

# **Modelling potential environmental impacts of science activity in Antarctica**

David O’Sullivan

University of Auckland, Geospatial Stuff  
david.osullivan@auckland.ac.nz

Fraser J. Morgan

Manaaki Whenua Landcare Research

## **Abstract**

We use GPS data collected on a science expedition in Antarctica to estimate hiking functions for the speed at which humans traverse terrain differentiated by slope and by ground cover (moraines and rock). We use the estimated hiking functions to build weighted directed graphs as a representation of specific environments in Antarctica. From these we estimate using a variety of graph metrics—particularly betweenness centrality—the relative potential for human environmental impacts arising from scientific activities in those environments. We also suggest a simple approach to planning science expeditions that might allow for reduced impacts in these environments.

## **1 Introduction**

Antarctica is often considered a vast, desolate continent devoid of human activity. However, there are small, scientifically significant areas where substantial, regular human activity occurs. Globally unique ice-free areas within Antarctica, such as the McMurdo Dry Valleys, are of particular scientific interest and so a significant amount of activity occurs in this specially managed area. Human activity in Antarctica is highly dependent on the infrastructure that supports people to survive in this environment. Outside of a reduced period of activity caused by COVID-19, human activity on the continent continues to grow (Tejedo et al. 2024) and so the actual, and future potential impact of these activities continues to grow. The McMurdo Dry Valleys are close to the largest research base on the continent, McMurdo Station, and the pressures on this unique environment from human activity are increasing. Understanding the scope and implications of human activity in key parts of Antarctica is critical to supporting environmental management of the continent.

The history of understanding the role and impacts of human activity in Antarctica starts with the Sixth Antarctic Treaty Consultative Meeting (ATCM) in 1970. A recommendation of the final report of this meeting was for member countries to encourage research on how human activity impacts Antarctica and its ecosystems (ATCM 1970, Recommendation VI-4). By the mid-1980s, it was recognised that scientific activity, particularly the logistics required to support that activity, required additional scrutiny, and also that scientific activity could pose a greater threat to Antarctica’s ecosystems than tourism (Benninghoff and Bonner 1985). At this point, the need for more understanding and the ability to forecast the environmental impact of human activity was recognised (ATCM 1985, Paragraph 46).

Moving into the mid-1990s, the multi-disciplinary scope of large programmes of research such as the McMurdo Dry Valleys Long Term Ecological Research (MCM LTER) programme, increased the

number of researchers in the field while also supporting more research into the history of human activity in the Dry Valleys (Priscu and Howkins 2016). With the establishment of the Committee for Environmental Protection (CEP) in 1998, human activities (including science activities) were to be monitored, but attention moved away from the understanding and scope of the activities, and became more focused on potential resulting impacts of those activities. For example, the current 5-year Work Plan for the CEP (REF), has several items associated with human activities but these are focused on managing the impact of human activities rather than quantifying the current level of activity, or forecasting cumulative impacts of the activity.

More recent research has focused either on a continental-scale picture of human activity (Leihy et al. 2020), or detailed mapping of the infrastructural footprint and impacts that the numerous Antarctic stations have on the continent (Brooks et al. 2018; Brooks et al. 2019). These approaches tackle the challenge of understanding human activity from contrasting spatial perspectives. The present research aims to complement these two techniques, by using high resolution mapping of human activity in the dry valleys to extrapolate how scientists navigate in these landscapes, or could potentially navigate in these landscapes.

In the next section we outline our approach and discuss related work. We then describe data sources, both general environmental and topographic, and the detailed GPS data from which our findings are derived. Procedures for deriving hiking functions and applying them to build hiking networks as representations of the dry valleys landscapes are explained, followed by examples of analyses possible using this approach, particularly the application of measures of graph betweenness centrality as proxies for potential impact. We also outline by which these methods might be used to plan for reduced impact from scientific expeditions to these environments. The paper concludes with a discussion of our findings, prospects for further work, and some general conclusions.

## 2 Our approach and related work

We chose to explore the question of where human impacts are likely to be strongest using an approach closely related to work on patterns of human movement in archaeology (Verhagen et al. 2019) where likely and potential movement paths of humans across landscapes have been used to infer the settlement structure and human geography of large-scale landscapes. Closely related work in biology investigates the structure and geography of animal transportation networks (Perna and Latty 2014). Both approaches rely on the idea that humans or animals move around in an environment in time or energy efficient ways. These approaches rely on hiking functions that relate speed of movement across a terrain to its slope.

Hiking functions must be applied in some context where locations across a landscape are connected to one another. Because hiking functions are asymmetric, with estimated speed of movement up slopes different than estimated speeds down the same slope, landscape must be represented in a way that allows for this asymmetry. We therefore represent terrain in our landscapes as directed graphs (or network) of connections between locations regularly distributed in planimetric space across the landscape of interest. Because the graphs are directed the costs associated with movement between two locations can be different depending on the direction of movement.

Additionally, we associate with locations (i.e., vertices in the graph) the ground cover at the location, which also affects the speed at which it can be traversed. Because the ground cover in the Antarctic environments under study can be broadly categorised into only two types, moraine and rock, we use the ground cover of a location to switch between two estimated hiking functions, rather than the more widely used approach of penalising movement on different ground covers by applying cost factors. We consider previous work on hiking functions and directed graphs in more detail below.

## 2.1 Hiking functions

Prisner and Sui (2023) provide an overview of a variety of functions that have been used to model how hiking times and speeds vary with terrain slope. They consider longstanding rules of thumb (Naismith 1892), and later modifications thereof (Langmuir 1984), along with more recent such guidance from the Swiss and German Alpine Clubs (Winkler et al. 2010; Deutscher und Österreichischer Alpenverein 2011). These functions estimate the time taken to travel 1km, referred to as *pace*, based on slope expressed as *rise over run*, that is change in elevation divided by horizontal distance. They are all piecewise functions with sharp changes in estimated pace at specific slopes.

Alongside these hiking pace functions Prisner and Sui (2023) also present hiking speed functions (generally referred to as simply *hiking functions*) from Tobler 1993 and more recent, related but more firmly empirically grounded alternatives offered by Márquez-Pérez et al. (2017), Irmischer and Clarke (2018), and Campbell et al. (2019). Another hiking function not discussed by Prisner and Sui (2023) is presented by Rees (2004). These hiking functions are all continuous in the slope of the terrain so that  $v = f(\theta)$ , where  $v$  is the speed, and  $\theta$  is the slope. They can all be parameterised to control the maximum speed attainable, the slope at which maximum speed is attained (expected to be a shallow downhill slope), and the rate at which speed falls off with increasing slope.

The functional forms of some hiking functions are shown in Table 1 and graphed in Figure 1.

Table 1: Functional forms of hiking functions

Description	Equation	Examples
Exponential	$ae^{b \theta-c }$	Tobler (1993), Márquez-Pérez, Vallejo-Villalta, and Álvarez-Francoso (2017)
Gaussian	$ae^{-b(\theta-c)^2}$	Irmischer and Clarke (2018), Campbell, Dennison, Butler, and Page (2019)
Lorentz	$\frac{a}{[b+d(\theta-c)^2]}$	Campbell, Dennison, Butler, and Page (2019)
Quadratic	$a + b\theta + c\theta^2$	Rees (2004)

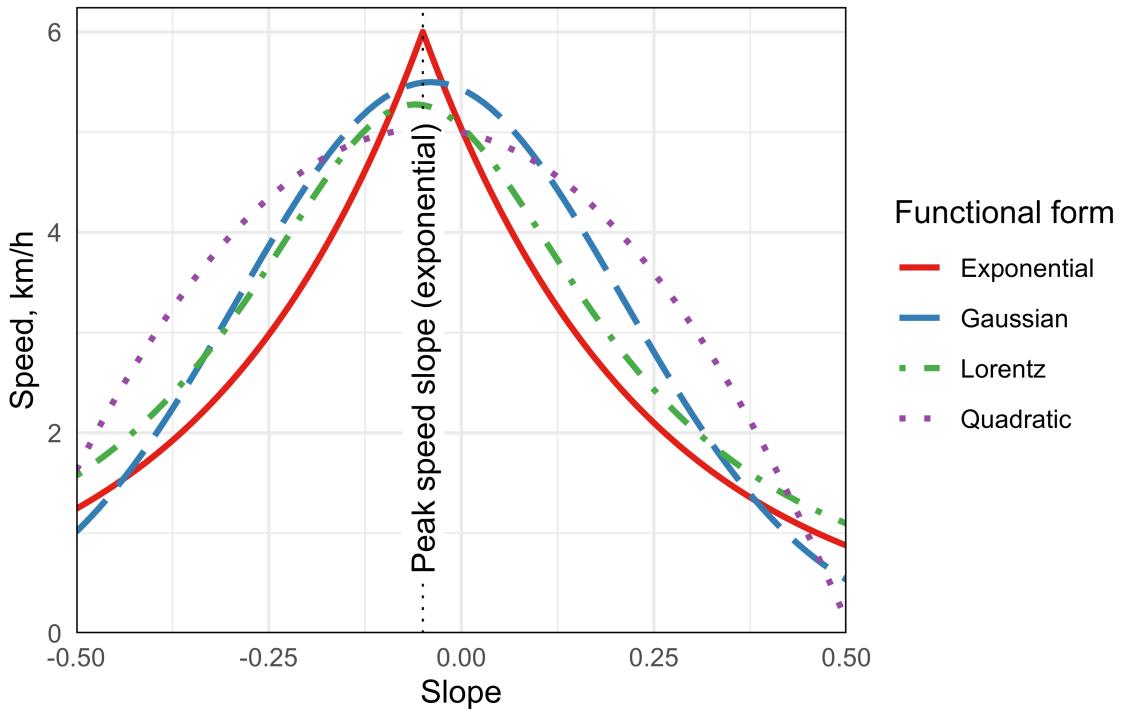


Figure 1: Example hiking functional forms: Exponential (Tobler 1993), Gaussian (Irmischer and Clarke 2018), Lorentz (Campbell, Dennison, Butler, and Page 2019), and Quadratic (Rees 2004).

The parameterisations of the functions in Figure 1 have been chosen for illustrative purposes only, although the parameter values for the exponential function shown are such that  $v = 6e^{-3.5|\theta|+0.05}$  as suggested in Tobler (1993). These parameters give *Tobler's hiking function* but, as has been noted elsewhere, (Herzog 2010; Campbell, Dennison, Butler, and Page 2019), are based on a poorly specified fit to secondary data on hiking speeds found in Imhof 1950. Nevertheless, these parameter values are widely applied in the literature.

All these hiking functional forms are somewhat *ad hoc*. They exhibit desirable, expected properties: a peak speed at a slope near zero, which we expect to be slightly negative (i.e., downhill), and continuously falling speeds as the slope deviates from the slope of peak speed. However, there is no theoretical basis for the specific functional forms listed in Table 1. More principled approaches might be developed based on the literature on the physiology of human movement (Minetti et al. 2002). In general, approaches based on minimising energy use yield similar results to empirical speed-slope functions, although it is worth noting that they more reliably generate zig-zag or ‘switchback’ movement behaviour on steep slopes (Llobera and Sluckin 2007). However, these are hard to implement, and we have adopted empirically-derived and locally-specific hiking functions following Márquez-Pérez, Vallejo-Villalta, and Álvarez-Francoso (2017), Irmischer and Clarke (2018), and Campbell, Dennison, Butler, and Page (2019). This choice is based on available data and the goals of our study, where the relative cost of different potential routes in the landscape is more important than exact prediction of routes. In practice, almost any function with the peaked form of those shown in Figure 1 is suited to our requirements.

It is commonplace in many applications to also incorporate a penalty on movement contingent on land cover, especially for off-track or off-road movement. For example the speed attainable off-track in forested terrain might be only half that attainable in grasslands. Unsurprisingly, there are no widely agreed land cover penalties, but see for example, those compiled by Herzog (2020). A hiking function derived by Wood et al. (2023) includes the local gradient of the terrain (i.e., the maximum slope at each location, not the slope in the direction of movement) as a covariate in the estimated function. It is possible that this kind of approach including other spatial covariates, such as land cover (which would be a categorical variable in most cases) in the estimation of complex hiking functions might be more widely applied in future work. In our application, because there are only two kinds of navigable land cover—moraine (or gravel) and rock—we chose instead to estimate two hiking functions, one for each land cover, and estimate movement costs conditional on the land cover at each location. This also has the advantage of allowing for differing effects of slope on speed due to land cover, where for example, gravel might allow more rapid movement on the level surfaces, but more rapidly reduce speed on slopes. Details of this approach are reported and discussed in Section 4.1.

## 2.2 Representing the landscape

As has been noted, hiking functions are usually asymmetric, with the highest attainable speed at a slight downhill slope. This asymmetry means that unless analysis is focused on assessing movement cost from a single origin or to a single destination, it is necessary to represent the landscape in a way that can accommodate asymmetry. We therefore represent the landscape as a directed graph  $G(V, E)$  of vertices and directed edges connecting them. In this representation, graph vertices  $v_i$  are locations with associated elevation and land cover. Vertices minimally have spatial coordinates  $(x_i, y_i)$ , an elevation  $z_i$ , and a land cover,  $C_i$ :

$$v_i = (x_i, y_i, z_i, C_i) \quad (1)$$

Graph edges  $e_{ij} = (v_i, v_j)$  are directed connections between vertices for which a change in elevation between the start and end vertex can be calculated, and a slope derived, based on the elevation difference, and the horizontal distance between the vertices. Thus the slope  $\theta_{ij}$  of edge  $e_{ij}$  is given by

$$\theta_{ij} = \frac{z_j - z_i}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \quad (2)$$

The obvious way to derive such a graph from spatial data is to assign each cell in a raster digital elevation model (DEM) to a graph vertex, so that graph vertices are arranged in a regular square grid or lattice, and this approach has been widely adopted. **CITATIONS NEEDED HERE** However, this still requires a decision about how to define graph edges, which can be thought of in terms of *allowed moves* between vertices. Three possibilities on a square lattice—Rook's, Queen's, and Knight's moves—are shown in Figure 2. Another option shown in the figure is to lay out a regular hexagonal grid of vertex locations and link each vertex to its six nearest neighbours.

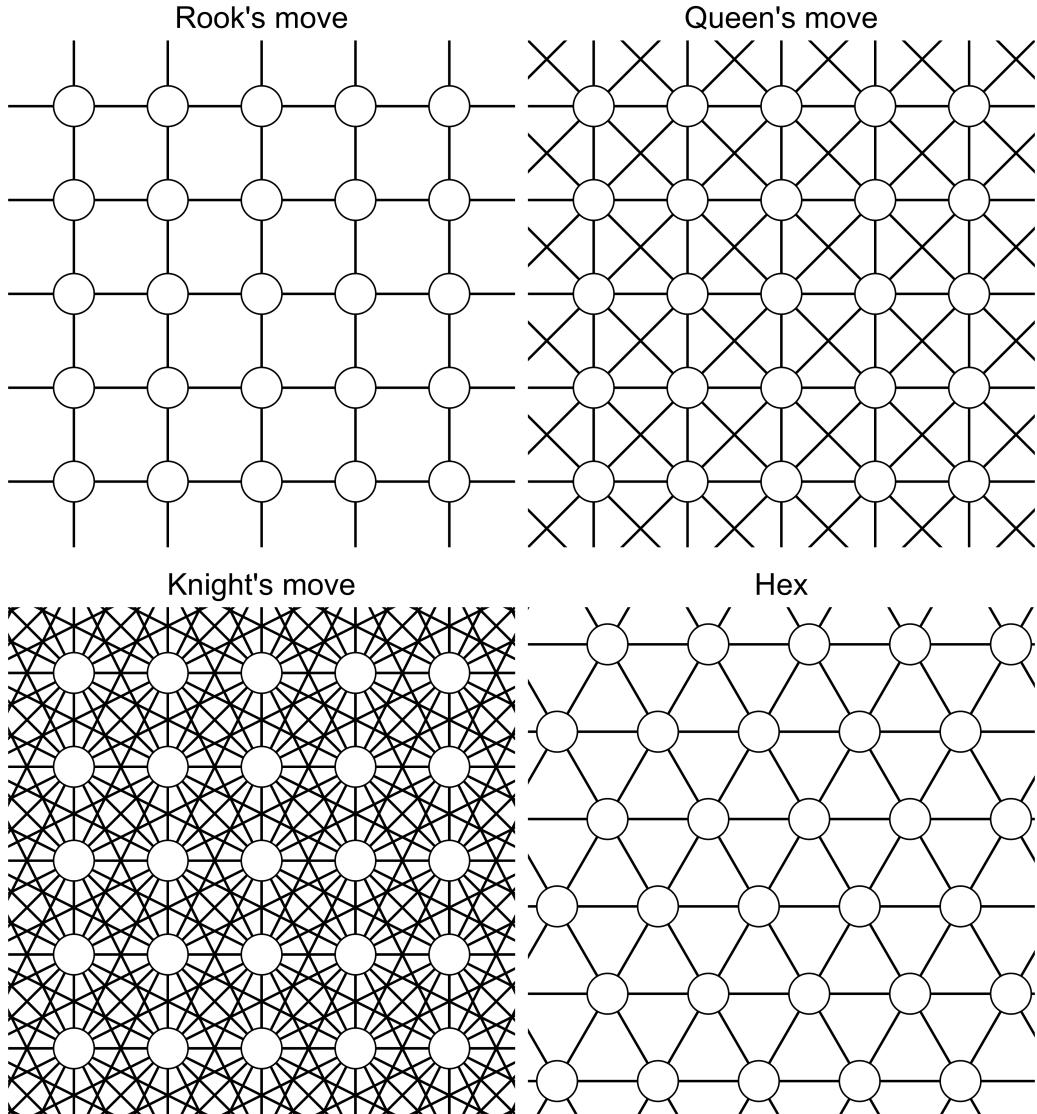


Figure 2: Possible graph lattices.

The advantage of a square lattice is that it is easy to derive vertex elevations assuming a DEM at the desired lattice spacing is available. A hexagonal lattice on the other hand usually requires that elevation values be interpolated from a DEM to the lattice locations.

Both square and hexagonal lattices lead to geometric artifacts when the allowed moves are used to determine contours of equal travel time or *isochrones* from an origin point on a flat surface with uniform movement speeds. In this situation, the Rook's move produces diamond-shaped isochrones, while the Queen's move produces octagons. In general isochrones will be polygons with as many sides as the number of neighbours of each vertex. While the 16-gons of the Knight's move lattice are likely to be close enough to circular for many purposes, the resulting graph is denser than the alternatives. Etherington (2012) suggests that by combining results from ran-

domly generated planar graphs such geometric artifacts can be removed, but this substantially increases the computational requirements. Another computationally intensive approach might draw on methods for estimating geodesics on complex triangular meshes (Martínez et al. 2005). Because these geometric effects of graph structure are masked when we introduce varying movement speeds due to the slope of each edge, we do not consider these more complex approaches necessary in our application. Nevertheless, more circular base isochrone shapes are to be preferred, all other things equal.

A further complication when a lattice includes edges of varying length as in the Queen's and Knight's move lattices, is bias in the estimation of slopes leading to lower estimated movement costs for longer edges. This is because longer edges (e.g., the knight's moves) may 'jump' across intervening segments of varying slope, smoothing them to a single slope estimate. Such smoothing will usually result in shorter edge traversal times along such edges than those that would accumulate along intervening sections of varying slope. This problem is discussed in a slightly different context by Campbell et al. 2019.

Any choice of graph structure is necessarily a compromise, and estimates of movement rates are always an approximation (Goodchild 2020). We consider the differing edge lengths introduced by all the square lattices other than the Rook's case to be problematic, and favour the hexagonal lattice structure over the simple square lattice because a base hexagonal isochrone is preferable to a diamond shape.

Having settled on a hexagonal lattice for the graph structure, at a chosen resolution we set out a regular hexagonal grid of locations across the study area, and assign to each point an elevation by interpolation from a DEM. Because the DEM is at finer resolution than the hexagonal lattice, the choice of interpolation method is not a major concern. Our approach applies bilinear interpolation based on elevation values in the DEM cell the graph vertex falls in and its four orthogonal neighbours. Based on the difference in elevation of the vertices at each end of each edge we estimate a slope using Equation 2, and also traversal times using our estimated hiking function. The important point here is that our graph is *directed* so that different traversal times  $t_{ij}$  and  $t_{ji}$  are estimated for moving between vertex  $v_i$  and  $v_j$  depending on the direction of travel.

As noted in the previous section, we estimate two hiking functions, one for moraine and one for rock land cover. When an edge connects locations of two different land cover types the estimated traversal time is the mean of the traversal times for each land cover.

### 3 Data sources

#### 3.1 Antarctic geospatial data

Geospatial data for Antarctica were obtained from sources referenced in Cox et al. (2023a), Cox et al. (2023b), and Felden et al. (2023). The study area was defined to be the Skelton and Dry Valleys basins, as defined by the NASA Making Earth System Data Records for Use in Research Environments (MEaSURES) project (Mouginot and University Of California Irvine 2017) and shown in Figure 3(a). The Skelton basin was included because while the expedition GPS data was ostensibly collected in the McMurdo Dry Valleys, it actually extends into that basin as shown in Figure 3(b).

Elevation data from the Reference Elevation Model of Antarctica (REMA) project Howat et al. (2022), and geology from GeoMAP Cox, Smith Lyttle, Elkind, Smith Siddoway, Morin, Capponi, Abu-Alam, Ballinger, Bamber, Kitchener, Lelli, Mawson, Millikin, Dal Seno, Whitburn, White, Burton-Johnson, Crispini, Elliot, Elvevold, Goodge, Halpin, Jacobs, Mikhalsky, Martin, Morgan, Smellie, Scadden, and Wilson (2023a) are shown in Figure 3(c). The five largest areas of contiguous non-ice surface geology across the study area shown in Figure 3(d) were chosen to be the specific sites for more detailed exploration using the methods set out in this paper. These range in size from around 320 to 2600 square kilometres.

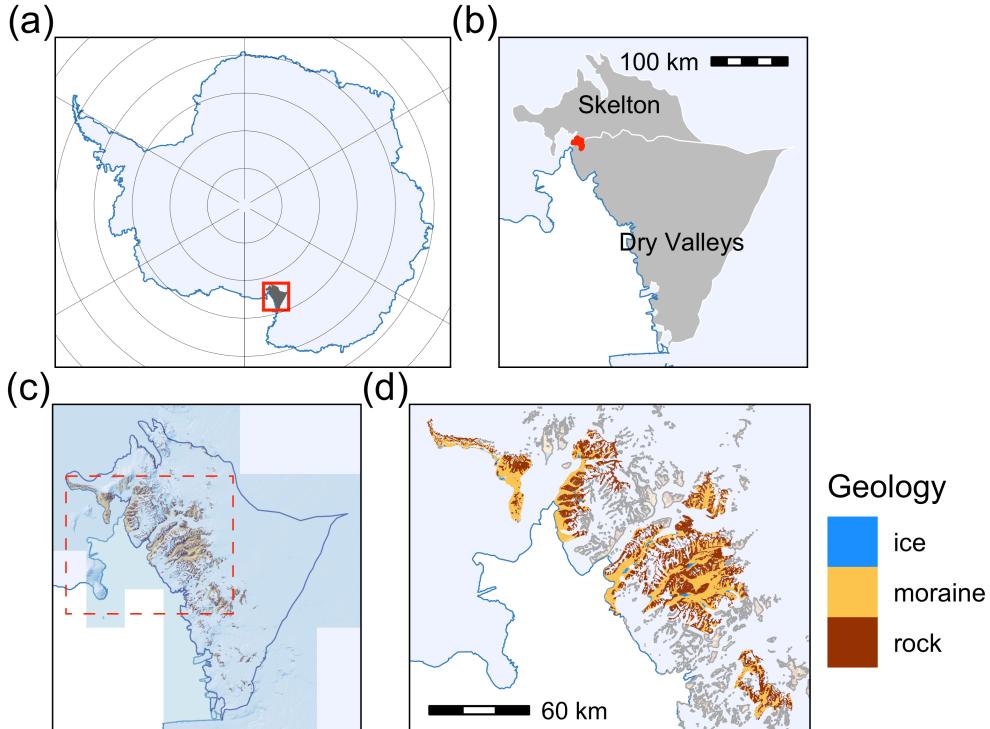


Figure 3: The study area: (a) Study area location in Antarctica; (b) Skelton and Dry Valleys basins; (c) Study area elevation (hillshade) and surface geology; and (d) Five sub-regions of contiguous surface geology.

### 3.2 GPS data from an expedition

GPS data recording how scientists moved around the McMurdo Dry Valleys was collected using QStarz Q1100P GPS Tracking Recorders between 2016 and 2018 (see Figure 4). These units were selected for their long battery life (40+ hours), chipset sensitivity, and simple design and use.



Figure 4: The Qstarz Q1100P GPS Tracking Recorders used to record the movement of research staff while in the McMurdo Dry Valleys.

Each unit was set to record a location every 30 seconds. Scientists turned on the devices when leaving base camp in the morning and turned them off when returning at the end of the day. The units were attached to the top of a backpack to ensure units had good access to the GPS signal.

Participants were sourced through existing networks of scientists supported by the New Zealand Antarctic Programme who were going to Antarctica for research purposes. All participants were asked if they wished to carry a GPS recorder. The study was entirely voluntary. Over the duration of the two primary field events, 40 people used the devices. Social Ethics approval was granted by the Manaaki Whenua Landcare Research Social Ethics committee (SE 1617/05).

Sample output of the GPS data is shown in Figure 5.

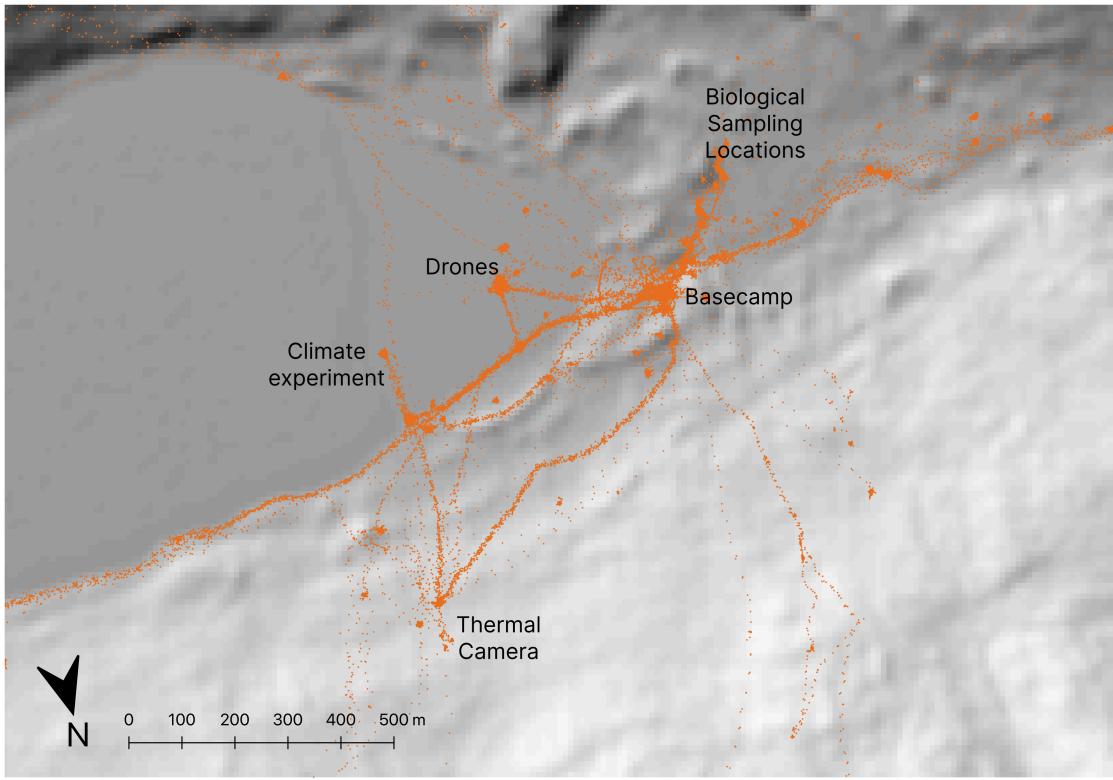


Figure 5: 30 second locations recorded by ~14 researchers during a single research trip to the Miers Valley. The image shows the basecamp, locations of research equipment and/or activities, movement among these locations and to access areas further afield.

It was necessary to process GPS data to make them better suited for use in the estimation of hiking functions. The first processing step was to confirm the plausibility of the data, particularly the device-generated speed distance between fixes, and elevations associated with fixes. The challenges of post-processing GPS data are well documented and relate to issues with GPS drift which can lead to estimated non-zero movement speeds as a result of noise in the signal. The raw GPS data included distance since last fix, speed, and elevation estimates and it was determined in all cases that the device generated results for these measurements were likely to be more reliable than post-processing the raw latitude-longitude fixes to calculate the values.

The second processing step was to remove fixes associated with faster movement on other modes of transport than walking. Wood, Mackaness, Simpson, and Armstrong (2023) cite a number of previous works that base detection of trip segments based on recorded speeds. This method was trivially applicable to our data to a limited degree as scientists arrive at the expedition base camp and occasionally also travel on helicopters on trips to more remote experimental sites.

The third, more challenging processing step was to deal with sequences of fixes associated with non-purposeful movement when scientists were in or around base camp, at rest stops, or at experimental sites. Crude filters removed fixes with low recorded distances between fixes (less than 2.5 metres), high turn angles at the fix location (greater than 150°), and fixes recorded on ice-

covered terrain, but this did not clean the data sufficiently for further analysis. An additional filtering step was to count fixes (across all scientists) in square grid cells and remove all fixes in grid cells with more than 50 fixes.

This left one persistent concern: an over-representation of consecutive fixes recorded at exactly the same elevation, resulting in many fixes with estimated slopes of exactly 0, and leading to a clearly evident dip in estimated movement speeds at 0 slope (Figure 6(a)). It is likely that these fixes are associated with GPS device drift, so it was decided to remove all fixes where estimated slope was exactly 0. Figure 6(b) shows the improvement in even a crudely estimated hiking function derived from local scatterplot (LOESS) smoothing. Note that such functions are likely overfitted and not used further in our analysis where we favour more easily parameterised functions such as those discussed in Section 2.1.

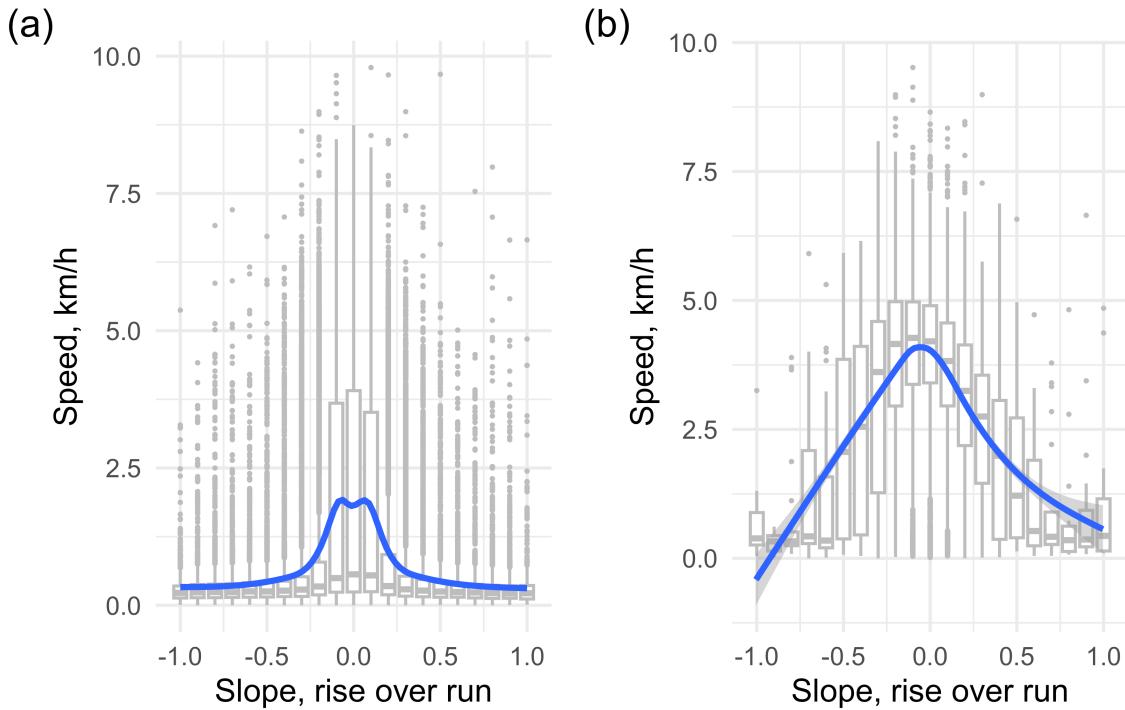


Figure 6: GPS data and crudely estimated hiking functions before and after filtering the to remove fixes associated with non-purposive movement: (a) Boxplots by slope of speed, with smoothed estimated hiking function showing a ‘dip’ due to over-representation of 0 slope fixes; and (b) After filtering the estimated hiking function no longer has a dip.

## 4 Methods and results

### 4.1 Hiking functions

We fit three alternative functional forms to the cleaned GPS data: exponential (Tobler 1993), Gaussian (Irmischer and Clarke 2018), and Lorentz (Campbell, Dennison, Butler, and Page 2019)

using the Levenburg-Marquardt algorithm (Moré 1978) as provided by the `nlsLM` function in the `minpack.lm` R package (Elzhov et al. 2022). The raw data and fitted curves are shown in Figure 7.

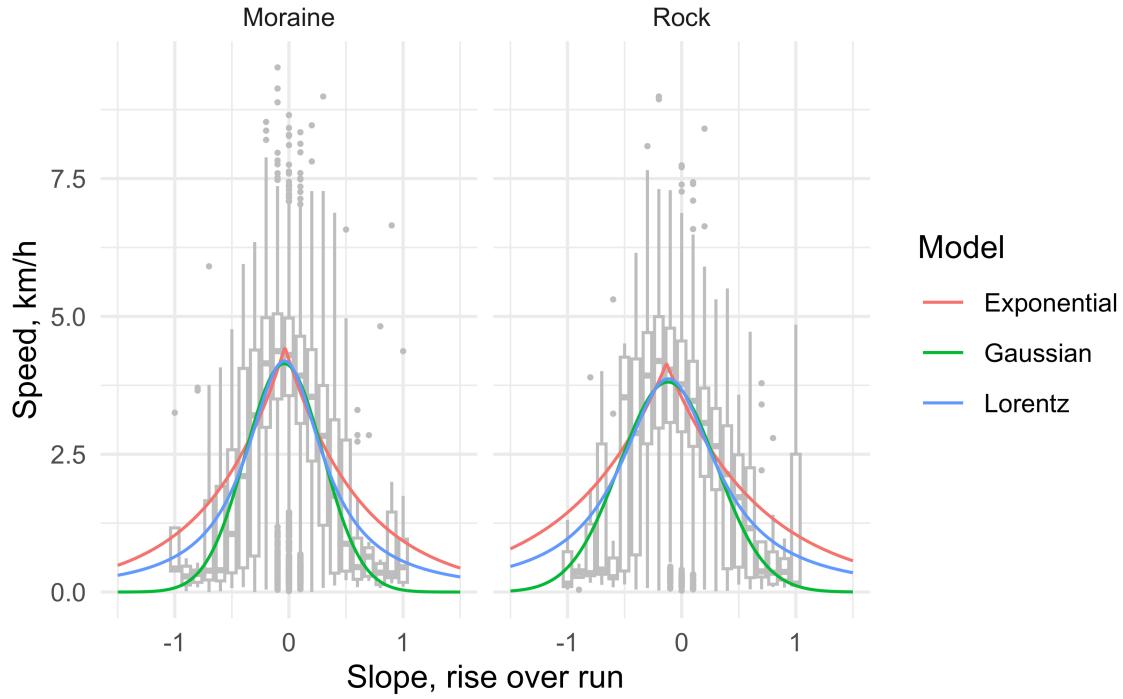


Figure 7: Three possible hiking functions applied to GPS data split by land cover.

The Lorentz function offers a marginal improvement in the model fit in comparison with the Gaussian function, while both are clearly better than the exponential form. However, the improvement offered by the Lorentz function over the Gaussian is marginal: residual standard error 1.489 vs 1.491 on Moraine, and 1.487 vs 1.488 on Rock, and inspection of the curves shows that estimated hiking speeds for the Gaussian functions are much closer to a plausible zero on very steep slopes. We therefore chose to adopt Gaussian hiking functions for the remainder of the present work.

In previous work researchers have applied a ground cover penalty cost to a base hiking function to estimate traversal times. We instead, as shown, estimate different hiking functions for the two ground cover types present. The peak speed on rock is attained on steeper downhill slopes than on moraines, perhaps indicative of the greater care required on downhill gravel slopes. Meanwhile the highest speeds on level terrain are attained on moraines.

Plotting both hiking functions along with an additional model fitted to all the data on the same axes confirms that the fitted functions are sufficiently different to retain separate models for each ground cover (see Figure 8). Plotting both functions in the same graph makes clearer the difference in maximum speed and slope at maximum speed associated with each ground cover.

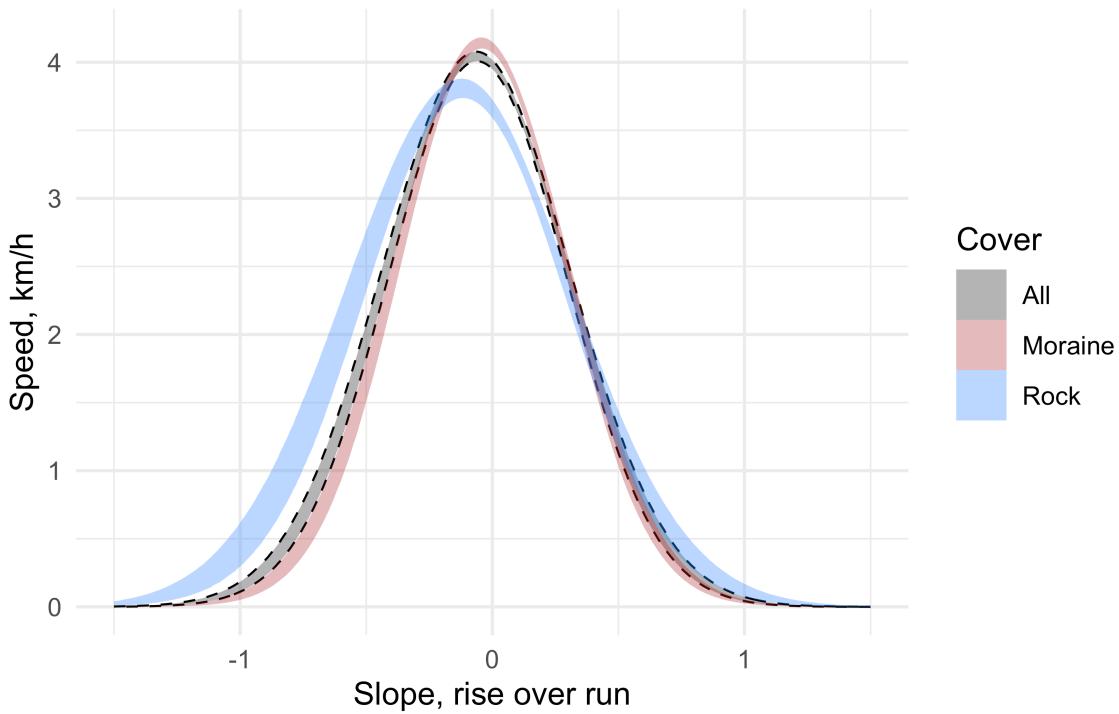


Figure 8: The hiking functions for All, Moraine and Rock ground covers compared, including 95% confidence intervals derived by Monte-Carlo simulation.

The estimated hiking functions associated with the two land covers are

$$\begin{aligned}
 v_{\text{moraine}} &= 4.17 \exp \left[ -\frac{(\theta + 0.0526)^2}{0.236} \right] \\
 v_{\text{rock}} &= 3.76 \exp \left[ -\frac{(\theta + 0.119)^2}{0.365} \right]
 \end{aligned} \tag{3}$$

where the different maximum speeds (in kilometres per hour) and slopes of maximum speed are apparent.

#### 4.2 Landscapes as graphs

We developed R code (R Core Team 2025) to build graphs (i.e. networks) with hexagonal lattice structure and estimated traversal times for graph edges derived from our hiking functions. Graphs are manipulated as `igraph` package (Csárdi and Nepusz 2006; Csárdi et al. 2024) graph objects for further analysis. An important decision in constructing graphs is choice of the spacing of the hexagonal lattice, and also of the underlying DEM from which graph vertex elevations are derived. Given the extent of the study sites (see Figure 3(d)) it was decided that a hexagonal lattice (see Figure 2) with hexagons equivalent in area to 100 metre square cells as appropriate. The hexagon centre to centre spacing of this lattice is given by

$$100\sqrt{\frac{2}{\sqrt{3}}} \approx 107.5 \text{ metres} \quad (4)$$

Given this lattice resolution we interpolated vertex elevations from a 32m resolution DEM from the REMA project (Howat, Porter, Noh, Husby, Khuvis, Danish, Tomko, Gardiner, Negrete, Yadav, Klassen, Kelleher, Cloutier, Bakker, Enos, Arnold, Bauer, and Morin 2022) by bilinear interpolation using the R `terra` package (Hijmans 2024). It would be straightforward to derive vertex elevations from a more detailed DEM if required.

Edge weights (i.e. estimated traversal costs) are assigned by calculating the slope of each directed edge, the estimated hiking speed for that slope, and thus finding how long it should take for an edge to be traversed. If the estimated traversal time of an edge in the nominal 100 metre lattice is greater than 30 minutes then it is removed from the graph, along with its ‘twin’ edge in the opposite direction. Removing edges in both directions is partly a practical matter as it simplifies the operation of many graph algorithms, but can also be justified on the basis that a slope steep enough to be a barrier to ascent is unlikely to be traversed when descending.

After removal of all such edges only the largest connected graph component is retained so that the resulting hiking network representation is fully connected with no isolated vertices unreachable from elsewhere in the network remaining. A map of the fifth largest study area’s hiking network is shown in Figure 9. This network includes 30,697 vertices and 174,798 directed edges. The largest of the five study sites (see Figure 3(d)) results in a network containing almost a quarter of a million vertices and over 1.4 million directed edges. A hiking network can be used to explore many connectivity properties of the environment. For example, for a chosen origin point, a *shortest path tree* can be derived showing the route (Figure 9(c)).

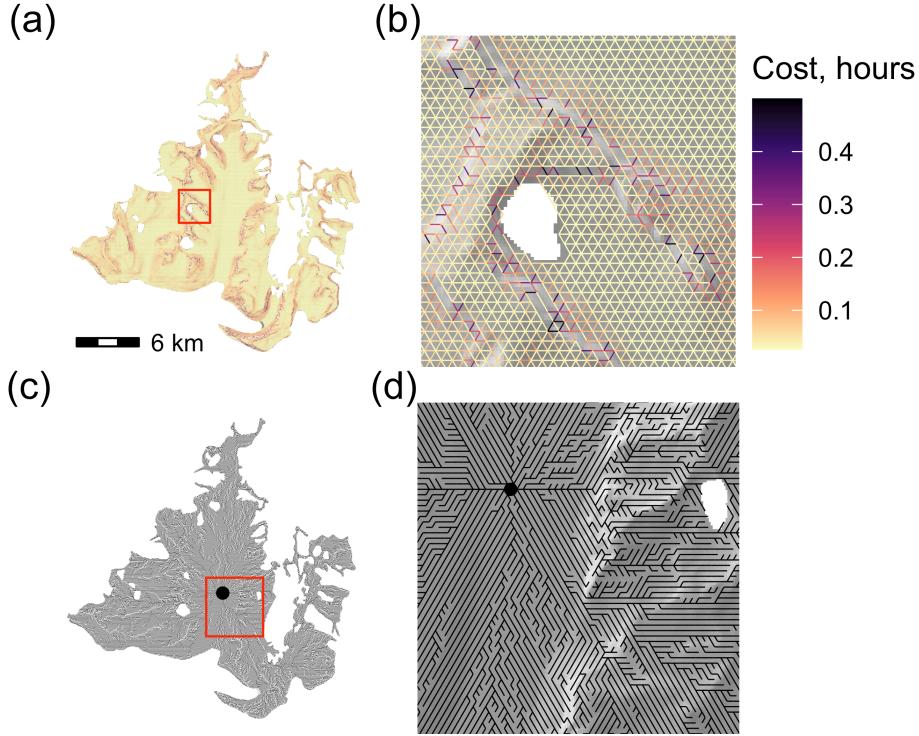


Figure 9: A hiking network derived for the fifth largest study site shown in Figure 3(d): (a) Map of a hiking network coloured by estimated traversal times of edges; (b) Red-outlined area zoomed in revealing edges removed from graph due to steep slopes; (c) Shortest path tree of a hiking network from the indicated origin; and (d) Red-outlined area zoomed in view of the shortest path tree.

### 4.3 Betweenness centrality

The connectivity properties of a hiking graph can be described using a range of measures of graph structure and used to reveal the relative likelihood of different parts of a terrain being frequently traversed. The particular structural measure we focus on is *betweenness centrality*, which is explained below.

In a graph  $G = (V, E)$  a path  $P$  is an ordered sequence of vertices,  $P = (v_1, v_2, \dots, v_n)$  such that each consecutive pair of vertices  $v_i$  and  $v_{i+1}$  is adjacent, i.e. connected by a directed edge  $e_{i,j}$ . The *length* of a path is the number of edges it contains. The *weighted length* or *cost* of a path is the sum over its edges of a *weight* or *cost* associated with each edge, which we here denote  $w_{i,j}$ , i.e., the length  $L$  of a path  $P$  is given by

$$L(P) = \sum_{i=1}^{n-1} w_{i,i+1} \quad (5)$$

The *shortest path* from  $v_i$  to  $v_j$  is the path  $P = (v_i, \dots, v_*, \dots, v_j)$  starting at  $v_i$  and ending at  $v_j$  passing through intervening vertices ( $v_*$ ) such that  $L(P)$  is minimised. In a regular lattice such as our hiking networks there are many shortest paths of equal length if the weight associated with each edge is equal, as it would be if we based it solely on the length of the edge between

each pair of vertices. When we consider edge traversal costs derived from the slope of each edge and a hiking function, then the shortest path will be unique, or one of only a small number of possibilities of equal total cost.

Shortest paths are used to develop many measures of vertex centrality in graphs (Freeman 1978). One such measure of vertex centrality is *betweenness centrality*. The betweenness centrality of a vertex is the total number of times that vertex is on shortest paths between every other pair of vertices in the network. If we denote the number of shortest paths or *geodesics* between two vertices  $v_j$  and  $v_k$  by  $g_{jk}$ , then each appearance of a vertex  $v_i$  on the shortest path between those vertices only contributes  $1/g_{jk}$  to the betweenness centrality of  $v_i$ . Formally, if we denote the number of times vertex  $v_i$  appears on shortest paths between  $v_j$  and  $v_k$  by  $g_{jk}(v_i)$ , then its betweenness  $b_i$  is given by

$$b_i = \sum_{k=1}^n \sum_{j=1}^n \frac{g_{jk}(v_i)}{g_{jk}} \forall i \neq j \neq k \quad (6)$$

Betweenness centrality is particularly relevant to applications where we are interested in how important vertices are to movement across the graph. An edge betweenness centrality measure can also be calculated on similar principles, but is not considered further here, in part because visualization is challenging given the potential for different scores for the two directed edges between each pair of vertices. Vertex betweenness on the other hand yields a single value for each vertex.

As might be expected betweenness centralities are computationally demanding to calculate. The time complexity of early algorithms was  $\mathcal{O}(n^3)$ , where  $n$  is the number of vertices in the graph (Brandes 2001). An implementation of Brandes's algorithm 2001 with time complexity  $\mathcal{O}(nm + n^2 \log n)$ , where  $m$  is the number of edges in the graph, is provided in the `igraph` package (Csárdi, Nepusz, Traag, Horvát, Zanini, Noom, and Müller 2024). Even with this improvement in performance, in our application such computational complexity is a strong motivation for working at a nominal 100 metre resolution. Halving the resolution to 50 metres would increase the number of graph vertices four-fold, and lead to a substantial increases in the times taken to calculate betweenness centralities.

Results of betweenness centrality for our example hiking network are shown in Figure 10. ‘All-paths’ betweenness centralities can be standardised relative to a theoretical maximum of  $(n - 1)(n - 2)$ , no straightforward standardisation is possible for the radius-limited case. Since we are primarily interested in betweenness as a measure of the *relative* vulnerability of different locations to human impacts, we linearly rescaled betweenness scores in all cases with respect to the range of scores across the graph being analysed. The rescaled betweenness,  $b'_i$  for vertex  $v_i$  is given by

$$b'_i = \frac{b_i - b_{\min}}{b_{\max} - b_{\min}} \quad (7)$$

This approach is also applicable to radius-limited betweenness scores (see Section 4.4).

When all-paths betweenness is calculated, bottlenecks between large sub-areas are strongly highlighted as is the case for the ‘inner bend’ of the relatively narrow neck of navigable terrain that connects the east and west sub-areas of this site. The other feature of this map is the identification of a distinctive structure of ‘arterial’ routes.

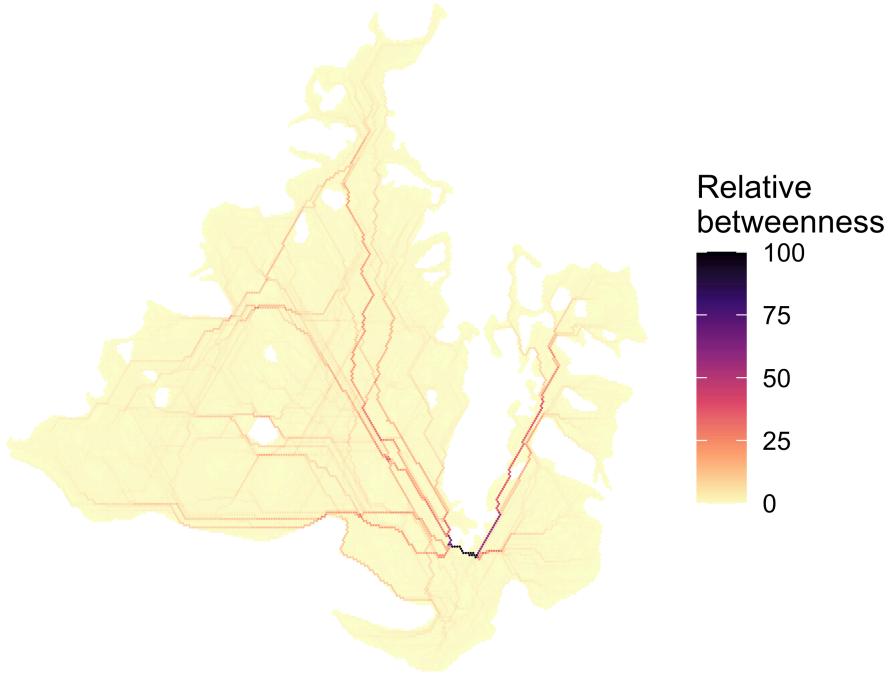


Figure 10: Relative vertex betweenness of vertices in the graph from Figure 9.

#### 4.4 Betweenness centrality limited by radius

The *igraph* implementation of betweenness centrality provides an option to *radius limit* betweenness calculations, meaning that only paths shorter than a specified radius (expressed in cost units, here of time) are considered in counting the appearance of vertices on shortest paths. This approach is particularly useful for large graphs where the time complexity of calculating radius limited betweenness scales more favourably than cited above, because the number of shortest paths that must be identified is greatly reduced. We informally confirmed the intuition that for a given radius limit the time take to compute betweenness scales approximately linearly with the number of vertices in the graph, since the number of shortest paths local to each vertex, and on which it might be found is similar.

Results for a series of radius limits are shown in Figure 11. The results here are interesting. At very short radii (the 30 minutes case) many locally important paths are identified as potentially relatively heavily trafficked. As the radius increases the pattern emphasizes potential paths that are important across wider areas, eventually tending toward the arterial structure of the all-paths case. We discuss these results more fully in Section 5.

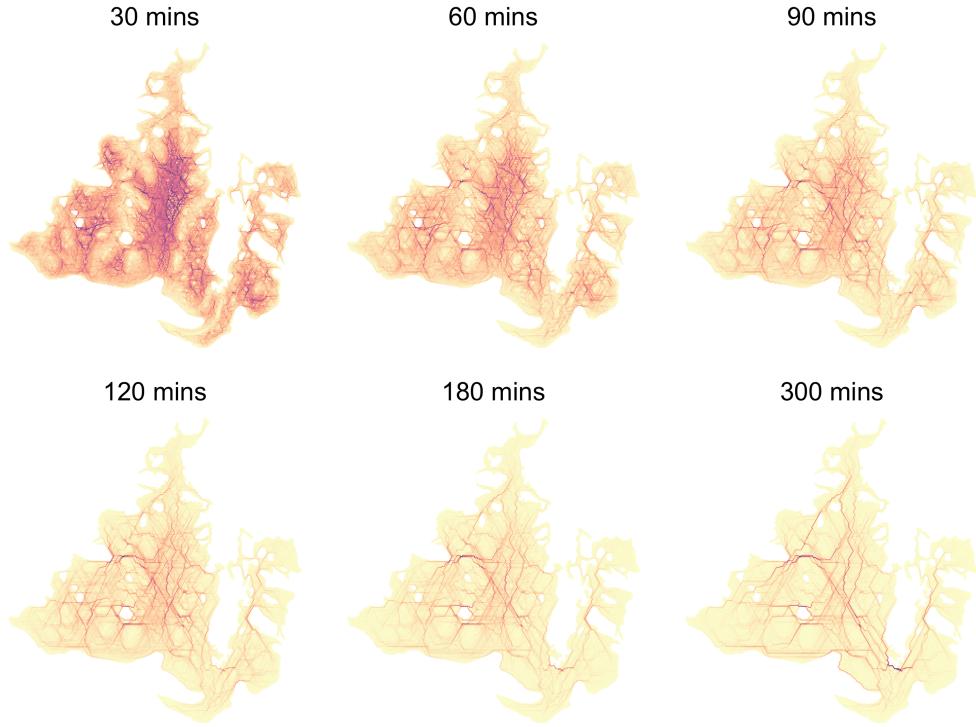


Figure 11: Vertex betweenness maps based on different radius limits. As in Figure 10 the visualization colouring is based on linear scaling of raw betweenness scores.

#### 4.5 Impact minimizing networks

In this section we explore developing the betweenness centrality for a limited set of locations, which might represent a base camp, rest sites, and experimental sites for a season's expedition.

This is relatively straightforward and like radius limited betweenness involves limiting the shortest paths considered to only those among a chosen subset of vertices in the graph. One slight simplification we make for performance reasons is to weight every appearance of a vertex on a geodesic as equal, even if multiple equal length shortest paths exist between two vertices. In practice, this is unlikely to have much effect on the resulting estimates of relative betweenness, since with real-valued edge costs two paths of exactly equal cost are unlikely. An example result is shown in Figure 12. One refinement shown in the figure is that sites have been weighted by relative importance so that a site such as an expedition base camp can be expected to originate more trips than other sites.

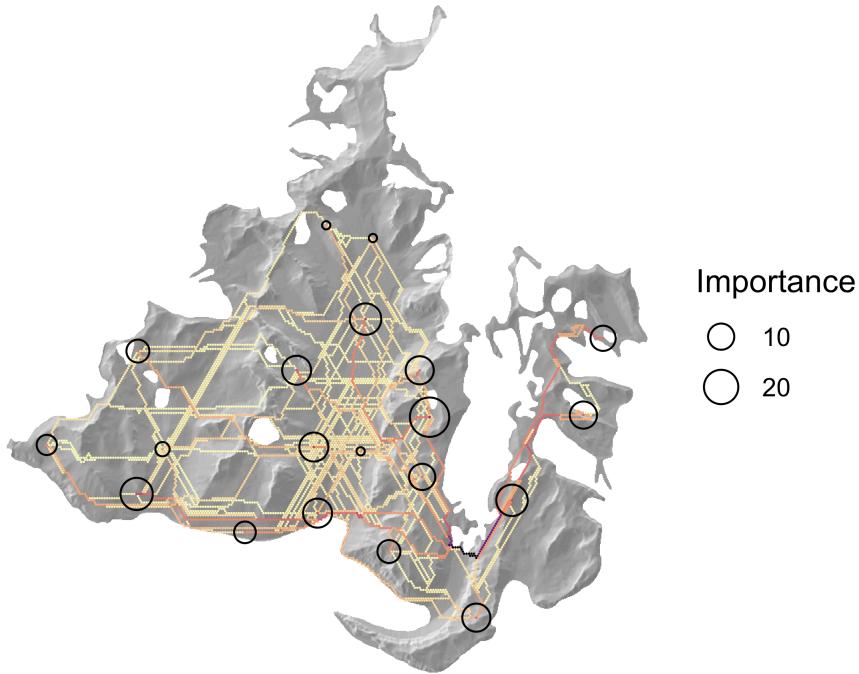


Figure 12: The network of shortest paths among a weighted set of expedition sites.

A further analytical possibility is to use the hiking network to develop a set of routes among the sites that might minimise total impact, by confining movement to a limited set of paths. This approach is consistent with work in ecology and park and wildlife management which suggests that path networks can reduce the impact of human movement in many environments (Cole 1995a; Cole 1995b; Kidd et al. 2015; Marion et al. 2016; Piscová et al. 2023; Tomczyk and Ewertowski 2023).

The approach taken is first to first determine the lengths of all the shortest paths among all the sites. Then a *minimum spanning tree* for the sites is calculated, which is a graph with minimum total edge length that connects all the sites (Prim 1957). One refinement here, which we do not account for is that algorithms for finding the minimum spanning tree of a directed graph (referred to as a *minimum-cost arborescence* or *optimal branching*) are more complex (Korte and Vygen 2018), and not at present supported in *igraph*. While there is some asymmetry in paths between locations depending on the direction of travel, it is not extreme, and so we approximate a solution by finding the minimum spanning tree links (i.e. direct connections) among sites treating our network as undirected, and then determine paths over the ground to make the identified connections between sites based on the directed hiking network. The resulting path network for the same set of sites as in Figure 12 is shown in Figure 13.

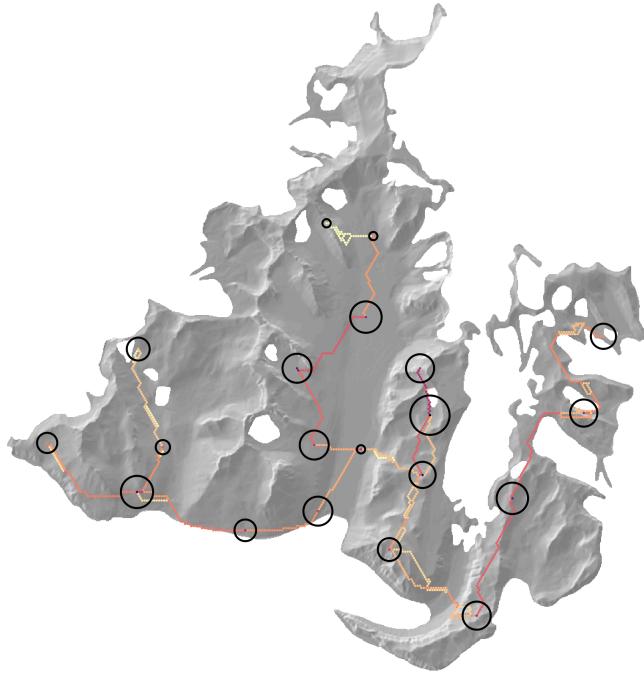


Figure 13: A potential impact minimising route network among sites.

## 5 Discussion

- Limitations of the method:
  - Data concerns, esp. given improvements in GPS data collection since 2016/18
  - Would higher resolution network be useful? (we think probably not)
  - Possible limitations of a regularly spaced lattice: extend to a lattice based on e.g., TIN?
  - Reliance on shortest paths as a guide to behaviour
  - Some minor issues regarding the difference between directed and undirected graphs (to check using networkx)
- Relevance of radius limits on betweenness: the implicit model of movement in this approach (short horizons, no existing paths, ‘keep going for half an hour in this general direction...’)
- Question of embedding paths, i.e., locking in more damage to some parts of the landscape vs others (this leads to the question in conclusion re method for less damaging plans)

## 6 Conclusions

To cover:

- Novelty of fitting land cover specific hiking functions
- A method for exploring potential impacts in a novel environment
- A method for proposing less damaging plans for expeditions
- Other?

## Bibliography

- ATCM. 1970. Antarctic Treaty Report of Sixth Consultative Meeting Tokyo. [https://documents.ats.aq/ATCM6/fr/ATCM6\\_fr001\\_e.pdf](https://documents.ats.aq/ATCM6/fr/ATCM6_fr001_e.pdf)
- ATCM. 1985. Antarctic Treaty Report of Thirteenth Consultative Meeting Brussels. [https://documents.ats.aq/ATCM13/fr/ATCM13\\_fr001\\_e.pdf](https://documents.ats.aq/ATCM13/fr/ATCM13_fr001_e.pdf)
- Benninghoff, W. S., and W. N. Bonner. 1985. *Man's Impact on the Antarctic Environment: A Procedure for Evaluating Impacts from Scientific and Logistic Activities*. Scientific Committee on Antarctic Research
- Brandes, U. 2001. A Faster Algorithm for Betweenness Centrality. *The Journal of Mathematical Sociology* 25(2): 163–177. doi: 10.1080/0022250X.2001.9990249
- Brooks, S. T., J. Jabour, and D. M. Bergstrom. 2018. What Is 'Footprint' in Antarctica: Proposing a Set of Definitions. *Antarctic Science* 30(4): 227–235. doi: 10.1017/S0954102018000172
- Brooks, S. T., J. Jabour, J. van den Hoff, and D. M. Bergstrom. 2019. Our Footprint on Antarctica Competes with Nature for Rare Ice-Free Land. *Nature Sustainability* 2(3): 185–190. doi: 10.1038/s41893-019-0237-y
- Campbell, M. J., P. E. Dennison, B. W. Butler, and W. G. Page. 2019. Using Crowdsourced Fitness Tracker Data to Model the Relationship between Slope and Travel Rates. *Applied Geography* 106: 93–107. doi: 10.1016/j.apgeog.2019.03.008
- Cole, D. N. 1995a. Experimental Trampling of Vegetation. I. Relationship Between Trampling Intensity and Vegetation Response. *The Journal of Applied Ecology* 32(1): 203–204. doi: 10.2307/2404429
- Cole, D. N. 1995b. Experimental Trampling of Vegetation. II. Predictors of Resistance and Resilience. *The Journal of Applied Ecology* 32(1): 215–216. doi: 10.2307/2404430
- Cox, S. C., B. Smith Lyttle, S. Elkind, C. S. Smith Siddoway, P. Morin, G. Capponi, T. Abu-Alam, M. Ballinger, L. Bamber, B. Kitchener, L. Lelli, J. F. Mawson, A. Millikin, N. Dal Seno, L. Whitburn, T. White, A. Burton-Johnson, L. Crispini, D. Elliot, S. Elvevold, J. W. Goodge, J. A. Halpin, J. Jacobs, A. P. Martin, E. Mikhalsky, F. Morgan, P. Scadden, J. Smellie, and G. S. Wilson. 2023b. A Continent-Wide Detailed Geological Map Dataset of Antarctica. *Scientific Data* 10(1): 250–251. doi: 10.1038/s41597-023-02152-9
- Cox, S. C., B. Smith Lyttle, S. Elkind, C. S. Smith Siddoway, P. Morin, G. Capponi, T. Abu-Alam, M. Ballinger, L. Bamber, B. Kitchener, L. Lelli, J. F. Mawson, A. Millikin, N. Dal Seno, L. Whitburn, T. White, A. Burton-Johnson, L. Crispini, D. Elliot, S. Elvevold, J. W. Goodge, J. A. Halpin, J. Jacobs, E. Mikhalsky, A. P. Martin, F. Morgan, J. Smellie, P. Scadden, and G. S. Wilson. 2023a. The GeoMAP (v.2022-08) Continent-Wide Detailed Geological Dataset of Antarctica. PANGAEA. doi: 10.1594/PANGAEA.951482
- Csárdi, G., and T. Nepusz. 2006. The igraph Software Package for Complex Network Research. *Interjournal Complex Systems* 1695: 1–9

- Csárdi, G., T. Nepusz, V. Traag, S. Horvát, F. Zanini, D. Noom, and K. Müller. 2024. *igraph*: Network Analysis and Visualization in R. <https://CRAN.R-project.org/package=igraph>
- Deutscher und Österreichischer Alpenverein. 2011. *Wegehandbuch Des Alpenvereins*. Deutscher Alpenverein
- Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker. 2022. Minpack.Lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, plus Support for Bounds. Comprehensive R Archive Network. doi: 10.32614/CRAN.package.minpack.lm
- Etherington, T. R. 2012. Least-Cost Modelling on Irregular Landscape Graphs. *Landscape Ecology* 27: 957–968. doi: 10.1007/s10980-012-9747-y
- Felden, J., L. Möller, U. Schindler, R. Huber, S. Schumacher, R. Koppe, M. Diepenbroek, and F. O. Glöckner. 2023. PANGAEA - Data Publisher for Earth & Environmental Science. *Scientific Data* 10(1): 347–348. doi: 10.1038/s41597-023-02269-x
- Freeman, L. C. 1978. Centrality in Social Networks Conceptual Clarification. *Social Networks* 1(3): 215–239. doi: 10.1016/0378-8733(78)90021-7
- Goodchild, M. F. 2020. Beyond Tobler's Hiking Function. *Geographical Analysis* 52(4): 558–569. doi: 10.1111/gean.12253
- Herzog, I. 2010. Theory and Practice of Cost Functions. In *CAA'2010 Fusion of Cultures: Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, eds. F. Contreras, M. Farjas, and F. J. Melero, 375–382. Granada, Spain: BAR Publishing
- Herzog, I. 2020. Spatial Analysis Based on Cost Functions. Edited by Mark Gillings, Piraye Hacıgüzeller, and Gary Lock. *Archaeological Spatial Analysis: A Methodological Guide*. Routledge
- Hijmans, R. J. 2024. terra: Spatial Data Analysis. <https://github.com/rspatial/terra>
- Howat, I., C. Porter, M.-J. Noh, E. Husby, S. Khuvis, E. Danish, K. Tomko, J. Gardiner, A. Negrete, B. Yadav, J. Klassen, C. Kelleher, M. Cloutier, J. Bakker, J. Enos, G. Arnold, G. Bauer, and P. Morin. 2022. The Reference Elevation Model of Antarctica - Mosaics, Version 2. Harvard Dataverse. doi: 10.7910/DVN/EBW8UC
- Imhof, E. 1950. *Gelände Und Karte*. Zurich: Eugen Rentsch Verlag
- Irmischer, I. J., and K. C. Clarke. 2018. Measuring and Modeling the Speed of Human Navigation. *Cartography and Geographic Information Science* 45(2): 177–186. doi: 10.1080/15230406.2017.1292150
- Kidd, A. M., C. Monz, A. D'Antonio, R. E. Manning, N. Reigner, K. A. Goonan, and C. Jacobi. 2015. The Effect of Minimum Impact Education on Visitor Spatial Behavior in Parks and Protected Areas: An Experimental Investigation Using GPS-based Tracking. *Journal of Environmental Management* 162: 53–62. doi: 10.1016/j.jenvman.2015.07.007

- Korte, B., and J. Vygen. 2018. Spanning Trees and Arborescences. Edited by Bernhard Korte and Jens Vygen. *Combinatorial Optimization: Theory and Algorithms*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-56039-6\_6
- Langmuir, E. 1984. *Mountaineering and Leadership: A Handbook for Mountaineers and Hillwalking Leaders in the British Isles*. Edinburgh: Scottish Sports Council
- Leihy, R. I., B. W. T. Coetzee, F. Morgan, B. Raymond, J. D. Shaw, A. Terauds, K. Bastmeijer, and S. L. Chown. 2020. Antarctica's Wilderness Fails to Capture Continent's Biodiversity. *Nature* 583(7817): 567–571. doi: 10.1038/s41586-020-2506-3
- Llobera, M., and T. Sluckin. 2007. Zigzagging: Theoretical Insights on Climbing Strategies. *Journal of Theoretical Biology* 249(2): 206–217. doi: 10.1016/j.jtbi.2007.07.020
- Marion, J. L., Y.-F. Leung, H. Eagleston, and K. Burroughs. 2016. A Review and Synthesis of Recreation Ecology Research Findings on Visitor Impacts to Wilderness and Protected Natural Areas. *Journal of Forestry* 114(3): 352–362. doi: 10.5849/jof.15-498
- Martínez, D., L. Velho, and P. C. Carvalho. 2005. Computing Geodesics on Triangular Meshes. *Computers & Graphics* 29(5): 667–675. doi: 10.1016/j.cag.2005.08.003
- Minetti, A. E., C. Moia, G. S. Roi, D. Susta, and G. Ferretti. 2002. Energy Cost of Walking and Running at Extreme Uphill and Downhill Slopes. *Journal of Applied Physiology* 93(3): 1039–1046. doi: 10.1152/japplphysiol.01177.2001
- Moré, J. J. 1978. The Levenberg-Marquardt Algorithm: Implementation and Theory. Edited by G. A. Watson. *Numerical Analysis*. Berlin, Heidelberg: Springer. doi: 10.1007/BFb0067700
- Mouginot, J., and University Of California Irvine. 2017. MEaSURES Antarctic Boundaries for IPY 2007-2009 from Satellite Radar, Version 2. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: 10.5067/AXE4121732AD
- Márquez-Pérez, J., I. Vallejo-Villalta, and J. I. Álvarez-Francoso. 2017. Estimated Travel Time for Walking Trails in Natural Areas. *Geografisk Tidsskrift-Danish Journal of Geography* 117(1): 53–62. doi: 10.1080/00167223.2017.1316212
- Naismith, W. W. 1892. Excursions: Cruach Ardran, Stobinian, and Ben More. *Scottish Mountaineering Club Journal* 2(3): 133–134
- Perna, A., and T. Latty. 2014. Animal Transportation Networks. *Journal of The Royal Society Interface* 11(100): 20140334–20140335. doi: 10.1098/rsif.2014.0334
- Piscová, V., M. Ševcík, A. Sedláč, J. Hrisko, F. Petrović, and T. Slobodová. 2023. Resistance of Lichens and Mosses of Regenerated Alpine Communities to Repeated Experimental Trampling in the Belianske Tatras, Northern Slovakia. *Diversity* 15(2): 128–129. doi: 10.3390/d15020128
- Prim, R. C. 1957. Shortest Connection Networks and Some Generalizations. *Bell System Technical Journal* 36(6): 1389–1401. doi: 10.1002/j.1538-7305.1957.tb01515.x
- Priscu, J. C., and A. Howkins. 2016. Environmental Assessment of the McMurdo Dry Valleys: Witness to the Past and Guide to the Future. *Special Publication, LPRES-PRG* 2

- Prisner, E., and P. Sui. 2023. Hiking-Time Formulas: A Review. *Cartography and Geographic Information Science* 50(4): 421–432. doi: 10.1080/15230406.2023.2197625
- R Core Team. 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing
- Rees, W. 2004. Least-Cost Paths in Mountainous Terrain. *Computers & Geosciences* 30(3): 203–209. doi: 10.1016/j.cageo.2003.11.001
- Tejedo, P., D. Cajiao, T. O'Neill, J. A. Lamphere, and D. Liggett. 2024. Tourism in Antarctica: Facts, Concerns, and Challenges. doi: 10.48361/4GWG-DB92
- Tobler, W. R. 1993. *Three Presentations on Geographical Analysis and Modeling: Non-Isotropic Geographic Modeling; Speculations on the Geometry of Geography; and Global Spatial Analysis*
- Tomczyk, A. M., and M. W. Ewertowski. 2023. Landscape Degradation and Development as a Result of Touristic Activity in the Fragile, High-mountain Environment of Vinicunca (Rainbow Mountain), Andes, Peru. *Land Degradation & Development* 34(13): 3953–3972. doi: 10.1002/lrd.4729
- Verhagen, P., L. Nuninger, and M. R. Groenhuijzen. 2019. Modelling of Pathways and Movement Networks in Archaeology: An Overview of Current Approaches. Edited by Philip Verhagen, Jamie Joyce, and Mark R. Groenhuijzen. *Finding the Limits of the Limes*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-04576-0\_11
- Winkler, K., H.-P. Brehm, and J. Haltmeier. 2010. *Bergsport Sommer: Technik, Taktik, Sicherheit*. SAC-Verlag
- Wood, A., W. Mackaness, T. I. Simpson, and J. D. Armstrong. 2023. Improved Prediction of Hiking Speeds Using a Data Driven Approach. *PLOS ONE* 18(12): e295848. doi: 10.1371/journal.pone.0295848