Landscape Ecol (2014) 29:639–653
DOI 10.1007/s10980-014-0011-5

RESEARCH ARTICLE

Contrasting the ability of data to make inferences regarding dispersal: case study of the Red-cockaded woodpecker (*Picoides borealis*)

Douglas J. Bruggeman · Thorsten Wiegand · Jeffrey R. Walters · Fernando González Taboada

Received: 5 April 2013 / Accepted: 28 February 2014 / Published online: 14 March 2014

Abstract Dispersal is a critical biological process that contributes to the persistence of species in complex and dynamic landscapes. However, little is known about the ability of different types of data to reveal how species interact with landscape patterns during dispersal. Further, application of process-based, landscape-scale models able to capture the influence of land use and climate change are limited by this lack of dispersal knowledge. Here we highlight a method for building such models when dispersal parameters are unknown, but information on the mating system and survival are available. We applied a common statistical framework, rooted in information theory, to contrast the ability of abundance, movement, and genetic data to estimate dispersal parameters for endangered Red-cockaded woodpecker (RCW), using an individual-based, spatially-explicit population model. Dispersal was modeled as a multifaceted process in which foray distance, long-distance dispersal, competition for mates, and landscape permeability were treated as uncertain. We found that movement data are three-times more powerful than abundance data collected at the same spatial and temporal scales. However, habitat occupancy data collected over much a shorter time scale but at regional spatial scales were very effective for estimating dispersal. We also found that one-year of abundance data provided a similar reduction in uncertainty as genetic differences among breeding groups estimated using a 24-year pedigree. Substituting population genetic data for movement and abundance data often led to the same parameter values, but not always. Our study highlights important differences in the information content of data commonly collected in the field.

Electronic supplementary material The online version of this article (doi:10.1007/s10980-014-0011-5) contains supplementary material, which is available to authorized users.

D. J. Bruggeman
Department of Fisheries & Wildlife, Michigan
State University, East Lansing, MI 48824, USA

D. J. Bruggeman (✉)
Ecological Services and Markets, 8 Bay St.,
Asheville, NC 28801, USA
e-mail: dougbrugg@gmail.com

T. Wiegand
Department of Ecological Modelling, UFZ Helmholtz
Centre for Environmental Research, Leipzig, Germany

J. R. Walters
Department of Biological Sciences,
Virginia Tech, Blacksburg,
VA 24061-0406, USA

F. G. Taboada
Departamento de Biología de Organismos y Sistemas,
Facultad de Biología, Universidad de Oviedo,
Oviedo 33071, Spain

Keywords Individually-based · Spatially-explicit population model · Inverse modeling · Monitoring programs · Bird banding data · Likelihood · Landscape genetics

Introduction

Dispersal is a critical biological process that facilitates the exploitation of new habitats, mating opportunities, and the exchange of genetic characters that may contribute to increased fitness (Clobert et al. 2001). Recent work has shed light on the many different interpretations of dispersal (Nathan et al. 2008, Lowe and Allendorf 2010). Dispersal is often inferred by analyzing patterns of abundance (Palmer et al. 2000), movement (Revilla et al. 2004), and genetic variation at broad geographic extents (Cushman and Lewis 2010). While inferring processes from observed patterns is common to all scientific disciplines, in ecology an erroneous description of the process, as multiple processes can lead to the same pattern (Grimm et al. 2005), could lead to misinformed conservation decisions.

Even if perfect data on dispersal could be obtained, they should be treated as case specific because dispersal is an inherently stochastic process (Nathan 2003). However, estimating the level of uncertainty that appropriately describes dispersal is difficult. Further, gaining an understanding of the consequences of using different data types to estimate this uncertainty is critical. Given the complexity of dispersal processes, it is reasonable to assume that data sources will differ in their ability to predict different components of dispersal. The contribution different data types make to understanding dispersal is largely unknown, because sites with multiple data sources are rare (Watts et al. 2007; Lowe and Allendorf 2010). Direct monitoring of dispersal through mark-recapture techniques or radio telemetry is common, but may be subject to bias due to small sample sizes (Finnegan et al. 2011) or limited spatial extent of study areas (Cooper et al. 2008). Population genetic techniques are often promoted as a more cost-effective approach for sampling large areas. The ability of population genetic data to provide indirect evidence of dispersal has long been debated (Nei 1986; Rousset 1997; Whitlock and McCauley 1999; Watts et al. 2007). In a few recent cases, dispersal was characterized using both radio telemetry and genetic data collected in the field, but these studies applied divergent statistical techniques suited to each data type (Cushman and Lewis 2010; Finnegan et al. 2011; Reding et al. 2013). Therefore, direct comparison of the ability of data types to estimate dispersal could not be made.

The nascent field of landscape genetics provides statistical tools to estimate the influence of landscape factors on dispersal from genetic characters (Sork and Waits 2010). These tools can be broadly classified as direct and indirect methods. Direct methods are useful for detecting recent dispersal events and use genotypes to assign individuals to likely regions of origin (Scribner et al. 2005). Indirect methods use differences in allele frequencies to estimate dispersal among population units (e.g., Fst, Wright 1951; Dst, Nei 1973; Bray-Curtis percent dissimilarity, Cushman et al. 2006). Statistical estimates of differences in allele frequencies are affected by both the processes of genetic drift and gene flow, which may be occurring heterogeneously across the landscape (Bruggeman et al. 2010). However, indirect methods require an assumption of migration-drift equilibrium (MDE) (Cushman et al. 2006; McRae 2006), in which rates of genetic drift are negligible or homogeneous over space such that genetic differences among population units reflect rates of dispersal. However, no method is available to estimate the strength of MDE without making specific assumptions regarding modes of dispersal (Rousset 1997; Hutchison and Templeton 1999).

Individual-based, and spatially-explicit population models (IB-SEPMs) provide a valuable tool for developing (Bruggeman et al. 2010) and evaluating (Lloyd et al. 2013) statistical methods in landscape genetics (Epperson et al. 2010). Most studies demonstrate how differences in allele frequencies among units in hypothetical populations arise from landscape features that impede movement over evolutionary time scales (Landguth et al. 2010; Jaquiéry et al. 2011; Lloyd et al. 2013). However, Cushman and Landguth (2010) developed an IB-SEPM of black bears that included empirical estimates of the permeability of landscape patterns to movement and concluded that genetic data, when analyzed with partial Mantel tests, have high power for identifying the correct dispersal model. In contrast, both Jaquiéry et al. (2011) and Lloyd et al. (2013) concluded that genetic data provide weak signals of the influence landscape patterns have on dispersal.

IB-SEPMs are very flexible tools that can best contribute to the development of ecological theory by forcing expectations of the models to confront data from real systems (Grimm et al. 2005). We adopt this approach in inverse modeling here and contrast the ability of alternative dispersal assumptions included in an IB-SEPM to replicate demographic, movement, and genetic data collected in the field. We use an IB-SEPM for the Federally endangered Red-cockaded woodpecker (RCW) (Letcher et al. 1998; Supplemental material, SM), allowing us to incorporate both a realistic mating system and movement mechanisms of actual species. If simulation studies ignore some dispersal mechanisms, then the patterns generated by these mechanisms will not be correctly identified (Nathan et al. 2008). Therefore, in addition to estimates of landscape permeability (Spear et al. 2010), we also allowed other dispersal mechanisms to vary, such as forays, long-distance movement (Kesler et al. 2010), and landscape influences on breeding opportunities. IB-SEPMs include both spatial and temporal heterogeneity, so no assumptions regarding MDE need to be made. Here we demonstrate how dispersal parameters can be estimated using traditional genetic statistics within an IB-SEPM but without assuming MDE. By doing so, we provide an explicit example of how patterns of drift can prevent inference of dispersal parameters.

Data describing movement (i.e., a connectivity matrix), abundance (i.e., habitat occupancy), and 24-year pedigree were collected from coastal North Carolina, USA (Fig. 1; SM ODD). Pedigree analysis was used to estimate Nei's minimum pairwise genetic distance among breeding groups (Dij; Nei 1973) and number of unique alleles within breeding groups (Ai) using a gene drop. Dij and Ai are commonly used genetic statistics and may differ in sensitivity to gene flow and genetic drift (Keyghobadi 2007). Because no tissues could be obtained for genetic analysis, results from the gene drop through the 24-year pedigree were used as "observed data".

To pursue our objective of contrasting the information content of different data types, we developed a common statistical framework. We estimated −log likelihood functions ($-\log[L\{S|O\}]$) to quantify the deviation between the observed data (O) and the data generated by each dispersal parameterization (S). Each dispersal parameterization was generated for inverse modeling by taking a systematic sample of the entire uncertain parameter space for each uncertain dispersal parameter applied to the IB-SEPM. We then estimated for each dispersal parameterization the resulting −log likelihood values of the different data types and determined the parameterizations that best approximated one or more types of observed data (Wiegand et al. 2004; Martinez et al. 2011; Hartig et al. 2011). This approach allowed us to learn how the uncertainty in each dispersal parameter is reduced when considering the information provided by the different types of observed data in our inverse approach.

To contrast information content of data types, we structured data analysis in three phases. The objective of the first phase was to identify the best attributes of genetic loci to include in the analysis by performing spatial autocorrelation analysis. The objective of the second phase was to perform a preliminary evaluation of information content among estimates of model fit (i.e., $-\log[L\{S|O\}]$-values) for movement, demographic, and genetic data using Spearman rank correlation. The objective of the third phase was to contrast the information content of data collected at different levels of biological complexity and spatial extent by testing five hypotheses (Table 3). Because multiple processes can produce the same patterns, $-\log[L\{S|O\}]$-values from multiple data types were included in hypothesis tests to remove parameterizations not able to closely approximate observed data. The hypotheses were tested by comparing the number of accepted parameterizations and the range of parameter values that result when different $-\log[L\{S|O\}]$-values were used to reduce the state space of parameters. For example, we test whether movement and abundance data collected over the same time period are equally useful for estimating dispersal parameters (Table 3). This hypothesis is tested by comparing the number of parameterizations accepted by filters A and B (Table 3).

Methods

Spatial autocorrelation analysis

A six-generation pedigree, rather than tissue samples, was used to derive genetic indices. While empirical genetic data reflecting the accumulated influence of natural history traits and landscape patterns over evolutionary time scales would be ideal, by using a

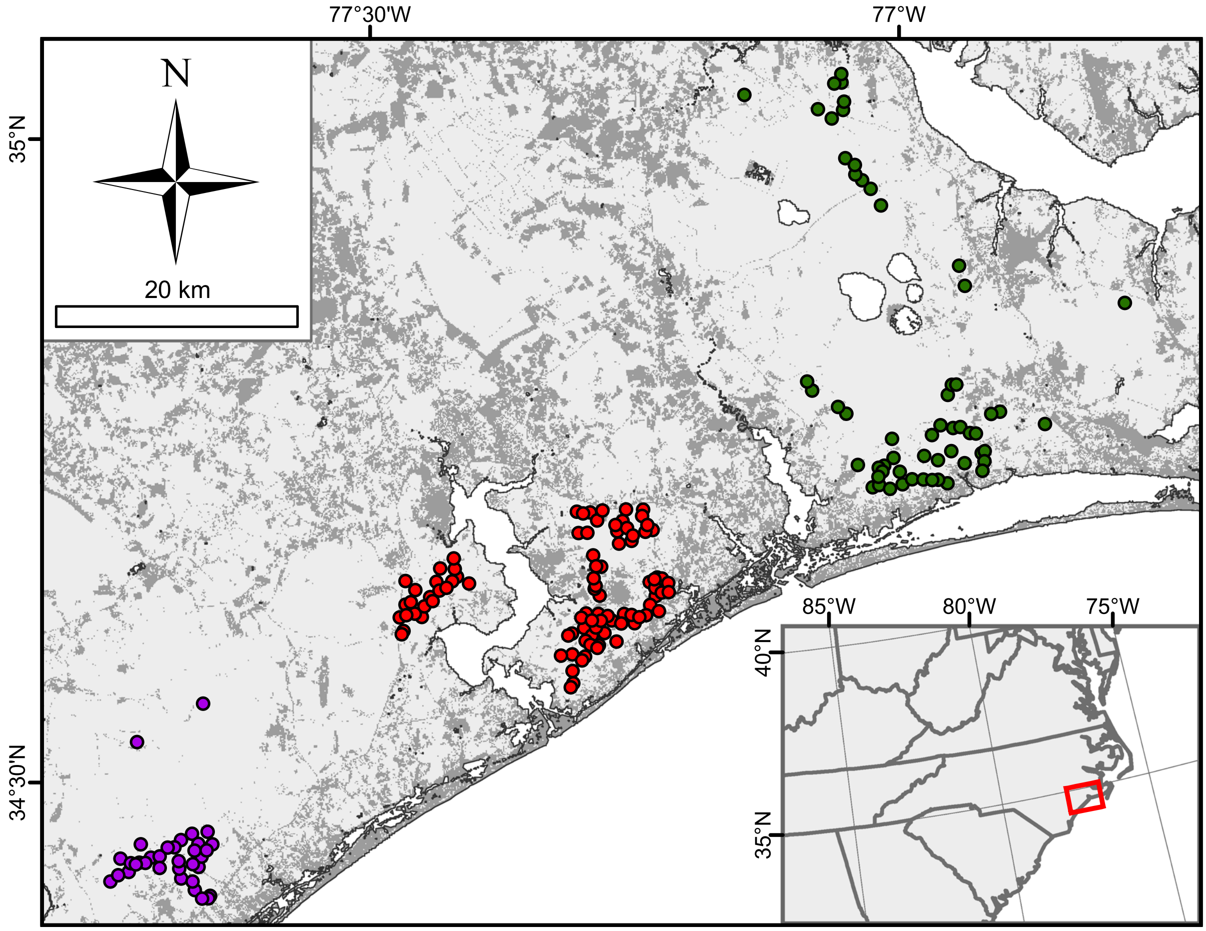


Fig. 1 Red-cockaded woodpecker territories in 2009 in Coastal North Carolina. Holly Shelter State Game Lands (HS, *purple*), Camp Lejeune (CL, *red*), and Croatan National Forest (CNF, *green*). Forested areas are *light gray*, non-forested gaps are *dark gray*, and water is *white*

pedigree to estimate genetic indices we were able to reduce the influence of historical dispersal and demographic events that would be reflected in empirical genetic data and cloud signatures of recent movements (Watts et al. 2007). Further, recent studies suggest that genetic differences due to landscape effects on dispersal may arise in as few as five generations (Cushman and Landguth 2010).

IB-SEPMs have recently been used to demonstrate that the more loci and greater number of alleles per locus increases the power of genetic analysis to detect the correct dispersal process over evolutionary time scales (Landguth et al. 2012). However, our study focuses on comparing the ability of data collected over ecological time scales, including demographic and movement data, to infer dispersal, and we are uncertain how many loci and how many alleles per loci should be included in the pedigree analysis. Past studies (Cushman and Landguth 2010; Landguth et al. 2012) have estimated statistical power based on the number of iterations of an IB-SEPM that report results in agreement with the underlying dispersal process coded into the simulation. In contrast, here, we estimate genetic summary statistics (Dij and Ai) within the simulation, and can report them per iteration or average them across iterations of a Monte Carlo simulation. When running simulations over shorter time-scales, we demonstrate how we can increase power of the analysis through the number of iterations included when averaging summary statistics.

To determine the sufficient number of loci and number of alleles that will allow us to detect the influence of spatial structure on movement when using a six-generation pedigree, we applied spatial autocorrelation analysis. We simulated 21 loci across a range of allelic richness. Locus 1 assumed an Infinite Alleles Model (IAM) of genetic variation—two unique alleles for each founder at the start of the simulation [i.e., total alleles = 2 × number of founders (SM ODD)]. Across the remaining 20 loci, four-levels of allelic richness were simulated to occur, each at five loci. The levels of allelic richness were set to provide five to two alleles/locus to emulate allelic richness observed at microsatellite loci for the species (Fike et al. 2009). This was achieved by sampling alleles from a normal distribution with standard deviations including 1, 0.5, 0.25, and 0.05. So, each of the four-levels of allelic richness were simulated to occur at five loci.

To avoid confounding correlations among alleles within a locus (Epperson 2004), Moran's I was calculated based on the frequency of the most common allele in a breeding group at each locus. Allele frequency data were estimated by a gene drop through a pedigree 1-, 200-, and 10,000-times from 1986 to 2009. The gene drop simulates the transmission of alleles from parents to offspring assuming Mendelian inheritance (i.e., their offspring had an equal probability of inheriting each of the two alleles). For loci simulating observed levels of allelic richness at microsatellite loci, we also averaged the frequency of the most common allele across the five independent loci that had the same level of allelic richness. Moran's I was calculated using Passage (Rosenberg and Anderson 2008) using 10 distance classes and a binary connection matrix.

IB-SEPM

Details of the RCW IB-SEPM can be found in SM following the Overview, Design concepts, and Details (ODD) protocol (SM ODD), which is intended to provide a standard approach for describing agent-based models across studies to increase transparency (Grimm et al. 2010). Briefly, the IB-SEPM was constructed based on a 24-year banding program for RCWs at Marine Corps Base Camp Lejeune (CL; Fig. 1). The RCW IB-SEPM simulates the cooperative breeding system of RCWs. Breeding groups consist of male and female breeders, fledglings, and, helpers who are usually male and full or half-sibs to the fledglings (Walters et al. 1988). Male helpers play a critical role in population dynamics by participating in the defense of the territories, feeding of nestlings, and inheriting their natal territory upon the death of the male breeder. Male helpers will preferentially inherit their natal territory upon the death of the breeding male, outcompeting floaters and helpers in adjacent territories. In contrast, we are more uncertain of the role female helpers play, but review of bird banding data indicate they do not inherit their mother‘s role as a breeder in their natal territory—so this was excluded from the model. Floaters of both sexes are also present in the region, which move continuously seeking a breeding vacancy in a territory (Walters et al. 1988).

The dispersal kernel was based on observations from another site so we allowed these parameters to vary (SM ODD). Alternative dispersal models were assembled by identifying ranges for seven parameter values used to simulate dispersal based on previous studies and expert opinion (Table 1). Thus, we did not need to implement different dispersal models to represent alternative hypotheses, but they were covered exclusively by the parameterization of our general dispersal model. Values from these ranges were sampled randomly with replacement using a Latin hypercube until 600,000 parameterizations were assembled. A systematic sample of the parameter space was required here because our aim was to compare the ability of different summary statistics to provide information regarding dispersal.

The population was simulated over an eight county area in coastal North Carolina covering an approximately one million hectare area, commonly referred to as the Onslow Bight Landscape (Fig. 1). Landscape classification in the Onslow Bight was performed by Dr. Aaron Moody's lab at University of North Carolina. FEMA Lidar data collected in 2001 and the National Land Cover Database were used to

Table 1 Model parameters estimated by inverse modeling for male and female individuals and their initial ranges

Parameter	Range
Foray distance of helpers	1–60 [cells]
Foray distance of floaters	1–60 [cells]
Seasonal dispersal speed	1–20 [cells]
Strength of terrestrial gap avoidance	0 to −0.005
Strength of water gap avoidance	0 to −0.005
Sensitivity to terrestrial gaps during competition	0 or 1
Sensitivity to water gaps during competition	0 or 1

Cell size = 1 ha, so 60 cells = 6 km. See SM ODD for parameter definitions identify four cover types including forested, open, water, and wetlands (SM ODD).

Observed data

Two sites that implemented independent monitoring programs provided data (Fig. 1). CL has maintained a bird-banding program since 1986 that provides a complete census and determination of reproductive success for the entire population each year (Walters et al. 1988). Holly Shelter State Game Lands (HS) monitors nesting success each year, but to include data comparable to CL we supplemented their efforts for one year to generate a spatially-explicit pattern of abundance. Therefore, the data set provides patterns that differed markedly in spatial and temporal scale and level of biological detail (Table 2).

Estimating model fit

We reduced the raw observed data into "summary statistics" that capture different aspects of the observed population dynamics (Martinez et al. 2011, Wiegand et al. 2004; Table 2). We then estimated the same summary statistics from the simulation by calculating the mean of multiple simulated runs to eliminate the internal model stochasticity, thereby treating the stochastic simulation model as deterministic on the level of the simulated summary statistics (Martinez et al. 2011). We estimate likelihood functions $L[S|O]$ to describe the deviation between the summary statistics of observed data (O) and data generated by each dispersal parameterization (S) (Csillery et al. 2010; Wood 2010). Summary statistics generated by the IB-SEPM were collected in Spring (SM ODD) for the same time intervals and territories present in the observed patterns (Table 2). Averages for all simulated data were taken across the 200 Monte Carlo iterations for each dispersal parameterization. For count data, including the number of Potential Breeding Groups (PBGs), group size, and connectivity, we assumed a Poisson error distribution. For example, the negative log likelihood ($-\log[L]$) for group size on CL was estimated as:

$$-\log(\{S|O\}) = \sum_{t=1997}^{2009} \sum_{i=1}^{106} -\log\left(\frac{S_{ti}^{O_{ti}} e^{-S_{ti}}}{O_{ti}!}\right)$$

where t = year and i = territory. The temporal and spatial extent of data included in the $-\log[L]$ estimates for connectivity, PBGs, and group size at other sites varied as described in Table 2.

Table 2 Description of observed summary statistics available from each landowner and their associated temporal and spatial scale

Observed pattern (abbreviation)	Camp Lejeune (CL)	Holly Shelter (HS)
Number of Potential Breeding Groups (PBG)	1997–2009	NA
Group Size (Group)	1997–2009 Territories 1–106	2009 Territories 1–34
Connectivity (Conn)[a]	1997–2009 Territories 1–106	NA
Minimum pair wise genetic distance (Dij)[b]	1986–2009 Territories 1–106	NA
Unique Alleles per breeding group—(Ai)	1986–2009 Territories 1–106	NA

NA not available

[a] Connectivity is summarized as a symmetrical matrix in which the values represent the number of birds born in territory "i" who became breeders in territory "j", summed over 13 years of field work (SM)

[b] Nei (1973)

For genetic data, in which the observed data were generated by a gene drop iterated 10,000 times, a Gaussian distribution was used.

$$-\log(L\{\mu^S|\mu^O,\sigma^O\}) = \sum_{i=1}^{106} \sum_{j=1,i\neq j}^{106} -\log\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\mu_{ij}^O-\mu_{ij}^S\right)}{2\sigma^2}}\right]$$

where μ^S equals the average Dij across 200 Monte Carlo iterations and μ^O and σ^O equal the mean and standard deviation for Dij from the 10,000 iterations of the observed gene drop. The same approach was applied to number of unique alleles per breeding group on CL, except it was summed across the vector describing unique alleles per territory rather than a territory by territory matrix. Both the observed and simulated genetic summary statistics assume that founders are unrelated, which removes any potential bias that would result from potential levels of relatedness among founders.

Table 3 Hypotheses tested using inverse modeling

Hypotheses	Filters compared					Result
H1 M vs. D: movement and demographic data collected over the same time period are equally useful for estimating dispersal parameters	A	CL Conn (1,861)	vs.	B	CL Group CL PBG (5,995)	Reject
H2 D long vs. D short: the temporal and spatial scale of demographic data does not affect estimates of dispersal parameters	B		vs.	C	HS Group (29,987)	Reject
H3 rD: short-term demographic data at a regional scale does not improve estimates of dispersal parameters	D	CL Group CL PBG CL Conn (119)	vs.	E	HS Group CL Group CL PBG CL Conn (1)	Reject
H4 Dij vs. D,M: substituting genetic distance for movement and demographic patterns does not lead to same dispersal parameters	E		vs.	F	CL Dij (30,005)	Accept
H5 Dij, Ai vs. D,M: adding unique alleles to genetic distance does not improve dispersal estimates	G	CL Dij CL A (13,892)	vs.	E, F		Reject

Hypotheses are tested by comparing the number of parameters accepted when different summary statistics are included in rejection filters at the 0.05 percentile (reported under the definition for each filter). For example, rejection filter C includes only summary statistic from HS describing group size (HS Group)

The goal is then to find parameterizations that minimize the deviation between the summary statistics of the observed and simulated patterns.

Hypothesis testing

In order to test the five hypotheses designed to contrast the information content of different summary statistics, we constrained the parameter state space, from the initial 600,000 dispersal parameterizations tested, by applying rejection "filters". By applying rejection filters, we require simulated dispersal parameters to closely approximate the observed summary statistics, as estimated by their associated $-\log[L]$-values. By including one or more summary statistics into a filter (Table 3), it provides a method for contrasting the ability of observed data to estimate dispersal parameters. For example, to contrast the ability of movement versus abundance data to estimate dispersal, parameters that results from applying filters A (i.e., requiring parameterizations to closely approximate the observed connectivity matrix on CL) versus B (i.e., requiring parameterizations to closely approximate the observed abundance data on CL) are compared (Table 3). To contrast the value of adding 1 year of abundance data from HS versus 13 years of abundance data from CL, filters B versus C are compared; etc.

Filters were applied in two ways. First, to contrast the power of summary statistics, we filtered parameterizations by selecting only those with $-\log[L]$-values within the 0.05 percentile. Summary statistics that accept many parameterizations show low information content (i.e., weak patterns; Wiegand et al. 2003) because they are only able to weakly constrain the dispersal parameter space. In contrast, summary statistics that reject many parameterizations show high information content (Wiegand et al. 2003). When more than one summary statistic is included within a rejection filter it may be harder to find parameterizations able to satisfy the 0.05 percentile for multiple summary statistics and the parameter space will be further reduced. The hypotheses were tested by comparing the number of accepted parameterizations that result when different restriction filters were applied. Second, we contrast differences in parameter values that result when different summary statistics are applied. This was performed by adjusting the

Table 4 Results for sensitivity to terrestrial and water gaps during competition, which was modeled as a binary variable

Filter Patterns	A CL Conn	B CL Group CL PBG	C HS Group	D CL Group CL PBG CL Conn	E HS Group CL Group CL PBG CL Conn	F CL Dij	G CL Dij CL A
Percentile	0.0037	0.0013	0.000015	0.021	0.0682	0.000015	0.00009
F Terrestrial	0	0	0.3	0	0	0.1	0.1
M Terrestrial	0	0.2	0.7	0	0.5	0.2	0.1
F Water	0.5	0.6	0.4	0.6	0.8	0.4	0.4
M Water	0.6	0.5	0.5	0.8	0.4	0.8	0.6

1 = presence of a gap prevents a competitor from winning a breeding vacancy; 0 = gaps do not affect competition. Results presented as the average across the ten parameterizations (i.e., 0.5 = gap sensitivity turned off in half of accepted models)

F female, *M* male, *CL* Camp Lejeune, *HS* Holly Shelter, *PBG* number of potential breeding groups, *Group* group size, *Conn* connectivity, *Dij* minimum pair wise genetic distance, *A* unique alleles per breeding group

percentiles applied to the $-\log[L]$-distributions for each filter (Table 4) to accept the ten best parameterizations.

Spearman rank correlations

First we estimated relationships among estimates of model fit $-\log[L]$-values) for all summary statistics. Correlations among estimates of model fit were calculated across the 600,000 dispersal parameterizations tested. Positive correlations tell us that estimates of model fit point to the same dispersal parameterizations and that there are no severe trade-offs which would point to errors in the process structure of the model. However, high correlations among estimates of model fit do not mean that different statistics are completely redundant; though they point to the same parameterizations/processes, they might differ in their ability to characterize uncertainty.

Next we examined the relationship between process uncertainty and the model's ability to approximate observations from the field. Therefore, we examined the relationship between parameter values and estimates of model fit for all 600,000 parameterizations. If the estimates of model fit and parameter value are not correlated then the parameter value has little influence on the ability of the model to approximate the observed pattern characterized by the summary statistic. Then we developed seven filters that remove parameterizations not able to closely approximate one or more patterns (Table 3). The filters were used to select the 200 parameterizations reporting the smallest negative log-likelihood values for one or more patterns. The correlation between parameter values and estimates of model fit may change across the application of different filters if the summary statistics differ in their ability to fit different aspects of dispersal.

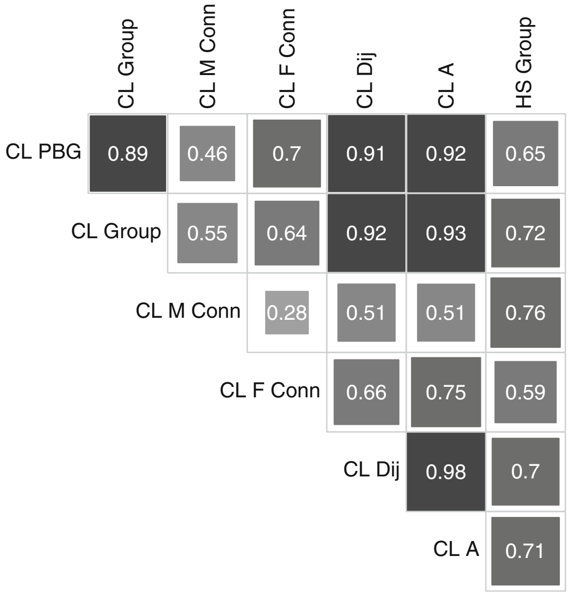
Results

Spatial autocorrelation analysis

When only one iteration of the Infinite Alleles Model (IAM) locus was used no spatial autocorrelation was detected (Table SM1). However, spatial autocorrelation resulted when 200 and 10,000 iterations were used. The Moran's I coefficients were very similar across these two treatments. The positive spatial autocorrelations were observed at distances >2.5 and ≤5.8 km, which corresponds to median distances of forays by juvenile females at another site (Kesler et al. 2010). At lag distances >12.9 km, negative autocorrelations were observed, indicating that allele frequencies were less similar than would be expected if they were randomly distributed—indicative of isolation by distance.

However, when locus 2 was evaluated no spatial autocorrelation was detected using 200 or 10,000 iterations. This is not surprising. We are sampling alleles from a normal distribution and then simulating mating over 6 generations so that we can contrast results with ecological data collected at short time scales. Therefore, after 6 generations gene correlations still reflect initial conditions. In contrast, under the IAM locus the simulation begins without any gene correlations. Assuming one-iteration, significant autocorrelations were detected with locus 2 but these reflect spurious correlations because it was not observed at the higher iterations.

When we averaged the frequency of the most common allele across five loci that sampled the same initial allele distribution as locus 2 independently, we find similar results to the one locus case. Therefore, when iterating 200 or 10,000 times, adding 4 more loci does little to increase power. However, compared to the one drop case adding 4 loci did remove the spurious correlation. We then evaluated whether averaging across five loci with lower levels of allelic richness assuming 10,000 iterations (i.e., loci 3–5, each simulated independently five times) would help, but found very similar results to locus 2. Therefore, these results highlight that when simulating over ecological time-scales (e.g., six generations) using one IAM-locus is able to detect realistic spatial autocorrelation but simulating 20 loci intended to emulate microsatellite loci do not. All subsequent analysis is only concerned with results from locus 1.



	CL Group	CL M Conn	CL F Conn	CL Dij	CL A	HS Group
CL PBG	0.89	0.46	0.7	0.91	0.92	0.65
CL Group		0.55	0.64	0.92	0.93	0.72
CL M Conn			0.28	0.51	0.51	0.76
CL F Conn				0.66	0.75	0.59
CL Dij					0.98	0.7
CL A						0.71

Fig. 2 Correlation matrix (Spearman rank coefficients) between $-\log[L]$-values that describe the goodness of fit for different summary statistics. Summary statistics available from Camp Lejeune (CL) include demographic patterns (group: breeding group size, and PBG: number of potential breeding groups), movement patterns (M Conn: male connectivity, and F Conn: female connectivity), and genetic patterns (Dij: average minimum pair wise genetic distance, and A: number of alleles in a breeding group). Data available from Holly Shelter (HS) included only breeding group size

Correlations among $-\log[L]$-values

We find strong relationships between estimates of model fit $-\log[L]$-values) for demographic and genetic data on CL, and weaker relationships between estimates of model fit for movement and genetic indices (Fig. 2). This suggests both genetic indices are more strongly associated with the process of genetic drift than migration, because parameterizations able to replicate demographic patterns were also successful at replicating genetic patterns. The lower rank correlations associated with patterns describing movement suggests connectivity provides complementary information relative to all other data sources at CL. It also suggests that there is often movement that does not lead to reproduction. We observed stronger relationships between estimates of model fit for female connectivity and genetic variation. Estimates of model fit for female connectivity were more strongly associated with patterns of allelic richness than genetic divergence. In other words, parameters able to closely approximate patterns of female connectivity were also able to approximate patterns of allelic richness. This suggests that allelic richness may be a stronger indicator of female movement than genetic divergence, which is perhaps not surprising because the mating system is characterized by female biased-dispersal that would result in new alleles being introduced to the breeding group thanks to longer distance female movement. At the regional level, group size in 2009 at HS was moderately correlated with data from CL, also indicating data complementarity.

Correlations among parameters and estimates of model fit

When all 600,000 parameterizations were evaluated for females, the strongest correlation was observed between floater foray distance and all summary statistics except male connectivity (Fig. SM1; see SM ODD for definitions of parameters). Therefore, female floater forays contribute critically to approximating abundance and the correct approximation of movement patterns. For males, male helper foray distance contributes more critically to approximating abundance on CL, but male floater foray distance contributes most critically to approximation of movement patterns. In contrast, male floater foray distance contributes more to approximating abundance on HS. Strength of gap avoidance during floating behaviors and gap sensitive competition were least important parameters for approximating observed patterns for both sexes.

When the best 200 parameterizations were selected the correlation coefficients often decreased relative to results from 600,000 parameterizations, because parameters contributing to a steeper slope were removed (i.e., the 200 best parameters tended to provide a more similar fit to observed data). When filter A was applied (i.e., only movement patterns on CL), the relationships between parameters and movements patterns was low, but the correlation among parameters and demographic and genetic summary statistics was higher. This indicates the 200 parameters resulting from filter A are doing a better job of approximating movement patterns but generating a poor fit for demographic and genetic patterns. This trend in which the application of summary statistics in filters reduces the correlation between summary statistics and parameters was observed across all seven filters (Fig. SM1).

Hypothesis testing

Movement versus demographic data (H1)

By applying rejection filters at the 0.05 percentile, we found that movement data were able to reduce uncertainty 3-times more than demographic data (filter A led to 1,861 and B to 5,995 parameterizations); thus, we reject H1 (M vs. D; Table 3). When percentiles were adjusted to select the 10 best parameterizations, results indicated that movement and demographic data collected at the same temporal- and spatial-scales would often lead to different dispersal estimates (Table 4; Fig. 3).

Under filter A, a considerable reduction in uncertainty, as indicated by small error bars, was observed for floater foray distance, dispersal speed, and terrestrial gap sensitive competition for both sexes (Table 4; Fig. 3). Therefore, a narrow range of values for these parameters is required to replicate connectivity patterns on CL. However, other parameters indicate a broader range of accepted values, indicating that either they are not critical for replicating movement patterns at the scale of CL or interactions with other parameters.

When demographic patterns were used to find the 10 best (filter B), uncertainty increased for dispersal speed and male floater foray distance, but it decreased for many parameters left uncertain by movement patterns, indicating that different processes contribute to replicating different patterns (Fig. SM1). For example, female helper foray distance is effectively turned off by filter B, but was left uncertain by filter A. Female helpers only occur rarely in the field, but this result indicates that allowing female helpers to compete for breeding vacancies, as was observed under filter A, increases error in the demographic trajectory. Indeed, a positive correlation between female helper foray distance and the summary statistic for the number of Potential Breeding Groups (PBGs) was observed under filter A (i.e., the longer the female helper dispersal distance the greater the values for $-\log[L]$ for PBGs on CL), however the relationship under filter B is weaker (Fig. SM1). The IB-SEPM assumes that female helpers have a higher survival rate than floaters, that competition favors older females, and middle-aged females produce more offspring (SM ODD). This shift in parameter values indicates that the processes associated with female helpers are likely still poorly understood, because permitting some female helpers to compete helps replicate movement patterns (filter A).

Similarly, strength of water gap avoidance for females shifts toward zero under filter B. However, under both filters A and C (i.e., only HS patterns) some avoidance of water gaps is indicated. Therefore, to replicate patterns of movement on CL and habitat occupancy on HS a reduction in vagility due to water helps (Fig. 1), otherwise too many birds from eastern CL colonize western CL. The relationship between strength of water gap avoidance for females and the summary statistic for female connectivity under filter B indicates increasing errors in the connectivity matrix on CL (Fig. SM1) as strength of water gap avoidance approaches zero.

Temporal and spatial scale of demographic data (H2)

We reject H2: D long versus D short because filter C leads to accepting far more parameterizations (29,987)

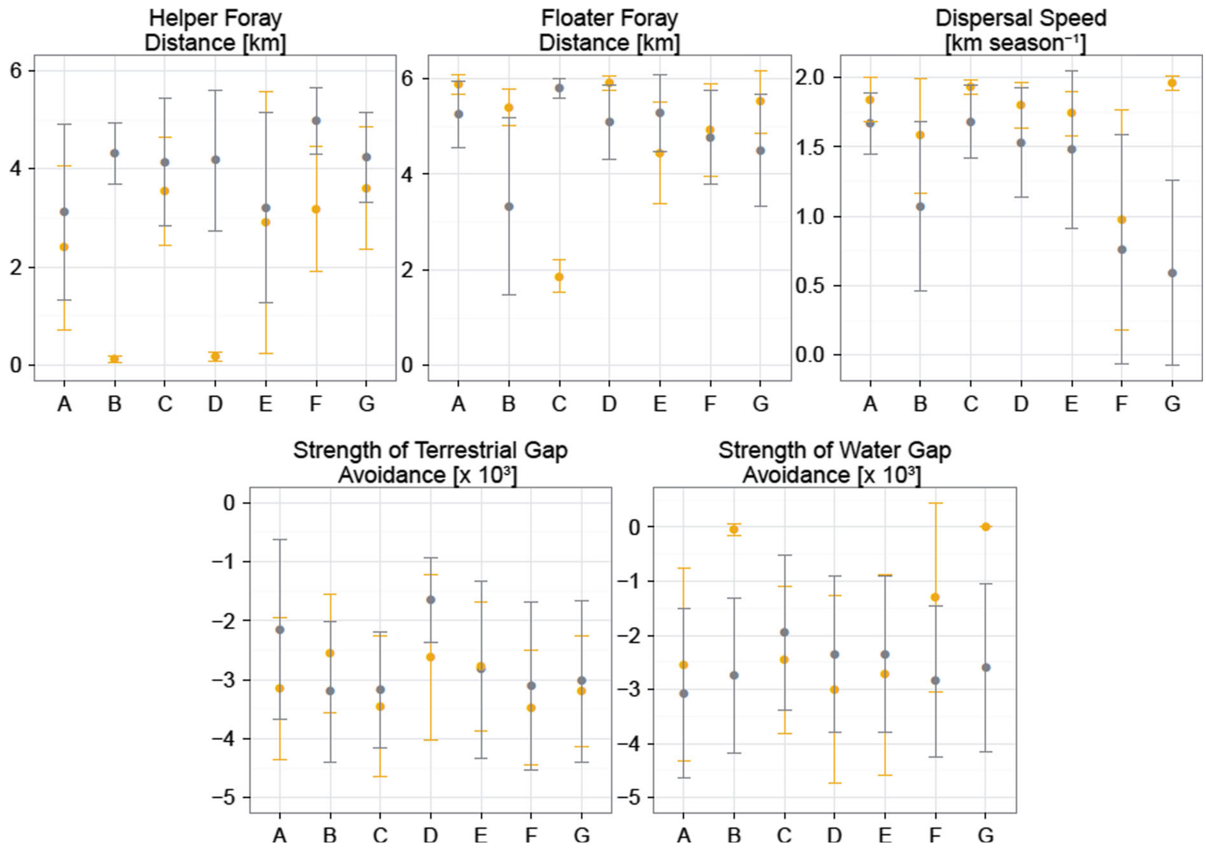


Fig. 3 Mean and 2 × standard error for the ten-best parameterizations that result when different rejection filters (Table 2) were applied. Five of the seven uncertain parameters are displayed, the other two are reported in Table 4. *Gray* males; *yellow* females

than filter B; therefore, the temporal- and spatial-scale of demographic data can have drastic effects on dispersal estimates. For example, female floater foray distance decreases under filter C indicating more restricted female forays help replicate patterns of habitat occupancy on HS, but male foray distances were similar to those observed under filter A (i.e., movement data only). Similarly, filter C indicates terrestrial gap sensitive competition is selected more often than observed in filters A and B, especially for males. Therefore, dispersal behaviors may be affected by local environmental differences not yet included in the IB-SEPM.

Regional, short-term demographic data versus local movement and demographic data (H3)

While 1-year of demographic data were useful on their own, by including HS Group summary statistic with movement and demographic summary statistics collected at CL, we were able to reduce the number of parameterizations from 119 to one using a rejection filter with a 0.05 percentile (filter D vs. E). Thus matching movement and demographic patterns simultaneously at regional scales was difficult. The inclusion of demographic data collected at regional spatial scales but over short time scales can improve dispersal estimates, indicating the infeasibility of hypothesis H3 rD (Table 3). When the percentile of filter E was adjusted to accept the 10 best parameterizations, we observe increased uncertainty in female helper and floater foray distance and male terrestrial gap sensitive competition compared to filter D. Therefore, to replicate patterns at regional scales a broader range of values are required for some parameters. Female dispersal speed, male floater foray distance, and female gap sensitive competition display consistent narrow ranges across filters A, C, D, and E; these parameters are critical for replicating observed patterns.

Genetic distance versus demographic and movement data (H4)

We found that substituting genetic distance patterns (Dij; filter F), derived from a 24-year pedigree, for movement and demographic data, using a 0.05 percentile led to a far greater number of accepted parameterizations (30,005) compared to filter E, but roughly the same number of parameterizations as resulted from only one-year of demographic data (29,987, filter C). Therefore, hypothesis H4 Dij versus D,M cannot be rejected (Table 3). Parameter values were similar under filter F or E when considering the best ten parameters sets in most cases, but not all. The biggest exception was dispersal speed for both sexes, which suggests that Dij may not be sensitive enough to detect dispersal distance when applied using inverse modeling. Also certainty regarding water gap sensitive competition for females decreased, indicating more uncertainty regarding sensitivity of females to water gaps. However, application of filter F moved median female strength of water gap avoidance estimate close to zero, but under filter E the median value remained close to the median of the original range tested (Fig. 3). This change in parameter values can best be understood by realizing that under filter E we were more certain that water prevented competition for breeding vacancies (Table 4), but under filter F this was found to be more uncertain. Therefore, there are tradeoffs in parameter values for female strength of water gap avoidance and water gap sensitive competition.

Combing genetic distance and unique alleles (H5)

Application of allelic richness patterns (Ai) and Dij (filter G) removed an additional 13,892 parameterizations compared to filter F when a rejection filter of 0.05 percentile was applied. Because application of filter G further reduced number of parameterizations at the 0.05 percentile compared to filter F we reject H5 (Dij, Ai vs. D,M). Therefore, in spite of the strong rank correlation between Dij and Ai (Fig. 2), allelic richness still provided complementary information. Examining the relationship between summary statistics and parameter values for Dij and Ai when all 600,000 parameterizations are considered (Fig. SM1), indicates that the source of this complementarity derives from stronger associations between Ai and the parameters female floater foray distance, dispersal speed, and terrestrial gap sensitive competition, and stronger associations between Dij and strength of terrestrial gap avoidance for males. When filter G was adjusted to find the ten best parameters, uncertainty in parameter values decreased for female floater perceptual distance, dispersal speed, and strength of water gap avoidance. Interestingly, even with the inclusion of both genetic patterns uncertainty regarding male dispersal speed did not decrease, which suggests genetic summary statistics may not be sensitive enough to detect rare longer distance movement of the philopatric sex. Application of filter G moved parameter estimates (Fig. 3) closer to those observed under filter E for floater foray distance for both sexes, male helper foray distance and female dispersal speed, These results suggest that by applying both Dij and Ai within rejection filters, we are able to select parameters able to replicate both patterns of genetic drift and gene flow observed across CL and improve parameter estimation.

Discussion

Inverse modeling techniques for IB-SEPMs based on summary statistics provide a useful hypothesis testing framework for including new processes and observed patterns to learn how life history traits respond to environmental heterogeneity. Three key components allowed us to address hypotheses in a rigorous manner. First, we initialized the model with as much observed data as possible which made comparisons to field observations possible (SM ODD). Second, stochastic components of reproduction, survival, and movement were allowed to interact. Third, information theoretical approaches allowed us to contrast the contribution different types of data make to revealing dispersal processes.

Movement data were the most powerful on their own, especially critical for estimating foray distance and dispersal speed. Abundance data on their own often led to likely erroneous parameter estimates, e.g., female helper foray distance and strength of water gap avoidance. Therefore, forcing dispersal parameters to only approximate abundance, without simultaneously controlling for aspects of movement, can introduce bias toward parameters that have stronger impacts on population growth. Including demographic data, from both sites, and movement data did not always decrease range of parameter values able to replicate data collected in the field. On the contrary, we demonstrated that under filter E uncertainty in parameter values often increased. This result highlights that dispersal is indeed a stochastic process and that mechanistic dispersal models able to generate results comparable to multiple types of empirical data are critical for understanding the underlying uncertainty associated with dispersal. Using genetic distance as a summary statistic led to parameter values that were similar to that observed under filter E, except for dispersal speed for both sexes and strength of water avoidance for females. Forcing parameterizations to also approximate patterns of allelic richness corrected the problem for female dispersal speed. Because Ai was slightly more correlated with PBGs than Dij, these results indicate that Ai and Dij provide complementary information useful for replicating patterns of genetic drift and gene flow respectively.

We demonstrated that basic census data are a more powerful predictor of dispersal than genetic data derived from a six-generation pedigree. However, we also found that when we adjusted rejection filters to accept the ten best parameterizations that genetic data can be very useful for estimating dispersal parameters in lieu of movement and demographic data; as one would expect based on population genetic theory. Whether applying one IAM locus to a six-generation pedigree provides a weaker or stronger signature of genetic drift and gene flow than can be obtained by microsatellite loci isolated from tissues collected in the field is uncertain. One may argue that microsatellite data capture a longer history of landscape influences on dispersal. But separating recent from historic influences of landscapes change using microsatellite data can be difficult (Balkenhol et al. 2009). Also, the greater number of alleles included in the IAM locus compared to even many microsatellite loci may create stronger signatures of drift and gene flow. Ultimately, we are unable to test this without tissue samples from the field. Granted, had genetic data been available at a bigger geographic extent, genetic data may have been more powerful (Anderson et al. 2010). We believe the application of our comparative statistical framework is critical given the increasing use of genetic data to infer dispersal (Sork and Waits 2010). The value of landscape genetic analysis can only be appreciated by contrasting the ability of genetic data to estimate landscape processes compared to other data types.

We used long-term movement, demographic, and genetic data of the RCW together with a detailed IB-SEPM to evaluate the ability of different data types to determine parameters and processes of dispersal. Though we only varied parameters related to movement, we show they can have strong effects on reproductive output, and therefore the model's ability to approximate demographic data. The ability of data to reduce uncertainty in ecological processes is limited by the capacity of the model to replicate the actual ecological processes occurring in nature. If model structure is poor representation of the actual process then the model will be unable to replicate observed data. Therefore, use of IB-SEPMs to estimate dispersal parameters should also play careful consideration to reproductive mechanisms. The comparative statistical framework described here allowed us to assess both parameter and structural uncertainty and will be used to inform future data collection and model development. The remaining uncertainty observed for many parameters indicates that either important processes still need to be included in the IB-SEPM or behaviors are variable, a conclusion supported by recent empirical data (Kesler et al. 2010).

Acknowledgments We are indebted to Marine Corps Base Camp Lejeune (W. Rogers, C. Ten Brink, G. Haught, and S. Cohen), Croatan National Forest (R. Powell, L. Thornhill, and G. Kauffman), Holly Shelter State Game Lands (T. Hughes, K Shugart, and D. Allen), and Steve Simon for help with data. We also thank K. Convery for collecting field data at Holly Shelter and I. Martinez for discussions and help with code. We thank four anonymous reviewers who contributed greatly to the presentation of this work. This research was supported by SERDP RC-1656. T. W. was supported by the ERC advanced grant 233066.

References

Anderson C, Epperson BK, Fortin MJ, Holderegger R, James P, Rosenberg MS, Scribner K, Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol Ecol 19:3565–3575

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH (2009) Identifying future research needs in landscape genetics: where to from here? Landscape Ecol 24:455–463

Bruggeman DJ, Wiegand T, Fernandez N (2010) The relative effects of habitat loss and fragmentation on population genetic structure of the Red-cockaded Woodpecker (*Picoides borealis*). Mol Ecol 19:3679–3691

Clobert J, Danchin E, Dhondt AA, Nichols JD (eds) (2001) Dispersal. Oxford University Press, USA

Cooper CB, Daniels SJ, Walters JR (2008) Can we improve estimates of juvenile dispersal and survival. Ecology 89:3349–3361

Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 25:410–418

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. Mol Ecol 19:3592–3602

Cushman SA, Lewis JS (2010) Movement behavior explains genetic differentiation in American black bears. Landscape Ecol 25:1613–1625

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Nat 168:486–499

Epperson BK (2004) Multilocus estimation of genetic structure within populations. Theor Popul Biol 65:227–237

Epperson BK, Mcrae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, James PMA, Murphy M, Manel S, Legendre P, Dale MRT (2010) Utility of computer simulations in landscape genetics. Mol Ecol 19:3549–3564

Fike JA, Athrey G, Bowman R, Leberg PL, Rhodes OE (2009) Development of twenty-five polymorphic microsatellite markers for the endangered red-cockaded woodpecker (*Picoides borealis*). Conserv Gen 10:1021–1023

Finnegan LA, Wilson PJ, Price GN, Lowe SJ, Patterson BR, Fortin M-J, Murray DL (2011) the complimentary role of genetic and ecological data in understanding population structure: a case study using moose (*Alces alces*). Eur J Wildl Res 58:415–423

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thuke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310:987–991

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF (2010) The ODD protocol: a review and first update. Ecol Mod 221:2760–2768

Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A (2011) Statistical inference for stochastic simulation models—theory and application. Ecol Lett 14:816–827

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898–1914

Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N (2011) Inferring landscape effects on dispersal from genetic distances: how far can we go? Mol Ecol 20:692–705

Kesler DC, Walters JR, Kappes JJ (2010) Social influences on dispersal and the fat-tailed dispersal distribution in red-cockaded woodpeckers. Behav Ecol 21:1337–1343

Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. Can J Zool 85:1049–1064

Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4179–4191

Landguth EL, Fedy BC, Oyler-McCance S, Garey AL, Emel SL, Mumma M, Wagner HH, Fortin MJ, Cushman SA (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Mol Ecol Res 12:276–284

Letcher BH, Priddy JA, Walters JR, Crowder LB (1998) An individual-based, spatially-explicit simulation model of the population dynamics of the endangered red-cockaded woodpecker, *Picoides borealis*. Biol Conserv 86:1–14

Lloyd MW, Campbell L, Neel MC (2013) The power to detect recent fragmentation events using genetic differentiation methods. PLoS ONE 8(5):e63981

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? Mol Ecol 19:3038–3051

Martınez I, Wiegand T, Camarero JJ, Batllori E, Gutiérrez E (2011) Disentangling the formation of contrasting tree-line physiognomies combining model selection and Bayesian parameterization for simulation models. Am Nat 177: E136–E152

McRae BH (2006) Isolation by resistance. Evolution 60(8): 1551–1561

Nathan R, Perry G, Cronin JT, Strand AE, Cain ML (2003) Methods for estimating long-distance dispersal. Oikos 103:261–273

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci USA 105:19052–19059

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M (1986) Definition and estimation of fixation indices. Evolution 40:643–645

Palmer M, Swan C, Nelson K, Silver P, Alvestad R (2000) Streambed landscapes: evidence that stream invertebrates respond to the type and spatial arrangement of patches. Landscape Ecol 15:563–576

Reding DM, Cushman SA, Gosselink TE, Clark WR (2013) Linking movement behavior and fine-scale genetic structure to model landscape connectivity for bobcats (*Lynx rufus*). Landscape Ecol 28:471–486

Revilla E, Wiegand T, Palomares F, Ferreras P, Delibes M (2004) Effects of matrix heterogeneity on animal dispersal: from individual behavior to metapopulation-level parameters. Am Nat 164:E130–E153

Rosenberg MS, Anderson CD (2008) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods Ecol Evol 2:229–232

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Scribner KT, Blanchong JA, Bruggeman DJ, Epperson BK, Lee C-Y, Pan Y-W, Shorey RI, Prince HH, Winterstein SR, Luukkonen DR (2005) Geographical genetics: conceptual foundations and empirical applications of spatial genetic data in wildlife management. J Wildl Manage 69:1434–1453

Sork VL, Waits L (2010) Contributions of landscape genetics—approaches, insights, and future potential. Mol Ecol 19: 3489–3495

Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol Ecol 19:3576–3591

Walters JR, Doerr PD, Carter JH III (1988) The cooperative breeding system of the red-cockaded woodpecker. Ethology 78:275–305

Watts PC, Rousset F, Saccheri IJ, Leblois R, Kemp SJ, Thompson DJ (2007) Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercurial*: Odonata: Zygoptera) populations: analysis of 'neighbourhood size' using a more precise estimator. Mol Ecol 16:737–751

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST $\neq$ 1/(4NM + 1). Heredity 82: 117–125

Wiegand T, Jeltsch F, Hanski I, Grimm V (2003) Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. Oikos 100:209–222

Wiegand T, Revilla E, Knauer F (2004) Dealing with uncertainty in spatially explicit population models. Biodivers Conserv 13:53–78

Wood SN (2010) Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466:1102–1104

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354