# A network for simulating pre-colonial migration in the Americas

**Gereon A. Kaiping**
Department of Geography, Universität Zürich, Switzerland
gereon.kaiping@geo.uzh.ch

## Abstract

Because history is inaccessible to experimentation, agent-based and other simulations are a main source to explore theories about pre-historical humanity. Continent-scale migrations are of great interest in this context. With advances in computing and GIS, tracking entire populations migrating across continents become accessible in simulation. In this paper, I present a network representing North and South America for such tasks. The nodes roughly follow a hexagonal grid and represent small territories around a focal point. They are annotated with the carrying capacity for hunter-gatherers per ecoregion in the vicinity. The edge weights represent the travel times between the focal points on foot or by boat. I validate the network by comparing its predicted optimal path between Nashville, TN and Natchez, MI with the route of the historical Natchez Trace.

**Funding** Swiss National Science Foundation (SNSF) Sinergia Project "Out of Asia" CRSII5_183578; NCCR Evolving Language, SNSF Agreement #51NF40_180888

## 1 Introduction

Agent-Based Models (ABM) can be a useful tool for the investigation of complex systems not otherwise accessible to the researcher. ABMs have been fruitfully applied to study the evolution of cooperation and culture. In the wider historical sciences, ABMs can help archaeologists and paleontologists interpret scarce data [2, 4]. However, most of these models are very abstract in either the interactions between the agents or their representation of geography. Easily available data structures, such as weighted graphs with metadata, could help advance computational models towards complex interactions in the context of real or at least realistic geography.

The human settlement history of the pre-historic Americas is unclear. While most evidence points to humans moving from East Beringia into the wider North America around 14'000 yBP, a small number of paleontologists highlight finds which suggest that humans may have lived in the Americas much earlier. Agent-Based Models with a detailed account of human migration could help to shed light on the human history of the Americas. In this short paper, I describe a network that can be used to model the dispersal of humans over North and South America. While the ultimate goal would be to provide a dataset that allows the modelling of ancient migrations, here I restrict myself to contemporary topography and ecology as a proof-of-concept that is at least partially accessible to evaluation.

## 2 Design Choices

The network I describe in this paper should be suitable for ABMs of human migration on the continental scale. For such models, long-term interactions between humans are an important aspect, and the network needs to provide an easy framework to facilitate these interactions. In a network structure, co-locality between agents is a useful condition for social interactions. Nodes will therefore represent an area of a radius where regular interactions between humans tend to happen. Following [10], an interaction radius of 5 km seems appropriate, so each node should represent a territory of extending ca. $r = 5$ km around a central location (thus having an area of roughly $\pi r^2 = 78$ km$^2$), which I shall refer to as 'core location' of the territory to distinguish from locations that are central given other measures.

It has been shown that language families spread easier east-west than north-south, due to climate varying much more with latitude than with longitude. Local biome can also play a role in the dispersal of a language family, as argued eg. by [7, 6], and Eco-Cultural Niche Modelling [1] has a variety of archaeological applications. This suggests that including a measure of ecology in addition to resource availability or population capacity will be necessary for generating realistic migration patterns. [15] argue that models should take into account not only the amount (e.g., in terms of caloric content) of resources, but also their accessibility. This suggests that the topography of resources around core locations is important.

I construct a cost function that describes the effort moving from one core location to another, in terms of raw energy expenditure and missed foraging opportunities due to time spent on the trail. For land-based travel, I estimate walking times using an empirical function that improves upon earlier heuristics such as Tobler's hiking function. [17] use very specific metabolic costs depending on load, body weight, and other properties of the walker, which are however narrowly clipped and proportional to the time travelled outside the central range of slopes. I therefore parameterize the migration network in terms of travel time, instead of giving metabolic costs, according to the empirical formulas derived from [9]. Similar to [11], I derive travel times along rivers from a constant canoeing speed plus or minus river flow speeds, and over sea using the same constant canoeing speed not adjusted for currents.

## 3 Method and Data

The migration network is implemented as an SQLite database, generated using Python. It includes travel times across terrain and over water (rivers, lakes, and sea). Terrestrial travel times are computed using an empirical formula depending on slope, which was estimated using navigation speeds of male US cadets on cross-country navigation tasks [9], applied to the 30 arc second GMTED2010 digital elevation model. I compute the geodesic distance between all pixels in an 8-neighbourhood (to N, NE, E, SE, S, SW, W, NW).

Together with the elevation difference from the DEM, this gives the slope that governs the travel speed, which is adjusted using terrain coefficients generalized from [12] (who collected data for Kenya) to all terrestrial ecoregions according to `https://ecoregions2017.appspot.com/`. In this way I derive travel times between all neighbouring pixels at a 30 arc-second resolution, which can be used as input for subsequent steps. Traversing rivers takes time. If a river segment crosses a pixel, I add a time penalty for leaving that pixel. Because the penalties from [11] refer to established travelling infrastructure, I instead estimate the possibility and time for wading using the river reaches taken from GloRiC 1.0 [5].

Travel by water is assumed to follow a basic canoeing speed. Modern paddlers seem to be able to maintain speeds of above 4 knots for longer times, but in general, enthusiast forums on the internet give numbers of 3 to 4 knots. The experiments cited in [8, 11] give lower speeds. Due to river navigation obstacles playing a role in wilderness navigation, but less so in leisure canoeing, I follow the latter and assume a basic long-term effective cruise speed of 3.0 km/h. For following coasts or short island hops, this velocity is taken unmodified up to a

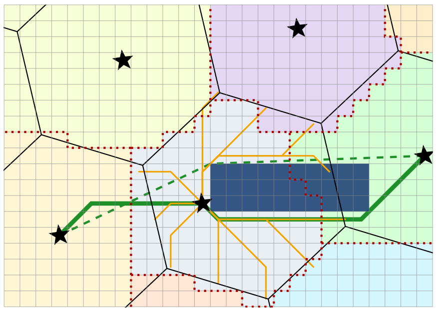

**Figure 1** Core locations (black stars) are generated according to the minimum sum of shortest paths from H3 hexagons boundary points, so obstacles (blue rectangle) can push the core locations away from the hexagon centres. Shortest paths from core locations then give the network edges and Voronoi cells (shaded, with red dotted border). As such, the optimal path in the network (green line) is only an approximation of the theoretical optimum (green dashed line).

maximum distance of 150 km (which corresponds to about 50h of gross travel time). Along rivers and streams, this is modified by the flow speed $v$ of the river, which is added when moving downriver and subtracted when moving upriver. Unfortunately, the mean speed of flow is available only for a minority of American rivers. A rough estimate can be derived following [14]:

$$W = 2.71 \cdot Q^{0.557} \qquad D = 0.349 \cdot Q^{0.341} \tag{1}$$

$$R = D \cdot W/(2D + W) \qquad v = \frac{1}{n} \cdot R^{2/3} \cdot S^{1/2} \tag{2}$$

where $Q$ is the discharge [m$^3$/s] and $S$ is the slope [1]. While they suggest that knowing the river roughness $n$ vastly improves the speed estimates, data is also lacking for those quantities, so I use their reported mean value of 0.044. Rivers are taken from GloRiC 1.0 [5]. While GloRiC does not include data on the slope directly, it includes the $\log_{10}$ of stream power $P$ [kg m/s$^2$], which is related to slope $S$ and discharge $Q$ by $P = \rho g Q S$ (with the density of water $\rho = 1000$ kg/m$^3$ and $g = 9.81$ kg m/s$^2$. Following [13, 18], I include all river reaches with a discharge of 8 m$^3$/s or more and a slope of 10% or less.

Together, this provides a theoretical basis to calculate the gross travel time between any two terrestrial or coastal 30" pixels on foot or by boat. Unfortunately, this calculation is computationally expensive, so I will instead derive a network of locations with surrounding territories and the distances between each of them and their nearest neighbours. In order to generate largely regular terrestrial territories with an area of about 100 km$^2$ each, I start with the H3 grid [3], a hexagonal, hierarchical global grid that is easily available as Python library. Core locations should however represent typical transit locations in the territory instead of the hexagon centre points, which may lie in impassable terrain. To mitigate this, I collect for each hexagon the corner points and the center of each edge. The core location for each grid cell is then the node (pixel) that minimizes the sum of travel times to and from each of those boundary points, and as such minimizes distortions when considered as proxy for the whole hexagon. I derive a 30" raster of ecoregions, and of population densities following [16], compatible with the DEM and the travel time rasters.

Territory and resources are then allocated to the nodes by calculating the generalized Voronoi decomposition of the 30" raster, according to the travel time (outbound plus return) from the core location. Reducing the granularity of the DEM raster to larger Voronoi cells permits, at the cost of some over-estimation of distances, the computation of migration routes

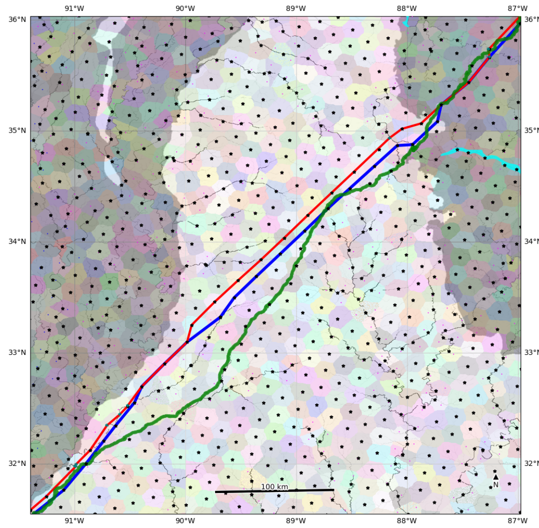


**Figure 2** Fastest path between the end points of the Natchez Trace (blue, red) compared to the modern-day Natchez Trace Parkway (green). The coloured background shows the Voronoi tessellation with stars at the core locations, the shading describes the terrain: Darker pixels, eg. rivers, are more difficult to cross.

on the continent scale while at the same time giving a convenient proxy for interactions between agents that share a cell. Figure 1 shows a simplified example of this process. In this manner, every pixel is associated to the core location from where it is most accessible. For each pixel, the population capacity (area represented by the pixel times population density according to the raster) is then allocated to the corresponding ecoregion in the node attributes corresponding to the core location. By tracking the distances computed from the core locations to the Voronoi boundary points and back individually, the Voronoi tessellation also gives the pairwise travel time for each pair of adjacent Voronoi cells, and thus the weights of the migration graph. However, in most short-distance cases the shortest path between the core locations of two Voronoi cells which are not directly adjacent will be shorter than the sum of the Delaunay edges. So, I compute the shortest walking paths not just between every neighbouring Voronoi cells, but to every node within a hexagonal radius of 2, and also to all end points of river reaches and coasts and all points where a river reach crosses a Voronoi cell boundary within that radius. This gives a more dense network which is better able to represent the topography.

The resulting network, as well as all the Python code to generate it and raster files containing the results of intermediate steps, is available from `https://doi.org/10.5281/zenodo.5167698`. The network contains 137 522 territorial nodes (each with population capacities per ecoregion) and 405 855 nodes representing water boundary points (on river reaches or the coastline, each without population capacities). The terrestrial edges between core points represent travel times between 2.4 h and 16.8 h (5% and 95% quantiles), with a mean of 10.9 h. Edges across the sea or large lakes have a broader range, from 1.7 h up to 50 h (the limit imposed by 150 km at 3 km/h). Due to the resolution of the river network, distances reflected by it are much shorter, with 95% of them shorter than 1.5 h. The average terrestrial area of a territory in the network is $104 \pm 34$ km$^2$, and thus appropriate for the desired resolution.

## 4 Evaluation

The Natchez Trace is a historic trail of 710 km from Nashville, Tennessee, to Natchez, Mississippi, going back thousands of years. It became obsolete (and is nowadays a US national parkway) when steamboats became an alternative for travelling upriver.

To test the generated network, I compute the path from the node at 86.979° W, 35.967° N, which contains the start of the trail near Nashville, to the node at 91.400° W, 31.579° N at the end of the trail near Natchez without the use of rivers, and in the reverse direction with the use of rivers. I compare these two optimal paths according to the network with the route represented by today's Natchez Trace Parkway.

The results are shown in Figure 2. The optimal path is the very similar in both directions, but the red path uses river nodes (teal '+') in a few isolated places, essentially crossing the rivers with less deviation. The paths deviate from the shortest hexagon path from Nashville to Natchez, but less than the actual Natchez Trace. This suggests that my implementation incorporates some, but not all environmental features that influence cross-country navigation. The best route is inferred with a travel time of around 1050 hours in both directions. Compared to an aerial distance of 643 km, this is a very low speed of 0.62 km/h. The maximum navigation speed is 2.8 km/h (0.78 m/s, [9]) and the area mostly contains forest biomes (brighter center area in the figure), where the terrain coefficient is 1, which means that the roughness of the terrain is a major factor slowing down the movement.

The inferred path follows the actual Natchez Trace Parkway about half of its length. A noticeable discrepancy is that the real Natchez Trace turns to the south following along the Yockanookany river and passes through the modern-day City of Jackson, the optimal paths in the network instead turns to the north and proceed on the other side of the Big Black River. The maximal distance between the inferred and the real path is about 60km. I speculate that this is not by chance: The location of the modern-day City of Jackson likely has properties (such as its location on the navigable Pearl River, or ecological factors making it suitable for settlement) that were favourable for both the near passing of the Trace and the foundation of the city, but which are not considered in this simple shortest-path analysis.

## 5 Discussion

The migration network presented here aggregates estimations and data about the natural environment, and human movement within it, from a variety of sources. The evaluation shows that terrain plays a major role in the selection of optimal paths in the absence of established routes, and that it can potentially reflect the route choices of pre-historic Americans. This network is created to be used as the underlying structure for an agent-based model investigating the settlement and language diversity in the Americas. Beyond this and immediate application and other agent-based model studies, the underlying data could also be used for investigations similar in archaeology and human history: [11] describes the difficulty of aggregating data and using general-purpose software to estimate pre-historic pathways. Hopefully this paper, together with the accompanying Python code and raster data, can contribute to filling this gap.

## References

1 William E. Banks, F. D'Errico, Harold L. Dibble, Leonard Krishtalka, Dixie West, Deborah Olszewski, A. Townsend Peterson, G. Anderson David, G. Christopher Gillam, Anta Montet-White, Michel Crucifix, Curtis W. Marean, María-Fernanda Sánchez-Goñi, Barbara Wohlfarthh,

and Marian Vanhaeren. Eco-Cultural Niche Modeling: New Tools for Reconstructing the Geography and Ecology of Past Human Populations. *PaleoAnthropology*, 2006:68–83, 2006.

2 Juan A. Barceló and Florencia del Castillo. *Simulating Prehistoric and Ancient Worlds*. Springer, New York, NY, October 2016.

3 Isaac Brodsky. H3: Uber's Hexagonal Hierarchical Spatial Index. https://eng.uber.com/h3/, June 2018.

4 Wendy H. Cegielski and J. Daniel Rogers. Rethinking the role of Agent-Based Modeling in archaeology. *Journal of Anthropological Archaeology*, 41:283–298, March 2016. `doi:10.1016/j.jaa.2016.01.009`.

5 Camille Ouellet Dallaire, Bernhard Lehner, Roger Sayre, and Michele Thieme. A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environmental Research Letters*, 14(2):024003, January 2019. `doi:10.1088/1748-9326/aad8e9`.

6 Christopher Ehret. Bantu history: Big advance, although with a chronological contradiction. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44):13428–13429, November 2015. `doi:10.1073/pnas.1517381112`.

7 Rebecca Grollemund, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti, and Mark Pagel. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 112(43):13296–13301, October 2015. `doi:10.1073/pnas.1503793112`.

8 Leslie C. Hazell. An Analysis of Log Raft Open Water Performance and Crew Capability to Move Megaliths Pre-classic Olmec Used for Colossal Head Sculptures. *Journal of Maritime Archaeology*, 8(1):139–152, June 2013. `doi:10.1007/s11457-013-9109-1`.

9 Ian J. Irmischer and Keith C. Clarke. Measuring and modeling the speed of human navigation. *Cartography and Geographic Information Science*, 45(2):177–186, March 2018. `doi:10.1080/15230406.2017.1292150`.

10 Robert L. Kelly. *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*. Cambridge University Press, Cambridge, 2013. `doi:10.1017/CBO9781139176132`.

11 Patrick Livingood. No Crows Made Mounds: Do Cost-Distance Calculations of Travel Time Improve Our Understanding of Southern Appalachian Polity Size? In Sarah L. Surface-Evans and Devin Alan White, editors, *Least Cost Analysis of Social Landscapes: Archaeological Case Studies*. University of Utah Press, Salt Lake City, 2012.

12 Francesca Pozzi and Timothy P. Robinson. Accessibility mapping in the Horn of Africa: Applications for livestock policy. 2008.

13 Stewart B. Rood, Chris George, Bobbi George, and Wilco Tymensen. Instream flows for recreation are closely correlated with mean discharge for rivers of western North America. *River Research and Applications*, 22(1):91–108, 2006. `doi:10.1002/rra.888`.

14 K. Schulze, M. Hunger, and P. Döll. Simulating river flow velocity on global scale. *Advances in Geosciences*, 5:133–136, December 2005. `doi:10.5194/adgeo-5-133-2005`.

15 James Steele and Marcy Rockman. "Where do we go from here?" – Modelling the decision-making process during exploratory dispersal. In Marcy Rockman and James Steele, editors, *The Colonization of Unfamiliar Landscapes: The Archaeology of Adaptation*, page 15. Routledge, 2003. `doi:10.4324/9780203422908`.

16 Miikka Tallavaara, Miska Luoto, Natalia Korhonen, Heikki Järvinen, and Heikki Seppä. Human population dynamics in Europe over the Last Glacial Maximum. *Proceedings of the National Academy of Sciences*, 112(27):8232–8237, July 2015. `doi:10.1073/pnas.1503784112`.

17 Brian M. Wood and Zoë J. Wood. Energetically optimal travel across terrain: Visualizations and a new metric of geographic distance with anthropological applications. In *SPIE Proceedings: San Jose, CA*, volume 6060, page 60600F. International Society for Optics and Photonics, January 2006. `doi:10.1117/12.644376`.

18 Peggy Zinke, Dag Sandvik, Ingrid Nesheim, and Isabel Seifert-Dähnn. Comparing Three Approaches to Estimating Optimum White Water Kayak Flows in Western Norway. *Water*, 10:1761, November 2018. `doi:10.3390/w10121761`.