



Water for a Healthy Country

REVIEW OF MODELLING ALTERNATIVES FOR CLLAMM FUTURES

Rebecca Lester & Peter Fairweather

May 2008



Water for a Healthy Country

REVIEW OF MODELLING ALTERNATIVES FOR CLLAMM FUTURES

Rebecca Lester & Peter Fairweather

May 2008

Corresponding author:

Rebecca Lester

School of Biological Sciences

Flinders University

GPO Box 2100

Adelaide SA 5001

rebecca.lester@flinders.edu.au

Ph: +61 8 8201 3436

Fax: +61 8 8201 3015

ISSN: 1835-095X

ISBN: 978 0 643 09589 2

The Water for a Healthy Country Flagship is a research partnership between CSIRO, state and Australian governments, private and public industry and other research providers. The Flagship aims to achieve a tenfold increase in the economic, social and environmental benefits from water by 2025.

The Australian Government, through the Collaboration Fund, provides \$97M over seven years to the National Research Flagships to further enhance collaboration between CSIRO, Australian universities and other publicly funded research agencies, enabling the skills of the wider research community to be applied to the major national challenges targeted by the Flagships initiative.

© Commonwealth of Australia 2007 All rights reserved.

This work is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced by any process without prior written permission from the Commonwealth.

Citation: Lester, R.E. & Fairweather P.G., 2008. Review of modelling alternatives for CLLAMM Futures. CSIRO: Water for a Healthy Country National Research Flagship

DISCLAIMER

CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

For more information about Water for a Healthy Country Flagship visit or the National Research Flagship Initiative at www.csiro.au.

Foreword

The environmental assets of the Coorong, Lower Lakes and Murray Mouth (CLLAMM) region are currently under threat as a result of ongoing changes to the hydrological regime of the Murray River. While a number of initiatives are underway to halt or reverse this environmental decline, such as the Murray-Darling Basin Commission's "Living Murray", rehabilitation efforts are hampered by the lack of knowledge about the links between flows and ecological responses in the system.

The Coorong, Lower Lakes and Murray Mouth program is a collaborative research effort with the aim to produce a decision-support framework for environmental flow management for the CLLAMM region. This involves understanding the links between the key ecosystem drivers for the region (such as water level and salinity) and key ecological processes (maintenance and improvement of bird habitat, fish recruitment, etc). A second step will involve the development of tools to predict how ecological communities will respond to manipulations of the "management levers" for environmental flows in the region. These include flow releases from upstream reservoirs, the Lower Lakes barrages, and the Upper South-East Drainage scheme, and dredging of the Murray Mouth. The framework will attempt to evaluate the social, economic and environmental trade-offs for different scenarios of manipulation of management levers, as well as different future climate scenarios for the Murray-Darling Basin.

One of the most challenging tasks in the development of the framework will be how to predict the response of ecological communities to future changes in environmental conditions in the CLLAMM region. The CLLAMMecology Research Cluster is a partnership between CSIRO, the University of Adelaide, Flinders University and SARDI Aquatic Sciences that is supported through CSIRO's Flagship Collaboration Fund. CLLAMMecology brings together a range of skills in theoretical and applied ecology with the aim to produce a new generation of ecological response models for the CLLAMM region.

This report is part of a series summarising the output from the CLLAMMecology Research Cluster. Previous reports and additional information about the program can be found at <http://www.csiro.au/partnerships/CLLAMMecologyCluster.html>

Table of Contents

1. Introduction.....	1
1.1. Background to the CLLAMM Futures project	1
1.2. Ecological modelling	2
2. Alternative ecosystem states.....	3
2.1. Overview of theory.....	3
2.2. Past & current applications	3
2.3. Potential for CLLAMM Futures	4
3. Modelling techniques	5
3.1. Hysteresis modelling.....	5
3.1.1. Overview of theory	5
3.1.2. Past & current applications	6
3.1.3. Potential for CLLAMM Futures	6
3.2. Bifurcation plots from chaos theory	7
3.2.1. Overview of theory	7
3.2.2. Past & current applications	7
3.2.3. Potential for CLLAMM Futures	8
3.3. State and transition modelling	9
3.3.1. Overview of theory	9
3.3.2. Past & current applications	9
3.3.3. Potential for CLLAMM Futures	10
3.4. Classification and regression trees.....	11
3.4.1. Overview of theory	11
3.4.2. Past & current applications	11
3.4.3. Potential for CLLAMM Futures	12
3.5. Multivariate regression trees.....	13
3.5.1. Overview of theory	13
3.5.2. Past & current applications	13
3.5.3. Potential for CLLAMM Futures	13
3.6. Boosted regression trees.....	14
3.6.1. Overview of theory	14
3.6.2. Past & current applications	14
3.6.3. Potential for CLLAMM Futures	15
3.7. Multivariate adaptive regression splines.....	16
3.7.1. Overview of theory	16
3.7.2. Past & current applications	16
3.7.3. Potential for CLLAMM Futures	17
3.8. Structural equation modelling	18
3.8.1. Overview of theory	18
3.8.2. Past & current applications	18
3.8.3. Potential for CLLAMM Futures	19
3.9. Bayesian belief networks	20
3.9.1. Overview of theory	20
3.9.2. Past & current applications	20
3.9.3. Potential for CLLAMM Futures	21
3.10. Gaussian error propagation	22
3.10.1. Overview of theory	22
3.10.2. Past & current applications	22
3.10.3. Potential for CLLAMM Futures	22
3.11. Individual based modelling	23
3.11.1. Overview of theory.....	23
3.11.2. Past & current applications	23
3.11.3. Potential for CLLAMM Futures	24
3.12. Levels of evidence	25

3.12.1.	Overview of theory	25
3.12.2.	Past & current applications	25
3.12.3.	Potential for CLLAMM Futures	26
3.13.	Summary of techniques	27
	Advantages	27
	Disadvantages	27
	Applicability to Futures.....	27
4.	Application to CLLAMM Futures.....	30
4.1.	Classification and regression trees (CART)	30
4.2.	Structural equation modelling (SEM).....	33
4.3.	Bayesian belief networks (BBN)	35
5.	Summary of useful techniques	38
6.	A note regarding modelling responses of key species.....	40
	References	41

1. Introduction

This document is a review of the possible modelling alternatives that may be of use for the CLLAMM Futures project. It outlines the theory behind each identified possibility, lists the previous applications in which each modelling alternative has been applied and places each in the context of the objectives of the CLLAMM Futures project. Examples of those techniques that offer the most promise are also included to familiarise readers with the application of those techniques where necessary.

The objective of this document is to identify those modelling techniques that are likely to be of use for CLLAMM Futures, given the constraints of the data available (both collected prior to and over the lifetime of CLLAMM Futures) and the limited number of 'levers' available for manipulation by managers. In producing this document, we aim for a degree of transparency in the choice of suitable modelling techniques and to provide an opportunity for the input of managers, research peers and other key stakeholders in the project. It is our intention that the outcomes of this project will be available to assist in efforts to manage the region using the principles of adaptive ecological assessment and management (Holling 1978).

The scope of the document is limited to ecosystem-scale models and analyses that are the principle aim of CLLAMM Futures. Analysis techniques that will, no doubt, be of use in producing outputs such as nutrient budgets, food webs and key species response curves are not considered here. These outputs form part of other projects and fall outside the scope of the CLLAMM Futures, which is focussed on synthesising these outputs into response models for the ecosystem as a whole.

1.1. Background to the CLLAMM Futures project

CLLAMM Futures is one of four projects that form the CLLAMMecology Research Cluster (CLLAMMecology). CLLAMMecology is a collaborative research effort undertaken by the University of Adelaide, Flinders University, SARDI Aquatic Sciences, the Department of Environment and Heritage, AWQC, DWLBC and CSIRO Water for a Healthy Country. CLLAMMecology is focused on developing ecosystem level understanding of the Coorong, Lower Lakes and Murray Mouth (CLLAMM) region. It contends that ecosystem response is a result of climate and management actions interacting with ecosystem drivers which affect key species response, productivity and trophodynamics to determine habitat availability. These can then be used to predict future outcomes under a range of conditions and to assess management-defined "water benefits" (including a viable commercial fishery and the aesthetic value of the area, for example) associated with a variety of management scenarios. The other three projects that, with CLLAMM Futures comprise the cluster are Key Species Responses, Productivity and Trophodynamics, and Dynamic Habitats.

The CLLAMM Futures project aims to develop a set of models at the scale of the ecosystem to integrate the new knowledge generated by the other CLLAMMecology projects with existing knowledge. These models developed will be system-wide models that will be used to investigate the likelihood of a number of future scenarios, including possible management options and alternative states within the Coorong region. The models will be explicit in both space and time and will allow prediction of the response of the ecosystem to the identified system drivers (including climatic forcing and management options, for example). The models will be based on a combination of statistical and analytical modelling approaches and will eventually inform the creation of a set of ecosystem-health indicators for use by NRM stakeholders in monitoring the region.

The aim of CLLAMM Futures, to construct a single, system-wide response model for the condition of the region, represents a challenge for environmental science due to the plethora of ways of characterising such a multivariate system (Fairweather 1999a). This report is a summary of some of these ways of analysing and modelling multivariate systems and reviews

the usefulness of each to the current application – the prediction of ecosystem-scale response of the CLLAMM region to future scenarios.

1.2. Ecological modelling

Ecological modelling is a relatively new discipline that is rapidly developing as technology allows for greater computational capacity. New methods of representing and modelling multivariate, ecological data are currently being developed and numerous options exist for any one modelling problem (e.g. see Sutherland 2006 for a review of some of the possible approaches). For a complex modelling application, such as that associated with CLLAMM Futures, the best approach is likely to include a combination of several different modelling techniques.

A number of modelling techniques have been identified as of interest for CLLAMM Futures. These include methods for combining multivariate data into fate pathways, coherent typologies and modelling frameworks, many of which have shown promise when applied to environmental problems in aquatic environments.

Achieving the objective of the CLLAMM Futures project, to construct a single, system-wide response model for the region, involves meeting a number of considerable challenges. The dataset that is available for use in modelling is disparate in nature and collection methods, includes qualitative, semi-quantitative and quantitative data, is patchy in both space and time and does not cover all parameters of interest. The nature of the data, along with the complexity of the system as a whole, with estuarine, marine and hyper-saline regions, means that constructing a valid deterministic model is unlikely to be feasible. There is simply insufficient information for any deterministic model to adequately capture the relevant processes and to be adequately validated. For this reason, the CLLAMM Futures project is likely to focus on developing a statistical, or stochastic model, based on probabilities and analysis of existing data sets, rather than attempting to mathematically characterise all the interacting processes occurring.

When attempting to model complex ecosystems, some theoretical constructs of how an ecosystem may be structure are useful for simplifying the task. Several of the modelling techniques identified as being of potential use for CLLAMM Futures rely on the existence of alternative stable states. This construct is introduced before the review of modelling techniques is described.

2. Alternative ecosystem states

A number of the modelling techniques reviewed herein rely on the concept of alternative ecosystem states. This concept is introduced here, and will be referred to in the sections referring to hysteresis, bifurcation and chaos and state and transition modelling.

2.1. Overview of theory

The concept that some ecosystems are capable of supporting more than one stable community for a given set of environmental parameters has a long history in ecology, with Lotka (1956) referring to literature from the late 19th century as indicative of the phenomenon. These communities have been termed alternative stable states (Petraitis and Dudgeon 2004). For a system to support alternative stable states, it must have more than one equilibrium point, or basin of attraction across a given set of parameter values (Petraitis and Dudgeon 2004). Most authors agree that changes between states will occur via hysteresis, where small changes in parameter values result in large changes in state variables. The processes that drive changes between states, however, may be different from those that maintain existing states (Petraitis and Dudgeon 2004). It should be noted that the equilibria referred to as an alternative stable state refers to the biological condition of the ecosystem, as opposed to mathematical equilibria in a given model, although these may coincide (that is, a stable state, may coexist with a steady state, but this is not part of the definition of the concept). This is due to the difficulties associated with defining equilibria in nature (Petraitis and Dudgeon 2004).

While the theory behind alternative stable states is relatively straightforward, demonstrating their existence is not. There are problems associated with the definition of what constitutes the same habitat or environments, as well as the time scales over which a state could be said to be stable and self-sustaining. It can be difficult to differentiate between various states that are symptomatic of a system in decline (or recovery) and true alternative stable states. It is almost impossible to determine whether states are true alternative stable states without experimentation (Petraitis and Dudgeon 2004). Spatial heterogeneity has also been demonstrated to alter the response of communities to changes in environmental conditions, further complicating the notion of stable communities existing under the same conditions (van Nes and Scheffer 2005)

The existence of alternative stable states, however, has important implications for the management and rehabilitation of an ecosystem. The existence of alternative stable states suggests that the current condition of a system is a function of its past, and that previous management actions may have long-lasting consequences for a system. It also suggests that it may be difficult or impossible to return a system to a previous state, given that recreating the environmental conditions that existed prior to the change in state will not necessarily cause a recreation of that previous state.

2.2. Past & current applications

While many studies have attempted to demonstrate the existence of alternative stable states in practice, for the most part, they have fallen short on one or more criteria (Petraitis and Dudgeon 2004). Such an example is the analysis of the long-term recovery of Lake Veluwe, Netherlands from eutrophication (Ibelings *et al.* 2007). Clear hysteresis was observed (Ibelings *et al.* 2007), but the mechanism used to define the various alternative states was not stated.

The closest example of a definitive example of alternative stable states is the work that has been done on mussel beds and seaweed stands in North American coastal environments (Petraitis and Dudgeon 2004; 2005). The most recent study on this system indicates that post-disturbance recruitment by seaweed or mussels is largely serendipitous and that apparent

consumer control is a result, not a cause of community structure (Petraitis *et al.* unpublished data).

2.3. Potential for CLLAMM Futures

The use of the alternative stable states concept within CLLAMM Futures may not be possible. Replicated experimental manipulations of the type described by Petraitis and Dudgeon (2004) are not likely to be possible within the Coorong, and the timeframes needed to demonstrate states to be self-sustaining do not fall within the timeline of the project. However, it may be possible and desirable to identify various alternative ecosystem states for the system, regardless of whether or not they meet the criteria for stable states. That is, there may be states within the system that represent various degrees of stress or degradation from a healthy ecosystem to the reverse, regardless of whether they persist over time and are actually alternative equilibria. These states could be useful for simplifying the complex ecosystem to allow modelling of future events, despite lacking a formal definition as 'stable'. It should be recognised that if these alternative states are, in fact, true alternative stable states, then this may have implications for the overall management of the system, and affect the ease with which the system can be pushed between the various states in any effort to rehabilitate the system.

3. Modelling techniques

There are numerous modelling techniques available for the representation of complex, multivariate ecological systems. This report reviews 12 such techniques, identifying the advantages and disadvantages of each, and likely applications within the CLLAMM Futures project, and the wider CLLAMMecology cluster. Those techniques that are most likely to be of use are further explored using examples in the 'Application to CLLAMM Futures' section, later in the report. The techniques reviewed here are:

- 2.1) Hysteresis modelling;
- 2.2) Bifurcation plots from chaos theory;
- 2.3) State and transition modelling;
- 2.4) Classification and regression trees;
- 2.5) Multiple regression trees;
- 2.6) Boosted regression trees
- 2.7) Multivariate adaptive regression splines
- 2.8) Structural equation modelling;
- 2.9) Bayesian belief networks;
- 2.10) Gaussian error propagation;
- 2.11) Individual based modelling; and
- 2.12) Levels of evidence.

Other techniques were excluded from this review as they were obviously unlikely to be of use to CLLAMM Futures. For example, inferring ecological health (as used in AUSRIVAS, for example) and habitat suitability modelling approaches were excluded due to the lack of baseline data against which to compare the singular CLLAMM region. This exclusion is not a reflection on the general utility of these techniques, but rather due to their lack of applicability in this instance.

3.1. Hysteresis modelling

3.1.1. Overview of theory

Hysteresis is the phenomenon whereby small changes in a parameter variable cause catastrophic shift in the state variable (Petraitis and Dudgeon 2004). These catastrophic shifts cannot be reversed by a corresponding small reversal in the value of the parameter variable (Petraitis and Dudgeon 2004); that is, the trajectory of recovery is significantly different from the path of decline (Ibelings *et al.* 2007). The presence of hysteresis in a system has important implications for management (Webster and Harris 2004). Once thresholds are crossed, considerable effort is likely to be required to return the system to its previous state, should this be possible at all anymore (Webster and Harris 2004).

Hysteresis is the mechanism by which shifts between alternative stable states are purported to occur in ecological communities (Scheffer 1998). These shifts occur in systems that are subject to a bifurcation fold, or range of parameter values over which more than one stable state can occur (Petraitis and Dudgeon 2004). However, hysteresis, in itself, is not sufficient to demonstrate the existence of alternative stable states in a system (Petraitis and Dudgeon 2004).

The phenomenon of hysteresis in ecological systems makes the behaviour of those systems more difficult to model and predict (Petraitis and Dudgeon 2004). This is because of the distinct non-linearity between parameter and state variables, and the need to understand the behaviour of the system when parameter values change in both directions. It is also entirely possible that the trajectory taken by a system will differ for a second or subsequent parameter perturbation.

Hysteresis modelling is the name given to models that incorporate hysteresis in their action. It does not refer to a single analytical technique capable of detecting and accounting for hysteresis within a biological system. It appears to be the result of a modelling effort, rather than the initial objective. Also, hysteresis modelling is not undertaken using a single approach. All articles reviewed used different methods to investigate the degree and effect of hysteresis on the system studied. To our knowledge, there is also a lack of synthesis across the methods with no review of the various approaches, or justification for their use.

A range of modelling approaches and software have been used to demonstrate hysteresis (or lack thereof) in various system (e.g. MATLAB and Solver in Microsoft Excel) but, to our knowledge, no software packages have been designed with hysteresis modelling as an objective.

3.1.2. Past & current applications

Modelling that has demonstrated hysteresis in a system includes work simulating the biogeochemical response of a coastal lagoon to nitrogen loading, including processes in both the sediment and water column (Webster and Harris 2004). Hysteresis was observed in the response of macrophyte and phytoplankton abundance to changes in nutrient loading (Webster and Harris 2004). Hysteresis was also observed in the recovery of Lake Veluwe, Netherlands, from high levels of phosphate (Ibelings *et al.* 2007). Here, recovery of macrophyte abundance and water clarity was conditional on the presence of zebra mussels in the system (Ibelings *et al.* 2007).

Modelling has not always demonstrated hysteresis adequately. The decay characteristics of a new pesticide for use in oyster production were expected to show hysteresis (Felsot and Ruppert 2002). The results of a test application failed to show any such pattern. Hysteresis was able to be demonstrated in the rise of phosphate concentrations in the River Swale, UK over various flood peaks (Bowes *et al.* 2005). From the patterns in the hysteresis for each event, the authors inferred the source of the phosphate and made recommendations regarding suitable management actions. However, the results appear to be over-interpreted and the recommendations are somewhat unjustified.

In addition to these applications of the theory, van Nes and Scheffer (2005) demonstrated that spatial heterogeneity of a system can have implications for the hysteresis displayed by that system, and that a smooth gradient of environmental change leads to smaller hysteresis than random patchiness.

3.1.3. Potential for CLLAMM Futures

As an approach, hysteresis modelling appears to have little potential for use within CLLAMM Futures. There is no recognised approach to demonstrating hysteresis in a system and it is unlikely that data will be available currently for the full range of parameter values that may exhibit hysteresis, particularly for the recovery trajectory given the lack of environmental flows in the CLLAMM region.

However, it will be important to consider the effects of hysteresis and how they may affect the models that are produced for the region to increase the accuracy of predictions made. That is, we need to be aware of the possibility of hysteresis occurring in the CLLAMM ecosystem and how that may affect the developed models. Consideration will need to be given to thresholds values of parameters, and for combinations of effects and how these may affect the various components of the system. In particular, the effect of hysteresis should be considered for the transition between the various hypothesised states for the Coorong, although it is unclear how this can be explicitly incorporated at this stage. It is possible that the biogeochemical model may be able to model any hysteresis in the cycling of nutrients in the region (c.f. Webster and Harris 2004), which may possibly be able to be used as an input for models of the broader system.

3.2. Bifurcation plots from chaos theory

3.2.1. Overview of theory

A bifurcation is the division of a time series into two paths. A series of bifurcations in a non-linear system can lead to the development of chaos. Chaotic behaviour has been observed in a number of otherwise deterministic ecological models. It refers to the development of apparently random behaviour as a parameter in a deterministic model passes through a threshold (Hansell *et al.* 1997). Exact prediction for a system that exhibits chaotic behaviour is not possible because of the sensitivity of the system to the initial values used (Hansell *et al.* 1997). The presence of bifurcations can affect our understanding of an ecosystem, specifically its predictability, but can also affect local extinction rates and biodiversity (Morozov *et al.* 2004).

In practice, conclusive evidence of chaotic behaviour is yet to be demonstrated for ecological data, although it has been shown under laboratory conditions (Morozov *et al.* 2004). However, bifurcations are hypothesised to occur in systems with alternative stable states, where small changes in parameter values can result in large shifts in state variables (Petraitis and Dudgeon 2004). The large shift occurs at a bifurcation fold, which is a range of values for the parameter of interest for which more than one state is stable (Petraitis and Dudgeon 2004). Reversal of state changes are not possible with corresponding small reversal in the parameter value, rather the parameter value must be pushed to the other edge of the bifurcation fold (i.e. hysteresis) (Petraitis and Dudgeon 2004). Terminology used to describe bifurcations and their effects include 'tipping points', 'catastrophes', 'regime shifts' and 'phase changes'.

Of prime interest to managers is the ability to detect in advance when a bifurcation is likely to occur in a system. In most instances, bifurcations are only recognised after the fact, but some techniques can be used to provide clues to future possible bifurcations. When ecosystems are known to have chaotic inputs (e.g. climate), they should be modelled stochastically because they are sufficiently complex to make a full deterministic model impractical (Hansell *et al.* 1997). By stochastic modelling, we refer to the inclusion of random fluctuation within the model, which often means that statistical modelling is the most practical option. As such, bootstrapped time-series analyses are one of the main tools available to understand the underlying dynamics of a system. Approaches including reconstructed past dynamics may provide clues to future behaviour. Periodic influences can also be removed to reveal underlying structure (Hansell *et al.* 1997). In particular, the behaviour of a time series near bifurcation points can be useful in providing clues to predict other bifurcations (Dennis *et al.* 2001), because the frequency of individual variations changes significantly before the onset of bifurcation (Hansell *et al.* 1997). This is attributed to the conceptual basin around an attractor becoming shallower and the increased likelihood that random perturbations will shift the system to a new basin (Hansell *et al.* 1997). One tool to identify chaotic-type behaviour in stochastic models is the stochastic Lyapunov exponent which identifies strange attractors in a stochastic time series. (Dennis *et al.* 2001)

A bifurcation continuation program called LOCBIF is one piece of software that has been used to explore bifurcation plots. We are unaware of others specifically designed to identify bifurcations.

3.2.2. Past & current applications

Bifurcation plots from chaos are usually used for deterministic models such as population models or trophic food chains. For example, Peel *et al.* (2005) demonstrated that a three-level trophic food chain showed chaotic oscillation, and bifurcation diagrams indicated hysteresis was also present in the system. The authors attributed these characteristics to the control of a subtle food-prey-predator interaction on the system (Peel *et al.* 2005). Models such as that constructed by Dennis *et al.* (2001) to describe population dynamics in flour beetles also exhibit chaotic behaviour. Dennis *et al.* (2001) also attempted to introduce a stochastic element into

their model, and investigated the behaviour of that where the deterministic model was chaotic. The Allee effect in population models has also been shown to induce chaotic behaviour for systems that are explicitly modelled spatially (Morozov *et al.* 2004). Usually, spatially-explicit models will not show chaotic behaviour, but the inclusion of an Allee effect (i.e. a positive relationship between population density and per-capita birth rate) alters the outcome of the modelling (Morozov *et al.* 2004).

Other past applications include the use of bifurcation and hysteresis dynamics to generate bursting oscillations of the type observed in neurons, in β -cells of the pancreas and for population dynamics (Francoise and Piquet 2005). Bifurcation analyses have also been used extensively when modelling genetic behaviour. For example, they were used to characterise circadian oscillations in *Drosophila* with a view to gaining a better understanding of the mechanisms behind conditions like seasonal affective disorder. (Tsumoto *et al.* 2006). Bifurcations have also been identified in the spectra of manatee calls (Mann *et al.* 2006).

Examples of applications that examine the structure of noise around a bifurcation are rare, but have been advocated by Hansell *et al.* (1997), Dennis *et al.* (2001) and Greenman and Benton (2003).

3.2.3. Potential for CLLAMM Futures

Given the focus on statistical modelling within CLLAMM Futures, rather than a deterministic approach, it is unlikely that bifurcation plots from chaos will be useful. For the most part, stochastic models are not considered to exhibit chaotic behaviour, so the use of techniques such as bifurcation to describe it are not likely to be required, although, as for hysteresis, an awareness of the possibility would be useful.

There is some potential for components of the models that are more deterministic in nature (e.g. biogeochemical modelling, food web analyses) to exhibit chaotic behaviour, in which case these techniques may be of use. There is also potential to explore the equivalent of chaotic behaviour in the stochastic models (strange attractors) using techniques such as the stochastic Lyapunov exponent. This may be useful in exploring the behaviour of systems that are close to a bifurcation, and assist in the prediction of future changes in state.

3.3. State and transition modelling

3.3.1. Overview of theory

State and transition (S&T) models draw on the theory of alternative states for a community (Wilkinson *et al.* 2005), although the states do not necessarily need to be true alternative stable states. The technique represents a system as a box-and-arrow flow diagram, which defines discrete system states, and the events and processes that cause the system to move between these states (transitions) (Plant and Vayssieres 2000). A transition is defined as the crossing of some parameter threshold so that the system cannot persist partway between states (Plant and Vayssieres 2000). Any given state will persist in the system until a transition condition is met, forcing the system to a new state (Wilkinson *et al.* 2005). Transition conditions can be based on natural or anthropogenic events and processes (Wilkinson *et al.* 2005). S&T models are designed to be a framework for allocating a system to various states. The states and transitions are inputs to the model, rather than outputs (Perry and Enright 2002), and can be theoretical, empirical, based on expert opinion or some combination thereof (Jackson and Bartolome 2002).

S&T models have several advantages, including that they do not assume a system is in equilibrium; their ability to incorporate both biotic and abiotic parameters simultaneously, and their applicability across a range of scales (Jackson and Bartolome 2002). They are also able to be linked to GIS, so as to be spatially explicit (Plant and Vayssieres 2000), and allow for a mosaic of states across a landscape (Wilkinson *et al.* 2005). Another advantage is that transition rules can explicitly account for hysteresis in the parameter values that will drive movement between states versus back again. However, S&T models are limiting in their ability to simulate gradual changes in state, and are not able to simulate changes associated with an evolving system, which may hinder the ability of the model to simulate the system accurately (Plant and Vayssieres 2000).

Few published examples of S&T models described a base for the model. The exception was Plant and Vayssieres (2000) who wrote their model as C++ code.

3.3.2. Past & current applications

The most advanced applications of S&T models are arguably those that have been applied to rangeland management in the USA (e.g. see Bestelmeyer *et al.* 2004; Westoby *et al.* 1989). S&T models have been produced to incorporate knowledge from a variety of sources, usually in a qualitative way, to enable managers to explore the effects of different management options (Briske *et al.* 2003).

More generally, S&T models have often been applied using descriptive models based on expert opinion, and have been criticised for this (Petraitis and Dudgeon 2004). However, some examples of S&T models applied to natural resource management include the development of a model to determine when mine rehabilitation sites are on track to meet rehabilitation regulations and lease relinquishment conditions (Grant 2006), the application of an S&T model for the effective allocation of resources in tropical forest restoration within the Galapagos National Park, particularly regarding the control of exotic species (Wilkinson *et al.* 2005), to describe changes in community structure in a Californian grassland ecosystem (Jackson and Bartolome 2002) and to combine data on community composition, plant characteristics and ecosystem properties to assist in the management of sub-alpine grasslands in the French Alps (Quetier *et al.* 2007).

S&T models have also been extended to be spatially explicit so as to test scenarios. An S&T model was linked to GIS to model specific scenarios of change in oak woodlands in the USA (Plant and Vayssieres 2000). The authors used a dynamic simulation of spatial and aspatial model components linked in an alternating sequence to develop an understanding of the consequences of various management actions (Plant and Vayssieres 2000). Spatially-explicit S&T modelling has also been used to determine whether the present condition of an ecosystem

was natural, or as a result of anthropogenic activities, such as for Mont Do in New Caledonia (Perry and Enright 2002). The final S&T model used repeated simulation to determine the likelihood that current conditions were a result from changes in the burning regime, and modelled different aspects of the system on two different spatial scales; some at the landscape scale, and some locally (Perry and Enright 2002).

3.3.3. Potential for CLLAMM Futures

S&T modelling has significant potential for use within CLLAMM Futures. It is able to simplify the complex ecosystem to a series of states with explicit rules governing the movement between states. The states and transition rules are inputs to the model, so are not limited to be based on expert opinion and theory as is common in past applications, but can be based on analyses of data collected both prior to and during CLLAMMecology. By reducing the natural system to a series of states and a list of defined transitions between each, we are able to simplify the ecosystem to a point where meaningful and manageable analyses can occur. This can help overcome the common problem that all but the simplest of ecosystems are too complicated to be effectively modelled in a deterministic fashion.

The ability to link S&T modelling to GIS to make it spatially explicit is highly attractive in the application to the CLLAMM region, with its distinct environmental gradients, and the ability of S&T to be run as a series of simulations for various management activities would provide a mechanism by which CLLAMM Futures models could link to the output of the Dynamic Habitats project.

One of the major limitations of the S&T approach is that possible states may exist outside the realm of the data we have available at any time. That is, there may be states that are more degraded than we have evidence of, or natural and healthy, but simply haven't occurred within the timeframe of data collection. How to best capture and include these states is a challenge that is not trivial.

3.4. Classification and regression trees

3.4.1. Overview of theory

Classification and regression trees (CART) is a technique that is well-suited to the analysis of ecological data (De'ath and Fabricius 2000). CART simultaneously combines the use of classification trees which are used for categorical dependent variables with that of regression trees for use on continuous dependent variables (Wilkinson 2004). CART explains the variation in a univariate response variable through repeated splitting of the data into increasingly homogenous groups based on a variety of exploratory variables (De'ath and Fabricius 2000). This can be applied to identify the conditions that define the differences between several fates. It has the capacity to deal with non-linear relationships with high-order interactions and cases with missing values (De'ath and Fabricius 2000).

The resultant tree can be displayed graphically for ease of interpretation (De'ath and Fabricius 2000). The graphical display can include the number of observations in each node, the values of the explanatory variables for each group, the number of cases misclassified, and the proportion of variance explained (De'ath and Fabricius 2000; Wilkinson 2004). A variety of loss functions appropriate to different situations are available upon which to develop the tree (Wilkinson 2004).

CART has the advantage of exploring all possible combinations of splits, so it will find the optimal split for each variable (Wilkinson 2004). This is superior to the algorithm used by CHAID (another classification technique), which is designed to be less computationally-expensive, and can only classify categorical variables (Wilkinson 2004). In contrast, CART has the ability to handle a range of data types, and can be used both for descriptive purposes and for prediction of the classification of new cases (De'ath and Fabricius 2000). The ability of CART to use mixtures of continuous and categorical variables allows the identification of site, time and treatment interactions that may not be apparent when using linear regression techniques (Jackson and Bartolome 2002).

One of the limitations of the approach is that different commercial packages will produce different outputs from the same dataset as they are based on different splitting algorithms (Wilkinson 2004). This problem is compounded by a lack of documentation in some packages outlining their individual algorithms (Wilkinson 2004). There is also a tendency for overly-complex trees to be developed, so tree pruning is needed (Breiman *et al.* 1984). Cross-validation is also an important step in determining the predictive error of the tree for new data (Wilkinson 2004). Procedures for undertaking this involving boot-strapping and jack-knifing are relatively well accepted.

CART analysis can be undertaken with several software packages, including DTREG and CART. SYSTAT and R are also able to undertake the analyses.

3.4.2. Past & current applications

CART in ecological analyses is typically used as a technique to relate the distribution of target species to a suite of habitat variables. Examples of this application include the description of blue crab habitat based on environmental characteristics of Chesapeake Bay (King *et al.* 2005), the classification of various soft coral communities according to habitat type (De'ath and Fabricius 2000), the prediction of lakes vulnerable to invasion (based on lake attributes) by rainbow smelt in eastern North America (Mercado-Silva *et al.* 2006), evaluating temporal relationships between environmental variables and aggregation patterns in Peruvian fish (anchovy and sardines) (Gutierrez *et al.* 2007), or vegetation change in Acacia landscapes in Queensland, Australia (Seabrook *et al.* 2007), and comparison of habitat preferences for a range of marine mammals in Scotland (MacLeod *et al.* 2007).

The *post hoc* exploration of variables in Bayesian hierarchical modelling is another previous application of CART. Large-scale parameters such as climate, landscape composition and human disturbance were investigated for *post hoc* inclusion in a model describing species-habitat relationships of birds in prairie to hardwood transition areas of eastern North America (Thogmartin and Knutson 2007). CART has also been applied to define differences in groups identified by principal components analysis in terms of environmental parameters for bird communities in northeastern US lakes (Allen and O'Connor 2000) and between groups identified by CHAID to determine that drought characteristics were the best predictors for fish species composition in German rivers (Wolter and Menzel 2005).

On at least one occasion, CART has been used to assess transitions between different community states. Jackson and Bartolome (2002) used CART to identify the climatic and habitat characteristics that lead to changes in vegetation community in Californian grasslands. The technique correctly allocated 63% of the observed transitions, with location and year being the most important predictors of transition (Jackson and Bartolome 2002). Management parameters were also included, but did not prove to be predictive of most community transitions (Jackson and Bartolome 2002). The advantages of the technique in this application included the ability to identify site x time x treatment interactions that may have been masked in linear or parametric analyses and the ability to incorporate controlling variables such as climate before more localised parameters describing management actions (Jackson and Bartolome 2002).

3.4.3. Potential for CLLAMM Futures

CART has a number of potential applications within the CLLAMM Futures project. It could be used to parameterise the hypothesised alternative states for the CLLAMM region based on environmental variables. CART could be used to compare the various alternative states and allow the defining environmental variables to be drilled down to those that are most critical in the demarcation between each. CART could also provide important information regarding threshold values for those environmental variables that are likely to lead to transitions between states.

Within CLLAMM Futures, CART could also be used to relate the distribution of key species to those environmental parameters and identify the components that comprise 'suitable habitat' for those species. It could also be used to evaluate the distribution of key species or key communities within the various states (e.g. waterfowl) with respect to environmental variables. This would allow changes in space and time to be explored explicitly, and predictions to be made for each of the management scenarios investigated.

An example of a CART analysis is provided in the 'Application to CLLAMM Futures' section of the report below.

3.5. Multivariate regression trees

3.5.1. Overview of theory

Multivariate regression trees (MRT) are an extension of the univariate regression trees used in applications such as CART. The univariate dependent variable in CART is, in effect, replaced by a multivariate response in MRT (De'ath 2002). In MRT, clusters of dependent variables modelled as a multivariate response represent species assemblages, with the community composition related to environmental conditions (De'ath 2002).

As for CART, MRT can be used for exploratory or predictive applications, and is suited for complex ecological data that involves high-order interactions, non-linear relationships and missing data (De'ath 2002). It also operates by creating clusters of sites by splitting the dependant variables based on the environmental conditions of each site (De'ath 2002). Some form of cross-validation process is commonly used to determine an appropriate sub-tree that is not over-fitted (Larsen and Speckman 2004).

MRT is robust to patchy ecological data which can include skew and kurtosis, missing values, non-linearity and interaction between independent variables (De'ath 2002). MRT is invariant to monotonic transformation of dependent variables and a general absence of model assumptions leads to more robustness (De'ath 2002). Existing MRTs can also be used to predict species compositions for new sites, or sites with a new set of environmental conditions (De'ath 2002).

Disadvantages of MRT include sensitivity to some outliers (De'ath 2002), and a selection bias towards variables with more potential split points (Hsiao and Shih 2007), although methods for dealing with both possibilities exist.

MRT analyses can be conducted in S, S-plus or R software packages.

3.5.2. Past & current applications

MRT has been used to characterise communities based on environmental variables in several different ecosystems. It was used to relate the abundances of 12 species of hunting spider to environmental variables, showing that the presence of twigs and water described the majority of variation within those communities (De'ath 2002). It was also used to relate the abundances of 12 co-occurring plant species to numerous categorical variables in Missouri (Larsen and Speckman 2004) and MRT was used in conjunction with PERMANOVA to aid in the planning of marine protected areas by determining the habitat characteristics associated with various reef fish assemblages in the French Mediterranean (Claudet *et al.* 2006).

MRT has also been applied to the Great Barrier Reef coral communities, relating composition to position on the reef and depth and identifying indicator species (DeVantier *et al.* 2006) and to characterise patterns in zooplankton communities based on the surrounding biotic and abiotic characteristics in Alberta, Canada (Norlin *et al.* 2006).

3.5.3. Potential for CLLAMM Futures

As for CART, MRT has significant potential for use within CLLAMM Futures. MRT provides the added flexibility of analysing a multivariate response to multiple environmental variables (i.e. it has the ability to deal with several key species at once) and could be used to predict species composition for the future scenarios under investigation. It could also be used to validate the proposed alternative states for the system, or to develop alternatives, should those proposed prove inconsistent with the available data. MRT also has the ability to identify species indicative of particular communities, which could be used by managers to monitor the presence and distribution of both desirable and undesirable states in the future.

3.6. Boosted regression trees

3.6.1. Overview of theory

Boosted regression trees (BRT) are an extension of classification and regression tree analyses that allow greater predictive accuracy than conventional CART models (De'ath 2007).

Several methods for reducing the predictive error of CART models exist, including model averaging, shrinkage methods and a technique known as 'bagging'. These typically reduce the variance component of predictive error, leading to greater predictive accuracy of the final overall model adopted (De'ath 2007).

Another method for increasing the predictive accuracy of trees is to use a technique known as 'boosting' (De'ath 2007). This uses a set of simple (usually relatively weak) predictors (i.e. single-split classification trees which artificially limit the output to a tree with a single split in one independent variable) and creates a series of trees, with each subsequent tree created from a re-weighting of the previous dataset (De'ath 2007). After the formation of a given tree, the are re-weighted based on the predictive accuracy of that tree, with those cases that were incorrectly classified given a higher weight than those that were correctly classified (De'ath 2007). In this way, those cases that are difficult to correctly classify become increasingly weighted, and are therefore more likely to be correctly predicted in later trees.

For any given iteration, some set proportion (usually 40-60%) of the training data is used to increase the predictive accuracy of the final model (De'ath 2007). The final model is similar to a regression model, where the individual model terms are actually simple CART models (Leathwick *et al.* 2006). A method of model averaging can also be incorporated into the analysis, leading to aggregated boosted trees (ABT), which have been shown to have still greater predictive accuracy than BRT (De'ath 2007).

BRT and ABT retain many of the desirable characteristics of other tree analyses. They are able to include numerous types of response variables (including numeric, categorical and censored data), predictor variables (numeric and categorical), as well as a variety of loss functions (such as Gaussian, binomial, Poisson and robust). They are also able to identify interactions between predictor variables, by using CARTs that contain more than a single predictive split (i.e. two or more splits). However, the accurate representation of such interactions can be difficult in some instances (Leathwick *et al.* 2006). Another potential pitfall for BRT models is that the model will continue to grow until the training set is perfectly represented (i.e. overfitted), which will decrease the predictive accuracy of the model for new datasets (Leathwick *et al.* 2006).

BRT and ABT can be implemented within the R software environment for statistical computation.

3.6.2. Past & current applications

BRTs were used to analyse species richness data for oceanic fishes surrounding New Zealand (Leathwick *et al.* 2006). Environmental and trawl characteristics were related to species richness and depth was found to be the most important predictor of richness, with variation in chlorophyll a concentrations, zones of mixing and temperature fluctuations also identified as important predictors (Leathwick *et al.* 2006). Models constructed using BRT were compared with those produced using generalised additive models and the BRT models gave superior predictive accuracy (Leathwick *et al.* 2006).

BRT was also applied to determine the freshwater residence of barramundi, to determine the contribution of various nursery habitats of different salinities (Cappo *et al.* 2005). Scales taken from the fish were analysed for a variety of elements, and BRT was used to classify each fish as either 'freshwater' or 'estuarine' (Cappo *et al.* 2005). The technique was successful both on the scale as an entity, but also on the annuli within the scales, suggesting that the technique could

be used to understand the links between barramundi growth and freshwater flows (Cappo *et al.* 2005).

BRTs were also used as a part of a suite of tree analysis methods to classify the agricultural practices relevant for measurement of carbon sequestration (Bricklemyer *et al.* 2007). Satellite imagery was used to identify tillage practices, soil disturbances and other management practices that were likely to affect carbon budgets, and BRT demonstrated the efficacy of these data for that purpose (Bricklemyer *et al.* 2007).

3.6.3. Potential for CLLAMM Futures

BRT is not likely to be of use in the CLLAMM Futures project as a method of increasing the predictive power of models produced using CART or MRT. This is due to the 'black box' format of the output, where the contributing trees are not able to be examined individually. This format precludes the predictive variables and cut-off point from being identified and incorporated into further analyses. Given the CLLAMM Futures focus on combining output from more than one modelling technique, this will limit the utility of the technique.

3.7. Multivariate adaptive regression splines

3.7.1. Overview of theory

Multivariate adaptive regression splines (MARS) is a form of localised regression modelling. MARS divides the data into a number of potentially overlapping regions, each of which is described by truncated splines known as 'basis functions' (Deconinck *et al.* 2007). These basis functions are paired, describing the space on either side of a threshold value, known as a 'knot' (Deconinck *et al.* 2007). For each descriptive variable, a knot and pair of basis functions are defined and are then combined into an overall model (Deconinck *et al.* 2007). This model describes complex, non-linear relationships within the data (Leathwick *et al.* 2006). It can be applied to one or more response variables of interest. Cross-validation and pruning of the resultant models is required (Deconinck *et al.* 2007).

A major advantage of MARS is the significantly shorter computation time compared with other analyses such as generalised additive modelling and generalised linear modelling (Leathwick *et al.* 2006). MARS is considered superior to regression trees as it applies a smoothing algorithm that is more likely to detect global patterns in the data (Prasad *et al.* 2006).

The predictive ability of MARS and its utility for ecological datasets is unclear. Some authors describe comparable predictive ability with more time-consuming techniques (Leathwick *et al.* 2006), while others suggest MARS provided a significantly worse prediction than other analyses (Prasad *et al.* 2006). Prasad *et al.* (2006) found the technique unstable, vulnerable to localised patterns in the data, sensitive to extrapolation and difficult to fit effectively. (Leathwick *et al.* 2006) however obtained good predictive ability and did not mention computational difficulties in fitting the models. Further clouding the issue, (Prasad *et al.*, 2006) considered the output difficult to interpret from an ecological perspective, while (Leathwick *et al.* 2006) listed this as a strength of the technique.

The inconsistencies in the literature also extend to the best software with which to apply MARS. Some authors use the R platform (Leathwick *et al.* 2006) while others found it deficient and used proprietary software called 'MARS' (Prasad *et al.* 2006).

3.7.2. Past & current applications

Many of the applications of MARS in ecology involve relating landscape or habitat characteristics to species distributions. The current and possible future distributions of tree species in the eastern USA was modelled based on present and future climate scenarios, but techniques other than MARS (i.e. BRT and random forests) were found to be superior (Prasad *et al.* 2006). An extensive set of data describing fish distributions was related to habitat characteristics in New Zealand rivers and streams with more success (Leathwick *et al.* 2006), landscape variables were used to predict the distribution of forest types in Michigan (Henne *et al.* 2007) and the technique was been applied to relate landscape characteristics to predict the size and shape of habitat patches (Ferrarini *et al.* 2005).

Outside of ecology, MARS has also been used to predict the absorption of drugs via the gastrointestinal system in humans (Deconinck *et al.* 2007).

3.7.3. Potential for CLLAMM Futures

There are a number of inconsistencies in the scientific literature regarding the use of MARS in ecological applications. At least some authors consider it a useful tool for predicting species distributions based on environmental variables but others disagree, suggesting its utility may vary between situations. The technique is potentially useful for CLLAMM Futures, but further investigation using CLLAMM Futures datasets would be required to verify this. As there are other techniques that perform similar functions where the benefits are less ambiguous, we may not undertake such an investigation unless other options are less useful than is anticipated.

3.8. Structural equation modelling

3.8.1. Overview of theory

Structural equation modelling (SEM) is a multivariate statistical procedure that allows researchers to evaluate the application of theoretical constructs to ecological systems. It provides a mechanism to reliably test hypothesised relationships between variables, accounting for measurement error and for indirect, as well as direct, effects (Ferguson 2006).

SEM is an extension of path analysis that incorporates aspects of factor analysis and regression analysis (Malaeb *et al.* 2000) and seeks to minimise the differences between the variance and covariance matrices of the model under investigation and the observed data (Reckhow *et al.* 2005). A combination of continuous and categorical data can be incorporated (Malaeb *et al.* 2000).

The ability to explicitly account for measurement error and variance that is unaccounted for is a significant advantage of SEM and allows more accurate predictions of complex relationships than multiple regression analysis, for example (Malaeb *et al.* 2000). It also has the advantage that confirmatory tests are possible, making it less purely descriptive than other multivariate techniques like cluster analysis, discriminant function analysis and MANOVA. The structure of the model itself can also be tested, adding to the generality of findings (Grace 2006).

The limitations of SEM include difficulty in incorporating space and time explicitly in the model, and difficulty in the interpretation of causality on the basis of correlations or covariations alone (Stoner and Joern 2004). There are, however, mechanisms for dealing with these issues that can be explored. Observations made by Petraitis *et al.* (1996) regarding the pitfalls of path analysis (a special case of SEM), are likely to be applicable. They warn that undetected collinearity and small sample sizes can have large effects on the resultant path diagram, and therefore, the interpretation of the analysis.

A number of software packages are available to undertake SEM. These include EQS, AMOS, Mplus and LISREL. The various packages have similar functionality, but differ in user interface design and the level of specialist knowledge required to operate each.

3.8.2. Past & current applications

SEM has had limited previous application in ecological systems. In an estuarine setting, it has been used to relate environmental variables collected by the US Environmental Protection Agency with invertebrate characteristics (Malaeb *et al.* 2000). It performed well in this instance, accounting for 81% and 69% of the observed variability in biodiversity and growth potential (abundance), respectively (Malaeb *et al.* 2000).

It has been used to compare competing explanations for differences in evolutionary strategy in pinniped lactation, demonstrating that presence of predators has resulted in shorter lactation than in predator-free areas, and that the availability of safe, terrestrial breeding habitats has encouraged the formation of harems, increased sexual dimorphism and lowered fat content in milk (Ferguson 2006).

SEM modelling was also used to demonstrate that inter-specific competition plays a greater role than habitat composition in the density of the Italian roe deer (*Capreolus capreolus italicus*) (Focardi *et al.* 2006).

SEM was applied to identify the best predictors of water quality in four case-study water bodies (Reckhow *et al.* 2005). This allowed managers to identify critical levels of each, quantify the risks associated with failing to meet targets and to concentrate future resources on monitoring those parameters that prove the most informative (Reckhow *et al.* 2005).

The analysis technique has also determined that local, habitat-scale factors and management actions influence insect communities in tallgrass prairies in North America, over and above landscape-scale factors, highlighting the need for managers who are planning conservation intervention to focus on the smaller scale (Stoner and Joern 2004).

3.8.3. Potential for CLLAMM Futures

For the CLLAMM Futures project, there is scope to use SEM as a method to quantify the conceptual models currently being used by NRM agencies. These conceptual models are being re-developed to be expressed in terms of the availability of suitable habitat, but the significant drivers of suitable habitat have yet to be identified for each key species. The findings from these analyses regarding the most important drivers for each key species can then be used as a basis for developing indicator sets for each of the stated management objectives for the region.

SEM is demonstrated in an example in the 'Application to CLLAMM Futures' section of the report below.

3.9. Bayesian belief networks

3.9.1. Overview of theory

Bayesian belief networks (BBN) is a technique that applies a Bayesian analysis approach to an influence diagram, using past experience and empirical data to test a hypothesised series of relationships between variables (Wooldridge and Done 2004). BBNs are also known as Bayes nets in some instances.

In a BBN, a series of nodes represent either independent or dependant variables, with the links between them representing the relationships between those variables. These links may be based on a combination of empirical, statistical or mechanistic evidence and can either reflect direct causal links or the overall effect of a series of associations (Wooldridge *et al.* 2005). A wide range of data types can be incorporated into a BBN, including the output from a range of other modelling techniques, as a meta-model (Varis and Kuikka 1997), or expert opinion on parameters for which empirical data are not available (McCann *et al.* 2006). The influence of any one parameter on another can be adjusted for either the strength of the interaction between the two or the certainty with which both are measured. The level of dependence between parameters creates the network (Tattari *et al.* 2003).

One of the main advantages of BBNs is that they are able to incorporate new information as it becomes available and can update their predictions based on this information (Wooldridge *et al.* 2005). This means that the BBN can be interactively refined over time as new data become available. The technique is also able to propagate uncertainty and measurement error through the network and explicitly present the precision of the eventual predictions (McCann *et al.* 2006). BBNs can also be used to identify the most likely set of parameters that will lead to a given outcome by solving the conditional probabilities for that outcome in reverse through the network structure (McCann *et al.* 2006).

There are a number of limitations associated with the use of BBNs. They require that a full probability structure is specified for each parameter and relationship within the network, which can be difficult to produce (McCann *et al.* 2006). They also require that continuous variables are expressed as discrete increments, and conditional probability tables can become complex quickly where there are a large number of states or parents for each node (McCann *et al.* 2006). There may be an over-reliance on expert knowledge to produce these conditional probability tables, which, without a structured approach, can lead to bias (McCann *et al.* 2006). Also, BBNs are poor at explicitly representing spatial and temporal dynamics and are unable to include direct feedback loops (McCann *et al.* 2006). BBNs can be built using Netica software.

Another technique that may be of use in conjunction with BBNs is information gap theory, which investigates the robustness of a decision to the level of uncertainty in the network (Regan *et al.* 2005).

3.9.2. Past & current applications

BBNs have usually been used in situations where decisions need to be made regarding the management of some ecosystem. For example, they have been used to develop a model of the response of coral reefs to bleaching events, with a view to identifying where best to direct management effort to achieve a desirable outcome (Wooldridge *et al.* 2005).

A similar aim of assessing policy scenarios lay behind the construction of a BBN describing caribou seasonal range recovery in British Columbia (McNay *et al.* 2006).

Expert opinion is often the basis (or one of the bases) of BBNs. For example, a BBN was constructed to determine whether exploitation of an aquifer on the Iberian Peninsula was sustainable (Matin de Santa Olalla *et al.* 2005).

Another BBN was constructed to measure uncertainty regarding the functioning of established buffer zones with respect to water protection and biodiversity in Finland (Tattari *et al.* 2003).

BBNs can also be used to characterise dependencies regarding the best measures of a system's response. (Wooldridge and Done 2004) used the tool to refine their beliefs about the best monitoring surrogate for measuring the likelihood of a bleaching event occurring.

The BBN approach has also been used to combine competing methods for estimating a sustainable annual take from salmon fisheries in the Baltic Sea (Varis and Kuikka 1997). This approach explicitly incorporated three different modelling approaches and combined their estimates to give a more comprehensive estimate of the risks associated with various take limits. The models that were incorporated included deterministic, non-linear steady state and statistical models (Varis and Kuikka 1997).

3.9.3. Potential for CLLAMM Futures

BBN has significant potential for use in the CLLAMM Futures project, both because of its utility when assessing management options and in its ability to combine various forms of data. It has the potential to be the framework within which the other various modelling techniques can be combined into an ecosystem-scale model to evaluate the response of the system as a whole. The approach has the capacity to be used to predict the effect of various management and climatic scenarios and it may be useful in inferring the probabilities of alternative model outcomes for a given set of causal conditions or values of key ecological parameters.

In order of BBN to be a useful technique for CLLAMM Futures, the difficulties associated with the inclusion of feedback loops will need to be addressed. The degree of this difficulty will depend on whether feedback loops can be adequately described within the S&T framework, or whether they need to be addressed separately in the BBN. Further investigation into this issue is needed.

An example BBN is provided below in the 'Application to CLLAMM Futures' section of the report.

3.10. Gaussian error propagation

3.10.1. Overview of theory

Gaussian error propagation (GEP) is an error analysis technique that has recently been introduced to the ecological literature. GEP is a technique that enables error and uncertainty to be determined from the interaction of multiple variables (Lo 2005) and represents an alternative to more common methods in ecology, such as Monte Carlo simulation analyses and bootstrapping (Lo 2005). A simple form of GEP is commonly used in physics.

GEP is a flexible method that is less computationally expensive than methods like Monte Carlo, and has additional power in that it allows the structure of errors to be analysed (Lo 2005). The technique can be used to produce error budgets and sensitivity indices, partition error between model components and investigate the precision of the model (Lo 2005). Conditions resulting in minimum error and optimal experimental design can also be calculated (Lo 2005).

GEP is a particularly useful technique in studies that involve scaling of responses to larger or smaller scales, but is constrained to models where output variables are able to be expressed as a mathematical function of input variables and parameters (Lo 2005).

No specific software packages have been identified to conduct GEP analyses.

3.10.2. Past & current applications

There are few examples of GEP in the non-ecological scientific literature, and only one ecological example was identified. GEP was applied to a data-set describing the amount of woody debris downed during an ice-storm in Quebec, Canada (Lo 2005). It has also been used to calculate standard deviations for modelled fluxes from a thermodynamic soil-vegetation model and to identify limitations due to stochastic errors and parameterisation weaknesses (Molders 2005). During that analysis, GEP was also used to identify parameters where higher measurement accuracy would have the greatest impact on error budgets (Molders 2005). GEP has also been used in environmental physics to compare two models for ocean bottom pressure in a study to estimate the circulation and climate of the Antarctic ocean (Zlotnicki *et al.* 2007).

3.10.3. Potential for CLLAMM Futures

GEP is unlikely to be of use for CLLAMM Futures as it is limited to models where mathematical equations can be constructed for all output variables. CLLAMM Futures modelling is likely to involve several analysis techniques linked together, so it is unlikely that variables will be able to be represented by single equations in the manner required for GEP.

3.11. Individual based modelling

3.11.1. Overview of theory

The premise behind individual-based modelling (IBM) is that individuals form the basic units of ecological systems. Therefore, the characteristics and behaviour of individuals can be used to define the properties of the system that they comprise (Grimm and Railsback 2005). This is consistent with an evolutionary approach to the response of communities to changing environmental conditions, given that adaptation occurs at the level of the individual (Grimm and Railsback 2005; Reuter *et al.* 2005). As such, IBMs are a hierarchical representation of an ecosystem that can provide a broadly applicable framework for prediction (Reuter *et al.* 2005). IBMs are also referred to as entity- or agent-based models.

IBMs are defined as models that explicitly consider the growth and development of individuals, the relationship between individuals and their resources, incorporate variation amongst individuals and where populations are represented as integers (Grimm and Railsback 2005). IBMs are built on the principles of game theory, foraging theory and ecological theory (West and Caldow 2006). They use decision rules to model the behaviour of individuals, such that the model is likely to respond to environmental change in a manner analogous to that in a real ecosystem (West and Caldow 2006). IBMs to illustrate spatial behaviours require information such as individual space use, available resources, and survival, dispersal and breeding characteristics (Hodder *et al.* 2007). By linking such interacting causal factors, the driving forces behind higher level dynamics become apparent (Reuter *et al.* 2005) and can be applied to new situations in a predictive fashion that is often not possible with empirical models where prediction is limited to the conditions under which the model was developed (West and Caldow 2006). IBMs are also able to integrate information from a variety of spatial and temporal scales and to assist in the development and testing of hypotheses (Reuter *et al.* 2005).

The disadvantages of IBMs include that they can be quite complex in comparison to other types of models which leads to difficulties in the understanding, testability and generality of the resultant models, as well as requiring significant amounts of data and computation time (Grimm and Railsback 2005). It can also be difficult to effectively communicate the model given that the mechanisms of an IBM often cannot be adequately described by mathematical equations (Grimm and Railsback 2005). Successful development of IBMs also often requires higher levels of computer programming expertise than is common in ecologists (Grimm and Railsback 2005).

IBMs are typically coded programs of their own, so dedicated software is not applicable, but such models could be developed in a modelling environment such as MATLAB or R.

3.11.2. Past & current applications

IBMs have been applied to a number of situations including modelling food web interactions, population and dispersal dynamics and nutrient allocation (Reuter *et al.* 2005).

The technique was used to analyse fish tracking data to gain an understanding of the spatial behaviour of pike in the River Frome, UK (Hodder *et al.* 2007). This modelling provided information about the home ranges of the fish, the use of floodplains during high flows and provided insights into the habitat preferences, demography and territoriality of the fish (Hodder *et al.* 2007).

A series of models, including IBMs, were developed during the Ecosystem Research Project Bornhoved Lakes District in Germany (Reuter *et al.* 2005). The IBM was applied to provide information on a variety of processes such as space utilisation, dispersal, food web interactions and the allocation of nutrients (Reuter *et al.* 2005).

Bird migration patterns and use of individual mudflats have also been modelled using IBMs. A fully parameterised and field-tested model described the behaviour of an Oystercatcher

population that overwintered in Devon, UK (Stillman *et al.* 2000). This model took 20 years to develop, but was able to accurately predict changes in mortality due to density-dependent factors (Stillman *et al.* 2000). That model has also been adapted to include prey populations of shellfish, but the amount of data available was insufficient for accurate predictions to be made (West and Caldow 2006). Another IBM also focused on bird migration, with the effect of plant propagule dispersal by ducks linked to a genetics model to calculate changes in gene frequency of the plants (Lurz *et al.* 2002).

3.11.3. Potential for CLLAMM Futures

IBM is unlikely to be of use in the CLLAMM Futures project due to the nature of the data required. Much of the CLLAMM Futures modelling will be undertaken with existing data and data collected by other projects, which does not meet the stringent requirements of IBM. There is limited scope to include the types of information required in ongoing fieldwork due to the set objectives of the other projects, and the Key Species project in particular. The level of detail required for each key species in the region is unlikely to be obtainable in the scope of this project and the complexity of the ecosystem is beyond what is likely to be able to be modelled effectively.

To use such an approach, it would be preferable to plan data collection around the model requirements, which should occur at the project planning stage, and is not possible at this stage of CLLAMMecology. However, the possibility of using the technique should be considered again in the planning stages of any 'Son of CLLAMMecology' project, should one arise. The technique may also be useful for some aspects of Key Species Response modelling, where sufficient data are available for particular species of interest.

3.12. Levels of evidence

3.12.1. Overview of theory

In many instances in ecology, conditions will preclude the rigorous experimental testing of the effect of a purported impact. In these instances, monitoring and experimentation may still be possible, but the inferences drawn from these data will not be definitive (Downes *et al.* 2002). Drawing together several sets of data, in a 'levels of evidence' analysis can enable additional confidence in the conclusions of the study than would otherwise be possible (Downes *et al.* 2002).

The 'levels of evidence' approach is based on the approach that has been used for many years in epidemiological studies, where experimentation is usually impossible and/or unethical. These studies are based strongly on correlations between disease and hypothesised causes. They are based on nine criteria that were proposed by Hill (1965), including the strength and consistency of association, temporality of the effect, biological or ecological plausibility and available experimental evidence, amongst others. Hill (1965) deliberately did not assign importance to the individual criteria, envisaging that different criteria would play varying roles in different studies.

The technique involves identifying criteria against which evidence will be assessed, deciding how effects will be deemed to be 'detected' or 'undetected', deciding on the characteristics of the studies to be included and conducting two reviews of the relevant literature and data; one looking for a given effect and one in instances where the effect has been absent but other characteristics occur in common with the situation of interest. Predictions can then be made and tested explicitly using the information gathered, and monitoring programs and experiments can be designed to fill any identified gaps (Downes *et al.* 2002).

A similar technique is known as 'weight of evidence'. This technique has a focus on data to confirm a hypothesis, rather than falsifying alternative hypotheses (Downes *et al.* 2002), so is somewhat less rigorous.

A dedicated software package is not necessary for this analysis.

3.12.2. Past & current applications

No ecological literature was found that referred to a 'levels of evidence' approach, although Bennett and Adams (2004) recommended the approach in their review of approaches to assess ecological effects of forest harvesting.

There were, however, a number of studies that used a 'weight of inference' approach. Guidelines for using such an approach in marine systems where multiple stressors exist have been proposed (Adams 2005), with many of the same limiting considerations as raised by Downes *et al.* (2002). 'Weight of inference' has been applied to an autecology matrix summarising individual fish species with the water and habitat qualities associated with each and with their life history characteristics for northern Taiwan (Suen and Herricks 2006) and to determine robust salinity tolerances for stream macroinvertebrates by combining two common types of tolerance information from Queensland, Australia (Horrigan *et al.* 2007).

Another application of the 'weight of evidence' method was in combining the data from a long-term monitoring and experimental study of brown trout in Switzerland (Burkhardt-Holm and Scheurer 2007). From this analysis, the decline in catch and health of the species was attributed to parasite-induced proliferative kidney disease, rather than another postulated cause or a combination of causes (Burkhardt-Holm and Scheurer 2007).

3.12.3. Potential for CLLAMM Futures

A 'level of evidence' approach is not likely to be of significant potential for CLLAMM Futures. It may, however, be of substantially more value within the Key Species project, particularly during the design phase of experiments. This would allow each sub-project to objectively assess the evidence for impact for their taxa, and to explicitly design experiments to fill these gaps. As CLLAMM Futures is not collecting substantial new field data, 'levels of evidence' is less likely to be useful, although it may have a role to play for knowledge gaps that are not filled by the Key Species project where the scientific literature is the best source of information to include in modelling. Another potential application may be to combine the results of various monitoring and experimental data arising during CLLAMMecology. Again, however, this may be best undertaken by Key Species project members.

3.13. Summary of techniques

The following table summarises the techniques explored in this review, listing their advantages, disadvantages and their applicability for CLLAMM Futures.

Table 1. Summary of the relative advantages, disadvantages and applicability of the techniques investigated

	Advantages	Disadvantages	Applicability to Futures
Hysteresis modelling	<ul style="list-style-type: none"> Leads to greater understanding of state change and system predictability 	<ul style="list-style-type: none"> Describes an outcome, rather than a particular technique 	<ul style="list-style-type: none"> Appropriate data to model hysteresis is unlikely to be available
Alternative stable states	<ul style="list-style-type: none"> Leads to greater predictability of future states, as systems with alternative stable states show memory of previous states 	<ul style="list-style-type: none"> Difficult to adequately demonstrate true alternative stable states 	<ul style="list-style-type: none"> The alternative state concept is useful, although it is unlikely that stability will be demonstrated
Bifurcation plots from chaos theory	<ul style="list-style-type: none"> Able to describe the differences in parameter values across a threshold value Understanding one bifurcation can assist in predicting others so can make systems more predictable 	<ul style="list-style-type: none"> Requires data on the behaviour of systems both in the decline and recovery phase Largely limited to use in deterministic models 	<ul style="list-style-type: none"> Unlikely to be useful given Futures' focus on statistical, rather than deterministic modelling
State and transition modelling	<ul style="list-style-type: none"> Able to combine data from different sources to describe state and/or transitions Do not assume equilibrium in a system Can be linked to GIS to become spatially explicit 	<ul style="list-style-type: none"> Have limited ability to account for gradual change Are not able to incorporate evolutionary change 	<ul style="list-style-type: none"> Could be used as the structure for the various hypothesised ecosystem states, combining output from other techniques to define the various states and transitions
Classification and regression trees	<ul style="list-style-type: none"> Is able to use a combination of categorical and continuous variables Is well suited to ecological data due to a lack of assumptions about distributions Has easy-to-understand graphical output 	<ul style="list-style-type: none"> A lack of consensus in algorithms mean that different programs can result in different output trees Tends to produce overly-complex trees unless pruned appropriately 	<ul style="list-style-type: none"> Can be used to parameterise the hypothesised alternative ecosystem states Able to identify threshold points between states
Multiple regression trees	<ul style="list-style-type: none"> Extends CART analyses to situations that have multiple response variables Able to assess high-order, complex interactions between predictor variables Robust to patchy ecological data 	<ul style="list-style-type: none"> As per CART, is prone to over-fitting Can be sensitive to outliers and be biased towards variables with more potential split points. 	<ul style="list-style-type: none"> Is likely to be of use in predicting species composition under tested scenarios May be able to validate hypothesised ecosystem states

	Advantages	Disadvantages	Applicability to Futures
Boosted regression trees	<ul style="list-style-type: none"> • Greater predictive accuracy than CART because a combination of CART models are fitted to progressively re-weighted data • Similar advantages to other tree bases analyses, like the ability to incorporate both categorical and continuous variables. 	<ul style="list-style-type: none"> • The output is complex, with up to hundreds of simple trees combined in one analysis with individual trees unable to be examined individually • Prone to over-fitting unless the output is carefully pruned 	<ul style="list-style-type: none"> • Unlikely to be useful given the 'black box' nature of the output that will preclude the results from being incorporated into other analyses.
Multivariate adaptive regression splines	<ul style="list-style-type: none"> • Able to detect global and linear structures within datasets • Combination of regression equations leads to greater predictive power than traditional regression equations. • Found to be effective in combination with other regression analysis techniques like GAM and GLM. 	<ul style="list-style-type: none"> • Lack of agreement about utility in ecological situations. The ability to incorporate output into other analyses is unclear and may vary between analytical packages. • Has the potential to be over-influenced by local processes, making findings unstable 	<ul style="list-style-type: none"> • The technique may be useful, but verification with CLLAMM Futures data is needed to confirm, given varying findings in the scientific literature.
Structural equation modelling	<ul style="list-style-type: none"> • Tests hypothesised relationships between variables • Able to propagate error and uncertainty through the model • Able to incorporate theoretical constructs, as well as measured variables 	<ul style="list-style-type: none"> • Can be difficult to include spatial and temporal scales into the model • Can be difficult to assess causality from co-variations and correlations 	<ul style="list-style-type: none"> • Could be used to quantify the conceptual models used by managers of the CLLAMM region
Bayesian belief networks	<ul style="list-style-type: none"> • Represents a system as a series of interactions • Able to combine a variety of data and analysis types • New information can be incorporated as it becomes available 	<ul style="list-style-type: none"> • Require a full probability structure to be specified • Poor at incorporating spatial or temporal variability • Continuous variables are not treated neatly • Historically, tend to rely overly on expert opinion 	<ul style="list-style-type: none"> • Could be used as the platform to tie various other analyses together
Gaussian error propagation	<ul style="list-style-type: none"> • Enables detailed study of error sources within the model • Computationally less expensive than techniques like Monte Carlo simulations 	<ul style="list-style-type: none"> • Output variables must be expressed as functions of input variables 	<ul style="list-style-type: none"> • Models will not be able to be expressed as a single mathematical function

	Advantages	Disadvantages	Applicability to Futures
Individual based modelling	<ul style="list-style-type: none"> • Scale consistent with evolutionary adaptation • Findings are likely to be predictive because they are based on theory and are not empirical 	<ul style="list-style-type: none"> • Highly data intensive • Involves high level programming skills • Often involves studies over decades • Can be difficult to communicate findings 	<ul style="list-style-type: none"> • Appropriate data will not be available
Levels of evidence	<ul style="list-style-type: none"> • Combines multiple studies to strengthen the evidence for or against a hypothesis • Rigorous method of hypothesis falsification • Identifies areas for further research 	<ul style="list-style-type: none"> • Time consuming technique 	<ul style="list-style-type: none"> • Unlikely, but may be useful in identifying knowledge gaps for possible CLLAMM II

4. Application to CLLAMM Futures

This section gives some examples of the modelling techniques that were identified above as being of particular interest for use in CLLAMM Futures. A brief description of the application of the technique is given. Where possible, data from the CLLAMM region have been used to create these examples, otherwise data pertaining to egret colonies in NSW (Baxter and Fairweather 1994; 1998) were utilised instead.

Examples are given here for CART, SEM and BBN.

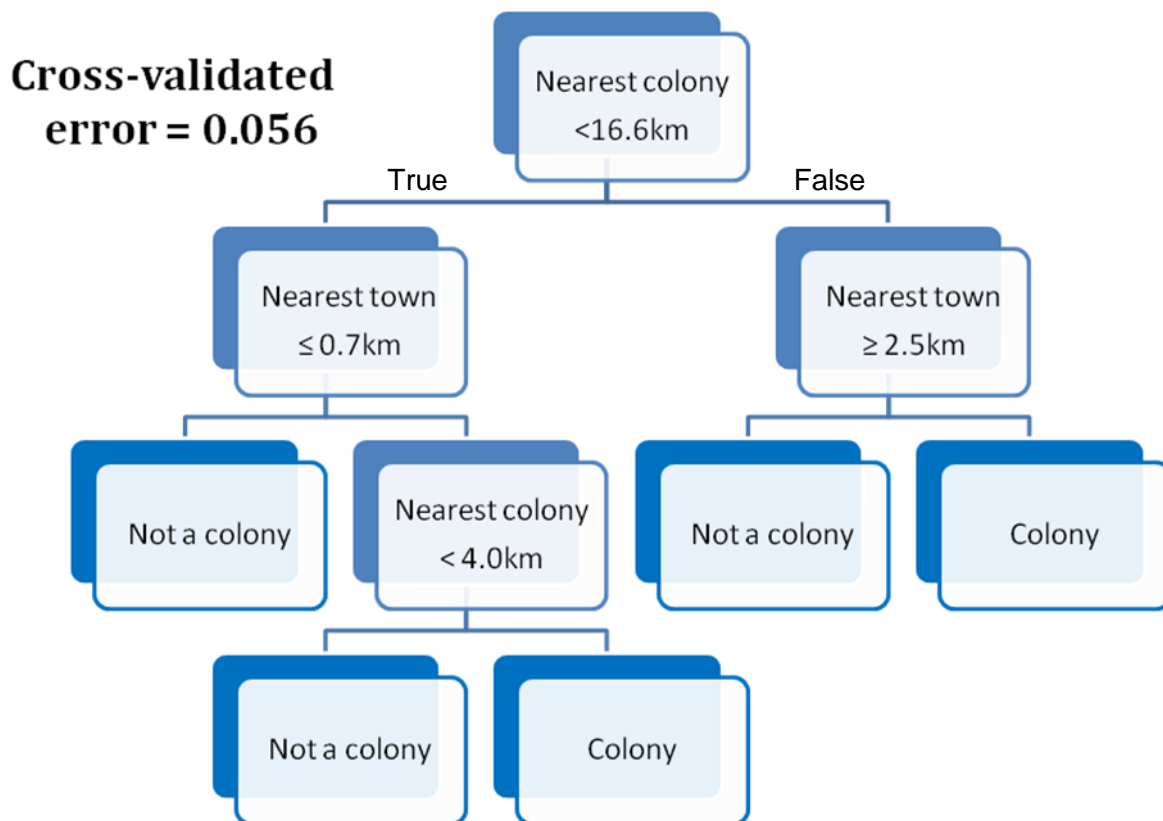
4.1. Classification and regression trees (CART)

A dataset describing the characteristics of a number of egret colonies in coastal New South Wales is used to give an example of the application of classification and regression tree analyses. These data, and other analyses on them are described in Baxter and Fairweather (1994; 1998). The study involved surveying 65 wetlands, of which only 13 were the location of multi-species egret colonies. The characteristics of each wetland were recorded, and CART was applied to these data to identify those features that were characteristic of wetlands containing a colony compared to those that did not.

Features included in the analysis were the distance to the nearest egret colony, the distance to the nearest town, the nearest large town, and the nearest coastline. The land surrounding the wetland (within a radius of 20km, that is, the distance egrets are known to travel during foraging trips) was also characterised, with variables describing the amount of surrounding freshwater wetlands, the surrounding grazed land, and the amount of mangrove, saltmarsh and seagrass habitats. Wetland characteristics such as area, elevation and the dominant tree type were also included. The wetland's status as a colony or not a colony was used as the dependent variable.

As the condition of supporting a colony or not was a categorical variable, a classification tree was applied to the dataset. The resulting classification tree (Figure 1) is read as a decision tree. Where a condition is true, the reader should follow the left hand side of the next branch. This continues until a terminal node (here either 'Colony' or 'Not a colony') is reached.

Figure 1. Example egret colony classification tree



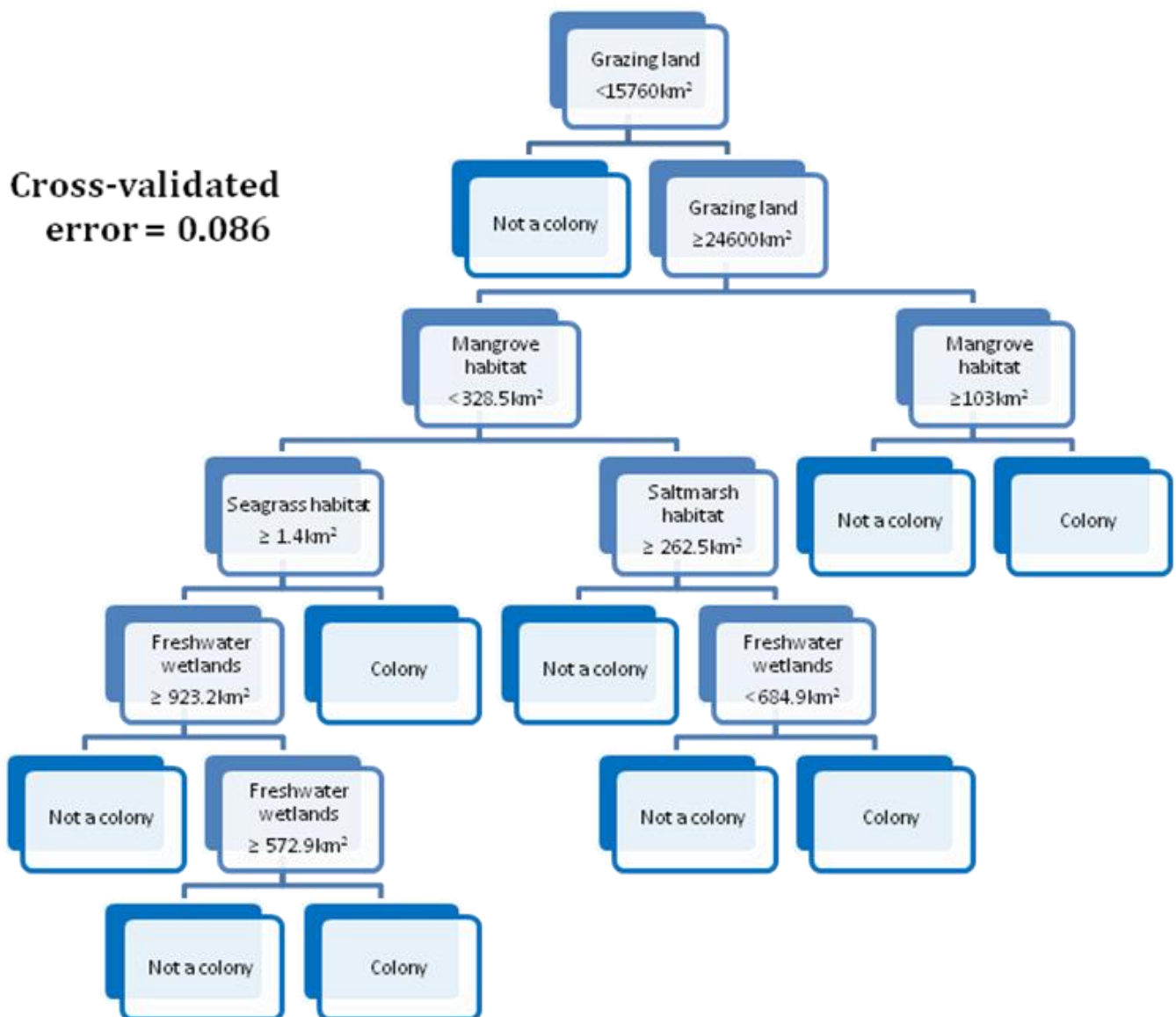
The distance to the nearest colony and the distance to the nearest town were identified as the most useful variables in classifying a wetland as supporting an egret colony or not. Egret colonies tended to not be immediately adjacent to one another and not adjacent to nearby towns. Cut-off values were estimated from the dataset. Where colonies were more than 16.6km apart, only wetlands that were more than 2.5km from a town were predicted to support a colony. Where colonies were less than 16.6km apart, only wetlands that were more than 0.7km from the nearest town and more than 4km from the nearest colony were likely to support a colony.

A measure of the error associated with the resultant classification is the cross-validated error. This is an estimate of the error that would be expected if this tree were used to classify new data, collected in the same manner as the initial dataset (e.g. additional wetlands in the same region). The cross-validated error for this tree is quite low, with only 5.6% of cases expected to be misclassified.

A second classification tree was developed describing the likelihood of a wetland supporting an egret colony based on the surrounding habitat characteristics (Figure 2). This demonstrates that a range of variables can be used in CART to describe a single dependent variable, and that the tree developed can be tailored to meet specific needs or data sources. In this instance, the second model would be relevant for wetlands where, for example, the nearest egret colony was unknown.

Regression tree analyses are very similar to that for classification trees, but are performed where the dependent variable (here 'Egret colony') is a continuous variable. For example, regression analyses would be more appropriate when investigating factors affecting the size of an egret colony (continuous variable).

Figure 2. Alternative example of an egret colony classification tree using surrounding habitat characteristics



4.2. Structural equation modelling (SEM)

SEM was applied to a dataset describing the performance of *Ruppia tuberosa* in the southern lagoon of the Coorong. The data were collected by David Paton between 1990 and 1993 at three sites (Pelican Point, Villa dei Yumpa and Ti Tree Crossing) with all surveys conducted in July. At each site, five replicate samples were taken at each of five set water depths. Each sample was assessed for the number of shoots, turions (of both Type I and Type II) and seeds. Replicate samples taken at each depth were combined to give a sample size of 44. The length of shoots was also measured and the proportion that were live was calculated.

In this analysis, relationships were explored between the various measures of *R. tuberosa* performance. It was found that variables describing shoot length, the number of shoots and the number of Type II turions interacted to give an acceptable measure of *R. tuberosa* performance (labelled '*Ruppia*' in Figure 3, and referred to as a 'latent' (i.e. composite) variable below). The arrows linking '*Ruppia*' to each of the individual performance measures indicate the standardised correlation of each measured variable (e.g. shoot length) with the latent '*Ruppia*' variable. Associated with each is also an error term which should be interpreted as a combination of measurement error and unexplained variance in the relationship between that variable and '*Ruppia*'. An r^2 value for each relationship is also given, signifying the proportion of variance explained by that section of the model. Of the three values, the number of shoots is the best estimator for the value of the latent '*Ruppia*' variable, with the number of Type II turions also correlating well, and some correlation with the length of the shoots also recorded. The error terms for the length of shoots has the highest value, suggesting that there is unexplained variance for that variable. This may indicate that the length of the shoots is responding to slightly different environmental factors than the number of shoots and the number of Type II turions. For example, the length of time the area had been inundated may be an important driver of shoot length which could be uncoupled from the actual water depth at the time of sampling and therefore is unrepresented in this version of the model.

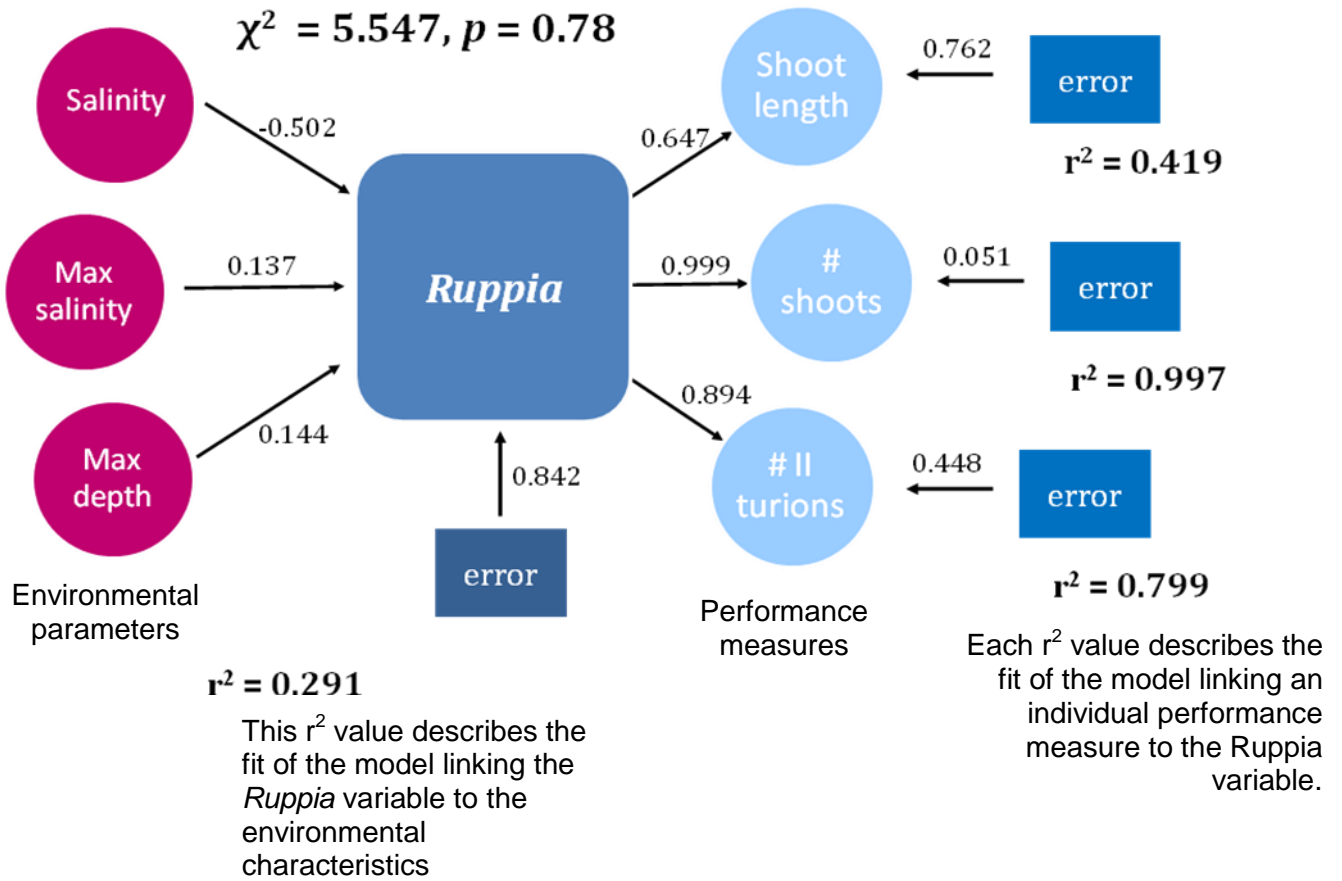
The number of Type I turions, the number of seeds and the proportion of live shoots were removed from the model as they did not correlate well with this latent '*Ruppia*' variable, again suggesting that they are responding to a disparate set of environmental factors, or that they respond very differently to the same environmental factors. For example, it is likely that the environmental factors needed for successful germination are different from those needed to produce seeds, so it is not surprising that the number of shoots and the number of seeds respond differently, particularly when it is considered that the seeds would have been produced the previous year. Removal of variables is based on the χ^2 value for the model and the r^2 values for each performance variable.

The latent '*Ruppia*' variable was then related to the various environmental parameters that have been measured. Six parameters were included as possibilities in this example analysis. Measured salinity (at the time of sampling), maximum recorded salinity (over the year of sampling), measured depth (at the location of sampling), maximum depth (across the site at the time of sampling) and minimum depth (also across the site at the time of sampling) and turbidity (measured as secchi depth) were explored. Three of these variables were able to satisfactorily explain some of the variance in the latent '*Ruppia*' variable. Again, standardised correlation coefficients, an associated error term and an r^2 value are given. A negative term (in the case of 'Salinity') simply indicates a negative relationship between the variable in question and the latent '*Ruppia*' variable (i.e. a negative correlation).

The overall fit of the model is measured using χ^2 and an associated p -value. Basically, the p value describes the likelihood that the model structure describes the data more effectively than would occur by chance. This interpretation of χ^2 and p are quite different from that usually applied to statistics. In this instance, the model presented here is 78% more likely to describe the data than would be expected to occur by chance.

It should be emphasised that this example is a preliminary model to describe this dataset, produced specifically as an example of SEM and that further modelling will be undertaken to include additional predictor variables (e.g. flow over the barrages) and to describe other aspects of the *R. tuberosa* population (e.g. propagule production).

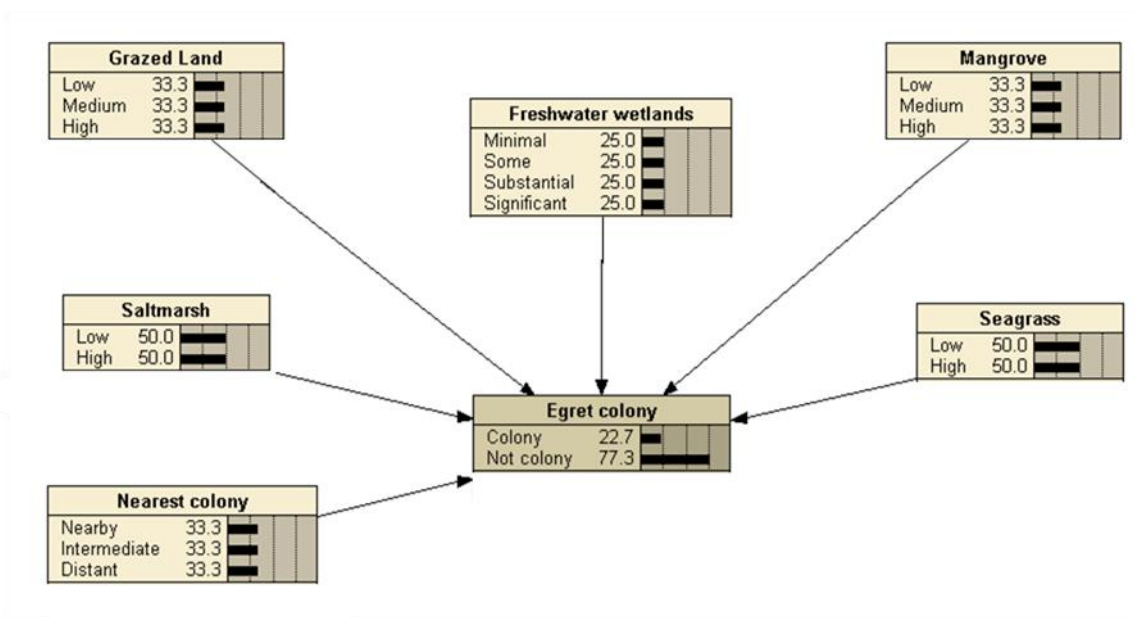
Figure 3. Example *Ruppia tuberosa* structural equation model



4.3. Bayesian belief networks (BBN)

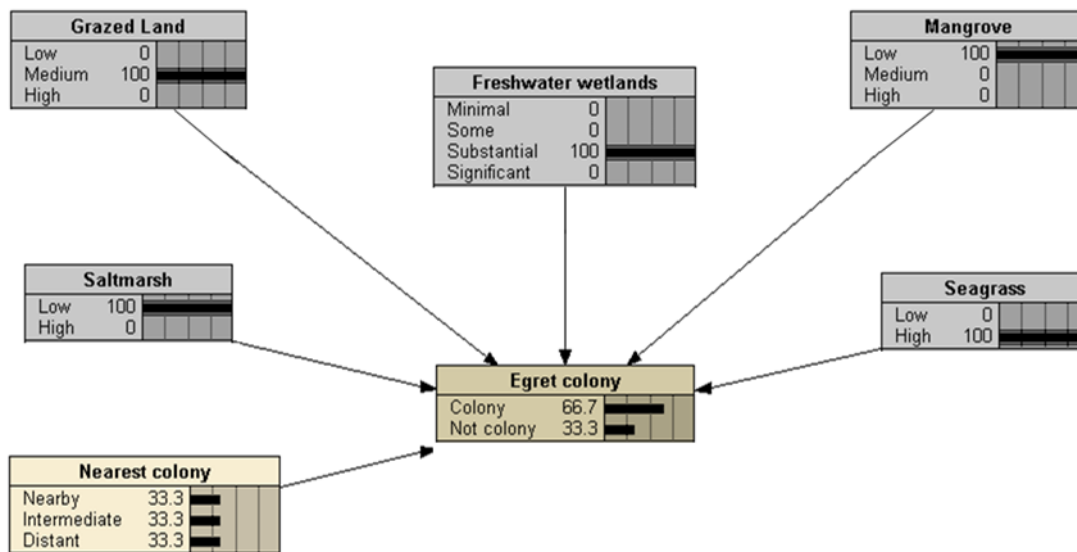
One of the strengths of BBN is its ability to incorporate information from a range of sources and other models. Here, a BBN is constructed for the same egret colony dataset used to demonstrate the use of CART. The classification tree shown in Figure 2, classifying wetlands as egret colonies or not based on their surrounding habitat, has been incorporated into the BBN. Also included is the knowledge that egret colonies are unlikely to be in adjacent wetlands, which is not demonstrated in that particular analysis, but is implied in Figure 1. The resultant BBN is shown in Figure 4.

Figure 4. Example egret colony Bayesian belief network based on local habitat condition (uninformative priors)



This BBN includes all of the information displayed in Figure 2, including the threshold values determining the likelihood that a particular wetland does or does not support an egret colony. The initial BBN has no information on the habitat conditions of an individual wetland (that is, it has uninformative prior probabilities). This means that the model assigns equal likelihoods to the various states for each habitat characteristic. For example, there are three possible categories for 'Grazed Land' – low, medium and high, so the model has an initial likelihood of 33% for the wetland falling into any one of those categories, given that no additional information is available. Information regarding individual cases (i.e. new wetlands) can then be entered into the BBN, and the 'Egret colony' node will describe the likelihood of an egret colony for that particular case. As is demonstrated in Figure 5, the calculations can be done with incomplete information. That is, in Figure 5, information has been entered to describe the wetland in terms of the surrounding habitat (grazed land, freshwater wetlands, mangroves, and saltmarsh or seagrass habitats) but not regarding the distance to the nearest colony. That is left with an equal probability of being nearby, intermediate or distant.

Figure 5. Application of egret colony Bayesian belief network with incomplete information regarding surrounding habitat

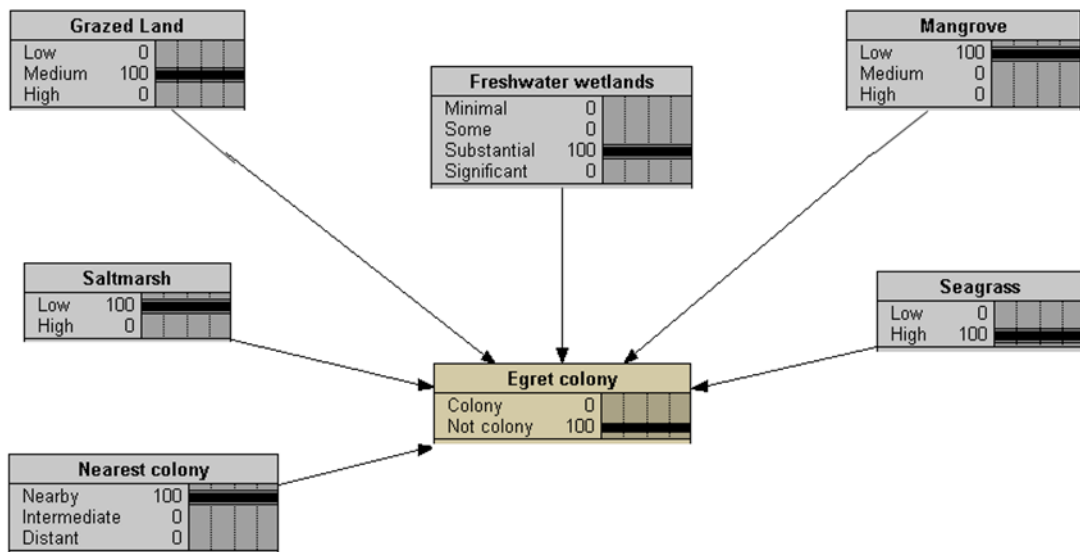


In Figure 5, grey boxes represent those for which the condition is known. That is, information regarding the amount of grazing land, the area of freshwater wetlands, mangrove habitat, saltmarsh habitat and seagrass habitat surrounding the new wetland have been entered. Because the status of these variables is known, there is a 100% chance of the wetland being in the nominated state, as shown in the grey boxes. The status of the wetland regarding other nearby colonies has not been entered (and is presumably unknown), so the default condition of equal probabilities remains. Based on the probability distributions entered for known egret colonies, the model predicts that there is a 67% likelihood that the wetland in question will support a colony, as is shown by the 'Egret colony' node.

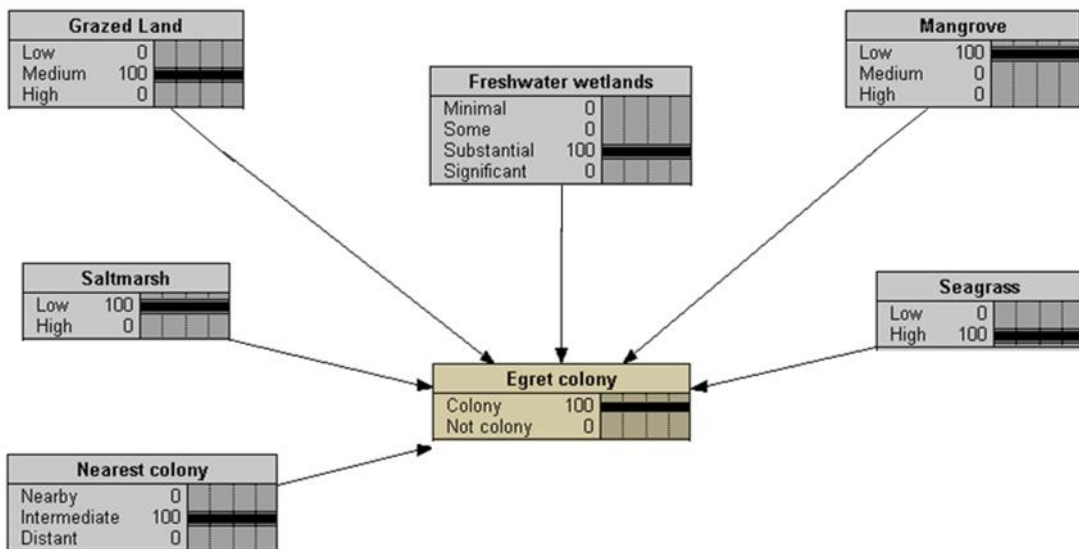
Additional information can be added to the BBN at any time, and the likelihoods are updated to take this new information into account. For example Figure 6 gives the outcome when the distance to nearby egret colonies is entered as nearby (Figure 6a) or intermediate (Figure 6b). Should the nearest colony have been distant, the wetland of interest would also have been classified as a colony.

Figure 6. Addition of new information to egret colony Bayesian belief network

a) indicating that the wetland has a nearby egret colony



b) indicating that distance to the nearest egret colony is intermediate

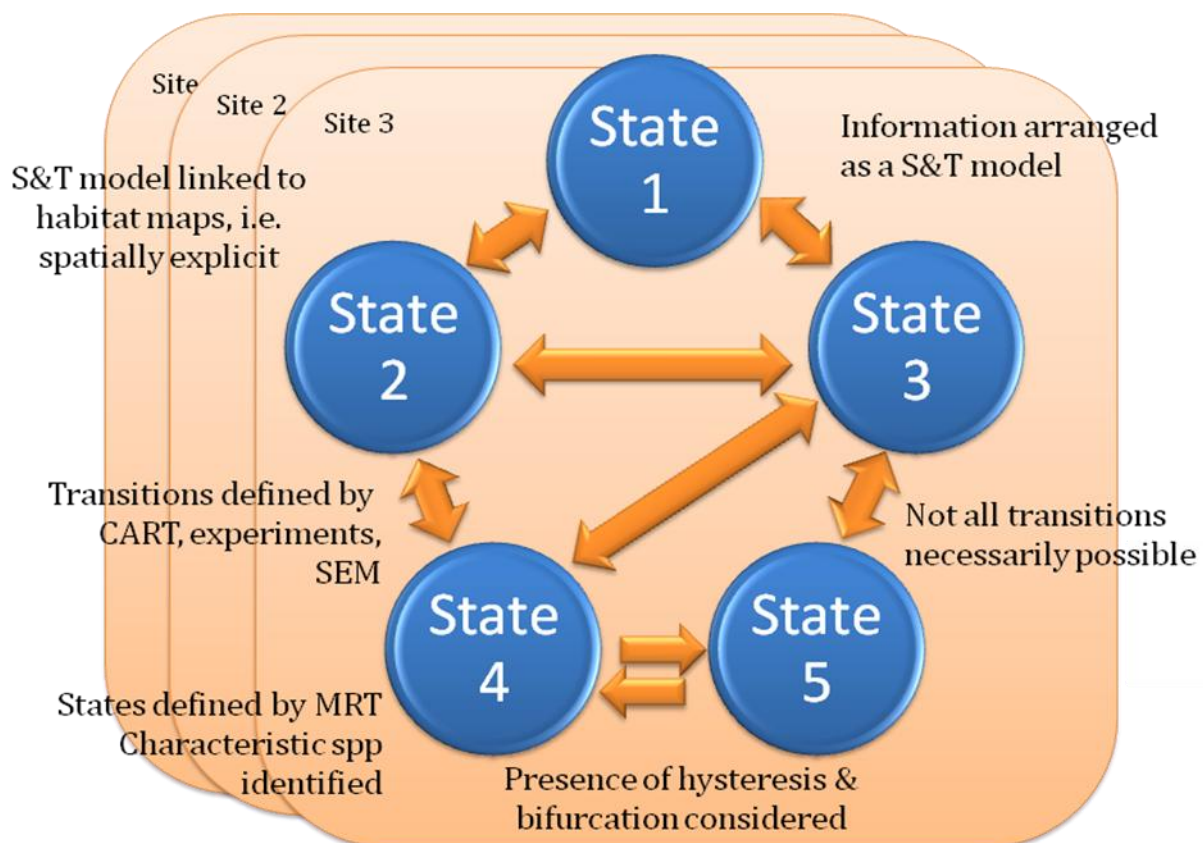


Additional nodes can be incorporated into a BBN as knowledge on the topic increases. That is, the results of additional analyses can be progressively included in a BBN to give increasingly accurate predictions of the state of one or more dependent variables. BBNs offer a unique opportunity to include information from the literature, expert opinion and from a wide range of univariate and multivariate analysis techniques.

5. Summary of useful techniques

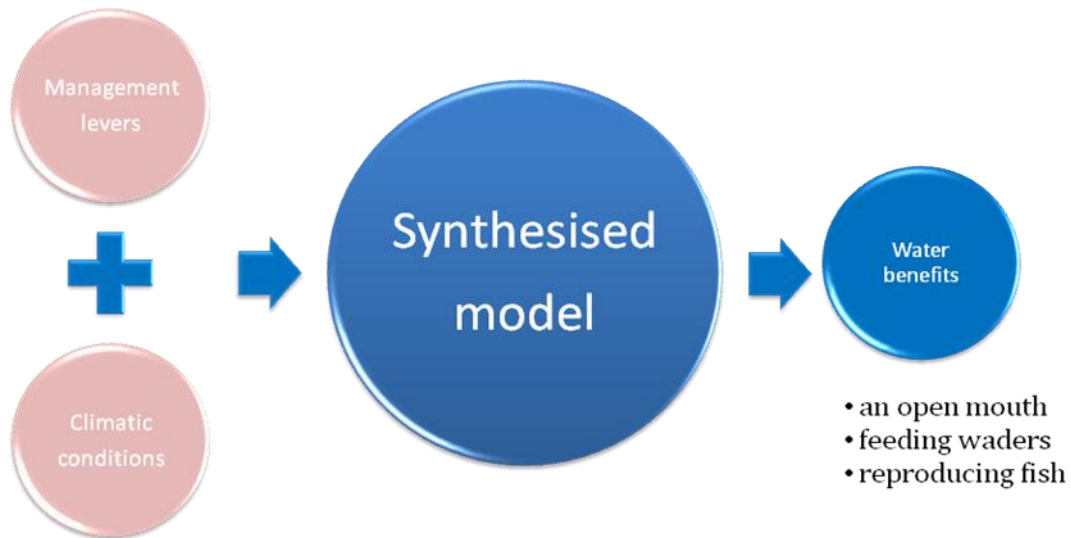
A number of analytical techniques appear promising for the CLLAMM Futures project. CART, MRT and SEM seem to be good initial choices, given their ability to classify and rank independent variables. This will allow the various states postulated for the CLLAMM region to be objectively defined and tested, and also for the parameterisation of the existing conceptual models currently used by managers. The combined information from these analyses could then be used to construct an S&T model describing the ecosystems of the CLLAMM region. It should be possible to make this model explicit in time and space by connecting it to the habitat maps produced within the Dynamic Habitat project. BBNs represent one possibility for constructing the S&T model, given that it will include information from many different analyses, of varying degrees of reliability, including data from experiments as well as from condition monitoring, along with information from the literature and expert opinion where data are unavailable. The range of data sources, along with the extreme patchiness of data coverage in space and time, rule out many potential alternatives and make BBN a good choice of platform. A synthesis of how modelling techniques may be applied within CLLAMM Futures is shown in Figure 7.

Figure 7. Synthesis of modelling techniques identified as potentially useful for CLLAMM Futures



This synthesis of modelling techniques will be used as the basis of scenario modelling. The various management levers and forcing factors (such as climate) will be manipulated to suit a given scenario. These will form the inputs into the synthesised model (Figure 7) and will drive the physical conditions that determine the ecosystem states at each site for each time step. Analyses of the outcomes of the model runs will determine the likelihood of the water benefit objectives being met for that scenario. This process is illustrated in Figure 8.

Figure 8. Scenario modelling using the synthesised model



An important aspect of constructing the S&T model as a composite of the results of various analyses will be to gain an estimate of the error associated with the model. As described above, Gaussian error propagation is unlikely to be able to be applied to the CLLAMM Futures model, but this does not remove the need to understand the errors inherent in the model and the implications of those errors for the predictions of the model. Error in the model can either be due to a lack of sufficient data or a lack of realism in the ability of the model to predict the ecological reality (Barry and Elith 2006). Sensitivity analyses should be able to provide an estimate of the effect of error on the model predictions. Barry and Elith (2006) suggest that an important check on the realism of models is that they are constrained to be consistent with current ecological knowledge. Austin *et al.* (2006) and Austin (2007) also agree that consistency with ecological theory, along with the skills of the analysts, are important in ensuring the model predictions approach the 'truth'. This topic is one that will need to be developed further as model construction proceeds.

6. A note regarding modelling responses of key species

Many approaches to producing response models for the key performance indicators of particular focal fauna rely upon restrictive assumptions regarding the datasets included and/or the form of the relationships that responses by species take. The data available to CLLAMMecology researchers is likely to be heterogeneous, incomplete and patchy in space and time. For these reasons, much of the above discussion about ecosystem-scale modelling will also be applicable to CLLAMMecology autecological modelling as well.

In addition to considering the techniques listed above, we have also examined the assumptions and data requirements of a range of modelling software that may be appropriate when constructing response models for key species. The program Hyperniche has a range of advantages that may mean it is useful for CLLAMMecology.

Hyperniche relates environmental variables to response in a target species, using either presence/absence or abundance data for that target species. It uses non-parametric analyses that do not make assumptions about the distributions of independent variables, and combines variables in a multiplicative manner that allows the inclusion of interactions between variables. The advantage of using a non-parametric, multiplicative approach to regression is that it acknowledges that any given species will be responding to more than one environmental parameter at any given time (McCune 2006). Response to a single parameter will always be affected by the condition of other parameters and the species may vary from being absent in the system to some unique carrying capacity.

The modelling approach used by Hyperniche has advantages over linear and logistic regression, which constrain the distribution of the response and parameter variables to simplistic forms that are unlikely to apply in reality (McCune 2006). Generalised additive models offer improvements over linear and logistic regression, but are still unable to detect interactions between parameter values, and only model those interactions that are specifically incorporated by the modeller (McCune 2006).

Approaches such as that used in Hyperniche do have a number of disadvantages too. They require reasonably sizeable amounts of data to empirically fit a robust model (McCune 2006). The method of multiplicative incorporation of parameters also means that the output is not in the form of a single equation, and the response of a species to a parameter must be interpreted from graphical output (McCune 2006). A final drawback of approaches such as that used in Hyperniche is that it has only been applied to univariate response of a single species, rather than being available for use on communities of key species.

Despite the disadvantages associated with Hyperniche, we believe that it will be a useful tool for the members of the Key Species Response team and will assist in interpreting the response of those key species to the environmental parameters of the CLLAMM region.

References

- Adams SM (2005) Assessing cause and effect of multiple stressors on marine systems. *Marine Pollution Bulletin* **51**, 649-657.
- Allen AP, O'Connor RJ (2000) Hierarchical correlates of bird assemblage structure on northeastern USA lakes. *Environmental Monitoring and Assessment* **62**, 15-37.
- Austin MP (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* **200**, 1-19.
- Austin MP, Belbin L, Meyers JA, Doherty MD, Luoto M (2006) Evaluation of statistical models used for predicting plant species distributions; Role of artificial data and theory. *Ecological Modelling* **199**, 197-216.
- Barry S, Elith J (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology* **43**, 413-423.
- Baxter GS, Fairweather PG (1994) Phosphorus and nitrogen in wetlands with and without egret colonies. *Australian Journal of Ecology* **19**, 409-416.
- Baxter GS, Fairweather PG (1998) Does available foraging area, location or colony character control the size of multispecies egret colonies? *Wildlife Research* **25**, 23-32.
- Bennett LT, Adams MA (2004) Assessment of ecological effects due to forest harvesting: approaches and statistical issues. *Journal of Applied Ecology* **41**, 585-598.
- Bestelmeyer BT, Herrick JE, Brown JR, Trujillo DA, Havstad KM (2004) Land management in the American Southwest: A state-and-transition approach to ecosystem complexity. *Environmental Management* **34**, 38-51.
- Bowes M, J, House WA, Hodgkinson RA, Leach DV (2005) Phosphorus-discharge hysteresis during storm events along a river catchment: the River Swale, UK. *Water Research* **39**, 751-762.
- Breiman L, Freidman JH, Olshen RA, Stone CJ (1984) 'Classification and regression trees.' (Wadsworth: Belmont)
- Breine JJ, Maes J, Quataert P, Van den Bergh E, Ilse S, Gerlinde V, Belpaire C (2007) A fish-based assessment tool for the ecological quality of the brackish Schelde estuary in Flanders (Belgium). *Hydrobiologia* **575**, 141-159.
- Bricklemeyer RS, Lawrence RL, Miller PR, Battogtokh N (2007) Monitoring and verifying agricultural practices related to soil carbon sequestration with satellite imagery. *Agriculture, Ecosystems and Environment* **118**, 201-210.
- Briske DD, Fuhlendorf SD, Smeins FE (2003) Vegetation dynamics on rangelands: a critique of the current paradigms. *Journal of Applied Ecology* **40**, 601-614.
- Burkhardt-Holm P, Scheurer K (2007) Application of the weight-of-evidence approach to assess the decline of brown trout (*Salmo trutta*) in Swiss rivers. *Aquatic Sciences* **69**, 51-70.
- Cappo M, De'ath G, Boyle S, Aumend J, Olbrich R, Hoedt F, Perna C, Brunskill G (2005) Development of a robust classifier of freshwater residence in barramundi (*Lates calcarifer*) life histories using elemental ratios in scales and boosted regression trees. *Marine and Freshwater Research* **56**, 713-723.
- Claudet J, Pelletier D, Jouvenel J-Y, Bachet F, Galzin R (2006) Assessing the effects of marine protected area (MPA) on a reef fish assemblage in a northwestern Mediterranean marine reserve: Identifying community-based indices. *Biological Conservation* **130**, 349-369.
- Courtenay GC, Gladstone W, Schreider M (2005) Assessing the response of estuarine intertidal assemblages to urbanised catchment discharge. *Environmental Monitoring and Assessment* **107**, 375-398.

- De'ath G (2002) Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology* **83**, 1105-1117.
- De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* **88**, 243-251.
- De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178-3192.
- Deconinck, E, Coomans, D, Vander Heyden, Y (2007) Exploration of linear modelling techniques and their combination with multivariate adaptive regression splines to predict gastro-intestinal absorption of drugs. *Journal of Pharmaceutical and Biomedical Analysis*, **43**, 119-130.
- Dennis B, Desharnais RA, Cushing JM, Henson SM, Constantino RF (2001) Estimating chaos and complex dynamics in an insect population. *Ecological Monographs* **71**, 277-303.
- DeVantier LM, De'ath G, Turak E, Done TJ, Fabricius KE (2006) Species richness and community structure of reef-building corals on the nearshore Great Barrier Reef. *Coral Reefs* **25**, 329-340.
- Downes BJ, Barmuta LA, Fairweather PG, Faith DP, Keough MJ, Lake PS, Mapstone BD, Quinn GP (2002) 'Monitoring ecological impacts: concepts and practice in flowing waters.' (Cambridge University Press: Cambridge)
- Fairweather PG (1999a) Determining the 'health' of estuaries: Priorities for ecological research. *Australian Journal of Ecology* **24**, 441-451.
- Fairweather PG (1999b) State of environment indicators of 'river health': exploring the metaphor. *Freshwater Biology* **41**, 211-220.
- Felsot AS, Ruppert JR (2002) Imidacloprid residues in Willapa Bay (Washington State) water and sediment following application for the control of burrowing shrimp. *Journal of Agricultural and Food Chemistry* **50**, 4417-4423.
- Ferrarini, A, Rossi, P, Rossi, O (2005) Ascribing ecological meaning to habitat shape by means of a piecewise regression approach to fractal domains. *Landscape Ecology*, **20**, 799-809.
- Ferguson SH (2006) The influences of environment, mating habitat and predation on evolution of pinniped lactation strategies. *Journal of Mammalian Evolution* **13**, 63-82.
- Focardi S, Aragno P, Montarnaro P, Riga F (2006) Inter-specific competition from fallow deer *Dama dama* reduces habitat quality for the Italian roe deer *Capreolus capreolus italicus*. *Ecography* **29**, 407-417.
- Francoise J-P, Piquet C (2005) Hysteresis dynamics, bursting oscillations and evolution to chaotic regimes. *Acta Biotheoretica* **53**, 381-392.
- Grace JB (2006) 'Structural equation modeling and natural systems.' (Cambridge University Press: Cambridge)
- Grant CD (2006) State-and-transition successional model for bauxite mining rehabilitation in the jarrah forest of Western Australia. *Restoration Ecology* **14**, 28-37.
- Greenman JV, Benton TG (2003) The amplification of environmental noise in population models: causes and consequences. *The American Naturalist* **161**, 225-239.
- Grimm V, Railsback SF (2005) 'Individual-based modeling and ecology.' (Princeton University Press: Princeton and Oxford)
- Gutierrez M, Swartzman G, Bertrand A, Bertrand S (2007) Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983-2003. *Fisheries Oceanography* **16**, 155-168.
- Hansell RIC, Craine IT, Byers R (1997) Predicting change in non-linear systems. *Environmental Monitoring and Assessment* **46**, 175-190.
- Henne, PD, Hu, FS, Cleland, DT (2007) Lake-effect snow as the dominant control of mesic forest distribution in Michigan, USA. *Journal of Ecology*, **95**, 517-529.

- Hill AB (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* **58**, 295-300.
- Hodder KH, Masters JEG, Beaumont WRC, Gozlan RE, Pinder AC, Knight CM, Kenward RE (2007) Techniques for evaluating the spatial behaviour of fish. *Hydrobiologia* **582**, 257-269.
- Holling CS (Ed.) (1978) 'Adaptive environmental assessment and management.' International series on applied systems analysis (John Wiley & Sons: Chichester, New York, Brisbane, Toronto)
- Horrigan N, Dunlop JE, Kefford BJ, Zavahir F (2007) Acute toxicity largely reflects the salinity sensitivity of stream macroinvertebrates derived using field distributions. *Marine and Freshwater Research* **58**, 178-186.
- Hsiao W-C, Shih Y-S (2007) Splitting variable selection for multivariate regression trees. *Statistics & Probability Letters* **77**, 265-271.
- Ibelings BW, Portielje R, Lammens EHRR, Noordhuis R, van den Berg MS, Jooisse W, Meijer ML (2007) Resilience of alternative stable states during the recovery of shallow lakes from eutrophication: Lake Veluwe as a case study. *Ecosystems* **10**, 4-16.
- Jackson RD, Bartolome JW (2002) A state-transition approach to understanding nonequilibrium plant community dynamics in Californian grasslands. *Plant Ecology* **162**, 49-65.
- King RS, Hines AH, Craige FD, Grap S (2005) Regional, watershed and local correlates of blue crab and bivalve abundances in subestuaries of Chesapeake Bay, USA. *Journal of Experimental Marine Biology and Ecology* **319**, 101-116.
- Larsen DR, Speckman PL (2004) Multivariate regression trees for analysis of abundance data. *Biometrics* **60**, 543-549.
- Leathwick JR, Elith J, Francis MP, Hastie T, Taylor P (2006) Variation in dermal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* **321**, 267-281.
- Leathwick, JR, Elith, J, Hastie, T (2006) Comparative performance of generalised additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, **199**, 188-196.
- Lo E (2005) Gaussian error propagation applied to ecological data: Post-ice-storm-downed woody biomass. *Ecological Monographs* **75**, 451-466.
- Lotka AJ (1956) 'Elements of Mathematical Biology' (Dover Publications: New York)
- Lurz PWW, Shirley MDF, Rushton SP, Sanderson RA (2002) Modelling the consequences of duck migration patterns on the genetic diversity of aquatic organisms: a first step towards a predictive tool for wetland management. *Acta Oecologica* **23**, 205-212.
- MacLeod CD, Weir CR, Pierpoint C, Harland EJ (2007) The habitat preferences of marine mammals west of Scotland (UK). *Journal of the Marine Biological Association of the United Kingdom* **87**, 157-164.
- McCune B (2006) Non-parametric habitat models with automatic interactions. *Journal of Vegetation Science* **17**, 819-830.
- Malaeb ZA, Summers JK, Pugesek BH (2000) Using structural equation modeling to investigate relationships among ecological variables. *Environmental and Ecological Statistics* **7**, 93-111.
- Mann DA, O'Shea TJ, Nowacek DP (2006) Nonlinear dynamics in manatee vocalisations. *Marine Mammal Science* **22**, 548-555.
- Matin de Santa Olalla FJ, Dominguez A, Artigao A, Fabeiro C, Ortega JF (2005) Integrated water resources management of the hydrogeological unit "Eastern Manch" using Bayesian belief networks. *Agricultural Water Management* **77**, 21-36.
- McCann RK, Marcot BG, Ellis R (2006) Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research* **36**, 3053-3062.

- McNay RS, Marcot BG, Brumovsky V, Ellis R (2006) A Bayesian approach to evaluating habitat for woodland caribou in north-central British Columbia. *Canadian Journal of Forest Research* **36**, 3117-3133.
- Mercado-Silva N, Olden JD, Maxted JT, Hrabik TR, van der Zanden MJ (2006) Forecasting the spread of invasive rainbow smelt in the Laurentian Great Lakes region of North America. *Conservation Biology* **20**, 1740-1749.
- Molders N (2005) Plant- and soil-parameter-caused uncertainty of predicted surface fluxes. *Monthly Weather Review* **133**, 3498.
- Morozov A, Petrovskii S, Li B-L (2004) Bifurcations and chaos in a predator-prey system with the Allee effect. *Proceedings of the Royal Society of London* **271**, 1407-1414.
- Norlin JI, Bayley SE, Ross LC (2006) Zooplankton composition and ecology in western boreal shallow-water wetlands. *Hydrobiologia* **560**, 197-215.
- Peel AB, Deutsch PA, Peacock-Lopez E (2005) Complex dynamics in a three-level trophic system with intraspecies interaction. *Journal of Theoretical Biology* **232**, 491-503.
- Perry GLW, Enright NJ (2002) Spatial modelling of landscape composition and pattern in a maquis-forest complex, Mont Do, New Caledonia. *Ecological Modelling* **152**, 279-302.
- Petraitis PS, Dudgeon S, R. (2004) Detection of alternative stable states in marine communities. *Journal of Experimental Marine Biology and Ecology* **300**, 343-371.
- Petraitis PS, Dudgeon S, R. (2005) Divergent succession and implications for alternative states on rocky intertidal shores. *Journal of Experimental Marine Biology and Ecology* **326**, 14-26.
- Petraitis PS, Dunham AE, Niewiarowski PH (1996) Inferring multiple causality: the limitations of path analysis. *Functional Ecology* **10**, 421-431.
- Petraitis PS, Methratta ET, Rhile EC, Vidargas NA (unpublished data) The breakdown of resilience and confirmation of alternative community states in a marine ecosystem.
- Plant RE, Vayssieres MP (2000) Combining expert system and GIS technology to implement a state-transition model of oak woodlands. *Computers and Electronics in Agriculture* **27**, 71-93.
- Prasad, AM, Iverson, LR, Liaw, A (2006) Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181-199.
- Quetier F, Thebault A, Lavorel S (2007) Plant traits in a state and transition framework as markers of ecosystem response to land-use change. *Ecological Monographs* **77**, 33-52.
- Reckhow KH, Arhonditsis GB, *et al.* (2005) A predictive approach to nutrient criteria. *Environmental Science and Technology* **39**, 2913-2919.
- Regan HM, Ben-Haim Y, Langford B, Wilson WG, Lundberg P, Andelman SJ, Burgman MA (2005) Robust decision-making under severe uncertainty for conservation management. *Ecological Applications* **15**, 1471-1477.
- Reuter H, Holder F, Middelhoff U, Jopp F, Eschenbach C, Breckling B (2005) The concepts of emergent and collective properties in individual-based models - Summary and outlook of the Bornhoved case studies. *Ecological Modelling* **186**, 489-501.
- Scheffer M (1998) 'Ecology of Shallow Lakes.' (Chapman & Hall: London)
- Seabrook L, McAlpine C, Fensham R (2007) Spatial and temporal analysis of vegetation change in agricultural landscapes: A case study of two brigalow (*Acacia harpophylla*) landscapes in Queensland, Australia. *Agriculture, Ecosystems and Environment* **120**, 211-228.
- Stillman RA, Goss-Custard JD, West AD, Durrell SEA, Le Vd, Caldow RW, McGroarty S, Clarke RT (2000) Predicting mortality in novel environments: tests and sensitivity of a behaviour-based model. *Journal of Applied Ecology* **37**, 564-588.

- Stoner KJL, Joern A (2004) Landscape vs. local habitat scale influences to insect communities from tall grass prairie remnants. *Ecological Applications* **14**, 1306-1320.
- Suen J-P, Herricks EE (2006) Investigating the causes of fish community change in the Dahan River (Taiwan) using an autecology matrix. *Hydrobiologia* **568**, 317-330.
- Sutherland GD (2006) Predicting the ecological consequences of environmental change: a review of the methods. *Journal of Applied Ecology* **43**, 599-616.
- Tattari S, Schultz T, Kuussaari M (2003) Use of belief network modelling to assess the impact of buffer zones on water protection and biodiversity. *Agriculture, Ecosystems and Environment* **96**, 119-132.
- Thogmartin WE, Knutson MG (2007) Scaling local species-habitat relations to the larger landscape with a hierarchical spatial count model. *Landscape Ecology* **22**, 61-75.
- Tsumoto K, Yoshinaga T, Iida H, Kawakami H, Aihara K (2006) Bifurcations in a mathematical model for circadian oscillations of clock genes. *Journal of Theoretical Biology* **239**, 101-122.
- Van Looy K, Vanacker S, Jochems H, De Blust G, Dufrene M (2005) Ground beetle habitat templates and riverbank integrity. *River Research and Applications* **21**, 1133-1146.
- van Nes EH, Scheffer M (2005) Implications of spatial heterogeneity for catastrophic regime shifts in ecosystems. *Ecology* **86**, 1797-1807.
- Varis O, Kuikka S (1997) Joint use of multiple environmental assessment models by a Bayesian meta-model: the Baltic salmon case. *Ecological Modelling* **102**, 341-351.
- Webster IT, Harris GP (2004) Anthropogenic impacts on the ecosystems of coastal lagoons: modelling fundamental biogeochemical processes and management implications. *Marine and Freshwater Research* **55**, 67-78.
- West AD, Caldow RW (2006) The development and use of individual-based models to predict the effects of habitat loss and disturbance on waders and waterfowl. *Ibis* **148**, 158-168.
- Westoby M, Walker BH, Noy-Meir I (1989) Opportunistic management for rangelands not at equilibrium. *Journal of Range Management* **42**, 266-274.
- Wilkinson L (2004) Classification and regression trees. In 'Statistics I'. (Ed. S 11) pp. I-35-I-56. (SYSTAT Software: Richmond)
- Wilkinson SR, Naeth MA, Schmiegelow FKA (2005) Tropical forest restoration within Galapagos National Park: Application of a state-transition model. *Ecology and Society* **10**, 28.
- Wolter C, Menzel R (2005) Using commercial catch statistics to detect habitat bottlenecks in large lowland rivers. *River Research and Applications* **21**, 245-255.
- Wooldridge S, Done T (2004) Learning to predict large-scale coral bleaching from past events: A Bayesian approach using remotely sensed data, in-situ data, and environmental proxies. *Coral Reefs* **23**, 96-108.
- Wooldridge S, Done T, Berkelmans R, Jones R, Marshall P (2005) Precursors for resilience in coral communities in a warming climate: a belief network approach. *Marine Ecology Progress Series* **295**, 157-169.
- Zlotnicki V, Wahr J, Fukumori I, Song YT (2007) Antarctic circumpolar current transport variability during 2003-05 from GRACE. *Journal of Physical Oceanography* **37**, 230-244.

Acknowledgements

This research was supported by the CSIRO Flagship Collaboration Fund and represents a collaboration between CSIRO, the University of Adelaide, Flinders University and SARDI Aquatic Sciences.

We also acknowledge the contribution of several other funding agencies to the CLLAMM program and the CLLAMMecology Research Cluster, including Land & Water Australia, the Fisheries Research and Development Corporation, SA Water, the Murray Darling Basin Commission Living Murray program and the SA Murray-Darling Basin NRM Board. Other research partners include Geoscience Australia, the WA Centre for Water Research, and the Flinders Research Centre for Coastal and Catchment Environments. The objectives of this program have been endorsed by the SA Department of Environment and Heritage, SA Department of Water, Land and Biodiversity Conservation, SA Murray-Darling Basin NRM Board and Murray-Darling Basin Commission.

We also would like to thank Dr Dan Rogers and Dr Nicky Grigg for their constructive comments on an earlier version of this report.

