Using Spatial Analysis of Emigrant Trails to Predict Likely Corridors Where Emigrants May Have Traveled

The southwest corner of Wyoming is filled with beautiful and remarkable terrain. There are long north-southrunning mountain ranges, mountain passes, several large rivers and streams, and rolling hills. It is rugged terrain with elevations ranging from 6,037 feet to 10,298 feet. Woven through this area's mountain passes and along its streams are 953.2 miles of mid-nineteenth-century emigrant trails.

By analyzing the spatial characteristics of those trails and their spatial relationship to the terrain they cross, we gain a greater understanding of emigrant trails at large. A spatial accounting of these trails—that is, the characteristics and nature of the space they occupy—reveals that emigrants rarely climbed hills greater than 5 degrees of slope and almost never more than 10 degrees; emigrants favored smooth terrain far more often than not; and more than 30 percent of the emigrant trails are within 1,000 feet of a stream, with over 80 percent within 10,000 feet. Moreover, using what we learn from this area's trail analysis, we can create computer-generated corridors between geographic points that model where emigrants would likely travel anywhere within the greater emigrant trail system. Historic trail researchers can use these corridors as tools to locate areas where trail segments or historic trails may be.

This article has two parts. First, a spatial analysis of the emigrant trails and the terrain they crossed is presented. The focus of the analysis is on the slope of the trails, how rugged those trails were, and how far from streams the trails tended to be. From this analysis, we gain an understanding of the characteristics of trails emigrants used. Second, using the data obtained from the spatial analysis, we will create a corridor where emigrants would likely travel between any two geographic points.

The spatial analysis will cover the trails and terrain contained in a rectangle extending from near the southern Wyoming border to a few miles north of Cokeville and from the western border of Wyoming to near the South Pass.¹ Within this study

The geographic coordinates for this area are: 111.05° west longitude along the western edge, 42.42° north latitude along the northern edge, (continued)



area is a network of trails including the California, Oregon, Mormon, and Cherokee trails, and the Sublette, Blacks Fork, Dempsey-Hockaday, Slate Creek, Kinney, and Hams Fork cutoffs. The Lander Trail is also in this area, but it is excluded from analysis because of the engineering that went into its making. Although we will confine our *analysis* to this area, we are not restricted in producing likely corridors anywhere within the greater trail system. To that end, as an example, a likely emigrant trail corridor is produced in the Donner Pass area.

Entering this southwest Wyoming area from the east, a mid-nineteenth-century emigrant spent weeks following a labyrinth of trails heading west. Emigrants were presented with a variety of choices of where to travel. They could stay on the main trails, or they could travel on a cutoff, a cutoff modification, or even a cutoff that cuts off a cutoff. Eventually, emigrants exited the network at one of two points: north of present-day Cokeville in the Bear River valley near Thomas Fork, or a southern point near Yellow Creek along Wyoming's western border. Essentially, the sum of this trail network forms a large, complex manifold of trails (see figure 1).

To analyze terrain and trails in this area, geographic information system (GIS) technology is used. A GIS is "a collection of computer hardware, software, and geographic data for capturing, storing, updating, manipulating, analyzing, and displaying all forms of geographically referenced information."² Within our study area, we will analyze the emigrant trails based on three characteristics: slope, ruggedness, and distance from streams.

^{108.82°} west longitude along the eastern edge, and 41.07° north latitude along the southern edge.

Anne Kelly Knowles, ed., *Past Time, Past Place: GIS for History* (Redlands, Calif.: ESRI Press, 2002), 186.

OPPOSITE FIGURE 1. Southwest Wyoming emigrant trails. The study area is within the black rectangle. The background is a subdued image representing slope, with darker grey for higher slope. ALL MAPS BY ROBERT E. DAVIS.

RIGHT FIGURE 2. Portion of the Dempsey-Hockaday Cutoff following low slope areas on a ridgeline as it makes its way through Commissary Ridge. An electronic USGS 1:24,000 scale is the basemap.

The backbone of terrain analysis is a computer file called a Digital Elevation Model (DEM). A DEM represents the surface of the earth. The DEM used in this study is from the U.S. Geological Survey (USGS).³ This DEM can be imagined as a giant checkerboard of 30-by-30-foot squares laid down atop the entire study area.⁴ These checkerboard, or matrix-type files are sometimes called rasters. Each square of the DEM is coded with the elevation of the surface of the earth it lies atop. Those elevations are used for spatial analysis. For instance, we compute slope by subtracting the elevation from two adjacent squares (rise), then dividing by the distance between the centers of the squares, 30 feet (run).⁵ The DEM used to analyze this study area has 377,406,810 squares, or cells, so a system of software, hardware, and data—a GIS—is needed to turn that mass of information into something manageable and understandable.

Using GIS,⁶ we compute the slope between each of those 377 million cells. In so doing, we create another raster of the



same size and spatial characteristics, but with slope as the value coded into each cell. Shading is assigned to the cells based on this slope value. Combined, the shaded cells of the raster create an overlay, or layer, that the GIS can display on a map with other layers, such as a USGS topographic map (see figure 2).

Just by looking at figure 2, we can see that the trail tends to follow areas with lower slope. While a visual interpretation like this important, we need a more complete numerical analysis to gain a fuller understanding of the relationship between slope and these trails. To do this, we use another computer file called a shapefile. Shapefiles are a series of short connected lines, the sum of which represents a longer line, or in this case, an emigrant trail. The shapefile used was created by the U.S. Bureau of Land Management using digitizing and Global Positioning Systems (GPS).⁷ The shapefile contains two-dimensional

³ U.S. Geological Survey, "The National Map: Elevation," accessed January 2, 2016, online at http://nationalmap.gov/elevation.html.

⁴ The precise size of a square is 29.64×29.64 feet.

⁵ Throughout this paper, I refer to slope in degrees.

⁶ Environmental Systems Research Institute (ESRI), ArcGIS, version 10.3 (ESRI, 2015), www.esri.com. Throughout this analysis, ESRI's ArcMap GIS software was used.

⁷ U.S. Bureau of Land Management, "Historical trails," accessed January 2, 2016, online at http://www.blm.gov/wy/st/en/resources/public_room/gis/ metadata/historical_trails.html. This shapefile is topologically incorrect. A small number of lines representing trails do not join with the lines preceding and following them. Furthermore, there are small "spurs" and "slivers" of trail segments that go nowhere. There are other minor problems as well. Nevertheless, the sum of the problems represents less than 0.3 percent of the total length of trail in our study area. Rather than "fix" those small errors, the author let them stand so as not to disturb the original U.S. BLM data.









spatial information—direction and length—for each short line.⁸ We add more information to each line, such as slope, through GIS techniques that extract information from underlying layers. To add slope to our shapefile, we divide the shapefile into 14,000 small segments, each about 300 feet long. Slope information is then derived from our new slope layer and added as an attribute to each line segment. In so doing, we move from a visual interpretation of emigrant trails to numerical analysis of their characteristics.

Grouping the 14,000 line segments by ranges of slope (0–5 degrees, 5–10 degrees, and greater than 10 degrees) and calculating the number of trail miles within each group, we develop a histogram showing the number of trail miles in each group (see figure 3). Of the 953.2 miles of trail in the study area, there are 903 miles with slope less than 5 degrees, 46 miles with slopes between 5 and 10 degrees, and only 5 miles with slopes in excess of 10 degrees. Indeed, 95 percent of the trails in the study have slopes less than 5 degrees. Emigrants reluctantly assailed a slope of greater than 5 degrees, and almost never over 10 degrees. Emigrants' reluctance to travel over high slopes is demonstrated by lore as well as numerical analysis. Areas of high slope are infrequent enough along the entire emigrant trail network that they were given names: California Hill, Windlass Hill, and Roller Pass are examples.

The next statistic that helps shed light on this trail system is a measure of ruggedness. Here, ruggedness is defined as the variance in elevation surrounding a geographic point. Since we will use the DEM to determine ruggedness, that geographic point is the center of each of the DEM's cells. If a cell's adjacent cells have very different elevations from its elevation, then ruggedness is said to be high. If the adjacent cell's elevations are the same or very nearly the same as the cell's elevation, then the ruggedness is said to be low. For instance, if a cell's elevation is 8,000 feet, and the adjacent 8 cells have elevations of 8,006 feet, 8,010 feet, 7,994 feet, 8,008 feet, 8,006 feet, 8,003 feet, 7,995 feet, and 8,009 feet, we see that variance between cells is quite high, and therefore rugged. Standard deviation is a statistical tool used to understand variance. In the above case, the standard deviation of the cell and its adjacent cells is 5.6. On the other hand, if the surrounding cells have more moderate elevations of 8,001 feet, 7,999 feet, 7,998 feet, 8,000 feet, 8,001 feet, 7,999 feet, 8,001 feet, and 8,000 feet, we see that variance is lower, as is the standard deviation of 0.99.

A GIS is used to calculate the standard deviations of each of the 377 million cells and its adjacent neighbors. In so doing, a new raster is created with standard deviation as each cell's value. As with slope, standard deviation (ruggedness) is derived from the new raster and added as an attribute to each small segment of our trail shapefile. A histogram of trail ruggedness (see

⁸ Because of these shapefile attributes—direction and length—shapefiles are sometimes called vectors.

TABLE 1. SPATIAL DATA BY EMIGRANT TRAIL WITHIN THE STUDY AREA ⁹					
Trail	Length (mi.)	Max. Elev. (ft.)	Min.Elev. (ft.)	Mean Slope (degrees)	Ruggedness
Oregon-California Trail	389.6	8,108	6,072	1.27	0.37
Sublette Cutoff	180.2	8,458	6,204	2.29	0.53
Southern Cherokee Trail	90.2	7,969	6,044	2.63	0.58
Northern Cherokee Trail	69.5	7,220	6,179	1.25	0.37
Slate Creek Cutoff	67.8	7,838	6,349	1.05	0.28
Hams Fork Cutoff	53.1	7,687	6,264	1.24	0.37
Kinney Cutoff	52.1	6,600	6,293	0.64	0.17
Blacks Fork Cutoff	33.7	6,554	6,263	0.83	0.26
Dempsey-Hockaday Cutoff	16.9	8,731	7,250	3.04	0.80
					0

figure 4) shows most of the trails have low ruggedness, with standard deviations of less than 1.0.

We can separate each of the named trails and analyze their length, minimum and maximum elevation, average slope, and ruggedness. As part of this process, the maximum and minimum elevation along a trail were revealed. The maximum elevation is 8,731 feet near the start of the Dempsey-Hockaday cutoff of the Sublette Trail. (This is much higher in elevation than Roller Pass in the Sierra Nevada.) The minimum elevation is 6,044 feet just north of Currant Creek Ridge (near Buckboard Crossing) along the Southern Cherokee trail. Table 1 lists spatial statistics by trail.

The above table represents some trail statistics based on

the physical characteristics of the terrain. Other information about the trails is derived from their spatial relationship to other features of the landscape. For instance, we find almost all (83 percent) of the trails lie within 10,000 feet of a centerline of a stream, with almost a third (31.2 percent) within 1,000 feet of a stream. To derive this spatial information, a shapefile describing the centerline locations of streams was obtained from the USGS National Hydrography Dataset.¹⁰ Using a GIS, the stream centerlines were increased in width to 2,000 feet; or to put it another way, the stream centerlines were "buffered" 1,000 feet both left and right (see figure 5). The GIS is able to determine where the buffered streams intersect with the emigrant trail shapefile and to calculate how many miles of trail are intersected. From this we find there are 297.8 miles of emigrant trail in Table 1 within 1,000 feet of a stream centerline.

⁹ Andy Mitchell, The ESRI Guide to GIA Analysis. Modeling Suitability, Movement, and Interaction, Volume 3 (Redlands, Calif.: ESRI Press, 2012), 218. All distances are Euclidean vice planar. That is, the distance from two points, A and B, is calculated along the slope between them.

U.S. Geological Survey, "National Hydrography Dataset," accessed January 2, 2016, online at http://nhd.usgs.gov/data.html.

Another histogram is developed showing the combined length of trail based on distance from a stream centerline. The histogram reveals the emigrants' bias for staying close to streams, as they were vital in providing water and grass. In fact, fully 98 percent of the trails are within 35,000 feet, or 6.6 miles, or roughly a half day's wagon travel of a stream. The only parts of trails not within 35,000 feet of a stream are 14 miles along the Sublette Cutoff and 3 miles along the northern elbow of Slate Creek Cutoff.

We now have three sets of information that help describe both the terrain and the trails within our study area: slope,





ruggedness, and distance from streams. We have learned from this analysis that emigrants favored crossing smooth terrain with low slope and close to streams. Using this set of information, we can model, or predict, how emigrants might move between any two geographic points within the study area. However, to model human movement across terrain, we must accept the premise that humans will generally follow the path of least *cost* (we define "costs" as impediments to movement). We see this in our daily lives as we tend to go around obstacles, follow smooth paths, and move between two points directly.

Defining what the costs are, classifying those costs, grouping

those separate costs into a single cost associated with movement through terrain, and then creating a corridor between two geographic points are the four general steps to modeling movement across terrain.¹¹ A single element of cost, such as slope, is developed into a cost layer. Each cost layer is classified into a standardized rubric of 1, 2, and 3, for low, medium, and high costs. Grouping the separate costs into a single cost layer creates a unified cost surface describing the combined cost of movement. Finally, we create a corridor between two geographic points where human movement would be most likely.

Our efforts of analysis have already created the cost layers. They are slope, ruggedness, and distance from streams. While a greater distance from a stream may not constitute a greater cost

11 Mitchell, ESRI Guide, 214.

TOP FIGURE 5. Portion of the Dempsey-Hockaday Cutoff showing buffered streams and those parts of the trail intersecting the buffered streams. An electronic USGS 1:24,000 scale is the basemap.

LEFT FIGURE 6. Number of miles of emigrant trail within various distances from the centerline of a stream.



to movement in and of itself, emigrants acted as if it did. Emigrants overwhelmingly stayed close to streams (see figure 6). Other costs increased if they did not: the cost of having less water, less food for the animals, and less firewood all accumulated the farther they ventured from streams. Because of this, we consider movement away from streams more costly.

Next, we classify these costs into a standardized scheme. Each cost layer is classified on a scale of 1 to 3, where 1 is a low cost and 3 is a high cost. For instance, slopes between 0 and 5 degrees are classified as a "1," where slopes greater than 10 degrees are classified as "3." The histograms associated with these costs were used to determine the break points between classifications. Table 2 summarizes our classification scheme for the cost layers.

We create a single "cost surface" raster by multiplying the classified values (1, 2, or 3) from each cost layer. The cost-surface raster has the same spatial dimensions as our original DEM but with the product of the cost layers coded into each cell. In

an area of low slope, low ruggedness, and within 2,000 feet of a stream, the cost of crossing a 30 by 30-foot cell would be 1 unit $(1 \times 1 \times 1)$, and this value is coded to that cell. On the other hand, if the slope and ruggedness were both low, but the distance from a stream were 4,000 feet, then the cost associated with crossing a cell would be 2 $(1 \times 1 \times 2)$. A value of 27 indicates an inhospitable cell where the slope is high (3), the ruggedness is high (3), and it is a long way from a stream (3). Throughout our new 377-million cell raster, each cell's value is assigned a number between 1 and 27 based on the product of the slope, ruggedness, and distance from a stream.¹² Assigning shades to the cells—from white for a value of 1 to black for a value of 27—reveals a visual interpretation of how difficult the terrain is to cross (see figure 7).

Finally, using the cost surface, we can determine the corridor requiring the least cost to traverse between two geographic points. In other words, we use GIS to determine the easiest corridor (or path) between points A and B. We call this the least-cost corridor. The GIS models the movement from one geographic point to another, adding the value of each cell from the cost surface along the way, then determines the sequence of cells between the two geographic points with the least accumulated cost. As an example, we can use the GIS to create a least-cost corridor between a geographic point on the Sublette Cutoff just to the east of Slate Creek Ridge and another point on the eastern side of Rocky Gap. Figure 8 shows that the leastcost corridor determined by the GIS very closely follows the existing trails.

Corridor modeling will work in other places along the emigrant-trail network as well. After all, what we are modeling is not the trails so much as the emigrants' decision making as they made their way across terrain. Decisions regarding slope, distance from streams, and ruggedness would likely be the same in the southwest corner of Wyoming as in the Donner Lake area, for example, or any other area along the trails. Indeed, using the same GIS techniques, a least-cost corridor was created from the east Donner Lake area across the Sierra Nevada passes to the Lake Van Norden meadows (these meadows were

¹² Not all the numbers between 1 and 27 are included. Excluded are the prime numbers 5, 7, 11, 13, 17, and 19. Also, the numbers 10, 14, 15, 16, and 20 through 26 cannot be a product of any combination of 1, 2, and 3. This leaves us with 1–4, 6, 8, 9, 12, 18, and 27.

FIGURE 7. Visual representation of the cost required to traverse a piece of terrain. Each cell of a raster is assigned a value between 1 and 27, determined by the cost, or effort, required to traverse that cell.

FIGURE 8. Computer generated least-cost corridor between two points (start and end) in the Rocky Gap area. It is important to keep in mind that the computer knew nothing about the trail between these two points. Note the subdued cost surface raster in the background.







FIGURE 9. Least-cost corridor between the eastern Donner Lake area and the meadows around Lake Van Norden. Using the analysis gained from the study of trails in southwest Wyoming, the GIS plotted a corridor very near the emigrant trails in California. USGS 1:24,000 topographical map is used as a basemap.

the first level, grassy area emigrants encountered immediately to the west of the Sierra passes along the Truckee Route), utilizing the criteria of slope, ruggedness, and distance away from streams—derived from the study area in Wyoming.

When the GIS first developed a corridor model across the Sierra Nevada passes, it "saw" Donner Lake as a magnificently flat piece of terrain and promptly shot right across it, then made its way up the steep slopes to cross the ridgeline, and then meandered to Lake Van Norden. To keep the GIS from modeling a path across lakes, another cost layer was built, assigning a cost of 1,000 to each cell over a body of water.¹³ This essentially made lakes "no-go" terrain. Running the model again, this time with the lake no-go layer added, the GIS created a least-cost corridor following Coldstream Canyon, Emigrant Canyon, over Roller Pass, and then to Lake Van Norden (see figure 9). This least-cost corridor closely follows the trails as they are understood today; however, the corridor is not perfect. On the west side of Roller Pass, the known trails make a couple of switchbacks to soften the descent, while the GIS-least-cost corridor simply flows down the hill. Nevertheless, the GIS correctly predicted the major parts of the route: Coldstream Canyon, Emigrant Canyon, and Roller Pass. So, at least in this area of California, the data derived from southwest Wyoming is effective in predicting at least one emigrant trail corridor.

This method of corridor prediction is only a tool. It should be used like any other tool, to achieve an end—a reasonable corridor prediction—in concert with knowledge of the subject and common sense. Refinements to this tool and how it is wielded are sometimes necessary to achieve results that make sense in an historical context. For instance, if corridor modeling results in a corridor that fits poorly with the bulk of the historical narrative around it, that is, other maps, written sources, or physical artifacts, the corridor should be questioned. Perhaps adjustments can be made to corridor modeling that will result in a least-cost corridor that fits better with the bulk of other historical information. There are two methods of adjusting the tool: narrowing or expanding the inputs and weighing the inputs.

In our analysis of southwest Wyoming trails, we analyzed slope, ruggedness, and distance from a stream. Those three analysis topics were then used as inputs to corridor modeling.

¹³ Another shapefile containing water bodies was downloaded from the USGS National Hydrography Dataset to accomplish this.

TABLE 3. POTENTIAL COST LAYERS

Potential cost layer	Technique	
A well, trading post, fort, ford, or other locale	These features should be modeled as geographic points and a corridor fror one point to another built between them.	
Avoiding multiple river crossings	Buffer a river shapefile by several hundred feet and assign a moderate cost within that buffered area. Doing this has the added benefit of preventing the corridor from following the centerline of a meandering stream.	
Water bodies	Assign a very high cost for the surface of a lake or other water body.	
The edge of a river valley	Areas at the edge of a river's flood basin tend to be flat in the direction of the river and have a sharp incline perpendicular to the river at the valley edge. A layer that shows low slope and high standard deviation will mark these areas. ¹⁴	
An area of high resources	In the emigrant context, these areas tend to be large areas of grass. These can be modeled either as a locale, or if the area is large, build a cost layer where the grassy area has a low classification value (1) and all other areas have a higher classification value. Since these areas tend to be coincident with rivers and streams, just the stream distance cost layer should suffice.	
Areas of high altitude	A cost layer is built where there is a greater cost for traveling over areas of high altitude. This technique can be used to model travel across areas of high latitude as well.	
Ridgelines	Build a ridgeline shapefile by using the GIS's hydrological tools. Mark ridgelines with a high or low classification value depending on circumstances.	
	Robert E. Davis, "Southwest Wyoming Nineteenth Century Emigrant Trail Analysis" (Tucson: University of Arizona, 2014), 21. The author gratefully acknowledges Dr. Gary Christopherson of the University of Arizona Geography Department who suggested this technique.	

There's no requirement that all three cost layers be used for the modeling to work. In many cases, especially where water and grass are plentiful thought the area, better results may be achieved by eliminating the importance of nearness to streams. Often just the slope cost-layer is information enough for the the numbers 1, 2, and 3 to classify slope (see Table 2), we could choose 1, 2, and 1,000. Doing so would essentially prevent GIS from developing a corridor where the slope was greater than 10 degrees. Care should be taken when assigning weights. Too much weight added to one cost layer while minimizing the

GIS to determine a corridor that one might reasonably take. This should come as no surprise, as our analysis demonstrated that emigrants were severely limited in the amount of slope their wagons could assail. Other layers can be built to represent other influences on travel; as we found with the Sierra Nevada example, a fourth input was needed (lakes as no-go terrain) for the model to create a leastcost corridor that made sense. While not particularly germane for the gold rush emigrants, a layer could be built to place a higher cost on winter travel at higher elevations or latitudes. Additional layers can be built to

Though I do not think the terrain she described was so miraculous, by and large, emigrants did find those shorter, grassy, level trails. What is miraculous is that they started the journey at all.

weight of another may have the same effect as eliminating a cost layer altogether.

Modeling movement over terrain is as much art as science. While the computer will strictly follow mathematics and logic, humans making their way across terrain may not. Additionally, humans conducting the analysis will always affect the input variables, as we have done here by dividing the costlayer categories where we did and adding another layer for lakes. Furthermore, conditions on the ground when the trails were first cut (high water or low water in a river basin, for instance) may have dictated a

increase cost for each stream or river crossed. Table 3, while not exhaustive, shows some additional layers that can be built to represent cost.

"Weighing" in this context is adjusting the relative importance of one or more of the cost layers. Since the cost-surface layer is based on the product of the cost layers, simply multiplying a cost layer by a constant will not work, as that procedure is the same as multiplying the entire cost surface by the constant, and the resulting corridor will show no practical difference from the unweighted corridor.¹⁵ What can be done, however, is to assign different values for each classification of a cost layer. Instead of 1, 2, 3 for classification values, 1, 4, 9 could be used if we want a layer to have an exponential weight. Or, we could weigh a layer in an effort to eliminate a corridor at the higher end of the cost layer. For instance, instead of using different path from the GIS logically created based on today's geospatial information. As such, these least-cost corridors are sometimes imperfect. Nevertheless, in the above examples, the GIS did a remarkably good job of finding essentially the same corridor the mid-nineteenth-century emigrants did.

Emigrant-trails researcher and author Irene Paden points out in *Prairie Schooner Detours* that wagon masters "prayed for ... a miraculous piece of road—grassy, level, and shorter than the one by which everyone else was plodding along."¹⁶ In one clause she aptly points out what I've been trying to capture in these last forty-five hundred words. Though I do not think the terrain she described was so miraculous, by and large, emigrants did find those shorter, grassy, level trails. What is miraculous is that they started the journey at all.

Recall the associative law of multiplication whereby the grouping of numbers of variables makes no difference to the product. So, $A \times B \times (C \times 2) = 2 \times (A \times B \times C)$.

¹⁶ Irene D. Paden, Prairie Schooner Detours (New York: Macmillan, 1949), ix.