Multi-resolution Spatial Statistics

Iao Hoi (Isaac) Leong

1 Background

Considering the following spatial process,

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

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

$$\mathbf{x}_u = W\mathbf{x}_v, \text{ where } W = \begin{bmatrix} \frac{1}{9} & \cdots & \frac{1}{9} & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \frac{1}{9} & \cdots & \frac{1}{9} & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & 0 & 0 & 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{9} & \cdots & \frac{1}{9} \end{bmatrix} \in \mathbb{R}^{m \times (9*m)} \tag{2}$$

Now, we can model $x(\cdot)$ at the BAU level as

$$x(\mathbf{v}) = \mathbf{r}(\mathbf{v})^\top \boldsymbol{\alpha} + \eta(\mathbf{v}) \tag{3}$$

where $\mathbf{r}(\mathbf{v})$ is a vector of known covariates, and the random process $\eta\mathbf{v}$ is a mean-zero Gaussian process with the stationary and isotropic convariance function $\gamma_\eta(h; \boldsymbol{\xi})$ with parameter $\boldsymbol{\xi}$. Moreover, $\eta(\cdot)$ is independent of $\varepsilon(\cdot)$. For simplicity, we can assume for now, $\mathbb{E}[x(\mathbf{v})] = \alpha_0$ constant (i.e. $\mathbf{r}(\mathbf{v}) = 1$). Therefore, using Equation (3), we have,

$$\mathbf{x}_v \sim \text{MVN}(\alpha_0 \mathbf{1}, \Sigma_\eta(\boldsymbol{\xi})) \tag{4}$$

Then, using the relationship $\mathbf{x}_u = W\mathbf{x}_v$ with Equation (3),

$$\mathbf{x}_u \sim \text{MVN}(\alpha_0 W\mathbf{1}, W\Sigma_\eta(\boldsymbol{\xi})W^\top) \tag{5}$$

2 MLE and Concentrated Log-likelihood

Now, given that we have observed $\mathbf{x}_u$ it is possible to perform MLE to estimates the parameters α_0 and $\boldsymbol{\xi}$. From the definition of multivariate normal distribution, the log-likelihood is,

$$\begin{aligned} L(\alpha_0, \boldsymbol{\xi}) = &-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(\left|W\Sigma_\eta(\boldsymbol{\xi})W^\top\right|\right) \\ &- \frac{1}{2}(\mathbf{x}_u - \alpha_0 W\mathbf{1})^\top \left(W\Sigma_\eta(\boldsymbol{\xi})W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0 W\mathbf{1}) \end{aligned} \tag{6}$$

Now, taking the derivative with respect to α_0, we have,

$$\begin{aligned} \frac{\partial L(\alpha_0, \boldsymbol{\xi})}{\partial \alpha_0} &= -\frac{1}{2}\left(\frac{\partial}{\partial \alpha_0}(\mathbf{x}_u - \alpha_0 W\mathbf{1})\right)^\top \left(2(W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}\right)(\mathbf{x}_u - \alpha_0 W\mathbf{1}) \\ &= (W\mathbf{1})^\top \left(W\Sigma_\eta(\boldsymbol{\xi})W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0 W\mathbf{1}) \end{aligned}$$

But, notice $W\mathbf{1}_{9m} = \mathbf{1}_m$ by definition of W from Equation (2).

$$= \mathbf{1}^\top (W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1})$$

Then, by setting it to zero, we obtain

$$\begin{aligned} 0 &\triangleq \mathbf{1}^\top (W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}(\mathbf{x}_u - \tilde{\alpha}_0\mathbf{1}) \\ \tilde{\alpha}_0(\boldsymbol{\xi}') &= \frac{\mathbf{1}^\top (W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}\mathbf{x}_u}{\mathbf{1}^\top (W\Sigma_\eta(\boldsymbol{\xi})W^\top)^{-1}\mathbf{1}} \end{aligned}$$

2.1 Constraint on Covariance Function

Now, 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[1],

$$\Sigma_\eta(\boldsymbol{\xi}) = \sigma^2 R_\eta(\boldsymbol{\xi}') \tag{7}$$

where $\boldsymbol{\xi}'$ is just $\boldsymbol{\xi}$ with σ^2 excluded. Then, by recognising the same trick as above, we can rewrite Equation (6) as

$$\begin{aligned} L(\alpha_0, \sigma^2, \boldsymbol{\xi}') = &-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(\left|\sigma^2 W R_\eta(\boldsymbol{\xi}')W^\top\right|\right) \\ &- \frac{1}{2}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top \left(\sigma^2 W R_\eta(\boldsymbol{\xi}')W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1}) \end{aligned} \tag{8}$$

Now, applying the same technique as for α_0 to simplify $W\mathbf{1}$ and linear algebra identities,

$$\begin{aligned} \frac{\partial L(\alpha_0, \sigma^2, \boldsymbol{\xi}')}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\Bigg[-\frac{1}{2}\log\left(\sigma^{2m}\left|W R_\eta(\boldsymbol{\xi}')W^\top\right|\right) \\ &\qquad - \frac{1}{2}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top \left(\sigma^2 W\Sigma_\eta(\boldsymbol{\xi}')W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1})\Bigg] \\ &= \frac{\partial}{\partial \sigma^2}\left[-\frac{m}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top (W R_\eta(\boldsymbol{\xi}')W^\top)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1})\right] \\ &= -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top (W R_\eta(\boldsymbol{\xi}')W^\top)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1}) \end{aligned}$$

Then, by setting the above to 0, we have

$$\begin{aligned} 0 &\triangleq -\frac{m}{\tilde{\sigma}^2} + \frac{1}{\tilde{\sigma}^4}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top (W R_\eta(\boldsymbol{\xi}')W^\top)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1}) \\ \tilde{\sigma}^2(\boldsymbol{\xi}') &= \frac{1}{m}(\mathbf{x}_u - \alpha_0\mathbf{1})^\top \left(W R_\eta(\boldsymbol{\xi}')W^\top\right)^{-1}(\mathbf{x}_u - \alpha_0\mathbf{1}) \end{aligned}$$

Therefore, we can define the mean and partial sill as a function of the other parameters $\boldsymbol{\xi}'$.

$$\begin{aligned} \tilde{\alpha}_0(\boldsymbol{\xi}') &= \left(\mathbf{1}^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}\right)^{-1}\left(\mathbf{1}^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{x}_u\right) \\ \tilde{\sigma}^2(\boldsymbol{\xi}') &= \frac{1}{m}(\mathbf{x}_u - \tilde{\alpha}_0(\boldsymbol{\xi}')\mathbf{1})^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \tilde{\alpha}_0(\boldsymbol{\xi}')\mathbf{1}) \end{aligned} \tag{9}$$

in which we define the matrix $Q(\boldsymbol{\xi}') \triangleq W R_\eta(\boldsymbol{\xi}')W^\top$.

[1] The constraint that the covariance satisfy Equation (7) is quite general. Some examples include Matérn class, Gaussian model, the powered exponential family and the spherical family.

2.2 Concentrated Log-likelihood

Now, by substituting these two solutions (Equation (9)) back in the log-likelihood function, we can now derive the constrained log-likelihood function.

$$\begin{aligned} L(\boldsymbol{\xi}') = &-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(\left(\tilde{\sigma}^2(\boldsymbol{\xi}')\right)^m |Q(\boldsymbol{\xi}')|\right) \\ &- \frac{1}{2}\left(\tilde{\sigma}^2(\boldsymbol{\xi}')\right)^{-1}(\mathbf{x}_u - \tilde{\alpha}_0(\boldsymbol{\xi}')\mathbf{1})^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \tilde{\alpha}_0(\boldsymbol{\xi}')\mathbf{1}) \end{aligned}$$

However, notice that the last term can be simplified as another $m\tilde{\sigma}^2(\boldsymbol{\xi}')$ term,

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

Then, one can use optimisation algorithm such as Newton's Method to optimise with respect to $\boldsymbol{\xi}'$ to find the estimates and substitute back in Equation (9).

3 Restricted Maximum Likelihood Method

Now, simple MLE method typically have a downward bias for the estimate of variance. One way to mitigate this problem is to use restricted maximum likelihood (REML). Specifically, one way to avoid the bias is to not estimate α_0 when estimating the parameters for the covariance function so that we do lose degree of freedoms for that part. Mathematically, we just need to find a matrix $K \in \mathbb{R}^{m\times(m-1)}$ such that $\mathbf{1}_m \in \ker K^\top$. Also, for later convenience, we will also assume K is orthonormal. Then, if we 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.

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

where we define $V_\eta(\boldsymbol{\xi}) = W\Sigma_\eta(\boldsymbol{\xi})W^\top$ for the ease of notation. Then, the log-likelihood function 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 [1], 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,

$$\begin{aligned} L(\boldsymbol{\xi}) = &-\frac{m-1}{2}\log(2\pi) - \frac{1}{2}\log|V_\eta(\boldsymbol{\xi})| + \frac{1}{2}\log(m) \\ &- \frac{1}{2}\log\left(\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{1}_m\right) - \frac{1}{2}(\mathbf{x}_u - \hat{\alpha}_0\mathbf{1}_m)^\top V_\eta(\boldsymbol{\xi})^{-1}(\mathbf{x}_u - \hat{\alpha}_0\mathbf{1}_m) \end{aligned} \tag{10}$$

where $\widehat{\alpha}_0 \triangleq \left(\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{x}_u$, in which precise expression is an artifact during simplification and linear algebra manipulation but appears to be the same as the one derived from normal MLE method (Equation (9)). Then, the estimation of the parameters $\boldsymbol{\xi}$ can simply be done by maximising the restricted log-likelihood function with Newton's method.

3.1 Concentration

Similar to MLE, we can also use concentration to avoid numerically estimating the value of σ^2 with similar assumption as Section 2.1. With assumption Equation (7) and same definition of $Q(\boldsymbol{\xi}')$, the restricted log-likelihood can now be expressed as,

$$\begin{aligned} L(\sigma^2, \boldsymbol{\xi}') = &\frac{m-1}{2}\log(2\pi) - \frac{1}{2}\log\left|\sigma^2 Q(\boldsymbol{\xi}')\right| + \frac{1}{2}\log(m) \\ &- \frac{1}{2}\log(\sigma^{-2}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m) - \frac{1}{2\sigma^2}(\mathbf{x}_u - \widehat{\alpha}_0\mathbf{1}_m)^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \widehat{\alpha}_0\mathbf{1}_m) \end{aligned} \tag{11}$$

It is tempting to apply the same procedure to Equation (11) as in Section 2.1, but it is important to note that $\widehat{\alpha}_0$ is now dependent on σ^2 by its definition. However, since the definition of α_0 is a "fraction" and, hence, homogeneous in $V_\eta(\boldsymbol{\xi})$. Therefore, simply factoring out the partial sill would result in equivalent definition of σ^2 but independent of σ^2. Specifically,

$$\begin{aligned} \widehat{\alpha}_0 &= \left(\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top V_\eta(\boldsymbol{\xi})^{-1}\mathbf{x}_u \\ &= \left(\sigma^2\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\left(\sigma^2\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{x}_u\right) = \left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{x}_u \end{aligned}$$

With now alternative definition, we now can apply the procedure and set the partial derivative to zero,

$$\begin{aligned} \frac{\partial L(\sigma^2, \boldsymbol{\xi}')}{\partial\sigma^2} &= \frac{\partial}{\partial\sigma^2}\left(-\frac{m}{2}\log(\sigma^2) + \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\overbrace{(\mathbf{x}_u - \widehat{\alpha}_0\mathbf{1}_m)^\top Q(\boldsymbol{\xi}')^{-1}(\mathbf{x}_u - \widehat{\alpha}_0\mathbf{1}_m)}^{\text{RSS}}\right) \\ &= -\frac{m-1}{2\sigma^2} + \frac{1}{2\sigma^4}\,\text{RSS} \triangleq 0 \implies \widehat{\sigma}^2(\boldsymbol{\xi}') = \frac{1}{m-1}\,\text{RSS} \end{aligned}$$

Now, substitute back into Equation (11), we have the following simplified expression,

$$\begin{aligned} L(\boldsymbol{\xi}') = &\frac{1}{2}\log(m) + \frac{m-1}{2}\log(2\pi) - \frac{m-1}{2} \\ &- \frac{m-1}{2}\log(\widehat{\sigma}^2(\boldsymbol{\xi}')) - \frac{1}{2}\log|Q(\boldsymbol{\xi}')| - \frac{1}{2}\log\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')\mathbf{1}_m\right) \end{aligned} \tag{12}$$

Equation (12) is vastly simplified because the residual sum of squares (RSS) was cancelled by the estimate of σ^2.

3.2 Variance of the Estimated Constant Mean – $\widehat{\alpha}_0$

Since estimate of $\widehat{\alpha}_0$ has the same analytical form, we shall derive the variance here once.

$$\begin{aligned} \text{Var}(\widehat{\alpha}_0) &= \text{Var}\left(\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{x}_u\right) \\ &= \left(\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\right)\text{Var}(\mathbf{x}_u)\left(\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\right)^\top \\ &= \sigma^2\left(\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\right)Q(\boldsymbol{\xi}')\left(Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\right) \\ &= \sigma^2\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1}\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)\left(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')^{-1}\mathbf{1}_m\right)^{-1} \\ &= \sigma^2(\mathbf{1}_m^\top Q(\boldsymbol{\xi}')\mathbf{1}_m)^{-1} \end{aligned}$$

, where the second line is from the identity $\text{Var}(A\mathbf{x}) = A\text{Var}(\mathbf{x})\,A^\top$.

4 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 of Equation (3). 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 $k \times k$[2]. However, we will first focus on performing kriging at BAU level for the ease of mathematical derivation.

4.1 Kriging at BAU Level

This leads to the distribution $\mathbf{x}_s \sim \text{MVN}(\alpha_0 \mathbf{1}, W_{k\times k}\Sigma_\eta(\boldsymbol{\xi})W_{k\times k})$[3] based on Equation (4). Given the model specified with constant mean parameter assumption, we will perform Ordinary Kriging to obtain the "interpolated" points. The search of BLUP can be formulated as the following optimisation problem,

$$\begin{aligned} \text{minimise} \quad & \mathbb{E}\left[\left(x(\mathbf{v}_0) - \underbrace{\left(\lambda_0 + \sum_{i=1}^{m} \lambda_i x_u(\mathbf{u}_i)\right)}_{\hat{x}(\mathbf{v}_0)}\right)^2\right] \\ \text{subject to} \quad & \mathbb{E}[x(\mathbf{v}_0)] = \mathbb{E}[\hat{x}(\mathbf{v}_0)] \end{aligned} \tag{13}$$

Note that where we have used $x_u(\cdot)$ to denote it being the realisation from the 3×3 resolution. Then, unpacking the unbiasedness constraint with notice that $\mathbb{E}[x(\mathbf{v}_0)] = \alpha_0$,

$$\alpha_0 = \mathbb{E}\left[\lambda_0 + \sum_{i=1}^{m} \lambda_i x_u(\mathbf{u}_i)\right] = \lambda_0 + \sum_{i=1}^{m} \lambda_i \mathbb{E}[x_u(\mathbf{u}_i)] = \lambda_0 + \alpha_0 \sum_{i=1}^{m} \lambda_i$$

Since α_0 is unknown arbitrary parameters, in order for the equation to hold for any α_0, $\lambda_0 = 0$ and $\sum_{i=1}^{m} \lambda_i = 1$, or, equivalently in matrix form, $\mathbf{1}^\top \boldsymbol{\lambda} = 1$. We will also simplify the objective function as the following to obtain a convex optimisation problem[4],

$$\mathbb{E}\left[\left(x(\mathbf{v}_0) - \lambda_0 - \sum_{i=1}^{m} \lambda_i x_u(\mathbf{u}_i)\right)^2\right] = \mathbb{E}\left[\left(x(\mathbf{v}_0) - \sum_{i=1}^{m} \lambda_i x_u(\mathbf{u}_i)\right)^2\right]$$

By setting $Y_0 \triangleq x(\mathbf{v}_0) - \alpha_0$ and $Y_i \triangleq Y_i = x_u(\mathbf{u}_i) - \alpha_0$

$$\begin{aligned} &= \mathbb{E}\left[\left(Y_0 - \sum_{i=1}^{m} \lambda_i Y_i\right)^2\right] = \text{Var}\left(Y_0 - \sum_{i=1}^{m} \lambda_i Y_i\right) \\ &= \text{Cov}(Y_0, Y_0) - 2\text{Cov}\left(Y_0, \sum_{i=1}^{m} \lambda_i Y_i\right) + \text{Cov}\left(\sum_{i=1}^{m} \lambda_i Y_i, \sum_{i=1}^{m} \lambda_i Y_i\right) \\ &= \mathcal{C}_v(0) - 2\mathbf{c}^\top \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top V_\eta(\hat{\boldsymbol{\xi}})\boldsymbol{\lambda} \end{aligned}$$

[2]Note that the above specified the data to be of resolution 5×5 but for illustration purpose, we will make a slightly more generalised assumption.

[3]Note that $W_{5\times5}$ is defined similar to previous W matrix to take the average, and we will use $W_{3\times3}$ to disambiguate between the two resolutions.

[4]Convexity in the primal variable is easy to see given the quadratic form after simplification.

In which we have $V_\eta(\boldsymbol{\xi}') \triangleq W\Sigma_\eta(\boldsymbol{\xi}')W^\top$ following Section 3 and define $\mathbf{c}$ as the covariance vector between Y_0 and $Y_1, \ldots, Y_m$ and have the following analytical form at the i-th element,

$$\begin{aligned}\mathrm{Cov}(Y_0, Y_i) &= \mathrm{Cov}(x(\mathbf{v}_0) - \alpha_0, x_u(\mathbf{u}_i) - \alpha_0) = \mathrm{Cov}(x(\mathbf{v}_0), x_u(\mathbf{u}_i)) \\ &= \mathrm{Cov}\left(x(\mathbf{v}_0), \frac{1}{9}\sum_{\mathbf{v}\in N_{3\times 3}(\mathbf{u}_i)} x(\mathbf{v})\right) = \frac{1}{9}\sum_{\mathbf{v}\in N_{3\times 3}(\mathbf{u}_i)} \mathrm{Cov}(x(\mathbf{v}_0), x(\mathbf{v}))\end{aligned}$$

Then, we obtain the lagrange multiplier function with dual variable ν as,

$$\mathcal{L}(\boldsymbol{\lambda}, \nu) = \mathcal{C}_v(0) - 2\mathbf{c}^\top\boldsymbol{\lambda} + \boldsymbol{\lambda}^\top V_\eta(\hat{\boldsymbol{\xi}})\boldsymbol{\lambda} - 2\nu(\mathbf{1}^\top\boldsymbol{\lambda} - 1)$$

Its gradient with respect to the primal variable as,

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L} = -2\mathbf{c} + 2V_\eta(\hat{\boldsymbol{\xi}})\boldsymbol{\lambda} - 2\nu\mathbf{1}$$

Further, applying the KKT conditions [2], we simply need to solve the following system,

$$\begin{bmatrix} V_\eta(\hat{\boldsymbol{\xi}}) & -1 \\ \mathbf{1}^\top & 0 \end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda} \\ \nu\end{bmatrix} = \begin{bmatrix}\mathbf{c} \\ 1\end{bmatrix} \tag{14}$$

Then, solving Equation (14) would simply yield that the BLUP,

$$\begin{aligned}\widehat{\boldsymbol{\lambda}} &= V_\eta^{-1}(\hat{\boldsymbol{\xi}})(\mathbf{c} + \nu\mathbf{1}) \\ \widehat{\nu} &= \frac{1 - \mathbf{1}^\top V_\eta^{-1}(\hat{\boldsymbol{\xi}})\mathbf{c}}{\mathbf{1}^\top V_\eta^{-1}(\hat{\boldsymbol{\xi}})\mathbf{1}}\end{aligned} \tag{15}$$

Hence, substituting Equation (15) back in the linear predictor form, we have

$$\widehat{x}(\mathbf{v}_0) = \widehat{\alpha}_0 + \mathbf{c}^\top V_\eta^{-1}(\hat{\boldsymbol{\xi}})(\mathbf{x}_u - \widehat{\alpha}_0\mathbf{1})\,, \tag{16}$$

in which the simplification requires the exact form of $\widehat{\alpha}_0$ to be identical to the one in Equation (9). Then, the final objective value, which is also an estimate of the variance, is,

$$\begin{aligned}\mathbb{E}\left[(x(\mathbf{v}_0) - \widehat{x}(\mathbf{v}_0))^2\right] &= \mathcal{C}_v(0) - 2\mathbf{c}^\top\widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\lambda}}^\top V_\eta(\hat{\boldsymbol{\xi}})\widehat{\boldsymbol{\lambda}} \\ &= \mathcal{C}_v(0) - \mathbf{c}^\top V_\eta^{-1}(\hat{\boldsymbol{\xi}})\mathbf{c} + \nu^2\,\mathbf{1}^\top V_\eta^{-1}(\hat{\boldsymbol{\xi}})\mathbf{1}\,,\end{aligned}$$

where the new location only matters for the covariance vector $\mathbf{c}$ with the existing points.

4.2 Kriging at General Resolution – $k \times k$

With the result at BAU level above, we would now show that, given the assumption that the values in the resolution box is being collected in an average manner, the BLUP for any higher resolution box is just simply the average for each point. Note that although it is possible to directly obtain the BLUP by solving an optimisation problem similar to Equation (13), being able to derive the estimation from the fine-grained grid would allow for more flexible prediction. Specifically, we want to show that the average predictor of the BLUPs at the fine-grained level

$$\overline{x}_{k\times k}(\mathbf{s}_0) = \frac{1}{k^2}\sum_{i=1}^{k^2}\widehat{x}_{\mathrm{BAU}}(\mathbf{v}_j) \tag{17}$$

is also the BLUP predictor for the new coarse resolution observation $\mathbf{x}_s$.

First, we will show that it is indeed a linear predictor. By expanding, the analytical formulation from Equation (16), we have

$$\overline{x}_{k\times k}(\mathbf{s}_0) = \frac{1}{k^2}\sum_{j=1}^{k^2}\left({}_j\lambda_0 + \sum_{i=1}^{m} {}_j\lambda_i x_{3\times 3}(\mathbf{u}_i)\right),$$

where ${}_j\boldsymbol{\lambda}$ is the kriging weight for location j in the $k \times k$ grid.

$$= \underbrace{\left(\frac{1}{k^2}\sum_{j=1}^{k^2} {}_j\lambda_0\right)}_{\overline{\lambda}_0} + \sum_{i=1}^{m}\underbrace{\left(\frac{1}{k^2}\sum_{j=1}^{k^2} {}_j\lambda_i\right)}_{\overline{\lambda}_i} x_{3\times 3}(\mathbf{u}_i),$$

which is now clear that it is still of a linear predictor form.

Then, we will now tackle the difficult part – showing that it is indeed a solution to the following optimisation problem.

$$\begin{aligned} \text{minimise} \quad & \mathbb{E}\left[\left(x_{k\times k}(\mathbf{s}_0) - \left(\lambda_0 + \sum_{i=1}^{m}\lambda_i x_{3\times 3}(\mathbf{u}_i)\right)\right)^2\right] \\ \text{subject to} \quad & \mathbb{E}[x(\mathbf{v}_0)] = \mathbb{E}[\hat{x}(\mathbf{v}_0)] \end{aligned} \tag{18}$$

First, before showing optimality, we will quickly check that the average predictor indeeds follow the unbiasedness constraint, which is easy to see given the BAU level BLUPs are unbiased.

$$\begin{aligned} \mathbb{E}[\overline{x}_{k\times k}(\mathbf{s}_0)] &= \mathbb{E}\left[\frac{1}{k^2}\sum_{j=1}^{k^2}\widehat{x}_{3\times 3}(\mathbf{v}_j)\right] \\ &= \frac{1}{k^2}\sum_{j=1}^{k^2}\mathbb{E}[\widehat{x}_{3\times 3}(\mathbf{v}_j)] = \frac{1}{k^2}\sum_{j=1}^{k^2}\alpha_0 = \alpha_0 = \mathbb{E}[x_{k\times k}(\mathbf{s}_0)] \end{aligned}$$

Then, to show optimality, we can use KKT condition because this problem can transformed into a convex optimisation problem in similar procedures shown in Section 4.1 as follows,

$$\mathbb{E}\left[\left(x_{k\times k}(\mathbf{s}_0) - (\lambda_0 + \sum_{i=1}^{m}\lambda_i x_{3\times 3}(\mathbf{u}_i))\right)\right]$$

similarly let $Y_0' \triangleq x_{k\times k}(\mathbf{s}_0) - \alpha_0$ and everything else same as Section 4.1

$$\begin{aligned} &= \mathbb{E}\left[\left(Y_0' - \sum_{i=1}^{m}\lambda_i Y_i\right)^2\right] = \mathrm{Var}\left(Y_0' - \sum_{i=1}^{m}\lambda_i Y_i\right) \\ &= \mathrm{Var}(Y_0') - 2\mathrm{Cov}\left(Y_0', \sum_{i=1}^{m}\lambda_i Y_i\right) + \mathrm{Var}\left(\sum_{i=1}^{m}\lambda_i Y_i\right) \end{aligned}$$

Hence, the only difference would be the definition of the covariance vector. Specifically, for the i-th element of the covariance vector $\mathbf{c}'$,

$$[\mathbf{c}']_i = \mathrm{Cov}(Y_0', Y_i) = \mathrm{Cov}(x_{k\times k}(\mathbf{s}_0) - \alpha_0\,, x_{3\times 3}(\mathbf{u}_i) - \alpha_0) = \mathrm{Cov}(x_{k\times k}(\mathbf{s}_0)\,, x_{3\times 3}(\mathbf{u}_i))$$
$$= \mathrm{Cov}\left(\frac{1}{k^2}\sum_{j=1}^{k^2} x_{\mathrm{BAU}}(\mathbf{v}_j')\,, \frac{1}{9}\sum_{j=1}^{9} x_{\mathrm{BAU}}(\mathbf{v}_j)\right) = \frac{1}{k^2}\sum_{j=1}^{k^2}\mathrm{Cov}\left(x_{\mathrm{BAU}}(\mathbf{v}_j')\,, \frac{1}{9}\sum_{q=1}^{9} x_{\mathrm{BAU}}(\mathbf{v}_q)\right)$$
$$= \frac{1}{k^2}\sum_{j=1}^{k^2}\mathrm{Cov}\big({}_jY_0\,, Y_i\big) = \frac{1}{k^2}\sum_{j=1}^{k^2}\mathbf{c}_j \qquad \text{(where } {}_jY_0 \triangleq x_{\mathrm{BAU}}(\mathbf{v}_j') - \alpha_0 \text{ and } \mathbf{c}_j \text{ similarly.)}$$

Therefore, the Lagrangian multiplier can be written as,

$$\mathcal{L}(\boldsymbol{\lambda}, \nu) = \mathrm{Var}(Y_0') - 2\mathbf{c}'^{\top}\boldsymbol{\lambda} + \boldsymbol{\lambda}^{\top}V_\eta(\hat{\boldsymbol{\xi}})\boldsymbol{\lambda} - 2\nu(\mathbf{1}^{\top}\boldsymbol{\lambda} - 1),$$

where $\mathbf{c}'$ is the new covariance vector we just defined. Then, we only need to check the stationarity condition in order to apply the KKT condition. Specifically, we want to show that our average predictor $\overline{\boldsymbol{\lambda}}$ is a solution to the following equation.

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{\lambda}, \nu) = -2\mathbf{c}' + 2V_\eta(\hat{\boldsymbol{\xi}})\boldsymbol{\lambda} - 2\nu\mathbf{1} = 0 \tag{19}$$

However, using the definition of $\mathbf{c}'$, we have

$$\mathbf{c}' + \nu\mathbf{1} = \nu\mathbf{1} + \frac{1}{k^2}\sum_{j=1}^{k^2}\mathbf{c}_j = \frac{1}{k^2}\sum_{j=1}^{k^2}(\mathbf{c}_j + \nu\mathbf{1})$$
$$= \frac{1}{k^2}\sum_{j=1}^{k^2} V_\eta(\hat{\boldsymbol{\xi}})\,{}_j\widehat{\boldsymbol{\lambda}} = V_\eta(\hat{\boldsymbol{\xi}})\left(\frac{1}{k^2}\sum_{j=1}^{k^2}{}_j\widehat{\boldsymbol{\lambda}}\right) = V_\eta(\hat{\boldsymbol{\xi}})\overline{\boldsymbol{\lambda}},$$

which shows that $\overline{\boldsymbol{\lambda}}$ is indeed a solution to Equation (19) hence supporting optimality of the average predictor, concluding that $\overline{\boldsymbol{\lambda}}$ is indeed a solution of Equation (18) and, thus, BLUP for the coarse-level kriging.

5 Simulation Methods & Results

In order to test the methodologies above, we shall test on synthetic data by simulation. For this report, we will be experimenting on a 100×100 grid. Given the detailed description on the analysis pipeline, this sections focuses on obtaining the synthetic data and discussion on the simulation results. First, we can generate the base Gaussian random field at BAU level by following the distribution as outlined in Equation (4). Overall, this would be one of the most computational intensive stage in the analysis pipeline, particularly if naive inversion of the covariance matrix is formed since the matrix is relatively large (10000×10000). In order to make this more computationally efficient, an efficient way would be to use Cholesky Decomposition when sampling from this high dimensional Gaussian distribution. At the same time, given the current setup of Equation (4), nugget effect, which is common in other spatial statistic application, is neglected for simplicity.

After the basic grid is generated, we would then proceed to generate the synthetic observation data which also need to take into account of resolution. First, as an experiment configuration, 400 points would be drawn to form the synthetic observation sample. However, since each observation is assumed to have a resolution of 3×3, for each sampled point out of the set of 400 points, the actual observation is the average of that point with the neighbouring eight points.

In the following section, we begin by assessing parameter recovery under the two estimation approaches introduced in Sections 2 and 3 (MLE and REML). For each simulation replicate, we fit the covariance parameters using both methods and summarise Monte Carlo bias, variance, and RMSE across replicates. Following, we will show the results of at higher number of simulations.

Table 1: Monte Carlo performance for covariance parameter estimation for the first R replicates. The true values are: $\alpha_0 = 2, \sigma^2 = 2, \phi = 5$

		$R = 100$				$R = 300$				$R = 500$			
Param.	Method	Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE
σ^2	MLE	1.9710	−0.0290	0.1975	0.1987	1.9644	−0.0356	0.1851	0.1882	1.9716	−0.0284	0.1740	0.1762
	REML	1.9955	−0.0045	0.2040	0.2030	1.9888	0.0379	0.1719	0.1912	1.9959	−0.0041	0.1797	0.1796
ϕ	MLE	4.9624	−0.0376	0.6541	0.6519	4.9388	−0.0612	0.6576	0.6594	4.9317	−0.0683	0.6353	0.6384
	REML	5.0617	0.0617	0.6810	0.6804	5.0379	−0.0112	0.6854	0.6853	5.0301	0.0301	0.6623	0.6622
α_0	MLE	1.9921	−0.0079	0.1558	0.1552	2.0074	0.0074	0.1719	0.1718	2.0066	0.0066	0.1672	0.1671
	REML	1.9922	−0.0078	0.1558	0.1552	2.0075	0.0075	0.1719	0.1718	2.0065	0.0065	0.1672	0.1671

Table 2: Monte Carlo performance for parameter estimation using REML for higher number of simulations, R. The true values are: $\alpha_0 = 2, \sigma^2 = 2, \phi = 5$

	$R = 600$				$R = 800$				$R = 1000$			
Param.	Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE
σ^2	1.9950	−0.0050	0.1968	0.1967	1.9990	−0.0188	0.1978	0.1977	1.9978	−0.0022	0.1977	0.1976
ϕ	4.9999	−0.0001	0.6878	0.6872	5.0189	−0.0010	0.7029	0.7027	5.0165	0.0165	0.6976	0.6974
α_0	2.0122	0.0122	0.1784	0.1967	2.0065	0.0065	0.1796	0.1796	2.0104	0.0104	0.1784	0.1786

5.1 Results

The comparison results are as shown in Table 1. Overall, the Monte-Carlo estimation of all three parameters are relatively to the true values, in particular with decreasing standard deviation with larger number of simulations. Meanwhile, the REML estimation has systematically achieved lower bias, as theoretically supported, than the MLE estimates for the parameters relating to variance-covariance, which are namely σ^2 and ϕ. This effect is particularly profound for the variance parameter σ^2 where the bias almost drops by an order of magnitude. However, these REML estimations suffer from slightly lower precision represented by the larger standard deviation demonstrating potential trade-off. Given that the RMSE with REML estimator shows consistently larger values than MLE estimator, this suggests that estimation error for the variance parameters is primarily driven by variance.

In particular, the range parameter, ϕ, is the most difficult parameters to estimate with consistently higher standard deviation and RMSE. One possible explanation may be that it is the only parameter that was estimated via numerical procedures, hence resulting in larger uncertainties and errors overall.

Table 2 reports REML performance for larger numbers of Monte Carlo replications. Since R increases only the number of simulated replications (not the information in each dataset), the underlying sampling variability of the estimator is not expected to decrease with R. Instead, larger R reduces Monte Carlo error in the reported summaries, so small non-monotonic changes in SD/RMSE are expected. The differences observed between $R = 600, 800,$ and 1000 are within plausible Monte Carlo variability, indicating that the results have largely stabilised. With this simulation study established, we will now look at the application of this method in a realistic setting stemmed from astronomy.

6 Application

In order to test the methodologies developed in realistic settings, we will use two astronomical datasets. These two spatial datasets – MGS and GAL[5] – are collected at different spatial resolution and layouts. MGS provides higher resolution field of the covariate variable at resolution $82\text{pc} \times 82\text{pc}$. The GAL dataset has collected measurements of metallicity of a specific galaxy at lower resolution field – $184\text{pc} \times 184\text{pc}$.

6.1 Processing Pipeline

We begin by converting the raw sky-coordinate measurements into deprojected planar coordinates in the galaxy frame, producing a common spatial representation that can be used consistently throughout the subsequent modelling pipeline. The resulting coordinates are now in parsec[6], but in order to establish the BAU approach, we shall define the BAU size be 20pc so that the BAU grid for GAL and MGS are roughly 9×9 and 4×4 respectively.

Next, we fit the aggregated observation model on the MGS dataset to estimate the mean and covariance parameters required for prediction. Although both MLE and REML are available in our framework, we use REML in the application, as the simulation study suggests it is more stable and less biased for covariance estimation under support mismatch. With these fitted parameters, we then perform ordinary kriging to predict the MGS covariate field onto the GAL support, producing an aligned covariate surface for the downstream analysis.

Finally, we model the GAL metallicity observations using a stationary geostatistical model fitted by maximum likelihood. Before fitting, we remove observations with missing or non-finite values and retain only the corresponding coordinates so that the likelihood is evaluated on a clean and consistent set of locations. For this stage, we do not use our custom likelihood implementation; instead, we rely on the `geoR` [3] package, which provides a robust and convenient interface for MLE fitting, together with more detailed diagnostic outputs and model comparison metrics. We fit both a baseline model with a constant mean and a model including the aligned covariate, allowing us to assess whether the kriged MGS field improves explanatory power for the response.

6.2 Results

After removing missing and non-finite values in the GAL response, $n_{\text{GAL}} = 869$ locations remain for the downstream model fit. The MGS dataset contains $n_{\text{MGS}} = 1000$ observations and is used to estimate the spatial dependence structure and produce aligned predictions.

Fitting the aggregated observation model to MGS under an exponential covariance produced a moderate range estimate but a comparatively large marginal variance, which is consistent with the high dispersion in the raw MGS observations. The implied correlation matrix used for prediction was numerically well-behaved, with a non-negligible minimum eigenvalue and a moderate condition number (Table 3).

Table 3: MGS aggregated model fit (exponential covariance) and properties for the correlation matrix.

	$\hat{\alpha}_0$	$\hat{\phi}$	$\widehat{\sigma}^2$	$\lambda_{\min}(Q)$	$\kappa(Q)$
MGS fit	11.6	10.0	1043	9.41×10^{-4}	746

[5] Specifically, we are using the `CO21_mgsd` from MGS dataset and `Z_RS32` from the `GAL` dataset to demonstrate our method.
[6] Note that the data is saved in kiloparsec (kpc).

Table 4: Maximum likelihood fits for the GAL response using `geoR` with an exponential correlation model. Here σ^2 denotes the partial sill and τ^2 the nugget.

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\widehat{\sigma}^2$	$\hat{\phi}$	$\widehat{\tau}^2$	$\log L$	AIC	BIC	Practical range
Baseline (constant mean)	8.7213	–	0.0053	23.45	0.0004	1189	-2369	-2350	70.26
+ aligned covariate	8.7182	0.0002	0.0053	23.65	0.0005	1190	-2369	-2345	70.84

We then fit two geostatistical models for the GAL response using `geoR` via maximum likelihood: a baseline model with a constant mean and a linear-trend model including the aligned covariate obtained from kriging. The fitted covariance parameters are nearly unchanged between the two models, and the covariate coefficient is effectively zero. In terms of likelihood-based metrics, the improvement in maximised log-likelihood is marginal and does not translate into a meaningful AIC gain, while BIC increases due to the extra parameter (Table 4).

References

[1] L. R. LaMotte, "A direct derivation of the reml likelihood function," *Statistical Papers*, vol. 48, no. 2, pp. 321–327, 2007.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. [Online]. Available: http://www.amazon.com/Convex-Optimization-Stephen-Boyd/dp/0521833787%3FSubscriptionId%3D192BW6DQ43CK9FN0ZGG2%26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0521833787

[3] P. Ribeiro Jr. and P. Diggle, "geoR: a package for geostatistical analysis," *R-NEWS*, vol. 1, no. 2, pp. 15–18, 2001. [Online]. Available: http://cran.R-project.org/doc/Rnews