Universal Spatio-temporal Kriging

Universal kriging to obtain best linear unbiased predictor (BLUP)

We first clarify how universal kriging obtains BLUP as following (Goldberger, 1962). We first define an OLS model as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{V}) \tag{a1}$$

where, $\boldsymbol{y}$ is a $N \times 1$ vector of observation variable, $\boldsymbol{X}$ is a $N \times K$ matrix of explain variables, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of OLS residuals, $\boldsymbol{0}$ is a $N \times 1$ vector composing entries with value 0, $\boldsymbol{V}$ is a $N \times N$ variance-covariance matrix.
We also define an estimation model as follows:

$$y_0 = \boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0, \qquad E(\varepsilon_0 \varepsilon) = \boldsymbol{c} \tag{a2}$$

where, y_0 is a scalar of observation value at point 0, $\boldsymbol{x}_0$ is a $K \times 1$ vector of explain variables at point 0, ε_0 is a scalar of residual of prediction at point 0, $\boldsymbol{c}$ is a $K \times 1$ vector of covariance between ε_0 and ε.

Now, letting $\boldsymbol{a}$ to be a $N \times 1$ of weighting vector, we have the following three formulas based on the characteristics of BLUP:

best estimator:

$$min[var(y_0 - \hat{y}_0)] \tag{a3}$$

linear estimator:

$$\hat{y}_0 = \boldsymbol{a}'\boldsymbol{y} \tag{a4}$$

unbiased estimator:

$$E(y_{\boldsymbol{0}} - \hat{y}_0) = 0 \tag{a5}$$

From the Eq. (a1), (a4), and (a5), we get the follow:

$$\begin{aligned} E(y_0 - \hat{y}_0) &= E(\boldsymbol{x}_{\boldsymbol{0}}'\boldsymbol{\beta} + \varepsilon_{\boldsymbol{0}} - \boldsymbol{a}'\boldsymbol{X\beta} - \boldsymbol{a}'\boldsymbol{\varepsilon}) \\ &= (\boldsymbol{a}'\boldsymbol{X} - \boldsymbol{x}_{\boldsymbol{0}}')\boldsymbol{\beta} \\ &= 0 \end{aligned} \tag{a6}$$

To fulfill Eq. (a6) regardless of $\boldsymbol{\beta}$, we get the follow:

$$\boldsymbol{a}'\boldsymbol{X} - \boldsymbol{x}_0' = \boldsymbol{0} \tag{a7}$$

We can also rewrite the variance as follows:

$$var(y_0 - \hat{y}_0) = \boldsymbol{a}'\boldsymbol{V}\boldsymbol{a} - \boldsymbol{2}\boldsymbol{a}'\boldsymbol{c} + \sigma^2 \tag{a8}$$

Under these conditions above, we can get the necessary and sufficient condition to make $var(y_* - \hat{y}_*)$ minimize by employing Lagrange's undetermined multiplier ($\boldsymbol{\lambda}$) and variance of the regression (σ^2).

That is, both of the partial differences the following Eq. (a8) by $\boldsymbol{a}$ and $\boldsymbol{\lambda}$ become 0:

$$f(\boldsymbol{a}, \boldsymbol{\lambda}) = \boldsymbol{a}'\boldsymbol{V}\boldsymbol{a} - \boldsymbol{2}\boldsymbol{a}'\boldsymbol{c} + \sigma^2 + \boldsymbol{2}(\boldsymbol{a}'\boldsymbol{X} - \boldsymbol{x}_0')\boldsymbol{\lambda} \tag{a9}$$

$$\begin{aligned} \frac{\partial f(\boldsymbol{a}, \boldsymbol{\lambda})}{\partial \boldsymbol{a}} &= 2\boldsymbol{V}\boldsymbol{a} - 2\boldsymbol{c} + 2\boldsymbol{X}\boldsymbol{\lambda} \\ \frac{\partial f(\boldsymbol{a}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= 2(\boldsymbol{X}'\boldsymbol{a} - \boldsymbol{x}_0) \end{aligned} \tag{a10}$$

Now, we solve the following Eq.:

$$\begin{bmatrix} \boldsymbol{V} & \boldsymbol{X} \\ \boldsymbol{X}' & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{x}_0 \end{bmatrix} \tag{a11}$$

Then, we get the follow:

$$\hat{\boldsymbol{a}}' = \boldsymbol{c}'[\boldsymbol{I} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{V}^{-1} + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V} \tag{a12}$$

Inserting Eq. (a12) to Eq. (a4), we get the follows:

$$\begin{aligned} \hat{y}_0 &= \hat{\boldsymbol{a}}'\boldsymbol{y} \\ &= \boldsymbol{c}'[\boldsymbol{I} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{V}^{-1}\boldsymbol{y} + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{y} \end{aligned} \tag{a13}$$

Rewriting Eq. (a13) with $\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$, we finally get BLUP as follows:

$$\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} + \boldsymbol{c}'\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}}\right) \tag{a14}$$

As above, we can get BLUP through the three characteristics, namely, the best, linear, and unbiased estimator. However, we need to know $\boldsymbol{c}$ and $\boldsymbol{V}$ beforehand. Kriging allows us to estimate $\boldsymbol{c}$ and $\boldsymbol{V}$. Thus, we can gain empirical BLUP (EBLUP) by structuralizing a *covariogram* or a covariance function (C).

Spatio-temporal kriging

We assume a Gaussian spatio-temporal random field (Z) defined by a spatial domain ($\mathcal{S}$) and temporal domain ($\mathcal{T}$). In the domain $\mathcal{S} \times \mathcal{T}$, there a sample ($\boldsymbol{z}$) with N space-time coordinates; that is, $\boldsymbol{z} = \left((\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_N, t_N)\right)$ and $(\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_N, t_N) \in \mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{R}$. Assuming second-order stationary, where $\operatorname{var}\left(Z(\boldsymbol{s}, t)\right) < \infty, for\ all\ \boldsymbol{s} \in \mathcal{S}, t \in \mathcal{T}$, we define a mean function (μ) and a spatio-temporal covariance function (C_{st}) as follows:

$$\mu(\boldsymbol{s}, t) = E\left(Z(\boldsymbol{s}, t)\right) \tag{b1}$$

$$C_{st}(\boldsymbol{h}, u) = \operatorname{cov}\left(Z(\boldsymbol{s}, t), Z(\boldsymbol{s}', t')\right) \tag{b2}$$

where, $\boldsymbol{h}$ is spatial distance, u is temporal distance. for any point $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{S} \times \mathcal{T}$ with $\boldsymbol{h} = \|\boldsymbol{s} - \boldsymbol{s}'\|$ and $u = |t - t'|$. Although it is difficult to estimate C_{st}, assuming it depends only the function of $\boldsymbol{h}$ and u, spatio-temporal kriging characterizes it with semi-variogram (γ):

$$C(\boldsymbol{h}, u) = \gamma(\boldsymbol{h}^r, u^r) - \gamma(\boldsymbol{h}, u) \tag{b3}$$

where, $\boldsymbol{h}^r$ and u^r are range of spatial distance and time lag where the value of the semi-variogram converges.

Theorical semi-variogram (γ_0) and empirical semi-variogram ($\hat{\gamma}$) are defined as follows:

$$\gamma_0(\boldsymbol{h}, u) = \frac{1}{2} E\left[\left(Z(\boldsymbol{s}, t) - Z(\boldsymbol{s}', t')\right)^2\right] \tag{b4}$$

$$\hat{\gamma}(\boldsymbol{h}, u) = \frac{1}{2|\mathbb{N}(\boldsymbol{h}, u)|} \sum_{k=1}^{\mathbb{N}(\boldsymbol{h}, u)} [Z(\boldsymbol{s}, t) - Z(\boldsymbol{s} + \boldsymbol{h}, t + u)]^2 \tag{b5}$$

where, $\mathbb{N}(\boldsymbol{h}, u)$ is the number of pairs of spatio-temporal lag.
Fitting to the empirical semi-variogram ($\hat{\gamma}$), we apply the *sum metric model* developed by Bilonick (1988) and revisited by Snepvangers et al. (2003), which shows the best fit of all. In addition, the sum metric model allows to easily interpret the impacts of spatial distance separately: time lag, and their interactions (Heuvelink et al. 1996). The sum metric model composes spatial, temporal, and spatio-temporal terms, as follows:

$$\gamma(\boldsymbol{h}, u) = \gamma_s(\boldsymbol{h}) + \gamma_t(t) + \gamma_{st}\left(\sqrt{\boldsymbol{h}^2 + (\alpha \cdot t)^2}\right) \tag{b6}$$

where γ_s, γ_t, and γ_{st} are spatial, temporal, and their combined semi-variogram, respectively; α is geometric anisotropy ratio.

Universal spatio-temporal kriging for PLPs

To evaluate the interpolation accuracy, cross-validation is an optimal way in kriging (Stone 1974; Geisser 1975). Cross-validation extracts a sample from a dataset whose sample size is n, in turn, to compare the observed value of the extracted point's value and interpolated value from rest n-1 samples, repeating n times (Wackernagel, 1995). This method is called the leave-one-out cross-validation. However, the leave-one-out cross-validation requires bulky calculation loads, especially when the sample size is large. Hence, we apply the k-fold cross-validation, which is a simplified way of leave-one-out cross-validation. In the k-fold cross-validation, we divide the dataset into k groups randomly and repeat k times to compare the selected groups' observed values and interpolated values with the rest k-1 groups. Burman (1989), Zhang (1993), and Shao (1993) evaluate the leave-one-out and the k-fold cross-validation in terms of accuracy and calculation loads. As a result, an efficient size of k is between 5 and 20. In this study, we set k as 5, considering the relatively large sample size.

Applying the 5-fold cross-validation, we calculate root mean squared errors (RMSE) by following the steps. First, we divide all PLP samples into five groups at random according to the characteristics. Next, we conduct spatio-temporal kriging with the same semi-variogram model for a group called kriged group. Finally, we calculate RMSE to compare the interpolated values of the kriging group and observation values of the rest four groups, called observation group, using the following Eq.:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\left(\ln\left(\widehat{PLP_i}\right) - \ln(PLP_i)\right)^2}{N}}$$

$$eRMSE = e^{RNSE}$$

where, $\hat{y}$ and y are interpolated values in the kriged group and observation values in the observation group, i and N are each observation point and the total number of the observation points. When $eRMSE = 1$, the interpolation has no error; $eRMSE = 1.1$, the error is 10%, on average.

References

Bilonick, R. A. (1988). Monthly hydrogen ion deposition maps for the northeastern U.S. from July 1982 to September 1984. *Atmospheric Environment*, *22*(9), 1909–1924. https://doi.org/10.1016/0004-6981(88)90080-7

Burman, P. (1989). A Comparative Study of Ordinary Cross-Validation , v-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika*, *76*(3), 503–514.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*(350), 320–328. https://doi.org/10.1080/01621459.1975.10479865

Goldberger, A. S. (1962). Best Linear Unbiased Prediction in the Generalized Linear Regression Model. *Journal of American Statistical Association*, *57*(298), 369–375.

Heuvelink, G. B. M., Musters, P., & Pebesma, E. J. (1996). Spatio-temporal kriging of soil water content. In E. Y. Baafi & N. A. Schofield (Eds.), *Geostatistics Wollongong* (2nd ed., pp. 1020–1030). Kluwer Academic Publishing.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486–494. https://doi.org/10.1080/01621459.1993.10476299

Snepvangers, J. J. J. C., Heuvelink, G. B. M., & Huisman, J. A. (2003). Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, *112*(3–4), 253–271. https://doi.org/10.1016/S0016-7061(02)00310-5

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, *36*(2), 111–147.

Wackernagel, H. (1995). *Multivariate Geostatistics: An Introduction with Applications*. Springer. https://doi.org/https://doi.org/10.1007/978-3-662-03098-1

Zhang, P. (1993). Model selection via multifold cross validation. *Annals of Statistics*, *21*(1), 299–313. https://doi.org/10.1214/aos/1176348654