

Report Cover Page

ACERA Project:		
06/03		
Title		
New spatial analysis methods for improved hazard/risk identification		
Author(s) / Address (es)		
Dr Jane Elith ¹ , Dr Daniela Leonte ² , and members of the working group. ¹ School of Botany; The University of Melbourne; Parkville Victoria 3010 ² School of Safety Science, University of NSW, Sydney, NSW 2052		
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)		
Project report		
Summary		
<p>This workshop aimed to bring together people with diverse but relevant expertise, to explore current best practice in spatial analysis and assess its relevance to the broad range of situations relevant to biosecurity risk assessments. Twenty five people from Australia and overseas attended, with expertise in spatial statistics, Bayesian analysis, statistical and mathematical modelling, invasion ecology and application of models in land management and conservation. The focus was methods for predicting distributions and movement of taxa that are not at equilibrium with their environment, such as invasive species or species whose distributions are changing due to climate change. The broad classes of methods usually applied to these problems include: (i) the same types of correlative models that are used to model current species distributions, optionally with modifications to represent the lack of equilibrium in the system or to represent physiological responses; (ii) models that explicitly focus on dispersal, spread and other population processes; and (iii) models that define important physiological constraints and identify those regions that satisfy the species requirements.</p> <p>The group made substantial progress in a number of areas, and several participants will continue work after the group publish it in international journals. Several interesting ideas emerged that may lead to ongoing collaboration. Effort was spread over several topics:</p> <ol style="list-style-type: none"> 1. Clarifying where applications of correlative models are useful, and exploring how to fit them in a way most suited to non-equilibrium situations. This will lead to one review paper and at least one methodological paper. 2. Comparing two alternative approaches to surveillance planning for the detection of weeds. 3. Exploring the use of Bayesian hierarchical models for modelling species distributions, in a form that allows incorporation of important processes such as dispersal. As a case study, considering its use in predicting the invasion front for Cane toads in north western Australia. 4. Developing realistic simulated data that enables assessment of alternative approaches for modelling distributions of invading species. 		
ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: () Yes () No	Date:



New spatial analysis methods
for improved hazard/risk identification

ACERA Project No. 06/03

Jane Elith, Daniela Leonte
& group members

Final report

February 2008



Acknowledgements

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and the Applied Environmental Decision Analysis research facility (AEDA).

This report is the result of a working group held at AMSI, Carlton, Melbourne attended by Agathe Leriche, Art Diggle, Ben Phillips, Bob Sutherst, Brendan Wintle, Brian Kay, Cindy Hauser, Daniela Leonte, Dave Richardson, David Warton, Greg Hood, Jane Elith, John Hearne, John Leathwick, Josie Hughes, Julian Fox, Kim Lowell, Mark Burgman, Mark Stanaway, Matt Wand, Michael Kearney, Noel Cressie, Rob Reeves, Simon Barry, and Simon Ferrier. **This report is the result of their collective knowledge and efforts.**

Disclaimer

This report has been prepared by consultants for the Australian Centre of Excellence for Risk Analysis (ACERA) and the views expressed do not necessarily reflect those of ACERA. ACERA cannot guarantee the accuracy of the report, and does not accept liability for any loss or damage incurred as a result of relying on its accuracy.

Table of contents

Acknowledgements	3
Disclaimer	4
Table of contents.....	5
List of Tables.....	6
List of Figures	7
1. Executive Summary	7
2. Introduction	9
3. Using statistical models for predicting the distribution of species not at equilibrium.	11
4. Use of Bayesian hierarchical models for predicting aspects of Cane toad invasion	13
5. Integrating a weed dispersal simulation model with a surveillance allocation model.....	17
6. Conclusions.....	18
Appendix 1: Participants	19
Appendix 2: Description of the meeting and its activities.....	20
Appendix 3: Summary of paper: "Use of correlational models to predict species distributions: black art or hard science?"	23
Appendix 4: Testing adaptations to statistical species distribution models for species not at equilibrium	24
Appendix 5: Using geographic predictors in models of species invasions	25
Appendix 6: Integrating species distribution modelling and invasion dispersion	27

List of Tables

Table 1: Examples of the use of a range of statistical models for predicting species invasions or their distributions Page 12

List of Figures

- Figure 1: The Cane toad and its distribution over time, and recent numbers in the north-west of Northern Territory Page 13
- Figure 2: Notes showing development of ideas for modelling evolving spread rates - a work in progress. Page 15
- Figure A4: Predictions of the potential distribution of the Cane toad under current and future climates using three different modelling methods and modifications to the modelling data. Page 24
- Figure A6.1: Tables to be part of the manuscript Page 27
- Figure A6.2: Examples of the data produced from the simulator Page 28

1. Executive Summary

This is the report of a five-day working group held in Melbourne in late 2007. The aims of the working group were:

"..to bring together people with diverse but relevant expertise, so that we can explore current best practice in spatial analysis and assess its relevance to the broad range of situations relevant to biosecurity risk assessments. We will identify candidate methods that show promise, produce case studies demonstrating the effectiveness of these in typical applications, write a report outlining the potential and limitations of a range of methods, and scope out promising topics for future research."

Twenty-five people from Australia and overseas attended, with expertise in spatial statistics, Bayesian analysis, statistical and mathematical modelling, invasion ecology and application of models in land management and conservation. Our focus was relevant methods for predicting distributions and movement of taxa that are not at equilibrium with their environment, such as invasive species or species whose distributions are changing due to climate change. The broad classes of methods usually applied to these problems include: (i) the same types of correlative models that are used to model current species distributions, optionally with modifications to represent the lack of equilibrium in the system or to represent physiological responses; (ii) models that explicitly focus on dispersal, spread and other population processes; and (iii) models that define important physiological constraints and identify those regions that satisfy the species requirements.

The group made substantial progress in a number of areas, and several participants will continue work after the group and publish it in international journals. Several interesting ideas emerged that may lead to ongoing collaboration. Effort was spread over several topics:

1. Clarifying where applications of correlative models are useful, and exploring how to fit them in a way most suited to non-equilibrium situations. This will lead to one review paper and at least one methodological paper.
2. Comparing two alternative approaches to surveillance planning for the detection of weeds.
3. Exploring the use of Bayesian hierarchical models for modelling species distributions, in a form that allows incorporation of important processes such as dispersal. As a case study, considering its use in predicting the invasion front for Cane toads in north western Australia.
4. Developing realistic simulated data that enables assessment of alternative approaches for modelling distributions of invading species.

The report details progress on the topics and future directions.

2. Introduction

The growing international movement of both people and goods brings increased risks of transfer of species to locations where they do not naturally occur. This carries with it sometimes considerable penalties both economically, through impacts on productivity of agriculture, forestry and fisheries (e.g., from weeds and pathogens), and ecologically, through the alteration of natural ecosystems and loss of their associated biodiversity values. Robust information about the potential distribution of such organisms is required to identify taxa likely to become problematic if introduced, and to manage those that are already established. For the latter, predictive models can be useful both for facilitating targeted surveillance to prevent further spread, and for assessing the impacts of alien species on natural ecosystems. In the wider context, information on species distributions is increasingly used to underpin the management of natural resources, including identification of representative reserves, assessment of threats, setting targets for ecosystem restoration, and managing the sustainable use of harvested species.

There are numerous types of models used for predicting the current distributions of species or their potential distributions. Whilst there is some consensus that well-implemented correlative models provide reasonable predictions of current distribution (Dormann 2007, Elith *et al.* 2006, Guisan and Thuiller 2005, Latimer *et al.* 2006), there is still considerable debate about how best to model species that are not currently at equilibrium. The clearest examples of this lack of equilibrium within the contexts discussed here are (i) invasive species, where a species has not yet reached all suitable habitats in a region, and (ii) climate change, where, as a result of changing environment, there is some doubt about the long term suitability of a given location for a species.

Some main contenders for modelling species not at equilibrium are: (i) the same types of correlative models that are used to model current species distributions, optionally with modifications to represent the lack of equilibrium in the system or to represent physiological responses (Araujo & New 2007, Hartley *et al.* 2006); (ii) models that explicitly focus on dispersal, spread and other population processes (BenDor *et al.* 2006); and (iii) models that define important physiological constraints and identify those regions that satisfy the species requirements (Kearney and Porter 2004). Whilst research teams tend to focus on one or other of these, there is a lack of communication across fields that might produce synergies across the spectrum of model types, and coherent communication of the strengths and weaknesses of alternative approaches. There are also situations early in invasions or disease spread where there are insufficient data to model the distribution of the species or the locations of the disease, and in these development of tools capable to predict the distribution of a species at the very early stage of incursion is important for rapid responses.

This working group aimed to address these issues by bringing together experts in spatial modelling and invasion biology, and with them explore current best practice in spatial analysis and assess its relevance to the broad range of situations relevant to biosecurity risk assessments. Our strategy was to invite international and Australian researchers and students (Appendix 1), and to first explore their thoughts about the most interesting current questions and practices in modelling species not at equilibrium. Notes on these discussions are presented in Appendix 2. We then worked on a subset of topics:

- Applying statistical models to non-equilibrium species
- Methods for testing and comparing models for non-equilibrium situations
- Surveillance
- Forecasting with Bayesian hierarchical models (case study: Cane toad invasions)

As outlined in the following sections we made considerable progress in identifying the potential and limitations of a range of methods, and in scoping promising topics for future research. One of the aims of the project was to provide guidelines for the sorts of problems amenable to analysis with spatial distribution models, and review alternative methods for those requiring different approaches. We decided that a useful way to achieve this is via the papers outlined in this report, because in general the topics required more work than could be completed within the group, and the output will have a broad an international audience. This report summarises the topics and progress to date, and points to ongoing work.

References

Araujo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22:42-47.

BenDor, T. K., S. S. Metcalf, L. E. Fontenot, B. Sangunett, and B. Hannon. 2006. Modeling the spread of the Emerald Ash Borer. *Ecological Modelling* 197:221-236.

Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. *Basic and Applied Ecology* 8:387-397.

Elith, J., C. H. Graham, *et al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993-1009.

Hartley, S., R. Harris, and P. J. Lester. 2006. Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters* 9:1068-1079.

Kearney, M., and W. P. Porter. 2004. Mapping the fundamental niche: physiology, climate and the distribution of nocturnal lizards across Australia. *Ecology* 85:3119-3131.

Latimer, A. M., S. S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16:33-50.

3. Using statistical models for predicting the distribution of species not at equilibrium.

Leathwick, Ferrier, Sutherst, Leriche, Richardson, Warton, Wand, Barry, Burgman

In this context, we define statistical models as those that rely on the relationships between records of a species occurrence, and environmental conditions, at locations through the region of interest to model the distribution of the species. Examples range from envelope methods such as BIOCLIM or CLIMATE, that define the broad envelope of suitable climatic conditions, to regression methods such as generalised linear models (GLMs) and generalised additive models (GAMs), to machine learning methods such as artificial neural networks (ANN), boosted regression trees (BRT) and GARP (the genetic algorithm for rule-set prediction). Examples of their application to modelling situations where species are not at equilibrium include (i) those predicting the potential distribution of species that occur in a region and are spreading, (ii) those that predict the distribution of a species from records in other regions or countries; and (iii) predictions of future distributions of species given projected climate change scenarios (Table 1).

The reason for using statistical models for predicting distributions is that they are relatively quick and straightforward, and are suited to the types of data typically available. Alternatives such as CLIMEX (a hybrid correlation – mechanistic method; see Table 1 for example) or process-based models are more complex and in most cases, more data-demanding, and it is not clear that their predictions are always more reliable than the simpler methods.

The working group that discussed use of statistical models decided to:

1. clarify the theoretical framework underlying use of species distribution models for species at equilibrium, and identify key issues in using them to model species not at equilibrium;
2. test adaptations to the usual methods that are consistent with typical properties of data for invading species;
3. test regression methods with spatial terms for their utility for predicting the spread of species;
4. create a method for simulating data in a realistic way so it can be used to compare and test alternative methods;
5. develop a suite of evaluation statistics.

Considerable progress was made on all these topics both at the working group and since then. Three manuscripts are in preparation (see Appendices 3 to 6), and are likely to be submitted by mid 2008. The topics are likely to be directly relevant to biosecurity issues, and ongoing contact between working group members and Government departments will provide the opportunity for adoption of new ideas.

Table 1: Examples of the use of a range of statistical models for predicting species invasions or their distributions

Species and location	Aim of modelling	Modelling method(s)	Reference
Invasive ants, New Zealand	To model the potential distribution of ants using existing museum records within NZ	BIOCLIM DOMAIN MAXENT	Ward, D. F. 2007. Modelling the potential geographic distribution of invasive ant species in New Zealand. <i>Biological Invasions</i> 9 :723-735.
Purple loosestrife, USA	To use records within USA to model potential distribution, and test how many years of data were necessary to predict well.	DOMAIN GARP	Welk, E. 2004. Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: Purple loosestrife.. in North America. <i>Ecological Modelling</i> 179 :551-567.
Mediterranean fruit fly and gypsy moth, world-wide	To model the global distribution of the species and understand the main correlates of their distribution	Artificial neural network	Gevrey, M., and S. P. Worner. 2006. Prediction of global distribution of insect pest species in relation to climate by using an ecological informatics method. <i>Journal of Economic Entomology</i> 99 :979-986.
Cane toad, Australia	To predict the potential distribution of the Cane toad	Regression models on subsets of data stratified by time	Urban, M. C., B. L. Phillips, D. K. Skelly, and R. Shine. 2007. The Cane toad's... increasing ability to invade Australia is revealed by a dynamically updated range model. <i>Proceedings of the Royal Society B-Biological Sciences</i> 274 :1413-1419.
Tick, Africa and New World	To use data from Africa to find climate that is suitable and predict to New World	DOMAIN, MAXENT	Estrada-Pena, A., R. G. Pegrarn, N. Barre, and J. M. Venzal. 2007. Using invaded range data to model the climate suitability for <i>Amblyomma variegatum</i> .. in the New World. <i>Experimental and Applied Acarology</i> 41 :203-214.
Eragrostis in Arizona	To model the potential distribution of the species in Arizona and New Mexico using records from invasions so far	Regression, GARP	Schussman, H., E. Geiger, T. Mau-Crimmins, and J. Ward. 2006. Spread and current potential distribution of an alien grass, <i>Eragrostis lehmanniana</i> Nees, in the southwestern USA: comparing historical data and ecological niche models. <i>Diversity and Distributions</i> 12 :582-592.
Fire ant in Oceania	To use range data from USA to understand climate responses of the ant and predict into new regions	CLIMEX	Sutherst, R. W., and G. Maywald. 2005. A climate model of the red imported fire ant, <i>Solenopsis invicta</i> Buren (Hymenoptera : Formicidae): Implications for invasion of new regions, particularly Oceania. <i>Environmental Entomology</i> 34 :317-335.

4. Use of Bayesian hierarchical models for predicting aspects of Cane toad invasion

Hughes, Cressie, Leonte, Wintle, Phillips, Kearney, Reeves, Stanaway, Hood, Diggle, Elith

From early discussions in the group it became clear that there was widespread interest in the use of Bayesian hierarchical models for modelling species not at equilibrium. These models have relatively recently been applied in ecology, and are discussed in relation to modelling of species distributions in Latimer *et al.*(2006). The group had strong representation of researchers well-versed in Bayesian statistical modelling (Cressie, Reeves, Wintle, Stanaway, Leonte). This, combined with a relevant problem and the availability of experts and their data on Cane toad invasions (Phillips, Kearney), provided an outstanding opportunity to explore state-of-the-art modelling methods that might be useful for (1) prediction of the time of arrival of the invasion front at selected locations in northern WA; (2) an understanding of invasion dynamics since the toads were first introduced in 1935 (Figure 1). We describe these later, but first provide relevant background.

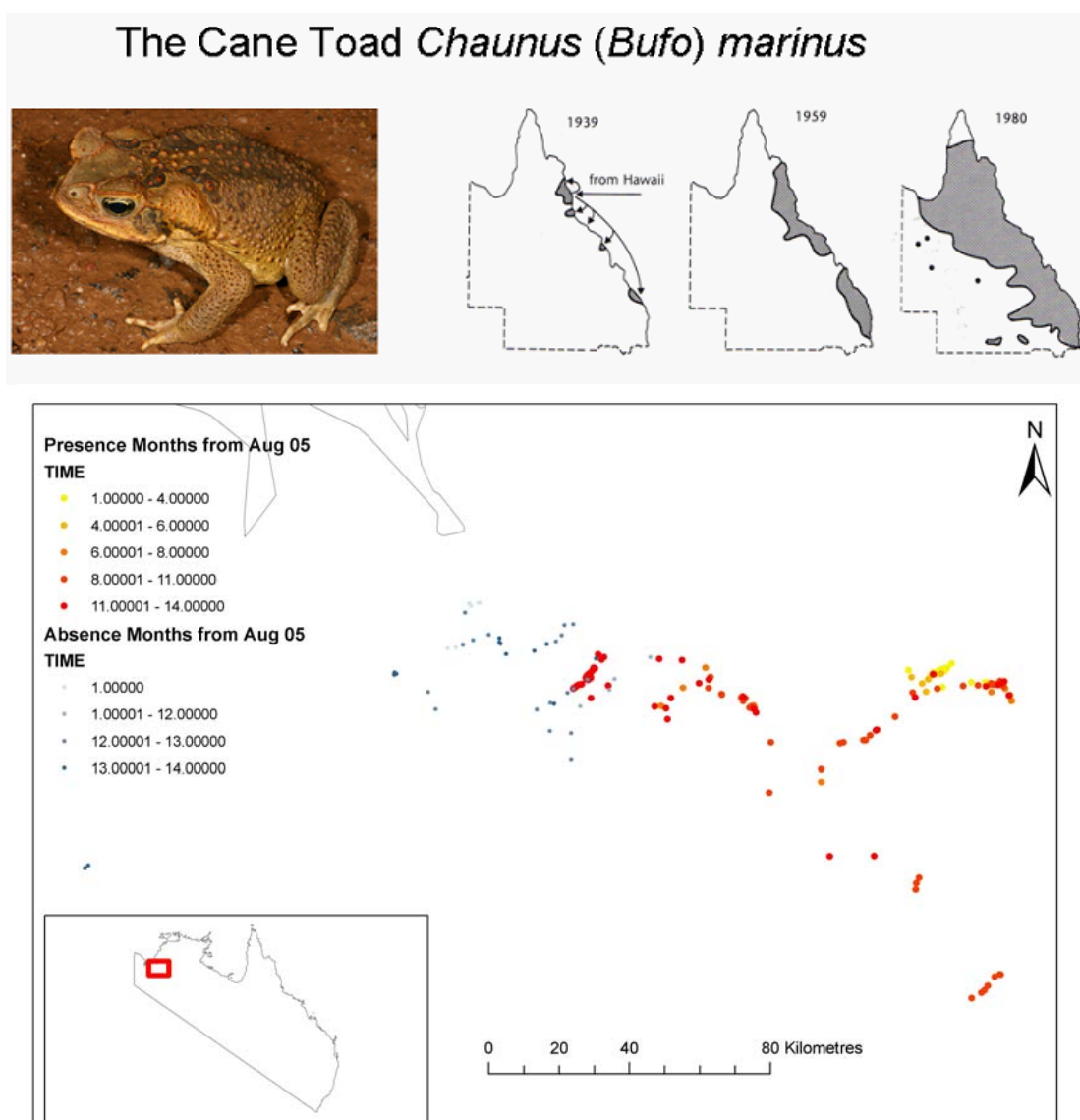


Figure 1: The Cane toad and its distribution over time (upper panel; courtesy of B.Phillips) and recent numbers in the north west of northern territory (lower panel).

Once an invasive species has established, so that eradication is impractical, a key challenge for applied research is predicting the rate of spread into new areas. For some important invasive species, intense monitoring effort has yielded a useable record of population abundance over time and space. In these cases, we have a unique opportunity to understand spread processes, and possibly predict future spread. However, even in cases where sufficient data exists, inferring underlying spread processes and parameters from space-time population data remains a technical challenge. Our working sub-group focused on understanding hierarchical Bayesian models in this context.

Predicting future spread accurately enough to be useful for monitoring and management action is difficult because dispersal and population processes such as survival and reproduction interact to produce the spread patterns we observe. Thus, observed spread rate typically varies with habitat suitability. For example, the spread rate of Cane Toads declines in hot dry areas where toads are unable to survive. The hierarchical Bayesian analysis method we considered (Wikle et al. 1998, Wikle 2003, Wikle and Hooten 2006, Cressie et al. 2007, Hooten et al. 2007) is promising because it allows relatively complex, biologically informed models to be fitted directly to observed data. A rich vein of ecological theory has focused on developing and analyzing models of population dynamics with spread, but coherent statistical treatment of these models has proved more challenging. This hierarchical Bayesian approach allows, at least in some cases, the posterior probability distributions of the parameters of biologically meaningful models to be jointly estimated. The probability that an organism will establish at a given time and location (i.e. a projection of spread with uncertainty) is one example of a useful result that could be obtained using this method.

These models are not straightforward. The first challenge for our sub-group was to better understand the conceptual basis for hierarchical Bayesian analysis. We focused on a related problem, published as an analysis of the invasion of collared dove in North America (Hooten et al. 2007), and understanding how this example might be adapted to the spread of Cane Toads across Australia. Considerable effort was devoted to synthesizing and exploring the Cane Toad data, and considering what kind of population model would be warranted given the data, Cane Toad biology, and the purpose of the analysis. An added complication in the Cane Toad invasion is that Cane Toad spread rate has increased over time, so models would ideally include a biologically and mathematically sound model to describe this variable spread rate. This is a challenging task.

For the first question, that of predicting local arrival times and places, we remain interested in applying these Bayesian hierarchical models to a relatively small region on the current Cane Toad invasion front (Figure 1, lower panel), where monitoring data is sufficient to support a model of population density (rather than just presence-absence), and the trend of increasing spread rate over time can be ignored. Unfortunately, actually fitting such a model remains a significant technical challenge. To help better understand the nature of this challenge, at the working group we attempted to repeat the collared dove analysis in R (<http://www.r-project.org/>) and WinBUGS (Lunn et al. 2000). This showed promise, though WinBUGS was slow and the analysis might have to be limited if kept within WinBUGS. Nevertheless, this project will further increase our understanding of the potential and current limitations hierarchical Bayesian analysis, and if successful, could provide some short-term projections of spread. Likely participants include Wintle, Elith, Hughes, Cressie, Leonte and Phillips, if time allows.

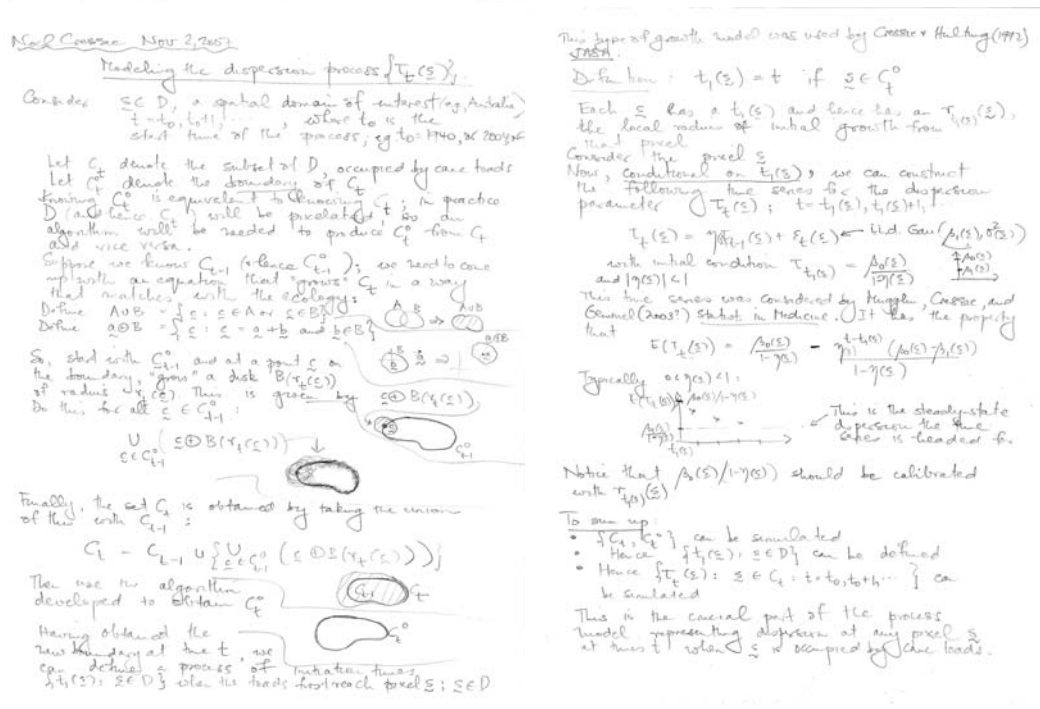


Figure 2: Notes (author: N.Cressie) showing development of ideas for modelling evolving spread rates - a work in progress.

With respect to the second analysis, that of invasion over the whole Cane toad range, our current understanding is that it would not be possible to fit the more complex variable spread model using any standard Bayesian Markov chain Monte Carlo (MCMC) techniques, and the most promising options for that model would be either by working with statisticians able to write full conditional probability statements and combine these into a model, or to use Approximate Bayes Computation (ABC) (see Beaumont et al. 2002, Sisson 2007, Sisson et al. 2007). We remain interested in analyzing a model that includes an evolving spread rate using ABC techniques. This is a more ambitious undertaking, as ABC techniques have not yet been well developed or established in ecology. One group member (Reeves) is currently working on computationally efficient model implementation, which is a crucial first step in the ABC approach. Continuing work on the Cane Toad application of these methods would require significant further commitment from various group members. Whether or not work continues on the Cane Toad example, this workshop has been invaluable for increasing our understanding of available analysis techniques, and thus helping to improve the quality of future research.

References

Beaumont, M. A., W. Y. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**:2025-2035.
 Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2007. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling.
 Cressie, N. and Hulting, F.L., 1992. A spatial statistical analysis of tumor growth. *Journal of the American Statistical Association*, 87(418), June, 272 – 284.
 Hooten, M. B., C. K. Wikle, R. M. Dorazio, and J. A. Royle. 2007. Hierarchical spatio-temporal matrix models for characterizing invasions. *Biometrics* **63**:558-567.

-
- Latimer, A. M., S. S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. *Ecological Applications* **16**:33-50.
- Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**:325-337.
- Mugglin, A.S., Cressie, N. and Gemmell, I., 2002. Hierarchical statistical modeling of influenza epidemic dynamics in space and time, *Statistics in Medicine*, 21(18), September, 2703 – 2721.
- Sisson, S. A. 2007. Genetics and stochastic simulation do mix! *American Statistician* **61**:112-119.
- Sisson, S. A., Y. Fan, and M. M. Tanaka. 2007. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **104**:1760-1765.
- Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. Hierarchical Bayesian space-time models. *Environmental & Ecological Statistics* **5**:117-154.
- Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84**:1382-1394.
- Wikle, C. K., and M. B. Hooten. 2006. Hierarchical Bayesian spatio-temporal models for population spread. *in* J. Clark and A. E. Gelfand, editors. *Hierarchical Modelling for the Environmental Sciences*. Oxford University Press, Oxford.

5. Integrating a weed dispersal simulation model with a surveillance allocation model

Cindy Hauser & Julian Fox

We spent some time examining how the recent research brought to the workshop by Julian and Cindy might be combined to improve surveillance planning for the detection of weeds. Julian presented a spatially-explicit simulation model of a weed's incursion and spread across a landscape with surveillance and control actions taken over the landscape. The software developed has the capability to import GIS layers that affect weed dispersal and germination, such as wind, roads, rivers and habitat type. Cindy's research focussed on the optimal allocation of surveillance resources over a heterogeneous landscape. The probability of weed presence (a function of habitat and likely dispersal) and the capacity of available surveillance methods to detect the weed determine the amount of surveillance effort to be allocated to each candidate site in a landscape. We found that an integrated model of weed germination and dispersal with optimal surveillance and control could be developed, but it would require substantial modification of the existing models and was beyond the scope of the workshop.

There were two key differences in the structure of Julian and Cindy's models. First, Julian's model includes the stage structure and abundance of the weed population at each location, while Cindy's model records only the presence or absence of detectable individuals (i.e. plants but not seed banks). Second, Julian's model assumed that if a location is surveyed, any plants (but not seeds) present at the site are detected with certainty. By contrast, Cindy modelled detection as an increasing function of surveillance effort and cost e.g. the more time spent by a paid observer at a site, the higher the probability of detecting the weed if it is indeed present.

We determined that under Julian's model of certain detection, the optimal allocation of surveillance effort would be to target locations with the highest probability of weed presence. The probability of weed presence could be estimated from simulations of dispersal and germination using Julian's model. Under Cindy's model of variable detection, locations with a high probability of weed presence would again be prioritised, but the precise amount of survey effort to be allocated to each is a more complex function of the capacity of the surveillance method to detect the weed in the local habitat.

Julian's original model with full stage-structure and abundance recorded at each location is too complex for Cindy's model. However multiple simulations of Julian's model could be used to parameterise a simpler presence/absence model of weed spread across the landscape. Two complicating factors are:

1. The selection of an appropriate time length for each increment in the discrete-time simulation. Mean generation time is one option;
2. Dispersal should be a function of the habitat types encountered by the seed, as well as distance from the source plant. Modelling the dispersal and germination of a seed that encounters multiple habitat types on its path is problematic, as recognised by another group at the workshop.

6. Conclusions

The project deliverables were:

The working group will produce guidelines and recommendations for the development and application of analytical methods for spatial analysis... Success may be measured by

- *Successful involvement of leaders in spatial analysis in the workshop.*
- *Completion of a report documenting the outcomes of the meeting*
- *Generation of on-going interest-driven research in the areas initiated by the working group*
- *Uptake of the group's recommendations by Government.*

With the successful completion of the working group and the work reported here, the first three points have been achieved. Members of the working group have ongoing contacts with the Government and uptake of ideas and methods is likely.

Appendix 1: Participants

Title	First name	Surname	City/State	Organisation	Relevant Speciality
Dr	Simon	Barry	Canberra	CSIRO	Statistical methods in biosecurity settings
Prof	Mark	Burgman	Melbourne	University of Melbourne, ACERA	Risk analysis
Prof	Noel	Cressie	Ohio, USA	Ohio State University	Spatial statistics
Dr	Art	Diggle	Perth	Dept of Agriculture	Weeds (spread, resistance, modelling)
Dr	Jane	Eliith	Melbourne	University of Melbourne	Species distribution modelling
Dr	Simon	Ferrier	Canberra	CSIRO	Species and biodiversity modelling; conservation planning
Dr	Julian	Fox	Brisbane	University of Queensland	GIS methods for modelling weed spread
Dr	Cindy	Hauser	Melbourne	University of Melbourne	Optimal monitoring
Prof	John	Hearne	Melbourne	RMIT	Mathematics; model simplification
Dr	Greg	Hood	Canberra	Bureau of Rural Statistics	Statistical modelling and simulation
Ms	Josie	Hughes	Toronto, Canada	University of Toronto	Spatial modelling (PhD student)
Prof	Brian	Kay	Brisbane	Queensland Inst of Medical Research	Surveillance, eradication and control strategies for arboviruses and their
Dr	Michael	Kearney	Melbourne	University of Melbourne	Ecophysiological species models
Dr	John	Leathwick	NZ	National Institute for Water and Atmospheric Research	Species distribution models and their use in conservation & resource
Dr	Daniela	Leonte	NSW	University of NSW	Spatial statistical analysis; risk analysis
Dr	Agathe	Leriche	Canberra	ENSIS	Use of CLIMEX
Prof	Kim	Lowell	Melbourne	DPI and University of Melbourne	Spatial systems modelling
Dr	Ben	Phillips	Sydney	University of Sydney	invasion biology and evolution and ecology of Cane toads; evolution and
Dr	Rob	Reeves	Brisbane	Queensland University of Technology	Bayesian modelling
Prof	David	Richardson	South Africa	Stellenbosch University, Centre of Invasion Ecology	Species' invasions, spatial simulation models
Mr	Mark	Stanaway	Brisbane	Queensland University of Technology	Surveillance systems for pests (PhD student)
Dr	Bob	Sutherst	Brisbane	University of Queensland	Mechanistic models for invasions (CLIMEX)
Prof	Matt	Wand	NSW	University of NSW	Statistical modelling
Dr	David	Warton	Sydney	University of NSW	Statistical methods for analysing ecological data
Dr	Brendan	Wintle	Vic	University of Melbourne	Spatial analytical methods and their application in monitoring

Appendix 2: Description of the meeting and its activities

Daniela Leonte

Preamble

The Australian Centre for Excellence in Risk Analysis (ACERA) organised a five-day spatial statistics workshop in Melbourne, between October 29th and November 2nd, 2007. The workshop is the materialization of ACERA Project No. 06/03.

The workshop venue was provided by the Australian Mathematical Sciences Institute (AMSI), with activities being facilitated by Professor Mark Burgman. Project leaders were Dr Jane Elith, Dr Brendan Wintle, Professor Matt Wand, and Dr Daniela Leonte.

Workshop-related activities

Activities related to the workshop occurred before, during, and after the event. Approximately one month prior to the workshop, participants were subscribed to a group email address, which facilitated communication and transfer of information. Following distribution of the workshop agenda, participants were invited to send relevant research articles published by them or their research group, to the rest of the workshop participants. This aimed to familiarize participants with the research of their colleagues and to give them an understanding of the research interests and expertise of each workshop attendees.

During the workshop a number of activities were undertaken, both generic and project/topic specific. All workshop attendees took part in the generic activities, which occupied the first and parts of the second, fourth and fifth days of the workshop. The remaining time was spent in project/topic specific activities, to which each workshop participant subscribed, given his/her interests and ability to contribute to these activities.

Generic workshop activities

The generic activities that made the subject of the first day of the workshop included:

1. Brief individual presentations on a problem/tool of relevance to the workshop, with which individual participants were familiar or involved;
2. Discussions on possible projects, including:
 - description of available data sets;
 - brainstorming themes/questions that could be usefully tackled during the workshop;
 - review of existing methods that could be used in the context of a specific data set and theme/question; and
 - achievable workshop outputs, both for the immediate and more distant future.
3. Establishment of tentative working groups.
4. Outline of workshop milestones and desired outputs.

The participants' presentations occurred over the first half-day. These highlighted the broad range of expertise and interests of the audience, the complex and dynamic nature of biosecurity risks assessment and management processes, as well as the numerous mathematical and computer tools available to aid biosecurity decision-making.

In general, the expertise and interest of participants fell in one, or a number of the following categories:

- Biodiversity assessment and conservation planning;

-
- Spread and distribution of native (eg., forests, lizards) and invasive species (e.g., mosquitoes, weeds, Cane toads, crop diseases);
 - Ecological population dynamics under climate change conditions;
 - Ecosystem – farms interactions and land use mapping;
 - Epidemics characterisation (early detection, spread modelling, emergency response);
 - Decision theory, including the value of information, stochastic optimization and resource allocation;
 - Development and/or use of generic statistical tools applicable to biosecurity problems, including Bayesian hierarchical models, splines, multivariate analysis, spatial partitions, Approximate Bayesian Computation (ABC), Bayesian learning and Bayesian decision theory;
 - Computer tools for biosecurity applications (eg., CLIMEX, WinBUGS, GIS).

The presentations highlighted the broad range of modelling tools employed by workshop participants in their research, including spatial and spatio-temporal Bayesian hierarchical models, spatial smoothers, generalized dissimilarity models, generalized additive models, machine learning methods (e.g. boosted regression trees, multivariate adaptive regression splines), Bayesian decision theory, ecophysiological models, and ABC.

Data available for research were also mentioned. These included spatio-temporal data on the spread of number of ecological species (i.e., Canadian insects, rock snot, Cane toad and Spiraling white fly).

Based on the availability of data and the expertise and interests among the workshop participants, a number of themes/questions were highlighted as important for topics for further research:

- How to include biology into dispersal/invasion models. Current tools rely on physical equations and statistical analysis of monitoring data, but fail to account for biological attributes that are critical to the dynamics of spread models. It was recognized that their absence from current models is a significant reason behind the sometimes poor predictive performance and lack of agreement between models.
- Suitable methods for deriving absence data for correlative models. Particularly for invasions, observations of absence only indicate absence of the species at that time, not lack of suitable habitat. They are dependent on the progression of the invasion front, so research is needed on how to present data so that correlative models applied to invasions are robust. There is also a need to address these problems via a search for a possible dynamic structure in the mean of the error. A related issue, for some species, is the relevance of the question: ‘When is zero really zero?’. For example, should larvae and eggs be accounted for in simulations of the spread of a pest?
- Modelling communities for climate change. It was commented that current climate change / species distribution models aim to incorporate population dynamics, but do not account for species interactions, making them unsuitable for biodiversity change assessment.
- Assessing systems in- and not-at- equilibrium. This was highlighted as a critical question requiring research and the development of tools capable to model dynamic spatial systems over time.
- Robust tools for extrapolation and validation. Because extrapolating to unknown conditions is a difficult task, using different models on the same data tends to lead to very

different predictions. One difficulty with species not at equilibrium is that there are rarely useful data for evaluating which model is most realistic – i.e., there is no final stable situation against which the models can be assessed. Usually, the only records that exist are those representing the current locations. This led to discussion of the use of simulated data for assessing the performance of different models. As outlined in the main body of the report, there is likely to be a data simulator published as a direct outcome from the workshop.

- Rapid response/dynamic allocation for ‘first contact’. This refers to the development of tools capable to predict the distribution of a species at the very early stage of incursion (first contact problem: If a species is found at one location, where did it come from and where would it go?). An obvious example at the time was the spread of equine influenza, though unfortunately the data were too sensitive to be made available for the group. However, there was general useful discussion around the problem of how to model such data.

The discussions highlighted that any products/models developed from research on these themes/questions would be likely most useful if they were scientifically credible in terms of their predictive power, be informed by broad expertise across disciplines, and practical in terms of the usability of the product.

Project/topic specific activities

The generic activities conducted during the initial stage of the workshop led to the establishment of three project/topic specific activities, which were pursued by participant sub-groups for the remaining three-and-a-half days of the workshop. These are presented in the main body of this report.

Appendix 3: Summary of paper: "Use of correlational models to predict species distributions: black art or hard science?"

J. R. Leathwick, S. Barry, J. Elith, S. Ferrier, J. Hearne, J. Hughes, M. Kearney, D. Richardson, B. Sutherst, and M.J.Fortin

Leathwick is lead author on this paper, which will be submitted to an international journal. We present here the introduction and a brief summary of likely take-home messages, and will make the paper available once it is accepted for publication.

Introduction

Increasing use is being made of correlationally-based species distribution models (SDM) in ecology. These models are distinguished by use of a statistical or envelope fitting procedure to make inferences about the relationship between a species and its environment, based on observations from some landscape of interest. These models are then commonly used to make environment-based predictions at new geographic locations.

Initial uses focussed predominantly on analysis of natural species distributions in which assumptions could be made that the species of interest exhibited at least some degree of equilibrium with environment. A number of studies have shown that such analyses of species relationships with environment under equilibrium conditions can be used to predict/recover the underlying geographic distributions with reasonable accuracy. This has in turn encouraged use of SDM to predict or forecast the potential spread of species under novel conditions, including when assumption of equilibrium are clearly unrealistic. These include prediction of both the potential geographic distribution of a species when introduced to new locations (e.g., invasive species), and of likely changes in the geographic distributions of species in the same location under changing environmental conditions (e.g., global change).

The validity of some of these recent applications has been subject to criticism on both theoretical and pragmatic grounds, with suggestions that the underlying assumptions of SDM invalidate their application for such purposes. While much of the current literature related to use of SDM focuses on improving the mechanics of fitting an SDM, e.g., through the use of novel and more powerful statistical methods, in this paper we explore instead the underlying assumptions of SDM and typical features of the data on which they are based. We then review the implications of these underlying assumptions and features, using this as a basis to indicate not only where predictions from SDM are most likely to be robust, but also where they are most likely to fail to correctly predict.

Take-home message: Our current preoccupation with model fitting methods often misses consideration of the much wider issues that determine the robustness of predictions derived from species distribution models (SDM). Here we argue that robust use of SDM for prediction requires not only a good statistical model fitting technique. It also depends on a wider set of contextual factors that include (i) robust choice of functional predictors from a very large set of inter-correlated candidates; (ii) management of the implications of correlations between predictors; (iii) recognition of the crucial role of competitive context; and (iv) awareness of variation in the degree to which the available observations support the fitted model through environmental space. The implications for best practice will be discussed.

Appendix 4: Testing adaptations to statistical species distribution models for species not at equilibrium

Jane Elith and Michael Kearney

In work started before the working group, progressed during it, and now ongoing, we are testing alternative methods for modelling species invasions, using Cane toad data (see elsewhere in this report) as a case study. Our intention is to submit it to an international journal by June 2008, and we will forward a copy when complete. We are using several modelling techniques and several ways of dealing with the data, and have found substantial impacts on models of current distribution and on forecasts for distributions under climate change. Some indicative maps are shown below. The methods are have potential for application in government applications.

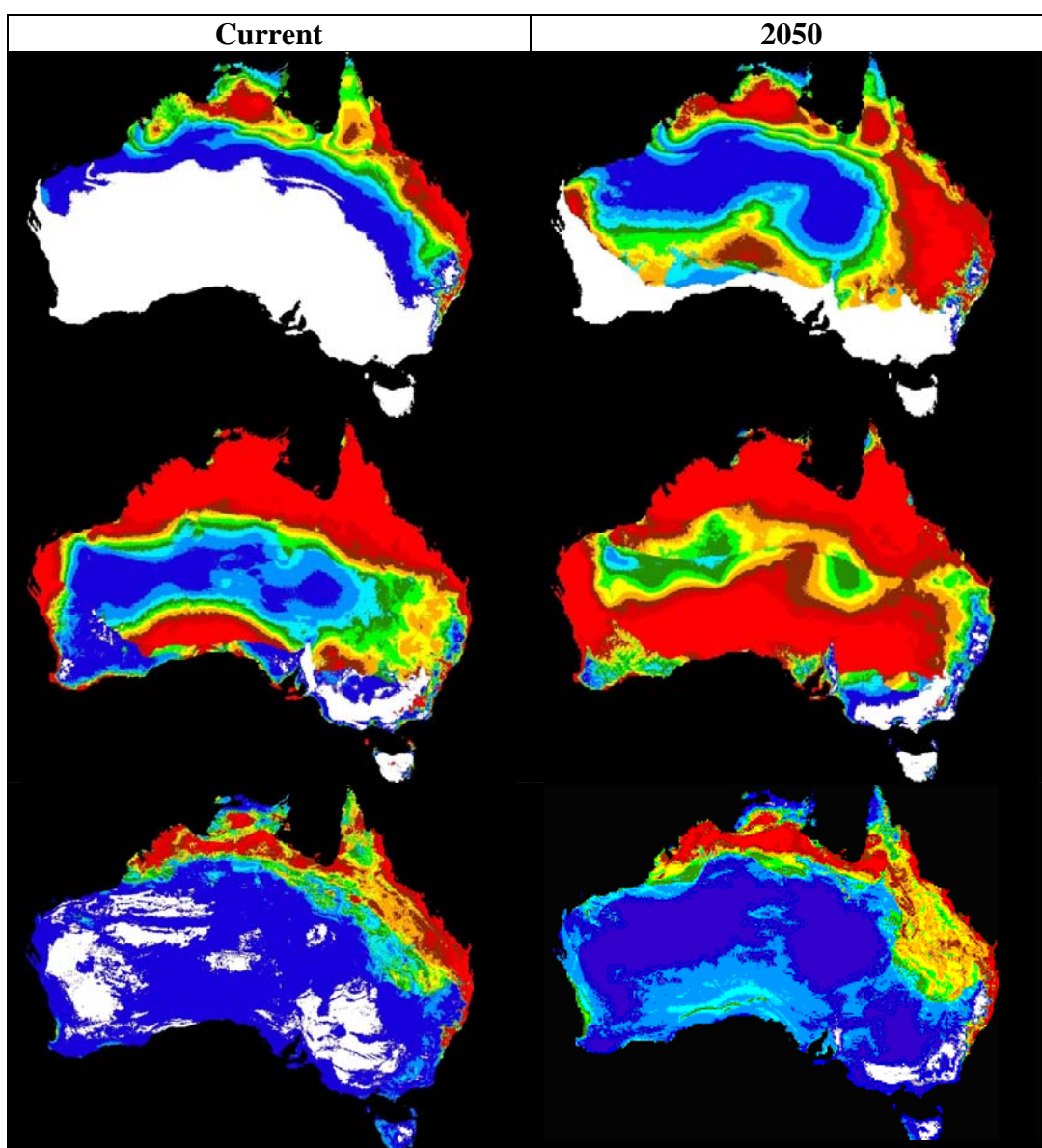


Figure A4: Predictions of the potential distribution of the Cane toad under current climate (left) and future (right) using three different modelling methods and modifications to the modelling data. Red is high suitability to blue (low) to white (unsuitable). The take-home message is that the modelling approach has a substantial impact on the result.

Appendix 5: Using geographic predictors in models of species invasions

David Warton, Simon Ferrier, Matt Wand, John Leathwick.

The focus of our sub-group was to explore methods of species distribution modelling when the species is at disequilibrium, i.e. it is found in only a part of the geographical range over which it is believed that environmental conditions are suitable for its occurrence. An example is an invasive species during early stages of invasion. The purpose of modelling is (i) to understand which environmental factors are important to the invasive species (ii) to predict the final distribution of the invasive species based on existing distributional information.

A particular method was proposed by Simon Ferrier for species distribution modelling in this situation - including "geographic variables" (i.e. latitude and longitude) as predictor variables, such that species absences due to distance from invasion source can be distinguished from absences due to unsuitable environments. One can then predict the final distribution of a species by changing the values of geographic predictors to essentially remove geography from the final model, and hence obtain predictions of species distribution based solely on environmental suitability. In a generalised additive modelling (GAM) framework, the idea of incorporating geographical space as a bivariate smooth surface within a GAM framework was inspired largely by Trevor Hastie's brief reference to this possibility in his GAM chapter of the *Statistical Model in S* book (1992), and applied routinely by some in the species-modelling community as far back as the mid-1990s (Ferrier et al 2002). More recently comparable approaches have been described as "geoadditive models" by Kammann & Wand, and these will be used in this study. The working group has developed a novel approach to obtaining predictions from the final model to disentangle the effects of environment and geography.

We would like to evaluate this proposed approach and compare it to alternatives (e.g. ignoring geography in model-fitting). To do this we propose a simulation approach. All working group members discussed simulation design, and what attributes of invasion we would like to modify in simulations (e.g. single vs multiple start points for invasion, nature of expansion of the invasion front). We decided to base simulations on real data so that our focus remained close to end-user applications (see next report).

David Warton and Matt Wand were responsible for (i) developing code for data pre-processing and analysis, (ii) fitting geoadditive models to a subsample of data, and obtaining predictions of the final distribution of invasive species (as proposed by Simon Ferrier) (iii) producing statistics that measure predictive power of the fitted model(s).

By the close of the workshop, Simon Barry had generated output from a couple of different simulation scenarios. David Warton and Matt Wand had imported one such dataset into R, subsampled for analysis, fitted geoadditive models, and obtained predictions of the final distribution of invasive species, based on information from part of the species range.

In future work, David Warton will finish producing statistics that measure predictive power of the fitted models (stage iii above). The simulation and analysis code will then be used (Leathwick?) to analyse data under different simulation settings and for different model fits

in order to draw more general conclusions about the efficacy of geosadditive models and related methods as tools for modelling species distributions under disequilibrium.

References

Chambers, J. M., and T. Hastie. 1992. *Statistical models in S*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, Calif., USA.

Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modelling. *Biodiversity and Conservation* **11**:2275-2307.

Kammann, E. E., and M. P. Wand. 2003. Geosadditive models. *Applied Statistics* **52**:1-18.

Appendix 6: Integrating species distribution modelling and invasion dispersion

Simon C. Barry and Jane Elith

This below is the beginning of a paper to be submitted to an international journal that will describe the program written to simulate data, and provide the code for general use. We will include a case study. To be submitted by June 2008.

1. Introduction

In this short note we describe software that we have developed to simulate the spread of a species through the environment. The novelty of the approach that we describe is that it uses as input standard output from regression based species distribution modelling. This we argue gives additional realism to scenarios produced thus leading to more reliable inference.

There are numerous abstract simulations of weed spread using diverse techniques such as cellular automata and continuous space kernels. These approaches have their place in the development of abstract theory but we were concerned that the models be as close to an observed reality as possible. We therefore considered using the predictive model from a species that was in approximate equilibrium with the environment and a spread kernel to produce a model that would be consistent with the species distribution but show realistic spread dynamics. In Section 2 we describe the algorithm. In Section 3 we consider an application of the approach. In Section 4 we discuss other approaches and applications of the technique.

2. Method

We consider the standard data used in species prediction modelling. We have an array of cells (pixels) with r rows and c columns. We start with the standard output of quality regression approaches. This is typically a probability p_{ij} where i denotes the row and j denotes the column. These probabilities are bounded by zero and one and can be interpreted empirically as the proportion of sites with similar environmental conditions (as defined by the terms in the model) that are occupied. The simulation approach also needs a kernel function which defines how the species disperses from occupied locations. This kernel is parametrised as a two column table with each row defining a dispersion event. The types of entries that can be in each row is given in Table 1 (Fig. A6.1).

column 1	column 2	Description
x	y	Attempt to disperse by offset of x rows and y columns
-9999	k	Disperse in a random direction as an exponential random variable with parameter k
-9998	k	Disperse in a random direction as an uniform random variable with parameter k

TABLE 1. Description of kernel row entries.

-1	-1
0	-1
1	-1
1	0
1	1
0	1
-1	1
-1	0

TABLE 2. Kernel table for all opposite and diagonal neighbours of a cell in the grid. This starts from top left and goes clockwise if the coordinates are the usual cartesian orientation.

Figure A6.1: Tables to be part of the manuscript

The kernel table is built up as a multiple number of rows. For example to have dispersal to each of the opposite and diagonal members of a cell the kernel table would be that given in Table 2 (Fig. A6.1). The simulation starts by defining a starting pixel(s). The algorithm then proceeds as follows: (1) Iterate through occupied sites (2) Iterate through dispersal table. (3) For each dispersal event (ie row) occupy new site with probability. An example of the output is provided in Fig. A6.2.

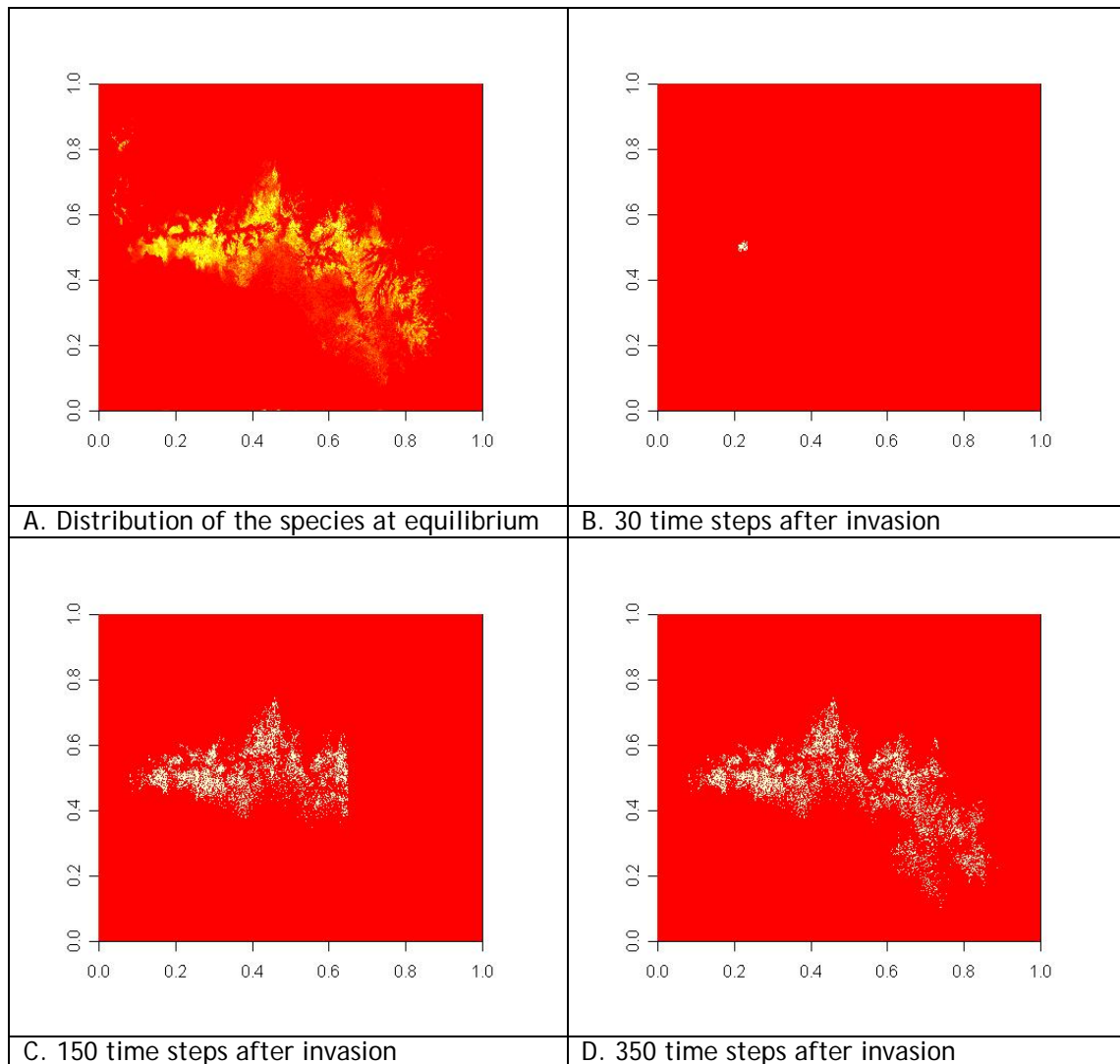


Figure A6.2: Examples of the data produced from the simulator