Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A data-driven framework for inferring past movement decisions from archaeological routes

Joseph Lewis

jl2094@cam.ac.uk

University of Cambridge

Research Article

Keywords: movement, Roman Sardinia, Roman roads, step selection, human behaviour, least-cost path

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8948011/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A data-driven framework for inferring past movement decisions from archaeological routes

Dr Joseph Lewis*

University of Cambridge, Cambridge, United Kingdom*

jl2094@cam.ac.uk*

0000-0002-0477-1756

*Corresponding Author: Dr Joseph Lewis

Keywords

movement; Roman Sardinia; Roman roads; step selection; human behaviour; least-cost path

Abstract

The spatial organisation of movement is a fundamental structuring principle of past societies, reflecting their technological capabilities and socio-economic priorities. Yet the processes that governed movement decisions in the past remain poorly understood. Here, a data-driven step-selection analysis framework is introduced that statistically infers the decision-making processes that governed the formation of past routes. Applied to Roman roads in Sardinia, the approach reveals a robust and generalisable engineering strategy characterised by a systematic avoidance of steep slope gradients. Substantial road-specific variation, however, indicates local adaptations to topographic constraints and logistical motivations. These results show that Roman road formation in Sardinia was governed by culturally specific decision-making processes rather than universal physiological costs. This approach provides a new framework for reconstructing the processes underlying past movement.

Introduction

Movement constitutes a fundamental structuring principle of past societies. As a primary mechanism for the transmission of resources and information, the spatial organisation of movement reflects both the technological capabilities and the socio-economic priorities of a past society (Trombold 1991). Inferring the decision-making processes that governed movement is therefore central for investigating past human-environment interactions and forms of socio-political complexity. Routes, including both trackways and roads, used by those in the past provide a material proxy for inferring these decision-making processes (Earle 1991). While the spatial trajectory of routes can now be reconstructed with increasing precision (e.g. De Soto et al. 2025), inference, however, remains difficult. This difficulty is partly conceptual. Traditionally, routes have been treated as static infrastructure that link sites in the landscape, with emphasis placed on their overall geometry and not the processes through which they were formed (Gorenflo and Bell 1991). Yet, a single route is the cumulative outcome of a series of locally constrained decisions embedded within specific landscape and cultural contexts. At this local scale, movement is a trade-off between physical constraints imposed by the landscape and cultural motivations such as the transportation of resources. By

conceptualising routes in this way, movement can be understood as a sequential decision-making process and provides the basis for evaluating how these local choices shaped past route formation.

For the past three decades, the quantitative study of past movement has been dominated by least-cost paths calculated using cost functions derived from physiological experiments (Herzog 2020). Here, a least-cost path is the optimal path that minimises an accumulated cost between a pre-determined origin and destination, while cost functions mathematically represent the cost of traversing slope gradient. Costs used in least-cost path analyses are most commonly how much time it takes to traverse a slope gradient or how much energy is expended (Lewis 2024). Examples of cost functions include the time-based Tobler's Hiking Function (Tobler 1993) and the energy-based sixth degree polynomial introduced by Herzog (2014). While least-cost paths derived using these cost functions have proven valuable for both predicting and explaining past routes (Lewis 2021, 2024), cost functions derived from physiological experiments have significant foundational limitations when used to infer the decision-making processes that governed past route formation.

First, by calculating a least-cost path using a cost function derived from physiological experiments, it is assumed that route formation minimised an *a priori*, generalised physiological cost—such as time or energy expenditure—rather than being structured by culturally specific motivations. By enforcing a universal physiological cost function onto the past, such analyses risk obscuring the decision-making processes influencing past movement, failing to account for 1) the possibility of a different functional form and 2) other specific cultural factors such as the influence of slope gradient on wheeled vehicles or the ideological projection of power through linear trajectories. The former would be indicated by a preference for shallower slope gradients, the latter an indifference to slope gradient.

Second, LCP analyses are typically deterministic and predict a single 'optimal' path that fails to incorporate statistical uncertainty. While recent developments have addressed 'environmental' uncertainty in the input data—principally by accounting for measurement error in the Digital Elevation Model (Lewis 2021)—the cost functions themselves are rarely defined as statistical models. Instead, parameters in cost functions, e.g. the coefficient governing speed decay in Tobler's Hiking Function, are typically presented as point estimates fitted to experimental data (Lewis 2024). These point estimates are subsequently used in archaeological analyses as fixed, universal constants, ignoring the statistical uncertainty in the original cost function model fit. As these parameters are treated as absolute values rather than those with associated uncertainty, traditional LCP analyses lack the parameter variance necessary to quantify the reliability of the cost function itself. As a consequence, predicted LCPs fail to propagate parameter uncertainty in cost functions, masking the potential range of LCPs that are statistically plausible given the original data used to generate the cost function.

Third, LCP studies frequently ignore inter-route behavioural variability by using a single cost function when aiming to explain multiple routes, effectively assuming that the formation process of these routes was identical. While a number of studies have explicitly addressed high behavioural variability in

physiological data by splitting data and estimating cost functions at different percentiles (e.g. Campbell et al. 2019), the majority of archaeological analyses continue to homogenise this variability. Herzog (2022), for example, used a single, shared cost function to model multiple trade routes, finding that a high LCP model fit for one route does not automatically imply a high model fit for another route. Crucially, this homogenisation can obscure the specific functional or cultural motivations—such as military strategy, trade, or pastoralism—that underpinned the formation of individual routes. By failing to account for this heterogeneity, traditional LCP analyses treat distinct strategic choices as statistical noise rather than evidence of varying function or agency.

Lastly, traditional LCP analyses often validate the use of a specific cost function by comparing the final geometry—the predicted least-cost path—to the empirical route (Lewis 2024). This focus on the 'product' of movement rather than the decision-making processes that governed movement is prone to equifinality, whereby the empirical route can be sufficiently predicted using different decision-making processes. Lewis (2024), for example, showed that minimising time or energy can predict the same least-cost path. More generally, an LCP might also achieve a high spatial fit by chance or due to the constraints of terrain such as a narrow valley, even if the cost function fails to capture the true decision-making process of route formation.

As an alternative to the use of cost functions derived from physiological experiments and to better represent the dynamic processes influencing route formation, this paper proposes the use of step-selection analysis (SSA), an approach adapted from movement ecology (Thurfjell et al. 2014). In contrast to using physiologically-derived cost functions, an SSA models an empirical route as a series of sequential, locally-defined movement decisions. At each 'step' along the route trajectory, the approach compares the selected step against a set of alternative locations that were accessible but not selected (**Figure 1**). Here, potential steps are limited to eight adjacent cells, or the 'Moore neighbourhood'. By analysing these selection decisions, the SSA estimates which factors and their values were consistently selected for or avoided—revealing the decision-making processes that governed the formation of past routes.

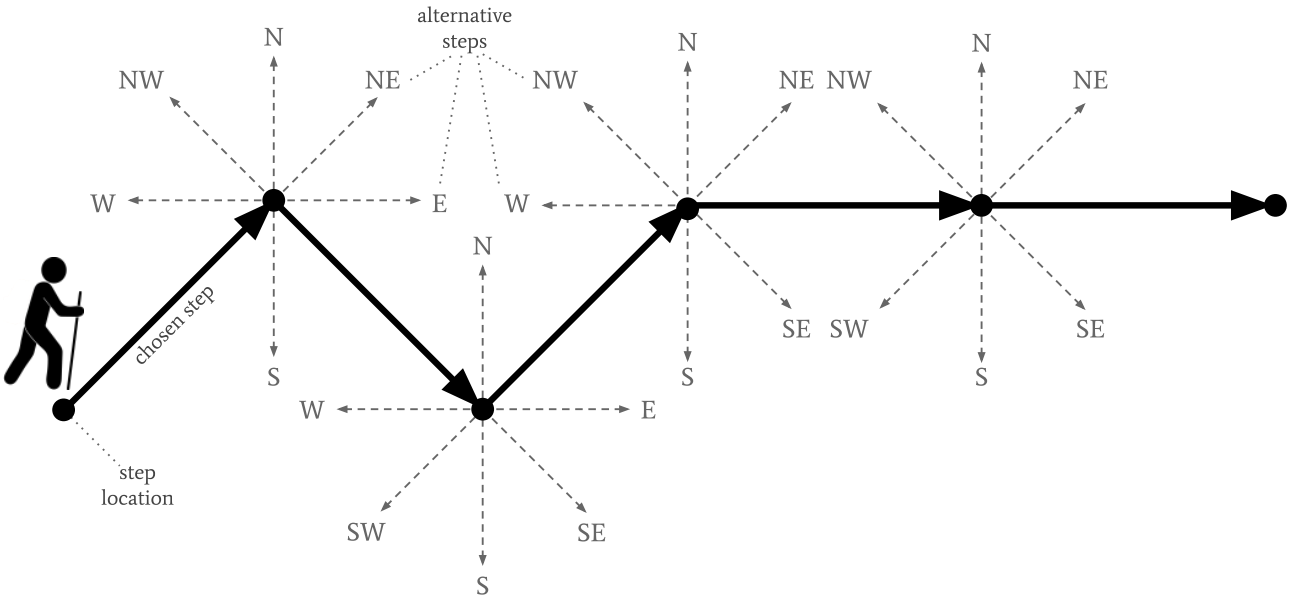


Figure 1. Example of a route trajectory in a step-selection analysis. The route trajectory is discretised into distinct selected steps that connect successive step locations (bold black arrows). For each step, a set of available steps (eight adjacent cell, or 'Moore neighbourhood') defines the choice set or stratum. The step-selection analysis compares the attributes of the selected step against these available alternative steps to statistically infer the movement selection process

Here, step-selection analysis is applied to the road network of Roman Sardinia (**Figure 2**), a landscape characterised by rugged terrain and complex colonial dynamics (Dommelen 1998). Following the Roman annexation in 238 BCE, Roman roads on the island were progressively constructed to consolidate military control over those in the interior and secure the transportation of resources, principally grain, lead, and silver to *Olbia*, the port closest to the *Lazio* coast (Lewis 2025; Mastino 2005). This system of roads largely formalised pre-existing Nuragic transhumance routes and Phoenician-Punic roads (Barreca 1974; Tetti 1985). From the reign of Augustus in the Imperial period, many of these trackways were systematically standardised into roads spanning 5-8 metres wide, surfaced with gravel and consolidated with pebbles and stones, with their margins delineated by unfinished blocks of stone (Mastino 2005, pp. 333–340).

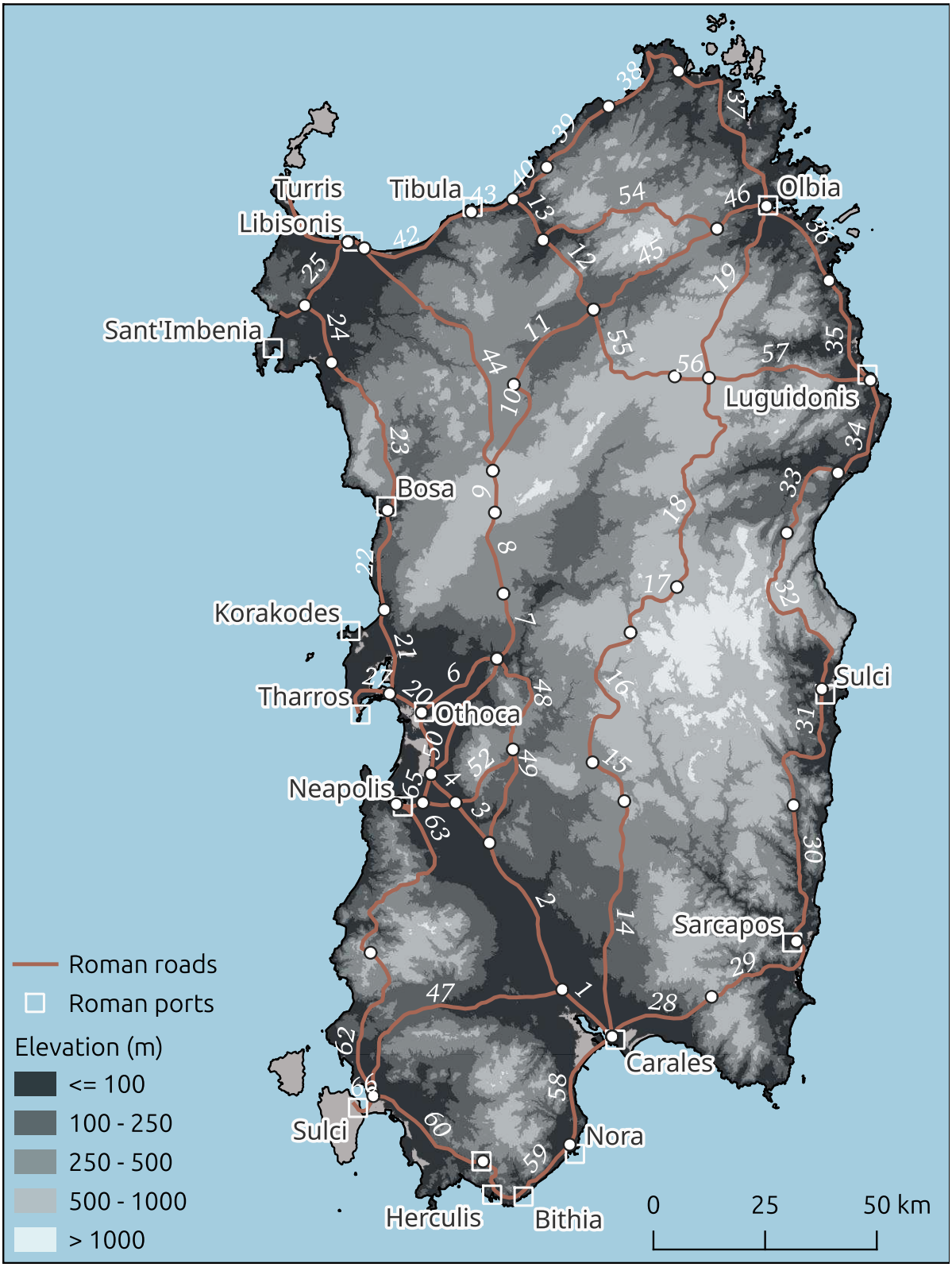


Figure 2. Study region of Sardinia with the Roman road system (Lewis 2025). Roman road system digitised from Mastino (2005). Individual Roman road segment IDs denoted by white text. Start and end of a Roman road segment (white circles) is defined using the location of Roman overnight stopping places *(mansiones)* and Roman cities from Mastino (2005) and where Roman roads intersect. Key Roman port cities are labelled

2. Materials and Methods

All data and R scripts are available at https://github.com/josephlewis/sardinia_roads_prediction and permanently archived at https://doi.org/10.5281/zenodo.18961598.

2.1 Roman roads in Sardinia

Step-selection analysis was used to infer the influence of slope gradient on the formation process of 66 Roman road segments in Sardinia. Roman roads were digitised from Mastino (2005) and are available at Brughmans et al. (2024). A total of 525,552 distinct movement steps were modelled, with 65,694 of those representing selected movement steps. The number of unique choice sets or stratum (i.e. both selected, and available but not selected steps), within a single Roman road ranged from 35 and 2,680, with a mean value of 995.

2.2 Influence of slope gradient

To incorporate the selection of specific slope gradients by Roman roads in Sardinia, mathematical slope gradient was calculated for every used and available step along each of the 66 Roman road segments (**Figure 3**). Mathematical slope gradients were derived from the 30m Digital Terrain Model for Continental Europe (2020). Slope gradient is widely recognised as influencing the trajectories of past routes including Roman roads, and is supported by both empirical evidence from Roman road sections and previous least-cost path analyses (Chevallier 1976; Lewis 2021; Parcero-Oubina et al. 2023). Given that the selection of slope gradient is assumed to decrease for both steeper ascents and steeper descents, the absolute value of slope gradient was used. This transformation of slope gradient from anisotropic to isotropic treats positive and negative slope gradients as equally 'costly' deviations from a 0° slope gradient and focuses the analysis on the overall magnitude of selection rather than the direction of vertical change.

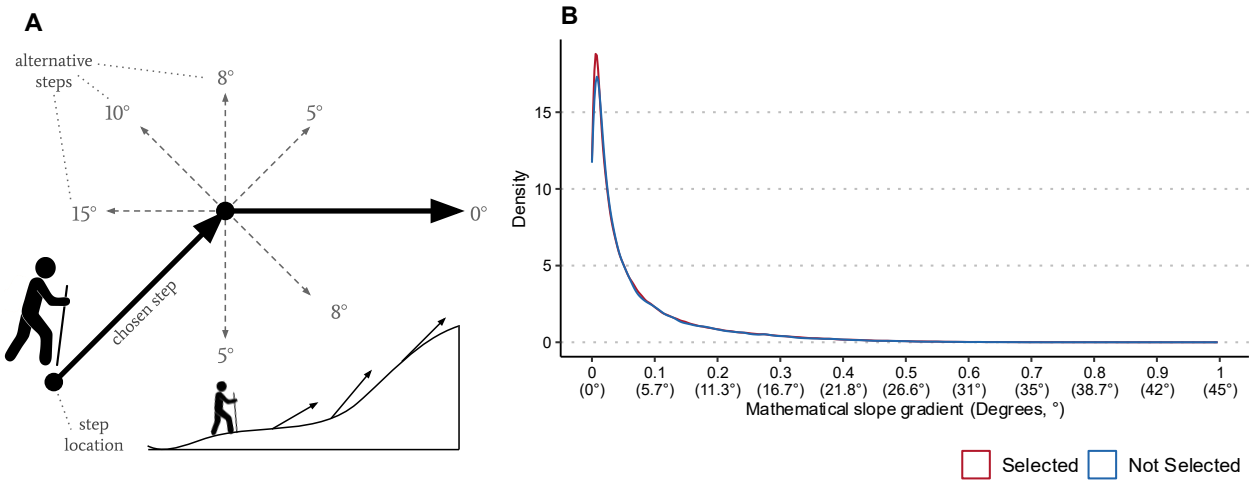


Figure 3. Schematic representation of mathematical slope gradient selection (**A**). Slope gradient is calculated as the change in elevation between two adjacent step locations. Mathematical slope gradients that are selected and not available but not selected by Roman roads in Sardinia (**B**)

2.3 Step selection analysis

Following the approach of Muff et al. (2020), the selection process was estimated using a Poisson model. Let y_{ijk} represent the binary observation (1 for selected step, 0 for available but not selected step) for step i within a choice-set (stratum) j on Roman road k.

$$y_{ijk} \sim \text{Poisson}(\lambda_{ijk})$$

The linear predictor for the log-intensity, $\log(\lambda_{ijk})$, was formulated as:

$$\log(\lambda_{ijk}) = \gamma_j + f(x_{ijk}) + u_k \cdot x_{ijk}$$

where:

- γ_j is the stratum-specific intercept. To mathematically equate this Poisson model to a conditional logistic regression and prevent shrinkage to the global mean, the intercepts were assigned a large, fixed precision ($\tau = 10^{-6}$). This parameterisation ensures that the baseline availability within each distinct choice-set is evaluated independently.
- $f(x_{ijk})$ is a global, non-linear second-order random walk (RW2) smooth function for the selection of mathematical slope gradient x.
- u_k is an independent and identically distributed (IID) random slope coefficient for Roman road k, capturing road-specific deviations from the global selection of mathematical slope gradient

To interpret the fitted model in terms of movement selection, the slope-related terms of the linear predictor were exponentiated to derive the relative selection weight, $w(x)$, independent of the baseline stratum availability:

$$w(x) = \exp(f(x) + u_k \cdot x)$$

All non-linear smooth functions were subsequently normalised so that a 0° slope gradient corresponds to a relative selection weight of 1.

For efficient Bayesian computation, the model was fit using the R packages *INLA* (Rue et al. 2009) and *inlabru* (Bachl et al. 2019). The smoothness of the RW2 function ($f(x_{ijk})$) was governed by a default log-Gamma priors ($1{,}5 \times 10^{-5}$), while the route-specific random slopes u_k were regularised using Penalised Complexity (PC) priors (Simpson et al. 2017). Specifically, the PC prior for the standard deviation of the random slopes was set to $P(\sigma_u > 1) = 0.05$, penalising extremes deviations from the global model unless strongly supported by the data.

Two candidate models of increasing complexity were evaluated to determine the most parsimonious model for predicting Roman road formation (**Table 1**). The '*M1 (complete pooling)*' assumes that the formation process of all Roman roads in Sardinia was identical, while the '*M2 (partial pooling)*' model allows the formation process of each road to vary. Models were compared using the Watanabe-Akaike Information Criterion (*WAIC*) (Watanabe 2010). WAIC is a fully Bayesian criterion that estimates the out-of-sample predictive accuracy of a model, penalising for the effective number of parameters to prevent

over-fitting. Lower WAIC values signify a higher model predictive accuracy. Model selection was determined by calculating the WAIC for each of the two candidate models.

Table 1. Candidate model parameters and WAIC scores for SSA estimation of Roman roads

Model Name	**Model components**	**WAIC**	**Δ**WAIC
M2 (partial pooling)	Global Slope (RW2) + Route Random Slope (IID) + Stratum Intercepts	582,912.6	0.0
M1 (complete pooling)	Global Slope (RW2) + Stratum Intercepts	582,956.1	43.5

Random Slope parameters are included to incorporate variation among Roman roads in Sardinia. Delta ($\boldsymbol{\Delta}$)WAIC values are the difference in WAIC between the model with the highest out-of-sample predictive accuracy and the stated model.

2.4 Estimating predictive accuracy

The predictive accuracy of the optimal '*M2 (partial pooling)*' model was assessed using a posterior predictive ranking procedure (**Figure 4**). This approach evaluated whether the model could correctly predict the selected step from the set of available and available but not selected steps within each stratum. Performance was evaluated via two complementary approaches. First, internal goodness-of-fit was assessed by the ability of the model to reconstruct selected steps within the full dataset used for training, characterising its capacity to describe movement steps across all 66 modelled Roman road segments. Second, external predictive accuracy was assessed using a Leave-One-Out-Cross-Validation (*LOOCV*) approach. In this approach, the optimal model was iteratively re-fitted on 65 $(N-1)$ road segments to predict the selected steps on the excluded road segment using only the global smooth function. While internal goodness-of-fit characterises the explanatory power of the model on the training data, the *LOOCV* approach assesses the generalisability of the model by evaluating its predictive performance on novel Roman roads entirely excluded from the modelling process.

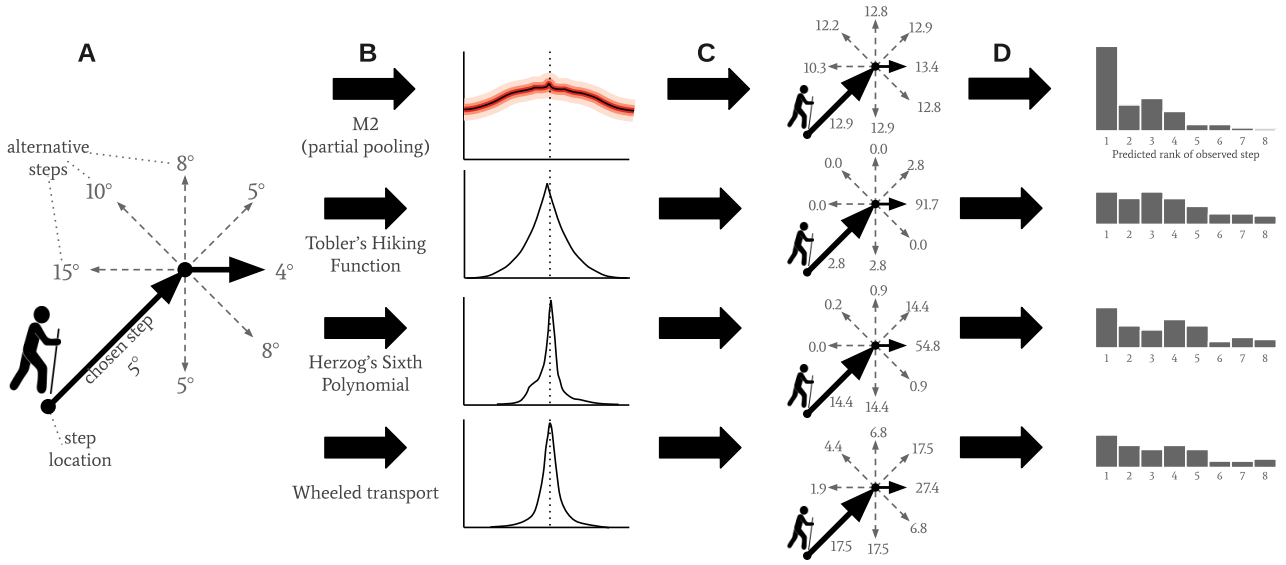


Figure 4. Schematic of the predictive ranking procedure. For each stratum, the empirical step (black solid line) is compared against the seven available but not selected steps (dashed lines) (**A**). Compared model function curves (**B**). All steps are assigned a 'score' based on the model curve (**C**). Steps are ranked 1-8 based on this score. The rank of the empirical Roman road step is used to built the predictive rank accuracy distribution (**D**)

To account for parameter uncertainty in the selection for slope gradient, 1,000 samples were generated from the joint posterior distribution of the optimal '*M2 (partial pooling)*' model. For each generated sample, the following procedure was performed:

i. *Response curve reconstruction*: the global non-linear *RW2* response curve for slope gradient was reconstructed based on the parameter values of that specific posterior draw;

ii. *Scoring*: for each stratum (one selected step and seven available but not selected steps), a predicted selection weight was calculated for each step by mapping its slope gradient value to the reconstructed response curve;

iii. *Ranking*: The steps within each stratum were ranked based on these predicted selection weights, where Rank 1 represents the highest predicted relative probability of selection and Rank 8 the lowest.

This procedure produced a distribution of predicted ranks for the empirical data across all 1,000 posterior draws. Predictive accuracy was summarised by calculating the frequency with which the selected step was assigned Rank 1. These results were compared against a random chance baseline of 12.5% (1 in 8).

To benchmark the predictive accuracy of the optimal M2 model, an identical ranking analysis was conducted using three established anisotropic cost functions commonly used in least-cost path analysis: the time-based Tobler's Hiking Function (1993), Herzog's (2014) energy-based sixth degree polynomial, and a wheel-based cost function with a 12% critical slope also proposed by Herzog (2013). For these three

deterministic cost functions, conductance (the inverse of cost) was calculated for all selected and available but not selected steps based on their slope gradient values. Steps were ranked within their strata according to their conductance, with Rank 1 indicating the most conducive, or least costly step for movement. The distributions of ranks derived from these cost functions were compared to the posterior predictive ranks of the M2 model to determine if an empirically derived 'selection weight function'—reflecting the specific decision-making processes of Roman roads in Sardinia—yielded higher predictive accuracy than the three generalised physiological baselines.

3. Results and Discussion

Model comparison identified that the 'M2 *(partial pooling)*' model achieved the highest out-of-sample predictive accuracy (lowest WAIC) and was therefore chosen as the optimal model. The improvement in fit of the *M2 model* over the 'M1 *(complete pooling)*' model (**Δ**WAIC = 43.7) indicates strong evidence against a single, globally uniform selection of slope gradient. Roman roads in Sardinia, instead, exhibit significant road-specific heterogeneity, where selection of slope gradient values varied between different roads—likely reflecting local adaptations to specific topographic constraints or logistic motivations.

3.1 Selection of slope gradient

The optimal ('M2 *(partial pooling)*' model characterises a typical Roman road formation in Sardinia as strongly selecting against steeper slope gradients (**Figure 5**). The relative selection weight for a 45° slope gradient is 80% lower than for a 0° slope gradient (56–91% reduction, 95% Credible Interval), with this sharp decrease reflecting a deliberate engineering strategy to avoid steeper slope gradients. This strategy is also evident at shallower gradients: the relative selection weight for a 6° slope gradient—a value typical of Roman roads in Italy (Tuppi 2014)—is, for example, 15% lower than for a 0° slope gradient (10–20% reduction, 95% CI), further demonstrating a pronounced sensitivity to the selection of shallower slope gradients.

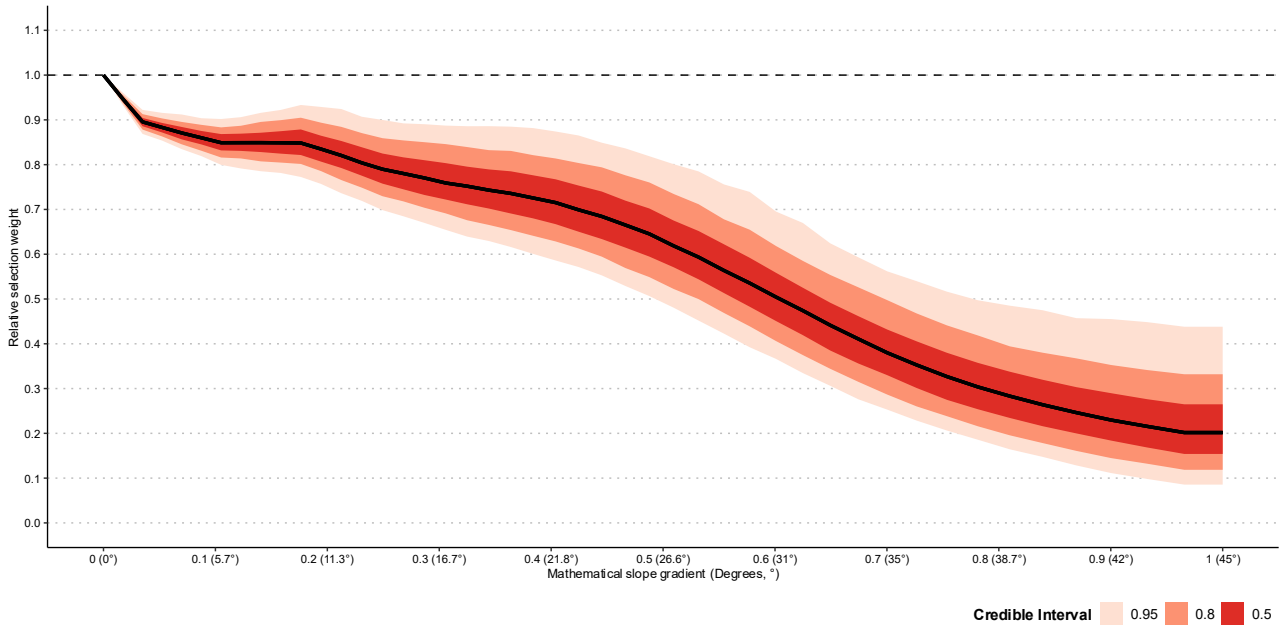


Figure 5. Predicted relative selection weight of mathematical slope gradient for a typical Roman road in Sardinia, derived from the optimal '*M2 (partial pooling)*' model. The y-axis represents the relative selection weight ($w(x)$), normalised such that a 0° slope gradient is equal to 1. The fitted curve shows a non-linear selection for shallower slope gradients

The inferred mean selection weight function for Roman roads exhibits a distinct functional form compared to established physiological and wheel-based cost functions (**Figure 6**). While all functions share a general decline in selection as slope gradient increases, the mean selection rate of decay for a typical Roman road is markedly different. In contrast to the three established cost functions, the typical Roman selection curve exhibits a 'broader' profile with a more gradual decline. Even at 45° slope gradient, the relative selection weight for Roman roads remains notably higher (~0.2) than the near-zero conductance predicted by the three established functions. This suggests that the empirical "engineering tolerance" for slope gradient in Roman Sardinia was systematically broader than the constraints predicted by either physiological or wheel-based cost functions.

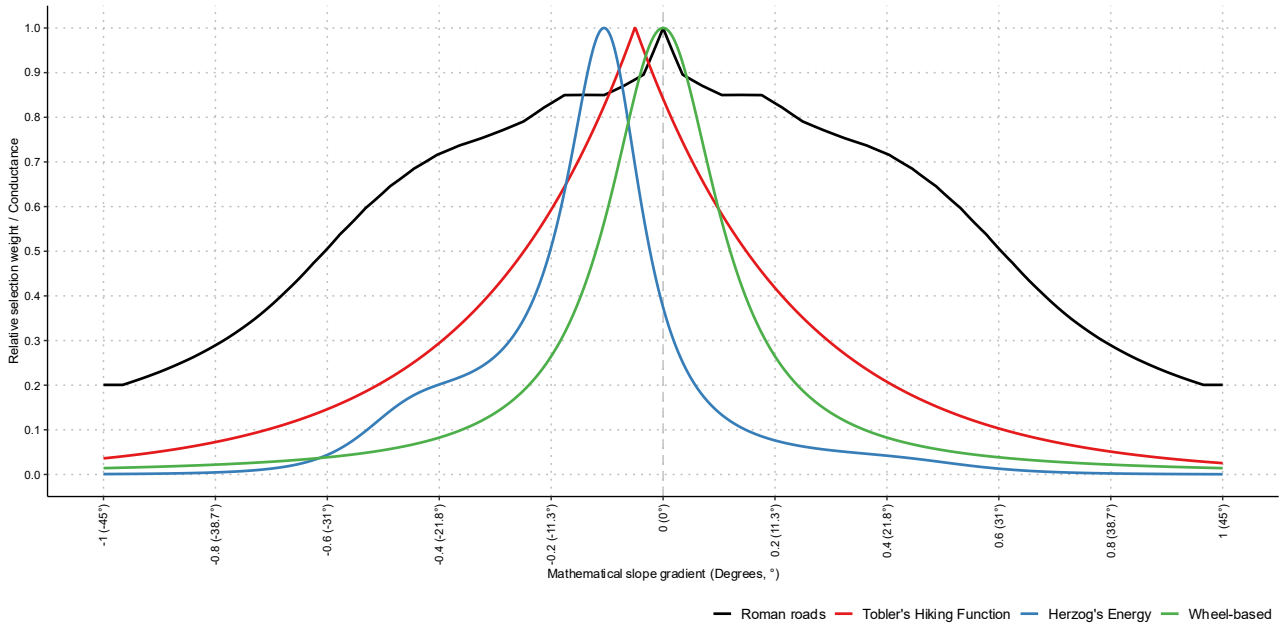


Figure 6. Comparison of Roman road selection weights against physiological cost functions. Predicted relative mean selection weight of mathematical slope gradient for a typical Roman road in Sardinia, derived from the optimal '*M2 (partial pooling)*' model (black line). Conductance values for Tobler's Hiking Function (red line), Herzog's energy-based polynomial (blue), and a wheel-based cost function with a 12% critical slope (green). All functions are normalised such that their peak values equal 1.

Predicted relative selection weights for individual Roman roads also reveals a distinction between the typical engineering strategy and local adaptations. While the typical Roman road shows a sharp decrease in selection of steeper slope gradients, road-specific relative selection weight predictions display significant heterogeneity (**Figure 7**).

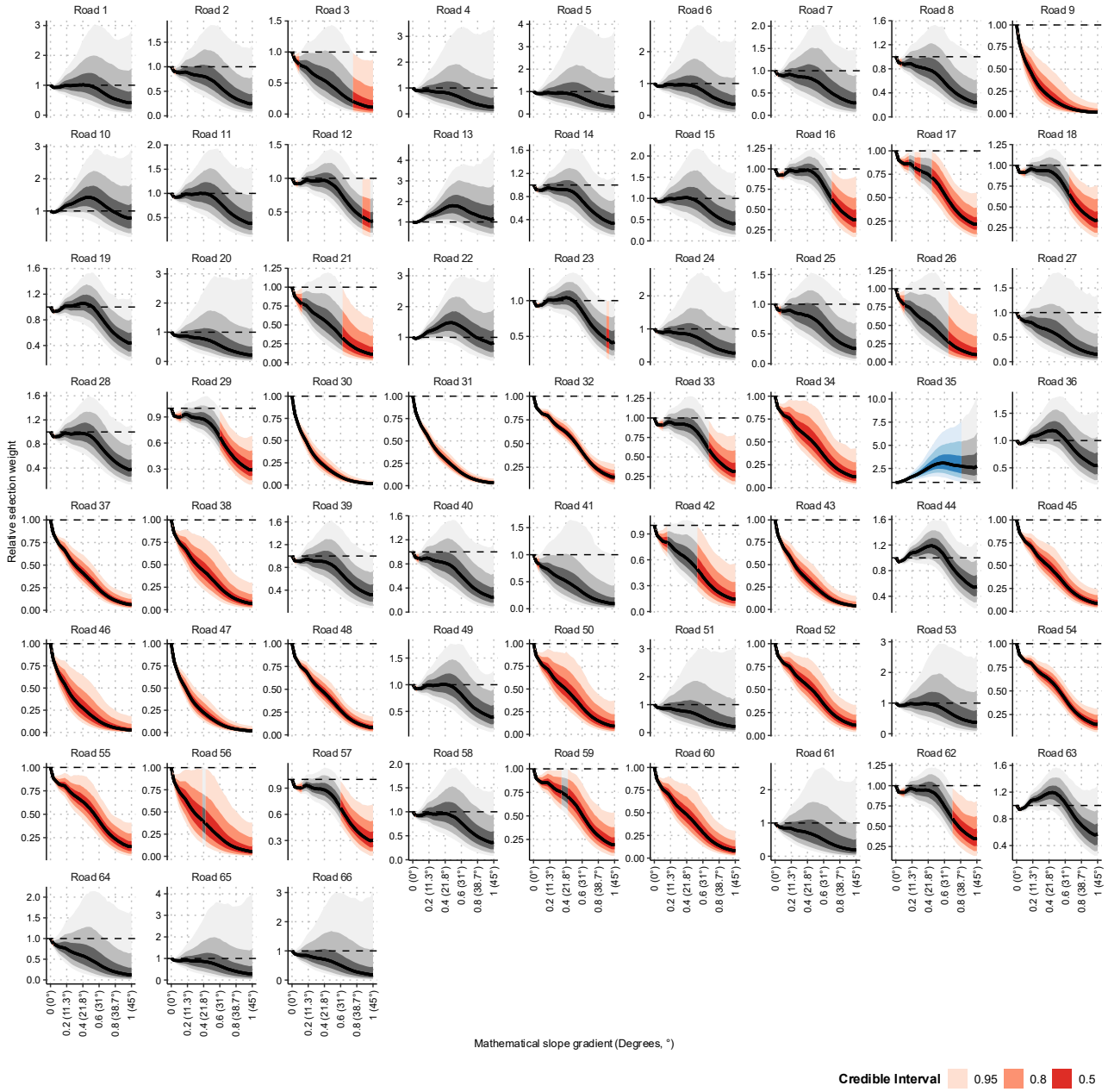


Figure 7. Heterogeneity in selection behaviour across Roman roads in Sardinia. Predicted relative selection weight (w) for individual roads are derived from the optimal '*M2 (partial pooling)*' model, with values normalised to a 0° slope gradient baseline ($w = 1$). Selection curves are categorised by their 95% credible intervals: above baseline (blue) indicates a selection preference for steeper slope gradients relative to the 0° slope gradient baseline; below baseline (red) indicates a selection avoidance in favour of the 0° baseline; and at baseline (grey) indicates neutral selection, where selection does not credibly differ from the 0° baseline

Aggregating the selection behaviour of individual Roman roads in Sardinia (**Figure 7**) across the full posterior distribution reveals three distinct engineering strategies: 1) slope-preferential roads (above baseline); 2) slope-averse roads (below baseline); and 3) slope-neutral roads (at baseline) (**Figure 8**). *Roman road segments 26*, *42*, and *43* in the north-west of Sardinia, for example, show an aversion to slope

gradients steeper than the 0° slope gradient baseline. These slope-averse roads are likely influenced by the logistical difficulties when transporting heavy goods, such as grain for exportation. The use of these roads for transporting heavy goods is supported by the port at *Turris Libisonis*, also in the north-west Sardinia. Serving as a key collection centre for products destined for export, the port and its nearby granaries (*horrea*) functioned as an interface between the countryside and overseas markets (Mastino and Vismara 1994). Evidence of commercial traffic is also visible in the nearby *Su Crucifissu Mannu* area, where traces of cart wheels are preserved in a road surface (Mastino and Vismara 1994). Likewise, *road segment 29* shows an aversion to steeper slope gradients. Historically, *road 29* connected the mining area at *Ferraria* to the port at *Sarcapos* (Mastino 2005). Crucially, the selection for shallower slope gradients for this road also reflects the specific logistical difficulties of transporting heavy goods. Lastly, the slope-averse *road segment 47* connecting the city of *Sulci* to the port at *Carales* in the south-west of Sardinia was used for the transportation of lead silver and wheat before being distributed from *Carales* to *Italy* (Atzori 2006; Lewis 2024).

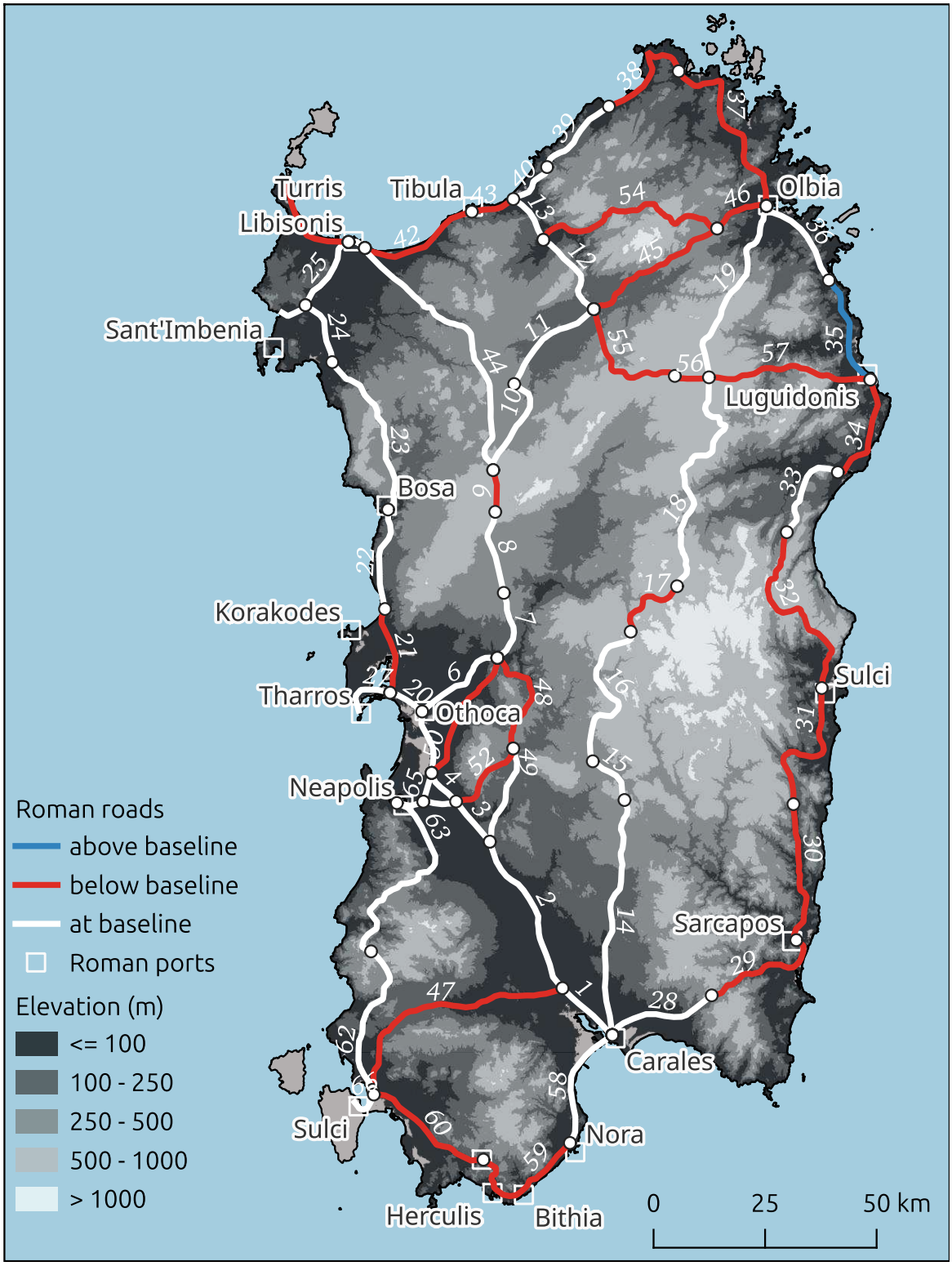


Figure 8. Spatial distribution of aggregated selection behaviour across Roman roads in Sardinia. Individual road segments are coloured by their mean relative selection weight (*w*), derived from the optimal 'M2 (partial pooling)' model. Routes are categorised by the 95% credible interval of these aggregated mean weights: above baseline (blue) indicates a mean selection preference for steeper slope gradients relative to the 0° slope gradient baseline; below baseline (red) indicates a mean selection avoidance in favour of the 0° baseline; and at baseline (white) indicates mean neutral selection, where selection does not credibly differ from the 0° baseline

The presence of all three engineering strategies along the 362 km eastern coastal road unit connecting *Portus Tibulas* and *Carales* also provides empirical support for heterogeneity in decision-making processes. The presence of the slope-preferential *Roman road segment 35* along this road unit, where steeper slope gradients are more likely to have been selected than a 0° slope gradient, supports the previous hypothesis by Mastino (2005, pp. 341–352) that this road unit was conceived not as a single entity from the beginning but was formed during the gradual conquest of Sardinia. That is, if it were conceived as a single entity, it would be expected that the decision-making process would be similar for all road segments in the road unit, for example with the selection for shallower slope gradients to facilitate the use of wheeled vehicles. The identification of *Roman road segment 35* as being slope-preferential instead indicates a prioritisation of directness over the avoidance of steeper slope gradients. This is corroborated by the *road segment 35* being more direct than the current modern road, as well as its failure to deviate for waterbodies, including river deltas (Mastino 2005, pp. 345–346). The slope-preference of *Road 35* perhaps reflects the persistence of the pre-Roman trackway that the Roman road subsequently formalised.

3.2 Estimated predictive accuracy

The posterior predictive ranking procedure reveals a sharp disparity in the ability of various functions to predict the selected step of Roman roads in Sardinia (**Figure 9; Supplementary Figures 1–2**). The empirically derived relative 'selection weight function' yields the highest predictive accuracy, assigning a correct 'Rank 1' prediction to the selected step in 42.2% of cases in both the internal goodness-of-fit approach and external Leave-One-Out-Cross-Validation approach (**Figure 9A,C**), and 42.3% based on the relative mean selection weight function (**Figure9B**). This is more than three times the value expected under random chance (12.5%), indicating that the non-linear selection weights sufficiently infers the decision-making process governing the formation of Roman roads in Sardinia. The high predictive accuracy further indicates that the avoidance of steep slopes constituted a systematic engineering strategy of Roman road formation. The comparable predictive accuracy observed under both validation approaches (**Figure 9A,C**) also demonstrates that these inferred selection weights sufficiently explain the 66 Roman road segments as well as generalise to predict the decision-making processes governing the formation of Roman roads not included in the model fitting process.

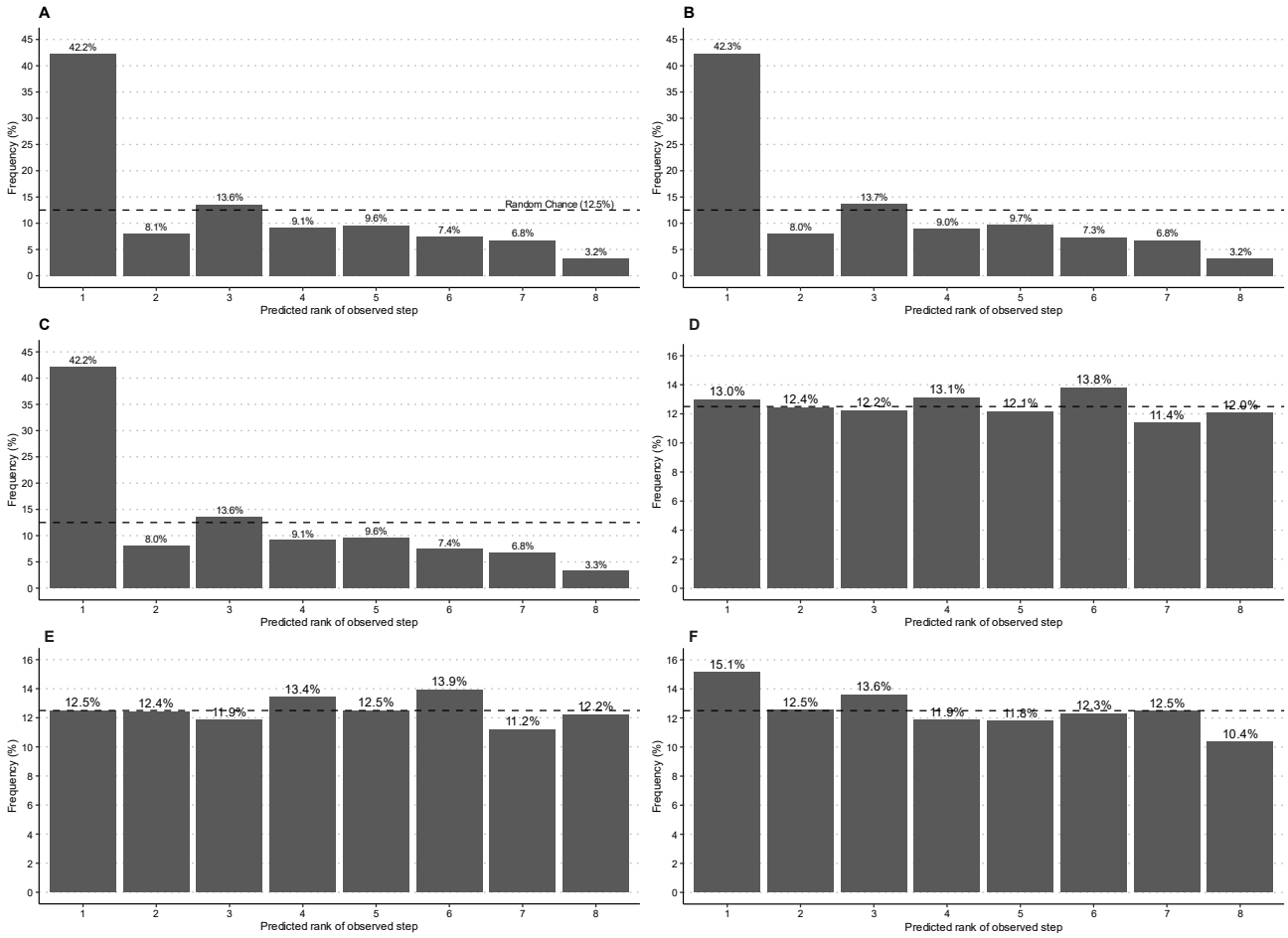


Figure 9. Predictive accuracy of model-derived selection weights versus generalised cost functions. Distributions of predicted ranks for observed Roman road steps from the internal goodness-of-fit of the empirically derived RW2 selection weight function (**A**); the goodness-of-fit using only the posterior mean selection weight function (**B**); the external leave-one-out cross-validation of the selection weight function (**C**); Tobler's hiking function (**D**); Herzog's energy-based sixth-degree polynomial (**E**); and Herzog's wheel-based cost function (**F**). Ranks are determined by the predicted relative probability of selection (**A–C**) or relative conductance (**D–F**) within each stratum, where Rank 1 represents the highest predicted probability of selection. The horizontal dashed line indicates the random chance baseline (12.5%). Both internal and external validation from the empirically derived model (**A–C**) demonstrates a high predictive accuracy (42.2% Rank 1), whereas the physiological baselines (**D–F**) perform similarly to a random probability of selection

In contrast, the three established anisotropic cost functions—Tobler's Hiking Function, Herzog's energy polynomial, and the wheeled-based cost function (**Figure9D–F**)—exhibit predictive accuracies that are similar to those that would be produced by random chance (13.0%, 12.5%, and 15.1%, respectively). This indicates that the generalised physiological costs of movement representing time, energy expenditure, and the difficulty of using a wheeled vehicle were not the primary determinants in the decision-making processes governing the formation of Roman roads in Sardinia. Instead, it is suggested that the formation of Roman roads followed a culturally specific selection process that was different than physiological

expectations. Specifically, while the physiological cost functions impose severe, immediate penalties on any slope gradient steeper than 0°, the empirically-derived selection weight functions reveal a shallower decline, and even instances of slope-preference, which more accurately reflects the varied selection processes of Roman road formation. The poor predictive accuracy of the wheel-based cost function in particular also indicates that the selection process of Roman roads was not solely a response to minimising mechanical cost, but a more complex prioritisation of slope gradient.

5. Conclusion

This study demonstrates that the formation of past routes was a complex and culturally specific process rather than dictated by physiological cost. By applying step-selection analysis to Roman roads in Sardinia, this study provides a data-driven alternative to the use of cost functions derived from physiological experiments. The results reveal that while a typical Roman road in Sardinia was slope-averse, likely to accommodate the logistic difficulties of transporting heavy goods, individual Roman roads adapted to local topographic and cultural contexts, resulting in significant road-specific heterogeneity. Crucially, the empirically derived selection weight function modelling the selection processes of Roman roads in Sardinia outperformed anisotropic cost functions in predictive accuracy, highlighting that generalised physiological costs of movement were not the primary determinants in Roman road formation. The three-fold increase in predictive accuracy over physiological functions suggests that the decision-making processes governing Roman road formation can be reconstructed with high explanatory power. Instead of imposing severe penalties on slope gradients steeper than 0° as predicted by physiological cost functions, the step-selection analysis reveals typical slope gradient selection as a shallower decline, with one road segment showing a preference for steeper slope gradients. Ultimately, whereas universal physiological cost functions obscure the decision-making processes of past movement, empirically derived selection weights reveal the complex trade-offs between landscape constraints, technological capabilities, and socio-economic priorities—providing a new framework for investigating the socio-political complexity of past societies.

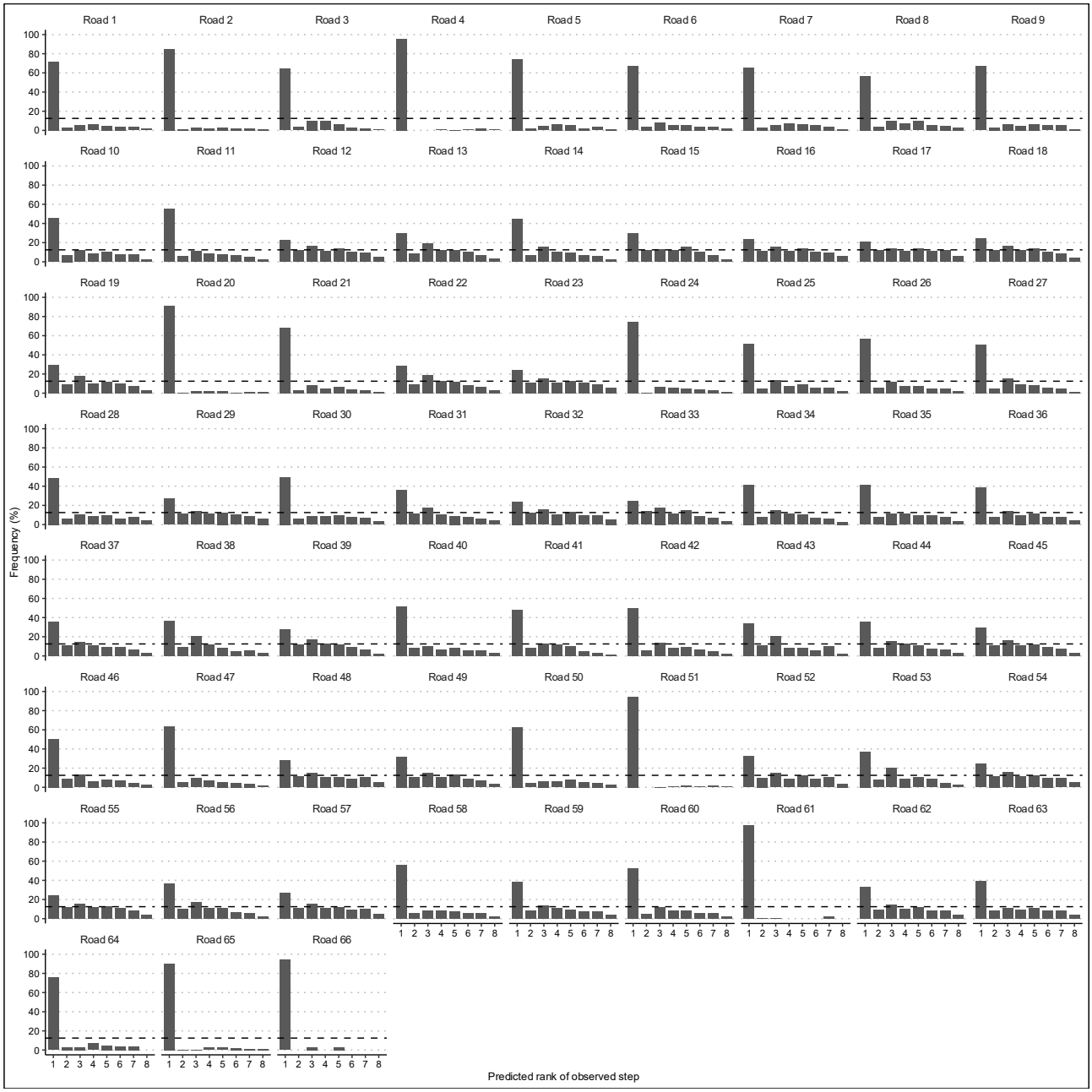
Funding declaration

No funding was received to assist with the preparation of this manuscript.

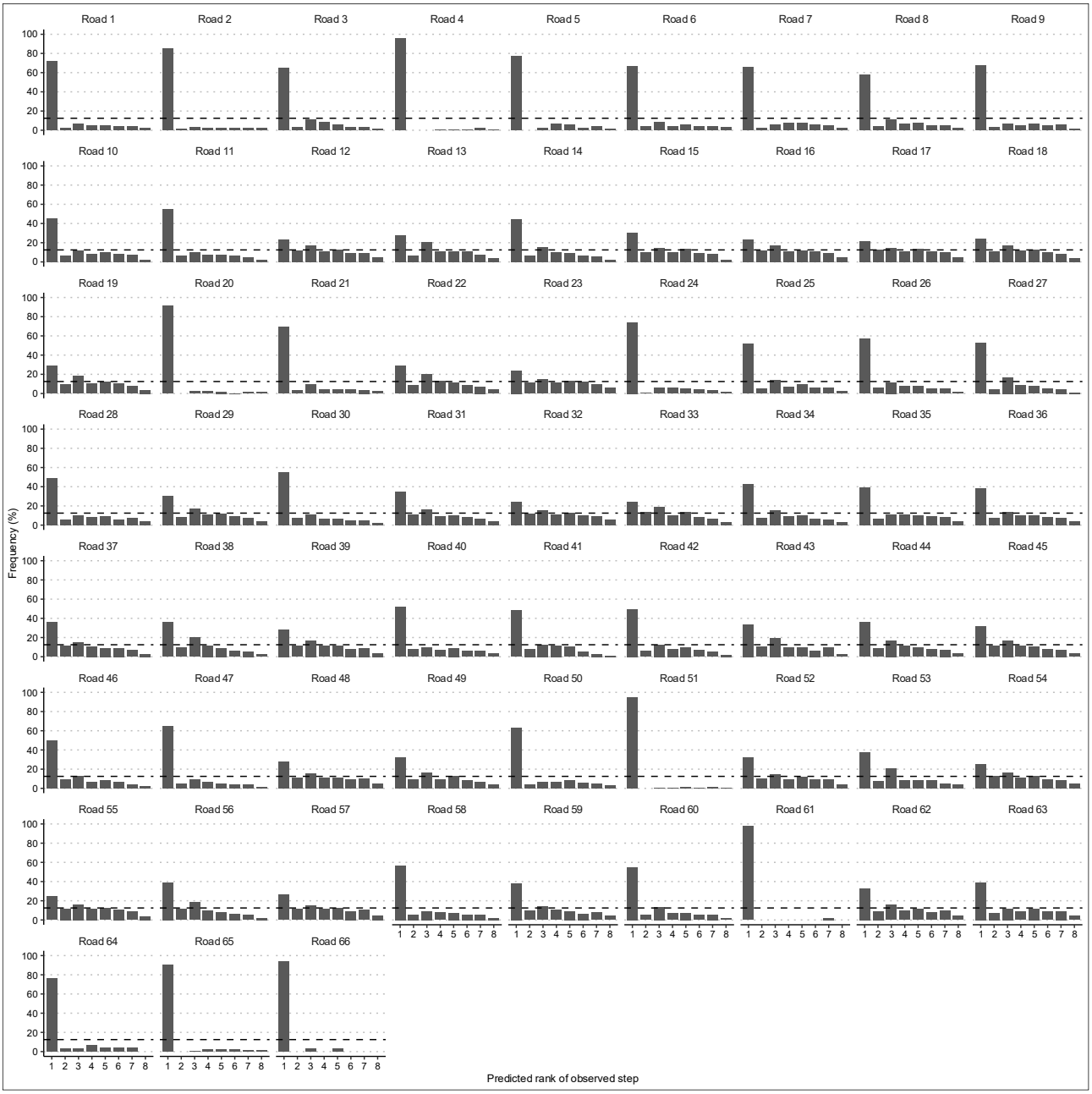
Acknowledgements

I would like to thank Dr. Enrico Crema for his supervision during my PhD, from which this work was developed. I am also grateful to the attendees of CAA-UK 2025 at the University of Cambridge for their valuable feedback on an early version of this paper.

Supplementary Figures



Supplementary Figure 1. Predictive accuracy of the empirically derived selection weight function across individual Roman roads. Distributions of predicted ranks for observed Roman road steps obtained from the posterior predictive ranking procedure using the RW2 selection weight function including both fixed and road-specific random effects. Ranks are determined by the predicted relative probability of selection within each step stratum, where Rank 1 represents the highest predicted probability of selection among the eight available alternatives. The horizontal dashed line indicates the random chance baseline (12.5%)



Supplementary Figure 2. External predictive accuracy of the empirically derived selection weight function under leave-one-out cross-validation. Distributions of predicted ranks for observed Roman road steps obtained from the posterior predictive ranking procedure when each road segment is sequentially excluded from the model fitting process. Panels show results for each of the 66 Roman road segments, where predictions are generated using selection weights estimated from the remaining roads. Ranks are determined by the predicted relative probability of selection within each step stratum, where Rank 1 represents the highest predicted probability of selection among the eight available alternatives. The horizontal dashed line indicates the random chance baseline (12.5%)

References

Atzori, S. (2006). *La strada romana "a Karalibus Sulcos."* Mogoro, Sardinia: PTM.

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*(6), 760–766. https://doi.org/10.1111/2041-210X.13168

Barreca, F. (1974). *La Sardegna Fenicia E Punica*. Sassari, Sardinia: Chiarella.

Brughmans, T., de Soto, P., Pažout, A., & Bjerregaard Vahlstrup, P. (2024). Itiner-e: the digital atlas of ancient roads. https://itiner-e.org.

Campbell, M. J., Dennison, P. E., Butler, B. W., & Page, W. G. (2019). Using crowdsourced fitness tracker data to model the relationship between slope and travel rates. *Applied Geography*, *106*, 93–107. https://doi.org/10.1016/j.apgeog.2019.03.008

Chevallier, R. (1976). *Roman roads*. London: Batsford.

De Soto, P., Pažout, A., Brughmans, T., Vahlstrup, P. B., Auir, Á., Bongers, T., et al. (2025). Itiner-e: A high-resolution dataset of roads of the Roman Empire. *Scientific Data*, *12*(1), 1731. https://doi.org/10.1038/s41597-025-06140-z

Dommelen, P. A. R. van. (1998). *On colonial grounds: a comparative study of colonialism and rural settlement in first millennium BC west central Sardinia*. Leiden: Faculty of Archaeology, University of Leiden.

Earle, T. (1991). Paths and roads in evolutionary perspective. In C. D. Trombold (Ed.), *Ancient road networks and settlement hierarchies in the New World* (pp. 10–16). Cambridge, England: Cambridge University Press.

Gorenflo, L. J., & Bell, T. L. (1991). Network analysis and the study of past regional organisation. In C. D. Trombold (Ed.), *Ancient road networks and settlement hierarchies in the New World* (pp. 80–98). Cambridge, England: Cambridge University Press.

Hengl, T., Leal Parente, L., Krizan, J., & Bonannella, C. (2020, September 28). Continental Europe Digital Terrain Model at 30 m resolution based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and background layers. Zenodo. https://doi.org/10.5281/ZENODO.4724549

Herzog, I. (2013). Theory and practice of cost function. In *Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain,*

April 2010, British Archaeological Reports International Series 2494 (pp. 375–382). Oxford, United Kingdom: Archaeopress.

Herzog, I. (2014). Least-cost Paths – Some Methodological Issues. *Internet Archaeology*, (36). https://doi.org/10.11141/ia.36.5

Herzog, I. (2020). Spatial Analysis Based On Cost Functions. In M. Gillings, P. Hacıgüzeller, & G. Lock (Eds.), *Archaeological Spatial Analysis* (1st ed., pp. 333–358). Routledge. https://doi.org/10.4324/9781351243858-18

Herzog, I. (2022). Issues in Replication and Stability of Least-cost Path Calculations. *Studies in Digital Heritage*, *5*(2), 131–155. https://doi.org/10.14434/sdh.v5i2.33796

Lewis, J. (2021). Probabilistic Modelling for Incorporating Uncertainty in Least Cost Path Results: a Postdictive Roman Road Case Study. *Journal of Archaeological Method and Theory*, *28*(3), 911–924. https://doi.org/10.1007/s10816-021-09522-w

Lewis, J. (2024). Explaining Known Past Routes, Underdetermination, and the Use of Multiple Cost Functions. *Journal of Archaeological Method and Theory*, *31*(3), 854–874. https://doi.org/10.1007/s10816-023-09621-w

Lewis, J. (2025). Sardinia and the Roman Road System Accessibility before and after the 'Augustan Revolution'. In S. España-Chamorro & M. del C. M. Escobar (Eds.), *Landscapes and the Augustan Revolution: The Transformation of the Western Provinces between the Republic and the Early Empire* (1st ed.). London: Routledge. https://doi.org/10.4324/9781003356288

Mastino, A. (2005). *Storia della Sardegna antica*. Nuoro, Italy: Il maestrale.

Mastino, A., & Vismara, C. (1994). *Turris Libisonis*. Sassari, Sardinia: C. Delfino.

Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, *89*(1), 80–92. https://doi.org/10.1111/1365-2656.13087

Parcero-Oubina, C., Smart, C., & Fonte, J. (2023). Remote Sensing and GIS Modelling of Roman Roads in South West Britain. *Journal of Computer Applications in Archaeology*, *6*(1), 62–78. https://doi.org/10.5334/jcaa.109

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, *32*(1). https://doi.org/10.1214/16-STS576

Tetti, V. (1985). Antiche vie romane della Sardegna e cursus publicus: note e riferimenti toponomastico. In *Archivio Storico Sardo di Sassari* (pp. 79–119).

Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, *2*(1), 4. https://doi.org/10.1186/2051-3933-2-4

Tobler, W. (1993). *Three presentations on geographical analysis and modeling* (No. Technical Report 93–1). Santa Barbara,CA: National Center for Geographic Information and Analysis. https://escholarship.org/uc/item/05r820mz. Accessed 17 February 2026

Trombold, C. D. (Ed.). (1991). *Ancient road networks and settlement hierarchies in the New World*. Cambridge, England: Cambridge University Press.

Tuppi, J. (2014). Approaching Road-Cuttings as Instruments of Early Urbanization in Central Tyrrhenian Italy. *Papers of the British School at Rome*, *82*, 41–72. https://doi.org/10.1017/S006824621400004X

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Mach. Learn. Res.*, *11*, 3571–3594.