

## Utility of computer simulations in landscape genetics

BRYAN K. EPPERSON,\* BRAD H. MCRAE,† KIM SCRIBNER,‡ SAMUEL A. CUSHMAN,§  
MICHAEL S. ROSENBERG,¶ MARIE-JOSÉE FORTIN,\*\* PATRICK M. A. JAMES,†† MELANIE  
MURPHY,‡‡ STÉPHANIE MANEL,§§ PIERRE LEGENDRE¶¶ and MARK R. T. DALE\*\*\*

\*Department of Forestry, Michigan State University, East Lansing, MI 48824, USA, †The Nature Conservancy, 1917 1st Avenue, Seattle, WA 98101, USA, ‡Department of Fisheries & Wildlife and Department of Zoology, Michigan State University, East Lansing, MI 48824, USA, §USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, USA, ¶School of Life Sciences and Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, PO Box 874501, Tempe, AZ 85287-4501, USA, \*\*Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada M5S 3G5, ††Department of Biological Sciences, University of Alberta, Edmonton, Alberta Canada T6G 2E9, ‡‡Biology Department, Colorado State University, 1878 Campus Delivery, Fort Collins, CO 80523-1878, USA, §§UMR 151 UP/IRD, Laboratoire Population–Environnement–Développement, Université de Provence, centre Saint-Charles, Case 10, 3, place Victor-Hugo, 13331 Marseille Cedex 3, France, ¶¶Département de Sciences Biologiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, Québec, Canada H3C 3J7, \*\*\*Office of the Provost, University of Northern British Columbia, 3333 University Way, Prince George, British Columbia, Canada V2N 4Z9

### Abstract

Population genetics theory is primarily based on mathematical models in which spatial complexity and temporal variability are largely ignored. In contrast, the field of landscape genetics expressly focuses on how population genetic processes are affected by complex spatial and temporal environmental heterogeneity. It is spatially explicit and relates patterns to processes by combining complex and realistic life histories, behaviours, landscape features and genetic data. Central to landscape genetics is the connection of spatial patterns of genetic variation to the usually highly stochastic space–time processes that create them over both historical and contemporary time periods. The field should benefit from a shift to computer simulation approaches, which enable incorporation of demographic and environmental stochasticity. A key role of simulations is to show how demographic processes such as dispersal or reproduction interact with landscape features to affect probability of site occupancy, population size, and gene flow, which in turn determine spatial genetic structure. Simulations could also be used to compare various statistical methods and determine which have correct type I error or the highest statistical power to correctly identify spatio-temporal and environmental effects. Simulations may also help in evaluating how specific spatial metrics may be used to project future genetic trends. This article summarizes some of the fundamental aspects of spatial–temporal population genetic processes. It discusses the potential use of simulations to determine how various spatial metrics can be rigorously employed to identify features of interest, including contrasting locus-specific spatial patterns due to micro-scale environmental selection.

*Keywords:* individual-based models, landscape ecology, population genetics, simulations, spatial statistics

*Received 14 November 2009; revision received 8 April 2010; accepted 14 April 2010*

### Introduction

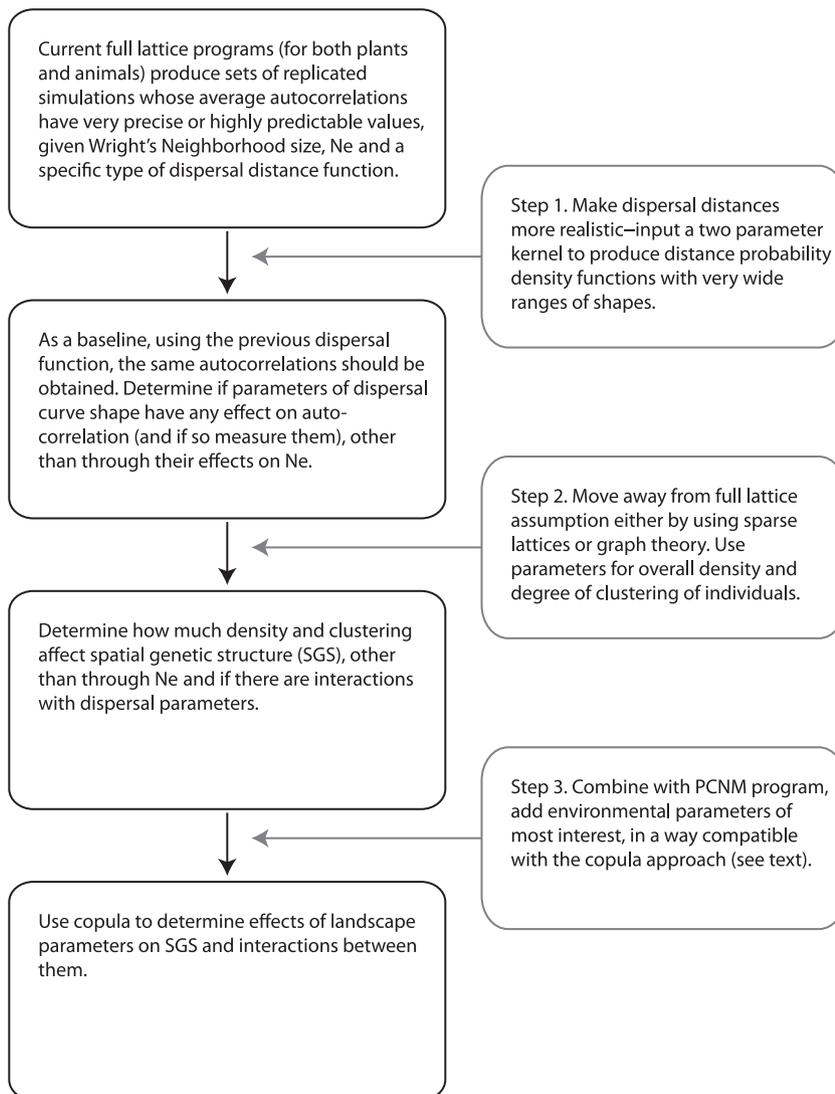
Computer simulations have had a rich and influential history in the fields of population genetics, evolution, ecology, and conservation. Particularly important are

Correspondence: Bryan K. Epperson  
E-mail: epperson@msu.edu

stochastic space–time simulations, which are probabilistic and in which data change in response to changing model states or inputs over time. Such simulations can incorporate stochasticity, spatial heterogeneity, individual variation, adaptive traits, historical effects and many other complexities in ways that would be intractable for mathematical models. They also allow system-level dynamics and patterns to emerge from parameterization of individual behaviour. Because random events and traits of individual organisms are key drivers of population dynamics and evolutionary processes, simulation approaches to understanding complex natural systems were adopted as soon as sufficient computing power became available (Turner *et al.* 1982; Grimm & Railsback 2005). Simulation models have been used in genetics decology for multiple objectives, including successfully advancing theory (e.g. source–sink dynamics, Wiegand *et al.* 1999), testing how robust analytical

predictions are to deviations from model assumptions (e.g. the effects of deviation from an island model on estimates of  $Nm$ , Slatkin & Barton 1989), and evaluating the performance of statistical tests (e.g. Legendre 2000). Moreover, their flexibility allows simulations to be tailored to specific, applied cases (e.g. Noon & McKelvey 1992; Taylor *et al.* 2000; Schumaker *et al.* 2004; Real & Biek 2007).

Most simulations fall into two broad categories: generic and specific. Generic simulations use broad models and parameters (usually derived from the literature, but often encompassing a wide range of values) and are generally used to examine the outcomes of specific models under a variety of conditions (parameter options; Fig. 1) or to test the ability of analytical methods to recover/estimate the input parameters that went into a model. In contrast, specific simulations attempt to mimic a specific observed population or condition



**Fig. 1** An example of a flow chart for Monte Carlo simulation studies that can elucidate various landscape genetic effects, by progressively adding complications to specifications. Initially, Moran  $I$  statistics may be used, but any other spatial metric could also be studied. Each step (right hand side boxes) uses a suite of sets of replicated simulations to address how dispersal, density and ultimately landscape features may affect spatial genetic structure, SGS, in increasingly complex processes of landscape genetics, resulting in increased knowledge (left hand side boxes). Each step provides information about how parameters behave and how much specific factors may be ignored in pursuit of general results. Step 1 should define large areas of parameter space over which the shape of the dispersal function has little or no effect on SGS. Future researchers, for example undertaking Step 2, could decide whether or not to ignore dispersal shape and just arbitrarily choose one, although it would be prudent to have some dispersal shape variants in case there are interactions between dispersal shape and density and clustering in affecting SGS. If there are no interactions, or if they can be managed, then favourable conditions are set for dealing with dispersal function, density, and clustering in carrying out Step 3. It is especially important to have the precise predictions used in Step 1, since with Monte Carlo simulations it can be difficult to definitively debug programs.

(Gauffre *et al.* 2008; Novembre & Stephens 2008). They are generally used to test under what conditions (or models or parameters) the observed data could have been generated or to predict what may happen to a specific population in the future. Both generic and specific simulations have proven quite useful for population genetics and landscape ecology.

Landscape genetics, as an amalgam of population genetics and landscape ecology (Manel *et al.* 2003; Holderegger & Wagner 2006; Storfer *et al.* 2007), can benefit from simulation modelling approaches, given its explicit focus on spatial complexity. Indeed, landscape genetics aims at understanding the processes governing spatial patterns of genetic variation, in response to environmental spatial heterogeneity. Landscape genetic studies combine the power of increasingly available spatially referenced multilocus genotypic data (e.g. Joost *et al.* 2007; Herrera & Bazaga 2008) with geographic information systems (GIS) and remote sensing data, greatly expanding our ability to understand sources of spatial genetic structure (Manel *et al.* 2010a). Simulations are especially useful and relevant to the challenges of landscape genetics based on several considerations. First, fitness and dispersal variability among individuals have numerous critical consequences for the dynamics, structure, and evolution of populations. Second, because landscape genetics focuses on the effects of landscape properties on genetic structure (Storfer *et al.* 2007), it is appropriate that models explicitly and realistically incorporate spatial heterogeneity and local interactions. Third, population history can also have strong effects on the structure of natural populations (Knowles 2009), confounding attempts to eluci-

date more contemporary landscape effects. Finally, the applied nature of many landscape genetic investigations, for example in conservation biology (Bruggeman *et al.* 2009), means that practitioners often grapple with 'real world' questions in complex, nonideal, and idiosyncratic systems. All of the above elements can be difficult or impossible to accommodate in purely mathematical modelling frameworks, whereas the flexibility and generality of simulations allows considerable complexity and biological detail to be included.

The historical progression of models and statistical methods in ecology and genetics, from aspatial to spatially implicit (where it is implicit that there are variations among individuals or populations over space, but the locations or the spatial relationships among them are not specified) to spatially explicit (Table 1), has led landscape genetics to a modelling framework which both is spatially explicit (i.e. spatial locations of individuals are monitored) and incorporates realistic landscape features and heterogeneity in a meaningful way. In this article, we review many relevant theoretical developments and examine the general roles that stochastic space-time simulations could play in landscape genetics (i.e. ranging from developing analytical models, testing hypothesis, making predictions, studying a specific or applied system, to comparing statistical methods). We then identify specific outstanding biological questions and critical processes that should be addressed, using simulation-based approaches. We review existing computer programs and suggest guidelines for future simulation studies in landscape genetics. Throughout the article, focus is maintained on continuous space processes or approximations of them, rather than discrete

**Table 1** Sequences of developments in population genetics, population biology, spatial statistics and landscape genetics, from aspatial to spatially explicit

Level	Population genetics	Population biology	Statistical geography	Landscape genetics
Aspatial	Gene drift, Gene flow, Gene mutation	Demography (Leslie matrix)	Regression, summary statistics	
Spatially implicit*	Wright's Island Model	Metapopulation	Regression with environmental variables	
Spatially explicit†	Isolation-by-distance, Isolation-by-barrier; stepping-stone	Spatially explicit landscape + individual based models (IBM) ‡	Univariate: spatial: CAR, SAR, autologistic, geographical weighted regression; spatio-temporal: STARMA, hierarchical Bayesian state-space, etc. multivariate: PCNM/Moran eigenvector maps, Mantel tests	Spatially explicit landscape + IBM + genetics

\*Models or methods where it is implicit that there are variations over locations, but locations are not assigned to specific places in space.

†Models or methods where individuals or populations are assigned to specific locations or the distances between them are defined.

‡While spatially explicit population-based models continue to contribute to these fields, the trends appear to be toward individual based models.

population models, for two reasons. First, most existing theoretical population genetic results on continuous models are paralleled by those for discrete populations (e.g. Epperson 2003). Second, salient features of discrete population models can be accommodated in continuous models by including a clustering factor (e.g. Barton & Wilson 1995) without the often arbitrary delineation of populations.

### Historical context and existing theoretical results relevant to simulations in landscape genetics

In order to understand the future of simulation studies in landscape genetics, it is helpful to review the past developments of models and results in the fields of population genetics, landscape ecology, and statistical geography (Table 1).

#### *Population genetics*

In population genetics, spatially explicit studies of genetic structure began with Wright's (1943) model of isolation by distance (IBD), in terms of decreasing levels of inbreeding within increasingly large areas (Table 1). Malécot (1948) reformulated Wright's model in terms of pairwise probabilities of identity by descent, which presaged modern methods. Both used simple Wright-Fisher discrete-generation lifecycles with added mating by proximity. Subsequently, several hundred mathematical papers, some with variant models or approaches (see e.g. Nagylaki 1989 for a review), developed many important mathematical relationships about the expected values of spatial metrics in stochastic spatial-temporal population genetic processes (e.g. Nagylaki 1986; Soboleva *et al.* 2003). However, it also became clear that the most mathematically difficult processes are those that exist in two spatial dimensions, especially at fine spatial scales; precisely the cases that are of most interest in landscape genetics today. It appears likely that with the added complexity of processes in landscape genetics, purely mathematical models will not lead to analytical solutions.

A second wave of theoretical studies relevant to simulations of landscape genetics began when relatively fast computers became available, starting with Turner *et al.* (1982) simulating Wright's approach (and using *F*-statistics) and Sokal & Wartenberg (1983) beginning a series of Monte Carlo simulation studies of the same models, but using Moran's *I*-statistics (which are closely related to other pairwise measures of identity by descent, Barbujani 1987; Hardy & Vekemans 1999). Since then, simulation studies have generally kept all of the assumptions required for discrete-generation life cycles and the assumption that individuals existed on points

of a lattice. They also usually modelled isotropic dispersal, represented very large populations (typically 10 000 diploid individuals), and included stochastic dispersal and stochastic Mendelian segregation. These types of simulation studies have well characterized the amount of spatial genetic structure that is expected based on basic dispersal characteristics under simple IBD processes. Importantly, they have also shown that realizations of IBD are themselves highly stochastic, or 'noisy' (e.g. Epperson 2003). We can expect even noisier correlations to be measured in empirical samples. It should also be noted that IBD can form the basis of null hypotheses for the operation of landscape factors rather than the usually unrealistic null hypothesis of a random distribution (Sokal & Wartenberg 1983; Fig. 1). Further, it has been shown recently that noise-to-signal ratio, as measured for example by coefficients of variation of Moran's *I*, tends to increase with spatial scale (Epperson submitted), which has implications for how best to design sampling schemes (e.g. sample within or beyond the range of IBD) in simulation studies, and how to combine spatial scales in statistical measures. For example, the detection or characterization of the effects of a particular landscape variable on spatial genetic structure may require greater numbers of replicate simulations, samples, or loci for larger scales. Similar general relationships between spatial scale of genetic structure and scale of movement (Anderson *et al.* 2010) could be expected in more complex landscapes, although sometimes altered.

In addition, IBD simulation studies have revealed the spatial genetic effects of mutations (Epperson 2005), random immigration and simple natural selection (reviewed in Epperson 2003), as well as some knowledge about clines due to microenvironmental gradients and local adaptation (Sokal *et al.* 1997). Some knowledge has also been gained regarding the effects of barriers to gene flow (e.g. Dupanloup *et al.* 2002; Guillot *et al.* 2005; Latch *et al.* 2006; Gauffre *et al.* 2008; Murphy *et al.* 2008; Frantz *et al.* 2009) and other environmental heterogeneities (Travis & Ezard 2006; Real & Biek 2007). In future contributions to landscape genetics, these types of simulations could be modified to incorporate more realistic life cycles, move away from the assumption of a (full) lattice, and include landscape features (Fig. 1).

Landscape genetics is becoming increasingly multilocus (e.g. Manel *et al.* 2010a), yet little is known about how multilocus genotypes behave in spatio-temporal systems. One two-locus simulation study produced some rather surprising results (Epperson 1995): the level of linkage disequilibrium (LD), was very small population-wide, whether or not the two loci were linked, although LD was quite large at smaller spatial scales.

We may expect that LD will generally change with spatial scale in more realistic landscapes as well, and that the relationship of LD with recombination rates will be complex. Such complexity could confound multilocus characterizations in simulation studies, for example of marker-trait associations in a landscape setting (Manel *et al.* 2010a). Finally, although much of the focus has been on pairwise measures, it should be noted that simulation studies can also determine spatial patterns in terms of genetic diversity measures such as heterozygosity and  $F_{ST}$  (e.g. Miller & Lacy 2005; Landgruth and Cushman 2009).

### *Landscape and population ecology*

Some of the earliest simulations in ecology were of forest stand dynamics (e.g. Botkin *et al.* 1972), but they quickly expanded into wildlife and fisheries management, movement ecology, theoretical ecology, conservation biology, metapopulation biology, and other ecological subdisciplines (e.g. Fahrig & Paloheimo 1988; Noon & McKelvey 1992; Palmer 1992; Turner *et al.* 1993; Zollner & Lima 1999; Tracey 2006). The advent of the field of landscape ecology (Turner 1989), coupled with increased availability of remote sensing data and computing power, fuelled the development of spatially explicit population models (Turner 1995; Table 1) and landscape disturbance models (LANDIS, He & Mladenoff 1999). Some were population-oriented, but increasingly such simulators follow individuals through their lifetimes, with daily to yearly updates, keeping track of sex, age, and events such as birth, mating, dispersal, and death (e.g. VORTEX, Lacy 1993; Miller & Lacy 2005), with transitions between life stages determined by Leslie matrices (Caswell 2000; Table 1). In more advanced simulators, individuals interact with habitat patches, other individuals, stressors, barriers, and/or other landscape elements, allowing the effect of changing landscape features on populations to be studied (e.g. Schumaker 1996; Schumaker *et al.* 2004). Increasingly, such models are incorporating adaptive behaviour (e.g. Railsback & Harvey 2002; Goss-Custard *et al.* 2006). Although many are constructed for specific purposes, some are designed as general and user-configurable tools (e.g. HexSim; Schumaker 2009). Simulations have also been used to incorporate both environmental control and spatial autocorrelation in spatially explicit predictions of community composition (e.g. Legendre *et al.* 2002, 2005).

Individual-based models (IBM) are beginning to combine stochastic ecological and genetic processes in a spatially explicit setting, as is required for landscape genetics (Table 1). By focusing explicitly on individuals, such models are freed from the Fisherian view of popu-

lations and statistics (e.g. using the normal distribution and a 'random sample' to estimate a population mean), and instead their results can be directly compared to spatial distributions of individual genotypes in empirical studies. For example, AMELIE (Kuparinen & Schurr 2007) is a plant population model that focuses on individual plant genotypes and incorporates flexible life histories, reproductive systems, and demography, with user-defined pollen and seed dispersal kernels. CDPop (Landguth & Cushman 2009) simulates dispersal, mating, and genetic exchange as probabilistic functions based on distance matrices rather than fixed kernels. These matrices are in turn derived from landscape ecological models, for example, least-cost path models which impose 'costs' imposed by movements through a 'resistant' landscape (Balkenhol *et al.* 2009; Spear *et al.* 2010). This provides a framework that allows comparison of multiple alternative models of landscape resistance directly with IBD and isolation by barriers into panmictic populations, enabling linkage between empirical analysis and simulation modelling. The modelling framework is specifically designed to provide a simulation framework that is directly amenable to causal modelling approaches for evaluating alternative landscape resistance hypotheses (e.g. Cushman *et al.* 2006).

### *Statistical geography*

Developments in spatial statistics and stochastic space-time process theory originally designed for general variables in statistical geography (Table 1) also have implications for analysing simulations in landscape genetics. The simplest model, known as Space-Time Autoregression Moving Average (STARMA; Haining 1979), happens to subsume stepping stone models of genetic drift and migration (Epperson 1993); hence STARMA theorems and results are directly applicable to population genetics. The formal STARMA concept of a well-defined lag structure (Hooper & Hewings 1981) or spatial 'regularity' (Epperson 2003), the degree to which spatially determined interactions among locations are repeated and scaled up over the landscape, dictates that spatial statistical analyses are more complicated in systems with less spatial regularity. Here it should be noted that spatial statistical modelling can be applied to somewhat irregular structures, for example, by using spatial eigenfunction analysis (Borcard & Legendre 2002; Dray *et al.* 2006). Regularity considerations, together with the high stochasticity or noisiness of space-time population genetic processes, also dictate that spatial genetic relationships must be replicated in a spatially regular way many times or insufficient statistical power will result, although to some degree this could be offset by using large numbers of genetic markers (Anderson *et al.* 2010).

Researchers can design, control and contrast spatial regularity in simulations and statistics, in order to improve statistical characterization of the effects of landscape features on spatial genetic structure.

STARMA theory, as well as simulations of IBD, has also shown that the manner in which stochastic inputs enter a system can have profound effects on spatial genetic structure. Inputs can occur through chance effects of dispersal distances and through genetic transmission, and it is critical that simulations include both. The distances that genes disperse (based on the inherent and biological tendencies of movements of individual animals or plant propagules) are typically represented by exponential or more complex long-tailed (large variance) distributions (e.g. Gregory 1968; Klein *et al.* 2003). In addition, at least for diploids, genetic transmission, or the process by which genes are transmitted from parents to offspring is also highly stochastic, even if following simple mating rules and Mendelian segregation. Further, genetic transmission in diploids can produce stochastic events that arise at locations rather than during movements (in STARMA, genetic drift vs. stochastic migration). Hence, it is critical that simulation studies in landscape genetics represent the appropriate ploidy level.

In addition, STARMA studies show that the manner in which stochastic events are propagated or 'percolate' (O'Neill *et al.* 1988) through the system largely determines the spatial genetic structure (Epperson 1993). Most critical perhaps is the effective number of spatial dimensions (i.e. whether habitat is essentially linear or arrayed in two or even three dimensions). Stochastic inputs in systems with one dimension tend to propagate less effectively and tend to produce higher expected correlations (Epperson 1993), but have greater variability in correlations, that is, are noisier (Fix 1994) than in two- or three-dimensional systems. Thus for example when gene flow is channelled through a corridor on a two-dimensional landscape, the spatial genetic structure in and near the corridor would differ and resemble that of a one-dimensional migration system. Terrestrial plants and animals can usually effectively occupy only one or two dimensions, whereas aquatic plants and animals, as well as soil-dwelling organisms, could occupy three-dimensional habitats. Thus it is important that both actual and effective spatial dimensionality are properly considered in simulations; for example, existing lattice-based models could be easily modified to 1D or 3D habitats.

It is also important to point out that the relationships among movement, life history characteristics, and spatial genetic structures depend on the fact that the latter build over time and generally do not arise instantaneously (e.g. spatial ARMA; Haining 1979). However,

available evidence indicates that most spatial genetic autocorrelation is generated over surprisingly short time periods, roughly 20 or so generations, as was shown most clearly in a coalescence analysis of IBD (Barton & Wilson 1995). Furthermore, changes in landscape structure, such as a new barrier, can change spatial genetic structure within as few as five generations (Murphy *et al.* 2008). Such considerations are particularly important when attempting to use simulations to distinguish effects of contemporary gene flow from historical events or processes in simulation studies.

Gene flow is generally considered solely as a 'smoothing' or homogenizing force (e.g. Wright 1965). However, both STARMA theory (e.g. Bennett & Haining 1985) and simulations (Epperson 2007) show that dispersal also can be the *cause* of spatial autocorrelation. There are circumstances where an increase in the rate of dispersal can increase the amount of autocorrelation (Epperson 2007).

## General roles of simulations in landscape genetics

### *Testing model assumptions*

One of the most crucial roles of simulations is in testing the robustness of process models to their assumptions, that is, the predictive capabilities of mathematical and analytical models, and the power of statistical methods, when applied to complex 'real-world' systems. Any model (analytical, simulation, and otherwise) makes simplifying assumptions (Box 1979), unless it were 'an entire reconstruction of the actual system—whereupon it ceases to be a model' (Burgman & Possingham 2000). Simple models remain useful, however, for capturing some salient aspects of complex systems or problems (Grimm 1999), developing theory, testing hypotheses, and for making predictions. More complex models tend to be more difficult to understand and derive general lessons from. Simulation studies can be structured to sequentially determine whether simple models (and their assumptions) are adequate for specific questions or whether additional complexity is necessary to draw conclusions about a particular type of system, by sequentially adding components or violating assumptions (see, e.g. Fig. 1).

For example, simple models of IBD with two spatial dimensions originally assumed an animal model with equal movement of the sexes (Wright 1943). Variant simulations have shown that summary measures of dispersal, particularly the total variance of parental dispersal distances as figured in Wright's neighbourhood size,  $N_e$ , are strongly predictive of global spatial autocorrelation. Predictive ability holds true even if

sexes differ greatly in dispersal ability or if plant models of propagule dispersal are used instead (Epperson 2007). Other studies have shown that the shape of the dispersal curve is usually not important (most recently by Lee & Hastings 2006). In addition, nonuniform density of individuals scarcely affects spatial genetic measures. However, when there is extreme clustering of individuals, global autocorrelation is (mildly) affected (Doligez *et al.* 1998). These results for IBD should not necessarily be taken to mean that these factors can always be safely ignored, especially where other simplifying assumptions are also violated.

Other simplifying assumptions about mating system and subtler aspects of life history are yet to be much explored in landscape genetics, many of which could be addressed through simulation modelling. These assumptions include: (i) that individuals do not vary in salient characteristics such as behaviour, (ii) that space and spatial processes are homogeneous (e.g. see Slatkin 1993; Rousset 1997), (iii) that demographic stochasticity (e.g. Engen *et al.* 2005) is inconsequential, and (iv) that populations are at equilibrium. In reality, individuals are often irregularly distributed across heterogeneous landscapes and face spatial heterogeneity in habitat quality, stressors, dispersal barriers, and dynamic interactions with predators, prey and competitors. Individuals may vary in genetic traits and in experience and learning, between sexes and during their life cycles (DeAngelis & Mooij 2005). Often-assumed genetic equilibrium is the exception, particularly in landscapes recently altered by human activities. Instead, phenomena such as fluctuating population sizes, range expansions, genetic bottlenecks, and recent contact zones are the norm. In such complex and contingent situations, simulation modelling offers a way to test how far applications of an analytical model can be stretched before it must be revised or discarded in favour of more complex or alternative modelling approaches (Slatkin & Barton 1989).

There is ample empirical evidence that subtle aspects of life history can alter spatial genetic structure. For example, in plants, different age groups can have differing spatial genetic structures, due to population self-thinning (Epperson & Alvarez-Buylla 1997) or during the early phases of population establishment (Epperson 2000); clonal reproduction can also drastically alter spatial genetic structure (e.g. Chung & Epperson 1999). In animals, it has been shown that home range and territoriality, natal habitat bias (e.g. Sacks *et al.* 2004), and social and feeding behaviour (e.g. Blanchong *et al.* 2006) can all come into play. Finally, although spatial genetic structure for systems with only one effective spatial dimension have been well-characterized theoretically, it appears underappreciated that many real land-

scapes, while essentially mostly existing in two (or possibly three) spatial dimensions, have areas where processes behave more as if they were in a single spatial dimension (Rousset 1997; McRae 2006). As noted, spatial genetic structures are strongly influenced by the effective number of dimensions.

#### *Characterizing properties of statistical estimators*

Another important role for simulation studies is in characterizing the properties of statistical estimators. Examples of questions or analyses include (i) tests for departure from spatial randomness, (ii) measures of distances over which there is positive autocorrelation, (iii) identification of specific locations of populations or individuals that comprise nonrandom associations, (iv) study of demographic variables (sex, ages, etc.) that are associated with structure, and (v) identification of landscape features associated with spatial genetic structure. Analyses that have been developed for (i–iv) include exploratory type analyses, model-based analyses and clustering methods. The degree of sophistication in methods has increased over time (Table 1) in accordance with the complexity of the research questions, and we can expect further increases. It is unlikely that appropriate statistics will have simple sampling distributions; hence generally their properties must be determined from simulations.

#### *Applications to real systems*

Lastly, simulations are of practical use when analysing landscape genetic data in real systems. By tailoring simulations to specific cases, they can address why field data collected to test a hypothesis may have spatial genetic distributions that differ from theoretical predictions. Hence they can be used to help decide whether or not to reject a hypothesis or determine if the assumptions of the mathematical models were substantially violated, if important processes were not considered, if the field sampling scheme was inappropriate, or if the utility of the genetic markers used was limited. Simulation can also help determine whether observed spatial genetic structure can be attributed to contemporary vs. historical isolating events, or provides evidence of recent range contraction or expansion (Currat & Excoffier 2004; Leblois *et al.* 2006; Wegmann *et al.* 2006; Cornuet *et al.* 2008). Many applied disciplines (e.g. wildlife management and conservation) can benefit from directly conducting simulations, such as understanding causal relationships between landscape features and gene flow. Managers could input the current state as an initial condition together with all known demographic and landscape parameters (with

appropriate levels of uncertainty) into models and run them to forecast spatial patterns under projected landscape and climate change. Simulations may also be used to optimize sampling schemes in advance of conducting empirical studies; for example, a manager of a system might consider doing a simulation study before investing time and resources into collecting data on landscape features and genetics.

## Outstanding issues in landscape genetics

### *Critical aspects of biological realism*

Landscape genetic studies could and ultimately should often involve a very large array of biological (including genetic) as well as landscape variables and parameters. As landscape genetic models add the effects of landscape features on biological processes such as individual dispersal and mating systems they must integrate these with the biological realism of individual behaviour and population dynamics. As noted, individual-based and other spatially explicit simulation modelling generally allow a systematic and theoretic relaxation of assumptions from ideal models, such as Wright–Fisher systems, with increasing biological realism and a variety of spatial dependence structures (e.g. Fig. 1).

Although landscape genetics generally focuses on how gene flow and spatial genetic structure are affected by landscape features, there are many other ecological and evolutionary processes affecting spatial genetic structures that could be investigated in a landscape simulation study. Genetic metrics used by landscape geneticists as dependent variables are typically based on variances in gene frequencies among populations [Wright's (1965) fixation indices] or pairwise genetic correlations (Cockerham 1969), including spatial genetic correlations between individuals (Cockerham 1973). Accordingly, models are needed that explicitly account for the manner in which genes are transmitted among individuals within and among groups, and for differential probabilities of dispersal among groups, (e.g. sex-bias or bias based on different population sizes, Scribner *et al.* 2001), leading to greater accuracy in estimation (Chesser 1991a,b, 1998; Sugg & Chesser 1994; Sugg *et al.* 1996).

In this context, for animals, an emerging view in landscape genetics is that it is essential to integrate information regarding behavioural ecology (Sugg *et al.* 1996) with other aspects of the mating system and social structure that influence gene transmission (Clobert *et al.* 2009). Demes may themselves be structured into sub-units (social group, family, clan, etc.), characterized by specific mating systems (monogamy, polygamy, pro-

miscuity, etc.), and connected by possibly sex-specific dispersal patterns. Factors that affect individual variation in reproductive success are complex and can change over time (Scribner & Chesser 2001). Various social systems have evolved in many animal species to make it common for males and/or females to disperse away from natal areas. In some cases, observations on genetic transmission (e.g. parent–offspring genotyping) can be combined with spatial population genetics (Robledo-Arnuncio *et al.* 2006), and these could be further combined with landscape features in order to more fully characterize biological processes.

For plants, species exhibit a wide variety of mating systems, including regular systems of frequent self-fertilization, negative assortative mating due to various kinds of incompatibility systems, and biparental inbreeding due to the clustering of related individuals within small neighbourhoods (e.g. Clegg 1980). Currently, it is largely unknown how spatial variations in these processes and environmental heterogeneity influence spatial genetic structure. In summary, future simulation studies of landscape genetics could profitably incorporate a very large array of biological as well as landscape variables.

### *Statistical properties*

Landscape genetics could further use simulations to provide precise results on the statistical properties of the various spatial statistics [e.g. Moran's  $I$ , Wombling (Womble 1951), landscape metrics] as used to evaluate the effects of landscape features (Fortin *et al.* 2003) on connectivity and genetic demographic parameters (e.g. migration rates, effective population size, mating system). This can be done by determining how such parameters affect spatial distributions of genetic variation (in contrasting sets of replicated simulations) and testing the various spatial metrics now available (many of these are discussed in other papers, e.g. Storfer *et al.* 2010) or those that will become available in the future. One of the strongest examples from existing work is the relationship between shortest distance spatial autocorrelation of genotypes,  $I_1$ , and overall amounts of dispersal in large continuous populations undergoing IBD, where a monotonic decrease of  $I_1$  with increasing variance of dispersal distances is found (Epperson 2007). Together with the values for stochastic and statistical variances (Epperson 2003), it is possible to estimate the dispersal variance from values of  $I_1$  observed in nature and determine the uncertainty of such estimates. In many other cases, it should be possible to set up multifaceted stochastic spatial-temporal processes (i.e. containing several, possibly interacting specific biological processes and landscape features) and statistically characterize

spatial measures designed to detect or measure said processes and features (Fig. 1). Most critically, studies could determine the type 1 and type 2 error rates of hypothesis tests constructed from spatial measures (e.g. Murphy *et al.* 2008; Legendre & Fortin 2010).

It is essential to evaluate the effects of landscape patterns on spatial population genetic processes across a wide range of appropriate spatial scales (Murphy *et al.* 2010; Anderson *et al.* 2010; Manel *et al.* 2010b). Contrasting sets of simulations could be used to examine how mismatched scales (e.g. grain vs. support; see Anderson *et al.* 2010) at the organismal, sampling, and analysis levels affect conclusions drawn from the analysis. Each organism will respond to environmental conditions at characteristic scales depending on its ecology, vagility and behaviour (Thompson & McGarigal 2002). In landscape genetic analyses incorrect specification of the thematic content, resolution, grain and extent over which organisms experience and respond to environmental variation may result in error in attribution of observed patterns in spatial genetic structure (Anderson *et al.* 2010). Recently, Cushman & Landguth (2010) used the CDPop model to evaluate how misspecification of the 'thematic resolution' and spatial grain of landscape patterns affects ability to detect and attribute effects of landscape on patterns of gene flow processes. They found that error in specifying the scale of landscape patterns can greatly impair the ability of landscape genetic analysis to correctly identify driving processes.

#### *Theoretical modelling of landscape effects on adaptation and natural selection*

Another set of important advances in landscape genetic simulation centres on the addition of selection to simulation models (Balkenhol *et al.* 2009), allowing exploration of the combined effects of gene flow and selection in complex landscapes, in order to better understand both evolutionary and ecological genetic processes. Previous studies used simple models that track both demography and the evolution of quantitative traits to study adaptation along a species range (e.g. Kirkpatrick & Barton 1997). More recently, Kramer *et al.* (2008) simulated adaptation by coupling an ecological and a genetic model tailored to European beech (*Fagus sylvatica* L.), with the aim of predicting how current management will affect adaptation. Gavrillets & Vose (2005, 2007) and Gavrillets *et al.* (2007) constructed an individual-based model to study speciation via local adaptation in spatially heterogeneous habitats. Neuenschwander *et al.* (2008) introduced an individual-based program to simulate the effect of selection and other genetic processes in structured populations located in hetero-

geneous habitats; however, landscape was not explicitly incorporated.

It is relatively straightforward to add selection to individual based models, in the form of spatial layers indicating differential mortality or fecundity as functions of the underlying environment and individual genotypes. Selection and gene flow can be combined by integrating presently separate simulators of IBD (e.g. Epperson 1990) with simulators of landscape spatial structure with multiple variables and at multiple spatial scales using principal coordinates of neighbour matrices, PCNMs (Borcard *et al.* 2004; Fig. 1; PCNM eigenfunctions are a special case of Moran's eigenvector maps, or MEM, Dray *et al.* 2006). PCNM can model the effects of many environmental variables, including both abiotic and biotic, as well as landscape features or environmental heterogeneity. Such simulation programs could also be combined with simulators of different degrees of habitat fragmentation and matrix quality (Bender & Fahrig 2005).

Further, selected (or functional) diversity arises from adaptive evolution due to natural selection, whereas diversity at neutral genetic loci is determined solely by the effects of genetic drift, mutation, or migration. Recent advances in genotyping techniques associated with increasing computational capacities and new statistical methods, including the use of simulation models, afford many opportunities to determine the effects of landscape features on contrasting spatial structures for functional vs. neutral genetic variation (Holderegger *et al.* 2006; Holderegger & Wagner 2008; Manel *et al.* 2010a,b). Awareness of the plausible trajectories of population change is increasingly important in situations of accelerated anthropogenic changes in population connectivity, demographic structure and abundance (Manel *et al.* 2010a).

#### *Integration of simulation studies with an empirical study of selection*

By integrating simulations with empirical data on large numbers of genetic markers, it may become possible to statistically detect loci under natural selection (Black *et al.* 2001; Luikart *et al.* 2003), especially important loci responsible for local adaptation, in a non-model species (Manel *et al.* 2010a). As noted above, the effects of evolutionary forces such as genetic drift or migration are replicated across all selectively neutral loci. In contrast, natural selection can result in an atypical pattern of variation at comparatively few loci, which can be interpreted as a possible signature of selection (Schlötterer 2003; Nielsen 2005; Storz 2005; Vasemägi & Primmer 2005). Correlative and other methods based on atypical differentiation of such loci have been developed to

identify signatures of selection throughout the genome without prior information regarding the traits or genes involved in the adaptation process (Vitalis *et al.* 2001; Schlotterer 2003; Beaumont & Balding 2004; Joost *et al.* 2007; Storz 2005; Foll & Gaggiotti 2008; Manel *et al.* 2010a,b). This approach offers a practical means to investigate selection in heterogeneous landscapes. Further, complex suites of contrasting simulations could ultimately provide an estimator of the probability that any particular gene is responding to environmental selection rather than being selectively neutral, in part by determining the proper 'experiment-wise' error rates.

### *Modelling uncertainty*

Simulation results could be used to evaluate uncertainty in estimated model parameters as a source of variation for model predictions. One important first step to understand model behaviour is to elucidate relationships between variation in parameters and the resulting variation in model outcomes. Similar needs are present for sensitivity analysis—for example, determining the relative importance of contributions of different parameters to the variation in model outcomes (Saltelli 2000; Fieberg & Jenkins 2005; Cariboni *et al.* 2007).

### **Future simulation studies**

Future simulation programs that have the power and flexibility required for addressing issues in landscape genetics will likely develop along several lines. These could include both the somewhat simpler models that are based on discrete or overlapping generations as well as more complex IBM-type models that operate in essentially continuous time. The former focuses more on spatial patterns of genetic variation accumulated in time, and the latter is a somewhat more mechanistic approach (Klein *et al.* 2003). Most useful approaches will include limited dispersal, since some limits to dispersal almost always pertain to study organisms and such limits are intrinsic to processes that determine spatial genetic structure. Less obvious perhaps is that generally the distances that individuals (or propagules for plants) disperse should be simulated as highly stochastic, as for example determined by a kernel that controls the programmed probability density function (or possibly probability mass function) for dispersal distances (Fig. 1).

Future simulation programs should also allow some relaxation from the lattice assumption and the uniform density it mimics (but see Doligez *et al.* 1998). However, this should be done carefully, since other complexities can then come into play, such as spatially based conspecific competition, which can be extremely

complicated to model in and of itself and would further complicate efforts to model microenvironmental selection. Competition may also conflict with factors that regulate the overall size of a population. Significant improvements in simulated realism which avoid such issues may be achieved by employing methods that use point locations such as graph-theoretic approaches (Urban & Keitt 2001; Garroway *et al.* 2008) or by using sparse lattices (Doligez *et al.* 1998). In addition, dispersal modelling must incorporate more complex dispersal rules, especially location-dependent rules, which can be constructed to represent partial or complete barriers to gene flow or corridors (for examples), yet maintain inherent dispersal characteristics. Models should allow some age-structure and similar deviations from discrete-generation assumptions. Finally, it should be recognized that the simulated state of spatial genetic structure may depend on its initial condition, how much time has passed, and the timing of changes in environment (see Anderson *et al.* 2010).

Landscape variations in habitat suitability and complex distributions of microenvironments that may also act differentially on different genotypes are critically important aspects of landscape genetics, and the future for simulations that combine these is promising. One very generalizable method for modelling spatial distributions of environmental factors at multiple spatial scales is the principle coordinate analysis of neighbour matrices (PCNM) approach (Borcard & Legendre 2002; Dray *et al.* 2006). We are currently building programs that combine PCNM with simulations of modified lattice-based IBD processes (Fig. 1), using R (R Development Core Team 2008). These programs are multi-allelic, allow either animal or plant mating systems (including self-fertilization), and include mutations and immigration events. Studies conducting such simulations could make significant progress in separating IBD effects from environmental effects on spatial genetic structure, a significant problem in landscape genetics. Similarly, such computer programs could be further combined with simulations of different degrees of habitat fragmentation and matrix quality (Bender & Fahrig 2005). Further, by contrasting theoretical spatial structures for loci subjected to environmental selection with those for loci that are neutral but undergoing all of the other same demographic features, it may be possible to develop methods for detecting natural selection. In addition, selection models should include multilocus selection, and should allow for epistasis and pleiotropy. Computer programs that incorporate all of the features discussed above would provide an important complement to the other approaches.

Combining individual-based models from ecology and genetics would also increase flexibility in the

incorporation of spatial and ecological details and processes. Such models may be particularly useful for specific management questions; an example would be determining how mortality risks during movement through novel, human-modified landscapes will influence gene flow and perhaps selection for new behaviours of a species facing future land use and climate change. Combining genetic simulators, simulators of landscape spatial structure [e.g. dynamic landscape simulators such as LANDIS (He & Mladenoff 1999) and SELES (Fall & Fall 2001; James *et al.* 2007)], and existing individual-based movement models can allow tracking of gene frequencies through time as populations change and evolve in response to changing conditions, while retaining high levels of biological detail. Table 2 lists many of the programs that may be useful (either in whole or in part) in a simulation study of landscape genetics. Selection could be added to models such as CDPop, in the form of spatial layers indicating differential mortality or fecundity as functions of environmental factors and individual genotypes.

For example, HexSim and SimuPop (Peng & Kimmel 2005) represent two of the more feature-rich simulators in ecology and population genetics, respectively. HexSim is an IBM that incorporates individual experiences that

affect traits like behaviour and fitness, such as exposure to modelled environmental stressors like pesticides (Schumaker 2009). Individual traits are heritable, allowing for incorporation of genetic components. Similarly, SimuPop is a highly flexible population genetic simulator that operates on individuals and incorporates realistic models of mutation, recombination, quantitative traits, selection, and other processes, and features pedigree tracing and calculation of a wide range of statistics such as gene frequencies, heterozygosity, and linkage disequilibrium measures (Peng & Kimmel 2005). Combining the strengths of such software packages would allow highly flexible landscape genetic simulations, including selective pressures that act on different life history stages, and providing outputs that can be compared to a wide range of empirical data (both ecological and genetic).

If various simulation approaches are compared, it seems likely that modified-lattice types of simulations and others with more regular structures will be more amenable to drawing analytical conclusions because of their relative simplicity and spatial regularity, particularly for large populations. More specifically, they may allow one to more easily determine basic properties of the relationship of process parameters and factors to resulting spatial-temporal genetic structure. On the

**Table 2** List of many of the relevant softwares

Category	Name	Description	Reference
Nonspatial	MS	Generate genotypes under Wright–Fisher neutral model	Hudson (2002)
	Ecogene	Mating system, dispersal,	Degen & Scholz (1998)
Semi-spatial	Aquasplatche	Simulate genetic diversity in linear habitats	Neuenschwander 2006
	EasyPop	Generate genotypic data with some degree of population structure (hierarchical stepping stone, IBD)	Balloux <i>et al.</i> (2004)
	IBDsim	Simulate genotypic data under isolation by distance	Leblois <i>et al.</i> (2009)
	Metasim	Individual based simulation framework for complex population dynamics	Strand (2002)
	quantiNEMO	Individual based program to investigate the effects of mutation, selection, recombination and drift on quantitative traits connected by migration through heterogeneous habitat.	Neuenschwander <i>et al.</i> (2008)
	SPLATCHE*	Simulate genetic diversity of sampled genes in a heterogeneous habitat. Coalescent framework with a unit of space equal to habitat	Curat <i>et al.</i> (2004)
	Fully spatial, without landscape heterogeneity	Mostly 'in house'	Monte Carlo space–time processes of isolation by distance, some with additional processes
Fully spatial with landscape heterogeneity	CDPop	Simulate genotypic data with variety of dispersal and mating systems. Based on landscape resistance	Landguth & Cushman (2009)
	EcoGenetics	Simulate stepping-stone dynamics as a function of landscape resistance	Hirzel <i>et al.</i> , (unpublished)
	SimSSD	Simulation of spatially structured data with or without spatial autocorrelation and/or deterministic spatial structure, with or without influence of explanatory variables	Legendre <i>et al.</i> (2002, 2005)

other hand, individual-based models will likely be more easily applicable as management tools, especially for small populations. As such models become more accessible and user-friendly, they may also play an increasing role as heuristic tools; this is because system dynamics emerge from simple behavioural rules at the individual level. Users can vary the input parameters and landscape variables and observe how more complex population-level processes and patterns emerge.

As models of relationships between suites of landscape features and measures of genetic differentiation become more complex, the numbers of variables and parameters in the models increase and the relationships among the effects of parameter values may not be linear. Hence, we can expect complex multidimensional parameter spaces to be the norm. Future simulation studies will need to use more efficient designs, in terms of combinations of parameter values to be simulated, than they have in the past in order for researchers to have confidence in the inferences made. Grimm & Railsback (2005) provide guidelines for model formulation, parameterization, testing, and analysis, and practical strategies for managing model complexity and data requirements, conducting sensitivity analyses, and running large numbers of replicates. Recently, methods for efficient parameter combination design developed in computer sciences have been applied to simulations in the biological sciences (Ragavendran 2009). One of these methods is the copula approach (Nelsen 2006)—using a function that defines the joint distribution of sets of parameter values for variables, based on their univariate marginal distributions. It is used to construct simulation studies that efficiently cover the entire parameter space, to estimate the strengths of effects of parameters and to detect linear and nonlinear dependencies among them. For example, it could be used to determine if different parameter combinations can give similar spatial structures. Lastly, landscape geneticists can take advantage of emerging model documentation standards (Grimm *et al.* 2006) to better convey details about increasingly complex simulation frameworks.

We believe that the field of landscape genetics could benefit substantially from a community effort to share simulated data sets. Access to communal benchmark simulated and empirical data sets would help standardize developments of new statistical methods for landscape genetic studies. As new statistics come on line, their performance could be directly compared to others, across a set of data designed to address any given array of biological processes included in the simulated stochastic spatial-temporal processes. Ultimately, such data sets could be employed to evaluate uncertainty, help inform sample design for applications in empirical studies, and aid management decisions.

## Acknowledgements

This work was conducted as a part of the An Interdisciplinary Approach to Advancing Landscape Genetics Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. MSR was also supported by the National Evolutionary Synthesis Center (NESCent: NSF #EF-0905606) and NSF #DBI-0542599. The authors also thank Ashok Ragavendran for helpful comments on the subject matter, and Lisette Waits, Doug Bruggeman, and an anonymous reviewer of an earlier version of the article.

## References

- Anderson CD, Epperson BK, Fortin M-J *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565–3575.
- Balkenhol N, Gugerli F, Cushman SA *et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455–463.
- Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, **13**, 3021–3031.
- Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. *Genetics*, **117**, 777–782.
- Barton NH, Wilson I (1995) Genealogies and geography. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, **349**, 49–59.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Bender DJ, Fahrig L (2005) Matrix structure obscures the relationship between interpatch movement and patch size and isolation. *Ecology*, **86**, 1023–1033.
- Bennett RJ, Haining RP (1985) Spatial structure and spatial interactions: modelling approaches to the statistical analysis of geographic data. *Journal of the Royal Statistical Society (Series A)*, **148**, 1–36.
- Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.
- Blanchong JA, Scribner KT, Epperson BK, Winterstein SR (2006) Changes in artificial feeding regulations impact white-tailed deer fine-scale spatial genetic structure. *Journal of Wildlife Management*, **70**, 1037–1043.
- Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, **85**, 1826–1832.
- Botkin DB, Janak JS, Wallis JR (1972) Some ecological consequences of a computer model of forest growth. *Journal of Ecology*, **60**, 849–872.
- Box GEP (1979) *Robustness in the Strategy of Scientific Model Building*. Academic Press, New York.
- Bruggeman DJM, Jones L, Scribner KT, Lupi F (2009) Relating tradable credits for biodiversity to sustainability criteria at a landscape-scale. *Landscape Ecology*, **24**, 775–790.

- Burgman MA, Possingham HP (2000) Population viability analysis for conservation: the good, the bad and the undescribed. In: *Demography and Viability of Fragmented Populations* (eds Young AG, Clarke GM), pp. 97–112. Cambridge University Press, Cambridge, UK.
- Cariboni J, Gatelli D, Liska R, Saltelli A (2007) The role of sensitivity analysis in ecological modelling. *Ecological Modelling*, **203**, 167–182.
- Caswell H (2000) *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd edn. Sinauer Associates, USA.
- Chesser RK (1991a) Gene diversity and female philopatry. *Genetics*, **127**, 437–447.
- Chesser RK (1991b) Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics*, **129**, 573–583.
- Chesser RK (1998) Relativity of behavioral interactions in socially structured populations. *Journal of Mammalogy*, **79**, 713–724.
- Chung MG, Epperson BK (1999) Spatial genetic structure of clonal and sexual reproduction in populations of *Adenophora grandiflora* (Campanulaceae). *Evolution*, **53**, 1068–1078.
- Clegg MT (1980) Measuring plant mating systems. *BioScience*, **30**, 814–818.
- Clober J, Le Galliard J-F, Cote J, Meylan S, Massot M (2009) Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology Letters*, **12**, 197–209.
- Cockerham CC (1969) Variance of gene frequencies. *Evolution*, **23**, 72–84.
- Cockerham CC (1973) Analysis of gene frequencies. *Genetics*, **74**, 679–700.
- Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Curat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology*, **2**, 421. doi 10.1371/journal.pbio.0020421.
- Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592–3602.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486–499.
- DeAngelis DL, Mooij WM (2005) Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 147–168.
- Degen B, Scholz F (1998) Ecological genetics in forest ecosystems under stress as analysed by the simulation model ECO-GENE. *Chemosphere*, **36**, 819–824.
- Doligez A, Baril C, Joly HI (1998) Fine-scale spatial genetic structure with nonuniform distribution of individuals. *Genetics*, **148**, 905–919.
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.
- Engen S, Lande R, Sæther BE, Weimerskirch H (2005) Extinction in relation to demographic and environmental stochasticity in age-structured models. *Mathematical Biosciences*, **195**, 210–227.
- Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics*, **124**, 757–771.
- Epperson BK (1993) Spatial and space-time correlations in systems of subpopulations with genetic drift and migration. *Genetics*, **133**, 711–727.
- Epperson BK (1995) Spatial structure of two-locus genotypes under isolation by distance. *Genetics*, **140**, 365–375.
- Epperson BK (2000) Spatial genetic structure and non-equilibrium demographics within plant populations. *Plant Species Biology*, **15**, 269–279.
- Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, New Jersey.
- Epperson BK (2005) Estimating dispersal from short distance spatial autocorrelation. *Heredity*, **95**, 7–15.
- Epperson BK (2007) Plant dispersal, neighbourhood size and isolation by distance. *Molecular Ecology*, **16**, 3854–3865.
- Epperson BK (2010) Spatial correlations at different spatial scales are themselves highly correlated in isolation by distance processes. *Molecular Ecology Resources*, doi:10.1111/j.1755-0998.2010.02886.x.
- Epperson BK, Alvarez-Buylla ER (1997) Limited seed dispersal and genetic structure in life stages of *Cecropia obtusifolia*. *Evolution*, **51**, 275–282.
- Fahrig L, Paloheimo J (1988) Effect of spatial arrangement of habitat patches on local population size. *Ecology*, **69**, 468–475.
- Fall A, Fall J (2001) A domain specific language for models of landscape dynamics. *Ecological Modelling*, **141**, 1–18.
- Fieberg J, Jenkins KJ (2005) Assessing uncertainty in ecological systems using global sensitivity analyses: a case example of simulated wolf reintroduction effects on elk. *Ecological Modelling*, **187**, 259–280.
- Fix AG (1994) Detecting clinal and balanced selection using spatial autocorrelation analysis under kin-structured migration. *American Journal of Physical Anthropology*, **95**, 385–397.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
- Fortin M-J, Boots B, Csillag F, Rempel TK (2003) On the role of spatial stochastic models in understanding landscape indices in ecology. *Oikos*, **102**, 203–212.
- Frantz AC, Do Linh San E, Pope LC, Burke T (2009) Using genetic methods to investigate dispersal in two badger (*Meles meles*) populations with different ecological characteristics. *Heredity*, **104**, 493–501.
- Garroway CJ, Bowman J, Carr D, Wilson PJ (2008) Applications of graph theory to landscape genetics. *Evolutionary Applications*, **1**, 620–630.
- Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, **17**, 4619–4629.
- Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18040–18045.

- Gavrilets S, Vose A (2007) Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Molecular Ecology*, **16**, 2910–2921.
- Gavrilets S, Vose A, Barluenga M, Salzburger W, Meyer A (2007) Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular Ecology*, **16**, 2893–2909.
- Goss-Custard JD, Burton NHK, Clark NA *et al.* (2006) Test of a behavior-based individual-based model: response of shorebird mortality to habitat loss. *Ecological Applications*, **16**, 2215–2222.
- Gregory PH (1968) Interpreting plant disease dispersal gradients. *Annual Review of Phytopathology*, **6**, 189–212.
- Grimm V (1999) Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling*, **115**, 129–148.
- Grimm V, Railsback SF (2005) *Individual-Based Modelling and Ecology*. Princeton University Press, Princeton, NJ.
- Grimm V, Berger U, Bastiansen F *et al.* (2006) A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, **198**, 115–126.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.
- Haining RP (1979) Statistical tests and process generators for random field models. *Geographical Analysis*, **11**, 45–64.
- Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.
- He HS, Mladenoff DJ (1999) Spatially explicit and stochastic simulation of forest-landscape fire disturbance and succession. *Ecology*, **80**, 81–99.
- Herrera CM, Bazaga P (2008) Adding a third dimension to the edge of a species' range: altitude and genetic structuring in mountainous landscapes. *Heredity*, **100**, 275–285.
- Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landscape Ecology*, **21**, 793–796.
- Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience*, **58**, 199–207.
- Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**, 797–807.
- Hooper PM, Hewings GJD (1981) Some properties of space-time processes. *Geographical Analysis*, **13**, 203–223.
- Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.
- James PMA, Fortin M-J, Fall A, Kneeshaw D, Messier C (2007) The effects of spatial legacies following shifting management practices and fire on boreal forest age structure. *Ecosystems*, **10**, 1261–1277.
- Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.
- Kirkpatrick M, Barton NH (1997) The strength of indirect selection on female mating preferences. *Proceedings of the National Academy of Sciences, USA*, **94**, 1282–1286.
- Klein EK, Lavigne C, Foueillassar X, Gouyon PH, Laredo C (2003) Corn pollen dispersal: quasi-mechanistic models and field experiments. *Ecological Monographs*, **73**, 131–150.
- Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology and Systematics*, **40**, 593–612.
- Kramer K, Buiteveld J, Forstreuter M *et al.* (2008) Bridging the gap between ecophysiological and genetic knowledge to assess the adaptive potential of European beech. *Ecological Modelling*, **216**, 333–353.
- Kuparinen A, Schurr FM (2007) A flexible modelling framework linking the spatio-temporal dynamics of plant genotypes and populations: application to gene flow from transgenic forests. *Ecological Modelling*, **202**, 476–486.
- Lacy RC (1993) VORTEX: a computer simulation model for Population viability analysis. *Wildlife Research*, **20**, 45–65.
- Landguth EL, Cushman SA (2009) CDPOP: an individual-based, cost-distance spatial population genetics model. *Molecular Ecology Resources*, **10**, 156–161.
- Latch EK, Applegate RD, Rhodes Jr OE (2006) Genetic composition of wild turkeys in Kansas following decades of translocations. *Journal of Wildlife Management*, **70**, 1698–1703.
- Leblois R, Estoup A, Streiff R (2006) Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Molecular Ecology*, **15**, 3601–3615.
- Leblois R, Estoup A, Rousset F (2009) IBDSim: a computer program to simulate genotypic data under isolation by distance. *Molecular Ecology Resources*, **9**, 107–109.
- Lee CT, Hastings A (2006) Non-equilibrium genetic structure is insensitive to the shape of the dispersal distribution. *Evolutionary Ecology Research*, **8**, 279–293.
- Legendre P (2000) Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, **67**, 37–73.
- Legendre P, Fortin M-J (in press) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, doi:10.1111/j.1755-0998.2010.02866.x.
- Legendre P, Dale MRT, Fortin M-J, Gurevitch J, Hohn M, Myers D (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, **25**, 601–615.
- Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.
- Malécot G (1948) *Les Mathématiques de l'Hérédité*. Masson, Paris.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.
- Manel S, Joost J, Epperson BK *et al.* (2010a) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.
- Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010b) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3835.
- McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

- Miller PS, Lacy RC (2005) *VORTEX. A stochastic simulation of the simulation process. Version 9.50 user's manual*. Conservation Breeding Specialist Group (IUCN/SSC), Apple Valley, MN, USA.
- Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography*, **31**, 685–697.
- Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone national park with landscape genetics. *Ecology*, **91**, 252–261.
- Nagylaki T (1986) Neutral models of geographical variation. In: *Stochastic Spatial Processes* (ed. Tauta P), pp. 216–237. Springer-Verlag, Berlin.
- Nagylaki T (1989) Gustave Malecot and the transition from classical to modern population genetics. *Genetics*, **122**, 253–268.
- Nelsen R (2006) *An Introduction to Copulas*. Springer-Verlag, New York.
- Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552–1553.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Noon BR, McKelvey KS (1992) Stability properties of the spotted owl metapopulation in southern California. In: *The California Spotted Owl: A Technical Assessment of its Current Status* (ed. Verner J), pp. 187–206. US Forest Service General Technical Report PSW-133.
- Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, **40**, 646–649.
- O'Neill RV, Krummel JR, Gardner RH *et al.* (1988) Indices of landscape pattern. *Landscape Ecology*, **1**, 153–162.
- Palmer MW (1992) The coexistence of species in fractal landscapes. *American Naturalist*, **139**, 375–397.
- Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.
- R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org>.
- Ragavendran A (2009) *Improving risk assessment of transgene invasion: assessing the role of uncertainty in predictions*. PhD thesis, Michigan State University.
- Railsback SF, Harvey BC (2002) Analysis of habitat-selection rules using an individual-based model. *Ecology*, **83**, 1817–1830.
- Real LA, Biek R (2007) Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *Journal of the Royal Society Interface*, **4**, 935–948.
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, **173**, 1033–1045.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.
- Saltelli A (2000) *Sensitivity Analysis*. John Wiley & Sons, Chichester.
- Schlötterer C (2003) Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 32–38.
- Schumaker NH (1996) Using landscape indices to predict habitat connectivity. *Ecology*, **77**, 1210–1225.
- Schumaker NH (2009) HexSim User's Guide. ONLINE. USEPA Western Ecology Division. Available from <http://www.epa.gov/wed/pages/models/hexsim/index.htm>.
- Schumaker NH, Ernst T, White D, Baker J, Haggerty P (2004) Projecting wildlife responses to alternative future landscapes in Oregon's Willamette Basin. *Ecological Applications*, **14**, 381–400.
- Scribner KT, Chesser RK (2001) Group-structured genetic models in analyses of the population and behavioral ecology of poikilothermic vertebrates. *Journal of Heredity*, **92**, 180–189.
- Scribner KT, Petersen MR, Fields RL *et al.* (2001) Sex-biased gene flow in spectacled eiders (Anatidae): inferences from molecular markers with contrasting modes of inheritance. *Evolution*, **55**, 2105–2115.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**, 1349–1368.
- Soboleva TK, Shorten PR, Pleasants AB, Rae AL (2003) Qualitative theory of the spread of a new gene into a resident population. *Ecological Modelling*, **163**, 33–44.
- Sokal RR, Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics*, **105**, 219–237.
- Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. *Biological Journal of the Linnean Society*, **60**, 73–93.
- Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Strand AE (2002) METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Molecular Ecology Notes*, **2**, 373–376.
- Sugg DW, Chesser RK (1994) Effective population sizes with multiple paternity. *Genetics*, **137**, 1147–1155.
- Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. *Trends in Ecology and Evolution*, **11**, 338–342.
- Taylor BL, Chivers SJ, Sexton S, Dizon AE (2000) Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation approaches. *Conservation Biology*, **14**, 1287–1297.
- Thompson CM, McGarigal K (2002) The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). *Landscape Ecology*, **17**, 569–586.

- Tracey JA (2006) Individual based modeling as a tool for conserving connectivity. In: *Connectivity Conservation: Maintaining Connections for Nature* (eds Crooks KR, Sanjayan M), pp. 343–368. Cambridge University Press, Cambridge, UK.
- Travis JMJ, Ezard THG (2006) Habitat geometry, population viscosity and the rate of genetic drift. *Ecological Informatics*, **1**, 153–161.
- Turner MG (1989) Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics*, **20**, 171–197.
- Turner MG (1995) Usefulness of spatially explicit population models in land management. *Ecological Applications*, **5**, 12–16.
- Turner ME, Stephens JC, Anderson WW (1982) Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. *Proceedings of the National Academy of Sciences, USA*, **79**, 203–207.
- Turner MG, Yegang W, Romme WH, Wallace LL (1993) A landscape simulation model of winter foraging by large ungulates. *Ecological Modelling*, **69**, 163–184.
- Urban D, Keitt T (2001) Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**, 1205–1218.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.
- Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, **174**, 2009–2020.
- Wiegand T, Moloney KA, Naves J, Knauer F (1999) Finding the missing link between landscape structure and population dynamics: a spatially explicit perspective. *American Naturalist*, **154**, 605–627.
- Womble W (1951) Differential systematics. *Science*, **28**, 315–322.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
- Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.
- Zollner PA, Lima SL (1999) Illumination and the perception of remote habitat patches by white-footed mice. *Animal Behaviour*, **58**, 489–500.