Bayesian Composite Gaussian Processes

Casey B. Davis

The Ohio State University

davis.3266@osu.edu

February 20, 2014

Overview

1. Introduction: Overview of Gaussian Processes (GPs)
2. Other Methods: Universal Kriging and Composite Gaussian Processes
3. Bayesian Composite Gaussian Processes
4. Examples: A Comparison of CGP and BCGP Through Examples

Definition and Basic Properties of GPs

- A stochastic process, $Y(\mathbf{x}), \mathbf{x} \in \mathcal{X} \subset \Re^d$ with underlying probability space $(\Omega, \mathcal{B}, P)$, is a Gaussian process (GP) if, for any $\mathbf{x}_1, \ldots, \mathbf{x}_n, n \geq 1$ in $\mathcal{X}$, the joint distribution of the vector $\mathbf{Y} = \left(Y\left(\mathbf{x}_1\right), \ldots, Y\left(\mathbf{x}_n\right)\right)^\top$ has a multivariate normal distribution

$$\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, \mathbf{C}\right),$$

  where $\boldsymbol{\mu} = \left(\mu\left(\mathbf{x}_1\right), \ldots, \mu\left(\mathbf{x}_n\right)\right)^\top$, and $\mathbf{C}$ is an $n \times n$ covariance matrix created by a (valid) covariance function, $C\left(\mathbf{x}, \mathbf{x}'\right) = Cov\left(Y\left(\mathbf{x}\right), Y\left(\mathbf{x}'\right)\right)$, such that $C_{ij} = Cov\left(Y\left(\mathbf{x}_i\right), Y\left(\mathbf{x}_j\right)\right)$.
- A GP is fully specified by its mean function, $\mu(\mathbf{x}) = E\left(Y\left(\mathbf{x}\right)\right)$, and its (valid) covariance function.
- It is also common to work with (valid) correlation functions, $R\left(\mathbf{x}, \mathbf{x}'\right) = Cor\left(Y\left(\mathbf{x}\right), Y\left(\mathbf{x}'\right)\right)$

Common Correlation Functions

- The power exponential family has the form

$$R(\mathbf{h}) = exp\left(-\sum_{j=1}^{d}\theta_j|h_j|^{p_j}\right), 0 < p_j \leq 2 \text{ and } \theta_j > 0, j = 1, \ldots, d.$$

- A special case of the power exponential family is when $p_j = 2, j = 1, \ldots, d$. This is the *Gaussian* correlation function

$$R(\mathbf{h}) = exp\left(-\sum_{j=1}^{d}\theta_j h_j^2\right)$$

- The Gaussian correlation function leads to smooth sample paths that are continuous and infinitely differentiable.

Properties of GPs

- A covariance funtion is *stationary* if, for any translation $\mathbf{h} \in \Re^d$ such that $\mathbf{x}+\mathbf{h}, \mathbf{x}'+\mathbf{h} \in \mathcal{X}$, $C(\mathbf{x}, \mathbf{x}') = C(\mathbf{x}+\mathbf{h}, \mathbf{x}'+\mathbf{h})$.
- A GP is *stationary* if, for any $\mathbf{x}_1, \ldots, \mathbf{x}_n, n \geq 1$ in $\mathcal{X}$, and any translation $\mathbf{h} \in \Re^d$ such that $\mathbf{x}_1+\mathbf{h}, \ldots, \mathbf{x}_n+\mathbf{h} \in \mathcal{X}$, $\left(Y\left(\mathbf{x}_1\right), \ldots, Y\left(\mathbf{x}_n\right)\right)^\top$ and $\left(Y\left(\mathbf{x}_1+\mathbf{h}\right), \ldots, Y\left(\mathbf{x}_n+\mathbf{h}\right)\right)^\top$ have the same mean vector and covariance matrix.
- In this case, $Var\left(Y(\mathbf{x})\right) = \sigma^2 \; \forall \, \mathbf{x} \in \mathcal{X}$. Then $C(\cdot,\cdot) = \sigma^2 R(\cdot,\cdot)$.

Properties of GPs

- A GP is *nonstationary* if either the mean function is not constant or the covariance function is nonstationary.
- A common technique for generating a nonconstant mean function is to let the mean depend on $\mathbf{x}$ much like a regression model. This process has the form

$$Y(\mathbf{x}) = \sum_{i=1}^{p} f_i(\mathbf{x})\beta_i + Z(\mathbf{x}) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}),$$

  where $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_p(\mathbf{x}))^\top$ is a vector of known regression functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown regression coefficients, and $Z(\mathbf{x})$ is a zero-mean stationary Gaussian process.
- A GP may also be nonstationary if the covariance between two locations depends not only on the orientation and distance between the points, but on the location of the points in $\mathcal{X}$.

Example of a Stationary GP *vs.* a Nonstationary GP

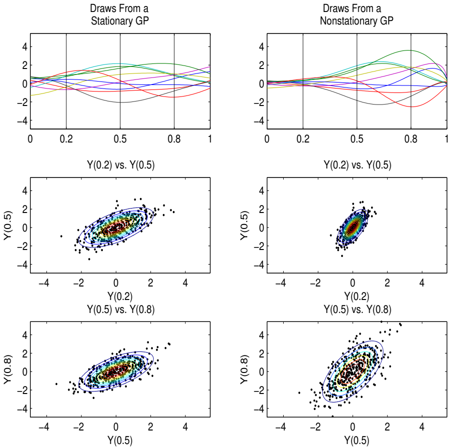

Figure : *This figure shows 10 draws from each of a stationary GP and nonstationary GP, along with the values of $Y(0.2)$ and $Y(0.5)$ and $Y(0.5)$ and $Y(0.8)$ plotted against each other for 400 draws from each process.*

Bayesian Approaches to Modeling the Mean Function

- A fully Bayesian approach to modeling with Gaussian processes involves setting a prior for the mean function, $\mu(\mathbf{x})$, assuming a covariance function, $C(\mathbf{x}, \mathbf{x}')$, and assuming priors for the hyperparameters involved in $C(\cdot, \cdot)$.
- A mean function is often specified as a constant, $\mu(\mathbf{x}) = \mu$, or a linear model

$$\mu(\mathbf{x}) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\beta},$$

  where $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_p(\mathbf{x}))^\top$ is a vector of known regression functions and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown coefficients.
- The priors on the parameters in this mean function are usually an improper uniform or a normal distribution.

Bayesian Approaches to Modeling the Correlation Function

- A covariance function must be assumed. Often, the Gaussian correlation function is assumed because of its smoothness properties.
- In this case the correlation parameters are often given Gamma priors.
- Improper priors on the correlation parameters often produce and improper posterior distribution.
- If the Gaussian correlation function is stationary ($C(\cdot,\cdot) = \sigma^2 R(\cdot,\cdot)$), then σ^2 is often given an Inverse Gamma prior.

Computational Methods

- Inferences are made using the posterior distribution, $(\cdot \mid \mathbf{y})$, where $\mathbf{y} = (y(\mathbf{x}_1), \ldots, (\mathbf{x}_n))^\top$ is the observed training data.
- When the posterior is difficult to work with directly, Markov Chain Monte Carlo (MCMC) methods can be used to sample from the posterior.
- The inversion of the $n \times n$ covariance matrix of the training data, $\mathbf{C}$, causes problems, particularly from being ill-conditioned (oftentimes overcome by adding a small nugget, σ_ϵ^2, to the diagonal elements) or from being large (it is computationally expensive to invert a large matrix).

Nonstationary GPs

- Stationary GP models are fairly common, even when they are not appropriate.
- Stationarity is often a strong assumption to make. It is not uncommon for the mean and/or the variance to change throughout the input space.
- A nonstationary GP allows for one or both of the mean and covariance to vary across the input space, $\mathcal{X} \subset \Re^d$, where d is the number of dimensions of $\mathbf{x}$.

Specification and Prediction for the Regression plus Stationary GP Model

- A common method for specifying a nonstationary GP is

$$Y(\mathbf{x}) = \sum_{i=1}^{p} f_i(\mathbf{x})\beta_i + Z(\mathbf{x}) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}),$$

  where $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_p(\mathbf{x}))^\top$ is a vector of known regression functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown regression coefficients, and $Z(\mathbf{x})$ is a zero-mean stationary GP.
- The best linear unbiased predictor is

$$\widehat{y}_{UK}(\mathbf{x}^*) = \mathbf{f}_*^\top \widehat{\boldsymbol{\beta}} + \mathbf{r}^\top(\mathbf{x}^*)\,\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\widehat{\boldsymbol{\beta}}),$$

  where $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n))^\top$, $\mathbf{f}_* = (f_1(\mathbf{x}^*), \ldots, f_p(\mathbf{x}^*))^\top$, $\mathbf{R}$ is a correlation matrix for the training data,
  $\mathbf{r}(\mathbf{x}^*) = (R(\mathbf{x}^* - \mathbf{x}_1), \ldots, R(\mathbf{x}^* - \mathbf{x}_n))^\top$, $\widehat{\boldsymbol{\beta}} = \left(\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{y}$, the usual generalized least squares predictor.

Example of the Regression plus Stationary GP Model

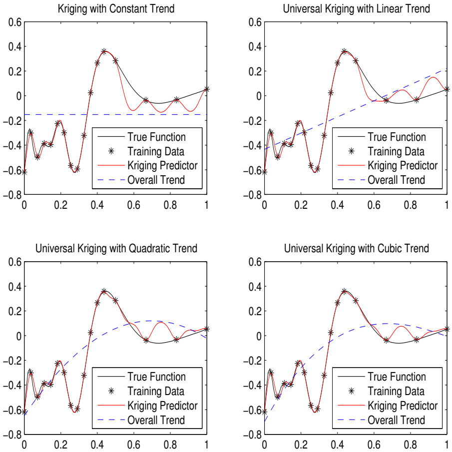

Figure : *This figure shows an example of the regression plus GP approach.*

Strenths and Weaknesses of the Regression plus Stationary GP Model

Strengths:

- This model allows for a nonstationary mean.
- A correctly specified trend can make predictions accurate.
- Relatively easy computationally.

Weaknesses:

- This model does not allow for a nonstationary covariance.
- The correct trend is generally unknown. A wrongly specified trend can make for very bad predictions.

CGP Description

- The CGP model allows for a flexible, non-constant mean and a nonstationary covariance.
- This model involves two processes, a “global” process and a local “process,” and a “volatility function.”
- The global process is smooth and stationary and captures the overall trend. It acts as the mean of the process. The local process makes local adjustments to the overall trend.
- The volatility function allows the volatility of the local process to change throughout the input space.

CGP Model

- Given the process parameters, $\mathbf{\Lambda}$, the CGP model is expressed as a sum of two Gaussian processes as follows:

$$Y(\mathbf{x}) = Y_g(\mathbf{x}) + \sigma(\mathbf{x}) Y_\ell(\mathbf{x})$$
$$[Y_g(\mathbf{x}) \mid \mathbf{\Lambda}] \sim GP(\mu, \tau^2 g(\cdot))$$
$$[Y_\ell(\mathbf{x}) \mid \mathbf{\Lambda}] \sim GP(0, \ell(\cdot)),$$

  where $g(\cdot)$ and $\ell(\cdot)$ are Gaussian correlation functions with unknown correlation parameters $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$. Without loss of generality, write $\sigma^2(\mathbf{x}) = \sigma^2 v(\mathbf{x})$.
- $Y_g(\mathbf{x})$ is the global process, $Y_\ell(\mathbf{x})$ is the local process, and $v(\mathbf{x})$ is the volatility function.
- $Y_g(\mathbf{x})$ and $Y_\ell(\mathbf{x})$ are independent.

CGP Prediction

The best linear unbiased predictor of $Y(\mathbf{x}^*)$ given the observed training data $\mathbf{y} = (y(\mathbf{x}_1), \ldots, y(\mathbf{x}_n))^\top$ is

$$\begin{aligned}\widehat{y}_{CGP}(\mathbf{x}^*) &= \widehat{\mu} + \mathbf{C}_*^\top \mathbf{C}^{-1}(\mathbf{y} - \widehat{\mu}\mathbf{1}) \\ &= \widehat{\mu} + \left(\mathbf{g}(\mathbf{x}^*) + \lambda v^{1/2}(\mathbf{x}^*)\mathbf{V}^{1/2}\boldsymbol{\ell}(\mathbf{x}^*)\right)^\top \left(\mathbf{G} + \lambda \mathbf{V}^{1/2}\mathbf{L}\mathbf{V}^{1/2}\right)^{-1} \\ &\quad \times (\mathbf{y} - \widehat{\mu}\mathbf{1})\end{aligned}$$

where $\lambda = \frac{\sigma^2}{\tau^2} \in [0, 1]$, $\mathbf{V} = diag\{v(\mathbf{x}_1), \ldots, v(\mathbf{x}_n)\}$, and

$$\widehat{\mu} = \left(\mathbf{1}^\top \left(\mathbf{G} + \lambda \mathbf{V}^{1/2}\mathbf{L}\mathbf{V}^{1/2}\right)^{-1} \mathbf{1}\right)^{-1} \mathbf{1}^\top \left(\mathbf{G} + \lambda \mathbf{V}^{1/2}\mathbf{L}\mathbf{V}^{1/2}\right)^{-1} \mathbf{y}.$$

CGP Prediction

- This predictor can be broken into two pieces:

$$\widehat{y}_{CGP}(\mathbf{x}^*) = \widehat{y}_g(\mathbf{x}^*) + \widehat{y}_\ell(\mathbf{x}^*),$$

$$\widehat{y}_g(\mathbf{x}^*) = \widehat{\mu} + \mathbf{g}^\top(\mathbf{x}^*)\left(\mathbf{G} + \lambda\mathbf{V}^{1/2}\mathbf{L}\mathbf{V}^{1/2}\right)^{-1}(\mathbf{y} - \widehat{\mu}\mathbf{1})$$

$$\widehat{y}_\ell(\mathbf{x}^*) = \lambda v^{1/2}(\mathbf{x}^*)\ell^\top(\mathbf{x}^*)\mathbf{V}^{1/2}\left(\mathbf{G} + \lambda\mathbf{V}^{1/2}\mathbf{L}\mathbf{V}^{1/2}\right)^{-1}(\mathbf{y} - \widehat{\mu}\mathbf{1}).$$

- $\widehat{y}_g(\mathbf{x}^*)$ can be thought of as a “global” predictor and $\widehat{y}_\ell(\mathbf{x}^*)$ can be thought of as a “local” predictor.

CGP Examples

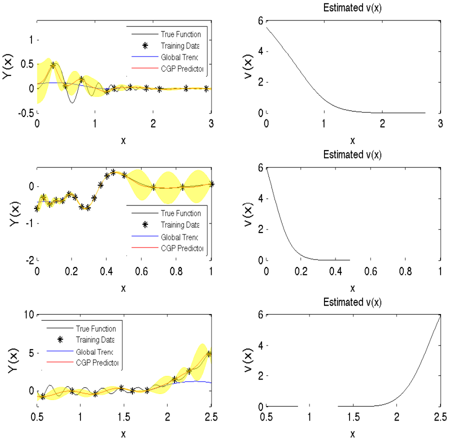

Figure : *This figure shows three plots of some non-stationary functions (black), along with training data for each, an overall prediction (red), and a prediction of the global trend (blue).*

Strengths and Weaknesses of the CGP Model

Strengths:

- Allows both the mean and covariance structure to vary throughout the input space.
- The correlation matrix is automatically well-conditioned due to the fact that the diagonal elements are automatically inflated.
- Does not require the addition of a "nugget," and so remains an interpolator.

Weaknesses:

- Does not handle noisy data well.
- Does not utilize prior knowledge about the process.

Nonstationary BCGP Description

- The nonstationary BCGP model allows for a flexible, non-constant mean and a nonstationary covariance.
- This model involves three processes, a "global" process, a local "process," and an "error" process.
- The global process is smooth and stationary and captures the overall trend. It acts as the mean of the process. The local process makes local adjustments to the overall trend. The error process allows for measurement error.
- The variance of the process changes throughout the input space.

Nonstationary BCGP Model

- This model treats the deterministic function, $y(\mathbf{x})$, as a realization of a Gaussian process, $Y(\mathbf{x})$. The model is as follows:

$$[Y(\mathbf{x}) \mid \boldsymbol{\Lambda}] \sim GP(\mu, C(\cdot, \cdot)),$$

where $\boldsymbol{\Lambda} = (\mu, w, \boldsymbol{\theta}, \boldsymbol{\kappa}, \sigma_\epsilon^2, \phi(\cdot), \mu_V, \boldsymbol{\theta}_V)^\top$, μ is an overall mean, and the covariance function is

$$C(Y(\mathbf{x}), Y(\mathbf{x}')) = \begin{cases} \sqrt{\phi(\mathbf{x})\phi(\mathbf{x}')}(wg(\mathbf{x} - \mathbf{x}') + \\ (1 - w)\ell(\mathbf{x} - \mathbf{x}')) & , \mathbf{x} \neq \mathbf{x}' \\ \phi(\mathbf{x})(1 + \sigma_\epsilon^2) & , \mathbf{x} = \mathbf{x}' \end{cases} .$$

- $\phi(\mathbf{x})$ is a positive function that allows the variance to change throughout the input space, and the other parameters are the same as in the stationary model.

An Equivalent Model Description

- The model may be rewritten as follows:

$$Y(\mathbf{x}) = Y_g(\mathbf{x}) + Y_\ell(\mathbf{x}) + Y_{\sigma_\epsilon^2}(\mathbf{x})$$

where $Y_g(\mathbf{x})$ is a mean μ Gaussian process with covariance

$$C_g(Y_g(\mathbf{x}), Y_g(\mathbf{x}')) = \begin{cases} \sqrt{\phi(\mathbf{x})\phi(\mathbf{x}')}wg(\mathbf{x}-\mathbf{x}') & , \mathbf{x} \neq \mathbf{x}' \\ \phi(\mathbf{x})w & , \mathbf{x} = \mathbf{x}' \end{cases},$$

$Y_\ell(\mathbf{x})$ is a mean 0 Gaussian process with covariance

$$C_\ell(Y_\ell(\mathbf{x}), Y_\ell(\mathbf{x}')) = \begin{cases} \sqrt{\phi(\mathbf{x})\phi(\mathbf{x}')}(1-w)\ell(\mathbf{x}-\mathbf{x}') & , \mathbf{x} \neq \mathbf{x}' \\ \phi(\mathbf{x})(1-w) & , \mathbf{x} = \mathbf{x}' \end{cases},$$

and $Y_{\sigma_\epsilon^2}(\mathbf{x})$ is a mean zero Gaussian (white noise) process and covariance

$$C_{\sigma_\epsilon^2}(Y_{\sigma_\epsilon^2}(\mathbf{x}), Y_{\sigma_\epsilon^2}(\mathbf{x}')) = \begin{cases} 0 & , \mathbf{x} \neq \mathbf{x}' \\ \phi(\mathbf{x})\sigma_\epsilon^2 & , \mathbf{x} = \mathbf{x}' \end{cases},$$

where $Y_g(\mathbf{x})$, $Y_\ell(\mathbf{x})$, and $Y_{\sigma_\epsilon^2}(\mathbf{x})$ are mutually independent.

Nonstationary BCGP Model Priors

The priors for this for this model will be assumed to be:

$$
\begin{aligned}
p(\mu) &\propto 1 \\
w &\sim Truncated\ Beta(a, b, \xi, \psi) \\
\kappa_i &\sim Gamma(\gamma_i, \eta_i),\ \kappa_1, \ldots, \kappa_d \text{ mut. indep.} \\
\theta_i \mid \kappa_i &\sim Truncated\ Beta(0, \kappa_i, \nu_i, \omega_i),\ \theta_1, \ldots, \theta_d \text{ mut. indep.} \\
\sigma^2_\epsilon &\sim Gamma(\delta, \lambda) \\
Log\ \phi(\mathbf{x}) \mid \mu_V, \boldsymbol{\theta}_V &\sim GP(\mu_V, g_{\theta_V}(\cdot)) \\
\mu_V &\sim N(\mu_{\mu_V}, \sigma^2_{\mu_V}) \\
\theta_{V_i} &\sim Gamma(\alpha_{\theta_{V_i}}, \beta_{\theta_{V_i}}),\ \theta_{V_1}, \ldots, \theta_{V_d} \text{ mut. indep.}
\end{aligned}
$$

Examples of Draws from This Process

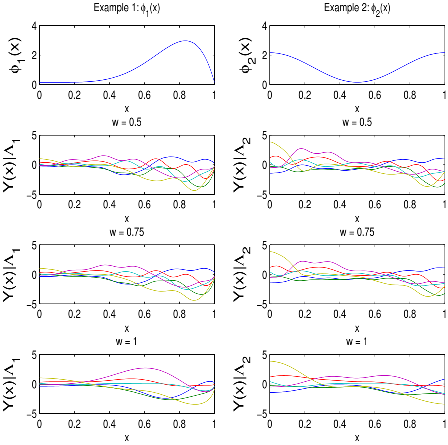

Figure : *This figure shows two different fixed* $\phi(\mathbf{x})$ *functions and six draws from this process for each* $w \in \{0.5, 0.75, 1\}$.

An Example Showing $Y(\mathbf{x})$ as the Sum of the Global, Local, and Error Processes

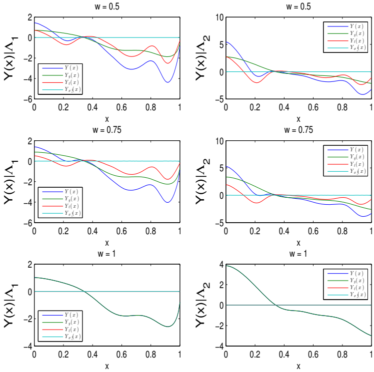

Figure : *This figure shows a draw from the process (blue) with two different fixed* $\phi(\mathbf{x})$ *functions for each* $w \in \{0.5, 0.75, 1\}$*, along with the global (green), local (red), and error (cyan) processes.*

Computational Methods

- The posterior distribution is used to make inferences. Let $\boldsymbol{\Lambda} = \left(\mu, w, \boldsymbol{\theta}, \boldsymbol{\kappa}, \sigma_\epsilon^2, \mathbf{V}, \mu_V, \boldsymbol{\theta}_V\right)^\top$, where $\mathbf{V} = (\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_n))^\top$.

$$
\begin{aligned}
[\boldsymbol{\Lambda} \mid \mathbf{y}] &= \frac{[\mathbf{y} \mid \boldsymbol{\Lambda}]\,[\boldsymbol{\Lambda}]}{[\mathbf{y}]} \\
&\propto [\mathbf{y} \mid \boldsymbol{\Lambda}]\,[\boldsymbol{\Lambda}] \\
&= [\mathbf{y} \mid \boldsymbol{\Lambda}]\,[\mu]\,[w]\left[\sigma_\epsilon^2\right] \prod_{i=1}^{d} [\theta_i \mid \kappa_i]\,[\kappa_i]\,[\mathbf{V} \mid \mu_V, \boldsymbol{\theta}_V]\,[\mu_V] \prod_{i=1}^{d} [\theta_{V_i}]
\end{aligned}
$$

- However, this posterior distribution is difficult to work with directly. In particular, it is difficult to integrate over this posterior distribution.
- Use MCMC to sample from the posterior.

Point Predictions

- This process is a Gaussian process. The joint distribution of $Y(\mathbf{x}^*)$ and $\mathbf{Y}$ also follows a multivariate normal distribution as follows:

$$\left[\left(\begin{array}{c} Y(\mathbf{x}^*) \\ \mathbf{Y} \end{array}\right) \mid \boldsymbol{\Lambda}\right] \sim N_{1+n}\left[\left(\begin{array}{c} \mu \\ \mu\mathbf{1} \end{array}\right), \left(\begin{array}{cc} \phi(\mathbf{x}^*)(1+\sigma_\epsilon^2) & \mathbf{C}_*^\top \\ \mathbf{C}_* & \mathbf{C} \end{array}\right)\right],$$

- A Rao-Blackwellized minimum mean squared prediction error (MSPE) point prediction for $y(\mathbf{x}^*)$ given the data is

$$\begin{aligned} \widehat{y}_{BCGP}(\mathbf{x}^*) &= E(Y(\mathbf{x}^*)|\mathbf{y}) \\ &= E_{\Lambda|\mathbf{y}}(E(Y(\mathbf{x}^*)|\mathbf{y}, \Lambda)) \\ &= E_{\Lambda|\mathbf{y}}\left(\mu + \mathbf{C}_*^\top \mathbf{C}^{-1}(\mathbf{y} - \mu\mathbf{1})\right) \\ &\approx \frac{1}{n_{mcmc}} \sum_{i=1}^{n_{mcmc}} \left(\mu^{[i]} + \mathbf{C}_*^{[i]\top} \mathbf{C}^{[i]^{-1}}(\mathbf{y} - \mu^{[i]}\mathbf{1})\right), \end{aligned}$$

where n_{mcmc} is the number of draws taken from the posterior $[\boldsymbol{\Lambda} \mid \mathbf{Y}]$.

Equivalent Model Point Predictions

The MSPE predictor for the equivalent model can be shown to be:

$$\begin{aligned}\widehat{y}(\mathbf{x}^*) &= E(Y(\mathbf{x}^*)|\mathbf{y}) \\ &= E_{\Lambda|\mathbf{y}}(E(Y(\mathbf{x}^*)|\mathbf{y},\Lambda)) \\ &\approx \frac{1}{n_{mcmc}} \sum_{i=1}^{n_{mcmc}} \left(\widehat{y}_g^{[i]}(\mathbf{x}^*) + \widehat{y}_\ell^{[i]}(\mathbf{x}^*) + \widehat{y}_{\sigma_\epsilon^2}^{[i]}(\mathbf{x}^*) \right)\end{aligned}$$

where

$$\begin{aligned}\widehat{y}_g(\mathbf{x}^*) &= \mu + \mathbf{C}_{g_*}^\top \mathbf{C}^{-1}(\mathbf{y} - \mu\mathbf{1}) \\ \widehat{y}_\ell(\mathbf{x}^*) &= \mathbf{C}_{\ell_*}^\top \mathbf{C}^{-1}(\mathbf{y} - \mu\mathbf{1}) \\ \widehat{y}_{\sigma_\epsilon^2}(\mathbf{x}^*) &= \mathbf{C}_{\sigma_\epsilon^2{}_*}^\top \mathbf{C}^{-1}(\mathbf{y} - \mu\mathbf{1})\end{aligned}$$

Strengths and Weaknesses of the BCGP Model

Strengths:

- Allows both the mean and covariance structure to vary throughout the input space.
- Can be a near-interpolator for deterministic data by setting the priors on the nugget so that the nugget will always be small.
- Can handle noisy data by setting the priors on the nugget accordingly.
- Can utilize prior knowledge about the process.

Weaknesses:

- The priors on the correlation parameters can be difficult to specify.
- When the number of dimensions is large, the MCMC algorithm can be time-consuming.
- Currently, the software requires manual calibration of the proposal widths.

Interpolation: Example 1

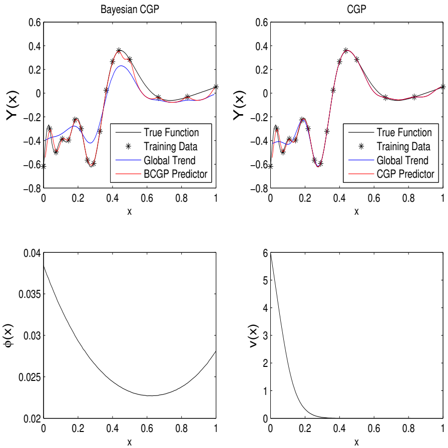

Figure : *Plots of the function* $y(x) = sin\left(30(x-0.9)^4\right) cos\left(2(x-0.9)\right) + \frac{(x-0.9)}{2}$ *with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

Interpolation: Example 2

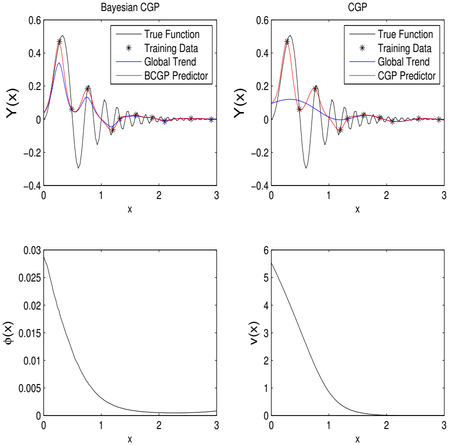

Figure : *Plots of the function* $y(x) = e^{-2x} sin(4\pi x^2)$ *with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

Interpolation: Example 3

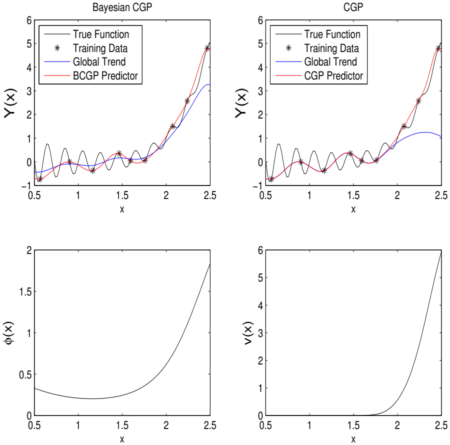

Figure : *Plots of the true function* $y(x) = \frac{sin(10\pi x)}{2x} + (x-1)^4$ *with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

Interpolation: Example 4

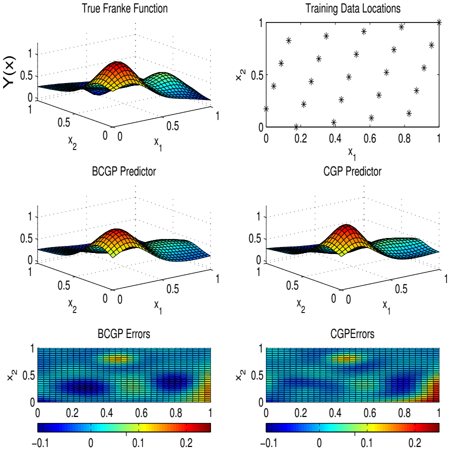

Figure : *Plots of the Franke function and the 24-run maximin Latin hypercube design, the BCGP and CGP predictors, and their respective error plots.*

Extrapolation: Example 1

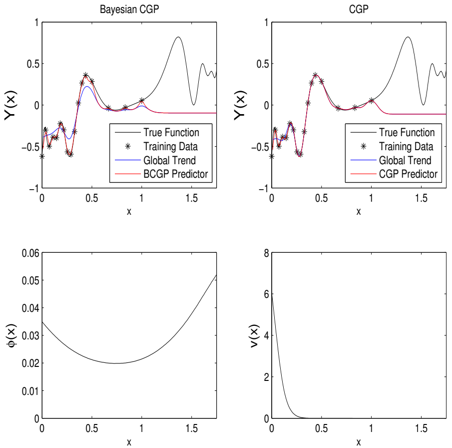

Figure : *Plots of the function* $y(x) = sin\left(30(x-0.9)^4\right) cos\left(2(x-0.9)\right) + \frac{(x-0.9)}{2}$ *for* $x \in [0, 1.75]$ *with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

Extrapolation: Example 2

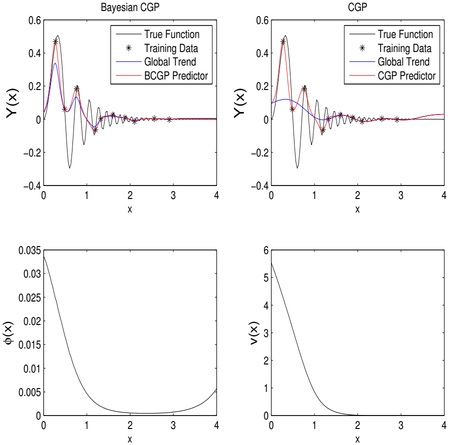

Figure : *Plots of the function* $y(x) = e^{-2x} sin(4\pi x^2)$ *for* $x \in [0, 4]$ *with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

100 Draws from a Stationary Process Using Relatively Uninformative Priors

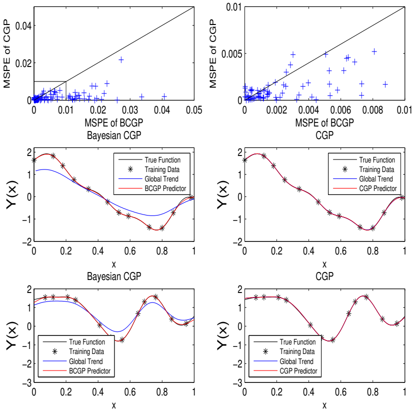

Figure : *MSPEs of the two predictors for 100 draws from a stationary process, along with two example draws. The priors on w, θ, κ, and θ_V were given the default parameters, making the priors relatively uninformative.*

100 Draws from a Stationary Process Using Relatively Informative Priors

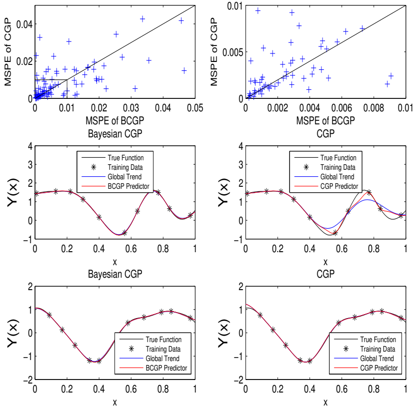

Figure : *MSPEs of the two predictors for 100 draws from a stationary process, along with two example draws. The priors on w, θ, and κ were specified so that the ranges of likely values for w, θ, and κ are very narrow.*

Example with Noisy Data

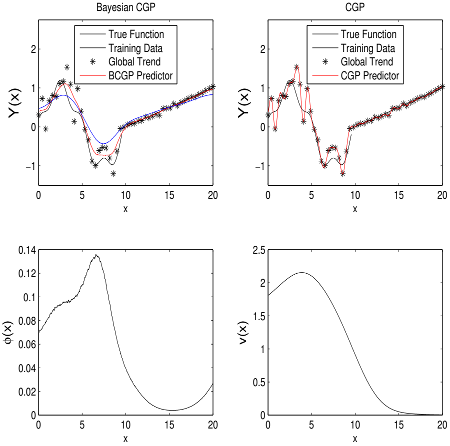

Figure : *Plots of the true function with the training data and the BCGP and CGP global(blue) and overall(red) predictors.*

Thank You