Generalized additive mixed models for spatial networks

Nicolas Gauthier

School of Human Evolution and Social Change, Arizona State University

What are GAMMs?

Generalized additive models (GAMs) and their variants Generalized additive *mixed* models (GAMMs) are a flexible family of regression model that can capture the complexity of the archaeological record.

Let's break it down:

- *Generalized* – Model non-normal distributions such as counts or proportions
- *Additive* – Combine linear and non-linear relationships
- *Mixed* – Control for network, spatial, and temporal autocorrelation

What can I do with them?

In archaeology, they're particularly useful for fitting so-called "maximum-entropy" spatial interaction models. These models estimate the *flow* of goods, information, or people between spatially-structured populations as a function of the origin site, destination site, and the space between them:
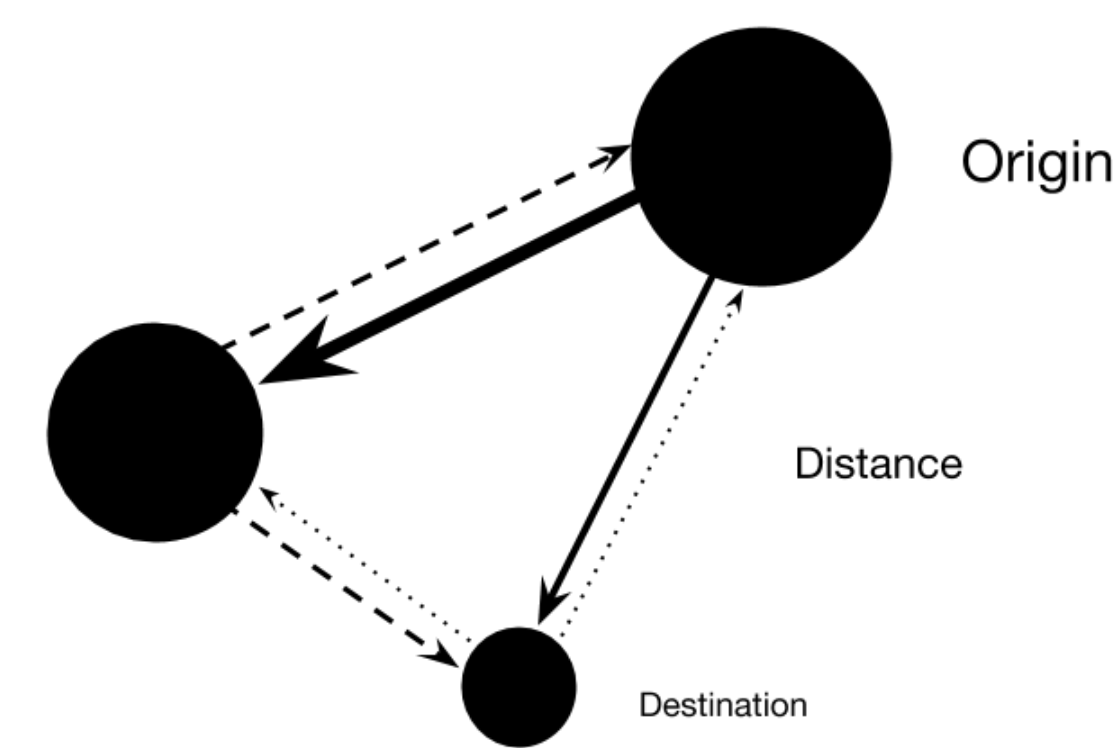
$$flow = f(origin) \times f(destination) \times f(distance)$$

The difficult part comes when we have to define the $f()$s. Economic geographers often use a generalized *linear* model (GLM), which requires them to definte the $f()$s ahead of time. In a GAM, the $f()$s are estimated directly from the data using splines.

How do they work?
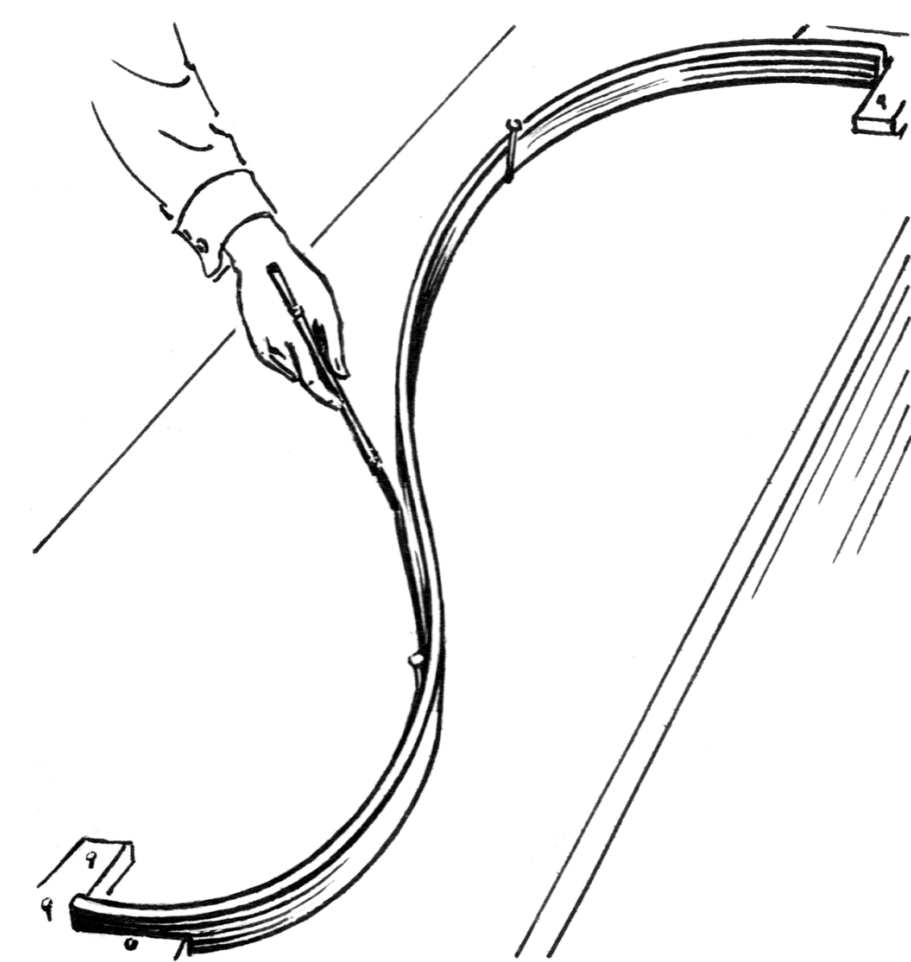
Real-world splines are flexible strips of metal or wood used to draw curves. Mathematical splines are complex curves made of many smaller, simpler curves. *Penalized* regression splines can estimate $f()$ from the data, limiting overfitting by penalizing the "wiggliness" of the function.

Case Studies

Ceramic distribution in Roman Britain

Percent late Romano-British pottery produced in Oxford and New Forest

Pottery from origin site
- 5%
- 10%
- 15%
- 20%

100 km
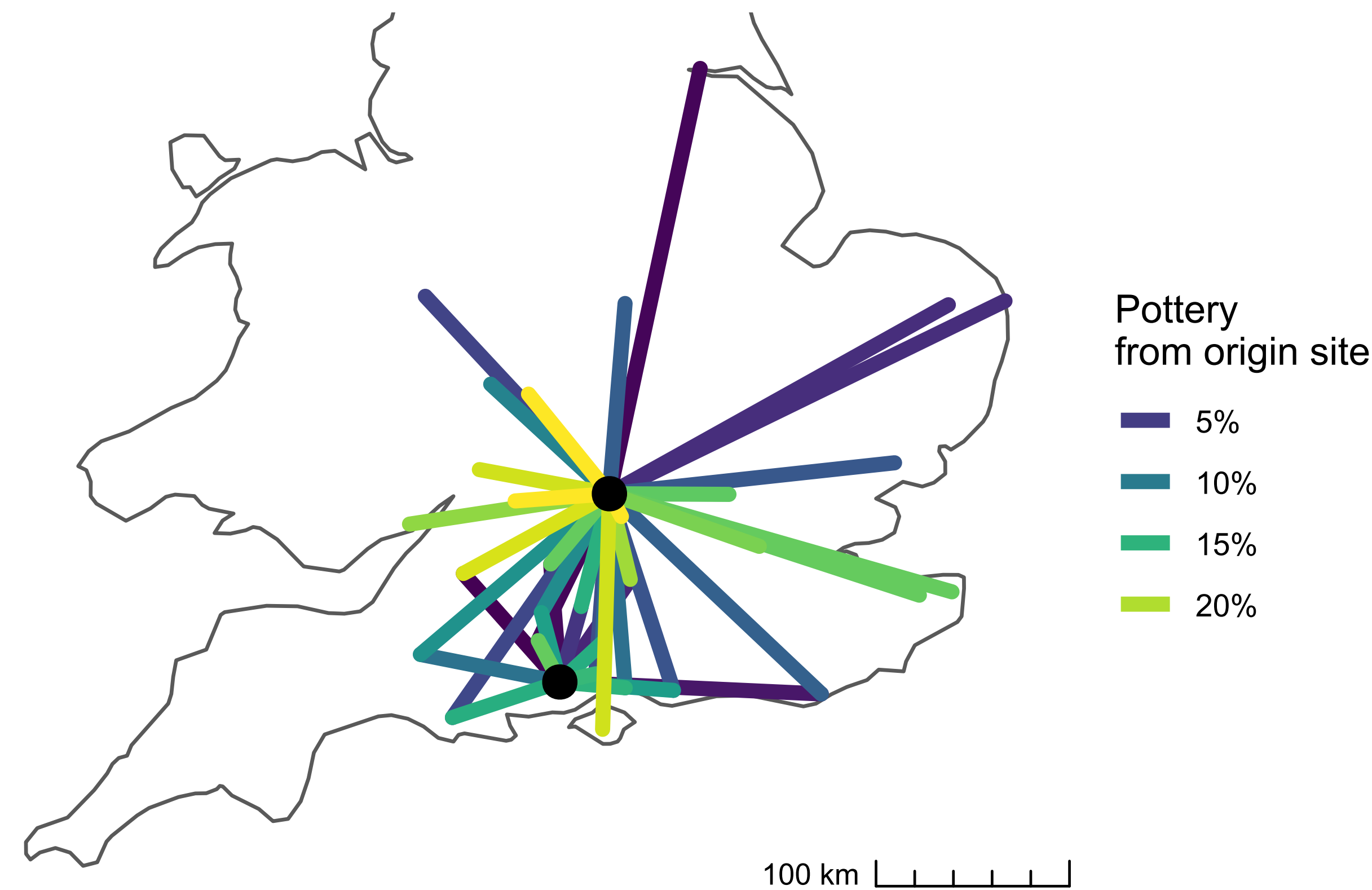
The Oxford Pots dataset can be found in the R package `archdata`. *David L. Carlson and Georg Roth (2018). archdata: Example Datasets from Archaeological Research. R package version 1.2.*

Coastal Chumash marriages in southern California

Marriages in mission registers, population estimated from baptisms

Baptisms ● 100 ● 200 ● 300   Marriages 0 10 20

30 km
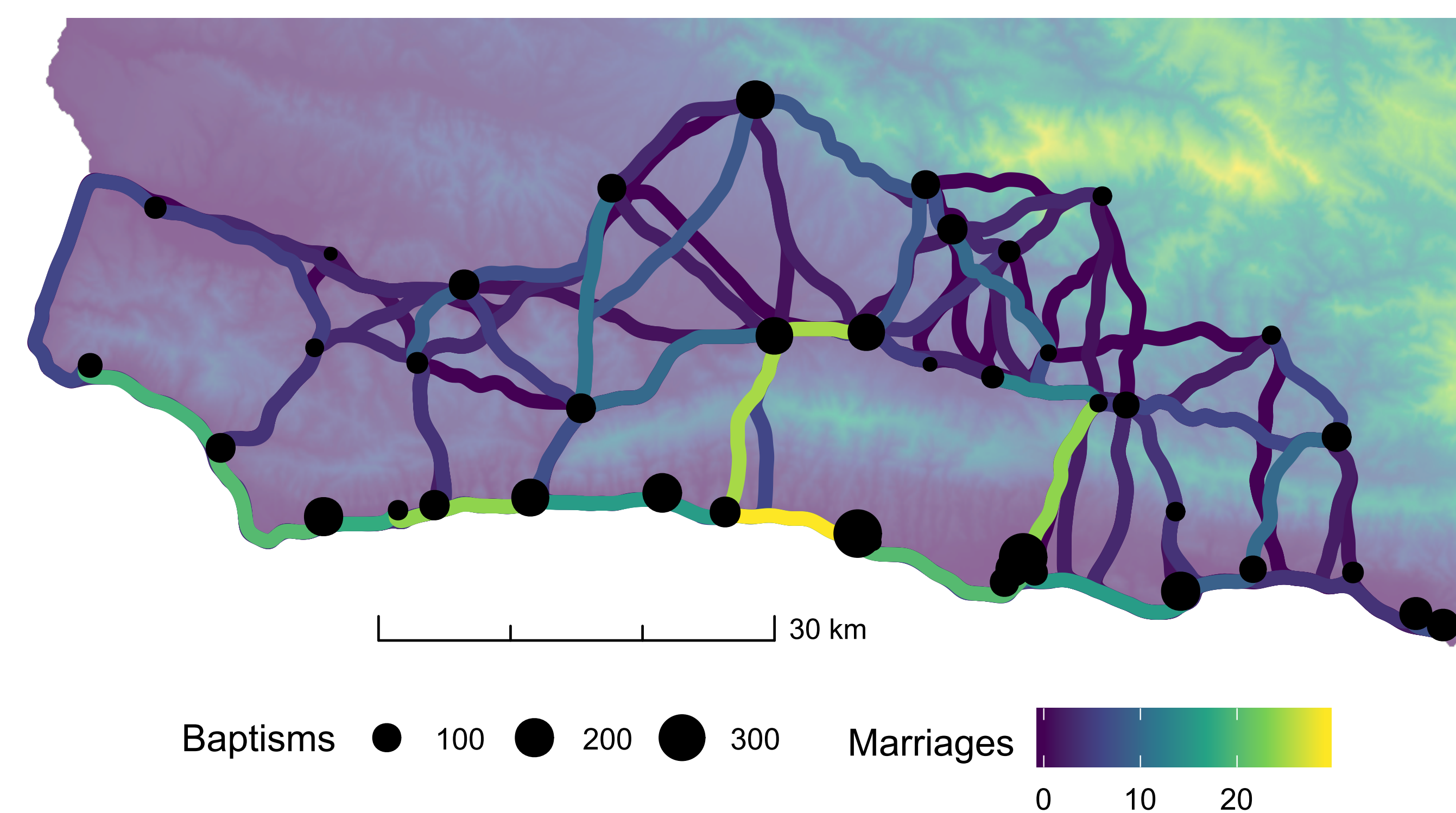
Chumash marriage data can be found in John Johnson's PhD thesis. *John R. Johnson, "Chumash social organization: An ethnohistorical perspective". PhD thesis. UC Santa Barbara, 1988.*
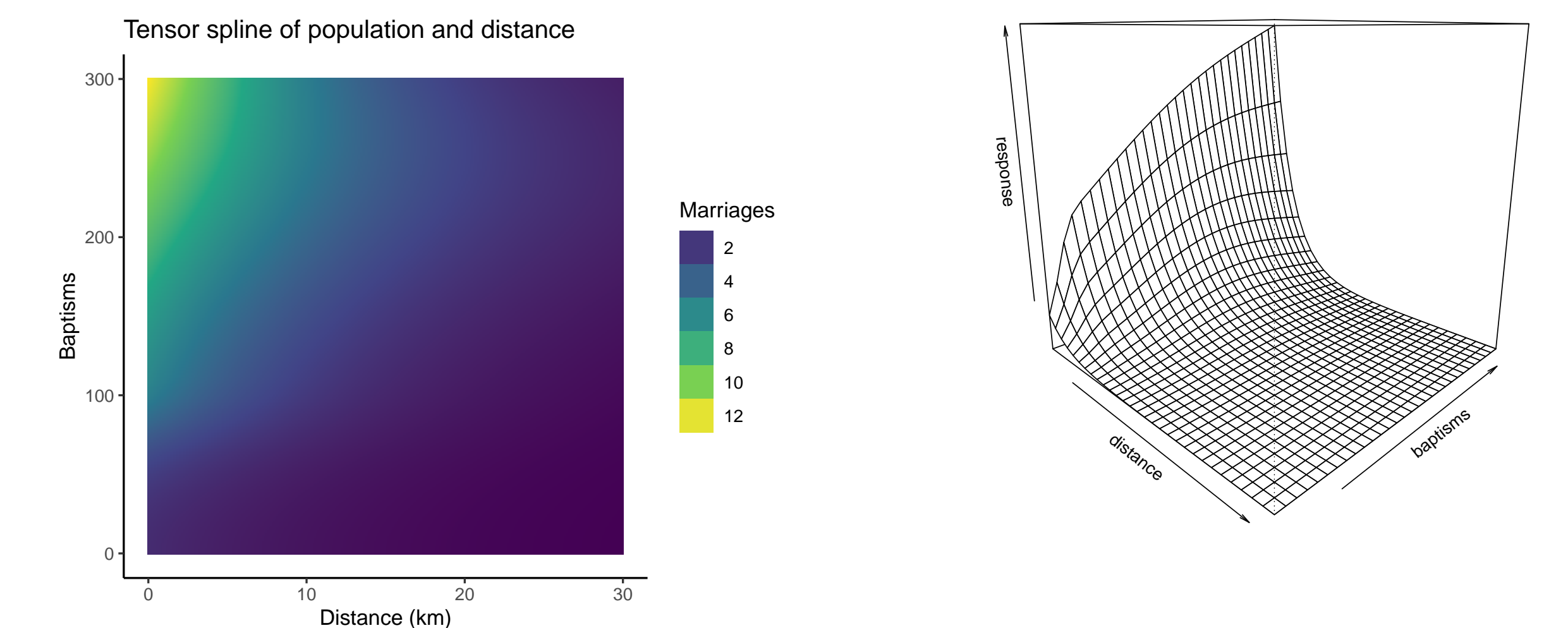
Results

Late Romano-British Pottery

Distance decay functions on land and water

Transport
- Land
- Water
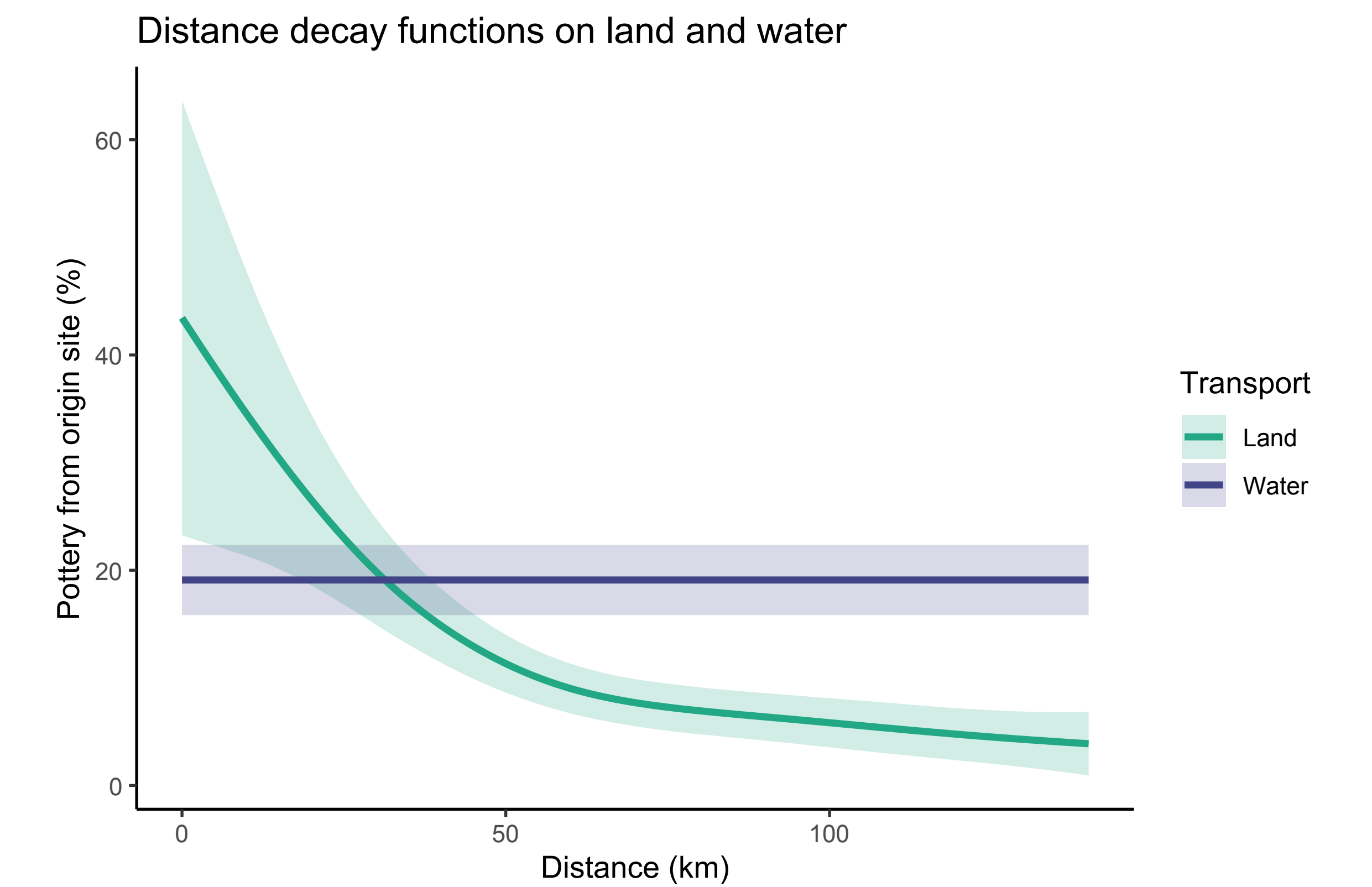
GAMs learn functions directly from data. Here we find that the amount of pottery moved over land decays logarithmically, but pottery moved by water does not.

Mission-period Chumash

Tensor spline of population and distance

Here we can use a "tensor" spline to model the nonlinear interaction between distance and population size that influences Chumash marriage patterns.

What's next?

Efficient estimation of GAMs and GAMMs is available in `R` with the package `mgcv`, or with a Bayesian implementation in `brms`. Scan this QR code or go to `https://github.com/nick-gauthier/gam-networks` for the code and data to generate this poster along with more detailed worked examples.