

This article was downloaded by: [Colorado State University]

On: 15 March 2013, At: 12:16

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:

<http://amstat.tandfonline.com/loi/uasa20>

Circuit Theory and Model-Based Inference for Landscape Connectivity

Ephraim M. Hanks^a & Mevin B. Hooten^{b,c}

^a Department of Statistics, Colorado State University, Fort Collins, CO, 80523

^b U.S. Geological Survey Colorado Cooperative Fish and Wildlife Research Unit

^c Fish, Wildlife, and Conservation Biology and Statistics, Colorado State University, Fort Collins, CO, 80523

Accepted author version posted online: 12 Sep 2012. Version of record first published: 15 Mar 2013.

To cite this article: Ephraim M. Hanks & Mevin B. Hooten (2013): Circuit Theory and Model-Based Inference for Landscape Connectivity, *Journal of the American Statistical Association*, 108:501, 22-33

To link to this article: <http://dx.doi.org/10.1080/01621459.2012.724647>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://amstat.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Circuit Theory and Model-Based Inference for Landscape Connectivity

Ephraim M. HANKS and Mevin B. HOOTEN

Circuit theory has seen extensive recent use in the field of ecology, where it is often applied to study functional connectivity. The landscape is typically represented by a network of nodes and resistors, with the resistance between nodes a function of landscape characteristics. The effective distance between two locations on a landscape is represented by the resistance distance between the nodes in the network. Circuit theory has been applied to many other scientific fields for exploratory analyses, but parametric models for circuits are not common in the scientific literature. To model circuits explicitly, we demonstrate a link between Gaussian Markov random fields and contemporary circuit theory using a covariance structure that induces the necessary resistance distance. This provides a parametric model for second-order observations from such a system. In the landscape ecology setting, the proposed model provides a simple framework where inference can be obtained for effects that landscape features have on functional connectivity. We illustrate the approach through a landscape genetics study linking gene flow in alpine chamois (*Rupicapra rupicapra*) to the underlying landscape.

KEY WORDS: Conditional autoregressive models; Gene flow; Landscape genetics; Nonstationary spatial covariance.

1. INTRODUCTION

Circuit theory has been successfully used to study connectivity in a wide range of fields, including molecular chemistry (Zhu and Klein 1996; Klein et al. 2004), collaborative recommendation (Fouss et al. 2007), communications network analysis (Tizghadam and Leon-Garcia 2010, 2011), social network analysis (Kunegis, Lommatzsch, and Bauckhage 2009), and random walks on a graph (Chandra et al. 1996; Volchenkov 2011). Circuit theory has also seen extensive recent use in landscape ecology, where it has been theoretically linked to animal movement and gene flow in heterogeneous landscapes (McRae 2006; Cushman et al. 2006; McRae and Beier 2007; McRae et al. 2008; Urban et al. 2009; Cushman and Landguth 2010; Dyer, Nason, and Garrick 2010; Lookingbill et al. 2010; Owen-Smith, Fryxell, and Merrill 2010; Rayfield, Fortin, and Fall 2010; Saura and Rubio 2010). In these latter cases, the landscape is specified as a raster grid with connectivity between grid cells determined by landscape characteristics and modeled based on circuit theory (Figure 1). Circuit theory provides a flexible framework for modeling nonstationary connectivity, and shows promise for predicting effects of landscape and environmental change on connectivity (e.g., Storfer et al. 2007; Spear et al. 2010).

A key challenge in modeling landscape connectivity using circuit theory is to estimate the relative resistance values of various landscape characteristics (e.g., Spear et al. 2010). In applications of circuit theory other than landscape ecology, resistance values are typically known, or all resistors in the circuit are assumed to have equal resistance. In these cases, the focus is typically on exploratory analysis of the connectivity implied by viewing the system as a circuit, rather than on estimating resistance values based on observations. In contrast, the goal in landscape ecological applications of circuit theory is to understand the impact that

different landscape characteristics have on connectivity. Observations are typically second order, and come in the form of an observed pairwise distance matrix representing the process under study (e.g., spatial gene flow in landscape genetics). The most common approach used to estimate resistance values for different landscape characteristics is to choose between a set of prespecified candidate resistance values for each landscape covariate hypothesized to have an effect on connectivity (e.g., Cushman et al. 2006; Cushman, McKelvey, and Schwartz 2009). Each set of candidate resistance values is used to create a hypothesized resistance distance matrix between the observed spatial locations in the study, where the resistance distance is computed based on circuit theory. The correlations between each of the hypothesized distance matrices and the observed distance matrix are computed, and the set of candidate resistance values that results in the highest correlation to the observed distance matrix is chosen (e.g., Cushman et al. 2006; Cushman, McKelvey, and Schwartz 2009; Wang, Savage, and Bradley Shaffer 2009; Shirk et al. 2010), with significance assessed through Mantel permutation tests (e.g., Legendre and Fortin 2010).

One major drawback of this approach is that there is no obvious way to assess the uncertainty in the parameter estimates for the resistance values of the landscape covariates. This is a critical point, as the results of spatial connectivity studies are being used to influence policy decisions (e.g., Theobald, Crooks, and Norman 2011) and predict the results of landscape change over time (e.g., Spear et al. 2010).

Our goal is to put the estimation of resistance values from observed genetic distance matrices within a model-based framework. Recent work by McCullagh (2009) shows that observed squared-Euclidean distance matrices can be modeled using the generalized (or intrinsic) Wishart distribution with a spatial covariance matrix as a parameter. However, it is not immediately obvious how to parameterize a covariance matrix in a way that models connectivity based on circuit theory. As circuits are based on a graph or network of nodes, it seems natural to

Ephraim M. Hanks is Doctoral candidate in the Department of Statistics, Colorado State University, Fort Collins, CO 80523 (E-mail: hanks@stat.colostate.edu). Mevin B. Hooten is Assistant Unit Leader at the U.S. Geological Survey Colorado Cooperative Fish and Wildlife Research Unit and Assistant Professor in Fish, Wildlife, and Conservation Biology and Statistics at Colorado State University, Fort Collins, CO 80523 (E-mail: mevin.hooten@colostate.edu). The use of trade names or products does not constitute endorsement by the U.S. government.

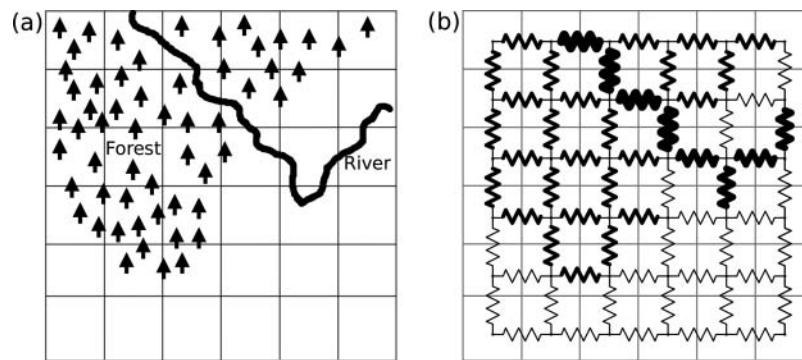


Figure 1. Illustration of the use of circuits to model spatial connectivity. The landscape is seen as a raster grid where connectivity between adjacent grid cells is a function of local landscape characteristics, and is computed using circuit theory. In the hypothetical landscape (a), both the forest and the river impede connectivity. In the corresponding circuit (b), the raster cells are seen as being nodes connected by resistors, with thicker resistors graphically representing higher resistance and reduced conductance between nodes.

consider spatial statistical models with discrete spatial support, such as Gaussian Markov random field (GMRF) models (e.g., Rue and Held 2005). In what follows, we make use of intrinsic conditional autoregressive (ICAR) models (Besag 1974; Besag and Kooperberg 1995) for modeling spatial connectivity based on circuit theory. We show that the GMRF with ICAR covariance structure can be seen as a spatial model for a Gaussian process with connectivity defined by circuit theory, and that the expected squared difference between observations from an ICAR model is exactly the resistance distance defined by circuit theory. The ICAR covariance matrix is also identifiable in the generalized Wishart model for observed distance matrices, making it an appealing choice for modeling spatial structure when the observations come as pairwise distances. This link between circuit theory and ICAR spatial models provides a computationally efficient framework for modeling circuits with Gaussian error and making inference on resistance values in the circuit based on observed data. Circuits and GMRFs both have a natural graph structure, but to the best of our knowledge this link between GMRFs with ICAR structure and circuits has not been described in the literature.

Our article is organized as follows. In Section 2, Graphs, Circuits, and Resistance, we review graph and circuit theory. We then review how circuit theory is used to model gene flow in Section 3, Landscape Genetics. In Section 4, Distance and Covariance, we review the relationship between a covariance matrix and the mean squared distance between observations. We also review how a symmetric positive semidefinite (SPSD) matrix can be used to calculate the resistance distance of a graph, and discuss the generalized Wishart model for observed squared distance matrices. In Section 5, Gaussian Markov Random Fields for Circuits, we show that an ICAR covariance induces the circuit theory resistance distance on a graph. This provides a framework for modeling observations with connectivity defined by circuit theory. We also discuss modeling in the case of partial or repeated observation of the nodes in the circuit. In Section 6, Application, we illustrate our approach through a simulation example and then use our approach to examine gene flow in alpine chamois (*Rupicapra rupicapra*) in the Bauges mountains of France. Finally, in Section 7, Discussion, we discuss possible extensions to our approach.

2. GRAPHS, CIRCUITS, AND RESISTANCE

An undirected graph (\mathbf{G}, \mathbf{A}) is a collection of m “nodes” $\mathbf{G} = \{G_1, G_2, \dots, G_m\}$ and the “edges” or “edge weights” $\mathbf{A} = \{\alpha_{ij}\}$ that are the connections between the nodes. If nodes i and j are first-order neighbors, denoted $i \sim j$, then $\alpha_{ij} > 0$ is a measure of their connectivity, with larger values leading to stronger connectivity between nodes. If nodes i and j are not first-order neighbors ($i \not\sim j$), then $\alpha_{ij} = 0$. The graph is said to be undirected if the edge weights are symmetric: $\alpha_{ij} = \alpha_{ji}$.

An electrical circuit can be represented by an undirected graph where the nodes \mathbf{G} are connected by a set of resistors \mathbf{A} . Ohm’s law expresses the relationship between the resistance (R) of a resistor and the current (I) that flows through the resistor when a voltage (V) is applied as $V = I \cdot R$. The conductance α_{ij} of the resistor connecting neighboring nodes i and j is the current I that would flow through the resistor if it were removed from the circuit and a one-volt charge ($V = 1$) was applied across the resistor. The resistance R of the resistor is the inverse of the conductance: $1/\alpha_{ij}$.

A voltage applied across any two nodes i and j , which may or may not be directly connected (first-order neighbors) in the circuit, results in a current flow through the circuit. The effective resistance Γ_{ij} between nodes i and j is then defined as the resistance of a single resistor that would admit the same current flow if the voltage were applied across it (e.g., Dorf and Svoboda 2004).

The effective resistance Γ_{ij} is a distance metric on (\mathbf{G}, \mathbf{A}) (Klein and Randić 1993) that incorporates all possible pathways through the graph, where distance decreases with the addition of new pathways throughout the graph.

3. LANDSCAPE GENETICS

Landscape genetics is the study of the effects of landscape on genetic diversity (Manel et al. 2003). Understanding how the landscape affects the genetic connectivity of a population of organisms can provide insight into the proper management of the population, the potential spread of disease through the population (e.g., Wheeler, Waller, and Biek 2010; Wheeler and Waller 2010), and the effects of climate or landscape change on population connectivity (e.g., Spear et al. 2010). Linking genetic

information (e.g., microsatellite data) to the landscape using a formal statistical model has been a continual challenge in the field of landscape genetics (e.g., Storfer et al. 2007; Spear et al. 2010). Spatial clustering methods (e.g., Guillot et al. 2005; Chen et al. 2007; Durand et al. 2009) are commonly used to identify genetic boundaries, but are typically based on the assumption of a homogeneous underlying landscape, with any link to the underlying landscape made post hoc (e.g., Wheeler, Waller, and Biek 2010). Approaches based on isolation by distance (Wright 1943; Spear et al. 2005; Broquet et al. 2006) hypothesize that gene flow is based on a random walk in a homogeneous landscape, and thus that a genetic distance metric computed from observed genetic information is correlated with the straight-line distance between observed locations.

As an alternative to straight-line distance, various effective distances, including the resistance distance from circuit theory, have been used to reflect the varying permeability of a heterogeneous landscape to gene flow. In these cases, the landscape is viewed as an undirected graph with raster cells $\{G_1, G_2, \dots, G_m\}$ for nodes and edge weights $\{\alpha_{ij}, i = 1, \dots, m, j = 1, \dots, m\}$ defined by the landscape characteristics of the raster cells. McRae (2006) showed that there are theoretical links between the resistance distance of the circuit representing the heterogeneous landscape and the genetic distance between animal populations. The fixation index F_{st} is a measure of genetic dissimilarity between subpopulations based on comparing the diversity of randomly chosen alleles within each subpopulation to the diversity found within the entire population (Holsinger and Weir 2009). If we assume that each node in the graph (e.g., each raster cell G_i on a landscape) contains a sub-population, then pairwise F_{st} values can be computed for each pair of nodes. The major finding of McRae (2006) is that under a random walk model of animal movement and gene flow on an undirected graph (\mathbf{G}, \mathbf{A}) , the resistance distance of an analogous circuit is proportional to linearized F_{st} . That is, if F_{st}^{ij} is the pairwise fixation index between the i th and j th node in \mathbf{G} , then the resistance distance Γ_{ij} between those two nodes is

$$\Gamma_{ij} \propto \frac{F_{st}^{ij}}{1 - F_{st}^{ij}} \quad (1)$$

(see McRae 2006 for details). Thus, resistance distance is proportional to a common formulation of genetic distance between subpopulations. This relationship between gene flow and circuit theory is known as the “isolation by resistance” hypothesis (Cushman et al. 2006; McRae 2006; McRae and Beier 2007; McRae et al. 2008; Cushman and Landguth 2010; Spear et al. 2010).

While the circuit theory model of gene flow (McRae 2006) is formulated at a population level, with the assumption that each grid cell on a landscape contains a subpopulation, it is often applied to individual-level genetic data (e.g., Cushman et al. 2006; Cushman and Landguth 2010). In effect, the observed genetic information of an individual is viewed as a realization of genetic information in the theoretical subpopulation in the grid cell. Pairwise genetic distances between individuals can be thought of as observations of the resistance distance

between the grid cells where the individuals were observed, with noise.

As an alternative to the resistance distance, some landscape genetic studies compare the “least-cost-path” (LCP) distance to the genetic distance between populations. In an LCP analysis, the effective distance between two nodes is the smallest possible resistance of a single path through the graph that connects the two nodes. Thus, using LCP distance to model animal movement and gene flow is based on the assumption that animals know the landscape fully and are able to pick the “best” path between any two locations of their choosing. An analysis based on resistance distance, instead of LCP distance, assumes that the effective distance between two locations is based on a weighted average of all possible paths through the graph that connect the two locations. Using resistance distance to model animal movement and gene flow is based on the assumption that animals move through the landscape at random, with movement choices being based only on the local environment, as reflected in the connectivity of the graph. At small temporal scales, movement is affected by memory, interactions with other animals, and many other factors not directly related to the local environment. At large temporal scales (i.e., genetic time), the assumption that the environment drives population-level movement and gene flow may be more plausible, lending credence to the use of resistance distance. Studies of observed genetic data have found that the circuit-based resistance distance is generally better correlated with observed genetic distance than is LCP distance (e.g., McRae and Beier 2007). However, the LCP distance can be computed much more efficiently than the resistance distance, allowing analysis of larger landscapes at finer resolutions.

Inference on resistance values of landscape characteristics is typically achieved using a correlation analysis. An observed genetic distance matrix is compared with a number of possible resistance distance matrices that result from various hypothesized resistance values for landscape characteristics. The hypothesized resistance values that result in the highest correlation with the observed genetic distance are selected.

We seek to improve the main features of this approach in two ways. First, little or nothing is learned through this approach about the uncertainty pertaining to the estimated resistance values. Second, the process of estimating the resistance values requires a grid search of the parameter space, and consequently a large number of candidate resistance values may need to be examined. For example, if the goal is to estimate the resistance to spatial gene flow due to p distinct landscape characteristics, a grid of hypothetical resistance values for each characteristic would need to be specified. If we consider a modest 10 hypothetical values for each resistance value, we would need to calculate the resistance distance and resulting correlation with observed data for each of the 10^p combinations of hypothesized values. The number of hypothetical values could grow much higher, depending on the degree of precision required.

Our goal is to put the estimation of resistance values from observed genetic distance matrices within a model-based framework. After developing a suitable model, Bayesian statistical methods will allow for estimation of the posterior distribution of the resistance values of landscape characteristics, providing

point estimates through posterior means and estimates of uncertainty through posterior credible intervals.

4. DISTANCE AND COVARIANCE

Covariance matrices and distance matrices both contain information about the connectivity between locations indexed by the matrix rows and columns. In this section, we review a link between covariance matrices and distance metrics, show how an improper (rank-deficient) covariance matrix can be used to induce the resistance distance $\mathbf{\Gamma}$ on the graph (\mathbf{G}, \mathbf{A}) , and review recent developments in the modeling of observed distance matrices.

4.1 Induced Distance Matrices

We will refer to the space of symmetric positive semi-definite (SPSD) matrices of dimension $m \times m$ as $\text{SPSD}(m)$. Let $\mathbf{G} = \{G_1, \dots, G_m\}$ be a set of m locations (e.g., nodes in a graph) and let $\mathbf{\Sigma} \in \text{SPSD}(m)$ be an SPSPD matrix with components (Σ_{ij}) , $i, j \in \{1, 2, \dots, m\}$. It is helpful to think of Σ_{ij} as the covariance between the i th and j th locations in \mathbf{G} . Also, let \mathbf{e}_j be the $m \times 1$ column vector with the j th element equal to 1 and all other elements equal to 0. The metric defined by

$$D_{ij} = \Delta_{ij}(\mathbf{\Sigma}) = (\mathbf{e}_i - \mathbf{e}_j)' \mathbf{\Sigma} (\mathbf{e}_i - \mathbf{e}_j) \quad (2)$$

$$= \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}$$

is a critical transformation in both circuit theory and in the modeling of observed distance matrices. In Section 4.2, we will show how a particular matrix $\mathbf{\Sigma}$ results in D_{ij} being the resistance distance of the graph. In Section 4.3, we will follow McCullagh (2009) and use the transformation (2) to formalize the modeling of observed distance matrices. We first list some properties of (2) that will be important in our effort to model observed genetic distance matrices using covariance matrices based on circuit theory.

If we collect the $\{D_{ij}\}$ in an $m \times m$ matrix \mathbf{D} , the transformation (2) can be written in matrix form. Let $\mathbf{d}_{\Sigma} = \text{diag}(\mathbf{\Sigma})$ be a column vector with elements equal to the diagonal elements of $\mathbf{\Sigma}$, and $\mathbf{1}$ be a column vector with all elements equal to 1. Then

$$\mathbf{D} = \Delta(\mathbf{\Sigma}) = -2\mathbf{\Sigma} + \mathbf{d}_{\Sigma} \mathbf{1}' + \mathbf{1} \mathbf{d}_{\Sigma}' \quad (3)$$

is a linear transformation from $\text{SPSD}(m)$ to the space of $m \times m$ matrices that are negative definite on contrasts. We will say that the covariance matrix $\mathbf{\Sigma}$ induces the distance matrix $\mathbf{D} = \Delta(\mathbf{\Sigma})$ on the set of locations \mathbf{G} .

This transformation Δ has an intuitive interpretation in the case where $\mathbf{\Sigma}$ is a covariance matrix. Let \mathbf{y} be an m -dimensional random variable with common mean $\boldsymbol{\mu}$ (i.e., $\mu_i = \mu$, $i = 1, 2, \dots, m$) and covariance matrix $\mathbf{\Sigma}$. Then the induced distance $\Delta_{ij}(\mathbf{\Sigma})$ is a variogram: the expected squared distance between the observations at locations i and j :

$$E[(y_i - y_j)^2] = \text{var}(y_i) + \text{var}(y_j) - 2\text{cov}(y_i, y_j)$$

$$= \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}$$

$$= \Delta_{ij}(\mathbf{\Sigma}).$$

While this interpretation only holds when $\mathbf{\Sigma}$ is a covariance matrix, the induced matrix $\Delta(\mathbf{\Sigma})$ is a distance for a class of rank-deficient SPSPD matrices. To see this, we first examine the null space or kernel of the operator Δ , which is the set of

$\mathbf{\Sigma} \in \text{SPSD}(m)$ such that $\Delta(\mathbf{\Sigma})$ is the zero matrix. McCullagh (2009) noted that the null space of Δ is the space of additive symmetric matrices $\mathcal{S} = \{\mathbf{S}_{m \times m} : S_{i,j} = v_i + v_j, \mathbf{v} \in \mathbb{R}^m\}$. We clarify this statement in the case where $\mathbf{\Sigma} \in \text{SPSD}(m)$.

Proposition 1. For $\mathbf{\Sigma} \in \text{SPSD}(m)$, the transformation $\Delta(\mathbf{\Sigma})$ is a linear transformation with null space equal to $\{\mathbf{\Sigma} : \mathbf{\Sigma} = c\mathbf{1}\mathbf{1}', c \in \mathbb{R}\}$. (Proof given in Appendix A.)

We will denote the null (column) space of a matrix $\mathbf{\Sigma}$ as $\mathcal{N}(\mathbf{\Sigma})$. If $\mathbf{\Sigma}$ is of full rank, or if $\mathcal{N}(\mathbf{\Sigma})$ is the space spanned by the $\mathbf{1}$ vector, then $\Delta(\mathbf{\Sigma})$ is a distance metric on $\mathbf{G} \times \mathbf{G}$. That is, $\mathbf{D} = \Delta(\mathbf{\Sigma})$ is symmetric ($D_{ij} = D_{ji}$), satisfies the triangle inequality ($D_{ij} + D_{jk} \geq D_{ik}$), and is zero only for the distance from a location to itself ($D_{ij} = 0$ iff $i = j$).

If $\mathcal{N}(\mathbf{\Sigma})$ is the space spanned by the $\mathbf{1}$ vector, then $\mathbf{\Sigma}$ is orthogonal to the null space of Δ and the transformation Δ is invertible. Note that this invertibility does not hold for any $\mathbf{\Sigma}$ of full rank. This will be important in Section 4.3 where we discuss the identifiability of covariance matrices in the generalized Wishart model for distance matrices induced by the transformation Δ in (3).

4.2 Inducing the Resistance Distance of a Graph

The effective resistance matrix $\mathbf{\Gamma} = (\Gamma_{ij})$ of a circuit represented by the undirected graph (\mathbf{G}, \mathbf{A}) can be computed by constructing an SPSPD matrix that induces $\mathbf{\Gamma}$. Kirchoff's current law (e.g., Dorf and Svoboda 2004) states that the electrical current flowing into a node must equal the current flowing out of the node. Applying Kirchoff's current law to the graph (\mathbf{G}, \mathbf{A}) leads to the formation of the Laplacian matrix (e.g., Babić et al. 2002) of the graph:

$$\mathbf{Q} = \begin{bmatrix} \sum_{j \neq 1} \alpha_{1j} & -\alpha_{12} & -\alpha_{13} & \cdots \\ -\alpha_{21} & \sum_{j \neq 2} \alpha_{2j} & -\alpha_{23} & \cdots \\ -\alpha_{31} & -\alpha_{32} & \sum_{j \neq 3} \alpha_{3j} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (4)$$

The Laplacian matrix \mathbf{Q} is of dimension $m \times m$, and each row sums to zero. The null space $\mathcal{N}(\mathbf{Q})$ is the space spanned by the $\mathbf{1}$ vector, and \mathbf{Q} is singular with rank $m - 1$. The generalized inverse of the Laplacian \mathbf{Q} is \mathbf{Q}^- , which is also of rank $m - 1$. As shown by Klein and Randić (1993), the effective resistance between the nodes of (\mathbf{G}, \mathbf{A}) is the distance induced by the SPSPD matrix \mathbf{Q}^- :

$$\Gamma_{ij} = (\mathbf{e}_i - \mathbf{e}_j)' \mathbf{Q}^- (\mathbf{e}_i - \mathbf{e}_j). \quad (5)$$

Generalized inverses are not unique, but the correspondence between the null space of the Laplacian matrix \mathbf{Q} and the null space of the transformation (2) results in a useful invariance.

Proposition 2. The resistance distance $\Gamma_{ij} = (\mathbf{e}_i - \mathbf{e}_j)' \mathbf{Q}^- (\mathbf{e}_i - \mathbf{e}_j)$ is invariant to the choice of generalized inverse \mathbf{Q}^- of the Laplacian matrix \mathbf{Q} . (Proof in Appendix A.)

This invariance was not noted by Klein and Randić (1993), though various generalized inverses have been used to compute the resistance distance (e.g., Babić et al. 2002; McRae 2006).

4.3 Modeling-Induced Distance Matrices

The observations in our landscape genetics example are genetic distance matrices, which we are viewing as distance matrices induced from an underlying spatial covariance motivated by circuit theory. The distributional properties of distance matrices induced by the transformation (3) have been studied by McCullagh (2009). We briefly review the results of McCullagh (2009) here, and then discuss identifiability in models for observed distance matrices.

Let $\mathbf{y}_i \sim N(\mu\mathbf{1}, \Sigma)$, $i = 1, 2, \dots, \nu$ be ν independent realizations from a Gaussian process with common mean μ and covariance matrix Σ . If we concatenate the observations as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_\nu]$, then $\mathbf{S} = \mathbf{Y}\mathbf{Y}'$ is Wishart distributed with ν degrees of freedom and covariance matrix Σ :

$$\mathbf{S} = \mathbf{Y}\mathbf{Y}' \sim \mathcal{W}_\nu(\Sigma).$$

McCullagh (2009) defined the generalized Wishart distribution as follows. \mathbf{V} is said to come from a generalized Wishart distribution with ν degrees of freedom, covariance matrix Σ , and kernel \mathcal{K} :

$$\mathbf{V} \sim \mathcal{GW}_\nu(\mathcal{K}, \Sigma),$$

if for any linear transformation \mathbf{L} such that the null space of \mathbf{L} is \mathcal{K} ($\mathcal{N}(\mathbf{L}) = \mathcal{K}$), $\mathbf{L}'\mathbf{V}\mathbf{L}$ is Wishart distributed:

$$\mathbf{L}'\mathbf{V}\mathbf{L} \sim \mathcal{W}_\nu(\mathbf{L}'\Sigma\mathbf{L}).$$

McCullagh (2009) then showed that induced distance matrices (3) can be modeled using the generalized Wishart distribution. Let $\mathbf{S} \sim \mathcal{W}_\nu(\Sigma)$ and consider the induced distance matrix $\mathbf{D} = \Delta(\mathbf{S})$. Consider an $m \times (m - 1)$ matrix \mathbf{L} with full column rank and $\mathcal{N}(\mathbf{L})$ spanned by $\mathbf{1}$. For example, \mathbf{L} could be a full-rank matrix of contrast vectors:

$$\mathbf{L} = [\mathbf{e}_2 - \mathbf{e}_1, \mathbf{e}_3 - \mathbf{e}_1, \dots, \mathbf{e}_m - \mathbf{e}_1].$$

Then $\mathbf{L}'\mathbf{1} = \mathbf{0}$ and it is easy to see from (3) that

$$\begin{aligned} \mathbf{L}'(-\mathbf{D})\mathbf{L} &= \mathbf{L}'(2\mathbf{S} - \mathbf{d}_\Sigma\mathbf{1}' - \mathbf{1}\mathbf{d}'_\Sigma)\mathbf{L} \\ &= \mathbf{L}'(2\mathbf{S})\mathbf{L}, \end{aligned}$$

which is Wishart distributed with covariance matrix $\mathbf{L}'(2\Sigma)\mathbf{L}$. Then, by definition,

$$-\mathbf{D} = -\Delta(\mathbf{S}) \sim \mathcal{GW}_\nu(\mathbf{1}, 2\Sigma). \tag{6}$$

The likelihood of an observed distance matrix \mathbf{D} under this model can be obtained by evaluating the likelihood $[\mathbf{L}'(-\mathbf{D})\mathbf{L} \mid \nu, \mathbf{L}'(2\Sigma)\mathbf{L}]$ in the Wishart distribution $\mathbf{L}'(-\mathbf{D})\mathbf{L} \sim \mathcal{W}_\nu(\mathbf{L}'(2\Sigma)\mathbf{L})$, where the bracket “[.]” notation indicates a probability distribution. Alternately, McCullagh showed that this likelihood can be obtained equivalently by using a projection $\mathbf{A} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$ onto the space of $m \times m$ matrices orthogonal to the null space of the transformation Δ . The resulting expression for the likelihood does not rely on the choice of \mathbf{L} :

$$[-\mathbf{D} \mid 2\Sigma, \nu] \propto |\Sigma^{-1}\mathbf{A}|^{\nu/2} \cdot \exp\left\{\frac{1}{4}\text{tr}(\Sigma^{-1}\mathbf{A}\mathbf{D})\right\}, \tag{7}$$

where the determinant $|\cdot|$ is the product of all nonzero eigenvalues.

As the transformation $\Delta(\mathbf{S})$ is invariant to the addition of a constant matrix $c\mathbf{1}\mathbf{1}'$ to \mathbf{S} , the likelihood (7) is invariant to the addition of a constant matrix $c\mathbf{1}\mathbf{1}'$ to Σ , and Σ is not identifiable in general. While McCullagh (2009) was not concerned with identifiability (see McCullagh 2009, p. 635), our goal is to parameterize and estimate a spatial covariance matrix Σ that describes gene flow under a circuit theory model for connectivity. Identifiability of Σ is important in this endeavor.

The fact that the generalized Wishart distribution is defined on distance matrices induced by applying the transformation (3) to observations from a Wishart distribution implies that the covariance matrix Σ in the generalized Wishart model (6) will be identifiable if the transformation (3) is invertible. The following proposition follows directly from Proposition 1, where we describe the null space of the transformation (3).

Proposition 3. In the generalized Wishart model (6), Σ is identifiable if and only if the space spanned by $\mathbf{1}$ is contained in $\mathcal{N}(\Sigma)$.

This implies that only a rank-deficient covariance matrix Σ will be identifiable. We also note that the deficiency implied here is identical to the null space of the Laplacian matrix (4) of the graph under circuit theory. We now specify a GMRF with a covariance matrix that will be identifiable in the generalized Wishart model for distance matrices and that will induce the resistance distance on the graph, matching the second-order structure defined by circuit theory.

5. GAUSSIAN MARKOV RANDOM FIELDS FOR CIRCUITS

GMRFs are used extensively in statistical modeling, especially for spatial and temporal processes (Rue and Held 2005; Lindström and Lindgren 2008; Bolin et al. 2009). GMRFs on a graph \mathbf{G} are characterized by the conditional independence of any two nodes that are not first-order neighbors when conditioned on all other nodes in \mathbf{G} . This conditional independence results in a sparse precision matrix, and sparse matrix methods can typically be applied to any statistical analysis using GMRFs (Rue 2001; Rue and Held 2005; Rue, Martino, and Chopin 2009). This allows for highly efficient computation, and can make inference possible at a finer spatial or temporal resolution than might be possible using a non-Markovian random field.

Our goal is to specify a GMRF on the graph (\mathbf{G}, \mathbf{A}) with a covariance matrix that induces the resistance distance matrix $\mathbf{\Gamma}$. The result is a statistical model for observations from a graph with Gaussian error and expected squared distance between observations equal to the resistance distance of the graph.

Consider a random variable $\mathbf{y} = [y_1, y_2, \dots, y_m]'$ on the m nodes $\mathbf{G} = \{G_1, G_2, \dots, G_m\}$ and define the conditional distribution of y_i given all other values in \mathbf{y} as

$$y_i \mid \mathbf{y}_{-i} \sim N\left(\sum_{j \sim i} w_{ij} y_j, \kappa_i\right) \tag{8}$$

$$w_{ij} = \frac{\alpha_{ij}}{\sum_{l \sim i} \alpha_{il}}, \quad \kappa_i = \frac{1}{\sum_{l \sim i} \alpha_{il}}.$$

Conditionally, each observation is normally distributed with mean equal to a weighted average of all nodes that are first-order neighbors $j \sim i$, where the weights are proportional to the conductance α_{ij} of the edge connecting the nodes.

The conditional specification in (8) satisfies the Markov property and is a GMRF. In particular, \mathbf{y} follows an intrinsic Gaussian conditional autoregressive (ICAR) model on the graph (\mathbf{G}, \mathbf{A}) (Besag and Kooperberg 1995; Rue and Held 2005). Using Brooks Lemma (Besag 1974; Besag and Kooperberg 1995), the joint distribution of \mathbf{y} is

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{Q}^-), \quad (9)$$

where the precision matrix \mathbf{Q} is exactly the Laplacian matrix (4) of the graph (\mathbf{G}, \mathbf{A}) . Then the generalized inverse \mathbf{Q}^- is an SPSD matrix that induces the resistance distance Γ on the graph (\mathbf{G}, \mathbf{A}) , and the mean squared distance between observations from the ICAR process (9) is the resistance distance of the graph.

The ICAR model (9) is proper under the constraint that $\mathbf{y}'\mathbf{1} = c$ for some constant $c \in \mathbf{R}$. If we observe v independent realizations $\mathbf{y}_i \sim N(\mathbf{0}, \mathbf{Q}^-)$, $i = 1, 2, \dots, v$, with $\mathbf{y}'_i\mathbf{1} = 0$, and concatenate the observations as in Section 4.3: $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_v]$, then $\mathbf{S} = \mathbf{Y}\mathbf{Y}' \sim \mathcal{W}_v(\mathbf{Q}^-)$ is proper under the constraint that $\mathbf{S}\mathbf{1} = \mathbf{0}$. Under this constraint, \mathbf{S} is orthogonal to the null space of the transformation Δ (3), and the transformation to $\mathbf{D} = \Delta(\mathbf{S})$ is invertible. The elements of \mathbf{D} are the sum of the squared difference in observations $D_{jk} = \sum_{i=1}^v (y_{ji} - y_{ki})^2$ subject to the constraints that $\mathbf{y}'_i\mathbf{1} = 0$. Under this model, each element D_{jk} of the distance matrix \mathbf{D} has mean $v\Gamma_{jk}$, the resistance distance Γ_{jk} between the j th and k th nodes in the graph \mathbf{G} multiplied by the number of pairwise observations v .

This provides a natural model for observed distance matrices with connectivity modeled by circuit theory:

$$-\mathbf{D} \sim \mathcal{GW}_v(\mathbf{1}, 2\mathbf{Q}^-), \quad (10)$$

where \mathbf{Q} is the Laplacian of the graph (\mathbf{G}, \mathbf{A}) , or, equivalently the ICAR precision matrix of the graph (\mathbf{G}, \mathbf{A}) . From Proposition 3, \mathbf{Q} is identifiable, and the observations in \mathbf{D} can be thought of as observations of the resistance distance, with Gaussian noise.

5.1 Partial and Repeated Observation

We have assumed so far that the circuit in question is fully observed. That is, an observed effective distance D_{ij} is obtained for each pair of nodes in \mathbf{G} . In practice, however, it will be common to have observations from only a small subset of the nodes, and multiple observations may come from a single node. In a spatial analysis, for example, the study domain may include thousands of raster cells, each of which is considered a node in the resistance surface (\mathbf{G}, \mathbf{A}) . If we obtain a set of spatially referenced observations, multiple observations may lie within one raster cell, and, depending on the number of observations in the study, only a small fraction of the raster cells (perhaps dozens or hundreds) will contain observations. In the case where only a subset of the nodes is observed, we want to preserve the sparse nature of the precision matrix \mathbf{Q} and the associated computational efficiency of the GMRF.

The precision matrix \mathbf{Q} is sparse, but the generalized inverse \mathbf{Q}^- will typically be dense. For large m , it is impossible to even store a dense $m \times m$ matrix in many standard computing

environments, so we wish to avoid the express calculation of \mathbf{Q}^- for the whole graph.

Consider the case where we have only n observations from the m nodes in \mathbf{G} , with $m \gg n$. While we will relax this assumption later, let us initially assume that the observations arise from distinct nodes with no two observations from the same node. Without loss of generality, let the observed nodes be indexed by the first n rows and columns of \mathbf{Q} . The goal is to compute the $n \times n$ precision matrix Φ of the observed nodes. Since the generalized Wishart distribution is invariant to the choice of generalized inverse of \mathbf{Q} (Proposition 2), we can pick a convenient form for the generalized inverse by computing the inverse of the full-rank $(m-1) \times (m-1)$ upper-left submatrix of \mathbf{Q} and setting all entries in the m th row and column equal to zero. In this formulation, the precision matrix Φ of the observed nodes can be obtained through partitioning the generalized inverse in block form and computing a Schur complement (e.g., Harville 2008). The advantage of this method is that we can use sparse matrix methods that are computationally efficient and do not require the explicit computation of large dense matrices (Rue 2001). Details of this computation are given in Appendix B.

In the case where multiple observations are obtained from a single node of \mathbf{G} , a nugget effect can be introduced to the covariance structure (Besag, York, and Mollié 1991; Lindström and Lindgren 2008; Cross et al. 2010). Let $\mathbf{y} = [y_1, y_1, \dots, y_{n_{\text{obs}}}]^T$ be the vector of n_{obs} observations from the graph (\mathbf{G}, \mathbf{A}) , and let $s_i \in G$ be the node of the i th observation, $i = 1, 2, \dots, n_{\text{obs}}$. Since there may be repeated observations in some nodes, the total number of nodes of \mathbf{G} for which we have observations is $n_{\text{nodes}} \leq n_{\text{obs}}$. Let $\Sigma_{\text{nodes}} = \Phi^{-1}$ be the $n_{\text{nodes}} \times n_{\text{nodes}}$ covariance matrix of the observed nodes, and let \mathbf{K} be an $n_{\text{obs}} \times n_{\text{nodes}}$ matrix with $K_{ij} = 1$ if $s_i = j$, that is, if the i th observation y_i comes from the j th node of \mathbf{G} , and $K_{ij} = 0$ otherwise. Then the covariance Ψ of the n_{obs} observations is

$$\Psi = \mathbf{K}\Sigma_{\text{nodes}}\mathbf{K}^T + \tau\mathbf{I},$$

where connectivity between the nodes of \mathbf{G} is modeled based on circuit theory (Σ_{nodes}) and τ is a spatial nugget parameter representing the variability in observations obtained from the same node in \mathbf{G} .

We can then model an observed distance matrix \mathbf{D} as arising from a generalized Wishart distribution with covariance matrix of the observed locations Ψ ,

$$-\mathbf{D} \sim \mathcal{GW}(\mathbf{1}, 2\Psi). \quad (11)$$

The covariance matrix Ψ is dependent on the edge weights $\{\alpha_{ij}\}$, and may contain a nugget effect τ to account for repeated observations.

5.2 Modeling Resistance

We now turn to the parameterization of the edge weights $\{\alpha_{ij}\}$ of the graph (\mathbf{G}, \mathbf{A}) in terms of landscape covariates. In doing so, it will be helpful to consider a few links between circuit theory and random walks on the graph. In a circuit, the edge weights are the conductances between neighboring nodes in the graph. If we consider a continuous-time random walk on the graph (\mathbf{G}, \mathbf{A}) , the edge weights $\{\alpha_{ij}\}$ are proportional to the rate at which a random walker transitions from node i to node j . As landscape covariates are typically available in gridded form, each node in \mathbf{G}

is a grid cell in the study area. The rate of transition from cell i to cell j could be driven by many factors, including the environment at the starting cell i , the environment at the neighboring cell j , and local environmental gradients (Hooten et al. 2010). To maintain the links to circuit theory, however, the transition rates must be symmetric ($\alpha_{ij} = \alpha_{ji}$) and nonnegative. This ensures that (\mathbf{G}, \mathbf{A}) is an undirected graph, and also ensures that the conditional specification (8) of the ICAR model results in a GMRF. As the edge weights must be symmetric, we cannot model drift due to migration or other directional factors using circuit theory. Forcing the edge weights to be symmetric implies that we are modeling connectivity of a system that is stationary. While this assumption may be unreasonable for individual organisms at short time scales (e.g., Hanks et al. 2011), there is more support for it at a population level and at long time scales (e.g., genetic time).

To specify symmetric edge weights, we consider an average of the landscape covariates at adjacent cells. This is motivated by considering the straight-line path between the centers of the two adjacent grid cells. Half of this path lies in each cell, and we can hypothesize a transition rate (or edge weight) that would be related to an average of the landscape covariates in the two cells. If \mathbf{x}_i is a vector of p landscape characteristics of the i th cell, and $\boldsymbol{\beta}$ is a p -vector of parameters related to the effect that each landscape characteristic has on gene flow, then the edge weights $\{\alpha_{ij}\}$ could be modeled as

$$\alpha_{ij} = \begin{cases} \exp \left[\frac{1}{d_{ij}} \left(\frac{\mathbf{x}'_i + \mathbf{x}'_j}{2} \right) \boldsymbol{\beta} \right], & j \in \mathcal{N}(i) \\ 0, & j \notin \mathcal{N}(i), \end{cases} \quad (12)$$

where d_{ij} is the straight-line distance between the centroids of cells i and j . In this model, connectivity is a function of the landscape characteristics of both cells ($\mathbf{x}_i, \mathbf{x}_j$), the distance between the cells (allowing for differing distances between neighbors in, for example, a queen's neighborhood), and the conductance parameters $\boldsymbol{\beta}$. Using an exponential link function ensures that the $\{\alpha_{ij}\}$ are nonnegative, and allows for convenient interpretation of the conductance parameters $\boldsymbol{\beta}$. If $\beta_h < 0$, then an increase in the h th landscape characteristic results in a greater resistance to gene flow, while if $\beta_h > 0$, then an increase in the h th landscape characteristic results in less resistance to gene flow.

Finally, assigning prior distributions to the coefficients $\boldsymbol{\beta}$ and nugget parameter τ ,

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (13)$$

$$\tau \sim \text{IG}(a, b), \quad (14)$$

results in a statistical model (11)–(14) that can be fit using Bayesian techniques (e.g., Markov chain Monte Carlo, MCMC). In both the simulation example and data analysis that follow, we have specified a diffuse, mean-zero prior for $\boldsymbol{\beta}$: $\boldsymbol{\beta} \sim N(\mathbf{0}, 10^6 \mathbf{I})$ and a relatively diffuse inverse-gamma prior for τ : $\tau \sim \text{IG}(10, 100)$. Samples from the posterior distribution of $\boldsymbol{\beta}$ and τ can be obtained using random walk Metropolis–Hastings updates in an MCMC sampler. After proposing a value for $\boldsymbol{\beta}$ or τ , the proposed value is used to compute the resulting precision matrix $\boldsymbol{\Psi}$ of the observations. The candidate $\boldsymbol{\beta}$ or τ value can then be accepted or rejected with the typical Metropolis–Hastings probability (e.g., Gelman et al. 2004).

One potentially useful extension to (12) is a varying-coefficient model (Hastie and Tibshirani 1993). For simplicity, consider one covariate x . The edge weights for neighboring cells could be modeled as

$$\alpha_{ij} = \exp \left[\frac{1}{d_{ij}} \left(\frac{\beta(x_i) + \beta(x_j)}{2} \right) \right],$$

where the functional regressor $\beta(x)$ is typically modeled as a linear combination of n_{spl} spline basis functions $\{\phi_k, k = 1, \dots, n_{\text{spl}}\}$: $\beta(x) = \sum_{k=1}^{n_{\text{spl}}} \gamma_k \phi_k(x)$. This specification gives a flexible framework for examining nonlinear relationships between conductance and the covariate x .

6. APPLICATION

We now apply our approach, first in a study of simulated data, and then to observed genetic data.

6.1 Simulation Example

Our approach for inference on gene flow across a heterogeneous landscape can be illustrated through a simulation example. A 100×100 landscape grid was created with a mix of continuous and discrete landscape characteristics (Figure 2). True values were assigned to the parameters $\boldsymbol{\beta}$ pertaining to the effect that each landscape characteristic has on gene flow, as well as for the nugget parameter τ . To simulate locations of organisms on the landscape, 300 completely spatial random locations were chosen and linked to grid cells. The covariance matrix $\boldsymbol{\Psi}_{\text{sim}}$ of these 300 locations was based on a rook's-neighborhood connectivity between grid cells (i.e., first-order spatial neighbors) and the conductances $\{\alpha_{ij}\}$ specified by the landscape characteristics and true $\boldsymbol{\beta}$ values. We simulated 10 realizations from a mean-zero Gaussian random field with covariance $\boldsymbol{\Psi}_{\text{sim}}$, and calculated the squared-Euclidean distance between these realizations. The resulting pairwise distances between locations are proportional to the linearized pairwise F_{ST} values that would be observed from the analysis of microsatellite alleles from 10 loci of each organism.

The simulated pairwise distance matrix was used to fit the model (11)–(14). We used an MCMC algorithm to obtain 10,000 posterior samples of $\boldsymbol{\beta}$ and τ from four separate chains with distinct starting values. The posterior distributions for $\boldsymbol{\beta}$ and τ both indicate agreement with the true specified values (Table 1), providing evidence that the proposed model is able to recover resistance parameters in similar situations to that of our application data discussed in the following section.

Table 1. Simulation example results for connectivity parameters in the resistance surface

Parameter	True value	Posterior mean	Lower 95% CI bound	Upper 95% CI bound
Intercept	2	1.364	0.744	2.015
Covariate 1	−5	−4.127	−6.340	−1.939
Covariate 2	0	0.040	−0.971	1.067
Covariate 3	1	1.386	0.815	1.901
Nugget	0.5	0.519	0.474	0.621

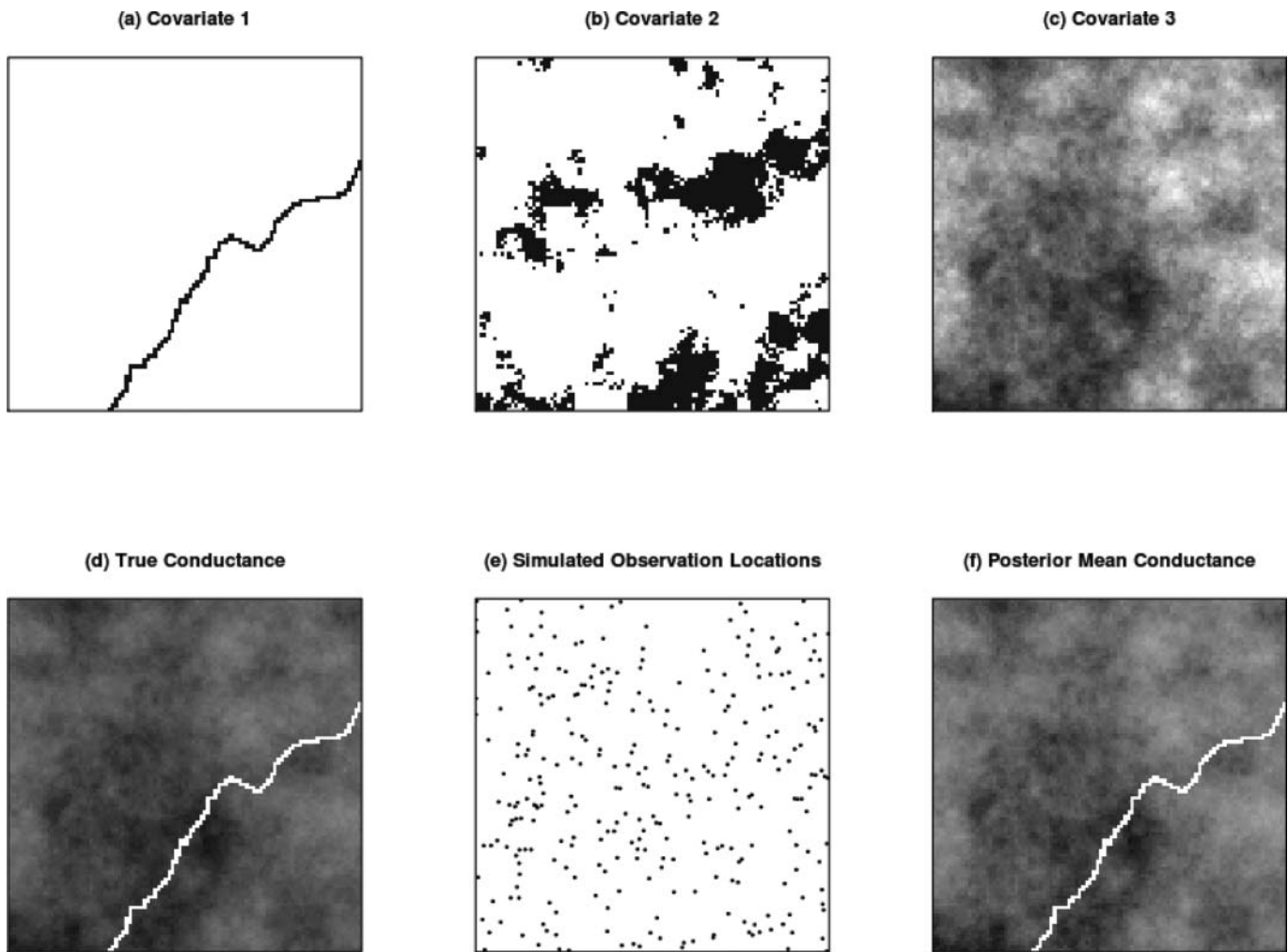


Figure 2. We simulated a resistance surface using two discrete cover types (a) and (b) and one continuous covariate (c). In the simulation study, Covariate 1 was set to impede connectivity (a), Covariate 2 was set to have no effect on connectivity (b), and Covariate 3 was set to facilitate connectivity (c). The resulting landscape conductance is shown in (d). Observations from the landscape were simulated (e), and inference on the contribution of each landscape covariate to connectivity was made (f).

For comparison, we also employed a correlation analysis, the most common existing method in landscape genetics studies (e.g., Cushman et al. 2006; Cushman, McKelvey, and Schwartz 2009; Wang and Xia 2009; Shirk et al. 2010). This method requires specifying hypothesized resistance values, then computing the correlation between the observed genetic distance matrix and the resistance distance matrix that is the result of the hypothesized resistance values. We note that since this method relies on correlation, the intercept term in our model can be seen as a constant multiple of the distance matrices, and will have no effect on correlation. For the remaining three covariates, we employed a grid search of possible parameter values. Models with integer-valued parameters ranging from -6 to 3 were tried, resulting in 1000 candidate models for gene flow. The resulting 1000 hypothesized distance matrices were tested for correlation with the observed (simulated) genetic distance matrix, with the correlation being expressed as Mantel's z statistic as computed in the “vegan” package (Oksanen et al. 2012) in the R statistical computing environment (R Development Core Team 2012). The five models with the highest correlation are shown in Table 2. The model with the true parameter values used in the simulation

was ranked 13th out of the 1000 hypothesized models, and is shown in bold in Table 2.

The correlation approach was able to identify the large negative effect of Covariate 1, and correctly identifies that Covariate 2 has no effect, but had some trouble with Covariate 3. This highlights the need for a measure of uncertainty about the estimates of parameter values, as is obtained from the posterior

Table 2. Simulation study results for the correlation method

Model rank	Mantel's z	Covariate 1	Covariate 2	Covariate 3
1	0.1450	-4	0	-1
2	0.1423	-4	0	-2
3	0.1421	-4	0	0
4	0.1416	-5	0	-1
5	0.1416	-5	0	0
⋮	⋮	⋮	⋮	⋮
13	0.1378	-5	0	1

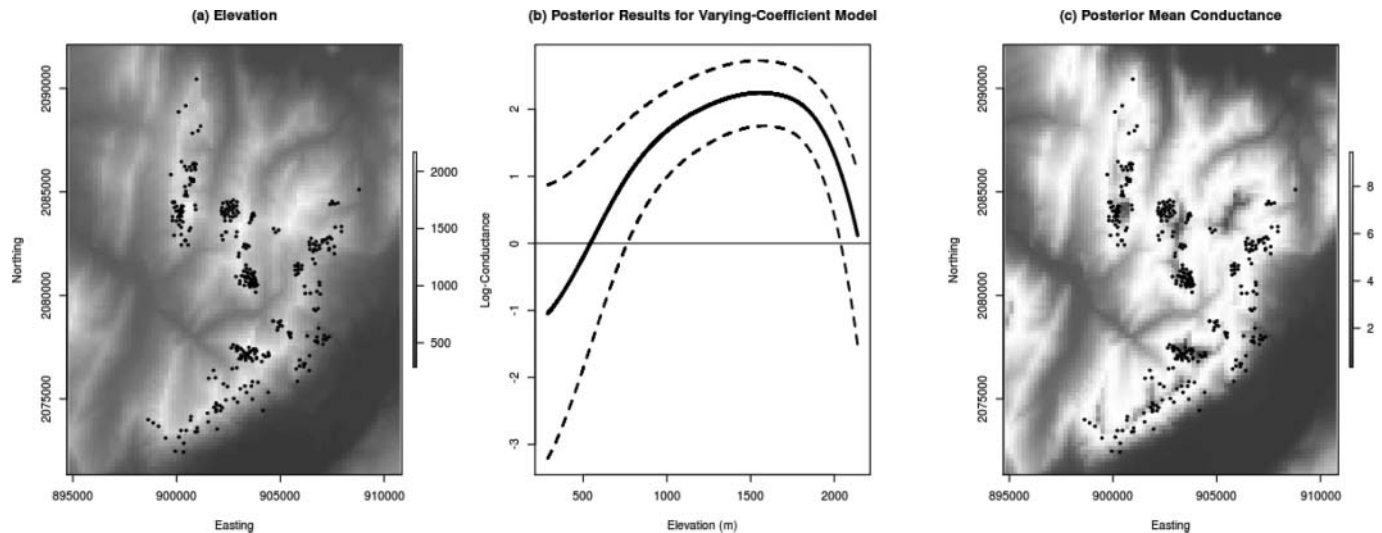


Figure 3. Spatial locations (a) of 335 alpine chamois (*Rupicapra rupicapra*) in the Bauges mountains, France. A landscape genetic analysis of these chamois using a varying-coefficient model for the effect of elevation reveals that gene flow is impeded by low-elevation valleys and mountain peaks and facilitated by mid- to high-elevation terrain (b). The posterior mean conductance of the landscape to gene flow is shown in (c), with lower values corresponding to increased resistance to gene flow.

credible intervals of resistance parameters β from the generalized Wishart model (Table 1).

6.2 Spatial Gene Flow in Alpine Chamois

The alpine chamois (*Rupicapra rupicapra*) is native to the mountains of Europe, and is a conserved species in France. Chamois live at moderately high altitudes and have adapted to living in steep, rocky terrain. Jombart (2012) recently studied spatial gene flow among chamois in the Bauges mountains of France (Figure 3(a)). Their analysis, based on spatial principal component analysis (Jombart et al. 2008), concludes that gene flow in chamois may be impeded by lower elevation terrain (e.g., mountain valleys), possibly due to the increased risk of predation. The analysis of Jombart (2012) does not explicitly incorporate the landscape; rather, the results of the spatial principal component analysis are overlaid on the landscape, and post hoc inference is obtained for the effects of landscape characteristics (elevation in this case) on gene flow. To illustrate our approach to inference for gene flow based on circuit theory and the isolation by resistance hypothesis, we reanalyze the data of Jombart (2012) using our Bayesian model (11)–(14). This approach allows us to explicitly model and estimate the resistance that landscape characteristics (like elevation) have on gene flow in alpine chamois, something not possible using the approach of Jombart (2012).

Jombart (2012) studied microsatellite allele data from nine loci for 335 individual chamois. These data, together with spatial locations for each individual animal and elevation data on a 104×80 grid, where each grid cell measures 200 m square, were obtained from the “adegenet” package (Jombart 2008) in the R statistical computing environment (R Development Core Team 2012).

To examine the effect of elevation on gene flow in alpine chamois, we compared four competing models using deviance information criterion (DIC). The first model contained only an

intercept term, hypothesizing no effect of elevation on gene flow. The second model contained an intercept term, as well as a linear effect for elevation in (12). The third model contained an intercept term, as well as linear and quadratic terms for elevation. The fourth model is a varying-coefficient model in elevation, in which we specified a B-spline basis expansion for the effect of elevation, using seven equally spaced knots. This results in a flexible model of the effect that elevation has on gene flow. The varying-coefficient model could be difficult to implement using the correlation approach common in landscape genetics studies, as multiple hypothesized values for each of the parameters related to the spline basis functions would have to be specified.

We found the grid cell each animal was located in using the “raster” package (van Etten 2012) in R, and computed pairwise genetic distances between the chamois in the study using the mismatch genetic distance of Smouse and Peakall (1999). The resulting genetic distance matrix was used to fit each of the four models. For each model, we obtained 20,000 posterior samples using an MCMC algorithm from two separate chains with distinct starting values. Convergence was assessed using the potential scale reduction factor \hat{R} (Gelman et al. 2004), which compares the within-chain variance to the between-chain variance. The resulting \hat{R} values were less than 1.1 for all parameters in each of the four models, indicating convergence to the stationary posterior distribution.

To compare the four models, we used the DIC of Spiegelhalter et al. (2002). The DIC of the varying-coefficient model for elevation was -2208 , much lower than the DIC of the intercept-only model (DIC = 8939), the linear model for elevation (DIC = 7244), or the quadratic model for elevation (DIC = 5963). This indicates strong support for the varying-coefficient model of the effect of elevation on gene flow. Posterior inference for the log-conductance of a grid cell as a function of elevation is shown in Figure 3(b), with upper and lower 95% credible limits. The posterior mean conductance of the study area is plotted

in Figure 3(c). This plot reflects the nonlinear relationship between elevation and gene flow. Higher elevation terrain (from about 1000 m to 1750 m) generally facilitates gene flow in alpine chamois, while lower elevations (valleys), and very high elevations (mountain peaks) impede gene flow. Jombart (2012) found that higher elevation terrain generally facilitates gene flow in alpine chamois using their spatial principal component analysis. Our model-based approach expands on their results by providing rigorous uncertainty estimates about the parameters related to gene flow, something not typically obtained in landscape genetics studies.

7. DISCUSSION

Using circuit theory to model connectivity on a landscape requires viewing the continuous landscape as a graph (\mathbf{G}, \mathbf{A}) . While this is a simplification, recent advances in GMRF theory have shown that GMRFs can be used to approximate solutions to stochastic partial differential equations on a continuous surface (Lindgren, Rue, and Lindström 2011). The connectivity implied by circuit theory on a graph is analogous to a diffusion process on a continuous surface with heterogeneous diffusivity. Statistical models based on diffusion processes have been used extensively to model spatio-temporal processes on continuous surfaces (e.g., Hooten and Wikle 2008; Wikle and Hooten 2010), and could be used to model many of the systems that are currently modeled using circuit theory. The main advantage of viewing the landscape as a discrete graph is the ability to use computationally efficient GMRFs like the ICAR model (7) used in our analysis.

The link between circuit theory, GMRFs, and genetic distance allows for an efficient approach to simulate pairwise genetic distances on a heterogeneous landscape (e.g., Landguth and Cushman 2010). Using (9), simulated resistance distances can be transformed to simulated linearized pairwise F_{st} values, which can provide significant information about population genetic structure across the landscape. This may be particularly useful in predicting the effects of landscape change on subpopulation separation. For example, inference concerning parameter values β for a resistance surface related to gene flow could be made using the model (11)–(14). The effect of changes to the landscape, such as the addition of a road or the removal of existing open space, on subpopulation separation could then be predicted using inferred values of β and a modified resistance surface reflecting the changes to the landscape.

While simulation is possible at extremely high resolutions, inference becomes increasingly computationally intensive as the landscape resolution increases. In practice, we have found that inference for 250×250 landscape grids is feasible with standard computing resources (e.g., a machine with 4 GB RAM and a 1.67 GHz quad-core processor), but inference for larger grids requires increased computing capabilities. This resolution is adequate for many studies, but larger landscapes may not be well approximated by such a coarse discretization. Examining methods for increasing computational efficiency and thus allowing inference at higher resolutions is the subject of ongoing research.

The link between circuit theory and GMRFs has many potential applications. For example, agent-based models for a binary

process on a graph (Hooten and Wikle 2010) provide a highly flexible framework for modeling dynamic systems with discrete support. The evolution of agent-based systems is based on transition probabilities between neighboring nodes, and could be equivalently modeled as a resistance surface by envisioning a circuit with resistors connecting nodes. Circuit theory provides a flexible model for evaluating connectivity, and the link we have presented between circuit theory and GMRFs provides a natural framework that can provide inference for resistance values in systems modeled by circuits, and thus the methodology we propose here has wide applicability.

APPENDIX A: PROOFS OF PROPOSITIONS

For Proposition 1, note that $\Sigma \in \text{SPSD}(m)$ implies that there is a singular value decomposition $\Sigma = \sum_{i=1}^m \lambda_i \mathbf{u}_i \mathbf{u}_i'$, where $\{\lambda_i \geq 0\}$ are the eigenvalues and $\{\mathbf{u}_i\}$ are orthonormal eigenvectors. Then we can rewrite (2) as

$$\begin{aligned} D_{ij} &= \Delta_{ij}(\Sigma) = (\mathbf{e}_i - \mathbf{e}_j)' \Sigma (\mathbf{e}_i - \mathbf{e}_j) \\ &= (\mathbf{e}_i - \mathbf{e}_j)' \left(\sum_{k=1}^m \lambda_k \mathbf{u}_k \mathbf{u}_k' \right) (\mathbf{e}_i - \mathbf{e}_j) \\ &= \sum_{k=1}^m \lambda_k [(\mathbf{e}_i - \mathbf{e}_j)' \mathbf{u}_k]^2. \end{aligned}$$

As $\lambda_k \geq 0$, each term in the sum above is nonnegative. From this, Σ is in the null space of the transformation Δ if and only if $\Delta_{ij}(\Sigma) = 0$ for all $i = 1, \dots, m$ and $j = 1, \dots, m$, which in turn is true if and only if either $\lambda_k = 0$ or $\mathbf{u}_k = c\mathbf{1}$, $c \in \mathbb{R}$ for each $k = 1, \dots, m$. Since the \mathbf{u}_k are orthogonal, we conclude that $\Delta_{ij}(\Sigma) = 0$ for all i and j if and only if $\Sigma = \lambda \mathbf{1}\mathbf{1}'$.

Proposition 2 is a direct consequence of the lemma that $\mathbf{a}'\mathbf{Q}^{-}\mathbf{a}$ is invariant to the choice of generalized inverse \mathbf{Q}^{-} if and only if \mathbf{a} is orthogonal to $\mathcal{N}(\mathbf{Q})$ (see Graybill 1983, pp. 134–135; Seber 2008, pp. 130–131). The null space of the Laplacian \mathbf{Q} is the space spanned by $\mathbf{1}$, and each contrast $\mathbf{e}_i - \mathbf{e}_j$ is orthogonal to $\mathbf{1}$.

Proposition 3 is a direct consequence of Proposition 1.

APPENDIX B: DETAILS OF THE COMPUTATION OF THE PRECISION MATRIX

Consider the case where we have n observations y_i , $i = 1, \dots, n$ from the m nodes in \mathbf{G} , where $m \gg n$ and let the observed nodes be indexed by the first n rows and columns of \mathbf{Q} . Let \mathbf{Q}^* be the $(m-1) \times (m-1)$ upper-left submatrix of \mathbf{Q} . Then, since \mathbf{Q} is of rank $m-1$, \mathbf{Q}^* is of full rank and invertible. The inverse of \mathbf{Q}^* then forms the $(m-1) \times (m-1)$ upper-left submatrix of the generalized inverse \mathbf{Q}^{-1} , with entries in the m th row and m th column of \mathbf{Q}^{-1} being set to zero. In this formulation, the covariance matrix of the n observed nodes in \mathbf{G} is just the $n \times n$ upper-left submatrix of $(\mathbf{Q}^*)^{-1}$. This submatrix can be found by partitioning \mathbf{Q}^* in block form:

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_{11}^* & \mathbf{Q}_{12}^* \\ \mathbf{Q}_{21}^* & \mathbf{Q}_{22}^* \end{bmatrix}, \quad (\text{B.1})$$

where \mathbf{Q}_{11}^* is the $n \times n$ block of \mathbf{Q} corresponding to the n observed nodes of \mathbf{G} , \mathbf{Q}_{22}^* is the $(m-n-1) \times (m-n-1)$ block of \mathbf{Q} corresponding to the $m-1$ unobserved nodes, after the removal of the m th row and column, and $\mathbf{Q}_{12}^* = (\mathbf{Q}_{21}^*)^T$ is the $n \times (m-n-1)$ block of \mathbf{Q}^* relating the observed nodes to the unobserved nodes. The precision matrix Φ of the n observed locations is the Schur complement (Harville 2008) of \mathbf{Q}_{22}^* : $\Phi = \mathbf{Q}_{11}^* - \mathbf{Q}_{12}^* (\mathbf{Q}_{22}^*)^{-1} \mathbf{Q}_{21}^*$. As \mathbf{Q}_{22}^* is of dimension $(m-n-1) \times (m-n-1)$, we also wish to avoid the explicit

calculation of the inverse $(\mathbf{Q}_{22}^*)^{-1}$ when computing Φ_{11} in (12). This can be accomplished by first computing the Cholesky decomposition: $\mathbf{Q}_{22}^* = \mathbf{U}^T \mathbf{U}$, which can be found efficiently using sparse matrix methods (Bates and Maechler 2011). We then compute $\mathbf{V} = (\mathbf{U}^T)^{-1} \mathbf{Q}_{21}^*$ by solving the n linear equations $\mathbf{U}^T \mathbf{V} = \mathbf{Q}_{21}^*$. As \mathbf{U}^T is lower triangular, this can be solved efficiently using forward substitution. The precision matrix of the observed nodes is then obtained by

$$\Phi = \mathbf{Q}_{11}^* - \mathbf{V}^T \mathbf{V}. \quad (\text{B.2})$$

[Received January 2012. Revised August 2012.]

REFERENCES

- Babić, D., Klein, D., Lukovits, I., Nikolić, S., and Trinajstić, N. (2002), "Resistance-Distance Matrix: A Computational Algorithm and its Application," *International Journal of Quantum Chemistry*, 90, 166–176. [25]
- Bates, D., and Maechler, M. (2011), *Matrix: Sparse and Dense Matrix Classes and Methods*, R package version 0.9996875-3. Available at <http://matrix.r-forge.r-project.org> [32]
- Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, Series B*, 36, 192–236. [23,27]
- Besag, J., and Kooperberg, C. (1995), "On Conditional and Intrinsic Autoregressions," *Biometrika*, 82, 733. [23,27]
- Besag, J., York, J., and Mollié, A. (1991), "Bayesian Image Restoration, With Two Applications in Spatial Statistics," *The Annals of the Institute of Statistical Mathematics*, 43, 1–20. [27]
- Bolin, D., Lindström, J., Eklundh, L., and Lindgren, F. (2009), "Fast Estimation of Spatially Dependent Temporal Vegetation Trends Using Gaussian Markov Random Fields," *Computational Statistics & Data Analysis*, 53, 2885–2896. [26]
- Broquet, T., Ray, N., Petit, E., Fryxell, J. M., and Burel, F. (2006), "Genetic Isolation by Distance and Landscape Connectivity in the American Marten (*Martes Americana*)," *Landscape Ecology*, 21, 877–889. [24]
- Chandra, A., Raghavan, P., Ruzzo, W., Smolensky, R., and Tiwari, P. (1996), "The Electrical Resistance of a Graph Captures its Commute and Cover Times," *Computational Complexity*, 6, 312–340. [22]
- Chen, C., Durand, E., Forbes, F., and Francois, O. (2007), "Bayesian Clustering Algorithms Ascertain Spatial Population Structure: A New Computer Program and a Comparison Study," *Molecular Ecology Notes*, 7, 747–756. [24]
- Cross, P., Heisey, D., Scurlock, B., Edwards, W., Ebinger, M., and Brennan, A. (2010), "Mapping Brucellosis Increases Relative to Elk Density Using Hierarchical Bayesian Models," *PLoS One*, 5, e10322. [27]
- Cushman, S., and Landguth, E. (2010), "Scale Dependent Inference in Landscape Genetics," *Landscape Ecology*, 25, 967–979. [22,24]
- Cushman, S., McKelvey, K., and Schwartz, M. (2009), "Use of Empirically Derived Source-Destination Models to Map Regional Conservation Corridors," *Conservation Biology*, 23, 368–376. [22,29]
- Cushman, S. A., McKelvey, K. S., Hayden, J., and Schwartz, M. K. (2006), "Gene Flow in Complex Landscapes: Testing Multiple Hypotheses With Causal Modeling," *The American Naturalist*, 168, 486–99. [22,24,29]
- Dorf, R., and Svoboda, J. (2004), *Introduction to Electric Circuits*, Hoboken, NJ: Wiley. [23,25]
- Durand, E., Jay, F., Gaggiotti, O., and François, O. (2009), "Spatial Inference of Admixture Proportions and Secondary Contact Zones," *Molecular Biology and Evolution*, 26, 1963. [24]
- Dyer, R., Nason, J., and Garrick, R. (2010), "Landscape Modelling of Gene Flow: Improved Power Using Conditional Genetic Distance Derived From the Topology of Population Networks," *Molecular Ecology*, 19, 3746–3759. [22]
- Fouss, F., Piroette, A., Renders, J., and Saerens, M. (2007), "Random-Walk Computation of Similarities Between Nodes of a Graph With application to Collaborative Recommendation," *IEEE Transactions on Knowledge and Data Engineering*, 19, 355–369. [22]
- Gelman, A., Carlin, B. P., Stern, H., and Rubin, D. B. (2004), *Bayesian Data Analysis* (2nd ed.), Princeton, NJ: Chapman and Hall/CRC. [28,30]
- Graybill, F. (1983), *Matrices With Applications in Statistics*, Belmont, CA: Wadsworth Inc. [31]
- Guillot, G., Estoup, A., Mortier, F., and Cosson, J. F. (2005), "A Spatial Statistical Model for Landscape Genetics," *Genetics*, 170, 1261–1280. [24]
- Hanks, E., Hooten, M., Johnson, D., and Sterling, J. (2011), "Velocity-Based Movement Modeling for Individual and Population Level Inference," *PLoS ONE*, 6, e22795. [28]
- Harville, D. (2008), *Matrix Algebra From a Statistician's Perspective*, New York: Springer-Verlag. [27,31]
- Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society, Series B*, 55, 757–796. [28]
- Holsinger, K., and Weir, B. (2009), "Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting *f_{st}*," *Nature Reviews Genetics*, 10, 639–650. [24]
- Hooten, M., and Wikle, C. (2008), "A Hierarchical Bayesian Non-Linear Spatio-Temporal Model for the Spread of Invasive Species With Application to the Eurasian Collared-Dove," *Environmental and Ecological Statistics*, 15, 59–70. [31]
- Hooten, M. B., Johnson, D. S., Hanks, E. M., and Lowry, J. H. (2010), "Agent-Based Inference for Animal Movement and Selection," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 523–538. [28]
- Hooten, M. B., and Wikle, C. K. (2010), "Statistical Agent-Based Models for Discrete Spatio-Temporal Systems," *Journal of the American Statistical Association*, 105, 236–248. [31]
- Jombart, T. (2008), "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers," *Bioinformatics*, 24, 1403–1405. [30]
- (2012), "A Tutorial for the Spatial Analysis of Principal Components (spca) Using Adegenet 1.3-4," *Vignette for the R package 'adegenet'*. Available at <http://adegenet.r-forge.r-project.org> [30]
- Jombart, T., Devillard, S., Dufour, A. B., and Pontier, D. (2008), "Revealing Cryptic Spatial Patterns in Genetic Variability by a New Multivariate Method," *Heredity*, 101, 92–103. [30]
- Klein, D., Palacios, J., Randić, M., and Trinajstić, N. (2004), "Random Walks and Chemical Graph Theory," *Journal of Chemical Information and Computer Sciences*, 44, 1521–1525. [22]
- Klein, D., and Randić, M. (1993), "Resistance Distance," *Journal of Mathematical Chemistry*, 12, 81–95. [23,25]
- Kunegis, J., Lommatzsch, A., and Bauckhage, C. (2009), "The Slashdot Zoo: Mining a Social Network With Negative Edges," in *Proceedings of the 18th International Conference on World Wide Web*, ACM, pp. 741–750. [22]
- Landguth, E., and Cushman, S. (2010), "Cdpop: A Spatially Explicit Cost Distance Population Genetics Program," *Molecular Ecology Resources*, 10, 156–161. [31]
- Legendre, P., and Fortin, M. (2010), "Comparison of the Mantel Test and Alternative Approaches for Detecting Complex Multivariate Relationships in the Spatial Analysis of Genetic Data," *Molecular Ecology Resources*, 10, 831–844. [22]
- Lindgren, F., Rue, H., and Lindström, J. (2011), "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach," *Journal of the Royal Statistical Society, Series B*, 73, 423–498. [31]
- Lindström, J., and Lindgren, F. (2008), *A Gaussian Markov Random Field Model for Total Yearly Precipitation Over the African Sahel*, Technical Report 8, Centre for Mathematical Sciences, Lund University, Lund Sweden. [26,27]
- Lookingbill, T., Gardner, R., Ferrari, J., and Keller, C. (2010), "Combining a Dispersal Model With Network Theory to Assess Habitat Connectivity," *Ecological Applications*, 20, 427–441. [22]
- Manel, S., Schwartz, M., Luikart, G., and Taberlet, P. (2003), "Landscape Genetics: Combining Landscape Ecology and Population Genetics," *Trends in Ecology & Evolution*, 18, 189–197. [23]
- McCullagh, P. (2009), "Marginal Likelihood for Distance Matrices," *Statistica Sinica*, 19, 631–649. [22,25,26]
- McRae, B. (2006), "Isolation by Resistance," *Evolution*, 60, 1551–1561. [22,24,26]
- McRae, B., Dickson, B., Keitt, T., and Shah, V. (2008), "Using Circuit Theory to Model Connectivity in Ecology, Evolution, and Conservation," *Ecology*, 89, 2712–2724. [22,24]
- McRae, B. H., and Beier, P. (2007), "Circuit Theory Predicts Gene Flow in Plant and Animal Populations," *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19885–19890. [22,24]
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2012), *Vegan: Community Ecology Package*, R package version 2.0-3. Available at <http://vegan.r-forge.r-project.org> [29]
- Owen-Smith, N., Fryxell, J. M., and Merrill, E. H. (2010), "Foraging Theory Upscaled: The Behavioural Ecology of Herbivore Movement," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2267–2278. [22]
- Rayfield, B., Fortin, M., and Fall, A. (2010), "The Sensitivity of Least-Cost Habitat Graphs to Relative Cost Surface Values," *Landscape Ecology*, 25, 519–532. [22]

- R Development Core Team (2012), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [29,30]
- Rue, H. (2001), "Fast Sampling of Gaussian Markov Random Fields," *Journal of the Royal Statistical Society, Series B*, 63, 325–338. [26,27]
- Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, in *Monographs on Statistics and Applied Probability* (Vol. 104), Boca Raton, FL: Chapman & Hall. [23,26,27]
- Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society, Series B*, 71, 319–392. [26]
- Saura, S., and Rubio, L. (2010), "A Common Currency for the Different Ways in Which Patches and Links can Contribute to Habitat Availability and Connectivity in the Landscape," *Ecography*, 33, 523–537. [22]
- Seber, G. (2008), *A Matrix Handbook for Statisticians* (Vol. 746), Hoboken, NJ: Wiley. [31]
- Shirk, A., Wallin, D., Cushman, S., Rice, C., and Warheit, K. (2010), "Inferring Landscape Effects on Gene Flow: A New Model Selection Framework," *Molecular Ecology*, 19, 3603–3619. [22,29]
- Smouse, P., and Peakall, R. (1999), "Spatial Autocorrelation Analysis of Individual Multiallele and Multilocus Genetic Structure," *Heredity*, 82, 561–573. [30]
- Spear, S., Balkenhol, N., Fortin, M., McRae, B., and Scribner, K. (2010), "Use of Resistance Surfaces for Landscape Genetic Studies: Considerations for Parameterization and Analysis," *Molecular Ecology*, 19, 3576–3591. [22,23,24]
- Spear, S. F., Peterson, C. R., Matocq, M. D., and Storfer, A. (2005), "Landscape Genetics of the Blotched Tiger Salamander (*Ambystoma tigrinum melanostictum*)," *Molecular Ecology*, 14, 2553–2564. [24]
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64, 583–639. [30]
- Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., and Waits, L. P. (2007), "Putting the 'Landscape' in Landscape Genetics," *Heredity*, 98, 128–142. [22,24]
- Theobald, D., Crooks, K., and Norman, J. (2011), "Assessing Effects of Land use on Landscape Connectivity: Loss and Fragmentation of Western us Forests," *Ecological Applications*, 21, 2445–2458. [22]
- Tizghadam, A., and Leon-Garcia, A. (2010), "On Traffic-Aware Betweenness and Network Criticality," In *INFOCOM IEEE Conference on Computer Communications Workshops, IEEE*, pp. 1–6. [22]
- (2011), "Robust Network Planning in Nonuniform Traffic Scenarios," *Computer Communications*, 34, 1436–1449. [22]
- Urban, D., Minor, E., Treml, E., and Schick, R. (2009), "Graph Models of Habitat Mosaics," *Ecology Letters*, 12, 260–273. [22]
- van Etten, R. J. H. J. (2012), *Raster: Geographic Analysis and Modeling With Raster Data*, R package version 1.9-92. Available at <http://raster.r-forge.r-project.org> [30]
- Volchenkov, D. (2011), "Random Walks and Flights Over Connected Graphs and Complex Networks," *Communications in Nonlinear Science and Numerical Simulation*, 16, 21–55. [22]
- Wang, H., and Xia, Y. (2009), "Shrinkage Estimation of the Varying Coefficient Model," *Journal of the American Statistical Association*, 104, 747–757. [29]
- Wang, I., Savage, W., and Bradley Shaffer, H. (2009), "Landscape Genetics and Least-Cost Path Analysis Reveal Unexpected Dispersal Routes in the California Tiger Salamander (*Ambystoma californiense*)," *Molecular Ecology*, 18, 1365–1374. [22]
- Wheeler, D. C., and Waller, L. A. (2010), "Spatial Analysis of Genetic Population Structure in Cougars," Technical Report 10-01, Department of Biostatistics and Bioinformatics, Rollins School of Public Health. [23]
- Wheeler, D. C., Waller, L. A., and Biek, R. (2010), "Spatial Analysis of Feline Immunodeficiency Virus Infection in Cougars," *Spatial and Spatio-temporal Epidemiology*, 1, 151–161. [23]
- Wikle, C., and Hooten, M. (2010), "A General Science-Based Framework for Dynamical Spatio-Temporal Models," *Test*, 19, 417–451. [31]
- Wright, S. (1943), "Isolation by Distance," *Genetics*, 28, 114. [24]
- Zhu, H., and Klein, D. (1996), "Graph-Geometric Invariants for Molecular Structures," *Journal of Chemical Information and Computer Sciences*, 36, 1067–1075. [22]