

A new cost-distance model for human accessibility and an evaluation of accessibility bias in permanent vegetation plots in Great Smoky Mountains National Park, USA

Jobe, R. Todd^{1*} & White, Peter S.²

¹*Geography Department, CB#3220 Saunders Hall, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3220, USA;* ²*Department of Biology, CB#3280 Coker Hall, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA, Email peter.white@unc.edu;*

**Corresponding author; E-mail toddjobe@unc.edu*

Abstract

Question: Can a new cost-distance model help us to evaluate the potential for accessibility bias in ecological observations? How much accessibility bias is present in the vegetation monitoring plots accumulated over the last three decades in Great Smoky Mountains National Park?

Location: Great Smoky Mountains National Park, North Carolina and Tennessee, USA.

Methods: Distance, slope, stream crossings, and vegetation density were incorporated into a least-cost model of energetic expenditure for human access to locations.

Results: Estimated round-trip energy costs for the park ranged from 0 to $1.62 \times 10^5 \text{ J kg}^{-1}$. The estimated round-trip energetic expenditure for the surveys ranged from 53 to $1.51 \times 10^5 \text{ J kg}^{-1}$. Their distribution was more accessible than the random expectation. Ten (17%) of the vegetation types in the park are significantly under-sampled relative to their area, and 16 (29%) are over-sampled. Plots in 18 of the 40 vegetation types exhibited a significant positive correlation with accessibility.

Conclusions: The least-cost model is an improvement over previous attempts to quantify accessibility. The bias in plot locations suggests using a least-cost model to test for bias in cases in which human accessibility is confounded with other sources of ecosystem variation.

Keywords: Cost-distance; Great Smoky Mountains National Park; Human accessibility; Least-cost paths; Sampling bias.

Nomenclature: Weakley (2008).

Introduction

The ability of humans to reach a particular location in a landscape, termed here “accessibility”,

varies with the built infrastructure (e.g., roads and trails) and natural features of the landscape, such as terrain steepness, barriers to travel (e.g., stream crossings), and vegetation density. Accessibility is important because direct and indirect human impacts on ecosystems vary with this parameter (Trombulak & Frissell 2000). In addition, accessibility may itself be correlated with environmental features since topography and substrate affect construction costs, leading to a potential confounding of environmental and human effects. Finally, accessibility is potentially important to ecological studies because observations may be differentially carried out in accessible locations, as a way of reducing costs and increasing the efficiency of data collection (Greenwood 1996).

The potential significance of accessibility leads to an important question: does sampling in accessible locations create bias in the observations; for example, if accessibility is correlated with distribution of vegetation types, if there is turnover in composition from accessible to inaccessible locations of the same vegetation types, or if human effects themselves influence local environmental conditions or the behavior or distribution of organisms. The latter effect could be positive (disturbance-dependent and invasive species along road corridors) or negative (species that are harvested or those that avoid human contact).

Despite the obvious importance of accessibility, its quantification has been simplistic. Distance from a road is the typical measure, which ignores the influences of topography, water features, and vegetation. Yet, even with this crude measure, studies have shown that increased distance from roads results in increased frequency of native grasses (Gelbard & Harrison 2003) and changes in movement behavior for a wide variety of animals,

including black bear (Brody & Pelton 1989), grizzly bear (McLellan & Shackleton 1988), elk (Grover & Thompson 1986), mule deer (Rost & Bailey 1979), and wolves (Whittington et al. 2005); see also Trombulak & Frissell (2000) for a full review,

Cost-distance analysis is a more sophisticated approach that has been used extensively in wildlife biology to assess functional connectivity of landscapes for a variety of species (e.g., Schadt et al. 2002; Larue & Nielsen 2008). Costs are calculated by applying a least-cost path algorithm to a source raster and a resistance raster (Adriaensen et al. 2003). Human movements have also been analyzed using cost-distance approaches. Tobler (1993) was the first to use the Imhof (1950) "hiking function" to calculate the cost associated with traversing a landscape. The "hiking function" estimates the velocity of travel for hikers across different slopes. A few anthropological studies have used similar functions to estimate the frequency of interaction between tribes as a function of hiking time between villages (e.g., Van Leusen 2002; Hare 2004). However, the "hiking function" has three weaknesses. First, energetic expenditure is a better measure of the true physiological cost associated with hiking than velocity (Rose et al. 1994). Second, the "hiking function" is symmetric about a -1° slope so that walking speed is roughly identical for uphill and downhill as steepness increases. Treadmill experiments show that energetic expenditure is asymmetric between uphill and downhill slopes (Minetti et al. 1993, 1994, 2001; Minetti 1995; Minetti et al. 2002). Finally, the "hiking function" does not consider costs associated with traversing obstacles, such as streams and vegetation.

Here, we use empirical estimates of energetic costs associated with traversing different vegetation densities from Soule & Goldman (1972) and incorporate these into a new cost-distance model that estimates the energetic cost associated with hiking along the least-cost path from the nearest car-accessible road to any point on a landscape. This model incorporates not only surface distance, but also the isotropic (i.e., direction-independent) friction associated with landscape features such as trails, vegetation, and streams. Further, we include an anisotropic (i.e., direction-dependent) factor that estimates the energetic cost associated with hiking along slopes of different gradient. We use the model to assess accessibility bias of permanent vegetation plots in Great Smoky Mountains National Park (GSMNP), North Carolina and Tennessee, USA. We asked the following questions: What advantages does a more sophisticated model have over simple

distance? Which model parameters have the greatest influence on energetic costs and therefore suggest priorities for future model improvements? How severe is the bias towards accessible locations in permanent vegetation plots? Does this bias vary among vegetation types or along topographic gradients? The model of human accessibility we present represents the first estimate of round-trip energetic cost (in J kg^{-1}) that incorporates the anisotropic costs of slope walking as well as the isotropic costs of walking through vegetation and streams.

Materials and Methods

Study area

GSMNP is a 2×10^5 ha national park in the Southern Appalachian Mountains of North Carolina and Tennessee, USA (Fig. 1). Elevation varies from 184 to 2029 m. There are 845 km of roads within and surrounding the park, with a main road (Newfound Gap Road) running N-S through the center of the park (Fig. 1). There are 1295 km of hiking trails in park.

Data

We used a series of GIS layers including a digital elevation model (DEM, resolution 10 m^2) and vector data for roads, trails, and streams (B. Zank, pers. comm.). We used an overstory vegetation polygon layer developed by the University of Georgia Center for Remote Sensing and Mapping that was hand-delineated from $23 \text{ cm} \times 23\text{-m}$ resolution aerial photographs (CRMS, Madden et al. 2004). The digital representation of each vegetation type was at least 80% accurate, based on collected field data (Jenkins 2007). The final classification included 56 vegetation types following the US National Vegetation Classification System, part of the International Vegetation Classification System (NatureServe 2006).

One of the most significant physiognomic features of GSMNP, even to a casual observer, is the distribution of ericaceous shrubs of the genera *Rhododendron* and *Kalmia*, which reach the height of small trees in the park (Weakley 2008). Although their abundance varies, approximately 24% of the park can be considered to have moderate or heavy dominance of these taxa. These thickly tangled, evergreen stands, sometimes referred to as "hells", are extremely difficult to travel through, and represent a significant barrier to the accessibility of

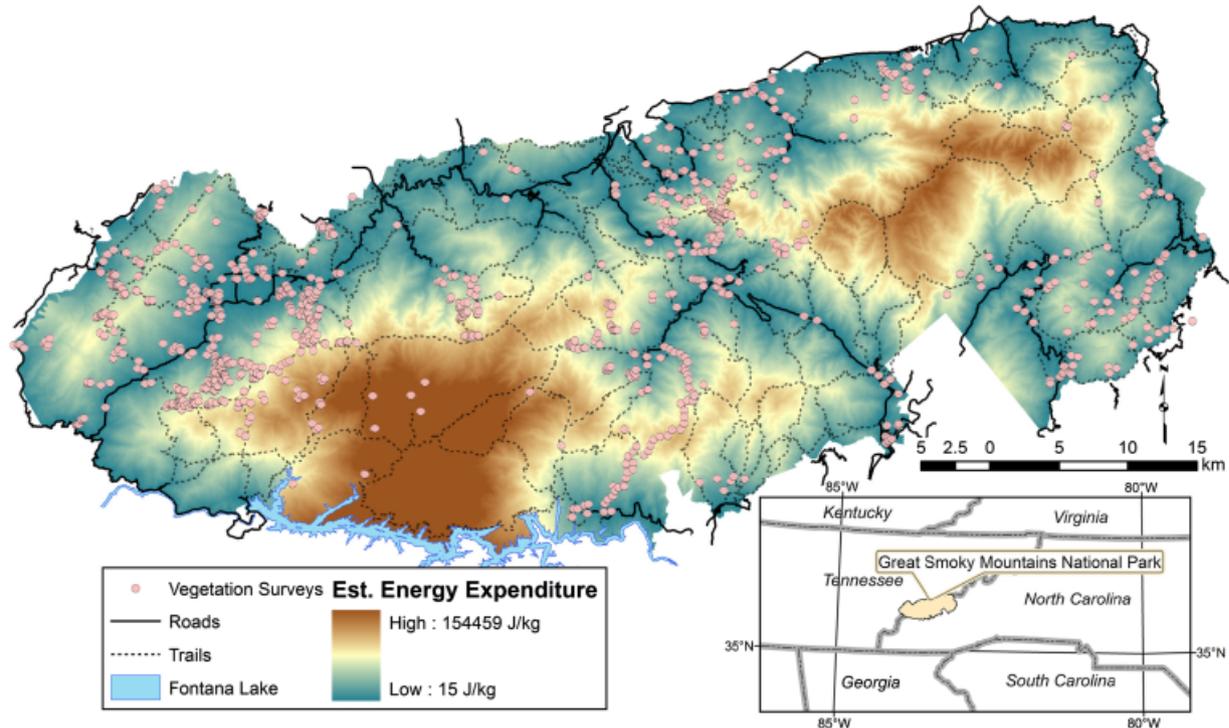


Fig. 1. Energetic cost estimates for GSMNP based on a least-cost path model of round-trip energetic expenditures. Paths were calculated from roads to every other point in the park. There are two major areas of inaccessibility: one located north of Fontana Lake and the other in the high-elevation eastern portion of the park. The collection of 1104 vegetation surveys that were used to assess sample correlation with accessibility is also shown.

interior regions of the park. For mapping the distributions of *Rhododendron* spp. (*R. maximum*, *R. catawbiense*, *R. minus*, and *R. carolinianum*) and *Kalmia latifolia*, we used an understory map developed by CRMS (Madden et al. 2004) that records the dominant evergreen species present in the understory. This classification gave a low, medium, and high density to the distribution of these two genera based on cover. Low densities had <50% cover of these species, medium had 50-80%, and heavy had >80% cover.

The vegetation plots used in this study were obtained from a recently compiled database of vegetation surveys conducted in GSMNP beginning in 1972 and continuing until 2004. The data are composed of 1104 individual survey plots and represent the vast majority of vegetation surveys conducted in the park. The geolocations of each survey unit in these data are known within approximately 50 m. This is a coarser resolution than the accessibility model itself, which is 10 m. We chose the smaller grain size for the model to reflect the best available GIS data. However, the resulting model is strongly autocorrelated, minimizing the probability of gross errors of plot accessibility for a 50-m error in location.

Constructing the energetic cost model

The energetic cost model is based on an accumulative least-cost path algorithm. The model was built in ArcGIS 9.2, using the *Pathdistance* function (ESRI 2006). The model is computed on a lattice of equal area cells. The extent of this lattice matched that of the park and the cell size was 100 m². The least-cost path algorithm computes the total energetic cost to travel from a focal area (the park) to a source area; in this case, the roads. The algorithm determines the least-cost path by starting from the focal cell and successively finding adjacent cells (in an 8-cell neighborhood) that are least costly. This process continues until a source cell is reached. Although the algorithm starts from an interior location, it ends up with the correct answer for the least-cost path from the nearest road. The energetic cost assigned to a focal cell is the sum of the individual cell costs along the least-cost path.

For a given cell, the energetic cost ($Cost\ J\ kg^{-1}$) of traversal is calculated from:

$$Cost = S \times I \times A \times E \quad (1)$$

where S is the surface distance (m), I is the sum of the isotropic costs, A is the sum of the anisotropic

costs, and E is a constant that converts surface distance in meters to energetic cost in J kg^{-1} . Surface distance (S) is the linear distance of travel across the surface of the Earth. Isotropic costs (I) are those energetic costs that do not change with the direction of travel. These include the friction associated with the trail surface, the vegetation, and any water features that are present. Anisotropic costs (A) are those costs that are dependent upon the direction of travel. The only anisotropic cost in this model was the cost associated with hiking on a slope. Slope costs are anisotropic because the perceived slope of a surface changes depending on the direction from which it is approached.

Isotropic costs

Three landscape classes isotropically influence the energetic cost of hiking: trails, vegetation, and streams. There is some history of calculating the increase in energetic costs associated with different terrains, especially in military applications (e.g., Soule & Goldman 1972). Imhof (1950) suggested that the effect of walking on- versus off-trail reduces speed by 60%. In a more detailed study, Soule & Goldman (1972) recorded energetic expenditure for test subjects walking on roads and trails, through light and heavy brush, and through swamps, and calculated energetic cost coefficients for each of these terrains. We have used these coefficients as the isotropic costs of vegetation and off-trail hiking in this model (Table 1). Unfortunately, Soule & Goldman (1972) do not explicitly define what is meant by light brush and heavy brush. For this model, we have assumed that any off-trail walking in GSMNP could be considered light brush in the context of Soule & Goldman (1972). We considered the effect of a medium to heavy density of *Rhododendron* and *Kalmia* to be equivalent to heavy brush in the context of Soule & Goldman (1972).

There is no previous research on the impact of streams on accessibility, as there is for vegetation and trail versus off-trail walking. High elevation, first-order streams are rarely difficult to cross, if they are flowing at all, while lower-elevation streams can

Table 1. Energetic cost coefficients for different terrains in the least-cost path model of human accessibility.

Terrain	Soule & Goldman (1972)	Coefficient
Trail	Dirt	1.2
Off-Trail	Light Brush	1.31
<i>Rhododendron</i> sp.- <i>Kalmia</i> sp.	Heavy Brush	1.59
Stream (Slope \rightarrow 0)	Swamp	1.87

be virtually impossible to cross. Taking this into account, we modeled the energetic expenditure associated with crossing a stream as a function of stream discharge. Stream discharge, which is typically measured using Manning’s equation (Gore 1996), is controlled by two factors: stream cross-sectional area and stream velocity. Following the general form of the Manning equation, we let stream area be proportional to the logarithm of accumulated flow from the upslope area. We let stream velocity vary with the tangent of slope. Steeper slopes have faster streams. Using these relationships, we estimated the proportional increase in energetic cost associated with stream crossing (C_{st}) to be:

$$C_{st} = a \ln(F) \tan(S) + b \tag{2}$$

where F is the accumulated flow of the stream, S is the slope, b is the baseline friction associated with crossing a stream of slope 0, as specified for swamps in Soule & Goldman (1972), and a is a fitted parameter used to standardize the minimum stream crossing friction to be equal to the friction of walking on a trail.

Anisotropic cost: Slope

Instead, of using the Imhof “hiking function” to calculate the cost of walking on slopes, we have synthesized recent work in biomechanics to generate a new hiking function based on energetic cost rather than velocity. There have been a variety of studies assessing the impact of slope on the energetic expenditure of walking and running (See Rose et al. 1994 for a comprehensive treatment). Only a few, however, focus on the energetic cost of walking on relatively steep terrain (e.g., Minetti 1995; Minetti et al. 2001, 1993, 1994, 2002). Minetti et al. (2002) determined the energetic cost associated with walking up and down slopes ranging from -24° to $+24^\circ$. They found that the slope with the smallest energetic cost was -6° . At slopes greater than this, costs rise sharply. At slopes less than this, costs rise slowly. To reflect these trends, we fitted two second-order polynomials to the treadmill experiments of Minetti et al. (2002):

$$W = \begin{cases} 20.9 \tan(S)^2 + 4.18 \tan(S) + 1.38 & -60 < S < -6 \\ 52.1 \times 10^3 \tan(S)^2 + 10.4 \tan(S) + 2.65 & -6 < S < 60 \end{cases} \tag{3}$$

where W is the energetic cost ($\text{J kg}^{-1} \text{m}^{-1}$) and S is the local slope in degrees. The conversion from surface distance to energetic costs (E in Eq. 1) is subsumed in Eq. 3. Extrapolation of the Minetti et al. (2002) functions was necessary since slopes in the park are often

greater than 24° , and people can traverse steeper slopes than these. Maximal slope in the model was 60° . Slopes steeper than this were considered impossible to traverse in that direction.

The functions that we derived from Minetti et al. (2002) consider the cost of traversing a slope in one direction only. It is round-trip costs, however, that are important for measuring accessibility. The one-way energetic cost for walking down the side of a mountain is low, and a one-way least-cost path would reflect that. The return trip along that same path, however, would have a tremendous energetic cost. To account for this, the calculated least-cost path minimized the energetic cost of both the inbound and outbound legs. A round-trip path of lower energetic cost may exist that follows a different path to and from a given location, but calculating these paths for all locations in the park is intractable (4×10^{14} calculations), given the extent and resolution of the remotely sensed data.

In the national park, a large area north of Fontana Lake is considered inaccessible because there are no road access points and least-cost paths are plotted from trails that intersect roads to the east and west of this area. If boat access was included in the model, then the lower elevation locations in this area would have lower energetic costs. However, the area has few vegetation plots because research teams have cars or trucks but not boats, so we have assumed this is a real limitation. The model could easily be expanded to include other modes of transportation.

Sensitivity analysis

We assessed the relative importance of the model parameters (both isotropic and anisotropic) by generating a series of least-cost path landscapes by varying each parameter. Each of the slope, trail, stream, and vegetation effects was given a high and a low value by multiplying the original parameters by 2 and 0.5, respectively. Given four parameters of three levels (low, original, high), we generated a factorized series of 81 landscapes of each unique parameter set. This factorization allowed the assessment of both independent and interaction effects between the parameters. We then compared the mean accessibilities of these varied landscapes to that of the original model.

Analysis of accessibility bias in the permanent vegetation plots

Each plot in the vegetation survey data was associated with an estimated energetic expenditure (in

J kg^{-1}). Since the geographic locations of vegetation samples are known to within only 50 m, we used the mean energetic cost of all sites within a radius of 50 m of a plot.

To assess whether the plots exhibited sampling bias for accessible locations, we compared the distribution of accessibility among plots to park-wide distribution accessibility using a one-sided Kolmogorov-Smirnov (KS) test. This procedure shows whether or not the distribution function of accessibility among the observed plots is consistently more accessible than the distribution function of accessibility for the entire park. Second, we tested whether each vegetation type was more accessible or less accessible than its random expectation given its area, using a randomization test where vegetation type assignment for each cell in the landscape was permuted 1000 times and the median accessibility of each vegetation type was calculated. Third, we tested the observed number of plots in each vegetation type against the null hypothesis that plots were randomly assigned to each vegetation type proportional to its area. This was done by comparing the observed number of plots to their expected abundance in a multinomial distribution, in which probabilities were the relative areas of each vegetation type. This procedure highlighted those vegetation types that are chronically under-sampled or over-sampled relative to their area in the park. We corrected for multiple comparisons using the false discovery rate (FDR; Storey et al. (2004); qvalues package in R, Dabney & Storey 2008), but the number of significant tests did not change even at high cutoff values, therefore those q-values are not reported. Finally, we tested whether observed plots within a vegetation type tended to be in accessible locations by generating 10,000 random samples of the same size as the number of observed plots within each vegetation type and comparing their distribution of accessibility to the observed distribution of accessibility among plots. Again, FDR was used for the multiple comparisons of vegetation types, and the results predicted no false positives given the original P -values and a large cutoff.

Results

The model

Estimated round-trip energy costs calculated from roads to all other points in the park ranged from 15 to $1.62 \times 10^5 \text{ J kg}^{-1}$ (Fig. 1). The distribution of energetic cost in the park was skewed toward

more accessible sites, with a mean of $4.54 \times 10^4 \text{ J kg}^{-1}$ and a standard deviation of $3.22 \times 10^4 \text{ J kg}^{-1}$. The area north of Fontana Lake in the SW part of the park is the largest contiguous area that is inaccessible, followed closely by an area in the NE part of the park. The estimated energetic costs can be measured in kilocalorie (kcal) expenditures if the weight of the person is known. Based on the model, an average male weighing 75 kg would expend on average 650 kcal traveling from the nearest road to a random site within the park, and would expend 2,768 kcal to travel to and from the most inaccessible site in the park along the least-cost path.

Sensitivity analysis of the model parameters revealed slope as the dominant contributor to the mean accessibility of the model (Fig. 2). The effect of varying costs associated with vegetation was similar to that of slope, but of a much lower magnitude (an approximately 3% change in mean accessibility). Interaction effects contributed little to mean accessibility relative to their independent effects and are not reported here. Doubling the resistance of slope increased energetic cost five-times more than doubling any of vegetation, trail, or stream resistances (Fig. 2). The greater sensitivity of the model to slope is significant because the relationship between energetic expenditure and slope has been well tested from treadmill experiments on human subjects (Minetti et al. 2002).

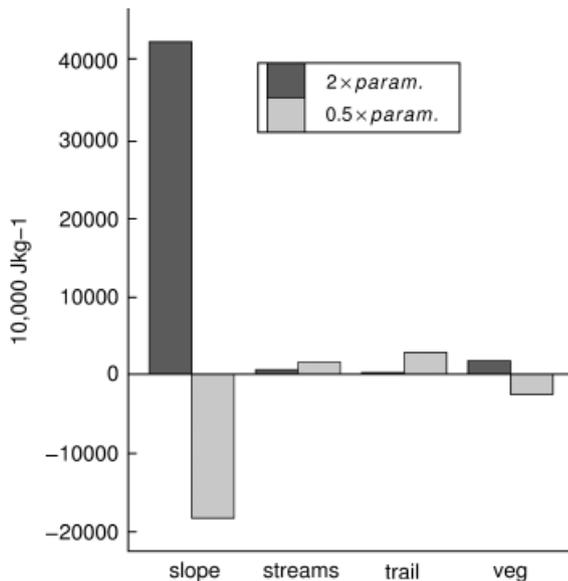


Fig. 2. The difference in mean landscape accessibility with relative changes in the costs associated with slope, trails, vegetation, and streams. Results are shown for each parameter multiplied by factors of 2.0 (dark bars) and 0.5 (light bars).

Accessibility bias

Topographic gradients are a major determinant of vegetation patterns in GSMNP (Whittaker 1956) and also effect how roads and trails are built. We analyzed the relationship between accessibility and local elevation, slope, and hillshade for 1×10^6 randomly chosen locations using a linear model (Table 2, Fig. 3). The interactions between elevation, slope, and hillshade were significantly predictive of accessibility, although the pure effects of elevation and hillshade were not. The accessible locations are clearly not a random sample of the park’s topographic gradients.

The estimated round-trip energetic expenditure to vegetation plots ranged from 53 to $1.51 \times 10^5 \text{ J kg}^{-1}$, with a mean energetic expenditure of $2.7 \times 10^4 \text{ J kg}^{-1}$. As expected, the distribution of sample units in the park is not a random sample of accessibilities in the park (Fig. 4, K-S test results: $D = 0.1802$ $P < 0.001$). There are nearly three times as many plots in highly accessible areas than would be expected given a random sample.

All vegetation types within the park exhibited a significant trend toward either accessibility or inaccessibility based on a random expectation (Appendix S1, Fig. 5). Vegetation type area did not predict the median energetic cost based on a linear regression ($R^2 = 0.009$, $P = 0.482$). The vegetation plots captured 40 of the 56 vegetation types within the park. Logistic regression of sampled and non-sampled vegetation types against area (km^2) revealed a significant trend for unsampled vegetation types to have smaller area (slope = 3.214, $P = 0.092$). Yet, of the unsampled vegetation types, only two had an expected number of plots of one or less. Ten (17%) of the vegetation types in the park are significantly under-sampled relative to their area, and 16 (29%) are over-sampled (Table 3). The most

Table 2. Linear regression results of energetic cost (J kg m^{-1}) as a function of elevation, slope and hillshade (azimuth 45, bearing 135) for a sample of 100,000 random points. Most of the explained variance is bound up in interactions between the topographic variables, rather than the pure effects.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34588.7521	3594.1111	9.62	0.0000
elev	5.5698	3.4122	1.63	0.1026
Shade	-145.3839	20.1727	-7.21	0.0000
Slope	-88.6227	117.7989	-0.75	0.4519
elev : shade	0.1084	0.0193	5.61	0.0000
elev : slope	0.3784	0.1105	3.42	0.0006
shade : slope	3.6384	0.6620	5.50	0.0000
elev : shade : slope	-0.0030	0.0006	-4.79	0.0000

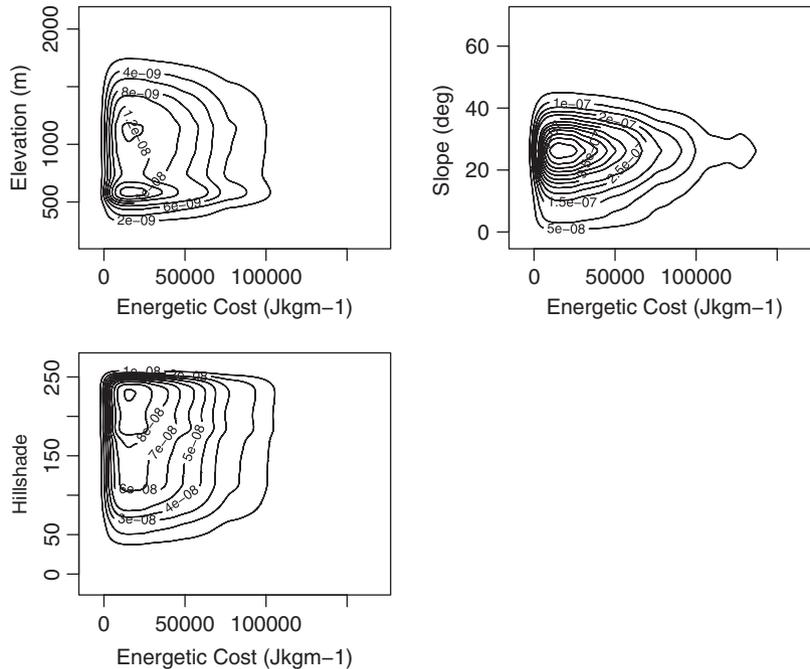


Fig. 3. Probability densities of energetic cost versus three topographic descriptors (elevation, slope, and hillshade) for 1×10^6 random locations in the park.

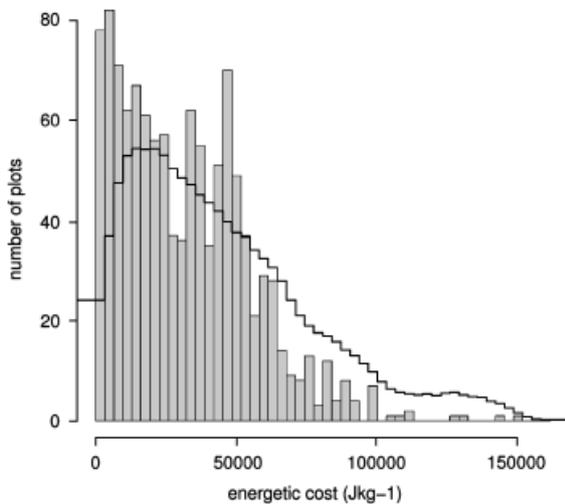


Fig. 4. Expected and observed frequencies of vegetation surveys by accessibility. Expected frequencies (black line) are the number of plots that would be observed if vegetation surveys were uncorrelated with accessibility. Plot frequencies (gray bars) are binned into 50 equal interval quantiles, each spanning 3230 J kg^{-1} . The data show a strong correlation with accessible sites, especially the most accessible sites.

significantly under-collected vegetation type was high-elevation Blue Ridge hemlock-northern hardwood forest. The most significantly over-collected vegetation type was also a high-elevation forest of the red spruce-Fraser fir type.

Even among vegetation types that were significantly over-sampled relative to their area, sampling bias for accessible sites remained. Fraser Fir forest, for example, is significantly over-sampled (Table 3). Its landscape distribution is inaccessible (Appendix S1), and the plots occurring within this vegetation type are significantly biased for more accessible locations (Table 4). As a result, although the vegetation types are over-sampled relative to its area, its largest inaccessible patch only has two plots (Fig. 5b).

Plots in 18 of the 40 vegetation types sampled exhibited significant correlation with accessible parts of the vegetation type (Table 4). Plots in two vegetation types, High-Elevation Red Oak Forest (Evergreen Shrub Type) and Montane Alluvial Forest (Cades Cove/Ocanaluftee) were significantly inaccessible, and both of these vegetation types were over-sampled relative to their area in the park. Six vegetation types were under-sampled, and biased for accessible locations in the places that were sampled.

Discussion

Assessment of the model

For both natural and social science studies of human interactions with conservation areas, this

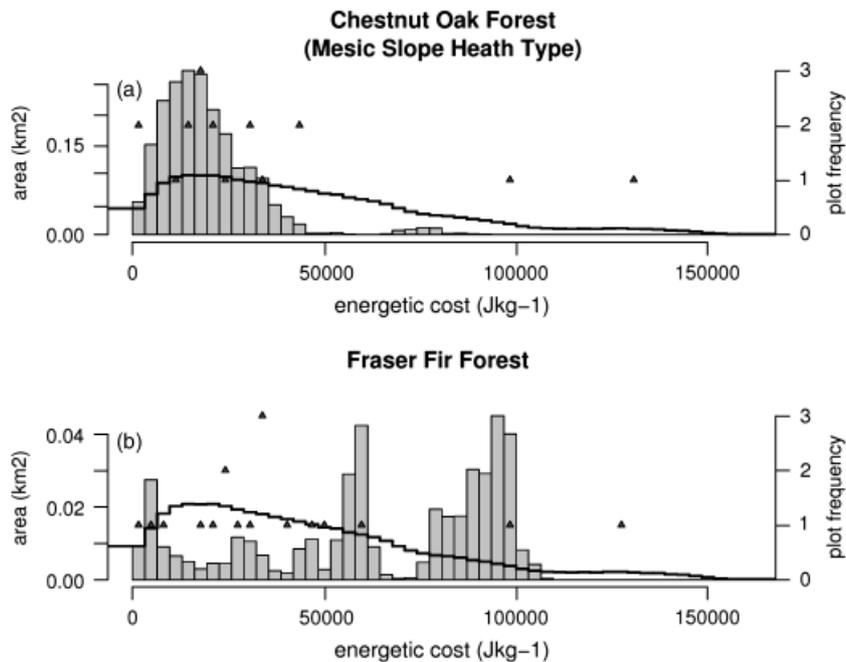


Fig. 5. The 50-bin histograms of the energetic cost distributions of two vegetation types (gray bars, left y-axis). The selected vegetation types are examples of significantly accessible (a), and inaccessible (b) vegetation types. Expected areas under an assumption of random accessibility are shown (black solid line, left y-axis). The numbers of vegetation plots occurring in the 50 accessibility bins are also shown (gray triangles, right y-axis). Fraser Fir Forest is over-sampled relative to its area; Chestnut Oak Forest is under-sampled relative to its area.

Table 3. Observed (obs) and expected (exp) number of vegetation plots for significantly over-sampled (+) and under-sampled (–) vegetation types in the park. Note: Two-tailed significance – **** < 0.1 , * < 0.05 , ** < 0.001 , *** < 0.0001 .

Vegetation type	Obs	exp	diff	P-value
Blue Ridge Hemlock–Northern Hardwood Forest	60	125	–	0.000***
Appalachian Montane Oak–Hickory Forest (Red Oak Type)	23	54	–	0.000***
Mountain Laurel (sparse canopy)	64	101	–	0.000***
<i>Rhododendron</i> sp. (sparse canopy)	1	13	–	0.000***
Red Spruce–Northern Hardwood Forest	2	14	–	0.000***
Blue Ridge Table Mountain Pine–Pitch Pine Woodland (Typic Type)	65	92	–	0.001**
Red Spruce–Fraser Fir Forest (Evergreen Shrub Type)	0	5	–	0.005*
Early Successional Appalachian Hardwood Forest	0	4	–	0.009*
Appalachian Montane Oak–Hickory Forest (Typic Acidic Type)	44	59	–	0.019*
Appalachian Montane Oak–Hickory Forest (Chestnut Oak Type)	2	7	–	0.021*
Southern Appalachian High-Elevation Rocky Summit and Cliff	1	0	+	0.034****
Grassy Bald (Southern Grass Type)	21	13	+	0.024*
Appalachian Montane Alluvial Forest	27	18	+	0.019*
Southern Appalachian Northern Hardwood Forest (Rich Type)	85	68	+	0.017*
Fraser Fir Forest	6	2	+	0.017*
Red Spruce–Northern Hardwood Forest (Herb Type)	10	4	+	0.005**
Chestnut Oak Forest (Xeric Ridge Type)	68	49	+	0.004**
Chestnut Oak Forest (Mesic Slope Heath Type)	37	22	+	0.002**
Southern Appalachian Northern Hardwood Forest (Typic Type)	28	13	+	0.000***
High-Elevation Red Oak Forest (Evergreen Shrub Type)	24	8	+	0.000***
Montane Alluvial Forest (Cades Cove/Ocanaluftee)	8	1	+	0.000***
Appalachian White Pine–Xeric Oak Forest	29	11	+	0.000***
Southern Appalachian Acid Cove Forest (Typic Type)	58	29	+	0.000***
Southern Appalachian Cove Forest (Rich Montane Type)	5	0	+	0.000***
Eastern White Pine Successional Forest	84	39	+	0.000***
Red Spruce–Fraser Fir Forest (Protected Hemlock Type)	6	0	+	0.000***
Red Spruce–Fraser Fir Forest (Shrub Type)	26	2	+	0.000***

model, the first to present round-trip energetic costs (in J kg^{-1}) and which incorporates vegetation density and stream crossings, represents an

improvement over the distance-based models in general use. Landscape estimates of accessibility also provide information for making conservation

Table 4. Vegetation types in which observed plots (obs) exhibit significant sampling bias for either accessible (–) or inaccessible (+) locations based on a randomization test. Note: Two-tailed significant – **** <0.1, * <0.05, ** <0.001, *** <0.0001.

Vegetation type	obs	bias	P-value
Southern Appalachian Beech Gap	67	–	0.000***
Pine Woodland (Xeric)	65	–	0.000***
Southern Appalachian Heath/Mountain Laurel Bald	21	–	0.000***
High-Elevation Blackberry Thicket	22	–	0.000***
Cultivated Meadow	27	–	0.000***
Rush Marsh	68	–	0.001**
Grassy Bald (Southern Grass Type)	39	–	0.001**
Appalachian Montane Alluvial Forest	4	–	0.002**
Blue Ridge Hemlock–Northern Hardwood Forest	10	–	0.002**
Fraser Fir Forest	104	–	0.002**
Appalachian Montane Oak–Hickory Forest (Red Oak Type)	37	–	0.002**
Red Spruce–Northern Hardwood Forest (Herb Type)	44	–	0.003**
Chestnut Oak Forest (Xeric Ridge Type)	28	–	0.004**
Chestnut Oak Forest (Mesic Slope Heath Type)	1	–	0.004**
Mountain Laurel (sparse canopy)	1	–	0.009*
Red Spruce–Northern Hardwood Forest	2	–	0.010*
Southern Appalachian Eastern Hemlock Forest (White Pine Type)	60	–	0.021*
Southern Appalachian Eastern Hemlock Forest (Typic Type)	58	–	0.024*
Early Successional Appalachian Hardwood Forest	7	–	0.031****
Appalachian Montane Oak Hickory Forest (Typic Acidic Type)	29	–	0.046****
High-Elevation Red Oak Forest (Evergreen Shrub Type)	6	+	0.022*
Montane Alluvial Forest (Cades Cove/Ocanaluftee)	24	+	0.003**

and management decisions, such as selecting restoration sites and balancing human use and species preservation goals in landscape planning.

The least-cost path is an improvement over the use of simple distance, as well as the typical human algorithm of minimizing off-trail distance, especially for more remote locations. This is illustrated by the increase in mean energetic cost as the cost associated with trails decreases. Making trails less costly (by one-half in the sensitivity analysis) increases the on-trail distances and reduces the off-trail distances for the least-cost model solutions. However, it usually increases total energetic cost compared to models with lower on-trail distances, because the off-trail portion is more costly than an alternative path that deviated from the trail at an earlier point. From this we can see that at a given distance, the ratio of on- and off-trail hiking affects the total energetic cost and the likelihood of human effects. Simple distance, then, is not an adequate measure of accessibility.

The most important parameter for estimating total energetic cost is the cost associated with walking up and down slopes. Our model is further supported by the fact that the correlation between energetic cost and slope is the most rigorously tested relationship in the model (e.g., Minetti et al. 2002). Nonetheless, the model may not include all factors responsible for the selection of a pathway for interior locations, because factors other than energetic cost may influence path choice. For example, footing is more precarious off-trail than on-trail, and a

hiker may try to minimize risk by staying on the trail longer than predicted by the energy-based model. We do not see an easy way of including those factors or of balancing them against energetic costs.

A second issue concerns the algorithm’s focus on the immediate neighborhood when calculating the least-cost path. The algorithm uses a radius of 10m for the next step in the path and, therefore, does not assess the larger-scale structure of the landscape. For instance, it may be possible to employ a series of switchbacks to get up a steep slope rather than merely find the shallowest gradient to follow along the slope. Nevertheless, the least-cost path is often an improvement over the typical algorithm of minimizing off-trail distance, especially for more remote locations, because we found that such choices generally increase mean energetic cost. The number of calculations needed to assess the larger-scale structure of the landscape is prohibitively large, since the number of potential pathways increases greatly as the search radius increases. An alternative may be to greatly increase the grain size of the data (in order to reduce the number of required calculations for larger search radii), but this choice would reduce resolution of the topographic data.

Sampling bias and accessibility

What are the implications of the accessibility sampling bias for data analysis? Reese et al. (2005) suggest that samples that occur near roads are

effective at predicting species distributions at a landscape scale. However, the generality of this result is dependent upon the spatial arrangement of species distributions and habitats relative to accessibility, the potential effects of humans on environment and species distributions, and beta diversity. We believe it is unwise to assume that these factors are unimportant. However, our model also creates a variable that can be used to structure field sampling in order to determine if bias in conclusions based on field observations is likely.

In GSMNP, most vegetation types and environments are not randomly arranged relative to accessibility. Some vegetation types have their greatest abundance in the more inaccessible areas of the park, while some are most abundant in accessible areas (Fig. 5). Even within broad vegetation types, such as all spruce-fir or all cove forests, correlations with accessibility were varied (Appendix S1). Thus, there is no reason to expect that sampling exclusively in accessible areas will capture a representative sample of all vegetation types in the park. The vegetation plots we have used do capture all the broad vegetation types, even though they were significantly biased for accessibility for two reasons: (1) sample size is large (1104 plots) and (2) the sample is aggregated from studies that may have targeted particular associations. A sample biased for accessibility that is smaller and random with respect to vegetation types would not capture those types at a rate equal to their area in the park.

Applications

As a general rule, habitat patches that are distant and isolated from each other tend to have differences in composition (Nekola & White 1999). This pattern holds true for GSMNP (Jobe 2008). Vegetation types exhibiting great spatial turnover in composition are more likely to differ between accessible and inaccessible areas. Therefore, exclusively accessible surveys will potentially capture only a portion of the landscape-scale variance in species composition for such vegetation types. While beta diversity would produce an expected bias in accessible plots (plots clustered in accessible locations would be less affected by species turnover than plots with greater spatial extent and dispersion), correlations of accessibility with environmental parameters (topography in this paper) would also mean that accessible plots would sample only a portion of the range of variation of any given vegetation.

Sophisticated models of accessibility are important because the correlations between accessibility and human effects are not always linear. The harvest of medicinal herbs, which is illegal in the park, serves as

example of such complexities. The goal of a poacher is to collect as many plants as possible without getting caught. They might choose to harvest plants in accessible areas because they can collect many herbs in a short time, and leave the park quickly. This would tend to deplete accessible relative to inaccessible populations. On the other hand, poachers may choose to harvest in inaccessible areas because they believe the probability of getting caught is low (the probability of finding undisturbed herb populations may also be high because of previous harvest in accessible areas). A practical application of the accessibility model would be to discern between those two foraging strategies by comparing the difference in population density between actual and potential distribution of herbs to accessibility.

Acknowledgements. This research was conducted while the author was supported by the W. C. Coker Fellowship from the Biology Department of the University of North Carolina at Chapel Hill and the Great Smoky Mountains National Park. The vegetation survey database for Great Smoky Mountains National Park used in these analyses was compiled by Jason D. Fridley, Robert K. Peet, Dean L. Urban, and Michael A. Jenkins, and Alan S. Weakley provided comments on an early draft.

References

- Adriaenssens, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H. & Matthysen, E. 2003. The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning* 64: 233–247.
- Anon. (ESRI) 2006. *ArcGIS 9.2, version 9.2*. ESRI, Redlands, CA, US.
- Anon. (NatureServe) 2006. Nature Serve Explorer: An online encyclopedia of life, NatureServe. Available at <http://www.natureserve.org/explorer/classeco.htm> (accessed 1 January 2008)
- Brody, A.J. & Pelton, M.R. 1989. Effects of Roads on Black Bear Movements in Western North-Carolina. *Wildlife Society Bulletin* 17: 5–10.
- Dabney, A. & Storey, J.D. 2008. qvalue: Q-value estimation for false discovery rate control, version 1.1.
- Gelbard, J.L. & Harrison, S. 2003. Roadless Habitats as Refuges for Native Grasslands: interactions with Soil, Aspect, and Grazing. *Ecological Applications* 13: 404–415.
- Gore, J.A. 1996. Discharge measurements and streamflow analysis. In: Hauer, F.R. & Lamberti, G.A. (eds.) *Methods in stream ecology*. pp. 53–74. Academic Press, San Diego, CA, US.
- Greenwood, J.J.D. 1996. Basic techniques. In: Sutherland, W.J. (ed.), *Ecological census techniques*. pp. 11–110. Cambridge University Press, Cambridge, UK.
- Grover, K.E. & Thompson, M.J. 1986. Factors influencing spring feeding site selection by Elk in the

- Elkhorn Mountains, Montana. *Journal of Wildlife Management* 50: 466–470.
- Hare, T.S. 2004. Using measures of cost distance in the estimation of polity boundaries in the Postclassic Yauatepec Valley, Mexico. *Journal of Archaeological Science* 31: 799–814.
- Imhof, E. 1950. *Gelaende und Karte*. Rentsch, Zurich, CH.
- Jenkins, M.A. 2007. *Thematic accuracy assessment: Great Smoky Mountains National Park*. National Park Service, Great Smoky Mountains National Park.
- Jobe, R.T. 2008. Estimating landscape-scale species richness: reconciling frequency- and turnover-based approaches. *Ecology* 89: 174–182.
- Larue, M.A. & Nielsen, C.K. 2008. Modelling potential dispersal corridors for cougars in midwestern North America using least-cost path methods. *Ecological Modelling* 212: 372–381.
- Madden, M., Welch, R., Jordan, T., Jackson, P., Seavey, R. & Seavey, J. 2004. Digital vegetation maps for the Great Smoky Mountains National Park. Final Report to the US Department of Interior National Park Service, CA 1443-CA-5460-98-019. Athens, GA, Center for Remote Sensing and Mapping Science, Department of Geography, The University of Georgia.
- McLellan, B.N. & Shackleton, D.M. 1988. Grizzly Bears and resource-extraction industries – effects of roads on behavior, habitat use and demography. *Journal of Applied Ecology* 25: 451–460.
- Minetti, A.E. 1995. Optimum gradient of mountain paths. *Journal of Applied Physiology* 79: 1698–1703.
- Minetti, A.E., Ardigo, L.P. & Saibene, F. 1993. Mechanical determinants of gradient walking Energetics in Man. *Journal of Physiology-London* 472: 725–735.
- Minetti, A.E., Ardigo, L.P. & Saibene, F. 1994. Mechanical determinants of the minimum energy-cost of gradient running in humans. *Journal of Experimental Biology* 195: 211–225.
- Minetti, A.E., Ardigo, L.P., Capodaglio, E.M. & Saibene, F. 2001. Energetics and mechanics of human walking at oscillating speeds. *American Zoologist* 41: 205–210.
- Minetti, A.E., Moia, C., Roi, G.S., Susta, D. & Ferretti, G. 2002. Energy cost of walking and running at extreme uphill and downhill slopes. *Journal of Applied Physiology* 93: 1039–1046.
- Nekola, J.C. & White, P.S. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26: 867–878.
- Reese, G.C., Wilson, K.R., Hoeting, J.A. & Flather, C.H. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications* 15: 554–564.
- Rose, J., Ralston, H.J. & Gamble, J.G. 1994. *Energetics of Walking*. v. 2nd ed. Williams & Wilkins, Baltimore, MD, US.
- Rost, G.R. & Bailey, J.A. 1979. Distribution of mule deer and elk in relation to roads. *Journal of Wildlife Management* 43: 634–641.
- Schadt, S., Knauer, F., Kaczensky, P., Revilla, E., Wiegand, T. & Trepl, L. 2002. Rule-based assessment of suitable habitat and patch connectivity for the Eurasian lynx. *Ecological Applications* 12: 1469–1483.
- Soule, R.G. & Goldman, R.F. 1972. Terrain coefficients for energy cost prediction. *Journal of Applied Physiology* 32: 706–708.
- Storey, J.D., Taylor, J.E. & Siegmund, D. 2004. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 66: 187–205.
- Tobler, W. 1993. Three Presentations on Geographical Analysis and Modeling: Non-isotropic modeling, speculations on the geometry of geography, global spatial analysis, National Center for Geographic Information and Analysis.
- Trombulak, S.C. & Frissell, C.A. 2000. Review of ecological effects of roads on terrestrial and aquatic vegetation types. *Conservation Biology* 14: 18–30.
- Van Leusen, P.M. 2002. *Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes*. University of Groningen, Groningen, NL.
- Weakley, A.S. 2008. *Flora of the Carolinas, Virginia, Georgia, northern Florida, and surrounding areas*. University of North Carolina Herbarium (NCU), North Carolina Botanical Garden, University of North Carolina at Chapel Hill, Chapel Hill, NC, US.
- Whittaker, R.H. 1956. Vegetation of the great smoky mountains. *Ecological Monographs* 26: 1–80.
- Whittington, J., St Clair, C.C. & Mercer, G. 2005. Spatial responses of wolves to roads and trails in mountain valleys. *Ecological Applications* 15: 543–553.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Vegetation types of GSMNP used for sorted by frequency in the park. Median energetic cost, and direction of correlation with accessibility for significantly accessible (–) and inaccessible (+) vegetation types are shown.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 21 February 2009;

Accepted 3 July 2009.

Co-ordinating Editor: V. Pillar.