**Advances in Applied Mathematics** 应用数学进展, 2023, 12(11), 4874-4883
Published Online November 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.1211480

# 多任务高斯过程回归的模型平均

**魏馨忆，邹晨晨***

青岛大学数学与统计学院，山东 青岛

收稿日期：2023年10月28日；录用日期：2023年11月23日；发布日期：2023年11月28日

## 摘 要

模型平均因其稳健性好、预测精度高等优点在当代统计学和计量经济学界有广泛应用。多任务高斯过程回归同时考虑了输入变量之间和输出变量之间的相关性，能够得到更为准确的预测值。考虑到输出之间的相关性，本文先构建多任务高斯过程回归模型(MGPR)，再进行模型平均及选择；不考虑输出之间的相关性时，分别构建单任务高斯过程回归(GPR)，再进行模型平均及选择。数值模拟结果表明，输出噪声为同方差时，MGPR的模型平均比GPR具有更小的损失风险，且MGPR的计算效率相对较高。输出噪声同方差时，采用中国少数民族地区城市化率、非农业人口比例及受教育年限的实际数据；输出噪声异方差时，采用苄醇转化率和转频率的实际数据，均表明MGPR的模型平均具有更小的损失风险。

## 关键词

多任务高斯过程回归，模型平均，损失风险，计算效率

# Model Averaging for Multi-Task Gaussian Process Regression

**Xinyi Wei, Chenchen Zou***

Department of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Oct. 28th, 2023; accepted: Nov. 23rd, 2023; published: Nov. 28th, 2023

## Abstract

**Model averaging, owing to its advantages of robustness and high predictive accuracy, finds extensive application in contemporary statistics and econometrics. Multi-task Gaussian process regression, considering both the interdependencies among input variables and the correlations among output variables, enables more precise predictions. Taking into account the inter-output correla-**

*通讯作者。

文章引用: 魏馨忆, 邹晨晨. 多任务高斯过程回归的模型平均[J]. 应用数学进展, 2023, 12(11): 4874-4883.
DOI: 10.12677/aam.2023.1211480

**tions, this paper constructs a Multi-Task Gaussian Process Regression model (MGPR) followed by model averaging and selection. Conversely, when disregarding inter-output correlations, single-task Gaussian Process Regression (GPR) models are individually constructed, followed by the same process of model averaging and selection. Numerical simulation results demonstrate that when the output noise exhibits homoscedastic variance, MGPR's model averaging exhibits lower loss risk compared to GPR, with relatively higher computational efficiency. For the case of homoscedastic output noise, actual data on urbanization rates in China's ethnic minority regions, non-agricultural population ratios, and years of education are employed. In the case of heteroscedastic output noise, real data on benzyl alcohol conversion rates and turnover frequencies are utilized, both indicating that MGPR's model averaging yields lower loss risk.**

## Keywords

**Multi-Task Gaussian Process Regression, Model Averaging, Loss Risk, Computational Efficiency**



## 1. 引言

多任务高斯过程能够处理多个相关输出变量，提高模型的灵活性和准确性。其通过协方差函数描述输出间的相关性，并共享相关性结构来获取信息。然而，单一的多任务高斯过程模型无法覆盖所有数据特征，因此引入模型平均方法可以减少模型的不确定性，提高预测性能。模型平均方法通过组合多个基本模型，综合优点并改善预测稳定性。这种方法在多个领域得到广泛应用，为解决复杂问题提供了强大而可靠的工具。

当前学术研究主要集中在贝叶斯模型平均和频率模型平均这两个方向。Zhang J.和 Taflanidis A. A. [1]提出一种高效的 BMA 克里金回归模型，通过将全局趋势的回归与提供局部校正的高斯过程近似相结合来制定预测。除了贝叶斯模型平均，学术界还关注了频率模型平均方法。权重计算准则是频率模型平均方法的关键问题。Wang Miaomiao 等人[2]提到基于 AIC 和 BIC 的平滑 AIC (S-AIC)和平滑 BIC (S-BIC)准则，通过极小化这些信息准则来确定权重。Priyam Mitra 等人[3]在没有对候选模型集设置限制的情况下建立了通用的频率模型平均框架，并通过最小化均方误差来获得自适应权重。Zhongqi Liang 等人[4]提到通过极小化 Mallows 准则来确定模型平均的权重。并证明在宽松条件下，利用最小化 Mallows 准则所确定的权重具有渐近最优性。Shaobo Jin 和 Sebastian Ankargren [5]将 FMA 方法应用于连续数据的结构方程模型，并提出了有效的置信区间和 $\chi^2$ 检验统计量。Carroll R.等人[6]提出简约模型平均法，将权重 1 分配给误差最小的准确模型，使其他准确模型的权重为 0。Zhang X.等人[7]探讨 MMA 和 JMA 方法对参数模型平均估计的分布特性。

多任务高斯过程在统计学和概率论中占据重要地位。Zexun Chen 等人[8]提供了对多任务高斯过程的精确定义，并介绍了其基本属性。Chen Z.等人[9]介绍多任务高斯过程的关键属性，如平稳性和独立性。多任务高斯过程近几年在各个领域应用逐渐广泛，并在多个方面不断改进，以提高预测精度。Xiaodan Hong 等人[10]通过多任务高斯过程回归和依赖性多任务高斯过程回归，考虑了相关噪声的影响并提升了建模性能。R. Kontar 等人[11]提出了一种可扩展和正则化的方法，以减小多任务高斯过程中的负迁移。J. Zapata 等人[12]通过引入部分可分性概念和函数图模型，实现了对多任务高斯过程的建模。Rong Zhu

等人[13]在本文中提及了如何在多个相关因变量的情况下，用模型平均与选择的方法对建立的模型提高预测精度。Zexun Chen 等人[14]提及高斯过程中如何设置超参数。Pengfei Wei 等人[15]在结构系统可靠性的研究中，基于多任务输出高斯过程的主动学习来提高系统可靠性。ChenZ 等人[16]提出 MV-TPR，并与 MV-GPR 作比较，证明在所考虑的数据集上的 MV-TPR 的有效性。MickaB 等人[17]提出在高斯过程中用潜在变量解决异方差问题，并运用大量实例证明自己的方法最优。Hansen 等人[18]提出 JMA 方法，通过删除一个样本的交叉验证法最小化误差向量。Tang Qinghu 等人[19]设计了两个输出任务相关的试验。

本文的其余部分结构如下：在第 2 节中，我们介绍多任务高斯过程回归模型、核、超参数以及五种模型选择及模型平均方法；在第 3 节中，通过数值模拟构建 MGPR 及 GPR，并分别进行模型平均及选择，最后将以上方法运用到第六次人口普查中部分地区数据及苄醇转化率和转频率的统计研究中；第 4 节是对文章的总结和讨论。

本文在以上文献的基础上，考虑现实情况中多个输出任务具有相关性，将模型平均引入多任务高斯过程回归中，这是本文的创新点。

## 2. 多任务高斯过程及相关介绍

### 2.1. 多任务高斯过程模型

如果 $f$ 是 $X$ 上的多任务高斯过程，具有向量型均值函数 $\mu: X \mapsto R^d$，协方差函数(也称为核) $k: X \times X \mapsto R$ 且半正定参数矩阵 $\Omega \in R^{d\times d}$，那么 $f \sim MGP(\mu, k, \Omega)$。并且在此文中，考虑将噪声项包含在核函数内的模型。

给定 $n$ 对样本观测值 $\{(x_i, y_i)\}_{i=1}^{n}, x_i \in R^p, y_i \in R^{1\times d}$，我们假设模型为：

$$f \sim MGP(\mu, k', \Omega) \tag{1}$$

$$y_i = f(x_i), i = 1, \cdots, n \tag{2}$$

其中 $k' = k(x_i, x_j) + \delta_{ij}\sigma_n^2$，且当 $i = j$ 时，$\delta_{ij} = 1$，否则 $\delta_{ij} = 0$，上式中的第二项代表的是随机噪声项。

在 $X_* = [x_{n+1}, \cdots, x_{n+m}]^{\mathrm{T}}$ 处的预测变量为 $f_* = [f_{*1}, \cdots, f_{*m}]^{\mathrm{T}}$，训练集 $Y = [y_1^{\mathrm{T}}, \cdots, y_n^{\mathrm{T}}]^{\mathrm{T}}$ 和预测变量 $f_*$ 的联合分为

$$\begin{bmatrix} Y \\ f_* \end{bmatrix} \sim MGP\left(0, \begin{bmatrix} K'(X,X) & K'(X_*,X) \\ K'(X_*,X) & K'(X_*,X_*) \end{bmatrix}, \Omega\right) \tag{3}$$

其中 $K'(X,X)$ 是一个 $n\times n$ 矩阵，第 $(i,j)$ 个元素是 $K'(X,X)_{ij} = k'(x_i, x_j)$，且 $K'(X_*,X)$ 是 $m\times n$ 矩阵，第 $(i,j)$ 个元素是 $K'(X_*,X)_{ij} = k'(x_{n+i}, x_j)$，$K'(X_*,X_*)$ 是 $m\times m$ 矩阵，第 $(i,j)$ 个元素是 $K'(X_*,X_*)_{ij} = k'(x_{n+i}, x_{n+j})$。因此，根据多任务高斯过程的条件分布，其预测分布为 $p(f_* \mid X, Y, X_*) \sim MGP(\hat{\mu}, \hat{\Sigma}, \hat{\Omega})$，其中

$$\hat{\mu} = K'(X_*,X)^{\mathrm{T}} K'(X,X)^{-1} Y \tag{4}$$

$$\hat{\Sigma} = K'(X_*,X_*) - K'(X_*,X)^{\mathrm{T}} K'(X,X)^{-1} K'(X_*,X) \tag{5}$$

$$\hat{\Omega} = \Omega \tag{6}$$

此外，期望和协方差如下所示：

$$E\left|f_*\right| = \hat{\mu} = K'(X_*,X)^{\mathrm{T}} K'(X,X)^{-1} Y \tag{7}$$

$$cov\left(vec\left(f_*^{\mathrm{T}}\right)\right)=\hat{\Sigma}\otimes\hat{\Omega}=\left[K'\left(X_*,X_*\right)-K'\left(X_*,X\right)^{\mathrm{T}}K'\left(X,X\right)^{-1}\right]\otimes\Omega \tag{8}$$

### 2.2. 核介绍

上述回归模型中有两个协方差阵，列协方差阵 $k$ 与行协方差阵 $\Omega$，但只有列协方差取决于输入，并且被视为内核，因为它包含有关我们希望学习和定义数据点之间接近度和相似性的函数的假设，行协方差阵取决于输出，构建多任务高斯过程回归，既要考虑到列协方差在又要考虑到行协方差阵。与传统 GPR 一样，核的选择对多任务高斯过程回归的性能有深远的影响。现有方法中已经提出了许多有用的核，如线性核、有理二次核和 matern。内核还可以通过 ARD 定义：

$$k_{SEard}=s_f^2\exp\left(-\frac{\left(x-x'\right)^{\mathrm{T}}\Theta^{-1}\left(x-x'\right)}{2}\right) \tag{9}$$

其中 $s_f^2$ 指的是信号方差，也可以认为是输出尺度的振幅，$\Theta$ 是一个对角矩阵，具有组成元素为 $\left\{\ell_i^2\right\}_{i=1}^{p}$，它表示每个相应输入维度的长度尺度，参数 $\ell_i$ 是第 $i$ 个输入的尺度。

### 2.3. 超参数介绍

参数估计需要从训练数据中估计核所涉及的超参数和行协方差矩阵。传统 GP 模型中使用的许多方法，例如最大似然估计(MLE)、最大后验(MAP)和马尔可夫链蒙特卡罗(MCMC)，可用于我们提出的模型。尽管蒙特卡罗方法可以执行 GPR，而不需要估计超参数，但由于蒙特卡罗方法的计算成本较高，常用的方法是通过 MLE 来估计它们。因此，我们考虑使用 MLE 进行参数估计。与传统的 GPR 模型相比，行协方差阵 $\Omega$ 是一个额外的参数。因此，未知参数包括核中的参数，噪声方差 $\sigma_n^2$ 及行协方差参数矩阵 $\Omega$。

因为 $\Omega$ 是半正定矩阵，可以被分解为 $\Omega=\Phi\Phi^{\mathrm{T}}$，其中

$$\Phi=\begin{bmatrix}\phi_{11} & 0 & \cdots & 0\\ \phi_{21} & \phi_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \phi_{d1} & \phi_{d2} & \cdots & \phi_{dd}\end{bmatrix} \tag{10}$$

为了保证 $\Phi$ 的唯一性，对角元素被限制为正，且令 $\varphi_{ii}=\ln\left(\phi_{ii}\right)$，$i=1,2,\cdots,d$。

在 MV-GPR 模型中，$Y\sim MGP\left(0,K',\Omega\right)$，其中 $K'$ 是带有噪声项的列协方差矩阵，有 $K'_{ij}=k'\left(x_i,x_j\right)$，即为 $K'=K+\sigma_n^2I$，其中 $K$ 是不带噪声项的列协方差阵，有 $K_{ij}=k\left(x_i,x_j\right)$。内核中的参数表示为 $K=K_\theta$，超参数集为 $\Theta=\left\{\theta_1,\theta_2,\cdots\right\}$，因此，

$$\frac{\partial K'}{\partial\sigma_n^2}=I_n,\frac{\partial K'}{\partial\theta_i}=\frac{\partial K}{\partial\theta_i} \tag{11}$$

其似然函数可表示为 $L=\frac{nd}{2}\ln\left(2\pi\right)+\frac{d}{2}\ln\det\left(K'\right)+\frac{n}{2}\ln\det\left(\Omega\right)+\frac{1}{2}tr\left(\left(K'\right)^{-1}Y\Omega^{-1}Y^{\mathrm{T}}\right)$。

利用似然函数对超参数进行估计有：

$$\frac{\partial L}{\partial\sigma_n^2}=\frac{d}{2}tr\left(\left(K'\right)^{-1}\right)-\frac{1}{2}tr\left(\alpha_{K'}\Omega^{-1}\alpha_{K'}^{\mathrm{T}}\right) \tag{12}$$

$$\frac{\partial L}{\partial\theta_i}=\frac{d}{2}tr\left(\left(K'\right)^{-1}\frac{\partial K_\theta}{\partial\theta_i}\right)-\frac{1}{2}tr\left(\alpha_{K'}\Omega^{-1}\alpha_{K'}^{\mathrm{T}}\frac{\partial K_\theta}{\partial\theta_i}\right) \tag{13}$$

$$\frac{\partial L}{\partial \phi_{ij}}=\frac{n}{2}tr\left[\Omega^{-1}\left(E_{ij}\Phi^{\mathrm{T}}+\Phi E_{ji}\right)\right]-\frac{1}{2}tr\left(\alpha_{\Omega}\left(K'\right)^{-1}\alpha_{\Omega}^{\mathrm{T}}\left(E_{ij}\Phi^{\mathrm{T}}+\Phi E_{ji}\right)\right) \tag{14}$$

$$\frac{\partial L}{\partial \varphi_{ii}}=\frac{n}{2}tr\left[\Omega^{-1}\left(J_{ii}\Phi^{\mathrm{T}}+\Phi J_{ii}\right)\right]-\frac{1}{2}tr\left(\alpha_{\Omega}\left(K'\right)^{-1}\alpha_{\Omega}^{\mathrm{T}}\left(J_{ii}\Phi^{\mathrm{T}}+\Phi J_{ii}\right)\right) \tag{15}$$

其中 $\alpha_{K'}=\left(K'\right)^{-1}Y$，$\alpha_{\Omega}=\Omega^{-1}Y^{\mathrm{T}}$，$E_{ij}$ 是 $d\times d$ 单位阵，$J_{ii}$ 与 $E_{ij}$ 形式相同，但对角线上元素为 $\mathrm{e}^{\varphi_{ii}}$。因此，采用共轭梯度法用来最小化负对数似然函数，得到参数估计。且由于随机噪声被纳入核函数，噪声方差与其他超参数一起估计。

## 2.4. 模型选择及模型平均方法

实际情况中，输入变量之间及输出变量之间存在相关性。考虑输出任务之间不存在相关性时，构建 GPR 的模型平均；考虑输出任务之间的相关性时，构建 MGPR 的模型平均，我们想得到哪种模型的计算效率及预测准确率较高，故我们将这两种模型进行比较。

线性回归模型中应用较多的是以下几种传统的模型选择和模型平均的方法，本文为了研究分析各种模型平均及选择方法在 MGPR 及 GPR 下的预测结果，考虑了以下模型选择及模型平均方法：

$$\text{AICc:}\quad AIC_m=n\log\left(\sigma_m^2\right)+2n\left(k_m+p+1\right)/\left(n-k_m-p-2\right) \tag{16}$$

其中 $\sigma_m^2=\frac{1}{n}\left(Y-F_m\hat{\beta}_m\right)'\hat{C}^{-1}\left(Y-F_m\hat{\beta}_m\right)$，$P$ 为与自相关函数相关的参数个数，$C$ 为单任务输出或者多任务输出时的协方差矩阵。具有最小 AICc 值挑选出来作为最优模型。

$$\text{(2)}\quad w_m=\exp\left(-AIC_m/2\right)\Big/\sum_{i=1}^{M}\exp\left(-AIC_i/2\right)\text{，其中 }AIC_m=n\log\left(\hat{\sigma}_m^2\right)+2\left(k_m+p+1\right) \tag{17}$$

$$\text{(3)}\quad w_m=\exp\left(-BIC_m/2\right)\Big/\sum_{i=1}^{M}\exp\left(-BIC_i/2\right)\text{，其中 }BIC_m=n\log\left(\hat{\sigma}_m^2\right)+\log\left(n\right)\left(k_m+p+1\right) \tag{18}$$

(4) MMA：采用同方差线性回归模型提出的模型平均方法。MMA 方法通过极小化 Mallows 准则得到权重：

$$w=\arg\min C_n\left(w\right) \tag{19}$$

$$C_n\left(w\right)=\left\|A\left(w\right)y\right\|^2+2\sigma^2 tr\left(P\left(w\right)\right) \tag{20}$$

其中 $A_k=I_k-P_k$，$A\left(w\right)=\sum_{i=1}^{k}w_iA_i$，$P_k=X_k\left(X_k'X_k\right)^{-1}X_k$ 为模型 $k$ 的投影矩阵。

(5) MMMA：

$$w=\arg\min C_n\left(w\right) \tag{21}$$

$$C_n\left(w\right)=\left\{Vec\left(Y\right)-Vec\left(\hat{\mu}\left(w\right)\right)\right\}'\hat{C}^{-1}\left\{Vec\left(Y\right)-Vec\left(\hat{\mu}\left(w\right)\right)\right\} \tag{22}$$

其中 $\hat{C}^{-1}$ 指的是多个任务构建多任务高斯过程回归得到的列协方差矩阵。

(6) JMA：在异方差情形下提出的模型平均方法。

$$w=\arg\min CV_n\left(w\right) \tag{23}$$

(24)

记 $\tilde{u}\left(w\right)=\sum_{m=1}^{M}w^m\tilde{u}^m$，$\hat{e}^m=y-\hat{u}^m$，$\tilde{e}=\left(\tilde{e}^1,\tilde{e}^2,\cdots,\tilde{e}^M\right)$，令 $P^m=x^m\left(\left(x^m\right)'x^m\right)^{-1}\left(x^m\right)'$，

$h_{ii}^m = x_i^m\left(\left(x^m\right)' x^m\right)^{-1}\left(x_i^m\right)'$ 为 $P^m$ 中的第 $i$ 个对角元素，令对角矩阵 $D_m = \left(1-h_{ii}^m\right)^{-1}$，可得 $\tilde{e}^m = D_m\hat{e}^m$。其中，$\tilde{u}^m = \left(\tilde{u}_1^m, \tilde{u}_2^m, \cdots, \tilde{u}_n^m\right)'$，$\tilde{u}_i^m$ 是从样本中删除第 $i$ 组观测值 $(x_i, y_i)$ 用最小二乘估计得到的 $\hat{u}^m$ 估计值，有 $\hat{u}^m = P^m y$。

(7) MT-JMA：将 JMA 扩展到多任务上的模型平均方法：

$$w = \arg\min C_n(w) \tag{25}$$

$$C_n(w) = \left\{Vec(Y) - Vec(\tilde{u}(w))\right\}' \hat{C}^{-1}\left\{Vec(Y) - Vec(\tilde{u}(w))\right\} \tag{26}$$

其中 $\hat{C}^{-1}$ 同 MMMA 方法指代的一样。$\tilde{u}(w)$ 为 JMA 方法计算步骤相同所得到的估计值。

为了评估估计量，我们计算损失风险：

$$Risk = \sqrt{\frac{1}{n}\left(y-\hat{y}\right)^2} \tag{27}$$

其中 $y$ 是真实值，$\hat{y}$ 是多任务高斯过程回归得到的候选模型进行模型平均及模型选择后的预测值。

## 3. 数值模拟

输出噪声同方差时，考虑输出间的相关性，即采用 MGPR，通过数值模拟比较函数在 aicc、MMMA、S-AIC、S-BIC 这四种方法下的损失风险。在不考虑输出间的相关性时，分别构建 GPR 时，通过数值模拟比较函数在 aicc、MMA、S-AIC、S-BIC 这四种方法下的损失风险。并将 MGPR 与 GPR 的进行不同模型平均及选择方法后得到的损失风险进行比较。

### 3.1. 参数设置

输出噪声同方差时，为了衡量多任务与单任务在预测值精度这个方面的性能，我们考虑如下的一个已知函数模型。这个已知函数模型是定义在 12 维($p = 12$)的输入空间$[0,1]^{12}$上的，模型中的前六个变量 $x_1, x_2, \cdots, x_6$ 对计算机试验输出结果的影响逐渐减弱，后六个变量的系数为零，即后六个变量为与输出没有相关性的变量。真模型表达成以下形式：

$$y_i(x) = 0.4x_1 + 0.3x_2 + 0.2x_3 + 0.1x_4 + 0.05x_5 + 0.01x_6 + \varepsilon_i(x),\quad i = 1,2 \tag{28}$$

其中 $\left[\varepsilon(1), \varepsilon(2)\right] \sim MGP\left(0, k_{SE}, \Omega\right)$。$k_{SE}$ 中的超参数为：

$$\left[\ell_i, s_f^2\right] = \left[1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 1.001, 0.1\right].$$

且 $\Omega = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$，响应值 $y_1$、$y_2$ 分别通过上式产生。在 Matlab 中，通过拉丁超立方抽样生成维数 $p = 12$，样本量分别为 $N = 120$，140，160，180 的样本 X，我们假设候选模型是嵌套的。即第 $m$ 个模型包含前 $m$ 个回归项。

使用损失风险来衡量多任务高斯过程模型平均后的平均模型对于随机产生的测试集 G 的预测精度，损失风险越小，预测精度越高。测试集 G 是使用计算机产生的 1000 个样本点。重复 500 次实验，计算平均的 Risk。Risk 越小时，我们认为 MGPR 进行模型平均比单个任务的 GPR 进行模型平均更有优势。

### 3.2. 试验结果

使用 aicc、MMMA、MMA、S-AIC、S-BIC 方法进行分析，仿真结果见表 1~4：

**Table 1.** Risk of model averaging and selection (×10)
**表 1.** 模型平均及选择的损失风险(×10)

| $n$ = 120 | Task One of MGPR | Task Two of MGPR | Task One of GPR | Task Two of GPR |
|---|---|---|---|---|
| aicc | 1.6157 | 1.6182 | **1.3762** | **1.2084** |
| S-AIC | **1.1796** | **1.1489** | 1.3755 | 1.2074 |
| S-BIC | **1.1794** | **1.1485** | 1.3707 | 1.1991 |
| MMMA (The first two)<br>MMA (The latter two) | **1.1800** | **1.1582** | 1.3864 | 1.2186 |

**Table 2.** Risk of model averaging and selection (×10)
**表 2.** 模型平均及选择的损失风险(×10)

| $n$ = 140 | Task One of MGPR | Task Two of MGPR | Task One of GPR | Task Two of GPR |
|---|---|---|---|---|
| aicc | 1.6833 | 1.6760 | **1.5795** | **1.4261** |
| S-AIC | **1.1571** | **1.1022** | 1.5795 | 1.4259 |
| S-BIC | **1.1563** | **1.1022** | 1.5758 | 1.4210 |
| MMMA (The first two)<br>MMA (The latter two) | **1.1470** | **1.0928** | 1.5972 | 1.4382 |

**Table 3.** Risk of model averaging and selection (×10)
**表 3.** 模型平均及选择的损失风险(×10)

| $n$ = 160 | Task One of MGPR | Task Two of MGPR | Task One of GPR | Task Two of GPR |
|---|---|---|---|---|
| aicc | 1.6978 | 1.7062 | **1.4574** | **1.4631** |
| S-AIC | **1.2131** | **1.2891** | 1.4570 | 1.4636 |
| S-BIC | **1.2130** | **1.2888** | 1.4532 | 1.4559 |
| MMMA (The first two)<br>MMA (The latter two) | **1.2237** | **1.2925** | 1.4684 | 1.4758 |

**Table 4.** Risk of model averaging and selection (×10)
**表 4.** 模型平均及选择的损失风险(×10)

| $n$ = 180 | Task One of MGPR | Task Two of MGPR | Task One of GPR | Task Two of GPR |
|---|---|---|---|---|
| aicc | 1.7306 | 1.7482 | **1.5996** | **1.3543** |
| S-AIC | **1.2769** | **1.2111** | 1.6003 | 1.3546 |
| S-BIC | **1.2774** | **1.2120** | 1.5941 | 1.3530 |
| MMMA (The first two)<br>MMA (The latter two) | **1.2768** | **1.2161** | 1.6146 | 1.3707 |

(1) 将 MGPR 与 GPR 分别进行模型平均及模型选择，可以得到结论：不论样本量大小是多少，在两种情况中都能够进行比较的模型平均方法中，多个任务同时输出的表现要优于单个任务分别输出。因此可以得知，在已经考虑到输入变量之间的相关性的前提下，再考虑输出之间相关性时，可以提高预测结果的准确性。

(2) 由上述所有表可以得出，先构建 MGPR，再进行模型选择及模型平均得到的 Risk 中，大部分情况下，S-AIC 与 S-BIC 的损失风险十分接近。

(3) 在两个任务分别构建 GPR 后，再进行模型选择及模型平均得到的结果中，可以得出 MMA 方法

的表现要差于其他模型平均方法，这是因为 MMA 方法未考虑输入变量之间的相关性，而其他模型平均方法考虑到输入变量间的相关性，这表明，考虑输入变量之间的相关性可以降低预测的损失风险。

### 3.3. 时间比较

本节在不同样本量下，分别将两个任务构建 MGPR 及 GPR，将得到的候选模型采用上述提到的五种模型选择及模型平均方法，同时输出两个任务的预测风险，对以上过程重复 500 次，得到两个过程所耗费的时间。将以上不同情况的耗费时间进行比较，得到结果见表 5：

**Table 5.** Time spent (minute)
**表 5.** 耗费时间(分)

| $n$ | aicc of MGPR | aicc of GPR | S-AIC of MGPR | S-AIC of GPR | S-BIC of MGPR | S-BIC of GPR | MMMA of MGPR | MMA of GPR |
|---|---|---|---|---|---|---|---|---|
| 120 | **34.71** | 39.99 | **41.34** | 45.48 | **42.18** | 43.00 | **39.12** | 39.78 |
| 140 | **43.38** | 46.59 | **48.82** | 52.88 | **52.46** | 52.80 | **44.75** | 49.65 |
| 160 | **49.85** | 56.06 | **60.62** | 67.54 | **55.15** | 61.75 | **46.86** | 55.92 |
| 180 | **65.93** | 69.01 | **70.03** | 73.35 | **73.76** | 75.25 | **64.74** | 69.18 |

由表 5 可知，在不同样本量及不同模型平均及选择的算法下，构建 MGPR 所耗费的时间比构建 GPR 所耗费的时间要短。因此，结合上一节结论可知，在多个输出具有相关性的前提下，选择 MGPR 构建模型，并进行模型平均，可以降低损失风险，提高预测精度。

### 3.4. 实证分析

#### 3.4.1. 输出噪声同方差

本节采用噪声同方差下提及的模型选择及模型平均方法，分析中国 77 个少数民族地区的数据，收集 2010 年中国第六次全国人口普查的数据。我们将城市化率、非农业人口占比以及受教育年限视为三个输出变量，我们考虑了 6 个与输出变量相关的输入变量:少数民族人口占比、城市和农村收入比的比例、固定资产人均投资、人均 GDP、第二产业增加值占比和第三产业增加值占比。

我们考虑嵌套候选模型，总共 6 个候选模型。为了比较上述提到的模型选择及平均方法的性能，我们使用任何 $T$ 个种族少数区域数据来进行模型建立，其中 $T = 50$。然后，我们通过 77-$T$ 个民族少数地区数据来预测风险。这个过程完成 $N = 500$ 次，我们计算模型选择及平均方法的损失风险。结果见表 6：

**Table 6.** Risk of model averaging and selection
**表 6.** 模型平均及选择的损失风险

| | Years of education of MGPR | Urbanizationrate of MGPR | Rate of nonagricultural population of MGPR | Years of education of GPR | Urbanizationrate of GPR | Rate of nonagricultural population of GPR |
|---|---|---|---|---|---|---|
| aicc | **0.9696** | **0.6896** | 0.9575 | 1.0533 | 0.8306 | **0.9411** |
| WAIC | **0.8562** | **0.7020** | **0.8951** | 1.0526 | 0.8298 | 0.9385 |
| WBIC | **0.8469** | **0.6711** | **0.8861** | 1.0535 | 0.8291 | 0.9382 |
| MMMA (The first three) MMA (The latter three) | **0.8866** | **0.6275** | **0.7951** | 1.0568 | 0.8382 | 0.9434 |

从表 6 可得出，构建 MGPR 后进行模型平均或选择得到的预测精度总体上要优于 GPR，表明在多个任务同时输出的情况中，既考虑输入变量间的相关性又考虑输出变量间的相关性可以降低损失风险。

在该实例中，构建 MGPR 后，采用 MMMA 方法得到的损失风险总体程度上比 S-AIC、S-BIC 方法得到的损失风险要低。而构建 GPR 得到的结果中，总体上是 MMA 方法的损失风险要高于其他模型平均方法，表明输出为单个任务时，考虑输入变量之间的相关性可以降低损失风险。

### 3.4.2. 输出噪声异方差

采用上述提及的噪声异方差下的模型选择及模型平均方法，对本节使用的数据进行分析。本节数据来源于文献[19]，共 38 个样本。通过实验研究了 5 个工艺因素(反应温度、氧气分压、苯甲醇浓度(以 10 ml 甲苯稀释的 mmol 为单位)、Mn 的百分比和 K:Mn 的比)在苄醇转化率和转频率(TOF)内稀释的影响。苄醇转化率和 TOF 被视为过程响应变量。需要注意的是，苄醇转化率和 TOF 高度相关(相关系数 0.65)，即可以使用 MGPR。

我们考虑嵌套候选模型，总共 5 个候选模型。为了比较上述提到的模型选择及平均方法的性能，我们使用 27 个样本数据来进行模型建立，使用剩余的 11 个数据进行预测。这个过程完成 500 次，我们计算模型选择及平均方法的损失风险。结果见表 7：

**Table 7.** Risk of model averaging and selection
**表 7.** 模型平均及选择的损失风险

| | Benzyl alcohol conversion of MGPR | Turnover frequency of MGPR | Benzyl alcohol conversion of GPR | Turnover frequency of GPR |
|---|---|---|---|---|
| aicc | **0.4269** | **0.9945** | 0.5825 | 1.1921 |
| WAIC | **0.4258** | **0.9912** | 0.4850 | 1.2431 |
| WBIC | **0.4253** | **0.9912** | 0.4489 | 1.2886 |
| MMMA (The first two) MMA (The latter two) | **0.4855** | **1.1870** | 0.5338 | 1.1933 |
| MT-JMA (The first two) JMA (The latter two) | **0.4850** | **1.1874** | 0.5090 | 1.1924 |

由表 7 可得，大多数情况下，采用的模型平均方法，总体上是 MGPR 得到的预测风险要低于 GPR 且在单个任务构建 GPR 时，可以得出 JMA 方法优于 MMA 方法，与上述数值模拟结论一致。

## 4. 总结和讨论

在本文中，考虑到现实情况大多数为多个相关的输出变量，可以构建 MGPR，并将不同的候选模型进行模型选择及平均，然后将不同任务分别构建 GPR，同样将候选模型进行模型选择及模型平均，将二者得到的损失风险进行比较。数值结果表明，均值函数为线性时，噪声同方差时，构建多任务高斯过程回归模型后，再进行模型平均的损失风险要低于多个任务分别进行高斯过程回归后模型平均的损失风险；而对于模型选择方法，MGPR 与 GPR 没有明显的优势。对于 GPR 建模来讲，MMA 方法表现较差，表明在输出任务之间存在相关性时，采用 MGPR 建模后进行模型平均得到的损失风险更低，预测结果更优。

## 参考文献


[1] Zhang, J. and Taflanidis, A.A. (2019) Bayesian Model Averaging for Kriging Regression Structure Selection. *Probabilistic Engineering Mechanics*, **56**, 58-70. https://doi.org/10.1016/j.probengmech.2019.02.002

[2] Wang, M.M., Zhang, X.Y., Wan, A.T.K. and Zou, G.H. (2019) On the Asymptotic Distribution of Model Averaging Based on Information Criterion. *Statistics*. https://doi.org/10.48550/arXiv.1910.12208

[3] Priyam, M., Heng, L, Ritwik, M., Hua, L. and Minge, X. (2019) Ageneral Framework for Frequentist Model Averaging. *Science China Mathematics*, **62**, 205-226. https://doi.org/10.48550/arXiv.1802.03511

[4] Liang, Z.Q., Chen, X.L. and Zhou, Y.Q. (2022) Mallows Model Averaging Estimation for Linear Regression Model with Right Censored Data. *Acta Mathematicae Applicatae Sinica*, *English Series*, **38**, 5-23. https://doi.org/10.1007/s10255-022-1054-z

[5] Jin, S.B. and Ankargren, S. (2019) Frequentist Model Averaging in Structural Equation Modelling. *Psychometrika Society*, **84**, 84-104. https://doi.org/10.1007/s11336-018-9624-y

[6] Carroll, R., Zhang, X., Zou, G. and Liang, H. (2020) Parsimonious Model Averaging with a Diverging Number of Parameters. *Journal of the American Statistical Association*, **115**, 126-143. https://doi.org/10.1080/01621459.2019.1604363

[7] Zhang, X. and Liu, C.A. (2019) Inference after Model Averaging in Linear Regression Models. *Econometric Theory*, **35**, 816-841. https://doi.org/10.1017/S0266466618000269

[8] Chen, Z., Fan, J. and Wang, K. (2023) Multivariate Gaussian Processes: Definitions, Examples and Applications. *Metron*, **81**, 181-191. https://doi.org/10.1007/s40300-023-00238-3

[9] Chen, Z., Fan, J. and Wang, K. (2020) Remarks on Multivariate Gaussian Process. https://doi.org/10.48550/arXiv.2010.09830

[10] Hong, X., Huang, B., Ding, Y., Guo, F., Chen, L. and Ren, L. (2017) Multi-Model Multivariate Gaussian Process Modelling with Correlated Noises. *Journal of Process Control*, **58**, 11-12. https://doi.org/10.1016/j.jprocont.2017.08.004

[11] Kontar, R., Raskutti, G. and Zhou, S. (2021) Minimizing Negative Transfer of Knowledge in Multivariate Gaussian Processes: A Scalable and Regularized Approach. *Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3508-3522. https://10.1109/TPAMI.2020.2987482

[12] Zapata, J., Oh, S.Y. and Petersen, A. (2022) Partial Separability and Functional Graphical Models for Multivariate Gaussian Processes. *Biometrika*, **109**, 665-681. https://doi.org/10.1093/biomet/asab046

[13] Zhu, R., Zou, G. and Zhang, X. (2018) Model Averaging for Multivariate Multiple Regression Models. *Statistics*, **52**, 205-227. https://doi.org/10.1080/02331888.2017.1367794

[14] Chen, Z. and Wang, B. (2018) How Priors of Initial Hyperparameters Affect Gaussian Process Regression Models. *Neurocomputing*, **275**, 1702-1710. https://doi.org/10.48550/arXiv.1605.07906

[15] Wei, P., Liu, F. and Tang, C. (2018) Reliability and Reliability-Based Importance Analysis of Structural Systems Using Multiple Response Gaussian Process Model. *Reliability Engineering & System Safety*, **175**, 183-195. https://doi.org/10.1016/j.ress.2018.03.013

[16] Chen, Z., Wang, B. and Gorban, A.N. (2020) Multivariate Gaussian and Student-*t* Process Regression for Multi-Output Prediction. *Neural Computing and Applications*, **32**, 3005-3028. https://doi.org/10.1007/s00521-019-04687-8

[17] Binois, M., Robert, B.G. and Mike, L. (2018) Practical Heteroskedastic Gaussian Process Modeling for Large Simulation Experiments. *Journal of Computational and Graphical Statistics*, **27**, 808-821. https://doi.org/10.48550/arXiv.1611.05902

[18] Hansen, B.E. and Jeffrey, S.R. (2012) Jackknife Model Averaging. *Journal of Econometrics*, **167**, 38-46. https://doi.org/10.1016/j.jeconom.2011.06.019

[19] Tang, Q., Chen, Y., Zhou, C., Chen, T. and Yang, Y. (2009) Statistical Modelling and Analysis of the Aerobic Oxidation of Benzyl Alcohol over K-Mn/C Catalysts. *Catalysis Letters*, **128**, 210-220. https://doi.org/10.1007/s10562-008-9740-x