

# Bayesian Geographical Multi-Dimensional Scaling

Hayato Nishi<sup>a, \*</sup>, Yasushi Asami<sup>b</sup>

<sup>a</sup> The University of Tokyo, [h.nishi@ua.t.u-tokyo.ac.jp](mailto:h.nishi@ua.t.u-tokyo.ac.jp)

<sup>b</sup> The University of Tokyo, [asami@csis.u-tokyo.ac.jp](mailto:asami@csis.u-tokyo.ac.jp)

\* Corresponding author

**Keywords:** Similarity maps, Time distance, Multi-dimensional scaling, Bayesian statistics

Abstract:

Multi-dimensional scaling (MDS) is a popular method of visualizing the similarity of individuals in a dataset. When dissimilarities between individuals in a dataset are measured, MDS projects these individuals into the (typically two- or three-dimensional) map. In this map, because similar individuals are projected to be close to one another, distances between individuals correspond to their dissimilarities. In other words, MDS makes a similarity map of a dataset.

Some of the dissimilarities and distances have a strong relation to the geographical location. For example, time distances are similar to geographical distances, and regional features will be similar if the regions are close together. Therefore, it will be useful to compare the MDS projection and geographical locations. However, because MDS projection is not concerned with the rotation, parallel translation, and similarity expansion, it might be difficult to compare the projection to the actual geographical locations. When geographically related similarities are visualized, projected locations should be bound to the geographical locations.

In this article, we propose Bayesian Geographical Multidimensional Scaling (BGMDs), in which geographical restrictions of projections are given from a statistical point of view. BGMDs gives not only geographically bound projections, but also incorporates the uncertainty of the projections.

The statistical model of BGMDs is as follows: for each individual  $n = (1, \dots, N)$  in a dataset, its geographical location is  $\mathbf{x}_n = \{x_{n,i}\} (i = 1, 2 \text{ or } i = 1, 2, 3)$ , and the observed dissimilarity between  $n$  and  $m$  is  $d_{n,m}$ . Under these conditions, the unobserved latent location (or projected location)  $\mathbf{z}_n$  follows the Gaussian distribution  $z_{n,i} \sim \mathcal{N}(x_{n,i}, \lambda^{-1})$ , and the latent distance between  $n$  and  $m$  is defined by  $\delta_{n,m} = \delta(\mathbf{z}_n, \mathbf{z}_m)$ , which is the Euclidean distance between latent locations. Because observed dissimilarity will include some noise, it also follows the Gaussian distribution  $d_{n,m} \sim \mathcal{N}(\delta_{n,m}, \tau^{-1})$ .

Once the precision (or inverse of the variance) hyperparameters  $\lambda$  and  $\tau$  are given, the posterior distributions  $p(\mathbf{z}_n | \mathbf{X}, \mathbf{d})$  show the projections by BGMDs. However, the posterior distributions cannot be derived analytically. Therefore, we approximate them with Gaussian distributions using the algorithm *Bayes by Backprop*.

The hyperparameters can be chosen to maximize the marginal likelihood of the BGMDs model. However, because it is as difficult to evaluate the marginal likelihood as it is the posterior distributions, we use the evidence lower bound (ELBO), which is the lower bound of the logarithm of the marginal likelihood, instead of the marginal likelihood.

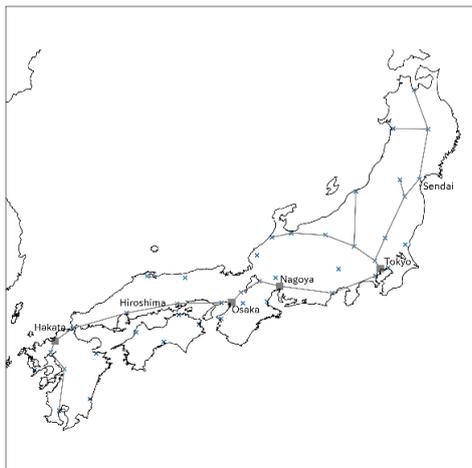


Figure 1. Geographical locations of major stations in Japan. (Coastlines are from GSHHG(A Global Self-consistent, Hierarchical, High-resolution Geography Database))

To show the usefulness of BGMDs, we used it to visualize the time distances between major stations in Japan. To catch the overall accessibility condition in Japan, one major station was chosen from each prefecture in Japan, except for Hokkaido and Okinawa, which are far from other prefecture. Time distances were located using <https://www.jorudan.co.jp/>, which is one of the most popular route search websites. The route search allowed all public traffic modes. The unit of time distance was modified to make the mean of time distance equal the mean of geographical distance. Figure 1 shows the geographical locations of the selected stations. Stations in prefectures with bullet trains (*Shinkansen*) service are joined with grey lines, and grey squares show the prefectures in which there are major airports.

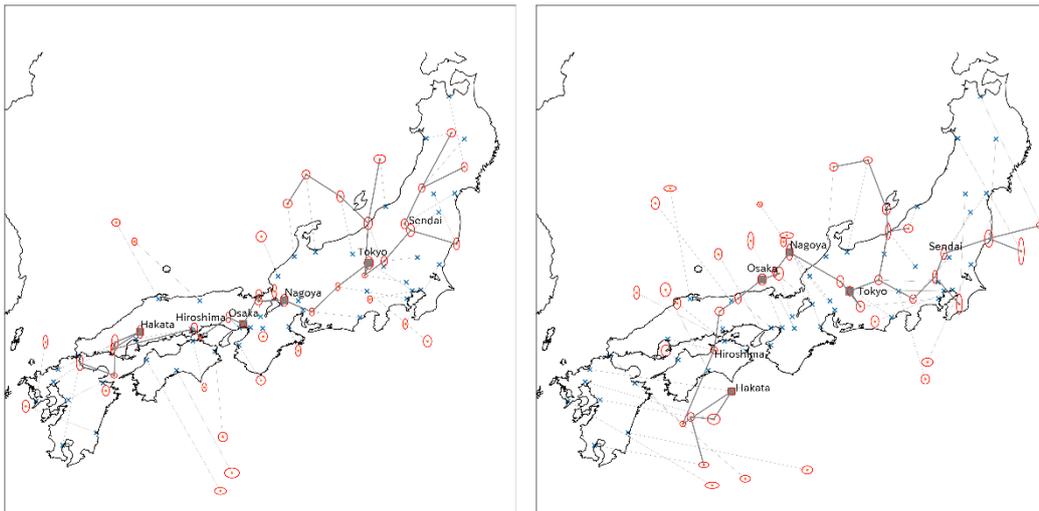


Figure 2. Time distance map created using BGMDs (left: without airplanes, right: with airplanes)

Figure 2 shows the BGMDs projection. The left map leaves out air travel time distance and the right one includes them. In each case, the units of time distances differ. Blue crosses show original geographical locations, and red points show projections. Red circles demarcate standard deviation ellipses of posterior distributions. As a result of a grid search for pairs of  $(\lambda, \tau)$  in which each takes a value from (0.1, 1, 10, 100),  $(\lambda, \tau) = (1, 1)$  was the best in both cases. On the left side of Figure 2, most of the stations move toward the center of Japan, but stations which are not served by *Shinkansen* move away from the center of Japan. This implies that the *Shinkansen* improve accessibility between prefectures. On the other hand, in the map on the right, major airports were allocated to be close to one another. As a result, Hakata station and the stations around it move closer to the center of Japan.

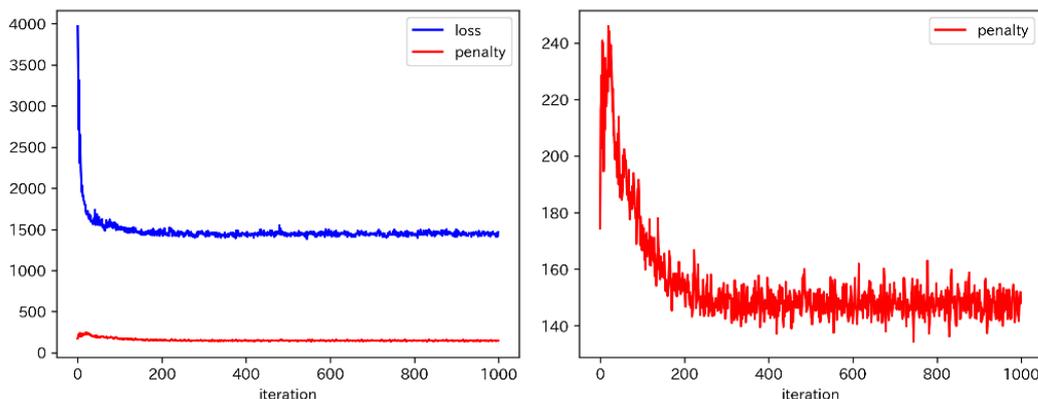


Figure 3. Learning process of BGMDs.

Figure 3 shows how the loss changed over the BGMDs learning process. The loss is the sign inversion of ELBO, and the penalty is the portion of the loss related to the geographical restriction term. The right side of Figure 3 is the zoom up of the penalty term. At the beginning of the learning process, BGMDs is not seriously concerned with the geographical restriction, but after the loss is saturated, the projected locations move to be closer to the geographical location.