What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA

Alexandrine Daniel, Paul Savary, Jean-christophe Foltête, Gilles Vuidel, Bruno Faivre, Stéphane Garnier, Aurélie Khimoun

Received: 15 May 2024 | Revised: 10 September 2024 | Accepted: 12 September 2024

DOI: 10.1111/1755-0998.14024

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA

Alexandrine Daniel[1] | Paul Savary[2] | Jean-Christophe Foltête[3] | Gilles Vuidel[3] | Bruno Faivre[1] | Stéphane Garnier[1] | Aurélie Khimoun[1]

[1]Biogéosciences, UMR 6282 CNRS, Université de Bourgogne, Dijon, France

[2]Department of Biology, Concordia University, Montreal, Quebec, Canada

[3]ThéMA, UMR 6049 CNRS, Université Bourgogne-Franche-Comté, Besançon, France

Correspondence
Alexandrine Daniel, Biogéosciences, UMR 6282 CNRS, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France.
Email: alexandrine.daniel92@gmail.com

Funding information
Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation; the French Investissements d'Avenir program, project ISITE-BFC, Grant/Award Number: contractANR-15-IDEX-0003

Handling Editor: Kimberly J. Gilbert

Abstract

Modelling population connectivity is central to biodiversity conservation and often relies on resistance surfaces reflecting multi-generational gene flow. ResistanceGA (RGA) is a common optimization framework for parameterizing these surfaces by maximizing the fit between genetic distances and cost distances using maximum likelihood population effect models. As the reliability of this framework has rarely been studied, we investigated the conditions maximizing its accuracy for both prediction and interpretation of landscape features' permeability. We ran demo-genetic simulations in contrasted landscapes for species with distinct dispersal capacities and specialization levels, using corresponding reference cost scenarios. We then optimized resistance surfaces from the simulated genetic distances using RGA. First, we evaluated whether RGA identified the drivers of the genetic patterns, that is, distinguished Isolation-by-Resistance (IBR) patterns from either Isolation-by-Distance or patterns unrelated to ecological distances. We then assessed RGA predictive performance using a cross-validation method, and its ability to recover the reference cost scenarios shaping genetic structure in simulations. IBR patterns were well detected and genetic distances were predicted with great accuracy. This performance depended on the strength of the genetic structuring, sampling design and landscape structure. Matching the scale of the genetic pattern by focusing on population pairs connected through gene flow and limiting overfitting through cross-validation further enhanced inference reliability. Yet, the optimized cost values often departed from the reference values, making their interpretation and extrapolation potentially dubious. While demonstrating the value of RGA for predictive modelling, we call for caution and provide additional guidance for its optimal use.

KEYWORDS
gene flow, genetic simulation, landscape genetics, resistance optimization, ResistanceGA

Stéphane Garnier and Aurélie Khimoun contributed equally to this work.

1 | INTRODUCTION

Dispersal is defined as the movement of an individual from its native population to another location for breeding. This key ecological process directly affects evolutionary dynamics by moderating gene flow. The resulting movement of individuals and genes shapes species and genetic diversities and influences biotic interactions (Frankham, 2015; Richardson et al., 2016; Schlägel et al., 2020; Spielman et al., 2004). Considering dispersal is therefore critical in a context of human-driven habitat fragmentation and species range shifts (Crispo et al., 2006; Crooks et al., 2017; Manel & Holderegger, 2013). Accordingly, spatial models of connectivity (Table 1) have been crucial for identifying population connectivity drivers and for implementing sound conservation policies (Correa Ayram et al., 2016; Newmark et al., 2023; Rudnick et al., 2012).

Connectivity models often represent the landscape as a resistance surface (Table 1) embedding the dispersal propensity, physiological cost and mortality risk incurred by individuals across heterogeneous environments (Diniz et al., 2020; Zeller et al., 2012). As such, resistance surfaces commonly consist of a map of discrete landscape features, each associated with a corresponding resistance value (Table 1) (Spear et al., 2010, 2015). Connectivity models then translate these resistance assumptions and our understanding of dispersal into connectivity estimates for detecting landscape barriers, mapping corridors or computing connectivity metrics (Dutta et al., 2022; Foltête et al., 2014). The latter approach involves the calculation of cost distances, that is, the sum of resistance values along the least-cost path between two populations (Table 1). Hence, assigning a resistance value to each landscape feature is a critical modelling decision, with far-reaching consequences for the reliability of connectivity analyses.

Although many connectivity studies rely upon expert opinion to assign resistance values (Spear et al., 2010), several methods have been developed to infer them from empirical data such as genetic or animal movement data (Dutta et al., 2022; Peterman et al., 2019; Peterman & Pope, 2021; Vanhove & Launey, 2023; Zeller et al., 2012). In this context, the ResistanceGA (RGA) framework has received a great deal of interest (Peterman, 2018). This resistance surface optimization method is based on genetic data and assumes that genetic distances reflect gene flow, and consequently landscape functional connectivity (Zeller et al., 2017). Increasingly used since its release in 2018 (114 publications using it to this date, see S1.1: Data S1) (Antunes et al., 2023;

TABLE 1 Definitions of terms used in this paper.

Term	Definition
Resistance value	Numerical value associated with a landscape feature representing the movement propensity, physiological cost, and mortality risk incurred by individuals dispersing across this feature
Cost scenario	A list of mapped landscape features in conjunction with their respective resistance values
Resistance surface	Landscape map that makes spatially explicit a cost scenario, *i.e.*, a resistance assumption about animal dispersal movements across heterogeneous environments
Cost-distance	The sum of cost values along the least-cost path between two populations, calculated on a resistance surface
Functional connectivity	Inversely related to landscape matrix resistance, the functional connectivity of a habitat patch could be seen as the amount of reachable habitat from that patch. From an individual's perspective, this is the resistance of the surrounding landscape matrix
Pruning	In graph theory, the selection of a limited number of connections among graph nodes. In contrast to a pruned graph, a complete graph has all its nodes connected to each other
Effective-dispersal-scale	At this spatial scale, the links between habitat patches are only modelled if migrants have actually moved along them. Therefore, the effective dispersal scale dataset included population pairs that were expected to exchange migrants
Sampling-scale	Spatial scale that corresponds to the scale of the sampling design. At this scale, all the links between habitat patches are modelled, leading to a complete landscape graph
Predictive modelling	A model that attempts to predict an unobserved pattern or process by analysing data from an observed pattern. As opposed to the explanatory model, which consists of identifying the variables that explain part of the variance of an observed process, to improve our understanding of that process
Predictive performance	Ability of a model to predict data not considered for its calibration (*e.g.*, out-of-bag or validation data), as evaluated using a K-fold cross-validation or a leave-one-out cross-validation method and quantified by an indicator such as a validation R^2
Transferability	The relevance of resistance values inferred by the ResistanceGA workflow on a given landscape and for a given set of populations to characterize movement resistance among new populations, in the same (*interpolation*) or in another (*extrapolation*) landscape
Interpolation	Use of resistance values inferred by ResistanceGA from a set of populations in a given landscape to predict genetic distances among a new set of populations in the same landscape
Extrapolation	Use of resistance values inferred by ResistanceGA in a given landscape to predict genetic distances among populations in another landscape
Training and test datasets	Terms used in cross-validation methods to designate the part of the dataset used to parameterize the model (training, in-bag or calibration data set) and the part used to evaluate the predictive quality of this parameterized model (test, validation or out-of-bag data set)

Atzeni et al., 2023; McCluskey et al., 2022; Zeller et al., 2023), it infers dispersal costs by maximizing the statistical fit between genetic and cost distances. Although not without computational constraints, its power stems from the use of a genetic optimization algorithm, which efficiently explores the whole cost parameter space.

Yet, the validity of the inferences made with this flexible optimization tool remains understudied (but see Beninde et al., 2023; Peterman et al., 2019; Winiarski et al., 2020). For instance, the encouraging results of Peterman et al. (2019) and Winiarski et al. (2020), based on the correlations between true and RGA optimized resistance surfaces, called for studies relying on direct simulations of genetic processes. Later, the study of Beninde et al. (2023) explored a broad range of genetic distance metrics and resistance scenarios and showed that RGA better recovered the true resistance surface when few landscape features restricted gene flow. However, these works have not explicitly examined the influence of the ecological profile and degree of habitat specialization of the species under study. For instance, we can reasonably assume that the interaction between dispersal capacity and the level of habitat specialization determines the spatial scale at which populations are linked by dispersal events and, hence, our ability to infer dispersal drivers based on genetic data. Although this has not yet been done in the context of resistance optimization, accounting for this spatial scale by removing from the analyses pairwise genetic distances mainly driven by stochastic genetic drift could improve the reliability of inferences (Savary et al., 2021a).

More generally, Beninde et al. (2023), Winiarski et al. (2020) and Peterman et al. (2019) called for a more critical look on the resistance surfaces optimized with RGA from empirical data. Of particular concern is the overfitting effect inherent to the optimization process, which may impede the accurate assessment of the relative resistance values of each landscape feature. On the one hand, unsupervised data-driven approaches are prone to overfitting when several parameter sets can result in the same pattern (Paris et al., 2004). Nonetheless, very few studies have tackled this issue (Palm et al., 2023; Pless et al., 2021) by testing whether bootstrapping or cross-validation could limit overfitting. On the other hand, if the data-driven nature of the stochastic optimization algorithm (Scrucca, 2013) does not necessarily make it ideal for understanding the processes at play, this makes it perfectly suited to predictive modelling (Table 1), learning from observed patterns to predict unobserved patterns or processes (Shmueli, 2010). This would answer the call for landscape genetics research to contribute to the prediction of genetic responses to landscape changes (Balkenhol et al., 2009; Storfer et al., 2007; Van Strien et al., 2014).

To provide evidence-based guidelines regarding the use and misuse of optimization frameworks in landscape genetics, we evaluated the performance of RGA across a wide range of realistic landscape structures, ecological profiles, sampling designs and spatial scales. For that purpose, we simulated gene flow using categorical resistance surfaces reflecting the level of habitat specialization and dispersal capacities of distinct virtual species. We then optimized cost distances among populations from these surfaces based on the simulated genetic distances using the RGA algorithm. Using graph-based methods, we performed these optimizations at both the spatial scale of the whole sampling area and at the scale of gene flow effect on genetic differentiation (Balkenhol et al., 2020; Savary et al., 2021a; Van Strien, 2017).

Our simulation framework aimed first at assessing whether the optimization approach correctly detects that ecological distances drive genetic differentiation. Second, we assessed the ability of the optimized cost distances to predict genetic differentiation using a cross-validation method to prevent overfitting (Daniel et al., 2023). Finally, we measured the congruence between the optimized costs and the true cost values shaping the simulated genetic patterns.

2 | MATERIALS AND METHODS

2.1 | Overall approach

To evaluate the RGA optimization framework, we performed demo-genetic simulations in real landscapes for virtual species varying in their dispersal capacities (Figure 1). We then used RGA and assessed its ability to recover the 'true' drivers of genetic differentiation (landscape resistance and corresponding resistance values) and predict the resulting pairwise genetic differentiation.

2.2 | Real landscape sampling

We selected 30 real landscapes (40×40 km, 100 m cell resolution) in metropolitan France, maximizing variations in the amount and spatial configuration of the four land cover types considered (forest, grassland, agricultural and urban areas). We used the 2018 Theia OSO landcover map (Inglada et al., 2018). See S1.2: Data S1 for more details about landscape sampling.

2.3 | True resistance surfaces parameterization and cost-distances calculation

We designed two cost scenarios (Table 1) characterizing the resistance of each land cover type to dispersal, scaling the contrast in resistance values with a negative exponential function, following Keeley et al. (2016) (See S1.3: Data S1 for details). The resulting cost scenarios represented two contrasted ecological profiles: (i) a forest-specialist species experiencing high resistance across all land cover types but forest: {$Cost_{forest}=1$; $Cost_{grassland}=700$; $Cost_{agricultural}=900$; $Cost_{urban}=1000$}; (ii) a generalist species with a smoother range of resistance values: {$Cost_{forest}=1$; $Cost_{grassland}=50$; $Cost_{agricultural}=200$; $Cost_{urban}=1000$}. These true cost scenarios served as a proxy for species specialization levels.

In each of the 30 landscapes, we located 80 virtual populations by randomly sampling 80 forest pixels more than 500 m apart, at least 2 km from the edge of the raster, in forest patches of more than 25 ha. Then, we computed the least-cost paths and corresponding cost distances among populations on each true resistance surface, using the graph4lg package (Savary et al., 2021b).

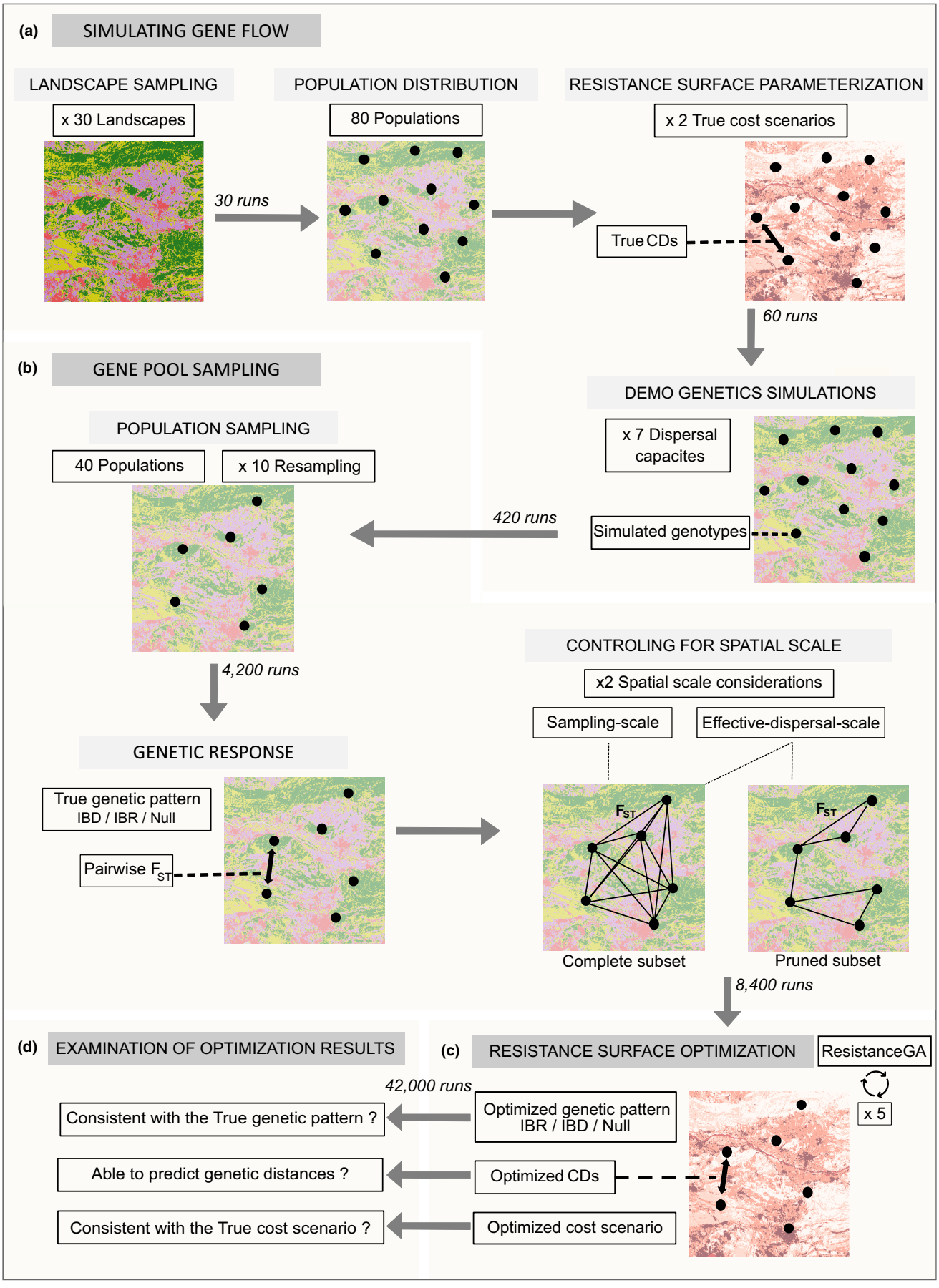


FIGURE 1 Overall methodology for assessing the ability of optimized cost distances to reliably reflect landscape effects on genetic structure. (a) Workflow for simulating gene flow among virtual populations on real landscapes, depending on two scenarios of movement costs and seven dispersal capacities. (b) Preparation of genetic inputs for the optimization process by sampling 10 times 40 populations and evaluating their related pairwise genetic distance. Two genetic distance datasets were considered depending on spatial scale. (c) Optimization of cost distances according to these genetic inputs using the RGA workflow. (d) Evaluation of RGA inferences according to the sensitivity of the algorithm to genetic pattern, the predictive power of optimized cost distances and the ability to recover the true cost scenario. CDs, cost distances; IBD, Isolation-by-Distance; IBR, Isolation-by-Resistance; Null, no spatial structure.

2.4 | Demo-genetics simulations and gene pool sampling

We used the PopGenReport package (Adamack & Gruber, 2014) to simulate gene flow and resulting individual genotypes. Dispersal cost was conditioned by the two 'true' cost scenarios. To test for the effect of dispersal capacities, we considered seven different maximum dispersal distances. Local population sizes remained constantly equal to 25 individuals throughout simulations. Further details about simulation parameters are given in S1.4: Data S1. We ran a total of 420 simulations (30 landscapes × 2 levels of specialization × 7 dispersal capacities) over 200 generations each to obtain a steady pattern of genetic differentiation. Using individual genotypes, we estimated pairwise genetic distances among populations with pairwise F_{ST} (Weir & Cockerham, 1984).

To test for the effect of sampling design on RGA performance, we randomly sampled 40 of the 80 populations in each landscape (10 sampling iterations) and extracted the corresponding F_{ST} matrices between the sampled populations. Every combination of a landscape, a specialization level, a dispersal capacity level and a sampling design (4200 combinations) will be referred to as a 'run' hereafter.

2.5 | Spatial scale of landscape influence on genetic structure

We tested whether matching the spatial scale at which dispersal influences genetic differentiation with that of the population pairs considered in the analyses would improve inferences (Savary et al., 2021a; Van Strien, 2017). We considered two types of pairwise matrices when modelling genetic distances, differing in the spatial scale of the connections considered. The 'sampling-scale' dataset included all population pairs (Table 1). In these matrices, some pairwise genetic distances may mostly reflect genetic drift rather than dispersal, especially for short-distance dispersers. Thus, for the 'effective-dispersal-scale' dataset (Table 1), we excluded from each pairwise matrix (run in our analysis) pairwise genetic distances mostly driven by drift effects. To prune (Table 1) these matrices and conserve population pairs whose differentiation reflects dispersal influence, we relied on a graph-theoretical method. Based on the conditional independence principle, this method is supposed to identify links between populations directly exchanging migrants (Dyer & Nason, 2004; see S1.5: Data S1 for more details).

2.6 | RGA workflow and new implementations

The RGA algorithm optimizes resistance surfaces from genetic distances using a genetic algorithm, which efficiently explores the parameter space until maximizing the statistical fit between pairwise cost distances and genetic distances. Each of the 60 resistance surfaces (30 landscapes × 2 levels of specialization) was optimized based on 140 F_{ST} matrices (7 dispersal capacities × 10 samplings × 2 spatial scales; Figure 1). Each optimization was replicated five times to assess the stability of the algorithm inferences (42,000 runs in total).

The optimization seeks to maximize an objective criterion measuring the fit of maximum likelihood population effect (MLPE) models. These models of pairwise genetic distances as a function of pairwise cost distances account for the non-independence inherent to pairwise data (see S1.6: Data S1 for more details on RGA workflow). To prevent model optimization from overfitting the data, we implemented a new objective criterion within the RGA workflow. We adapted the Leave One Out Cross-Validation (LOOCV) method to the pairwise context and computed a validation R^2 to quantify the prediction error. When fitting the models, we iteratively removed one of the 40 populations, and used the calibrated model to predict the genetic distances involving this population. The mean of the predicted 'out-of-sample' genetic distances was compared with the observed genetic distances to assess predictive accuracy, following Daniel et al. (2023). To implement this approach, we adapted the RGA algorithm (code available online: https://gitlab.com/psavary3/rga).

2.7 | Reliability of cost inferences from RGA optimization

To assess the reliability of the cost inferences made with RGA, we carried out a multi-criteria analysis, described by Figure 2 and the following sections.

2.7.1 | RGA optimization sensitivity and specificity to simulated genetic patterns

First, we checked whether the RGA optimization identified the correct drivers of genetic differentiation, that is, ecological distances in our simulations (leading to an 'Isolation-by-Resistance' pattern). For that purpose, we compared the AIC values deriving from the

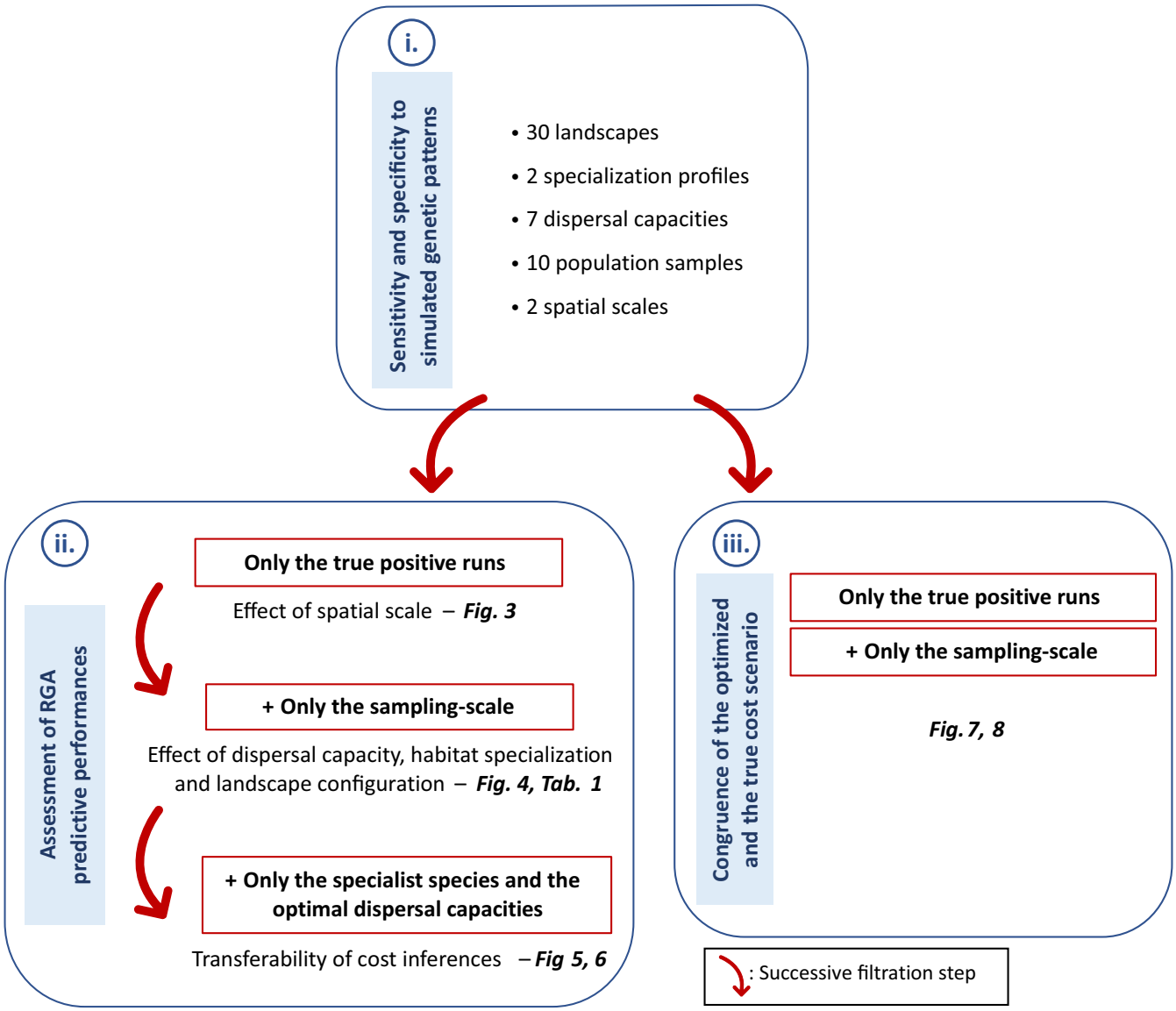


FIGURE 2 Diagram summarizing the successive stages of data filtering throughout the analyses. The filtration rules are shown in the red boxes. The arrows indicate the successive filtration rules.

MLPE models explaining genetic distances with either true or optimized cost distances ('Cost distances model'), Euclidean distances ('Euclidean model') or a constant ('Null model'). We distinguished three possible outcomes indicating which pattern is detected:

(i) Isolation-by-Resistance (IBR) when:
$AIC_{Euclidean\ model} - AIC_{Cost\ distances\ model} > 2$
(ii) Isolation-by-Distance (IBD) when:
$AIC_{Euclidean\ model} - AIC_{Cost\ distances\ model} < 2$ and $AIC_{null\ model} - AIC_{Euclidean\ model} > 2$
(iii) Null pattern when:
$AIC_{Cost\ distances\ model} - AIC_{null\ model} > -2$ and $AIC_{Euclidean\ model} - AIC_{null\ model} > -2$

Consequently, we classified the optimization runs into four categories: true positives (TP: IBR detected from both true and optimized cost distances), true negatives (TN: IBR never detected), false positives (FP: IBR detected from optimized but not from true cost distances) and false negatives (FN: IBR detected from true but not from optimized cost distances; see S10.a: Data S1). We then computed the sensitivity (TP/(TP + FN)) and the specificity (TN/(TN + FP)) indices for each spatial scale. These indices allowed us to assess whether the type of pairwise connections considered (sampling vs. effective-dispersal subset) significantly affected inferences (see S10.b: Data S1). Then, we modelled the sensitivity or specificity indices as a function of habitat specialization (2 levels), dispersal capacities and landscape structure (9 FRAGSTAT metrics). We used the glmmTMB package (Brooks et al., 2017) to fit generalized linear mixed models with beta distribution (beta GLMM). The focal landscape was set as a random effect (30 levels, intercept-only) to control for the non-independence of the data. We checked for variable collinearity and performed a

stepwise model selection based on AIC criteria, using the R package Buildmer (Voeten, 2020).

In the following analyses, we only focused on the true-positive runs (Figure 2), and, hereafter, all the beta GLMMs were run using the same mixed-effect model specification and selection.

2.7.2 | Assessment of RGA predictive performance

Effect of spatial scale on predictive performance

First, we tested the effect of restricting pairwise connections to the effective-dispersal-scale on the predictive performance of optimized cost distances. For each run, for comparative purposes, we computed the validation R^2 measuring predictive accuracy by considering only the population pairs shared by both the sampling and the effective-dispersal-scale datasets. We then ran beta GLMM explaining validation R^2 by a binary categorical variable distinguishing sampling (complete) and effective-dispersal (pruned) datasets.

Effect of dispersal capacity, habitat specialization and landscape configuration on predictive performance

Second, we compared the predictive accuracy (validation R^2) of optimized cost distances across dispersal capacities, specialization levels and landscape structure variations, using a beta GLMM. The effective-dispersal-scale dataset included runs with varying numbers of populations. Therefore, for the sake of reliable comparison, we only considered for this analysis and the following the True Positives runs for the sampling-scale dataset, as they shared the same number of population pairs (Figure 2).

Transferability of cost inferences

For assessing predictive power, we focused on specialist species and considered only the dispersal distance that maximized the fit between genetic and cost distances (Figure 2). This reduced the number of parameters to consider and reflected the conditions optimizing RGA performance. To assess the transferability of cost inferences, we optimized a cost scenario on a landscape and we evaluated its predictive power for genetic distances between the 40 remaining out-of-sample populations, assessing its interpolation ability. To assess the extrapolation ability of the optimized scenario, we assessed the predictions of genetic differentiation based on cost-distances computed under this scenario for 40 out-of-sample populations from another randomly selected landscape. We then fitted beta GLMM explaining validation R^2 with a binary variable representing whether the costs were inferred in the same landscape (interpolation) or in a different one (extrapolation).

Finally, we compared the predictions based on optimized cost distances to those based on true cost distances. For each landscape, we fitted a beta GLMM explaining the validation R^2 with a categorical variable distinguishing the true and optimized scenarios.

2.7.3 | Congruence between optimized and true cost scenarios

We tested whether the optimized cost values corresponded with the true ones, and whether they correctly identified landscape barriers and permeable landscape features. We considered both the true specialist and generalist cost scenarios as references for comparison (cf. section II.). Additionally, to assess the sensitivity of the algorithm to variations in land cover resistance rankings, we considered two other scenarios ranking the urban, agricultural and grassland resistances in reverse order compared to the specialist scenario: (1) $\{Cost_{forest}=1;\ Cost_{grassland}=900;\ Cost_{agricultural}=700;\ Cost_{urban}=1000\}$; (2) $\{Cost_{forest}=1;\ Cost_{grassland}=1000;\ Cost_{agricultural}=900;\ Cost_{urban}=700\}$.

We then assessed the proportion of 1200 optimization runs (30 landscapes × 10 samplings × 4 scenarios) assigning each land cover to each resistance rank, for simulations performed with the dispersal capacities maximizing the fit between genetic and cost distances, using sampling-scale datasets. Considering the true scenarios, we would ideally expect 100% of the runs assigning rank 1, 2, 3 and 4 to forest, grassland, agricultural and urban areas, respectively.

We also quantified the similarity between the true and optimized scenarios by calculating the Spearman correlations of land cover ranks between these two scenarios (see S1.12: Data S1 for the correspondence table between recovery index and land cover ranking). We assessed the contrast between the maximum and minimum costs for each optimized scenario as cost.ratio $= \log_{10}\left(\frac{\text{opt.costmax}}{\text{opt.costmin}}\right)$ and compared it to that of the true scenarios $\left(\log_{10}\left(\frac{1000}{1}\right)=3\right)$ by the difference 3 − cost.ratio. The closer this difference to zero, the better the recovery of the true contrast. The correlations of land cover rankings and the contrast difference were referred to as recovery indices.

We ran beta GLMM and Linear Mixed Models (LMM) to test for the influence of specialisation level, dispersal capacities and landscape structure (9 FRAGSTAT metrics) on these recovery indices. When modelling the contrast, we used the lmer and MuMin R packages for the LMM and model selection, respectively (Bartoń, 2013). Finally, we tested whether the recovery indices affected the predictive quality (validation R^2) of the optimized cost scenario, running a beta GLMM.

3 | RESULTS

3.1 | Landscape sampling and demo-genetic simulations

124 out of the 1000 sampled landscapes met land cover proportion criteria, and we sampled 30 of them along the two first PCA axes (accounting for 61% of the variance, see S1.7: Data S1 for more information).

After 200 generations of simulation, the detected genetic patterns depended on the interaction between species dispersal capacities, their specialization profile and the spatial scale of the analysis (S1.8: Data S1). For the generalist species, the detected genetic structure was similar when analysing data at the sampling- and effective-dispersal-scales. For a dispersal capacity of 1000 cost units (cu), we mostly detected an IBR pattern (sampling-scale: 83%, effective-dispersal-scale: 77%). Beyond this dispersal distance, a null pattern was detected for 82% of the simulations. For the specialist species and the sampling-scale dataset, 47% of the simulations resulted in a pattern of IBD at a small dispersal capacity (1000 cu). Simulations mainly resulted in an IBR pattern up to an optimal scale of 20,000 cu before gradually shifting towards a null pattern beyond that scale (S1.8: Data S1). In certain regions of the parameter space (specialist species with low dispersal capacities in particular), a bimodal distribution of F_{ST} values resulted in the identification of spurious IBD patterns (see S1.9: Data S1 for more details). These spurious patterns explained by a predominant effect of genetic drift on genetic differentiation at the sampling-spatial scale tended to be replaced by a null pattern in the effective-dispersal-scale dataset (only 7% of IBD patterns at 1000 cu; S1.8.C: Data S1).

3.2 | Optimization sensitivity and specificity to simulated genetic patterns

The sensitivity (i.e., TP/(TP+FN)) of RGA to the simulated genetic pattern (IBR, IBD, or null pattern) was 92% for the effective-dispersal-scale dataset and 90% for the sampling-scale dataset (S1.10.a: Data S1). A chi-squared test (see S1.10.b: Data S1) showed that the lower sensitivity value of the sampling-scale dataset comes from a slightly higher number of false negatives when using the sampling-scale dataset and indicates that pruning the distance matrices improved the identification of the main driver of genetic patterns ($X^2 = 5.3$, df = 1, $p = .02$).

The specificity indices (i.e., TN/(TN+FP)) were 88% and 92% at the effective-dispersal- and sampling-scales, respectively (see S1.10.a: Data S1). The better performance of the sampling scale dataset stems from a higher number of false positives at the effective-dispersal-scale ($X^2 = 16.11$, df = 1, $p = 5.1e^{-5}$ S1.10.b: Data S1). We did not detect any significant effect of landscape, specialization level and dispersal capacity on the sensitivity and specificity. Next, we only report the results obtained with the true-positive runs (see Figure 2.ii.).

3.3 | Assessment of RGA predictive performances

3.3.1 | Effect of spatial scale on predictive performance

In the effective-dispersal-scale dataset, 49% of runs associated with a generalist scenario and a dispersal capacity of 1000 cu showed drift-driven genetic differentiation and had therefore undergone a pruning stage. Those with a specialist scenario were pruned in decreasing proportion with increasing dispersal distances (e.g., 85% of the links pruned at 1000 cu, 65% at 5000 cu, 50% at 10,000 cu and 28% at 15,000 cu).

The accuracy of F_{ST} predictions was significantly better when the model was calibrated with the effective-dispersal-scale dataset rather than with the sampling-scale dataset (estimate for sampling-scale ± SE = −1.5 ± 0.1, p-value $<2e^{-16}$, see S1.11: Data S1). However, for comparative purposes, we focused on the runs optimized with the sampling-scale dataset in subsequent analyses (Figure 2.ii.).

3.3.2 | Effect of dispersal capacity, habitat specialization and landscape configuration on predictive performance

The predictive performance of the IBR models (assessed by the validation R^2) depended on dispersal capacities, as evidenced by a quadratic effect of dispersal distances in cost units, and was affected by their interaction with specialization levels (Table 2). Genetic differentiation was slightly better predicted for specialist species than for generalist species, with optimum median values of validation R^2 reaching 0.48 at 15,000 and 20,000 cu for the specialist species, and 0.38 at 1000 cu for the generalist species (Table 2, Figure 3a,b). For both species, forest aggregation improved the predictive performance (estimate = 4.7, $p = 1.4e^{-4}$), whereas grassland aggregation decreased the predictive performance (estimate = −2.8, $p = .01$; Table 2).

Overall, the strength of the IBR pattern, assessed by AIC differences between Euclidean distance and cost distance models, reached an optimum at 20,000 cu for the specialist species and at 1000 cu for the generalist species (Figure 3c,d). This AIC difference

TABLE 2 Results of statistical models explaining the predictive performance of optimized IBR models from the sampling-scale dataset (assessed using the validation R^2 index as the response variable) as a function of the level of specialization, the dispersal capacity and two landscape structure metrics.

Response variable: Validation R^2	
Random effect: 1 | landscape	
Fixed effect	**Estimate (SE)**
Generalist scenario	−0.94 (0.09) ***
Dispersal capacity	6.4 (1.4) ***
Dispersal capacity^2	−8.4 (1.0) ***
Forest aggregation	4.7 (1.2) ***
Grassland aggregation	−2.8 (1.1) *
Generalist sc. * Disp.	−39.4 (3.4) ***
Generalist sc. * Disp.^{2}	ns

Note: Significance: *$p < .05$; **$p < .01$; ***$p < .001$.
Abbreviations: Disp., dispersal capacity; ns, non-significant.

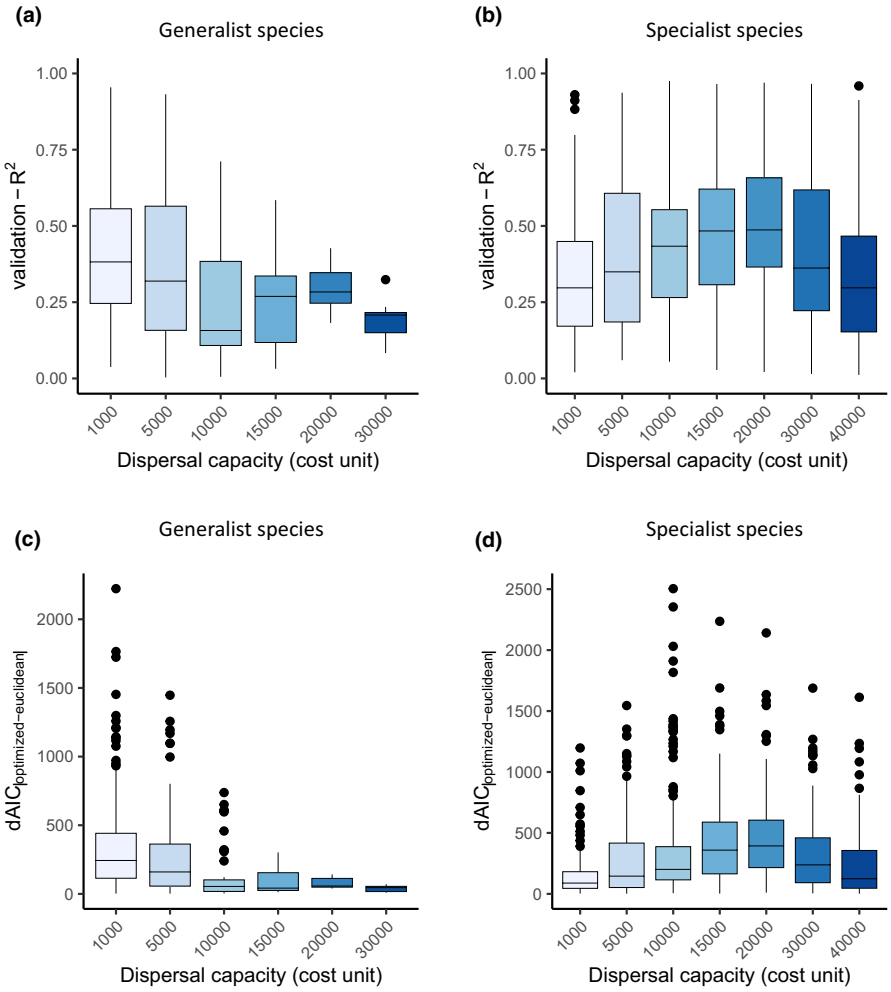


FIGURE 3 Performance criteria as a function of dispersal capacities and specialization levels. Only sampling-scale datasets coming from true positive runs have been considered here. Each box corresponds to a dispersal capacity level and includes the model performance criteria for 30 landscapes and their 10 corresponding population samples. (a, b). Predictive accuracy as assessed by the validation R^2 of the IBR models explaining the genetic distances by the optimized cost-distances for the generalist (a) and specialist species (b). The higher the validation R^2, the better the predictive power of IBR models. (c, d). Strength of the IBR patterns as assessed by the delta AIC (dAIC) between the Euclidean distance models and the IBR models for the generalist (c) and specialist species (d). A positive value indicates that the IBR model performs better than the Euclidean model. The higher the dAIC, the stronger the IBR signal. As the data were filtered to select only the true positive runs, the IBR models always have more support than the Euclidean models (i.e., only positive dAIC are shown here).

was affected by dispersal capacities in the same way as the validation R^2 (quadratic relationship, Figure 3b,c). Hence, both the predictive power and the IBR strength reached an optimum at greater dispersal distances for the specialist species than for the generalist species. Next, we relay the results obtained at the optimal dispersal capacity, maximizing the intensity of the IBR signal, for the specialist virtual species (Figure 2.ii.).

3.3.3 | Transferability of cost inferences

Transferring cost inference to compute cost distances among populations located in another landscape and predict their F_{ST} resulted in poorer extrapolated predictions as compared with interpolated predictions (GLMM parameter estimate of extrapolation effect on validation $R^2 \pm SE = -2.27 \pm 0.12$, $p = 2e^{-16}$). Overall, interpolations led to predictions with a mean validation R^2 of .44, whereas extrapolations had a lower mean validation R^2 of .12 (Figure 4). This indicates that resistance values informed predictions more reliably in the calibration landscape than in other landscapes.

In addition, large variations in predictive performances within and among landscapes (median validation R^2 ranging from 0 to .7 across landscapes) suggested that certain landscape structures improved RGA performances (Figure 4). Similarly, the variability of R^2 validation across the 10 population samples within the same landscape (i.e., within-box variations in Figure 4) also depended on the landscape. This may be related to differences in the configuration of sampled populations, since all other variables remained equal. This suggests that RGA performance is more sensitive to sampling design in certain landscapes.

Finally, the predictive capacity of optimized cost scenarios was much better than that of true cost scenarios (true cost scenario estimate $\pm$ SE $= -0.78 \pm 0.06$, $p < 2e^{-16}$, Figure 5), which indicates that the optimization consistently resulted in overfitting.

3.4 | The congruence between optimized and true cost scenarios

The ranking of optimized costs for land cover types remained almost identical for the four ecological profiles. Yet, this ranking differed from that expected under the true cost scenarios for 90% of the optimized runs (Figure 6). Forest areas were correctly assigned the lowest resistance (i.e., rank 1) in about 50% of the cases, and the second lowest otherwise. Urban and grassland areas were assigned to each of the four ranks in equal proportions, while agricultural areas were

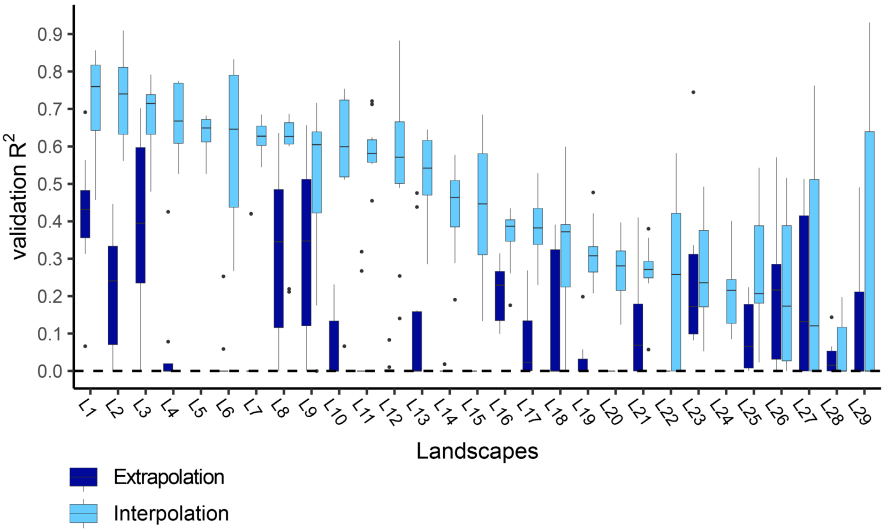


FIGURE 4 Evaluation of the interpolation and the extrapolation capacities of the optimized IBR models for a specialist species. Only the true positive runs and dispersal capacity maximizing the intensity of the IBR signal at sampling-scale are displayed here. The figure shows the validation R^2 of IBR models as a function of the model calibration landscapes. Each pair of light and dark blue boxes represents runs calibrated on the same landscape (displayed in the x axis), for their 10 related samples of populations. Light blue bars correspond to interpolation results, showing validation R^2 for data located in the calibration area, whereas dark blue bars correspond to extrapolation results, showing validation R^2 for data located on another landscape, different from the calibration area. Only 29 of the 30 landscapes are represented. Indeed, one landscape has less than six population samples at the selected dispersal distance that resulted in true positive runs. It was discarded to ensure enough runs per box (here $6 \leq n \leq 10$). The remaining 29 landscapes are ranked in descending order according to the median value of the validation R^2 associated with the interpolated models. L1, Landscape 1; and so on for the 29 landscapes.

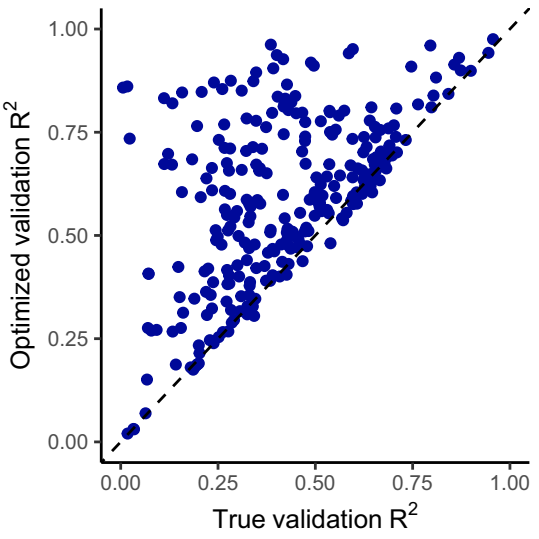


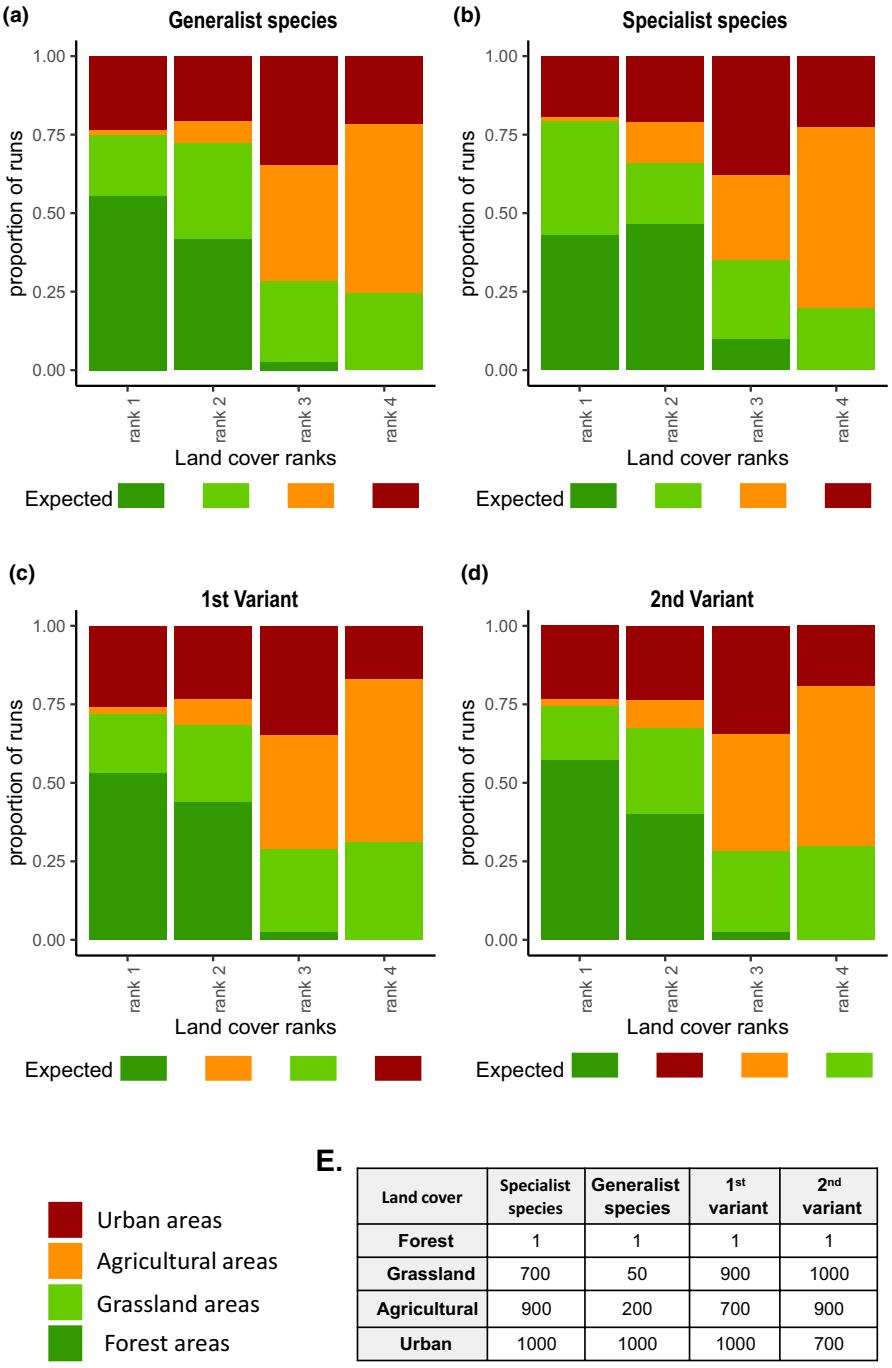
FIGURE 5 Comparison of predictive performances (assessed by the validation R^2 of IBR models) between the true and the optimized cost scenarios for a specialist species. Only the true positive runs with a dispersal capacity maximizing the intensity of the IBR signal at the sampling-scale are shown. Each point corresponds to a combination of landscape × population sample. The points above the $y = x$ line indicate better predictive power of the models using optimized cost distances as compared with the models using the true cost distances used for simulating the data (i.e., the true scenarios), reflecting the effect of overfitting related to the optimization process.

mostly associated with the highest resistance rank (rank 4 in 53% of the cases). The median contrast between the lowest and highest optimized cost values was consistent with the true scenarios (i.e., $\log_{10}(1000) = 3$; Figure 7a). Moreover, we found a correlation of 0.51 between the land cover recovery index (i.e., the Spearman correlation between true and optimized scenarios) and the contrast recovery index (i.e., $3 - \log_{10}$(optimized contrast)). This suggested that the closer the optimized classification of land cover types is to the true ranking, the closer the contrast in optimized values is to the true contrast.

The models explaining the two recovery indices by landscape and ecological variables only evidenced an effect of the specialization profile. The recovery of land cover ranks showed significantly poorer performance with variant 1 ($p < .001$), and the contrast was poorly assessed with the generalist profile ($p < .001$). Finally, the ranking of the land cover types and the contrast values did not affect the validation R^2 of optimized cost scenarios.

4 | DISCUSSION

We demonstrated that the RGA approach properly detects the process shaping genetic structure (i.e., IBR) and leads to accurate genetic differentiation predictions. However, the optimized costs do not always reflect the actual permeability of landscape features to gene flow. We provide guidance for future uses of RGA in landscape



E.

Land cover	Specialist species	Generalist species	1st variant	2nd variant
Forest	1	1	1	1
Grassland	700	50	900	1000
Agricultural	900	200	700	900
Urban	1000	1000	1000	700

FIGURE 6 Assignment of each land cover type to the 4 ranks of relative resistance in the optimized cost scenarios for (a) a generalist species, (b) a specialist species, (c, d) virtual species with ecological profiles derived from variations in the specialist profile (the order of land use resistance has been reversed as compared with the specialist scenario). Each bar represents the proportion of each land cover type assigned to every resistance rank, across the 30 landscapes and their 10 related population samples from the sampling-scale dataset. The 'Expected' line below each bar plot corresponds to the land cover type expected under the true cost scenario, represented by its colour. The recovery of the land cover ranking is accurate if the expected colours match the dominant colours in the corresponding bar plot. (e) Cost values assigned to every land cover type in the four true cost scenarios considered. Each column of the table (i.e., each true cost scenario) has to be compared with the bar plot describing the related optimized cost scenario. For example, for (b) and rank 2, we expected all runs to be grassland (light green), but we see that only about 50% of the runs are grassland. The remaining 50% is divided into about 20% forest (dark green), 20% urban (red) and 10% agricultural (orange).

genetics, emphasizing the importance of the set of population pairs included in analyses and the use of cross-validation approaches preventing overfitting.

4.1 | The spatial scale of landscape influence on genetic structure depends on the topology of the effective dispersal network

The detection of an IBR pattern depended on the interaction between species specialization level and dispersal capacity. These two parameters affect the scale at which the genetic pattern emerges, and therefore the scale to consider for properly detecting it. For instance, in species experiencing moderate movement costs across the landscape (called 'generalist' in our study), IBR patterns can only emerge and be detected if this species covers short distances overall. In other words, the dispersal limitation responsible for IBR patterns (Orsini et al., 2013) is caused by the interplay of movement costs and dispersal distances. This dual cause of dispersal limitation also explains why IBR patterns were detected for larger dispersal distances with the 'specialist' cost profile. This result is consistent with other studies that have demonstrated the impact of species specialization on effective dispersal and its consequences for population genetic differentiation (Harris & Reed, 2002; Khimoun et al., 2016). A spatially structured

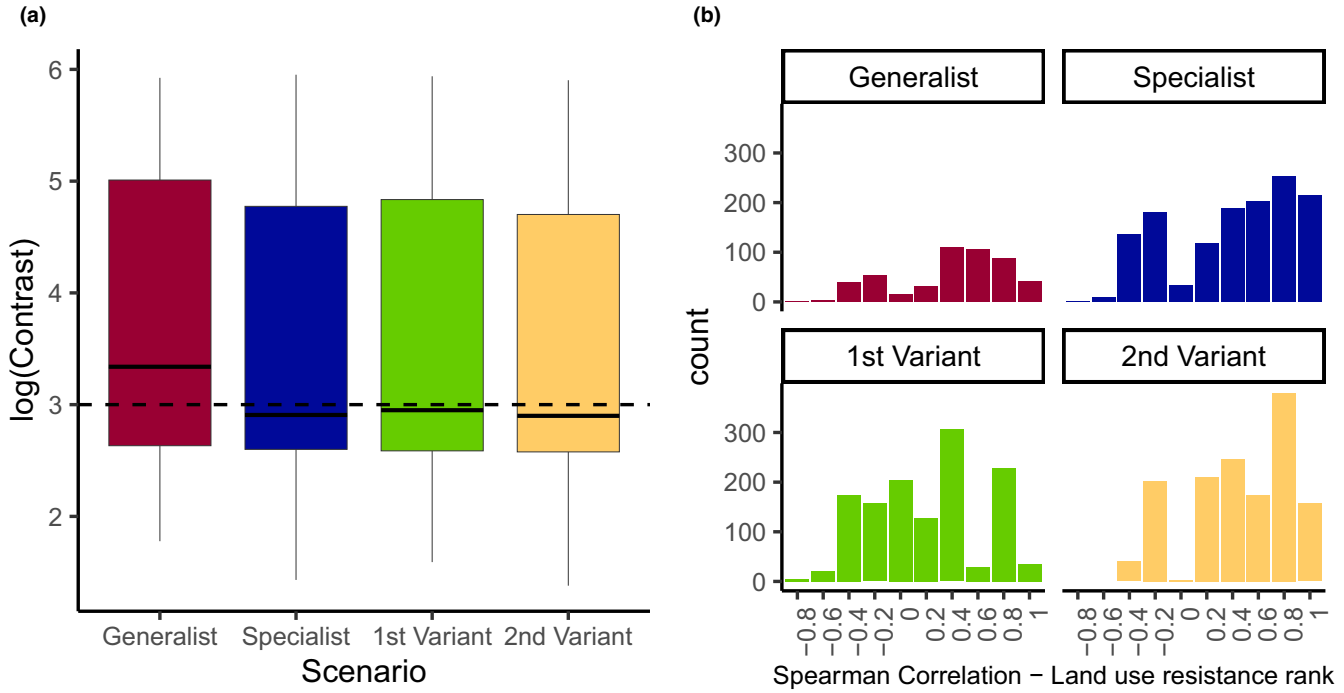


FIGURE 7 (a) Characterization of the optimized cost scenarios according to their contrast as a function of the ecological profile. A log contrast of 3 (dotted line) represents a contrast of 1000, that is, the contrast of the simulated cost scenarios and that we would expect for the optimized scenarios. A log contrast greater than 3 means that the contrast is overestimated, while a log contrast less than 3 means that the contrast is underestimated. (b) Distribution of the ranking consistency with true cost scenarios of optimized runs (i.e., Spearman correlation between the ranking of land covers in the optimized cost scenario and in the true cost scenario) according to the ecological profile. The closer the correlation value is to 1, the greater the agreement between the order of resistance of the land covers in the real scenarios and the optimized scenarios. See S1.12: Data S1 for the significance of Spearman correlation values for land cover ranking in optimized cost scenarios.

pattern is an essential pre-requisite for optimizing resistance surfaces, as subsequent analyses assume its presence. Furthermore, matching the scale at which both drift and dispersal shape genetic structure (i.e., the effective-dispersal scale here) is needed to align inferences with the process under study (Savary et al., 2021a). This scale also depends on species dispersal abilities (cost profile, dispersal distances), landscape and population configuration and defines the topology of the effective dispersal network (Savary et al., 2021a).

4.2 | Pruning enhances RGA predictive performance under strong dispersal limitation

We demonstrated that RGA predictive performances were maximized when considering the subset of population pairwise connections matching the effective-dispersal-scale. In other words, better results were obtained from a reduced dataset. It is expected from theory that low gene flow will lead to isolated populations whose genetic differentiation is mainly driven by stochastic genetic drift (Hutchison & Templeton, 1999) and consequently difficult to predict from their location on the landscape. Conversely, if gene flow is too strong, landscape constraints to dispersal may no longer be a limiting factor, thus making their effects difficult to detect. Accordingly, the effective dispersal scale to consider for selecting pairwise population connections allowing for the detection and assessment of landscape effects on gene flow depends on the degree to which gene flow is restricted (Savary et al., 2021b).

Considering the above, we introduced a new approach to infer cost values, involving a pruning method (Dyer & Nason, 2004) restricting the pairwise connections to those matching the effective-dispersal scale. We showed that pruning the distance matrices improved RGA performances as it ensured that genetic differentiation was only modelled between populations supposed to be linked by substantial dispersal and gene flow (Dyer & Nason, 2004; Keller et al., 2013; Murphy et al., 2010; Van Strien et al., 2015). Daniel et al. (2023) had already shown a similar positive effect of reducing the genetic dataset to capture better the spatial scale of effective dispersal using empirical data. These results confirm the conclusion of previous studies (Savary et al., 2021b; Van Strien, 2017), suggesting that considering the topology of dispersal networks might improve resistance surface optimization.

4.3 | Sampling design and landscape structure affect optimization performance

We found that in each landscape, the predictive performances of the optimized cost values varied greatly across the 10 sets of 40

sampled populations, out of 80 simulated. This suggests a significant influence of the population sampling design on the inference outcome, potentially related to the capacity of this sampling to properly capture the dispersal network. Indeed, the spatial distribution of the whole set of populations determines the dispersal network shaping genetic patterns (McRae, 2006; Van Strien, 2017). Accordingly, when some populations are absent from the sampled set, landscape genetic analyses cannot reliably capture the spatial drivers of genetic structure because part of the spatial signal is missing (Naujokaitis-Lewis et al., 2013; Van Strien, 2017). This was further supported by our finding of a significant relationship between the validation R^2 and the aggregation of forest areas, a factor that influences the spatial distribution of the sampled populations. Van Strien et al. (2015) showed that population topology was tightly linked to habitat distribution, and our results suggest that some landscape configurations and compositions might be more resilient to partial sampling. Interestingly, RGA performances seemed to be enhanced in coarse-grained and aggregated landscapes. One explanation could be that the topology of the dispersal network is easier to capture when land cover patches are aggregated, as sampling each aggregate reduces the risk of missing central populations in the dispersal network.

4.4 | Optimization is not causation: Good predictive abilities at the expense of accurate causal interpretation

Many studies have used RGA cost inferences to rank landscape features according to their resistance to gene flow (Antunes et al., 2023; Khimoun et al., 2017; Mapelli et al., 2020; Martin et al., 2023; Mulvaney et al., 2021; Reyne et al., 2023; Ruiz-Lopez et al., 2016). These interpretations often served as the basis for a mechanistic understanding of the landscape effect on dispersal. However, we evidenced a frequent mismatch between the optimized and true cost rankings across land cover types. Only 10% of the optimized runs showed a land cover cost ranking in line with the simulated reality. Surprisingly, the low accuracy of cost assignments remained fairly identical, regardless of the type of true cost scenario shaping genetic structure. Moreover, this cost ranking often seemed to reflect landscape composition rather than its actual permeability to gene flow. For instance, we found that the least frequent land cover types in the landscape, that is, grassland and urban areas, were equally likely to be assigned any one of the 4 resistance ranks. They probably served as adjustment variables in the optimization process, mainly driven by model goodness-of-fit.

Besides, the more different the optimized ranking from the true ranking, the larger the overestimation of the optimized contrast relative to the true contrast. Therefore, our results call for great caution when interpreting the resistance of landscape features. Similarly, without discussing the issues in detail, Peterman et al. (2019) and Beninde et al. (2023) pointed out possible difficulties in deriving reliable cost values with the RGA framework. The overfitting effect inherent to the optimization process (Peterman et al., 2019; Winiarski et al., 2020; Yates et al., 2018) was mentioned as a possible cause of mismatch. Given that optimized cost scenarios had better predictive performance than the true cost scenarios, we can reasonably expect that overfitting in the calibration of cost values leads to a strong dependence on landscape configuration.

Furthermore, the dependence of the inference on the topology of the dispersal network might exacerbate the adverse effect of overfitting on inference accuracy. Indeed, the good predictive accuracy of RGA models during cost calibration, even for population pairs excluded from the calibration dataset, dropped when the models were extrapolated to population pairs located in another landscape. As cost inference is highly dependent on landscape composition and population topology due to overfitting, optimized cost scenarios are unlikely to provide reliable parameters for predicting the genetic structure of populations from another landscape.

Despite the above-mentioned limitations, our results emphasize the RGA's good predictive performances, thereby reflecting the common duality between explanatory and predictive modelling (Shmueli, 2010; Yates et al., 2018). Based on a data-driven process and an optimization algorithm, RGA provides accurate predictions for genetic differentiation, which could then lead to a better understanding of the effect of population topology or landscape configuration on genetic structure. However, the causal link between the cost values obtained in the inference and the landscape effect on gene flow is potentially dubious. This echoes the fact that predictive models, such as RGA, rarely provide insight into the underlying causal mechanism (Shmueli, 2010).

Some very promising landscape genetics approaches, for example, using deep and machine learning, have recently been implemented to predict genetic connectivity. Kittlein et al. (2022) showed that convolutional neural networks could provide highly accurate predictions for small-scale genetic differentiation and diversity, while Pless et al. (2021) conducted a least-cost transect analysis to predict gene flow for *Aedes aegypti* vector regulation. As soon as the common limitations of data-driven approaches are acknowledged, predictive models could be of great interest to capture complex patterns and relationships, otherwise difficult to predict using theory-based models (Lucas, 2020; Murphy et al., 2010; Shmueli, 2010; Vanhove & Launey, 2023).

4.5 | The use of RGA in landscape genetics: Conditions and prospects

We outline below a set of guidelines regarding the use of RGA and list them in the order in which they should be considered when designing a landscape genetic study.

First, we call for preliminary assessments of spatial genetic structuring before the use of optimization approaches (e.g., through genetic clustering analyses or the study of IBD patterns). When the genetic structuring is weak, great care should be taken when interpreting the optimization results.

Second, one needs to capture the spatial scale of the observed genetic signal properly (Savary et al., 2021a). To this end, we recommend performing inferences at multiple spatial scales, either using Moran's Eigenvector Maps (Dray et al., 2006; Galpern et al., 2014) or through iterative analyses with multiple pruning thresholds (Van Strien et al., 2015), and selecting the one that leads to the best predictive performance of the optimized IBR model.

Third, the reliability of RGA inferences may be improved by favouring exhaustive sampling designs to correctly model topological effects and limit their confounding effect (Van Strien, 2017). If this is too costly, preliminary sensitivity analyses, for example, based on genetic simulations, could determine the ideal set of populations to sample in the focal landscape (Naujokaitis-Lewis et al., 2013). After empirical data have been sampled and used for optimization, the stability of the optimized resistance surface when refitting it with a subset of the populations can also be assessed with a new bootstrap procedure implemented in RGA (Peterman et al., 2019).

Fourth, attention must be paid to the structure of the studied landscape and the thematic resolution. When some land cover types are poorly represented, or for some land cover configurations, the sensitivity of RGA inferences to landscape structure may prevent interpreting the optimized cost values. In that case, highly unstable cost values and rankings point towards unreliable inferences. Here too, simulating gene flow in the focal landscape and assessing the algorithm's ability to capture the correct cost scenario might prevent spurious conclusions.

Fifth, it is strongly recommended to limit the overfitting of the optimized model and ensure that it does not fit the noise in the data rather than the targeted signal (Lucas, 2020; Peterman et al., 2019; Shmueli, 2010; Winiarski et al., 2020). To do this, out-of-sample performance can be assessed, as in the present and a few previous studies (see Daniel et al., 2023; Van Strien et al., 2014). This also provides an assessment of the transferability of the inferences. Additionally, when sufficient data is available to perform independent K-fold cross-validation, accounting for spatial autocorrelation for defining the training and test datasets (i.e., spatial cross-validation) can lead to more accurate extrapolated predictions than random cross-validation (Palm et al., 2023).

Finally, alternative methodological choices potentially improving the inference of landscape resistance remain to be explored in the context of landscape genetic optimizations. For instance, Beninde et al. (2023) called for the adoption of eigenvector-based estimates of pairwise genetic distances to reliably infer the effect of landscape features on gene flow. Moreover, using individual-based instead of population-based genetic indices or resistance distances instead of cost distances to perform similar analyses could provide insights into the use of RGA in a wider range of contexts. However, these choices should not fundamentally change our conclusions, as they are partially supported by other studies using other ecological (Beninde et al., 2023; Peterman et al., 2019) or genetic distances (Beninde et al., 2023; Winiarski et al., 2020). Furthermore, reproducing our analyses for continuous landscape representations (Peterman et al., 2019; Vanhove & Launey, 2023) or using more recent gradient algorithms (such as the Radish package, Peterman & Pope, 2020) or gradient forests (Vanhove & Launey, 2023) could broaden its conclusions in a useful way.

Although our guidelines and future tests of these approaches might improve the reliability of optimization results, they cannot lead to any improvements in the presence of significant noise in the data or when a key spatial process or covariate is missed (Lucas, 2020). If such a covariate is of great importance, including it in explanatory models is likely to improve their performance (Keller et al., 2013; Savary et al., 2021a; Van Strien, 2017; Van Strien et al., 2014).

In conclusion, the strength of the RGA workflow is its excellent ability to predict genetic distances, although its lack of transferability limits the prospective use of its inference to predict the impact of landscape change on gene flow. This characteristic can still be very useful for operational studies in conservation and population genetics (Van Strien et al., 2014). It may even be interesting to use the genetic distances predicted by RGA to inform operational models. For example, we could imagine using RGA's cost inference to weight the links of habitat network models and derive relevant functional connectivity metrics, design dispersal corridors, or identify restoration areas (Foltête et al., 2014).

AUTHOR CONTRIBUTIONS

A.D., P.S., A.K. and S.G. designed the project. P.S. and G.V. developed the R functions for the optimization workflow. P.S. and A.D. designed the R functions for the genetic simulations. A.D. performed the analysis and wrote the manuscript, with significant contributions and remarks from all co-authors.

ACKNOWLEDGEMENTS

We thank Julien Pergaud for his help with the computer programming aspect. Simulations and optimization analyses were carried out on the computing cluster of the University of Bourgogne, with Henri Gaulard as our cluster correspondent. We are very grateful to Erin Landguth and two anonymous reviewers for the relevance of their comments, which notably improved this manuscript. This work is supported by the French Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, and is part of the CANON project steered by Stéphane Garnier and funded by the French Investissements d'Avenir program, project ISITE-BFC (contract ANR-15-IDEX-0003).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data and code are available online: https://doi.org/10.5281/zenodo.8180746. The modified version of ResistanceGA package is also available online: https://gitlab.com/psavary3/rga.

ORCID

Alexandrine Daniel https://orcid.org/0000-0003-3379-281X
Paul Savary https://orcid.org/0000-0002-2104-9941

Jean-Christophe Foltête https://orcid.org/0000-0003-4864-5660
Bruno Faivre https://orcid.org/0000-0002-2493-8381

REFERENCES

Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, *5*, 384–387. https://doi.org/10.1111/2041-210X.12158

Antunes, B., Figueiredo-Vázquez, C., Dudek, K., Liana, M., Pabijan, M., Zieliński, P., & Babik, W. (2023). Landscape genetics reveals contrasting patterns of connectivity in two newt species (*Lissotriton montandoni* and *L. vulgaris*). *Molecular Ecology*, *32*, 4515–4530. https://doi.org/10.1111/mec.16543

Atzeni, L., Wang, J., Riordan, P., Shi, K., & Cushman, S. A. (2023). Landscape resistance to gene flow in a snow leopard population from Qilianshan National Park, Gansu, China. *Landscape Ecology*, *38*, 1847–1868. https://doi.org/10.1007/s10980-023-01660-8

Balkenhol, N., Gugerli, F., Cushman, S. A., Waits, L. P., Coulon, A., Arntzen, J. W., Holderegger, R., Wagner, H. H., & Participants of the Landscape Genetics Research Agenda Workshop 2007. (2009). Identifying future research needs in landscape genetics: Where to from here? *Landscape Ecology*, *24*, 455–463. https://doi.org/10.1007/s10980-009-9334-z

Balkenhol, N., Schwartz, M. K., Inman, R. M., Copeland, J. P., Squires, J. S., Anderson, N. J., & Waits, L. P. (2020). Landscape genetics of wolverines (*Gulo gulo*): Scale-dependent effects of bioclimatic, topographic, and anthropogenic variables. *Journal of Mammalogy*, *101*, 790–803. https://doi.org/10.1093/jmammal/gyaa037

Bartoń, K. (2013). *MuMIn: Multi-model inference*, *R package version*, *1.10.0*.

Beninde, J., Wittische, J., & Frantz, A. C. (2023). Quantifying uncertainty in inferences of landscape genetic resistance due to choice of individual-based genetic distance metric. *Molecular Ecology Resources*, *24*, e13831. https://doi.org/10.1111/1755-0998.13831

Brooks, M., Bolker, B., Kristensen, K., Mächler, M., Magnusson, A., Skaug, H., Nielsen, A., Berg, C., & van Benthem, K. (2017). *glmmTMB: Generalized linear mixed models using Template Model Builder*.

Correa Ayram, C. A., Mendoza, M. E., Etter, A., & Salicrup, D. R. P. (2016). Habitat connectivity in biodiversity conservation: A review of recent studies and applications. *Progress in Physical Geography: Earth and Environment*, *40*, 7–37. https://doi.org/10.1177/0309133315598713

Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., & Hendry, A. P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, *15*, 49–62. https://doi.org/10.1111/j.1365-294X.2005.02764.x

Crooks, K. R., Burdett, C. L., Theobald, D. M., King, S. R. B., Di Marco, M., Rondinini, C., & Boitani, L. (2017). Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7635–7640. https://doi.org/10.1073/pnas.1705769114

Daniel, A., Savary, P., Foltête, J., Khimoun, A., Faivre, B., Ollivier, A., Éraud, C., Moal, H., Vuidel, G., & Garnier, S. (2023). Validating graph-based connectivity models with independent presence–absence and genetic data sets. *Conservation Biology*, *37*, e14047. https://doi.org/10.1111/cobi.14047

Diniz, M. F., Cushman, S. A., Machado, R. B., & De Marco Júnior, P. (2020). Landscape connectivity modeling from the perspective of animal dispersal. *Landscape Ecology*, *35*, 41–58. https://doi.org/10.1007/s10980-019-00935-3

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3–4), 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Dutta, T., Sharma, S., Meyer, N. F. V., Larroque, J., & Balkenhol, N. (2022). An overview of computational tools for preparing, constructing and using resistance surfaces in connectivity research. *Landscape Ecology*, *37*, 2195–2224. https://doi.org/10.1007/s10980-022-01469-x

Dyer, R. J., & Nason, J. D. (2004). Population graphs: The graph theoretic shape of genetic structure. *Molecular Ecology*, *13*, 1713–1727. https://doi.org/10.1111/j.1365-294X.2004.02177.x

Foltête, J.-C., Girardet, X., & Clauzel, C. (2014). A methodological framework for the use of landscape graphs in land-use planning. *Landscape and Urban Planning*, *124*, 140–150. https://doi.org/10.1016/j.landurbplan.2013.12.012

Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*, 2610–2618. https://doi.org/10.1111/mec.13139

Galpern, P., Peres-Neto, P. R., Polfus, J., & Manseau, M. (2014). MEMGENE: Spatial pattern detection in genetic distance data. *Methods in Ecology and Evolution*, *5*(10), 1116–1120. https://doi.org/10.1111/2041-210x.12240

Harris, R. J., & Reed, J. M. (2002). Behavioral barriers to non-migratory movements of birds. *Annales Zoologici Fennici*, *39*, 275–290.

Hutchison, D. W., & Templeton, A. R. (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*(6), 1898. https://doi.org/10.2307/2640449

Inglada, J., Vincent, A., & Thierion, V. (2018). *Theia OSO Land Cover Map 2018*. https://doi.org/10.5281/zenodo.3613415

Keeley, A. T. H., Beier, P., & Gagnon, J. W. (2016). Estimating landscape resistance from habitat suitability: Effects of data source and nonlinearities. *Landscape Ecology*, *31*, 2151–2162. https://doi.org/10.1007/s10980-016-0387-5

Keller, D., Holderegger, R., & Van Strien, M. J. (2013). Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology*, *22*, 2467–2482. https://doi.org/10.1111/mec.12265

Khimoun, A., Eraud, C., Ollivier, A., Arnoux, E., Rocheteau, V., Bely, M., Lefol, E., Delpuech, M., Carpentier, M.-L., Leblond, G., Levesque, A., Charbonnel, A., Faivre, B., & Garnier, S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, *25*, 3831–3844. https://doi.org/10.1111/mec.13733

Khimoun, A., Peterman, W., Eraud, C., Faivre, B., Navarro, N., & Garnier, S. (2017). Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology*, *26*(19), 4906–4919. https://doi.org/10.1111/mec.14233

Kittlein, M. J., Mora, M. S., Mapelli, F. J., Austrich, A., & Gaggiotti, O. E. (2022). Deep learning and satellite imagery predict genetic diversity and differentiation. *Methods in Ecology and Evolution*, *13*, 711–721. https://doi.org/10.1111/2041-210X.13775

Lucas, T. C. D. (2020). A translucent box: Interpretable machine learning in ecology. *Ecological Monographs*, *90*, e01422. https://doi.org/10.1002/ecm.1422

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology and Evolution*, *28*, 614–621. https://doi.org/10.1016/j.tree.2013.05.012

Mapelli, F. J., Boston, E. S. M., Fameli, A., Gómez Fernández, M. J., Kittlein, M. J., & Mirol, P. M. (2020). Fragmenting fragments: Landscape genetics of a subterranean rodent (Mammalia, Ctenomyidae) living in a human-impacted wetland. *Landscape Ecology*, *35*, 1089–1106. https://doi.org/10.1007/s10980-020-01001-z

Martin, S. A., Peterman, W. E., Lipps, G. J., & Gibbs, H. L. (2023). Inferring population connectivity in eastern massasauga rattlesnakes (*Sistrurus catenatus*) using landscape genetics. *Ecological Applications*, *33*, e2793. https://doi.org/10.1002/eap.2793

McCluskey, E. M., Lulla, V., Peterman, W. E., Stryszowska-Hill, K. M., Denton, R. D., Fries, A. C., Langen, T. A., Johnson, G., Mockford, S. W., & Gonser, R. A. (2022). Linking genetic structure, landscape genetics, and species distribution modeling for regional conservation of a threatened freshwater turtle. *Landscape Ecology*, *37*, 1017–1034. https://doi.org/10.1007/s10980-022-01420-0

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*, 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

Mulvaney, J. M., Matthee, C. A., & Cherry, M. I. (2021). Species–landscape interactions drive divergent population trajectories in four forest-dependent Afromontane forest songbird species within a biodiversity hotspot in South Africa. *Evolutionary Applications*, *14*, 2680–2697. https://doi.org/10.1111/eva.13306

Murphy, M. A., Dezzani, R., Pilliod, D. S., & Storfer, A. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, *19*, 3634–3649. https://doi.org/10.1111/j.1365-294X.2010.04723.x

Naujokaitis-Lewis, I. R., Rico, Y., Lovell, J., Fortin, M.-J., & Murphy, M. A. (2013). Implications of incomplete networks on estimation of landscape genetic connectivity. *Conservation Genetics*, *14*, 287–298. https://doi.org/10.1007/s10592-012-0385-3

Newmark, W. D., Halley, J. M., Beier, P., Cushman, S. A., McNeally, P. B., & Soulé, M. E. (2023). Enhanced regional connectivity between western north American national parks will increase persistence of mammal species diversity. *Scientific Reports*, *13*, 474. https://doi.org/10.1038/s41598-022-26428-z

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, *22*, 5983–5999. https://doi.org/10.1111/mec.12561

Palm, E. C., Landguth, E. L., Holden, Z. A., Day, C. C., Lamb, C. T., Frame, P. F., Morehouse, A. T., Mowat, G., Proctor, M. F., Sawaya, M. A., Stenhouse, G., Whittington, J., & Zeller, K. A. (2023). Corridor-based approach with spatial cross-validation reveals scale-dependent effects of geographic distance, human footprint and canopy cover on grizzly bear genetic connectivity. *Molecular Ecology*, *32*, 5211–5227. https://doi.org/10.1111/mec.17098

Paris, G., Robilliard, D., & Fonlupt, C. (2004). Exploring overfitting in genetic programming. In P. Liardet, P. Collet, C. Fonlupt, E. Lutton, & M. Schoenauer (Eds.), *Artificial Evolution* (pp. 267–277). Springer.

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*, 1638–1647. https://doi.org/10.1111/2041-210X.12984

Peterman, W. E., & Pope, N. S. (2020). The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology*, *30*(1), 37–47. https://doi.org/10.1111/mec.15716

Peterman, W. E., Winiarski, K. J., Moore, C. E., Carvalho, C. d. S., Gilbert, A. L., & Spear, S. F. (2019). A comparison of popular approaches to optimize landscape resistance surfaces. *Landscape Ecology*, *34*, 2197–2208. https://doi.org/10.1007/s10980-019-00870-3

Pless, E., Saarman, N. P., Powell, J. R., Caccone, A., & Amatulli, G. (2021). A machine-learning approach to map landscape connectivity in *Aedes aegypti* with genetic and environmental data. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2003201118. https://doi.org/10.1073/pnas.2003201118

Reyne, M. I., Dicks, K., Flanagan, J., Nolan, P., Twining, J. P., Aubry, A., Emmerson, M., Marnell, F., Helyar, S., & Reid, N. (2023). Landscape genetics identifies barriers to Natterjack toad metapopulation dispersal. *Conservation Genetics*, *24*, 375–390. https://doi.org/10.1007/s10592-023-01507-4

Richardson, J. L., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, *25*, 849–863. https://doi.org/10.1111/mec.13527

Rudnick, D., Ryan, S., Beier, P., Cushman, S., Dieffenbach, F., Epps, C., Gerber, L., Hartter, J., Jenness, J., Kintsch, J., Merenlender, A., Perkl, R., Preziosi, D., & Trombulak, S. (2012). The role of landscape connectivity in planning and implementing conservation and restoration priorities. *Issues in Ecology*, *16*, 1–23.

Ruiz-Lopez, M. J., Barelli, C., Rovero, F., Hodges, K., Roos, C., Peterman, W. E., & Ting, N. (2016). A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity*, *116*, 167–176. https://doi.org/10.1038/hdy.2015.82

Savary, P., Foltête, J., Moal, H., Vuidel, G., & Garnier, S. (2021a). Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, *21*, 1167–1185. https://doi.org/10.1111/1755-0998.13333

Savary, P., Foltête, J., Moal, H., Vuidel, G., & Garnier, S. (2021b). graph4lg: A package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, *12*, 539–547. https://doi.org/10.1111/2041-210X.13530

Schlägel, U. E., Grimm, V., Blaum, N., Colangeli, P., Dammhahn, M., Eccard, J. A., Hausmann, S. L., Herde, A., Hofer, H., Joshi, J., Kramer-Schadt, S., Litwin, M., Lozada-Gobilard, S. D., Müller, M. E. H., Müller, T., Nathan, R., Petermann, J. S., Pirhofer-Walzl, K., Radchuk, V., … Jeltsch, F. (2020). Movement-mediated community assembly and coexistence. *Biological Reviews*, *95*, 1073–1096. https://doi.org/10.1111/brv.12600

Scrucca, L. (2013). GA: A package for genetic algorithms in *R*. *Journal of Statistical Software*, *53*, 1–37. https://doi.org/10.18637/jss.v053.i04

Shmueli, G. (2010). To explain or to predict? *Statistics Science*, *25*, 289–310. https://doi.org/10.1214/10-STS330

Spear, S. F., Balkenhol, N., Fortin, M.-J., Mcrae, B. H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis. *Molecular Ecology*, *19*, 3576–3591. https://doi.org/10.1111/j.1365-294X.2010.04657.x

Spear, S. F., Cushman, S. A., & McRae, B. H. (2015). Resistance surface modeling in landscape genetics. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Eds.), *Landscape genetics* (pp. 129–148). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118525258.ch08

Spielman, D., Brook, B. W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 15261–15264. https://doi.org/10.1073/pnas.0403809101

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. P. (2007). Putting the 'landscape' in landscape genetics. *Heredity*, *98*, 128–142. https://doi.org/10.1038/sj.hdy.6800917

Van Strien, M. J. (2017). Consequences of population topology for studying gene flow using link-based landscape genetic methods. *Ecology and Evolution*, *7*, 5070–5081. https://doi.org/10.1002/ece3.3075

Van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes: Considerations for landscape genetics. *Heredity*, *114*, 27–37. https://doi.org/10.1038/hdy.2014.62

Van Strien, M. J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F., & Bolliger, J. (2014). Landscape genetics as a tool for conservation planning: Predicting the effects of landscape change on gene flow. *Ecological Applications*, *24*, 327–339. https://doi.org/10.1890/13-0442.1

Vanhove, M., & Launey, S. (2023). Estimating resistance surfaces using gradient forest and allelic frequencies. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13778

Voeten, C. C. (2020). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 1.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370. https://doi.org/10.2307/2408641

Winiarski, K. J., Peterman, W. E., & McGarigal, K. (2020). Evaluation of the R package 'ResistanceGA': A promising approach towards the accurate optimization of landscape resistance surfaces. *Molecular Ecology Resources*, *20*, 1583–1596. https://doi.org/10.1111/1755-0998.13217

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., … Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology and Evolution*, *33*, 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, *27*, 777–797. https://doi.org/10.1007/s10980-012-9737-0

Zeller, K. A., Vickers, T. W., Ernest, H. B., & Boyce, W. M. (2017). Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. *PLoS One*, *12*, e0179570. https://doi.org/10.1371/journal.pone.0179570

Zeller, K. A., Wultsch, C., Welfelt, L. S., Beausoleil, R. A., & Landguth, E. L. (2023). Accounting for sex-specific differences in gene flow and functional connectivity for cougars and implications for management. *Landscape Ecology*, *38*, 223–237. https://doi.org/10.1007/s10980-022-01556-z

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Daniel, A., Savary, P., Foltête, J.-C., Vuidel, G., Faivre, B., Garnier, S., & Khimoun, A. (2024). What can optimized cost distances based on genetic distances offer? A simulation study on the use and misuse of ResistanceGA. *Molecular Ecology Resources*, *00*, e14024. https://doi.org/10.1111/1755-0998.14024