Documentation for the Types of Gaussian Process Regression

Arjun Babu Nellikkattil*

(Dated: February 12, 2025)

I. SINGLE INPUT SINGLE OUTPUT GAUSSIAN PROCESS REGRESSION

- The default GPRs seen in texts and that used by Lickley and Fletcher (2024)
- Each grid has a GPR model independent of other grid points.
- One independent variable $\mathbf{x}$ (year) used to predict one dependent variable $f(\mathbf{x})$ (5 year mean anomaly of annual precipitation from historical period).
- A total of 33 CMIP6 simulations were used to design the regressor.
- Leave one out strategy is used to validate the method. i.e, in each iteration, one of the 33 models are used as "training data" and the other 32 is used to construct the priors.
- The hyperparameters are optimized by maximizing the sum of log marginal likelihood for the training period.

A. Parameter Definitions

- $\mathbf{x}$: Independent variable for seen observations. In this case *years*
- $f(\mathbf{x})$: Seen observations. Mean precipitation anomaly at a location from out of sample GCM (OOS GCM).
- $\mathbf{x}_*$: Years for which we are making predictions.
- $f(\mathbf{x}_*)$: Unseen or predicted precipitation anomaly.

* arjun.nellikkattil@georgetown.edu

- $\mathcal{M}$: Prior mean. Calculated as the multimodel mean precipitation anomaly from other GCMs (all the other 32 GCMs except the OOS GCM) for the selected location. $\mathcal{M}(\mathbf{x_i}) = \frac{1}{m}\sum_{g=1}^{m} P_g(\mathbf{x}_i)$

- K: The kernel used for GPR regression.

- σ_f^2: Hyperparameter 1. Variance.

- ℓ^2: Hyperparameter 2. Length scale.

- σ_w^2: Hyperparameter 3. Variance of white noise.

- σ_n^2: Observational noise. Calculated as the variance of historical precipitation anomaly from long term mean.

- $\mathcal{C}$: Prior covariance matrix. The covariance matrix calculated from the other GCMs. $\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j) = cov(P(\mathbf{x}_i), P(\mathbf{x}_j)) = \frac{1}{m-1}\sum_{g=1}^{m}(P_g(\mathbf{x}_i) - \bar{P}(\mathbf{x}_i))(P_g(\mathbf{x}_j) - \bar{P}(\mathbf{x}_j))$

B. The Regression

A Gaussian process is defined as;

$$p(f|\mathbf{x}) = \mathcal{N}\big(f|\mathcal{M}, K\big) \tag{1}$$

The joint distribution of seen and unseen data can be written as;

$$\begin{bmatrix} f(\mathbf{x}) \\ f(\mathbf{x}_*) \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mathcal{M}(\mathbf{x}) \\ \mathcal{M}(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{x},\mathbf{x}) + \sigma_n^2 I & K(\mathbf{x},\mathbf{x}_*) \\ K(\mathbf{x}_*,\mathbf{x}) & K(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix} \right) \tag{2}$$

1. Kernel

Where the kernel, K is defined as;

$$K(x_i, x_j) = \sigma_f^2 e^{\left(-\frac{1}{2\ell^2}(x_i - x_j)^2\right)} + \sigma_w^2 I + \mathcal{C}(x_i, x_j) \tag{3}$$

The posterior mean is calculated as;

$$f(\mathbf{x}_*) = \mathcal{M}(\mathbf{x}_*) + K(\mathbf{x}_*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}[f(\mathbf{x}) - \mathcal{M}(\mathbf{x})] \tag{4}$$

The uncertainty in the predicted mean is given by;

$$\epsilon^2 = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}, \mathbf{x}_*)[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1} K(\mathbf{x}_*, \mathbf{x}) \tag{5}$$

2. Hyperparameter Optimization

The hyperparameters, σ_f^2, ℓ^2, and σ_w^2 are obtained by maximizing the sum of log marginal likelihood in the training dataset.

The log marginal likelihood is calculated 33 times. In each iteration one of the 33 CMIP6 models is used as the OOS GCM and the other 32 is considered other GCMs. As mentioned above, other GCMs are used to make prior mean and covariance.

The summed log marginal likelihood is given by;

$$\begin{aligned} sml = -\frac{1}{2}\sum_{g=1}^{m}(f_g(\mathbf{x}) - \mathcal{M}_g(\mathbf{x}))^T[K_g(\mathbf{x}, \mathbf{x}) + \sigma_{n,g}^2 I]^{-1}(f_g(\mathbf{x}) - \mathcal{M}_g(\mathbf{x})) \\ - log|K_g(\mathbf{x}, \mathbf{x}) + \sigma_{n,g}^2 I| \end{aligned} \tag{6}$$

II. MULTIPLE INPUT SINGLE OUTPUT GAUSSIAN PROCESS REGRESSION

- GPR takes multiple inputs and computes one output (Bijl 2019, Section 2.4.1). Examples can be seen in works like Gilford et al. (2020), Hay et al. (2015).

- Each grid has a GPR model that takes into account the spatio-temporal covariance around the grid point.

- Three independent variables $\mathbf{x}$ (year, longitude, latitude) used to predict one dependent variable $f(\mathbf{x})$ (5 year mean anomaly of annual precipitation from historical period).

- A total of 33 CMIP6 simulations were used to design the regressor.

- Leave one out strategy is used to validate the method. i.e, in each iteration, one of the 33 models are used as "training data" and the other 32 is used to construct the priors.

- The hyperparameters are optimized by maximizing the sum of log marginal likelihood for the training period.

A. Parameter Definitions

- $\mathbf{X} = (\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3)$: Independent variable for seen observations. In this case *year, longitude, latitude.*

- $f(\mathbf{X})$: Seen observations. Mean precipitation anomaly at the central grid in the out of sample GCM (OOS GCM).

- $\mathbf{X}_*$: Years, longitude, latitude for which we are making predictions. When making predictions for a location, longitude and latitude remains the same as years changes.

- $f(\mathbf{X}_*)$: Unseen or predicted precipitation anomaly.

- $\mathcal{M}$: Prior mean. Calculated as the multimodel mean regional precipitation anomaly from other GCMs (all the other 32 GCMs except the OOS GCM).
  $\mathcal{M}(\mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3k}) = \frac{1}{m}\sum_{g=1}^{m} P_g(\mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3k})$.

- K: The kernel used for GPR regression.

- ν: Variance parameter for the linear kernel.

- σ_f^2: Hyperparameter 1. Variance.

- ℓ_1^2: Hyperparameter 2. Length scale along the time dimension.

- ℓ_2^2: Hyperparameter 3. Length scale along longitude.

- ℓ_3^2: Hyperparameter 4. Length scale along latitude.

- σ_w^2: Hyperparameter 5. Variance of white noise.

- σ_n^2: Observational noise. Calculated as the variance of historical precipitation anomaly from long term mean.

- $\mathcal{C}$: Prior covariance matrix. The covariance matrix calculated from the other GCMs.
$\mathcal{C}(x_{1i}, x_{2j}, x_{3k}, x_{1p}, x_{2q}, x_{3r}) = cov(P(x_{1i}, x_{2j}, x_{3k}), P(x_{1p}, x_{2q}, x_{3r}))$
$= \frac{1}{m-1} \sum_{g=1}^{m} (P_g(x_{1i}, x_{2i}, x_{3i}) - \bar{P}(x_{1i}, x_{2i}, x_{3i}))(P_g(x_{1p}, x_{2q}, x_{3r}) - \bar{P}(x_{1p}, x_{2q}, x_{3r}))$

B. The Regression

A Gaussian process is defined as;

$$p(f|\mathbf{X}) = \mathcal{N}\big(f|\mathcal{M}, K\big) \tag{7}$$

The joint distribution of seen and unseen data can be written as;

$$\begin{bmatrix} f(\mathbf{X}) \\ f(\mathbf{X}_*) \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mathcal{M}(\mathbf{X}) \\ \mathcal{M}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \tag{8}$$

1. *Kernel*

The definition of the kernel opens up possibilities for us to incorporate more prior knowledge into the regression models.

A simple anisotropic version of the kernel that uses a squared exponential covariance function (Rasmussen and Williams 2005, Equation 4.9) with different values of length scales can be written as;

$$K(x_1, x_2, x_3) = S(x_1, x_2, x_3, \sigma_f, \ell_1, \ell_2, \ell_3) + \sigma_w^2 I + \mathcal{C}(x_1, x_2, x_3) \tag{9}$$

The above equation can be expanded to its full form as;

$$\begin{aligned} K(x_{1i}, x_{2j}, x_{3k}, x_{1p}, x_{2q}, x_{3r}) = \sigma_f^2 e^{-\frac{1}{2}\left(\frac{(x_{1i}-x_{1p})^2}{\ell_1^2} + \frac{(x_{2j}-x_{2q})^2}{\ell_2^2} + \frac{(x_{3k}-x_{3r})^2}{\ell_3^2}\right)} \\ + \sigma_w^2 I + \mathcal{C}(x_{1i}, x_{2j}, x_{3k}, x_{1p}, x_{2q}, x_{3r}) \end{aligned} \tag{10}$$

This also opens up the possibility of combining different kernels along different dimensions as the relationship we intend to capture might be different across the dimensions Gilford et al. (2020), Hay et al. (2015).

A combination of particular interest is a kernel that uses a linear kernel (L) in the time dimension and a Matern kernel (M) in the spatial dimensions. The kernel can be written as;

$$K(x_1, x_2, x_3) = L(x_1, \nu_1) * M(x_2, x_3, \sigma_f, \ell_2, \ell_3) + \sigma_w^2 I + \mathcal{C}(x_1, x_2, x_3) \tag{11}$$

Where ν_1 is the variance parameter for the linear kernel.

The posterior mean is calculated as;

$$f(\mathbf{X}_*) = \mathcal{M}(\mathbf{X}_*) + K(\mathbf{X}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1}[f(\mathbf{X}) - \mathcal{M}(\mathbf{X})] \tag{12}$$

The uncertainty in the predicted mean is given by;

$$\epsilon^2 = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}, \mathbf{X}_*)[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1} K(\mathbf{X}_*, \mathbf{X}) \tag{13}$$

2. Hyperparameter Optimization

The hyperparameters, σ_f^2, ℓ_1^2, ℓ_2^2, ℓ_3^2, and σ_w^2 are obtained by maximizing the sum of log marginal likelihood in the training dataset.

The log marginal likelihood is calculated 33 times. In each iteration one of the 33 CMIP6 models is used as the OOS GCM and the other 32 is considered other GCMs. As mentioned above, other GCMs are used to make prior mean and covariance.

The summed log marginal likelihood is given by;

$$\begin{aligned} sml = -\frac{1}{2}\sum_{g=1}^{m} (f_g(\mathbf{X}) - \mathcal{M}_g(\mathbf{X}))^T [K_g(\mathbf{X}, \mathbf{X}) + \sigma_{n,g}^2 I]^{-1} (f_g(\mathbf{X}) - \mathcal{M}_g(\mathbf{X})) \\ - log|K_g(\mathbf{X}, \mathbf{X}) + \sigma_{n,g}^2 I| \end{aligned} \tag{14}$$

III. MULTITASK GAUSSIAN PROCESS REGRESSION

- GPR takes multiple inputs and computes multiple output (Bijl 2019, Section 2.4.2). Examples of implementation can be found in Bonilla et al. (2007), Chen et al. (2023).

- The main difference is that, the kernel considers the covariance between the two outputs in addition to the independent variables. In this regard it is similar to the initial version of multivariate GPR that we implemented.

H. Bijl. *Gaussian Process Regression Techniques With Applications to Wind Turbines; A student version of the thesis*. PhD thesis, 2019.

E. V. Bonilla, K. Chai, and C. Williams. Multi-task gaussian process prediction. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL https://proceedings.neurips.cc/paper_files/paper/2007/file/66368270ffd51418ec58bd793f2d9b1b-Paper.pdf.

Z. Chen, J. Fan, and K. Wang. Multivariate gaussian processes: definitions, examples and applications. *METRON*, 81(2):181–191, Jan. 2023. ISSN 2281-695X. doi:10.1007/s40300-023-00238-3. URL http://dx.doi.org/10.1007/s40300-023-00238-3.

D. M. Gilford, E. L. Ashe, R. M. DeConto, R. E. Kopp, D. Pollard, and A. Rovere. Could the last interglacial constrain projections of future antarctic ice mass loss and sea-level rise? *Journal of Geophysical Research: Earth Surface*, 125(10), Oct. 2020. ISSN 2169-9011. doi:10.1029/2019jf005418. URL http://dx.doi.org/10.1029/2019JF005418.

C. C. Hay, E. Morrow, R. E. Kopp, and J. X. Mitrovica. Probabilistic reanalysis of twentieth-century sea-level rise. *Nature*, 517(7535):481–484, Jan. 2015. ISSN 1476-4687. doi:10.1038/nature14093. URL http://dx.doi.org/10.1038/nature14093.

M. Lickley and S. Fletcher. Bayesian estimation of advanced warning time of precipitation emergence. *Earth's Future*, 12(2), Jan. 2024. ISSN 2328-4277. doi:10.1029/2023ef004079. URL http://dx.doi.org/10.1029/2023EF004079.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Nov. 2005. ISBN 9780262256834. doi:10.7551/mitpress/3206.001.0001. URL http://dx.doi.org/10.7551/mitpress/3206.001.0001.