Explaining known past routes and the use of multiple cost functions

Joseph Lewis

Department of Archaeology, University of Cambridge, Cambridge, UK

Jl2094@cam.ac.uk

0000-0002-0477-1756

Keywords: least-cost path analysis; cost functions; explanation; movement; postdiction; Roman road

“Our truth is the intersection of independent lies” – Levens (1966, p. 423)

Abstract

Explaining material traces of movement as proxies for past movement is fundamental for explaining the processes behind why people in the past traversed the landscape in the way that they did. For this, least-cost path analysis and the use of cost functions have become commonplace. Despite its prevalence, current approaches misrepresent what cost functions are, their relationship to the hypotheses that they aim to represent, and their role in explanation. As a result, least-cost paths calculated from these cost functions are liable to spurious results and limit their credibility when used for explaining known past routes. Using the idea of multiple model idealisation, I show that least-cost path results from time- and energy-based cost functions are robust when used concurrently. Robust probabilistic least-cost paths are applied for the explanation of a Roman road in south-western Sardinia. Results suggest that the road did not follow the route that minimised time taken, or energy expended, with its route instead influenced by pre-existing routes and settlements. Through the application of robust least-cost path analyses, it is possible for our hypotheses to have explanatory power for uncovering why people in the past moved where they did.

Introduction

Explaining the rationale behind past movement is often limited to explaining material traces of past movement. From roads to footpaths to trails, the material traces of past movement are thought to have preserved the decision-making processes of people when traversing the landscape in the past (Snead et al. 2009). Without direct records of these processes, it is by explaining past routes that we can aim to uncover why people in the past moved where they did. For proposed explanations to be reliable, it is however necessary that results are robust and sufficiently represent the hypotheses that is being assessed.

A common method for explaining known past routes is the use of least-cost path (LCP) analysis (e.g. Fonte et al. 2017; Güimil-Fariña and Parcero-Oubiña 2015; Herzog 2010, 2022; Lewis 2021). In this approach, a cost function that express the cost of traversing a specific slope gradient, e.g. as measured in time-taken or energy expended, is used to calculate an LCP (Herzog 2010). When the path of the calculated LCP and the known past route is similar, it is suggested that the resultant LCP, and the hypothesis that the cost function
aims to represent, reflects the processes that resulted in the known past route. For example, if an LCP calculated using a time-based cost function, under the hypothesis that humans minimise time when traversing the landscape, shares similarity with a known past route, it can be suggested that the path of the known past route was chosen to minimise time.

Whilst the comparison of multiple LCPs derived using different cost functions can be used to assess which cost function, and by extension hypothesis, best explains the route of the known past route, this approach misrepresents what cost functions are, their relationship to the hypotheses that they aim to represent, and their role in explanation. Using the idea of multiple model idealisation (MMI) as proposed by Levins (1966) and further developed by Weisberg (2007), this paper argues that focus should shift from comparing single cost functions and their ability to explain known past routes to the comparison of the hypotheses that multiple cost functions share.

First, the theoretical and methodological basis of using cost functions for explaining routes is examined in light of MMI, followed by two cases studies: (1) a tactical simulation, where LCP results from multiple cost functions sharing the same hypothesis are shown to be robust and therefore sufficiently represent the hypotheses that they aim to represent, and (2) the comparison of hypotheses as mediated through multiple cost functions for the explanation of a Roman road in south-western Sardinia.

Cost Function Idealisation and Multiple Models

All cost functions are idealisations, that is they intentionally misrepresent – through different assumptions, approximations, and simplifications – the relationship between an associated cost and slope gradient (e.g. Campbell et al. 2017, 2017, 2019; Davey et al. 1994; Herzog 2014; Irmischer and Clarke 2018; Langmuir 1984; Llobera and Sluckin 2007; Márquez-Pérez et al. 2017; Minetti et al. 2002; Naismith 1892; Rees 2004; Tobler 1993). These idealisations include the functional form of the cost function, e.g. the double exponential used by Tobler’s Hiking function (Tobler 1993) or the sixth degree polynomial by Herzog (2014); the set of parameters chosen within the cost function, e.g. the offset parameter to represent the anisotropic property of slope (e.g. Campbell et al. 2019; Tobler 1993); and the specific parameter values used within the cost functions themselves.

Whilst the choice of functional form, set of parameters, and parameter values when creating idealised cost functions is often made to best fit the data at hand and maximise predictive accuracy, i.e. the MAXOUT idealisation (Weisberg 2007), it does not guarantee that the resultant cost function is useful for explanation. Idealised cost functions are derived from data with their own biases, for example the number of study participants, e.g. the unknown number used by Tobler’s Hiking Function (Herzog 2010; Tobler 1993), or the fitness level of the participants, e.g. the United States Military Academy cadets used by Irmischer and Clarke (2018). With each cost function representing a specific idealisation of the relationship between cost and slope gradient, varying trade-offs in their accuracy, precision, generality, and simplicity of representation are made (sensu Levins 1966; Weisberg 2007). As a result, no single cost function can simultaneously maximise all...
properties. To be able to explain known past routes using cost functions and have their results be epistemologically reliable for explanation, it is therefore necessary that multiple cost functions are used, each making varying trade-offs but sharing the same hypothesis. Only then will resultant LCPs be robust, with outcomes reflecting not the specific simplifying assumptions of the cost function made during the idealisation process, but the shared essential of the cost function, that is the hypotheses that the multiple cost functions share (sensu Levins 1966; Weisberg 2006).

**Case Study 1: Robustness of Cost Functions, a Tactical Simulation**

**The Robustness of Cost Functions**

When using cost functions to explain known past routes, it is necessary that the proposed modelling assumptions are robust: that the results depend on the shared essentials of the cost functions. If multiple cost functions representing the same hypothesis, e.g. the relationship between time and slope gradient, each similar but distinct in their idealisation, are able to generate similar results, then the shared hypothesis can be deemed robust (Weisberg 2006). With robustness, increased support can be placed on the hypothesis as being sufficiently represented in the resultant LCPs.

**Materials and Methods**

To test the robustness of cost functions, a simulated Digital Elevation Model (DEM) of 1km by 1km, with a spatial resolution of 1m, was used. Generated using the spectral synthesis method (Saupe 1988) as implemented in GRASS (GRASS Development Team 2022), the simulated DEM is a fractal surface with a fractal dimension of 2.40. The fractal dimension denotes the complexity of the surface, with greater values representing more topographic variability (Tate and Wood 2001). Given that the complexity of real landscapes ranges from a fractal dimension of 2.20 to 2.60 (Hofierka et al. 2009), the value of 2.40 is deemed sufficiently representative.

The robustness of the hypotheses represented by time- and energy-based cost functions were assessed using the following approach:

1. Two random points are selected within the extent of the simulated DEM. These represent the origin and destination used when calculating the least-cost paths
2. 256 (16x16) least-cost paths using the `leastcostpath` R package (Lewis 2022) were calculated from the origin and destination using sixteen different cost functions (Figure 1):
   a. time-based Tobler’s on-path and off-path Hiking functions (Tobler 1993), the modified Tobler’s Hiking function (Márquez-Pérez et al. 2017), the Irmischer-Clarke male and female on-path and off-path cost functions (Irmischer and Clarke 2018), and the cost functions proposed by Rees (2004), Davey et al. (1994), Garmy et al. (2005), Kondo and Seino (2010), Naismith (1892), and Campbell et al. (2019) (50th percentile).
b. the energy-based cost functions proposed by Herzog (2014), Llobera and Sluckin (2007), and Minetti (2002);

3. Each calculated route assigned its ‘hypothesis type’, that is time or energy

4. Spatial separation between the 256 least-cost paths calculated using the Path Deviation Index (Jan et al. 2000)

**Figure 1** Sixteen cost functions estimating time taken and energy expenditure as a function of mathematical slope. Uphill/downhill slopes are denoted by positive/negative slope values, respectively. The cost functions are scaled and limited to slopes of ±40% for ease of comparison. Campbell 2019 is based on the 50th percentile

Having calculated 256 least-cost paths and their accompanying PDI values for 1,000 simulations, the 256,000 routes were filtered to those that resulted in perfect agreement, i.e. zero spatial separation, resulting in a final number of 29,326 least-cost paths. The hypothesis that multiple cost functions aim to represent is deemed
robust if the majority of the resultant least-cost paths result in the same outcome, irrespective of the exact cost function used – provided that the cost functions share the same hypothesis, i.e. time or energy.

Results and Discussion

Both hypotheses represented by multiple time- and energy-based cost functions are shown to be robust (Figure 2). That is, the majority of time- and energy-based cost functions produce the same results, irrespective of which idealised cost function sharing the same hypothesis is used. As a result, it is shown that multiple least-cost paths calculated using cost functions but sharing the same hypothesis can be deemed to sufficiently represent the corresponding hypothesis. For example, if least-cost paths calculated using multiple time-based cost functions are used concurrently, the outcome can be deemed to sufficiently reflect the outcome of the hypothesis ‘humans minimise time-taken when traversing the landscape’. Whilst this robustness property is necessary when aiming to explain known past routes via cost functions (see below case study), it is also useful when predicting where people might have moved in the past. In these cases, the past route is unknown and must therefore be inferred. Given the robustness property when using multiple cost functions, it is thus possible to calculate robust least-cost paths under the assumption that ‘humans minimised time/energy when traversing the landscape’.

![Figure 2 Robustness of cost functions based on least-cost path ‘type’ and associated Hypothesis ‘type’.

Increased percentage agreement between the same least-cost path Type and Hypothesis Type indicates that the hypothesis that the cost functions aim to represent is robust. With increased agreement between least-cost path Type and Hypothesis Type, more support can be placed on the hypothesis as being sufficiently represented in the resultant least-cost paths.

Case Study 2: Explaining the ‘a Karalibus Sulcos’ Roman road in Sardinia

Having shown that the hypotheses represented by multiple time- and energy-based cost functions are robust, the use of multiple cost functions via probabilistic least-cost paths was applied for the explanation of the ‘a Karalibus Sulcos’ Roman road in Sardinia.

Background
Following Nuragic transhumance routes and Punic-era roads, the ‘a Karalibus Sulcos’ Roman road connected the city of Sulci in the south-west of Sardinia to the port at Carales in the south (Atzori 2006; Mastino 2005, pp. 382–385; Meloni 1990, pp. 350–353). Providing an internal and more direct route than the road along the coast through Teluga and Nura (Figure 3), the road was used to transport lead silver and wheat stored at Sulci to the port at Carales before being distributed to the peninsula (Atzori 2006, pp. 11–13). Given its economic role, it can be hypothesised that the ‘a Karalibus Sulcos’ Roman road followed a route that minimised an accumulated cost from Sulci to Carales. In this case, the outcome of the hypotheses that ‘humans minimise time taken or energy expended when traversing the landscape’ will be assessed for its ability to explain the Roman road.

Figure 3 Overview of Roman Sardinia, the Roman road system, and the ‘a Karalibus Sulcos’ Roman road. Road stations and overview of the Roman road system following Mastino (2005)

Materials and Methods

The ‘a Karalibus Sulcos’ Roman road was based on that recorded by Atzori (2006, pp. 61–111). With a digitised version of the road by Atzori (2006) not available, the road was digitised by the current author. The topography of Sardinia was represented using the TINITALY 10m resolution DEM (RMSE, 4.3m) (Tarquini et al. 2007). Like the tactical simulation above, thirteen time- and three energy-based cost functions were used concurrently to represent the hypothesis of minimising time taken and minimising energy expended, respectively. Probabilistic least-cost paths were calculated from the intersection of the ‘a Karalibus Sulcos’ Roman road and the Roman road along the western coast to the modern-day town of Decimomannu (Figure 3).

The uncertainty in the LCPs as a result of random error in the DEM was propagated through the analysis, with each LCP calculated fifty times with different realisations of random error (Lewis 2021). In contrast to Lewis (2021), autocorrelation in random error was not accounted for, with completely random unfiltered error fields used instead (Wechsler and Kroll 2006). The use of unfiltered error fields has two advantages: (1) reduction in computational burden as autocorrelation does not need to be calculated; and (2) no assumptions are made about the spatial relationships of the random error. As a result of the second
advantage, the effect of the random error fields can be deemed as the worst-case scenario (Wechsler and Kroll 2006).

Results and Discussion

With multiple cost functions sharing the same hypothesis shown to be robust, the time- and energy probabilistic least-cost paths can be deemed to sufficiently reflect the outcome of these hypotheses. As a result, the hypothesis that ‘humans minimise time-taken when traversing the landscape’ is unable to explain the ‘a Karalibus Sulcos’ Roman road, i.e. the time-based probabilistic least-cost path and the known Roman road share little similarity in their route (Figure 4, A).

Figure 4 Time- and energy-based probabilistic least-cost paths (A and B respectively) from the origin at destination of the ‘a Karalibus Sulcos’ Roman road. Probabilistic least-cost paths include all least-cost paths calculated from cost functions sharing the same hypothesis. Probabilistic least-cost paths can be deemed to sufficiently reflect the outcome of the hypotheses.

Whilst the most probable route of the energy-based probabilistic least-cost path (Figure 4, B) also shares little similarity to the ‘a Karalibus Sulcos’ Roman road, the probability of the energy-based probabilistic least-cost path sharing a similar route to the Roman road is not zero. Thus, it is possible that the route of the ‘a Karalibus Sulcos’ Roman road was chosen to minimise energy expenditure. Alternatively, the LCPs following the route of the Roman road might be the result of modern-day roads captured within the high-resolution DEM. As discussed by Herzog and Yépez (2015), DEMs – unless corrected for – can contain modern-day
activities that post-date the ancient landscape. Overcoming this is difficult however, with reconstructions of past landscapes often only available at low resolution and thus not capturing small features (Herzog and Yépez 2015).

More generally, the lack of agreement between the time- and energy-based probabilistic least-cost paths suggests that the ‘a Karalibus Sulcos’ Roman road was not constructed to minimise time taken or energy expended. Instead, it can be suggested that the use of pre-existing routes and settlements therein played a greater role in the placement of the Roman road. Whilst dating of settlements along the ‘a Karalibus Sulcos’ Roman road is difficult (Atzori 2006, pp. 31–60), Tronchetti (1995) argues for settlement continuity from the Punic to the Roman Republican period within south-west Sardinia. Later, during the Imperial period, many of these settlements would have developed infrastructure related to the Cursus Publicus, the state system for land transportation (Atzori 2006, pp. 31–60). The re-use and formalisation of pre-existing routes is not however unique to the ‘a Karalibus Sulcos’ Roman road, with this process occurring throughout Roman Sardinia (Barreca 1974, pp. 65–68; Tetti 1985), and the Roman Empire more widely, e.g. the Etruscan road system in Italy (Ward Perkins 1957), the ‘road of Hercules’ from Spain to Italy (Campedelli 2013), and the route across the Stainmore Pass in Britain (Robinson 2001). By making use of pre-existing routes and settlements when constructing the formalised Roman road system, the Romans could more easily control the movement of goods and local traffic across Roman Sardinia (Atzori 2016).

As a corollary, this also suggests that the pre-existing routes did not follow a route that minimised time taken, or energy expended. Given that the pre-existing routes from Sulci to Carales are thought to have been the result of transhumance, the seasonal movement of shepherds and their flocks between different regions, it is possible that these routes followed a path most preferable for their movement.

**Conclusion**

Using the idea of multiple model idealisation, this paper has shown that the hypotheses ‘humans minimise time/energy when traversing the landscape’ when represented by multiple time- and energy-based cost functions are robust. As a result, multiple LCPs calculated using cost functions that share the same hypothesis, i.e. minimising time or energy, produce results that are deemed to sufficiently represent the shared hypothesis.

The use of time- and energy-probabilistic least-cost paths as a method for operationalising the idea of multiple model idealisation reduces the chance of spurious results. With this, the probabilistic LCPs, shown to sufficiently represent the outcome of the hypotheses ‘humans minimise time/energy when traversing the landscape’, can credibly be used to assess their ability in explaining known past routes. It is only by assessing the outcome of hypotheses and not individual cost functions against known past route that we can place more support on these hypotheses operationalised via multiple cost functions and LCP analysis as having explanatory power for uncovering why people in the past moved where they did.

**Data and Materials Availability**
All data and code necessary to reproduce the analyses are available on the GitHub repository

TBC

Statements and Declarations

I have no conflicts of interest to disclose.
References


https://doi.org/10.11141/ia.36.5


https://doi.org/10.1080/15230406.2017.1292150


https://doi.org/10.1007/s10816-021-09522-w


https://doi.org/10.1080/00167223.2017.1316212


https://doi.org/10.1152/japplphysiol.01177.2001


https://doi.org/10.1016/j.cageo.2003.11.001


Tarquini, S., Isola, I., Favalli, M., & Battistini, A. (2007). TINITALY, a digital elevation model of Italy with a 10
meters cell size. Istituto Nazionale di Geofisica e Vulcanologia (INGV).

https://doi.org/10.13127/TINITALY/1.0

Scale in Geographical Information Systems (pp. 35–51). New York: John Wiley & Sons.


Barbara, CA: National Center for Geographic Information and Analysis.

https://escholarship.org/uc/item/05r820mz. Accessed 22 December 2019


Photogrammetric Engineering & Remote Sensing, 72(9), 1081–1090.

https://doi.org/10.14358/PERS.72.9.1081


https://doi.org/10.1086/518628


https://doi.org/10.5840/jphil20071041240