Multi-resolution Spatial Statistics

Iao Hoi (Isaac) Leong   Supervisor: Dr Tingjin Chu

THE UNIVERSITY OF MELBOURNE

Motivation

Many spatial datasets are collected at incompatible locations and resolutions being the average of the local neighbourhood.

If this mismatch is ignored, simple upsampling can blur spatial structure and give uncertainty statements that are overconfident.

This project is motivated by scientific settings where the covariate is intrinsically aggregated by the measurement process, so the analysis should respect the support on which it is observed. One example is the following two datasets as shown in Figure 1 captured for astronomical research.
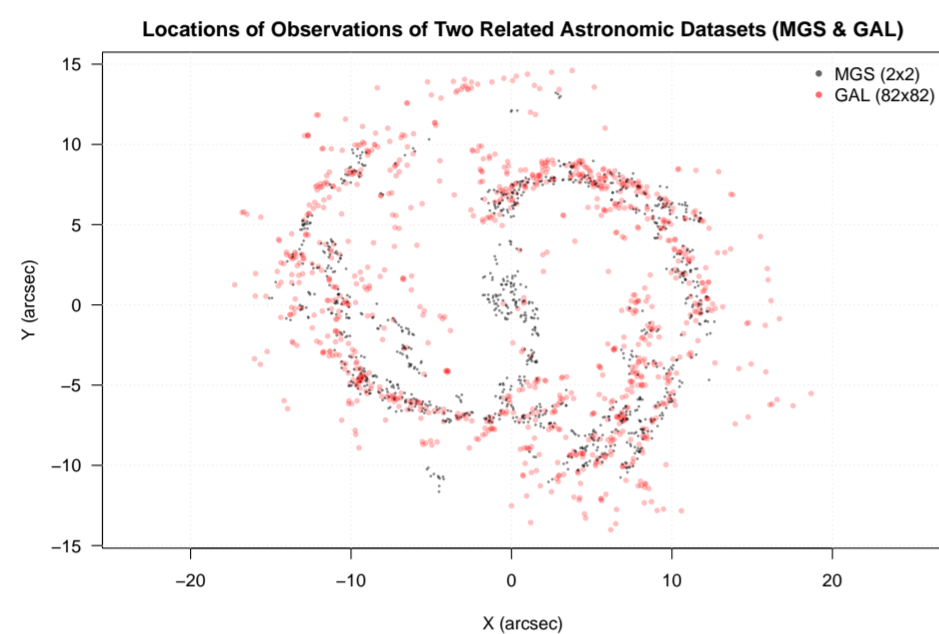

Figure 1: Observation locations for two metallicity datasets with support mismatch: MGS is sampled at finer resolution, while GAL is measured on a much coarser footprint. Both record gas-phase metallicity across the same galaxy using different indicators.

Our goal is to build a principled framework for relating fine-scale targets to multi-resolution covariates, as a foundation for later prediction and application to astronomical data.

Background

Consider the following spatial process

$$y(\mathbf{s}) = \beta_0 + \beta_1 d(\mathbf{s}) + \beta_2 x(\mathbf{s}) + \varepsilon(\mathbf{s}), \qquad (1)$$

where both $y(\cdot)$ and $d(\cdot)$ are observed at $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, while $x(\cdot)$ is observed at other locations at $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ with different resolution in a 100×100 domain. We assume that $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ has a resolution of 5×5 while $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ has a resolution of 3×3 for simplicity.

We divide the 3×3 area to 1×1 basic area unit (BAU) by modelling the coarsing 3×3 resolution as an average of the BAU. Then $\mathbf{x}_u = \begin{bmatrix} x(\mathbf{u}_1) & x(\mathbf{u}_2) & \cdots & x(\mathbf{u}_m), \end{bmatrix}^\top$ and similar for $\mathbf{x}_v$ which denotes $x(\cdot)$ at all BAUs, we can put it in matrix form as,

$$\mathbf{x}_u = W\mathbf{x}_v, \text{ where } W = \underbrace{\begin{bmatrix} \frac{1}{9} & \cdots & \frac{1}{9} & 0 & 0 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \frac{1}{9} & \cdots & \frac{1}{9} & 0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & 0 & 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{9} & \cdots & \frac{1}{9} \end{bmatrix}}_{\mathbb{R}^{m \times (9*m)}}$$

Then model $x(\cdot)$ at BAU level, as $x(\mathbf{v}) = \mathbf{r}(\mathbf{v})^\top \boldsymbol{\alpha} + \eta(\mathbf{v})$ would give the following distribution,

$$\mathbf{x}_v \sim \text{MVN}(\alpha_0 \mathbf{1}, \Sigma_\eta(\boldsymbol{\xi})) \implies \mathbf{x}_u \sim \text{MVN}(\alpha_0 \mathbf{1}, W\Sigma_\eta(\boldsymbol{\xi})W^\top)$$
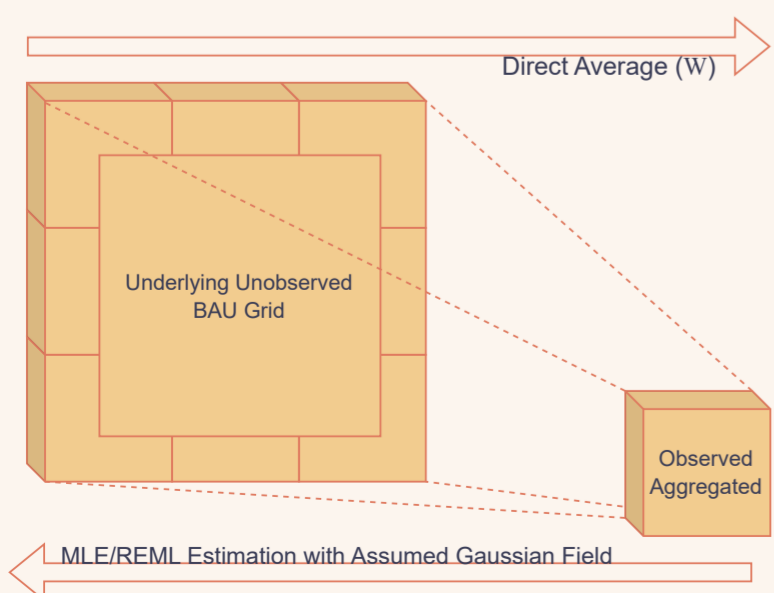

Figure 2: Approach for going from observed aggregated data and underlying BAU grids.

MLE

Since the aggregated data $\mathbf{x}_u$ is observed, we can use likelihood methods to estimate the unknown parameters $\alpha 0$ and $\boldsymbol{\xi}$. Sepcifically, the log-likelihood is:

$$L(\alpha_0, \boldsymbol{\xi}) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(|W\Sigma_\eta(\boldsymbol{\xi})W^\top|\right)$$
$$- \frac{1}{2}(\mathbf{x}_u - \alpha_0 W\mathbf{1})^\top \left(W\Sigma_\eta(\boldsymbol{\xi})W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0 W\mathbf{1}) \qquad (2)$$

Then, we by setting the gradient with respect to α_0 obtain the following estimator for $\widehat{\alpha}_0(\boldsymbol{\xi})$ in terms of the covariance function parameters:

$$\tilde{\alpha}_0(\boldsymbol{\xi}') = \frac{\mathbf{1}^\top(W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}\mathbf{x}_u}{\mathbf{1}^\top(W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}\mathbf{1}}$$

Without further assumption, it is difficult to further simplify constraint the log-likelihood. However, if the covariance function admits a separable variance (partial sill) parameter σ^2, then, the covaraince matrix $\Sigma(\boldsymbol{\xi})$ can be expressed as

$$\Sigma_\eta(\boldsymbol{\xi}) = \sigma^2 R_\eta(\boldsymbol{\xi}')$$

where $\boldsymbol{\xi}'$ is just $\boldsymbol{\xi}$ with σ^2 excluded. Then, by recognising the same trick as above, we can rewrite the log-likelihood in Equation (2) as:

$$L(\alpha_0, \sigma^2, \boldsymbol{\xi}') = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(|\sigma^2 W R_\eta(\boldsymbol{\xi}')W^\top|\right)$$
$$- \frac{1}{2}(\mathbf{x}_u - \alpha_0 \mathbf{1})^\top \left(\sigma^2 W R_\eta(\boldsymbol{\xi}')W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0 \mathbf{1}) \qquad (3)$$

Then by setting the gradient of Equation (3) with respect to σ^2, we obtain the following estimator for σ^2:

$$\tilde{\sigma}^2(\boldsymbol{\xi}') = \frac{1}{m}(\mathbf{x}_u - \alpha_0 \mathbf{1})^\top \left(W R_\eta(\boldsymbol{\xi}')W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0 \mathbf{1})$$

Hence, we obtain the final concentrated log-liklihood as,

$$L(\boldsymbol{\xi}') = -\frac{m}{2}\log(2\pi) - \frac{m}{2} - \frac{m}{2}\log\left(\tilde{\sigma}^2(\boldsymbol{\xi}')\right) - \frac{1}{2}\log\left|Q(\boldsymbol{\xi}')\right|$$

Then, we use optimisation algorithm such as Newton's Method to optimise with respect to $\boldsymbol{\xi}'$

Restricted Maximum Likelihood

MLE method typically have a downward bias for variance-related estimates. One way to mitigate is to use restricted maximum likelihood (REML). Specifically, we first ignore hide α_0 when estimating variance parameters, then recover it later to avooid losing degree of freedoms.

Mathematically, we just need to find $K \in \mathbb{R}^{m \times (m-1)}$ such that $\mathbf{1}_m \in \ker K^\top$, and w.l.o.g assume K is orthonormal. Then, define $\mathbf{x}_u^* = K^\top \mathbf{x}_u$, we obtain that the expectation on the new variable would be zero,

$$\mathbb{E}[\mathbf{x}_u^*] = K^\top \mathbb{E}[\mathbf{x}_u] = \alpha_0 K^\top \mathbf{1} = 0.$$

Then, we get that $\mathbf{x}_u^*$ follows multivariate distribution.

$$\mathbf{x}_u^* \sim \text{MVN}(\mathbf{0}_{m-1}, K^\top V_\eta(\boldsymbol{\xi})K)$$

where we define $V_\eta(\boldsymbol{\xi}) = W\Sigma_\eta(\boldsymbol{\xi})W^\top$ for the ease of notation. Then, the log-likelihood can be written as:

$$L(\boldsymbol{\xi}) = -\frac{m-1}{2}\log(2\pi) - \frac{1}{2}\log\left|K^\top V_\eta(\boldsymbol{\xi})K\right|$$
$$- \frac{1}{2}\mathbf{x}_u^\top K(K^\top V_\eta(\boldsymbol{\xi})K)^{-1}K^\top \mathbf{x}_u$$

Now, using the same procedure as outlined by LaMotte (2007), it can be shown that such restricted log-likelihood function is independent of the choice of matrix K. Specifically, it can be written in the form of,

$$L(\boldsymbol{\xi}) = -\frac{m-1}{2}\log(2\pi) - \frac{1}{2}(\mathbf{x}_u - \hat{\alpha}_0 \mathbf{1}_m)^\top V_\eta(\boldsymbol{\xi})^{-1}(\mathbf{x}_u - \hat{\alpha}_0 \mathbf{1}_m)$$
$$- \frac{1}{2}\log\left(\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{1}_m\right) - \frac{1}{2}\log|V_\eta(\boldsymbol{\xi})| + \frac{1}{2}\log(m)$$

where $\widehat{\alpha}_0 \triangleq \left(\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{x}_u$. The precise expression is an algebraic artifact but appears to be the same as above. Then, the estimation of the parameters $\boldsymbol{\xi}'$ can simply be done by maximising the likelihood.

Concentrated REML

Similar to MLE, we can also use concentration to avoid numerically estimating the value of σ^2 with similar assumption, the restricted log-likelihood can now be wrtten as:

$$L(\sigma^2, \boldsymbol{\xi}') = -\frac{m-1}{2}\log(2\pi) - \frac{1}{2}\log|\sigma^2 Q(\boldsymbol{\xi}')| + \frac{1}{2}\log(m)$$
$$- \frac{1}{2}\log(\sigma^{-2}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m) - \frac{1}{2\sigma^2}(\mathbf{x}_u - \widehat{\alpha}_0 \mathbf{1}_m)^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \widehat{\alpha}_0 \mathbf{1}_m)$$

We only need to concetrate σ^2 as α_0 is not to be estimated here.

$$\widehat{\sigma}^2(\boldsymbol{\xi}') = \frac{1}{m-1}(\mathbf{x}_u - \widehat{\alpha}_0 \mathbf{1}_m)^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \widehat{\alpha}_0 \mathbf{1}_m)$$

Kriging

Now, in order to the result at correct resolution, we first need to perform kriging to obtain the best linear unbiased predictor (BLUP) with respect to MSE based on the estimated model.

Ultimately, we would like to perform kriging at resolution the same as $y(\cdot)$ at location $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ with resolution $k \times k$. However, we will first focus on performing kriging at BAU level for the ease of mathematical derivation.
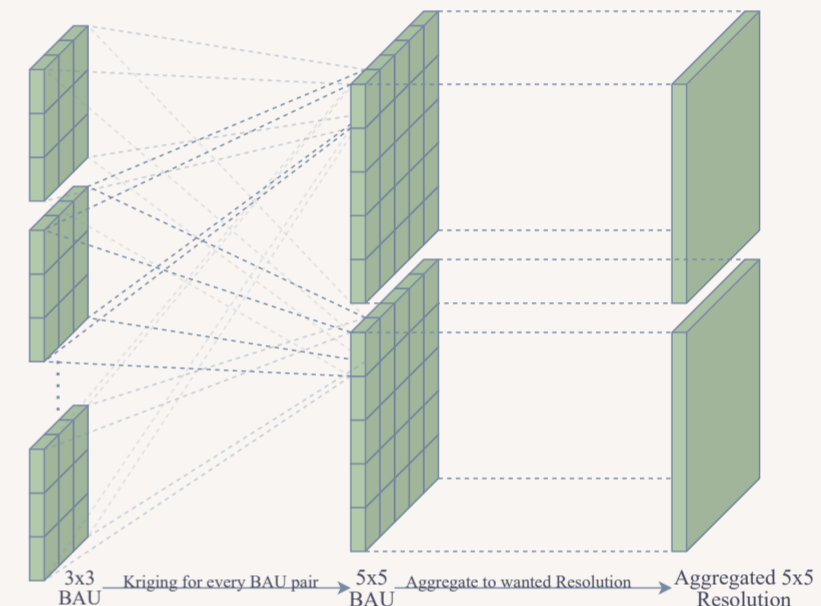

Figure 3: Kriging to the new locations and resolution.

As shown in Figure 3, we first need to krig for every BAU in 5×5 neighbourhood at new locations. This essentially needs to solve the following optimisation problem.

$$\text{minimise} \quad \mathbb{E}\left[\left(x(\mathbf{v}_0) - \underbrace{\left(\lambda_0 + \sum_{i=1}^m \lambda_i x_u(\mathbf{u}_i)\right)}_{\hat{x}(\mathbf{v}_0)}\right)^2\right]$$
$$\text{subject to} \quad \mathbb{E}[x(\mathbf{v}_0)] = \mathbb{E}[\hat{x}(\mathbf{v}_0)]$$

Then, using KKT condition from Boyd and Vandenberghe (2004), the solution is:

$$\widehat{x}(\mathbf{v}_0) = \widehat{\alpha}_0 + \mathbf{c}^\top V_\eta^{-1}(\widehat{\boldsymbol{\xi}})(\mathbf{x}_u - \widehat{\alpha}_0 \mathbf{1}),$$

However, not to forget that we actually need the BLUP for the new resolution, $k \times k$, i.e. solve the following problem,

$$\text{minimise} \quad \mathbb{E}\left[\left(x_{k \times k}(\mathbf{s}_0) - \left(\lambda_0 + \sum_{i=1}^m \lambda_i x_{3 \times 3}(\mathbf{u}_i)\right)\right)^2\right]$$
$$\text{subject to} \quad \mathbb{E}[x(\mathbf{v}_0)] = \mathbb{E}[\hat{x}(\mathbf{v}_0)]$$

Then, applying KKT conditions with algebraic manipulations, we show that the "average predictor" is the optimal predictor.

$$\overline{x}_{k \times k}(\mathbf{s}_0) = \frac{1}{k^2}\sum_{i=1}^{k^2} \widehat{x}_{\text{BAU}}(\mathbf{v}_j)$$

Acknowledgements

References

L. R. LaMotte, "A direct derivation of the reml likelihood function," *Statistical Papers*, vol. 48, no. 2, pp. 321–327, 2007.
S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.