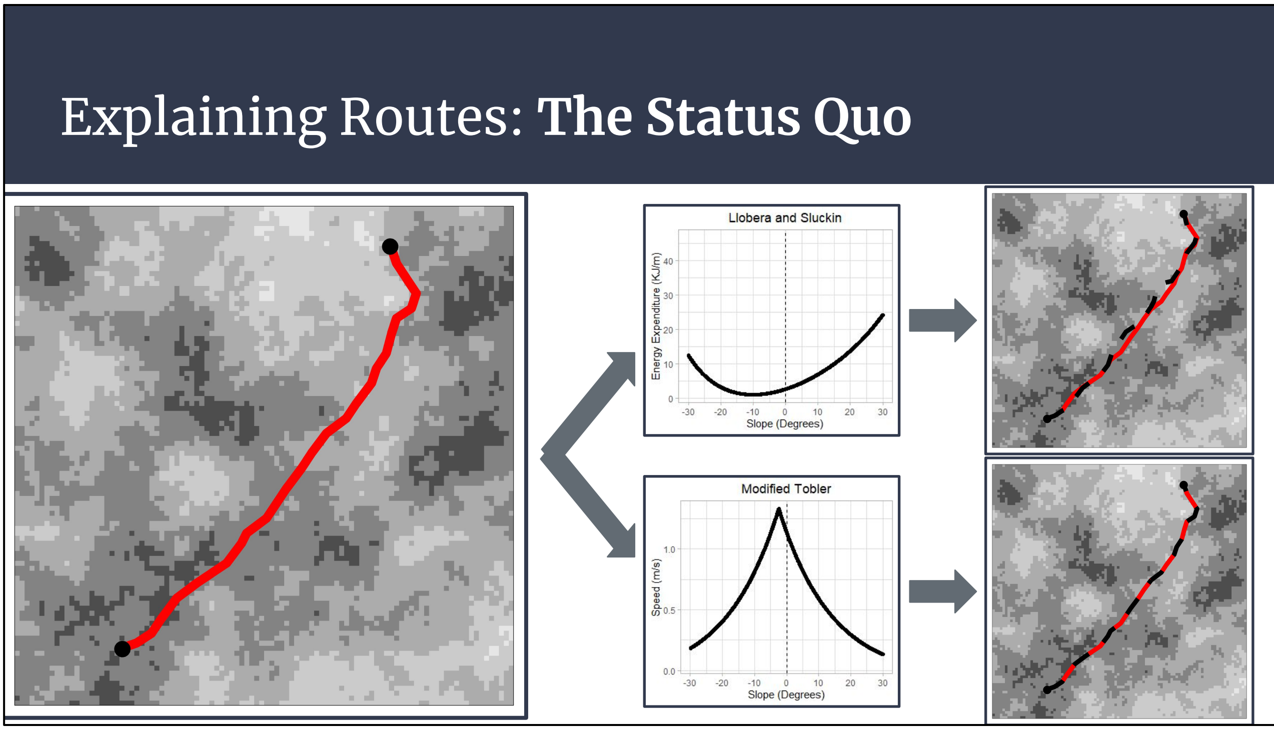
Towards a Coherent Framework for Explaining Known Historical Routes

Joseph Lewis
@josephlewis1992

Hi, my name is Joseph Lewis and today I'm going to be introducing a new framework for how we model and explain routes in the past. I'll be using Roman roads in Britain as the case study but the approach is applicable to all known routes regardless of their spatial or temporal positioning.

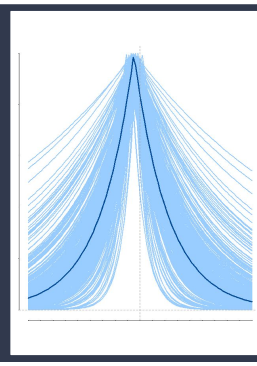


First to set the scene. Here is what I believe to be the status quo when modelling and explaining known routes. In this approach, we have a known route in red that we want to explain. We hypothesise that this route was the product of humans minimising some cost. In the case of movement through a landscape, this is normally the cost in terms of time taken or energy expended. Using a cost function that numerically expresses the relationship between slope and time taken or energy expended, we can calculate the path in black that minimises this cost. For example, the top right is the minimisation of energy expended based on the cost function produced by Llobera and Sluckin whilst the bottom right is the minimisation of time taken. Having calculated these paths, we can compare them against the known route. If there is similarity in their alignment, we can suggest that the proposed hypothesis - for example in this case time taken - is able to sufficiently explain the route at hand, and thus the hypothesis of minimising time is not rejected.

And whilst this approach is fine, I believe we can do much more when using models to explain known routes

Explaining Routes: **Current Issues**

Normative vs. Descriptive function?



Which slope-based cost function?

Incorporating factors other than slope?

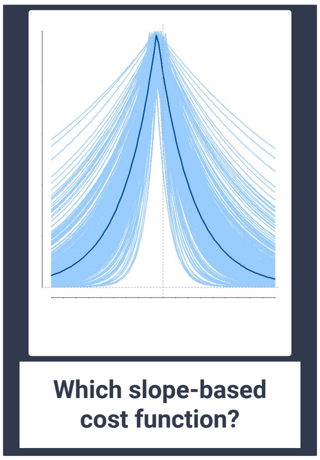
So why is a new framework needed? I've identified three issues that I believe need to addressed in order to more effectively model and explain known routes. The first is what function does the model play? Is it normative whereby we model what people ought to do based on some general law, or descriptive where we aim to model what people actually did? The second is the issue with choosing a cost function when modelling routes. Which one do we choose? Should we even aim to select one? And lastly, how we incorporate additional factors other than slope into the model?

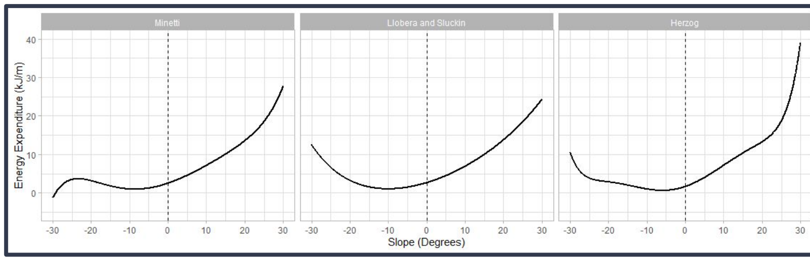
The Role of Models

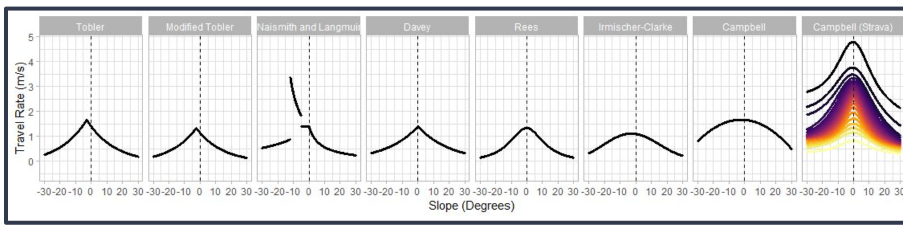
Normative vs.
Descriptive function?

First is what role should the model play when aiming to explain routes? Here we have a university campus. Within this campus there are paths cutting across the field. However, these paths were not laid out based on how people ought to move across the field, but rather reflecting how people actually moved across the field. And this is how we need to need to rethink our modeling when aiming to explain routes. We are not interested in how people ought to move, but how they actually moved. Therefore the models we create to explain routes need to be descriptive, culturally dependent, and reflect the historical, political, and social dynamics in which the routes were created

Rethinking Cost Functions





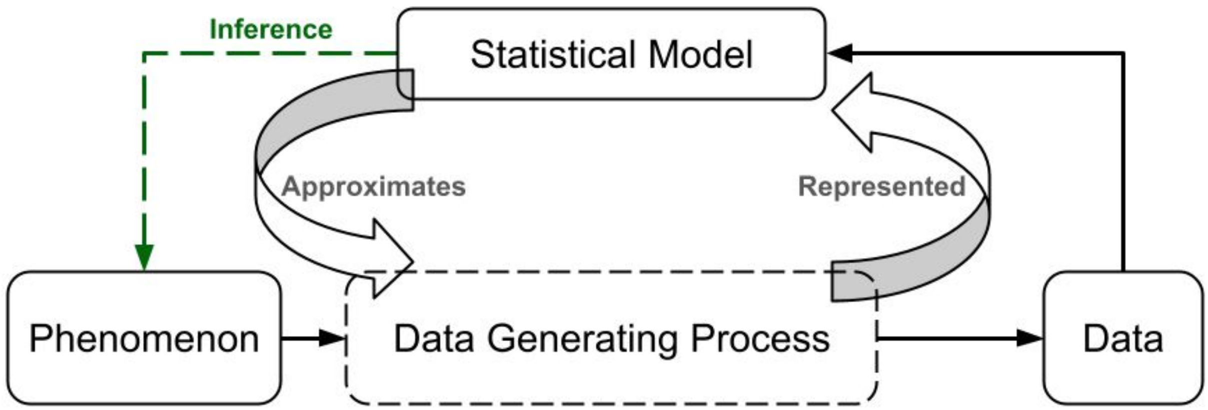


Second, is how we think about cost functions. Here we have a range of cost functions that numerically express the relationship between slope and energy expended or travel rate. When choosing a cost function to model routes, it's common for defaults to prevail: for example when modelling routes under the hypothesis of minimising time Tobler's Hiking function is commonly used..

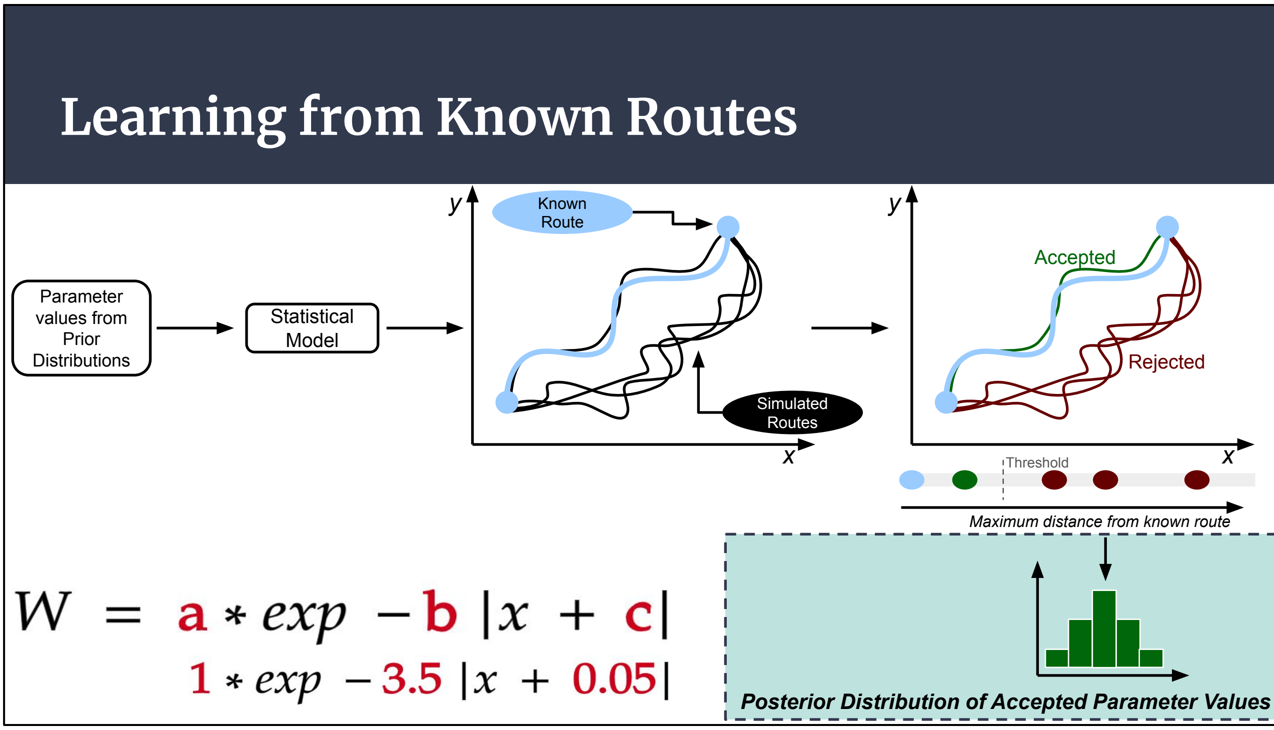
When we plot these time-based cost functions - most based on small sample sizes with little variability in participants- we can see that they're quite similar. And then there's the strava cost function in the bottom right. Based on a large number of data points from people of different sexes, ages, and fitness levels, this cost function shows the variability that can be expected in the relationship between slope and time. And if we see this level of variability in the relationship, should we even be thinking about which cost function to choose? All cost functions are approximations with their own biases. So, I argue that when trying to explain routes, we should learn the shape of the cost function from the known route, with the only decision we need to make being whether the route was generated under the hypothesis of minimising energy expended or time

And lastly, how do we incorporate additional factors other than slope? So we know that factors other than slope influences movement, but we lack a systematic approach for their incorporation within the model that we create. For example, when incorporating factors such as snow, marshlands, and rivers do we use presence/absence, snow/no snow or do we include it as continuous variables,height of snow? At what point does the factor become an issue? Is the effect of the factor on movement linear or exponential? In light of time I won’t have time to discuss this but I believe we can use Multicriteria decision analysis to overcome these issues

The Need for Generative Modelling



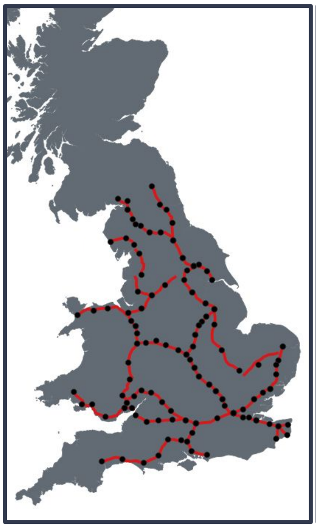
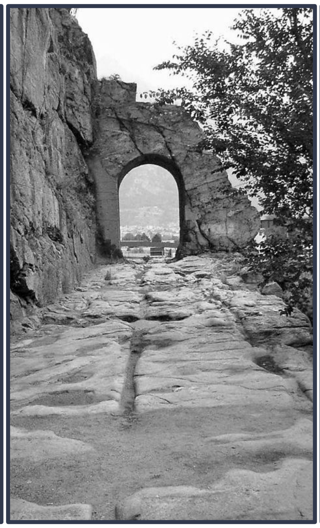
Now that I've outlined three issues, how do we overcome them? For this, I think we need to first reconceptualise how we understand the relationship between modelling and explanation. When aiming to explain phenomena, we need to think about the data generating process - that is, the process that resulted in the phenomena that we see in the world. We can never access this data generating process nor the phenomenon directly but what we can do is collect data on this phenomenon. Using a statistical model that approximates the data generating process we can work from the data that we collect back to the phenomenon. If the statistical model is able to sufficiently explain the data at hand, we can infer that the statistical model approximates the data generating process that created the phenomenon.

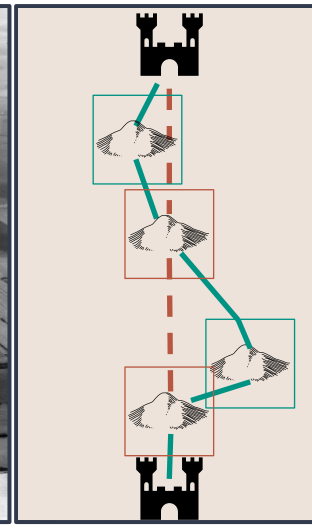


How do we operationalise generative modelling when explaining routes? For this we can use Approximate Bayesian Computation, or ABC for short. Within the ABC framework, we have a statistical model with some parameters. For example, the double exponential function used for Tobler's Hiking function has three parameters, a,b,c. But unlike Tobler's Hiking function we're unsure what these values are. So we set what is called a prior distribution on each of these parameters. These are probability distributions reflecting our knowledge of the what parameter values could be prior to seeing the data

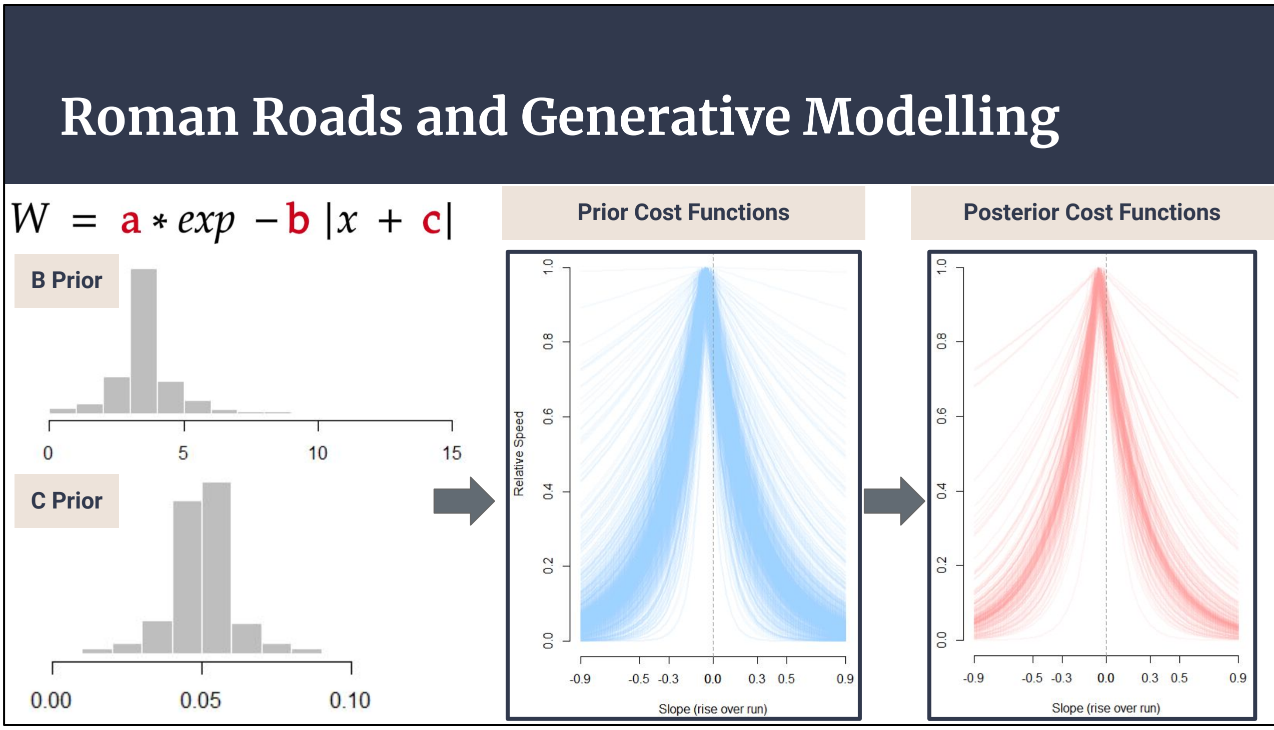
By inputting these parameter values into the statistical model we produce simulated data. We then repeat this process with each simulations based on different parameter values. These simulations are then compared against the known route that we're trying to model. If the distance is below a threshold value, we say that the simulated data is close enough to be deemed equal to the known route. We then take those accepted parameter values to get to what we're after: THE POSTERIOR DISTRIBUTION of what the parameter values are more likely to be given the data that we're modelling.

Roman Roads and Generative Modelling



Now for its application. For this, I'll be aiming to explain all Roman roads in Britain that were recorded within the Antonine Itinerary, a late third century document of routes across the Roman Empire. When thinking about the data generating process of Roman roads I have identified three factors of relevance. First I've included the effect of slope on time. This is to reflect that Roman roads and their positioning within the landscape could be impacted by slope. For example, based on a systematic analysis of Roman roads in Roman Britain, it appears that the target for the maximum gradient was around 12.5%. Second, I've included straightness. Again this is to reflect that where possible Roman roads were constructed to be straight. The third factor however introduces nuance to this straightness. Rather than a single straight line between termini, it is more common that roads were constructed from sighting points taking into account any obstacles between two locations. For this, I've included hills within the model.

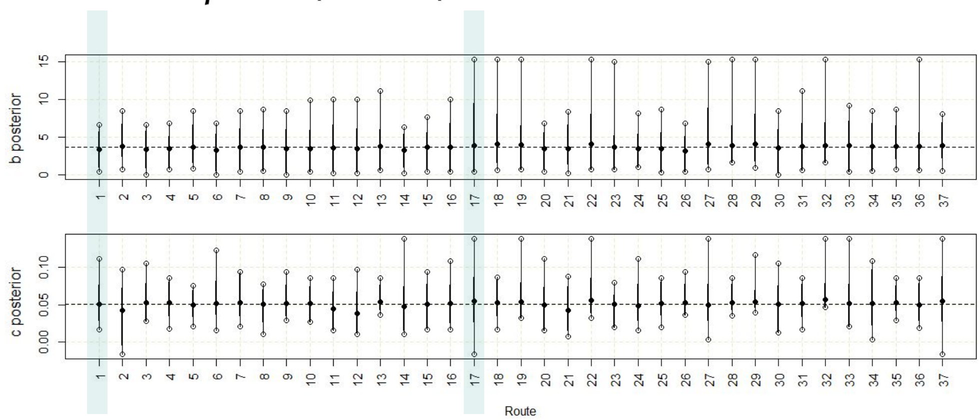


Here's an example of how we can learn the most likely parameter values from the known routes. Using the double exponential again I've set prior distributions on b and c. From values drawn from these two distributions, we can simulate what we would expect the cost function to be prior to seeing the data. We can see that although most of the simulated cost functions share a similar shape the prior distributions of b and c are wide enough to allow variability within the simulated cost functions, for example by allowing increasing slope to have less of an impact on speed as shown by the more horizontal blue lines

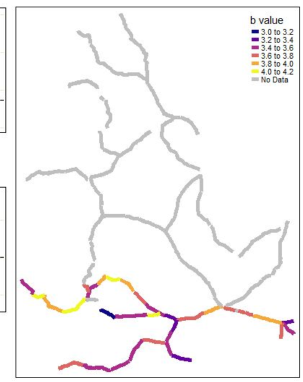
After accepting simulations that are within a threshold from the known route shown with the red cost functions, we get a better understanding of the shape of the most likely cost function given the route that we're modelling. And because we're modelling each route separately each cost function that best explains a particular route can vary.

Modelled Roman Roads

$$W = \mathbf{a} * exp\ -\mathbf{b}\ |x + \mathbf{c}|$$



b value



And these are how they can vary between a selection of modelled roads. For this I've taken the closest 10% of simulated routes for each road. Showing b and c again, we can see that whilst most roads have a similar mean shown with the black point, the range of parameter values for each road varies. For example, the parameter b for road 1 is much tighter than road 17. Therefore this suggests that road 17 is not as sensitive to changes in parameter b, with a wider range of values able to explain the road.

Nonetheless, through this ABC framework approach, we can (1) quantify the level of uncertainty around parameter values, (2) assess how uncertainty differs across modelled roads; (3) identify when a particular generative model is unable to explain roads, and suggesting that the model is missing important information, and (4) we can plot these parameter values and start to explore their spatial distribution. From this, we can develop a more nuanced understanding of the road network within Roman Britain, how it integrated itself within pre-existing social and political structures, as well as uncovering the heterogeneity in how Romans dealt with situations when constructing their roads

Next Steps for my Research

- **Iteratively modify and assess outputs against known Roman roads**
- **Explore posterior parameters and their spatial relationships**
- **Explain posterior parameter values by proposing hypotheses:**
  - **e.g. only topography? time of road construction? sites connected?**

As my PhD is still ongoing, the next steps for my research will involve:

(1) First is to Iteratively modify the model and assess outputs against the Roman roads. This will aim to create a better generative model that can sufficiently explain more of the Roman roads

(2) With these parameter values, I can start to explore their spatial relationships. For example, do we see similar parameter values in certain areas?

(3) And three, propose hypotheses for explaining these parameter values. For example, does the time of construction influence the maximum slope traversable? Are we able to see changes in Roman policy as the conquest of Britain develops? Can we identify how the Romans incorporated themselves within pre-existing structures through their roads? And with this, we can fundamentally change our understanding of the role Roman roads played within the developing province

Thank you
Any questions?

Joseph Lewis
@josephlewis1992

CAMBRIDGE TRUST | Open-Oxford-Cambridge | UNIVERSITY OF CAMBRIDGE | CAA2022 OXFORD iNSIDE FORMATION