66	Clarke 1866	a=6378206.4	b=6356583.8
clrk80	Clarke 1880 mod.	a=6378249.145	rf=293.4663
CPM	Comm. des Poids et Mesures 1799	a=6375738.7	rf=334.29
delmbr	Delambre 1810 (Belgium)	a=6376428.	rf=311.5
engelis	Engelis 1985	a=6378136.05	rf=298.2566
evrst30	Everest 1830	a=6377276.345	rf=300.8017
evrst48	Everest 1948	a=6377304.063	rf=300.8017
evrst56	Everest 1956	a=6377301.243	rf=300.8017
evrst69	Everest 1969	a=6377295.664	rf=300.8017
evrstSS	Everest (Sabah & Sarawak)	a=6377298.556	rf=300.8017
fschr60	Fischer (Mercury Datum) 1960	a=6378166.	rf=298.3
fschr60m	Modified Fischer 1960	a=6378155.	rf=298.3
fschr68	Fischer 1968	a=6378150.	rf=298.3
helmert	Helmert 1906	a=6378200.	rf=298.3

Acronym	Name	Radius	Parameter2
hough	Hough	a=6378270.0	rf=297.
intl	International 1909 (Hayford)	a=6378388.0	rf=297.
krass	Krassovsky, 1942	a=6378245.0	rf=298.3
kaula	Kaula 1961	a=6378163.	rf=298.24
lerch	Lerch 1979	a=6378139.	rf=298.257
mprts	Maupertius 1738	a=6397300.	rf=191.
new_intl	New International 1967	a=6378157.5	b=6356772.2
plessis	Plessis 1817 (France)	a=6376523.	b=6355863.
SEasia	Southeast Asia	a=6378155.0	b=6356773.3205
walbeck	Walbeck	a=6376896.0	b=6355834.8467
WGS60	WGS 60	a=6378165.0	rf=298.3
WGS66	WGS 66	a=6378145.0	rf=298.25
WGS72	WGS 72	a=6378135.0	rf=298.26
WGS84	WGS 84	a=6378137.0	rf=298.257223563
sphere	Normal Sphere (r=6370997)	a=6370997.0	b=6370997.0



## APPENDIX E: ISMS INSTALLATION

This document outlines the installation process for ISMS on Mandrake 9.1. You will need to have root access to the server to carry out certain of the steps.

### E.1 INSTALLING DEPENDENT SOFTWARE

Use the 'urpmi' tool on the command line to install all the requisite software packages. Use the following session log as a guide:

#### E.1.1 mySQL installation

```

\end{scriptsize}
[root@pc2-rdng3-4-cust67 root]# urpmi mysql
One of the following packages is needed:
 1- MySQL-4.0.12-1mdk.i586
 2- MySQL-Max-4.0.12-1mdk.i586
What is your choice? (1-2) 1
To satisfy dependencies, the following packages are going to be installed (14 MB):
MySQL-4.0.12-1mdk.i586
MySQL-client-4.0.12-1mdk.i586
MySQL-common-4.0.12-1mdk.i586
libmysql12-4.0.12-1mdk.i586
Is this OK? (Y/n) y

Preparing...
 1:libmysql12
 2:MySQL-client
 3:MySQL-common
 4:MySQL

```

#### E.1.2 PostgreSQL installation

```

[root@pc2-rdng3-4-cust67 root]# urpmi postgresql
installing /var/cache/urpmi/rpms/postgresql-7.3.2-5mdk.i586.rpm

```

```

Preparing...
 1:postgresql
*****

[root@pc2-rdng3-4-cust67 root]# urpmi postgresql-server

installing /var/cache/urpmi/rpms/postgresql-server-7.3.2-5mdk.i586.rpm

Preparing...
 1:postgresql-server
*****

```

### E.1.3 LaTeX installation

LaTeX is used for generating reports as pdf documents.

```

[root@pc2-rdng3-4-cust67 root]# urpmi tetex
installing /var/cache/urpmi/rpms/tetex-2.0.1-6mdk.i586.rpm

Preparing...
 1:tetex
*****

[root@pc2-rdng3-4-cust67 root]# urpmi tetex-latex
installing /var/cache/urpmi/rpms/tetex-latex-2.0.1-6mdk.i586.rpm

Preparing...
 1:tetex-latex
*****

```

### E.1.4 Installation of Proj4

```

[root@pc2-rdng3-4-cust67 root]# urpmi proj
installing /var/cache/urpmi/rpms/proj-4.4.5-2mdk.i586.rpm

Preparing...
 1:proj
*****

```

### E.1.5 Installation of PHP with mySQL drivers

The PHP web application language is used to implement all the core ISMS logic. It is installed as an apache module. Additional modules for PHP database connectivity need to be installed to facilitate communication with the MySQL, PostgreSQL and Sybase TDS (Microsoft SQL Server) backends. Setup up of TDS connectivity is not described in this document.

```

[root@pc2-rdng3-4-cust67 root]# urpmi apache2-mod_php
To satisfy dependencies, the following packages are going to be installed (2 MB):
apache2-2.0.45-5mdk.i586

```

```

apache2-common-2.0.45-5mdk.i586
apache2-mod_php-2.0.45_4.3.1-1mdk.i586
apache2-mod_ssl-2.0.45-5mdk.i586
apache2-modules-2.0.45-5mdk.i586
libapr0-2.0.45-5mdk.i586

```

```

installing /var/cache/urpmi/rpms/apache2-modules-2.0.45-5mdk.i586.rpm
/var/cache/urpmi/rpms/apache2-common-2.0.45-5mdk.i586.rpm
/var/cache/urpmi/rpms/apache2-mod_ssl-2.0.45-5mdk.i586.rpm
/var/cache/urpmi/rpms/apache2-2.0.45-5mdk.i586.rpm
/var/cache/urpmi/rpms/apache2-mod_php-2.0.45_4.3.1-1mdk.i586.rpm
/var/cache/urpmi/rpms/libapr0-2.0.45-5mdk.i586.rpm

```

```

Preparing...
 1:apache2-modules
 2:libapr0
 3:apache2-common
 4:apache2-mod_ssl
 5:apache2
 6:apache2-mod_php

```

```
[root@pc2-rdng3-4-cust67 root]# urpmi php-mysql php-pgsql
```

Now the MySQL and PostgreSQL modules for php are installed:

```

/var/cache/urpmi/rpms/php-mysql-4.3.1-1mdk.i586.rpm
/var/cache/urpmi/rpms/php-pgsql-4.3.1-1mdk.i586.rpm
Do you want to continue installation ? (y/N) y
installing /var/cache/urpmi/rpms/php-mysql-4.3.1-1mdk.i586.rpm
/var/cache/urpmi/rpms/php-pgsql-4.3.1-1mdk.i586.rpm

```

```

Preparing...
 1:php-mysql
 2:php-pgsql

```

By default (on Mandrake 9.1), no php.ini file is installed. To install this required file, do the following:

```
[root@pc2-rdng3-4-cust67 conf]# urpmi php-ini
```

```
installing /var/cache/urpmi/rpms/php-ini-4.3.1-1mdk.noarch.rpm
```

```

Preparing...
 1:php-ini

```

When models are executed, they are run using the CGI (Common Gateway Interface) Binary version of php. To install this file, do the following:

```
[root@pc2-rdng3-4-cust67 root]# urpmi php-cgi
To satisfy dependencies, the following packages are going to be installed (1 MB):
libphp_common430-430-13mdk.i586
php-cgi-4.3.1-13mdk.i586
Is this OK? (Y/n) y

installing /var/cache/urpmi/rpms/php-cgi-4.3.1-13mdk.i586.rpm
/var/cache/urpmi/rpms/libphp_common430-430-13mdk.i586.rpm

Preparing...
 1:libphp_common430
 2:php-cgi
```

Note that if your distribution does not install the PHP cgi binary to `/usr/bin/php-cgi`, you should make a symlink to this name, or modify the first line of the `ISMS/bin/runGARP.php` file with an appropriate path.

## E.1.6 Optional installation of the Grass GIS

```
[root@pc2-rdng3-4-cust67 root]# urpmi grass
To satisfy dependencies, the following packages are going to be installed (108 MB):
expect-8.4.2-1mdk.i586
grass-5.0.2-1mdk.i586
libfftw2-2.1.5-1mdk.i586
libgdal0-1.1.8-2mdk.i586
libgrass5_0-1.0.0-1mdk.i586
libproj0-4.4.5-2mdk.i586
libpython2.2-2.2.2-7mdk.i586
python-2.2.2-7mdk.i586
python-base-2.2.2-7mdk.i586
tcl-8.4.2-1mdk.i586
tk-8.4.2-1mdk.i586
tkinter-2.2.2-7mdk.i586
Is this OK? (Y/n) y
installing /var/cache/urpmi/rpms/libfftw2-2.1.5-1mdk.i586.rpm
/var/cache/urpmi/rpms/tcl-8.4.2-1mdk.i586.rpm
/var/cache/urpmi/rpms/libpython2.2-2.2.2-7mdk.i586.rpm
/var/cache/urpmi/rpms/grass-5.0.2-1mdk.i586.rpm
/var/cache/urpmi/rpms/libgrass5_0-1.0.0-1mdk.i586.rpm
/var/cache/urpmi/rpms/libproj0-4.4.5-2mdk.i586.rpm
/var/cache/urpmi/rpms/libgdal0-1.1.8-2mdk.i586.rpm
/var/cache/urpmi/rpms/tkinter-2.2.2-7mdk.i586.rpm
/var/cache/urpmi/rpms/tk-8.4.2-1mdk.i586.rpm
/var/cache/urpmi/rpms/expect-8.4.2-1mdk.i586.rpm
/var/cache/urpmi/rpms/python-2.2.2-7mdk.i586.rpm
/var/cache/urpmi/rpms/python-base-2.2.2-7mdk.i586.rpm
```



```

Preparing...
 1:libfftw2
 2:tcl
 3:libpython2.2
 4:libgrass5_0
 5:libproj0
 6:libgdal0
 7:tk
 8:grass
 9:expect
10:python-base
11:python
12:tkinter

```

## E.1.7 Start system services

Ensure apache, MySQL and PostgreSQL are running:

```

[root@pc2-rdng3-4-cust67 root]# /etc/init.d/httpd restart
Starting httpd2: [ OK ]
[root@pc2-rdng3-4-cust67 root]# /etc/init.d/mysql start
Starting MySQL Server [ OK ]
[root@pc2-rdng3-4-cust67 root]# /etc/init.d/postgresql start
Starting postgresql service: [ OK ]

```

## E.2 POSTGRESQL CONFIGURATION

### E.2.1 Create a non-privileged user

```

[root@pc2-rdng3-4-cust67 root]# su - postgres
-bash-2.05b\$ /usr/bin/createuser apache
Shall the new user be allowed to create databases? (y/n) n
Shall the new user be allowed to create more new users? (y/n) n
CREATE USER

```

### E.2.2 Create a privileged user

```

-bash-2.05b\$ createuser -a -d -P -E aps02ts
Enter password for user "aps02ts":
Enter it again:
-bash-2.05b\$ exit

```

```
logout
```

### E.2.3 Allow TCP/IP access to PostgreSQL

Edit `/var/lib/pgsql/data/postgresql.conf` and uncomment the `tcpip_socket` parameter and set it to true:

```
#
#         Connection Parameters
#
tcpip_socket = true
#ssl = false

#max_connections = 32
#superuser_reserved_connections = 2
```

Edit the `/var/lib/pgsql/data/pg_hba.conf` and add a line for localhost connections using md5 encrypted password connections:

```
# TYPE DATABASE USER IP-ADDRESS IP-MASK METHOD
#local all all trust
#host all all 127.0.0.1 255.255.255.255 trust
host all all 127.0.0.1 255.255.255.255 md5
```

### E.2.4 Install sample specimen database

```
\end{scriptsize}
aps02ts@pc2-rdng3-4-cust67:~/public_html/$ psql ISMS_data -f ~/ISMS/ISMS_data.sql
```

```
psql:/home/aps02ts/ISMS/ISMS_data.sql:5: ERROR:
CREATE DATABASE: database "ISMS_data" already exists
```

```
You are now connected as new user aps02ts.
```

```
psql:/home/aps02ts/ISMS/ISMS_data.sql:33: NOTICE:
CREATE TABLE / PRIMARY KEY will create implicit index 'temptaxon_pkey'
for table 'temptaxon'
```

```
CREATE TABLE
CREATE SEQUENCE
```

```
psql:/home/aps02ts/ISMS/ISMS_data.sql:54: NOTICE:
CREATE TABLE / PRIMARY KEY will create implicit index 'tbldateprecision_pkey'
```

```

for table 'tbldateprecision'
CREATE TABLE
CREATE SEQUENCE
psql:/home/aps02ts/ISMS/ISMS_data.sql:75: NOTICE:
CREATE TABLE / PRIMARY KEY will create implicit index 'tblidprecision_pkey'
for table 'tblidprecision'
CREATE TABLE
CREATE SEQUENCE
psql:/home/aps02ts/ISMS/ISMS_data.sql:96: NOTICE: CREATE TABLE / PRIMARY KEY
will create implicit index 'tbllocalityprecision_pkey' for table 'tbllocalityprecision'
CREATE TABLE
CREATE SEQUENCE
psql:/home/aps02ts/ISMS/ISMS_data.sql:140: NOTICE:
CREATE TABLE / PRIMARY KEY will create implicit index 'tblspecimen_pkey' for
table 'tblspecimen'
CREATE TABLE
CREATE VIEW
CREATE VIEW
CREATE INDEX
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
CREATE TRIGGER
setval
-----
      1
(1 row)

```

## E.3 MYSQL CONFIGURATION

### E.3.1 Creating a mySQL user

```

GRANT ALL PRIVILEGES ON * . *
TO "isms"@"localhost"
IDENTIFIED BY "*****"
WITH GRANT OPTION
MAX_QUERIES_PER_HOUR 0
MAX_CONNECTIONS_PER_HOUR 0
MAX_UPDATES_PER_HOUR 0 ;

```

### E.3.2 Loading the ISMS database

```
mysql < /tmp/ISMS.sql
```

## E.4 ISMS APPLICATION DEPLOYMENT

### E.4.1 Allowing Apache access to the 'AT' command scheduler

ISMS uses the at command scheduler to batch process model runs to prevent browser timeouts. Many distributions (including Mandrake 9.1) do not allow all users access to at by default. Because of this, the http daemon needs to be added to the `/etc/at.allow` configuration file:

```
[root@pc2-rdng3-4-cust67 etc]# echo "apache" >> /etc/at.allow
```

Be sure to substitute 'apache' with the user name under which your http process runs. Once the ISMS tarball has been extracted to your web root directory, you should ensure that permissions are granted recursively to the apache process. On Mandrake, this process is run under the user account 'apache', so the following command should be issued:

```
chown -R apache.apache /home/aps02ts/public_html/
```



## APPENDIX F: INDEPENDENT VARIABLE PREPARATION - A DETAILED ACCOUNT

The ISMS application has an administration module which allows the administrator to create projection definitions (e.g. Transverse Mercator, lon 19), extent definitions (e.g. Western Cape, cell size 500m, projection Transverse Mercator, minimum and maximum x and y coordinates). Additionally the administration module allows independent layers to be uploaded to the ISMS/GARP server and associated with an extent. The following section describes how to prepare a typical series of independent variable layers using the ArcView GIS application.

The preparation of the independent variable layers in ArcView requires the ArcView Spatial Analyst extension in order to output the results in ASCII grid format. It is also possible to use ArcInfo for this purpose, although the procedure using ArcView is described here. You will need to have your independent data layers available in either ESRI shapefile format or as Spatial Analyst grid files.

To start the process, use the 'View: Properties' dialogue to set the projection for your view. All the data layers output will then adopt this projection.

I recommend making a mask file that will be used to define the analysis environment and to screen out any unwanted data during the data preparation process. This mask layer can also be used during the GARP modelling process if your study area is a non-rectangular shape. One simple approach to creating a mask is to take a continuous vector coverage that has been clipped to your study area (e.g. vegetation or geology), use the 'Theme: Convert to Grid' menu option and save this to a temporary file. Be sure to specify the cell size that you would like to use for the final GARP analysis. Also ensure that the 'Output Grid Extent' is set to 'same as [your clipped layer name]'. The resulting grid will likely have numerous grid values assigned. Use the 'Analysis: Reclassify' menu to reassign all non-'no data' values to 1. Assigning 1 to non-masked areas will allow the mask layer to be used with another layer to carry out grid arithmetic without altering the numerical Z values of the other layer. As will be seen later, a GARP mask layer requires a value of 255 for areas to be masked out, so the mask layer will need to be adjusted before exporting to ISMS/GARP.

Once the Spatial Analyst mask grid has been reclassified (1=non-masked, no data=masked), this mask layer should be used as a basis for setting the Spatial Analyst 'Analysis Environment'. This will ensure that all future grids created in this view have identical geographical extents, cell sizes and are masked by the mask layer. First save the mask layer in a new directory (created using windows explorer). Give the mask grid a meaningful filename (e.g. 'mask'). Next, choose 'Analysis: Properties' and set the 'Analysis Extent', 'Analysis Cell Size' and 'Analysis Mask' all to the name of the mask layer you have created. Make a note of the left, right, top, bottom, number of rows and columns and cell size parameters shown on this dialogue.

Next you should perform the grid generation process for each independent variable layer. This may consist of simple data conversion using the 'Theme: Convert to Grid' function, or more complex processes such as deriving a DEM, slope steepness and slope aspect layers from a vector layer containing contours. Ensure that a consistent naming convention is used.

A typical dataset may consist of vegetation, geology, slope steepness, aspect, terrain height, rainfall data, temperature data (both of the aforementioned interpolated from point source weather station data) and other correlates such as latitude, longitude, etc. These datasets can be categorised into two kinds:

1. categorical data (e.g. geology) stored as an integer ranging from 0-255;
2. continuous data (e.g. temperature) stored as a real number.

Categorical data with a high number of discrete classes should be avoided or compensated for to cater for this 8-bit limitation.

## F.1 PREPARING THE MASK LAYER

The layer to be used to mask dependent variable observations outside of the study area requires special treatment. The process begins in ArcView/ArcInfo where the mask layer that has been used to 'clip' independent variable layers is reclassified so that all cells inside the area of interest are coded with a value of 255 and all cells beyond the area of interest are coded as null (the latter should already be the case). Now the File: Export dataset option in ArcView is used to write the data to a space delimited ascii export grid.

Following this, the mask layer should be copied to the directory on the GNU/Linux/Unix box that contains the GARP binaries (using SCP, SAMBA or FTP).

```
cat maskfile.asc | ./rasteriz -layer 2
```

This will provide output similar to the listing supplied below. This information can be used to assemble a 'paramete' file before carrying out the actual rasterisation.

### F.1.1 Obtaining the file properties of an ASCII grid file

The GARP rasteriz application can be used to interrogate an ASCII Grid file for its properties.

```
cat maskfile.asc | ./rasteriz -layer 2
RASTERIZE - point coverage to gray layer
Columns = 241 Rows = 207
Raster cell size is 7128.3697x7128.3697
Writing to layer00
Missing val -9999
Min is 255.000000, max is 255.000000
Columns -795282.250000 922654.830566 7128.369629
Rows -3846910.750000 -2371338.236816 7128.369629
No of Cols 241 No of Rows 207
```

### F.1.2 A sample 'paramete' listing for the mask file

```
Columns -795282.26310 922654.82830 7128.36970
Rows -3846910.63440 -2371338.11190 7128.36970
Variable 0 Taxon 255 255 c mask %2.0f
```

### F.1.3 Carrying out the rasterisation process

```
cat maskfile.asc | ./rasteriz -layer 3
RASTERIZE - point coverage to gray layer
Columns = 241 Rows = 207
Raster cell size is 7128.3697x7128.3697
Writing to layer00
Processing 207 rows and 241 columns
```

No of data points written 49887

The resulting file will be called 'layer00' - as with any other file that is output from the rasteriz routine. The output file should be renamed to 'mask'. This is a special purpose name - the GARP application will look for a file by this name if it encounters the parameter:

Maskvalues 255

during the parameter file parsing process. Finally you should verify the output is as expected by using the ImageMagick 'display' command. For the above file the syntax would be as follows:

```
display -size 241x207 gray:mask
```

## F.2 RASTERISING CATEGORICAL DATA

A similar process is used to import categorical data into the GARP binary file format. First file property data are obtained:

```
cat sageol_aea.asc | ./rasteriz -layer 2
```

```
RASTERIZE - point coverage to gray layer
Columns = 1318 Rows = 1019
Raster cell size is 500.0000x500.0000
Writing to layer00
Missing val -9999
Min is 0.000000, max is 999.000000
Columns -795282.250000 922654.830566 7128.369629
Rows -3846910.750000 -2371338.236816 7128.369629
No of Cols 241 No of Rows 207
```

Next the 'paramete' file to be used for the import process is defined:

```
Columns -795282.26310 922654.82830 7128.36970
Rows -3846910.63440 -2371338.11190 7128.36970
Variable 0 Geol 0 999 c geol_id %2.0f
```



'c' in this case indicates the data is category data, and 0 999 indicates the range of possible values for this dataset. Following the parameter file preparation, the dataset can be rasterised:

```
cat sageol_aea.asc | ./rasteriz -layer 3
```

```
RASTERIZE - point coverage to gray layer  
Columns = 241 Rows = 207  
Raster cell size is 7128.3697x7128.3697  
Writing to layer00  
Processing 207 rows and 241 columns  
No of data points written 49887
```

Finally, it is possible to verify the process completed successfully using the ImageMagick 'display' command:

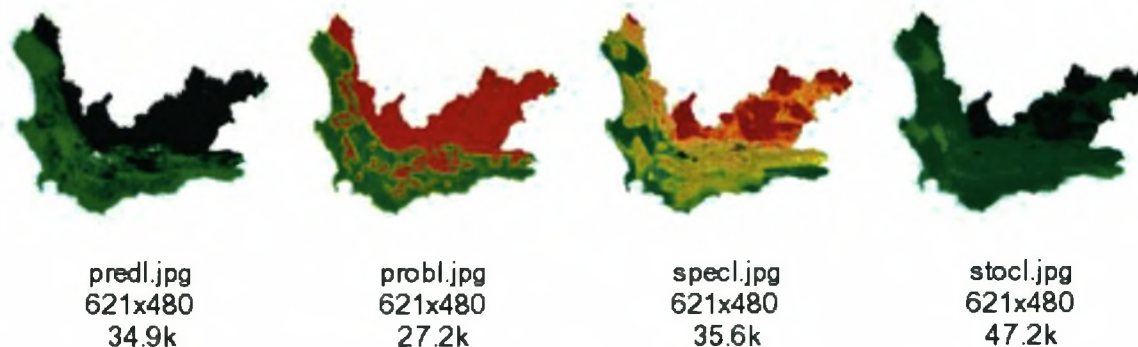
```
display -size 241x207 gray:layer00
```

## **APPENDIX G: SAMPLE OUTPUT FROM AN ISMS MODEL ITERATION**

Tim Sutton ISMS Model Iteration Report (76) for *Xenopus laevis laevis*

This is an automatically generated report produced by the Integrated Species Modelling System (ISMS). ISMS is a user-friendly front-end to the GARP modelling system. ISMS is written by Tim Sutton. GARP is written by Dr David Stockwell.

## G.1 MODEL ENVIRONMENT



Taxonomy		
	Target Taxon:	Xenopus laevis laevis
	Surrogate Absence Taxon:	
Model Duration		
	Start Time:	2003-04-25 11:26:19
	End Time:	2003-04-25 11:27:37
Data Precision		
	Taxonomic ID Precision:	Species
	Temporal Precision:	Unknown
	Spatial Precision:	Nearest Minute

### G.1.1 Extents of the modelled area

Name	Min	Max	Cell Size
Columns	-330101.00000	328899.00000	500.00000
Rows	-4061519.00000	-3552019.00000	500.00000

### G.1.2 Parameters used

Parameter	Value	Description
Atomicflag	1	Check this to use atomic rules.
Bioclimflag	1	Check this to use bioclimatic envelope rules.
Garpflag	1	Check this to use rules produced by the GARP modelling system.
Maskvalues	255	Enter the cell value for areas to be masked.

## G.1.3 Layers used

ID	Name	Min Val	Max Val	Type	Unit	Frmt	Description
1	veglr	1.00	23.00	c	catno	c	Low and Robelo Vegetation Types
2	apan	685.50	2822.9	r	mm	r	Apan Evaporation Index
3	wcdem	0.00	2247.0	r	m	r	Digital Elevation Model
4	slopes	0.00	67.35	r	perc	r	Slope steepness
5	geology	1.00	12.00	c	catno	c	Dominant geology
6	seadist	0.00	268708	r	m	r	Distance from the sea
7	catch	1.00	485.00	c	catno	c	Quaternary catchment number
8	solrad	13.84	24.19	r	perc	r	Solar radiation
9	avgrain	6666.1	3264.5	r	mm	r	Average rainfall
10	avgmxt	15.23	28.70	r	degc	r	Average annual maximum temperature
11	aspect	1.00	359.95	r	deg	r	Slope aspect
12	latitude	2500.0	488000	r	m	r	Relative latitude
13	longitude	9999.0	630500	r	m	r	Relative longitude
14	avgmint	3.18	13.56	r	degc	r	Average annual minimum temperature
15	rivdist	0.00	41067.	r	m	r	Distance from 1:500 000 rivers
16	mintapr	2.99	15.01	r	degc	r	Average minimum temperature for April
17	mintaug	-1.13	10.07	r	degc	r	Average minimum temperature for August
18	mintdec	7.67	16.54	r	degc	r	Average minimum temperature for December
19	mintfeb	9.17	18.25	r	degc	r	Average minimum temperature for February
20	mintjan	7.66	17.87	r	degc	r	Average minimum temperature for January
21	mintjun	-3.79	10.51	r	degc	r	Average minimum temperature for June
22	mintjul	-3.29	11.04	r	degc	r	Average minimum temperature for July



23	mintmar	7.18	16.60	r	degc	r	Average minimum temperature for March
24	mintmay	-0.48	13.09	r	degc	r	Average minimum temperature for May
25	mintnov	5.77	14.24	r	degc	r	Average minimum temperature for November
26	mintoct	2.16	12.37	r	degc	r	Average minimum temperature for October
27	mintsept	0.13	11.14	r	degc	r	Average minimum temperature for September
28	maxtapr	15.15	29.29	r	degc	r	Average maximum temperature for April
29	maxtaug	11.19	22.10	r	degc	r	Average maximum temperature for August
30	maxtdec	18.44	34.94	r	degc	r	Average maximum temperature for December
31	maxtfeb	18.00	37.39	r	degc	r	Average maximum temperature for February
32	maxtjan	18.77	36.29	r	degc	r	Average maximum temperature for January
33	maxtjul	8.22	20.56	r	degc	r	Average maximum temperature for July
34	maxtmar	18.94	34.09	r	degc	r	Average maximum temperature for March
35	maxtjun	7.82	21.80	r	degc	r	Average maximum temperature for June
36	maxtmay	11.79	25.30	r	degc	r	Average maximum temperature for May
37	maxtnov	15.47	32.59	r	degc	r	Average maximum temperature for November
38	maxtoct	13.76	28.89	r	degc	r	Average maximum temperature for October
39	maxtsept	12.60	24.89	r	degc	r	Average maximum temperature for September

## G.2 MODEL RESULTS

### G.2.1 Cartographic outputs



Figure G.1: predI (Predicted presence / absence)

'In the case of predictions the byte has the value 253 [shown in green] when present and 1 [black in image] when absent. The value zero is an unpredicted area where no rule applies and 254 is a masked area, outside the area of interest. The uncertainty is similarly expressed, with the probability value scaled between 1 and 253.' (Payne & Stockwell 1999)

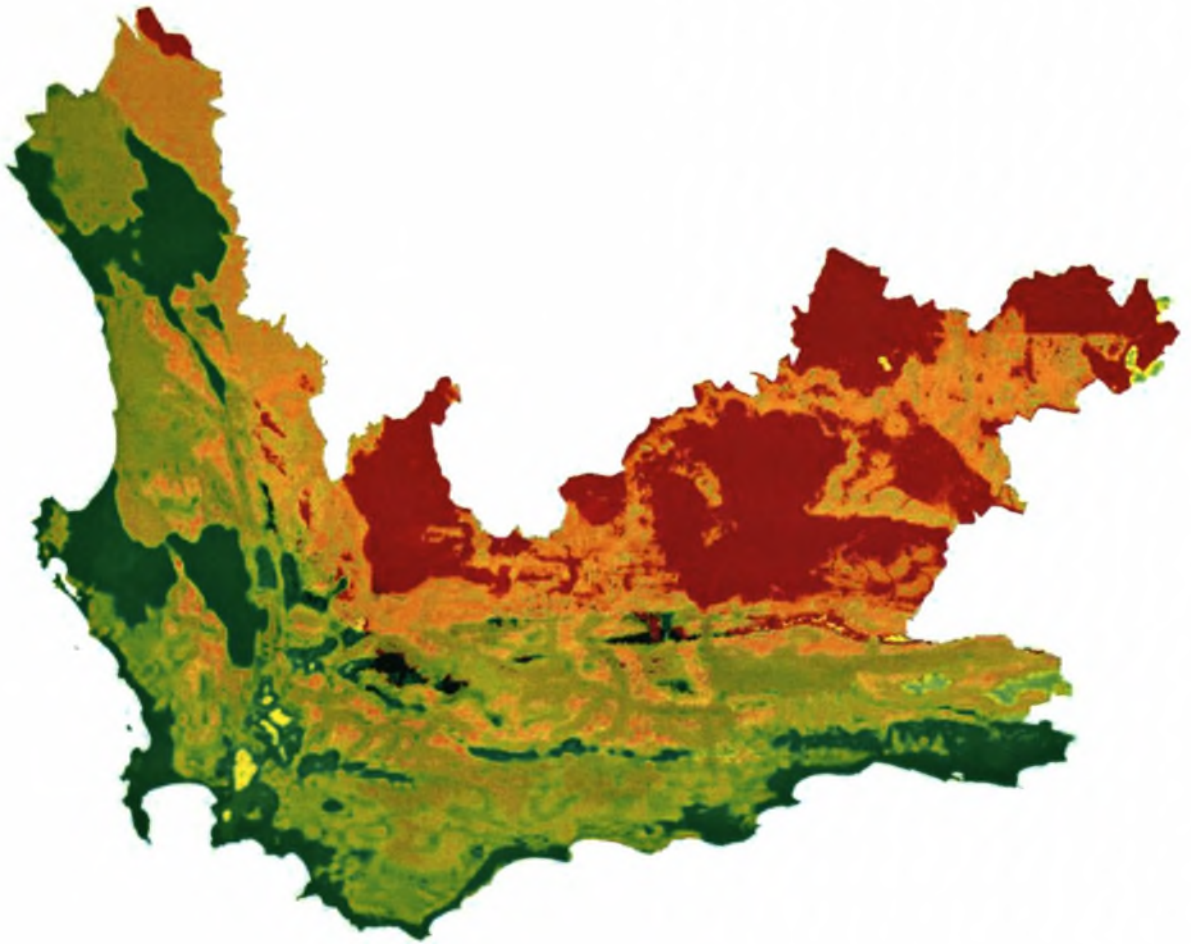


Figure G.2: specI (Rule specialisation)

'In outputting areas where rules conflict, a colour visualisation using green for predicted presences and red for absences, and a chromatic combination of the two (brown to yellow) where rules contradict each other (predict presence and absence). The layer, specI, is produced by combining two grey scale layers. The red layer is the maximum posterior probability of absence and the green layer the maximum posterior probability of presence. [...] Referring to the legend, pure green is produced when a prediction of presence is made with no rules predicting absence. Pure red areas are produced when predictions of absence are made with no rules predicting presence. Brown and yellow areas are called conflict areas: yellow where rules have predicted both presence and absence with a high probability, while brown areas show conflict with low intensities.' (Payne & Stockwell 1999)





Figure G.3: probI (Probability of occurrence)

'A probability surface for the prediction is used in outputting probability of occurrence. The layer probI represents the probability of occurrence scaled into the range of a single byte. The probability of occurrence can be derived from the most accurate rules that applies at a site by the rules:

- if predicted value is present  $P =$  posterior probability [mapped as intensity of green],

- if predicted value is absent  $P = 1 -$  posterior probability [mapped as red].'  
(Payne & Stockwell 1999)





Figure G.4: stocI - This is an experimental undocumented output from the GMS.

'In practice this method was found to give unrealistic sharp transitions between areas of high predicted probability of occurrence and areas of low predicted probability of occurrence. A method that yield a smoother probability surface is to apply rules stochastically according to their probability. When no rule applies the area is unpredicted.' (Payne & Stockwell 1999)

## G.2.2 Initial outputs

INITIAL - producing initial model (prelim) from training set (train)

## G.2.3 Presample outputs

PRESAMPLE - from data to train and test sets

Warning: data point is masked

Warning: data point is masked

Warning: data point is masked

Warning: data point is masked

Warning: data point is masked

Warning: data point is masked

Warning: data point is masked

518121 points with pred 0

76 points with pred 254

presences 76 absences 0 background 518121 no\_of\_data 518197

2500 points to train and 2500 points to test

## G.2.4 Verify outputs

VERIFY - predictive accuracy of rules using training data

Cut value = 0.000000 0

Results for all rules

	Actual		
	Present	Absent	Background
Predict Present	1010	0	240
Absent	0	0	0
Background	545	0	694

Conflicts 0 0

Unpredicted 0 0

NoMasked	0.000
Unpredicted	0.004
Predicted	0.996
Accuracy	0.685 s.d. 0.009
Overall Acc.	0.686 s.d. 0.009

### G.2.5 Explain outputs

EXPLAIN - producing model (model) from training set (train)  
Generations: 38

### G.2.6 Translate outputs

Key to rules:  
r = logistic rule  
m = bioclim rule  
d = range rule  
a = atomic rule  
! = negated rule

VERIFY - predictive accuracy of rules using training data

Cut value = 0.000000 0

Results for all rules

Confusion Matrix:		Actual		
		Present	Absent	Background
Predict	Present	1010	0	240
	Absent	0	0	0
	Background	545	0	694

Conflicts 0 0

Unpredicted 0 0

NoMasked	0.000
Unpredicted	0.004
Predicted	0.996
Accuracy	0.685 s.d. 0.009
Overall Acc.	0.686 s.d. 0.009

TRANSLATE - model into natural language

No-rule number, Type, Prior-prior, Post-accuracy,

Sig-Significance, Cov-coverage, Use-usage

No	Type	Prior	Post	Sig	Cov	Use
----	------	-------	------	-----	-----	-----

4	d	0.51	0.70	11.49	0.38	0.299
---	---	------	------	-------	------	-------

IF apan=[1573,2502]mm

AND

wcdem=[ 0,1341]m

AND

catch=[ 5,202]catno

AND

solrad=[17.5218,22.4718]perc

AND

avgmint=[7.20071,11.79581]degc

AND

mintaug=[3.16407,8.25498]degc

AND

mintdec=[9.70344,14.54162]degc

AND

mintfeb=[11.14391,15.91719]degc

AND

mintjan=[9.71814,15.85221]degc

AND

mintjun=[1.35348,8.02304]degc

AND

mintjul=[2.09083,8.77439]degc

AND



```
mintnov=[8.64913,13.00130]degc
```

```
AND
```

```
mintsept=[3.08921,9.61688]degc
```

```
AND
```

```
maxtdec=[21.17913,30.44000]degc
```

```
AND
```

```
maxtmar=[23.85028,29.11984]degc
```

```
AND
```

```
maxtoct=[17.04913,22.61075]degc
```

```
THEN Taxon=PRESENT
```

```
0 r 0.52 0.85 16.57 0.25 0.200
```

```
IF + veglr*0.32 catno + apan*0.21 mm + wcdem*0.38 m
+ slopes*0.26 perc - seadist*0.14 m + catch*0.30 catno
- solrad*0.16 perc + avgrain*0.13 mm + aspect*0.30 deg
+ latitude*0.18 m + longitud*0.02 m - avgmint*0.36 degc
- rivdist*0.46 m - mintapr*0.41 degc - mintaug*0.29 degc
+ mintdec*0.14 degc + mintfeb*0.07 degc - mintjan*0.06 degc
- mintjun*0.07 degc - mintjul*0.21 degc - mintmay*0.23 degc
+ mintnov*0.30 degc - mintoct*0.46 degc - maxtaug*0.39 degc
- maxtdec*0.28 degc - maxtfeb*0.40 degc - maxtjan*0.31 degc
+ maxtjul*0.08 degc - maxtmar*0.10 degc - maxtjun*0.09 degc
- maxtmay*0.32 degc - maxtnov*0.26 degc - maxtoct*0.17 degc
```

```
THEN Taxon=BACKGROUND
```

```
5 r 0.49 0.62 10.76 0.70 0.109
```

```
IF + veglr*0.32 catno - apan*0.46 mm + slopes*0.26 perc
- geology*0.35 catno + seadist*0.01 m - catch*0.49 catno
+ avgrain*0.13 mm - avgmaxt*0.02 degc + aspect*0.30 deg
+ latitude*0.08 m + longitud*0.02 m - avgmint*0.36 degc
- rivdist*0.32 m - mintapr*0.41 degc - mintaug*0.20 degc
+ mintdec*0.14 degc + mintfeb*0.07 degc + mintjan*0.06 degc
- mintjun*0.07 degc + mintmar*0.27 degc - mintmay*0.07 degc
+ mintnov*0.30 degc - mintoct*0.46 degc - mintsept*0.28 degc
+ maxtapr*0.21 degc - maxtaug*0.39 degc - maxtdec*0.30 degc
- maxtfeb*0.40 degc - maxtjan*0.31 degc + maxtjul*0.08 degc
```

```
- maxtmar*0.15 degc - maxtmay*0.32 degc + maxtnov*0.04 degc
- maxtoct*0.17 degc
THEN    Taxon=PRESENT

1    d    0.51  0.96  13.15  0.09  0.073
IF  veglr=[ 2,23]catno
AND
apan=[1784,2561]mm
AND
wcdem=[107,1190]m

AND
slopes=[0.2662,26.6206]perc
AND
geology=[ 2,11]catno
AND
seadist=[48856,219852]m

AND
catch=[ 5,416]catno
AND
solrad=[18.6673,22.9627]perc
AND
avgrain=[6572,6101]mm

AND
avgmaxt=[20.55411,26.78332]degc
AND
aspect=[19,344]deg
AND
latitude=[63907,382457]m

AND
longitud=[39430,608427]m
AND
avgmint=[6.87249,11.46759]degc
AND
rivdist=[2759,11849]m
```

AND

mintapr=[7.78850,12.44447]degc

AND

mintaug=[2.19016,7.76802]degc

AND

mintdec=[10.89546,14.75198]degc

AND

mintfeb=[12.57949,16.27609]degc

AND

mintjan=[11.69557,15.97328]degc

AND

mintjun=[0.56217,7.62739]degc

AND

mintjul=[1.97755,8.37790]degc

AND

mintmar=[10.41929,15.22237]degc

AND

mintmay=[4.93727,9.97909]degc

AND

mintnov=[9.01739,13.06826]degc

AND

mintoct=[6.15522,10.87684]degc

AND

mintsept=[4.26419,8.65949]degc

AND

maxtapr=[21.57727,25.99253]degc

AND

maxtaug=[15.32976,20.67696]degc

AND

maxtdec=[23.98348,31.41826]degc

AND

maxtfeb=[25.81731,32.40838]degc

```
AND
maxtjan=[23.75593,34.00478]degc
AND
maxtjul=[14.21929,18.95043]degc

AND
maxtmar=[24.68862,29.83842]degc
AND
maxtjun=[12.68261,21.02640]degc
AND
maxtmay=[17.07652,23.16403]degc

AND
maxtnov=[22.77814,28.93593]degc
AND
maxtoct=[17.82656,28.41158]degc
AND
maxtsept=[18.67213,21.87822]degc
THEN    Taxon=BACKGROUND

9    d    0.49  0.64  8.39  0.31  0.056
IF    veglr=[ 5,17]catno
AND
geology=[ 3,12]catno
AND
avgmaxt=[21.03328,24.65368]degc

AND
latitude=[40879,313374]m
AND
longitud=[88481,574091]m
AND
rivdist=[487,11525]m

AND
mintaug=[2.45577,8.29925]degc
AND
mintjun=[1.63609,7.96652]degc
```



```
AND
mintmay=[3.81091,10.46182] degc

AND
mintoct=[4.78312,11.48217] degc
AND
mintsept=[2.91514,9.61688] degc
AND
maxtdec=[24.17913,29.13565] degc

AND
maxtjun=[15.44545,19.70024] degc
AND
maxtoct=[16.92953,26.97632] degc
THEN Taxon=PRESENT

6 m 0.51 0.63 9.02 0.56 0.054
IF veglr=[ 3,22] catno
AND
apan=[1556,2561]mm
AND
wcdem=[53,1430]m

AND
slopes=[0.2662,20.2316]perc
AND
geology=[ 3,12] catno
AND
seadist=[24428,219852]m

AND
catch=[ 5,405] catno
AND
solrad=[17.0309,22.5536]perc
AND
avgrain=[6585,5631]mm

AND
```

avgmaxt=[17.67909,26.09119]degc

AND

aspect=[ 1,361]deg

AND

latitude=[8257,315292]m

AND

longitud=[88481,608427]m

AND

avgmint=[5.96988,11.95992]degc

AND

rivdist=[ 0,24997]m

AND

mintapr=[6.22067,12.72953]degc

AND

mintaug=[1.79174,8.43206]degc

AND

mintdec=[9.52814,14.54162]degc

AND

mintfeb=[10.96447,15.91719]degc

AND

mintjan=[9.51636,15.85221]degc

AND

mintjun=[0.33609,8.24913]degc

AND

mintjul=[1.07130,9.17087]degc

AND

mintmar=[10.04696,14.88727]degc

AND

mintmay=[3.54273,10.73000]degc

AND

mintnov=[7.07565,13.06826]degc

AND

mintoct=[4.78312,11.48217]degc

```
AND
mintsept=[2.91514,9.61688]degc

AND
maxtapr=[18.05625,27.27798]degc
AND
maxtaug=[14.81229,20.50447]degc
AND
maxtdec=[20.85304,30.44000]degc

AND
maxtfeb=[21.75537,32.86822]degc
AND
maxtjan=[21.95546,31.51182]degc
AND
maxtjul=[12.75605,19.24308]degc

AND
maxtmar=[21.51490,30.91628]degc
AND
maxtjun=[12.40632,20.08703]degc
AND
maxtmay=[16.54253,23.43103]degc

AND
maxtnov=[18.65040,29.47727]degc
AND
maxtoct=[16.92953,26.31850]degc
AND
maxtsept=[15.80609,22.50972]degc
THEN    Taxon=PRESENT

28  d    0.49  0.77  5.14  0.03  0.038
IF  apan=[1573,2705]mm
AND
wcdem=[ 0,1892]m
AND
slopes=[-0.2662,0.0000]perc
```

AND

geology=[ 1,12]catno

AND

solrad=[18.6264,22.9627]perc

AND

avgrain=[6478,6115]mm

AND

longitud=[73766,537302]m

AND

mintapr=[5.98312,12.96708]degc

AND

mintaug=[2.63285,7.67949]degc

AND

mintdec=[10.72016,14.85715]degc

AND

mintjan=[12.42198,15.56972]degc

AND

mintjul=[1.12794,8.66111]degc

AND

mintmar=[10.82885,14.99897]degc

AND

mintoct=[5.50953,10.99790]degc

AND

mintsept=[4.04660,8.74652]degc

AND

maxtapr=[21.57727,26.27198]degc

AND

maxtfeb=[25.66403,32.63830]degc

AND

maxtnov=[21.22178,31.10131]degc

AND

maxtoct=[19.97945,26.31850]degc



THEN Taxon=BACKGROUND

7 r 0.48 1.00 8.76 0.03 0.027

IF + veglr\*0.32 catno - apan\*0.46 mm + wcdem\*0.38 m  
 + slopes\*0.26 perc - geology\*0.35 catno + seadist\*0.01 m  
 + catch\*0.30 catno - solrad\*0.16 perc + avgrain\*0.13 mm  
 - avgmxt\*0.43 degc + aspect\*0.30 deg + latitude\*0.08 m  
 + longitud\*0.09 m - avgmint\*0.36 degc - rivdist\*0.32 m  
 - mintapr\*0.41 degc - mintaug\*0.20 degc + mintdec\*0.14 degc  
 + mintfeb\*0.07 degc + mintjan\*0.06 degc - mintjun\*0.07 degc  
 + mintjul\*0.03 degc + mintmar\*0.33 degc - mintmay\*0.22 degc  
 + mintnov\*0.30 degc - mintoct\*0.46 degc - mintsept\*0.28 degc  
 + maxtapr\*0.21 degc - maxtaug\*0.39 degc - maxtdec\*0.28 degc  
 - maxtfeb\*0.40 degc - maxtjan\*0.31 degc + maxtjul\*0.08 degc  
 - maxtmar\*0.15 degc + maxtjun\*0.21 degc - maxtmay\*0.32 degc  
 - maxtnov\*0.26 degc - maxtoct\*0.17 degc

THEN Taxon=PRESENT

15 d 0.48 0.68 6.86 0.12 0.024

IF veglr=[ 5,16]catno

AND

wcdem=[44,1430]m

AND

slopes=[0.0000,15.4399]perc

AND

geology=[ 2,11]catno

AND

seadist=[76470,161438]m

AND

catch=[ 3,445]catno

AND

solrad=[17.6445,23.0036]perc

AND

avgrain=[6653,5362]mm

AND

latitude=[69664,372862]m

AND

longitud=[63956,554470]m

AND

avgmint=[6.25708,11.63170]degc

AND

mintapr=[7.69348,12.49198]degc

AND

mintaug=[1.39332,7.85656]degc

AND

mintdec=[11.07075,14.68186]degc

AND

mintfeb=[12.54360,16.27609]degc

AND

mintjan=[12.30091,15.65044]degc

AND

mintjun=[-0.05957,8.19261]degc

AND

mintjul=[2.54395,8.15134]degc

AND

mintmay=[4.34727,10.14000]degc

AND

mintnov=[9.35217,12.93435]degc

AND

mintoct=[6.47806,11.11897]degc

AND

mintsept=[1.00036,10.05206]degc

AND

maxtapr=[21.52138,26.32787]degc

AND

maxtaug=[15.58850,20.59071]degc

AND

maxtfeb=[25.81731,32.56166]degc

```
AND
maxtjan=[25.55640,32.34281]degc
AND
maxtjul=[13.24379,19.53573]degc

AND
maxtmar=[24.56885,29.95818]degc
AND
maxtjun=[14.50609,19.70024]degc
AND
maxtmay=[17.98431,22.63004]degc

AND
maxtnov=[21.42478,30.96597]degc
AND
maxtoct=[19.50103,26.85672]degc
THEN    Taxon=BACKGROUND

8    d    0.50  0.69  8.73  0.20  0.019
IF    veglr=[ 3,18]catno
AND
wcdem=[80,1288]m
AND
slopes=[0.0000,15.7061]perc

AND
catch=[20,366]catno
AND
aspect=[14,361]deg
AND
mintdec=[11.14087,14.68186]degc

AND
mintjan=[12.46233,15.56972]degc
AND
mintjun=[1.74913,7.23174]degc
AND
mintmar=[11.05225,14.77557]degc
```

```

AND
mintmay=[4.77636,10.08636]degc
AND
mintoct=[3.49174,9.26261]degc
AND
mintsept=[4.78640,8.61597]degc

AND
maxtaug=[15.58850,20.54759]degc
AND
maxtdec=[24.17913,31.28782]degc
AND
maxtfeb=[24.97427,33.78790]degc

AND
maxtjul=[11.68300,20.16980]degc
AND
maxtmay=[16.22213,24.07182]degc
AND
maxtnov=[22.98115,28.66526]degc

AND
maxtoct=[20.15885,26.13909]degc
THEN Taxon=BACKGROUND

2   r   0.50  0.78  12.48  0.19  0.017
IF   + apan*0.21 mm + slopes*0.26 perc - seadist*0.14 m
      + catch*0.30 catno + avgrain*0.13 mm + aspect*0.30 deg
      + latitude*0.18 m + longitud*0.46 m - avgmint*0.21 degc
      - mintapr*0.23 degc - mintaug*0.29 degc - mintdec*0.18 degc
      - mintjan*0.13 degc - mintjun*0.07 degc - mintjul*0.21 degc
      - mintmay*0.23 degc + mintnov*0.30 degc - mintoct*0.46 degc
      - maxtdec*0.28 degc - maxtfeb*0.40 degc - maxtjan*0.31 degc
      + maxtjul*0.08 degc - maxtmay*0.10 degc - maxtjun*0.09 degc
      - maxtmay*0.32 degc - maxtnov*0.27 degc - maxtoct*0.17 degc
THEN Taxon=BACKGROUND

```



```
17  r    0.51  0.88  6.47  0.03  0.017
IF  + veglr*0.32 catno + slopes*0.26 perc + catch*0.06 catno
    - avgrain*0.06 mm + aspect*0.30 deg + longitud*0.02 m
    - avgmint*0.36 degc - rivdist*0.32 m + mintdec*0.14 degc
    + mintmar*0.29 degc - mintmay*0.22 degc - mintoct*0.46 degc
    - maxtaug*0.39 degc + maxtjul*0.08 degc + maxtjun*0.21 degc
THEN Taxon=PRESENT

24  d    0.49  0.66  5.32  0.10  0.011
IF  veglr=[ 4,16]catno
AND
apan=[1573,2485]mm
AND
wcdem=[80,1199]m

AND
slopes=[1.5972,14.6413]perc
AND
geology=[ 3,12]catno
AND
catch=[ 3,399]catno

AND
solrad=[16.9082,22.9627]perc
AND
avgrain=[6545,5819]mm
AND
avgmaxt=[18.69067,26.03794]degc

AND
aspect=[27,333]deg
AND
latitude=[50474,290346]m
AND
avgmint=[7.73407,12.00095]degc

AND
rivdist=[ 0,24997]m
```

AND  
mintapr=[6.45822,12.72953]degc  
AND  
mintaug=[2.85419,8.07790]degc  
  
AND  
mintdec=[10.22933,14.47150]degc  
AND  
mintfeb=[11.93348,15.66597]degc  
AND  
mintjan=[11.37273,15.48901]degc  
  
AND  
mintjun=[2.37087,7.85348]degc  
AND  
mintjul=[1.07130,9.17087]degc  
AND  
mintmar=[10.19589,14.73834]degc  
  
AND  
mintmay=[3.59636,10.56909]degc  
AND  
mintnov=[7.77870,13.06826]degc  
AND  
mintoct=[6.72020,11.07862]degc  
  
AND  
mintsept=[3.65494,9.39929]degc  
AND  
maxtapr=[19.90059,26.27198]degc  
AND  
maxtaug=[15.41601,20.28885]degc  
  
AND  
maxtdec=[21.89652,30.37478]degc  
AND  
maxtfeb=[25.66403,30.33909]degc  
AND

maxtjan=[25.00241,30.54233]degc

AND

maxtjul=[15.14601,18.95043]degc

AND

maxtmar=[21.51490,31.21569]degc

AND

maxtjun=[14.17455,19.81075]degc

AND

maxtmay=[17.23672,23.00383]degc

AND

maxtoct=[17.52755,26.19889]degc

THEN Taxon=PRESENT

22 a 0.51 1.00 5.41 0.01 0.009

IF slopes=52.7087

AND

geology= 2

AND

seadist=195424

AND

catch=485

AND

avgrain=4017

AND

rivdist=16070

AND

mintapr=5.41300

AND

mintjan=10.28312

AND

mintmar=9.26506

AND

mintmay=1.88000

```
AND
maxtapr=22.24795
AND
maxtaug=16.79593

AND
maxtmar=28.52103
AND
maxtnov=25.14652
AND
maxtoct=22.19213
THEN    Taxon=PRESENT

25  a    0.51  1.00  5.30  0.01  0.009
IF  apan=1682
AND
geology= 2
AND
avgmaxt=20.87356

AND
latitude=397808
AND
longitud=331286
AND
avgmint=6.38016

AND
rivdist=39606
AND
mintfeb=12.32826
AND
mintjan=11.09024

AND
mintjun=1.35348
AND
mintnov=8.24739
```



AND  
maxtapr=20.96249

AND  
maxtfeb=26.89028

AND  
maxtmay=16.11534

AND  
maxtnov=23.18415

THEN Taxon=PRESENT

27 a 0.55 1.00 5.16 0.01 0.008

IF seadist=48856

AND  
catch=464

AND  
solrad=23.6991

AND  
longitud=223373

AND  
rivdist=26621

AND  
mintapr=5.84059

AND  
mintaug=1.26051

AND  
mintjul=-0.85447

AND  
mintsept=2.47996

AND  
maxtjan=29.98834

AND  
maxtmar=28.64079

AND  
maxtjun=10.52759

```
AND
maxtmay=16.22213
AND
maxtnov=25.48486
AND
maxtoct=22.25194
THEN    Taxon=PRESENT

30  a    0.50  1.00  5.10  0.01  0.007
IF  apan=1649
AND
wcdem=2194
AND
solrad=19.1991

AND
aspect=-0
AND
avgmint=5.51858
AND
rivdist=40580

AND
mintaug=1.48186
AND
mintdec=10.15921
AND
mintjan=10.44455

AND
mintjun=-0.22913
AND
mintmay=2.89909
AND
maxtaug=15.58850

AND
```

maxtdec=27.44000

AND

maxtjul=11.63423

AND

maxtmay=16.91632

AND

maxtoct=20.81668

THEN Taxon=PRESENT

40 a 0.49 1.00 2.72 0.00 0.007

IF veglr= 4

AND

apan=1758

AND

slopes=63.0907

AND

geology= 3

AND

seadist=195424

AND

catch=410

AND

avgmaxt=20.66059

AND

longitud=331286

AND

avgmint=6.70838

AND

rivdist=37658

AND

mintaug=3.20834

AND

mintfeb=12.47182

```
AND
mintjun=1.57957
AND
mintnov=8.41478
AND
mintoct=5.02526

AND
maxtaug=14.38107
AND
maxtjan=26.66439
THEN    Taxon=PRESENT

10  d    0.51  0.99  8.06  0.03  0.006
IF  veglr=[ 3,20]catno
AND
apan=[1733,2350]mm
AND
wcdem=[53,1092]m

AND
slopes=[4.2593,7.9862]perc
AND
geology=[ 7,11]catno
AND
seadist=[24428,268708]m

AND
catch=[32,190]catno
AND
solrad=[17.4400,22.3900]perc
AND
avgrain=[6585,5537]mm

AND
avgmaxt=[21.29949,23.69534]degc
AND
aspect=[82,332]deg
```



AND

latitude=[40879,432350]m

AND

longitud=[113007,556923]m

AND

avgmint=[6.58530,11.79581]degc

AND

rivdist=[2759,8603]m

AND

mintapr=[9.26130,12.49198]degc

AND

mintaug=[3.16407,7.85656]degc

AND

mintdec=[9.52814,15.69858]degc

AND

mintfeb=[14.08684,15.70186]degc

AND

mintjan=[11.29202,16.21542]degc

AND

mintjun=[1.57957,7.06217]degc

AND

mintjul=[2.26075,8.32127]degc

AND

mintmar=[10.75439,14.47771]degc

AND

mintmay=[5.47364,10.24727]degc

AND

mintnov=[9.15130,12.93435]degc

AND

mintoct=[8.49585,10.43293]degc

AND

mintsept=[3.65494,9.31225]degc

AND

maxtapr=[19.90059,25.93664] degc

AND

maxtaug=[16.27846,19.90075] degc

AND

maxtdec=[24.30956,28.22261] degc

AND

maxtfeb=[26.35379,28.95956] degc

AND

maxtjan=[25.55640,28.46486] degc

AND

maxtjul=[15.58498,19.09676] degc

AND

maxtmar=[23.85028,30.91628] degc

AND

maxtjun=[14.67186,19.64498] degc

AND

maxtmay=[18.78530,22.79024] degc

AND

maxtnov=[22.64281,29.47727] degc

AND

maxtoct=[20.39806,25.78028] degc

AND

maxtsept=[18.57498,22.02395] degc

THEN Taxon=PRESENT

32 a 0.49 1.00 4.57 0.01 0.006

IF apan=1725

AND

wcdem=2238

AND

solrad=17.5218

AND

avgrain=3950

AND  
avgmaxt=20.76707  
AND  
aspect=180

AND  
latitude=395889  
AND  
avgmint=6.62632  
AND  
mintjan=11.29202

AND  
mintjul=1.75099  
AND  
mintmar=10.34482  
AND  
mintmay=4.18636

AND  
mintnov=8.44826  
AND  
mintoct=5.02526  
AND  
maxtaug=14.55356

AND  
maxtjan=26.80289  
AND  
maxtmar=26.78447  
AND  
maxtjun=11.02490

AND  
maxtoct=20.09905  
THEN Taxon=PRESENT

14 d 0.50 0.60 7.21 0.55 0.001

```
IF veglr=[ 3,18]catno
AND
apan=[1437,2730]mm
AND
wcdem=[80,1297]m

AND
slopes=[0.0000,31.6785]perc
AND
geology=[ 1,12]catno
AND
avgrain=[6599,6048]mm

AND
avgmint=[7.32379,11.30348]degc
AND
mintaug=[1.30478,7.98937]degc
AND
mintjun=[-0.34217,8.41870]degc

AND
mintmay=[3.16727,10.30091]degc
AND
mintnov=[8.64913,13.26913]degc
AND
maxtmar=[23.31134,31.99415]degc

AND
maxtmay=[15.26095,22.89704]degc
THEN Taxon=BACKGROUND

38 d 0.50 1.00 2.84 0.00 0.001
IF apan=[1742,2578]mm
AND
slopes=[0.0000,15.9723]perc
AND
catch=[ 5,424]catno
```



AND

avgmint=[6.25708,11.63170]degc

AND

mintapr=[7.78850,12.44447]degc

AND

mintaug=[3.03126,7.54668]degc

AND

mintfeb=[12.25648,16.49142]degc

AND

mintjan=[12.22020,15.69079]degc

AND

mintjun=[1.63609,7.23174]degc

AND

mintjul=[0.61818,8.88767]degc

AND

mintmay=[3.16727,10.30091]degc

AND

maxtdec=[23.39652,32.00521]degc

AND

maxtfeb=[26.12387,32.17846]degc

AND

maxtmar=[29.77854,29.77854]degc

AND

maxtmay=[16.75613,23.64462]degc

THEN Taxon=BACKGROUND

39 r 0.50 1.00 2.83 0.00 0.001

IF + veglr\*0.32 catno + wcdem\*0.38 m + slopes\*0.26 perc  
 - geology\*0.35 catno + seadist\*0.01 m + catch\*0.30 catno  
 + avgrain\*0.44 mm - avgmaxt\*0.02 degc + aspect\*0.30 deg  
 + latitude\*0.08 m + longitud\*0.02 m - avgmint\*0.36 degc  
 + mintdec\*0.14 degc + mintfeb\*0.07 degc + mintjan\*0.06 degc  
 - mintjun\*0.07 degc - mintmay\*0.22 degc + mintnov\*0.30 degc  
 - mintsept\*0.28 degc + maxtapr\*0.21 degc - maxtaug\*0.39 degc  
 - maxtdec\*0.28 degc + maxtjul\*0.08 degc - maxtmar\*0.15 degc

```
+ maxtjun*0.21 degc - maxtoct*0.17 degc  
THEN Taxon=BACKGROUND
```

```
3 d 0.50 0.69 11.62 0.39 0.000
```

```
IF apan=[1573,2502]mm
```

```
AND
```

```
wcdem=[ 0,1341]m
```

```
AND
```

```
catch=[ 5,202]catno
```

```
AND
```

```
solrad=[17.5218,22.4718]perc
```

```
AND
```

```
avgmint=[7.20071,11.79581]degc
```

```
AND
```

```
mintapr=[2.99000,15.05751]degc
```

```
AND
```

```
mintaug=[3.16407,8.25498]degc
```

```
AND
```

```
mintdec=[9.70344,14.54162]degc
```

```
AND
```

```
mintfeb=[11.14391,15.91719]degc
```

```
AND
```

```
mintjan=[9.71814,15.85221]degc
```

```
AND
```

```
mintjun=[1.35348,8.02304]degc
```

```
AND
```

```
mintjul=[2.09083,8.77439]degc
```

```
AND
```

```
mintnov=[8.64913,13.00130]degc
```

```
AND
```

```
mintsept=[3.08921,9.61688]degc
```

```
AND
```

```
maxtdec=[21.17913,30.44000]degc
```

AND

maxtmar=[23.85028,29.11984]degc

AND

maxtoct=[17.04913,22.61075]degc

THEN Taxon=PRESENT

11 r 0.53 0.97 7.70 0.03 0.000

IF + veglr\*0.32 catno - apan\*0.46 mm + wcdem\*0.38 m  
 + slopes\*0.26 perc - geology\*0.35 catno + seadist\*0.06 m  
 + catch\*0.30 catno - solrad\*0.16 perc + avgrain\*0.13 mm  
 - avgmaxt\*0.43 degc + aspect\*0.30 deg + latitude\*0.08 m  
 + longitud\*0.09 m - avgmint\*0.36 degc - rivdist\*0.32 m  
 - mintapr\*0.41 degc - mintaug\*0.20 degc + mintdec\*0.14 degc  
 + mintfeb\*0.07 degc + mintjan\*0.06 degc - mintjun\*0.07 degc  
 + mintjul\*0.03 degc + mintmar\*0.31 degc - mintmay\*0.22 degc  
 + mintnov\*0.30 degc - mintoct\*0.46 degc - mintsept\*0.28 degc  
 + maxtapr\*0.21 degc - maxtaug\*0.39 degc - maxtdec\*0.28 degc  
 - maxtfeb\*0.40 degc - maxtjan\*0.31 degc + maxtjul\*0.08 degc  
 - maxtmar\*0.15 degc + maxtjun\*0.21 degc - maxtmay\*0.32 degc  
 - maxtnov\*0.26 degc - maxtoct\*0.17 degc

THEN Taxon=PRESENT

12 d 0.50 0.61 7.67 0.50 0.000

IF apan=[1742,2578]mm

AND

slopes=[0.0000,15.9723]perc

AND

catch=[ 5,424]catno

AND

avgmint=[6.25708,11.63170]degc

AND

mintapr=[7.78850,12.44447]degc

AND

mintaug=[3.03126,7.54668]degc

AND

mintfeb=[12.25648,16.49142]degc

AND

mintjan=[12.22020,15.69079] degc

AND

mintjun=[1.63609,7.23174] degc

AND

mintjul=[0.61818,8.88767] degc

AND

mintmay=[3.16727,10.30091] degc

AND

maxtdec=[23.39652,32.00521] degc

AND

maxtfeb=[26.12387,32.17846] degc

AND

maxtmay=[16.75613,23.64462] degc

THEN Taxon=BACKGROUND

13 r 0.50 0.87 7.37 0.04 0.000

IF - veglr\*0.21 catno + slopes\*0.26 perc + seadist\*0.01 m  
 + catch\*0.30 catno + avgrain\*0.13 mm + latitude\*0.08 m  
 - avgmint\*0.36 degc + mintdec\*0.14 degc + mintfeb\*0.07 degc  
 + mintjan\*0.06 degc - mintjun\*0.07 degc - mintmay\*0.22 degc  
 + mintnov\*0.30 degc + maxtapr\*0.21 degc + maxtjul\*0.08 degc  
 - maxtmar\*0.15 degc

THEN Taxon=PRESENT

16 d 0.51 0.61 6.66 0.49 0.000

IF apan=[1750,2578]mm

AND

avgmaxt=[19.96846,26.73008] degc

AND

latitude=[79259,344077]m

AND

mintapr=[6.45822,12.77704] degc

AND

mintdec=[7.91542,16.22447] degc



```
AND
mintfeb=[12.86660,16.09664]degc

AND
mintjun=[0.73174,7.51435]degc
AND
mintjul=[0.56154,8.94431]degc
AND
mintmar=[10.71715,15.03621]degc

AND
mintnov=[9.45261,12.86739]degc
AND
maxtfeb=[23.82466,35.55063]degc
AND
maxtjul=[13.53644,19.29186]degc

AND
maxtmar=[24.14968,30.73664]degc
AND
maxtjun=[13.73249,20.08703]degc
AND
maxtmay=[16.86292,23.37763]degc

AND
maxtnov=[21.56012,30.62763]degc
THEN    Taxon=BACKGROUND

18  d    0.50  0.58  6.32  0.56  0.000
IF  veglr=[ 3,18]catno
AND
apan=[1488,2738]mm
AND
wcdem=[80,1297]m

AND
slopes=[0.0000,31.6785]perc
AND
```

```
geology=[ 1,12]catno
AND
avgrain=[6599,6048]mm

AND
avgmint=[7.32379,11.30348]degc
AND
mintaug=[1.30478,7.98937]degc
AND
mintjun=[-0.34217,8.41870]degc

AND
mintjul=[-3.29000,10.87008]degc
AND
mintmay=[3.16727,10.30091]degc
AND
mintnov=[8.68261,13.23565]degc

AND
maxtmar=[23.31134,31.99415]degc
AND
maxtmay=[15.26095,22.89704]degc
THEN    Taxon=BACKGROUND

19  d    0.50  0.59  5.78  0.41  0.000
IF  veglr=[ 3,18]catno
AND
wcdem=[80,1288]m
AND
slopes=[0.0000,15.7061]perc

AND
catch=[20,366]catno
AND
aspect=[14,361]deg
AND
mintdec=[11.14087,14.68186]degc
```

```
AND
mintjan=[12.46233,15.56972] degc
AND
mintjun=[1.74913,7.23174] degc
AND
mintmar=[11.05225,14.77557] degc

AND
mintmay=[4.77636,10.19364] degc
AND
mintsept=[4.78640,8.61597] degc
AND
maxtapr=[15.26178,28.50755] degc

AND
maxtaug=[15.58850,20.54759] degc
AND
maxtdec=[24.17913,31.28782] degc
AND
maxtfeb=[24.97427,33.78790] degc

AND
maxtjul=[11.68300,20.16980] degc
AND
maxtmay=[16.22213,24.07182] degc
AND
maxtnov=[22.98115,28.66526] degc

AND
maxtoct=[20.15885,26.13909] degc
THEN    Taxon=BACKGROUND

20    d    0.50  0.59  5.52  0.38  0.000
IF    slopes=[0.2662,20.2316] perc
AND
seadist=[24428,219852] m
AND
```

```
solrad=[17.6445,22.3900]perc

AND
latitude=[48555,301860]m
AND
longitud=[88481,574091]m
AND
avgmint=[6.58530,11.83684]degc

AND
mintjan=[10.68668,15.48901]degc
AND
mintjun=[2.37087,7.79696]degc
AND
mintmay=[5.25909,10.19364]degc

AND
mintnov=[8.85000,13.00130]degc
AND
mintoct=[6.47806,11.11897]degc
AND
maxtapr=[18.05625,27.27798]degc

AND
maxtjul=[12.75605,19.24308]degc
AND
maxtmar=[21.57478,30.79652]degc
AND
maxtjun=[15.22443,19.70024]degc

AND
maxtmay=[16.91632,23.43103]degc
THEN    Taxon=PRESENT

21    d    0.49  0.57  5.50  0.48  0.000
IF    veglr=[ 3,22]catno
AND
wcdem=[ 0,1288]m
```



AND

seadist=[24428,268708]m

AND

avgrain=[6518,5846]mm

AND

avmaxt=[21.03328,24.65368]degc

AND

latitude=[8257,357510]m

AND

longitud=[98292,547112]m

AND

mintapr=[7.78850,12.53949]degc

AND

mintdec=[11.00063,14.40138]degc

AND

mintjul=[2.03419,8.77439]degc

AND

mintmar=[10.67992,14.62664]degc

AND

mintnov=[7.77870,13.00130]degc

AND

mintsept=[3.08921,9.61688]degc

AND

maxtaug=[15.41601,20.28885]degc

AND

maxtfeb=[22.29186,32.86822]degc

AND

maxtjan=[21.95546,31.92731]degc

AND

maxtmar=[23.55087,30.13783]degc

THEN Taxon=PRESENT

23 d 0.50 0.59 5.40 0.40 0.000

```
IF veglr=[ 3,18]catno
AND
wcdem=[80,1288]m
AND
slopes=[0.0000,15.7061]perc

AND
catch=[20,366]catno
AND
aspect=[14,361]deg
AND
mintdec=[11.14087,14.68186]degc

AND
mintjan=[12.46233,15.56972]degc
AND
mintjun=[1.74913,7.23174]degc
AND
mintmar=[11.05225,14.77557]degc

AND
mintmay=[4.77636,10.08636]degc
AND
mintsept=[4.78640,8.61597]degc
AND
maxtaug=[15.58850,20.54759]degc

AND
maxtdec=[23.98348,31.48348]degc
AND
maxtfeb=[24.97427,33.78790]degc
AND
maxtjul=[11.68300,20.16980]degc

AND
maxtmay=[16.22213,24.07182]degc
AND
maxtnov=[22.98115,28.66526]degc
```

AND

maxtoct=[20.15885,26.13909]degc

THEN Taxon=BACKGROUND

26 d 0.49 0.56 5.28 0.46 0.000

IF apan=[1792,2561]mm

AND

wcdem=[ 0,1661]m

AND

seadist=[ 0,244280]m

AND

catch=[22,359]catno

AND

solrad=[18.5036,22.9627]perc

AND

aspect=[21,361]deg

AND

mintdec=[11.17593,14.64680]degc

AND

mintfeb=[12.79482,16.16842]degc

AND

mintjun=[1.74913,7.23174]degc

AND

mintjul=[2.20411,8.32127]degc

AND

maxtfeb=[25.43411,33.32806]degc

AND

maxtjan=[25.27941,32.61980]degc

AND

maxtjul=[13.48767,19.34063]degc

AND

maxtmar=[24.20957,30.55700]degc

AND

maxtjun=[13.67723,20.19755]degc

```
AND
maxtoct=[18.54419,27.93316] degc
THEN    Taxon=BACKGROUND

29  d    0.50  0.60  5.11  0.28  0.000
IF  apan=[1573,2705]mm
AND
wcdem=[ 0,1892]m
AND
geology=[ 1,12]catno

AND
solrad=[18.6264,22.9627]perc
AND
avgrain=[6478,6115]mm
AND
longitud=[73766,537302]m

AND
mintapr=[5.98312,12.96708] degc
AND
mintaug=[2.63285,7.67949] degc
AND
mintdec=[10.72016,14.85715] degc

AND
mintjan=[12.42198,15.56972] degc
AND
mintjul=[1.12794,8.66111] degc
AND
mintmar=[10.82885,14.99897] degc

AND
mintoct=[5.50953,10.99790] degc
AND
mintsept=[4.04660,8.74652] degc
AND
```

maxtapr=[21.57727,26.27198]degc

AND

maxtfeb=[25.66403,32.63830]degc

AND

maxtnov=[21.22178,31.10131]degc

AND

maxtoct=[19.97945,26.31850]degc

THEN Taxon=BACKGROUND

31 d 0.50 0.57 4.87 0.51 0.000

IF apan=[1750,2578]mm

AND

avgmaxt=[19.96846,26.73008]degc

AND

latitude=[79259,344077]m

AND

mintapr=[6.45822,12.77704]degc

AND

mintfeb=[12.86660,16.09664]degc

AND

mintjun=[0.73174,7.51435]degc

AND

mintjul=[0.56154,8.94431]degc

AND

mintmar=[10.71715,15.03621]degc

AND

mintnov=[9.45261,12.86739]degc

AND

maxtfeb=[23.82466,35.55063]degc

AND

maxtjul=[13.53644,19.29186]degc

AND

maxtmar=[24.14968,30.73664]degc



```
AND
maxtjun=[13.73249,20.08703]degc
AND
maxtmay=[16.86292,23.37763]degc
AND
maxtnov=[21.56012,30.62763]degc
THEN    Taxon=BACKGROUND

33  d    0.49  0.56  4.42  0.46  0.000
IF  wcdem=[89,1199]m
AND
slopes=[0.0000,31.6785]perc
AND
geology=[ 2,12]catno

AND
avgmaxt=[20.92680,25.82498]degc
AND
latitude=[71583,370943]m
AND
longitud=[83576,517682]m

AND
mintapr=[7.45593,12.49198]degc
AND
mintaug=[2.54431,7.67949]degc
AND
mintjan=[12.17984,15.73115]degc

AND
mintoct=[6.63949,10.67506]degc
AND
mintsept=[2.69755,9.00763]degc
AND
maxtfeb=[25.97059,32.25510]degc

AND
maxtjan=[23.75593,34.14328]degc
```

```
AND
maxtmar=[24.74850,29.77854]degc
AND
maxtmay=[15.79494,24.49901]degc

AND
maxtoct=[20.39806,25.78028]degc
THEN    Taxon=BACKGROUND

34  a    0.48  1.00  4.27  0.01  0.000
IF  seadist=48856
AND
catch=464
AND
longitud=223373

AND
rivdist=26621
AND
mintapr=5.84059
AND
mintaug=1.26051

AND
mintjul=-0.85447
AND
mintsept=2.47996
AND
maxtjan=29.98834

AND
maxtmar=28.64079
AND
maxtjun=10.52759
AND
maxtmay=16.22213

AND
```

```
maxtnov=25.48486
AND
maxtoct=22.25194
THEN    Taxon=PRESENT

35  d    0.47  0.52  3.83  0.51  0.000
IF  apan=[1750,2578]mm
AND
avgmaxt=[20.02170,26.78332]degc
AND
latitude=[79259,344077]m

AND
mintapr=[6.45822,12.77704]degc
AND
mintfeb=[12.86660,16.09664]degc
AND
mintjun=[0.73174,7.51435]degc

AND
mintjul=[0.56154,8.94431]degc
AND
mintmar=[10.71715,15.03621]degc
AND
mintnov=[9.45261,12.86739]degc

AND
maxtdec=[18.70087,34.80956]degc
AND
maxtfeb=[23.82466,35.55063]degc
AND
maxtjul=[13.53644,19.29186]degc

AND
maxtmar=[24.14968,30.73664]degc
AND
maxtjun=[13.73249,20.08703]degc
AND
```

maxtmay=[16.86292,23.37763]degc

AND

maxtnov=[21.56012,30.62763]degc

THEN Taxon=BACKGROUND

36 d 0.50 0.55 3.16 0.41 0.000

IF veglr=[ 4,22]catno

AND

slopes=[0.2662,19.4330]perc

AND

geology=[ 3,12]catno

AND

seadist=[24428,219852]m

AND

avgmaxt=[20.50087,25.45229]degc

AND

aspect=[46,361]deg

AND

latitude=[27447,313374]m

AND

rivdist=[487,11525]m

AND

mintoct=[6.19557,11.24004]degc

AND

maxtapr=[18.55925,27.27798]degc

AND

maxtjan=[21.95546,31.51182]degc

AND

maxtmar=[22.17360,30.13783]degc

AND

maxtmay=[18.41150,22.09605]degc

AND

maxtoct=[17.46775,26.19889]degc

```
THEN    Taxon=PRESENT

37  d    0.49  0.59  2.89  0.09  0.000
IF  veg1r=[ 4,16]catno
AND
apan=[1573,2485]mm
AND
wcdem=[18,1492]m

AND
slopes=[0.2662,20.2316]perc
AND
geology=[ 3,12]catno
AND
catch=[ 7,229]catno

AND
solrad=[17.5218,22.4718]perc
AND
avgrain=[6532,5819]mm
AND
avgmaxt=[19.06336,27.63518]degc

AND
aspect=[27,333]deg
AND
latitude=[71583,355591]m
AND
avgmint=[6.87249,11.46759]degc

AND
rivdist=[649,12499]m
AND
mintapr=[6.45822,12.72953]degc
AND
mintaug=[3.16407,7.90083]degc

AND
```



mintdec=[9.70344,14.54162]degc  
AND  
mintfeb=[11.75403,15.77364]degc  
AND  
mintjan=[11.37273,15.48901]degc  
  
AND  
mintjun=[1.35348,8.02304]degc  
AND  
mintjul=[1.07130,9.17087]degc  
AND  
mintmar=[10.19589,14.73834]degc  
  
AND  
mintmay=[3.59636,10.56909]degc  
AND  
mintnov=[7.77870,13.06826]degc  
AND  
mintsept=[4.52530,7.91968]degc  
  
AND  
maxtapr=[19.90059,26.27198]degc  
AND  
maxtaug=[15.41601,20.28885]degc  
AND  
maxtdec=[21.89652,30.37478]degc  
  
AND  
maxtfeb=[25.66403,30.33909]degc  
AND  
maxtjan=[25.00241,30.54233]degc  
AND  
maxtjul=[13.14625,19.24308]degc  
  
AND  
maxtmar=[21.51490,31.21569]degc  
AND  
maxtjun=[14.28506,19.75549]degc

AND

maxtmay=[17.23672,23.00383]degc

AND

maxtoct=[17.52755,26.19889]degc

THEN Taxon=PRESENT

## APPENDIX H: GRASS PROJECT SETUP

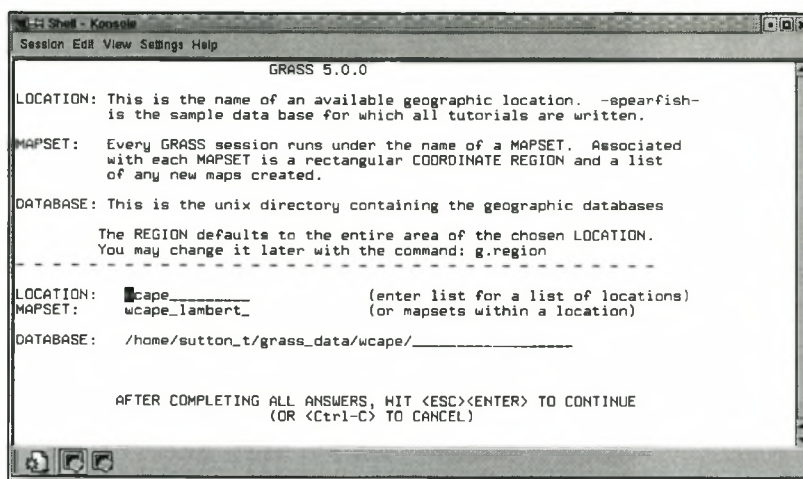


Figure H.1: Setting up a new location in Grass.

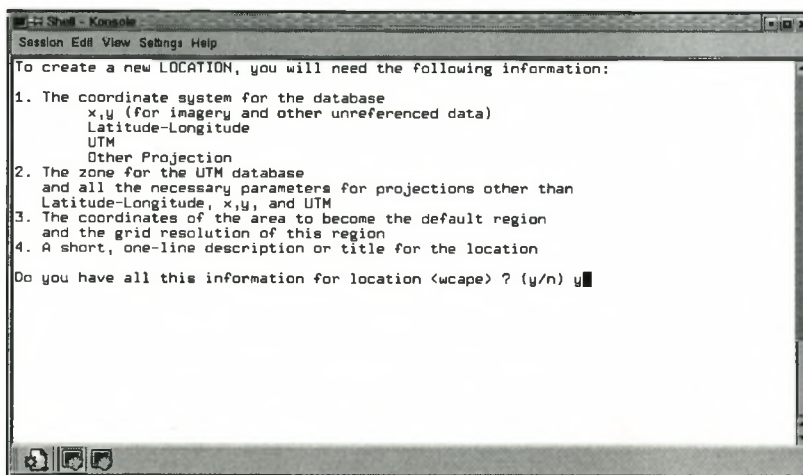


Figure H.2: New location confirmation in Grass.

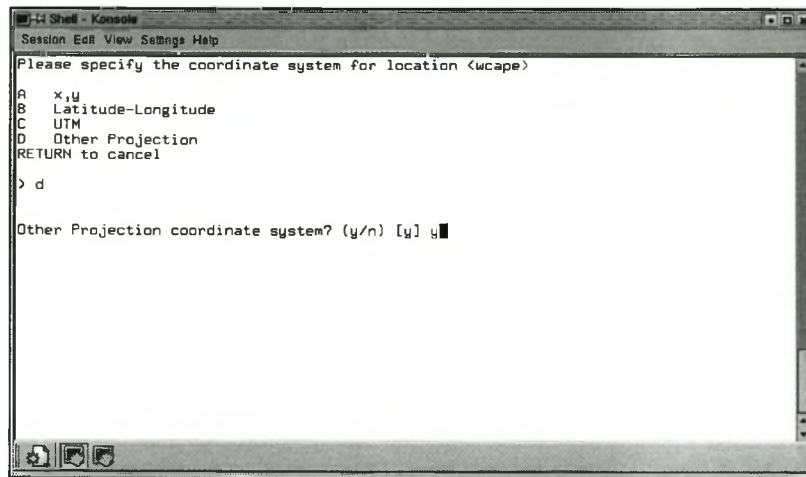


Figure H.3: Coordinate system selection for the new project.

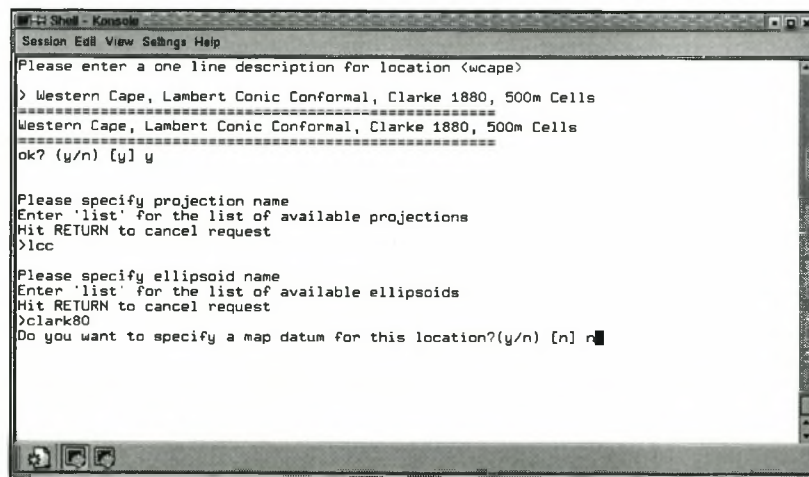
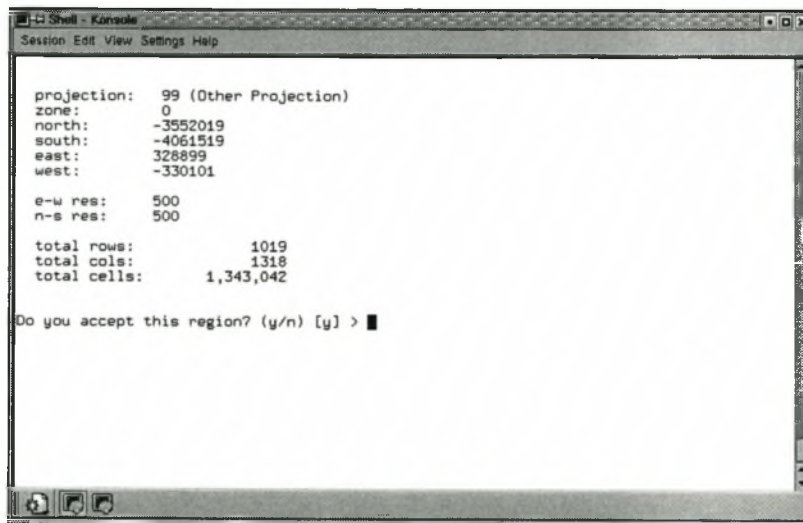


Figure H.4: Setting the location description, projection and ellipsoid in Grass.



The image shows a terminal window titled "Shell - Konsole" with a menu bar containing "Session Edit View Settings Help". The terminal displays the following text:

```
projection: 99 (Other Projection)
zone: 0
north: -3552019
south: -4061519
east: 328899
west: -330101

e-w res: 500
n-s res: 500

total rows: 1019
total cols: 1318
total cells: 1,343,042

Do you accept this region? (y/n) [y] > █
```

Figure H.5: Defining the region of interest for the new location.