Spatial dynamics of anthropogenically altered dispersal patterns

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Ecology

at Massey University, Albany, New Zealand.

Luca Bütikofer, 2017

Abstract

The thesis put forward in this study is that anthropogenic disruption of natural dispersal is of central importance for the conservation of biodiversity. The main rationale being that dispersal, as a fundamental life history trait, is directly linked to, and affected by three of the main threats to biodiversity: Invasive species, habitat fragmentation and climate change.

In terms of biological invasions, the introduction of exotic taxa into a foreign habitat is performed by people, and the spread of these species within the invasion range is often a result of human aided dispersal. Surprisingly though, not much emphasis is put on anthropogenic dispersal when modelling biological invasions.

Habitat fragmentation is the direct consequence of habitat destruction. In a fragmented landscape, remnant natural patches host populations with varying degrees of inter-connectivity. Dispersal through degraded habitat is the key to maintaining viable meta-populations in a fragmented landscape. To facilitate dispersal in highly fragmented landscapes such as cities, a clear knowledge of the roles played by different land cover features (e.g. density of trees or traffic) is the starting point to make informed urban planning decisions.

The main issue with climate change is not that climate conditions are disappearing, but that they are moving. As a consequence of this, organisms also need to move along with the climatic conditions they are adapted to. This process, named range shift, is being observed worldwide on several taxa. Researchers, however, put very little attention to the problems arising when "climate migrants" encounter dispersal barriers (e.g. anthropogenic land cover) on their path.

The main body of work presented here consists in both the practical analysis of specific systems and the development of conceptual and methodological frameworks to address general issues. Among the former, I used population and landscape genetics approaches to analyse the case of Copper skinks (*Oligosoma aeneum*) in Auckland (New Zealand) as a recent and intensive habitat fragmentation scenario, with the goal of detecting the land cover types that allow the highest dispersal in urban settings. I also analysed the case of the Australian *Litoria* frogs (*L. aurea*, *L. raniformis* and *L. ewingii*) in New Zealand with the intent of quantifying the anthropogenic dispersal component of their invasion. As per the latter, I developed a novel method for estimating the anthropogenic component of biological invasions, infer their natural dispersal parameters, and forecast future developments of biological invasions. I also developed the novel concept of C-trap, whereby the shape and spatial orientation of the interface between natural and anthropogenic land cover types may originate traps for climate migrants, and used it to determine where, on a global scale, their high densities can further threaten endangered, endemic animal species.

My results identified an early stage fragmentation scenario for urban Copper skinks in Auckland. Relatively high population genetic structure formed as a result of segregation by a motorway. However, the roles of other land cover types in controlling population connectivity could not be determined. Of all *Litoria* frog populations recorded in New Zealand, about 30% were found to be of anthropogenic origin. Even though the borders of the distribution range occupied so far are unlikely to expand, the density of populations within the range is expected to rise with new, suitable patches being found. Tests on virtual species with the novel methodology for the modelling of biological invasions showed good performance (accuracy in estimating natural dispersal kernel and anthropogenic contribution), also in the presence of challenging limitations in the input

dataset. Finally, the global scale application of the C-traps concept to endangered, terrestrial species highlighted the high potential for conservation issues among climate migrants in Eastern Europe and Southern Asia.

Although it is challenging to quantify the consequences for biodiversity conservation of anthropogenic alterations to natural dispersal, results from the applied studies confirm the initial thesis on their critical role as biodiversity threats; while the novel concepts and methodologies provide a valuable tool for better incorporating these aspects in ecological studies and conservation management.

Acknowledgements

I would like to express my gratitude to the many people and associations that supported me, and the research I carried out during the years of my doctoral studies. Firstly, I wish to extend my thanks to Massey university and INMS for the invaluable financial support provided in the form of scholarships and research funds. I also wish to thank Forest and Bird, the Society for the Research on Amphibians and Reptiles in New Zealand and Dianne Gleeson for their generous contributions to the research presented in chapter three. Secondly, several people supported my research in many ways: primarily, my thanks are due to my main supervisor Weihong Ji, whose guidance and wisdom I much appreciated in innumerable occasions, both at a professional and personal level. In addition to Weihong, I was very fortunate to have a very big team of co-supervisors and collaborators, who came to my rescue in several critical situations. Even though my thesis is written in the first-person singular, it is undeniably an effort of many people. I can't thank enough Roberto Sacchi, Marco Mangiacotti and Stefano Scali for sharing their expertise on spatial ecology – contributing to chapters two and four – and for their welcoming nature and exceptional availability to be on very long Skype calls at unusual times. I am very obliged to Beatrix Jones, Daniel Welsh, (chapters two-a and two-b) Dianne Gleeson and Bernd Gruber (chapter three) who led me through the slippery territories of statistics, coding and genetic analyses. I also want to thank Manu Barry for her availability and prompt, concrete responses to all my queries, and for taking me on board Massey University's Reptiles Facility.

A PhD is a huge project to work on, mostly very enjoyable, at times very challenging. It demands plenty of time and attention, removing it from those who would deserve it the most. I offer my most sincere thanks to my family – papà, mamma, Marco, nonna and to

the most recent addition Shalini – for having put up with the most unpleasant version of myself and still offer me their support and closeness.

There has actually been a great deal of people around me who offered me their support. Not only they kept me sane; they comforted, inspired, nourished, amused me, and made me laugh: Ricky Giotto your closeness and support in spite of living at the antipodes, was that of a true brother, and I am very fortunate of have such a loyal friend as you. The New Zealand Sahaj Sanga, these people of all ages and backgrounds have been my home away from home, my other family, it would have been unreasonable to expect what I got from these uncles, aunts, brother and sisters; yet, for the whole length of my stay in New Zealand, their care has been an everyday reality. Peer students and researchers from the Ecology and Microbiology buildings, friends with whom to share ideas, jokes, field trips, lunches, and free time, thank you all for the happiness of your company. The people of Wandsworth, Chelsham Road, Anna and Aaron: yet another warm-hearted family with whom to share my time writing up this thesis, and another instance of unforeseen, extraordinary encouragement. Concluding, this section would be incomplete without a mention of my deepest gratitude to Shri Mataji Nirmala Devi, whose simple teachings reached the core of myself, ultimately inspiring also countless aspects of this work.

Table of contents

Chapter 1: Introduction.

Anthropogenic contribution to biodiversity loss

About four billion species are estimated to have inhabited the earth since its formation. Of those, about 99% have gone extinct (Novacek 2011). Extinction is a normal ecological phenomenon which is in dynamic equilibrium with speciation. When considering the fluctuating numbers of marine genera found in the fossil record of the Phanerozoic eon (542 million years ago (mya) to now), five mass extinction (or extinction events), can be identified (Table 1, Raup & Sepkoski 1982). The average number of extinctions per time unit during an extinction event is much higher than that in between them (background extinction rate). Similarly, extinction rates recorded in the last few decades are much higher – by up to 100 times – than the background extinction rate, even under the most conservative of estimations (Ceballos *et al.* 2015). Even though it is generally difficult to attribute ancient extinction events to specific causes, it is clear that the current (sixth) mass extinction is caused by anthropogenic activity (IPCC 2014).

	families		genera	
mass extinction	observed extinction (%)	calculated species loss (%)	observed extinction (%)	calculated species loss (%)
End-Ordovician (439 Ma)	26 ± 1.9	84 ± 7	60 ± 4.4	85 ± 3
Late Devonian (367 Ma)	22 ± 1.7	79 ± 9	57 ± 3.3	83 ± 4
End-Permian (245 Ma)	51 ± 2.3	95 ± 2	82 ± 3.8	95 ± 2
End-Triassic (208 Ma)	22 ± 2.2	79 ± 9	53 ± 4.4	80 ± 4
End-Cretaceous (65 Ma)	16 ± 1.5	70 ± 13	47 ± 4.1	76 ± 5

Table 1: Extinction intensities for the five mass extinctions of the Phanerozoic eon (Jablonski 1994).

There are many recognised anthropogenic threats to biodiversity (Salafsky *et al.* 2008). Among the most consequential ones, three have relevant links to anthropogenic alterations of natural dispersal: biological invasions, habitat fragmentation and climate

change induced range shifts. In brief, biological invasions occur when mankind amplifies the natural dispersal capabilities of a species; habitat fragmentation, on the contrary, causes a general reduction in natural dispersal; lastly, the case of range shifts is more complex as pressure to disperse is forced on species by climate change, while anthropogenic land covers prevent it from taking place. These three different scenarios, in which people interfere with the natural dispersal of species, are the focus of this thesis. For the rest of this chapter I will start by briefly discussing natural dispersal from an eco-evolutionary perspective. I will then elaborate on the anthropogenic alterations of dispersal mentioned above, introducing methodologies and approaches to investigate such phenomena, with an emphasis on those used in the following chapters. Research gaps will finally be presented together with the relevant questions, and the systems used to address them.

Natural dispersal, eco-evolutionary perspectives

Dispersal can be defined as any movement of organisms that generates gene flow (Clobert *et al.* 2012). It typically occurs when dispersers move between site of birth and site of reproduction (natal dispersal) or between successive sites of reproduction (breeding dispersal). Dispersal is conventionally divided in three stages: emigration, transfer and immigration. Dispersal can also be classified as active or passive: active dispersal implies a certain level of control from the dispersing organism; while passive dispersal depends on external factors such as wind (anemochory), water flow (hydrochory), or movements of other animal species (zoochory). Even though there is a component of specialisation in this, the same organism can be subject to both types of dispersal – such as birds in a storm or amphibian larvae during floods.

Dispersal is a spatial phenomenon and, as such, it occurs on a landscape, which is typically characterised by a varying degree of habitat suitability for any given species. For practical reasons, it is easier to consider such a landscape as an unsuitable medium containing patches of suitable habitat. This is clearly a simplification of reality since habitat suitability is normally not discrete and patches' borders would be fuzzy. Moreover, this view limits dispersal and landscape connectivity as inter-patch movements, thus ignoring movements within the patch. This is also a problem of scale since fine scale patches can be clustered inside broader scale patches (like trees in a forest and forests on a mountain range). Even though studies where the landscape is modelled as a continuous variable exist (e.g. Cornell & Ovaskainen 2008), the simplification of the landscape in suitable patches in an unsuitable matrix is very convenient, as it allows us to address many dispersal-related problems in terms of patch quality, patch size and inter-patch distance (Benton & Bowler 2012). Patch quality and size dictating the patch carrying capacity, and inter-patch distance reflecting the landscape's connectivity, or fragmentation level.

As for many other life-history traits, dispersal depends on complex interactions between organisms and environment. For this reason, causality in dispersal is manifold, and its variability depends on several factors (Bowler & Benton 2005). The ultimate causes of dispersal generally referred to are kin selection, bet hedging, and seeking for a better habitat.

Kin selection can promote natal dispersal so that competition for resources occurs with non-related individuals rather than with siblings (Hamilton & May 1977). However, even if avoidance of kin competition can promote dispersal, it also has the opposite effect of reducing the benefits of kin cooperation. The balance between these two forces has been proposed as a determining factor in the formation of kin societies, so that the likelihood of cooperative behaviour would increase for situations where the cost of

dispersal is particularly high (Baglione 2003). Similarly, dispersal also reduces the chances of inbreeding and parent-offspring competition. In this latter case, behaviours of parental aggression (Clobert *et al.* 2001), territory bequeathal (Berteaux & Boutin 2000) and offspring movement (Matthysen *et al.* 2010, Bonte *et al.* 2007) are proved to promote natal dispersal.

Biological bet hedging refers to the increased likelihood of survival in adverse conditions at the cost of decreased fitness in favourable conditions (Philippi & Seger 1989). For example, in favourable conditions the seeds of a tree would immediately sprout; however, to hedge its bets, trees also produce some dormant seeds that would germinate, for instance, after a drought would have killed the freshly sprouted plants. Similarly, by dispersing one's offspring into different habitats, the variability of the expected offspring fitness increases. This is stronger when the stochastic variability in habitat quality of the overall landscape is higher (Den Boer 1968).

Variation in habitat quality over space and time can promote dispersal in order for organisms to reach more favourable patches. Habitat quality, in this context, can be seen as intrinsic – thus pertaining to all the factors affecting the patch carrying capacity – or it can be related to population dynamics (e.g. overcrowding or lack of mates) – which would decrease realised patch quality without modifications on the patch quality *per se*. Most of the research on the causes of dispersal has classically focussed on the former case (Bowler & Benton 2005). While the latter case mostly applies to small populations, where stochastic fluctuations of demographic parameters are more pronounced. In the extreme circumstances of ephemeral or successional habitats, where variability in habitat quality is such as to cause local habitat extinction, dispersal becomes a good strategy to cope with the expected degradation of the patch over time (Denno *et al.* 1996).

Dispersal data is typically summarised in what is known as a dispersal kernel: a probability density function describing the chance of a disperser to reach growing distances from its source. The probability of reaching very long distances is typically lower than that of short distances and, therefore, the typical shapes of a dispersal kernel has a positive skew, with its shape varying from negative exponential to Gaussian to leptokurtic (fat tailed) depending on the speed at which the chance of reaching long distances drops. Information in a dispersal kernel can be summarised in terms of Long and Short Distance Dispersal (LDD and SDD respectively). SDD is described by the left side of the curve, while LDD events are summarized by the function's tail (the fatter the tail the more LDD in the kernel). LDD events are rare and stochastic (Shigesada & Kawasaki 2002), normally generated by passive dispersal, also in actively dispersing species (Shigesada & Kawasaki 1997). Collection of LDD information in the field is more challenging than that of SDD since it requires more extensive surveys. However, LDD has been recognised as the main player in various ecological processes such as in the rate of spread of expanding/invading populations, in climate shifts, in fragmented habitats, and in the maintenance of metapopulations viability (Nathan *et al.* 2012). Natural propensities of dispersal, however, can change considerably in the presence of anthropogenic disturbance (Trakhtenbrot *et al.* 2005).

Anthropogenic alterations to dispersal

Dispersal, being a central element of population dynamics, is also crucial for the conservation of biodiversity (Macdonald & Johnson 2001). Among the main obstacles to our understanding of dispersal are a lack of empirical information (Reed 1999) and of experiments clarifying the causality of dispersal and the interplay of the different dispersal promoters (Clobert *et al.* 2001). To compensate the lack of empirical data, modelling is used to fill the gaps by identifying potential variables and processes most relevant to

population viability (Macdonald & Johnson 2001). In spite of these difficulties though, it is clear that dispersal is central to some of the most relevant biodiversity threats. As global climate changes, dispersal parameters of many taxa are challenged by their need to keep track of their shifting climates. This process is supposed to be particularly challenging for sessile species (e.g. plants and corals) which are required to disperse at comparatively high speeds (Clark 1998; Pearson 2006; Corlett & Westcott 2013). Additionally, climate change induced range shifts may render current reserves pointless by shifting the protected ecosystem outside their boundaries (Heller & Zavaleta 2009; Robillard *et al.* 2015). Habitat destruction and subsequent fragmentation is certainly having dramatic consequences to population connectivity, by depriving organisms with suitable land cover for dispersal. It can also affect population connectivity indirectly. For example the pesticide acephate (acetylphosphoramidothioic acid O, S-dimethyl ester) has been discovered to affect migratory orientation of the white-throated sparrow (*Zonotrichia albicollis*, Vyas *et al.* 1995). Furthermore, habitat selection based on conspecific attraction of dispersers can be tricked by the reduction of population densities caused by habitat degradation (Reed 1999).

Anthropogenically increased dispersal: the problem of Invasive species

Invasive species are human-introduced, non-native species expanding their distribution range into novel environments – typically at the expenses of native. Anthropochory, the transportation of species by people, is in fact at the very core of the problem of biological invasions. The application of the IUCN-CMP classification of biodiversity threats (Salafsky *et al.* 2008) to 737 conservation projects across the world shows that the most common threat to biodiversity is "invasive non-native/alien species".

Invasive species are indeed one of the most important threats to biodiversity worldwide (Shigesada & Kawasaki 1997).

Because of the increase in worldwide movements of people, many previously un-crossable barriers virtually fell causing a reduction in the isolation of different land masses. Steep increases of human transportation of alien species notably occurred at the end of the Middle Ages (transoceanic sailing) and at the beginning of the Industrial revolution (railways). However, a huge and much more recent boost in the introduction of non-native species occurred in the last few decades with the beginning of the era of globalisation (Hulme 2009).

Different introduced species are subject to different dispersal strategies; for example, stowaway species like the Norway rat (Moors *et al.* 1992) or the House mouse (Eileen & Baker 1994) have been moved unknowingly; pigs (Larson *et al.* 2007) and chickens (Liu *et al.* 2006) were distributed as a food source; freshwater fishes (Gozlan *et al.* 2010, Britton & Orsi 2012) have been translocated for recreational purposes; Red-eared turtles have been dispersed by pet traders (Cadi & Joly 2004); songbirds were introduced for the acclimatisation of European settlers in New Zealand, and so on. The common pattern is the anthropogenic nature of their increased dispersal. Once established in a new ecosystem, invasive species have various impacts on it: some act as new predators or grazers, others as pathogens, some hybridise with closely related native species and others compete with natives for resources. Whatever the strategy may be, invasive species have an enormous economic cost mainly as a result of crop losses, prevention, or spread control (Olson 2006). Table 2 summarizes the costs of invasive species estimated for several nations. Take the U.S.A. as an example: the costs related to invasive species for 2005 (including microbes) amounts to a total of 133.6 billion dollars.

Country	Type of Invasive		
	Plant	Animal	Microbial
Australia (in $AU)	4 billion[a]	491.5 million[b] (9 vertebrates) 703.9 million[c] (10 vertebrates, includes environmental costs)	
Canada (in $CAN)	38.21 million[d] (leafy spurge and knapweed)	101.3 million[d] (3 invertebrates) 14–16 million[e] (emerald ash borer)	1.5 million[d] (Dutch elm disease) 73.34 million[e] (potato wart fungus) 1,000,000[e] (BSE)
Germany (in €)	103 million[f] (8 species)	60.2 million[f] (6 species)	5 million[f] (Dutch elm disease)
New Zealand (in $NZ)	100 million[g]	270 million[h] (vertebrates) 2 billion[i] (invertebrates)	
United States (in $US)	34.5 billion[j]	59.4 billion[j]	39.7 billion[j]

Table 2: Annual economic impact of terrestrial invasive species in different nations (Olson 2006).

Three distinct steps are recognisable in biological invasions: introduction, establishment and spread (Elton 1958, Phillips *et al.* 2008). The first step is always performed by people, while the last is almost always supported by people (Hodkinson & Thompson 1997); the second, being mostly related to population growth, has little to do with dispersal. Many studies have addressed the first step, for example, by trying to define what makes a species an invasive one (Kolar & Lodge 2001) or by correlating introductions to main cargo ship routs (Levine & D'Antonio 2007, Holeck et al. 2004). The dispersal occurring before the actual introduction is better described by socioeconomics than ecology, while the spread step is the one that can be best explained by dispersal ecology.

Dispersal in biological invasions

As illustrated by Shigesada and Kawasaki (2002), three main dispersal strategies can be observed in biological invasions. When the disperser expands its range by SDD only,

the resulting expanding range can be approximated to a single circle of growing radius with the location of first introduction in its centre. On the contrary, when invaders spread by a mixture of SDD and LDD, in addition to the primary, previously described circle of growing radius, additional satellite colonies appear as a result of LDD. In this second scenario, the balance between LDD and SDD generates two types of spread. If the disperser heavily relies on SDD and LDD events are relatively close to the main colony, the satellite populations remain isolated for a short time only to soon coalesce into the parent population. On the other hand, when LDD generates satellite colonies far away from their parent – or when SDD is slow – smaller colonies stay independent for a long part of the invasion process.

In biological invasions, most of the spread is accomplished through LDD (Suarez *et al.* 2001); few individuals or propagules disperse for exceptionally long distances – often helped by rare natural phenomena such as floods, high winds, or anthropogenic dispersal – establishing new populations at the expansion front. Anthropogenic dispersal also fits into these LDD events, often spreading invaders over very long distances in very short times. For this reason, even if the natural dispersal scheme of an invasive species would be of the first described type, the human component can drastically change the dispersal scenario by increasing LDD rates.

The problem of dispersal in biological invasions is complex and several dispersal models have been used to address it (Shigesada & Kawasaki 1997; Lewis *et al.* 2013). Among these, the main questions asked by researchers are "what area is likely to be invaded?" and "at what speed will the invasion proceed?", thus dividing the task of modelling dispersal in biological invasions into a spatial and a temporal component.

Modelling the spatial component of biological invasions

Modelling the spatial component of biological invasions generally requires some sort of Habitat Suitability Model (HSM, Peterson 2003). HSMs use the correlation between the presence of a particular species at a given location, and the environmental features at the same location to model the species' distribution (Hirzel & Le Lay 2008; Gallien *et al.* 2010). The last decades have seen a steep increase in the development of new methods for modelling species distributions (Peterson 2006). One of the main differences among these methods is the type of data used to describe the current distribution. Presence/absence or density data, commonly derived from systematic surveys, is often modelled by generalised linear or additive models or ensembles of regression trees (random forests or boosted regression trees, Elith *et al.* 2011). However, the availability of this type of data is limited, often restricted in space and time and the collection of extensive long-term datasets is very resource-intensive. On the other hand, museums and national and international databases hold vast collections of presence-only data covering a wide range of temporal and geographical scales. These data collections are of great value for researchers and methods for the analysis of presence-only data have thus been developed (Dettmers & Bart 1999; Tsoar *et al.* 2007). In addition to presence localities, building HSMs typically requires the use of spatially explicit environmental variables (Peterson & Nakazawa 2008). These have to be selected in order to address the specific question the HSM is expected to answer (e.g. by covering the range of climatic variables expected to influence the distribution of the study species, or by choosing a spatial scale that reflects the home range of the study species). The development of Geographic Information Systems (GISs) and the availability of data from environmental survey agencies have resulted in an ample range of online geographic data providers. This gives researchers the chance to easily get access to appropriate sets of variables for the study species and its geographic distribution.

There are issues however in building HSMs on presence-only data for invasive species. HSMs assume the modelled species to be at equilibrium with the environment at the model training locations (Gallien *et al.* 2010). This assumption is mostly violated in virtually any biological invasion since the sighting locations in the invasion range do not represent a full realisation of the invaded ecological niche, especially at early stages of the invasion process. Elith *et al.* (2010) suggest five measures to mitigate the biases introduced by the violation of the species-environment equilibrium assumption. Among these, it is suggested that the modeller (1) controls for the location of novel habitat. In fact, sighting locations of an early stage invasion typically do not capture the full potential of the suitable habitat conditions, and training a model on such incomplete sampling results in unreliable projections on novel climatic conditions (i.e. combinations of environmental variables not yet experienced by the invader). Multivariate Environmental Similarity Surface (MESS) analysis is a tool that allows the detection of novel habitats and the environmental variables that fall outside the training range. Model predictions on the novel habitat detected by MESS analysis will require very careful consideration. (2) Another good practice is limiting the model complexity in order to avoid modelling artefacts (i.e. model over-fitting). Even though oversimplification is just as undesired as an overly complex model, the latter will erroneously consider every imperfection in the dataset as habitat choice. On the contrary, a smoother model response bypasses this problem. (3) Excluding background data (see appendix) that was not sampled in order not to include false absence data and (4) including mechanistic/physiological information when available. For example, Sullivan *et al.* (2012) proposed a method of differentially weighing the study area based on the probability of the study species dispersing to that location. The general pattern emerging from these points is that producing HSMs for invasive species in their invasive range is more challenging the earlier the invasion is, and

perhaps the only option for generating reliable HSMs of invading species in their invasion range, is to allow for some time to pass in order to gain access to more reliable raw data.

Modelling the temporal component of biological invasions

Addressing the temporal problem of a biological invasion means dealing with species dispersal. HSMs represent areas with high potential for colonisation. However, because of limitations in the invaders dispersal, certain areas may never get colonised or may be colonised only in many years' time. Engler & Guisan (2009) define "potentially *suitable* habitat" and "potentially *colonisable* habitat" as the suitable land that can be colonised in an unlimited dispersal scenario, and in the presence of dispersal limitations respectively. Contrary to the unrealistic, unlimited dispersal scenario, dispersal constraints are normal (e.g. gaps in suitable habitat, barriers, max dispersal distance and time to reach maturity in sessile species). The importance of including dispersal information when modelling range expansions has been realised by the scientific community (Engler *et al.* 2012), yet such implementation is not common. This is possibly due to the difficulty in retrieving dispersal information for the study species and to the lack of user friendly software (Engler *et al.* 2012).

Among the first attempts to model dispersal, reaction-diffusion models have been widely employed (Fisher 1937; Skellam 1951). These models use partial differential equations to describe isotropic dispersal (populations expanding with the same speed in any direction) in typically homogeneous habitats; in their basic formulation, migration rates are predicted by reproduction (population growth) and dispersal. Even though reaction-diffusion models can be adapted to heterogeneous habitats (e.g. Shigesada *et al.* 1985) they have been criticized for not considering LDD and therefore underestimating the rate of spread (Clark 1998). Another important factor in modelling dispersal is the scale of the phenomenon: for large populations, where a certain amount of stochasticity is

included, population-level models can be adequate; however, for smaller populations at lower densities, or at the outskirts of distribution ranges, individual-based models (IBMs) become a necessity. In IBMs, individuals are modelled in continuous space – or in a spatial grid – where each cell is either occupied or not. At the expense of higher computational time, IBMs have the advantage of including selection, adaptation, kin competition, demographic stochasticity and different intensities at the invasion front.

Dispersal often depends on environmental factors like wind and land cover which need to be taken into consideration by modellers. This is can be achieved with either a mechanistic or an empirical approach. In mechanistic models, multiple factors are parameterised in relation to their effects on dispersal; for example, dispersal capabilities and population growth rates are factors typically accounted for in mechanistic approaches (Phillips *et al.* 2008). However, the biggest challenge in estimating dispersal capabilities is the estimation of LDD due to its rare nature and consequent low detectability (Nathan *et al.* 2003, Driscoll *et al.* 2014). Moreover, both dispersal capabilities and population growth can vary considerably across time (e.g. due to climate change) and space (e.g. land cover type) during the invasion process (Walters *et al.* 2006), raising additional difficulties for extrapolation. Conversely, in empirical methods, dispersal is modelled in relation to certain variables (e.g. environmental variables). Similarly to HSMs, empirical approaches to model dispersal suffer from inaccuracy of extrapolation into novel habitats (Phillips *et al.* 2008) and caution is therefore required when projecting the modelled invasion process to the future or to novel habitat. Recently, Bayesian methods combining correlative methods with mechanistic parameters are being developed (Cook *et al.* 2007, Catterall *et al.* 2012). Although these recent methodologies are complex and computationally heavy, they represent a promising way forward in the forecast of biological invasions.

Dispersal can be measured in the field directly, for example with radio and satellite telemetry (Gillespie 2001; Webster *et al.* 2002), mark-recapture studies (Fujiwara *et al.* 2006) or by parental attribution through genetic analysis (Cain *et al.* 2000), or indirectly, where proxies for dispersal can be inferred from habitat occupancy data (Moilanen 2004; Risk *et al.* 2011), population genetics (Kranner *et al.* 2007; Hamrick & Trapnell 2011) or historic time series (Nathan *et al.* 2003; Kelly *et al.* 2014). Driscoll *et al.* (2014) found that the majority of published studies (22%) use occupancy data to infer dispersal, followed by expert opinion (14%) and theoretical models (9%) with the observed proportion being stable for the ten years previous to the study. Time series are a type of occupancy data where the spatial location of sightings is coupled with the time of first record. They can be ancient, for instance those for tree recolonization after glacial maxima (Clark 1998), or more recent, for example in records of biological invasions (Kelly *et al.* 2014). In these cases, dispersal is often modelled as nearest neighbour data, which is the minimum distance from each point and its closest other point. Potential problems with nearest neighbour data arise when there are time gaps in the survey data or when multiple dispersal modes are active simultaneously.

Since human dispersal, or anthropochory, has been recognised as the main reason behind biological invasions (Nathan 2006) – both in the first (introduction) and third (spread) phases – proper weight should be given to it. However, even though some authors addressed this topic in their research (Engler & Guisan 2009; Caley *et al.* 2015), it is mostly in the general form of LDD (see Chapter 2a) and dedicated methodology addressing the problem of anthropochory in biological invasions is not yet available. Being able to model biological invasions including human dispersal would deliver more realistic results, ultimately contributing to the management of one of the primary threats to biodiversity.

Anthropogenically reduced dispersal: the problem of habitat fragmentation

Contrarily to the case of biological invasions, people can act in an opposite way and drastically reduce the dispersal capabilities of species, especially LDD (Trakhtenbrot *et al.* 2005). This is the core problem of habitat fragmentation (Schtickzelle *et al.* 2006). Species are prevented from dispersing as they would need in order to maintain sufficient levels of genetic admixture. This reduction of dispersal occurs through the creation of barriers, which can be *hard* – completely preventing dispersal – or *soft* – allowing a reduced amount of dispersal to occur. In both cases they divide natural areas with zones of anthropogenic land cover.

Fragmentation transforms a single population in a continuous habitat, into sub-populations exploiting remnant patches of natural habitat. Dispersal is what allows the required gene flow among subpopulations to form and maintain a viable metapopulation. However, dispersal does not remain untouched by fragmentation. In the simple scenario of remnant patches of equal quality exploited by a "dumb" disperser (one that disperses entirely at random) a general reduction in dispersal is expected as a consequence of its increased cost (Schtickzelle *et al.* 2006). This affects the capacity of colonising available suitable patches and causes increments in kin competition and inbreeding. On the other hand, reduced dispersal also leads to enhanced local adaptation due to lower levels of dilution of locally adapted genes (Kawecki & Ebert 2004). The balance of dispersal and adaptation interplays with the pattern of habitat fragmentation in interesting ways. For instance, what has been described as the "paradox of Rockall", states that the most isolated patches can only be (counter-intuitively) colonised by the least dispersive taxa (Johannesson 1988). Indeed, high dispersal taxa can sporadically occupy the most isolated patches (high colonisation and subsequent local extinction) while low dispersal ones are

less likely to reach the same patch, but also more likely to adapt. In the presence of habitat fragmentation, extreme strategies are more likely to be selected for, with moderate dispersal being the worst strategy, as it has been observed for butterflies in fragmented British landscape (Thomas 2000). In the hypothetical case of three scenarios of a 50% habitat destruction of 1) 50% reduction in quality over the whole landscape, 2) complete loss of 50% of the landscape at random locations and 3) loss of half of the landscape in a continuous area, Dytham and Travis (2012) argue that the first scenario will cause an increase of dispersal rates due to increased kin interactions and stochasticity; the second scenario will cause a reduction in dispersal as a consequence of the increased inter-patch distance, and the third case will see no difference in dispersal rates.

The process of increasing the level of fragmentation in an originally natural area results in non-linear ecological consequences: studies on this phenomenon showed that few significant steps are recognisable (Baguette *et al.* 2012). 1) When the proportion of remnant natural land cover falls below 60%, the original natural habitat collapses and cannot be further considered a single unit. 2) When no more than 40% of the original habitat is left, there is a steep increase of the inter-patch distance. 3) The highest achievable number of remnant patches occurs when the proportion of natural habitat reaches down to 30%. As a consequence of this, dispersal in fragmented habitats also correlates with the proportion of natural land cover in a non-linear fashion. In fact, from an evolutionary standpoint, different dispersal strategies are selected by different levels of fragmentation and dispersal polymorphism is likely to evolve in landscapes with different fragmentation patterns (Baguette *et al.* 2012). At intermediate levels of fragmentation and patch persistence time, high emigration phenotypes are selected in clustered patches, while they are discarded in isolated populations (Travis & Dytham 1999). Similarly, low emigration phenotypes are selected only in high quality patches, while high emigration

ones are common to both low and high quality patches, indicating a generalist strategy in high dispersal phenotypes and a more specialised strategy in low dispersal ones (Mathias *et al.* 2001). With regard to immigration, simulation studies have shown that short-distance dispersal is selected for in clustered patches while both short and long distance dispersal are negatively correlated with increasing levels of fragmentation (Bonte *et al.* 2010).

The ecological significance of habitat fragmentation

Similarly to invasive species, habitat degradation is one of the biggest threats to biodiversity (Haddad et al. 2015). In general, habitat degradation refers to the spectrum of changes that gradually makes a natural ecosystem unsuitable to support its original biological community. These changes typically lead to habitat fragmentation, which transforms natural areas into smaller natural patches isolated from one another by areas of anthropogenically altered land cover. Habitat fragmentation is currently so widespread that half of the forests on earth do not have trees that lie more than 500m from their closest edge (Haddad et al. 2015).

An early framework used in the study of the effects of habitat fragmentation is the theory of island biogeography, for which remnant patches are seen as *islands* in a *sea* of anthropogenic land cover (MacArthur & Wilson 2015). Analysing habitat fragmentation through the loupe of island biogeography, however, makes it difficult to tell apart its effects in terms of fragmentation from those caused by habitat loss in itself (Fahrig 2003). In fact, the more pervasive effects of habitat loss seem to be a much stronger driver to biodiversity loss than decreasing fragment size and inter-patch distance. Observational and experimental studies however, have managed to highlight the complex and interactive effects that habitat fragmentation exerts on biodiversity, especially over long time periods (Haddad *et al.* 2015). Among the many ways in which habitat fragmentation

results in biodiversity loss (Fahrig 2003), the direct consequence of reducing species dispersal is of primary importance.

Modelling habitat fragmentation: the landscape genetics approach

Many approaches are available to measure habitat connectivity, for example field observations, radio- or Global Position System (GPS), tracking devices and experiments in captivity. However, molecular techniques have proven to be very useful tools for indirectly quantifying habitat connectivity through measures of gene flow. Studies combining molecular techniques with spatial analysis have formed a research field named landscape genetics. Here, methodologies from various disciplines like landscape ecology, spatial statistics and population genetics are used to understand the spatial characteristics of genetic variability (Storfer et al., 2010). This interdisciplinary field evolved as a consequence of the recent advances in molecular techniques and of the ever-increasing availability of high definition and broad scale spatial data. The term "landscape genetics" has been devised by Manel et al. in 2003 and, since then, scientists have prolifically contributed to the literature of this discipline, new protocols have been developed and applied on different taxa, geographic locations and research questions (see for reviews Holderegger & Wagner 2006, Storfer et al., 2007, Holderegger & Wagner, 2008). Questions generally addressed in the field of landscape genetics include the identification of barriers to dispersal or planning of dispersal corridors (Epps *et al.* 2007), the description of source-sink dynamics (Murphy *et al.* 2010), the identification of dispersal routes (Wang *et al.* 2009), the gaining of evolutionary insight (Joost *et al.* 2007) and the prediction of the spread of introduced species and diseases (Biek & Real 2010).

Currently, there is no unified protocol for landscape genetics analysis and different approaches have been used to link genetic information with spatial data (Wagner & Fortin 2012). One of the most widely employed methods is the production of resistance surfaces.

Landscape connectivity is a fundamental parameter controlling the viability of spatially structured populations. It is a twofold property of the landscape: the shape, size and position of landscape features defines what is called structural connectivity; whereas the way individuals respond to structural connectivity (e.g. behaviour patterns, mortality, etc.) defines what is called functional connectivity (Stevens *et al.* 2006). In order to understand how different landscape features influence its connectivity, a resistance surface can be computed. Resistance surfaces consist of raster maps where every pixel has a value indicating its permeability for the dispersal of a specific taxa. The assignment of resistance values to different landscape features typically requires either field data, expert opinion or model optimisation (Spear et al., 2010). In the first instance, dispersal information on different land cover types is normally gathered through independent observational studies. For the controversial expert opinion approach, land cover types are given relative resistance values based on the subjective views of people and institutions with expertise on the study species. Lastly, model optimisation attributes multiple resistance values to each land cover type and choses the set of values that better explains the landscape-genetics correlation.

Landscape genetics in the suburbs

Urbanisation is a rapidly spreading phenomenon. Since 2007, for the first time in history, more than half of the world's population lives in cities (United Nations 2014). In spite of this, little is known about the dispersal dynamics of such an environment. Published studies in the field of urban ecology mainly deal with birds and mammals, while those considering reptiles and amphibians are few (Adams 2005). Urban ecology studies identified a general degradation of ground cover, entrapment in window walls, increased predation, collection, and pollution as factors affecting herpetofauna in urban areas. The

majority of studies that deal with urban herpetofauna, however, tend to record losses of species from urban areas without causal attribution (Germaine & Wakeling 2001).

From a conservation angle, the typical response of ecosystems to urbanisation is the general reduction of biodiversity with a relative increase in density of the most urban-adaptable taxa. Specialist species are typically more affected than others, whilst generalists and introduced species often do well in urban areas. Current literature reports instances of increased biodiversity under moderate disturbance levels in partially developed areas, but it is important to distinguish between local and regional diversity: local diversity typically increases in low-level development areas because of the appearance of common species with generally stable or increasing populations; whilst regional diversity decreases because of the loss of species with already declining populations (Adams, 2005).

Storfer et at. (2010), in their review of landscape genetics work, have estimated the proportion of landscape ecology studies conducted in urban environments to be 7% of the published literature (a total of only 12 publications out of 173). It is a small number compared to those conducted in temperate forests (40%), meadows/shrub areas (28%) or freshwaters (21%) and it represents a major knowledge gap when considering the rate at with urbanisation is spreading worldwide and its ecological consequences. It is indeed crucial to better understand the dynamics of habitat fragmentation in highly fragmented landscapes such as cities. Even though ecological corridors are a possible compensatory measure to attenuate the effects of habitat fragmentation in suburbs, making the very spatial structure of a city permeable enough to dispersal without the introduction of corridors would be a preferable approach with benefits for the whole urban area, rather than just for the suburbs connected by the corridor.

Anthropogenic dispersal traps: the problem of climate change induced range shifts

In addition to invasive species and habitat fragmentation there is a third way in which anthropogenically altered dispersal affects biodiversity. Here the problems of dispersal barriers and increased dispersal interplay under the influence of climate change.

Climatic parameters are currently changing in many complex ways, and as a consequence of this, earth's climates are shifting their geographical locations. Species depend on certain climatic conditions for their survival, and if these conditions move, so must do the species that depend on them. The shifting of distribution ranges caused by such reasons is generally referred to as range shift (Parmesan 2006; Lenoir & Svenning 2015). However, if a dispersal barrier is found along the path of a shifting range, it will stop the dispersal of the biotic component (Robillard *et al.* 2015) ultimately causing a separation between a species and its ecological niche, eventually leading to the extinction of such taxa. This specific "climatic trap" acts by forcing species dispersal in the first place (through range shifts), then preventing it from happening due to a dispersal barrier.

The environmental significance of climate change

The United Nations Framework Convention on Climate Change, in its "Article 1" defines climate change as "a change of climate which is attributed directly or indirectly to human activity" (United Nations 1992) in addition to natural climate variability. However, climate change can also be defined as "a change in the state of the climate (…) that persists for an extended period, typically decades or longer" (as it is done by the Intergovernmental Panel on Climate Change (IPCC, 2014)) without necessarily attributing its causes to anthropogenic activity. The IPCC further defines climate variability as "*variations in the mean state and other statistics (…) of the climate on all spatial and*

temporal scales beyond that of individual weather events", therefore distinguishing climate change and climate variability by the time frame at which they occur.

There is strong evidence for the attribution of contemporary climate change to anthropogenic activities. Since the industrial revolution, due to population and economic growth, humans have increasingly released fossil carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) in the atmosphere. Current atmospheric concentrations of these so called "greenhouse gasses" are the highest in the last 800,000 years. This is extremely likely to have caused climate warming since the mid 20th century (IPCC 2014).

From 1880 to 2012, the average earth and ocean surface temperature has risen by 0.85 [0.65 to 1.06] °C, with the last three decades being increasingly warmer than any other one since 1850. The vast majority (90% from 1971 to 2010) of the energy increase is stored in the oceans, while only 1% is stored in the atmosphere. Oceans are also getting more acid because of the increasing dissolution of atmospheric CO_2; the PH of surface oceanic water since the industrial revolution decreased by 0.1 (26% increase in acidity). Higher salinity zones are getting more saline and lower salinity areas are getting fresher. Worldwide permafrost temperatures have been raising and glaciers are shrinking causing the mean sea level to rise (0.19m from 1901 to 2010). Precipitation patterns are also changing, the average precipitation over mid latitude land in the northern hemisphere has increased since 1901. Extreme weather events patterns also have changed: there is evidence that extreme cold temperatures are decreasing while warm temperature extremes have increased, as well as extreme high sea levels and heavy precipitation events (IPCC 2014).

Dispersal under climate shift scenarios

A rich fossil record dating back to the beginning of the Quaternary period (~2.5 mya till present) is available for scientists to study range shifts of the past. During this period, several fluctuations of the earth's temperature have occurred in cycles of about 100,000 years. Common patterns are recognizable from the analysis of the Quaternary fossil record (Galliard *et al.* 2012). 1) There is no evidence of synchronized range shifts and species responded differently to climate change possibly due to differences in climate tolerance, influence of geography and differences in dispersal capabilities. 2) Spread rates measured from pollen records are higher than estimations from current species, possibly because of the difficulty in measuring LDD from contemporary data and the effects of species persistence in unknown refugia of the past. 3) Range shifts have not been a continuous phenomenon but were rather characterised by latent stages interrupted by jump dispersal. 4) In spite of its drastic fluctuations in temperature, the Quaternary period is characterized by relatively low extinction rates, confirming that most species must be able to withstand climate fluctuations (Galliard *et al.* 2012).

Even though the speeds of climate fluctuations in the Quaternary are on average slower than the current climate change, there have also been instances of faster changes (Brewer *et al.* 2002). Dispersal, in addition to tolerance, is the main life-history trait allowing species to survive in the face of climate change, and range shifts are the manifestation of such dispersal.

Similarly to climate changes of the Quaternary period, range shifts are the most likely response for current climate change. In their study on plants, Corlett and Westcott (2013), state that range movements observed in the last decades seem to have tracked climate warming, at most, just partially. It is difficult however, to tell whether this phenomenon is caused by a lack of need of movement in response to a relatively small change in climatic

parameters, a delayed response, dependencies on other parameters (e.g. precipitation, geology or other species) or an actual lack of ability to track climate velocities (Corlett & Westcott 2013). On the contrary, Parmesan and Yohe in their long-term, multi-taxa (more than 1,700 species), global meta-analysis (2003) concluded that range shifts are very common (especially in the oceans) and with speeds averaging 6.1 km per decade towards the poles. Similarly, Mason *et al.* (2015) recorded a poleward shift of the leading-edge (i.e. cool) range margin for 1573 species from 21 animal groups in Great Britain over the past four decades. Latitudinal and altitudinal range shifts have also been well documented in literature (e.g. Parolo & Rossi 2008), even though most assessments are unidirectional (either poleward or upward) and lacking in multidirectional observations (Lenoir & Svenning 2015).

Modelling climate shifts

Climate is a complex system with many interacting components. Climate warming, the global increase in temperatures, is expected to generate a shift in the isolines of temperature (geographic lines connecting areas with the same temperature) towards the poles, higher altitudes and higher depths (for oceans). Other components of climate, however, (e.g. precipitation, winds and seasonality) may shift in different directions. When considering climate as a whole, numerous outcomes are possible with regards to climate change. Climate shift occurs where all components of a local climate shift in the same direction and with the same speed. However, if different components shift with different speeds and/or directions, the extent of a local climate may reduce or increase depending on the behaviour of such components in neighbouring areas. Two extreme scenarios are possible in such circumstances: novel and disappearing climates (Williams *et al.* 2007). Novel climates are new combinations of climatic variables that are not currently in existence, while their counterparts are climates that will not be found in the future (like for

species extinction, climates can disappear locally and still be present elsewhere, or disappear globally). Climate velocity is a useful tool to describe the direction and speed of shifting climates (Loarie *et al.* 2009). It is a metric obtained by dividing the forecasted temporal rate of temperature change (the future temperature for a grid cell minus the current one [ºC/year]) by the spatial rate of climate variability (the maximum difference between a cell and its neighbours [ºC/km]). The resulting measure, expressed in km/year, represents the speed at which isotherms shift for a specific grid cell. Burrows *et al.* (2014), in their study on climate warming (temperature only), classified the earth's surface in relation to characteristic behaviours of local climate velocities. *Absolute climate sinks* are areas where climate velocities converge or are blocked by coastlines; *climate sources* are areas where no climate velocity ends; *corridors* are areas with high numbers of velocities passing through; *divergence cells* see more velocities starting than ending, and vice-versa for *convergence cells*. Climate velocity as per Loarie (2009) has the advantage of being a simple measure of straightforward interpretation. However, climate velocities computed for different variables (e.g. temperature and precipitation) can have different directions and magnitudes, rendering very difficult to generate combined velocities. Additionally, artefacts may occur as a consequence of limited search radiuses, whereby velocities are overestimated on flat terrains and underestimated in cases of disappearing climates (e.g. on mountain tops, where high gradients generate low velocities but the local climate is disappearing). In addition to unidimensional (e.g. temperature only) methods, multivariate methods typically try to match the position of current climates (measured as the combination of two or more climatic variables) with their forecasted position in the future (Hamann *et al.* 2015). These more realistic approaches produce velocity vectors that vary in both orientation and speed compared to univariate methods, with orientations at times longitudinal or towards the equator (VanDerWal *et al.* 2013; Lenoir & Svenning 2015).

Modelling range shifts

When considering the effects of climate shifts on the biotic component of ecosystems, among the most important questions to answer is whether species can keep track with shifting climates. Since the speed at which mobile species can migrate is several orders of magnitude higher than that of shifting climates, this question primarily refers to passively dispersed organisms. Studies on the speed of plant migration after the last glacial maximum (Holocene) show speeds of about 1 km per year, possibly due to LDD and high latitude seed refugia (Pearson 2006). Estimated isotherms velocity average at 0.41 km per year, with minima at 0.08 km per year for tropical and subtropical coniferous forests and maxima at 1.26 km per year for flooded grassland (Loarie *et al.* 2009). With regards to temperature, montane ecosystems can act as refugia due to their slow climate velocities as a consequence of high spatial temperature gradients (difference in temperature per km). Generalising these findings, 71.2% of the globe would see plants keep up with local climate velocities (28.8% of the globe having velocities grater than 1 km per year (Loarie *et al.* 2009)).

Different areas can act in peculiar terms with regard to range shifts, affecting local biodiversity differently. Biodiversity in source areas is expected to shrink as there would be no taxa replacing those that leave. In *Converging cells*, biodiversity would increase and reshuffling of ecological interactions is expected to take place. *Absolute sinks* are where there will be high extinction potential unless fast adaptation would take place. *Corridors* – areas used by high numbers of climate migrants to follow their niche – will play a very important role for the conservation of biodiversity. Dispersal barriers, such as anthropogenically modified landscapes, can in fact easily transform corridors into absolute sinks. Despite the importance of such transformation though, land-use changes are typically not taken into consideration when modelling range shifts. The shape and spatial

orientation of the interface between natural and anthropogenic land cover types can originate traps for climate migrants. Primary importance should be given to further understanding the dynamics of those areas of high importance for climate migrants which coincide with intense anthropogenic habitat fragmentation, in order to provide reliable management directions for the preservation of biodiversity under climate change scenarios.

Research questions and expected outcomes

The research questions I aim to investigate in this thesis include:

How much does anthropogenic dispersal contribute to biological invasions? Can anthropogenic dispersal in biological invasions be modelled independently than natural dispersal? I expect anthropogenic dispersal to be central to biological invasions, especially with regard to LDD. I also expect it to be effectively distinguished from natural dispersal in a biological invasion time series due to its different mechanism of action. Consequently, I expect to be able to model it independently.

What are the key landscape features controlling dispersal in a highly fragmented (urban) landscape at a very fine scale? How can the overall structural connectivity of cities be incremented? I expect the density of natural elements such as trees, shrubs and gardens to increase connectivity (even if they are scattered and not organised in corridors) and the density of impervious surfaces to act in the opposite fashion.

What effects do anthropogenic land cover changes have on climate migrants? Are there particular species or areas of the world where these effects are particularly intense? I expect that, at specific locations, climate shifts would lead climate migrants against dispersal barriers that may prevent them from tracking their climatic niche. I also expect

these locations to be spatially clustered around areas of high climate velocity and high fragmentation levels.

Thesis structure

This thesis consists of six chapters. In chapter one (introduction) I examined the consequences in terms of biodiversity loss, of interfering with the natural dispersal of organisms. This was done by analysing three focal manifestations of the issue: biological invasions, habitat fragmentation, and range shifts. I reviewed the existing knowledge on these aspects and presented the aims, structure and objectives of the thesis.

Chapters 2a and 2b both deal with dispersal in biological invasions. In chapter 2a, I look into the possibilities of modelling biological invasions from invasion time series by accounting for anthropogenic dispersal. I do this by developing a novel methodology that allows us to estimate the natural dispersal kernel and the intensity of anthropochory in a biological invasion starting from its invasion time series. This constitutes an important contribution to the management of invasive species as it provides new and flexible means to better understand – and consequently manage – disparate invasion scenarios. These methodologies have been made freely available as an R package named *Biolinv* (https://cran.r-project.org/package=Biolinv), and chapter 2b describes this software, its functions and how to use it for an example analysis.

In chapter 3, I investigate the effects of reduced dispersal rates in a young and highly fragmented urban landscape. I aim to define the contribution of several landscape features to the permeability of urban areas and to the dispersal of land animals, using Copper skinks (*Oligosoma aeneum*) in Auckland's North Shore as a study system. First, I assess the degrees of isolation and structure patterns in the study species. Then, using a landscape genetics approach, I investigate the role that different urban land cover types

(e.g. % of grass, % of trees, roads width, housing densities, etc.) have on the dispersal of species between remnant natural patches. The significance of this goal lies in its application to urban planning, for example, in designing new suburbs with high permeability to animal dispersal, where populations in parks and reserves are granted the required gene flow for maintaining high viability.

In chapter 4 I explore the effects of anthropogenic land-use on the range shifts of climate migrants. I present the novel concept of C-traps, elaborate on its theoretical side, provide a simple method to spatially locate C-traps, and use such a method to determine where – at a global scale – their high densities can further threaten endangered, endemic, animal species. Climate change is considered among the most critical threats to biodiversity loss for the future (Heller & Zavaleta 2009). Understanding the dynamics it sets off on the ecological plane is of central importance for the prevention and mitigation of its effects.

In chapter 5 I summarise the main findings and draw the general conclusions from this study. I also summarise the contributions of this study to the fields of dispersal ecology and biodiversity conservation, and present potential future research araising from this thesis.

Chapter 2a: A new method for modelling biological invasions from early spread data accounting for anthropogenic dispersal.

Abstract

Biological invasions are one of the major causes of biodiversity loss worldwide. Even though human aided (anthropogenic) dispersal is the key element in the spread of invasive species, no framework published so far accounts for its peculiar characteristics, such as very rapid dispersal and independence from the existing species distribution.

Here I present a new method for modelling biological invasions using historical spatio-temporal records. This method first discriminates between data points of anthropogenic origin and those originating from natural dispersal, then estimates the natural dispersal kernel. I used the expectation-maximisation algorithm for the first step; I then used Rippley's K-function as a spatial similarity metric to estimate the dispersal kernel. This was done while accounting for habitat suitability and providing estimates of the inference precision.

Tests on simulated data showed good accuracy and precision for this method, even in the presence of challenging but realistic limitations of data in the invasion time series. These included gaps in the survey years and a low number of records. I also provided a real case application of my method by using the case of *Litoria* frogs in New Zealand.

This method is widely applicable across the field of biological invasions, epidemics and climate change induced range shifts, and provides a valuable contribution to the management of such issues. Functions to implement this modelling technique are made available as the R package Biolinv (https://cran.r-project.org/package=Biolinv).

Introduction

Biological invasions are increasingly common phenomena due to the intensification of transportation of both people and goods (Hodkinson & Thompson 1997). Human aided (anthropogenic) dispersal of invaders occurs at the initial stages of invasions (when species are introduced in new areas) and can persist during the invasion, with people acting as a dispersal vector during the colonisation of new areas.

The speed of natural dispersal compared to anthropogenic dispersal is very different. For instance, truck transportation of worm-snakes in potting mix is so much faster than the respective natural snake dispersal that it can be approximated to instantaneous colonisation of new localities. Despite this, when modelling biological invasions, anthropogenic dispersal is often attributed to long distance dispersal (LLD) and not modelled independently. In a typical dispersal kernel, the probability function asymptotically approaches zero with growing dispersal distances. However, anthropogenic dispersal may result in peaks at distances far beyond the maximum natural dispersal distance. Additionally, even though in some situations including anthropogenic dispersal in LDD may be accurate enough, in most cases the locality at which new introductions are made does not depend on the previously colonised locations. This is particularly true when the source of propagules is located in the native distribution range.

Recent modelling approaches that consider anthropogenic dispersal include Interacting Particle Systems (IPS) like in Engler & Guisan (2009) and Pitt *et al.* (2009) and Bayesian inference (Caley *et al.*, 2015; Catterall *et al.* 2012; Cook *et at.*, 2007). In both the IPS methods, dispersal is divided into Short Distance Dispersal (SDD) and LDD. The method of Pitt *et al.* has the four neighbouring cells of any occupied cell becoming themselves occupied at the subsequent time-step, leaving it to the grid scale to account for different SDD distances. The method of Engler *et al.* instead, uses a stepwise user defined

dispersal kernel to select the cells that will be colonised at the next time-step. In both methods LDD is modelled by sampling distances from occupied cells and is therefore dependent on the past colonisation. Engler *et al.* sample the LDD distance from a uniform distribution with user defined maximum at a user defined frequency. Pitt *et al.* sample the number of LDD events from a Poisson distribution with a user defined mean; the dispersal distances are then sampled from a Cauchy distribution. None of these methods provide tools to infer dispersal parameters. Cook *et al.* (2007) and the generalisation on their method presented in Catterall *et al.* (2012) adopt a Bayesian approach to infer the invasion dispersal kernel from an ongoing invasion time series. They do not consider LDD separately, and limit their dispersal kernel to the power-law and negative exponential probability functions respectively.

Here, I present a method for modelling biological invasions that accounts for natural and anthropogenic dispersal separately. Starting from existing spatio-temporal data of an ongoing biological invasion, my method initially discriminates points of natural origin from those originated through anthropochory. Based on this categorization, the natural sub-sample is used to infer the dispersal kernel. By estimating the proportion of anthropogenic introduction and the natural dispersal kernel, it is possible to make predictions for the future invasion process. This method accounts for habitat suitability and provides a measure of precision of the estimated dispersal kernel.

I tested my method on simulated data to assess whether the performance of all the steps of the process were accurate and reliable. I also tested the resilience of the method to variations in sample size, proportion of anthropogenic populations, and ecological niche width. A key feature is that the computations remain tractable even in the presence of gaps in data collection, and thus uncertainty about the exact time of the dispersal events.

I also used this method on the real case scenario of the introduction of three species of *Litoria* frogs from Australia to New Zealand: *L. aurea*, *L. raniformis* and *L. ewingii* have been introduced by acclimatisation societies in the 19th century and currently, *L. aurea* occupies the northern part of the North Island while *L. raniformis* and *L. ewingii* are found across the two main islands. *Litoria* frogs are explicitly excluded from the Wildlife Act of New Zealand by its fifth schedule but a limited protection is warranted them when residing within a conservation area. The Act also states that *Litoria* frogs cannot be released into freshwaters or exported from the country (Bishop 2008). Nonetheless, *Litoria* tadpoles are commonly sold as pets in New Zealand and are frequently traded online across the country. Unwanted *Litoria* adults are commonly released into freshwater bodies contributing to the spread of these species.

Sympatric populations of native and introduced frogs are common in New Zealand. Thurley & Bell (1994) recorded the occurrence of predation of *Leiopelma archeyi* by *Litoria aurea* in a sympatric population in the Whareorino forest. However, direct competition is normally considered unlikely because of the different microhabitat selection of the two genera. In contrast, indirect competition may occur through the dispersal of diseases by the more dynamic *Litoria* frogs. *Batrachochytrium dendrobatidis* has been found in all three *Litoria* species and in the native frog *Leiopelma archeyi* (Shaw *et al.* 2013). Currently, there is no evidence that *Litoria spp.* can infect *Leiopelma spp.* with chytrid. However, it is plausible that the more mobile and abundant *Litoria spp.* act in spreading chytrid (Bishop, 2008).

Methods

Given a dataset with spatial and temporal coordinates (time series) of an exotic species during a biological invasion, this algorithm can be used to answer two main questions:

What proportion of points of the invasion time series is of anthropogenic origin? What does the dispersal kernel of the naturally dispersed points look like?

Method for estimating the anthropogenic component

In order to discriminate between points of anthropogenic and natural origin, I used the expectation-maximisation (EM) algorithm (Dempster *et al.* 1977). My main assumptions were (1) that the subset of points of anthropogenic origin would be described by a Poisson point process, where the probability of having a new population of anthropogenic origin occurring at any spatial location does not depend on the distribution of the other populations. By contrast, (2) new natural populations were more likely to occur in proximity of other extant populations.

The EM algorithm was based on the nearest neighbour distances d_i: for each point in the time series the nearest neighbour distance with other points was computed. To avoid overestimating SDD and loosing track of human mediated dispersal events, I forced at least one nearest neighbour distance per year to be measured against points from past years. The dispersal kernel is then fit to the amended nearest neighbour distances.

Anthropogenic points were assumed to be uniformly distributed: to model their distance to the nearest existing point I computed the yearly probability density function for the nearest neighbour distance of a random point, $g_y(d)$ by measuring its nearest neighbour distance (10,000 replicates) with all the points in the time series till that year. A yearly approach was necessary since the increasing number of points in the time series made short distances more frequent as more populations accrued.

The nearest neighbour distribution for naturally dispersed points, $f(d, \sigma)$ was a single tail Gaussian distribution on positive numbers; an initial guess at the standard deviation was updated as the algorithm progressed. This distribution was a crude approximation of

the true nearest neighbour distribution, or indeed the nearest neighbour distribution implied by the dispersal kernel fitted later. Nevertheless, in simulations it worked well for categorizing natural vs anthropogenic points. The mixture distribution of nearest neighbour distances, $L_y(d)$ was then:

$$L_y(d) = (1-\pi)g_y(d) + \pi f(d,\sigma),$$

with π being the probability of being of natural origin. I used the EM approach (Dempster *et al.* 1977) to estimate σ, π, and a point specific probability of being of natural origin, W_i.

The EM algorithm started by using an initial guess for π and σ to estimate the W_i, based on the relative likelihood of the natural and anthropogenic distributions for its nearest neighbour distance

$$W_i = \frac{\pi f(d_i,\sigma)}{(1-\pi)g(d_i) + \pi f(d_i,\sigma)}$$

I then updated π:

$$\pi = \frac{\sum W_i}{n}$$

and σ:

$$\sigma = \sqrt{\frac{\sum(W_i {d_i}^2)}{\sum W_i}}$$

This process was iterated until the estimates of π and σ did not change more than 0.00001. This process was quite robust to the initial values of π and σ.

Dispersal probability distribution

In order to describe dispersal distance probability distributions I used the *One Dimensional Dispersal Kernel function* (Clark 1998), which is based on two variables: the shape parameter C and the distance parameter α:

$$f(x) = \frac{\mathrm{c}}{2\alpha\Gamma\left(1/c\right)} \exp\left(-\left|\frac{x}{\alpha}\right|^{c}\right)$$

Where Γ indicates the gamma function.

C accounts for different levels of kurtosis with the negative exponential and the Gaussian distributions being special cases (C=1 and C=2 respectively); C<1 makes for a fat tailed kernel. In order to choose which combinations of α and C to consider in my algorithm, I generated 496 single-generation datasets with all the combinations of α (regular sequence from 1 to 7 with 0.2 increment), C (regular sequence from 0.25 to 2.5 with 0.25 increment) and maximum dispersal distance (10). I then performed a K-means cluster analysis on the dissimilarity matrix obtained by computing the sum of squared differences of the Ripley's K-function (Ripley 1977) of all possible couples of datasets. The number of clusters was chosen by detecting a bend in the plot of the within-group sum of squares (y axis) over the number of clusters (x axis) (Appendix Fig. 6). This analysis highlighted four groups of inherently similar dispersal kernels (Figure 2 and Appendix Fig. 6). Within these groups, different α and C combinations could not be discriminated from one another. Therefore, I only estimated which cluster α and C belonged to rather than their actual values. The chosen representatives of each of these clusters are α = (10%, 15%, 25% and 50% of the maximum dispersal distance) and C= 2 (Gaussian distribution). Keeping C constant greatly reduced computation time while still representing of all the four identified clusters.

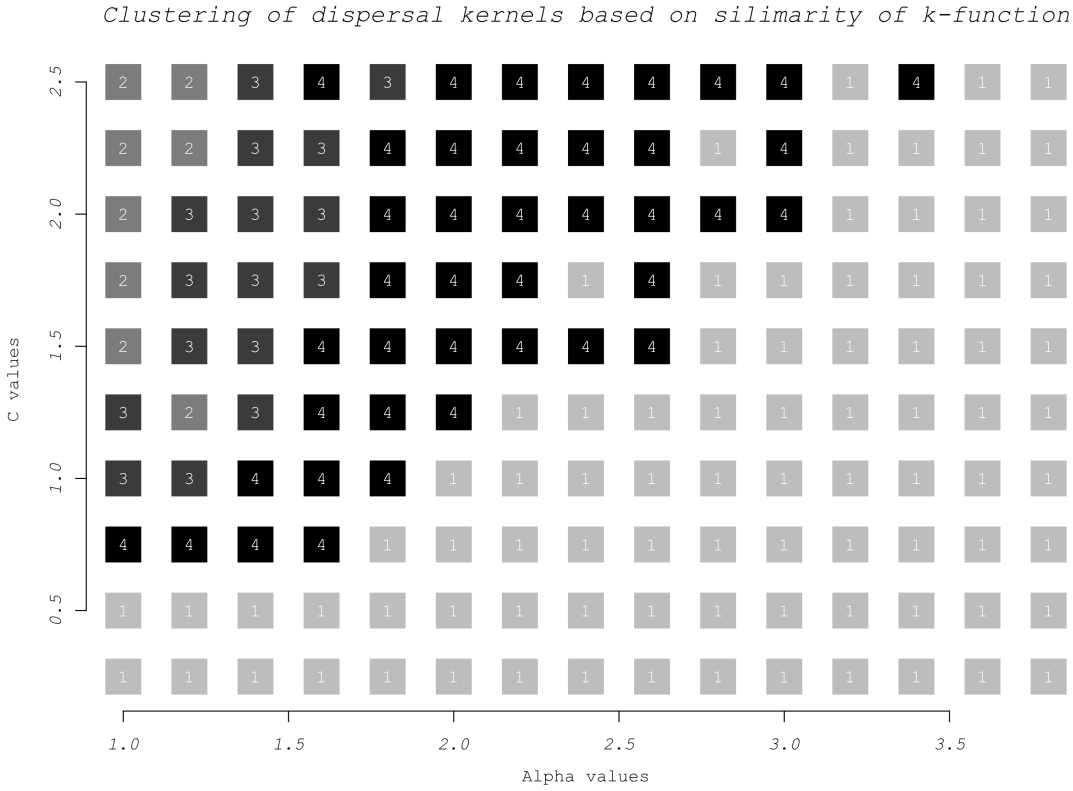


Figure 2: Clustering of dispersal kernels based on similarity of K-functions. Dispersal kernels produced with the One-dimensional dispersal kernel using different values of a and C cluster in four groups (x axis is clipped at 4; a>4 clusters in group 1). This image shows cluster membership of different couples of a and C values.

Method for estimating the natural dispersal kernel

The EM algorithm is capable of discriminating between natural and anthropogenic populations. However, the nearest neighbours do not necessarily correspond to the source population and for this reason it does not provide an accurate estimate of the dispersal distribution. Another important problem in modelling the dispersal process through nearest neighbour data of an opportunistic dataset, is the presence of gap years where no data has been collected (at least in some locations). I bypassed these problems by adopting a more phenomenological approach where I summarised the spatial properties of the final product of the invasion process with Ripley's K-function (Ripley 1977). This approach discards some of the temporal information of the invasion time series, but it allows

computationally efficient estimation of the dispersal kernel even in the presence of gap years.

To start, a threshold of 0.5 was chosen to classify points in the invasion time series as either of anthropogenic ($W_i < 0.5$) or natural ($W_i > 0.5$) origin. Then, the estimation of the dispersal kernel of the natural subset was performed by comparing the spatial structure of the observed data with a set of simulated data sets with varying dispersal kernels for the natural component. Many replicates were produced for each candidate kernel. The kernel for the set of replicates which was on average most similar to the original data, was considered as the best approximation of the true dispersal kernel.

As a measure of similarity, I used the total sum of squares between the Rippley's K-functions of the original time series and the one of the simulated ones. Rippley's K-function measures spatial aggregation and it reflects changes in the dispersal kernel of a time series. For each point in the data set the K-function counts the number of surrounding points within a series of growing search radiuses. In order to accurately estimate the natural dispersal kernel, points of anthropogenic origin did not have their positions simulated, but were copied as they are from the original to the simulated time series.

Simulated time series were generated sequentially, starting with a subset of the original time series representing the initial distribution (iteration zero) of the species at the year of introduction. For each iteration (year) recorded in the original time series, the points of anthropogenic origin were copied without any modification from the original time series. For that same year, the exact number of points of natural origin recorded in the original dataset was generated from the points which were already present from the previous generations in the virtual time series. Dispersal distances from the source point were sampled from the dispersal kernel being simulated. Points generated/copied in this

fashion were then added to the previous generation (initial distribution at the first cycle) and the process was repeated.

As a boundary for the point generations, I used extracted North and South Island shorelines from the CIA World Data Bank II (Becker & Wilks 1993, 1995), which are available through worldHires() command in R package *mapdata*. The same boundary was also used as *window* in the computation of the K-function, as available in package *spatstat* with Ripley's isotropic edge correction (Ripley 1988).

Habitat suitability

A species' ecological niche can influence aggregation in a time series: the narrower the niche, the more aggregated the points. In order to account for this phenomenon, it is possible to filter the newly generated natural points with a Habitat Suitability Map (or Model, HSM) where the new natural points are deleted with a probability corresponding to the inverse of the habitat suitability value of the HSM cell they fall into. This is done via an iterative process that ensures the number of simulated points matches the number of observed points in the data set. Each year an excess number of new points are generated and successively filtered as explained above. The number of points generated can be set in the algorithm as a multiple of the actual number of points for that year. After filtering, the remaining points are randomly sampled to generate the exact number needed for that year.

Jackknife resampling

Jackknife resampling can be used to reduce the effect of outlier points in the time series. This is particularly useful when analysing real world datasets, which come without replicates. Jackknifing can also provide a useful measure of uncertainty around the estimated α value for each dataset.

To test whether Jackknifing improved the performance of this algorithm, for two datasets (ten replicates) that performed poorly (α=4.5, % anthropogenic=10, n=100, no aggregation; α=15, % anthropogenic=30, n=400, with aggregation), a number of subsamples was taken, each containing 85% of the points in the original dataset. Points within the same subsample were sampled without replacement; however, the same point could be present in different subsamples. Each subsample was then processed as a normal dataset. The final α estimate was taken to be the average of the estimates of all subsamples derived from the same original dataset.

Algorithm testing

I tested the algorithm's accuracy and precision in estimating the proportion of points of anthropogenic origin and the parameters of the dispersal kernel by running it on simulated datasets. I simulated time series of virtual invasions with populations of anthropogenic origin being sampled randomly within New Zealand mainland, and naturally dispersed populations spreading from existing ones with a known dispersal kernel. The process of generating the virtual dataset was sequential: for each cycle (year) a precise number of naturally and anthropogenically dispersed populations was created. I generated two sets of virtual datasets, one for testing the EM algorithm and one to test the estimation of the dispersal kernel parameters.

In order to evaluate the classification obtained from the EM algorithm, I generated datasets with varying α values (4.5; 15), C values (0.3; 1; 2), proportion of anthropogenic points (0%; 10%; 30%). Each replicate of the virtual dataset was then duplicated and, for one copy, the year names were aggregated in groups of 5 (e.g. 1, 2, 3, …20 becomes 5, 5, 5, 5, 5, 10, 10, 10, 10, 10, 10, …, 20, 20, 20, 20, 20). This was done to simulate the periodic sampling effort of real world data collections. These datasets were only run

through the EM algorithm and all points were classified as either natural or anthropogenic. I then compared the EM-classified values against the original values to determine the proportion of points correctly classified. When the true proportion of anthropogenic points was non-zero, I also computed the ratio between the estimated and original proportions to assess over/under estimation.

To test the dispersal kernel estimation, I generated a set of datasets that varied in key attributes. For each parameter setting, ten replicates were generated. All combinations of the following were tried: α (4.5 or 15), total number of points (100, 400, 1000, 2000), percentage of points with anthropogenic origin (0, 0.05, 0.1, 0.3) and amount of aggregation (without aggregation, or aggregation into 5 year bins). In each case the number of generations (years) was (80), the maximum natural dispersal distance was 30km, and C=2. Every virtual dataset generated was then run through the algorithm and the quality of the α estimate was measured as the difference between the estimated and real value. α values used in the comparison datasets were 2; 3; 4.5; 7.5; 11; 15; 20; and 25; for each value 90 replicates were generated.

To test the effect of the HSM filtering I generated a virtual dataset where, for every cycle of the point generation algorithm, new points (anthropogenic and natural) were produced in higher quantity than needed and subsequently filtered with the HSM for *Litoria raniformis*. This virtual dataset was then processed with and without accounting for habitat suitability and the results were compared.

To test in what measure Jackknife resampling improved the performance of the algorithm I took two sets of ten replicate datasets that produced relatively inaccurate results and performed 10 85% Jackknife resampling on each replicate. All the Jackknife datasets were then processed as described in the "Jackknife resampling" section, and their α estimates compared with the results of the procedure without Jackknifing.

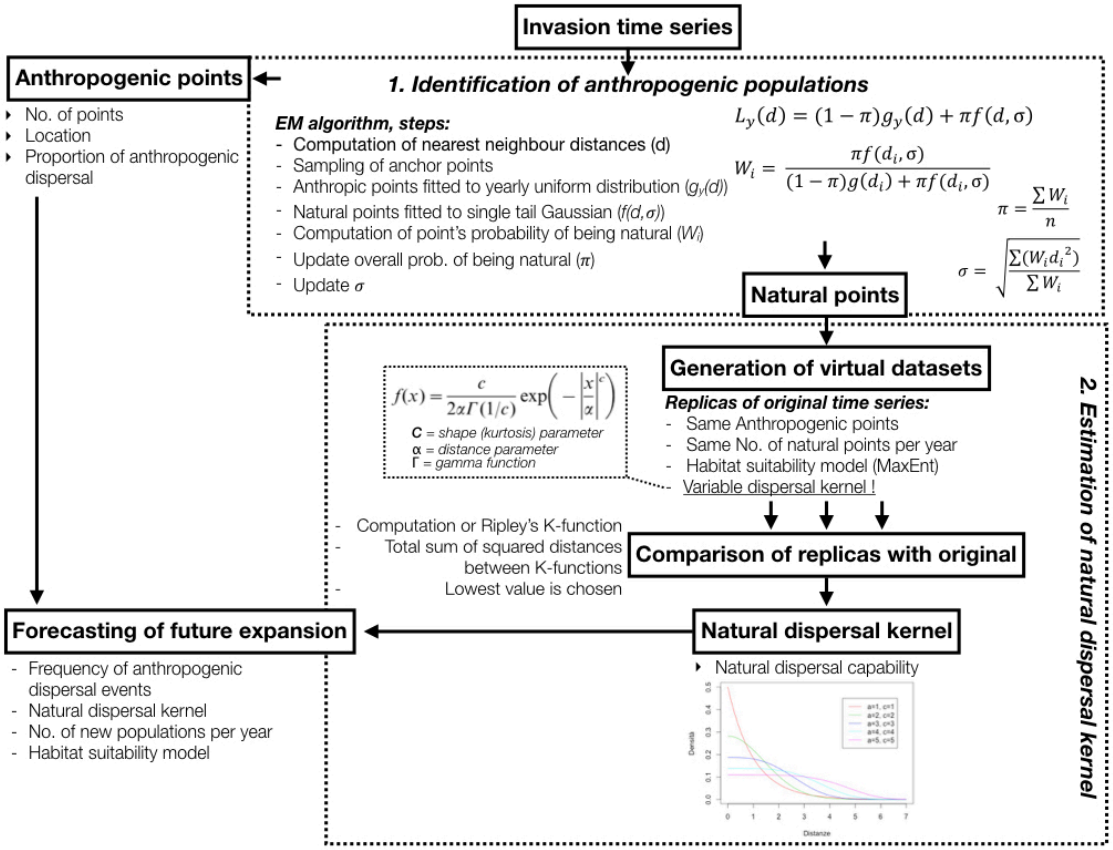


Figure 3: Flow chart of the modelling process. Invasion time series data is provided as input. The first step is to classify each point as either of natural or anthropogenic origin through the EM algorithm. This involves the computation of the nearest neighbour distance for each point to the closes from the same or previous years (with the exception of one "anchor point" per year forced to be sampled from points of previous years) and fitting these to either a single tail gaussian function (natural) or a yearly uniform distribution (anthropogenic). The second step is the estimation of the natural dispersal kernel. To achieve this a number of virtual datasets is generated, each having different dispersal kernel, but the same anthropogenic points as the real dataset. Then the similarity of each virtual dataset's K-functions is compared to the real dataset and the estimated dispersal kernel is retrieved from the closest one. By combining the information obtained in these two steps, it is possible to forecast the development of the biological invasion.

Applied example

To assess whether the expansion of *Litoria* species in New Zealand has already reached its maximum, a MaxEnt (MaxEnt version 3.3.3k; Phillips & Dudík 2008) Habitat Suitability Model (HSM) was built for each species based on the sighting locations retrieved from New Zealand's Department of Conservation's Amphibians and Reptiles Distribution Scheme (ARDS). The database records were selected in order to remove duplicate records (same combinations of species, latitude and longitude) and data with an

inaccuracy of more than one km. Accuracy in ARDS is classified into 7 classes; imprecision less than 5 km (model cell size) would be acceptable, but data with more than one km of error were classified in the next class of 10 km inaccuracy and were therefore discarded. After this process, there were 129 records for *L. aurea*, 254 for *L. raniformis*, and 323 for *L. ewingii*.

Pearson's statistic was used to test the spatial correlation of a total of 31 environmental variables that could potentially be included in this model: elevation; percentage of artificial forests; all 19 Worldclim biological variables; percentage of cropland; grassland (percentage of grassland; LUCAS 2nd edition); percentage of natural forest; density of roads; slope; solar radiation; percentage of urban area; percentage of water cover; percentage of wooded grassland. Only a subset of 14 variables with low pairwise correlations ($-0.7<p<0.7$) were kept for the analysis (Appendix Tab. 1). All environmental variables maps were projected in an equal area coordinate system (New Zealand Transverse Mercatore 2000 (NZTM2000). All maps were given the same spatial extent, pixel size and corner coordinates and were saved as ASCII files (No. of columns: 1000; No. of rows:1371; No. of cells: 1371000; Extent: 1089971, 2089971, 4823127, 6194127 (xmin, xmax, ymin, ymax); Cellsize: 1000 x 1000; Length measure of unit: meter).

To correct for the clustering of sighting locations in the model due to systematic sampling errors (i.e. sighting locations are more abundant close to cities or research sites), a background dataset of 10,000 random points stratified on a kernel density estimate of all the sighting locations of the three species together was used. Cross-validation of MaxEnt models were performed over 10 random sub-samplings of the original dataset. A Multivariate Similarity Surface (MESS) was computed to highlight the areas of novel habitat type relatively to the current distribution of the species.

EM algorithm and kernel estimation

Jackknife resampling was done on the cleaned ARDS time series for the three species of *Litoria* frogs in New Zealand. Ten Jackknifed datasets, each of which contained 85% of the data points, were obtained and processed in parallel.

The EM algorithm was run on the Jackknifed datasets and on the original full dataset. The 10,000 random points used for the estimation of the yearly probability density function for the nearest neighbour distances were sampled from the whole of New Zealand for *L. raniformis* and *L. ewingii*, while for *L. aurea* the points where sampled exclusively within the North Island, as there are no sighting locations for this species in the South Island. Points were considered of anthropogenic origin when their probability of being so (1 – prob. of being natural) was bigger than 0.5.

Once the origin of the population was classified as either natural or anthropogenic, the dataset was run through the dispersal kernel estimation algorithm. Spatial filtering was used to account for habitat suitability. The number of points produced each year was set to be 30 times the actual number needed. The MaxEnt models described above were used as the HSM. The number of replicates for the comparison datasets was set to 90 and their initial α values were 2, 3, 4.5, 7.5, 11, 15, 20 and 25. α values were updated to higher numbers ((2, 5, 10, 20, 30, 40, 50), and (20, 30, 40, 50,60,80,100)) for those datasets that required them in order to identify a minimum dissimilarity in the range of tested αs ("U" shaped lines in the similarity plots of the results section). The first ten rows of each Jackknifed dataset (the oldest ten points available) were used as the initial dataset on which to build the comparison datasets. The search radius for Rippley's K-function was set from zero to 30 km (and, if needed, 100 and 200km) at one km steps.

Results

Assessment of the EM classification shows an overall high proportion of correct classification, although yearly aggregation, C value, and the proportion of anthropogenic points have clear effects on the proportion of correct classification (Figure 4). A combination of high percentage of anthropogenic points and leptokurtosis causes the lowest success rates in the EM classification. In these conditions, the algorithm confuses some anthropogenic points with natural LDD events. Similarly, a high percentage of anthropogenic points confuses the EM classification in the presence of gap years. Figure 4 shows that, when the true anthropogenic proportion is non-zero, errors in classification typically result in underestimation of the proportion of anthropogenic points.

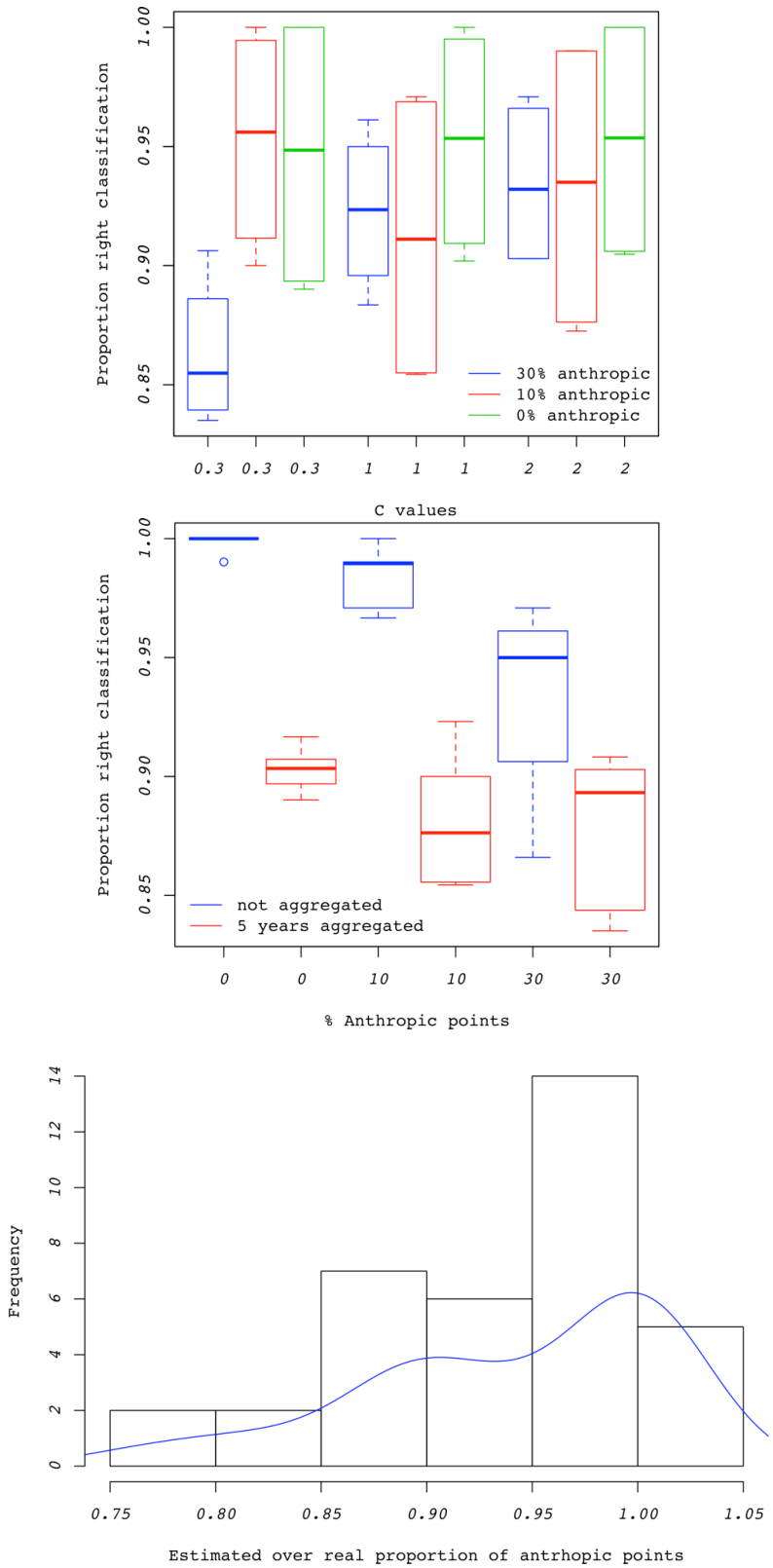


Figure 4: Proportions of rightly classified points with the EM algorithm in relation to the C values (top) and number of anthropogenic points (middle) used to build the virtual datasets. Bottom: histogram of the proportion of rightly classified anthropogenic points; the algorithm performs well and is biased towards underestimating the number of anthropogenic points.

Of the 32 variations in the virtual datasets (average of the replicates, Figure 5; Appendix Fig. 7) 18 α estimates fall within the right cluster of α and C combinations, eight are overestimated by one unit and six are overestimated by two units. Within the 16 virtual datasets with α=15, only two α estimates are off by one step, while all the others are exact. Virtual datasets with 400 points produce better estimates of α than those with 100. In Figure 6 ("*Effect of Alpha*"), outliers for simulated datasets with α values of 4.5 are only above – and only below for datasets with α of 15. This may be due to the sequence of tested α values (2; 3; 4.5; 7.5; 11; 15; 20; 25) being relatively longer when below for 15 and when above for 4.5. Uncertainty of the estimates is visualised in the dissimilarity plots (Figure 5) by the variability between replicates and by the depth of the "valley" in the average dissimilarity curve. Jackknife resampling improves the overall estimation of the dispersal kernel and reduces the variability between replicated datasets (Figure 7 "*A-D*"). This improvement was from two units overestimation to one unit overestimation for the first dataset and from one unit overestimation to exact estimate for the second dataset.

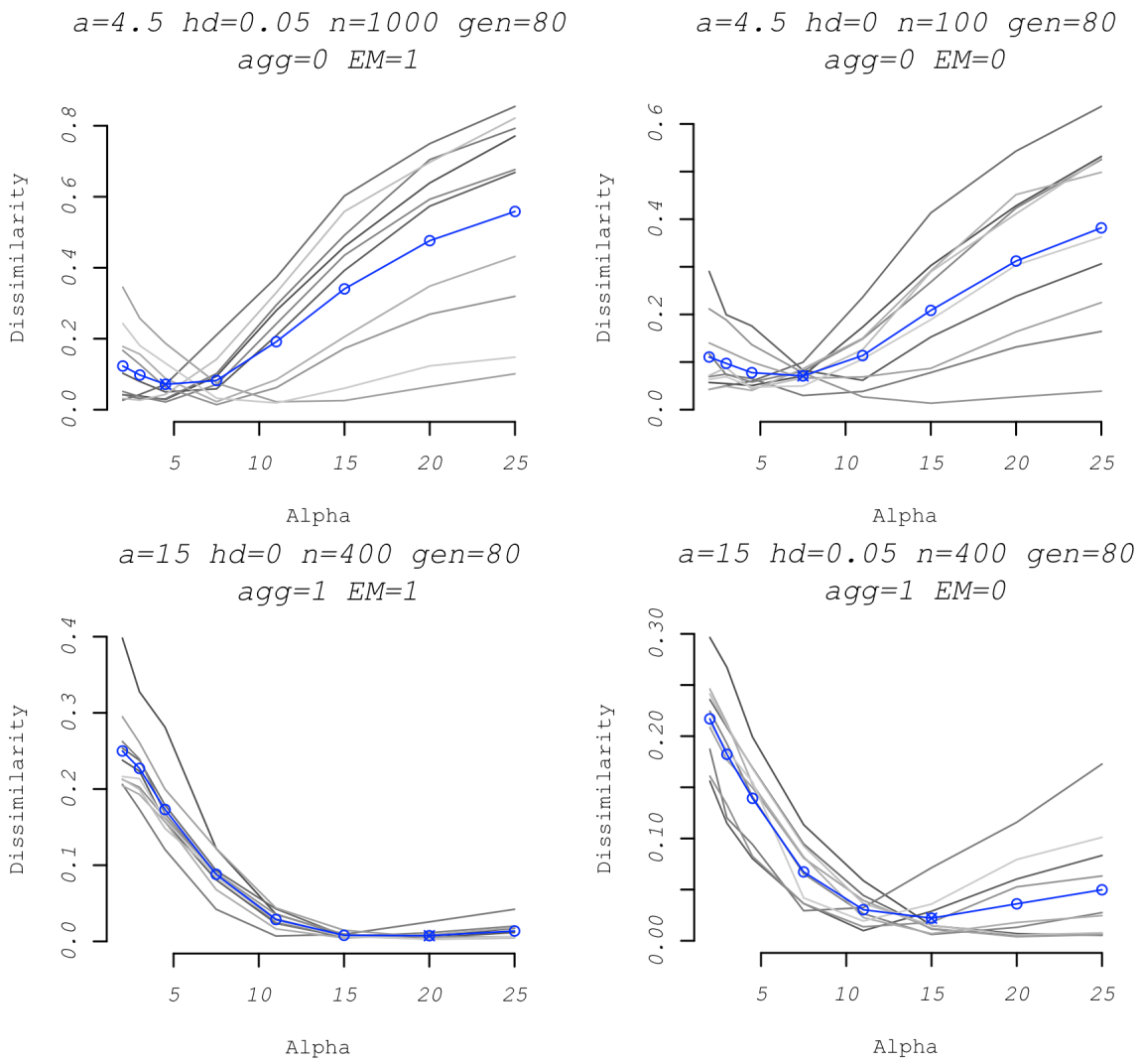


Figure 5: Four sample dissimilarity plots for the estimation of the alpha parameter out of 320 virtual datasets (see Appendix Fig. 7 for image with all 320 datasets). The parameters used to build the virtual datasets are on top of each plot: a= alpha; hd= proportion of anthropogenic points; n= No, of points; gen= Np. of generations; agg= presence (1) or absence (0) of gap-years; EM= real (0) or EM-estimated (1) classification of anthropogenic points. Each grey line represents a replicate, the blue line is the average of the grey ones. The lowest point is the best representative of the alpha value of the virtual dataset. Uncertainty of the estimates is visualised by the divergence of the grey curves at any alpha value and by the depth of the "valley".

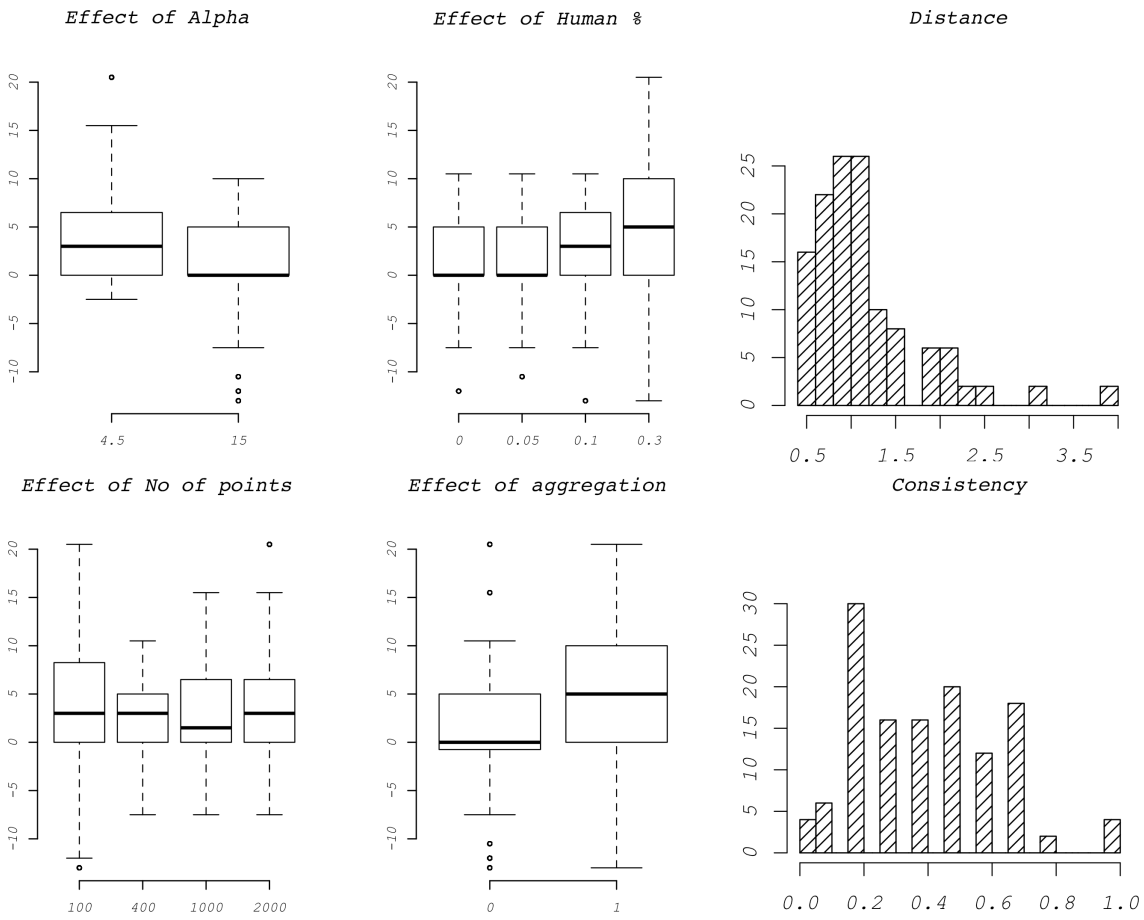


Figure 6: Performance of alpha estimation. Boxplots: Summary of the effect of different variables used to build the virtual datasets on the performance of the algorithm. As measure of performance on the y axis, estimated minus real alpha value is used. So, values above zero mean overestimation of alpha, while negative values mean underestimation. Distance: histogram of the average error in the alpha estimates measured as the number of alpha values away from the mean each replicate is. Consistency: histogram of the average consistency of the estimate in terms of proportion of replicates with the same estimated alpha value. These two histograms combined give a good idea of the performance of the algorithm: about half of the estimates are exact, and the remaining half rarely exceed one alpha value difference.

Increasing the proportion of anthropogenic populations augments the variability of the estimation but does not increase bias. A similar effect is generated by decreasing the number of points in the datasets: the main change is the reduction in precision, but not in accuracy, of the α estimate with datasets with only 100 points. Introducing in the dataset gaps years (aggregation), also causes a small increase of the estimation's variability (Figure 6, "*Effect of aggregation*").

The HSM filtering dramatically improves the accuracy of the algorithm. Spatial clustering, in virtual datasets where spreading is constrained by habitat suitability, is overestimated if the ecological niche of the species is not considered (Figure 7 "*E-F*").

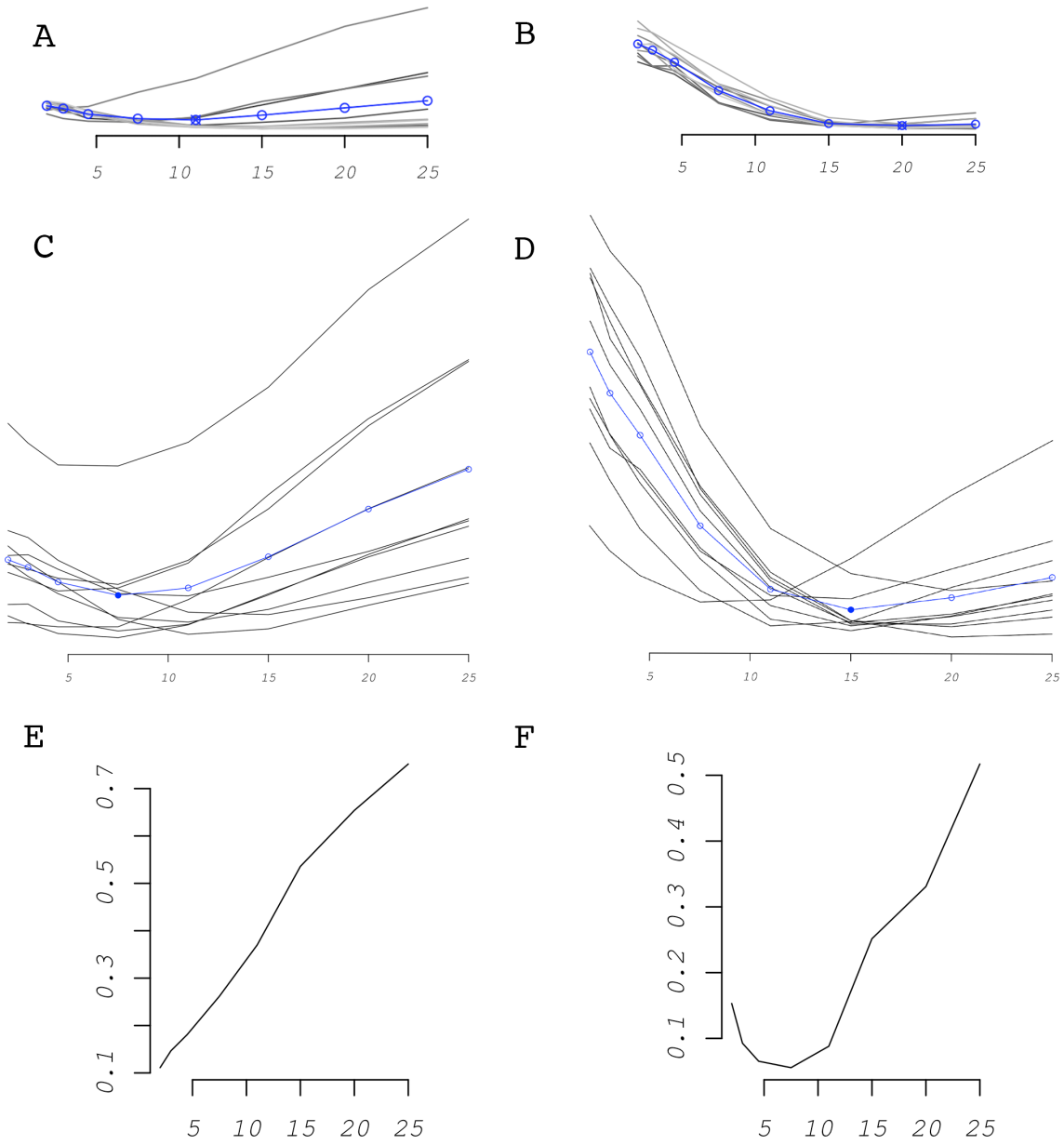


Figure 7: Performance of Jackknife re-sampling and HSM filtering. A-D: effect of Jackknife re-sampling on two datasets (ten replicates) that produced poor results (A & C: a=4.5; % anthropogenic=10; n=100; no aggregation. B & D: a=15; % anthropogenic=30; n=400; with aggregation). A & B: without Jackknife; C & D: with Jackknife. E & F: comparison of one virtual dataset (α=4.5; % anthropogenic=10; n=400; no aggregation) generated with L. raniformis HSM and analysed with (F) and without (E) HSM filtering.

Applied example

Results of MaxEnt modelling are available on the Appendix (MaxEnt results). The EM algorithm classified 19.86% (29 out of 146) of *L. aurea* populations, 40.22% (111 out of 276) of *L. raniformis* populations and 25.22% (86 out of 341) of *L. ewingii* populations as of anthropogenic origin (Figure 8, first row).

The estimated dispersal kernel for *L. aurea* is α= 30 km; *L. ewingii* is estimated to disperse with α= 60 km. Both these species present a "U" shaped similarity plot (Figure 8, fourth row "*aurea*" and "*ewingii*"). The dispersal kernel for *L. raniformis* is a positive linear function with estimated value of α: 2 km (Figure 8, fourth row "*raniformis without correction*"). However, the results for *L. raniformis* show more variation between Jackknifed datasets compared to the other two species. Some of them have a very different behaviour (almost opposite) to that of the majority of the other Jackknifed datasets. This difference in the estimation of the dispersal kernel for such similar species (especially *L. aurea*) seems to point to an artefact in the dataset. For this reason, 49 points classified as of natural origin in the dataset for *L. raniformis* were deleted. These points are very close to one another and follow a road in Taranaki (Makuri and Wawiri road, Figure 8; they all date 2001). They are obviously part of a survey, where all animals sighted along the road were passed to the ARDS and included in the general dataset. After deleting them, the new dataset was run as previously and results changed radically, with α now being estimated as 50 km and the similarity plot being "U" shaped, resembling the other species (Figure 8); the proportion of anthropogenic points is now 34.80% (instead of 40.22%) as 32 anthropogenic and 17 natural points were deleted (misclassification of few of these points occurred as a consequence of being an artefact); all the Jackknifed replicates now behave similarly (Figure 8, fourth row "*raniformis*" and "*raniformis without correction*"). It was not necessary to remove this survey artefact prior the building of the HSM because all the points fall in two

single raster cells and are therefore automatically reduced to two single points by the MaxEnt algorithm.

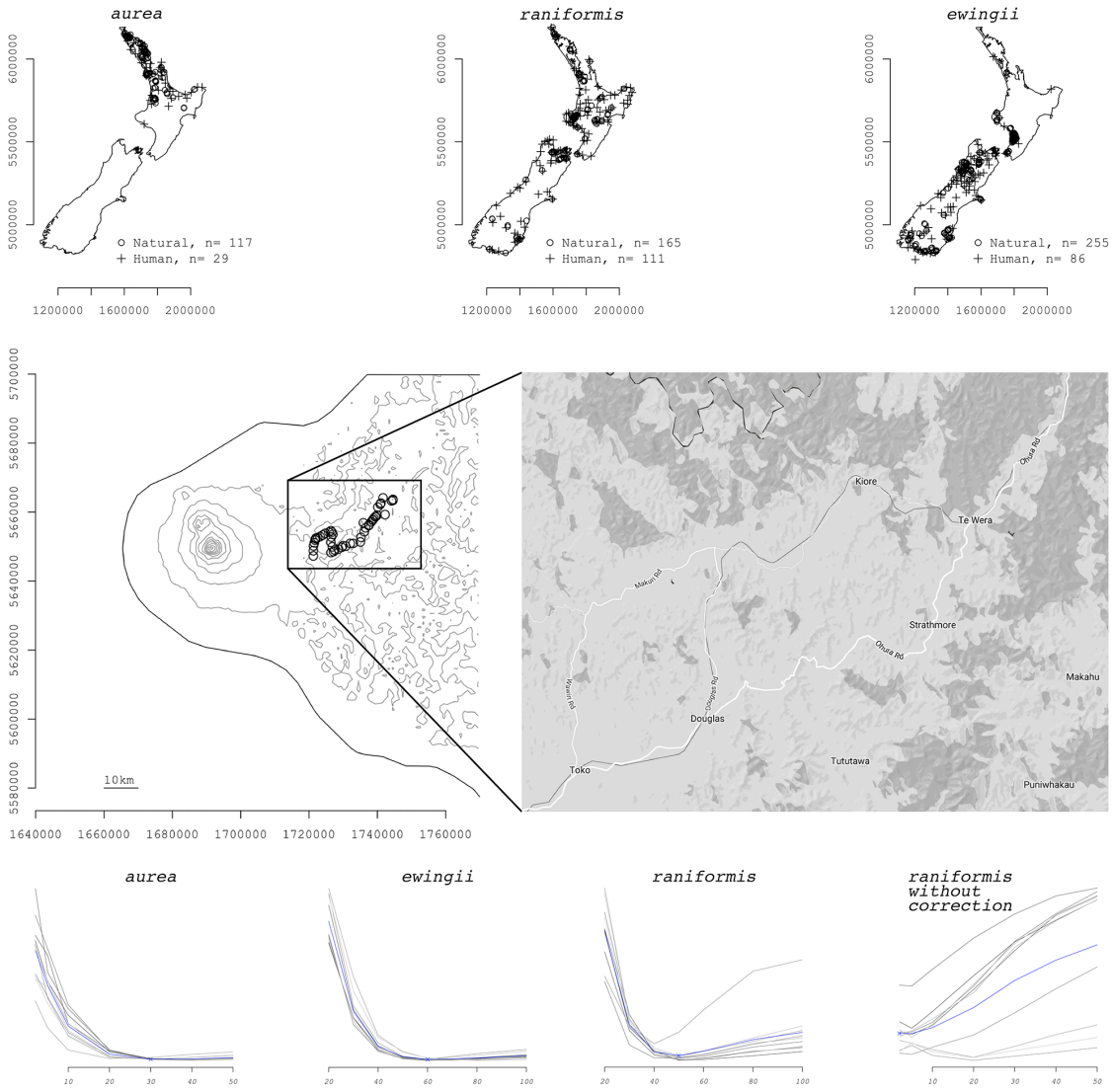


Figure 8: Applied example, Litoria spp. in New Zealand. First row: Maps of sighting locations for Litoria frogs in New Zealand after EM classification: populations of human origin as crosses and of natural origin as circles. Second row: Artefact in the dataset of L. raniformis that was removed. A line of very dense sighting locations along a road in Taranaki from 2001. On the right, an enlargement of the road followed during the survey. Third line: Average dissimilarity values between comparison and observed datasets for the Litoria spp. time series. Each line represents averaged similarity values of all the simulated datasets for each a value. Each line represents a Jackknifed dataset. HSM filtering was performed. The blue line is the average of the grey lines.

Discussion

This method's accuracy for both the classification of anthropogenic vs. natural populations and the estimation of the dispersal kernel is reliable and can be informative for datasets with as little as 100 records, even when there are gaps in the survey years. Single α estimates have a certain variation around their mean, but most values are just one step away along the sequence of tested α values (Figure 6). I suggest using the HSM implementation whenever possible as my virtual dataset test shows it drastically improves the accuracy of the algorithm.

The estimated anthropogenic contribution to the spread of *Litoria spp.* in New Zealand has been very high, with an average of ~27%, and peaking at ~35% in the recorded populations of *L. raniformis* originated through anthropogenic dispersal. Note that, even if only a minority of individual records are of anthropogenic origin, most of the large scale spread is anthropogenic in nature.

Jackknifing is also a very useful implementation as it 1) gives more accurate results, especially for datasets that are small or potentially flawed; 2) provides information on the variability of the estimated dispersal kernel and 3) helps detect the presence of inhomogeneous areas in the source dataset. In the case of *Litoria spp.* Jackknife analysis helped detect a survey artefact that otherwise would have caused a major bias in the estimation of the dispersal kernel. This was highlighted by the Jackknifed datasets behaving very differently from the others, when a good portion of the survey artefact was randomly removed by the Jackknife process.

Once the dispersal kernel and the proportion of new releases are known, the same method used to produce simulated datasets can be used to project any invasion to into the future. In the case of *Litoria spp.* it is not particularly useful as the boundaries for the potential distribution have been practically reached. Projecting the invasion would

generate new points whose density would merely reflect the HSM. Nevertheless, projecting an invasion into the future is of great importance, especially at the early stages of biological invasions. When doing this, it is fundamental to remember that the process my method is modelling is not necessarily the actual spreading of the study species, but rather the spatial point process in the sighting locations database. Only if the survey is thorough enough to approximate the real invasion, can we expect to be modelling the actual invasion. For this reason, especially when the sighting locations database is not particularly thorough, the estimated dispersal kernel is to be handled carefully outside the framework of phenomenological/statistical invasion modelling (e.g. in mechanistic dispersal models). In addition to biological invasions, this method can be used to model climate change induced range shifts or epidemics, as long as spatio-temporal data on the ongoing process is available.

Future developments of this method would include accounting for climate change both in the dispersal kernel estimation and in future projections and allowing for the use of a release probability map for both classifying anthropogenic vs. natural populations and for future projections. This latter case, would allow to use this method where the probability of release of propagules by people is not uniformly distributed across the landscape (e.g. clustered along rivers).

Chapter 2b: Biolinv, R package for modelling biological invasions.

Abstract

Biological invasions modelling has very important implications for the conservation of biodiversity and the prevention of the high costs that are associated with the mitigation measures against exotic species (Pimentel *et al.* 2005). People play a fundamental role in the spread of exotic species; however, little attention is being paid to anthropochory when modelling invasions. Biolinv provides functions to analyse past and ongoing biological invasions and to make future forecasts. It does this with a stochastic, non-mechanistic approach that estimates to the anthropogenic component, accounts for habitat suitability and provides measures of precision for its estimates. Here, I first give a very brief introduction to the methodology behind the package, I then describe the main functions and eventually provide example code for a typical application of the package.

Introduction

The biolinv R (R_Core_Team 2016) package provides tools for modelling and forecasting biological invasions with special consideration to anthropochory, the transportation of propagules by human. Anthropogenic dispersal is at the core of biological invasions. However, available software that considers anthropochory often models it as long distance dispersal. This implies that the new release sites for anthropogenic dispersal depend on the location of existent colonised sites, which is not necessarily true in common scenarios (e.g. when the source of propagules is the native distribution range or when anthropogenic dispersal is much faster than natural dispersal).

Starting from an ongoing invasion time series, Biolinv provides tools for classifying sighting locations that result from either natural or anthropogenic dispersal. It also estimates the anthropogenic contribution to the invasion, the invasion's natural dispersal kernel, and makes future forecasts on the invasion process. This is done while accounting for habitat suitability in the form of a Habitat Suitability Model (HSM, in raster format) and by providing a measure of precision in the estimates.

Core functions

The package is based on three main functions – namely `EM`, `simulacro` and `modsel`. These are used to run the EM algorithm, generating a virtual dataset and measuring spatial similarity respectively. The package also has several auxiliary functions that are primarily called by the main functions, but they can also be run on their own for specific tasks. Functions for the graphical representation of results are also available.

Function `EM`

The `EM` function is used to estimate which populations in a biological invasion's time series are of natural origin and which ones originated by anthropogenic dispersal. It uses nearest neighbour distances from the invasion time series as computed internally by the function.

The EM algorithm at the core of this function iterates between updating the distribution of nearest neighbour distances for natural points (assumed to be from a half normal (0, `sigma`), and updating the probability that each point is the result of natural dispersal (`Pnat`). Starting values for `Pnat` and `sigma` must be supplied, but the algorithm is robust to the exact values. The anthropogenic subset is assumed to come from a uniform distribution; the resulting distribution of nearest neighbour distances (with

neighbours assumed to be extant points) is computed by simulation, using the observed points as the extant points and the specified spatial boundary as described in chapter 2b. The output of this process is a vector of probabilities of being of natural origin (`Pnat`) for each data point. A threshold can then be used to classify points as natural or anthropogenic.

Function `simulacro`

The `simulacro` function is used to simulate an invasion time series based on spatial boundaries, a proportion of natural vs. anthropogenic points, a natural dispersal kernel and a HSM. It can do this wither by simulating "from scratch", or by copying the anthropogenic release sites from observed data. Time series are generated by an iterative process for which every cycle represents a time step (e.g. one year). For each iteration (time step), a number of points of anthropogenic and natural origin are generated. Anthropogenic points are either generated randomly across the geographic boundaries with the function `RPG` (Random Points Generator) or copied from a data frame that includes columns named `x`, `y`, `year` and `Pnat`. In the latter case, `x` and `y` are the points coordinates in a projected coordinate system; `year` must contain the same year numbers for which the point generation is performed and `Pnat` (probability of being of natural origin) must contain values between zero and one (values smaller than a threshold specified in the `simulacro` function will then be considered of anthropogenic origin). Natural points are generated in user defined, yearly numbers starting from extant points at that time step. Distances from existing points are sampled from a user defined One Dimensional Dispersal Kernel (function `fx`) and directions are sampled randomly. This part is done with the `NPG` (Natural Points Generator) function. Spatial filtering is done on both the natural and anthropogenic subsets through the function `spatFilter`, which

deletes points from the input argument `points` based on the value of the raster's cell (argument `MAP`) they fall into. In order to account for the reduction in the number of points after the filtering, an excess of points is initially generated (arguments `FACNAT` and `FACANTH` express factors by which the wanted number of points is multiplied before filtering). Eventual superfluous points are then randomly deleted to reach the user defined quantity.

Function `modSel`

`modSel` is used to compare several simulated time series with one observed invasion time series. It is a wrapper around the `spatSim` function which generates a measure of similarity between any two time series. This is done by computing Rippley's K-function for the two datasets and measuring the sum of their square differences.

Inputs

Requested input files are of two types: invasion time series and spatial boundaries.

Invasion time series

The time series must be provided in the form of a data frame: one record per row containing the columns `year`, listing the year of the sighting in a numerical form; `x` and `y`, listing longitude and latitude respectively of the sighting locations in a projected coordinate system.

Spatial boundaries

The spatial boundaries must be in the form of either a 1) RasterLayer object with cell values of either 0 (indicating the area where the computation occurs) or >=1 (indicating the area outside the study system) or 2) a SpatialPoligon(s) object. Like for the time series

and all other spatial information in the package, spatial boundaries must also be provided in a projected coordinate system. An example invasion time series can be called with the command:

```
data(frogs)
```

`frogs` contains `year` (year of the sighting), `y` (latitude in the New Zealand Transverse Mercatore coordinate system, length unit of measure is meter), `x` (longitude in the New Zealand Transverse Mercatore coordinate system, length unit of measure is meter). An example of spatial boundaries can be called with the command:

```
data(nzp)
```

`nzp` is a SpatialPolygons object with two features (the North and South islands of New Zealand).

Sample analysis

Using the input data presented before, I show the typical Biolinv workflow. This allows to classify the sighting locations in a time series as either of anthropogenic or of natural origin, to quantify the human contribution to the invasion, to estimate the natural dispersal kernel, and to project the time series in the future.

Jackknife on invasion time series

```
data(frogs)
frogsJK<- jackKnife(DF= frogs, N= 10)
```

The first step of the analysis is optional but I suggest implementing the Jackknife approach especially in the case of small time series (few localities or few years). Jackknifing generates multiple datasets from the original time series, all of which will need to be processed in the same way. Ideally the processing can be done in parallel; naturally, the

processing time increases considerably if Jackknifed datasets must be processed in series. All following instructions are supposed to be repeated for each of the Jackknifed datasets.

Running the EM algorithm

```
data(nzp)
data(frogs)
randp<- RPG(rpopn=1000, boundary=nzp, SP= 'random_frog')
frogsEM<-  EM(dataset=  frogs,  randompoints=  randp,  sigma=6,
pi=0.5)

# plot output

plotlacro(x= frogsEM, outline= nzp)
```

Function `EM` runs the EM algorithm. The `dataset` parameter is the time series (`frogs`). The output of the `EM` function is a copy of `dataset` with a new column, `frogs$Pnat`, which indicates the probability of that point being of natural origin. Argument `randompoints` can be generated with the function `RPG` (Random Points Generator), which is used to generate a Poisson point process within the spatial boundaries of the analysis. This dataset is used to compute the nearest neighbour distances used in the EM algorithm. `rpopn` is the number of desired points to be generated (1000 is a good number for most cases). `SP` is the name to be stored in the column `species` of the output.

The output of the `EM` function and its argument `dataset` can be conveniently plotted by the function `plotlacro`, where argument `x` is the data frame to be plotted and argument `outline` is the spatial boundaries for the analysis.

Setting up time series simulation

```
idst<- frogsEM[1:10,]
Cr<- frogsEM[-(1:10),]
yr<- unique(Cr$year)

nNoYear<- rep(NA,length(unique(Cr$year)))
hNoYear<- rep(NA,length(unique(Cr$year)))
```

```
for(i in 1:length(unique(Cr$year))){
   # Cr for that year:
 CrYear<- Cr[Cr$year==unique(Cr$year)[i],]
   # natural points for that year:
 nNoYear[i]<- nrow(CrYear[CrYear$Pnat>=.5,])
   # human points for that year:
 hNoYear[i]<- nrow(CrYear[CrYear$Pnat<.5,])
}

AV<- c(2,3,4.5,7.5,11,15,20,25)  #alpha values

frogsLacro<- simulacro(INIDIST=idst,YEARS=yr,
 BOUNDARY=nzp,NNAT=nNoYear,NANTH=hNoYear,
 FACNAT=10,
 A=AV,X=seq(.1,30,.1),
 TRUEANTH=TRUE,TRUEDB=Cr,PROB=.5,
 ITERATIONS=10,HSM=nzp)
```

The `simulacro` function generates multiple (or one) virtual datasets that have the same anthropogenic points as the input dataset (`frogs`) and the same number of natural points. These are generated with varying natural dispersal kernels. The simulated dataset that most resembles the `frogs` dataset will then be selected in the next paragraph with the `modSel` function. The `simulacro` function requires considerable processing power to generate all the required datasets. Therefore, I suggest performing an initial test run using a low number of Alpha values and replicates before running the full dataset generation process.

The `simulacro` function requires several arguments: `INIDIST` is the initial distribution from which every simulated dataset starts. In the example above, the first ten records of `frogs` have been used for this (`idst`). `YEARS` is a vector of unique year numbers used to correctly name the output simulated time series. `BOUNDARY` is a map of the geographic boundaries within which points are generated (same as in previous functions). `NNAT` is either a vector of (integer) numbers of natural points to be generated per year or the mean (integer) number of natural points to be generated each year. `NANTH` is the same but for the anthropogenic points. `FACNAT` and `FACANTH` are factors by which the number of natural and anthropogenic points respectively are multiplied

before the HSM filtering. They are only used if HSM is a raster file and should be set to higher values the lower the proportion of suitable cells in HSM. A is a vector of unique alpha values used to make virtual datasets. In the example, eight values are used, this means that eight datasets per replicate will be created by the simulacro function. Similarly, C is a vector of unique C values, however it is not necessary to set it to anything else than the default value of 2. X is a vector of possible dispersal distances used for simulating natural dispersal. This set of distances will be sampled from the One Dimensional Dispersal Kernel function with C= 2 and Alpha= A. TRUEANTH can be either "true" or "false". If "true", TRUEDB will be used to generate anthropogenic points, if "false" anthropogenic points will be generated with the RPG function in the quantities specified in NANTH. The use of this latter option is discouraged at this stage of the process as it is mainly meant for future forecasts of the invasion process (see below). Using TRUEDB will generate better datasets for the estimation of the dispersal kernel. TRUEDB is a data frame containing points of anthropogenic origin; must be in the same format as INIDIST. Rows where TRUEDB$Pnat>PROB will be ignored. PROB is a threshold value over which values in TRUEDB time series are considered of natural origin. ITERATIONS is the number of datasets with the same Alpha (A) and C values. For a full analysis, I suggest a higher number of iterations than the one in the example above (this function has tested positively with 90 replicates). HSM is the habitat suitability map. It can be a RasterLayer object, giving the probability of successful establishment of new populations, or a SpatialPolygons object, giving the boundaries within which new points are generated.

The output of the simulacro function is a list of lists of dataframes, each of which is a simulated time series. If DIR=TRUE one folder will also be created for each Alpha and C combination containing all the replicates datasets set in ITERATION. if DIR=FALSE

(default) the order in the list will follow the order of the C and Alpha values respectively as set in C and A.

Selecting best time series simulation

```
data(nzw)
frogsSum<- modSel(WIN= nzw, M0= frogsEM, M2= frogsLacro,
 AV= c(2,3,4.5,7.5,11,15,20,25), RAD= seq(0,30000,1000))

# plot output:

plotAlpha(SSIM= frogsSum, REP= 10)
plotAlpha(SSIM= frogsSum, REP= 10, BP=TRUE)
```

To select the simulated time series that is the most similar to my observed one (frogs) the modSel function is used. It takes five arguments. WIN is the spatial boundaries of the analysis (like nzp) in the form of a "window" object from package "spatstat". An example WIN object (nzw) is called in the first row. M0 is the output of the EM function (observed time series). M2 is the output of the simulacro function (simulated time series list). AV is a numeric vector of the Alpha values used in the simulated datasets in the same order as in the list of argument M2. It is used to save the output data frame. RAD is a numeric vector of search distances for the K-function. This parameter indicates the radiuses at which the K-function K(r) is evaluated. Fewer values speed up the computation time but produce less accurate results.

The output of this function is a data frame with two columns: dissimilarity contains the dissimilarity values computed by the spatSim function; compAlpha contains the input values from AV. This data frame can be conveniently visualised with the plotAlpha function. When argument BP is set as "true" all replicates are represented as boxplots, otherwise only the average line is plotted.

Future projections

```
idst<- frogsEM
```

```
yr<- seq(2015,2030,1)

hNoYear<-rep(c(1,0,0),length.out=length(yr))

AV<- 20

frogsLacro<- simulacro(INIDIST=idst,YEARS=yr,
                       BOUNDARY=nzp,NNAT=2,NANTH=hNoYear,
                       A=AV,X=seq(.1,30,.1),
                       TRUEANTH=FALSE,PROB=.5,
                       HSM=nzp,FACNAT=2)

f<- frogsLacro[[1]][[1]]
f<-f[f$year>2010,]

plot(nzp)
points(idst$x, idst$y,pch=3)
points(f[f$Pnat>=.5,'x'],f[f$Pnat>=.5,'y'],col=4)
points(f[f$Pnat<.5,'x'],f[f$Pnat<.5,'y'],col=2)
```

Function `simulacro` can also be used to project an invasion into the future. The main differences with respect to the use described above are the argument `YEARS`, that for future projections must contain the year's numbers for which to produce new points; and the argument `TRUEANTH`, which must be set as `=FALSE`. The mean number of anthropogenic and natural points to be generated per year can be computed by averaging the number of anthropogenic and natural points per year as classified by function `EM`. Note that values of less than one are not acceptable as arguments `NNAT` and `NANTH` represent numbers of points to be generated.

Conclusion and future implementations

Biolinv is a concrete step forward towards modelling and forecasting biological invasions, especially when accounting for anthropogenic dispersal.

Planned improvements in this package include the possibility of running the `EM` and `simulacro` functions on non-random anthropogenic release patterns, and allowing the HSM to change over time within the `simulacro` algorithm to account for climate change. The package is available on the CRAN website (https://cran.r-project.org/package=Biolinv).

Chapter 3: Early stage habitat fragmentation scenario for nature reserves in young New Zealand suburbs.

Abstract

The world's urbanisation is increasing the significance of urban ecology for the conservation of biodiversity. A consequence of this urbanisation is habitat fragmentation, which reaches its highest intensity in cities. However, in spite of its growing importance for conservation, little is known about the role played by different urban land covers with respect to their impact on connectivity. This chapter attempts to shed light on this interplay by studying the case of Copper skink populations of remnant vegetation patches in Auckland, and evaluates the effects of habitat fragmentation at a fine scale, urban setting. First, I use measures of pairwise genetic structure and clustering to assess the degree and pattern of isolation of the study species. Then, I employ a landscape genetics approach to attribute landscape resistance weights to multiple, urban landscape features.

I found urban copper populations being genetically structured (with high Gst values) and grouped into two main clusters separated by a motorway. However, no effect of different land cover types has been found to be affecting such structure.

My findings possibly describe an early stage of habitat fragmentation, where the genetic response to the creation of new urban barriers is beginning to manifest. Stronger obstacles to gene flow like the motorway already generated a genetic signal, while weaker barriers did not manifest yet. At this stage, the aggravation of the trend is likely to be prevented by implementing conservation strategies on Auckland's urban reserves.

Introduction

More than half of the global human population now lives in urban areas: the proportion of humans living in cities is expected to rise from 54% in 2014 to 66% by 2050. As a consequence of the urbanisation of humanity, challenges to sustainable development are expected to concentrate in urban areas, especially in fast urbanising countries (United Nations 2014). Among these challenges, mitigating the inherent habitat fragmentation that results from urbanisation is central for the conservation of biodiversity (Haddad et al. 2015).

The negative effects that habitat fragmentation has on biological communities are linked to a general reduction of landscape connectivity – the capacity of a landscape to facilitate dispersal – caused by urban barriers. To better understand the dynamics of such phenomenon, Landscape Genetics (LG) have effectively combined methodologies from spatial ecology and population genetics to specifically address landscape connectivity issues (Manel *et al.* 2003, Holderegger & Wagner 2008). LG studies quantified the beneficial effects of ecological corridors (e.g. under- or over- passes and greenways) in rural areas. For example, Mech & Hallett (2013) found corridors between unlogged habitats to improve connectivity of Red-backed voles in logging forests of north-eastern Washington, and Dixon *et al.* (2006) proved the efficacy of a regional corridor for connecting two Florida black bear populations. However, studies on urban settings remain rare. Similarly, most of the research effort in the field of LG is focused on predominantly broad scale, while finer scale studies on are uncommon (Storfer et al. 2010, LaPoint et al. 2015).

Even though LG applications to the urban environment are too few to generalize, recent studies have started showing how particular urban features correlate with functional connectivity in different taxa. New York City's canopy cover correlates

positively with gene flow in White-footed mouse populations (Munshi-South 2012). Also, 20% to 50% impervious surface cover type is proven to drastically reduce Yellow-faced bumble bee gene flow in California, and allocating the same amount (20-50%) of urban land cover to permeable materials would make for a sustainable urban landscape for the study species and possibly other ground nesting insects (Jha & Kremen 2013). Finally, vegetation gaps exceeding 30 meters were shown to be responsible for dramatically reducing the likelihood of movement of songbirds in Calgary, Canada (Tremblay & St. Clair 2009). Being able to predict the effect of land cover types on the landscape's connectivity in urbanised land is of central importance to the sustainable development of new urban areas and for mitigating the already existing effects of habitat fragmentation in extant suburbs. However, we are mostly, still far from being able to use specific urban features (like proportion of different land covers, or tree density) to manipulate habitat connectivity during urban planning.

A typical problem linked to studies on urban connectivity is that both the processes of urbanisation and adaptation to urbanised land take time. Due to the so called time lag problem (Epps & Keyghobadi 2015), the genetics of remnant populations do not immediately respond to a reduction in habitat connectivity. Only after a certain number of generations do these changes manifest in the genetic material. Subsequently, through the process of environmental filtering (Kraft *et al.* 2015), inbreeding depression causes populations that cannot adapt to disappear while new and urban adaptable species colonise the vacant niches (Niemela 1999). From a conservational standpoint, LG studies are most valuable when the study system is at an intermediate stage: where a genetic response is detectable but has not caused local extinctions yet. Such a genetic response can be used to infer the differential effects land cover types have on populations' gene flow and inform urban planners on the most effective ways to restore habitat connectivity.

First, using a population genetics approach, I addressed the question of whether or not urbanization has affected Auckland's Copper skinks and by what measure. Then, I investigated what key features of the urban landscape promote connectivity among Copper skink's remnant populations. For this purpose, a relatively young urban area was chosen as the system of this study: Auckland's North Shore (New Zealand). Urbanisation in the North Shore started in the first half of the 1900s and was mostly completed before the current century. This area displays several remnant vegetation patches scattered in a relatively young and homogeneously urbanized medium. These patches host a number of native lizards, among which the Copper skink (*Oligosoma aeneum* – GIRARD, 1858) is relatively abundant and widespread. Addressing this study system, I focus on three main research questions (Table 8): 1) Is gene flow in the sampled urban reserves is limited compared to that of rural sites? And if yes, is there spatial structuring in the surveyed populations? Since copper skinks are normally found in reserves and in tight association with vegetation, it is likely that urban land cover types (i.e. impervious surfaces) act as barriers causing a reduction in gene flow in the urban sites compared to the rural ones. 2) Did the hypothetically reduced gene flow among urban reserves cause Copper skinks' genetic diversity to shrink? The small size of urban populations is likely to shrink urban populations' genetic diversity compared to rural populations. 3) Which urban features are acting as barriers to gene flow and which promote connectivity? Impervious surfaces such as roads, houses and carparks possibly act as gene flow barriers while relatively more natural land cover types such as trees, shrubs and gardens possibly facilitate dispersal.

Methods

A preliminary, qualitative survey in most of the North Shore reserves was conducted to determine the general distribution of Copper skinks. Hand searching, pitfall and Artificial Cover Objects (ACOs) trapping were used to detect lizards' presence and gain

opportunistic insight into the species habitat selection (this, together with expert-based information helped optimising the second fieldwork session where genetic samples were collected). All reserves found to host Copper skinks were revisited in a second instance (except for Pin Oak Drive reserve, where only a newborn specimen was found in the preliminary survey, Figure 9) and less than two millimetres of tail tip was sampled as a source of genetic material. A total of 212 Copper skinks were sampled from 12 reserves (Table 3); among these, five came from a rescue operation carried out by Bioresearches ltd. in the Rosebank peninsula (an area of Auckland separated from the North Shore by the Auckland Harbour sea). Sampled reserves were widespread across the study area, covering zones with different functionalities (e.g. industrial or residential) and with different proportions of natural land cover types (e.g. trees, shrubs or gardens).

Control sites without urbanisation (i.e. rural), against which to compare the measures of genetic distance and structure of the urban populations, where chosen in the Shakespear and Waitakere Regional Parks. Waitakere Regional Park is a large, non-fragmented native forest west of Auckland; Shakespear Regional Park is located north of Auckland at the end of the Whangaparaoa peninsula, where remnant forest patches are fragmented by paddocks.

Figure 9: Section of aerial photograph of the Auckland region containing all sites surveyed for Copper skink presence. The three regions in the black squares (Shakespear regional park, North Shore and Waitakere regional park) are enlarged in the following pages.

Figure 10: Aerial image of the North Shore of Auckland with the names of all reserves sites for Copper skink presence. Copper skinks were found in the sites in white and not found in the sites in grey.

Figure 11: Aerial image of Shakespear regional park with the names of all sites surveyed for Copper skink presence. Copper skinks were found in the sites in white and not found in the sites in grey.

Figure 12: Aerial image of Waitakere regional park with the names of all sites surveyed for Copper skink presence. Copper skinks were only found in the "Dam Walk" site.

To sample Copper skinks' genetic material, lizards where captured either by hand searching or funnel traps. These were hand built out of a metal mesh, the main body of which was made of stainless steel (~2mm mesh size) while the entrances to the trap were made of a finer, metallic, window's insect mesh. The two entrances to the traps were removable and held in place by rubber bands. Traps were baited with cat food and loosely filled with vegetation to provide shelter to trapped animals, and were placed directly on the ground, below the soil litter, in areas with rich low vegetation (undergrowth), typically in connection with a plant, rock, or dead wood. Traps were spread throughout the reserve, wherever suitable habitat was found, in order to ensure the collection of a representative sample and avoid targeting small areas were groups of relatively closely related animals may reside. Trapping time for a reserve was two to three

weeks; during this period, traps were checked daily. The missing tail tip that was collected made recaptures easily recognisable for several weeks from the time of the sampling. All animals were released just after handling at the same location where they were found.

Samples (Table 3) were collected in three separate batches. I) Ten specimens were used to test PCR amplification. II) a successive, preliminary sample of 49 specimens from 5 urban reserves (including the 5 specimens from the Rosebank peninsula) and 2 specimens from Shakespear Regional Park was used to assess the ideal per-reserve sample size and to test for the presence of genetic structure. Since enough group separation was shown in the latter test, confirming the feasibility of the study. A third III) batch of 153 specimens was collected.

Location name	No.specimens	Land cover	Vegetation	Size [Ha]
Karaka	16	Urban	Natural forest	1.1
Little shoal bay	15	Urban	Natural forest/shrubs	29
Tuff crater	19	Urban	Natural-	14
Lakeroad	21	Urban	Natural-forest/shrubs	1.1
Watercare	19	Urban	Plantation	8.7
Centennial park	14	Urban	Natural/Plantation	28
Eskdale	19	Urban	Natural-forest	71
Saddleback raise	18	Urban	Natural-forest	12
Northcross	20	Urban	Natural-forest	4.3
Mountbatten	27	Urban	Natural-forest-shrubs	0.51
Kauri point	14	Urban	Natural-forest	70
Treetops	10	Urban	Natural-forest-shrubs	21
Pin oak drive	1	Urban	Natural-forest	12
Rosebank peninsula	9	Urban	Natural-forest	2
Waitakere damwalk	4	Rural	Natural-forest	2.2
Shakespear kauripoint	16	Rural	Natural-shrubs	0.7
Shakespear	17	Rural	Natural-forest	17

Table 3: Number of Copper and Ornate skinks found per site. 34 sites were surveyed with Coppers being found in 17.

Following collection, tissue samples were sent to EcoGene Genetics Laboratory at Tamaki (Auckland) for genotyping. 20 microsatellite markers developed by Berry *et al.* (2003) for *Oligosoma grande* were tested for the study species and ten markers successfully

amplified (Table 4). Microsatellites peak size data from my samples was binned with the software TANDEM (Matschiner & Salzburger 2009) and FlexiBin (Amos *et al.* 2007) and results were compared. I then removed suspected null alleles at each locus using the Brookfield method (Brookfield 1996) on each population independently with the Micro-checker software (Van Oosterhout *et al.* 2004).

3

Locus	No. alleles
Locus3	25
Locus4	18
Locus6	32
Locus7	29
Locus8	24
Locus10	26
Locus13	29
Locus14	3
Locus17	15
Locus19	2

Table 4: Number of amplified alleles per locus for the total Copper skinks sample.

The minimum per-reserve sample size was assessed via rarefaction analysis: the number of alleles per sample size per population was plotted and the minimum sample size was selected based on where the functions reached their asymptote (Figure 17).

To measure gene flow levels among the study populations (research question 1), I computed pairwise Gst (Nei 1977) and Gst' (Hedrick 2005) and D (Jost 2008) indexes, analysed the genetic structure of my sample with the Structure algorithm (Pritchard *et al.* 2000) and performed a per-population Principal Coordinates Analysis (PCoA, following Jombart *et al.* 2009) with R package PopGenReport (Adamack & Gruber 2014). Genetic diversity (research question 2) was measured by analysing per-population number of alleles, private alleles and allelic richness, calculated using the method of El Mousadik & Petit (1996). To understand the role of different urban land cover types as either dispersal

barriers or corridors (research question 3) I adopted a population-based LG approach (Figure 13) as samples were clustered within the reserves and no sample was available from outside reserves (see Gruber & Adamack, 2015). I compared the pairwise genetic distance matrix (Jost's D, Hedrick's G_{ST}' and Nei's G_{ST}) for the North Shore populations with a set of pairwise, least cost path distance matrixes. These latter matrixes were computed by measuring the lengths of pairwise least cost paths over different resistance surfaces. Resistance surfaces were rasters whose cell values represented the equivalent number of cells to add to the least cost distance whenever such a cell was crossed by the least cost path, for example: crossing a cell with a resistance value of ten is equivalent to crossing ten cells with a resistance value of one – this dictates whether it is worthwhile to cross barriers or disperse around them. Varying resistance values were attributed to different land cover types in order to find which set of values generated the resistance surface that better correlated the least costs path distances with the genetic distances. Firstly, single landscape features were given high resistance values while leaving the remaining cells with "zero" resistance (e.g. all roads=1000, other cells=1), pairwise least cost distances were then computed on the resulting resistance surface. Successively, combined resistance surfaces were produced by attributing different resistance values to different features (e.g. Motorway and houses = 1000, large roads= 500 and small roads =100) and new pairwise least cost paths were computed. Correlation between the pairwise least cost distances and the genetic distances was tested with Multiple Matrix Regression with Randomization analysis (R package PopGenReport's wrapper function based on MRM function of package ecodist) and with Partial Mantel tests ("wassermann" function from PopGenReport package) correcting for the effects of the pairwise Euclidean distances between populations.

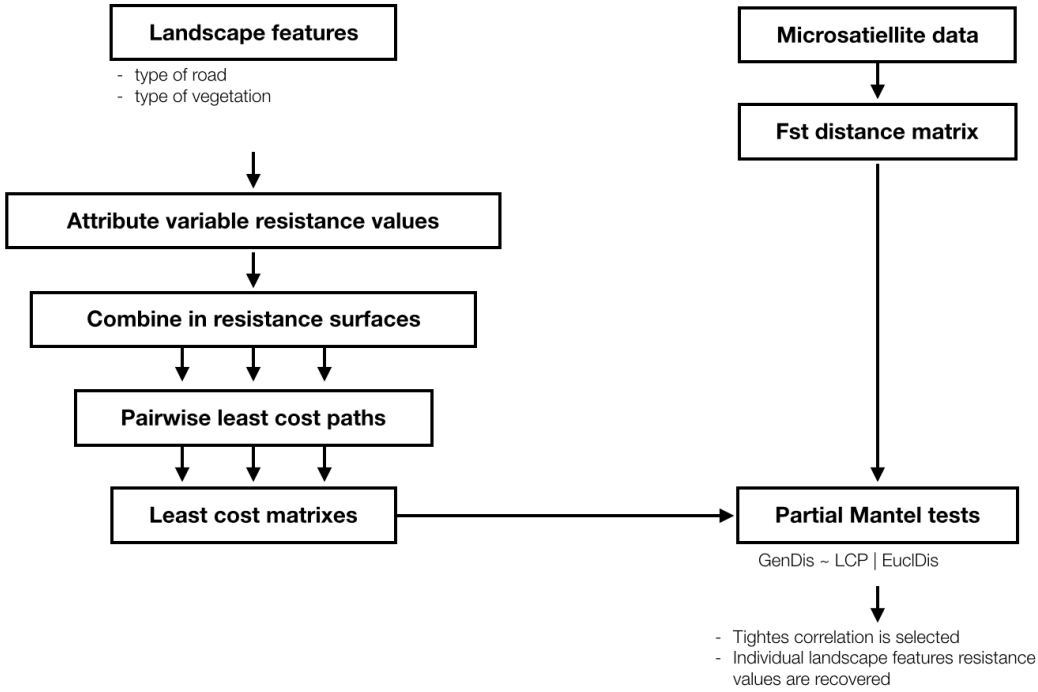


Figure 13: Flow chart of the landscape genetics analysis. Selected urban landscape features are attributed several different resistance values. Each set of values is combined into a resistance surface. The resulting resistance surfaces are used to compute pairwise resistance matrixes between reserves and compiled into resistance matrixes. Each of these resistance matrixes is compared to the genetic distance (Fst) matrix from the microsatellite data with a partial mantel test. The resistance matrix yielding the highest correlation to the genetic distance matrix is selected and the individual features' resistance values are retrieved.

The land cover types used in the LG analysis included vegetation height from a Light Detection And Ranging (LIDAR) derived map of the North Shore (Figure 14, bottom), obtained from Auckland Council's Biodiversity Group. This raster layer has cell values representing the land cover's height from the ground. Roads and buildings were subtracted from this map, the remaining cells were down sampled to 1 square meter and values were classified in to "grass" [0; 0.75]m, "shrubs"]0.75; 3]m and "trees"]3; max]m. Roads and buildings were derived from the Auckland Council map of impervious surfaces (Figure 14, top). This vector map was transformed into a raster layer by buffering all linear features (roads): motorways were given a six metres buffer (they were originally mapped with double line features, one per direction, resulting in a final feature at least 12 meters wide); "arterial" and "major" roads were also given a six metres buffer (these were

single line features); "medium" roads were given a five metres buffer and "minor" roads four meters. Buffer sizes mostly resulted in very close approximations of the actual road width when compared to aerial images. At the raster resolution used in the analysis, roads always had the minimum width required for preserving seamlessness.

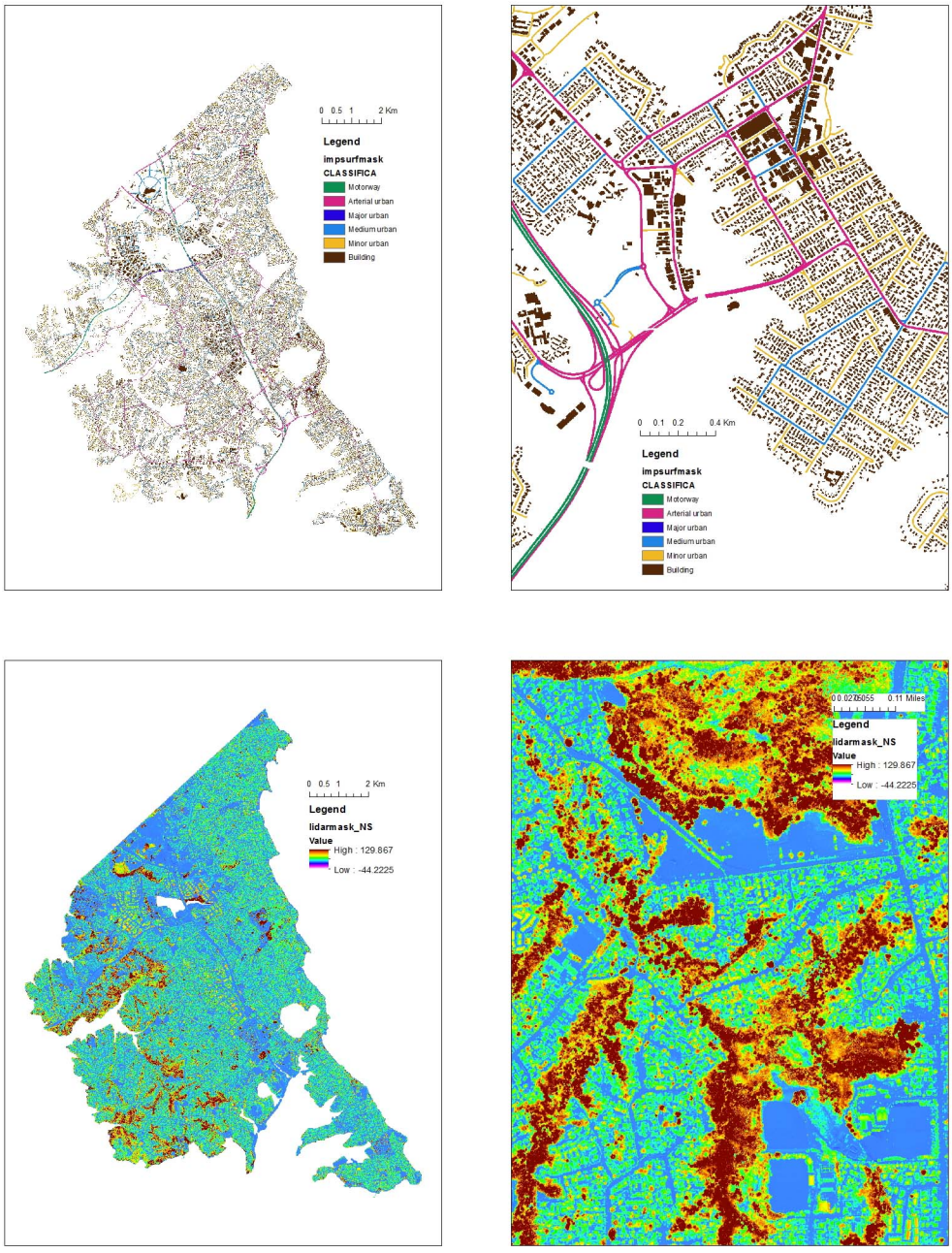


Figure 14: Example of landscape features maps. Top left: imperious surfaces – roads and buildings. Top right: higher magnification imperious surfaces. Bottom left: LIDAR derived vegetation height map (buildings were removed by attributing low values to pixels occupied by impervious surfaces). Bottom right: higher magnification LIDAR.

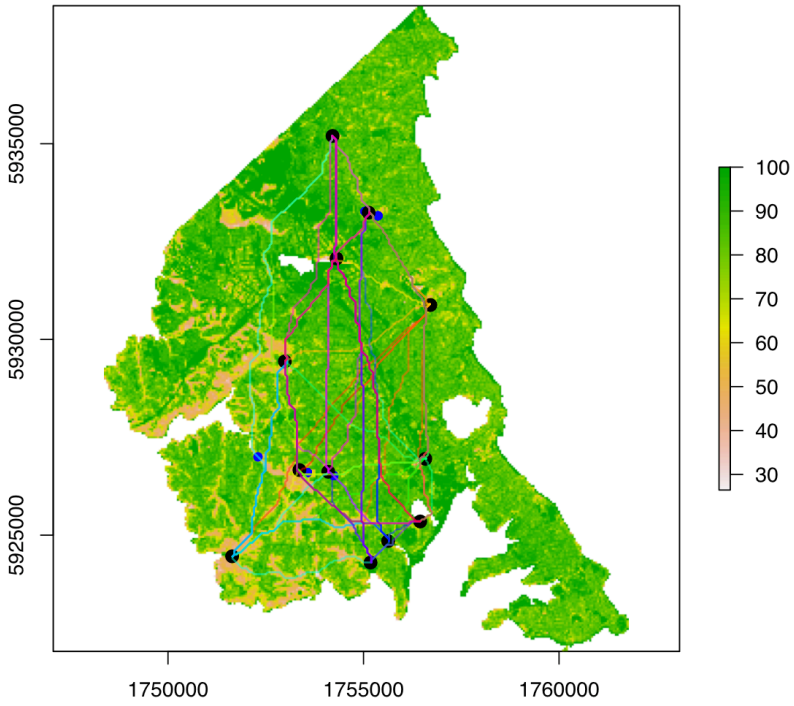


Figure 15: Example resistance surface derived from the LIDAR map of Figure 14 by attributing low resistance values to high vegetation (trees and shrubs) and high resistance values to low height vegetation and impervious surfaces.

Landscape maps were transformed into resistance surfaces (Figure 15) by first attributing varying values to different features and then combining different maps (e.g. roads map with vegetation map) by summing cell values, and scaling the resulting raster to values between 1 and 1000. Values were then changed to give different features varying importance as described above and in (Figure 13).

In order to break possible correlations between the pairwise Euclidean distances and the pairwise least cost distances, two additional resistance surfaces were made. A clustered map of road density (all road types apart from motorway and arterial roads) and one of building density were made by computing focal values for neighbourhoods of each cell using a moving window of Gaussian weights (sigma=40 for buildings and sigma=90 for roads; window size was three times sigma). This approach forced the least cost paths to deviate from the Euclidean path to avoid high resistance clusters lengthening therefore the

Euclidean distances and increasing the chances of breaking possible correlation with genetic distances. These clustered maps were then combined, resistance values for motorways and arterial roads were added, and the resulting cell values were scaled between 1 and 1000.

Results

Copper skinks were more easily found in urban reserves than in rural or natural areas such as the Waitakere or Shakespear regional parks. Three weeks of search and trapping in the Waitakere Ranges Regional Park only yielded four specimens, all of which were found in a single day in two adjacent funnel traps along the road to the Waitakere dam ("damwalk" site, Figure 9), off Scenic Drive. Capture rates in Shakespear Regional Park were higher than in the Waitakere, but Ornate and Moko skinks (*Oligosoma ornatum* and *O. moco*) were much more common catches than Copper skinks.

From a habitat selection point, Copper skinks are more abundant in ecotones, especially damp transition areas from forest to open vegetation (e.g. towards grassy road sides following forest margins or at the interface between forests and gardens). Towards the inside of forests, Copper skinks favour areas of thick, continuous leaf litter and patchy tree coverage. The highest capture rates occurred during warm humid days after abundant rainfall in spring time and early summer. The season with the highest catch rate is spring, followed by autumn. During the winter, Copper skinks are less active and more difficult to trap, however, hand searching still provides acceptable results. Sampling in urban areas during the central summer months yields the poorest results.

The rarefaction analysis performed on the first and second batches (I&II) of preliminary samples shows that, by 20 specimens the curve does not quite reach an asymptote, but the contribution of additional specimen becomes very small (Figure 17).

Therefore, for practical reasons, 20 was considered a sufficient sample size to encompass within-population genetic diversity. Results of the PCoA on the first and second batches (I&II) shows a clear structure in the genetics of the sampled populations (Figure 19). This was considered enough for supporting further landscape genetics analysis and justify the collection of the third batch of samples.

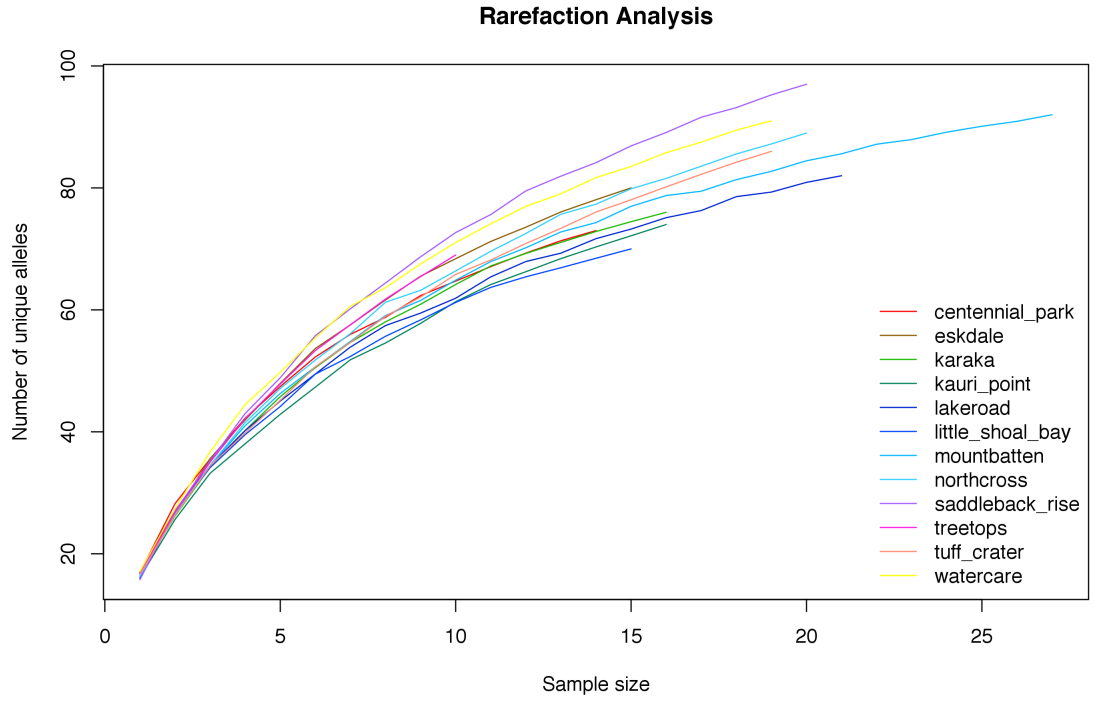


Figure 17: Rarefaction analysis conducted on all the sampled sites – number of unique alleles (y axis) per sample size (x axis) per reserve (colours). 20 specimens was considered a sufficient sample size because of the limited contribution of additional specimens.

Null alleles' tests highlight high percentages of null alleles for loci 3, 4 and 7 (Figure 18, white). This improved after removing suspected null alleles with the Microchecker algorithm (Figure 18, grey).

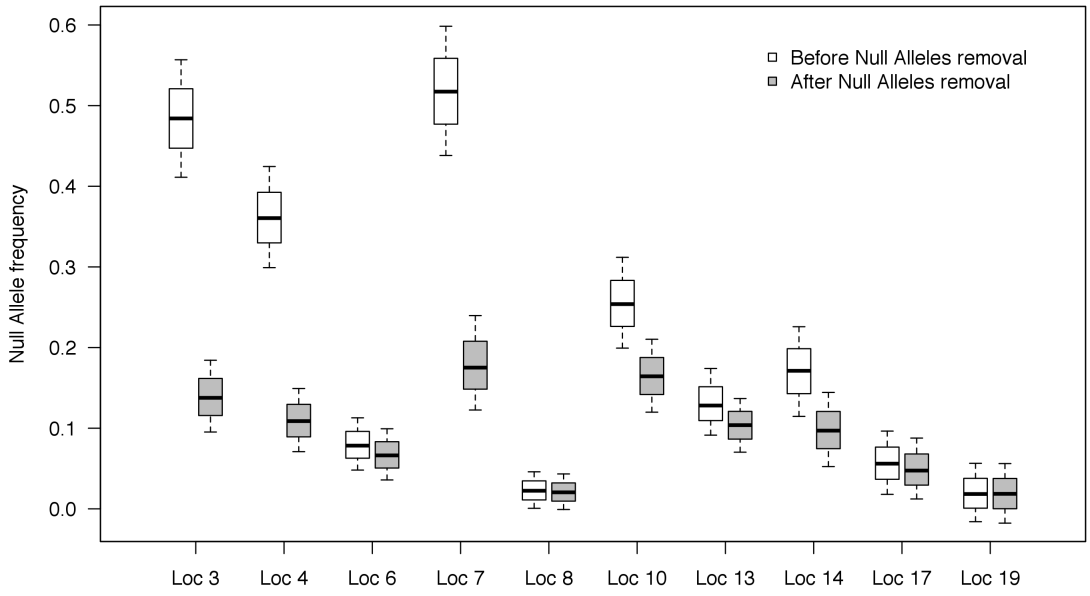


Figure 18: Null alleles per-locus for the North Shore (assuming HW equilibrium for the North Shore metapopulation) computed with the Brookfield method before (white boxes) and after (grey boxes) the removal of suspected null alleles with the Micro-Checker algorithm.

Answering research question 1), Structure's algorithm classifies the North Shore dataset into two main populations (Table 5, Figure 21). The three north-eastern populations of Northcross, Saddleback Rise and Watercare cluster together; the population of Centennial park acts as transitional population falling in between the two groups; and the remaining south-western populations group together in a second cluster.

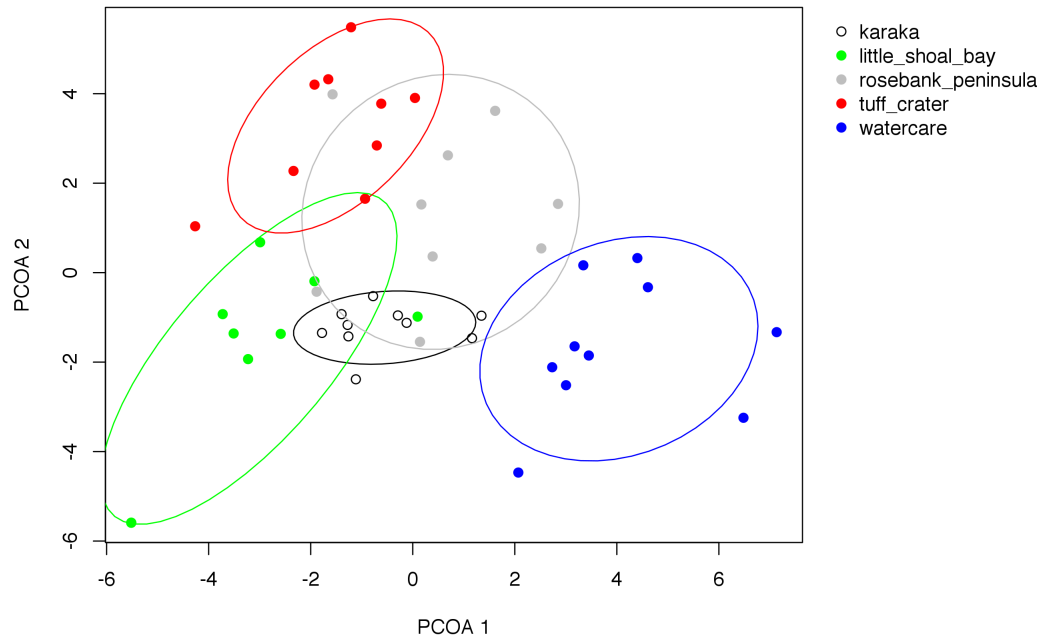


Figure 19: Principal Coordinates Analysis of the pairwise Fst values of the first and second batches on samples. At this stage good structure emerged from the collected data. However, newly sampled reserves fell in between the ones on this plot reducing the overall structure (see Figure 21. The ellipses are drawn at confidence interval levels of 0.7).

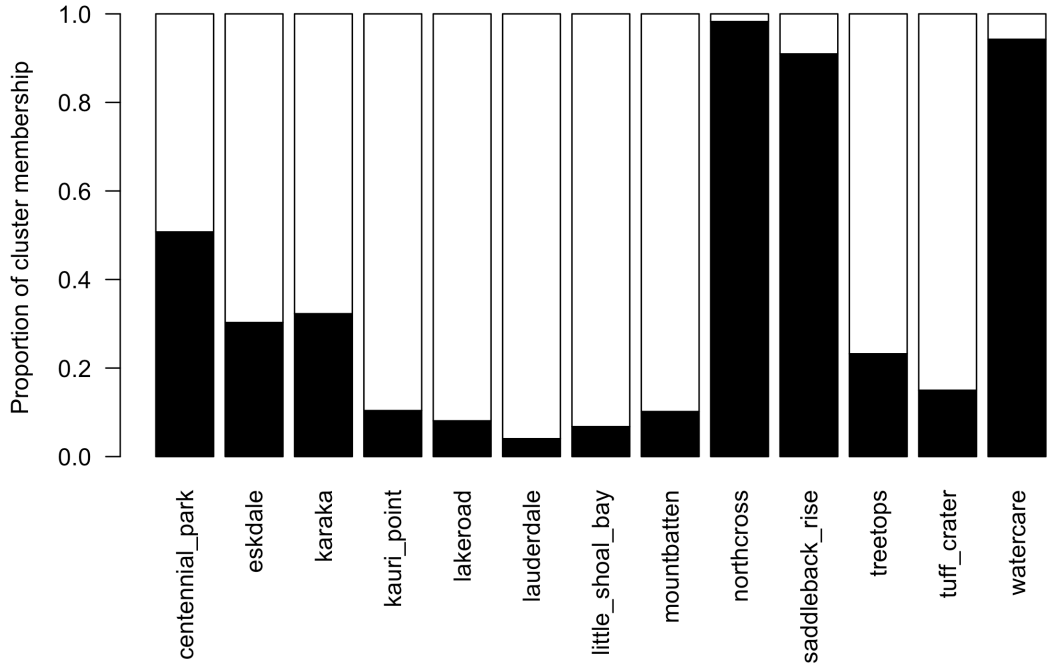


Figure 21: Proportion of membership to either of the two clusters identified by Structure algorithm for the sites in the North Shore. White portion: South West populations. Black portion: North East populations.

In spite of the neat structuring obtained from the PCoA performed on the first and second samples batches (Figure 19), the same analysis carried out for the full dataset

highlights only moderate genetic structuring, with newly sampled populations falling in-between those of the first batch (Figure 22). Shakespear Regional Park's populations are markedly distinct from other populations prior to the removal of null alleles, but they mix with other samples after removal. The few specimens from the Waitakere Ranges Regional Park are widespread across the plot, show high levels of genetic diversity compared to the average of the other populations. By removing rural sampling sites from the analysis, the structure within the urban populations becomes clearer. Similarly, to the results of the Structure algorithm, the north-eastern and south-west populations cluster on the two sides of the plot.

#K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-9715.8	0.9153	NA	NA	NA
2	10	-9325.61	6.5489	390.19	130.01	19.852287
3	10	-9065.43	5.9382	260.18	85.68	14.428591
4	10	-8890.93	21.2278	174.5	39.75	1.872543
5	10	-8756.18	38.3176	134.75	18.82	0.491159
6	10	-8640.25	25.8411	115.93	23.45	0.907469
7	10	-8547.77	14.8152	92.48	18.65	1.258846
8	10	-8473.94	29.5448	73.83	7.78	0.263329
9	10	-8392.33	45.6146	81.61	66.93	1.467294
10	10	-8377.65	43.8201	14.68	8.54	0.194888
11	10	-8354.43	87.6948	23.22	21.27	0.242546
12	10	-8309.94	89.8131	44.49	NA	NA

Table 5: Structure algorithm output. Each row reports different statistics averaged from for ten replicates of increasing number of clusters (#K). The highest Delta K value is selected to indicate the most likely number of clusters.

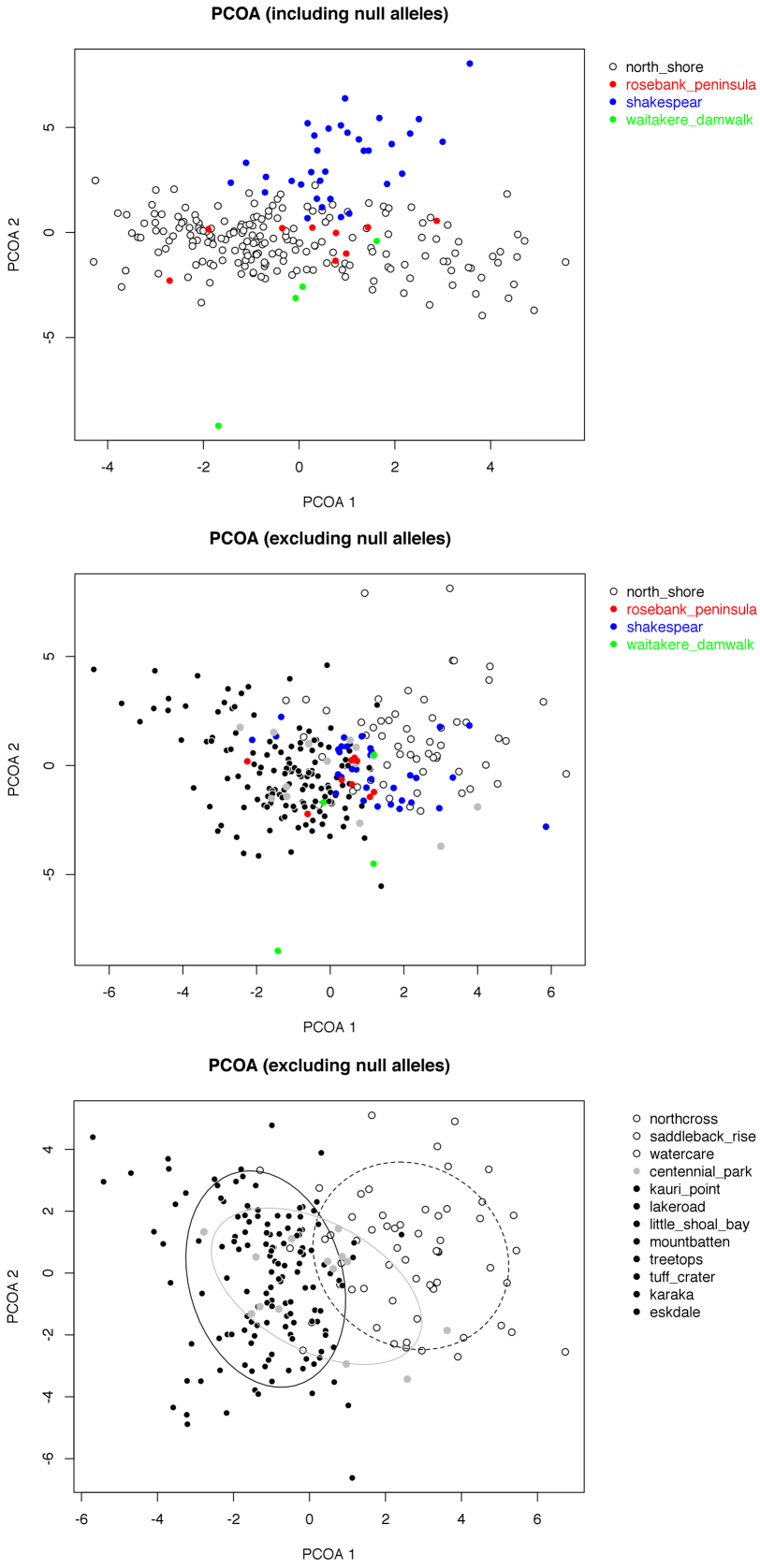


Figure 22: PCOA plots. Top: all samples before removing null alleles (note how samples from Shakespear regional park stand out here). Middle: same as top plot but after the removal of null alleles. Bottom: North Shore populations, colours are coded to reflect Structure's output – North East populations in white; South West populations in black; transitional population of Centennial Park in grey.

Regarding research question 2), the colour-coded table of allelic richness (Table 6) shows the highest levels of diversity occurring in the north-eastern populations and the lowest in the South-western populations, with transitional populations appearing in between these. The control sites have the lowest overall genetic richness, even though these populations have relatively high numbers of private alleles.

	No. indiv.	No. alleles	No. priv. alleles	Mean all. richness	Locus 3	Locus 4	Locus 6	Locus 7	Locus 8	Locus 10	Locus 13	Locus 14	Locus 17	Locus 19
waitakere_damwalk	4	40	1	3.12	3.9	3.13	3.73	2.91	4.1	3.2	4.63	1.97	2.6	1
shakespear_kauripoint_navy	16	68	3	3.44	3.5	3.69	4.79	3.81	4.46	3.68	3.96	1.95	2.68	1.87
shakespear_kowhaiglen	17	78	2	3.63	4.24	3.63	4.58	4.1	4.81	3.33	4.6	1.97	3.19	1.85
northcross	20	92	5	3.73	4.49	4.26	4.81	3.86	4.36	3.97	4.97	1.96	3.32	1.27
saddleback_rise	20	102	0	3.85	3.94	4.5	4.7	4.8	4.48	4.35	5.06	1.66	3.64	1.38
watercare	19	97	3	3.86	4.17	4.17	4.61	4.49	4.88	4.53	4.99	1.97	3.32	1.5
centennial_park	14	76	0	3.68	4.01	4.08	4.62	3.69	4.28	4.21	4.42	1.97	3.66	1.84
treetops	10	70	2	3.74	3.93	4.17	4.13	4.45	4.3	4.34	4.36	1.98	4.08	1.65
karaka	16	79	1	3.72	4.3	4.63	4.54	4.06	4.34	3.84	4.71	1.84	2.97	1.95
mountbatten	27	95	2	3.81	4.39	4.21	4.41	4.34	4.6	3.8	4.96	1.94	3.57	1.84
eskdale	15	83	1	3.81	4.2	4.68	4.81	4.21	4.58	4.09	4.81	1.89	3.11	1.68
tuff_crater	19	86	1	3.58	3.85	3.31	4.62	4.28	4.84	3.83	4.58	1.72	2.94	1.85
little_shoal_bay	15	72	1	3.64	4.2	3.99	4.78	3.7	4.48	3.76	4.56	1.97	3.03	1.98
lakeroad	21	88	2	3.61	4.12	3.91	4.81	3.34	4.25	4.69	4.31	1.94	2.84	1.85
kauri_point	16	78	0	3.55	3.98	3.62	3.53	4.59	4.27	4.77	4.38	1.87	2.99	1.46
rosebank_peninsula	9	60	1	3.46	2.85	4.31	4.59	4.16	4.65	3.86	3.92	1.96	2.46	1.88

Table 6: Allelic richness. Cells are colour-coded per column with red hues attributed to high allelic richness and green hues to low richness.

The colour coded table of genetic distance Gst' (Table 7) shows overall high levels of genetic differentiation (index values are restricted between zero and one, with zero representing complete admixture and one complete isolation). Gst (Nei), and D show the same patterns of Gst' (not shown): the north-eastern populations of the North Shore have high levels of differentiation compared to the South-western ones. The Waitakere samples are the most differentiated from all the others while the Shakespear populations are just comparatively less differentiated from all the North Shore samples. Rosebank peninsula samples are the second most diverse ones in the dataset.

	centennial_park	eskdale	karaka	kauri_point	lakeroad	little_shoal_bay	mountbatten	northcross	rosebank_peninsula	saddleback_rise	shakespear_kauripoint_vy	shakespear_kowhaiglen	treetops	tuff_crater	waitakere_damwalk	watercare
eskdale	0.15730428															
karaka	0.27947262	0.29383011														
kauri_point	0.27442521	0.11015054	0.35645487													
lakeroad	0.226821	0.26223555	0.25837225	0.31288438												
little_shoal_bay	0.26132402	0.18186362	0.27820008	0.34850616	0.19055534											
mountbatten	0.21505339	0.14429444	0.2577181	0.19994443	0.3181572	0.22535998										
northcross	0.35054381	0.25363262	0.32383768	0.26073187	0.31960408	0.42033468	0.23102972									
rosebank_peninsula	0.37831584	0.34340702	0.26657955	0.30982772	0.29413985	0.42042743	0.3446032	0.32208636								
saddleback_rise	0.24552828	0.3246622	0.25238911	0.37545519	0.2372094	0.42556224	0.32914791	0.21846162	0.27557281							
shakespear_kauripoint_vy	0.31801959	0.28568169	0.33245941	0.35510596	0.24027743	0.33896904	0.37437464	0.32229579	0.31490892	0.27761492						
shakespear_kowhaiglen	0.27934775	0.26628206	0.29265014	0.31471239	0.25597139	0.32272772	0.31124649	0.30177391	0.33526576	0.2325014	0.14587					
treetops	0.10874161	0.09361718	0.24326124	0.14881872	0.2350037	0.28781271	0.17808853	0.26708369	0.32253137	0.2767117	0.26193856	0.27455202				
tuff_crater	0.33423687	0.20936642	0.34614455	0.26412998	0.34428787	0.27108389	0.27394376	0.32516689	0.38026706	0.43454909	0.38734274	0.41244338	0.24996385			
waitakere_damwalk	0.43689186	0.36076561	0.5178801	0.34996646	0.47557835	0.51730605	0.352459	0.35010062	0.35645649	0.44221853	0.51216289	0.44111759	0.37729681	0.41875861		
watercare	0.19803597	0.206669	0.27870952	0.22254584	0.27870585	0.34416393	0.2694217	0.19676555	0.28240386	0.15955976	0.28777479	0.26091867	0.19919161	0.33144982	0.3498405	

Table 7: Pairwise Gst' (Hedrick) for all the populations. Cells are colour-coded per column with red hues attributed to high allelic richness and green hues to low richness.

For what concerns research question 3), the LG analysis did not find any effect of the landscape features on the genetics of urban Copper skink populations (Figure 23). No significant correlation was found between least cost paths and genetic data corrected for Euclidean distance under any of the tested genetic distance metrics (Jost's D, Hedrick's G_{ST}' and Nei's G_{ST}) for both the partial mantel test and the MMRR analyses. Correlations were found between all least cost paths and Euclidean distances (Figure 23), which even the use of the clustered maps of road and building densities was not enough to break.

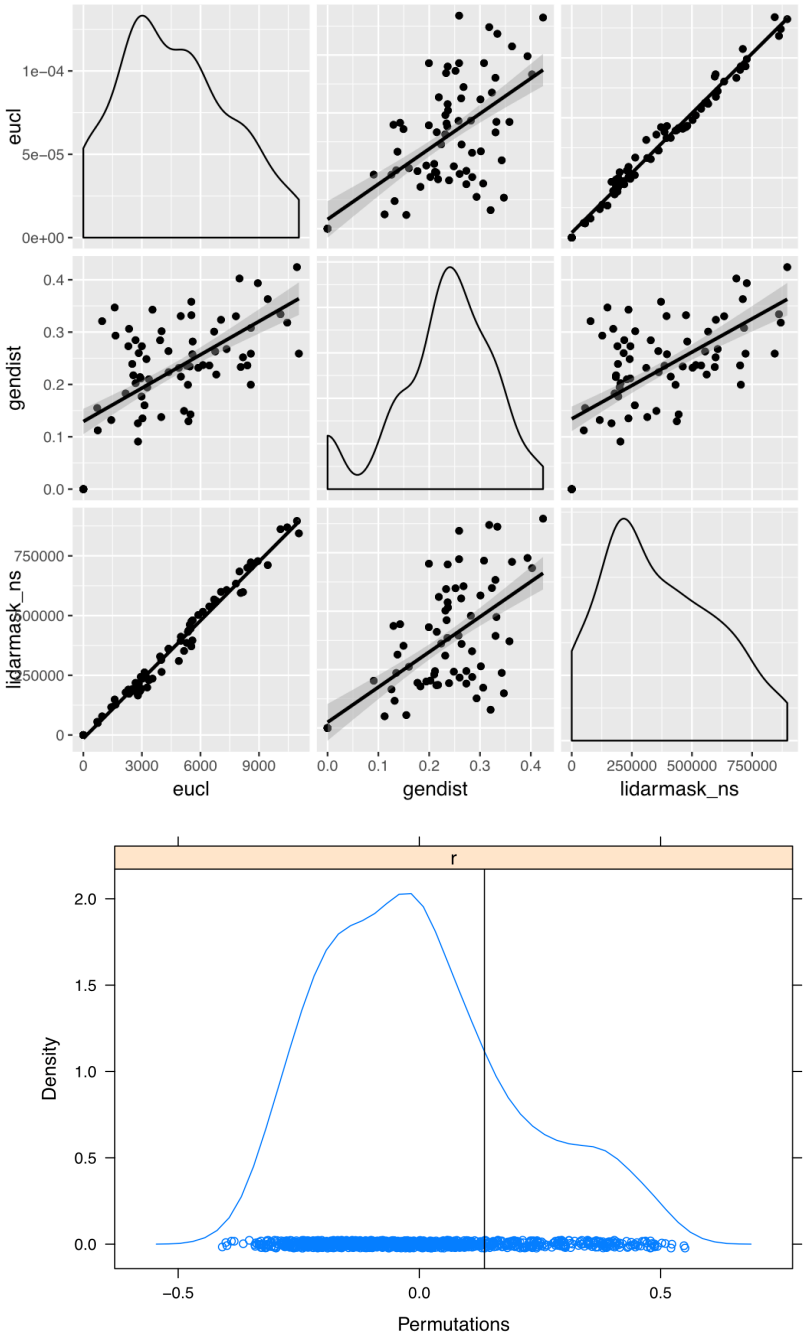


Figure 23: Example output of partial Mantel test. Top: correlation plots with regression lines and confidence intervals between the pairwise least cost paths matrix obtained from the resistance surface of Figure 15 (LIDAR derived vegetation cover, "lidarmask_ns") and the genetic distance matrix ("gendist"). Bottom: output of the partial Mantel test of model "gendist" ~ "lidarmask_ns" | "eucl", where the effect of the vegetation hight on the genetic distance is corrected by the Euclidean distance ("eucl"). This model, like all those obtained from all the tested resistance surfaces, is non-significative. Note that the high levels of correlation between least cost and Euclidean distances in the correlation plots on the top. This is a possible explanation for the lack of correlation.

Figure 15: Genetic structure of the North Shore populations as per the Structure algorithm. In cyan, the north-eastern populations of Northcross, Saddleback Rise and Watercare; in yellow the transitional populations of Centennial park, Karaka and Eskdale;in magenta the remaining South Western populations (Image © 2016 TerraMetrics, Image © 2016 CNESS / Astrium, Data SIO, NOAA, U.S. Navy, NGA, GEBCO).

Discussion

The results obtained from this study appear to indicate the early effects of habitat fragmentation (Table 8). Relatively high Gst' values show a considerable level of isolation between populations, especially in the South of the North Shore, where urbanization is older. The Rural sites of Shakespear Regional park also show a relatively high level of genetic isolation between themselves and low genetic diversity. This suggests that Copper

skinks may be affected by habitat fragmentation even if in small amounts, or if barriers are of semi-natural origin (like paddocks in the case of Shakespear Regional park). This level of genetic isolation is probably due to their secretive habits which keeps them away from open habitats and possibly limits their dispersal capabilities. Both the PCoA and Structure algorithm showed evidence of two main groups of populations within the North Shore: the north-eastern and the south-western. This split could be caused by the older age of urbanisation in the south-western North Shore – which isolated the south-western populations earlier than the others. The results are also consistent with the idea of the motorway being a major barrier to dispersal. The north-eastern populations are characterized by relatively high allelic richness, probably due to their proximity to rural areas.

Research Question	Methodology	Results	Conclusions
1) Is gene flow in the sampled urban reserves is limited compared to that of rural sites? If yes, is there spatial structuring in the surveyed populations?	Pairwise Gst (Nei 1977), Gst' (Hedrick 2005) and D (Jost 2008) indexes.	High genetic structure and low gene flow.	Gene flow is low on both rural and urban sites. Urban sites are structured in two main groups: North-eastern and South-western populations.
	Structure algorithm (Pritchard *et al.* 2000).	Two main clades. North-eastern and South-western populations.	
	Principal Coordinates Analysis (PCoA, Jombart *et al.* 2009).	Two main clusters. Again, North-eastern and South-western populations.	
2) Did the hypothetically reduced gene flow among urban reserves cause Copper skinks' genetic diversity to shrink?	Per-population number of alleles, private alleles and allelic richness (El Mousadik & Petit 1996).	Within the North-Shore, highest diversity in the North-eastern populations and lowest in South-western. Control sites were the overall lowest in diversity, but high in unique alleles.	Diversity shrunk in the more isolated, South-western populations compared to the North-eastern ones.
3) Which urban features are acting as barriers to gene flow and which promote connectivity?	Population-based Landscape Genetics (LG) approach (described in Figure 13).	No correlations found among populations genetics and landscape features.	Young age of urbanisation and other variables potentially masked the results of this analysis.

Table 8: Summary of research questions, methodology employed to answer them, results and their discussion.

The landscape genetics analysis did not highlight any correlation between landscape and genetic structure. This could be due to a few reasons, including: the potentially low number of surveyed reserves and sampled specimens per reserve; the short isolation time due to the young age of urbanized land in Auckland; the weak nature of the urban barrier; the spatial pattern of land cover types in the study area. The sample size option is unlikely to be the only explanation for the lack of a landscape effect, as the number of specimens per population appears to capture most of the genetic diversity present within each location. Sampling of more populations would have possibly increased the chances of detecting a signal but all suitable reserves were sampled and the number of reserves in the study area is limited. It seems likely that the analysed populations have not had

enough time to genetically respond to the urbanisation process. For example, Copper Skinks reach sexual maturity at the age of two-three years (Chapple *et al.* 2008) and the motorway was built between 1975 and 1983 (New Zealand Transport Agency 2008) which allowed 15-20 generations to pass prior to the start of this study. Many factors affect the lag time for microsatellite data, for example mutation rates, length of fragment, natural dispersal ratios and the degree of isolation caused by the barrier. However, Hedrick's Gst is considered to be quite sensitive to the appearance of new, strong barriers with their genetic signature becoming detectable within 15 generations (Landguth *et al.* 2010). Clearly this time can be prolonged in the case of weaker barriers, and changes in any of the above mentioned parameters. Likewise, the spatial distribution of the different land cover types can be partly responsible for the impossibility of detecting any effect of landscape features on the genetic structure of the sample. When Euclidean and least cost paths distances are correlated, genetic variability can be explained by distance alone, while a better distribution of land cover types (more spatially clustered rather than evenly distributed) would have helped break this correlation.

The findings of this study fit within the existing literature on urban herpetofauna. Noel *et al.* (2011) found comparatively low genetic diversity in urban populations of Blue-spotted salamanders compared to rural sites and attribute it to, among other causes, the high mortality rates and low reproductive outcome of the small urban population isolated by the urban matrix. Similarly, the Red-backed salamander (Noël *et al.* 2007) populations of the same area were found to be less genetically diverse and more structured than their rural counterparts. The same pattern was found in Common frogs populations (Hitchings & Beebee 1997), where, in spite of the absence of clear barriers, urban populations were much more differentiated and less diverse than nearby rural populations. Similarly,

genetic isolation of three lizard species in California was shown to be higher for populations in older urban areas and in the presence of road (Delaney *et al.* 2010).

Habitat fragmentation, however, is clearly not the only consequence of urbanisation that threatens urban herpetofauna. For example, Jellinek *et al.* (2004) found that lizard densities in remnant fragments of differing sizes near Hobart, Tasmania, are not affected by fragment size, but rather by environmental variables within the fragments. Gila monsters in the USA (Kwiatkowski *et al.* 2008), Common skinks in New Zealand (van Heezik & Ludwig 2012) and Blue-tongued skinks in Australia (Koenig *et al.* 2001) all make use of man-made structures in areas with low levels of urbanisation. So do Copper skinks, which have been found in gardens nearby remnant natural patches. Similarly, to Blue-tongued skinks, also Copper skinks may be more sedentary in urban areas and avoid road crossing, thus increasing isolation among populations. Predation, by reducing population size, can also be responsible low genetic diversity in urban areas. I expected higher diversity in Shakespear sites because of the presence of a predator-proof fence, however the fencing is relatively young (erected in 2011) and the effects of predator removal may have not yet generated a genetic response. Also, genetic diversity may be hindered by low populations densities caused by the possible competition with the more abundant Moko and Ornate skinks (which are absent or less common in urban areas).

Because of the young age of urbanisation in the North Shore, repeating this study either after a decade at least or by using more variable makers such as Single Nucleotide Polymorphisms (SNPs) could deliver clearer population genetics patterns. Given the high isolation and low genetic diversity found in this study, conservation measures such as translocations or opening of ecological corridors may be needed in the near future in order to maintain viable urban skinks populations. Apart from planned translocations, a good practice would be relocating individuals rescued from building sites to reserves

slightly away from the capture location. This would improve admixture levels and contribute to the maintenance of a viable, urban meta-population. Urban reserves, however, are complex ecosystems and, in order to have viable wildlife populating them, they not only need to be planned as an interconnected network, but attention should also be given to important ecological factors other than just habitat fragmentation. In fact, making micro-habitat in private gardens more available, or providing better predators control, would provide more widespread benefits that extend to the entire urban ecosystem.

Chapter 4: Climate migrants' survival threatened by "C" shaped anthropogenic barriers.

Abstract

The aim of this publication is firstly to raise awareness on the problematics related to the influence of anthropogenic dispersal barriers on climate migrants. In an effort to delineate this issue, I firstly present the notion of C-trap and the theoretical context from which it emerges. I also present a methodology which allows to implement the C-trap framework into conservation practices. As an example application, I look for potential C-traps across the world's terrestrial realm with the scale and spatial features to affect endangered, endemic animal species in the near future.

Climate change induced range shifts studies have raised widespread interest in the scientific community and concern among conservationists (Lenoir & Svenning 2015). However, even though studies on range shifts are very abundant, the role played by dispersal barriers is yet to be fully included into any modelling framework. Here, I introduce a novel concept where the interplay of range shifts and dispersal barriers of a particular spatial configuration can threaten the persistence of populations under climate change – I named this concept "C-trap".

After elaborating on the theoretical features of C-traps, I provide a simple method that combines spatially explicit environmental data and future climate projections to spatially locate them. As an example application, I then use such method to determine where high C-trap densities can further threaten the conservation of endangered, endemic animals across the world's terrestrial realm at a particular spatial scale, in a climate

change scenario. My methodology detected potential C-traps for the study system with areas of high density mostly located in east Europe, south Asia and North America.

Dispersal barriers add an additional dimension to range shift studies and can ultimately prevent otherwise successful climate migrants from tracking their shifting climatic niche. The methodology presented here is simple and flexible enough to be adapted to a wide range of taxa and geographical locations, and to be implemented further to account for the fast development of range shift modelling. I, therefore, encourage researchers to include the effects of anthropogenic dispersal barriers in range shifts modelling and in the planning of effective conservation strategies with reference to climate change.

Introduction

When considering climate warming (i.e. temperature only), range shifts cause species to move poleward, upwards on land, and downwards in the seas following isotherms (Loarie *et al.* 2009). However, when multiple climatic variables are considered, vectors of climate shift deviate from the temperature-only trend and examples of longitudinal shifts and even shifts towards the equator are found (Lenoir & Svenning 2015). In addition to that, recent studies observed range shift deviations from climate velocities (the speed and direction of climate shift at a location) and attributed this phenomenon to the significant role played by species-specific traits and anthropogenic influence for climate migrants (Lehikoinen & Virkkala 2016).

In addition to shifting, a climatic niche can also change its spatial extent. Both range expansions and contractions have been reported in the literature, although range contractions seem to be rarer or more difficult to detect (Thomas *et al.* 2006). Range contraction observations along with simulations based on future climate, highlight the risk of extinction, for example, in several mountain species (Gottfried et al., 2012; Parolo &

Rossi, 2008), where converging climate velocities cause a shrinkage of local climatic conditions. Burrows et al. (2014) called these convergence points absolute sink areas and divided them into costal and internal sinks, where costal sinks see converging climate vectors meeting a shoreline, while internal sinks are formed exclusively by converging climate vectors. Dynamics of climate sinks can be of two types: niche contractions, where there is a reduction in the area occupied by a particular climate (with the climate completely disappearing in extreme circumstances), and niche-distribution dissociation (Figure 16). Here I focus on this latter case, where the climatic niche does not shrink or disappear from the landscape but climate migrants are prevented from tracking it by dispersal barriers.

With the work presented in this chapter (and its subsequent publication) I aim at raising awareness on the problematics related to anthropogenic land use change for climate migrants. I will address this theme by firstly developing and exploring a novel conceptual framework based on the C-trap dynamics (detailed in the next chapter "Concept"), and, successively, by presenting an example application of this conceptual framework to endangered, terrestrial, animal species at the global scale.

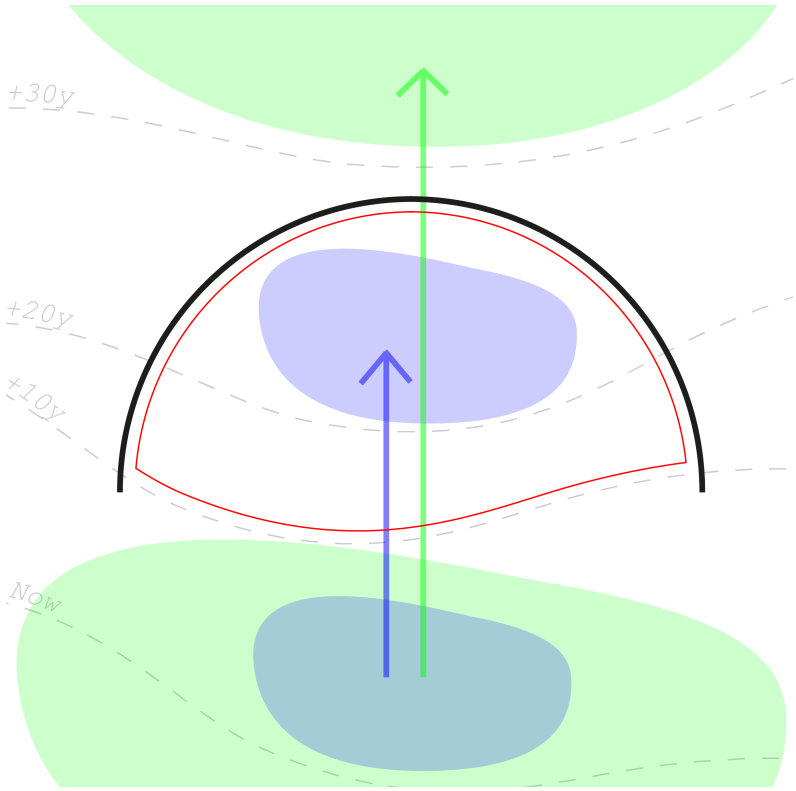


Figure 16: Diagram of a C-trap: in blue the distribution range, in green the climatic niche, in dashed climate isolines (shifting towards the top with years passing) in black the dispersal barrier, in red the trap. The C-trap is set when the lower maximum tolerable climate isoline (below the distribution range) reaches the opening of the "C" shaped barrier and blocks the distribution range in it. While the distribution range is blocked in the C-trap the climatic niche proceeds further generating niche-distribution dissociation and the consequent the local extinction.

Concept

I describe "C-trap" as a dispersal barrier, typically in the form of a "C", that generates niche-distribution range dissociation in climate migrants. These dispersal barriers lock climate migrants in an area while their climatic niche shifts elsewhere (Figure 16). The shape of this landscape feature is fundamental to make it a trap rather than just an obstacle: it must be oriented orthogonally to the climate shift vectors – thus preventing species from circumventing it – and it needs to be wide enough to contain the distribution range. The trap is set once the distribution range is contained in it and the maximum tolerable climate isoline (e.g. the maximum isotherm or precipitation isoline representing the trailing edge of the climatic niche) reaches the entrance of the trap, closing it. At this point, species cannot follow shifting climates any further, resulting in the shrinkage of their

distribution range by dissociation from its climatic niche. This process can potentially continue till the eventual extinction of the species, or local taxa (e.g. sub-species) (Figure 16).

The temporal and spatial scales of this phenomenon are particularly important. As the speed of climate shift increases, so does the area of the trap and the width of the distribution range it can trap. Also, the area of the trap (and, consequently, the trapped distribution range) increases with the length of time considered: the longer the time and the resulting climate shift, the deeper and more threatening the trap becomes. Similarly, the location of a C-traps relates to the intensity of its threat: C-traps laying along climate shift corridors – areas where climate velocity vectors converge to continue in a common direction at high speed – are likely to cause the highest potential for local extinction.

As opposed to the consequences of niche contraction, the effects of niche-range dissociations operated by C-traps are also taxa-specific. For example, a shore line can prevent certain land mammals from reaching another island with a suitable climate but it does not stop anemophilous plants. Similarly, some barriers completely stop certain species, but just slow down others (Spear *et al.* 2010). In this latter example, if the decrease in speed and the timeframe considered are big enough, conservation problems can still occur.

Landscape features that act as C-traps can be natural, such as shorelines or major rivers and lakes, or anthropogenic, such as urban areas or crops (assuming that cultivated areas will remain the same with regard to climate change, notwithstanding the changes to the cultivated species). Even if the land-use changes, its speed is expected to be negligible compared to that of climate change. Ordonez et al. (2014) in their study on the conterminous U.S.A., use land-use change metrics based on alternative economic incentives – as these are considered to be the main driver of land-use change in the next

50 years – to find that land-use change speeds are roughly an order of magnitude smaller than climate speed.

Normally, the effect of C-traps placed along shorelines is intrinsically included in climate shift studies by the geographic boundaries of the study areas themselves, as most environmental variables are mapped either on land or in the seas, providing an intrinsic global discontinuity. On the contrary, barriers of anthropogenic origin are rarely taken into consideration (Early & Sax 2011). As a consequence, it is hard to establish to what extent climate migrants are actually able to track their shifting climatic niche inland, where models have so far made the unrealistic assumption of infinite dispersal.

Methods

As an application of my concept, I located C-traps for land animals with small distribution ranges, that cannot disperse across cultivated land, at a global scale, for the next 50 years. To simplify my study system, I made a number of assumptions. 1) Climate change will occur following the Representative Concentration Pathways (RCP) 8.5, the *status quo* emission rate maintained for the next 50 years. This RPC account for the most drastic climate change scenario. It is the most conservative approach from a conservation standpoint, since some of the C-traps obtained under this model may not have existed under more optimistic RPCs. 2) This analysis was limited to the coming 50 years (average of climate predictions for 2061-2080). 3) The minimum size of C-traps was limited to 5,000 km^2, this is comparable with the maximum distribution range area for endangered species as defined by the IUCN Categories & Criteria for endangered species (version 3.1), B1 criterion. This ensured that C-traps found this way had the potential to affect local endangered species but excluded smaller C-traps.

To identify dispersal barriers, I combined spatial information on urban areas, cultivated land and shorelines at the global scale. I then fed the obtained boundary lines to a peak-finding algorithm to detect segments in the shape of "C" with a minimum depth of about 80 km to match point 3). Climate velocities were computed on the first principal component of 12 bioclimatic variable predictions for 2070 under RCP 8.8 following Hamann *et al.* method (2015, points 2) and 3)) and used them to delete those "C" shaped barriers that were not orthogonal to the climate velocities. Remaining "C" shapes were considered C-traps.

Climate velocity

Climate velocity vectors were obtained by using Hamman's method for univariate environmental variables (Hamann *et al.* 2015). The principal components analysis was performed on all bioclimatic variables from the WorldClim database for current and future conditions at ten minutes resolution (Hijmans *et al.* 2005). WorldClim rasters were resampled at a lower resolution of about 40 km of cell size at the equator. I used future climate variables built under the ACCESS1-0 global climate model for the RCP 8.5. Since almost all climatic variability was captured by the first principal component I used a univariate approach projecting the PC1 on present and future bioclimatic variables (Figure 17). Climate velocities were then transformed into a raster with angles and length values coded in each cell.

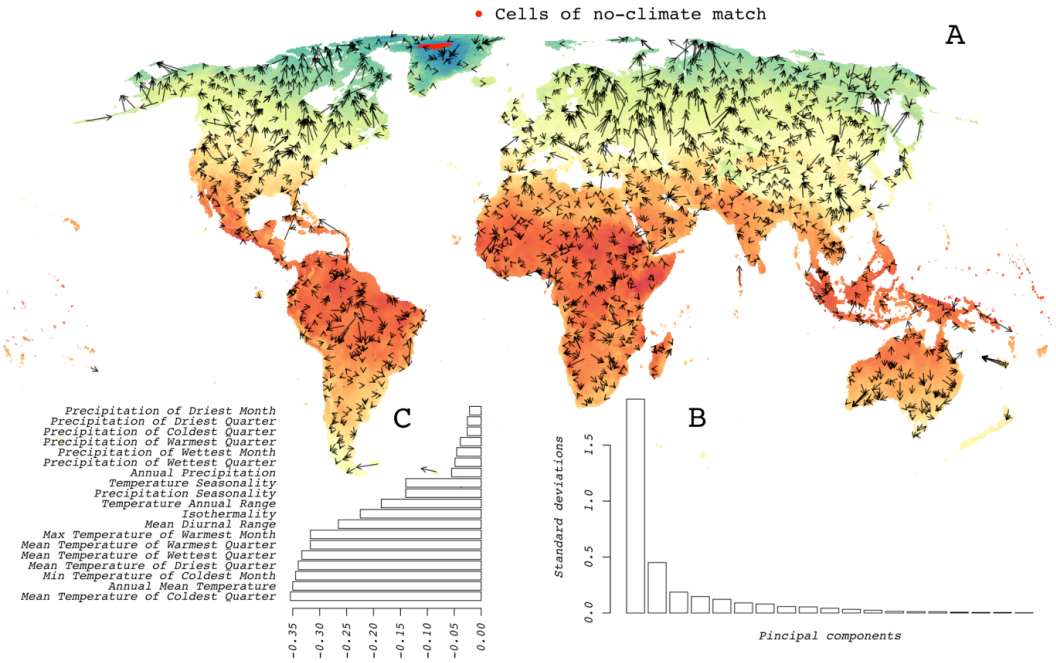


Figure 17: A: Map of the first principal component projected on current bioclimatic variables. Arrows indicate a sample of 1,800 climate shift vectors (distances of more than 1,000 km are omitted for clarity. In red, areas of climatic extinction – places with a climate that will not exist anymore by 2070. B: Variability captured by the principal components (histogram bars are PC1, …, PC19). C: Climatic variables loadings in the first principal component.

Anthropogenic dispersal barriers

I used raster files from EarthEnv Global 1 km Consensus Land Cover (Full Version 1.0, with DISCover) classes 7 (cultivated and managed vegetation, here also referred to as crops), 9 (urban/built-up) and 12 (open water) as a source of global data on anthropogenic and natural barriers. The original rasters were resampled at a lower resolution of about 30 km at the equator. Threshold consensus values were chosen to convert the original continuous variables into binary variables: water > 90%; crops > 65%; urban > 30% (cell values (consensus) in the original maps are the percentage of source land cover maps agreeing on a cell being classified as a specific land cover class). These proportions were selected to highlight particular features. For example, 90% water excluded smaller water bodies and rivers; 65% crops excluded less intensive cultivations; 30% urban excluded scattered buildings.

Two datasets were analysed in parallel: shorelines only and combined shoreline, crops and urban. This was necessary to detect traps formed by the combined effects of, for example, a cultivated area terminating against the coast. Areas which were covered by water, crops or urban centres were considered as barriers and converted into objects of class "sp::Polygons" in R (R Core Team, 2016). These polygons were subsequently transformed into "sp::Lines" and simplified using the Douglas-Peucker algorithm (Douglas & Peuker 2011) with a tolerance value of 80 km. This algorithm simplifies a line composed of segments producing a similar line with less segments. The tolerance value ensures that angles in the line features that are deeper than 80 km are preserved, giving traps a final minimum size comparable with the maximum distribution range IUCN criterion for endangered species. Simplified lines were then fed to a peak-finder algorithm based on the "pracma::findpeaks" function. This function detects peaks in one-dimensional numerical series; peaks in two-dimensional (spatial) line data needed to be decomposed in y (latitude) for north pointing peaks, -y for south pointing peaks, x (longitude) for east pointing peaks and -x for west pointing peaks. At this stage, any possible C-trap was simplified as a triangle (or peak) demarked by three points: the vertex and the previous and next points forming the peak edges.

In the situation of a barrier polygon containing a non-barrier polygon (a remnant natural patch inside a wider cultivated land), polygon boundaries (sp::Lines) were divided into "holes" and "solid", where "holes" lines were the borders of the internal polygon and "solid" lines the borders of the containing polygon. These lines represent barriers' boundaries, species could not *enter* solid lines, but they could not *exit* hollow lines. To determine whether a peak was open towards a hollow polygon (and therefore representing a real trap) or to a solid polygon (therefore not representing a trap) the centroid of the

peak triangle was tested on to see if it fell or not on the inside of the parental polygon with the "sp::point.in.polygon" function.

Filtering

Only peaks that were aligned with and pointing to the same direction as climate shift vectors were considered traps. To verify this criterion, I discarded all the peaks whose bearing differed from the one of the climate shift vector at the peak's vertex by more than *"90° - peak width"* (the angle of the peak corner in the peak triangle). This rule ensured that, with respect to the peak's edges, the component of climate change direction pointing inside the peak is bigger than the one pointing outside.

To discard traps entirely originated by the shoreline (with no anthropogenic component) I deleted all peak coordinates from the combined dataset that were also present in the shoreline dataset. Global density of traps of anthropogenic origin was obtained by computing a two-dimensional kernel density estimate of the traps coordinates.

I used a Robinson projection for all spatial data. This was a convenient compromise between retaining angles, areas, and shape values, all of which were of some importance for the computations. All computations were performed in R (R Core Team, 2016).

Results

The lengths of the computed climate velocity vectors have a positively skewed distribution with a mode of ~100 km. Under the projected scenario, few values of the current climate, as described in the first principal component, will disappear by 2070 (Figure 17 A). These values are found in northern Greenland, and represent the only case

of disappearing climates (Williams *et al.* 2007). All other current climates are expected to persist, somewhere on the globe, till 2070.

High speeds of projected climate shift vectors are mainly found on the temperate-cold regions of Eastern Europe, Canada and Northern/central Siberia. Climate vectors longer than 1,000 km were not plotted in Figure 17 for clarity, however, small oceanic islands often display very long climate shifts due to the obvious lack of land with similar climate at close distances.

A total of 742 C-traps have been detected in my application prior to filtering for climate velocities. Filtering reduces the total to 195 traps of anthropogenic origin. The area with the highest concentration of C-traps is Eastern Europe (Figure 17), followed by India, South-East China and Northern South-East Asia. The eastern European trap cluster is also located in an area of high climate velocity, which makes these traps faster and more dangerous. These high C-trap density areas are characterised by a pattern of managed-to-natural vegetation, with a patch size of about 5,000 km^2 as a consequence of the selected 80 km tolerance value in the Douglas-Peucker algorithm.

Discussion

I detected several hundred C-traps which could potentially threaten terrestrial, endangered animal species that do not disperse through anthropogenic land cover types within the next 50 years at the spatial scale examined here. Not all of these C-trap areas necessarily overlap with ranges of species with a limited distribution, however, where this condition does occur, it would be sensible to further investigate the conservation strategies for such species with respect to climate shifts. The effects of climate change on the conservation of species, especially if endemic to a C-trap area, can be of primary importance and measures to ensure connectivity between natural patches may need to be

accounted for in medium-term conservation strategies. Implementing the C-trap framework to the local scale and specific taxa also allows to set the trap's temporal and spatial scale more accurately: small distribution ranges, and high climate velocities make for a higher threat level, while the opposite is true for low climate velocities and widespread distributions. Also, the employment of taxon-specific barrier types would make for a more accurate identification of susceptible areas.

At large, the distribution of high C-trap density areas indicates that, with regard to climate change, a mixed matrix of barriers and suitable habitat can generate more conservation issues than a wide uniform barrier area. This is because the former can host endangered species which then risk facing climate-niche separation, while, for the latter, climate-niche separation is only a problem on the barrier's outer perimeter. Habitat fragmentation is therefore a possibly underestimated conservation problem when considered together with climate change: it increases the interface between natural and anthropogenic land cover types, thus raising the chances of C-traps appearing.

Clearly, not all species whose distribution range is inside a C-trap will get extinct. As illustrated by Soberon & Peterson (2005), the persistence of species is not bound to just climatic (abiotic) factors, but to biotic and dispersal components too. These components have to be carefully considered when assessing the limitations of the C-trap framework.

Concerning the biological and dispersal factors, it must be noted that different species in a community follow climate at different speeds and not all of them are stopped by the same barriers. Species that are not stopped by the modelled barrier may depend on others that are, and would therefore find themselves threatened by the lack of the biotic component of their ecological niche. Also, the distribution range of an endangered species may be restricted by non-climatic requirements, such as geological or biological components, while its climatic niche extends far outside the realised distribution. In this

case, the species can remain in the same geographical range so long as its large, shifting climatic niche does.

The main goal of this paper was to raise general awareness on the concept of C-traps and the effects of anthropogenic land-use changes on climate migrants. Although, my application highlights sensible areas for the conservation of species under climate change worldwide, more specific studies on particular locations and taxa are necessary to pinpoint delicate specific scenarios and fine-scale settings, and for delivering concrete conservation measures. For instance, in the east Indian C-trap cluster (see Figure 18) Tiger (*Panthera tigris*; LINNAEUS, 1758) and Indian elephant (*Elephas maximus*; LINNAES, 1758) populations are present in several reserves inside of or in close proximity to C-traps. Even though these species have wide, scattered, distribution ranges mostly restricted by land-use changes – thus suggesting a wide climatic tolerance – local populations could have specific climatic adaptations or dependencies on more climate-sensible species. Moreover, the conservation of species genetic diversity heavily depends on the conservation of sub-specific, localised taxa, which are even more likely to be subjected to the dynamics of C-traps in the near future.

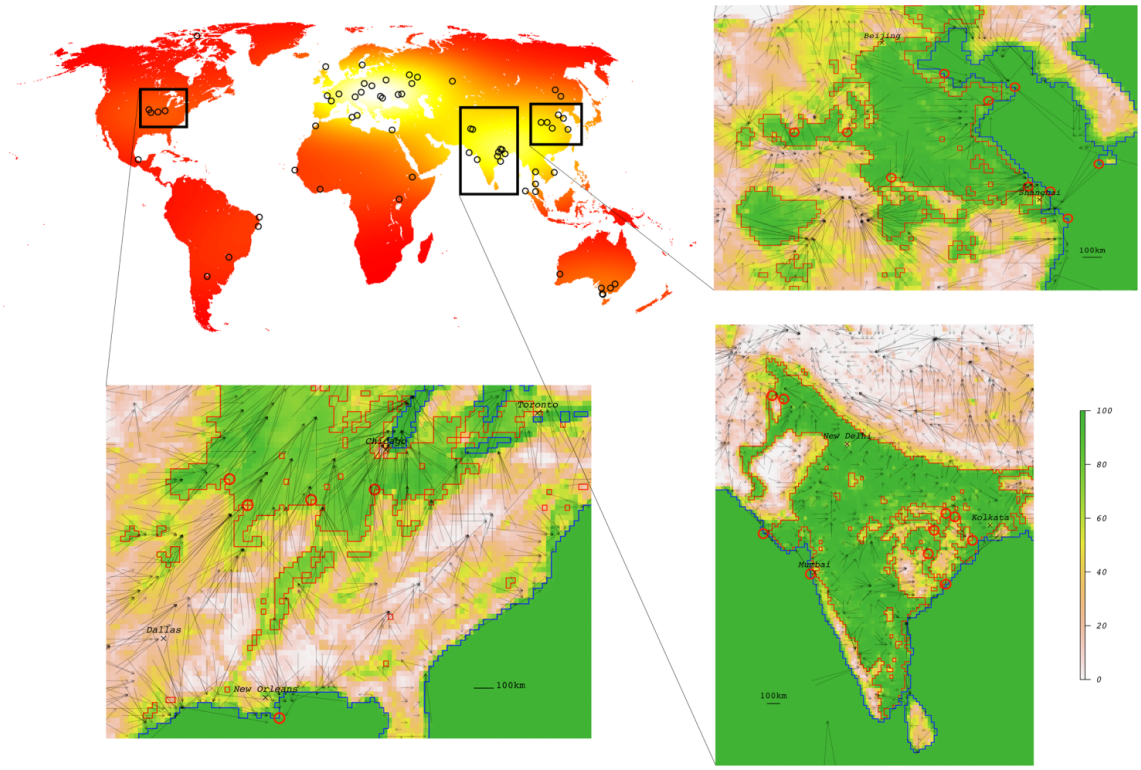


Figure 18: From top-left in clockwise order: Map of two-dimensional density kernel of anthropogenic C-traps; C-traps marked as black circles. Enlargement of east China; India; east U.S.A. Raster values on enlargement maps are the consensus of crops + open waters; red lines are boundaries of crops polygons and blue lines are boundaries of water polygons; C-traps are marked as red circles.

Among the possible developments of my methodology, the inclusion of a dynamic choice of C-trap depth in accordance with the length of local climate shift vectors would increase the detectability of additional traps in areas of particularly high climate velocity. Also, I choose the C-trap peak climate vector as a representative of the average direction of climate shift inside the C-trap. This is justified by the spatial autocorrelation of climate velocities, by the smoothing factor as set by the rounding of climate data in Hamann's *et al.* method, and by the compatible resolution of the climatic and land-use data. However, especially for finer scales applications, more precise descriptions of climate shift vectors would allow to delineate the closing front of the C-trap, thus allowing a precise description of the full trap's perimeter.

From a management perspective, it is somehow fortunate that these are barriers of anthropogenic origin, as it is still possible to intervene by creating corridors that follow climate change vectors (Robillard *et al.* 2015). However, the first step still remains the detection of C-traps themselves so that management measures can be promptly put into place in the most effective manner.

Chapter 5: Conclusions.

This thesis examined three ways in which human activities interfere with natural dispersal: facilitating dispersal in biological invasions; obstructing dispersal in habitat fragmentation; both facilitating and obstructing dispersal in the case of range shifts. The work presented here includes methodological, conceptual and applied approaches. It represents a two-fold contribution to understanding the implications – in terms of biodiversity loss – of the typical alterations to natural dispersal occurring in a human dominated landscape. On one hand, I use concrete examples to confirm and enforce the notion that anthropogenic interference with natural dispersal is a key element to biodiversity loss. On the other hand, I provide tools – methodologies and concepts – to deal with the harmful manifestations of this problem. I elaborate on these two aspects under the following two sub-headings.

Anthropogenic changes to natural dispersal as a vulnerable trait for the conservation of biodiversity

Dispersal *per se* has not been given much consideration in the fight against biodiversity loss. Nevertheless, the implications of losing the natural dispersal balance of organisms does result in major damages to ecosystems. In chapter one, I showed how much human aided dispersal promoted the biological invasion of *Litoria* frogs in New Zealand. 20%, 25% and 40% of *L. aurea*, *L. ewingii* and *L. raniformis* known populations in New Zealand were estimated to be established through anthropogenic dispersal. This is a huge contribution considering that the dispersal distance of anthropogenic populations is practically infinite within New Zealand's boundaries in comparison with natural dispersal (tadpoles of these species being traded online and dispersed through postal services). Consequently, chances are that new anthropogenic populations appear far away from

existing colonies, greatly increasing the speed of these animals' colonisation. In chapter two, I detected genetic signals indicating an early stage habitat fragmentation scenario, although I was unable to identify the contribution of different land cover types to the dispersal permeability of Auckland's suburbs for Copper skink populations. Copper skinks are just one of the many terrestrial, native species living in Auckland's North Shore reserves. Finally, in chapter three I highlighted several instances where anthropogenic land covers would affect dispersal of climate migrants worldwide. I did this by conducting an applied analysis of the global distribution of C-traps for land animal species with small distribution ranges. In the short term (next 50 years), local sub-specific taxa are likely to suffer the most from the effects of C-traps. Again, it is by ensuring that these susceptible areas are granted sufficient dispersal permeability that the detrimental effects of anthropogenic barriers on climate migrants will be mitigated.

The applied examples presented in the research chapters can be easily generalized to several systems worldwide. Biological invasions normally do involve anthropogenic dispersal, and land-use changes (be it urbanisation or intensively cultivated areas) involve habitat fragmentation. Therefore, being aware of the practical consequences of interfering with natural dispersal is a key element to the prevention for biodiversity loss.

Tools to address anthropogenic changes to natural dispersal

First, with the applied analyses described in the previous paragraphs, I highlighted the significance for biodiversity conservation of the anthropogenic alterations to the dispersal of organisms. In addition to that, I provided innovative tools to deal with such imbalances. On the biological invasion front, I proposed a modelling approach that gives weight to anthropogenic dispersal and made available free software for its implementation. My methodology succeeded in estimating the proportion of anthropogenic dispersal and the

natural dispersal kernel of simulated biological invasions under various scenarios, including fat tailed natural dispersal kernels and gaps in the virtual dataset. With regard to the issue of land-use change for range shifting taxa, I presented the novel concept of C-traps, describing the dynamics of this specific phenomenon. I also provided a spatially explicit methodology to detect the location of C-traps for arguably any given taxa and location, thus facilitating the implementation of the presented concepts into range shift modelling studies.

Limitations

The methodologies presented in this thesis are young and several improvements can potentially be made. At first, Biolinv R package is currently taking a single Habitat Suitability Model (HSM) that will then be used for the whole length of future simulations. However, if forecasting a biological invasion for a few decades, users may want to update their HSM to reflect the influence of climate change on the potential distribution of the study species. Even though the same outcome can be obtained by looping the `simulacro` function updating the HSM every few years and feeding the simulated distribution back as initial distribution, adding an automated option would be more practical and it would also encourage climate change aware modelling. Another useful update to the software would be allowing for user defined anthropogenic dispersal functions for the EM algorithm, so that not only random appearance of new anthropogenic points would be modelled. For example, it would be useful to allow for variables like proximity to roads to contribute to the EM classification. Additionally, the presented methodology does not account for evolutionary dynamics (Urban *et al.* 2008; Llewelyn *et al.* 2010; Shine 2012; Brown *et al.* 2014) or acclimatisation (Kolbe *et al.* 2010). This would be a very difficult parameter to include in the modelling protocol as it varies for different scenarios. The best approach to include evolutionary changes – similarly to

the case of climate change aware HSMs – would be to break the modelling process in time segments, to which different dispersal parameters are set accordingly to the expected evolutionary or acclimatisation shifts. Another option, useful in the case of extensive time series invasion data, is to discard the earliest records in order to avoid modelling the acclimatisation period with its lower dispersal rates. As an alternative, it would be possible to divide the time series into a few time segments to be modelled independently and extrapolate the growth of the dispersal rate for future predictions. A similar difficulty is found in of producing a reliable HSM (discussed in the introduction chapter). Basically, biological invasions are "evolving" processes where parameters change in time and space, making extrapolations particularly difficult, regardless of the modelling approach. For empirical/probabilistic models the problem is in the evolution of the predicted niche with every new bit of environmental space experienced by the invader; similarly, for deterministic/mechanistic models it is in the evolution/adaptation of the biological response to new environmental stimuli.

Concerning my work of chapter three, I obtained null results in the landscape genetic analysis and sustainable urban planning still awaits information on how to improve structural connectivity. However, significant developments have been done on the use of SNPs since 2012 (the beginning of data collection for this thesis work), and it would be interesting to process the collected tissue samples with this more sensitive methodology, as it may produce a more informative (i.e. non-null) landscape genetics outcome. Alternatively, to produce information on the connectivity weights of different land cover types, the same analysis could be performed on other terrestrial taxa such as non-flying insects.

Lastly, a limitation of the C-trap concept presented in chapter four is that range shifts of the next 50 years can be short when compared to the extent of species distribution

ranges. It is indeed relatively rare to find species whose entire distribution range is small enough to be entirely enclosed in a C-trap. Additionally, endangered species may have a disjunct distribution resulting from limitations in their habitat availability (e.g. due to land-use changes). Even though the realised distribution of such species can be very small, it is not restricted by its climatic requirements, and individual patches may not necessarily respond to climate shifts; the spatial distribution of their climatic requirements is much broader than the realised distribution range. Even though range shifts that follow local climate preferences may still occur in such circumstances, their conservation implications are likely to be more relevant for sub-specific, local taxa rather than whole species.

Future directions

Firstly, with regard to chapter four, the idea of C-traps appears to be conceptually robust. However, no specific applied example is given that tests how the described dynamics manifest in the field. This fulfils the goal of raising awareness on the problems resulting from combining range shifts and anthropogenic land-use changes – and my global application of the C-trap concept to species with small ranges also points at areas of high vulnerability to such dynamics. However, to generate concrete conservation guidelines, specific cases of C-traps for specific taxa and location need to be individually analysed. For example, in Chapter four I mention locations in India where C-traps may affect the survival of Elephant and Leopard populations. The realisation of such threats, however, depends on several ecological connections, such as the interaction of Leopards with their preys and the prey's sensitivity to shifting climates, or the link between Elephants and vegetation and the vegetation response to climate change. This would require focussing on specific scenarios.

Secondly, Although I demonstrated the efficacy of Biolinv R package in estimating the proportion of anthropogenic dispersal and natural dispersal kernel with virtual species, it would be interesting to compare its estimations with those obtained with other methods. One possibility for doing this would be to model well documented invasions such as the Cane toad in Australia (Urban *et al.* 2007, 2008; Kearney *et al.* 2008; Phillips *et al.* 2008; Kolbe *et al.* 2010; Tingley *et al.* 2013) and compare the outcomes with previous publications. Another interesting possibility for the analysis of biological invasions would be to model the time series as a point process. Natural and anthropogenic dispersal can be modelled as different point processes, for example, with the functions of R package spatstat. Such a modelling approach would be less computationally intensive than the one adopted in Biolinv, yet the problem of gap years would still need to be addressed.

An interesting alternative to genetic methodologies for addressing the problem of connectivity of natural patches in urban settings, was looked into during the first year of my PhD but didn't make it past that due to its costly nature. The project involved taking cloacal swabs of Copper skinks in addition to tissue samples and then analysing those by sequencing bacterial 16S ribosomal RNA (rRNA) genes hypervariable regions (Chakravorty *et al.* 2007). This procedure allowed for the identification of a remarkably diverse intestinal flora that would have the potential variability to be used as an alternative to genetic data in a "landscape microbial genetics" analysis. In addition to its high costs, extracting bacterial rRNA out of the very small swabs for Copper skinks could have been challenging. Nonetheless, this still remains an interesting concept worth investigating in the future and an analysis that could help shed light on the questions of chapter three. Another interesting study to complement a genetics based approach would be any field-based dispersal estimation study. In this case, radio tracking can be quite challenging given the small size of Copper skinks, but extremely light transmitters could be used on

Ornate skinks instead. Another option would be to attempt a mark recapture study on Copper skinks. However, recaptures are low in reptiles in general and Copper skinks in particular are already difficult to capture in the first place. Lastly, fluorescent powders have been employed to track short term dispersal on small animals and an attempt to apply such technique to Copper skinks could deliver interesting outcomes.

Finally, it is interesting to note how both invasive species and climate migrants are range-shifting species (Phillips *et al.* 2008) and as such, they both benefit from similar theoretical frameworks and methodologies. Indeed, the expanding front of a biological invasion is analogous to the advancing edge of a shifting range (Elith *et al.* 2010; Sorte *et al.* 2010). Perhaps the most compelling future development of the topics discussed here would be analysing the specific case of a climate change induced range shift in the presence of anthropogenically altered land cover. The methodologies developed in the second chapter could be used to simulate the yearly shift of a micro-endemic taxa and the C-traps framework would implement the effects of anthropogenic dispersal barriers. Clearly, this would only be meaningful at locations where the speed of climate shifts is higher than the dispersal capabilities of the study species. Otherwise it would be worthwhile to include Biolinv methodologies in forecasting the range shift as the whole process would proceed at the lower speed of the climate shift.

Bibliography

Adamack, A.T. & Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, **5**, 384–387.

Adams, L.W. (2005). Urban wildlife ecology and conservation : A brief history of the discipline. *Urban Ecosystems*, **8**, 139–156.

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. & Hill, A.V.S. (2007). Automated binning of microsatellite alleles: Problems and solutions. *Molecular Ecology Notes*, **7**, 10–14.

Baglione, V. (2003). Kin Selection in Cooperative Alliances of Carrion Crows. *Science*, **300**, 1947–1949.

Baguette, M., Legrand, D., Fréville, H., Dyck, H. Van & Ducatez, S. (2012). Evolutionary ecology of dispersal in fragmented landscape. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, G.T. Benton & J.M. Bullock), p. 497. Oxford University Press, Oxford.

Becker, R.A. & Wilks, R.A. (1995). *Constructing a geographical database*. New Jersey.

Becker, R.A. & Wilks, R.A. (1993). *Maps in S*. New Jersey.

Benton, T.G. & Bowler, D.E. (2012). Linking dispersal to spatial dynamics. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, G.T. Benton & J.M. Bullock), pp. 286–300. Oxford University Press, Oxford.

Berry, O., Gleeson, D.M. & Sarre, S.D. (2003). Microsatellite DNA markers for New Zealand skinks. *Conservation Genetics*, **4**, 411–414.

Berteaux, D. & Boutin, S. (2000). Breeding Dispersal in Female North American Red Squirrels. *Ecology*, **81**, 1311–1326.

Biek, R. & Real, L.A. (2010). The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, **19**, 3515–3531.

Bishop, P. (2008). Bell frog populations in New Zealand-good news or bad news? *Australian Zoologist*, **34**, 408–414.

Den Boer, P.J. (1968). Spreading of risk and stabilization of animal numbers. *Acta Biotheoretica*, **18**, 165–194.

Bonte, D., Van Belle, S. & Maelfait, J.P. (2007). Maternal care and reproductive state-dependent mobility determine natal dispersal in a wolf spider. *Animal Behaviour*, **74**, 63–69.

Bonte, D., Hovestadt, T. & Poethke, H.J. (2010). Evolution of dispersal polymorphism and local adaptation of dispersal distance in spatially structured landscapes. *Oikos*, **119**, 560–566.

Bowler, D.E. & Benton, T.G. (2005). Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, **80**, 205–225.

Brewer, S., Cheddadi, R., de Beaulieu, J.L. & Reille, M. (2002). The spread of deciduous Quercus throughout Europe since the last glacial period. *Forest Ecology and*

Management, **156**, 27–48.

Britton, J.R. & Orsi, M.L. (2012). Non-native fish in aquaculture and sport fishing in Brazil: Economic benefits versus risks to fish diversity in the upper River Paraná Basin. *Reviews in Fish Biology and Fisheries*, **22**, 555–565.

Brookfield, J.F.Y. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Brown, G.P., Phillips, B.L. & Shine, R. (2014). The straight and narrow path: the evolution of straight-line dispersal at a cane toad invasion front. *Proceedings of the royal society B*, **281**, 1–7.

Burrows, M.T., Schoeman, D.S., Richardson, A.J., Molinos, J.G., Hoffmann, A., Buckley, L.B., Moore, P.J., Brown, C.J., Bruno, J.F., Duarte, C.M., Halpern, B.S., Hoegh-Guldberg, O., Kappel, C. V., Kiessling, W., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Sydeman, W.J., Ferrier, S., Williams, K.J., Poloczanska, E.S., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Sydeman, W.J., Ferrier, S., Williams, K.J. & Poloczanska, E.S. (2014). Geographical limits to species-range shifts are suggested by climate velocity. *Nature*, **507**, 492–5.

Cadi, A. & Joly, P. (2004). Impact of the introduction of the red-eared slider (Trachemys scripta elegans) on survival rates of the European pond turtle (Emys orbicularis). *Biodiversity and Conservation*, **13**, 2511–2518.

Cain, M.L., Milligan, B.G. & Strand, A.E. (2000). Long-Distance Seed Dispersal in Plant Populations. *American journal of botany*, **87**, 1217–1227.

Caley, P., Ramsey, D.S.L. & Barry, S.C. (2015). Inferring the distribution and demography of an invasive species from sighting data: The red fox incursion into Tasmania. *PLoS ONE*, **10**, 1–18.

Catterall, S., Cook, A.R., Marion, G., Butler, A. & Hulme, P.E. (2012). Accounting for uncertainty in colonisation times: A novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. *Ecography*, **35**, 901–911.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M. & Palmer, T.M. (2015). Accelerated modern human – induced species losses: entering the sixth mass extinction. *Sciences Advances*, **1**, 1–5.

Chakravorty, S., Helb, D. & Burday, M. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, **69**, 330–339.

Chapple, D.D.G., Patterson, G.G.B., Bell, T. & Daugherty, C.H. (2008). Taxonomic revision of the New Zealand copper skink (Cyclodina aenea: Squamata: Scincidae) species complex, with descriptions of two new species. *Journal of Herpetology*, **42**, 437–452.

Clark, J.S. (1998). Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *The American naturalist*, **152**, 204–224.

Clobert, J., Baguette, M., Benton, G.T. & Bullock, J.M. (2012). *Dispersal Ecology and Evolution* (J. Clobert, M. Baguette, G.T. Benton & J.M. Bullock, Eds.). Oxford University Press, Oxford.

Clobert, J., Danchin, E., Dhondt, A.A. & Nichols, J.D. (2001). *Dispersal* (J. Clobert, E. Danchin, A.A. Dhondt & J.D. Nichols, Eds.). Oxford University Press, Oxford.

Cook, A., Marion, G., Butler, A. & Gibson, G. (2007). Bayesian inference for the spatio-temporal invasion of alien species. *Bulletin of Mathematical Biology*, **69**, 2005–2025.

Corlett, R.T. & Westcott, D.A. (2013). Will plant movements keep up with climate change? *Trends in Ecology and Evolution*, **28**, 482–488.

Cornell, S.J. & Ovaskainen, O. (2008). Exact asymptotic analysis for metapopulation dynamics on correlated dynamic landscapes. *Theoretical Population Biology*, **74**, 209–225.

Delaney, K.S., Riley, S.P.D. & Fisher, R.N. (2010). A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. *PLoS ONE*, **5**, 1–11.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*.

Denno, R.F., Roderick, G.K., Peterson, M.A., Huberty, A.F., Dobel, H.G., Eubanks, M.D., Losey, J.E., Langellotto, G.A., Huberty, A.F., Dobel, H.G. & Eubanks, M.D. (1996). Habitat Persistence Underlies Intraspecific Variation in the Dispersal Strategies of Planthoppers. *Ecological Monographs*, **66**, 389–408.

Dettmers, R. & Bart, J. (1999). A GIS Modeling Method Applied to Predicting Forest Songbird Habitat. *Ecological Applications*, **9**, 152–163.

Dixon, J.D., Oli, M.K., Wooten, M.C., Eason, T.H., M, J.W. & Paetkau, D. (2006). Effectiveness of a regional corridor in connecting two Florida black bear populations. *Conservation Biology*, **20**, 155–162.

Douglas, D.H. & Peuker, T.K. (2011). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica (The Canadian Cartographer)*, **10**, 112–122.

Driscoll, D. a., Banks, S.C., Barton, P.S., Ikin, K., Lentini, P., Lindenmayer, D.B., Smith, A.L., Berry, L.E., Burns, E.L., Edworthy, A., Evans, M.J., Gibson, R., Heinsohn, R., Howland, B., Kay, G., Munro, N., Scheele, B.C., Stirnemann, I., Stojanovic, D., Sweaney, N., Villaseñor, N.R. & Westgate, M.J. (2014). The Trajectory of Dispersal Research in Conservation Biology. Systematic Review (J. Hewitt, Ed.). *PLoS ONE*, **9**, e95053.

Dytham, C. & Travis, J.M.J. (2012). Modelling the effects of habitat fragmentation. *Dispersal Ecology and Evolution* (eds J. Colbert, M. Baguette, T.G. Benton & J.M. Bullock), p. 497. Oxford University Press, Oxford.

Early, R. & Sax, D.F. (2011). Analysis of climate paths reveals potential limitations on species range shifts. *Ecology Letters*, **14**, 1125–1133.

Eileen, A. & Baker, M. (1994). Stowaway Transport Rates of House Mice (Mus Domesticus) and Deermice (Peromyscus Maniculatus). *Proceedings of the Sixteenth Vertebrate Pest Conference*, 106–112.

Elith, J., Graham, C.H., Anderson, R.P., Dudı´k, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., McC.Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti- Pereira, R., Schapire, R.E., Sobero´n, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Kearney, M. & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Elton, C.S. (1958). The Ecology of Invasions by Animals and Plants. p. 196. The University of Chicago Press.

Engler, R. & Guisan, A. (2009). MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 590–601.

Engler, R., Hordijk, W. & Guisan, A. (2012). The MIGCLIM R package - seamless integration of dispersal constraints into projections of species distribution models. *Ecography*, **35**, 872–878.

Epps, C.W. & Keyghobadi, N. (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, **24**, 6021–6040.

Epps, C.W., Wehausen, J.D., Bleich, V.C., Torres, S.G. & Brashares, J.S. (2007). Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*, **44**, 714–724.

Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 487–515.

Fisher, R.A. (1937). The wave of advance of advantageous genes. *Annals of Eugenics*, **7**, 355–369.

Fujiwara, M., Anderson, K.E., Neubert, M.G. & Caswell, H. (2006). On the estimation of dispersal kernels from individual mark-recapture data. *Environmental and Ecological Statistics*, **13**, 183–197.

Galliard, J.F. Le, Massot, M. & Clobert, J. (2012). Dispersal and range dynamics in changing climates: a review. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, G.T. Benton & J.M. Bullock), p. 497. Oxford University Press, Oxford.

Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I. & Thuiller, W. (2010). Predicting potential distributions of invasive species: Where to go from here? *Diversity and Distributions*, **16**, 331–342.

Germaine, S.S. & Wakeling, B.F. (2001). Lizard species distributions and habitat occupation along an urban gradient in Tucson, Arizona, USA. *Biological Conservation*, **97**, 229–237.

Gillespie, T.W. (2001). Remote sensing of animals. *Progress in Physical Geography*, **25**, 355–362.

Gottfried, M., Pauli, H., Futschik, A., Akhalkatsi, M., Barančok, P., Alonso, B., Luis, J., Coldea, G., Dick, J., Erschbamer, B., Calzado, F., Rosa, M., Kazakis, G., Krajči, J., Larsson, P., Mallaun, M., Michelsen, O., Moiseev, D., Moiseev, P., Molau, U., Merzouki, A., Nagy, L., Nakhutsrishvili, G., Pedersen, B., Pelino, G., Puscas, M., Rossi, G., Stanisci, A., Theurillat, J.-P., Tomaselli, M., Villar, L., Vittoz, P., Vogiatzakis, I. & Grabherr, G. (2012). Continent-wide response of mountain vegetation to climate change. *Nature Climate Change*, **2**, 111–115.

Gozlan, R.E., Britton, J.R., Cowx, I. & Copp, G.H. (2010). Current knowledge on non-native freshwater fish introductions. *Journal of Fish Biology*, **76**, 751–786.

Gruber, B. & Adamack, A.T. (2015). landgenreport: A new R function to simplify landscape genetic analysis using resistance surface layers. *Molecular Ecology Resources*, **15**, 1172–1178.

Haddad, N.M., Brudvig, L. a., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, a. J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B. a., Nicholls, a. O., Orrock, J.L., Song, D.-X. & Townshend, J.R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, **1**, 1–9.

Hamann, A., Roberts, D.R., Barber, Q.E., Carroll, C. & Nielsen, S.E. (2015). Velocity of climate change algorithms for guiding conservation and management. *Global Change Biology*, **21**, 997–1004.

Hamilton, W. & May, R. (1977). Dispersal in stable habitats. *Nature*, **269**, 578–581.

Hamrick, J.L. & Trapnell, D.W. (2011). Using population genetic analyses to understand seed dispersal patterns. *Acta Oecologica*, **37**, 641–649.

Hedrick, P.W. (2005). A Standardized Genetic Differentiation Measure. *Evolution*, **59**, 1633–1638.

van Heezik, Y. & Ludwig, K. (2012). Proximity to source populations and untidy gardens predict occurrence of a small lizard in an urban area. *Landscape and Urban Planning*, **104**, 253–259.

Heller, N.E. & Zavaleta, E.S. (2009). Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biological Conservation*, **142**, 14–32.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hirzel, A.H. & Le Lay, G. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**, 1372–1381.

Hitchings, S.P. & Beebee, T.J.C. (1997). Substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations : implications for biodiversity conservation. *Heredity*, **79**, 117–127.

Hodkinson, D.J. & Thompson, K. (1997). Plant dispersal: the role of man. *Journal of Applied Ecology*, **34**, 1484–1496.

Holderegger, R. & Wagner, H.H. (2006). A brief guide to Landscape Genetics. *Landscape Ecology*, **21**, 793–796.

Holderegger, R. & Wagner, H.H. (2008). Landscape Genetics. *BioScience*, **58**, 199–207.

Holeck, K.T., Mills, E.L., MacIsaac, H.J., Dochoda, M.R., Colautti, R.I. & Ricciardi, A. (2004). Bridging Troubled Waters: Biological Invasions, Transoceanic Shipping, and the Laurentian Great Lakes. *BioScience*, **54**, 919.

Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, **46**, 10–18.

IPCC. (2014). *Climate Change 2014, Synthesis Report.*

Jablonski, D. (1994). Extinctions in the fossil record. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 2155–2155.

Jellinek, S., Driscoll, D. a. & Kirkpatrick, J.B. (2004). Environmental and vegetation variables have a greater influence than habitat fragmentation in structuring lizard communities in remnant urban bushland. *Austral Ecology*, **29**, 294–304.

Jha, S. & Kremen, C. (2013). Urban land use limits regional bumble bee gene flow. *Molecular Ecology*, **22**, 2483–2495.

Johannesson, K. (1988). The paradox of Rockall: why is a brooding gastropod (Littorina saxatilis) more widespread than one having a planktonic larval dispersal stage (L. littorea)? *Marine Biology*, **99**, 507–513.

Jombart, T., Pontier, D. & Dufour, A.-B. (2009). Genetic markers in the playground of multivariate analysis. *Heredity*, **102**, 330–341.

Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G. & Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kawecki, T.J. & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008). Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.

Kelly, R., Lundy, M.G., Mineur, F., Harrod, C., Maggs, C. a., Humphries, N.E., Sims, D.W. & Reid, N. (2014). Historical data reveal power-law dispersal patterns of invasive aquatic species. *Ecography*, no-no.

Koenig, J., Shine, R. & Shea, G. (2001). The ecology of an Australian reptile icon: how do blue-tongued lizards (Tiliqua scincoides) survive in suburbia? *WILDLIFE RESEARCH*, **28**, 215–227.

Kolar, C.S. & Lodge, D.M. (2001). Progress in invasion biology: predicting invaders. *Trends in ecology & evolution*, **16**, 199–204.

Kolbe, J.J., Kearney, M. & Shine, R. (2010). Modeling the consequences of thermal trait variation for the cane toad invasion of Australia. *Ecological Applications*, **20**, 2273–2285.

Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S. & Levine, J.M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, **29**, 592–599.

Kranner, I., Zorn, M., Turk, B., Wornik, S., Beckett, R.P., The, S., Phytologist, N. & Oct, N. (2007). Spatial Aspects of Seed Dispersal and Seedling Recruitment in Orchids. *The New Phytologist*, **176**, 237–241.

Kwiatkowski, M. a., Schuett, G.W., Repp, R. a., Nowak, E.M. & Sullivan, B.K. (2008). Does urbanization influence the spatial ecology of Gila monsters in the Sonoran Desert? *Journal of Zoology*, **276**, 350–357.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M. & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179–4191.

LaPoint, S., Balkenhol, N., Hale, J., Sadler, J. & van der Ree, R. (2015). Ecological connectivity research in urban areas. *Functional Ecology*, **29**, 868–878.

Larson, G., Albarella, U., Dobney, K., Rowley-conwy, P., Tagliacozzo, A., Manaseryan, N., Miracle, P., Wijngaarden-bakker, L. Van, Masseti, M., Bradley, D.G. & Cooper, A. (2007). Ancient DNA , pig domestication , and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Science*, **104**, 15276–15281.

Leathwick John, Morgan Fraser, Wilson Gareth, Rutledge Daniel, McLeod Malcom, J.K. (2002). *Land Environments of New Zealand: A Technical Guide*.

Lehikoinen, A. & Virkkala, R. (2016). North by North-West: Climate change and directions of density shifts in birds. *Global Change Biology*, **22**, 1121–1129.

Lenoir, J. & Svenning, J.C. (2015). Climate-related range shifts - a global multidimensional synthesis and new research directions. *Ecography*, **38**, 15–28.

Levine, J.M. & D'Antonio, C.M. (2007). Forecasting biological invasions with increasing international trade. *Conservation Biology*, **21**, 1666.

Lewis, M.A., Maini, P.K. & Petrovskii, S. V. (2013). *Dispersal, Individual Movement and Spatial Ecology: A Mathematical Perspective* (M.A. Lewis, P.K. Maini & S. V. Petrovskii, Eds.). Springer, Berlin.

Liu, Y.P., Wu, G.S., Yao, Y.G., Miao, Y.W., Luikart, G., Baig, M., Beja-Pereira, A., Ding, Z.L., Palanichamy, M.G. & Zhang, Y.P. (2006). Multiple maternal origins of chickens: Out of the Asian jungles. *Molecular Phylogenetics and Evolution*, **38**, 12–19.

Llewelyn, J., Phillips, B.L., Alford, R.A., Schwarzkopf, L. & Shine, R. (2010). Locomotor performance in an invasive species: Cane toads from the invasion front have greater endurance, but not speed, compared to conspecifics from a long-colonised area. *Oecologia*, **162**, 343–348.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009). The velocity of climate change. *Nature*, **462**, 1052–5.

MacArthur, R.H. & Wilson, E.O. (2015). *Theory of Island Biogeography*. Princeton University Press.

Macdonald, W.D. & Johnson, D.P.D. (2001). Dispersal in theory and practice: consequences for population biology. *Dispersal* (eds E.J.C. Danchin, A.A. Dhondt & J.D. Nichols), p. 452. Oxford University Press, Oxford.

Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mason, S.C., Palmer, G., Fox, R., Gillings, S., Hill, J.K., Thomas, C.D. & Oliver, T.H. (2015). Geographical range margins of many taxonomic groups continue to shift polewards. *Biological Journal of the Linnean Society*, **115**.

Mathias, A., Kisdi, É., Olivieri, I., Kisdi, E., Olivieri, I., Kisdi, É., Olivieri, I. & Kisdi, E. (2001). Divergent evolution of dispersal in a heterogeneous landscape. *Evolution; international journal of organic evolution*, **55**, 246–59.

Matschiner, M. & Salzburger, W. (2009). TANDEM: Integrating automated allele binning into genetics and genomics workflows. *Bioinformatics*, **25**, 1982–1983.

Matthysen, E., van Overveld, T., van de Casteele, T. & Adriaensen, F. (2010). Family

movements before independence influence natal dispersal in a territorial songbird. *Oecologia*, **162**, 591–597.

Mech, S.G. & Hallett, J.G. (2013). Evaluating the Effectiveness of Corridors : A Genetic Approach. *Conservation Biology*, **15**, 467–474.

Moilanen, A. (2004). SPOMSIM: Software for stochastic patch occupancy models of metapopulation dynamics. *Ecological Modelling*, **179**, 533–550.

Moors, P.J., Atkinson, I. a. E. & Sherley, G.H. (1992). Reducing the rat threat to island birds. *Bird Conservation International*, **2**, 93–114.

El Mousadik, A. & Petit, R.J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.

Munshi-South, J. (2012). Urban landscape genetics: Canopy cover predicts gene flow between white-footed mouse (Peromyscus leucopus) populations in New York City. *Molecular Ecology*, **21**, 1360–1378.

Murphy, M.A., Dezzani, R., Pilliod, D.S. & Storfer, A. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, **19**, 3634–3649.

Nathan, R. (2006). Long-distance dispersal of plants. *Science (New York, N.Y.)*, **313**, 786–8.

Nathan, R., Klein, E., Robledo-Arnuncio, J.J. & Revilla, E. (2012). Dispersal kernels: review. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, G.T. Benton & J.M. Bullock), p. 497. Oxford University Press, Oxford.

Nathan, R., Perry, G., Cronin, J.T., Strand, A.E. & Cain, M.L. (2003). Methods for estimating long-distance dispersal. *Oikos*, **103**, 261–273.

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics*, **41**, 225–233.

New Zealand Transport Agency. (2008). *Auckland Motorways*. Auckland.

Niemela, J. (1999). Is there a need for a theory of urban ecology? *Urban Ecosystems*.

Noël, S., Labonté, P. & Lapointeo, F.-J. (2011). Genomotype Frequencies and Genetic Diversity in Urban and Protected Populations of Blue-spotted Salamanders (Ambystoma laterale) and Related Unisexuals. *Journal of Herpetology*, **45**, 294–299.

Noël, S., Ouellet, M., Galois, P. & Lapointe, F.J. (2007). Impact of urban fragmentation on the genetic structure of the eastern red-backed salamander. *Conservation Genetics*, **8**, 599–606.

Novacek, M.J. (2011). *The Biodiversity Crisis: Losing What Counts*, America Mun. The New Press.

Olson, L. (2006). The economics of terrestrial invasive species: a review of the literature. *Agricultural and Resource Economics Review*, **1**, 178–194.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Ordonez, A., Martinuzzi, S. & Radelo, V.C. (2014). Combined speeds of climate and land-use change of the conterminous US until 2050. *Nature Climate Change*, **4**, 1–6.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change.

The Annual Review of Ecology, Evolution, and Systematics, **37**, 637–669.

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Parolo, G. & Rossi, G. (2008). Upward migration of vascular plants following a climate warming trend in the Alps. *Basic and Applied Ecology*, **9**, 100–107.

Pearson, R. (2006). Climate change and the migraiton capacity of species. *Trends in Ecology & Evolution*, **21**, 111–113.

Peter, N., James, S. & David, P. (2013). *Establishing New Zealand's LUCAS Land Use and Land Use-Change and Forestry 2012 Map*.

Peterson, A.T. (2003). Predicting the Geography of Species' Invasions Via Ecological Niche Modeling. *The Quarterly Review of Biology*, **78**, 419–433.

Peterson, A. (2006). Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3**, 59–72.

Peterson, A.T. & Nakazawa, Y. (2008). Environmental data sets matter in ecological niche modelling: An example with Solenopsis invicta and Solenopsis richteri. *Global Ecology and Biogeography*, **17**, 135–144.

Philippi, T. & Seger, J. (1989). Hedging One 's Evolutionary Bets, Revisited. **4**, 2–5.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, B.L., Chipperfield, J.D. & Kearney, M.R. (2008). The toad ahead: Challenges of modelling the range and spread of an invasive species. *Wildlife Research*, **35**, 222–234.

Phillips, S. & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 161–176.

Pimentel, D., Zuniga, R. & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Pitt, J.P.W., Worner, S.P. & Suarez, A. V. (2009). Predicting Argentine ant spread over the heterogenous landscape using a spatially-explicit stochastic model. *Ecol. Appl.*, **19**, 1176–1186.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R_Core_Team. (2016). R: A Language and Environment for Statistical Computing.

Raup, D.M. & Sepkoski, J.J. (1982). Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503.

Reed, J.M. (1999). Role of behavior in avian extinctions and endangerments. *Conservation Biology*, **13**, 232–241.

Ripley, B.D. (1977). Modelling Spatial Patterns. **39**, 172–212.

Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.

Risk, B.B., Valpine, P. De, Beissinger, S.R., Risk, B.B., Valpine, P. De & Beissinger, S.R. (2011). A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. *Ecology*, **92**, 462–474.

Robillard, C.M., Coristine, L.E., Soares, R.N. & Kerr, J.T. (2015). Facilitating climate-change-induced range shifts across continental land-use barriers. *Conservation Biology*, **29**, 1586–1595.

Salafsky, N., Salzer, D., Stattersfield, A.J., Hilton-Taylor, C., Neugarten, R., Butchart, S.H.M., Collen, B., Cox, N., Master, L.L., O'Connor, S. & Wilkie, D. (2008). A standard lexicon for biodiversity conservation: Unified classifications of threats and actions. *Conservation Biology*, **22**, 897–911.

Schtickzelle, N., Mennechez, G. & Baguette, M. (2006). Dispersal depression with habitat fragmentation in the bog fritillary butterfly. *Ecology*, **87**, 1650–1658.

Shaw, S.D., Skerratt, L.F., Haigh, A., Bell, B.D., Daglish, L., Bishop, P.J., Summers, R., Moreno, V., Melzer, S., Ohmer, M., Herbert, S., Gleeson, D., Rowe, L. & Speare, R. (2013). The distribution and host range of Batrachochytrium dendrobatidis in New Zealand, 1930–2010. *Ecology*, **94**, 2108–2111.

Shigesada, N. & Kawasaki, K. (1997). *Biological Invasions: Theory and Practice*. Oxford University Press, Oxford.

Shigesada, N. & Kawasaki, K. (2002). Invasion and the range expansion of species: effects of long-distance dispersal. *Dispersal Ecology* (eds J.M. Bullock, R.E. Kenward & R.S. Hails), p. 458. Blackwell Publishing Ltd, Oxford.

Shigesada, N., Kawasaki, K. & Teramoto, E. (1985). Traveling periodic waves in heterogeneous environments. *Theoretical Population Biology*, **30**, 143–160.

Shine, R. (2012). Invasive species as drivers of evolutionary change: Cane toads in tropical Australia. *Evolutionary Applications*, **5**, 107–116.

Skellam, J.G. (1951). Random dispersal in theoretical populations. *Biometrika*, **38**, 196–218.

Soberon, J. & Peterson, T.A. (2005). Interpretation of Models of Fundamental Ecological Niches and Species ' Distributional Areas. *Biodiversity Informatics*, **2**, 1–10.

Sorte, C.J.B., Williams, S.L. & Carlton, J.T. (2010). Marine range shifts and species introductions: Comparative spread rates and community impacts. *Global Ecology and Biogeography*, **19**, 303–316.

Spear, S.F., Balkenhol, N., Fortin, M.-J., McRae, B.H. & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular ecology*, **19**, 3576–91.

Stevens, V.M., Verkenne, C. & Vandewoestijne, S. (2006). Gene flow and functional connectivity in the natterjack toad. *Molecular Ecology*, **15**, 2333–2344.

Storfer, A., Murphy, M. a, Spear, S.F., Holderegger, R. & Waits, L.P. (2010). Landscape genetics: where are we now? *Molecular ecology*, **19**, 3496–514.

Suarez, a V, Holway, D. a & Case, T.J. (2001). Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 1095–100.

Sullivan, M.J.P., Davies, R.G., Reino, L. & Franco, A.M. a. (2012). Using dispersal information to model the species-environment relationship of spreading non-native species. *Methods in Ecology and Evolution*, **3**, 870–879.

Thomas, C.D. (2000). Dispersal and extinction in fragmented landscapes. *Proceedings of the*

Royal Society B: Biological Sciences, **267**, 139–145.

Thomas, C.D., Franco, A.M.A. & Hill, J.K. (2006). Range retractions and extinction in the face of climate warming. *Trends in Ecology and Evolution*, **21**, 415–416.

Thurley, T. & Bell, B. (1994). Habitat distribution and predation on a western population of terrestrial Leiopelma (Anura: Leiopelmatidae) in the northern King Country, New Zealand. *New Zealand journal of zoology*, **21**, 37–41.

Tingley, R., Phillips, B.L., Letnic, M., Brown, G.P., Shine, R. & Baird, S.J.E. (2013). Identifying optimal barriers to halt the invasion of cane toads Rhinella marina in arid Australia. *Journal of Applied Ecology*, **50**, 129–137.

Trakhtenbrot, A., Nathan, R., Perry, G. & Richardson, D.M. (2005). The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions*, **11**, 173–181.

Travis, J.M.J. & Dytham, C. (1999). Habitat persistence, habitat availability and the evolution of dispersal. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 723–728.

Tremblay, M.A. & St. Clair, C.C. (2009). Factors affecting the permeability of transportation and riparian corridors to the movements of songbirds in an urban landscape. *Journal of Applied Ecology*, **46**, 1314–1322.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.

United Nations. (1992). United Nations Framework Convention on Climate Change. *Fccc/Informal/84*, **1**, 270–277.

United Nations. (2014). *World Urbanization Prospects: The 2014 Revision, Highlights*.

Urban, M.C., Phillips, B.L., Skelly, D.K. & Shine, R. (2008). A Toad More Traveled: The Heterogeneous Invasion Dynamics of Cane Toads in Australia. *The American Naturalist*, **171**, E134–E148.

Urban, M.C., Phillips, B.L., Skelly, D.K. & Shine, R. (2007). The cane toad's (Chaunus [Bufo] marinus) increasing ability to invade Australia is revealed by a dynamically updated range model. *Proceedings. Biological sciences / The Royal Society*, **274**, 1413–9.

VanDerWal, J., Murphy, H.T., Kutt, A.S., Perkins, G.C., Bateman, B.L., Perry, J.J. & Reside, A.E. (2013). Focus on poleward shifts in species' distribution underestimates the fingerprint of climate change. *Nature Climate Change*, **3**, 239–243.

Vyas, N.B., Hill, E.F., Sauer, J.R. & Kuenzel, W.J. (1995). Acephate affects migratory orientation of the white-throated sparrow (Zonotrichia albicollis). *Environmental Toxicology and Chemistry*, **14**, 1961–1965.

Wagner, H.H. & Fortin, M.-J. (2012). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, **14**, 253–261.

Walters, R.J., Hassall, M., Telfer, M.G., Hewitt, G.M. & Palutikof, J.P. (2006). Modelling dispersal of a temperate insect in a changing climate. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2017–2023.

Wang, I.J., Savage, W.K. & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (Ambystoma californiense). *Molecular Ecology*, **18**, 1365–1374.

Webster, M.S., Marra, P.P., Haig, S.M., Bensch, S. & Holmes, R.T. (2002). Links between worlds: Unraveling migratory connectivity. *Trends in Ecology and Evolution*, **17**, 76–83.

Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences*, **104**, 5738–5742.

Appendix

Introduction to MaxEnt habitat suitability modelling

Among the methods developed for the analysis of presence-only data on vast geographic area, MaxEnt (Maximum Entropy, S. Phillips & Dudík, 2008; Phillips *et al.* 2006) is a machine learning technique that has been proved to be consistently reliable and competitive with the highest performing methods (Elith *et al.* 2006, Elith *et al.* 2011). It has been extensively used across many applications, including the study of basic ecology of species, evolution, and prediction of future distribution scenarios, invasion ecology and biosecurity (Table 7).

Primary purpose	Extent	Organisms	Refs
Predict current distributions as input for conservation planning, risk assessments or IUCN listing, or new surveys	Andes	Humming-birds	Tinoco *et al.* (2009)
	Global	Stony corals seamounts	Tittensor *et al.* (2009)
Understand environmental correlates of species occurrences, groups of species, or other	Norway	Macrofungi	Wollan *et al.* (2008)
	Portugal	European wildcat	Monterroso *et al.* (2009)
Predict potential distributions for invasive species, or explore expanding distributions	New Zealand	Ants	Ward (2007a)
	China	Nematode	Wang *et al.* (2007)
Predict species richness or diversity	California	Amphibians and reptiles	Graham & Hijmans (2006)
	Brazil	Myrtaceae 19 species	Murray-Smith *et al.* (2009)
Predict current distributions for understanding morphological / genetic diversity ("phylogeography", "phyloclimatic studies"), endemism and evolutionary niche dynamics	Global	Seaweeds	Verbruggen *et al.* (2009)
	Andes	Birds	Young *et al.* (2009)
	Madagascar	Bats	Lamb *et al.* (2008)
Hindcast distributions to understand patterns of endemism, vicariance, etc	NW Europe	Pond snails	Cordellier & Pfenninger (2009)
	Brazilian coast	Forests	Carnaval & Moritz (2008)
Forecast distributions to understand changes with climate change / land transformation; includes retrospective studies	Mediterr'n + surrounds	Cyclamen	Yesson & Culham (2006)
	Regional W. Australia	Banksia	Yates *et al.* (2010)
	Canada	Butterflies	Kharouba *et al.* (2009)
Test model performance against other methods	Patagonia	Insects	Tognelli *et al.* (2009)
	Local region in California	Rare plants	Williams *et al.* (2009)
	Regional to national	Many species	Elith *et al.* (2006)

Table 9: Examples of published studies using MaxEnt, showing variation in purpose, extent and study organism (from J. Elith et al. 2011).

Conceptually, MaxEnt modelling makes use of values of environmental variables recorded at the species' presence localities to estimate their probability density across the

landscape *f1*. Many of such distributions are possible but MaxEnt choses the one that is most similar to the probability density of the environmental variables sampled on a set of randomly distributed background points *f*. *f* is a null model for *f1* and in MaxEnt the distance between *f* and *f1* is considered to be the relative entropy of *f1* with respect to *f*. The closer *f* and *f1*, the higher the relative entropy of *f1* with respect to *f*. We can therefore say that MaxEnt chooses the *f1* which has the highest relative entropy with respect to *f*. For example, in the extreme case of a species not operating any habitat selection, *f* and *f1* would be equivalent, and the relative entropy of *f1* with respect to *f* would be null. A null model for a species distribution corresponds to its random and uniform distribution over the landscape, where its spatial entropy is minimised.

MaxEnt parameters

Testing on the *Litoria spp.* MaxEnt models, was done by 10-fold cross-validation. 10% of the sighting locations were kept aside for testing. Response curves were created for all the variables both as a single-variable MaxEnt model and with all variables together. Permutation importance of environmental variables was also computed: for each variable in turn, the values of that variable on training and background data are randomly permuted, the model is re-evaluated on the permuted data and the resulting drop in training AUC (Area Under the receiver operating characteristic Curve, the probability that any sighting location cell is given a higher habitat suitability value then any background cell) is plotted. After a trial run, "Autofeatures", "Threshold" and "Hinge" feature types were disabled to avoid overfitting. Jackknife resampling of the environmental variables was used to test the variables importance.

The used output format was "Raw" and the maximum number of iteration was fixed at 10,000 but never exceeded 1,500 for *L. raniformis*, 1640 for *L. ewingii* and 2040 for *L.*

aurea. Model performance was tested by computing the AUC for classification of location cells as sighting locations.

MaxEnt results

Habitat suitability models' performance, as indicated by the AUC values, is good: *L. aurea* 0.835 (σ 0.026); *L. raniformis* 0.773 (σ 0.027); *L. ewingii* 0.790 (σ 0.034). Results of the MESS analysis show that the only case for which there is a significative area of novel habitat is *L. aurea* (Appendix Fig. 1). As shown by the MaxEnt models (Appendix Fig. 2), this area does not overlap with any area of high habitat suitability.

Summarizing the information in Appendix Fig. 3-5 (Jackknife test of the gain contribution of single variables, response curves and variables permutation importance respectively), the distribution of *L. aurea* is mainly driven by the variable "solar", selecting therefore areas of high solar radiation. Precipitation seasonality ("bio15") and isothermality ("bio03") are other important factors, with this species selecting for stable environments in terms of temperature, but with a moderate seasonality in terms of rainfall. This probably reflects the need for water in the breeding season. *L. aurea* is limited in its altitude gradient to about 1000 m.a.s.l.. The most significative variables predicting the presence of *L. raniformis* are "natforest" and "roadkern", with the species avoiding natural forest and never reaching areas which are far from roads. Also, for this species, habitat suitability decreases with altitude, with a maximum of 2000 m.a.s.l. as the highest tolerable altitude. *L. raniformis* also seems to select flat terrains ("slope"). *L. ewingii* appears to be the more generalist of the tree species, not showing big differences among the importance given to different environmental variables. Its distribution is linked to cities of the least densely populated areas of the country (high "urban" values and low "roadkern" values). Also, this species selects low values of solar radiation ("solar") and precipitation

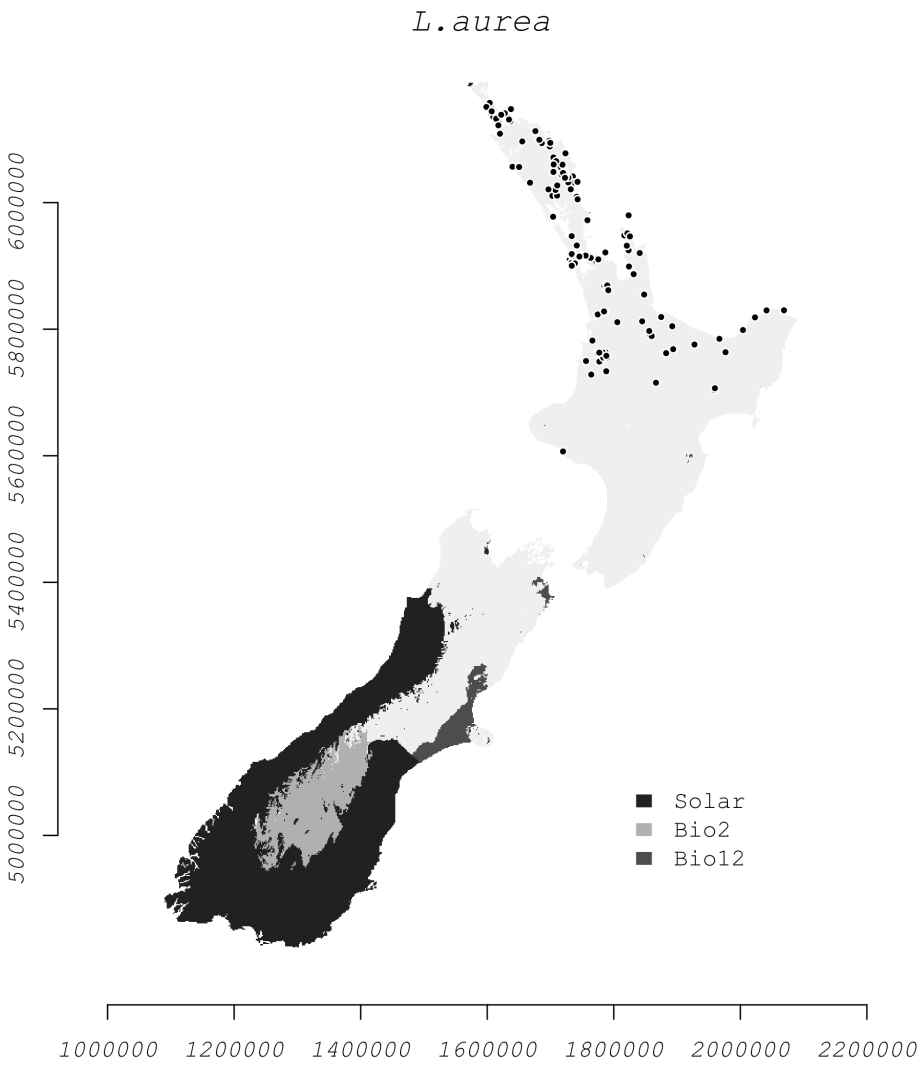
("bio15"). Presence of fresh water ("water") is selected here more than in the other two species.

Models of all the introduced *Litoria* species show areas of un-colonised, suitable habitat mainly within the current distribution range. *L. ewingii* is the species with the biggest area of possible range expansion, while *L. raniformis* presents an amost realised distribution range. Of course, this does not mean that population densities within the range cannot grow.

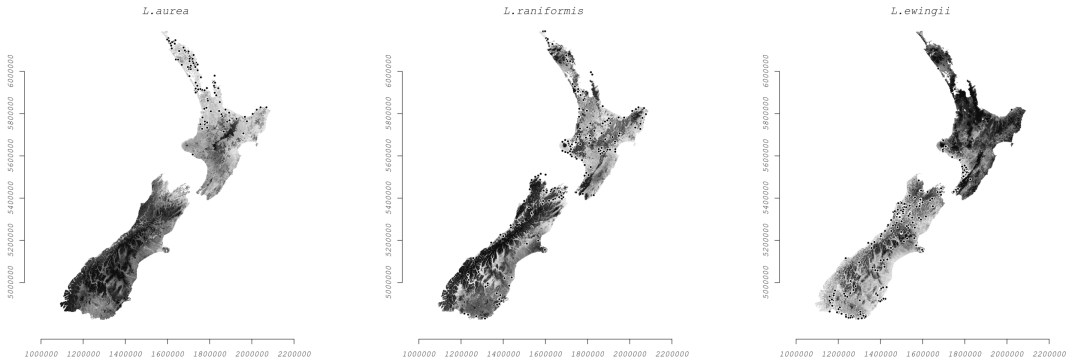
Tables and figures

Abbreviation	Meaning	Source	Type
Altitude	Altitude [m.a.s.l.]	Worldclim (Hijmans *et al.*2005)	Continous
Artforest	% of artificial forest	LUCAS 2nd edition (Newsome *et al.* 2013)	Continous
Bio02	Mean temperature diurnal range	Worldclim	Continous
Bio03	Isothermality	Worldclim	Continous
Bio12	Annual precipitation	Worldclim	Continous
Bio15	Precitpitation seasonality	Worldclim	Continous
Cropland	% of cropland	LUCAS 2nd edition	Continous
Natforest	% of natural forest	LUCAS 2nd edition	Continous
Roadkern	Kernel estimate of road density[0,1]	NZ Roads: Road Section Geometry (LINZ)	Continous
Slope	Land slope [%]	Worldclim	Continous
Solar	Solar radiation [MJ m^{-2} day^{-1}]	LENZ (Leathwick John *et al.* 2002)	Continous
Urban	% of urban area	LUCAS 2nd edition	Continous
Water	Presence/absence of water	LUCAS 2nd edition	Categorical
Wdgrassland	% of wooded grassland	LUCAS 2nd edition	Continous

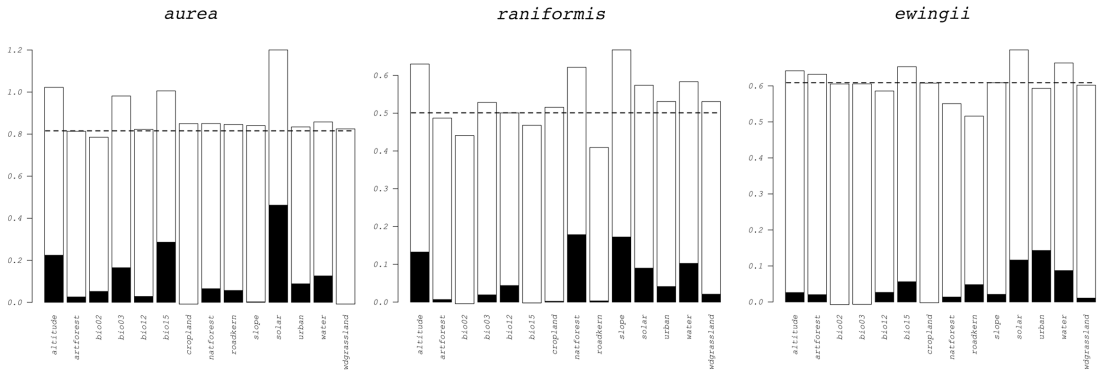
Appendix Tab. 1: Environmental variables used in the MaxEnt models of the Litoria spp.



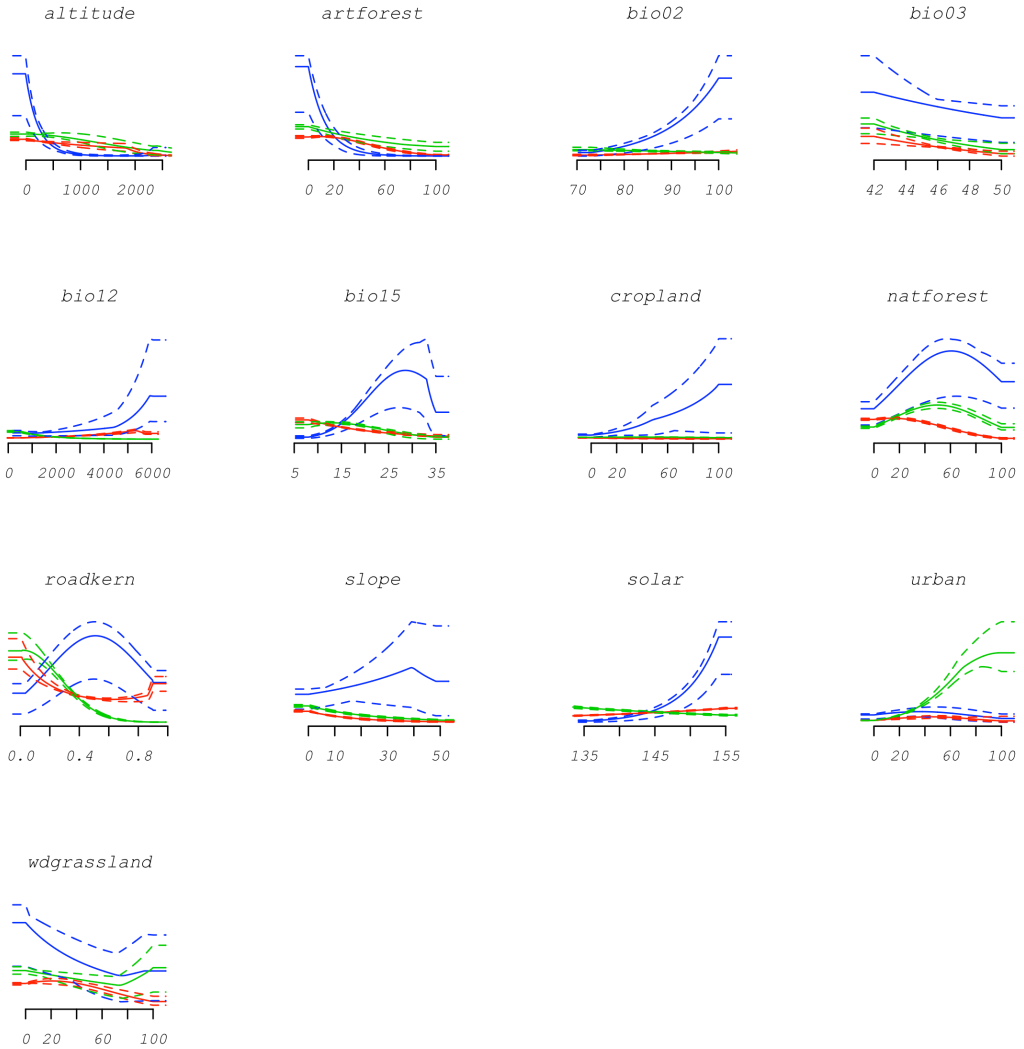
Appendix Fig. 1: Results of the MESS analysis for L. aurea. Shaded areas have at least one environmental variable outside the training range of the model. Different shades indicate the variable with the most dissimilar values compared to the training range.



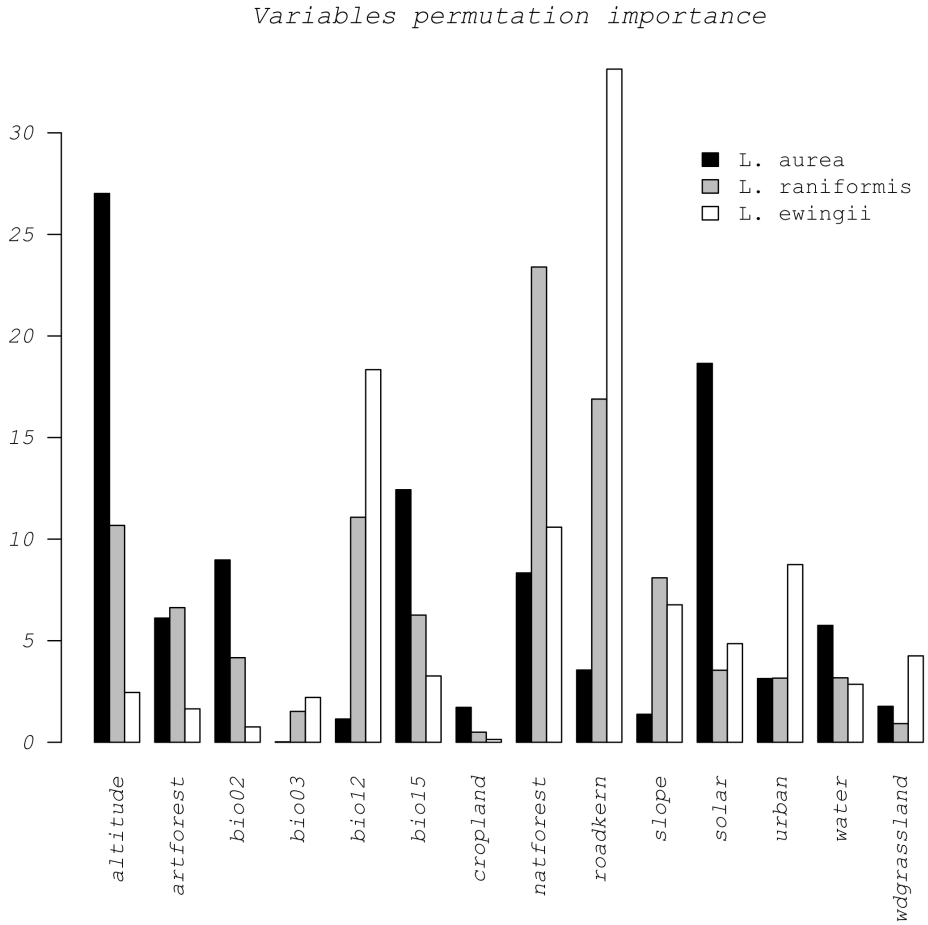
Appendix Fig. 2: MaxEnt models of Litoria frogs plotted on a quantile scale. Higher values are represented by lighter shades.



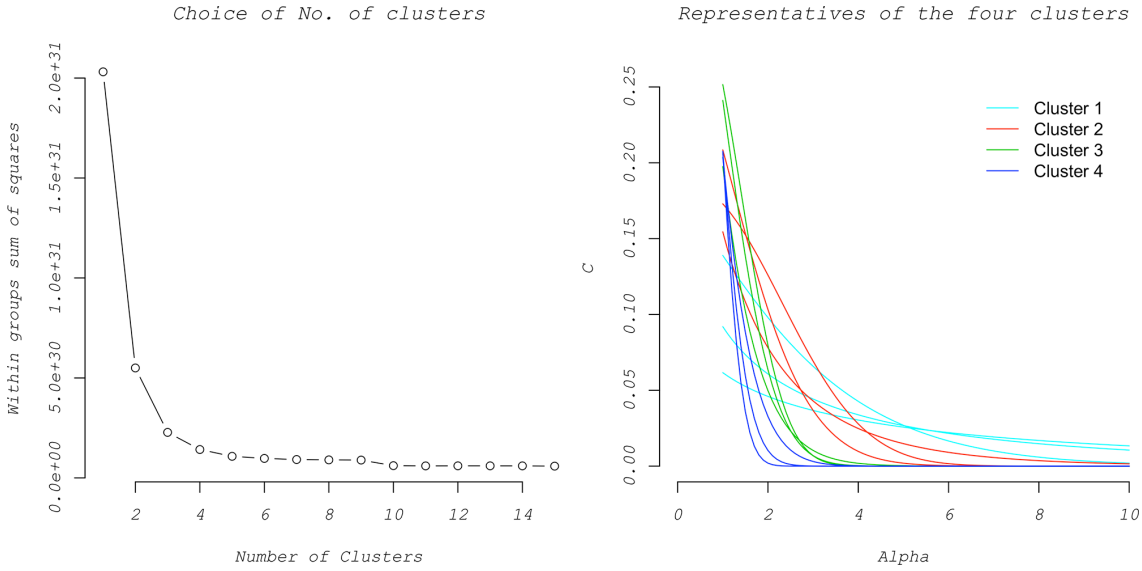
Appendix Fig. 3: Test gain of the model (on the 10% test data subsample) for each variable. White: on the model without the variable. Black: on a model with only that variable. Dashed line: Gain of model with all variables.



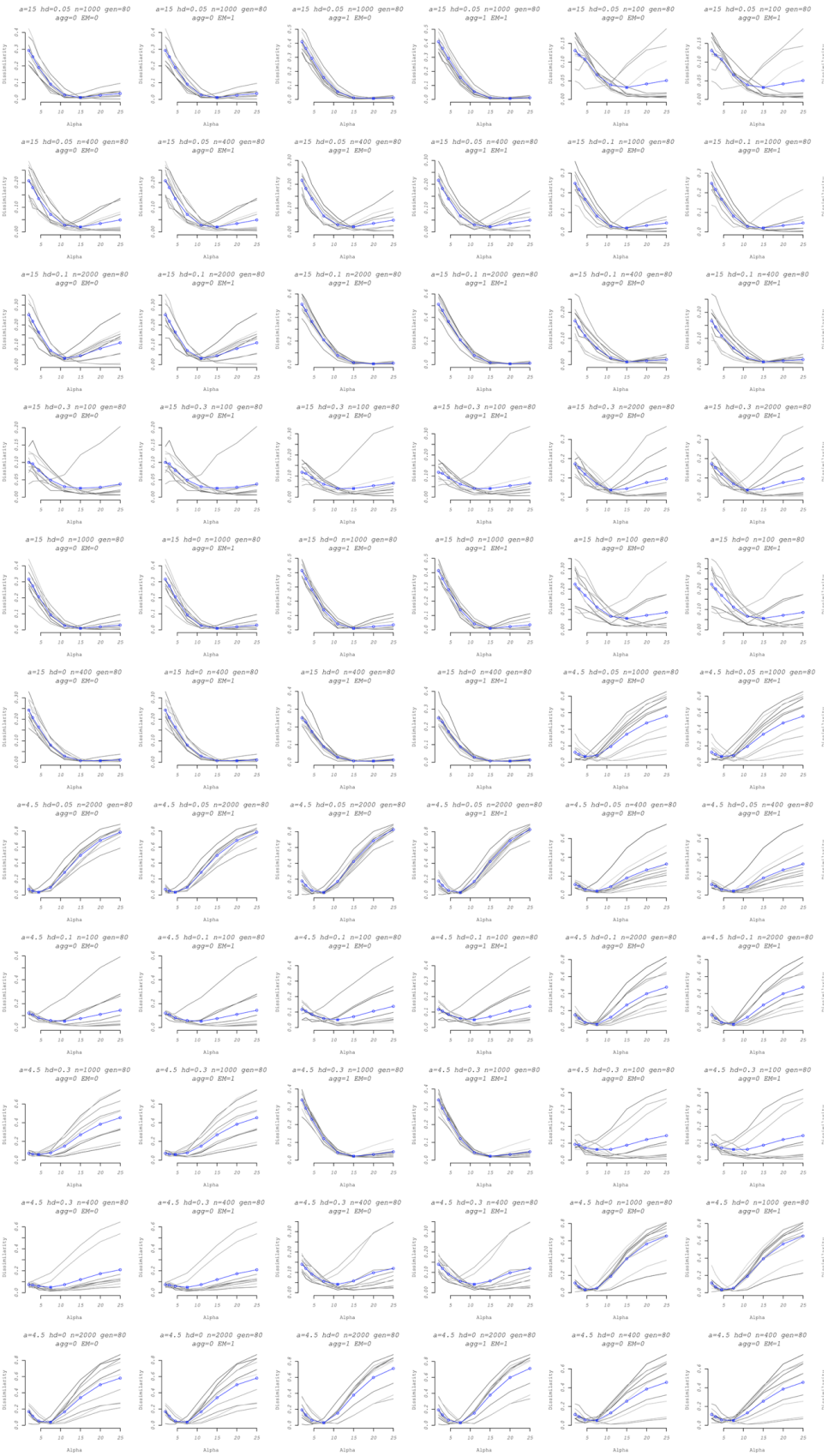
Appendix Fig. 4: Response curves for all environmental variables computed on MaxEnt models with a single variable per model. L. aurea in blue; L. raniformis in red: L. ewingii in green.



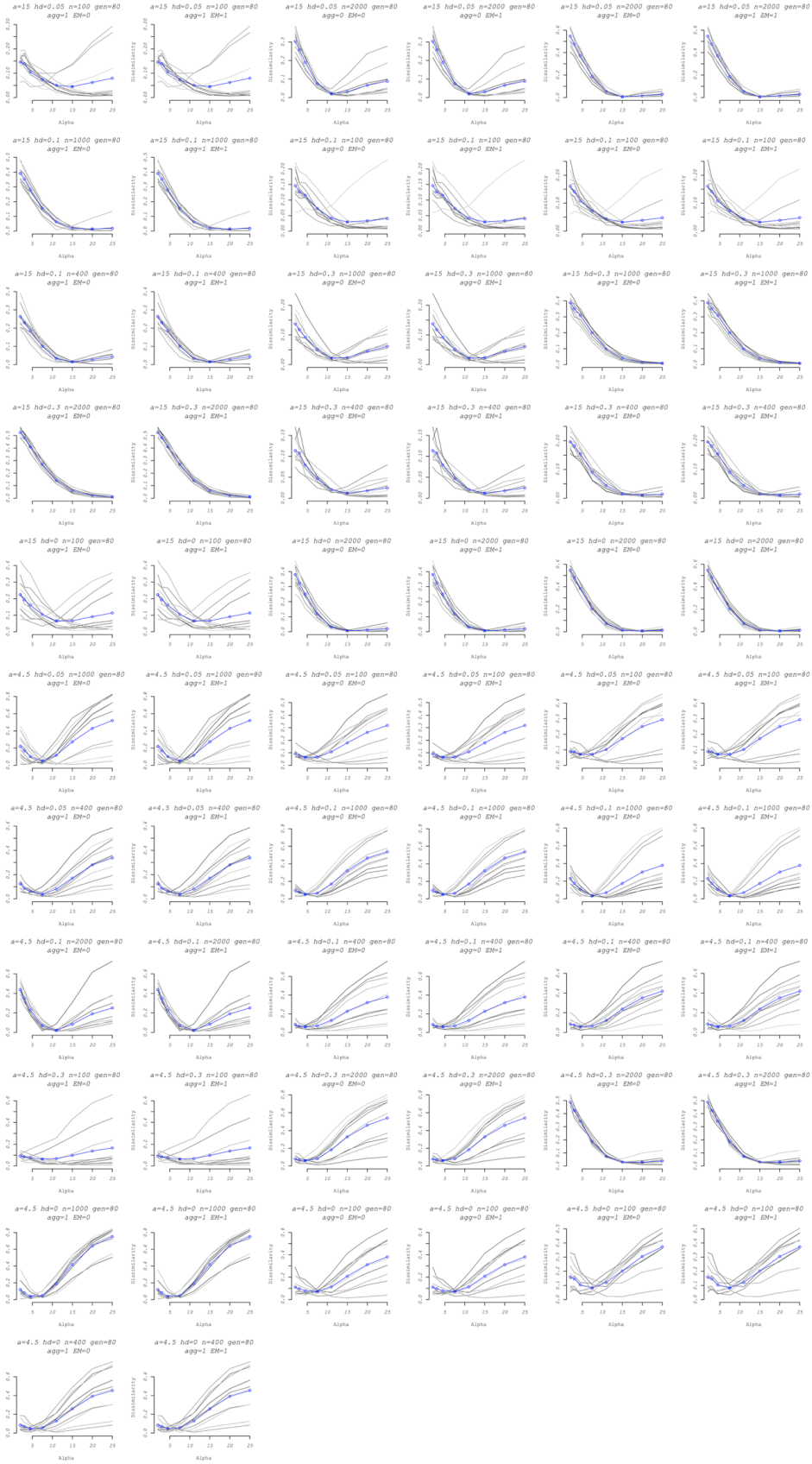
Appendix Fig. 5: Permutation importance of single environmental variables for the three species of Litoria.

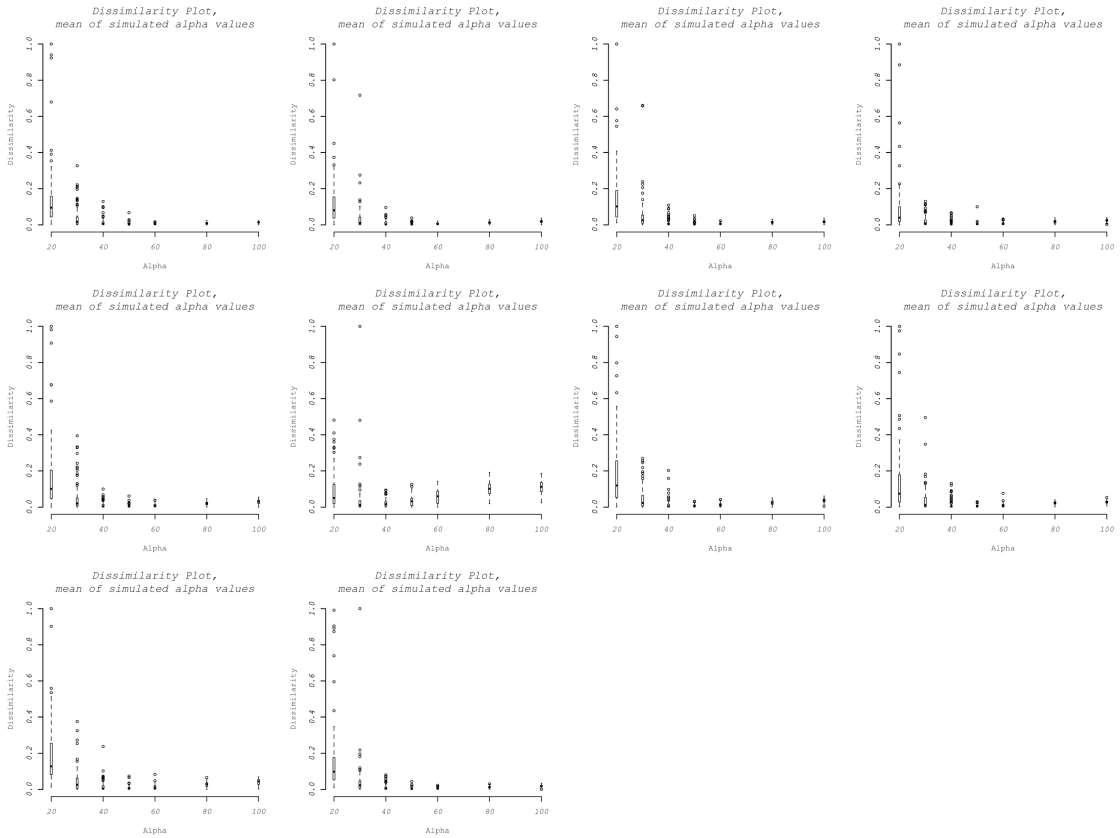


Appendix Fig. 6: Choice of No. of clusters. Plot of the within-groups sum of squares against number of groups for choosing the number of groups for the cluster analysis. "Representatives of the four clusters": Few dispersal kernels grouped in their respective cluster



Appendix Fig. 7: Similarity curves of a estimation for all 320 virtual datasets. Continues next page.


a=15 hd=0.05 n=100 gen=80 agg=1 EM=0
a=15 hd=0.05 n=100 gen=80 agg=1 EM=1
a=15 hd=0.05 n=2000 gen=80 agg=0 EM=0
a=15 hd=0.05 n=2000 gen=80 agg=0 EM=1
a=15 hd=0.05 n=2000 gen=80 agg=1 EM=0
a=15 hd=0.05 n=2000 gen=80 agg=1 EM=1
a=15 hd=0.1 n=1000 gen=80 agg=1 EM=0
a=15 hd=0.1 n=1000 gen=80 agg=1 EM=1
a=15 hd=0.1 n=100 gen=80 agg=0 EM=0
a=15 hd=0.1 n=100 gen=80 agg=0 EM=1
a=15 hd=0.1 n=100 gen=80 agg=1 EM=0
a=15 hd=0.1 n=100 gen=80 agg=1 EM=1
a=15 hd=0.1 n=400 gen=80 agg=1 EM=0
a=15 hd=0.1 n=400 gen=80 agg=1 EM=1
a=15 hd=0.3 n=1000 gen=80 agg=0 EM=0
a=15 hd=0.3 n=1000 gen=80 agg=0 EM=1
a=15 hd=0.3 n=1000 gen=80 agg=1 EM=0
a=15 hd=0.3 n=1000 gen=80 agg=1 EM=1
a=15 hd=0.3 n=2000 gen=80 agg=1 EM=0
a=15 hd=0.3 n=2000 gen=80 agg=1 EM=1
a=15 hd=0.3 n=400 gen=80 agg=0 EM=0
a=15 hd=0.3 n=400 gen=80 agg=0 EM=1
a=15 hd=0.3 n=400 gen=80 agg=1 EM=0
a=15 hd=0.3 n=400 gen=80 agg=1 EM=1
a=15 hd=0 n=100 gen=80 agg=1 EM=0
a=15 hd=0 n=100 gen=80 agg=1 EM=1
a=15 hd=0 n=2000 gen=80 agg=0 EM=0
a=15 hd=0 n=2000 gen=80 agg=0 EM=1
a=15 hd=0 n=2000 gen=80 agg=1 EM=0
a=15 hd=0 n=2000 gen=80 agg=1 EM=1
a=4.5 hd=0.05 n=1000 gen=80 agg=1 EM=0
a=4.5 hd=0.05 n=1000 gen=80 agg=1 EM=1
a=4.5 hd=0.05 n=100 gen=80 agg=0 EM=0
a=4.5 hd=0.05 n=100 gen=80 agg=0 EM=1
a=4.5 hd=0.05 n=100 gen=80 agg=1 EM=0
a=4.5 hd=0.05 n=100 gen=80 agg=1 EM=1
a=4.5 hd=0.05 n=400 gen=80 agg=1 EM=0
a=4.5 hd=0.05 n=400 gen=80 agg=1 EM=1
a=4.5 hd=0.1 n=1000 gen=80 agg=0 EM=0
a=4.5 hd=0.1 n=1000 gen=80 agg=0 EM=1
a=4.5 hd=0.1 n=1000 gen=80 agg=1 EM=0
a=4.5 hd=0.1 n=1000 gen=80 agg=1 EM=1
a=4.5 hd=0.1 n=2000 gen=80 agg=1 EM=0
a=4.5 hd=0.1 n=2000 gen=80 agg=1 EM=1
a=4.5 hd=0.1 n=400 gen=80 agg=0 EM=0
a=4.5 hd=0.1 n=400 gen=80 agg=0 EM=1
a=4.5 hd=0.1 n=400 gen=80 agg=1 EM=0
a=4.5 hd=0.1 n=400 gen=80 agg=1 EM=1
a=4.5 hd=0.3 n=100 gen=80 agg=1 EM=0
a=4.5 hd=0.3 n=100 gen=80 agg=1 EM=1
a=4.5 hd=0.3 n=2000 gen=80 agg=0 EM=0
a=4.5 hd=0.3 n=2000 gen=80 agg=0 EM=1
a=4.5 hd=0.3 n=2000 gen=80 agg=1 EM=0
a=4.5 hd=0.3 n=2000 gen=80 agg=1 EM=1
a=4.5 hd=0 n=1000 gen=80 agg=1 EM=0
a=4.5 hd=0 n=1000 gen=80 agg=1 EM=1
a=4.5 hd=0 n=100 gen=80 agg=0 EM=0
a=4.5 hd=0 n=100 gen=80 agg=0 EM=1
a=4.5 hd=0 n=100 gen=80 agg=1 EM=0
a=4.5 hd=0 n=100 gen=80 agg=1 EM=1
a=4.5 hd=0 n=400 gen=80 agg=1 EM=0
a=4.5 hd=0 n=400 gen=80 agg=1 EM=1
Dissimilarity
Alpha



Appendix Fig. 8: Dissimilarity values between comparison and observed datasets for the Litoria raniformis time series. Each boxplot represents similarity values of all simulated datasets with the same a value. Each plot is a Jackknifed dataset. HSM filtering was performed.