Quantitative approaches for environmental decision making

Ameneh Shobeirinejad
B.Eng. (Hons), M.Sc.

School of Information and Communication Technology
Griffith University

Submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy

July 2017
Abstract

Ecologists and environmental managers regularly need to make decisions about restoration and management with limited information and uncertainty about the outcomes of system interventions. Uncertainty is an inevitable component of environmental management and decision making (as well as other complex systems’ management); however, the significance of environmental problems necessitates the effort to quantify and, where possible, minimise the uncertainty around the outputs of models of a given system. Uncertainty about the model output stems from the uncertainty about the model input (arising from different sources) and the model structure. The latter is the more complex of the two and as such challenging to address.

There are a range of quantitative methods available to assist with environmental management and decision making. However, while finding one specific model to represent a system effectively is an ideal goal, the selection of the most suitable model among those available is a challenging task for the modeller. Model selection includes at least two aspects, the selection of appropriate variables and an appropriate model structure. Model structure describes the nature of the mathematical representation of the cause and effect relationships that are quantified by the model. Many approaches for variable selection already exist; however, methods to guide the quantitative selection of an appropriate model structure are not so well developed. Model structure selection is an important step in modelling, which needs not only to include the essential variables and processes of the system, but also to avoid unnecessary complexity that doesn’t improve the modelling results.

Selecting an appropriate model structure involves several related steps. One step is to determine the possible nature of the cause and effect relationships among the variables. A second step is to investigate carefully the amount and the quality of the available information, along with evaluating the uncertainty about the conditions, the variables and the modelling parameters, and then selecting one out of many model structures. For example, the modeller may decide to use deterministic models, or to embody some stochastic components in the model and assign some probability to the occurrence of some random variables. For the latter, they may decide to use a statistical model or other
probabilistic models, such as Bayesian networks. They may decide to represent the whole system with one model of the system, or may choose to break the large-scale system down into simpler ones and use different models to represent the more clear cause and effect relationships among variables.

Among the many aspects of model structure for a modeller to select, there is the choice between a single-level and a multi-level model structure. Single-level models represent the relationships between variables with fixed coefficients, and in case the data are grouped, ignore the group differences. Conversely, hierarchical models consider the relationships within and between levels of grouped data and can account for the variation between groups. Hierarchical models can include varying coefficients to quantify how relationships between variables at one level may depend on variables at other levels. The complexity and heterogeneity of environmental and ecological systems can benefit from the use of hierarchical models to accommodate more complexity and embrace different principles that might apply at different scales. However, compared to single-level models, hierarchical models are considerably more complex and more difficult to implement. Therefore, there is a strong need to understand the conditions where the additional work of fitting a hierarchical model is necessary.

In this research, I aimed to identify the statistical conditions under which hierarchical models provide a better fit to complex data than single level models. This involved the analysis of an empirical ecological dataset in tandem with a large simulation study of 70,000 datasets. The simulation study provided a way to analyse a large range of datasets with known structure, uncertainty and relationships among the variables, while the empirical study provided an avenue to test the approach in a real setting with noisy data.

For the simulation study, I set both single-level and hierarchical models’ structures as Poisson regression (due to the importance of this distribution in a large number of ecological studies) in a Bayesian framework. The Bayesian framework is a flexible approach that is used increasingly to quantify environmental and ecological processes, and guide decision making. A key feature of Bayesian approaches is the capacity to quantify uncertainty at all levels of the model and propagate that uncertainty through to the response or outcome variable. This ensures that uncertainty around predictions from the model, be they ecological responses to natural disturbances or management
interventions, is naturally included in the model output. Moreover, Hierarchical Bayesian models offer great promise in quantifying multiscale processes and developing complex probabilistic models that reflect underlying ecological processes. The results of the simulation study identified which of the 70,000 datasets where a hierarchical model fit better than a single-level model.

Based on the findings, I developed a statistical tool for model-structure selection that can be efficiently applied to a set of data, and recommend a model-structure with an accompanying reliability of recommendation. This tool provides a quantitative approach to inform users when the additional effort of hierarchical modelling would provide a better model fit and when the simpler single level model structures are appropriate. To demonstrate the applicability of the proposed model-structure selection tool, I applied the proposed tool to three empirical ecological datasets. I also developed both single-level and hierarchical models on these datasets and compared standard goodness fit metrics to the recommendation from my proposed tool. For all datasets, the proposed tool recommended, with a very high reliability of recommendation, the same model-structure selected as the better fit by modelling results.
Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Ameneh Shobeirinejad
July 2017
To my beautiful son Ali,
you are the light of my life.
Table of Contents

Abstract ................................................................................................................................. III
Statement of Originality ........................................................................................................ VII
Table of Contents ................................................................................................................ XI
List of Figures ..................................................................................................................... XV
List of Tables ...................................................................................................................... XIX
Abbreviations ..................................................................................................................... XXI
Acknowledgements ........................................................................................................... XXIII
Publications during candidature ...................................................................................... XXIV
Keywords ........................................................................................................................... XXV

Chapter 1 Introduction ........................................................................................................ 1
  1.1 Background .................................................................................................................. 3
  1.2 Research problem ....................................................................................................... 5
  1.3 Research Aims and Objectives .................................................................................. 6
  1.4 Thesis outline ............................................................................................................ 7

Chapter 2 Quantitative models for environmental management –
Mathematical models ........................................................................................................ 11
  2.1 Introduction ................................................................................................................ 13
  2.2 Mathematical Models ............................................................................................... 14
    2.2.1 Input-Output Models (IO Models) ...................................................................... 15
    2.2.2 Agent-Based Models (ABMs) .......................................................................... 20
    2.2.3 Artificial Neural Networks ............................................................................... 22
    2.2.4 Bayesian Models ............................................................................................... 25
  2.3 Applications, Strengths/ Limitations, Comparison, and Applicability of Different Model types in Decision Making .............................................................................. 29
    2.3.1 Applications of Models .................................................................................... 29
    2.3.2 Strengths/Limitations of Various Models ......................................................... 33
    2.3.3 Comparison of the Different Models ................................................................. 37
    2.3.4 Applicability of Models for Decision Makings .................................................. 39
5.1 Introduction ........................................................................................................................................................................ 99
5.2 Analyses of Modelling results ................................................................................................................................................ 100
  5.2.1 Models predictive performance ........................................................................................................................................ 101
  5.2.2 Models descriptive performance ........................................................................................................................................ 114
5.3 Comparing the measures’ results and selecting the most appropriate goodness-of-fit measure ........................................................................................................................................................................ 119
5.4 Partitioning study space based on WAIC results .......................................................................................................................... 121
5.5 Summary .................................................................................................................................................................................... 127

Chapter 6 Establishing a tool for model-structure selection ................. 129

6.1 Introduction ........................................................................................................................................................................ 131
6.2 Method to develop the model-structure selection tool ............................................................................................................ 132
  6.2.1 Selecting sub-regions ....................................................................................................................................................... 133
  6.2.2 Data cleaning .................................................................................................................................................................. 133
  6.2.3 Finding informative measures ........................................................................................................................................ 134
  6.2.4 Model-based cluster analysis ........................................................................................................................................ 136
  6.2.5 Developing the model-structure selection tool ............................................................................................................... 138
  6.2.6 Applying the proposed tool to case studies ..................................................................................................................... 142
  6.2.7 Demonstrating applicability of the proposed tool on simulated datasets ........................................................................... 145
6.3 Description of results ................................................................................................................................................................. 146
  6.3.1 Selecting sub-regions ....................................................................................................................................................... 146
  6.3.2 Data cleaning .................................................................................................................................................................. 152
  6.3.3 Applying informative measures ........................................................................................................................................ 152
  6.3.4 Model-based cluster analysis ........................................................................................................................................ 156
  6.3.5 The proposed model-structure selection tool ..................................................................................................................... 157
  6.3.6 Applying the proposed tool to the case studies ............................................................................................................... 160
  6.3.7 Demonstrating applicability of the proposed tool on generated datasets ........................................................................... 163
6.4 Discussion ......................................................................................................................................................... 164
6.5 Summary .................................................................................................................................................................................... 167

Chapter 7 Conclusions and recommendations ........................................ 169

7.1 Introduction ........................................................................................................................................................................ 171
7.2 Summary of this research ......................................................................................................................................................... 171
7.2.1 Literature review ....................................................................................................... 172
7.2.2 Proposed approach .................................................................................................. 173
7.2.3 Analysing the results of simulation study ................................................................. 173
7.2.4 Establishing the model-structure selection tool ...................................................... 174
7.2.5 Applicability of the proposed tool ......................................................................... 175

7.3 Main significance and contribution ............................................................................ 175
7.4 Research Limitations .................................................................................................. 176
7.5 Future research .......................................................................................................... 176

References ......................................................................................................................... 179

Appendix I .......................................................................................................................... 199
List of Figures

Figure 1-1 Thesis outline ...............................................................................................................8
Figure 2-1 Research outline (Literature review highlighted in red)...........................................14
Figure 2-2 The diagram of pendulum forces at start point with no friction. .........................17
Figure 2-3 The structure of a typical agent-based model (Wooldridge, 1997) ....................20
Figure 2-4 An Agent and its properties (Macal and North, 2005). ....................................21
Figure 2-5 a) A feed-forward neural network with a single hidden layer, b) Structure of a sample neuron. ..............................................................................................................................23
Figure 2-6 A sign function, a linear function, and a sigmoid function .............................23
Figure 2-7 Lung-cancer example (Korb and Nicholson, 2004)...........................................28
Figure 2-8 Types of reasoning (Korb and Nicholson, 2004)..............................................28
Figure 3-1 Four normal distribution with random different values of $\mu$ and $\sigma$ ..............48
Figure 3-2 Four Binomial distribution with random different values of $p$ and $n$ ..............49
Figure 3-3 Continuous uniform distribution with upper and lower values of $a$ and $b$ .......50
Figure 3-4 Four Poisson distribution with random different values of $\lambda$ ......................51
Figure 4-1 Research outline (Chapter 4 highlighted in red). ..............................................69
Figure 4-2 The simulation methodological framework; it involves several steps with each one building on the previous to follow a systematic process to evaluate the performance of each model structure under a range of conditions. .................................................................70
Figure 4-3 Simulated values of $\beta_0$ and $\beta_1$ over the range of values of $Z$, for the point values of $\gamma_{0i}$, $\gamma_{1i}$, $\gamma_{2i}$= 0, and $\sigma_1$, $\sigma_2$= 0.4 (x-axis shows $Z$ values, y-axis shows $\beta$ values). ..................75
Figure 4-4 Generated $Y$s against $X$s, ($Y(X)$), in 10 groups on the same point as Figure 4-3. Regression lines shown are the lines of best fit for each separate group. .........................76
Figure 4-5 The convergence and density plot of a random parameter .............................78
Figure 4-6 The schematic process of WAIC scoring for every dataset, and the sum of WAIC scores of the seven datasets at every point of study space ........................................82
Figure 5-1 Research outline (Chapter 5 highlighted in red). ............................................99
Figure 5-2 Chapter 5 framework; the steps of analysing the modelling results ..............100
Figure 5-3 The line plot of WAIC average probability scores and the histogram of WAIC binary scores; the WAIC average probability scores are continuous variables, whereas, the WAIC binary scores only hold two possible values ................................................................. 102

Figure 5-4 Coefficients plot of PCR/PLS regression for WAIC average probability scores .... 103

Figure 5-5 The regression tree of the WAIC average probability score. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average WAIC average probability scores in each leaf, and n shows the number of points in each leaf, the red number above the box shows the leaf number (sd refers to $\sigma$ and g refers to $\gamma$) .......... 105

Figure 5-6 The boxplot of the WAIC average probability scores across all leaves of the regression tree in Figure 5-5. Observations above 0.5 indicate points in the subspace where hierarchical models fit best, indicating that the first terminal leaf, number 4, contains datasets where the majority fit best with a single level model ...................................................... 105

Figure 5-7 The MSE over the number leaves; the Mean Squared Error decreases with the increase of number of leaves, and no pruning is applicable .............................................. 106

Figure 5-8 The classification tree of the WAIC binary score. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The leaves labelled as ‘Hierarch’ show the hierarchical sub-regions, and the leaf labelled as ‘Single-l’ represents the single-level leaf. The right-side number in each leaf shows the number of hierarchical points, and the left-side number shows the number of single-level points in the leaf .................................................................................................................. 107

Figure 5-9 The regression tree for DIC scores. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of DIC scores in each leaf, and n shows the number of points in each leaf, the red number above the box shows the leaf number ...... 108

Figure 5-10 The boxplot of DIC scores across all leaves of the regression tree in Figure 5-9. Observations above 0.5 indicate points in the subspace where hierarchical models fitted best, while the ones below 0.5 are the points on which single-level models fitted best ..... 108

Figure 5-11 The regression tree for average $R^2$ scores. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average $R^2$ scores in each leaf, and n shows the number of points in each leaf, the red number above the box shows the leaf number. ............................................................................................................................ 110

Figure 5-12 The boxplot of average $R^2$ scores across all leaves of the regression tree in Figure 5-11. Observations above 0.5 indicate points in the subspace where hierarchical models fitted best, indicating that the leaves number 5 & 6, contain datasets where the majority fit best with a single level model. .......................................................................................... 111

Figure 5-13 The regression tree for average RMSE scores. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average RMSE scores in each
leaf, and n shows the number of points in each leaf, the red number above the box shows
the leaf number ................................................................. 112

Figure 5-14 The boxplot of average RMSE scores across all leaves of the regression tree in
Figure 5-13. Observations above 0.5 indicate points in the subspace where hierarchical
models fitted best, indicating that the first terminal leaf, number 5, contains datasets where
the highest majority fit best with a single level model ................................................ 113

Figure 5-15 The regression tree of the descriptive R² scores. ‘sd’ refers to ‘σ’ and ‘g’ refers to ‘γ’.
The decimal fractions in the boxes are the median values of descriptive R² scores in each
leaf, and n shows the number of points in each leaf, the red number above the box shows
the leaf number ................................................................. 116

Figure 5-16 The boxplot of the descriptive R² scores across all leaves of the regression tree in
Figure 5-15. Observations above 0 indicate points in the subspace where hierarchical
models fitted best, indicating that the leaves number 7 and 11 contain datasets where the
majority fit best with a single level model ................................................................. 116

Figure 5-17 The regression tree of descriptive RMSE scores. ‘sd’ refers to ‘σ’ and ‘g’ refers to
‘γ’. The decimal fractions in the boxes are the median values of descriptive RMSE scores in
each leaf, and n shows the number of points in each leaf, the red number above the box
shows the leaf number ................................................................. 118

Figure 5-18 The boxplot of the descriptive RMSE scores across all leaves of the regression tree
in Figure 5-17. Observations above 0 indicate points in the subspace where hierarchical
models fitted best; indicating that hierarchical models fitted best on all points in the study
space ................................................................. 118

Figure 5-19 a) The posterior distribution centred near the pre-known value of beta0, b) The pre-
known value lied out of the acceptable range ................................................................. 119

Figure 5-20 The box plot of the WAIC average probability scores of each set of points in the 6D
study space located at the same points in the 3D space with γ₁₁, σ₀, and σ₁ axes. The x-axis
shows the WAIC average probability scores, and the y-axis shows the values of γ₁₁, σ₀, and
σ₁ of each set of points. By mapping the 6D study space over the 3D space, 128 points of
the 6D space are mapped over each point of the 3D space ................................................................. 122

Figure 5-21 The 3D plot of the model-type changes along the γ₁₁, σ₀, and σ₁ axes. Colours from
blue to yellow show the suitability of fitting single-level models on the points. The darker the
point is, the single-level models fitted best in more of the 125 points of the 6D space that
were mapped on this coloured point in 3D space ................................................................. 123

Figure 5-22 The 3D plot of the slow change of model-types over the regions from two sides. The
blue region represents single level sub-region where 88% of the datasets simulated from
this region were fit best with the single level model (Table 5-15). The white region of the
study space represents the hierarchical region, where more than 95% of datasets simulated from this region were fit best with the hierarchical model (Table 5-15).

Figure 6-1 Research outline (Chapter 6 highlighted in red)

Figure 6-2 The process of developing and applying the model-structure selection tool

Figure 6-3 The steps of proposed model-structure selection tool

Figure 6-4 Example datasets’ properties of single-level and hierarchical sub-regions

Figure 6-5 Plots of the expected response variable under the model against the actual ones. Error bars shown are the 0.025 and 0.975 percentiles of the MCMC samples for each expected value. The dashed line shows the one to one line of a perfectly fitting model. In each model, n=200.

Figure 6-6 Density plots of all informative measures in different sub-regions. The first column of plots show the informative measures for single-level sub-region (black line) and for every hierarchical sub-regions (six coloured lines), the second column of plots show the informative measures for single-level sub-region (black line) and for the integrated hierarchical sub-region, which is built by joining all six hierarchical sub-regions together (red line).

Figure 6-7 The scatter plot for each pair of measures. The symbols in the plot show the six measures; sd.sd: standard deviation of within-group standard deviations over mean, vt: total variance, cov: coefficient of variation, vb: between-group variance, vw: within-group variance, vr: variance ratio. The blue points show hierarchical and red ones show single-level datasets.

Figure 6-8 The scatter plot for each pair of measures in the dataset and the five cluster groups. Datasets in the five cluster groups are represented by different colours: cluster groups 1 to 5 (in Table 6-6) are represented by red, blue, green, brown, and orange, respectively.

Figure 6-9 Flowchart of the proposed model-structure selection tool

Figure 6-10 The values of each measure for the three fish species datasets in density plots of the six measures for the single-level (black line) and hierarchical (red line) datasets. Green point, blue point, and grey point represent Anguilla reinhardtii, Melanotaenia duboulayi, and Tandanus tandanus fish datasets, respectively.

Figure 6-11 Histograms of predicted probabilities of each set belonging to their pre-known best fitting model structure

Figure 7-1 Research outline (Chapter 7 highlighted in red)
List of Tables

Table 2-1 Strengths and limitations of models................................................................. 38
Table 2-2 Model applications.......................................................................................... 39
Table 2-3 Applicability for management and control .................................................... 40
Table 4-1 WAIC average probability scores of the points with the WAIC scores of the seven datasets at the points............................................................... 83
Table 4-2 WAIC binary scores of the same points in Table 4-1; it is notable as the techniques of scoring are slightly different, the same point might be labelled differently by the two techniques.............................................................. 83
Table 4-3 Response and predictor variables for datasets of predictive and descriptive performance check; the predictor variables are the same in all datasets, it is because the goal is to find the relationship between the attributes of the points (coordinates in the study space) and each of the measure results........................................................... 90
Table 5-1 The PCR/ PLS regression coefficients for WAIC average probability scores .......... 102
Table 5-2 The explained variance from each component .............................................. 103
Table 5-3 The MLR standardised coefficients for WAIC average probability scores; as the predictor variables were standardised, the standard errors by the MLR were the same... 104
Table 5-4 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the WAIC average probability scores regression tree................................................................. 106
Table 5-5 MLR standardised regression coefficients of DIC scores; as the predictor variables were standardised, the standard errors by the MLR were the same. ....................... 107
Table 5-6 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the DIC scores regression tree........................................................................................................ 109
Table 5-7 The MLR standardised regression coefficients for average $R^2$ scores; as the predictor variables were standardised, the standard errors by the MLR were the same........... 110
Table 5-8 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the average $R^2$ scores regression tree ........................................................................................................... 111
Table 5-9 The MLR standardised regression coefficients for average RMSE scores; as the predictor variables were standardised, the standard errors by the MLR were the same... 112
Table 5-10 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the average RMSE scores regression tree .............................................................. 113

Table 5-11 The MLR regression coefficients for the descriptive $R^2$ scores; as the predictor variables were standardised, the standard errors by the MLR were the same. .................. 115

Table 5-12 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the descriptive $R^2$ scores regression tree. ........................................................................................................ 117

Table 5-13 The MLR standardised regression coefficients for the descriptive RMSE scores; as the predictor variables were standardised, the standard errors by the MLR were the same.............................................................. 117

Table 5-14 The 10 sub-regions defined by the leaves of the regression tree of the WAIC results, the associated values of attributes, and the probability of best fit of single-level models in every sub-region........................................................................................................ 125

Table 5-15 Probability of best fit of single-level models based on values of attributes, as defined by the regression tree of the WAIC results................................................................. 125

Table 6-1 The variables and their definitions used for each fish species (Stewart-Koster, 2011) .. .................................................................................................................................................. 143

Table 6-2 Variables used in models .......................................................................................... 145

Table 6-3 Single-level and hierarchical subregions based on regression tree of WAIC average probability scores as described in Section 5.2.1.1.3. In the first column, the number of the leaf in the regression tree corresponding to the sub-region is listed in brackets. .......... 147

Table 6-4 Number of datasets in every sub-regions before and after data cleaning process... 152

Table 6-5 Sample rows of the 49842 rows of the dataset used to do the model-based clustering. The first six columns are the measures upon which the datasets of the last column is to be clustered .......................................................... 156

Table 6-6 The abundance of single-level and hierarchical datasets in each class .............. 157

Table 6-7 The probability of each fish species belonging to the 5 cluster groups produced by the tool .................................................................................................................................. 161

Table 6-8 The number of datasets in each set defined in 6.2.7, and the number of datasets in each set for which the tool correctly recommended the model structure with better fit determined by the modelling results .............................................................. 163
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABMs</td>
<td>Agent-Based Models</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>analysis of covariance</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AREA</td>
<td>Total sampled area of reach</td>
</tr>
<tr>
<td>AVD</td>
<td>Average depth across reach</td>
</tr>
<tr>
<td>AVL</td>
<td>Average water velocity across reach</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>CA</td>
<td>Cellular Automata</td>
</tr>
<tr>
<td>CFD</td>
<td>Computational Fluid Dynamics</td>
</tr>
<tr>
<td>CVD</td>
<td>CV of daily flow of prior spring/summer</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>DIC</td>
<td>Deviance Information Criterion</td>
</tr>
<tr>
<td>DISTM</td>
<td>Distance of reach to river mouth</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximisation</td>
</tr>
<tr>
<td>HB</td>
<td>Hierarchical Bayesian</td>
</tr>
<tr>
<td>HLM</td>
<td>Hierarchical Linear Models</td>
</tr>
<tr>
<td>IBM</td>
<td>Individual-based Modelling</td>
</tr>
<tr>
<td>IO Models</td>
<td>Input-Output Models</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>MDF</td>
<td>Mean daily flow 4 months before sampling</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple linear regression analysis</td>
</tr>
<tr>
<td>NS</td>
<td>Navier-Stokes</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Principal Components Regression</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PLANTS</td>
<td>Percentage of reach area covered by plants</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Squares</td>
</tr>
<tr>
<td>PMF</td>
<td>Probability Mass Function</td>
</tr>
<tr>
<td>RMASS</td>
<td>Percentage of stream bank covered by root masses</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root-Mean-Square Error</td>
</tr>
<tr>
<td>WAP</td>
<td>Water Allocation Plans</td>
</tr>
<tr>
<td>WULUM</td>
<td>Water Use and Land Use Model</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to extend my heartfelt and sincere thanks to Associate Professor Peter Bernus and Dr. Ben Stewart-Koster for their endless patience, invaluable advice, encouragement and continuous support. Their support has been critical in getting through and more importantly getting to the end!

I would also like to acknowledge Associate Professor Ed Boone, my brother Dr. Hamid Shobeirinejad, and Mr James McBroom for their very generously sharing their expertise and experience in statistical matters.

I would like to express my sincere appreciations to my true mentor, my wonderful father Dr. Ali Akbar Shobeirinejad for his unconditional love, guidance and support, and to my dear mother Mrs. Sarieh Aboulian, for her eternal love and devotion. This thesis was not possible without her support.

Also special thanks to my sister Dr. Maryam, my brother Dr. Hadi for their affection, and to my parents- in-law for their kind help and care.

I would like to deeply thank my incredibly supportive husband Dr. Ahmad Tavassoli for his endless love and encouragement.

In addition, I published work arising from Chapter 2 of this thesis under the supervision of the co-authors on that paper, Associate Professor Peter Bernus and Dr. Jarrod Trevathan. This publication is included as Appendix 1. My contribution to the paper is outlined at the beginning of Appendix 1.
Publications during candidature


Keywords

Chapter 1
Introduction
1.1 Background

Sustainability is a key issue of concern, affecting the future wellbeing of humankind, both on the micro- and macro levels (Laboy-Nieves et al., 2008). It is the ability to endure and has many interpretations. In the environmental context, for example, it is a property that characterises whether the desired output of a project (or program) is assured without imposing permanent and unacceptable changes to the environment (Comm and Mathaisel, 2012).

It has been acknowledged (United Nations Secretary, 2012) that the ever-increasing connectivity and dynamic interactions among complex large-scale systems, including ecosystems, economic systems and social systems, pose a significant risk to a sustainable future. To pursue a stable, constant path towards a sustainable future of a system, making decisions at all levels of management, i.e., operational, tactical, and strategic levels, is the critical challenge. To exemplify, I refer to some examples of decision making at different management levels in a few cases taken from the Australian context, such as managing monthly groundwater extraction per bore in the Daly river catchment as an operational level management objective, Water Allocation Plans (WAP) decision procedure to set annual water extraction limits from Tindall Aquifer at Katherine (Pantus et al., 2011) and annual closure of lake Currimundi regarding the biting midge winter larval densities (Tomlinson et al., 2010) as tactical level management objectives, and providing future climate scenarios to enable determination of long-term average sustainable diversion limits to maintain the availability of the Basin water resources in the presence of the risks imposed by climate change (CSIRO, 2009) as a strategic level management objective.

Using models of systems to make management / control decisions is a well-established paradigm, which originated in the engineering field, and was later introduced into environmental management. Such models have the vital and determining characteristics of the system and represent the system’s elements, the interconnectivities and interdependencies of the elements, along with their interactions with elements inside or outside of the system. Much effort has been
undertaken towards modelling the environment to assess scientific hypotheses, explain a system’s behaviour in response to changes in system properties, and to predict the future state of a system (Wainwright and Mulligan, 2005). These models enhance the understanding of environmental systems at various levels of detail and support decision making for management and control.

To construct a model of the system, the modeller faces two concerns; firstly, what kind of model should be built, and secondly, how to build it (Uusitalo et al., 2015). In modelling, both a theory and specific information are required; the theory clarifies the type of model needed to answer the concerns at hand, and the information specifies the parameters of the actual model.

Modelling a system can be based on different motivations/objectives (Bertuglia and Vaio, 2005):

1. To explain ‘why’ the system behaves/acts in the way it does
2. To find out how to control the system’s behaviour in a way that the system doesn’t produce undesirable behaviour
3. To find out how to change the system so as to avoid the possibility of it producing undesirable behaviour in the future

This leads to create a model with the following characteristics:

1. The capability of the model to reproduce the system’s behaviour.
2. The capability to predict the behaviour of the system with the purpose of managing and controlling the system in order to achieve/keep an optimum function of the system.
3. The capability to design changes to the system in cases that simple controlling tasks are not adequate, e.g. when the future trajectory of the system is risky, when the safe path is tapering in the longer term future, or when the control process is unacceptably costly, and so on.

While finding one specific model of a system to answer all these queries is an ideal goal, the selection of the most suitable model among all types of models with regards to the modelling objectives and the available data is a challenging task for the modeller. Therefore, model selection is an elaborate step in modelling, which needs not only to comprehend the essential variables and
Chapter 1  Introduction

processes of the system, but also to avoid unnecessary complexity that doesn't improve the modelling results (Jorgensen and Fath, 2011).

1.2 Research problem

Uncertainty is an inevitable component of environmental management (as well as other complex systems’ management) and decision making; whereas, the significance of environmental problems urges the effort to quantify and, where possible, minimise the uncertainty of the outputs of the models of the system (Burgman, 2005; Uusitalo et al., 2015). Uncertainty about the model output stems from the uncertainty about the model input and the uncertainty about the model structure (O'Hagan, 2012). Uncertainty about the model input also arises from different sources, e.g., initial conditions, measurement, model parameters. Several sources of uncertainty are recognised (Regan et al., 2002) and solutions have been proposed and implemented to deal with the uncertainty of each source (e.g. Borsuk et al., 2004; O'Hagan, 2006; Laskey et al., 2010; Jolma et al., 2014). Among all sources, the uncertainty about the model’s structure is the most complex one and very difficult to deal with (O'Hagan, 2012; Uusitalo et al., 2015).

The model-structure selection comprises determining the cause and effect relationships among the variables (Uusitalo et al., 2015). It includes carefully investigating the amount and the quality of the available information, along with evaluating the uncertainty about the conditions, the variables and the modelling parameters, and then selecting one out of many model structures. For example, the modeller may decide to use deterministic models, or to embody some stochastic components in the model and assign some probability to the occurrence of some random variables. For the latter, they may decide to use a statistical model or a probabilistic model, such as Bayesian networks, which are based on the probability distribution/ function to represent the probability. They may decide to represent the whole system with one model of the system, or may choose to break the large-scale system down into simpler ones and use different models to represent the more clear cause and effect relationships among variables.
Among the many aspects of model structure for a modeller to select, there is the choice between a single-level and a multi-level model structure. Single-level models represent the relationships between variables with fixed coefficients, and, in case the data are grouped, ignore the group differences. Conversely, hierarchical models consider the relationships within and between levels of grouped data and account for the variation between groups (Woltman et al., 2012); hierarchical models use varying coefficients to explain the relationships between variables at one level, and represent the model for the varying coefficients at another level (Gelman and Hill, 2007). The complexity and heterogeneity of ecosystems urges the use of hierarchical models to accommodate more complexity and embrace different principles that might apply at different scales (Caswell, 1988; Clark, 2005). However, compared to single-level models, hierarchical models are considerable more complex and more difficult to implement (Hodges, 2010). Therefore, there is a strong need to understand the conditions where the additional work of fitting a hierarchical model is necessary.

1.3 Research Aims and Objectives

The principal aim of this research is to establish a robust methodology for model-structure selection between the two single-level and hierarchical structures, by investigating the characteristics of any available data and proposing which of the two model structures is more suitable. The additional aim of this research is to put forward a practical procedure as a model-structure selection tool to provide recommendations according to input data.

The statement of research problem is reflected in the following questions:

1. Under which conditions is the additional work of fitting hierarchical models necessary?

2. Which characteristics of data are determining in selection of the more suitable model structure?

Therefore, the overarching aim of this research is to investigate emerging quantitative methods available to ecologists and environmental managers and
develop a decision support tool to assist the modeller to choose between single-level and hierarchical model structures for a particular dataset.

To achieve the abovementioned aim, the research will seek to identify the statistical conditions under which hierarchical models might be necessary. This will involve the analysis of an empirical ecological dataset in tandem with a large simulation study. The simulation study will provide a way to analyse a large range of datasets with known structure, uncertainty and relationships among the variables, while the empirical study will provide an avenue to test the approach in a real setting with noisy data (Sokolowski and Banks, 2009). Based on the findings, the research will lead to develop a practical test that is efficiently applied to a set of data, with accompanying recommendations, to inform users when the additional effort of hierarchical modelling would provide a better model fit and when the simpler single level model structures are appropriate.

The stages of this research are:

1. To determine the study space
2. To run a simulation study over the study space
3. To evaluate the performance of the proposed single-level and hierarchical statistical models in a sample space across a range of statistical conditions
4. To determine the measure of fit upon which the guidelines for model structure selection will be set
5. To apply sensitivity analyses on the attributes of model structures to classify the study space into sub-regions where hieratical or single-level models fit best
6. To establish a model-structure selection tool for model development decisions
7. To validate the proposed guide tool

1.4 Thesis outline

This report is structured as follows:
Chapter 2 investigates fundamental mathematical models available to quantify environmental processes and guide decision making. It identifies the advantages and disadvantages of each approach and provides guidelines for selecting among the extensive array of methods available to environmental decision makers.

Chapter 3 reviews statistical models along with Bayesian methods as a very powerful collection of tools that can be used to quantify ecological processes, and guide decision making. It describes hierarchical Bayesian models as a great promise in quantifying multiscale processes and developing complex probabilistic models that reflect underlying ecological processes. It also serves
to narrow the focus of subsequent research in the thesis, and identifies a gap in knowledge that is to become the specific research topic of this thesis.

Chapter 4 describes the proposed methodology of this research. A large simulation study is conducted and the detailed method employed is determined and described in this chapter. The simulation study provides a way to analyse a large range of datasets with known structure, uncertainty and relationships among the variables. According to the models’ parameters, a study space is defined, and single-level and hierarchical models are developed on all points throughout the study space. A number of goodness-of-fit measures are employed to evaluate and compare the models’ predictive and descriptive capacities.

Chapter 5 provides a comparative analysis of simulation study results. Consequently, some subspaces in the study space are identified as the regions in which single-level modelling is adequate, and in some other regions hierarchical modelling is recommended. In this chapter, statistical techniques are used to calculate measures that distinguish the subspaces. Sensitivity analysis is applied to inspect the impact of model parameters on the subspaces of the study space.

In Chapter 6, I develop the model-structure selection tool to recommend the best fitting model structure for any specific set of data. The tool is applied on three fish species datasets to demonstrate its applicability, and the tool’s recommendation is compared to the modelling results. In the end of the section, the advantages and limitations of the tool are discussed, and future research is suggested.

Chapter 7 summarises the findings of this research and discusses the implications and limitations of the study.
Chapter 2
Quantitative models for environmental management – Mathematical models
2.1 Introduction

In the field of ecology, intense research has been done during the last three decades in order to set up practically applicable models to represent complex macro- and micro ecosystems. These ecological models also have been applied in environmental management at all levels (Jorgensen and Fath, 2011). This has led to the abundance of various types of models with different names, intended to solve a wide spectrum of problems in this field, and categorised by application area (Jorgensen et al., 2011). By today, the selection of the appropriate model-type for a desired management goal has become a complex task itself. At least part of this complexity problem arises from the fact that almost all the modelling efforts in the field of environmental management have been done with specific ecological points of view, by experts who have faced very specific problems of particular ecosystems. However, as mathematics is the common base of all of these modelling efforts, it is a useful effort to discuss and analyse the basic model categories using their mathematical properties, to pinpoint the potentials and pitfalls of models applications in environmental management.

This section presents a survey of fundamental mathematical models employed in environmental modelling, and explores the strengths and weaknesses of each model type (Figure 2-1).
2.2 Mathematical Models

In reality, we mostly focus our attention on the system’s evolution over time (i.e., the system’s dynamics) and attempt to predict its future behaviour. A model is a simplified representation of reality; it never contains all features of the real system, but to be satisfactory, it has to represent the main important features in the content of the problem to be solved. As the system gets more complex, it gets more complicated to ascertain its essential features regarding the problem. When modelling a system with the purpose of parametric optimization of the system’s dynamic behaviour, the concept of the system’s state is fundamental; meaning that given the current situation of the system including its dynamic and
static properties, we characterise a predefined desired state of the system by some appropriate function, which expresses many measuring points, and propose the actions leading the system towards this goal. In describing a system and its dynamics, mathematics is the most qualified language. Indeed, the idea that the ‘book of nature’ is written in the mathematical language was established a few centuries ago (Bertuglia and Vaio, 2005) and the mathematical modelling is a discipline which plays a basic role in describing the dynamics of complex ecological, economic and social systems.

This study categorises the different types of applied mathematical models into four basic mathematical models, namely: Input-Output Models, Agent Based Models, Artificial Neural Networks and Bayesian Models. It is important to mention that no separate overview of fuzzy logic models was attempted due to the fact that this can be considered as a modality of multiple model types.

2.2.1 Input-Output Models (IO Models)

Input-Output models are a category of models describing the input/output behaviour of systems. The input/output view of systems originated in electrical engineering, where the design of electronic amplifier led to a focus on input/output behaviour. This way, the system was considered as a device that transformed inputs to outputs. Therefore, an IO model was conceptually a large table of inputs and outputs (Aström and Murray, 2010).

The IO concept is widely used in many engineering systems; it allows decomposing a complicated system into individual manageable components, each having a set of inputs and outputs. The concept of IO modelling promptly spread in other fields, such as economic and ecological modelling. In economy, the basic input-output (I/O) model presented by Wassily Leontief demonstrates the flow of commodities between economic sectors in a modern economy. It is suitable for studying interdependencies. The logic of the I/O model can be described by the following:

\[ \text{Total Input} (X_{\text{Input}}) = \text{Total Output} (X_{\text{Output}}) \]

\[ \text{Total Output} = \text{Output to intermediate Demands} + \text{Output to Final Demands} \]
Intermediate demands are the ones of the intermediate sectors of the same system and the final demand is the one of an end user (Haimes et al., 2005). This concept is basically very similar to the Kirchhoff’s current law proposed a century earlier in the field of electrical engineering (Paul, 2001).

The mathematical apparatus used to construct IO models, commonly used in control theory, are differential equations and finite difference equations, both of which have been developed to describe dynamic processes. By setting these equations to zero, the model represents the static situation, which would be of benefit for the time the data quality and quantity are not sufficient to develop a complete dynamic model, or when the modelling objective is to investigate a worst-case situation (Jorgensen and Fath, 2011). This research categorises all developed models based on differential or finite difference equations as IO models.

Generally, the model state is measured by certain magnitudes (state variables) that are dependent on time, and the model describes the system’s evolution over time using differential equations, when the time is considered a continuity, or finite difference equations, when the time varies in discrete intervals (Bertuglia and Vaio, 2005).

The following set of differential equations describes a dynamic system with a set of \( n \) state variables in continuous time:

\[
\frac{d}{dt} x_i = F_i(x_1,\ldots,x_n,t) \quad i = 1,\ldots,n
\]

The dynamic system in this case is called a flow. On the other hand, the system can be described by the following set of finite difference equations, which show the evolution of the system over discrete time intervals:

\[
x_i(k+1) = f_i(x_i(k),\ldots,x_n(k)) \quad i = 1,\ldots,n
\]

In this second case, the model of the system is called a map (Bertuglia and Vaio, 2005). Some examples of the most common IO models are briefly reviewed below.
2.2.1.1 Oscillating Pendulum

An oscillator is one of the simplest and most important models which has been a basis for further extended and developed models in other sectors such as some macro scale ecological or economic systems that experience periodicity and oscillation over time. The application of differential equations in clarifying the behaviour of such systems in both linear and nonlinear aspects is notable. Therefore, the simple basic pendulum model is discussed here. Figure 2-2 shows the pendulum at its start point where no friction is considered.

![Diagram of pendulum forces at start point with no friction.](image)

The differential equation describing the behaviour of this system is

\[
\frac{d}{dt^2} \theta = -\sin \theta \cdot \frac{g}{L}
\]

where \( g \) is the gravitational acceleration.

The above system of differential equations is nonlinear as it includes the term, \( \sin \theta \), however, given that \( \theta \approx \sin \theta \) for small angles, in a subspace of the pendulum’s state space the system can be modelled using a system of linear differential equations.

This is a typical case, illustrating that after approximations and assumptions, it is either possible to explicitly solve the equations, or if this is not possible, then numerical techniques must be applied. It is, however, an important question, that every modeller must be able to answer: is the so derived (approximate) solution adequately accurate and in what subspace of the systems state space is this the case?
2.2.1.2 Predator-Prey

The classic theme of the predator-prey model is a model of two interacting populations in an ecosystem where one feeds on the other. This type of system has been studied over decades and widely applied in a variety of areas. This is an example of a case in which the evolution of the system is described at discrete time intervals rather than continuously in time. Such systems are called discrete time systems and are described by difference equations.

The general fundamental principles of the basic model (called the Volterra-Lotka model) are still used as a reference for modelling in mathematical biology as well as other fields such as urban and regional science (Bertuglia and Vaio, 2005). The original form of the model assumes unlimited food supply for the prey and therefore no internal competition for that species, while the predator species in this model compete over the finite population of the prey species. The equations of the basic model are:

\[
\begin{align*}
  n_1(k+1) &= k_{11}n_1(k) - k_{12}n_1(k)n_2(k) \\
  n_2(k+1) &= k_{21}n_1(k)n_2(k) - k_{22}n_2(k)
\end{align*}
\]

Where

- \( n_1 \) is the number of members of preys
- \( n_2 \) is the number of members of predators
- \( k_{11} \) is the growth rate of the prey
- \( k_{12} \) is the mortality rate of the prey due to predation
- \( k_{21} \) is the reproduction rate of the predators by consuming preys
- \( k_{22} \) is the mortality rate of the predators

The two populations interact in a nonlinear way as the term \( n_1n_2 \) represents the encounter frequency between a prey and a predator which is proportional to the product of the prey population multiplied by the population of predators (which in itself is an approximation assumed to be admissible without influencing the model’s predictions in a substantial way).
2.2.1.3 Hydrodynamic Systems

Hydrodynamics is the study of motion of liquids, in particular water, and heat in oceans and coastal seas in order to determine the nature of marine ecosystems. The models of this type are also widely used to predict the transport of sediment. Hydrodynamic modelling is also used in meteorology, aerospace and automotive design, ventilation systems, and so on.

The common basis of these models is the numerical solution of differential equations governing the transport of mass (the law of conservation of mass), transport of momentum (Newton’s second law of motion) and transport of energy (the first law of thermodynamics) in moving fluids. The Navier-Stokes (NS) equations are the fundamental equations of fluid motion (Date, 2005) and are the basis of almost all hydrodynamic models. In all but the simplest cases an explicit solution of the NS differential equations is not possible; therefore, it is necessary to use finite difference equations, where the numerical solutions are acquired at discrete points rather than all points as in the continuous space. As a consequence, Computational Fluid Dynamics (CFD) problems are normally solved using finite elements methods (Glowinski and Pironneau, 1992), whereupon the space is subdivided into very small areas, and the behaviour of the fluid (flow) is described by a separate equation for each area. This approach emerged and has been widely used with the availability of computers since the early 60s. Today the method is extensively used in basic and applied research, in design of engineering tools, and in environmental models (Date, 2005).

In hydrodynamic models, it is a basic assumption that the model can be discussed at any desired precision with the help of finite difference equations, and also that the finite element mesh created to represent the model is extremely accurate. While such assumptions would be true in small-scale engineering applications (e.g., sewerage system pipelines) large-scale environmental cases may be totally different (e.g., a river catchment). This makes a significant difference for the model’s ability to make predictions. For example, an engineering hydrodynamic system is designed using mathematical equations of the surfaces enclosing the volume of liquid, therefore, when finite element approximation is necessary, this approximation can be as fine as
needed and constitutes a completely predictable 3D surface. As opposed to this, the accuracy and predictability of an environmental hydrodynamic model is more limited due to factors such as uncertain/fewer measurements and lower resolution and possibly due to natural changes to the shape and volume of the enclosing 3D surface. This state of affairs is not something that can be easily changed; therefore, additional methods are necessary to validate such models (such as validation against observable and measurable properties of the natural system).

2.2.2 Agent-Based Models (ABMs)

Individual-based Modelling (IBM) or Agent-based modelling (ABM) is a relatively new approach to modelling the dynamics of complex systems composed of interacting, autonomous agents. Agents have behaviours and interactions with other agents and with their environment, which in turn influence their behaviours. By modelling individuals as agents, the effects of the diversity that exists among individuals in their behaviours and attributes could be represented and observed. In such models, patterns, structures, and behaviours, that are not explicitly programmed into the models but arise through agent interactions, would emerge (Macal and North, 2010). The structure of a typical agent-based model is shown in Figure 2-3.

![Figure 2-3 The structure of a typical agent-based model (Wooldridge, 1997)](image)
ABMs can be considered as the extension of cellular automata (CA). CA are typically composed of cells on a two-dimensional grid. Each cell can be interpreted as an agent that interacts with a fixed set of neighbouring cells. The cell state is either ‘on’ or ‘off’ at any point in time. Cells change their states every finite time step, following simple rules of behaviour. The next state of the cell depends on the cell’s current state and the actual states of its neighbours. A CA is deterministic so that the same state for a cell and its neighbours always results in the same future state (Macal and North, 2005).

Agents are entities, which can be considered having a set of goals and capable of autonomously performing actions in a dynamic and unpredictable environment in order to meet their goals. They encompass heterogeneous behavioural instructions and are adaptive in terms of learning and modifying their behaviour based on their accumulated experience (Bandini et al., 2009). Figure 2-4 shows an agent and its properties.

![Figure 2-4 An Agent and its properties (Macal and North, 2005).](image)

An ABM consists of a set of agents, a set of agent relationships, and a topology for simulating agent behaviours and interactions. This topology defines the neighbours and the mechanisms of dynamic interactions. It usually includes a spatial grid or network of nodes (agents) and links (relationships). In complex systems, agents do not interact with all agents all the time, and local information is obtained from interactions with an agent’s neighbours and from the agent’s local environment. The set of local information resources of an agent can
change rapidly over the agent life (Weisbuch and Ryckebusch, 1991; Kauffman, 1993; Macal and North, 2010).

ABMs allow the explicit consideration of the structural heterogeneity of the components along with their spatiotemporal variation. They can represent individual behaviours and can model the population’s response as an emergent property. Therefore, these models are capable of realistically representing the overall dynamics of natural systems, and help a modeller to understand the role and importance of the individual in shaping the overall system (Macal and North, 2010).

2.2.3 Artificial Neural Networks

The first discussions on Artificial Neural Networks (ANN) began by a simulation of how neurons might work in the brain with a simple model of neural network implemented using electrical circuits. An ANN is an interconnected group of artificial neurons (processing units), which communicate by sending signals to each other over a large number of links, where each link has a numeric weight. Each node has a set of input links from other nodes, a set of output links to other nodes, a current activation level, and an activation function to compute the activation level in the next time step (Gupta et al., 2003).

The system is inherently parallel meaning that many units can perform computations at the same time. There are three types of units in a neural network; the input units, which receive data from outside of the system, the output units, which send data out of the system, and the hidden units, whose input and output signals remain within the system (Hu, 2000). Figure 2-5 shows a schematic of a neural network and a neuron structure.

The current activation level for every unit is equivalent to the output of the unit, a propagation rule determines the effective input of the unit from its external inputs, and an activation function \( f \) determines the new level of activation based on the effective input and the current activation (Basheer and Hajmeer, 2000).
ANNs have the ability to learn and to generalise; they learn how to perform their function on their own and determine their function based only upon sample inputs, and can produce reasonable outputs for new inputs.

Generally ANNs use some sort of threshold function as their activation functions, such as a sign function, a hard limiting threshold function, a linear or a sigmoid function (Figure 2-6) (Kröse and van der Smagt, 1996).

ANNs are classified into two categories based on the pattern of connections between the units and the propagated data (Kröse and van der Smagt, 1996):

- Feed-forward networks; data flow from inputs to outputs is forward and there is no feedback connection in the network, examples of feed-forward networks are the Perceptron and Adaline (Gupta et al., 2003).
- Recurrent networks; networks contain feedback connections, such as networks presented by Anderson (1977), Kohonen (1977), and Hopfield (1982).
Training is the act of presenting the system with inputs from some sample dataset and modifying the weights in order to achieve a better approximation of the desired output function. The learning process can be categorized in two distinct types:

- Supervised learning or Associative learning
- Unsupervised learning or Self-organisation

In supervised learning, the neural network is supplied with inputs and their desired matching outputs, and the weights are modified to reduce the difference between the actual and desired outputs.

On the other hand, in unsupervised training, only the inputs are supplied and the neural network adjusts its own weights in a way that similar inputs cause similar outputs. In this kind of network, the system is supposed to discover statistically important features of the input population, and identifies the patterns and differences in the inputs without any external assistance (Gupta et al., 2003).

One of the most common neural networks is the MLP, which is an extension of the perceptron, using a supervised learning technique called backpropagation for training the network. MLP has two important characteristics: being not limited to linear problems (the logistic function and the hyperbolic tangent are most widely used as its function); and its massive interconnectivity (any element of a given layer feeds all the elements of the next layer) (Hung et al., 2009).

Backpropagation is the most common method of obtaining the many weights in the network. The basic backpropagation algorithm is based on minimizing the error of the network using the derivatives of the error function. The method is simple, but slow. In this technique, the main idea is that the errors for the units of the hidden layer are determined by back-propagating the errors of the units of the output layer, hence, the method is called backpropagation. Choosing the appropriate learning rate is important, if it is too small, the convergence would be extremely slow, while having a too large learning rate the network may not converge at all.
As a popular unsupervised learning network, we can refer to the Kohonen network, which is an extension to the competitive learning network. Neurons in the Kohonen layer sum all the weighted inputs received, and the neuron with the largest sum outputs a 1, while the other neurons output 0 (Kohonen et al., 2000).

### 2.2.4 Bayesian Models

Reasoning under uncertainty is crucial for accurate analysis, synthesis, prediction, inference, and decision making. The sources of uncertainty include ignorance (e.g. which side of this coin is up), complexity (e.g. meteorology), physical randomness (e.g. which side of this coin will land up), and vagueness (e.g. which tribe am I closest genetically). In such reasoning, the significant role of Bayesian network is revealed. Bayesian networks were first introduced by Pearl in 1988. They represent causal assertions between variables as patterns of probabilistic dependence, and therefore, can be used for logical and holistic reasoning about complex systems (Borsuk, 2008).

A BN can be used to represent knowledge and observational inference. It is a way to distinguish deductive knowledge (e.g. mathematics) from inductive belief (eg. science). A BN is a compact representation of a joint probability distribution over a domain of variables of interest (Ken and Andy, 2011).

In reasoning under uncertainty about a set of events, it is possible to observe that some events are causally related to other events with certain probability. Conditional independence (the lack of direct dependency relationship) plays a very important role in a Bayesian network’s structure, as probabilistic inference can be made on the basis of information about the direct antecedent variables.

Here I present some axioms as the bases of BN concept:

**Probability (Kolmogrov’s) axioms:**

1. \( P(U) = 1 \)
2. \( \forall X \subseteq U, P(X) \geq 0 \)
3. \( \forall X, Y \subseteq U, \text{ if } X \cap Y = 0 \text{ then } P(X \cup Y) = P(X) + P(Y) \)
Conditional probability:

\[ P(X|Y) = \frac{P(X \land Y)}{P(X)} \]

Independence:

\[ X \perp Y \iff P(X|Y) = P(X) \]

and Bayes theorem is mathematically stated as:

\[
P(h|e) = \frac{P(e|h)P(h)}{P(e)}\]

\[
\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Probability of evidence}}
\]

where:

- \( h \) is the hypothesis,
- \( e \) is the evidence (the new data, that were not used in computing the prior probability),
- \( p(h) \) is the prior probability; the estimate of the probability of the hypothesis \( h \) before having the evidence (new data) \( e \),
- \( p(h|e) \) is the posterior probability; the probability of hypothesis \( h \) after observing the evidence \( e \),
- \( p(e|h) \) is the probability of observing the evidence \( e \) given the hypothesis \( h \) (the likelihood),
- \( p(e) \) the probability of the evidence \( e \).

A Bayesian Network is a Directed Acyclic Graph (DAG) (i.e. no directed cycles) in which the following holds:

- A set of nodes which represent variables.
- A set of directed edges or links represent probabilistic dependence between variables (dependence may mean mere correlation, or may be the representation of some type of causality based on physical law);
- Each node has a conditional probability table that quantifies the effects the parents have on the node. Note that these conditional probabilities in
environmental applications typically represent statistical findings about observed/measured data, but may also be based on expert opinion, or could even be the result of analysing a separate fine grained model (of appropriate type) of a particular aspect of the system of interest.

Nodes can represent discrete or continuous variables. Nodes without any parents (immediate predecessors) are called roots, and are described as unconditional (marginal) value distributions. Nodes without any children (immediate descendants) are called leaves and the non-leaf, nonroot nodes are called intermediate nodes.

In the absence of any explicit connecting edge between two nodes in a BN, the two variables are assumed independent, given the values of any intermediate nodes. In other words, the probability distribution of any variable in any state of the network can be determined by knowing only the values of its (immediate) parents, with no need to know the values of any other variables (Borsuk, 2008). This is referred to as the Markov property.

Figure 2-7 illustrates the BN of a medical diagnosis example. In this case, the factors affecting a patient’s chance of having cancer are ‘pollution’ and ‘smoking’. In the same way, cancer can be a cause of ‘dyspnoea’ and a ‘positive X-ray result’. As also shown in this figure, discrete nodes can be Boolean (e.g., cancer node with true or false value), ordered values (e.g., pollution node with low, medium, and high values), or it can also have integral value (e.g., age node with possible values from 1 to 120 not included in this example).
2.2.4.1 **Reasoning with BN**

There are four main types of inferences with Bayesian networks:

- Predictive or causal inferences (from cause to effects),
- Diagnostic or evidential inferences (from effects to cause),
- Inter-causal inferences (between causes of a common effect), and
- Mixed or combined inferences.

Figure 2-8 shows the four types of inferences.

When the structure and conditional probabilities of a BN have been specified, the network can be used to determine the probability distributions of specific target or query nodes, given findings (either deterministic or probabilistic observations) for other nodes (evidence nodes). When the query nodes are descendants of the evidence nodes, the process is called prediction. When they are ancestors of evidence nodes, it is called diagnosis.
BNs can also be used to determine the most probable explanation of the particular values for some variables, to explain away the alternative causes of an effect, to describe the effects of interventions (or external controls) on the system, and to support decisions about management actions in the face of uncertainty (Borsuk, 2008).

2.2.4.2 Extensions to BN: Bayesian Decision Network

BNs can be extended to support decision making. When it comes to decision making under uncertainty, the goal is to find out what action to take (plan to adopt) when the future state of the world is not known. A Bayesian solution would be to find the utility of each possible outcome (action-state pair) and take the action that maximizes expected utility (determined by the decision maker’s preferences regarding possible outcomes of various plans). This way, decision networks are produced as an extension of BNs with decision nodes and utility nodes to support decision making.

2.3 Applications, Strengths/ Limitations, Comparison, and Applicability of Different Model types in Decision Making

The previous section described four fundamental, and generic, mathematical model types that are the basis of most mathematical models being used today in environmental decision making. In this section, these model types are evaluated regarding several essential aspects of modelling; their functionalities and deficiencies are compared, and their determining attributes in the process of model selection for a particular case with specific goals are discussed.

2.3.1 Applications of Models

This section briefly describes the applications for each of the four types of models.

2.3.1.1 Applications of Input-Output Models

IO models are widely used in ecosystems and environmental management applications such as modelling water quality (Ke et al., 2016), water pollution
(Mannina and Viviani, 2010), fisheries management (Steinback and Thunberg, 2006), management of natural resources and natural parks (Emeo, 2012), and for modelling the effects of chemicals and toxic substances (Jorgensen and Fath, 2011).

2.3.1.1 Example

Pantus et al (2011) is an example in the Australian context; it is a report of the managing framework of an Australian river catchment. The complete system is a large complex environmental system which is made up of several subsystems and their interrelations. Due to the quality and quantity of available data, each subsystem is modelled using a separate kind of IO models such as the catchment water model, the groundwater extraction model, the habitat models and economics model. The authors have a management point of view and use the output of each model in a particular step of the management process. The aim of research is to facilitate the “what-if” scenarios evaluations for the managing communities and policy makers.

2.3.1.2 Applications of Agent-Based Models

The early use of ABMs in ecology goes back to a forest model (Botkin et al., 1972) and a fish cohort model (Deangelis et al., 1980). Other examples of applications of ABMs include: predicting the spread of epidemics (Bagni et al., 2002) and the threat of bio-warfare (Carley et al., 2006), modelling the adaptive immune system (Folck et al., 2007), understanding consumer purchasing behaviour (North et al., 2010), understanding the fall of ancient civilizations (Kohler et al., 2005), modelling the engagement of forces on the battlefield (Moffat et al., 2006), and traffic modelling (Helbing, 2001).

2.3.1.2.1 Example

Water Use and Land Use Model (WULUM) (Zellner, 2007) is an example of an ABM implemented to investigate the relationship between land use, water use and groundwater dynamics in the Monroe County in Michigan. In this case, the exact relations linking land-use and groundwater dynamics are not clear, therefore, the main objective of this modelling is to link these processes in an
integrated model, evaluate their effect on groundwater resources, and make policies towards the resource sustainability. The model consists of two integrated components: one representing the land-use processes and the other, the water-use processes and the groundwater dynamics. The land-use component includes the four main water-extracting sectors: residents, stone quarries, golf courses, and farmers. A lattice is used for the landscape, where each cell contains information including groundwater, forest, soil quality, roads, zoning restrictions and municipal water coverage.

The water component consists of two layers: the glacial deposits overlaying the bedrock aquifer. Cellular automata transitions rules are used to model the regional groundwater gradients with points of recharge and discharge. Under the same principle, farm cells may change to residential or golf locations, depending on land-use policies for location, existing development and zoning. The authors conclude that there are non-linear relationships between land-use diversity and intensity and groundwater levels that depend on the interaction between natural, policy and infrastructure variables.

2.3.1.3 Applications of Artificial Neural Networks

ANNs have been successfully applied to a broad spectrum of data-intensive applications. They have been employed to create models to predict rainfall-runoff (Hsu et al., 1995; Izadifar and Jahromi, 2007), stream flow (Zealand et al., 1999; Abrahart and See, 2000), or be used for groundwater management (Rogers and Dowla, 1994), water quality simulation (Maier and Dandy, 1999), rainfall forecasting (Hung et al., 2009), weather forecasting (Mandal and Prabaharan, 2006). ANNs are also useful for pattern recognition and classification, such as fish species identification (Lawson et al., 2001; Cabreira et al., 2009).

2.3.1.3.1 Example

(Zhang et al., 2012) uses an ANN to develop a model to identify bearing faults in wind turbines. The model is able to predict the fault in an average length of 1.5 h prior to its occurrence. The output of this research is of high efficiency as it can prevent costly damages to some high-value turbine components. The data
is collected from 24 wind turbines with ten second intervals for a period of four months. The types of data to be analysed were first selected upon experts’ knowledge, reducing from a hundred to fifty parameters, and then with applying three different data-mining algorithms the eighteen most relevant parameters were found and used in models. Five different NNs were selected and optimised as the best among a hundred randomly selected NNs. The number of neurons in hidden layer was in the range of 5 to 25 and the functions of tanh, exponential, identity, and logistics were used as activation functions in hidden and output layers. The results showed that the data with 10 second intervals were of too high-frequency and an interval of an hour in data collection would be sufficient for the purpose.

2.3.1.4 Applications of Bayesian Models

The application of Bayesian models for forward prediction is an obvious application of such models. However, another significant application of these models is in performing probabilistic inference, or diagnosis. Examples of application areas include weather forecasting (Madadgar and Moradkhani, 2014; Avilés et al., 2016), medicine (Mascaro et al., 2014; Rabinowicz et al., 2017), speech recognition (Yildiz et al., 2013; Norris et al., 2016), and robotics (Rubio et al., 2014; Chung et al., 2015).

2.3.1.4.1 Example

Hamilton et al. (2007) applies a BN to model an environmental problem in an Australian coastal area. The area has recently experienced an increase in the frequency and extent of an algal bloom with having a range of adverse effects. The size and the probability of the blooms occurrences are the result of a number of interactions between some key factors causing the bloom. To mitigate the risk of the occurrence of the problem, these main factors need to be identified. In this case as in many other complex environmental systems, the exact causes of the problem are not well understood, and the main source of information is the knowledge of experts. The BN model is used to integrate all the diverse knowledge from different fields. They have found that while the model is built in a local scale, it can be generalised to other cases. Besides,
upgrading the model with new data is easy, and furthermore, the model probabilistic framework makes it capable of dealing with the uncertainty inherent in complex environmental problems. According to the authors, it is an effective method to assess risk factors and a basis to produce appropriate management actions. However, in such a large BN model, investigating and comparing all possible scenarios is a much time-consuming and too complicated process.

2.3.2 Strengths/Limitations of Various Models

2.3.2.1 Input-Output Models

Input-Output models are based on causality. A fundamental concept of IO models is that they are mostly based on conservation principles (physical laws / laws of nature). They are easy to understand and develop. In the case of linear systems, the model is able to predict the accurate future behaviour of the system, and the situation of the system at any point in the space of states can be perfectly discussed.

However, having an explicit and direct linear relationship between cause and effect is rare in the natural world, if not impossible; in practice, linear models are the result of linearization of a problem, in which the approximations are supposed to be slight and the ignored factors and parameters to only have trivial effects. As a result, linear models are only realistic, and the mathematical approximations satisfactory, if the system is close to a stable equilibrium (Bertuglia and Vaio, 2005).

Therefore, even though the linearization process might operate acceptably for the status quo, the slight ignored factors may result in very large unpredictable effects in the longer future. Besides, the accuracy of the input data is very critical to obtain reliable outputs from mathematical equations. This means that IO models require an accurate set of state variable (data) measurements and any small ignored measurement errors have the potential to yield consequential inaccuracy in long term future predictions. Both of the aforementioned types of difficulties of IO models may result in either quantitative or qualitative errors in the long-term predictions of the system’s behaviour.
There is a third type of difficulty regarding a system’s behaviour prediction which is related to the nonlinearity of many systems. Nonlinear models may at times behave in a way that gives rise to unstable evolutive trajectories which are completely unpredictable (Bertuglia and Vaio, 2005). This happens when a dynamical system is highly sensitive to initial conditions; a small difference in initial conditions produces widely diverging outcomes of the system. Even though the system is deterministic and no random element is involved, the future behaviour of the system can still be unpredictable. This behaviour is known as chaotic behaviour. In terms of the system’s space of states, nonlinear equations can only be used safely to manoeuvre in a limited subspace of the large space of possible states, and therefore, the model is not necessarily valid for a possible future situation of the system when the system stands at a point far from the subspace that is known to be accurately described by the model. For the case of chaos (a chaotic subspace of the system’s space of states), a small change in the present state of the system may result in an unpredictable future state of the system, which means that even if the nonlinear model is truly representing the present state of the system and its vicinity, the only sincere prediction it can make for the system’s future behaviour would be its unpredictability. In practice, the three aforementioned issues, alone or together, limit the predictive ability of these types of models.

Of particular interest can be the use of control systems theory (Lin, 2007), to map the ‘troublesome’ areas of the state space, and allows the consideration of alterations to the system to change the location of chaotic subspaces if the system’s trajectory is desired (or predicted) to trend toward such areas.

It is important to note that while a system might be capable of producing chaotic behaviour in certain sub-space(s) of its state space, the system may still be predictable from many respects, and/or the system may still be controllable (Lin, 2007), provided that the control of the system steers the system’s trajectory so as to avoid such chaotic areas of the state space.

A further weakness of this type of model is the (practical or theoretical) difficulty of describing the complete complex system in a single model; instead, it is often necessary to view the complex system as a ‘system of (sub)systems’,
whereupon it is feasible to explain each subsystem using a type of IO model. One reason for this can be the lack of theory (knowledge of natural laws) connecting multiple aspects of a system. Therefore, a large complex system might have to be (and in environmental modelling often is) described by a number of different models, each explaining a part of the system, plus a further ‘super-model’ required to explain the interconnections among these subsystem models. Note that explaining all sub-models and their relations might not be feasible using differential/ difference equations and/or this type of model would not perform appropriately as the super-model. Besides, such a super-model, if possible, would carry all the limitations of all sub-models together. Also the same applies to ‘systems of systems’, where the contributing systems are not ‘subsystems’. The same problems may arise, but in even more accentuated manner, because the contributing systems have their own independent life trajectories. Such systems of systems may be better approached using agent-based models (see Section 2.2.2, however, individual agents may still be modelled using IO models, or indeed other suitable models types).

2.3.2.2 Agent-Based Models

Agent-Based Models are used when differences among individuals of the same species are of importance for understanding system behaviour. Therefore, ABMs represent properties of individuals, such as adaptation and spatial distribution. ABMs also represent the behaviour of a population of individuals. It is possible to use these models to explain and predict emergent behaviours in simple cases, whereupon the emergent behaviour is the result of the behaviour of individuals.

However, the larger the number of properties considered in modelling, the more complex the ABM would be. Correspondingly, a significant amount of data is required for model calibration and validation (Jorgensen and Fath, 2011). In dealing with complex systems, these models are significantly more demanding to develop than IO models regarding data requirements, technical skills, computer power, and development time. Furthermore in very large populations, the predictive ability of this type of model is very vulnerable as the success of ABMs depends on the precision of representing the agents and their
interactions, so it is possible that just small differences in the level of individual
detail can cause quite different emergent behaviour, and the validation question
becomes how do we know whether this is or is not the case?

2.3.2.3 Artificial Neural Networks

Artificial Neural Networks can be applied to a heterogeneous database or
observed / measured data to learn the system’s behaviour. As an ANN
simulates the behaviour of the system as a black-box, the possibly very
complicated internal effects and processes within the system do not need to be
clearly investigated, understood or evaluated. ANNs can be easily applied when
explaining the relations between the data with IO models (e.g. some differential
equations) seems impossible. This gives an advantage to ANNs as they take
into account all available data with no need to make approximations. ANNs can
be easily applied to a system without ignoring any supposedly trivial
information. Therefore, this type of model can enhance the accuracy and
reliability of the model's output.

We conclude that the above property improves the ability of making predictions
using these models. Assuming we have a sufficient amount of data
(measurements / observations) about the subspace of states in which the
system is currently manoeuvring, an ANN is able to predict well the possible
further situations of the system. This would normally be also true when the
system is heading at new points of the space of states not far away from the
known subspace, however, there is no guarantee that this is the case. This is
because a system’s behaviour may have been known in an observed subspace
of state variables, however, some points in that subspace may be close to a
hitherto unobserved tipping point, which fact could therefore remain
unpredictable by an ANN. This indicates the predictive ability of this model type
being limited to short-term and medium-term future.

Note the fact that ANNs are black-box models and since the learning process is
not about discovering causality-based facts, which imposes some restrictions to
the applicability of this type of models. In addition in some steps of ANN model
building, such as the ANN type selection, data type and parameter selection,
decision about the number of hidden layers and neurons, and the selection of
types of activation functions, expert opinion plays an important role – although
using different algorithms to choose the most influential data can be helpful in
some cases there is always some doubt about the repeatability (by others) of
the modelling exercise.

2.3.2.4 Bayesian Networks

Bayesian Networks Inference is very flexible, as it can enter evidence about any
node and update beliefs in any other nodes, which means that upgrading the
model with new data is easy. Furthermore, the model’s probabilistic nature
makes it appropriate to be utilised in a wide variety of cases, where uncertainty
dominates considerations and it is crucial to ‘get it right’.

However, we realise that a BN model is very sensitive in terms of the changes
to the framework: new findings and data might impose structural changes,
destroy the network’s hierarchical relations and may require the development of
a completely new BN with a different structure.

2.3.3 Comparison of the Different Models

Table 2-1 lists the strengths and limitations of the four model types. Table 2-2
shows the most common cases in terms of the application of each type of
models. As it shows, there are many common applications among different
model types. This is due to the fact that environmental management plans
almost always cover local case studies and a similar problem may be treated
very differently in different situations. The reason is that in large complex
systems, there are many interactions and interdependencies among the many
elements of the system, and a similar issue in different situations can result
from and be affected by completely different factors. Therefore in selecting a
modelling technique for managing every situation, there are many influential
factors to be considered rather than just the problem; such as the type of data,
the future perspective of the system, the desired output of the system, and so
on.
Table 2-1 Strengths and limitations of models

<table>
<thead>
<tr>
<th>Model</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| IO Model | • Often based on causality  
• Easy to understand and develop  
• Can be used to calculate effect of change on controllability | • Applies approximations and estimations  
• Often sensitive to measurements  
• Often valid only in a small subspace of the system’s state-space |
| ABM | • Can account for individuality  
• Can cover spatial distribution | • High complexity when having a number of properties  
• Can be sensitive to level of detail  
• Large amount of data to calibrate and validate |
| ANN | • Effective when the components of system and/or their interrelations cannot be clearly identified  
• Easy to apply  
• No approximation or data exclusion required  
• Applicable to heterogeneous datasets | • Is a black box and has no causality by nature  
• The expert opinion in:  
  ○ network type selection  
  ○ parameter selection  
  ○ number of hidden layers and neurons  
  ○ activation functions  
• Needs historic data |
| BN | • Applicable to heterogeneous datasets | • Sensitive to structure |
### Table 2-2 Model applications

<table>
<thead>
<tr>
<th>Model</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>IO Model</td>
<td>Water quality, water pollution, fisheries, natural resource/parks, chemicals, toxic substances, climate models</td>
</tr>
<tr>
<td>ABM</td>
<td>Fish cohort model, predicting the spread of epidemics and the threat of bio-warfare, modelling the adaptive immune system, understanding consumer purchasing behaviour</td>
</tr>
<tr>
<td>ANN</td>
<td>Rainfall-runoff, stream flow, groundwater management, water quality simulation, rainfall forecasting, weather forecasting, fish species identifications, short-term industrial management</td>
</tr>
<tr>
<td>BN</td>
<td>Weather forecasting, medicine, robotics</td>
</tr>
</tbody>
</table>

### 2.3.4 Applicability of Models for Decision Makings

The modeller faces two problems when constructing a model of the system:

- What kind of model should be built?
- How to build it?

For determining the type of model(s) that should be built, there are several essential factors: (i) what is the objective of modelling? (ii) is there a known theory (in terms of natural laws) which, in theory, could be used to build a model that produces answers for managers? Depending on the answer to this question some model types can be eliminated; in practice another limiting factor can be the availability of expertise (iii) what is necessary (in terms of data) to build such a model and are there data available, or is it feasible to obtain such data. Feasibility can be a limiting factor (due to cost, accessibility, as well as space and time constraints).

It is therefore clear that to build models for environmental management/ control both the properties of the specific system under study and the objectives of modelling are to be considered. The system properties are recognised by the
characteristics of the available data of different sources: measurements, simulations, experts’ opinions. The type of dataset (homogeneous/heterogeneous), data reliability (accuracy of measurements, certainty of experts’ opinions, and accuracy of simulations), data availability (sparse/large amount), and data accessibility (access to historical data/high frequency sampling) construct the pack of system properties to be identified in this step.

The other part to be clearly recognised is the set of modelling objectives. Different types of models support decision making at different levels of control and management. As Table 2-3 indicates, the objective of modelling can be limited to the operational level of management, which involves real-time control of the system. In this case, a ‘current model’ of the system is necessary. In this regard, the IO and ABM models can perform well. The aim of this level of control and management is to identify the system’s critical elements and their interrelations in order to keep the system’s performance indicators within acceptable limits.

<table>
<thead>
<tr>
<th>Model</th>
<th>Applicability</th>
<th>Prediction capability</th>
<th>Management level</th>
</tr>
</thead>
<tbody>
<tr>
<td>IO Model</td>
<td>(i) Real-time control; (ii) Systems Design (control system design)</td>
<td>(i) Present/short-term (ii) Long term</td>
<td>(i) Operational (ii) Under certain circumstances strategic (eg strategic risk mitigation)</td>
</tr>
<tr>
<td>ABM</td>
<td>(i) Real-time control (ii) Systems Design</td>
<td>Present/short-term</td>
<td>(i) Operational (ii) Tactical</td>
</tr>
<tr>
<td>ANN</td>
<td>(i) Real Time control (ii) Operational (iii) Optimisation of industrial or ecological systems</td>
<td>(i) Short term (ii) Medium</td>
<td>(i) Operational (ii) Tactical</td>
</tr>
<tr>
<td>BN</td>
<td>(i) Management (ii) Policy making</td>
<td>(i) Medium (ii) Long term</td>
<td>(i) Tactical (ii) Strategic</td>
</tr>
</tbody>
</table>
The next level of management, the tactical level, looks into the short-term future, and with the help of historical data and consequent behaviours of system, forecasts the future behaviour of the system. However, tactical level decision making and control is (by definition) limited to parametric manipulation of the system, whereupon strategic level decision making needs to be open to structural change (or at least needs to be able to consider such course of action, if no acceptable pathway of the present system seems possible).

Tactical level models are unable to produce long-term predictions, as they are limited to considering the control of the system as is. However, although the prediction validity is limited to a short-term or medium term future, a wide range of critical decisions can be made just in time to prevent serious damage or gain significant profits. For example, as Table 2-3 shows, ANNs are extremely efficient in optimising functionality in industry by making accurate short-term or medium term future predictions. For example, to identify faulty parts of the system before the damage occurs (Zhang et al., 2012).

At the strategic level, managing and controlling the system might require applying structural changes to the system due to a perspective of long-term future. Even though the status quo of a system may be satisfactory, the secure path (trajectory in the state space) in which the system is proceeding is expected to turn unsafe and an unsustainable/ uncertain future may be predicted for the system. Strategic level decision making must deal with uncertainty, and be able to consider alternative futures as well as the uncertainties inherent in the predictions of such futures.

In these cases, therefore, the model of the system needs to be able to be manipulated, possibly evaluating the effects of significant changes to the system and allow the production and evaluation of different scenarios and possible future situations to help experts make effective policies which ensure a sustainable future for the system. BN models are capable of working with such uncertainty in the systems. For example, Hamilton et al. (2007) use a BN model to integrate available information in order to assess risk factors to conduct scenario analysis. However, IO models can also be used (provided the system in question has a well understood set of physical laws based on which an IO
model can be constructed). Increasingly, decision making in environmental management will have to consider hybrid systems, in terms of the systems of interest being partly natural and partly built.

In addition to the above, the feasibility of model building and use in decision making requires several other conditions, the present article limits the investigation to the selection of model types, while acknowledging that many limitations exist in terms of data accessibility, time, and resource (money, expert human resource, computational power), as well as limitations of current scientific theories available for exploitation.

2.4 Summary

In this chapter, I discussed the major types of mathematical models which are currently in use in the field of environmental control and management, and explored the strengths and weaknesses of the models in dealing with different systems and performing at different levels of management.

While finding one specifically most suitable model type for a system to answer all the aforementioned queries is a difficult, mostly impractical task, in this chapter, I provided a summary of the advantages and limitations of each model type, in addition to the modelling objectives each model type would fit in.

In the next chapter, statistical models would be studied as another significant group of models employed in the field of environmental management and control.
Chapter 3

Quantitative models for environmental management – Statistical models
3.1 Introduction

In chapter 2, the main four mathematical models; namely, Input-Output Models, Agent Based Models, Artificial Neural Networks and Bayesian Models, were briefly reviewed and their applications in environmental management and decision-making were discussed and compared. In this chapter, statistical models are discussed as another extensively used category of models in the field of environmental management and decision-making.

The term ‘statistical models’ refers to models of probabilities (Kruschke, 2010). A statistical model is a non-deterministic model comprised of a linear or even non-linear combination of a deterministic and stochastic component (Ang and Tang, 2007).

Therefore, a simple linear statistical model can be shown as:

\[ y = X\beta + \varepsilon \]  

Where

\( y \) is the vector of response variables; \( y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \)

\( X\beta \) is the deterministic component; \( X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \), \( \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \)

\( \varepsilon \) is the stochastic component vector; \( \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \)

The deterministic component includes the predictor variables and the coefficients that describe the relationship between the predictor variables and the response variable. The stochastic part is referred to as the random part of a model and describes the form of the error distribution. The presence of the
stochastic part makes the model categorised as statistical, rather than just a mathematical model (Kéry, 2010).

Stochastic elements of the model are described through statistical distributions rather than specific values and are responsible for the variability in the response. Therefore, a statistical model employs a set of probability distributions on the sample space (McCullagh, 2002; Ang and Tang, 2007) to accommodate uncertainty and to describe stochastic events (Konishi and Kitagawa, 2007).

Similar to mathematical models (Chapter 2), statistical models are used to determine the structure of complex systems, display assumed causal relationships, and draw inferences about the system behaviour, as well as to make predictions and decisions about the system (Konishi and Kitagawa, 2007). The information from the observed data of the system is obtained and analysed to understand relationships among variables with uncertainty and approximate the most possibly accurate structure of the system. (Konishi and Kitagawa, 2007).

Some common distributions that are used to capture the variability in a response are: normal, binomial, uniform, and Poisson distributions. These four distributions are the most frequently employed distributions in ecological modellings (Kéry, 2010). The appropriate statistical distributions used in building the model are determined based on the type and characteristics of observed data. For example, response variables that are whole numbers (e.g. counts of species) require whole number distributions for the stochastic component, such as the Poisson distribution, whereas continuous response variables require continuous distributions such as the Normal distribution (Gelman and Hill, 2007). In the next section, before reviewing the statistical models, I first review these distributions.

### 3.2 Probability distributions

To represent events, we use random variables. A random variable is a mathematical tool to represent an event in an analytical form. In all cases, the
state of an event can be represented with the value or a set of values of a function that is called a random variable (Ang and Tang, 2007).

When an event occurs in a deterministic way it can be either invariable and shown with a fixed value, as \( x = a \), or as in general, it can vary depending on some factor(s) and the state of the event can be expressed with some function of those factors.

However in most events, some uncertainty exists and also our knowledge about the factors is incomplete. Therefore in such events, the state of event cannot be specified as a deterministic function or a fixed value, and probability distributions are employed to express the event (Konishi and Kitagawa, 2007).

When speaking of probability, we are referring to a numerical measure of the likelihood of occurrence of an event within a set of all possible alternative events (if there weren’t more than one possibility, the case would be deterministic). This set of all possibilities is called the sample space (Ang and Tang, 2007).

Therefore, the numerical values of the random variables, which represent events, are associated with specific probability measures related to the sample space. These probability measures can be assigned according to rules that are called probability distributions (Ang and Tang, 2007).

Probability distribution of a random variable \( (X) \) can always be expressed by its cumulative distribution function (CDF), which is defined as:

\[
F_x \equiv P(X \leq x)
\]

For a discrete random variable, the probability distribution can also be described in terms of a Probability Mass Function (PMF), which shows the probability of \( P(X = x_i) \) for all \( x_i \), and is denoted by:

\[
p_{X}(x_i) \equiv P(X = x_i)
\]

For a continuous random variable, the probability distribution is described with a Probability Density Function (PDF), which shows the probability of \( X \) in an interval \((a,b]\) and is denoted by:

\[
P(a < X \leq b) = \int_{a}^{b} f_{X}(x) \, dx
\]
3.2.1 Normal distribution

Normal or Gaussian distribution is certainly the best known and most commonly used probability distribution (Ang and Tang, 2007). The PDF for normal distribution for a continuous variable \( X \) is:

\[
f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right]
\]

The two parameters; \( \mu \) and \( \sigma \), completely specify the distribution. These parameters are the mean and the standard deviation of the random variable \( X \). A popular short notation for this distribution is \( N(\mu, \sigma) \). An extension of normal distribution is the log-normal distribution for the case in which data are known to be strictly positive (Ang and Tang, 2007).

Significance of the two parameters can be observed in Figure 3-1.

![Figure 3-1 Four normal distribution with random different values of \( \mu \) and \( \sigma \).](https://example.com/image)

3.2.2 Binomial distribution

Many cases involve the occurrence or recurrence of an unpredictable event in a sequence of discrete independent trials.

If the random variable \( X \) denotes the number of occurrences of an event among \( n \) trials, the probability of occurrence of the event in each trial is \( p \), and the
corresponding probability of non-occurrence is \((1-p)\), then the probability of exactly \(x\) occurrences among the \(n\) trials is described by the Binomial PMF as:

\[
P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,2,\ldots,n
\]

For the case with \(n=1\), this probability distribution is referred to as a Bernoulli distribution (Ang and Tang, 2007). The parameters of the distribution are \(n\) and \(p\) and \((n)\) is known as the Binomial coefficient. Figure 3-2 shows the Binomial distribution and its parameters.

![Figure 3-2 Four Binomial distribution with random different values of \(p\) and \(n\).](image)

### 3.2.3 Continuous uniform distribution

The random variable \(X\) in uniform or rectangular distribution is restricted to a finite interval with \(x=a\) as the minimum value \(X\) can take on, and \(x=b\) as the maximum value \(X\) can take on. Any intervals of equal lengths are equally likely to occur. The PDF of uniform distribution has constant density over the interval of possible values of \(X\) as follows:

\[
f(x) = \begin{cases} 
  \frac{1}{b-a} & a \leq x \leq b \\
  0 & \text{otherwise} 
\end{cases}
\]

\(a\) and \(b\) are the parameters of the distribution and the mean and variance of the distribution are:
\[
\bar{X} = (a + b) / 2
\]

\[
\sigma^2 = \frac{1}{12}(b - a)^2
\]

Figure 3-3 illustrates the general plot of continuous uniform distribution. Continuous uniform distribution is used when the data consists of measurements that are all equally likely to occur in a certain range of values, e.g. to specify ignorance in a prior, as an alternative to a flat normal distribution (Ang and Tang, 2007).

3.2.4 Poisson distribution

Poisson distribution is a discrete probability distribution that is appropriate to model the occurrences of an event when the event can randomly occur at any instant of time/ point of space, the occurrences of the event are statistically independent, the probability of two or more occurrences at the same interval is negligible, the mean occurrence rate of the event is constant, and the probability of an occurrence in an interval is proportional to the length of the interval (Ang and Tang, 2007). The PMF of Poisson distribution is as follows:

\[
P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda} \quad x = 0, 1, 2, \ldots
\]

With \( \lambda \) is the mean number of occurrences in a specified time and \( X \) is the actual number of occurrences in the specified time.
Lambda is the sole parameter of the Poisson distribution, and both mean and variance of the distribution are equal to $\lambda$ (this assumption is called equidispersion). However, real data often show overdispersion (where variance is greater than the mean) or underdispersion, and the dispersion test should be considered by the practitioners to determine if this assumption can be met (Serinaldi, 2013).

Poisson distribution is mostly used for counts. When discrete data are randomly distributed and a “counting window” is randomly placed to record the number of them, then that number is Poisson distributed (Kéry, 2010; Agresti and ProQuest, 2015). Figure 3-4 illustrates Poisson distributions with different values of $\lambda$.

![Figure 3-4 Four Poisson distribution with random different values of $\lambda$.](image)

### 3.3 Linear statistical models

Linear statistical models are the most common types of statistical models that have been widely used in the field of ecology. Such models specify a linear (additive) relationship between the response and some predictor variables; either they are continuous or discrete (Gelman and Hill, 2007; Kéry, 2010). Based on the type and characteristics of data and the objectives of modelling, different models can be employed.

‘Generalised Linear Models’ (Nelder and Wedderburn, 1972) is the term that
includes and unifies a large number of statistical models such as model of the mean, t-test, analysis of variance (ANOVA), analysis of covariance (ANCOVA), and regression (McCullagh and Nelder, 1989; Kéry, 2010). They declared a large number of statistical tools, which were previously thought to be quite separate, are united in a larger class of a linear model (Lee et al., 2006; Kéry, 2010).

A GLM consists of three components: a probability distribution of the data vector \( y \) (response variables), a linear predictor of \( X\beta \) with the coefficients vector \( \beta \) and predictors matrix \( X \), and a link function which is applied to map the linear predictor to the response. The link function explains the relationship between the mean of data distribution and the linear predictor (Gelman and Hill, 2007; Kéry, 2010; Kruschke, 2010):

\[
y = (y_1 \ldots y_n)
\]

\[
\hat{y} = f^{-1}(X\beta)
\]

\[
y_i \sim p(y | \hat{y})
\]

where \( y \) is the data vector, \( \hat{y} \) is the vector of mean of data, \( p(y | \hat{y}) \) is the data distribution, and \( f \) is the link function.

The most widely used statistical distributions in GLMs are Binomial, Poisson, and normal distributions, and the three mostly used link functions are the identity (as in equation 3-1), the logit, and the log (Kéry, 2010).

While the bases of these models are much similar, here the very common types of these models with ample applications in ecology are discussed.

### 3.3.1 Model of the mean

The model of the mean is perhaps the simplest model of all. This model is taking a simple average and estimates the mean of a normal population from a sample of measurements taken from that population. This model does not account for dependencies among measurements or populations with asymmetric distributions. The model can algebraically be shown as:

\[
X_i = \mu + \varepsilon_i
\]

Which indicates every measurement \( i \) of variable \( X \) is the addition of an overall
mean ($\mu$) with some individual deviation from that mean, which is called $\varepsilon_i$. $\varepsilon_i$ is the residual of measurement $X_i$.

To estimate the mean in a linear statistical model, an assumption about these residuals is also needed, e.g., the residuals are normally distributed around $\mu$ with a variance of $\sigma_2$, or $\varepsilon_i \sim \text{Normal}(0, \sigma_2)$ (Gelman and Hill, 2007; Kéry, 2010).

### 3.3.2 T-test

T-test is one of the most widely used linear statistical models (Kéry, 2010; Agresti and ProQuest, 2015); there are three main types of t-test; namely, the one-sample, the paired-samples, and the independent-samples tests (Field, 2013).

The independent-sample test is the most common type and includes the very first use of t-test by Gosset to compare the Guiness bears (Box, 1987). This test is a tool to check if two means (averages) are reliably different from each other. Given that $H_0 : \mu_2 - \mu_1 = \sigma$ is the null hypothesis, indicating that one mean is greater than the other by a specific amount, and $H_1 : \mu_2 - \mu_1 \neq \sigma$ is the alternative hypothesis, if $H_0$ is rejected in favour of $H_1$, it indicates the two groups are desirably different (Ryan, 2007).

T-test can be used when the two populations have (approximately) normal distributions, and the sample sizes are not large. In case of independent-samples t-test, the two sets of measurements should be independent of each other.

### 3.3.3 Analysis of variance (ANOVA)

The ANOVA is an extension of t-test for situations where there are more than two groups. There are different kinds of ANOVA: one-way, with a single factor, and two- or multiway, with two or more factors (Kéry, 2010; Agresti and ProQuest, 2015).

The ANOVA model assumes that the populations have (approximately) normal distributions, each probability distribution has the same variance, and the sets of measurements are independent of each other (Walpole et al., 2012).
The objective of one-way ANOVA is to test whether there are any real differences between the means of many groups. In this type of ANOVA model, the variation in the observed data is divided into two parts; one related to the group (the known causes), and the other related to natural (the error due to unexpected random causes); these two categories of variations are called the between, and the within variations, respectively (Kéry, 2010; Agresti and ProQuest, 2015).

Similar to t-test, when a comparison of several group means is carried out using samples from the independent sets of populations, the observed data is assumed to have the following construction:

Observation = group mean + error

By subtracting the overall mean, and working with deviations from this overall mean, each measurement can be shown as follows:

Observation = overall mean + group effect + error

Or

\[ y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \]  \hspace{1cm} (3-14)

Where:

- \( y_{ij} \) is the \( j \)th observation of the \( i \)th group,
- \( \mu \) is the overall population mean,
- \( \alpha_i \) is the effect of \( i \)th group,
- \( \varepsilon_{ij} \) is the random variations related to the \( ij \)th observation.

\( \varepsilon \)'s are independent and normally distributed random errors with zero mean and finite variance of \( \sigma^2 \), which is the same for all groups (Kotegoda and Rosso, 2008).

### 3.3.4 Normal linear regression

Regression analysis is the collection of statistical tools that are used to model relationships between variables that are related in a nondeterministic manner. This model shows the relationship between a continuous response (dependent variable) and one or more continuous predictor (independent) variables. The
link function is identity. This model is written algebraically as follows (equation 3-15 is the mean value of \( y \) given \( X \) using vector notation, and equation 3-16 is the same in index notation):

\[
\hat{y} = X\beta \\
y_i = \beta_0 + \beta_1x_{i1} + \ldots + \beta_kx_{ik} + \epsilon_i
\]

where \( \beta_k \)'s are called regression coefficients. As illustrated, the mean of \( y \) is a linear function of \( X\beta \); however, the actual observed value of \( y_i \) for fixed values of \( x_{ki} \) is determined by the mean value function (the linear model) plus a random ‘error term’, represented by the symbol \( \epsilon_i \). The errors \( \epsilon_i \) have independent normal distributions with mean 0 and standard deviation \( \sigma \).

An equivalent representation is:

\[
y_i \sim N(x_i\beta, \sigma^2)
\]

where \( X \) is an \( n \times k \) matrix with \( i^{th} \) row \( x_i \).

By fitting the model and using least squares the mean of \( \beta \) and \( \sigma \) are estimated. The least squares estimate is the \( \beta \) that minimises the sum of squared errors,

\[
\sum_{i=1}^{n}(y_i - x_i\beta)^2.
\]

The least squares criterion is useful as it minimises the error of the prediction when predicting an outcome using other variables. The least squares estimate is equivalent to the maximum likelihood estimate when the errors \( \epsilon_i \) are independent with equal variance and normally distributed (Gelman and Pardoe, 2006).

There are some assumptions to use the regression model: validity (e.g., the relevance of the data to the research purpose, the right choice of input for the model, the applicability of the results to other cases), additivity and linearity, independence of errors, equal variance and normality of errors (Gelman and Hill, 2007). In regression analyses, the collinearity of every pair of predictors needs to be tested. Collinearity is a high degree of correlation among predictor variables, and happens when predictors are linearly dependent (Montgomery et al., 2012).

The collinearity does not affect the prediction of the regression model; however, it affects the estimation of regression coefficients. The regression
coefficient represents the mean change in the response variable for each one unit change in a predictor variable when all of the other predictor variables are held constant. However, when predictor variables are correlated, it indicates that changes in one predictor are associated with changes in another predictor. It becomes difficult for the model to estimate the relationship between each predictor and the response variable independently, because the predictors tend to change in unison. However, when the predictors are highly correlated, they provide very similar information in the model. Therefore, using either one or both predictors might not affect the prediction of the model. As the correlation between two predictors increases, the differences in the predictions of models using either predictor become smaller (Seber et al., 2003; Montgomery et al., 2012).

As the bases of the special cases of the GLM models are similar, the linear regression model is discussed in this research as a very common type of GLMs with ample application in ecology.

### 3.3.5 Hierarchical Linear Models (HLM)

Hierarchical models are extensions of linear regression models in which data are multiscale and structured in groups, and coefficients vary by group (Raudenbush and Bryk, 2002; Gelman, 2006b).

More generally, a hierarchical model is a regression in which the intercept and/or the regression coefficients are given a probability model; a second-level model with its own parameters, called the hyperparameters of the model. The varying coefficients and a model for those varying coefficients are the fundamental parts of an HLM (Gelman, 2006b). The main characteristic of a hierarchical model which makes it recognised from a single-level regression is the modelling of the variation between groups. With grouped data, a hierarchical regression can be structured as one of the following models:

**varying-intercept model**, which stands for a model with a different intercept within each group, and can be represented as

$$y_{ji} = \beta_{0j} + \beta_ix_{ji} + \varepsilon_{ji}$$
**varying-slope model**, which stands for a model with a different slope within each group, and can be represented as

\[ y_{ji} = \beta_0 + \beta_{1j} x_{ji} + \varepsilon_{ji} \]  \hspace{1cm} (3-19)

**varying-intercept, varying-slope model**, which stands for a model with a different intercept and different slope within each group, and can be represented as

\[ y_{ji} = \beta_{0j} + \beta_{1j} x_{ji} + \varepsilon_{ji} \]  \hspace{1cm} (3-20)

In the abovementioned equations, \( j \) represents the group to which data belongs, \( y_{ji} \) is a measured response variable, \( x_{ji} \) is the individual-level predictor variable, \( \beta_{0j} \) and \( \beta_{1j} \) are the intercept and slope coefficients, respectively, which vary across groups.

HLM is composed of distinct regression models that describe different levels of hierarchical data and explains the relationships within the dataset by variables at different scales. Data can be modelled at the level they are collected, or any higher level.

An HLM can be represented in the form of a two-level, simple regression equations, all notations follows Raudenbush and Bryk (2002) as below:

\[ y_{ji} = \beta_{0j} + \beta_{1j} x_{ji} + \varepsilon_{ji} \]  \hspace{1cm} (3-21)

which is a varying-intercept, varying-slope model as described before.

In addition, the residual errors \( \varepsilon_{i} \) are assumed to be distributed normally with a mean of zero and a variance of \( \sigma^2 \) (as explained in section 3.3.1). Group-level coefficients are then forming the higher-level regressions:

\[ \beta_{0j} = \gamma_{00} + \gamma_{01} z_j + \omega_{0j} \]  \hspace{1cm} (3-22)

and

\[ \beta_{1j} = \gamma_{10} + \gamma_{11} z_j + \omega_{1j} \]  \hspace{1cm} (3-23)

where \( \gamma_{00} \) and \( \gamma_{10} \) are the second-level model coefficients for the intercept and slope, respectively. \( z_j \) is a second-level predictor, and \( \omega_{0j} \) and \( \omega_{1j} \) are the error terms.
3.3.6 Bayesian versus frequentist framework

Statistical models can be fitted either in a Bayesian or a frequentist framework. As explained in section 2.2.4, the Bayesian approach is to estimate the posterior probability via conditional probability using Bayes’ theorem. Similarly, Bayesian statistics is a statistical framework which incorporates data with subjective prior knowledge about parameter values to extract posterior probabilities of the parameter values (Bolker et al., 2009). In the absence of appropriate information, the prior probability of the model can be ‘uninformative’, meaning that the model parameters are estimated based on only the data with no complementary impact of a prior distribution (Ellison, 1996; Gelman et al., 2004).

Many studies have been conducted to compare the Bayesian and frequentist approaches, with varying conclusions. Some studies have argued that frequentist approaches are stronger and more reliable, because they rely on objective observation and avoid bias that may arise from subjective probability (e.g. Dennis (1996) and Knapik et al. (2011)). Other studies have practically showed the advantages of Bayesian over frequentist methods after implementing both approaches in on different case studies (Chaloner, 1987; Ludwig, 1996; Taylor et al., 1996; David et al., 2007; Liu et al., 2015). Still others have suggested that either approach may be most suitable, depending on the context and research goals (Browne and Draper, 2006; Berger, 2010; Röver et al., 2011). With this range of findings from quantitative studies, philosophical and practical differences become important in determining the best approach.

Ellison (1996) provided a comparative assessment of Bayesian statistics and frequentist null hypotheses approaches. He concluded that Bayesian statistics makes better use of prior information, provides stronger inferences from non-replicable datasets, and is more appropriate and meaningful in the field of environmental management and decision-making (Ellison, 1996). These differences between the Bayesian and frequentist approaches can be summarised as follows:
- The Bayesian approach produces an estimate of the posterior probability of the model given the observed data and any prior information (Gelman et al., 2004; Clark, 2005), whereas the frequentist estimates the probability of the observed data given the model, which is the likelihood component of Bayes theorem (Ellison, 1996) (equation 2-6 and equation 2-7):

\[
P(h | e) = \frac{P(e | h)P(h)}{P(e)},
\]

\[
(Posterior = \frac{Likelihood \times Prior}{Probability \ of \ evidence})
\]

In the frequentist statistics, the concept of probability is the probability of observing the data under the model; there is no probability attached to the hypotheses or any fixed but unknown values (Ellison, 1996; Gelman et al., 2014a). However, the probability in the Bayesian statistics is interpreted as the degree of belief in the model; it represents the state of knowledge of the observer, and updates the estimate of the probability that the model is the true underlying model, based on the given data (Gelman et al., 2014a).

- Frequentist statistics does not assume any prior information (Anderson et al., 2000; Fidler et al., 2004). However in ecology, usually there is a considerable prior knowledge about the interactions between organisms and their environment. This knowledge derived from previous studies can be explicitly used in the analysis and interpretation of new data by using a Bayesian approach (McCarthy and Masters, 2005). Therefore, the Bayesian approach makes use of additional sources of information via the inclusion of prior probabilities in the model (Ellison, 1996; Gelman et al., 2004).

- It is also important to note that simple Bayesian analyses based on uninformative prior distributions produce numerically equivalent results to non-Bayesian likelihood-based frequentist methods (Clark, 2005; McCarthy and Masters, 2005; Aksoy and Weesie, 2013; Gelman et al., 2014a). Note that in general as the size of the dataset increases, the
influence of the prior distribution on posterior inferences decreases (Gelman et al., 2014a). The posterior probability is a “compromise” between the prior information and the likelihood, and the likelihood (or data) is more dominant as the sample size increases (Gelman et al., 2014a).

- In both frequentist and Bayesian statistics, ‘data’ are the result of some observation of a stochastic system with some random processes. However, in frequentist statistics, the parameters used to describe these random processes are assumed to be fixed and unknown constants (Efron, 1978); in fact, classical statistics estimates a single point for a parameter, such as the fixed coefficients of predictor variables in a sample statistical model shown in equation 3-1. In contrast, Bayesian statistics estimates an entire distribution for any unknown parameter (Ellison, 1996; Kéry, 2010).

Frequentist approach provides a confidence interval for the estimation of parameters; e.g., a 95% confidence interval includes the true value of the parameter of interest for the 95% of the infinite repeats of the practice (Ellison, 1996). In contrast, in the Bayesian approach, the parameters are not fixed, and the credible interval indicates the portion of possible values of the parameter falling within the interval; e.g., a 95% credible interval represents the interval that contains 95% of possible values of the parameter of interest (Ellison, 1996; Gelman et al., 2014a).

- In frequentist statistics, uncertainty is addressed by calculating the frequency of various observation outcomes if the study were replicated infinite number of times (Ellison, 1996). However, such replication in ecological cases is mostly small (Hurlbert, 1984) and in some cases rarely existent (Hurlbert, 1984; Carpenter, 1990; Reckhow, 1990). Besides, if repeated, the very alike ecological cases are rarely independent (Ellison, 1996). In many case, inferences need to be derived only from the knowledge of a single dataset (Ellison, 1996; Kéry, 2010).

In Bayesian statistics, uncertainty is addressed using the posterior distribution of a parameter. As explained in section 2.2.4, the posterior distribution is the conditional probability distribution of the unknown
quantities, given the data, the model, and our prior knowledge of these quantities before conducting the analysis (Kéry, 2010; Gelman et al., 2014a).

The main practical advantages of the Bayesian approach emerge in hierarchical models (Gelman et al., 2014a). Hierarchical Bayesian (HB) modelling was a transformation in computational statistics which provides a framework to fit many examples of high-dimensional data (Gelfand and Smith, 1990; Carlin and Louis, 2000; Clark, 2005). Consequently, HB modelling is getting more popular for its capacity to accommodate complexity (Link et al., 2002; Clark, 2005; Clark and Gelfand, 2006; Lele et al., 2007; Kéry, 2010) by making it possible to break down the problem into several more explicit levels. The intuitive structure allows a better understanding of the model uncertainty and the processes to treat unknown quantities (Clark et al., 2005; Cressie et al., 2009; Gelman et al., 2014a).

HB modelling has a flexible structure to deal with uncertainty and variability, and can tolerate the errors in variables, random effects, hidden variables, and multiple data sets at different scales (Clark, 2003; Clark, 2005). The flexible framework means that hierarchical structured models can be constructed to resolve more complex problems with messy data, and the model parameters and variables represent a wider range of models rather than those of traditional models (Clark, 2005). Such models may embrace different sources of uncertainty “such as observation errors, uncertainty around ecological processes, and model parameter uncertainty” (Cressie et al., 2009; Stewart-Koster, 2011).

Applications of HB in ecology are growing rapidly. Examples are: in species abundance and distribution (Jabot and Chave, 2011; Aderhold et al., 2012; Zhang et al., 2014), in capture-recapture (Brooks et al., 2000; Tardella, 2002; Royle and Dorazio, 2012), in fisheries (Maoiléidigh et al., 2004; Tomberlin and Holloway, 2010; Tang et al., 2014), in riverine systems (Wyatt, 2002; Smith et al., 2009; Cha et al., 2010; Stewart-Koster et al., 2013), and in forestry (Green et al., 1999; Radtke et al., 2002; Chen et al., 2016).
3.3.7 Hierarchical Bayesian models vs single-level models

The literature reveals the advantages of hierarchical linear models, using either maximum likelihood or Bayesian techniques. However, there are a few concerns that caution ecologists against blindly using these models.

Complexity of hierarchical models is a disadvantage of fitting these models (Downes, 2010; Hodges, 2010; LaDeau, 2010). The mathematical structure of hierarchical models, and Hierarchical Bayesian models in specific, is notably more complex than single-level models and the implementation of such models is consequently more difficult (Hodges, 2010).

Also, hierarchical models comprise more parameters than single-level models (Clark et al., 2005); therefore, they can model much more complex relationships than the single-level models (e.g. Dorazio et al. (2010) and Heisey et al. (2010)). However, the inclusion of many parameters in the hierarchical models may lead to the problem of overfitting where there are too many parameters (the model is incorporating too much complexity) compared to the certain number of observations in the dataset; such models perfectly fit the dataset, but is not likely to fit new datasets (Babyak, 2004; Rushton et al., 2004).

The mathematical computation of Bayesian analysis has become feasible with MCMC algorithms. MCMC can be efficiently used to perform Bayesian analysis to calculate the posterior distributions (Polansky and Kirmani, 2003; Lele et al., 2007; Cressie et al., 2009). More detail on MCMC can be found in (Gelfand and Smith, 1990; Casella and George, 1992; Gilks et al., 1996; Robert and Casella, 2004). However, convergence of parameters in MCMC algorithm is slow for HB models (Gelman et al., 2014a).

These drawbacks of HB models have raised doubts about the applicability of such models in the field of ecology (Hodges, 2010; LaDeau, 2010). Instead, the environmental managers and, specifically, the ecologists generally tend to use simpler, faster, less expensive, and familiar approaches (Downes, 2010).

Latimer et al. (2006) conducted a study to compare HB models with simple ones. Besides, Stewart-Koster (2011) showed that the single-level Bayesian
model fitted a particular dataset almost as well as the hierarchical Bayesian model.

3.4 Complex modelling and the idea of hierarchy in ecology

The field of ecology centres around inspecting patterns of species distribution, abundance, and interaction of species (Andrewartha and Birch, 1964; Leibold et al., 2004; Diez and Pulliam, 2007). Different approaches to study ecology reveal that complexity and uncertainty are inevitable characteristics of ecological systems (Cressie et al., 2009; Jorgensen and Fath, 2011). Ecological systems are complex, non-linear and self-organised entities with large numbers of continuous interactions with one-another and with their biotic and abiotic surroundings occurring at multiple spatial and temporal scales (Levin, 1992; Rosenzweig, 1995; Chase and Leibold, 2002; Diez and Pulliam, 2007). Providing accurate knowledge about the processes driving a system and making reliable predictions of the system’s changes has been a challenge to ecologists (Carpenter, 2002; Peterson et al., 2003; Pielke Jr and Conant, 2003; Clark et al., 2005). Therefore, modelling ecological phenomena needs accurate accounting for multiple sources of uncertainty (Daszak et al., 2000; Clark et al., 2001; Beckage and Platt, 2003; Sacks et al., 2007).

Due to the complexity of ecosystems, representing the uncertainty parameter is a difficult task (Ibáñez et al., 2006; Cressie et al., 2009). Simple models mostly cannot cover the wide range of effects at different scales, while building up a more complicated model to accommodate all complexity and uncertainty parameters is rarely possible (Caswell, 1988; Clark, 2005). Such heterogeneity in ecological systems suggests that different principles might apply at different scales. To address this issue, the idea of applying hierarchy theory to model ecological systems without ignoring complexity was introduced (Von Bertalanffy, 1968; Allen and Hoekstra, 1992; Holling, 1992; Clark et al., 2005). Hierarchical linear models provide a natural statistical approach to estimate regression relationships and coefficients in multiscale/hierarchically structured data. HLM effectively deals with parameter variation across groups by giving a model for the parameters (Jackman, 2009).
During the last two decades, hierarchical models have shown great performance in elucidating ecological multiscale interactions. Numerous studies have employed hierarchical frameworks to model species distributions (Pearson et al., 2004; P Anderson et al., 2006; Elith and Leathwick, 2009; Latimer et al., 2009), species abundance (Laplanche, 2010; Boone et al., 2012; Arévalo-Frías and Mendoza-Carranza, 2015) species interactions (Vázquez and Simberloff, 2004), and riverine ecosystems (Ward and Tockner, 2001; Fausch et al., 2002; Thoms and Parsons, 2003; Thorp et al., 2006).

3.5 Summary and Research gap

Hierarchical Bayesian models are a powerful tool that can be used to quantify ecological processes, and guide decision making. HB models offer great promise in quantifying multiscale processes and developing complex probabilistic models that reflect underlying ecological processes.

Many studies have highlighted the performance of HB over non-hierarchical models. However, the disadvantages of HB models are a drawback to using these models; moreover, some studies have shown that hierarchical models are not always a better solution than the non-hierarchical ones in ecological problems. This reveals the strong need to understand the conditions where the additional work of fitting a hierarchical model provides a better model fit.

While both single-level and hierarchical models have been widely used in the fields of ecology and environment, I have found no study that is focusing on establishing a framework based on which one can decide which of the two approaches better fit a particular case.

In the view of this knowledge gap, the research in this thesis will seek to identify the statistical conditions under which hierarchical models might be necessary, in order to inform users when the additional effort of a hierarchical model development would be necessary and when the simpler single level model structures are appropriate.
Chapter 4

Proposed research methodology


4.1 Introduction

As discussed in the literature review (Chapter 2 and Chapter 3), environmental systems encompass complex interactions, which cause considerable uncertainty in every step of environmental modelling and eventually in environmental management that intends to use these models for decision making. Environmental problems are often about situations not formerly experienced and unrepeatable, and the outputs of modelling and management decisions can often only be expressed as better or worse, rather than right or wrong (Uusitalo et al., 2015).

The uncertainty in the environmental problems can arise from different sources. Regan et al. (2002), described different types of uncertainty, namely; measurement error, systematic error, natural variation, inherent randomness, model uncertainty, and subjective judgment. The model uncertainty is the uncertainty of the model parameters (e.g., the shape of functions, the omission of the less important variables), and the uncertainty of the model structure (the model structure selected to fit which reflects the relationship among the variables) (Uusitalo et al., 2015). While probabilistic approaches can effectively incorporate some types of uncertainty into the model calculations at different stages of modelling (Hoeting et al., 1999; Uusitalo et al., 2015), the uncertainty of the model structure remains as the 'notorious' type of uncertainty, which is inevitable and difficult to quantify (Regan et al., 2002; Uusitalo et al., 2015).

Typically, modellers select a model which is assumed to be appropriate for the case study, and credit its results as though it has generated the data in the first place (Hoeting et al., 1999). Unfortunately, ignoring the uncertainty in model structure may lead to drawing inferences and making decisions that are not as reliable as presumed (Hoeting et al., 1999). This is especially true when the model is used to predict system behaviour in an area of the state-space that has never been experienced before, because typical model validation techniques often use techniques that compare model behaviour with known system behaviour. Therefore to ensure as much uncertainty as possible is accommodated, the very first step in modelling, as well as the most important
one, is to make an accurate and credible selection of the type of model to be constructed that can be argued to lead to reliable inferences and decisions.

The purpose of this research is to establish a tool for finding the most probably appropriate model structure to understand the system and therefore to make management & control decisions about a system. Chapter 2 categorised the different types of models usually applied in the management & control of a system, and reviewed four basic categories of mathematical models: Input-Output Models, Agent Based Models, Artificial Neural Networks and Bayesian Models. At the end of the chapter, the strengths and limitations of each model category were discussed in detail and probabilistic models were found to be most useful for management & control decisions making, where they provide a more realistic picture of all possible outcomes of the system (Uusitalo et al., 2015). Chapter 3 specifically scrutinised statistical models as a sub-category of Input-Output models, which are widely used in the field of ecology. The conclusion of Chapter 3 showed how choosing Bayesian statistics, specifically Hierarchical Bayesian (HB) Models, can very successfully represent ecological systems and produce valid information to describe and predict the behaviour of such systems (Clark et al., 2005; Diez and Pulliam, 2007). However, the statement is usually not qualified with what the limits of the choice are.

A gap that was observed and explored in the literature review was the absence of a quantitative framework to determine the conditions in which a single-level model fails to faithfully represent the system and where an HB model is required. Equally, such a framework should identify where a single-level model is a good or reasonably good approximation of data that may otherwise be hierarchical (i.e., collected in a hierarchical fashion). To limit the scope of investigation of this thesis, this research will seek to identify the statistical conditions under which hierarchical models are recommended and those under which a simpler single level model will suffice to represent the system. The subsequent goal is to establish straightforward and robust guidelines that can be simply applied to a set of data, to inform users when the additional effort of developing a hierarchical model provides a better model fit and when the single-level model structures are appropriate and sufficient. Figure 4-1 illustrates a
proposed framework and methodology for this research.

**Figure 4-1 Research outline (Chapter 4 highlighted in red)**

### 4.2 Methodological research framework

To achieve the abovementioned goal of this research, I conducted a large simulation study, along with the analysis of an empirical ecological dataset. The simulation study is a way to analyse a large range of datasets with known structure, uncertainty and relationships among the variables, and leads to establish a statistical tool for model-structure selection. Whereas, the empirical study demonstrates the application of the proposed tool in a real setting with noisy data.

To achieve the objective of quantifying the statistical conditions where a
hierarchical model is required and those where a single-level model is sufficient, I shall follow a seven step process (Figure 4-2):

1. Defining the study space and the data generation model.
2. Generating datasets (simulated datasets) from the data generation model.
3. Fitting the two statistical models (a single-level and a hierarchical) to every simulated dataset.
4. Evaluating the models.
5. Analysing the results of the models’ evaluation from step 4.
6. Partitioning the study space.
7. Proposing a model-structure selection tool.

Figure 4-2 The simulation methodological framework; it involves several steps with each one building on the previous to follow a systematic process to evaluate the performance of each model structure under a range of conditions.
Each of these five steps involves a specific set of methods and analyses.

4.2.1 Defining the study space and the data generation model

I used a standard hierarchical statistical model structure as a data generation model to define the study space. This data generation model was a two-level Poisson regression. The Poisson distribution (being a non-negative, discrete whole number distribution) is appropriate and commonly used in modelling many ecological problems, such as modelling species abundances (Kéry, 2010). (It is to be noted that of course other similar studies could also be performed using different distributions).

The data generation model had the form:

\[ y_{ij} \sim \text{Poisson}(\lambda_{ij}) \]  
\[ \log(\lambda_{ij}) = \beta_{0j} + \beta_{1j} X_{ij} \]  
\[ \beta_{k} \sim N(\phi_{k}, \tau_k) \]

where, \(i\) refers to observations in the group, \(j\) refers to groups, \(y\) refers to observations, \(\lambda\) is the mean of the Poisson distribution of the observation \(y\), \(\beta\) refers to the coefficients of the linear regression of the link function, \(k=0\) refers to the intercept and \(k=1\) to the slope coefficient of \(\log(\lambda_{ij})\).

In the second level of the hierarchical model, the relationships between \(\beta_{k}\) in group \(j\) and the upper level predictor, \(Z_j\), are estimated. \(\beta_{kj}\) is normally distributed, with mean \(\phi_{kj}\) and precision \(\tau_k\), estimated by the following regression:

\[ \phi_{kj} = \gamma_{0k} + \gamma_{1k} Z_j \]  
\[ \gamma_{ik} \sim N(0, \tau_{ik}) \]  
\[ \tau_k \sim \text{Gamma}(\alpha_1, \alpha_2) \]

A Gamma distribution is used for the precision parameter (1/variance), \(\tau_k\), because it is the conjugate prior for Poisson distribution (Gelman, 2006b). The values for the parameters \(\alpha_1\) (shape) and \(\alpha_2\) (rate) determines the level of informativeness. For a relatively non-informative prior, \(\alpha_1\) and \(\alpha_2\) should be small. To finalise the Bayesian specification of the models, I used the following parameters for the relatively non-informative prior distributions (the relatively non-informative prior values for the gamma distribution (Gelman, 2006b)):
The gamma distribution has mean value equal to 4 (shape / rate) and a variance of 8 (shape / rate\(^2\)), and allows the precision to vary between 0 and 10. It gives us a reasonably non-informative prior which is also appropriate for computational purposes.

The study space, in which all simulated datasets were generated, was defined by several parameters of the data generation model. These included parameters that quantify the relationship to predictor variables and uncertainty around those relationships. Specifically, the parameters were:

\[\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma_0, \text{ and } \sigma_1.\]

Where \(\sigma_k\) is the square root of \(1/\tau_k\), and the \(\gamma\) parameters define the magnitude of the \(\beta\) parameters, which quantify the relationship between the response and the predictor variable.

To choose the parameters of the data generation model to define the study space, I firstly should determine where to put the variance component, which leads to generating a comprehensive collection of datasets that encompass as broad a range of conditions as possible. Based on the model structure, the variance can be inserted through the precision of the Normal distribution at two levels (parameter \(\tau\) in equations 4-3 and 4-5); either on the lower-level regression coefficients, \(\beta_s\), or the upper-level regression coefficients, \(\gamma_s\). I finally found that I need to factor the variance in by putting it on the draws from the normal distribution of \(\beta_s\); \(\gamma_s\) are the parameters that govern the mean of the regression parameters, \(\varphi_s\). The variance of the \(\gamma_s\) would be between the simulations.

Theoretically, as shown in Equations 4-4, the changing component of \(\varphi_{kj}\) in the hierarchical model is \(\gamma_{1k}Z_j\). If this component is set to zero, it leads to a constant \(\varphi_k\) across all groups, and consequently a constant normal distribution for \(\beta_k\)s, over all groups. The subsequent \(\beta_k\) values are then similar so as to be equivalent to a single-level data generation model.
The mean value of each of the $\beta$s, $\phi$, is dependent on two upper level regression coefficients, $\gamma$s. So, in building up the study space based on variations of $\beta$s, there is a surface of two axes of $\gamma$s for each $\phi_k$ (corresponding to each $\beta_k$). Also a third axis of $\tau$, is required to enclose the changes of the precision of each $\beta_k$. Therefore, the models are to be examined over the changes of $\tau$s of $\beta$s as well as of values of $\gamma$s that affect the mean values of $\beta$s (i.e., $\phi$s).

To assist with data simulation, I converted the precision, $\tau$, to the standard deviation, $\sigma$ (where $\tau = 1/\sigma^2$). Therefore, three axes of $\gamma_{0k}$, $\gamma_{1k}$, and $\sigma_k$ in the multi-dimensional study space are allocated to encompass the variations of each $\beta_k$. Based on the structure of hierarchical model, I inferred that small values of the varying attributes may lead to similar $\beta$ values, which is the nature of single-level models, while big values of the varying attributes can lead to very different $\beta$ values, which is the nature of hierarchical models. Therefore, I chose values for $\gamma$ and $\sigma$ in a way that the study space surely encloses the turning point. I assigned the following values to $\sigma$ and $\gamma$:

The value of each $\sigma$ changes between 0.05, 0.1, 0.2, and 0.3,

The value of each $\gamma$ changes between 0, 0.1, 0.25, 0.5, and 1.

These values were selected after careful consideration and preliminary analyses that showed either end of each gradient. For data generation model, I used one variable as the lower level regression variable and one as the upper level regression variable. Therefore, when building up the study space based on the coefficients of the lower level regression, there are two $\beta$s ($\beta_0$ and $\beta_1$), each creating three axes in the study space ($\gamma_{00}$, $\gamma_{10}$, and $\sigma$). Therefore, the study space is a 5x5x5x5x4x4 points in a 6 dimensional space of $\gamma_{00}$, $\gamma_{01}$, $\gamma_{10}$, $\gamma_{11}$, $\sigma_0$ and $\sigma_1$.

For clarity, I use the term ‘attributes’ to refer to these parameters of the data generation model through the rest of this research.
4.2.2 Generating the simulated datasets

The simulated datasets reflected a statistical design of grouped data. I assumed 10 groups of data (j=1…10), and for each of 10 groups, I generated 20 observations (i=1…20). The simulated datasets comprised two predictor variables (X and Z) and a response variable (y).

To ensure the findings of this research can be generalised and extended to different fields of study, it was necessary to make the statistical study be independent of the type and dimension of the data of any particular case study. Therefore, I standardised the predictor variables used to generate the simulated response variables, so that the findings of this research can be applicable to other case studies:

\[ X_{i,\text{new}} = \frac{(X_i - \bar{X})}{s} \]

where
- \( X_i \) = Each data point i
- \( \bar{X} \) = The average of all the sample data points
- \( s \) = The standard deviation of all sample data points
- \( X_{i,\text{new}} \) = The standardised data point i

This process brings all measurements to a normal distribution with mean equal to zero and variance equal to 1.

Accordingly, the first predictor variable, X (in equation 4-2), was randomly chosen from a standard normal distribution, while the second predictor variable, Z (in equation 4-4), was fixed with equal intervals between 0-1. This upper–level regression predictor variable represents some measurement that changes between the groups and is constant within each group. As the applied models had only one random predictor variable, checking the collinearity of the predictor variables was not required.

The simulated response variable, y, was generated given the randomly chosen X and Z values and the parameter values of the data generation model for the given point in the study space.

The process of data generation involved the following:
In the multi-dimensional study space, the values of the four \( \gamma \)'s \((\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11})\) and the two \( \sigma \)'s \((\sigma_0 \text{ and } \sigma_1)\) are known as the coordinates of each point of this study space. Having the values of the four \( \gamma \)'s of the point, and given the 10 values for \( Z_j \)'s, \( \phi \)'s \((\phi_{0j} \text{ and } \phi_{1j})\) were calculated by equation 4-4. By replacing each of the \( \phi \)'s and knowing the values of \( \sigma_0 \) and \( \sigma_1 \) of the point, the distribution for \( \beta_0 \) and \( \beta_1 \) of each group resulted. The \( \beta \) values were taken as random draws from equation 4-3 (here is where the noise is inserted). Having \( \beta \)'s and the lower-level regression predictor variables, \( X_{ij} \)'s (as in equation 4-2), the \( \lambda_{ij} \)'s were simulated. Then, \( y_{ij} \)'s were generated as random draws from a Poisson distribution with mean of \( \lambda_{ij} \)'s (as in equation 4-1).

Figure 4-3 shows an example of changes of generated \( \beta \)'s over \( Z \) on a point of the study space, and Figure 4-4 shows the subsequently generated response, \( y \), in ten groups of the same point. As expected, with no effect of the values of \( \gamma \)'s and the large value for \( \sigma \)'s, the generated \( \beta \)'s are scattered over \( Z \), as well as generated \( y \)'s over \( X \)'s in each group.

By repeating the abovementioned procedure, I generated seven simulated datasets at each point in the study space to ensure adequate replication throughout, and the results were not skewed by a single dataset at any given location of the study space.

![Figure 4-3 Simulated values of \( \beta_0 \) and \( \beta_1 \) over the range of values of \( Z \), for the point values of \( \gamma_{01}, \gamma_{10}, \gamma_{02}, \gamma_{12} = 0, \) and \( \sigma_1, \sigma_2 = 0.4 \) (x-axis shows \( Z \) values, y-axis shows \( \beta \) values).](image-url)
4.2.3 Fitting the two statistical models

As illustrated in Figure 4-2, at the third stage of the simulation study, the two single-level and of the hierarchical models were fitted to every simulated dataset over all points in the study space. Both model structures are Poisson Bayesian regressions with the logarithm function as the link function (as it is the most advantageous one for Poisson regression models (Kéry, 2010)).

The *single-level model* assumes that the relationships between the response variable and each predictor variable do not change. Equation 4-8 and 4-9 show the mathematical representation of the single-level model used in this research.

\[
y_i \sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \tag{4-9}
\]

where \( y_i \) represents the \( i \)th observation (e.g. the observed species count on the \( i \)th sampling occasion), and is assumed to follow a Poisson distribution with mean \( \lambda_i \). The logarithm of \( \lambda_i \), the expected value of the \( i \)th observation, is related linearly to a set of predictor variables, \( X_{ik} \), through the link function (McCullagh and Nelder, 1989).

For the Bayesian single-level model structure, normal prior distributions are used for the regression coefficients, \( \beta_k \), and Gamma distributions for the precision parameters, \( \tau_k \), as it is the conjugate prior for Poisson distribution (this specification follows Gelman (2006b) and Gelman et al. (2014a)).
\[ \beta_k \sim N(0, \tau_k) \]  
\[ \tau_k \sim \text{Gamma}(\alpha_1, \alpha_2) \]

Where, \( \tau_k \) is the precision of the Normal distribution (1/variance) and \( \alpha_1 \) and \( \alpha_2 \) are prior distribution parameters.

The *hierarchical model* follows the structure of the data generation model explained in 4.2.1. It is a two-level model which allows response-predictor variables’ relationships to vary among sampling groups, when they exist within the dataset. \( Z \), is the upper level predictor variable that describes the groups. Therefore, the regression coefficients, the response variables, and the predictors are indexed by the \( j \) groups, as \( \beta_{kj}, y_{ij}, \) and \( X_{kij} \).

\[ y_{ij} \sim \text{Poisson}(\lambda_y) \]  
\[ \log(\lambda_y) = \beta_0 + \beta_1 X_{ij} \]  
\[ \beta_{kj} \sim N(\phi_{kj}, \tau_k) \]

where, \( i \) refers to observations, \( j \) refers to groups, and \( k \) refers to predictor variables (\( k=0 \) represents the intercept). In the second level of the hierarchical model, the relationships between \( \beta_k \) in group \( j \) and the upper level predictor, \( Z_j \), are estimated. \( \beta_{kj} \) is normally distributed, with mean \( \phi_{kj} \) and precision \( \tau_k \), estimated by the following regression:

\[ \phi_{kj} = \gamma_{0k} + \gamma_{1k} Z_j \]  
\[ \gamma_{ik} \sim N(0, \tau_{ik}) \]  
\[ \tau_k \sim \text{Gamma}(\alpha_1, \alpha_2) \]

with the following parameters for the uninformative prior distributions (Gelman, 2006b):

\[ \alpha_1 = 2 \]  
\[ \alpha_2 = 0.5 \]  
\[ \tau_i = 0.001 \]

Given the above equations, it is demonstrated that in the hierarchical model, the mean values of \( \beta \)s, \( \phi_{ki} \), come from a Bayesian regression (equation 4-4).

To build up the models, five MCMC chains were run to achieve convergence, each with 12,000 iterations and no thinning. The first initial 2,000 samples of each chain were discarded as burn-in and the 10,000 additional samples were
saved for a total of 50,000 samples from which all inferences were made. I visually examined a random selection of trace plots from throughout the study space to verify that each chain achieves convergence. Figure 4-5 illustrates the convergence and density plots of a random parameter.

![Figure 4-5 The convergence and density plot of a random parameter](image)

To assess the convergence quantitatively, the potential scale reduction factor, $\hat{R}$, (Gelman et al., 2014a) was calculated for each parameter to verify that enough samples were obtained. Across all parameters, the calculated $\hat{R}$'s were less than 1.03, which indicated that enough samples were obtained and that the chains were converged.

The Bayesian software, JAGS (Plummer, 2003) was used to do the MCMC sampling for all models in this research. JAGS was accessed via the R statistical environment (R Core Team, 2015) using the R2jags (Su and Yajima, 2015), rjags (Plummer, 2016), and runjags (Denwood, 2016) packages.

4.2.4 Evaluating the models

After fitting the two models, single-level and hierarchical, the next step is to decide which model fits best to the dataset (Figure 4-2). To evaluate the fit of models, several goodness-of-fit tests were applied as the indicators of the models' predictive and descriptive capacities. Below, these goodness-of-fit tests are briefly described, while the actual results are analysed and further discussed in Chapter 5.
4.2.4.1 **Predictive performance check**

One way to assess models is to check the accuracy of their predictions. The predictive performance of model is the model’s capability to predict future data. After fitting a Bayesian model, its predictive accuracy is often needed to be measured either for evaluating the model or for model comparison and selection (Geisser and Eddy, 1979; Hoeting et al., 1999; Vehtari and Lampinen, 2002; Ando and Tsay, 2010; Vehtari and Ojanen, 2012).

The predictive performance of Bayesian models can be evaluated and compared in several ways. In point prediction, a single value is predicted for the unknown future observation, and the mean squared error is an applicable point prediction measure (Gelman et al., 2014a). In probabilistic prediction, the full uncertainty over the future observations is to be taken into account. Log-likelihood is a probabilistic prediction measure that is the basis of information criteria and the leave-one-out cross-validation (LOO-CV) technique for predictive performance check in this research (Gelman et al., 2014a).

I used the following information criteria and cross-validation technique to check the predictive performance of the models:

4.2.4.1.1 **Watanabe-Akaike or widely applicable information criterion (WAIC)**

WAIC is an information criterion to measure the predictive accuracy of a Bayesian model (Watanabe, 2010; Gelman et al., 2014b; Vehtari and Gelman, 2014). I used this criterion as it is a fully Bayesian approach and has been introduced as an improvement on the Deviance Information Criterion (DIC) for Bayesian models (Gelman et al., 2014a; Vehtari et al., 2016b); whereas, DIC is not fully Bayesian as it is based on a point estimate (Linde, 2005; Plummer, 2008). In model comparison, the model with the smaller WAIC fits better (Gelman et al., 2014b).

To apply the test, I used the `loo` package (Vehtari et al., 2016a) in R. As this package is mainly built up compatible with models in “Stan”, I extracted the MCMC data from JAGS and built the Poisson log-likelihood matrix, so that I could use this package on my models in JAGS.
The process of comparing model fit according to WAIC involved several steps. WAIC was calculated on each of the models at every dataset throughout the study space. The WAIC values of every pair of models fit for each simulated dataset at each point of the study space were compared, and the output contained two values: a comparison of the difference in expected predictive accuracy and the standard error of the difference (Vehtari et al., 2016a; Vehtari et al., 2016b). Knowing the outcome of the comparison of the expected predictive accuracy and the associated standard error, I calculated an approximate probability that the difference in predictive accuracy of the two models was greater than zero. Differences greater than zero indicate the second model in the comparison was the better fit (Vehtari et al., 2016a), which was the hierarchical model in our case. I used this approximate probability as a derived ‘WAIC score’ for every dataset in the study space. A WAIC score of 0 indicated that the single level model was the certainly better fit and a score of 1 indicated the hierarchical model was certainly the better fit.

Figure 4-6 shows an example WAIC distribution and scoring process. The first table in the figure, which is called “Expected predictive accuracy”, is showing the estimated WAIC. The second table, called “Standard error” is the estimated standard error of differences. Each cell of each table shows the corresponding results for every pair of models on a single dataset, and the rows of each complete table indicate the corresponding results for the seven datasets on each point of the study space; therefore, there are 7 columns and 10000 rows in each table. The pdf graph shows the approximation to the normal distribution with mean equal to the expected predictive accuracy of WAIC comparison result, and standard deviation derived from its standard error. The coloured area under the distribution represents the WAIC score of the dataset, which shows the strengths of the hierarchical modelling at that specific dataset. Columns 2 to 8 of the third table in the figure show the calculated WAIC score for all datasets throughout the study space.

As seven datasets have been generated on every point of the simulation study space, there were seven WAIC scores corresponding to each point, as
demonstrated in every row of the third table in Figure 4-6. I calculated the scores of points in the study space in two ways for later analysis:

1. The first scoring method was to take the average of these seven WAIC scores for each point in the study space. This result is called the WAIC average probability score.

2. The second scoring method was to classify every dataset as single-level or hierarchical, based on the WAIC score of the dataset. Based on the results of the seven datasets on this point I could label the given point in the study space as predominantly single-level or predominantly hierarchical (i.e., producing a binary classification of the points in the study space). This result is called the WAIC binary score for each point in the study space.

For the WAIC average probability scores, the seven WAIC scores of the seven datasets on each point were averaged, and a final score as the average probability score was given to every point. In Figure 4-6, for each selected row of the third table (indicating a point), all the seven columns (the seven WAIC scores of the seven datasets on that point) are averaged to produce the average probability score of that point. The scoring system was set in a way that the final WAIC average probability score showed the strength of hierarchical modelling on that point; whereas, the average probability score of 1 indicated that the hierarchical models were certainly the better fit to a specific point, while the average probability score of zero indicated the single-level models always fitted better according to WAIC test.

For the WAIC binary scores, I categorised the WAIC probability scores for each of the seven datasets of every point as either hierarchical or single-level according to the threshold of 0.5. Then the more frequent model structure over the seven datasets at each point was considered as the recommended model structure of that point. This led to have either of the two 0 and 1 values for every point, where 0 indicated the best fit of single-level models and 1 indicated the best fit of hierarchical models on the point.
Figure 4-6 The schematic process of WAIC scoring for every dataset, and the sum of WAIC scores of the seven datasets at every point of study space.

$X \sim N(\mu, \sigma^2)$

$p(X \geq 0) = 0.81643$
Table 4-1 and Table 4-2 show examples of calculating WAIC average probability scores and WAIC binary scores of the points by this process.

Table 4-1 WAIC average probability scores of the points with the WAIC scores of the seven datasets at the points

<table>
<thead>
<tr>
<th>set1</th>
<th>set2</th>
<th>set3</th>
<th>set4</th>
<th>set5</th>
<th>set6</th>
<th>set7</th>
<th>average probability score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.221313</strong></td>
<td>0.013961</td>
<td>0.413602</td>
<td>0.555851</td>
<td>0.012544</td>
<td>0.202323</td>
<td>0.000282</td>
<td><strong>0.20284</strong></td>
</tr>
<tr>
<td><strong>0.627394</strong></td>
<td>0.409187</td>
<td>0.993490</td>
<td>0.799577</td>
<td>0.248929</td>
<td>0.685222</td>
<td>0.080123</td>
<td><strong>0.549131</strong></td>
</tr>
<tr>
<td><strong>0.324564</strong></td>
<td>0.786029</td>
<td>0.449610</td>
<td>0.632882</td>
<td>0.989827</td>
<td>0.934345</td>
<td>0.937015</td>
<td><strong>0.722038</strong></td>
</tr>
<tr>
<td><strong>0.062953</strong></td>
<td>0.871506</td>
<td>0.042623</td>
<td>0.471623</td>
<td>0.806533</td>
<td>0.332089</td>
<td>0.089874</td>
<td><strong>0.382457</strong></td>
</tr>
<tr>
<td><strong>0.157649</strong></td>
<td>0.037077</td>
<td>0.553385</td>
<td>0.409752</td>
<td>0.710483</td>
<td>0.980127</td>
<td>0.530307</td>
<td><strong>0.482683</strong></td>
</tr>
<tr>
<td><strong>0.795111</strong></td>
<td>0.250465</td>
<td>0.430357</td>
<td>0.747484</td>
<td>0.791349</td>
<td>0.732868</td>
<td>0.133367</td>
<td><strong>0.554429</strong></td>
</tr>
<tr>
<td><strong>0.993765</strong></td>
<td>0.965456</td>
<td>0.896194</td>
<td>0.893920</td>
<td>0.935262</td>
<td>0.949105</td>
<td>0.982948</td>
<td><strong>0.945236</strong></td>
</tr>
<tr>
<td><strong>0.004654</strong></td>
<td>0.599368</td>
<td>0.695681</td>
<td>0.319090</td>
<td>0.771849</td>
<td>0.623477</td>
<td>0.070378</td>
<td><strong>0.440642</strong></td>
</tr>
</tbody>
</table>

Table 4-2 WAIC binary scores of the same points in Table 4-1; it is notable as the techniques of scoring are slightly different, the same point might be labelled differently by the two techniques

<table>
<thead>
<tr>
<th>set1</th>
<th>set2</th>
<th>set3</th>
<th>set4</th>
<th>set5</th>
<th>set6</th>
<th>set7</th>
<th>binary score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

4.2.4.1.2 Deviance information criterion (DIC)

I also applied DIC as it is known as the measure of choice in Bayesian application. There is a debate if WAIC is a better choice than DIC: while WAIC is fully Bayesian and is asymptotically equal to Bayesian leave-one-out cross-validation (Watanabe, 2010), the ease of computation of DIC in standard software is the key advantage of this criterion (Spiegelhalter et al., 2014). DIC is incorporated in the popular Bayesian modelling packages; such as BUGS and JAGS (Spiegelhalter et al., 1994, 2003; Spiegelhalter et al., 2002; Plummer, 2003). Similar to WAIC, in model comparison, the model with the smaller DIC fits better.
The DIC outcomes for each pair of the single-level and hierarchical models on every dataset were calculated and compared, and a score of zero or one was allocated depending if the single level or hierarchical model fit best on the given dataset. The average score was produced over the seven datasets per point throughout the simulation study space.

4.2.4.1.3 \( k \)-fold Cross-validation

Performing external validation is an appropriate approach to check a model's predictive performance; however, it is often desired to check the model before obtaining new data or waiting for the future to happen. Cross-validation is a measure of predictive accuracy that approximates external validation using already available data (Stone, 1977; Akaike, 1992). I used this technique due to the fact that the best way to measure a model’s fit is by external validation using a new independent replication of data (Gelman et al., 2014a), and cross-validation is a technique to achieve this (Vehtari et al., 2014). Cross-validation fits the model to the training data and then evaluates the predictive accuracy on a holdout set. A strong advantage of this technique is its capability of using the entire training set for testing (which is done gradually in several steps); this creates the largest possible test set (Rao et al., 2008). In model comparison, the model with the lower cross-validation error is deemed to fit better.

Cross-validation does not have the problem of overfitting; however, as it is all based on the available data, the true accuracy of the model’s predictions of the future data may be lower (Gelman et al., 2014a). Besides, this technique requires partitioning the data and fitting the model for many times, and thus is computationally expensive and time consuming when using MCMC to fit Bayesian models (Gelman et al., 2014a). In this research, I used 10-fold cross-validation, because 10 is a moderate value for \( k \) and makes the computation time plausible and more reasonable (Gelman et al., 2014a).

I applied 10-fold cross-validation by randomly partitioning the response data in every generated dataset into 10 subsets, each containing 20 data. A single subset is retained as the validation data for testing the model, and the remaining 9 subsets are used as training data. The model is fitted to the training
data, and the testing data is predicted by the model. This process is repeated 10 times with each of the 10 subsets used exactly once as the testing data.

I analysed the results for predictive performance check by applying $R^2$ and RMSE (Root-Mean-Square Error) on the predicted sets of data which were produced by the 10-fold cross-validation. For every dataset on each point of study space, the $R^2$ is calculated for each of the single-level and hierarchical models. From the two estimated values, the bigger one implies the more variation of data is explained by the model; therefore, the difference of the $R^2$ values are computed and an average $R^2$ score is produced over the seven datasets per point in the simulation study space. To calculate the total average RMSE scores, the same process was performed, except that in the comparison of two values, the smaller value shows the better fit of the model.

### 4.2.4.1.4 Leave-one-out cross-validation (LOO-CV)

The extreme case of cross-validation is the leave-one-out cross-validation (LOO-CV) with $n$ partitions, in which each holdout set represents a single data point (Stone, 1977; Shibata, 1989; Watanabe, 2010). The exact LOO-CV requires re-fitting the model with many different training sets. The approximate LOO-CV are easy to compute (Gelfand et al., 1992; Gelfand, 1996); importance sampling LOO-CV is a reliable and accurate approximate LOO-CV (Vehtari et al., 2014) (Vehtari and Gelman, 2015; Vehtari et al., 2016b; Vehtari et al., 2016c). Similar to WAIC test, I used the \texttt{loo} package in R to apply the LOO-CV test (Vehtari et al., 2016a). This package uses a Monte Carlo technique that takes samples from an approximating distribution, instead of the target distribution, and estimates some properties of the target distribution based on those samples (Ionides, 2008; Vehtari and Gelman, 2015). I applied both 10-fold cross-validation and LOO-CV to account for the bias that may arise from using only a single approach (James et al., 2013).

The procedure of applying of LOO-CV and processing the outputs were very much similar to those of WAIC; the output of comparing the LOO-CV results of the two models contained two values: a comparison of the difference in expected predictive accuracy and the standard error of the difference (Vehtari et
al., 2016a; Vehtari et al., 2016b). The same technique to calculate WAIC average probability score (explained in 4.2.4.1.1) was used to analyse the results and estimate the LOO-CV average probability scores for every point in the study space.

4.2.4.1.5 Mean squared error

Mean squared error is a measure of a model’s fit to the observed data. It can be summarised as:

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - E(y_i | y))$$

$y_i$ is the holdout observation that we want to predict, and $E(y_i | y)$ is the value of $y_i$ predicted by the model. The weighted version of this measure is also referred to as the Bayesian $\chi^2$ test (Gelman et al., 2004), and is represented as

$$\sum_{i=1}^{n} \frac{(y_i - E(y_i | y))^2}{\text{var}(y_i | y)}$$

In an MCMC simulation, $y_i$ is the holdout observation, $E(y_i | y)$ and $\text{var}(y_i | y)$ are the median and variance of the predicted values of $y_i$ across the MCMC steps, and the summation is across the holdout values (Gelman et al., 1996; Gelman et al., 2004; Lynch and Western, 2004; Gelman et al., 2014a). The mean squared error compares the average of the model errors to the variance of the predicted values. I applied the weighted version of this measure to evaluate my results.

I used this technique over the same 10 groups of holdout samples of the cross-validation. In this case, the results should be compared to $\chi^2 = 31.41$, considering the value of the $\chi^2$ table for 20 degrees of freedom and a p-value of 5%. If the observed deviations are significantly greater than this level, the models are appropriate to be used to make predictions.

4.2.4.2 Descriptive performance check

The descriptive performance of a model is the model’s performance to reproduce a system’s behaviour. I fitted both single-level and hierarchical models at each point of the study space. To evaluate how accurately each
model represents the system and compare which of the two models is more appropriate for the purpose of describing the system, the ability of the model to reproduce the available data using the model parameters should be checked (Gelman et al., 2014a). In order to do so, a couple of widely used goodness-of-fit measures were applied to the model; namely, R-squared, and RMSE. The fundamentals of these tests are briefly explained in this section.

4.2.4.2.1 The Coefficient of Determination - $R^2$

$R^2$ is a well-known statistical measure to assess a linear regression model. Technically, the linear regression is a method for estimating the parameters in a way that minimises the sum of the squared residuals (Weisberg, 2005). The $R^2$ measures the percentage of variation of the response variable that is explained by the model (Draper, 1998; Weisberg, 2005). Therefore, it measures how close the data are to the fitted regression line. In general if a model fits better, the predicted values are closer to the observed data, and so it leads to a higher $R^2$. However, $R^2$ can be increased artificially by adding the number of independent variables without actually improving the model’s fit to new data.

It is not possible to define $R^2$ for a Poisson response as the model relationship is between the predictor variables and the log link function rather than the response variable (McCullagh and Nelder, 1989). Therefore, I used a “pseudo-$R^2$” being the square of the Spearman’s correlation coefficient of the observed data and the median of the posterior distribution of the fitted values for the response variable (Stewart-Koster, 2011).

4.2.4.2.2 Root-mean-square error (RMSE)

This statistical measure is also known as the fit standard error, and is a measure of the average error. It squares the model errors (the difference between the actual and predicted values), averages them, and takes the square root to produce the result. It is a very common measure that makes an excellent general-purpose error metric for numerical predictions (Chai and Draxler, 2014). Similar to pseudo- $R^2$, I calculated the RMSE of every pair of hierarchical and single-level models on each dataset, compared the results, and then averaged
the results of the seven datasets on each point in order to produce the average RMSE scores, which is called it the descriptive RMSE scores.

The abovementioned two statistical measures were used to check the capabilities of the two model-types in reproducing the available datasets, and compare the capacities of the models in describing the datasets.

4.2.4.2.3 Posterior predictive checking

The posterior predictive approach to model checking is a technique for checking the fit of a Bayesian model to data (Rubin, 1981; Rubin, 1984; Gelman et al., 1996). If the Bayesian model fits, the replicated data generated under the model look similar to the observed data. In other words, the observed data should be credible under the posterior predictive distribution. This approach is a check for model consistency; any observed inconsistency (e.g., a very small or a very big posterior predictive p-value that indicates the observed data is located on either of the tails of the reference posterior predictive distribution) might be due to a poor model fit (Gelman et al., 2014a).

This standard technique to evaluate the fit of a Bayesian model to data is to compare the observed data with the posterior predictive distribution of the replicated data (produced by the model in MCMC simulation) (Gelman et al., 1996; Gelman et al., 2014a). Lack of fit of the data with respect to the posterior predictive distribution can be measured by the posterior predictive p-value, however, with the ease of graphical display of the distribution in related software, plotting the observed value against the distribution is recommended instead of only checking the p-value (Gelman et al., 1996). In a good fit of data, the posterior distribution will be centred near the observed value. A range of values between 0.05 and 0.95 can be considered reasonable for p-value (Lynch and Western, 2004; Gelman et al., 2014a).

I applied the posterior predictive checking approach on every single-level and hierarchical model on every dataset to check the descriptive performance of the model with respect to both the replicated data, and the known model parameters used to generate each simulated dataset (regression coefficients). As in the data generation step of the simulation, I am using the known attributes
of the data generation model, in order to generate the datasets, it gives me the possibility to check the model descriptive performance regarding the models’ coefficients, by comparing the parameters of fitted models against the known attributes of the point.

4.2.5 Analysing the results of models evaluations

A series of statistical techniques were used to analyse the results of the multiple model performance indicators and identify patterns in the model performance across the study-space. As explained before, every point of the study space is identified by its coordinates in the six-dimensional space. These coordinates are the parameters of the data generation model, which are called ‘attributes’ in this research (explained in 4.2.1). The six values of coordinates represent the values of the six attributes \( \gamma_{00}, \gamma_{01}, \sigma_0, \gamma_{10}, \gamma_{11}, \sigma_1 \) on that point. Therefore, the package of data to be analysed consisted of the attributes (coordinates) of each point, along with a set of scores corresponding to every goodness-of-fit measure for all points of the study space.

Regarding the predictive performance check, there are seven datasets:

- the WAIC average probability scores,
- the WAIC binary scores,
- DIC scores,
- LOO-CV average probability scores,
- 10-fold cross-validation \( R^2 \) scores,
- 10-fold cross-validation RMSE scores, and
- Mean squared error scores.

These are the response variables of each dataset, while the six values of the attributes of each point in the study space are the predictor variables in all of the seven datasets.

Similarly, regarding the descriptive performance check, there are three datasets:

- the \( R^2 \) scores,
- the RMSE scores, and
the posterior predictive checking scores.

These are the response variables of each dataset, while the six values of the attributes of each point in the study space are the predictor variables in all of the three datasets. Table 4-3 shows the response variable and predictor variables in each of these datasets.

Table 4-3 Response and predictor variables for datasets of predictive and descriptive performance check; the predictor variables are the same in all datasets, it is because the goal is to find the relationship between the attributes of the points (coordinates in the study space) and each of the measure results.

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Predictor variables (points attributes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIC average probability scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>DIC scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>LOO-CV average probability scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>10-fold-CV R² scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>10-fold-CV RMSE scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>PPL scores</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>( \gamma_{00} ), ( \gamma_{10} ), ( \gamma_{01} ), ( \gamma_{11} ), ( \sigma_0 ), ( \sigma_1 )</td>
</tr>
</tbody>
</table>

To analyse these datasets, four statistical techniques are used: Principal Component Regression, Partial Least Squares regression, multiple linear regression, and regression trees. As there are six predictor variables in these datasets, the analyses of the results would be very complicated, and interpreting the results and drawing inferences in order to set the model selection tool could prove to be difficult.
4.2.5.1 Multiple linear regression analysis (MLR)

Multiple linear regression analysis is an extension of simple linear regression and is used to predict the unknown value of the response variable from the known value of two or more predictor variables (Gelman and Hill, 2007). The MLR analysis allows determining the overall fit (with variance explained) of the model and the relative contribution of each of the independent variables to the total variance explained.

The coefficient value of the regression represents the mean change in the response given a one-unit increase in the predictor. However, when comparing the impact of coefficients in a regression, as the units vary between the different predictor variables, the regression predictor variables should be standardised before making any comparison. This process will put the different predictors on the same scale and allows comparing their coefficients directly. The standardised coefficient value represents the mean change in the response given a one standard deviation change in the predictor (Frost, 2016). Therefore, the MLR was applied on the standardised predictor variables.

4.2.5.2 Principal Component Regression (PCR)/ Partial Least Squares (PLS)

Principal Component Regression, firstly proposed by (Hotelling, 1957; Kendall, 1957), is a regression method based on Principal Component Analysis (PCA). PCA is a technique to identify patterns in data and highlight their similarities and differences (Smith, 2002; Abdi and Williams, 2010). This technique is very useful in high dimensional sets of data, where graphical representations are not attainable. It is a procedure to determine a smaller number of uncorrelated variables, called "principal components", from a set of data consisting of a large number of possibly interrelated variables, while retaining the maximum amount of variance present in the dataset (Jolliffe, 2002; Abdi and Williams, 2010; Jolliffe, 2014). The method reduces dimensionality in the case of a large number of variables and produces a smaller subset of synthetic variables which capture as much information as possible with high explained variance (Bair et al., 2006; Jolliffe, 2014). Thus, PCR is a "non-dependent" procedure and is applied without the consideration of the correlation between the dependent
variable and the independent variables (Maitra and Yan, 2008). PCR basically applies a PCA on independent variables, then runs a multiple linear regression on the selected components and computes the coefficients of the model that correspond to the response variable (Mosteller and Tukey, 1977; Mardia et al., 1979; Gunst and Mason, 1980).

Many studies show that, doing a PCR, the more significant components are the ones containing the more variance, and the deletion of components can be performed by considering the sizes of variances (Massy, 1965; Hocking, 1976; Mansfield et al., 1977; Gunst and Mason, 1980).

Partial Least Squares (PLS) is a supervised dimension reduction technique. The PLS is an appropriate alternative for PCR when we are interested in the predictive variables that capture as much information in the variables as well as in the relation between the predictive and response variables (Wold, 1975; Bair et al., 2006; Rosipal and Krämer, 2006; Maitra and Yan, 2008).

While PCR captures the maximum variance of the independent variables and MLR achieves the maximum correlation between independent variables and response, PLS does the two by maximizing the covariance of independent variables and response (Wold et al., 1984; Tobias, 1995; Maitra and Yan, 2008). Although in most cases, PCR and PLS produce similar results and prediction errors resulted by the two techniques have no significant difference (Wentzell and Montoto, 2003), I have tried both to see whether this is the case in my simulation data.

### 4.2.5.3 Regression Trees

The data to analyse are the points’ attributes (coordinates of the points in the study space) as predictor variables, and the outcome of each measure is the response. Therefore, the application of a classification technique is tenable in order to recognise the subspaces appropriate for the application of each model structure.

The Classification & Regression Trees methodology was first introduced in 1984 by Breiman et al. (Lewis, 2000), and refers to both types of decision trees: classification trees, and regression trees. Regression trees are used when the
response variable is continuous, whereas a classification tree is used when the response variable is categorical. A regression tree is a treelike diagram which splits the data into partitions and illustrates the choices available for decision making. The groups of data classified by the regression tree are shown in the terminal leaves of the tree.

Regression trees can be used to select the effective predictor variables for classification. In this research, this technique was used to classify the study space into sub-regions based on the six predictor variables (the ‘model attributes’). Regression trees have a simple structure and a fast calculating process to produce predicted values (Nisbet et al., 2009). They use a ‘winnowing process’ that separates important predictors from unimportant ones, and make classifications on that basis (Nisbet et al., 2009). The ability of regression tree generating algorithms to handle missing values (in addition to not having to make any parametric statistical assumptions, such as assumptions of normal data distribution, or linear relationships among the variables) has made them useful in many applications (Nisbet et al., 2009). The regression tree can gain its optimal size by pruning, which is the process of removing the parts that are less influential in classification (Izenman, 2008). The tree first grows to its largest size, and then pruning the tree results in the simplest possible model as the constraints and variables with a minimal predictive effect are pruned away (Breiman et al., 1984; Berk, 2008; Izenman, 2008). This process minimises the probability that an important structure in the dataset will be overlooked by stopping the tree growth too soon (Nisbet et al., 2009). Further details on this technique can be found in (Breiman et al., 1984; Berk, 2008; Izenman, 2008).

Regression trees can be used to select the effective predictor variables for classification; in this research, given a relatively large number of variables to analyse (six predictor variables), it would be useful to determine whether fewer variables could be used for classification and thus in the model selection tool. This can be achieved by reducing the dimension of analyses and simplify the problem by omitting the less influential variables. The application of regression
trees satisfied this need and classified the study space into sub-regions based on only a subset of the six predictor variables (the model attributes).

4.2.6 Partitioning the study space

In the next step, as in Figure 4-2, I classified the study space into sub-regions; this was performed using the output of regression tree generation. The study space was classified into single-level/ hierarchical/ or transitional sub-regions. Changes of the type of sub-regions in the study space were inspected against the changes of the attributes used for classification.

The name of the single-level/ hierarchical sub-regions indicates which model structure fitted best on most of the points of the sub-region. These sub-regions are the regions in which one of the two model structures was mostly recommended according to the results of the measures of the models' fit. In the transitional sub-regions, neither of the two model structures could be stated to fit best on a large majority of points of the sub-region.

4.2.7 Proposing a model-structure selection tool

In every subspace of the study space, the characteristics of the input data, the simulated datasets in this study, were studied and evaluated. Based on this information and knowing the corresponding model-type of the region, a model-structure selection tool was established as a basis to help modellers decide which model structure, namely, single-level or hierarchical, is more appropriate for any given dataset.

The use of the proposed tool was then demonstrated through three case studies in the field of ecology, based on three fish abundance data.

4.3 Summary

This thesis aims to provide a framework to determine whether, in a specific situation, a single-level model is appropriate or developing a hierarchical model is required considering the additional work and time it takes compared to developing a simple single level model.
To achieve the abovementioned objective this chapter described the details of a simulation study and the associated methods that can be used to determine guidelines for a model selection tool.

The goal of the simulation study was to determine, if we have a set of collected data, which type of the two models, namely, single-level and hierarchical, would be better.

For this study model structures were determined (selected) based on a selected domain where these models are frequently used in practice: both models were of Poisson regression in a Bayesian framework. According to these models’ parameters, a six dimensional study space was defined, and seven sets of data were generated throughout this space. Both types of model were fitted on every dataset, and a number of measures were employed to evaluate the predictive and descriptive capacities of these models.

In Chapter 5, the results would be analysed, consequently, and some subspaces in the study space were identified as the regions in which the single-level modelling was adequate, while in some other regions hierarchical modelling was recommended. Statistical techniques were used to calculate measures that distinguish the subspaces. Also the impact of model parameters on the subspaces of the study space was inspected.

In Chapter 6, the last step of this research would be carried out and reported. As this step involved many details and different techniques to identify and derive the important characteristics of simulated datasets in each sub-region of the study space, and to establish the proposed tool to select model-type according to data characteristics, a separate chapter is allocated to define the detailed process and report the results of this step.
Chapter 5
Comparative analysis of two statistical model structures
5.1 Introduction

After performing the simulations, which comprised the steps of generating data, fitting the models, applying different measures of fit on all the models and obtaining the outputs of each measure, the results are analysed and discussed in this chapter (Figure 5-1). The undertaken steps in this chapter are shown in Figure 5-2.
5.2 Analyses of Modelling results

According to the methodology described in section 4.2, the two model structures were fitted to the datasets on every point of the study space, and subsequently compared using several goodness-of-fit measures. The results of the measures of fit for the performance of both types of models were analysed with the four techniques explained in section 4.2.5; namely:

- Multiple Linear Regression (MLR),
- Principal Components Regression (PCR),
- Partial Least Squares (PLS), and
- Regression trees.

The results for each of these performance evaluations are discussed separately in this chapter.

In analysing the results of all measures of fit, neither PCR, nor PLS made any dimension reduction possible; and applying principal component analysis showed that there is no correlation among the points’ attributes (which are the predictor variables in the analyses at this chapter), and all resulted components accounted for the equal variances of data.
Therefore, applying the two (PCR / PLS) techniques in the case of this research was not useful, and produced the same results as MLR. This finding will only be illustrated for WAIC results, but is the same for all other measures’ results.

5.2.1 Models predictive performance

In this section, the results of every measure of each model’s predictive performance (explained in section 4.2.4.1) are analysed and individually explained. All results are compared and unified in section 5.3, and an integrated and consistent outcome for model predictive performance is provided. In most, but not all of the study space, hierarchical models provided an overall better fit than single-level models.

5.2.1.1 WAIC (average probability and binary score)

As explained in section 4.2.4.1.1, WAIC is one of the most powerful and reliable tests to compare two Bayesian models. The analyses of the outputs of different measures of fit in the following sections of Chapter 5 reveal that this measure is an appropriate choice to be used for the purpose of this research. Following the procedure explained in section 4.2.4.1.1, a WAIC average probability score and a WAIC binary score were calculated for every point of the study space.

Figure 5-3 shows the plot of WAIC average probability scores and of the WAIC binary scores. In this section, using the techniques explained in 4.2.5, the resulting scores are analysed.
Figure 5-3 The line plot of WAIC average probability scores and the histogram of WAIC binary scores; the WAIC average probability scores are continuous variables, whereas, the WAIC binary scores only hold two possible values

5.2.1.1.1 Principal component regression/ partial least squares (PCR/PLS)

In the first step, PCR and PLS regressions were applied. The two techniques produced the same outcomes; the calculated coefficients for WAIC average probability scores are shown in Table 5-1

Table 5-1 The PCR/ PLS regression coefficients for WAIC average probability scores

<table>
<thead>
<tr>
<th></th>
<th>WAIC average probability scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0983</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.0546</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>0.0482</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.0293</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1248</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.1079</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1212</td>
</tr>
</tbody>
</table>

The upper-level regression slope coefficient for the lower-level regression slope coefficient, $\gamma_{11}$, had the highest impact on WAIC scores, while the upper-level regression intercept coefficient for the lower-level regression slope coefficient, $\gamma_{01}$, had the least impact on the response variable. The standard deviation of
the distribution of the lower-level slope coefficient $\beta_1$, $\sigma_1$, and the standard deviation of the lower-level intercept $\beta_0$, $\sigma_0$, had also the second and third highest impacts on the response variable. The technique could explain 71% of the variance of the response variable (with an $R^2$ equal to 71%). Figure 5-4 shows the coefficients (including intercept) plot of PCR/PLS regression.

![Figure 5-4 Coefficients plot of PCR/ PLS regression for WAIC average probability scores](image)

The equal amounts of explained variance of predictor variables by each component (shown in Table 5-2) indicated that no dimension reduction could be applied without the loss of information and all components were significant. It means applying the PCA (the first step of PCR/PLS techniques) had no advantage, while the second step of PCR/ PLS is applying an MLR on the PCA results. Therefore, in this case, the PCR/PLS techniques performed as a simple MLR, and can be omitted from the analyses of the rest of measures of fit in this study, as well as the analysis of WAIC binary scores.

<table>
<thead>
<tr>
<th>Component</th>
<th>Explained variance of predictor variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component1</td>
<td>16.67%</td>
</tr>
</tbody>
</table>

5.2.1.1.2 Multiple linear regression (MLR)

There was a generally good fit of the MLR model for the WAIC average probability score ($R^2=0.71$). All of the predictor variables showed statistically significant relationships to the WAIC score. Both variance and slope parameters (i.e., the three variables of $\sigma_0$, $\sigma_1$, and $\gamma_{11}$) had comparatively large standardised
coefficients which showed the response variable was more affected by any changes in those three variables (Table 5-3).

Table 5-3 The MLR standardised coefficients for WAIC average probability scores; as the predictor variables were standardised, the standard errors by the MLR were the same.

<table>
<thead>
<tr>
<th></th>
<th>Estimated value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0983</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.0546</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>0.0482</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.0293</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1248</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.1079</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1212</td>
<td>0.001402</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>

5.2.1.1.3 Regression tree

As explained in section 4.2.5, I first applied the regression tree on my dataset with all six predictor variables (points’ attributes) and the WAIC average probability scores as the response variable.

The variables needed for the regression tree to partition the WAIC scores into classes that are significantly different to each other, were $\gamma_{11}$, $\sigma_0$, and $\sigma_1$ (Figure 5-5). These three variables are the ones with larger standardised coefficients in the MLR analysis. Each terminal leaf of the tree showed which model structure was fitting best in each subspace of the study space. The first leaf, identified a subspace where single-level models fit better than hierarchical models, being those datasets where the magnitude and variance of the group level slope and intercept coefficients was low ($\gamma_{11}<0.38$, $\sigma_0<0.15$, and $\sigma_1<0.15$) (Figure 5-6). All other leaves of the tree comprised subspaces where hierarchical models fit best on the majority of datasets, with leaf number 12 only containing the points that were fitted best with hierarchical models (Figure 5-6; Table 5-4).
Figure 5-5 The regression tree of the WAIC average probability score. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average WAIC average probability scores in each leaf, and $n$ shows the number of points in each leaf, the red number above the box shows the leaf number (sd refers to $\sigma$ and g refers to $\gamma$).

Figure 5-6 The boxplot of the WAIC average probability scores across all leaves of the regression tree in Figure 5-5. Observations above 0.5 indicate points in the subspace where hierarchical models fit best, indicating that the first terminal leaf, number 4, contains datasets where the majority fit best with a single level model.
Analysis of the change in Mean Squared Error from pruning the tree (with the plot of the output shown in Figure 5-7) resulted that no pruning should be applied on this regression tree, and the size of the tree remained with ten leaves.

![Figure 5-7 The MSE over the number leaves; the Mean Squared Error decreases with the increase of number of leaves, and no pruning is applicable.](Image)

<table>
<thead>
<tr>
<th>Leaf number</th>
<th>Number points best fitted with hierarchical models</th>
<th>Number of points best fitted with single-level models</th>
<th>% of the points best fitted with single-level models in the leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>179</td>
<td>1321</td>
<td>88.1</td>
</tr>
<tr>
<td>6</td>
<td>508</td>
<td>242</td>
<td>32.3</td>
</tr>
<tr>
<td>7</td>
<td>737</td>
<td>13</td>
<td>1.7</td>
</tr>
<tr>
<td>10</td>
<td>339</td>
<td>161</td>
<td>32.2</td>
</tr>
<tr>
<td>11</td>
<td>493</td>
<td>7</td>
<td>1.4</td>
</tr>
<tr>
<td>12</td>
<td>1000</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>16</td>
<td>570</td>
<td>180</td>
<td>24.0</td>
</tr>
<tr>
<td>17</td>
<td>744</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>18</td>
<td>1489</td>
<td>11</td>
<td>0.7</td>
</tr>
<tr>
<td>19</td>
<td>1994</td>
<td>6</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Then, I applied the classification tree on the WAIC binary scores dataset (Figure 5-8). The resulted classification tree had four leaves and performed the classification based on the three variables $\gamma_{11}$, $\sigma_0$, $\sigma_1$, and $\gamma_{10}$, same as the variables selected by the regression tree of WAIC average probability scores (Figure 5-5).

Similar to the results of the WAIC average probability scores, the only single-level sub-region is located where the values of the group level slope and intercept coefficients were low ($\gamma_{11}<0.38$, $\sigma_0<0.15$, and $\sigma_1<0.15$).
The leaves labelled as ‘Hierarch’ show the hierarchical sub-regions, and the leaf labelled as ‘Single-l’ represents the single-level leaf. The right-side number in each leaf shows the number of hierarchical points, and the left-side number shows the number of single-level points in the leaf.

5.2.1.2 Deviance information criterion (DIC)

Similar to WAIC average probability score analyses, MLR and regression tree were applied on DIC scores.

MLR resulted an $R^2$ equal to 60.9%; with the three largest standardised coefficients belonging to the three variables $\sigma_0$, $\sigma_1$, and $\gamma_{11}$. The DIC MLR coefficients are shown in Table 5-5, and the DIC regression tree is illustrated in Figure 5-9 and Figure 5-10. The results were fairly consistent with the WAIC average probability score results.

Table 5-5 MLR standardised regression coefficients of DIC scores; as the predictor variables were standardised, the standard errors by the MLR were the same.

<table>
<thead>
<tr>
<th></th>
<th>Estimated value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.7728</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.0535</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>0.0445</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.0287</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1288</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.1208</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1347</td>
<td>0.00188</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>
Figure 5-9: The regression tree for DIC scores. ‘sd’ refers to ‘\(\sigma\)’ and ‘g’ refers to ‘\(\gamma\)’. The decimal fractions in the boxes are the median values of DIC scores in each leaf, and \(n\) shows the number of points in each leaf, the red number above the box shows the leaf number.

Figure 5-10: The boxplot of DIC scores across all leaves of the regression tree in Figure 5-9. Observations above 0.5 indicate points in the subspace where hierarchical models fitted best, while the ones below 0.5 are the points on which single-level models fitted best.

Table 5-6 shows the number of hierarchical and single-level points based on DIC scores in every leaf of the tree (the points are labelled hierarchical if they
stand above the 0.5 line in Figure 5-10). The results were almost the same as the WAIC average probability score results.

Table 5-6 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the DIC scores regression tree

<table>
<thead>
<tr>
<th>Leaf number</th>
<th>Number of points best fitted with hierarchical models</th>
<th>Number of points best fitted with single-level models</th>
<th>% of the points best fitted with single-level models in the leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>180</td>
<td>1320</td>
<td>88.0</td>
</tr>
<tr>
<td>6</td>
<td>506</td>
<td>244</td>
<td>32.5</td>
</tr>
<tr>
<td>7</td>
<td>729</td>
<td>21</td>
<td>2.8</td>
</tr>
<tr>
<td>10</td>
<td>334</td>
<td>166</td>
<td>33.2</td>
</tr>
<tr>
<td>11</td>
<td>489</td>
<td>11</td>
<td>2.2</td>
</tr>
<tr>
<td>12</td>
<td>1000</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>16</td>
<td>576</td>
<td>174</td>
<td>23.2</td>
</tr>
<tr>
<td>17</td>
<td>739</td>
<td>11</td>
<td>1.5</td>
</tr>
<tr>
<td>18</td>
<td>988</td>
<td>12</td>
<td>1.2</td>
</tr>
<tr>
<td>19</td>
<td>2485</td>
<td>15</td>
<td>0.6</td>
</tr>
</tbody>
</table>

The selected attributes to classify the points, the number of leaves, and the number of hierarchical and single-level points located in every leaf in DIC scores regression tree were almost the same as those of WAIC average probability scores regression tree. It was expected, as the WAIC technique is built on the basis, and as an improvement, of the DIC technique. However, the results of MLR indicated a higher $R^2$ for WAIC average probability scores (equal to 71%) compared to DIC scores (equal to 60.9%).

5.2.1.3 10-fold cross-validation

By applying 10-fold cross-validation, as explained in section 4.2.4.1.3, a predicted value was obtained for every response data in each dataset. The predictive accuracy of the models was evaluated by both $R^2$ and RMSE methods.

According to the average $R^2$ scores analyses, MLR resulted in an $R^2$ equal to 50.6% with the three variables $\sigma_0$, $\sigma_1$, and $\gamma_1$, having the largest standardised coefficients (Table 5-7). The average $R^2$ regression tree performed the classification based on four variables ($\gamma_0$, $\gamma_1$, $\sigma_0$, and $\sigma_1$) (Figure 5-11), while the WAIC average probability scores regression tree classified the study space based on only three attributes ($\gamma_1$, $\sigma_0$, and $\sigma_1$). Figure 5-12 illustrates the boxplot for average $R^2$ scores.
Table 5-7 The MLR standardised regression coefficients for average $R^2$ scores; as the predictor variables were standardised, the standard errors by the MLR were the same.

<table>
<thead>
<tr>
<th></th>
<th>Estimated value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.7793</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.0489</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>-0.0023</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.0157</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.0942</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0819</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1113</td>
<td>0.00173</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>

Figure 5-11 The regression tree for average $R^2$ scores. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average $R^2$ scores in each leaf, and n shows the number of points in each leaf, the red number above the box shows the leaf number.
Figure 5-12 The boxplot of average $R^2$ scores across all leaves of the regression tree in Figure 5-11. Observations above 0.5 indicate points in the subspace where hierarchical models fitted best, indicating that the leaves number 5 & 6, contain datasets where the majority fit best with a single level model.

Table 5-8 shows the number of hierarchical and single-level points based on average $R^2$ scores in every leaf of the tree. Leaves number 5 and 6 represented the classes with best fit of single-level models in about 76% and 53% of datasets, respectively. The two sub-regions together were located where $\sigma_0 \& \sigma_1<0.15$, $\gamma_{11}<0.38$, which are the same conditions as of the single-level sub-region according to the regression tree of the WAIC average probability score.

According to the average RMSE scores analyses (as shown in Table 5-9 and Figure 5-13), both MLR (with an $R^2$ equal to 58.7%) and regression tree
revealed that $\gamma_{00}$ was an influential predictor variable and a significant attribute for classification. Figure 5-14 shows the boxplot of average RMSE scores.

Table 5-9 The MLR standardised regression coefficients for average RMSE scores; as the predictor variables were standardised, the standard errors by the MLR were the same.

<table>
<thead>
<tr>
<th></th>
<th>Estimated value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6760</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.0732</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>0.0263</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.0389</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.1187</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0940</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1353</td>
<td>0.00185</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>

Figure 5-13 The regression tree for average RMSE scores. ‘sd’ refers to ‘$\sigma$’ and ‘g’ refers to ‘$\gamma$’. The decimal fractions in the boxes are the median values of average RMSE scores in each leaf, and n shows the number of points in each leaf, the red number above the box shows the leaf number.
Figure 5-14 The boxplot of average RMSE scores across all leaves of the regression tree in Figure 5-13. Observations above 0.5 indicate points in the subspace where hierarchical models fitted best, indicating that the first terminal leaf, number 5, contains datasets where the highest majority fit best with a single level model.

Table 5-10 shows the number of points best fitted with hierarchical and single-level models based on average RMSE scores in every leaf of the tree. Leaf number 5 contains the highest percentage of points best fitted with single-level models relative to the other leaves. Also leaf number 6 has about the same number of points best fitted with single-level and hierarchical models, and leaf number 9 indicates the sub-region is about 58% single-level, while other sub-regions are more hierarchical.

<table>
<thead>
<tr>
<th>Leaf number</th>
<th>Number of points best fitted with hierarchical models</th>
<th>Number of points best fitted with single-level models</th>
<th>% of the points best fitted with single-level models in the leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>205</td>
<td>1295</td>
<td>86.3</td>
</tr>
<tr>
<td>6</td>
<td>247</td>
<td>253</td>
<td>50.6</td>
</tr>
<tr>
<td>9</td>
<td>252</td>
<td>348</td>
<td>58.0</td>
</tr>
<tr>
<td>10</td>
<td>446</td>
<td>154</td>
<td>25.7</td>
</tr>
<tr>
<td>11</td>
<td>676</td>
<td>124</td>
<td>15.5</td>
</tr>
<tr>
<td>12</td>
<td>985</td>
<td>15</td>
<td>1.5</td>
</tr>
<tr>
<td>13</td>
<td>615</td>
<td>285</td>
<td>31.7</td>
</tr>
<tr>
<td>17</td>
<td>808</td>
<td>92</td>
<td>10.2</td>
</tr>
<tr>
<td>18</td>
<td>1155</td>
<td>45</td>
<td>3.8</td>
</tr>
<tr>
<td>19</td>
<td>1930</td>
<td>70</td>
<td>3.5</td>
</tr>
</tbody>
</table>
5.2.1.4  \textit{LOO-CV}

As explained in section 4.2.4.1.4, the process to analyse the LOO-CV results were the same as of WAIC average probability scores. The two techniques, MLR, and regression tree were applied on the LOO-CV average probability scores dataset. The MLR standardised regression coefficients and the MLR $R^2$ value were the same as those of WAIC average probability scores with a difference of less than 2%. Also the leaves of the regression tree shared the same categories of the variables as those of the WAIC average probability scores analyses. Therefore, the LOO-CV results are not reported here, and, instead, the WAIC results are used in the next stages of this research.

5.2.1.5  \textit{Mean squared error}

The mean squared error values of 98.2\% of data fell above the critical value for the single-level models, and 95.6\% of data fell above the critical value for the hierarchical models. This result indicated that the model structures were suitable to be used for making predictions of system’s behaviour. Knowing that, the model structure with best fit can be determined by comparing the goodness-of-fit measures results.

5.2.2  \textit{Models descriptive performance}

In this section, the results of $R^2$ and RMSE measure for models descriptive capacity, as explained in section 4.2.4.2.1 and section 4.2.4.2.2, are analysed and explained individually. The results will be compared in section 5.3.

As explained later in this section and due to the results of descriptive measures of fit, hierarchical models are recommended in most of the study space. The dominance of best fit of hierarchical models based on these two measures may be because of the problem of overfitting; these two measures are affected by model parameters; which means that more complex models tend to show better fit according to these measures. This is more clearly seen in the results of RMSE, where the hierarchical models fitted best at every single point throughout the study space.
5.2.2.1 \( R^2 \)

MLR resulted in an \( R^2 \) equal to 42.1%. The MLR coefficients are shown in Table 5-11. Also the regression tree and boxplot of descriptive \( R^2 \) scores are illustrated in Figure 5-15 and Figure 5-16.

<table>
<thead>
<tr>
<th>Table 5-11 The MLR regression coefficients for the descriptive ( R^2 ) scores; as the predictor variables were standardised, the standard errors by the MLR were the same.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated value</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>( \gamma_{00} )</td>
</tr>
<tr>
<td>( \gamma_{01} )</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
</tr>
<tr>
<td>( \sigma_0 )</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
</tr>
</tbody>
</table>

In the regression tree, the descriptive \( R^2 \) scores set as the negative values indicated the best fit of single-level models, and the positive values indicated the best fit of hierarchical models. The classification was based on all six attributes, which would make the further inferences more complicated.

According to this measure of fit as shown in Table 5-12, only a very small portion of points were best fitted by single-level models. In addition, the two single-level classes had a low number of points with best fit of single-level models.
Figure 5-15 The regression tree of the descriptive $R^2$ scores. 'sd' refers to $\sigma$ and 'g' refers to $\gamma$. The decimal fractions in the boxes are the median values of descriptive $R^2$ scores in each leaf, and $n$ shows the number of points in each leaf, the red number above the box shows the leaf number.

Figure 5-16 The boxplot of the descriptive $R^2$ scores across all leaves of the regression tree in Figure 5-15. Observations above 0 indicate points in the subspace where hierarchical models fitted best, indicating that the leaves number 7 and 11 contain datasets where the majority fit best with a single level model.
Table 5-12 The number of points best fitted with hierarchical / single-level models and the percentage of points best fitted with single-level models in every leaf of the descriptive $R^2$ scores regression tree.

<table>
<thead>
<tr>
<th>Leaf number</th>
<th>Number of points best fitted with hierarchical models</th>
<th>Number of points best fitted with single-level models</th>
<th>% of the points best fitted with single-level models in the leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>10</td>
<td>152</td>
<td>93.8</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
<td>9</td>
<td>8.3</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>40</td>
<td>74.1</td>
</tr>
<tr>
<td>12</td>
<td>74</td>
<td>7</td>
<td>8.6</td>
</tr>
<tr>
<td>13</td>
<td>248</td>
<td>22</td>
<td>8.1</td>
</tr>
<tr>
<td>15</td>
<td>1694</td>
<td>106</td>
<td>5.9</td>
</tr>
<tr>
<td>16</td>
<td>864</td>
<td>36</td>
<td>4.0</td>
</tr>
<tr>
<td>18</td>
<td>900</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>19</td>
<td>1347</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>24</td>
<td>93</td>
<td>69</td>
<td>42.6</td>
</tr>
<tr>
<td>25</td>
<td>106</td>
<td>2</td>
<td>1.9</td>
</tr>
<tr>
<td>26</td>
<td>405</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>27</td>
<td>450</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>28</td>
<td>749</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>33</td>
<td>134</td>
<td>82</td>
<td>38.0</td>
</tr>
<tr>
<td>34</td>
<td>144</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>35</td>
<td>240</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>36</td>
<td>900</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>37</td>
<td>1000</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

5.2.2.2 RMSE

MLR resulted in an $R^2$ equal to 43.8%. The variable $\sigma_0$ had the smallest standardised coefficient, which was inconsistent with the results of other measures of fit (Table 5-13). As the regression tree and boxplot of the descriptive RMSE scores show (in Figure 5-17 and Figure 5-18), the partitioning was performed based on all six attributes, and the hierarchical models fitted best in all points of the study space.

Table 5-13 The MLR standardised regression coefficients for the descriptive RMSE scores; as the predictor variables were standardised, the standard errors by the MLR were the same.

<table>
<thead>
<tr>
<th></th>
<th>Estimated value</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.8669</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>0.4142</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>0.7855</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{01}$</td>
<td>0.2413</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.6735</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.1201</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.4</td>
<td>0.01376</td>
<td>&lt;2e-16</td>
</tr>
</tbody>
</table>
Figure 5-17 The regression tree of descriptive RMSE scores. ‘sd’ refers to \( \sigma \) and ‘g’ refers to \( \gamma \). The decimal fractions in the boxes are the median values of descriptive RMSE scores in each leaf, and \( n \) shows the number of points in each leaf, the red number above the box shows the leaf number.

Figure 5-18 The boxplot of the descriptive RMSE scores across all leaves of the regression tree in Figure 5-17. Observations above 0 indicate points in the subspace where hierarchical models fitted best; indicating that hierarchical models fitted best on all points in the study space.
5.2.2.3 Posterior predictive checking

As explained in section 4.2.4.2.3, I applied the posterior predictive checking approach on both single-level and hierarchical models for both the replicated data, and the model parameters (regression coefficients). Figure 5-19 illustrates two examples of the pre-known values of the model's coefficients against the MCMC output posterior distribution of the coefficient.

![Figure 5-19 a) The posterior distribution centred near the pre-known value of beta0, b) The pre-known value lied out of the acceptable range](image)

Applying the posterior predictive checking for a number of models' coefficients on every single-level and hierarchical model developed on each dataset in the study space, produced thousands of results. Every single result was checked; over 97% of observations lied within the acceptable area (between 0.05 and 0.95) under the posterior predictive distribution, while regarding the models' coefficients, this number decreased to not less than 70%.

5.3 Comparing the measures’ results and selecting the most appropriate goodness-of-fit measure

To enlarge the scope of the research and be comprehensive of most possible outcomes, a number of widely used measures of fit were applied on the models. So far, the result of WAIC average probability scores, WAIC binary scores, DIC scores, LOO-CV scores, average $R^2$ scores (for 10-fold cross-validation), average RMSE scores (for 10-fold cross-validation), descriptive $R^2$, and descriptive RMSE have been discussed. In this stage, it is necessary to
compare the measures and select an appropriate measure in order to build up a basis for further stages of this research.

Applying the MLR showed the relationship between the attributes (coordinates of points in the study space) in each of the measure’s results, while regression tree classified the study space based on the results of each measure. The application of both MLR and regression tree on all measures’ results revealed the followings:

- Using any goodness-fit-measures, the hierarchical models fitted best at most of the points in the study space.
- The MLR produced the largest value of $R^2$ for WAIC average probability scores; it meant the attributes (predictor variables of the MLR) could best explain the variance of the results of this measure than that of the results of other measures.
- WAIC binary scores were binary values, and, for classification of the study space based on those scores, a classification tree was used instead of a regression tree. The classification tree of WAIC binary scores used one more attribute for classification than the WAIC average probability scores’ regression tree; although the overall results of classification were consistent with WAIC average probability scores.
- DIC scores regression tree produced almost the same results as WAIC average probability scores, with a smaller number of points with best fit of single-level models in the single-level leaf of the tree.
- LOO-CV average probability scores produced almost the same results as WAIC average probability scores.
- The regression trees of 10-fold cross-validation (both average $R^2$ and RMSE scores) used an extra attribute for classification compared to the regression tree of the WAIC average probability score. Besides, the percentages of the points with best fit of single-level models in their single-level leaves were lower compared to that of the WAIC average probability scores.
- The descriptive $R^2$/RMSE had the smallest values of MLR $R^2$, and used all attributes for classification. The results of the descriptive $R^2$/RMSE
revealed that the hierarchical models fitted best at almost all the points in the study space. However, the superiority of hierarchical models in terms of $R^2$/RMSE results can mainly be due to the fact that adding more parameters to the model leads to yield a better $R^2$/RMSE result, regardless of whether this parameter addition improves the fit of the model (overfitting problem).

Considering abovementioned findings, I concluded WAIC average probability scores can be viewed as a reasonable appropriate measure as:

- Its results were consistent with all other measures.
- It used the least number of attributes for classification among other measures (three attributes) which is a very significant advantage in further analyses of this research (explained and applied in Chapter 6).
- It gained the highest $R^2$ in MLR analyses.
- It is a fully Bayesian criterion and theoretically is an appropriate choice for the model structures used in this research.

Consequently, WAIC (using the average probability scores method) was selected as the basis for further analyses in this research.

5.4 Partitioning study space based on WAIC results

As explained in section 5.2.1.1.3, the attributes used by the WAIC regression tree for classification of the study space were $\gamma_{ij}$, $\sigma_0$, and $\sigma_j$. Figure 5-5 and Figure 5-6 illustrated the regression tree and the distribution of WAIC average probability scores in each leaf of the tree. The more detailed plot over all changes of the three partitioning attributes, $\gamma_{ij}$, $\sigma_0$, and $\sigma_j$, is shown in Figure 5-20. It shows the gradual change of best fitting model structure over the space; every set of numbers on the left-hand axis shows the coordinates of the point in the 3D space in the order of $\gamma_{ij}$, $\sigma_0$, and $\sigma_j$ axes. Each horizontal box of the plot shows the distribution of the WAIC average probability scores of the points in 6D study space, that are located on the same point in this 3D space.
Figure 5-20 The box plot of the WAIC average probability scores of each set of points in the 6D study space located at the same points in the 3D space with $\gamma_{11}$, $\sigma_0$, and $\sigma_1$ axes. The x-axis shows the WAIC average probability scores, and the y-axis shows the values of $\gamma_{11}$, $\sigma_0$, and $\sigma_1$ of each set of points. By mapping the 6D study space over the 3D space, 128 points of the 6D space are mapped over each point of the 3D space.
Having set four values for each of $\sigma_0$, and $\sigma_1$ axes, and five values for $\gamma_{11}$ axis, there were 80 points in the 3D space; each representing 125 points of the 6D study space (see section 4.2.1). In each set of the 125 points, I estimated the percentage of points best fitted with single-level models, and allocated that percentage to one of the 80 points in the 3D space on which that set of 125 points was mapped. According to the percentage allocated to each point, I coloured the points in the 3D space. Figure 5-21 shows the plot of this 3D space. The colour of dots from blue to yellow show the suitability of the single-level models on the point; the very dark blue dots represent the points in which the single-level models fitted best with a probability of more than 95%.

![3D plot](image)

**Figure 5-21** The 3D plot of the model-type changes along the $\gamma_{11}$, $\sigma_0$, and $\sigma_1$ axes. Colours from blue to yellow show the suitability of fitting single-level models on the points. The darker the point is, the single-level models fitted best in more of the 125 points of the 6D space that were mapped on this coloured point in 3D space.

As the colour gets lighter, the probability of best fit of single-level compared to hierarchical models decreases. The light yellow dots represent the points in which hierarchical models fitted best with a probability of more than 95%. Figure 5-22 shows the same 3D space from two sides; the points are integrated into regions according to the leaves of the WAIC average probability scores.
regression tree in which they were classified (regression tree in Figure 5-5). As it shows, the space can be divided into four subspaces due to the suitability of the two model-types:

1. The dark blue region (leaf number 4 of the regression tree): the standard deviations of both the intercept and slope coefficients of the lower-level are low ($\sigma_0$ and $\sigma_1 < 0.15$), and the upper-level slope of the lower-level slope coefficient is also low ($\gamma_{11} < 0.375$); in this case, single-level fitted best, with a probability of 88%. The former condition ($\sigma_0$ and $\sigma_1 < 0.15$) can be stated as the precision of the lower-level coefficients are high (as shown in Table 5-14).

2. The two orange regions (leaves number 6 and 10 of the regression tree): $\sigma_1$ is low (less than 0.15), $\sigma_0$ is also low (less than 0.15) and $\gamma_{11}$ is medium (between 0.375 and 0.75), or $\sigma_0$ is medium (between 0.15 and 0.25) and $\gamma_{11}$ is low (less than 0.375); in both cases, single-level is less probable to best fit, with a probability of 32.2% (Table 5-14).

3. The yellow region (leaf number 16 of the regression tree): $\sigma_0$ and $\gamma_{11}$ are both low (less than 0.15 and 0.375, respectively), and $\sigma_1$ is medium (between 0.15 and 0.25); in this case, the probability of best fit of single-level level models further decreases to 24% and the hierarchical models seem to be more appropriate (Table 5-14).

4. The no-colour region (leaves number 7, 11, 12, 17, 18, and 19 of the regression tree): it can be described with the following two conditions: either, at least one of the three variable of $\sigma_0$, $\sigma_1$, and $\gamma_{11}$ are high (greater than 0.25, 0.25, and 0.75, respectively), or at least two of the three variables of $\sigma_0$, $\sigma_1$, and $\gamma_{11}$ are medium to high (greater than 0.15, 0.15, and 0.375, respectively); in either of the cases, the hierarchical models fitted best in more than 95% of points (Table 5-14).

The sub-regions 1 and 4 are called single-level and hierarchical sub-regions, respectively; they showed the strongest dominance of either of the two model structures. The sub-regions 2 and 3 are called transition sub-regions. In these tw sub-regions, the hierarchical models fitted best on a more number of points than the single-level models; however, the percentage of points fitted best with
Hierarchical models were much lower than sub-region 4 (Table 5-14). Table 5-15 shows the summary of the sub-regions of the study space based on WAIC results. In this table, different ranges on $\sigma_0$, $\sigma_1$, and $\gamma_{11}$ axes are categorised into groups of low, medium, and high.

Table 5-14 The 10 sub-regions defined by the leaves of the regression tree of the WAIC results, the associated values of attributes, and the probability of best fit of single-level models in every sub-region.

<table>
<thead>
<tr>
<th>Sub-region</th>
<th>$\sigma_0$</th>
<th>$\sigma_1$</th>
<th>$\gamma_{11}$</th>
<th>% of points best fitted with single-level models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-level sub-region (leaf#4)</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>88.1%</td>
</tr>
<tr>
<td>Transition sub-region (leaf#6)</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>32.27%</td>
</tr>
<tr>
<td>sub-region (leaf#10)</td>
<td>LOW</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>32.2%</td>
</tr>
<tr>
<td>sub-region (leaf#16)</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>24%</td>
</tr>
<tr>
<td>Hierarchical sub-region-1 (leaf#7)</td>
<td>HIGH</td>
<td>LOW</td>
<td>LOW</td>
<td>1.7%</td>
</tr>
<tr>
<td>sub-region-2 (leaf#11)</td>
<td>MEDIUM-HIGH</td>
<td>LOW</td>
<td>MEDIUM-HIGH</td>
<td>1.4%</td>
</tr>
<tr>
<td>sub-region-3 (leaf#12)</td>
<td>ALL</td>
<td>LOW</td>
<td>HIGH</td>
<td>0%</td>
</tr>
<tr>
<td>sub-region-4 (leaf#17)</td>
<td>LOW</td>
<td>HIGH</td>
<td>LOW</td>
<td>0.8%</td>
</tr>
<tr>
<td>sub-region-5 (leaf#18)</td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
<td>LOW</td>
<td>0.7%</td>
</tr>
<tr>
<td>sub-region-6 (leaf#19)</td>
<td>ALL</td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

$\sigma_0$ & $\sigma_1$: Low (<0.15), Medium (0.15< & <0.25), High (>0.25)  
$\gamma_{11}$: Low (<0.375), Medium (0.375< & <0.75), High (>0.75)

Table 5-15 Probability of best fit of single-level models based on values of attributes, as defined by the regression tree of the WAIC results.

<table>
<thead>
<tr>
<th>% of points best fitted with single-level models</th>
<th>$\sigma_0$</th>
<th>$\sigma_1$</th>
<th>$\gamma_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>88%</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
</tr>
<tr>
<td>32.2%</td>
<td>LOW</td>
<td>LOW</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>24%</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
</tr>
<tr>
<td>&lt;5%</td>
<td>HIGH</td>
<td>ALL</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>HIGH</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>ALL</td>
<td>HIGH</td>
</tr>
<tr>
<td></td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
</tr>
<tr>
<td></td>
<td>MEDIUM-HIGH</td>
<td>ALL</td>
<td>MEDIUM-HIGH</td>
</tr>
</tbody>
</table>
Figure 5-22 The 3D plot of the slow change of model-types over the regions from two sides. The blue region represents single level sub-region where 88% of the datasets simulated from this region were fit best with the single level model (Table 5-15). The white region of the study space represents the hierarchical region, where more than 95% of datasets simulated from this region were fit best with the hierarchical model (Table 5-15).
5.5 Summary

In this chapter, I analysed and compared a number of different measures of fit of the models. I then selected WAIC as a convenient choice of measure of fit to be used for this research, with the WAIC average probability scores as the basis for further analyses in this study. The results of WAIC average probability scores regression tree were used to classify the study space into subspaces of either mostly single-level or mostly hierarchical models regions.

Partitioning of the study space was performed over the attributes used in the classification of WAIC regression tree. The simultaneous changes of three attributes ($\sigma_0$, $\sigma_1$, and $\gamma_{11}$) were studied to identify the single-level and hierarchical subspaces, regarding the WAIC average probability scores regression tree.

Therefore, the subspaces created on WAIC results (including the regions of single-level models, the regions of hierarchical models, and the regions of the transition between the two), are the basis for the next chapter, where the characteristics of the data in each subspace are evaluated, and a tool is established to inform users when the additional effort of a hierarchical model would be worth, and when the single-level model structures are appropriate and sufficient.
Chapter 6
Establishing a tool for model-structure selection
6.1 Introduction

So far, I generated simulated datasets over my study space and fitted two model structures (single-level and hierarchical) for every dataset at each point in this study space (section 4.2). Subsequently, I derived different measures of fit for each model (section 5.2), and compared, analysed and synthesised the results (sections 5.3 and 5.4), to determine which model structure is appropriate in which regions of the study space.

In this chapter, I develop a tool to enable model developers to determine which model structure is appropriate for any given dataset, with an associated probability, before fitting either model (Figure 6-1). After developing this tool I apply it to an ecological case study to demonstrate how the proposed tool works.
6.2 Method to develop the model-structure selection tool

In this section, the process of developing the model-structure selection tool is described (Figure 6-2). The essence of this process is to identify some differentiating characteristics of datasets from the different sub-regions of the study space that were defined as being fitted best with either single-level or hierarchical model structures (Figure 5-22). By using these characteristics, the tool’s main function is that given a new dataset, it can inform the model developer and decision-maker which part of the study space the dataset most likely comes from. Therefore, users can select a suitable model structure.

The first step is to determine which datasets (out of all the simulated datasets, section 4.2.2) will be used to develop the tool. This step includes selecting the sub-regions of the study space and performing further data cleaning to exclude the outlier datasets in the sub-regions that would reduce the applicability of the tool. Having selected these datasets, a set of measures (based on simple
summary statistics) are developed and applied to find the differentiating characteristics between the two different types of datasets: single-level datasets, on which the single-level models fitted best, and hierarchical datasets, on which the hierarchical models fitted best.

In the next step, using model-based clustering, these single-level and hierarchical datasets are grouped based on the values of the informative measures to reveal the differentiating characteristics of each group; each group may contain both single-level and hierarchical datasets, and therefore, holds a probability for each type of datasets. The same algorithm is then used to predict where a new dataset is most probably located, based on the measures that are used in the tool.

Consequently, for a new given dataset, the analysis of the measures would reveal which model structure fits best before fitting any models.

6.2.1 Selecting sub-regions

In the previous chapter, the results of different measures of fit were analysed and investigated. As explained in section 5.3, the measure of WAIC was selected from the goodness of fit statistics to represent the fit of models, and provide the basis to develop the guidelines in this chapter. To ensure the guidelines are robust the analyses in this chapter are restricted to the regions of the study space where the better fit of either of the two model structures (single-level and hierarchical) is clearly dominant. The regions of the study space which hold a high uncertainty between the better fit of the two model structures (i.e. where each model fit best for some of the datasets) would not provide clear indications of the characteristics of datasets best fitted with either model structure.

6.2.2 Data cleaning

In addition to selecting a reduced set of sub-regions, a process of data cleaning was followed to remove datasets from a sub-region that were inconsistent with the results from the sub-region (i.e., the outlier datasets). For example, all datasets from the hierarchical sub-region that happened to be fitted best with a
single-level model were removed from the analysis at this stage. In this way, the datasets that are used to develop the guidelines are ‘representative’ of the type of dataset (single level or hierarchical) that defines each region.

6.2.3 Finding informative measures

The next step is to identify what information about a given dataset is required to determine to which sub-region of the study space it belongs without having to fit the two statistical model structures. To this end I calculated several summary statistics to describe the datasets within each sub-region for use in the analyses and development of the tool.

I trialled a large number of summary statistics and their combinations, and selected the ones that were most useful for developing the tool (i.e., those that provide more differentiating information between the single-level and hierarchical sub-regions).

The data are collected in a number of groups. The data spread in different groups in a single-level dataset (a dataset on which the single-level model fits best) follow the same trend, whereas, the groups of data in a hierarchical dataset (a dataset on which the hierarchical model fits best) follow different trends. Therefore, the groups of data in a hierarchical dataset can be expected to (more potentially) convey different characteristics. Among the measures I used, there is one measure that does not consider the groups of data in the datasets, while the rest of the measures do.

The proposed informative measures are as follows:


Coefficient of variation

Coefficient of variation is the standard deviation divided by the mean of a variable, thereby describing the amount of variability in the variable relative to its mean. As it is a unitless measure, it can be easily used to compare the spread of datasets with different units (Walpole et al., 2012).

\[
\mu = \frac{\sum x}{n}
\]
where $x$ is the variable with size $n$, $\mu$ is the mean of variable $x$, $\sigma$ is the standard deviation, and $cv$ is the coefficient of variation of variable $x$. This measure does not consider the groups in datasets.

**Standard deviation of within-group standard deviations over mean**

This measure is a combination of two summary statistics, the standard deviation and the mean. It takes the variation of group data as a characteristic of the group and reveals the dispersal of this characteristic over the dataset. To estimate this measure, the spread of data in each group is first calculated by the standard deviation; this is called the within-group standard deviation. Then the standard deviation of these values is calculated. Finally, the product is normalised by dividing by the mean of all data in the dataset (which is called the grand mean).

\[
\sigma_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n - 1}}
\]

\[
\sigma_w = \sqrt{\frac{\sum_{j=1}^{K} (\sigma_j - \bar{\sigma})^2}{K - 1}}
\]

where $K$ is the number of groups, $n$ is the number of data in each group, $N$ is the number of total data, $i$ refers to data in groups, $j$ refers to groups, $\bar{x}_j$ is the group mean, $\bar{x}$ is the grand mean, $\sigma_j$ is the within-group standard deviation, and $\sigma_w$ is the standard deviation of within-group standard deviations over mean.

**Within-group variance**

This measure is estimated by calculating the variance of data in each group, the within-group variance, taking the mean of the values, and then dividing it by the grand mean.
where \( K \) is the number of groups, \( n \) is the number of data in each group, \( N \) is the number of total data, \( i \) refers to data in groups, \( j \) refers to groups, \( \bar{x}_j \) is the group mean, \( v_{w} \) is the within-group variance.

### Between-group variance

I calculated the between-group variance as the variance of the mean values of every group of data, and then normalised the result by dividing it by the mean of all data in the dataset (Walpole et al., 2012).

\[
v_b = \frac{\sum_{j=1}^{K}(\bar{x}_j - \bar{x})^2}{K - 1} \tag{6-7}
\]

where \( K \) is the number of groups, \( j \) refers to groups, \( \bar{x}_j \) is the group mean, \( \bar{x} \) is the grand mean, \( v_b \) is the between-group variance.

### Total variance

The total variance is estimated by summation of the between-group and within-group variances.

\[
v_t = v_w + v_b \tag{6-8}
\]

### Variance ratio

The variance ratio shows the ratio of the between-group and within-group variances.

\[
v_r = \frac{v_b}{v_w} \tag{6-9}
\]

### 6.2.4 Model-based cluster analysis

The proposed measures described the characteristics of each of the single-level and hierarchical datasets. So far, by fitting the two model types, the datasets
were identified as hierarchical or single-level (section 5.4; Figure 5-22). The next step in the development of the guidelines is to again classify the datasets into new groups, which may or may not comprise only hierarchical or single-level datasets, based on the measures described in section 6.2.3. To ensure an objective classification in this step, I used model-based clustering with an information theoretic metric guiding model selection (Fraley and Raftery, 2002).

Model-based clustering is the automated grouping of related observations in a dataset (Fraley and Raftery, 2002). It uses a multivariate mixture model that classifies the observations in a dataset into groups according to a set of variables, in this case the informative measures described in section 6.2.3, using the multivariate normal distribution. The mixture likelihood approach maximises

$$L_{\text{mix}}(\theta_1, \ldots, \theta_G; \tau_1, \ldots, \tau_G \mid y) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(y_i \mid \theta_k)$$

where

- $y=(y_1, \ldots, y_n)$ is a vector of given observations,
- $i$ is the observation number,
- $n$ is in the number of observations,
- $k$ is the group number,
- $G$ is the number of groups,
- $f_k$ is the density of the $k$th group in the mixture model,
- $\theta_k$ is the parameters of the $k$th group in the mixture model,
- $\tau_k$ is the probability that an observation $y_i$ belongs to group $k$.

The probability of the group membership of each observation in the dataset is estimated by an unsupervised learning process (Banfield and Raftery, 1993). The outcome of this analysis is an estimated mean and variance of each variable for each cluster group. More details of the method can be found in Fraley and Raftery (2002), Fraley and Raftery (2005), and Fraley and Raftery (2007).

To apply the models-based clustering, I used the mclust (Fraley et al., 2017) package in R. This package is the most straightforward implementation of model-based clustering. It determines a maximum number of groups and
considers a set of mixture models, then implements the EM (Expectation-Maximisation) algorithms for maximum likelihood estimation for these models. It takes different covariance structures and different numbers of clusters (groups) into account. The covariance structure can be different due to the equal or varying volume, shape, and orientation of the covariances across groups. 14 different models with different characteristics are used in \textit{mclust} and the Bayesian Information Criterion (BIC) is used to determine the best fitting model and the number of groups (Scrucca et al., 2016; Fraley et al., 2017).

6.2.5 Developing the model-structure selection tool

The model-structure selection tool is built up based on the results of the model-based clustering and the classification of the sub-regions by the WAIC classification tree in section 5.2.1.1.3. The model-based cluster analysis of the simulated datasets (which were determined to be single-level/ hierarchical based on WAIC analysis), classifies these simulated datasets into $m$ groups using the informative measures (described in section 6.2.3). Having estimated the multivariate mixture model, the fitted model predicts the membership of any new dataset to the cluster groups according to the values of the informative measures. The \textit{mclust} classification of the new dataset is combined with the results from the classification of the sub-regions in section 5.4 (Figure 6-3) to derive a recommended model structure for the user to fit. Below, the steps are explained in details:
There are three probabilities that are combined for the tool:

1. The probability that a new dataset, $X$, is classified into each of the $m$ groups by mclust, $P_{X1}, \ldots, P_{Xm}$; the probability that dataset $X$ is in group $k$ ($k=1, \ldots, m$), given its characteristics (i.e. its informative measures). This is a direct output of the fitted mclust model, applied to the new dataset.
2. The probability that a simulated dataset in group $k$ (from the $m$ groups of the mclust analysis of the simulated datasets) is hierarchical or single level; $P_{k(h)}$ and $P_{k(s)}$.
3. The probability that the single-level (or hierarchical) sub-regions, as defined by the regression tree analysis in section 5.4, are indeed single-level (or hierarchical), calculated as the frequency of single-level (or hierarchical) datasets in the sub-regions that are defined as single-level (or hierarchical).
These three probabilities are combined to derive a probability that the new dataset, $X$, will be best fitted with a single level or hierarchical model in the following way:

```
\text{mclust} \text{ predicts the probability of the new dataset, } X, \text{ belonging to each of the } m \text{ groups } (P_{X1}, \ldots, P_{Xm}), \text{ and the dataset, } X, \text{ is classified as belonging to the group with the highest probability of membership.}
```

\begin{align*}
P(X \in \text{Group}_k) = P_{Xk} \quad \text{6-11}
\end{align*}

Each of the $m$ groups of the model-based clustering analysis of simulated datasets may contain either hierarchical or single-level datasets or both (as defined by the WAIC analysis). Therefore, for a dataset in an \text{mclust}-group, there is a probability of being single-level or hierarchical according to the frequency of single-level and hierarchical datasets in that group:

\begin{align*}
P_k(s) &= \frac{n_k(s)}{n_k}, \text{ for } k=1 \ldots m \quad \text{6-12} \\
P_k(h) &= \frac{n_k(h)}{n_k}, \text{ for } k=1 \ldots m \quad \text{6-13}
\end{align*}

where

- $k$ is the group number,
- $m$ is the number of groups,
- $n_k$ is the number of datasets in group $k$,
- $P_k(s)$ is the probability of a dataset of group $k$ being single-level,
- $n_k(s)$ is the number of single-level datasets in group $k$,
- $P_k(h)$ is the probability of a dataset of group $k$ being hierarchical,
- $n_k(h)$ is the number of hierarchical datasets in group $k$.

These two probabilities are combined to derive an interim estimate of which model structure fits best to the given dataset $X$:

\begin{align*}
P_X(s) &= \sum_{k=1}^{m} P_k(s) \times P_{Xk} \quad \text{6-14} \\
P_X(h) &= \sum_{k=1}^{m} P_k(h) \times P_{Xk} \quad \text{6-15}
\end{align*}

where

- $P_X(s)$ is the interim probability of dataset $X$ being located in the single-level sub-region of the study space,
\( P_X(h) \) is the interim probability of dataset \( X \) being located in the hierarchical sub-region of the study space.

The third probability, which is the underlying probability of the sub-region being either hierarchical or single level, is applied to derive the probability that the underlying model structure that produced the observed data is either single level or hierarchical (\( P(S=s|X) \) and \( P(S=h|X) \)):

\[
P(S = s | X) = P_{s\text{-region}} \times P_X(s) \tag{6-16}
\]

\[
P(S = h | X) = P_{h\text{-region}} \times P_X(h) \tag{6-17}
\]

where

- \( S \) is the model structure that produces the observed data and is either hierarchical or single-level,
- \( P_X(s) \) is as defined in equation 6-14,
- \( P_X(h) \) is as defined in equation 6-15,
- \( P_{s\text{-region}} \) is the frequency of single-level datasets in the single-level sub-region,
- \( P_{h\text{-region}} \) is the frequency of hierarchical datasets in the hierarchical sub-region.

It is notable that the sum of the two scores is less than one; this is because there is a third probability which I call the probability of the tool’s ignorance about dataset \( X \):

\[
P(X = \text{single-level}) + P(X = \text{hierarchical}) + P(X = \text{unknown}) = 1 \tag{6-18}
\]

where

- \( P(X=\text{unknown}) \) is the probability of the model-structure selection tool being not certain which model structure fits best on dataset \( X \).

The two final probabilities estimated in equation 6-16 and equation 6-17 are the scores estimated by the tool for the dataset \( X \) of being best fitted with single-level or hierarchical model structures. Therefore, for any given dataset \( X \), the tool selects the model structure which has the higher score, and reports the score as the associated reliability of the selection. Consequently, the output of the model-structure selection tool is one recommended model structure for the given dataset, with an associated reliability of the recommendation.
6.2.6 Applying the proposed tool to case studies

In this stage, the feasibility of the developed model-structure selection tool is demonstrated through three case studies and the results are reported. Also, both the single-level and hierarchical models were developed on each case study and the best fitting model was revealed. The results of modelling were compared with the outcome of the tool on every case study.

6.2.6.1 Case studies description

The case studies comprise distribution and abundance data of three fish species collected in the catchments of South East Queensland by researchers at the Australian Rivers Institute at Griffith University (Kennard, 2005; Kennard et al., 2006; Mackay, 2007). These data have been previously used to develop models for landscape ecology (Pusey et al., 2004; Kennard et al., 2006; Stewart-Koster et al., 2013). Subsequently, Stewart-Koster (2011) used these data to develop and compare both Bayesian single-level and hierarchical models and compare the fit of models on the datasets.

The data were collected at 17 sites in the Mary River between winter 1994 and winter 1997. A total of 168 observations were collected from all sites. A range of habitat variables were estimated for each sampling site. The fish species in the study include Melanotaenia duboulayi, Anguilla reinhardtii, and Tandanus tandanus. The set of predictor variables for each species in this study is taken from Stewart-Koster (2011). Table 6-1 shows the variables used for each fish species in the case study models. The selection of variables for each fish species was as the experts in the field understood their ecological relevance to fish species (Stewart-Koster, 2011). Further details on the data collection procedure and variable definitions are available in Kennard et al. (2006).
Table 6-1 The variables and their definitions used for each fish species (Stewart-Koster, 2011)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Melanotaenia duboulayi</th>
<th>Tandanus tandanus</th>
<th>Anguilla reinhardtii</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREA</td>
<td>Total sampled area of reach</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PLANTS</td>
<td>% reach area covered by plants</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMASS</td>
<td>% of stream bank covered by root masses</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>AVD</td>
<td>Average depth across reach</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AVL</td>
<td>Average water velocity across reach</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>MDF</td>
<td>Mean daily flow 4 months before sampling</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CVD</td>
<td>CV of daily flow of prior spring/summer</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>DISTM</td>
<td>Distance of reach to river mouth</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

In order to validate the outcomes of the tool, single-level and hierarchical models are developed on each fish species’ dataset and the results are compared by analysis of WAIC measure results on each model structure to each dataset.

6.2.6.2 The proposed tool’s outcomes

The modelling step revealed which model structure best fitted in each case of modelling according to WAIC measure. Subsequently, in this step, necessary summary statistics of the fish species datasets were calculated to and entered to the proposed tool. Then, the outcomes of the proposed tool for every fish species were compared to the modelling results.

6.2.6.3 Modelling results

To be consistent with the simulation study, I developed Bayesian Poisson regression models; the single-level model with two predictor variables, and the hierarchical model with one (same as the model structures explained in section 4.2.3). Following Stewart-Koster (2011), the first variable in Table 6-1 (AREA) was not assumed as a predictor variable; instead, it was used to adjust the species counts according to sampled reach area. The variable DISTM was the upper-level variable for the hierarchical models. Therefore, to develop the single-level models, the two predictor variables consisted of DISTM plus one other variable. In the hierarchical models, DISTM was the predictor variable of the upper-level regression of the model and the other predictor variable was in the lower-level regression.
The models I used to fit and compare are as following:

For the single-level models:

\[ y_i \sim Poisson(\lambda_i) \quad 6-19 \]
\[ \lambda_i = \mu_i \times AREA_i / 10 \quad 6-20 \]
\[ \log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 DISTM_i \quad 6-21 \]

where

- \( y_i \) is the \( i \)th observed data,
- \( \beta_k \sim N(0, \tau_k), \ k = 0,1 \)
- \( \tau_k \sim Gamma(\alpha_1, \alpha_2), \ k = 0,1 \)

and

- \( \tau_k \) = the precision of the Normal distribution
- \( \alpha_1 = 2, \ \alpha_2 = 0.5 \)

For the two-level hierarchical models:

\[ y_{ij} \sim Poisson(\lambda_{ij}) \quad 6-22 \]
\[ \lambda_{ij} = \mu_{ij} \times AREA_{ij} / 10 \]
\[ \log(\lambda_{ij}) = \beta_0 + \beta_{ij} X_{ij} \quad 6-23 \]
\[ \beta_{ij} \sim N(\phi_{ij}, \tau_k), \ k=0,1 \quad 6-24 \]

where, \( i \) refers to observations, \( j \) refers to groups.

The second level of the hierarchical model:

\[ \phi_{ij} = \gamma_{0k} + \gamma_{1k} DISTM_j, \ k=0,1 \quad 6-25 \]
\[ \gamma_{lk} \sim N(0, \tau_{lk}), \ l,k = 0,1 \quad 6-26 \]
\[ \tau_k \sim Gamma(\alpha_1, \alpha_2), \ k = 0,1 \quad 6-27 \]

and

- \( \alpha_1 = 2 \)
- \( \alpha_2 = 0.5 \)
- \( \tau_1 = 0.001 \)

Therefore, for each pair of variables (one is certainly the DISTM), one set of models (including the single-level and the hierarchical ones) were developed.
This way, four sets of models were developed for the \textit{A. reinhardtii} dataset, and three sets of models were developed for each of \textit{M. duboulayi}, and \textit{T. tandanus} datasets (Table 6-2).

The fit of models in every set were compared according to the WAIC value gained from the model.

<table>
<thead>
<tr>
<th>Table 6-2 Variables used in models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sets of models</td>
</tr>
<tr>
<td>----------------</td>
</tr>
</tbody>
</table>
| \textit{Anguilla reinhardtii} dataset | \begin{tabular}{l}
1 DISTM, AVD \\
2 DISTM, AVL \\
3 DISTM, MDF \\
4 DISTM, CVD \\
\end{tabular} |
| \textit{Melanotaenia duboulayi} dataset | \begin{tabular}{l}
1 DISTM, PLANTS \\
2 DISTM, AVD \\
3 DISTM, MDF \\
\end{tabular} |
| \textit{Tandanus tandanus} dataset | \begin{tabular}{l}
1 DISTM, RMASS \\
2 DISTM, AVD \\
3 DISTM, AVL \\
\end{tabular} |

\subsection*{6.2.7 Demonstrating applicability of the proposed tool on simulated datasets}

To test the applicability of the proposed tool in a bigger range, I applied it on several groups of generated datasets at the simulation stage. As explained in section 6.2.1, 3 sub-regions of the study space (represented by 3 leaves of the WAIC average probability score regression tree in Figure 5-5) have been omitted from the datasets' analyses in the stage of establishing the tool, as neither of the model structures showed a high probability of the best fit on the datasets locating in the sub-regions. Also, in the data cleaning step (explained in section 6.2.2), further datasets were removed in the analysed sub-regions. At this stage, I used all these omitted datasets to demonstrate the applicability and validity of the proposed tool in making reliable predictions.

In the simulation stage, both single-level and hierarchical models were fitted to all these datasets and it is already known which model structure fits best for each dataset. In this regard, for the purpose of this stage, I grouped the
datasets into the following sets and examined them separately using the proposed tool:

- Set 1: datasets with best fit of single-level models locating in the omitted sub-regions
- Set 2: datasets with best fit of hierarchical models locating in the omitted sub-regions
- Set 3: datasets with best fit of single-level models omitted from the six analysed hierarchical sub-regions
- Set 4: datasets with best fit of hierarchical models omitted from the analysed single-level sub-region

### 6.3 Description of results

In this section, the results relevant to every step of the proposed method (defined in section 6.2) are explained in detail.

#### 6.3.1 Selecting sub-regions

As explained in 6.2.1, to build up an indubitable and confident basis for the inferences, only the regions of the study space with a high dominance of either of the two model structures (single-level and hierarchical) were selected.

As described in section 5.2.1.1.3, the WAIC average probability scores regression tree classified the study space into ten sub-regions (represented by the ten leaves of the regression tree; Figure 5-5 and Figure 5-6). Three of the six attributes of the study space points were used to perform the classification, because the values of the other three attributes were not effective in this classification; these effective attributes were $\gamma_{1}, \sigma_{0}$, and $\sigma_{1}$ (Figure 5-5).

Out of the ten sub-regions of the study space, six sub-regions were classified as hierarchical regions, in which the hierarchical models were recommended in more than 95% of datasets, and one sub-region was recognised as single-level region, in which single-level models were recommended in 88% of datasets (Table 6-3).
The values of the three effective attributes according to every sub-region are shown in the table. The values are categorised into three ranges; high, medium, and low depending on the values of the parameters at each split in the regression tree.

The table shows the probability of the best fit of the model structure regarding the sub-regions. This probability was calculated as the frequency of datasets with best fit of the dominant model structure of the sub-region in that sub-region (section 5.4). All six hierarchical sub-regions were also joined together to build one integrated hierarchical sub-region. The frequency of hierarchical datasets in this integrated hierarchical sub-region was ~99%.

There are clear differences among the parameters that make up the datasets in the different sub-regions (Figure 6-4). I randomly selected a sample dataset of each sub-region for the purpose of demonstrating and comparing the sub-regions in this figure. The values of the attributes $\gamma_{11}$, $\sigma_0$, and $\sigma_1$ show the position of the datasets in the study space (presented in the first column of the figure).

The second and third columns show the values of the lower-level regression coefficients ($\beta$s) of the hierarchical models built on the same datasets; the hierarchical model gives out ten values of every $\beta$ as there were ten groups of data in every generated dataset of my simulation study.

---

### Table 6-3 Single-level and hierarchical subregions based on regression tree of WAIC average probability scores as described in Section 5.2.1.1.3. In the first column, the number of the leaf in the regression tree corresponding to the sub-region is listed in brackets.

<table>
<thead>
<tr>
<th>Sub-region</th>
<th>Standard deviation of the lower-level intercept coefficient ($\sigma_0$)*</th>
<th>Standard deviation of the lower-level slope coefficient ($\sigma_1$)*</th>
<th>Upper-level slope coefficient of the lower-level slope coefficient ($\gamma_{11}$)**</th>
<th>Probability of the best fit of the dominant model in sub-regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-level sub-region (leaf#4)</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>88.1%</td>
</tr>
<tr>
<td>Sub-region-1 (leaf#7)</td>
<td>HIGH</td>
<td>LOW</td>
<td>LOW</td>
<td>98.3%</td>
</tr>
<tr>
<td>Sub-region-2 (leaf#11)</td>
<td>MEDIUM-HIGH</td>
<td>LOW</td>
<td>MEDIUM-HIGH</td>
<td>98.6%</td>
</tr>
<tr>
<td>Sub-region-3 (leaf#12)</td>
<td>ALL</td>
<td>LOW</td>
<td>HIGH</td>
<td>100%</td>
</tr>
<tr>
<td>Sub-region-4 (leaf#17)</td>
<td>LOW</td>
<td>HIGH</td>
<td>LOW</td>
<td>99.2%</td>
</tr>
<tr>
<td>Sub-region-5 (leaf#18)</td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
<td>LOW</td>
<td>99.3%</td>
</tr>
<tr>
<td>Sub-region-6 (leaf#19)</td>
<td>ALL</td>
<td>MEDIUM-HIGH</td>
<td>MEDIUM-HIGH</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

* $\sigma_0$ & $\sigma_1$: Low (<0.15), Medium (0.15<$\sigma$<0.25), High (>0.25)
** $\gamma_{11}$: Low (<0.375), Medium (0.375<$\gamma$<0.75), High (>0.75)
The fourth column shows the relationship between lambda and the predictor variable $X$ for each group in each dataset. As defined in equation 4-2, in this simulation study, the logarithm of the mean of the Poisson distribution, $\lambda$, is the response variable of the lower-level regression.

As shown in Figure 6-4, in the single-level sub-region, the values of $\beta_0$s and $\beta_1$s were almost constant among different groups; this resulted in having similar patterns of the relationship between the log transformed of response and the predictor variables in different groups. All groups followed nearly the same trend in the single-level region, while the trends of different groups had varying intercepts and slopes in the hierarchical regions.

The six hierarchical sub-regions are labelled as hierarchical for different reasons; a big value (>0.75) of upper-level slope coefficient ($\gamma_{11}$) can create dispersed lower-level coefficients ($\beta$) even if the standard deviations of $\beta$ distributions are small (as in hierarchical sub-region-3 in Figure 6-4), whereas, a big value (>0.25) of standard deviation of either of $\beta$ distributions classifies the sub-region as hierarchical, even if the slope coefficient in upper-level regression is zero (as in hierarchical sub-regions 1 and 4 in Figure 6-4). As it can be seen in the table, having a value equal or greater than medium for at least two of the three effective attributes is another reason to label the sub-region as hierarchical (as in hierarchical sub-regions 2, 5, and 6 in Figure 6-4).
### Chapter 6  Establishing a tool for model-structure selection

<table>
<thead>
<tr>
<th>Sub-regions</th>
<th>X axis: site#, Y axis: $\beta_0$</th>
<th>X axis: site#, Y axis: $\beta_1$</th>
<th>X axis: predictor variable $x$, Y axis: Log$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-level sub-region</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=0$, $\sigma_0 = 0.05$, $\sigma_1 = 0.05$)</td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-1</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=0$, $\sigma_0 = 0.3$, $\sigma_1 = 0.05$)</td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-2</td>
<td><img src="image13" alt="Graph" /></td>
<td><img src="image14" alt="Graph" /></td>
<td><img src="image15" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=0.5$, $\sigma_0 = 0.3$, $\sigma_1 = 0.05$)</td>
<td><img src="image16" alt="Graph" /></td>
<td><img src="image17" alt="Graph" /></td>
<td><img src="image18" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-3</td>
<td><img src="image19" alt="Graph" /></td>
<td><img src="image20" alt="Graph" /></td>
<td><img src="image21" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=1$, $\sigma_0 = 0.05$, $\sigma_1 = 0.05$)</td>
<td><img src="image22" alt="Graph" /></td>
<td><img src="image23" alt="Graph" /></td>
<td><img src="image24" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-4</td>
<td><img src="image25" alt="Graph" /></td>
<td><img src="image26" alt="Graph" /></td>
<td><img src="image27" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=0$, $\sigma_0 = 0.05$, $\sigma_1 = 0.3$)</td>
<td><img src="image28" alt="Graph" /></td>
<td><img src="image29" alt="Graph" /></td>
<td><img src="image30" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-5</td>
<td><img src="image31" alt="Graph" /></td>
<td><img src="image32" alt="Graph" /></td>
<td><img src="image33" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=0.25$, $\sigma_0 = 0.3$, $\sigma_1 = 0.3$)</td>
<td><img src="image34" alt="Graph" /></td>
<td><img src="image35" alt="Graph" /></td>
<td><img src="image36" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-6</td>
<td><img src="image37" alt="Graph" /></td>
<td><img src="image38" alt="Graph" /></td>
<td><img src="image39" alt="Graph" /></td>
</tr>
<tr>
<td>($\gamma_{11}=1$, $\sigma_0 = 0.3$, $\sigma_1 = 0.3$)</td>
<td><img src="image40" alt="Graph" /></td>
<td><img src="image41" alt="Graph" /></td>
<td><img src="image42" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Figure 6-4** Example datasets’ properties of single-level and hierarchical sub-regions.
Figure 6-5 shows the expected response variable under the model against the actual ones. The dashed lines in the plots illustrate the one-to-one line of a perfect fitting model. The dispersion of points around the dashed lines indicates the fit of models.

As demonstrated, in the hierarchical sub-regions, data are gathered closer to the dashed lines in the hierarchical models compared to those in the single-level models, which shows the better fit of hierarchical models to the data in the sub-regions.

As opposed to the above, in the single-level sub-region, the single-level and the hierarchical model show very similar patterns, while the lower WAIC score in the single-level model indicates the better fit of this model in spite of the additional parameters in the hierarchical model.

Beside this, the error bars around each median expected value indicates that the uncertainty surrounding each fitted value is bigger in the hierarchical models than that of the single-level ones in all seven sub-regions. However, the accuracy of expected values is greatly improved in the hierarchical models in most of the study space.
### Sub-regions

<table>
<thead>
<tr>
<th>Sub-region</th>
<th>Single-level model</th>
<th>Hierarchical model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical sub-region-2</td>
<td>((\gamma_1=0.5, \sigma_0=0.3, \sigma_1=0.05))</td>
<td><img src="" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-3</td>
<td>((\gamma_1=1, \sigma_0 &amp; \sigma_1=0.05))</td>
<td><img src="" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-4</td>
<td>((\gamma_1=0, \sigma_0=0.05, \sigma_1=0.3))</td>
<td><img src="" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-5</td>
<td>((\gamma_1=0.25, \sigma_0=0.3, \sigma_1=0.3))</td>
<td><img src="" alt="Graph" /></td>
</tr>
<tr>
<td>Hierarchical sub-region-6</td>
<td>((\gamma_1=1, \sigma_0 &amp; \sigma_1=0.3))</td>
<td><img src="" alt="Graph" /></td>
</tr>
</tbody>
</table>

Figure 6-5 Plots of the expected response variable under the model against the actual ones. Error bars shown are the 0.025 and 0.975 percentiles of the MCMC samples for each expected value. The dashed line shows the one to one line of a perfectly fitting model. In each model, \(n=200\).
6.3.2 Data cleaning

The process of data cleaning involved removing a considerable number of datasets. Out of the total number of 10000 points of my study space (as in section 4.2.1), 1500 points were located in the single-level sub-region (Figure 5-5). Since seven datasets were generated on each point, $7 \times 1500 = 10500$ datasets were existing in the single-level sub-region (Table 6-4). The sub-region with the highest number of existing datasets was the hierarchical sub-region-6; after applying data cleaning, it had still the highest number of datasets. The hierarchical sub-region-3 had the highest probability of the best fit of the model throughout the sub-regions (100%), (Table 6-3), and lost the least number of datasets in the process of data cleaning (Table 6-4).

<table>
<thead>
<tr>
<th>Sub-region</th>
<th>Number of points of study space per region*</th>
<th>Number of datasets per region*</th>
<th>Number of datasets per region after data cleaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-level sub-region</td>
<td>1500</td>
<td>10500</td>
<td>7399</td>
</tr>
<tr>
<td>Hierarchical sub-region-1</td>
<td>750</td>
<td>5250</td>
<td>4440</td>
</tr>
<tr>
<td>Hierarchical sub-region-2</td>
<td>500</td>
<td>3500</td>
<td>3146</td>
</tr>
<tr>
<td>Hierarchical sub-region-3</td>
<td>1000</td>
<td>7000</td>
<td>6922</td>
</tr>
<tr>
<td>Hierarchical sub-region-4</td>
<td>750</td>
<td>5250</td>
<td>4680</td>
</tr>
<tr>
<td>Hierarchical sub-region-5</td>
<td>1500</td>
<td>10500</td>
<td>9753</td>
</tr>
<tr>
<td>Hierarchical sub-region-6</td>
<td>2000</td>
<td>14000</td>
<td>13502</td>
</tr>
<tr>
<td>Total hierarchical region</td>
<td>6500</td>
<td>45500</td>
<td>42443</td>
</tr>
<tr>
<td>All selected regions of study space</td>
<td>8000</td>
<td>56000</td>
<td>49842</td>
</tr>
</tbody>
</table>

* Number of datasets per region = Number of points of study space per region $\times$ 7

6.3.3 Applying informative measures

The selected six measures in section 6.2.3 were applied to each of the simulated datasets (after data cleaning) and the outputs formed a distribution of each measure in every one of the seven sub-regions defined by the WAIC regression tree analysis in section 5.2.1.1.3 (Figure 6-6).
<table>
<thead>
<tr>
<th>Coefficient of variation</th>
<th>Density plots of informative measures for single-level and all six hierarchical sub-regions</th>
<th>Density plots of informative measures for single-level and the integrated hierarchical sub-regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td><img src="image" alt="Coefficient of variation" /></td>
<td><img src="image" alt="Coefficient of variation" /></td>
</tr>
<tr>
<td>standard deviation of within-group standard deviations</td>
<td>(C)</td>
<td>(D)</td>
</tr>
<tr>
<td>within-group variance</td>
<td>(E)</td>
<td>(F)</td>
</tr>
<tr>
<td>Density plots of informative measures for single-level and all six hierarchical sub-regions</td>
<td>Density plots of informative measures for single-level and the integrated hierarchical sub-regions</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><a href="#">Image</a></td>
<td><a href="#">Image</a></td>
<td></td>
</tr>
<tr>
<td><a href="#">Image</a></td>
<td><a href="#">Image</a></td>
<td></td>
</tr>
<tr>
<td><a href="#">Image</a></td>
<td><a href="#">Image</a></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-6 Density plots of all informative measures in different sub-regions. The first column of plots show the informative measures for single-level sub-region (black line) and for every hierarchical sub-regions (six coloured lines), the second column of plots show the informative measures for single-level sub-region (black line) and for the integrated hierarchical sub-region, which is built by joining all six hierarchical sub-regions together (red line).

The density plots of the results (Figure 6-6) show that the hierarchical sub-regions were more right-skewed (i.e. the cover larger values of most of the measures). The density plots for the coefficient of variation and the standard
deviation of within group standard deviations had very little difference between single-level and hierarchical datasets (Figure 6-6 A-D). The overlap between the single-level and hierarchical curves decreases as we get closer to the bottom of the figure; the density plots for the variance ratio (Figure 6-6 K-L) have the largest difference between the two types of datasets, nonetheless there is considerable overlap among several of the hierarchical sub-regions. Although the density curves of these measures in different sub-regions are close with large overlaps, the measure provides informative details about the sub-regions.

Figure 6-7 shows the scatter plot for each pair of measures. Hierarchical and single-level datasets are represented by blue and red colours, respectively.

As shown in the figure, the single-level datasets are mostly surrounded by the hierarchical ones. This was expected as the measures cover bigger ranges of values in the hierarchical sub-region compared to the single-level sub-region. It
can be easily seen in Figure 6-6; the hierarchical sub-regions’ density curves have bigger dispersion than the single-level ones, and enclose a wider ranges of values.

### 6.3.4 Model-based cluster analysis

As explained in 6.2.4, I used \texttt{mclust} to categorise my datasets based on the values of the six informative measures (Table 6-5 shows 10 sample rows of the whole dataset). There were a total of 49842 datasets to be clustered; 42443 hierarchical and 7399 single-level datasets.

<table>
<thead>
<tr>
<th>standard deviation of within-group standard deviations over mean</th>
<th>Coefficient of variation</th>
<th>between-group variance</th>
<th>within-group variance</th>
<th>variance ratio</th>
<th>Dataset type as defined by WAIC analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8438</td>
<td>16.4603</td>
<td>0.6212</td>
<td>1.04002</td>
<td>15.4203</td>
<td>14.8269 hierarchical</td>
</tr>
<tr>
<td>2.6941</td>
<td>247.648</td>
<td>0.28125</td>
<td>17.2737</td>
<td>230.374</td>
<td>13.3367 hierarchical</td>
</tr>
<tr>
<td>1.2192</td>
<td>30.8113</td>
<td>0.55165</td>
<td>2.44781</td>
<td>28.3635</td>
<td>11.5873 hierarchical</td>
</tr>
<tr>
<td>1.1102</td>
<td>17.9289</td>
<td>0.59106</td>
<td>2.00528</td>
<td>15.9236</td>
<td>7.94088 hierarchical</td>
</tr>
<tr>
<td>0.0906</td>
<td>0.97872</td>
<td>1.02497</td>
<td>0.03243</td>
<td>0.94629</td>
<td>29.1829 single-level</td>
</tr>
<tr>
<td>0.1794</td>
<td>1.14019</td>
<td>0.98753</td>
<td>0.03344</td>
<td>1.10675</td>
<td>33.0985 single-level</td>
</tr>
<tr>
<td>0.2454</td>
<td>1.01073</td>
<td>1.05417</td>
<td>0.04320</td>
<td>0.96754</td>
<td>22.3975 single-level</td>
</tr>
<tr>
<td>0.0998</td>
<td>1.04555</td>
<td>0.99760</td>
<td>0.06947</td>
<td>0.97608</td>
<td>14.0498 single-level</td>
</tr>
<tr>
<td>0.1055</td>
<td>0.90488</td>
<td>1.15734</td>
<td>0.03292</td>
<td>0.87195</td>
<td>26.4834 single-level</td>
</tr>
<tr>
<td>0.1219</td>
<td>1.00240</td>
<td>0.97978</td>
<td>0.07775</td>
<td>0.92465</td>
<td>11.8930 single-level</td>
</tr>
</tbody>
</table>

The BIC for the model based clustering determined that the best fitting model had five cluster groups. Therefore, I conducted the next steps based on the five classes of datasets at this step.

Figure 6-8 shows the scatter plot for each pair of measures and the five cluster groups of datasets resulted by the clustering. The datasets are clustered in a six dimensional space, and the figure shows all possible 2-dimensional views of the cluster groups on the surfaces of pairs of measures. As the figure shows, looking at any pairs of measures makes only a few of the cluster groups more visible and separable, and no pair of measures can completely separate all of
the cluster groups. Four of the five cluster groups, consist entirely of hierarchical datasets, and in one class, almost 70% of the datasets were single-level datasets (Table 6-6). The red colour shows the datasets in cluster group 1, which contains all single-level datasets and a number of hierarchical datasets. All other 4 colours (blue, green, yellow, and orange) show the cluster groups only containing hierarchical datasets (as shown in Table 6-6).

<table>
<thead>
<tr>
<th>Class number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of single-level datasets</td>
<td>68.9%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>% of hierarchical datasets</td>
<td>31.1%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Based on the findings in this chapter, a model-structure selection tool is proposed in the next section to recommend the more appropriate model type.

### 6.3.5 The proposed model-structure selection tool

In section 6.2.56.3.4, the model-based classification was applied on the datasets over the study space, and five groups of datasets were determined. In
this section, the model-structure selection tool is developed to predict which model structure fits best for a specific dataset.

Given the number of classes and the abundance of single-level and hierarchical datasets in each class, the equations in section 6.2.5 can be rewritten more specifically as follows:

\[
P_k(s) = 68.9\%, \quad P_2(s) = P_3(s) = P_4(s) = P_5(s) = 0\%
\]
\[
P_k(h) = 31.1\%, \quad P_2(h) = P_3(h) = P_4(h) = P_5(h) = 100\%
\]

where

- \( P_k(s) \) is the probability of a dataset of group \( k \) being single-level,
- \( P_k(h) \) is the probability of a dataset of group \( k \) being hierarchical,
- \( k \) is the group number

Given a new dataset, \( X \), mclust predicts the probability of the dataset belonging to each of the 5 groups (\( P_{X1}, \ldots, P_{X5} \)). The proposed tool calculates the two probabilities of being hierarchical or single-level dataset for dataset \( X \):

\[
P_X(s) = \sum_{k=1}^{5} P_k(s) \times P_{X_k} \quad 6-28
\]
\[
P_X(h) = \sum_{k=1}^{5} P_k(h) \times P_{X_k} \quad 6-29
\]

where

- \( P_X(s) \) is the probability of dataset \( X \) being located in the single-level sub-region of the study space,
- \( P_X(h) \) is the probability of dataset \( X \) being located in the hierarchical sub-region of the study space.

Based on equations 6-16 and 6-17, and Table 6-3, the two final scores of the model structures fitting best for dataset \( X \):

\[
P(S = s \mid X) = 0.881 \times P_X(s) \quad 6-30
\]
\[
P(S = h \mid X) = 0.9934 \times P_X(h) \quad 6-31
\]

where

- \( S \) is the model structure that produces the observed data and is either hierarchical or single-level,
- \( P(S = s \mid X) \) is the probability of single-level model structure fitting best to dataset \( X \) estimated by the model-structure selection tool,
$P(S=h|X)$ is the probability of that hierarchical model structure fitting best to dataset $X$ estimated by the model-structure selection tool, $P_X(s)$ is as defined in equation 6-28, $P_X(h)$ is as defined in equation 6-15.

The findings of the \textit{mclust} analyses on the study space datasets was saved as two R objects (an \textit{mclust} and a dataframe). The further calculations of the tool were saved in a separate R file (named ‘model-structure-selection-tool’). The user needs to save any new dataset in a csv file in the same folder with the R file under the name of ‘newdataset.csv’, then, run the tool (the R file). The output of the proposed tool is another csv file (named ‘tool-recommendations’), which produces as output the recommended model structure for the dataset, along with a reliability of the recommendation (Figure 6-9).

Here, I summarise the findings and put them into a coherent whole as the instructions to use the proposed model-structure selection tool:

1. The prerequisite to have the hierarchical models as an option is to have data collected in a hierarchy. It should be possible to categorise the collected data into different groups, such as sites in an ecosystem or classes in a school (Gelman et al., 2014a). If the condition is satisfied,
Chapter 6  Establishing a tool for model-structure selection 160

the next steps need to be followed to get recommendations to develop either a single-level or a hierarchical model on the data.

2. By entering the dataset to the proposed tool (as a csv file), the tool derives the outcome in form of the six informative measures. This process is done by computing four summary statistics of the dataset:
   a. The mean value of each group of data
   b. The standard deviation value of each group of data
   c. The mean value of all data
   d. The standard deviation value of all data

3. The proposed tool produces the probability of the best fit of model structure based on WAIC as the goodness of fit for both single-level and hierarchical Bayesian Poisson regression framework. The results enable the decision maker to decide the model structure based on their objectives of modelling, as a trade-off between the model results reliability and the required modelling time and effort.

6.3.6 Applying the proposed tool to the case studies

In the beginning of this section, the results of applying the developed tool on the three case studies (6.2.6.1) is reported. Then, the modelling results of the three case studies are provided and compared with the proposed tool’s outcomes.

6.3.6.1 The proposed tool outcomes

For the three fish species datasets (M. duboulayi, T. tandanus and A. reinhardtii datasets), the instruction described in 6.3.5 was followed. To create a perception of the fish species datasets, the positions of the values regarding each informative measure (6.2.3) for the three datasets in the combined density plots of the single-level and hierarchical datasets is demonstrated in Figure 6-10.

The tool allocated two of the fish species (A. reinhardtii and M. duboulayi) datasets to the cluster group 5, and one fish species (T. tandanus) to cluster group 3 (Table 6-7). Both cluster groups 3 and 5 contain entirely hierarchical datasets (Table 6-6).
Figure 6-10 The values of each measure for the three fish species datasets in density plots of the six measures for the single-level (black line) and hierarchical (red line) datasets. Green point, blue point, and grey point represent Anguilla reinhardtii, Melanotaenia duboulayi, and Tandanus tandanus fish datasets, respectively.

Table 6-7 The probability of each fish species belonging to the 5 cluster groups produced by the tool

<table>
<thead>
<tr>
<th>Fish species</th>
<th>Probability of the data belonging to each cluster group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Anguilla reinhardtii</td>
<td>0%</td>
</tr>
<tr>
<td>Melanotaenia duboulayi</td>
<td>0%</td>
</tr>
<tr>
<td>Tandanus tandanus</td>
<td>0%</td>
</tr>
</tbody>
</table>
Therefore, according to equation 6-11:

\[
P_{(A.\text{rein})k} = 0 \text{ for } k=1,2,3,4, \text{ and } P_{(A.\text{rein})5} = 1
\]
\[
P_{(M.\text{dub})k} = 0 \text{ for } k=1,2,3,4, \text{ and } P_{(M.\text{dub})5} = 1
\]
\[
P_{(T.tandanus)k} = 0 \text{ for } k=1,2,4,5, \text{ and } P_{(T.tandanus)3} = 1
\]

By replacing the values in the equations 6-28 and 6-29 6-31, the reliability of the recommendation (the recommendation score) was calculated as:

\[
P_{(A.\text{rein})}(s) = P_5(s) \times P_{(A.\text{rein})5}, \quad P_{(A.\text{rein})}(h) = P_5(h) \times P_{(A.\text{rein})5}
\]
\[
P_{(M.\text{dub})}(s) = P_5(s) \times P_{(M.\text{dub})5}, \quad P_{(M.\text{dub})}(h) = P_5(h) \times P_{(M.\text{dub})5}
\]
\[
P_{(T.tandanus)}(s) = P_3(s) \times P_{(T.tandanus)5}, \quad P_{(T.tandanus)}(h) = P_3(h) \times P_{(T.tandanus)5}
\]

Knowing that \( P_5(h) \) and \( P_3(h) \) are both equal to 1, and \( P_5(s) \) and \( P_3(s) \) are both equal to 0 (Table 6-6):

\[
P_{(A.\text{rein})}(h) = P_{(M.\text{dub})}(h) = P_{(T.tandanus)}(h) = 1
\]
\[
P_{(A.\text{rein})}(s) = P_{(M.\text{dub})}(s) = P_{(T.tandanus)}(s) = 0
\]

Therefore, according to equations 6-30 and 6-31:

\[
P(S=s \mid A.\text{rein}) = P(S=s \mid M.\text{dub}) = P(S=s \mid T.tandanus) = 0
\]
\[
P(S=0 \mid A.\text{rein}) = P(S=0 \mid M.\text{dub}) = P(S=0 \mid T.tandanus) = 0.9934
\]

Therefore, according to the proposed tool outcome, all species were recommended to fit a hierarchical model with 99.34% reliability.

6.3.6.2 Modelling results

In order to validate the outcomes of the proposed tool, single-level and hierarchical models were developed on each fish species dataset and the results were compared.

I developed Bayesian Poisson regression models, as explained in 6.2.6.3; four sets of models for the A. reinhardtii dataset, and three sets of models were fitted for each of M. duboulayi and T. tandanus datasets. All the models fit (listed in Table 6-2) were hierarchical according to the WAIC analysis.
6.3.7 Demonstrating applicability of the proposed tool on generated datasets

The proposed model-structure selection tool was applied on the four sets of datasets explained in 6.2.7.

Across all known dataset types the probability of being classified to that type by the tool was quite high (Figure 6-11). Nonetheless, there were some datasets that were classified poorly, with only a low probability of being classified correctly, for example almost 200 datasets in sets 1 and 3 had a probability of being single-level equal to zero (or near zero). The tool correctly recommended single-level model structure for 85% of the datasets in set 1 and set 3, with an average reliability of the recommendation of 67%. Also, the tool recommended hierarchical model structure for 76% of the datasets in set 2 and set 4, with an average reliability of 82% (Table 6-8).

![Histogram of sets 1 & 3](image1.png)  ![Histogram of sets 2 & 4](image2.png)

Figure 6-11 Histograms of predicted probabilities of each set belonging to their pre-known best fitting model structure

<table>
<thead>
<tr>
<th>Number of datasets</th>
<th>Number of datasets with correct recommendation by the tool</th>
<th>Correct recommended model structure rate</th>
</tr>
</thead>
</table>
| **Set1**  
(single-level datasets in omitted sub-regions) | 4931 | 4500 | 91.26% |
| **Set2**  
(hierarchical datasets in omitted sub-regions) | 9069 | 7258 | 80.03% |
| **Set3**  
(single-level datasets in hierarchical sub-regions) | 3057 | 2293 | 75.01% |
| **Set4**  
(hierarchical datasets in single-level sub-region) | 3101 | 1981 | 63.88% |

Table 6-8 The number of datasets in each set defined in 6.2.7, and the number of datasets in each set for which the tool correctly recommended the model structure with better fit determined by the modelling results
6.4 Discussion

To my knowledge, no previous probabilistic framework was proposed in the literature to provide straightforward and robust guidelines to enable the users to select a credible model structure that can be argued to lead to reliable inferences and decisions. There are several model selection approaches existing in the literature that are widely used to assess fit of models in different fields of study (e.g., AIC, DIC, WAIC, LOO-CV). Each of these approaches represents slightly different procedures for model comparison, with many being developed to overcome some limitations of the others (Gelman et al., 2014a; Spiegelhalter et al., 2014). However, as these approaches are based on the likelihood of the model (Akaike, 1992; Spiegelhalter et al., 2002), the common feature of them is that the modeller needs to fit competing models to the data and compare them based on the chosen model selection approach. My proposed model-structure selection tool has the advantage of providing an assessment of probable comparative goodness of fit before fitting either model. This represents a considerable advance, as applied users can make an informed decision on when they can confidently apply a simpler model in easy to use, commercially available software without having to fit a complex model requiring some degree of programming, for which they may not be trained. Equally, users would know when the additional work to fit a hierarchical model would provide a better model fit, and when the single-level model structures are appropriate and likely to be sufficient. The proposed tool not only recommends the model structure, but also proposes a reliability score for its recommendation, upon which the user can make a decision based on the likely goodness of fit of the model.

The application of the three fish species datasets along with a big number of simulated datasets (which were not used in building up the tool) revealed that the proposed tool is competent in recommending the suitable model structure. The outcomes of the tool for the three case studies were perfectly consistent with the modelling results. The applicability of the tool was tested on a big number of datasets; these datasets were the simulated datasets removed in the data cleaning process (set 1 and set 2 in Table 6-8), and each could represent
a totally new case study. The correct recommended model structures for these simulated datasets locating in the omitted sub-regions had a considerably high rate. This high rate showed the perfect applicability of the tool to the new datasets that are not used in the development of the tool. This rate declined when the tool was applied to the outlier simulated datasets in the single-level/hierarchical sub-regions (set 3 and set 4 in Table 6-8). Nevertheless, the outcome of tool for these two sets was credible, knowing that these datasets were inconsistent with the results from the sub-regions upon which the tool has been developed.

The reliability of the recommendation was considerably lower for single-level model structure than that of the hierarchical model structure; this can be due to the fact that a big portion of simulated datasets in this research were recognised as hierarchical based on the WAIC analysis, and highest frequency of single-level datasets in the sub-regions (as defined by the WAIC regression tree in section 5.2.1.1.3) was ~88%, which was much lower than that of the hierarchical sub-regions. Knowing that the single-level models mostly fitted best where the values of attributes were low (in the single-level sub-region, as defined in 5.4), redefining the study space (in section 4.2.1) and allocating more low values to the attributes (based on the results of this research), may lead to produce a fairly balanced study space (which contains a bigger portion of single-level simulated datasets compared to the current study space). In such a new study space, the regression tree analysis may result in single-level sub-regions with higher frequency of single-level datasets in the sub-regions. This can increase the tool reliability of the recommendation for single-level model structure.

Besides, the proposed tool is limited to the single-level and hierarchical model structures used in this research. It is built up based on Bayesian Poisson regression, and has not been tested on other model structures. Besides, the regressions in model structures used in this research have limited number of predictors (two predictors in single-level model structure, and one predictor in each of the two levels of hierarchical model structure). This limits the applicability of the proposed tool to the cases with the same number of
predictors. This limitation was revealed by studying the fish species case studies; the modelling results on the same fish species datasets by Stewart-Koster (2011) showed that while hierarchical models fitted best for two fish species datasets, single-level model was as sufficient for the third fish species dataset. The regression models used by Stewart-Koster (2011) had all possible predictors (Table 6-1). It is notable that when the same datasets were modelled by only two predictors (one predictor in each level of hierarchical models), (Table 6-2), hierarchical models fitted best in all cases (Table 6-7). This result shows the number of predictors affects which model structure fits best. Therefore, in this research, the statistical conditions I determined for model structure selection is limited to the number of predictors I used, and the present tool’s outcomes are not reliable for model structure selection when more numbers of predictors are considered.

Also, the dispersion test was applied on the fish species case studies. All the three case studies had overdispersed Poisson distributions; therefore, alternative models such as Poisson mixture models might provide a better fit to data.

Another point of consideration is that many real world case studies would be required to validate the proposed tool in real applications; the few real world case studies used in this research are better treated as examples to demonstrate how to use the tool. The dispersion test is required to be considered for the real world case studies to ensure the outcomes of the tool can be reliably applied to the case studies.

In conclusion, the proposed tool is but a first step in establishing a framework to determine the best fitting model structure in a more general scenario. Similar research using the same procedure should be undertaken to expand the tool’s applicability to other model structures (i.e., types of models/ number of predictors).
6.5 Summary

In this chapter, the sub-regions of hierarchical and single-level models produced by WAIC results were used to construct the basis for making inferences. Based on summary statistics of datasets in each sub-region, six measures were selected; the measures summarised the single-level and hierarchical datasets.

The findings were used as the grounds to propose a model-structure selection tool to recommend the best fitting model structure for any specific set of data. The tool was applied on three fish species datasets to demonstrate how it can be applied, and the tool’s recommendation was compared to the modelling results. Also, the datasets of the study space, that were not used to develop the proposed tool, were used to test the proposed tool. The results showed the tool was able to give recommendations with almost the same predictive reliability as it claimed. In the end of the section, the advantages and limitations of the tool were discussed, and future research was suggested.
Chapter 7
Conclusions and recommendations
7.1 Introduction

This chapter presents the conclusions of this thesis and summarises the analyses, discussions and results presented, as indicated in Figure 7-1. Firstly, a brief thesis summary is provided in the next section. Then, the contribution to knowledge, limitations of this research and recommendations for further research are addressed in the following sections.

![Figure 7-1 Research outline (Chapter 7 highlighted in red)](image)

7.2 Summary of this research

Uncertainty is an inevitable component of ecological and environmental management and decision making, and the uncertainty about the model's
structure is the most difficult one to deal with. In particular (given the focus of this thesis), while hierarchical Bayesian models are very powerful in quantifying multiscale processes and developing complex probabilistic models to reflect underlying ecological processes, the additional effort to fit these models made the author ask when it is necessary to choose them over single-level models. This thesis established a framework to address the need to understand the conditions where fitting hierarchical models is necessary.

7.2.1 Literature review

In Chapter 1, the aim, objectives and the scope of the thesis were discussed. Chapter 2 and Chapter 3 carried out a comprehensive literature review on the fundamental mathematical and statistical models that are the bases of all types of models being used with the purpose of explaining, reproducing and predicting the behaviour of all kinds of systems; from the very elementary small-scale systems to complex large-scale systems. In Chapter 2, the mathematical models were categorised into four basic model types, and the strengths and limitations of each type of models, as well as their applications were briefly overviewed. In the end of the chapter, the characteristics of the system under review along with the capacities of the model were recognised as two important aspects to be considered for model selection, and a summary of the applicability of the four basic mathematical models based on the models capacities and modelling objectives was provided.

Chapter 3 reviewed the statistical models, in both Bayesian and frequentist context, and their application in the field of environmental management and decision making, more specifically on ecosystems. Reviewing the two types of single-level and multi-level models revealed the abundant applications of the two model structures, as well as the advantages and disadvantages of each model structure. While some literature compared the two model structures on individual real life cases (e.g., (Clark et al., 2005; Gelman, 2006a; Stewart-Koster, 2011; Gelman et al., 2014a; Gelman et al., 2014b), the pros and cons of using each of the two model structures highlighted a significant lack in literature
that would provide a guiding framework to recommend the more suitable model structure of the two,

The findings narrowed the goal of this research to develop a model-structure selection tool to identify the statistical conditions under which hierarchical models might be necessary, in order to inform users when the additional effort of fitting a hierarchical model provides a better model fit and when the simpler single-level model structures are appropriate.

7.2.2 Proposed approach

To achieve the goal of this research, a large simulation study was conducted and the detailed method to be employed was determined and described in Chapter 4.

A study space was defined, and several datasets were generated over each point of the study space. Both types of models, namely, single-level and hierarchical, were fitted on each dataset to examine under which conditions and characteristics of the data, each type of model fits best.

The models’ structures were set: both models were Poisson regression (due to the prevalence of this distribution in a large number of ecological processes) in a Bayesian framework. According to these models’ parameters, a six dimensional study space was defined, and seven sets of data were generated throughout the study space. To test the fit of each model, a number of measures were employed to evaluate the models’ predictive and descriptive capacities.

7.2.3 Analysing the results of simulation study

In chapter 5, the results of applying each measure to compare the fit of the two model structures on every dataset in the study space were analysed, and, consequently, some subspaces in the study space were identified as the regions in which single-level modelling was adequate, and in some other regions hierarchical modelling was recommended. Multiple linear regression analysis and regression trees were used to analyse the outputs of the simulation step.
Among different applied measures of fit, WAIC was revealed to be appropriate to be used for this research, and the WAIC average probability scores were selected to be the basis for further steps of this study. The results of WAIC were used to classify the study space into subspaces of either single-level or hierarchical models regions.

Three attributes ($\sigma_0$, $\sigma_1$, and $\gamma_{11}$) were identified to be crucial in classifying the single-level and hierarchical subspaces, regarding the WAIC results. Based on the values of these attributes, the whole study space was partitioned into the single-level region (the region in which single-level models fitted best in 88% of datasets), the hierarchical regions (six regions in which hierarchical models fitted best in more than 95% of datasets), and the regions of the transition between the two (both single-level and hierarchical models fitted best on a significant number of datasets in the region, and none of the two model types were mainly dominant).

### 7.2.4 Establishing the model-structure selection tool

The developed model-structure selection tool was provided in Chapter 6; the method section explained the detailed steps of the process taken to develop the tool, and the result section reported the output of each step.

In this chapter, only the hierarchical and single-level sub-regions of the study space produced by WAIC results were used to construct the basis for making inferences.

Based on summary statistics of datasets in each sub-region, six measures were selected to estimate the differentiating characteristics of the data in each subspace; the measures summarised the single-level and hierarchical datasets.

Using model-based clustering analysis, the single-level and hierarchical datasets were grouped into five classes based on their characteristics (the six measures’ outputs). The findings were used as the grounds to propose a model-structure selection tool to recommend the best fitting model structure for any specific set of data.
Any new dataset can be entered to the tool as a csv file. The tool applies the six measures on the dataset, and, based on the measures outputs, classifies the dataset into its pre-existing five groups. Then, the tool recommends one model structure for the dataset, in addition, it gives a reliability of its recommendation. Based on the reliability of the recommendation, and considering the objective of modelling, the modeller can decide either to go with the recommended model, or to select their favourite model structure, or to fit both model structures and compare the results.

7.2.5 Applicability of the proposed tool

To test the applicability of the proposed model-structure selection tool, the tool was applied on three fish species datasets. Also both model structures were fitted in each case to compare and validate the output of the proposed tool.

For all three fish species datasets, the tool recommended hierarchical modelling with a reliability of recommendation of 99.34%. The modelling results also confirmed the tool’s outputs and demonstrated the hierarchical modelling fitted best in all three datasets.

Also, the datasets of the study space, that were not used to develop the proposed tool, were used to test the proposed tool. The results showed the tool was able to give recommendations with almost the same predictive reliability as it claimed.

7.3 Main significance and contribution

To my knowledge, no previous study was conducted to propose a straightforward framework to advise the users about a credible model structure that can be argued to lead to reliable inferences and decisions. This thesis has taken the first step in establishing a reliable framework to guide the modeller to select the appropriate model type and to lessen the uncertainty about the model structure in environmental management and decision making.

The proposed model-structure selection tool can be efficiently applied to a dataset, to inform users when the additional effort of a hierarchical model
provides a better model fit and when the single-level model structures are appropriate and sufficient. The proposed tool not only recommends the model structure, but also proposes a percentage of reliability for its recommendation, upon which the user can make a decision as a trade-off between the aims of modelling and the effort they put into the modelling.

7.4 Research Limitations

The proposed model-structure selection tool is restricted by the single-level and hierarchical model structures used in this research, which were built up based on Bayesian Poisson regression. The applicability of the proposed tool is not tested on other model structures.

The limited number of regression predictors in model structures used in this research (two predictors in single-level model structure, and one predictor in each level of hierarchical model structure) limits the applicability of the proposed tool to the cases with the same number of predictors.

The proposed tool has not been validated with the real world cases. While the three studies are demonstrations of implementing the proposed tool, a large number of real world case studies would be required to validate the tool in real applications.

7.5 Future research

This thesis has provided a detailed framework for model structure selection between the two single-level and hierarchical model structures. However, there are areas that require further research. The main areas for further research are as follows:

1. The proposed model-structure selection tool is developed based on the results of the goodness-of-fit measure of WAIC. A further comparative analysis on the measures is recommended to get insights into the mathematical concepts of measures, their robustness and capabilities, and their suitability in different contexts. This knowledge will lead to determine the applicability of the tool proposed in this research, as well
as to conduct the development of similar tools based on further desired measures.

2. The single-level and hierarchical model structures of this research were defined in the context of Bayesian Poisson regression. Similar research would be possible and recommended for other modelling contexts.

3. The tool reliability of the recommendation for single-level model structure is considerably lower than that of hierarchical model structure. Using the results of this study as a preliminary result, to build up a new study space with more datasets in the single-level sub-region (revealed by this study), and then going through all analyses steps again on the new study space datasets, may lead to produce new single-level sub-regions with higher frequency of single-level datasets, and improve the tool reliability of the recommendation for single-level model structure.

4. The study revealed that the number of predictors was very influential and could alter the best fitting model structure, as discussed in the limitations of this research. It is recommended that similar process of running a large simulation and, consequently, developing a recommendation tool be undertaken for higher numbers of predictor variables. Having multiple predictor variables necessitates the test of collinearity as a part of statistical analysis.

5. This research used parametric models, but for future research the exploration of nonparametric models is also recommended. This is because parametric models assume some finite set of parameters to define the data distribution, as opposed to nonparametric models that assume an infinite-dimensional parameter space (e.g., a function in a regression problem, or a set of all densities in a density estimation problem) (Ghahramani, 2013). In a parametric model, the model parameters are estimated based on the available set of data with an assigned distribution, and predicting a future data value from the current state of the model is just based on the model parameters (e.g. the intercept and coefficients of a linear regression model). In contrast, the number of parameters in a nonparametric model grows with the amount of data, and the parameters are determined by the data, not the model.
So the nonparametric models do not assume a given probability distribution for the data, and are more flexible to embrace different data distributions (Bzdok and Yeo, 2017). Such models are more powerful for cases when there are not enough observations to determine the probability distribution of the data. Therefore, exploring such models would be very beneficial for research in ecology, psychology, neuroscience or any fields where ordinal or nominal data is employed and where the probability distribution of population cannot be surely estimated from the available observations (Orbanz and Teh, 2010; Bzdok and Yeo, 2017).

6. Ecological and environmental studies often include data with spatial and/or temporal dependence structures. Ecological data are collected across time as well as space, and the presence of a datum in a certain space or time is not probably a random event; by contrast, it is very likely that the datum at one point is affected by other data at different points of time and space (Carroll and Pearson, 2000; Diggle, 2002). However, the models considered in this work did not account for these type of dependence structures; this study used randomly generated, independent data, and the models were developed with the assumption that the model errors were independent. Therefore, it is recommended that in the future a similar simulation study be undertaken to consider these types of dependencies and investigate the more complex models.

7. A separate comprehensive study should be conducted to validate the proposed tool with many sets of data of real world cases.
References


References


Communications in Statistics—Simulation and Computation®, 36(6), 1309-1320.


James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An Introduction to Statistical Learning*, Springer.


Kröse, B. & Van Der Smagt, P. (1996) *An Introduction to Neural Networks*, University of Amsterdam.


References


Appendix I
<table>
<thead>
<tr>
<th>1</th>
<th>Publication Type</th>
<th>Referenced journal article</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Publication Title</td>
<td>The Application of Decision Making Models in Sustainable Management of Environmental Systems</td>
</tr>
<tr>
<td>4</td>
<td>Acknowledgements</td>
<td>The original idea for this paper came from my PhD project discussions with my principal supervisor A/Prof Peter Bernus. The idea of comparing the models and their applicability in the field of environmental management and control as the focus of my PhD project was co-developed by A/Prof Peter Bernus and me. This paper is part of my critical literature review of the most common types of mathematical models employed in the field of environmental management. I evaluated the advantages and disadvantages of models applications at all levels of management based on their mathematical properties, and wrote the manuscript under supervision from A/Prof. Peter Bernus during my candidature. Dr. Jarrod Trevathan edited and amended the paper for submission.</td>
</tr>
<tr>
<td>5</td>
<td>Comments</td>
<td>No comments</td>
</tr>
</tbody>
</table>
| 6 | Co-Authors' Signatures | Ameneh Shobeirinejad Date: 7/7/2017  
Peter Bernus Date: 8/7/2017  
Jarrod Trevathan Date: 12/7/2017 |
The Application of Decision Making Models in Sustainable Management of Environmental Systems

Ameneh Shobeirinejad *
Peter Bernus †
Jarrod Trevathan §

School of Information and Communication Technology
Griffith University
170 Kessels Road, Nathan (Brisbane)
Queensland, 4111
Australia

Abstract
Recent research has highlighted environmental sustainability as a key issue of concern affecting the future of human wellbeing. To pursue a sustainable future of a system, it is a critical challenge to make decisions at all levels of management. Decision making for controlling or steering a system requires models of the system that represent the determining characteristics of the system and have sufficient predictive power. Practitioners/researchers of environmental management use a large variety of (types of) models to support decision making, and it has become a complex task to select the appropriate model-type for a desired management goal. This paper considers the most common types of models and uses the mathematical properties of these model types to discuss the pros and cons of their use in environmental management / decision making. We explore how these mathematical properties influence the feasibility, viability and validity of the use of such models.

Keywords: Environmental modelling; environmental management; decision making; mathematical properties

1. Introduction
Sustainability is a key issue of concern, affecting the future wellbeing of humankind, both on the micro- and macro levels. It has been

*E-mail: a.shobeirinejad@griffith.edu.au (Corresponding author)
†E-mail: P.Bernus@griffith.edu.au
§E-mail: j.trevathan@griffith.edu.au
acknowledged [29] that the ever-increasing connectivity and dynamic interactions among complex large-scale systems, including ecosystems, economic systems and social systems, pose a significant risk to a sustainable future.

To pursue a stable, constant path towards a sustainable future of a system, making decisions at all levels of management, i.e., operational, tactical, and strategic levels, is a critical challenge. At the operational level of management, short-range (e.g., weekly, monthly) plans are developed. This can include short-term optimization of the performance of system functions and the direction of the use of resources. At the tactical level of management, short- and medium-range plans and schedules are developed, with the view of achieving desirable system characteristics and at the same time optimizing the use of resources. Typically, both of these management levels assume an essentially stable system structure. As opposed to this, on the strategic level of management, overall system-level goals, strategies, policies, and objectives as part of a strategy making process are developed, with the view of keeping essential system characteristics. At this level, transformation and intervention into the system structure may be needed. Also the strategic performance of the system and its overall direction in the environment are monitored at this level [25]. To exemplify, we refer to some examples of decision making at different management levels in a few cases taken from the Australian context:

- Managing monthly groundwater extraction per bore in the Daly River catchment (Northern Territory, Australia) as an operational level management objective;
- Water Allocation Plans (WAP) decision procedure to set annual water extraction limits from the Tindall Aquifer at Katherine [26], or the annual closure of lake Currimundi to control the biting midge winter larval densities [28] as tactical level management objectives;
- Providing future climate scenarios to enable determination of long-term average sustainable diversion limits to maintain the availability of the Basin water resources in the presence of the risks imposed by climate change [7] as a strategic level management objective.

Decision making for controlling or steering a system requires appropriately constructed models of the system. Such models have to be able to represent and predict the vital and determining characteristics of
APPLICATION OF DECISION MAKING MODELS

the system and represent the system’s elements, the interconnectivities and interdependencies of the elements, along with their interactions with elements inside or outside of the system.

The need to use models to ‘control’ systems, and to make management decisions, is a well-established paradigm, which originated in the engineering field (control engineering) and was later introduced to environmental management. These models are expected to enhance the understanding of environmental systems at various levels of detail and support decision making for management and control [4].

In the field of ecology, intense research has been undertaken during the last three decades in order to develop practically applicable models to represent complex macro- and micro ecosystems, and these ecological models also have been applied in environmental management [16] at all levels. There exists an abundance of various types of models with different names, intended to solve a wide spectrum of problems in this field, and categorised by application area. By today, the selection of the appropriate model-type for a desired management goal has become a complex task itself. At least part of this complexity problem arises from the fact that almost all the modelling efforts in the field of environmental management have been done with specific ecological points of view, by experts who have faced very specific problems of particular ecosystems. However, as mathematics is the common base of all of these modelling efforts, we believe that it is a useful effort to discuss and analyse the basic model categories using their mathematical properties, to pinpoint the potentials and pitfalls effecting environmental management applications. The paper aims at contributing to the understanding of how these mathematical properties influence the feasibility and viability of the use of such models for environmental decision making and control.

This paper is organised as follows: Section 2 provides an introduction to four frequently used types of mathematical models and discusses their properties. Section 3 discusses the models’ capacities to be used in different management situations and categorises the suitable model-type for the job, according to the strengths and weaknesses of each model type. The last section draws conclusions and discusses future work.

2. Mathematical Modelling

In describing a system and its dynamics, mathematics is the most qualified language. Indeed, the idea that the ‘book of nature’ is written in the language of mathematics was established a few centuries ago [3]
and mathematical modelling is a discipline which plays a basic role in describing the dynamics of complex ecological, economic and social systems.

This paper categorizes the different types of applied models into four basic types, namely: Input-Output Models, Agent-Based Models, Artificial Neural Networks and Bayesian Models. (Note that no separate overview of fuzzy logic models was attempted due to the fact that this can be considered as a modality of multiple model types.)

2.1 Input Output Models (IO Models)

IO models are a category of models describing the input/output behaviour of systems. The IO concept is widely referred to the model's ability to represent the evolving states of the system steadily over time by describing the flow of some variables with the help of mathematical equations. Such flows are governed by physical laws, such as the flow of energy governed by the conservation of energy.

The mathematical apparatus used to construct IO models, commonly used in control theory, are differential equations and finite difference equations, both of which have been developed to describe dynamic processes. By setting these equations to zero, the model represents the static situation, which would be of benefit for the time the data quality and quantity are not sufficient to develop a complete dynamic model, or when the modelling objective is to investigate a worst-case situation [16]. We categorize all models based on differential or finite difference equations as IO models.

Generally, the model state is measured by certain magnitudes (state variables) that are dependent on time, and the model describes the system's evolution over time using differential equations, when the time is considered a continuity, or finite difference equations, when the time varies in discrete intervals [3].

The following set of differential equations describes a dynamic system with a set of \( n \) state variables in continuous time:

\[
\frac{d}{dt} x_i = F_i(x_1,...,x_n,t) \quad t = 1,...,n
\]

The dynamic system in this case is called a flow. On the other hand, the system can be described by the following set of finite difference equations, which show the evolution of the system over discrete time intervals:
\[ x_i(k+1) = f_i(x_1(k), \ldots, x_n(k)) \quad i = 1, \ldots, n \]

In this second case, the model of the system is called a map [3]. Some examples of the most common IO models are briefly reviewed below.

2.1.1 Oscillating Pendulum

An oscillator is one of the simplest and most important models which has been a basis for further extended and developed models in other sectors such as some macro scale ecological or economic systems that experience periodicity and oscillation over time. The application of differential equations in clarifying the behaviour of such systems in both linear and nonlinear aspects is notable. Therefore, the simple basic pendulum model is discussed here. Fig. 1 shows the pendulum at its start point where no friction is considered.

The differential equation describing the behaviour of this system is \( \frac{d}{dt^2} \theta = -\sin \theta \cdot g/L \) (where \( g \) is the gravitational acceleration).

The above system of differential equations is nonlinear as it includes the term, \( \sin \theta \), however, given that \( \theta \approx \sin \theta \) for small angles, in a subspace of the pendulum’s state space the system can be modelled using a system of linear differential equations.

This is a typical case, illustrating that after approximations and assumptions, it is either possible to explicitly solve the equations, or if this is not possible, then numerical techniques must be applied. It is, however, an important question, that every modeller must be able to answer: is the

![Pendulum Diagram](image)

**Fig. 1**

Pendulum at start point with no friction
so derived (approximate) solution adequately accurate and in what subspace of the systems state space is this the case?

2.1.2 Predator-Prey

The classic theme of the predator-prey model is a model of two interacting populations in an ecosystem where one feeds on the other. This type of system has been studied over decades and widely applied in a variety of areas. This is an example of a case in which the evolution of the system is described at discrete time intervals rather than continuously in time. Such systems are called discrete time systems and are described by difference equations.

The general fundamental principles of the basic model (called the Volterra-Lotka model) are still used as a reference for modelling in mathematical biology as well as other fields such as urban and regional science [3]. The original form of the model assumes unlimited food supply for the prey and therefore no internal competition for that species, while the predator species in this model compete over the finite population of the prey species.

The two populations interact in a nonlinear way as the populations of the two species include the nonlinear term of encounter frequency between a prey and a predator, which is proportional to the product of the prey population multiplied by the population of predators (which in itself is an approximation assumed to be admissible without influencing the model's predictions in a substantial way).

2.1.3 Hydodynamic Systems

"Hydrodynamics" is the study of motion of liquids, in particular water, and heat in oceans and coastal seas in order to determine the nature of marine ecosystems. The models of this type are also widely used to predict the transport of sediment. Hydrodynamic modelling is also used in meteorology, aerospace and automotive design, ventilation systems, and so on.

The common basis of these models is the numerical solution of differential equations governing the transport of mass (the law of conservation of mass), transport of momentum (Newton's second law of motion) and transport of energy (the first law of thermodynamics) in moving fluids. The Navier-Stokes (NS) equations are the fundamental equations of fluid motion [8] and are the basis of almost all hydrodynamic models. In all but the simplest cases an explicit solution of the NS differential
APPLICATION OF DECISION MAKING MODELS

Equations is not possible; therefore, it is necessary to use finite difference equations, where the numerical solutions are acquired at discrete points rather than all points as in the continuous space. As a consequence, Computational Fluid Dynamics (CFD) problems are normally solved using finite elements methods [11], whereupon the space is subdivided into very small areas, and the behaviour of the fluid (flow) is described by a separate equation for each area. This approach emerged and has been widely used with the availability of computers since the early 60s. Today the method is extensively used in basic and applied research, in design of engineering tools, and in environmental models [8].

In hydrodynamic models, it is a basic assumption that the model can be discussed at any desired precision with the help of finite difference equations, and also that the finite element mesh created to represent the model is extremely accurate. While such assumptions would be true in small-scale engineering applications (e.g., sewerage system pipelines) large-scale environmental cases may be totally different (e.g., a river catchment). This makes a significant difference for the model's ability to make predictions. For example, an engineering hydrodynamic system is designed using mathematical equations of the surfaces enclosing the volume of liquid, therefore, when finite element approximation is necessary, this approximation can be as fine as needed and constitutes a completely predictable 3D surface. As opposed to this, the accuracy and predictability of an environmental hydrodynamic model is more limited due to factors such as uncertain/fewer measurements and lower resolution and possibly due to natural changes in the shape and volume of the enclosing 3D surface. This state of affairs is not something that can be easily changed; therefore, additional methods are necessary to validate such models (such as validation against observable and measurable properties of the natural system).

2.2 Agent-Based Models (ABMs)

Individual-based Modelling (IBM) or Agent-based modelling (ABM) is a relatively new approach to modelling the dynamics of complex systems composed of interacting, autonomous agents. Agents have behaviours and interactions with other agents and with their environment, which in turn influence their behaviours. By modelling individuals as agents, the effects of the diversity that exists among individuals in their behaviours and attributes could be represented and observed. In such models, patterns, structures, and behaviours, that are not explicitly programmed into the
models but arise through agent interactions, would emerge [22]. Fig. 2 shows the structure of a typical agent-based model.

ABMs can be considered as the extension of cellular automata (CA). CA are typically composed of cells on a two-dimensional grid. Each cell can be interpreted as an agent that interacts with a fixed set of neighbouring cells. The cell state is either 'on' or 'off' at any point in time. Cells change their states every finite time step, following simple rules of behaviour. The next state of the cell depends on the cell's current state and the actual states of its neighbours. A CA is deterministic so that the same state for a cell and its neighbours always results in the same future state [21].

Agents are entities, which can be considered having a set of goals and capable of autonomously performing actions in a dynamic and unpredictable environment in order to meet their goals. They encompass heterogeneous behavioural instructions and are adaptive in terms of learning and modifying their behaviour based on their accumulated experience [2]. Fig. 3 shows an agent and its properties.

An ABM consists of a set of agents, a set of agent relationships, and a topology for simulating agent behaviours and interactions. This topology defines the neighbours and the mechanisms of dynamic interactions. It usually includes a spatial grid or network of nodes (agents) and links (relationships). In complex systems, agents do not interact with all agents
all the time, and local information is obtained from interactions with an agent’s neighbours and from the agent’s local environment. The set of local information resources of an agent can change rapidly over the agent life [22].

ABMs allow the explicit consideration of the structural heterogeneity of the components along with their spatiotemporal variation. They can represent individual behaviours and can model the population’s response as an emergent property. Therefore, these models are capable of realistically representing the overall dynamics of natural systems, and help a modeller to understand the role and importance of the individual in shaping the overall system [22].

2.3 Artificial Neural Networks

The first discussions on Artificial Neural Networks (ANN) began by a simulation of how neurons might work in the brain with a simple model of neural network implemented using electrical circuits. An ANN is an interconnected group of artificial neurons (processing units), which communicate by sending signals to each other over a large number of links, where each link has a numeric weight. Each node has a set of input links from other nodes, a set of output links to other nodes, a current activation level, and an activation function to compute the activation level in the next time step.
The system is inherently parallel meaning that many units can perform computations at the same time. There are three types of units in a neural network; the input units, which receive data from outside of the system, the output units, which send data out of the system, and the hidden units, whose input and output signals remain within the system. Fig. 4 shows a schematic of a neural network and a neuron structure.

The current activation level for every unit is equivalent to the output of the unit, a propagation rule determines the effective input of the unit from its external inputs, and an activation function \( f \) determines the new level of activation based on the effective input and the current activation.

ANNs have the ability to learn and to generalise; they learn how to perform their function on their own and determine their function based
APPLICATION OF DECISION MAKING MODELS

only upon sample inputs, and can produce reasonable outputs for new inputs.

Generally ANNs use some sort of threshold function as their activation functions, such as a sign function, a hard limiting threshold function, a linear or a sigmoid function (Fig. 5) [19].

ANNs are classified into two categories based on the pattern of connections between the units and the propagated data [19]:

- Feed-forward networks; data flow from inputs to outputs is forward and there is no feedback connection in the network,
- Recurrent networks; networks contain feedback connections.

Training is the act of presenting the system with inputs from some sample dataset and modifying the weights in order to achieve a better approximation of the desired output function. The learning process can be categorized in two distinct types:

- Supervised learning or Associative learning
- Unsupervised learning or Self-organisation

In supervised learning, the neural network is supplied with inputs and their desired matching outputs, and the weights are modified to reduce the difference between the actual and desired outputs.

On the other hand, in unsupervised training, only the inputs are supplied and the neural network adjusts its own weights in a way that similar inputs cause similar outputs. In this kind of network, the system is supposed to ‘discover’ statistically important features of the input population, and identifies the patterns and differences in the inputs without any external assistance.

2.4 Bayesian Models

Reasoning under uncertainty is crucial for accurate analysis, synthesis, prediction, inference, and decision making. Bayesian Networks (BN) are models that support such reasoning. BNs represent causal assertions between variables as patterns of probabilistic dependence, and therefore, can be used for logical and holistic reasoning about complex systems [4]. A BN can be used to represent knowledge and observational inference.

A BN is a compact representation of a joint probability distribution over a domain of variables of interest [17].
In reasoning under uncertainty about a set of events, it is possible to observe that some events are causally related to other events with certain probability. Conditional independence (the lack of direct dependency relationship) plays a very important role in a Bayesian network's structure, as probabilistic inference can be made on the basis of information about the direct antecedent variables.

A BN is a Directed Acyclic Graph (i.e., no directed cycles) in which:

- A set of nodes represent variables;
- A set of directed edges or links represent probabilistic dependence between variables (dependence may mean mere correlation, or may be the representation of some type of causality based on physical law);
- Each node has a conditional probability table that quantifies the effects the parents have on the node. Note that these conditional probabilities in environmental applications typically represent statistical findings about observed / measured data, but may also be based on expert opinion, or could even be the result of analysing a separate fine grained model (of appropriate type) of a particular aspect of the system of interest.

Nodes can represent discrete or continuous variables. Nodes without any parents (immediate predecessors) are called roots, and are described as unconditional (marginal) value distributions. Nodes without any children (immediate descendants) are called leaves and the non-leaf, non-root nodes are called intermediate nodes.

In the absence of any explicit connecting edge between two nodes in a BN, the two variables are assumed independent, given the values of any intermediate nodes. In other words, the probability distribution of any variable in any state of the network can be determined by knowing only the values of its (immediate) parents, with no need to know the values of any other variables [4]. This is referred to as the Markov property.

Fig. 6 illustrates the BN of a medical diagnosis example. In this case, the factors affecting a patient's chance of having cancer are 'pollution' and 'smoking'. In the same way, cancer can be a cause of 'dyspnoea' and a 'positive X-ray result'. As also shown in this figure, discrete nodes can be Boolean (e.g., cancer node with true or false value), ordered values (e.g., pollution node with low, medium, and high values), or it can also have integral value (e.g., age node with possible values from 1 to 120).
2.4.1 Reasoning with BN

There are four main types of inferences with Bayesian networks:

- Predictive or causal inferences (from cause to effects),
- Diagnostic or evidential inferences (from effects to cause),
- Inter-causal inferences (between causes of a common effect), and
- Mixed or combined inferences.

Fig. 7 illustrates the four types of reasoning.

When the structure and conditional probabilities of a BN have been specified, the network can be used to determine the probability distributions of specific target or query nodes, given findings (either deterministic or probabilistic observations) for other nodes (evidence nodes). When the query nodes are descendants of the evidence nodes, the process is called prediction. When they are ancestors of evidence nodes, it is called diagnosis.

BNs can also be used to determine the most probable explanation of the particular values for some variables, to explain away the alternative causes of an effect, to describe the effects of interventions (or external
controls) on the system, and to support decisions about management actions in the face of uncertainty [4].

2.4.2 Extensions to BN: Bayesian Decision Network

BNs can be extended to support decision making. When it comes to decision making under uncertainty, the goal is to find out what action to take (plan to adopt) when the future state of the world is not known. A Bayesian solution would be to find the utility of each possible outcome (action-state pair) and take the action that maximizes expected utility (determined by the decision maker’s preferences regarding possible outcomes of various plans). This way, decision networks are produced as an extension of BNs with decision nodes and utility nodes to support decision making.

Applications, Strengths/Limitations, Comparison, and Applicability of Different Model types in Decision Making

The previous section described four fundamental, and generic, mathematical model types that are the basis of most models being used today in environmental decision making. In this section, these model
types are evaluated regarding several essential aspects of modelling; their functionalities and deficiencies are compared, and their determining attributes in the process of model selection for a particular case with specific goals are discussed.

3. Applications of Models

This section briefly describes the applications for each of the four types of models.

Applications of IO Models IO models are widely used in ecosystems and environmental management applications such as modelling water quality, water pollution, fisheries management, management of natural resources and natural parks [10], and for modelling the effects of chemicals and toxic substances.

Applications of Agent-Based Models The early use of ABMs in ecology goes back to a forest model [5] and a fish cohort model [9]. Other examples of applications of ABMs include: predicting the spread of epidemics [1], and traffic modelling [13].

Applications of Artificial Neural Networks ANNs have been successfully applied to a broad spectrum of data-intensive applications. They have been employed to create models to predict rainfall-runoff [15], stream flow [31], or be used for groundwater management [27], water quality simulation [23], rainfall forecasting [14], weather forecasting [24]. ANNs are also useful for pattern recognition and classification, such as fish species identification [6].

Applications of Bayesian Models The application of BNs for forward prediction is an obvious application of BN models. However, another significant application of BNs is in performing probabilistic inference, or diagnosis. Examples of application areas include weather forecasting (e.g., fog, hailstorm), medicine (e.g., death, injury, and disease), law, spam filtering, speech recognition, and robotics.

4. Strengths/Limitations of Various Models

IO Models These models are based on causality. A fundamental concept of IO models is that they are mostly based on conservation principles (physical laws / laws of nature). They are easy to understand and develop. In the case of linear systems, the model is able to predict the accurate future behaviour of the system, and the situation of the system at any point in the space of states can be perfectly discussed.
However, having an explicit and direct linear relationship between cause and effect is rare in the natural world, if not impossible; in practice, linear models are the result of linearization of a problem, in which the approximations are supposed to be slight and the ignored factors and parameters to only have trivial effects. As a result, linear models are only realistic, and the mathematical approximations satisfactory, if the system is close to a stable equilibrium [3].

Therefore, even though the linearization process might operate acceptably for the status quo, the slight ignored factors may result in very large unpredictable effects in the longer future. Besides, the accuracy of the input data is very critical to obtain reliable outputs from mathematical equations. This means that IO models require an accurate set of state variable (data) measurements and any small ignored measurement errors have the potential to yield consequential inaccuracy in long term future predictions. Both of the aforementioned types of difficulties of IO models may result in either quantitative or qualitative errors in the long-term predictions of the system’s behaviour.

There is a third type of difficulty regarding a system’s behaviour prediction which is related to the nonlinearity of many systems. Nonlinear models may at times behave in a way that gives rise to unstable evolutive trajectories which are completely unpredictable [2]. This happens when a dynamical system is highly sensitive to initial conditions; a small difference in initial conditions produces widely diverging outcomes of the system. Even though the system is deterministic and no random element is involved, the future behaviour of the system can still be unpredictable. This behaviour is known as chaotic behaviour. In terms of the system’s space of states, nonlinear equations can only be used safely to manoeuvre in a limited subspace of the large space of possible states, and therefore, the model is not necessarily valid for a possible future situation of the system when the system stands at a point far from the subspace that is known to be accurately described by the model. For the case of chaos (a chaotic subspace of the system’s space of states), a small change in the present state of the system may result in an unpredictable future state of the system, which means that even if the nonlinear model is truly representing the present state of the system and its vicinity, the only sincere prediction it can make for the system’s future behaviour would be its unpredictability. In practice, the three aforementioned issues, alone or together, limit the predictive ability of these types of models.

Of particular interest can be the use of control systems theory[20], to map the ‘troublesome’ areas of the state space, and allows the consideration
of alterations to the system to change the location of chaotic subspaces if the system’s trajectory is desired (or predicted) to trend toward such areas.

It is important to note that while a system might be capable of producing chaotic behaviour in certain sub-space(s) of its state space, the system may still be predictable from many respects, and/or the system may still be controllable [20], provided that the control of the system steers the system’s trajectory so as to avoid such chaotic areas of the state space.

A further weakness of this type of model is the (practical or theoretical) difficulty of describing the complete complex system in a single model; instead, it is often necessary to view the complex system as a ‘system of (sub)systems’, whereupon it is feasible to explain each subsystem using a type of IO model. One reason for this can be the lack of theory (knowledge of natural laws) connecting multiple aspects of a system. Therefore, a large complex system might have to be (and in environmental modelling often is) described by a number of different models, each explaining a part of the system, plus a further ‘super-model’ required to explain the interconnections among these subsystem models. Note that explaining all sub-models and their relations might not be feasible using differential/difference equations and/or this type of model would not perform appropriately as the super-model. Besides, such a super-model, if possible, would carry all the limitations of all sub-models together. In spite of these difficulties, IO models are frequently used in conjunction with other types of models in describing a complex system of systems.

Agent Based Models ABMs are used when differences among individuals of the same species are of importance for understanding system behaviour. Therefore, ABMs represent properties of individuals, such as adaptation and spatial distribution. ABMs also represent the behaviour of a population of individuals. It is possible to use these models to explain and predict emergent behaviours in simple cases, whereupon the emergent behaviour is the result of the behaviour of individuals.

However, the larger the number of properties considered in modelling, the more complex the ABM would be. Correspondingly, a significant amount of data is required for model calibration and validation [16]. In dealing with complex systems, these models are significantly more demanding to develop than IO models regarding data requirements, technical skills, computer power, and development time. Furthermore, in very large populations, the predictive ability of this type of model is very vulnerable as the success of ABMs depends on the precision of representing the agents and their interactions, so it is possible that just small differences in the level of individual detail can cause quite different
emergent behaviour, and the validation question becomes how do we know whether this is or is not the case?

Artificial Neural Networks ANNs can be applied to a heterogeneous database or observed / measured data to learn the system’s behaviour. As an ANN simulates the behaviour of the system as a black-box, the possibly very complicated effects and processes within the system do not need to be clearly investigated, understood or evaluated. ANNs can be easily applied when explaining the relations between the data with IO models (e.g. some differential equations) seems impossible. This gives an advantage to ANNs as they take into account all available data with no need to make approximations. ANNs can be easily applied to a system without ignoring any supposedly trivial information. Therefore, this type of model can enhance the accuracy and reliability of the model’s output.

We conclude that the above property improves the ability of making predictions using these models. Assuming we have a sufficient amount of data (measurements / observations) about the subspace of states in which the system is currently manoeuvring, an ANN is able to predict well the possible further situations of the system. This would normally be also true when the system is heading at new points of the space of states not far away from the known subspace, however, there is no guarantee that this is the case. This is because a system’s behaviour may have been known in an observed subspace of state variables, however, some points in that subspace may be close to a hitherto unobserved tipping point, which fact could therefore remain unpredictable by an ANN. This indicates the predictive ability of this model type being limited to short-term and medium-term future.

Note the fact that ANNs are black-box models and since the learning process is not about discovering causality-based facts, which imposes some restrictions to the applicability of this type of models. In addition in some steps of ANN model building, such as the ANN type selection, data type and parameter selection, decision about the number of hidden layers and neurons, and the selection of types of activation functions, expert opinion plays an important role – although using different algorithms to choose the most influential data can be helpful in some cases there is always some doubt about the repeatability (by others) of the modelling exercise.

Bayesian Models BN Inference is very flexible, as it can enter evidence about any node and update beliefs in any other nodes, which means that upgrading the model with new data is easy. Furthermore, the model’s probabilistic nature makes it appropriate to be utilised in a wide variety
of cases, where uncertainty dominates considerations and it is crucial to ‘get it right’.

However, we realise that a BN model is very sensitive in terms of the changes to the framework: new findings and data might impose structural changes, destroy the network’s hierarchical relations and may require the development of a completely new BN with a different structure.

5. Comparison of the Different Models

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| IO Model | • Often based on causality  
• Easy to understand and develop  
• Can be used to calculate effect of change on controllability | • Applies approximations and estimations  
• Often sensitive to measurements  
• Often valid only in a small subspace of the system’s state-space |
| ABM | • Can account for individuality  
• Can cover spatial distribution | • High complexity when having a number of properties  
• Can be sensitive to level of detail  
• Large amount of data to calibrate and validate |
| ANN | • Effective when the components of system and/or their interrelations cannot be clearly identified  
• Easy to apply  
• No approximation or data exclusion required  
• Applicable to heterogeneous datasets | • Is a black box and has no causality by nature  
• The expert opinion in:  
  • network type selection  
  • parameter selection  
  • number of hidden layers and neurons  
  • activation functions  
  • Needs historic data |
| BN | • Applicable to heterogeneous datasets | • Sensitive to structure |
Table 2
Model applications

<table>
<thead>
<tr>
<th>Model</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>IO Model</td>
<td>Water quality, water pollution, fisheries, natural resource/parks, chemicals, toxic substances, climate models</td>
</tr>
<tr>
<td>ABM</td>
<td>Fish cohort model, predicting the spread of epidemics and the threat of bio-warfare, modelling the adaptive immune system, understanding consumer purchasing behaviour</td>
</tr>
<tr>
<td>ANN</td>
<td>Rainfall-runoff, stream flow, groundwater management, water quality simulation, rainfall forecasting, weather forecasting, fish species identifications, short-term industrial management</td>
</tr>
<tr>
<td>BN</td>
<td>Weather forecasting, medicine, robotics</td>
</tr>
</tbody>
</table>

Table 1 lists the strengths and limitations of the four model types. Table 2 shows the most common cases in terms of the application of each type of models. As it shows, there are many common applications among different model types. This is due to the fact that environmental management plans almost always cover local case studies and a similar problem may be treated very differently in different situations. The reason is that in large complex systems, there are many interactions and interdependencies among the many elements of the system, and a similar issue in different situations can result from and be affected by completely different factors. Therefore in selecting a modelling technique for managing every situation, there are many influential factors to be considered rather than just the problem; such as the type of data, the future perspective of the system, the desired output of the system, and so on.

6. Applicability of Models for Decision Makings

The modeller faces three problems when constructing a model of the system:

1. What is the question that the decision maker(s) need to answer – i.e. what are the so-called ‘stakeholder concerns’?
2. What kind of model should be built?
3. How to build it?
APPLICATION OF DECISION MAKING MODELS

To deal with the first problem, stakeholder concerns are initially expressed in an informal way (i.e., in the way stakeholders formulate them). However, since these concerns and associated questions are usually not formulated in a precise and unambiguous way, environmental modellers need to clarify the terms, and disambiguate the questions.

For determining the type of model(s) that should be built, there are several essential factors: (i) is there a known theory (in terms of natural laws) which, in theory, could be used to build a model that produces answers for decision making? Depending on the answer to this question some model types can be eliminated; in practice another limiting factor can be the availability of expertise (ii) what is necessary (in terms of data) to build such a model and are there data available, or is it feasible to obtain such data. Feasibility can be a limiting factor (due to cost, accessibility, as well as space and time constraints).

It is therefore clear that to build models for decision making both the properties of the specific system under study and the objectives of modelling are to be considered. The system properties are recognised by the characteristics of the available data of different sources: measurements, simulations, experts’ opinions. The type of dataset (homogeneous/heterogeneous), data reliability (accuracy of measurements, certainty of experts’ opinions, and accuracy of simulations), data availability (sparse/large amount), and data accessibility (access to historical data/high frequency sampling) construct the pack of system properties to be identified in this step.

The other part to be clearly recognised is the set of modelling objectives. Different types of models support decision making at different levels of control and management. As Table 3 indicates, the objective of modelling can be limited to the operational level of management, which involves real-time control of the system. In this case, a ‘current model’ of the system is necessary. In this regard, the IO and ABM models can perform well. The aim of this level of control and management is to identify the system’s critical elements and their interrelations in order to keep the system’s performance indicators within acceptable limits.

The next level of management, the tactical level, looks into the short-term future, and with the help of historical data and consequent behaviours of system, forecasts the future behaviour of the system. However, tactical level decision making and control is (by definition) limited to parametric manipulation of the system, whereupon strategic level decision making needs to be open to structural change (or at least needs to be able to
Table 3
Applicability for decision making

<table>
<thead>
<tr>
<th>Model</th>
<th>Applicability</th>
<th>Prediction capability</th>
<th>Management level</th>
</tr>
</thead>
</table>
| IO Model | (i) Real-time control;  
(ii) Systems Design (control system design) | (i) Present/short-term  
(ii) Long term | (i) Operational  
(ii) under certain circumstances strategic (e.g., strategic risk mitigation) |
| ABM | (i) Real-time control  
(ii) Systems Design | Present/short-term | (i) Operational  
(ii) Tactical |
| ANN | Real Time control  
Operational optimisation of industrial or ecological systems | Short term/medium | Operational / Tactical |
| BN | Management/Policy making | Medium/long term | Tactical /Strategic |

consider such course of action, if no acceptable pathway of the present system seems possible).

Tactical level models are unable to produce long-term predictions, as they are limited to considering the control of the system as is. However, although the prediction validity is limited to a short-term or medium-term future, a wide range of critical decisions can be made just in time to prevent serious damage or gain significant profits. For example, as Table 3 shows, ANNs are extremely efficient in optimising functionality in industry by making accurate short-term or medium-term future predictions. For example, to identify faulty parts of the system before the damage occurs [32].

At the strategic level, managing and controlling the system might require applying structural changes to the system due to a perspective of long-term future. Even though the status quo of a system may be satisfactory, the secure path (trajectory in the state space) in which the system is proceeding is expected to turn unsafe and an unsustainable/uncertain future may be predicted for the system. Strategic level decision making must deal with uncertainty, and be able to consider alternative
APPLICATION OF DECISION MAKING MODELS

futures as well as the uncertainties inherent in the predictions of such futures.

In these cases therefore, the model of the system needs to be able to be manipulated, possibly evaluating the effects of significant changes to the system and allow the production and evaluation of different scenarios and possible future situations to help experts make effective policies which ensure a sustainable future for the system. BN models are capable of working with such uncertainty in the systems. For example, [12] use a BN model to integrate available information in order to assess risk factors to conduct scenario analysis. However, IO models can also be used (provided the system in question has a well understood set of physical laws based on which an IO model can be constructed). Increasingly, decision making in environmental management will have to consider hybrid systems, in terms of the systems of interest being partly natural and partly built.

In addition to the above, the feasibility of model building and use in decision making requires several other conditions, the present article limits the investigation to the selection of model types, while acknowledging that many limitations exist in terms of data accessibility, time, and resource (money, expert human resource, computational power), as well as limitations of current scientific theories available for exploitation.

7. Conclusion

In modelling for decision making, both theory and ‘specific information’ are required; where the theory defines the type of model needed to answer the concerns at hand, and the ‘specific information’ is what is needed to set the parameter values of the actual model. This paper discussed the major types of models which are currently in use in the field of environmental management decision making and explored the strengths and weaknesses of the models in dealing with different systems and performing at different levels of management.

While finding one specific model of a system to answer all the aforementioned queries is an ideal goal, it seems mostly impractical in real life. A simple example of this is the case in which the current model of the system is a linear, easily controllable model with some constant parameters, which are expected to change in the long-term future as a result of an inevitable structural change either within the system or externally imposed onto the system. In such a case, the model of the future system would be nonlinear and definitely different from the current one.
Therefore, separate models of the system would be required for the future and the transient state of the system.

The problem reveals the need for new types of theories and models, where the predictive capability of the models can be calibrated and validated in the vicinity of desired spatio-temporal trajectories, the viability of the system can be extrapolated, and options for structural (architectural) changes can be explored for strategic level decision-making.

At this stage, the types of decision making models in environmental management that we have identified fit into the four major categories that have been discussed in this paper. In the future we intend to develop a model type selection methodology which enables the modeller to choose the most suitable model to achieve the desired management goal for the specific system. This methodology would then be verified by applying it to several case studies.

References


Appendix I

APPLICATION OF DECISION MAKING MODELS


Received November, 2014