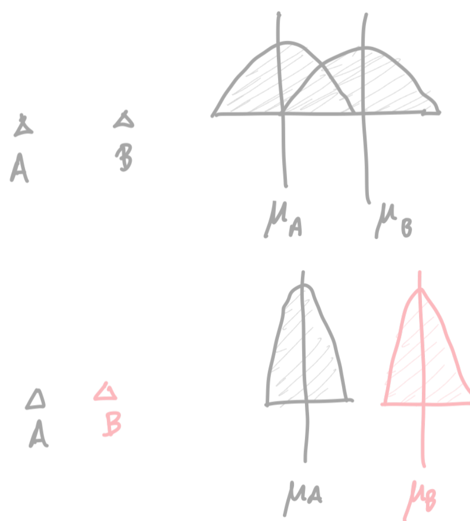
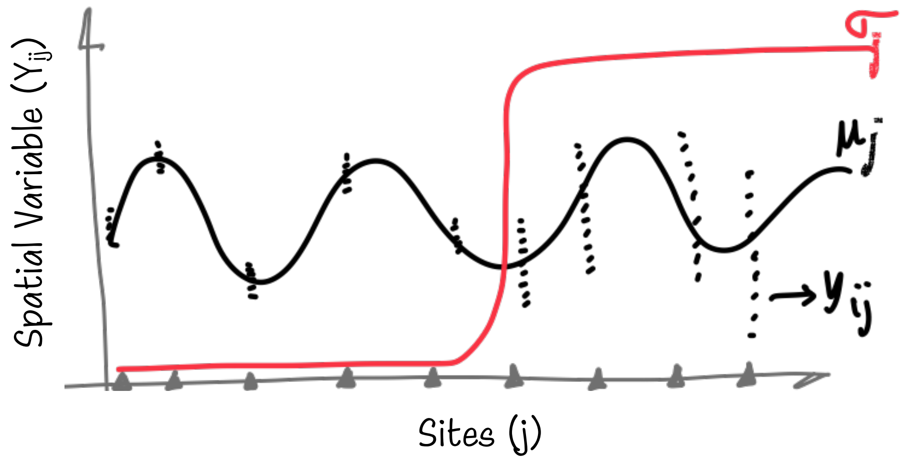
A simple introduction to **Uncertainty Projected Mapping (UPM)**



Anirban Chakraborty

Hosei University, Japan

Let's consider the following dataset shown in Fig1:



where, Y_{ij} = random observation Y at site *j* during event *i*

μ_j = *Known source mean of* Y_{ij} *at site j*

σ_j = *Known source standard deviation of* Y_{ij} *at site j*

Now. Let's answer these questions considering conventional mapping

Q1: How is do we map Y_{ij} ?   A1: We connect μ_j at all sites j

Q2: How is μ_j estimated? A2: Arithmetic mean of all Y_{ij} at site j

Next two questions are particularly important

Q3: Does the σ_j play any role in the mapping process? A3: No

Q4: If some sites have a significantly low sample size, does it affect the estimation of μ_j and, hence the mapping in any way? A4: No

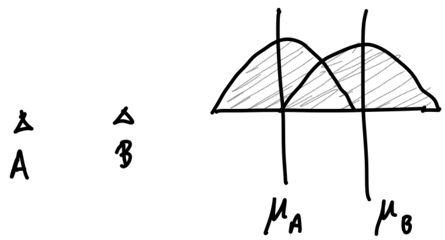
If you think the answers to Q3 and Q4 are not satisfactory, and expect σ_j and the sample size to affect the mapping, you might be interested in

Uncertainty Projected Mapping (UPM)

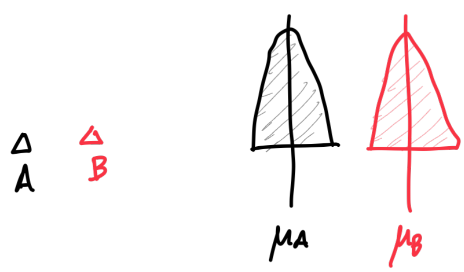
<u>**What's so special about UPM?**</u>

UPM decides on the mapped value at a site depending on whether the value is statistically significant compared to its neighbors.

When the value at a site cannot be differentiated from its neighbors (no statistical significance), UPM plots a smooth map. Thus, although μ_A and μ_B are distinct, UPM treats A and B as the same site.



When the value at a site can be differentiated from its neighbors (statistical significance), UPM plots the mapped value distinctly. Thus, μ_A and μ_B are distinct, and UPM also treats A and B as different sites.



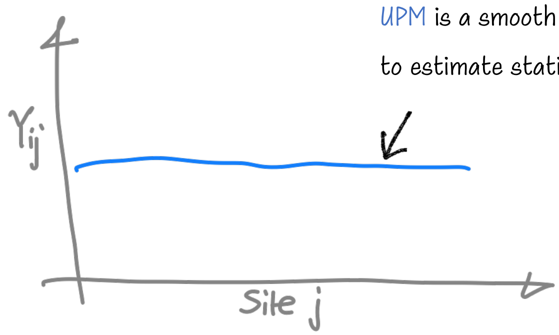
Just because the μ_j of two neighboring sites are different, it doesn't necessarily indicate the data distribution at the two sites are significantly different. UPM considers the statistical significance and accordingly decides on the mapped value at a site. **So, UPM has a statistical significance.**

Q: So, what would the UPM for Fig1 look like?

A: It **will depend on the sample size** of observations at the site.

When the sample size is **small**

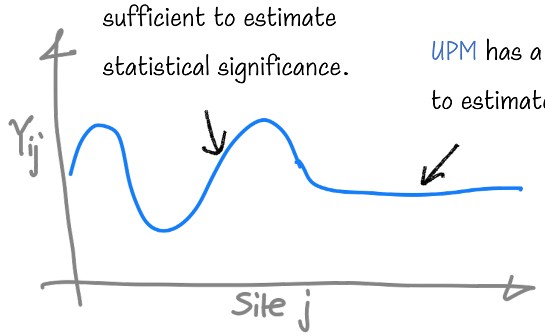
UPM is a smooth plot. Sample size is insufficient to estimate statistical significance for any site



When the sample size is **medium**

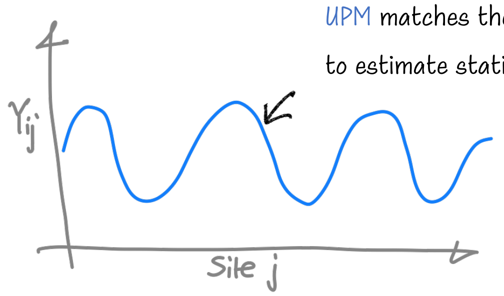
UPM has a local roughness. Sample size sufficient to estimate statistical significance.

UPM has a local smoothness. Sample size is insufficient to estimate statistical significance here.



When the sample size is **large**

UPM matches the Fig1. Sample size is sufficient to estimate statistical significance for all sites



<u>So, how does UPM work?</u>

UPM can be summarized by the following four equations

$$Y_{ij} \sim f(\mu_j, \sigma_j^2) \qquad (1)$$

$$\mu_j = \mu_g + \Delta\mu_j \qquad (2)$$

$$\Delta\mu_j = N\left(\frac{\sum_{m \in M_j} \Delta\mu_m}{N_j}, \frac{s_j^2}{N_j}\right) \qquad (3)$$

$$c = \sigma_j \times s_j \qquad (4)$$

In Equation (1), the function f could be a normal or log Normal distribution.

In Equation (2), μ_g is the group mean of Y_{ij} at sites j. $\Delta\mu_j$ is the fluctuation of μ_j from the μ_g at site j.

Equation (3), $\Delta\mu_j$ is modelled as a **conditional Autoregressive (CAR)** model. CAR gives a spatial structure to UPM by introducing a neighborhood M_j to the sites j. Estimation of $\Delta\mu_j$ depends on this neighborhood where N_j is the number of neighbors. When used as a prior in Bayesian Hierarchical modelling, CAR priors are usually modelled as a normal distribution with mean=0 and a variance.

What does the variance in a CAR model signify?

The variance $\frac{s_j^2}{N_j}$ is the dispersion of $\Delta\mu_j$ in the neighborhood of site j. When $\frac{s_j^2}{N_j}$ is high, the values of $\Delta\mu_j$ are significantly different from each other. When $\frac{s_j^2}{N_j}$ is low, the values of $\Delta\mu_j$ are very similar. Thus, controlling the S_j parameter can control the mapping.

And hence, in Equation (4) we introduce the constraint $c=\sigma_j \times s_j$ so that σ_j could control the mapping.

References on UPM:

1. Chakraborty, A., & Goto, H. (2018). A Bayesian model reflecting uncertainties on map resolutions with application to the study of site response variation. *Geophysical Journal International*, *214*(3), 2264-2276.

2. Chakraborty, A., & Goto, H. (2020). Visualizing data saturation process in mapping site amplification of earthquake ground motions. *Journal of Natural Disaster Science*, *40*(2), 14-25.

3. Chakraborty, A., Goto, H., & Sawada, S. (2024). Updating proxy-based site amplification map with in-situ data in Osaka, Japan: A Bayesian scheme based on uncertainty projected mapping. *Earthquake Spectra*, 2024;40(1):113-142.

.

References on CAR:

1. De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, *64*, 107-133.