arXiv:2301.02296v2 [stat.ME] 5 May 2023

使用贝叶斯加性回归树

JohnC. Yannotty、Thomas J. Santner、Richard J. Furnstahl 和 Matthew T. Pratola 俄亥俄州立大学

2023年5月8日

摘要

在现代计算机实验应用中，人们经常会遇到这样的情况：针对某个物理系统考虑多个模型，每个模型都通过计算机模拟器实现。在这种情况下，一个重要问题是确定用于预测和推断的最佳模拟器，或最佳模拟器组合。贝叶斯模型平均（BMA）和堆叠是两种用于处理模型不确定性的统计方法，它们通过简单的线性组合或加权平均汇总一组预测。贝叶斯模型混合（BMM）在此基础上加以扩展，通过定义输入依赖权重来捕捉各模拟器的局部行为。一种可行方法是使用灵活的非参数模型定义输入与权重函数之间的关系，从而学习各模拟器在不同局部区域的优势和不足。本文提出了一种基于贝叶斯加性回归树（BART）的 BMM 模型。本文将所提方法应用于组合有效场论（EFT）的预测，并以一项核物理应用为例进行说明。

关键词： 计算机实验；有效场论；模型堆叠；不确定性量化

1 引言

在统计学习问题中，人们通常会考虑一组合理的模型，每个模型都旨在解释所关注的系统。常见做法是根据一些预先指定的准则，选择表现最佳的模型。随后，便以所选模型为真实的数据生成机制，对感兴趣的量进行推断。由此得到的不确定性

量化忽略了底层模型结构所导致的变异性（Draper，1995）。对这类量相关的不确定性加以错误表述，最终可能导致错误的解释或不恰当的决策。典型建模方法的另一个缺点是，所得推断可能高度依赖于选择准则。换言之，不同的准则集合可能会导致截然不同的最终模型和推断结果。为考虑这些不确定性，可以通过某种方式整合这组模型所包含的信息。

任何模型集都可以分为 $\mathcal{M}$-封闭、 $\mathcal{M}$-完备或 $\mathcal{M}$-开放三类（Bernardo 和 Smith，1994）。这三类的区别在于，它们对真实模型 $\mathcal{M}_\dagger$ 及其与模型集之间关系所作的基本假设不同。 $\mathcal{M}$-封闭情形假设可以建立对 $\mathcal{M}_\dagger$ 的数学表示，并且该表示包含在模型集中。在这种情形下，模型选择是合适的，因为可以从所考虑的模型集合中恢复 $\mathcal{M}_\dagger$ 。 $\mathcal{M}$-完备情形同样假设可以构建 $\mathcal{M}_\dagger$，但它不包含在模型集中。例如，可能存在对 $\mathcal{M}_\dagger$ 的表达式，但与所考虑的模型相比，该表达式的计算开销可能很大，甚至难以计算。 $\mathcal{M}$-开放情形假设真实模型可能存在，但由于知识或资源不足，无法构建其数学表示。因此， $\mathcal{M}_\dagger$ 被排除在模型集之外。本研究的动机来自核物理领域的应用，这些应用往往属于 $\mathcal{M}$-开放类，因为人们可能尚未理解有关物理系统的根本真相。在此类情况下，人们或许希望利用模型集中包含的有关物理系统的已知信息以及实验数据，进一步理解核现象。

假设在研究某个特定系统时考虑了一组 K 模型。处理模型不确定性的一种方法是结合这些 K 模型中的信息。这可能涉及以某种加性方式，结合各个模型的点预测或概率密度函数。传统的频率学派方法和贝叶斯方法采用全局加权方案，其中每个模型的权重都是一个旨在反映整体（全局）模型性能的数值。例如，一种经典的全局加权方案是贝叶斯模型平均（BMA）（Raftery 等，1997），它通过凸组合来结合各个模型的后验密度。BMA 权重由各个模型后验概率给出，每个概率都可以解释为相应模型是真实数据生成模型的概率。因此，BMA 隐含地假设真实模型包含在模型集中，这使得该方法不适用于 $\mathcal{M}$-封闭情形（Bernardo 和 Smith，1994）。较新的贝叶斯全局加权方案采用模型堆叠方法，

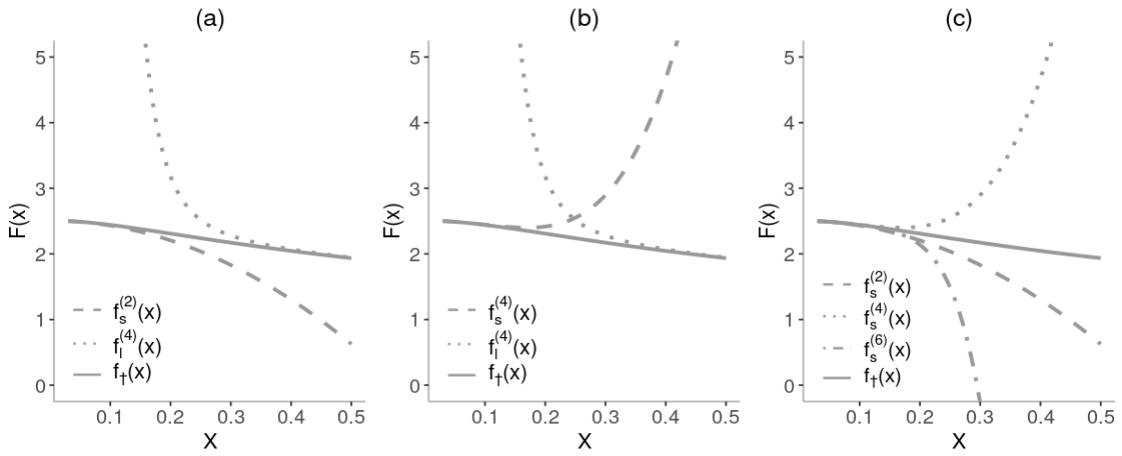


图1：三种不同的有效场论实验设置。每个面板都展示了真实物理系统（实线）以及所考虑的有效场论的均值预测（非实线）。

其中，模型权重的设定旨在最小化指定的后验期望损失。全局加权的这种决策理论视角可用于组合点预测（Le 和 Clarke，2017）或概率密度（Yao 等，2018）。在某些假设下，研究表明，堆叠方法在 $\mathcal{M}$-开放和 $\mathcal{M}$-封闭情形下都更为适用（Yao 等，2018）。

尽管全局加权方法有效，但当单个模型性能具有局部性时，它们仍可能无法很好地近似真实系统。在这种情况下，可以通过构造输入依赖权重，选择一种能够反映模型局部特征的加权方案。使用输入依赖权重时，预期单个模型在其预测性能较强的输入区域获得较高权重，而在性能较差的区域获得接近 0 的权重。对于 $\mathcal{M}$-开放或 $\mathcal{M}$-完备设置，局部加权方案更为合适，因为真实模型可能更适合描述为所考虑模型集的局部混合。

本研究的动机源于核物理中的一些问题，这些问题采用一种称为有效场论（EFT）的技术建模（Burgess，2020；Petrov 和 Blechman，2016；Georgi，1993）。有效场论旨在输入域的特定子区域内表现良好，但在输入域的其余部分则会发散。此类模型的典型例子是 Honda（2014）提出的零维 ϕ^4 理论配分函数的弱耦合和强耦合有限阶展开。这一问题的例子见图1，其中各种虚线和点线表示有限阶展开的均值预测，实线则表示真实物理系统。可以看出，这些模型在某些区域内能够高度准确地描述真实系统

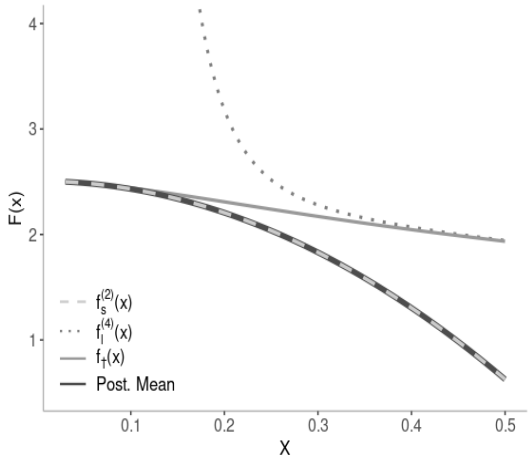


图2：将贝叶斯模型平均应用于二阶弱耦合展开和四阶强耦合展开时， $f_{\dagger}(x)$ 的后验均值预测。

，但无法提供全局准确的模型。大多数 EFT 问题都属于 $\mathcal{M}$-开放设定，因为整个定义域内系统的真实底层描述未知，因此并未包含在模型集中。相反，可以依据已知物理构建多个有效场论，从而在定义域的若干子集上还原真实系统。这引出了一个问题：如何结合多个有效场论的预测，以获得全局准确的预测？目前已有各种插值方法（Honda，2014），但尚无适用于有效场论的数据驱动方法。

为了说明为何属于 $\mathcal{M}$-开放类的问题可能不适合采用模型平均方案，可以考虑将贝叶斯模型平均应用于图1(a)所示、包含这两种展开的模型集。贝叶斯模型平均得到的后验均值预测对真实系统的估计效果不佳，如图2所示。从本质上说，贝叶斯模型平均选择了虚线模型，而未能利用模型集中各模型的局部优势。鉴于有效场论的特性以及此类问题所处的 $\mathcal{M}$-开放环境，仅对各模型的预测进行简单加权平均，不足以恢复真实物理系统。更好的方法是采用依赖输入的加权方案，利用每个模型的局部行为，从而获得恰当的均值预测并进行不确定性量化。这种方法属于一类统称为贝叶斯模型混合（BMM）的问题（Yao 等，2021）。

BMM 的一项关键挑战是定义输入与权重函数之间的关系。本文提出了一种贝叶斯树模型，将权重函数表示为一组树的和。这种表示依赖于树基函数，用于学习局部模型行为。

此外，这种灵活的非参数方法使用户无需为权重函数指定限制性更强的模型，例如广义线性模型。在保留贝叶斯加性回归树（BART）模型所具有的传统共轭性这一特性的同时，权重函数通过多元高斯先验进行正则化。该先验经过校准，使权重函数倾向于区间 $[0,1]$ ，且不施加任何其他约束。此外，该框架还提供了一种简单的策略，可在有相关先验信息时将其纳入局部模型性能的考量。总而言之，与传统模型平均方法相比，这种方法突显了候选模型的局部行为，并显著提升了预测、解释和不确定性量化的表现。

除了提出一种新颖的非参数 BMM 方法外，本研究还介绍了一种新的数据驱动方法，用于结合不同有效场论的预测。这不仅对系统预测很重要，也关系到由此得出的推断。具体而言，研究人员可以更好地了解每种有效场论的准确性，同时加深对底层物理系统的认识，尤其是在当前所考虑的有效场论无法充分解释的领域。

本文其余部分安排如下。第2节介绍与模型平均、模型混合和贝叶斯加性回归树相关的一些工作。第3节介绍有效场论的基本特征，第4节则概述所提出的、基于贝叶斯加性回归树的框架的具体内容。第5节展示三个用于说明问题的有效场论示例。最后，第6节详细讨论本文所呈现的结果。方法的完整推导见附录。有关有效场论的其他示例和信息见在线补充材料。

2 背景

本节概述本文讨论的主要统计方法。第2.1节详细介绍常见的模型平均和模型混合技术。第2.2节总结贝叶斯树模型的主要特征；这些特征对于本文提出的模型混合方法至关重要。

2.1 模型平均与模型混合

过去几十年来，人们广泛研究了应对模型不确定性的方法。该领域的大多数研究都通过均值或

密度估计来结合竞争模型。无论采用哪种方式，通常都会通过对所考虑模型各自的预测均值或密度进行线性组合来计算组合结果。该线性组合中的权重可能取决于各模型的输入，也可能不取决于这些输入；这些权重是利用训练数据集 $\mathcal{D} = \{(\boldsymbol{x}_1, y_1) \ldots, (\boldsymbol{x}_n, y_n)\}$ 学得的。现有许多用于估计模型权重的频率学派方法和贝叶斯方法。诸如堆叠 （Breiman，1996）和模型聚合 （Bunea 等，2007）等常用的频率学派方法通过最小化指定的损失函数来估计权重。此外，还可以选择对权重施加非负约束或和为一约束等限制，或采用正则化技术。其他频率学派方法则使用赤池信息准则（Burnham 等，1998）或 Mallows 的 CP（Hansen，2007）等评估指标来估计权重。除 Sill 等（2009）的方法外，这些方法通常都属于模型平均范畴，因为权重独立于模型输入。本节余下部分将进一步介绍常用的贝叶斯方法。

贝叶斯模型平均： 一种结合模型 $\mathcal{M}_1, \ldots, \mathcal{M}_K$ 的经典方法是贝叶斯模型平均（Raftery 等，1997）。设 Q 是一个感兴趣的量。Q 的后验密度定义为 $\pi(Q \mid \mathcal{D}) = \sum_{l=1}^{K} w_l \, \pi(Q \mid \mathcal{D}, \mathcal{M}_l)$，这是各模型对应后验密度的加权平均。每个权重均根据相应的后验模型概率定义，即 $w_l = \pi(\mathcal{M}_l \mid \mathcal{D})$，其中

$$\pi(\mathcal{M}_l \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathcal{M}_l)\pi(\mathcal{M}_l)}{\sum_{k=1}^{K} p(\mathcal{D} \mid \mathcal{M}_k)\pi(\mathcal{M}_k)}$$

而 $p(\mathcal{D} \mid \mathcal{M}_l)$ 是相对于第 l 个模型的数据边际似然。尽管贝叶斯模型平均很有用，但它也受到批评：过于强调对训练数据的拟合，而非样本外预测；渐近地选择单一模型（这在 $\mathcal{M}$-完备和 $\mathcal{M}$-开放情形下并不合适，例如图2）；并且对先验设定敏感。

贝叶斯均值堆叠： 近期研究将堆叠拓展到贝叶斯范式，将其作为一种均值估计方法（Clyde 和 Iversen，2013；Le 和 Clarke，2017）。给定 K 个竞争模型，输入为 $\tilde{\boldsymbol{x}}$ 时，未来观测 $\tilde{y}$ 的堆叠均值由各个模型预测器 $E[\tilde{y} \mid \tilde{\boldsymbol{x}}, \mathcal{D}] = \sum_{l=1}^{K} w_l \, f_l(\tilde{\boldsymbol{x}})$ 的线性组合构成，其中 $E[\tilde{y} \mid \tilde{\boldsymbol{x}}, \mathcal{D}, \mathcal{M}_l] = f_l(\tilde{\boldsymbol{x}})$。当各个模型未知时，堆叠分两步进行：(i) 根据训练数据集 $\mathcal{D}$，分别拟合各个模型 $\mathcal{M}_l, \quad l = 1, \ldots K,$ ；(ii) 给定拟合后的模型，为堆叠预测器估计权重 $\mathbf{w} = (w_1, \ldots, w_K)^\top$。

ˆ 第一步，对每个模型进行拟合，并在每个训练点处得到相应的均值预测 $f_l(\boldsymbol{x}_i)$。在实践中，会使用交叉验证技术来降低

堆叠预测器对训练数据过拟合的风险。第二步，将系数向量$\boldsymbol{w}=(w_1、.....、\ w_K)^\top$定义为指定后验期望损失的最小化解。此外，还可以对权重施加各种约束，例如单纯形约束、非负约束或和为 m 约束（Le 和 Clarke，2017）。其他方法包括通过惩罚项或先验进行正则化（Breiman，1996；Yang 和 Dunson，2014）。

贝叶斯完全堆叠：完全堆叠源于 BMA 的不足（Yao 等，2018）。这种贝叶斯堆叠模型强调预测，因为其权重的选择旨在最小化真实预测密度与堆叠预测密度之间的 Kullback-Leibler（KL）散度 $p(\tilde{y} \mid \tilde{\boldsymbol{x}}) = \sum_{l=1}^{K} w_l\ p(\tilde{y} \mid \tilde{\boldsymbol{x}},\ \mathcal{D},\ \mathcal{M}_l)$，其中 $\tilde{y}$是输入为 $\tilde{\boldsymbol{x}}$ 的未来观测。与均值堆叠类似，当各个模型未知时，可以使用留一法（LOO）交叉验证预测密度代替 $p(\tilde{y} \mid \tilde{\boldsymbol{x}},\ \mathcal{D},\ \mathcal{M}_l)$。给定训练数据后，权重被限制在一个 $K-$维单纯形 S_K 中，并通过 $\widehat{\boldsymbol{w}} = \text{argmax}_{\boldsymbol{w}\in S_K} \sum_{i=1}^{n} \log \sum_{l=1}^{K} w_l\ p(y_i \mid \boldsymbol{x}_i,\ \mathcal{D}^{(-i)},\ \mathcal{M}_l)$ 进行估计，其中 $\mathcal{D}^{(-i)}$ 表示排除数据对 $(\boldsymbol{x}_i,\ y_i)$ 后的训练集。

贝叶斯层次堆叠：层次堆叠（Yao 等，2021）是一种模型混合方法，它通过定义输入依赖权重来扩展完全堆叠，并以完全贝叶斯的方式估计这些权重。定义权重函数的一种方法是使用参数模型。首先，$K-1$ 无约束权重函数定义如下，

$$w_l^*(\boldsymbol{x}_i) = \mu_l + \sum_{j=1}^{J} \alpha_{lj}\ g_j(\boldsymbol{x}_i),$$

它们取决于超参数集合 $\{\alpha_{lj}\}$ 和 $\{\mu_l\}$，以及用户指定的基函数 g_j（$\boldsymbol{x}_i$），其中 $j=1,\ ...,\ J$ 和 $l=1,\ ...,\ K-1$。将 K^{th} 函数 w_K^*（$\boldsymbol{x}_i$）设为 0，作为基准。然后，对无约束权重应用 softmax 变换，使每个模型权重都限制在 K 维单纯形上，即

$$w_l(\boldsymbol{x}_i) = \frac{\exp\left(w_l^*(\boldsymbol{x}_i)\right)}{\exp\left(w_1^*(\boldsymbol{x}_i)\right) + \cdots + \exp\left(w_K^*(\boldsymbol{x}_i)\right)}, \quad l=1,\ldots,K.$$

上述讨论的方法概述了若干种整合多个模型信息的策略。在有效场论实验中，预测具有局部性，因此，贝叶斯层次堆叠这类依赖输入的加权方案更为适合。不过，指定所需的基函数可能并非易事。因此，本文提出的方法将在加性树基模型中采用均值堆叠的概念，以灵活且非参数的方式实现局部加权。

2.2 贝叶斯树模型

贝叶斯加性回归树（BART）在复杂高维系统建模中日益受到欢迎（Chipman 等人，2010）。这种加性方法将 m 棵树所作的预测相加，并通过贝叶斯回拟合算法加以实现（Hastie 和 Tibshirani，2000）。每棵树 T_j 由其结构（包括内部节点和终端节点）及其对应的一组终端节点参数 M_j所定义。内部节点根据指定的分裂规则对输入空间进行二元划分。节点 η被定义为内部节点的概率为 p（η为内部节点） $=\alpha$（$1+d_\eta$）$^{-\beta}$其中d_η 是 η 的深度， α 和 β 是调节参数。按此构造，该先验会对树的复杂度进行惩罚，从而确保每棵树都保持浅而简单的结构。给定 d 个不同的预测变量， x_1，.....， x_d，分裂规则的形式为 $x_v < c$，其中 $v \in \{1, ..., d\}$，切分点 c取自 $\mathbb{R}$的离散子集。在最简单的方法中，与每条分裂规则相关的预测变量和切分点均从离散均匀分布中随机选取。用于确定各节点类别的概率，以及内部节点的分裂规则，共同定义了每棵树的随机树生成先验。

m 树通过 MCMC 学习，在模拟的每次迭代中都会提出对每个树结构进行细微修改的方案。通常，这些树结构修改包括生长、死亡、扰动或旋转，如 Chipman 等人（1998 年）和 Pratola（2016年）所述。随后使用 Metropolis-Hastings 步骤接受或拒绝这些提议。为避免复杂的可逆跳跃马尔可夫链蒙特卡罗算法，该算法依赖于积分似然；积分似然是通过对给定树对应的终端节点参数进行积分而得到的。采用终端节点参数的条件共轭先验，可以得到该密度的闭式表达式。

给定树结构后，可以为每个终端节点参数指定先验分布。在贝叶斯加性回归树模型中，先验确保每棵树解释数据中一种微小但不同的变异来源。对于连续数据，贝叶斯加性回归树模型为终端节点参数指定高斯先验。假设数据已进行均值中心化，则分配给节点η_{pj}中终端节点参数μ_{pj} 的先验为 $\mu_{pj} \mid T_j \sim N(0, \tau^2)$，其中 $\tau = (y_{max} - y_{min})/(2k\sqrt{m})$，调节参数为 k。此外，还为方差指定缩放逆卡方先验，即 $\sigma^2 \sim \nu\lambda/\chi^2_\nu$。

—

传统的贝叶斯回归树模型可以扩展，以允许终端节点具有更复杂的结构。现有扩展包括线性回归（Chipman 等人，2002；Prado 等人，2021）和高斯过程（Gramacy 和 Lee，2008）。对于模型设置

模型混合方面，本研究采用了多元高斯终端节点模型。

3 使用有效场论进行模型混合

有效场论（EFT）将输入参数与物理相关尺度之比表示为一个或多个展开。计算机模型通过模拟器实现有效场论。物理系统的理论预测是各模拟器给出的近似，再加上一个差异项，该差异项旨在解释系统中其余尚未解释的部分。在处理观测数据时，可以利用这两个组成部分的特定性质。本节结合有效场论概述这些细节，补充材料中则提供了进一步讨论。

3.1 有效场论示例的动机

考虑这样一个有效场论示例：真实物理系统是零维 ϕ^4 理论的配分函数，其定义为

$$f_\dagger(x) = \int_{-\infty}^{\infty} \exp\Big(-\frac{u^2}{2} - x^2u^4\Big)\, du, \tag{1}$$

其中， x 表示耦合常数（Honda，2014）。该配分函数有两种有限阶展开，分别由式（2）和式（3）给出，对应于 n_s 或 $n_l \geq 1$，即

$$h_s^{(n_s)}(x) = \sum_{t=0}^{n_s} s_t x^t \quad \text{where} \;\; s_t = \begin{cases} \frac{\sqrt{2}\Gamma(t+0.5)}{(t/2)!}(-4)^{(t/2)} & t \text{ is even} \\ 0 & t \text{ is odd} \end{cases} \tag{2}$$

$$h_l^{(n_l)}(x) = \sum_{t=0}^{n_l} l_t x^{-t} \quad \text{where} \;\; l_t = \frac{\Gamma(0.5t+0.25)}{2t!}\Big(-\frac{1}{2}\Big)^t, \quad t = 0, ..., n_l. \tag{3}$$

式（2）中的弱耦合展开是一个以零为中心、阶数为 n_s 的渐近类泰勒级数。因此， $h^{(n_s)}{}_s(x)$ 在耦合常数较小时会给出高保真预测，并随着耦合常数增大而发散。式（3）中的强耦合展开则呈现相反的行为， $h^{(n_l)}{}_l(x)$，它是收敛的。使用这些有限阶展开对物理系统所作的示例预测见图1，并在第3.2节中详细讨论。

物理系统的弱耦合展开和强耦合展开理论预测

分别用式（4）和式（5）表示。

$$f_s^{(n_s)}(x) = h_s^{(n_s)}(x) + \delta_s^{(n_s)}(x) \tag{4}$$

$$f_l^{(n_l)}(x) = h_l^{(n_l)}(x) + \delta_l^{(n_l)}(x). \tag{5}$$

其中，截断误差 $\delta^{(n_s)}{}_s(x)$ 和 $\delta^{(n_l)}{}_l(x)$ 均采用高斯过程（GP）建模（Gra- macy，2020；Santner 等人，2018）。正如 Melendez 等人（2019）所述，两个截断误差模型中的参数均取决于相应有限阶展开（分别见式（2）和式（3））在稀疏点网格上的求值结果。差异模型还取决于物理量 $Q(x)$ 和 $y_{\text{ref}}(x)$，这些量根据领域专业知识选定。这些物理量与差异之间的关系总结于补充材料中。当 $Q(x)$ 和 $y_{\text{ref}}(x)$ 未知时，也可以改用 Semposki 等人（2022）所述的误差近似方法。

Honda（2014）这一示例所呈现的特征，在有效场论问题中十分常见。例如，物理系统可以表示为一个加性模型，其中包含有限阶展开及其诱导的截断误差。有限阶展开旨在为定义域的特定子区域提供高保真预测。在定义域中存在一个子区域，所有有限阶展开在该区域都无法给出准确的理论预测。总而言之，这个示例为一般实验环境中可能遇到的有效场论提供了一个原型。

3.2 有效场论实验的模型集

使用有效场论时，可能会遇到各种实验环境。第 3.1 节介绍的示例背景中阐述了这类情形。首先，考虑最基本的情况：模型集仅包含一个有效场论。尽管在局部区域表现良好，但只有一个有效场论时，真实系统的整体预测准确度较差。例如，假设模型集 $\mathcal{M}$ 包含二阶弱耦合展开 $f_s^{(2)}(x)$。由式 (2) 和 (4) 构造的均值预测如图 1(a) 中的虚线所示。显然，该模型只有在定义域左侧的子区域内才具有较强的预测准确度。

在条件允许时，可以考虑同一有效场论的不同有限阶近似。例如，考虑图 1(c) 中所示的二阶、四阶和六阶耦合展开。这三个模型在较小的耦合常数下非常相似，但在定义域的其余部分则截然不同。尽管每种展开的理论预测都不理想，我们仍然可以

利用现有信息来提高对物理系统的整体预测。例如，二阶和六阶展开（虚线和点划线）是凹函数，而四阶展开（点线）是凸函数。这表明真实物理系统位于所考虑的展开之间，并且可以通过对相应的预测重新加权来恢复。

第三种情况是考虑以定义域中不同区域为中心的有效场论。例如，模型集可以包含有限阶弱耦合展开（虚线）和四阶强耦合展开（点线），如图1(a)和图1(b)所示。加入强耦合展开后，便可以在定义域最右侧的子区域中考虑对物理系统的高保真近似。面板(a)中列出的模型集意味着真实系统位于这两种展开之间。这在两种有效场论都不准确的中间范围内尤其有用。与此同时，面板(b)中的模型集展示了一种有趣的情况：在中间范围内，物理系统低于两种有效场论。在这种情况下，可以利用观测数据中的信息来帮助还原真实系统。

在这个例子中，与强耦合展开的预测相比，弱耦合展开的预测精度下降得较为缓慢。因此，在整个定义域内，弱耦合展开通常表现出更好的整体预测性能。使用贝叶斯模型平均等全局加权方案来组合这两类有效场论时，由于弱耦合展开在整体模型性能上具有显著优势，最终的预测会偏向弱耦合展开。图2清楚地显示了贝叶斯模型平均解并不理想：它实际上与二阶弱耦合展开相匹配。因此，在有效场论的设定中，更理想的是采用能够捕捉各个模型局部行为的加权方案。

后续各节将讨论一组一般的 K 不同有效场论，记作$f_1(x) \ldots, f_K(x)$。在这个示例中， $f_l(x) = h_l(x) + \delta_l(x)$，其中 $h_l(x)$表示阶数为 N_l的弱耦合或强耦合展开， $l = 1, \ldots, K$。与此同时， $\delta_l(x)$是相应的截断误差，并按照补充材料中的描述使用高斯过程进行建模。

3.3 有效场论的预测

在进行模型混合之前，分别对这 K 个有效场论进行独立模拟。为方便起见，考虑第 l 个有效场论，记作 $f_l(\boldsymbol{x})$。假设该有效场论配有一组模拟器运行，其输入取自固定集合 $\boldsymbol{x}^c_{l1}$，……，$\boldsymbol{x}^c_{ln_l}$。有关每个有效场论的计算机实验设计信息，可参见 Melendez等人（2021）。模拟器运行是在指定输入处对有限阶展开 $h_l(\cdot)$ 进行求值。利用这些运行结果，可以

在每个固定输入处提取系数集合 N_l+1 $c_0(\cdot)$……， $c_{N_l}(\cdot)$。于是，第 l 个有效场论的训练集定义为 $\mathcal{D}_l=\left\{\left(\boldsymbol{x}_{l1}^c,\boldsymbol{C}(\boldsymbol{x}_{l1}^c)\right)，……，\left(\boldsymbol{x}_{ln_l}^c，\boldsymbol{C}(\boldsymbol{x}_{ln_l}^c)\right)\right\}$ 其中 $\boldsymbol{C}(\cdot)$ 表示指定模型输入处已知有限阶系数组成的向量。不同模型的系数和输入集合可能各不相同，因此集合 $\mathcal{D}_1$，……， $\mathcal{D}_K$ 所包含的信息也会有所不同。

如补充材料所述，利用有限阶系数拟合有效场论，以学习表征分配给截断误差的高斯过程的未知参数。可以从 $\mathcal{D}_l$ 中提取这些信息，这意味着拟合每个有效场论并不需要场观测。因此，无需使用任何观测数据，即可获得输入域内所需的理论预测。由此得到的后验预测分布是一个高斯过程，可按照 Melendez 等人（2019）的描述（另见补充材料），通过相应的均值函数和协方差函数来表征。随后，通过后验均值计算有效场论的预测。

4 贝叶斯加性模型混合树

4.1 定义混合模型

所提出的 BMM模型使用一组观测数据进行训练， Y_1，……， Y_n，假设这些数据是在固定输入 $\boldsymbol{x}_1$，……， $\boldsymbol{x}_n$ 处独立生成的，依据

$$Y_i=f_\dagger(\boldsymbol{x}_i)+\epsilon_i,\quad \epsilon_i\overset{iid}{\sim}N(0,\sigma^2)$$

其中， $f_\dagger(\boldsymbol{x}_i)$ 表示真实但未知的物理系统。给定某一点处的理论预测后， $f_1(\boldsymbol{x}_i)$……， $f_K(\boldsymbol{x}_i)$，数据可建模为

$$Y_i\mid\boldsymbol{f}(\boldsymbol{x}_i),\boldsymbol{w}(\boldsymbol{x}_i),\sigma^2\overset{ind}{\sim}N\left(\boldsymbol{f}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i),\sigma^2\right)\tag{6}$$

其中 $\boldsymbol{f}(\boldsymbol{x}_i)=\left(f_1(\boldsymbol{x}_i)……，\ f_K(\boldsymbol{x}_i)\right)^\top$ 和 $\boldsymbol{w}(\boldsymbol{x}_i)=(w_1(\boldsymbol{x}_i)……，\ w_K(\boldsymbol{x}_i))^\top$。这种表述是贝叶斯均值堆叠与依赖输入的加权方案的一个示例。实际应用中，每个模型的预测都是未知的，必须进行估计。

所提出的 BMM模型采用两步法来组合跨越 K 有效场论的预测。这意味着首先对每个有效场论分别进行拟合，并在学习权重函数 $w_1(\boldsymbol{x}_i)\ldots,w_K(\boldsymbol{x}_i)$之前，得到估计的预测 $f_l(\boldsymbol{x}_i)$，分别对应 $l=1\ldots,K$ 和 $i=1,\ldots,n$。所提出的两步法针对有效场论量身设计，利用了第3.3节中描述的数据来源以及补充材料中描述的性质。条件

在给定估计预测的条件下，观测数据的模型变为

$$Y_i \mid \hat{\boldsymbol{f}}(\boldsymbol{x}_i), \boldsymbol{w}(\boldsymbol{x}_i), \sigma^2 \overset{ind}{\sim} N\big(\hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i), \sigma^2\big)$$

其中 $\hat{\boldsymbol{f}}(\boldsymbol{x}_i) = \big(\hat{f}_1(\boldsymbol{x}_i)\ldots,\hat{f}_K(\boldsymbol{x}_i)\big)^\top$。随后，利用现场数据集学习权重函数。下一节将概述所提出的模型混合方案，该方案使用贝叶斯加性回归树（BART）定义权重函数。

4.2 使用贝叶斯加性回归树进行模型混合

The 权重函数 $\mathbf{w}(\boldsymbol{x}) = (w_1(\boldsymbol{x})\ldots, w_K(\boldsymbol{x}))^\top$ 通过树求和进行建模 s

$$\boldsymbol{w}(\boldsymbol{x}_i) = \sum_{j=1}^{m} \boldsymbol{g}(\boldsymbol{x}_i, T_j, M_j), \tag{7}$$

其中，$\boldsymbol{g}(\boldsymbol{x}_i, T_j, M_j)$ 是第 j 棵树在输入 $\boldsymbol{x}_i$ 处使用终端节点参数集 M_j 得到的 K 维输出。该方法利用从数据中学习得到的树基函数定义权重函数。可以通过改变树的数量或调整先验分布中的超参数，来控制权重函数的灵活程度。

在此处的贝叶斯加性回归树应用中，每个终端节点参数都是一个 K 维向量，并赋予多元高斯先验。该参数经过正则化，使每棵树对权重函数的变化仅产生较小影响。对于接下来的表述，令 η_{pj} 表示第 j 棵树上的第 p 个终端节点，并将其对应参数定义为 $\boldsymbol{\mu}_{pj} = (\mu_{pj1}, ..., \mu_{pjK})^\top$。现在假设观测值 $(\boldsymbol{x}_1, y_1)$ ...,$(\boldsymbol{x}_{n_p}, y_{n_p})$ 位于由 η_{pj} 定义的超矩形内，其中 n_p 是分配到该子区域的观测值数量。每个终端节点处的模型相当于拟合一个局部化的贝叶斯线性回归，其参数向量为 $\boldsymbol{\mu}_{pj}$。由于条件独立性，该节点中的似然定义为

$$L(r_1, ..., r_{n_p} \mid T_j, \boldsymbol{\mu}_{pj}, \sigma^2) = (2\pi\sigma^2)^{-n_p/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n_p} \Big(r_i - \hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{\mu}_{pj} \Big)^2 \Big)$$

其中，$\hat{\boldsymbol{f}}(\boldsymbol{x}_i) = \big(\hat{f}_1(\boldsymbol{x}_i) ..., \hat{f}_K(\boldsymbol{x}_i)\big)^\top$ 是来自各个有效场论的均值预测向量，而 r_i 是由 $r_i = y_i - \sum_{q \neq j} \hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i) g(\boldsymbol{x}_i, T_q, M_q)$。

给定树结构 T_j，终端节点参数 $\boldsymbol{\mu}_{pj}$ 服从共轭多元高斯先验，即

$$\boldsymbol{\mu}_{pj} \mid T_j \overset{ind}{\sim} N_K\Big(\boldsymbol{\beta}, \tau^2 \boldsymbol{I}_K\Big) \tag{8}$$

其中， $\boldsymbol{\beta}=(\beta_1,...,\beta_K)^\top$ 是一个 K 维均值向量， $\boldsymbol{I}_K$ 是单位矩阵。无论输入空间如何划分，该先验的均值都保持不变，因此它是非信息先验。

在模型混合中，每个模拟器可能在输入空间的某个子区域表现出色，而在另一个子区域表现欠佳。这种认识可以通过让超参数取决于分配给给定终端节点的输入空间划分，反映在$\boldsymbol{\mu}_{pj}$ 的先验分布中。因此，可以为$\boldsymbol{\mu}_{pj}$ 构造如下信息先验:

$$\boldsymbol{\mu}_{pj} \mid T_j \overset{ind}{\sim} N_K\Big(\boldsymbol{\beta}_{pj}, \tau^2 \boldsymbol{I}_K\Big)$$

其中 $\boldsymbol{\beta}_{pj}=(\beta_{pj1}$、.....、$\beta_{pjK})^\top$。这样，先验均值就可以随树划分而变化，从而体现局部模型性能。与此同时，假定的协方差结构意味着 K 向量的各分量μ_{pj1}、.....、μ_{pjK} 先验独立。

所提出的两种先验都是共轭先验，这在贝叶斯加性回归树中是一项重要的选择，因为这使得残差向量 $\boldsymbol{R}_{pj}=(r_1,$ ....., $r_{n_p})^\top$，的边际似然可以写成闭式表达式。此外，共轭先验还使终端节点参数和误差方差的全条件分布具有闭式表达式。这些分布的推导见附录。具体而言， T_j 中第 p 个终端节点的全条件分布为

$$\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2 \overset{ind}{\sim} N_K\left(\left(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\hat{\boldsymbol{F}}_{pj}+\frac{1}{\tau^2}\boldsymbol{I}_K\right)^{-1}\left(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj}+\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\boldsymbol{R}_{pj}\right), \left(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\hat{\boldsymbol{F}}_{pj}+\frac{1}{\tau^2}\boldsymbol{I}_K\right)^{-1}\right)$$

其中 $\hat{}\ \boldsymbol{F}_{pj}$ 是 $n_p \times K$ 设计矩阵，其第 i 行向量为向量 $\hat{}\boldsymbol{f}^\top(\boldsymbol{x}_i)$。 σ^2 的全条件分布为缩放逆卡方分布，即 $\sigma^2 \mid \cdot \sim \nu'\lambda'/\chi^2_{\nu'}$，其中

$$\nu' = n+\nu \quad \text{and} \quad \lambda' = \frac{1}{n+\nu}\Big(\sum_{i=1}^{n}\Big(y_i-\hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i)\Big)^2+\nu\lambda\Big),$$

其中 ν 和 λ 分别表示先验形状参数和尺度参数。

4.3 先验校准

首先考虑终端节点参数的先验。非信息先验和信息先验的超参数校准方式有所不同，但这两种方法都旨在确保每个模型权重 $w_l(\boldsymbol{x})$ 倾向于区间 $[0,1]$ ，且其中心值位于该区域内。此外，假设在固定输入下，函数 $w_1(\boldsymbol{x})\ldots,w_K(\boldsymbol{x})$ 先验独立。这使得可以对每个权重的先验进行边际校准。

4.3.1 非信息先验

考虑终端节点参数的非信息先验。在此设定下，$\boldsymbol{\mu}_{pj} \mid T_j \overset{iid}{\sim} N_K\left(\boldsymbol{\beta}, \tau^2 \boldsymbol{I}_K\right)$ 表示第 j 棵树中的第 p 个终端节点参数。首先，固定 $l \in \{1, ..., K\}$ 和 $i \in \{1, ..., n\}$，以校准 $w_l(\boldsymbol{x}_i)$ 的先验。由于终端节点参数相互独立且同分布，并具有对角协方差结构，因此在 $w_l(\boldsymbol{x}_i)$ 上诱导的先验对于其余权重与输入组合均相同。根据式 (7) 和式 (8)，第 l 个模型权重上的诱导先验为 $w_l(\boldsymbol{x}_i) \sim N(m\beta_l, m\tau^2)$。由于认为 $w_l(\boldsymbol{x}_i)$ 以较高概率满足 $\in [0, 1]$，因此将 $m\beta_l = 0.5$ 设为相应值是合理的。因此，$\beta_l = 0.5/m$。这样，无论输入位置如何，每个权重取到 0 或 1 这两个“极端”值的机会都相同。可选择先验标准差 τ，使得 $w_l(\boldsymbol{x}_i)$ 以较高概率满足 $\in [0, 1]$。为此，构造 $w_l(\boldsymbol{x}_i)$ 的置信区间，使得 $0 = 0.5 - k\tau\sqrt{m}$ 且 $1 = 0.5 + k\tau\sqrt{m}$。将第一个方程从第二个方程中相减，并求解 τ，得到 $\tau = 1/2k\sqrt{m}$。这种校准方法与 Chipman 等人 (2010) 提出的方法非常相似。主要区别源于问题背景：我们认为权重主要落在区间 $[0, 1]$ 内，而非数据的观测范围 $[y_{min}, y_{max}]$。

4.3.2 信息先验

在信息先验设定中，先验均值直接取决于给定的树所诱导的输入空间划分，即 $\boldsymbol{\mu}_{pj} \mid T_j \sim N_K\left(\boldsymbol{\beta}_{pj},\ \tau^2 \boldsymbol{I}_K\right)$。该先验专为有效场论设计，其中函数方差 $v_l(\boldsymbol{x}_i)$ 表示截断误差的严重程度。定义域中特定子区域内的方差越大，表明截断误差越大，也就是说，该有效场论对真实系统的近似效果越差。

根据对截断误差方差的这种解释，结合有效场论的一种策略是精度加权（Phillips 等人，2021）。例如，在 $\boldsymbol{x}_i$ 处，第 l 个有效场论的精度权重为

$$\beta_l(\boldsymbol{x}_i) = \frac{1/v_l(\boldsymbol{x}_i)}{1/v_1(\boldsymbol{x}_i) + ... + 1/v_K(\boldsymbol{x}_i)}.$$

精度权重 $\beta_l(\boldsymbol{x}_i)$ 可解释为权重函数 $w_l(\boldsymbol{x}_i)$ 的初始猜测，适用于$l = 1, \ldots, K$和 $i = 1, \ldots, n$。

由于终端节点参数的先验会随树结构而变化，因此每个$\boldsymbol{\beta}_{pj}$ 都是独立于其他终端节点参数单独选取的。给定以下内容的精度权重：

有效场论和 n 训练点，先验均值向量 β_{pjl} 的每个分量均通过以下方式选取：

$$\beta_{pjl} = \frac{1}{m\sum_{i=1}^{n}\mathbf{1}(\boldsymbol{x}_i \in \eta_{pj})}\sum_{i=1}^{n}\beta_l(\boldsymbol{x}_i)\ \mathbf{1}(\boldsymbol{x}_i \in \eta_{pj}),$$

其中，$\mathbf{1}(\boldsymbol{x}_i \in \eta_{pj})$ 是一个指示函数，用于表示 $\boldsymbol{x}_i$ 是否被分配到终端节点 η_{pj}。可以为每个终端节点参数设置长度为 $1/m$ 的置信区间，以确保每棵树都是弱学习器。具体做法是设置 $\tau = 1/2km$。

4.3.3 方差先验

为误差方差 σ^2指定一个超参数为 ν 和 λ 的共轭缩放逆卡方分布。要校准该先验，首先选择一个 ν 值，以反映分布所需的形状。ν 的常见取值范围为 3 到 10。在为 λ选择取值之前，需要先对误差方差进行初步估计，以便将先验设定在 σ^2的合理取值范围附近。给定模型集以及各训练点处对应的点预测$\hat{}$ $\boldsymbol{f}_l(\boldsymbol{x}_i)$，可通过设定 $\hat{}\sigma^2 = \max_{l=1,..,K}\left\{\min_{i=1,..,n}\left(y_i - \hat{}f_l(\boldsymbol{x}_i)\right)^2\right\}$，采用轻度数据信息先验。由于普遍认为每个模型都能在定义域的某个子区域内准确近似真实系统，因此可以预期，各模型的最小平方差集合将揭示有关真实误差方差的可靠信息。基于这些信息，一种策略是将 $\hat{}\sigma^2$ 设为某个 $\lambda\nu/\chi^2_\nu$ 分布的均值或众数。然后通过求解相应方程来确定 λ 的值。

5 个有效场论示例

本节将所提出的模型混合方法应用于三个不同的示例。第5.1节展示了基于BART的混合方法在两个单变量有效场论示例中的成功应用，这两个示例在第3节中介绍。第5.2节通过使用基于三角函数泰勒级数展开的模拟器，重点介绍一个多维示例。尽管最后这个示例不涉及真实的底层物理系统，但模型集考虑了与具有双重展开的有效场论相似的模拟器（见 Burgess（2020））。每个示例都突出了所提出的基于BART的混合模型的特定特征，例如用于权重的灵活基函数及其对应的先验正则化。

5.1 示例 1：混合单变量有效场论

本节将BART模型混合（BART-BMM）方法应用于一维定义域上的各种有效场论。作为比较，也将层次堆叠（HS）应用于同一组有效场论。在这两个有效场论示例中，按照以下方式独立生成20个观测值：

$$Y_i = f_\dagger(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

其中 $i = 1, ...,20$、 $\sigma = 0.005$ 和 $f_\dagger(x)$ 定义见 (1)。20 个训练点对应的输入在 0.03 至 0.50 的区间内均匀分布。误差标准差设为 0.005，以模拟受控实验环境。每个有效场论模型均使用 $n_c = 4$ 次对应有限阶展开的计算结果进行拟合。

5.1.1 示例 1a：混合两个有效场论

首先，考虑混合基于二阶弱耦合展开的有效场论 $f_s^{(2)}(x)$ 和基于四阶强耦合展开的有效场论 $f_l^{(4)}(x)$，如图 1(a) 所示。真实系统 $f_\dagger(x)$ 在整个定义域内均位于两个有效场论之间，因此，对两个有效场论的预测进行凸组合，适合用于恢复真实系统。BART-BMM 模型使用 10 棵树和 $k = 5.0$ 进行拟合。同时，HS 无约束权重函数定义为$w_1^*(x) = \mu_1 + \alpha_1 x$。BART-BMM 方法和 HS 的结果见图3。

就预测系统与真实$f_\dagger(x)$之间的均方根误差（RMSE）而言，与 HS 相比，BART-BMM 模型给出的均值预测更准确，其均方根误差分别为 0.0053 和 1.9460。HS 的均方根误差偏高，是因为其混合预测在定义域左侧部分发散。例如，HS 模型在区间 $[0.1, 0.5]$ 上的均方根误差降至 0.0717。此外，从图3可以明显看出，BART-BMM 对 $f_\dagger(x)$ 的预测不确定性低于 HS。

图3中的权重函数也呈现出类似的类S形形状，但HS解表现出高度的不确定性。两种方法之间最明显的差异体现在 $f_l^{(4)}(x)$的权重函数上（虚线）。具体而言，与HS结果相比，BART-BMM结果中的曲线在子区域 $[0.3, 0.4]$ 中的增长速度更快。这种较慢的增长速度导致HS在该子区域中的预测效果较差。另一个差异出现在 $[0.03, 0.15]$区域：在HS方法下， $f_l^{(4)}(x)$的权重接近0，但还不足以抵消$f_l^{(4)}(x)$所产生的均值预测大幅发散的影响。与此同时，由于

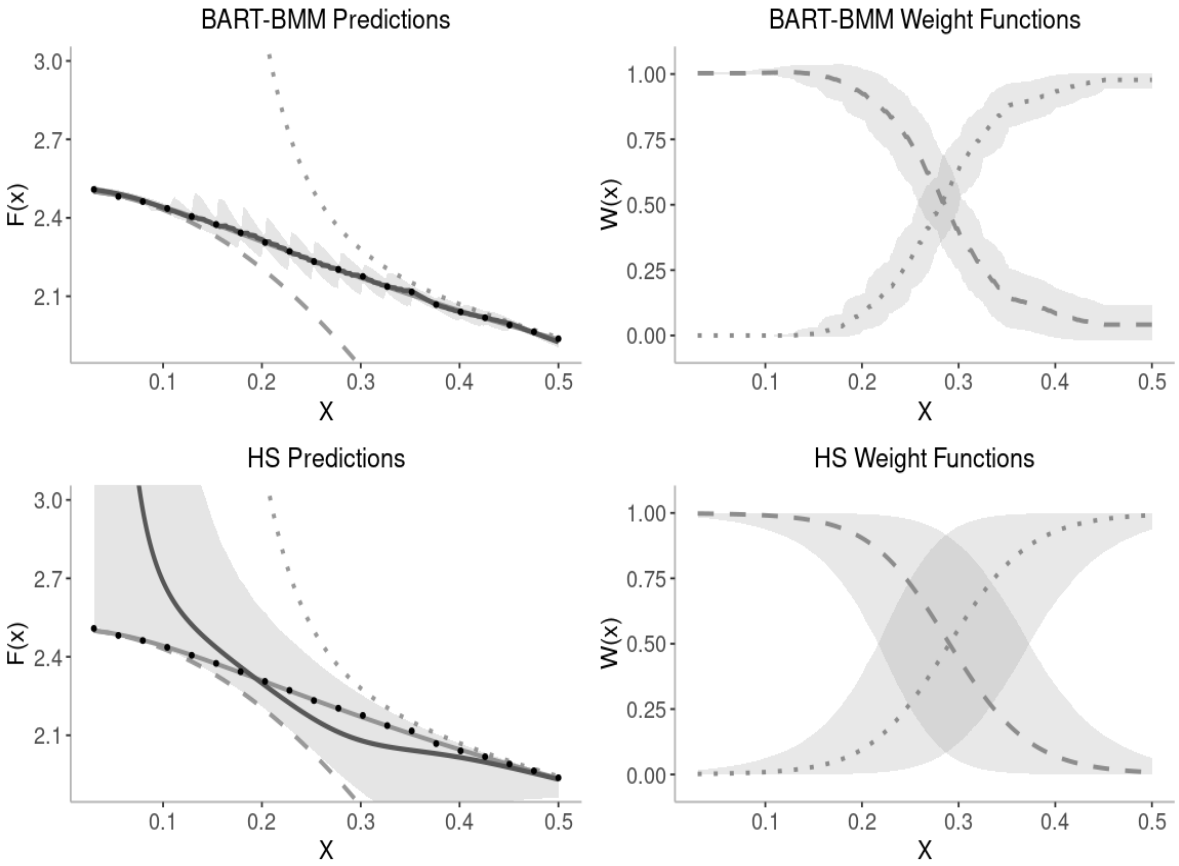


图3：混合 $f_s^{(2)}(x)$（虚线）和 $f_l^{(4)}(x)$（点线）时的预测均值（深灰色）和 95%可信区间（阴影部分）。结果来自包含 10 棵树的 BART-BMM模型，以及采用线性无约束权重函数的层次堆叠模型（下图）。

均值估计目标会直接重新加权单个模型的均值预测，且不存在单纯形约束，BART-BMM权重在不确定性极小的情况下被压缩至接近0。

BART-BMM的另一个优势是，权重函数通过树模型在马尔可夫链蒙特卡洛过程中不断学习。这与HS不同，HS需要预先指定无约束权重的基。在这个例子中，可以考虑采用不同的基函数，因为指定的线性基函数似乎不足以在整个定义域内获得高保真度的均值预测。

5.1.2 例1b：混合两个凸EFT

现在，考虑图1(b)所示的第二个模型集，它用$f_s^{(4)}(x)$替代了 $f_s^{(2)}(x)$。两个有效场论在中间范围内都高估了 $f_{\dagger}(x)$，因此，被限制在

单纯形内的权重无法恢复真实系统。在这种情况下，如下所示，将一个分段基函数分配给不受约束的 HS 权重，

$$w_1^*(x) = \mu_1 + \alpha_1 \mathbf{1}(x < 0.15) + \alpha_2 \mathbf{1}(0.15 \leq x < 0.25) + \alpha_3 \mathbf{1}(0.25 \leq x < 0.35).$$

选择这一基函数，是为了大致反映均值预测开始以不同速率变化的区域。对于分段基函数而言，其他划分方式同样有效。

BART-BMM 和 HS 的结果如图 4 所示。再次可以看到，在均值预测方面，BART-BMM 方法优于 HS，二者的均方根误差分别为 0.0057 和 0.1141。值得注意的是，由于模型权重受到单纯形约束，HS解无法准确预测定义域中间范围内的真实系统。与此同时，BART-BMM 方法采用了针对权重的先验正则化方法，不会施加如此严格的约束，因此能够在整个定义域内恢复该系统。

从这一 HS 结果来看，在预测定义域左右两侧的真实系统时，分段基函数似乎比线性基函数更有效。这也进一步引出了一个问题：如何选择分段基函数所产生的划分，因为不同的选择可能导致截然不同的结果。这个问题促使我们定义了基于 BART 的模型，该模型能够根据观测数据和模型集自适应地学习这些划分。

5.2 示例 2：多维模型混合

所提出的模型混合方法也适用于依赖多维输入的计算机实验。为说明这一点，考虑一个二维问题，其中真实的底层系统定义为

$$f_\dagger(\boldsymbol{x}) = \sin(x_1) + \cos(x_2),$$

其中 $\boldsymbol{x} = (x_1,\ x_2)^\top \in [-\pi, \pi] \times [-\pi, \pi]$. 从该真实系统中生成了 80 个训练点，观测误差标准差为 0.1。此外，还考虑了两个候选模型，其模拟器分别以 $s(x_i) := \sin(x_1)$ 和 $c(x_2) := \cos(x_2)$ 的泰勒级数展开为定义依据。对于此示例，模拟器定义为

$$h_1(\boldsymbol{x}) = \sum_{j=0}^{7} \frac{s^{(j)}(x_1)}{j!}(x_1 - \pi)^j + \sum_{k=0}^{10} \frac{c^{(k)}(x_2)}{k!}(x_2 - \pi)^k$$

$$h_2(\boldsymbol{x}) = \sum_{j=0}^{13} \frac{s^{(j)}(x_1)}{j!}(x_1 + \pi)^j + \sum_{k=0}^{6} \frac{c^{(k)}(x_2)}{k!}(x_2 + \pi)^k$$

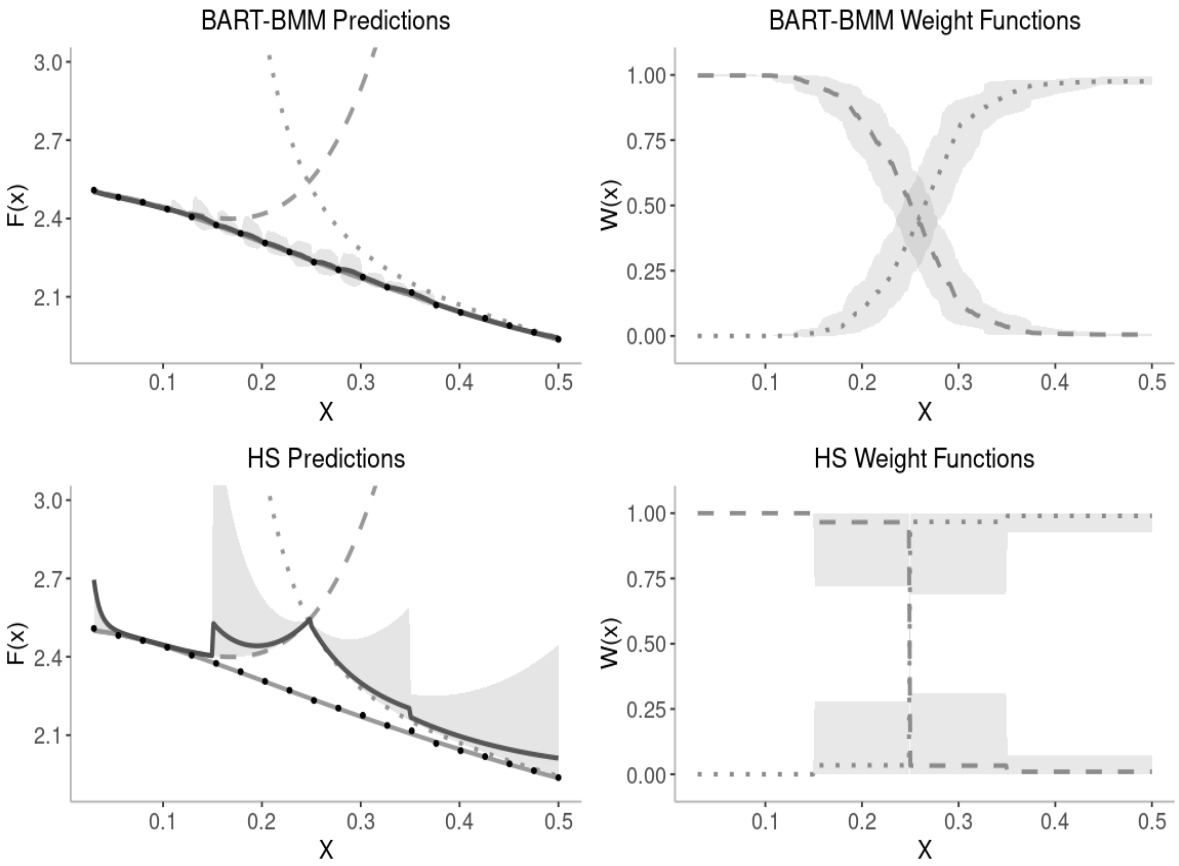


图 4：混合 $f_s^{(4)}(x)$（虚线）和 $f_l^{(4)}(x)$（点线）时的预测均值（深灰色）和 95%可信区间（阴影部分）。结果分别来自包含 10 棵树的 BART-BMM 模型（上图）和采用分段无约束权重函数的层次堆叠模型（下图）。

其中 $s^{(j)}(x_1)$ 和 $c^{(k)}(x_2)$ 分别表示 $\sin(x_1)$ 和 $\cos(x_2)$ 的第 j 阶和第 k 阶导数。请注意，第一个模拟器 $h_1(\boldsymbol{x})$ 将两个泰勒级数展开的中心都设在 π，因此它在定义域右上角能够对系统作出相对准确的预测，而向定义域的负值区域移动时则会逐渐偏离。与此同时，$h_2(\boldsymbol{x})$ 由以 $-\pi$ 为中心的泰勒级数展开构成，因此能够在负值区域作出准确预测。两个模拟器的一个关键区别在于，$h_2(\boldsymbol{x})$ 包含了对整个区间 $[-\pi, \pi]$ 内 $\sin(x_1)$ 的高度准确近似，因为其对应的泰勒级数展开由 7 个非零项组成。因此，尽管 $\sin(x_1)$ 和 $\cos(x_2)$ 的展开以 $-\pi$ 为中心，仍可预期 $h_2(\boldsymbol{x})$ 能够在整个矩形 $[-\pi, \pi] \times [-\pi, 0]$ 内准确预测 $f_\dagger(\boldsymbol{x})$。

每个模型的理论预测 $f_1(\boldsymbol{x})$ 和 $f_2(\boldsymbol{x})$ 可使用加性

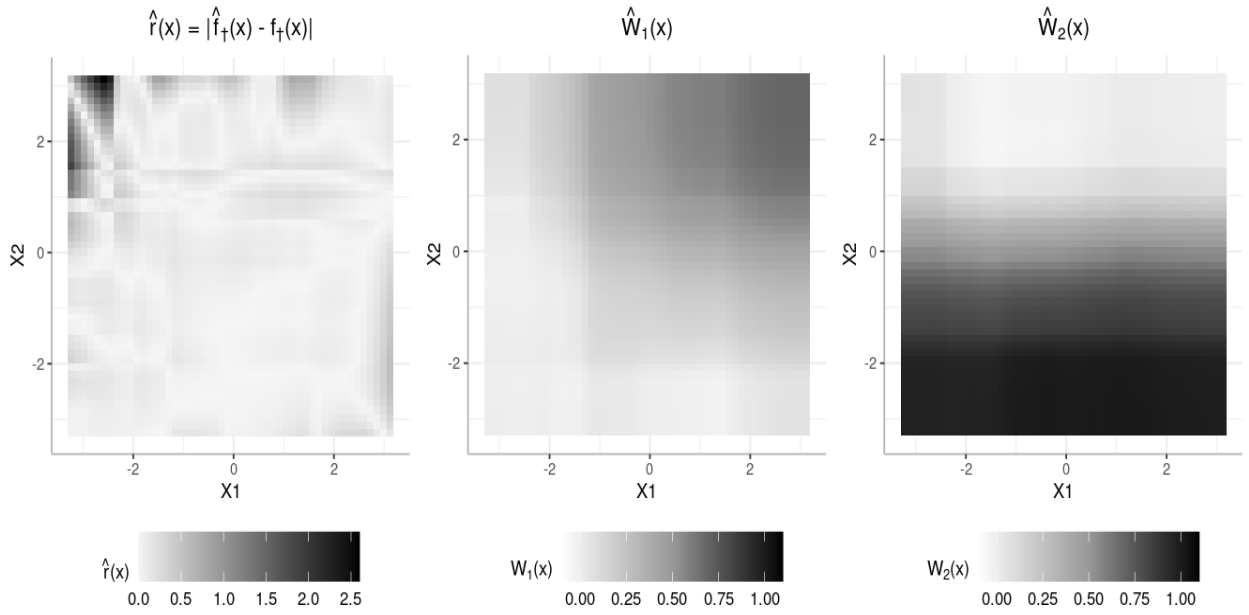


图5：（左）预测系统 $f_{\dagger}(\boldsymbol{x})$ 与真实系统$f_{\dagger}(\boldsymbol{x})$之间的均值差。（中） $h_1(\boldsymbol{x})$的平均权重函数。（右） $h_2(\boldsymbol{x})$的平均权重函数。

形式 $f_l(\boldsymbol{x}) = h_l(\boldsymbol{x}) + \delta_l(\boldsymbol{x})$ 来定义，其中 $\delta_l(\boldsymbol{x})$ 表示未知的高阶修正，而$l = 1, 2$。由于这个示例的性质，对于 $\delta_l(\boldsymbol{x})$ 并未假设任何模型。因此，$\hat{}$ 在每个训练点 $\boldsymbol{x}_i$ 处估计的理论预测为 $f_l(\boldsymbol{x}_i) = h_l(\boldsymbol{x}_i)$。请注意，在多维有效场论的设置中，第3节所讨论的策略仍然适用。

30棵树的BART-BMM模型的结果如图5所示。最左侧的图展示了平均残差的绝对值，$\hat{} \hat{} \ |f_{\dagger}(\boldsymbol{x}) - f_{\dagger}(\boldsymbol{x})|$，其中 $f_{\dagger}(\boldsymbol{x})$表示BMM模型的均值预测。从残差图来看， $f_{\dagger}(\boldsymbol{x})$在定义域的大部分区域都得到了充分恢复，均方根误差为0.2575。正如预期，定义域左上角的均值预测误差明显增大；该区域仅包含两个训练点，且两个模拟器都不准确。

第二幅和第三幅图展示了每个模拟器的后验平均权重函数。根据中间的图，随着 x_1 和 x_2 都增大，第一个模拟器的效用也会提高。这是意料之中的，因为 $h_1(\boldsymbol{x})$由两个以 π为中心的展开组成。请注意，由于模拟器在该区域略微高估了 $f_{\dagger}(\boldsymbol{x})$的峰值，因此 $w_1(\boldsymbol{x})$的均值在定义域的右上角未达到 1。与此同时， $w_2(\boldsymbol{x})$的后验均值表明，$h_2(\boldsymbol{x})$对于 $\boldsymbol{x} \in [-\pi, \pi] \times [-\pi, 0]$具有较高的效用，这与该模拟器所包含的展开的性质相符。此外， $h_2(\boldsymbol{x})$的预测似乎与数据高度吻合

，因此 $f_{\mathrm{t}}(\boldsymbol{x})$也与数据高度吻合；定义域下半部分的权重接近 1，便可看出这一点。

6 讨论

模型平均与混合有多种频率学派和贝叶斯方法。每种方法都涉及基于各个模型估计整体预测均值或预测密度。最终应根据希望确定的底层统计推断目标，在这两种目标之间作出选择。在计算机实验中，一个主要目标是恢复底层系统，这通常表示为观测数据加性模型中的均值函数。因此，在这种情境下，与旨在预测未来观测值 $\tilde{y}$ 的预测密度估计相比，均值估计方法更为可取。

示例 5.1 比较了所提出的均值估计方法与层次堆叠（HS）中的密度估计方法。在 HS 中，权重函数是在单纯形约束下，根据留一法（LOO）预测密度学习得到的。这些 LOO 密度纳入了有关给定 x 处各个有效场论（EFT）的均值和方差信息。在模型可能迅速发散的定义域部分，所得的 LOO 预测密度会向 0 收缩。相应地，权重函数也会趋近于 0；不过，它可能难以取得足够小的值，以消除发散均值的影响。相比之下，混合各个 EFT 的均值预测，似乎更容易收缩发散预测的影响。

权重函数的主要目的是重新缩放各个模型给出的预测，使这些预测的线性组合能够充分还原真实系统。鉴于应用于权重函数的先验正则化方法，所得数值的确切解释可能并不明确。不过，从这种正则化视角来看，可以得出结论：在特定子区域内接近 0 的权重函数表明，相应模型对于整体预测而言并非必需。与此同时，在某个特定区域内作为唯一局部专家的模型，其权重应接近 1。总体而言，适合对权重函数进行联合解释，尤其是在权重集中于远离 0 或 1 的数值附近的区域。每个示例中都观察到了这些特征。

通过权重函数之和的后验分布，可以进一步理解所提出的正则化方法的优势：$w_{\mathrm{sum}}(x) = \sum_{l=1}^{K} w_l(x)$，如

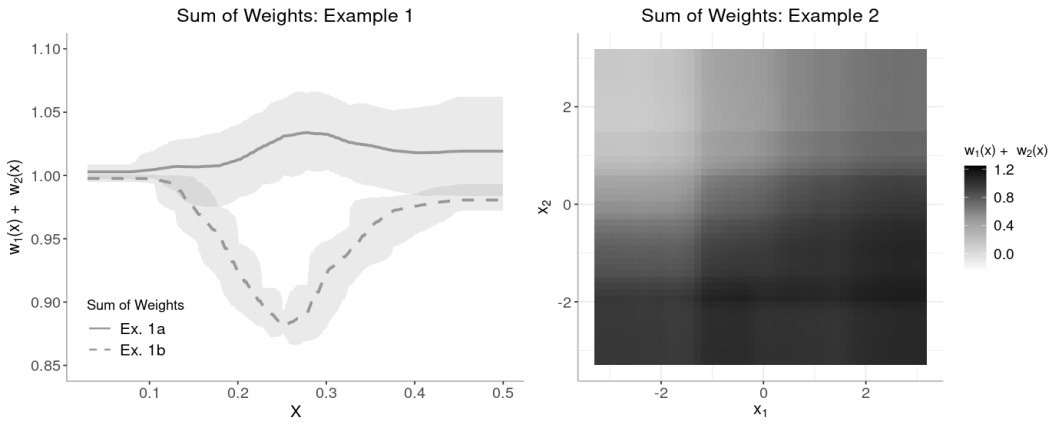


图 6：（左）示例 1a 和 1b 中权重函数之和的后验均值估计及 95%可信区间（阴影部分）（分别以实线和虚线表示）。（右）示例 2 中权重函数之和的后验均值估计。

图6所示。例1a中 $w_{\text{sum}}(x)$ 的后验分布（左图，实线）中心非常接近1，且不确定性相对较小。这是因为：(i) 先验正则化；(ii) $f_\dagger(x)$ 位于所选有效场论之间，这表明凸组合是合适的。尽管没有严格施加和为一性质，但在适合对竞争模型进行插值的情况下，这一性质似乎会自然出现。与此同时，例1b中 $w_{\text{sum}}(x)$ 的后验分布（左图，虚线）在定义域的中间范围显著低于1，因为两个有效场论都高估了真实系统，因此凸组合并不合适。在二维示例中也观察到了类似特征：在至少一个模拟器与真实系统吻合的区域，$w_{\text{sum}}(x)$ 的均值集中在1附近。与此同时，当两个模拟器都不准确时（即左上角），$w_{\text{sum}}(x)$ 的均值远低于1。根据这些观察，BART-BMM 方法的优势似乎在于不对权重施加单纯形约束等严格假设。

最后，权重函数可用于更好地理解与模型集相关的 $\mathcal{M}$-开放假设。当权重函数随输入发生明显变化时，可以初步确认 $\mathcal{M}$-开放设定。这一观察表明每个模型的表现具有局部性，因此可以确认真实系统不包含在该模型集中。如果权重函数近乎恒定，也可以查看 $w_{\text{sum}}(x)$ 的后验分布，了解权重之和是否固定在接近 1 的位置。这种情况可能表明，带有单纯形约束的模型平均也可能是合适的解决方案。仅凭这一点还不足以确认

或否定 $\mathcal{M}$-开放假设，但这可能表明模型集满足 $\mathcal{M}$-完备或 $\mathcal{M}$-封闭假设。最后还需考虑这样一种情况：某个模型的权重接近 1，而竞争模型的影响在定义域的某个子区域内被压缩至 0。尽管在考虑整个定义域时属于 $\mathcal{M}$-开放情形，但在所关注的子区域内，该情况可能表明模型集是 $\mathcal{M}$-封闭的。

总之，本文提出了一种贝叶斯树形框架，用于混合一组竞争模型的预测结果；这些模型各自旨在解释定义域某个子区域内的物理系统。该方法属于贝叶斯模型混合这一类问题，因为它定义了输入依赖权重，以反映各模型的局部行为。权重函数采用树之和进行建模，并通过多元高斯先验加以正则化。树基函数与正则化方法相结合，使权重能够以灵活的非参数方式进行学习，不受严格约束。利用权重函数，通过线性组合混合各个模型的预测结果。本文通过三个示例展示了这种混合方法的有效性，每个示例都考察了预测性能具有局部性的模型。与现有的全局和局部加权方案相比，利用各个模型的局部行为，显著提升了 $f_{\dagger}(x)$的后验预测和不确定性量化，并改善了对系统的整体解释。

致谢

JCY 和 RJF 的工作部分由美国国家科学基金会根据协议 OAC-2004601 资助。MTP 的工作部分由美国国家科学基金会根据协议 DMS-1916231、DMS-1564395、OAC-2004601 资助，部分由阿卜杜拉国王科技大学（KAUST）赞助研究办公室（OSR）根据资助编号 OSR-2018-CRG7-3800.3 资助。TJS 的工作部分由美国国家科学基金会根据协议 DMS-1564395（俄亥俄州立大学）资助。

参考文献

Bernardo, J. M. 和 Smith, A. F.（1994），贝叶斯理论，约翰·威立父子出版公司。Breiman, L.（1996），“堆叠回归”，*Machine Learning*24(1)，49–64。Bunea, F.、Tsybakov, A. B. 和 Wegkamp, M. H.（2007），“高斯回归的聚合”。

Burgess, C. P.（2020），《有效场论导论：有效地思考尺度层级》，剑桥大学出版社。Burnham, K. P.、Anderson, D. R.、Burnham, K. P. 和 Anderson, D. R.（1998），信息论方法的实际应用，施普林格。Chipman, H. A.、George, E. I. 和 McCulloch, R. E.（1998），“贝叶斯 CART 模型搜索”，《美国统计学会杂志》**93**(443)，935–948。Chipman, H. A.、George, E. I. 和 McCulloch, R. E.（2002），“贝叶斯树模型”，*MachineLearning* **48**(1)，299–320。Chipman, H.、George, E. 和 McCulloch, R.（2010），“BART：贝叶斯加性回归树”，《应用统计年鉴》**4**(1)，266–298。Clyde, M. 和 Iversen, E. S.（2013），“M 开放框架下的贝叶斯模型平均”，贝叶斯理论与应用第 484–498 页。Draper, D.（1995），“模型不确定性的评估与传播”，《英国皇家统计学会杂志：*B*辑（方法论）》**57**(1)，45–70。Georgi, H.（1993），“有效场论”，《核与粒子科学年度评论》**43**，209–252。Gramacy, R. B.（2020），《代理模型：应用科学中的高斯过程建模、设计与优化》，Chapman and Hall/CRC。Gramacy, R. B. 和 Lee, H. K. H.（2008），“贝叶斯树形高斯过程模型及其在计算机建模中的应用”，《美国统计学会杂志》**103**(483)，1119–1130。Hansen, B. E.（2007），“最小二乘模型平均”，《计量经济学》**75**(4)，1175–1189。Hastie, T. 和 Tibshirani, R.（2000），“贝叶斯回拟合（含评论及作者答辩）”，《统计科学》**15**(3)，196–223。Honda, M.（2014），“通过强耦合展开改进的微扰理论”，《高能物理学杂志》**2014**(12)，1–44。Le, T. 和 Clarke, B.（2017），“M 完备和 M 开放情境下堆叠的贝叶斯解释”，《贝叶斯分析》**12**(3)，807–829。

Melendez, J. A., Furnstahl, R. J., Phillips, D. R., Pratola, M. T. 和 Wesolowski, S.（2019年），《量化有效场论中的相关截断误差》，《物理评论*C*》**100**(4)。

Melendez, J., Furnstahl, R., Grießhammer, H., McGovern, J., Phillips, D. 和 Pratola, M.（2021），《设计最优实验：质子康普顿散射的一个应用》，《欧洲物理杂志*A*》**57**(3)，1–24。

Petrov, A. A. 和 Blechman, A. E.（2016年），有效场论，世界科学出版社。**网址：** *https://www.worldscientific.com/doi/abs/10.1142/8619*

Phillips, D., Furnstahl, R., Heinz, U., Maiti, T., Nazarewicz, W., Nunes, F., Plumlee, M., Pratola, M., Pratt, S., Viens, F. 等（2021），《加入 BAND 行列：一种用于量化核动力学中模型不确定性的贝叶斯框架》，《物理学杂志*G*：核与粒子物理》**48**(7)，072001。

Prado, E. B., Moral, R. A. 和 Parnell, A. C.（2021），“带模型树的贝叶斯加性回归树”，《统计学与计算》**31**(3),1–13。

Pratola, M. T.（2016），“贝叶斯回归树模型的高效 Metropolis–Hastings 提议机制”，《贝叶斯分析》**11**(3), 885–911。

Raftery, A., Madigan, D. 和 Hoeting, J.（1997），“线性回归模型的贝叶斯模型平均”，《美国统计学会杂志》**92**(437),179–191。

Ravishanker, N., Chi, Z. 和 Dey, D. K.（2021），《线性模型理论入门》，Chapman and Hall/CRC。

Santner, T. J., Williams, B. J. 和 Notz, W. I.（2018），《计算机实验的设计与分析》，第2版，施普林格。

Semposki, A.、Furnstahl, R. 和 Phillips, D.（2022），“无处不在的不确定性：使用贝叶斯模型混合在小型和大型g展开之间进行插值”，*arXiv* 预印本 *arXiv:2206.04116* 。

Sill, J.、Tak´acs, G.、Mackey, L. 和 Lin, D.（2009），“特征加权线性堆叠”，*arXiv* 预印本 *arXiv:0911.0460*。

Yang, Y. 和 Dunson, D. B.（2014），《极小极大最优贝叶斯聚合》，*arXiv* 预印本 *arXiv:1403.1345* 。

Yao, Y.、Pirˇs, G.、Vehtari, A. 和 Gelman, A.（2021），《贝叶斯层次堆叠：某些模型在（某些方面）有用》，《贝叶斯分析》**1**(1)，1–29。

Yao, Y., Vehtari, A., Simpson, D. 和 Gelman, A.（2018），“使用堆叠对贝叶斯预测分布进行平均”，《贝叶斯分析》**13**(3),917–1007。

附录

令 η_{pj} 表示第 p 棵树中的第 j 个终端节点。不失一般性，假设 (x_1, y_1) ...，(x_{n_p}, y_{n_p}) 位于由 η_{pj} 定义的超矩形内。此外，将每个残差定义为

$$r_i = y_i - \sum_{q \neq j} \hat{\boldsymbol{f}}^{\top}(\boldsymbol{x}_i)\, \boldsymbol{g}(\boldsymbol{x}_i, T_q, M_q), \quad i = 1, \ldots, n_p$$

将这些量收集成一个 n_p 维向量 $\boldsymbol{R}_{pj} = (r_1, ..., r_{n_p})^{\top}$。最后，用 $\hat{} \boldsymbol{F}_{pj}$ 表示一个矩阵，其第 l 列为 $(\boldsymbol{f}_l(\boldsymbol{x}_1) ..., \boldsymbol{f}_l(\boldsymbol{x}_{n_p}))^{\top}$。根据独立性和方差恒定假设，残差向量的模型及其相应的先验定义如下：

$$\boldsymbol{R}_{pj} \mid \boldsymbol{\mu}_{pj}, T_j, \sigma^2 \sim N_{n_p}\Big(\hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj}, \sigma^2 I_{n_p}\Big)$$

$$\boldsymbol{\mu}_{pj} \mid T_j \overset{ind}{\sim} N_K(\boldsymbol{\beta}_{pj}, \boldsymbol{\Sigma})$$

$$\sigma^2 \sim \lambda\nu/\chi^2_{\nu}$$

其中假设 $\boldsymbol{\Sigma} = \tau^2 I_K$。

边际似然

节点 η_{pj} 中残差的边际似然定义如下：

$$L(\boldsymbol{R}_{pj} \mid T_j, \sigma^2) = \int L(\boldsymbol{R}_{pj} \mid T_j, \boldsymbol{\mu}_{pj}, \sigma^2)\pi(\boldsymbol{\mu}_{pj} \mid T_j)\, d\boldsymbol{\mu}_{pj} \tag{9}$$

因此，

$$\begin{aligned}
L(\boldsymbol{R}_{pj} \mid T_j, \sigma^2) = &\int (2\pi\sigma^2)^{-n_p/2} \exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{R}_{pj} - \hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj})^{\top}(\boldsymbol{R}_{pj} - \hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj})\Big)\times \\
&(2\pi\tau^2)^{-K/2} \exp\Big(-\frac{1}{2\tau^2}(\boldsymbol{\mu}_{pj} - \boldsymbol{\beta}_{pj})^{\top}(\boldsymbol{\mu}_{pj} - \boldsymbol{\beta}_{pj})\Big)\, d\boldsymbol{\mu}_{pj} \\
= &(2\pi\sigma^2)^{-n_p/2}(2\pi\tau^2)^{-K/2}\times \\
&\int \Big\{ \exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{R}_{pj}^{\top}\boldsymbol{R}_{pj} - 2\boldsymbol{\mu}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj} + \boldsymbol{\mu}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj}\Big)\times \\
&\exp\Big(-\frac{1}{2\tau^2}(\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{\mu}_{pj} - 2\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{\beta}_{pj} + \boldsymbol{\beta}_{pj}^{\top}\boldsymbol{\beta}_{pj})\Big)\, d\boldsymbol{\mu}_{pj}\Big\} \\
= &(2\pi\sigma^2)^{-n_p/2}(2\pi\tau^2)^{-K/2} \exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{R}_{pj}^{\top}\boldsymbol{R}_{pj} - \frac{1}{2\tau^2}\boldsymbol{\beta}_{pj}^{\top}\boldsymbol{\beta}_{pj}\Big)\times \\
&\int \exp\Big(-\frac{1}{2}\boldsymbol{\mu}_{pj}^{\top}\Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}\boldsymbol{I}_K\Big)\boldsymbol{\mu}_{pj} + \Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}\Big)^{\top}\boldsymbol{\mu}_{pj}\Big)\, d\boldsymbol{\mu}_{pj}.
\end{aligned}$$

现在令 $\boldsymbol{A}^{-1} = \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K$ 和 $\boldsymbol{b} = \Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}\Big)$。将这些项代入上述表达式，得到

$$L(\boldsymbol{R}_{pj} \mid T_j, \sigma^2) = (2\pi\sigma^2)^{-n_p/2}(2\pi\tau^2)^{-K/2}\exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{R}_{pj}^{\top}\boldsymbol{R}_{pj} - \frac{1}{2\tau^2}\boldsymbol{\beta}_{pj}^{\top}\boldsymbol{\beta}_{pj}\Big)\times \tag{10}$$
$$\int \exp\Big(-\frac{1}{2}\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{A}^{-1}\boldsymbol{\mu}_{pj} + \boldsymbol{b}^{\top}\boldsymbol{\mu}_{pj}\Big)\, d\boldsymbol{\mu}_{pj}$$

利用 Santner 等人（2018）的引理 B.1，积分可简化为

$$\int \exp\Big(-\frac{1}{2}\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{A}^{-1}\boldsymbol{\mu}_{pj} + \boldsymbol{b}^{\top}\boldsymbol{\mu}_{pj}\Big)\, d\boldsymbol{\mu}_{pj} = (2\pi)^{K/2}|\boldsymbol{A}|^{1/2}\exp\Big(\frac{1}{2}\boldsymbol{b}^{\top}\boldsymbol{A}\boldsymbol{b}\Big). \tag{11}$$

然后，根据 (10) 和 (11)，边际似然可简化为

$$\begin{aligned} L(\boldsymbol{R}_{pj} \mid T_j, \sigma^2) &= (2\pi\sigma^2)^{-n_p/2}(\tau^2)^{-K/2}|\boldsymbol{A}|^{1/2}\exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{R}_{pj}^{\top}\boldsymbol{R}_{pj} - \frac{1}{2\tau^2}\boldsymbol{\beta}_{pj}^{\top}\boldsymbol{\beta}_{pj} + \frac{1}{2}\boldsymbol{b}^{\top}\boldsymbol{A}\boldsymbol{b}\Big). \\ &= (2\pi\sigma^2)^{-n_p/2}(\tau^2)^{-K/2}\left|\Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}\boldsymbol{I}_K\Big)^{-1}\right|^{1/2} \\ &\quad \times \exp\Big(-\frac{1}{2}\Big(\frac{1}{\sigma^2}\boldsymbol{R}_{pj}^{\top}\boldsymbol{R}_{pj} + \frac{1}{\tau^2}\boldsymbol{\beta}_{pj}^{\top}\boldsymbol{\beta}_{pj} - \boldsymbol{b}^{\top}\boldsymbol{A}\boldsymbol{b}\Big)\Big) \end{aligned}$$

where $\boldsymbol{b}^{\top}\boldsymbol{A}\boldsymbol{b} = \Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}\Big)^{\top}\Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K\Big)^{-1}\Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}\Big)$.

$\boldsymbol{\mu}_{pj}$ 的全条件分布

现在考虑终端节点参数 $\boldsymbol{\mu}_{pj}$ 的全条件后验分布。根据贝叶斯定理,

$$\pi(\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2) \propto L(\boldsymbol{R}_{pj} \mid T_j, \boldsymbol{\mu}_{pj}, \sigma^2)\pi(\boldsymbol{\mu}_{pj} \mid T_j)$$

假设 μ_{pj} 服从共轭先验，因此可以重新排列似然和先验中的各项，从而得到后验分布的正态核。此过程总结如下。

$$\begin{aligned} \pi(\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2) &\propto \exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{R}_{pj} - \hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj})^{\top}(\boldsymbol{R}_{pj} - \hat{\boldsymbol{F}}_{pj}\boldsymbol{\mu}_{pj})\Big)\times \\ &\qquad \exp\Big(-\frac{1}{2\tau^2}(\boldsymbol{\mu}_{pj} - \boldsymbol{\beta}_{pj})^{\top}(\boldsymbol{\mu}_{pj} - \boldsymbol{\beta}_{pj})\Big) \\ &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{pj}^{\top}\Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K\Big)\boldsymbol{\mu}_{pj} - 2\boldsymbol{\mu}_{pj}^{\top}\Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}\Big)\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{A}^{-1}\boldsymbol{\mu}_{pj} - 2\boldsymbol{\mu}_{pj}^{\top}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{b}\right)\right\} \end{aligned}$$

其中 $\boldsymbol{A}^{-1} = \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K$和 $\boldsymbol{b} = \frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^{\top}\boldsymbol{R}_{pj}$。前一个表达式可简化为

$$\pi(\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{\mu}_{pj} - A\boldsymbol{b})^{\top}\boldsymbol{A}^{-1}(\boldsymbol{\mu}_{pj} - A\boldsymbol{b})\Big)$$

这是一个均值为 $A\boldsymbol{b}$ 、协方差矩阵为 $\boldsymbol{A}$ 的多元高斯分布的核。因此可得

$$\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2 \overset{ind}{\sim} N_K\Big(A\boldsymbol{b}, \boldsymbol{A}\Big)$$

将 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 替换为各自的定义可得

$$\boldsymbol{\mu}_{pj} \mid \boldsymbol{R}_{pj}, T_j, \sigma^2 \overset{ind}{\sim} N_K\left(\Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K\Big)^{-1}\Big(\frac{1}{\tau^2}\boldsymbol{\beta}_{pj} + \frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\boldsymbol{R}_{pj}\Big), \Big(\frac{1}{\sigma^2}\hat{\boldsymbol{F}}_{pj}^\top\hat{\boldsymbol{F}}_{pj} + \frac{1}{\tau^2}I_K\Big)^{-1}\right)$$

σ^2 的全条件分布

最后， 考虑误差方差的全条件后验分布，其定义为 b y

$$\pi(\sigma^2 \mid \boldsymbol{Y}, T, M) \propto L(\boldsymbol{Y} \mid T, M, \sigma^2)\pi(\sigma^2)$$

其中 $\boldsymbol{Y} = (y_1, \ldots\ldots, y_n)^\top$、 $T = \{T_1, ..., T_m\}$ 和 $M = \{M_1, ..., M_m\}$。

进一步地，假设 σ^2 的共轭先验为 $\sigma^2 \sim \nu\lambda/\chi^2_\nu$ ，其概率密度函数定义为

$$\pi(\sigma^2) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\lambda^{\nu/2}(\sigma^2)^{-(\nu/2+1)}\exp\Big(-\frac{\nu\lambda}{2\sigma^2}\Big)$$

由于共轭性，全条件分布为

$$\pi(\sigma^2 \mid \boldsymbol{Y}, T, M) \propto (\sigma^2)^{-n/2}\exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\Big(y_i - \hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i)\Big)^2\Big\}(\sigma^2)^{-(\nu/2+1)}\exp\Big\{-\frac{\nu\lambda}{2\sigma^2}\Big\}$$

$$\propto (\sigma^2)^{-(n/2+\nu/2+1)}\exp\Big\{-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n}\Big(y_i - \hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i)\Big)^2 + \nu\lambda\Big)\Big\}$$

这是另一种缩放逆-χ^2 分布的核，即 $\sigma^2 \sim \nu'\lambda'/\chi^2_{\nu'}$ 其中

$$\nu' = n + \nu \quad \text{and} \quad \lambda' = \frac{1}{n+\nu}\Big(\sum_{i=1}^{n}\Big(y_i - \hat{\boldsymbol{f}}^\top(\boldsymbol{x}_i)\boldsymbol{w}(\boldsymbol{x}_i)\Big)^2 + \nu\lambda\Big)$$

补充材料

有效场论概述

有效场论通过无穷级数展开来描述物理系统，这些项按照幂次计数原理，依重要性递减的顺序排列（Burgess，2020年；Petrov 和 Blechman，2016年；Georgi，1993年）。将这些项求和即可得到系统的精确理论预测。实际上，我们只知道有限个低阶项。因此，可以用一个类泰勒级数来分解理论预测，其中包括已知的有限阶展开以及由此产生的截断误差。实验量的预测随后可以用加性模型表示

$$Y(\boldsymbol{x}) = f^{(N)}(\boldsymbol{x}) + \epsilon(\boldsymbol{x})$$
$$f^{(N)}(\boldsymbol{x}) = h^{(N)}(\boldsymbol{x}) + \delta^{(N)}(\boldsymbol{x})$$

其中， $\boldsymbol{x} \in \mathbb{R}^d$ 表示与系统相关的自变量， $h^{(N)}(\boldsymbol{x})$表示已知的 N阶有限阶展开， $\delta^{(N)}(\boldsymbol{x})$是相应的截断误差，而 $\epsilon(\boldsymbol{x})$是随机观测误差。有限阶展开的精度在定义域的某个子区域内可能有显著差异。例如，以零为中心的有限阶展开可能在定义域的较低区域内产生高保真近似。然而，预测的精度在定义域的较高区域内会迅速下降。

此外，假设有限阶展开可以建模为随机过程。首先，有限阶展开可以分解为

$$h^{(N)}(\boldsymbol{x}) = y_{\mathrm{ref}}(\boldsymbol{x}) \sum_{k=0}^{N} c_k(\boldsymbol{x}) Q^k(\boldsymbol{x}), \tag{12}$$

其中， $y_{\mathrm{ref}}(\boldsymbol{x})$设定了变化尺度， $c_0(\boldsymbol{x})$ ...， $c_N(\boldsymbol{x})$是无量纲可观测系数，而 $Q(\boldsymbol{x})$是无量纲展开参数。当根据理论论证已知变化尺度和展开参数时，系数 $c_0(\boldsymbol{x})$ ...， $c_N(\boldsymbol{x})$似乎表现为来自某个随机过程的一组独立同分布曲线（Melendez et al., 2019）。因此，系数的一种常用模型是高斯过程

$$c_k(\boldsymbol{x}) \mid \boldsymbol{\theta} \sim GP(\mu, \bar{c}^2 r(\boldsymbol{x}, \boldsymbol{x}'; \ell)) \tag{13}$$
$$\boldsymbol{\theta} = (\mu, \bar{c}^2, \ell),$$

其中， μ 表示常数均值函数，而 $r(\boldsymbol{x},\ \boldsymbol{x}';\ell)$表示协方差函数。一种常见假设是设定 $\mu = 0$，同时可为模型中的其余参数指定先验分布（Melendez et al., 2019）。

式（13）中的参数通过对以下 k 阶展开进行一组评估来学习： $k = 0$、……、 N，这些评估对应于 n_c 个设计输入 $\boldsymbol{x}_1^c$、……、 $\boldsymbol{x}_{n_c}^c$。定义在第 i 个设计点处对这些展开进行评估所得的集合为 $H(\boldsymbol{x}_i^c) = \{h^{(0)}(\boldsymbol{x}_i^c)……,\ h^{(N)}(\boldsymbol{x}_i^c)\}$。给定 $Q(\boldsymbol{x})$ 和 $y_{\mathrm{ref}}(\boldsymbol{x})$ 后，这些评估结果用于提取各设计点处观测到的有限阶系数 $C(\boldsymbol{x}_i^c) = \{c_0(\boldsymbol{x}_i^c)……,\ c_N(\boldsymbol{x}_i^c)\}$。利用 $C(\boldsymbol{x}_1^c)……,\ C(\boldsymbol{x}_{n_c}^c)$ 和式（13）构造似然。根据这些观测到的系数所得到的后验分布，对未知参数 $\boldsymbol{\theta}$ 进行估计。

截断误差涵盖了级数中其余未知项，因此 $\delta^{(N)}(\boldsymbol{x})$ 采用类似的因式分解方式建模

$$\delta^{(N)}(\boldsymbol{x}) = y_{\mathrm{ref}}(\boldsymbol{x}) \sum_{k=N+1}^{\infty} c_k(\boldsymbol{x}) Q^k(\boldsymbol{x}). \tag{14}$$

结合式（13）和式（14）以及多元正态分布的性质（Ravishanker et al., 2021），截断误差项所诱导的先验分布为

$$\delta^{(N)}(\boldsymbol{x}) \mid \boldsymbol{\theta}, Q \sim GP\big(m_\delta(\boldsymbol{x}), \bar{c}^2 R_\delta(\boldsymbol{x}, \boldsymbol{x}'; \ell)\big), \tag{15}$$

具有均值和协方差函数

$$m_\delta(\boldsymbol{x}) = \mu\ y_{\mathrm{ref}}(\boldsymbol{x})\ \frac{Q^{N+1}(\boldsymbol{x})}{1 - Q(\boldsymbol{x})} \tag{16}$$

$$R_\delta(\boldsymbol{x}, \boldsymbol{x}'; \ell) = y_{\mathrm{ref}}(\boldsymbol{x}) y_{\mathrm{ref}}(\boldsymbol{x}')\ \frac{[Q(\boldsymbol{x})Q(\boldsymbol{x}')]^{N+1}}{1 - Q(\boldsymbol{x})Q(\boldsymbol{x}')}. \tag{17}$$

(15)–(17) 中的未知参数源自 (12) 中的系数模型。因此，用于表征差异模型的均值和协方差函数，也通过有限阶展开在 n_c 设计点处的一组计算结果进行学习。这是有效场论的一项独特性质，因为学习模型差异并不需要观测数据。

当有限阶展开的计算成本较低时，理论预测 $f^{(N)}(\boldsymbol{x}) = h^{(N)}(\boldsymbol{x}) + \delta^{(N)}(\boldsymbol{x})$ 上的诱导先验为

$$f^{(N)}(\boldsymbol{x}) \mid \boldsymbol{\theta}, Q, \boldsymbol{h}^{(N)} \sim GP\big(m_{\mathrm{th}}(\boldsymbol{x}), \Sigma_{\mathrm{th}}(\boldsymbol{x}, \boldsymbol{x}')\big),$$

其中 $m_{\mathrm{th}}(\boldsymbol{x}) = h^{(N)}(\boldsymbol{x}) + m_\delta(\boldsymbol{x})$ 和 $\Sigma_{\mathrm{th}}(\boldsymbol{x}, \boldsymbol{x}') = \bar{c}^2 R_\delta(\boldsymbol{x}, \boldsymbol{x}'; \ell)$。在计算成本较高的情况下，可以使用高斯过程来模拟有限阶展开，其定义为

$$h^{(N)}(\boldsymbol{x}) \mid \boldsymbol{\theta}, Q \sim GP\big(m_N(\boldsymbol{x}), \bar{c}^2 R_N(\boldsymbol{x}, \boldsymbol{x}'; \ell)\big).$$

由此得到的理论预测先验分布是一个高斯过程，其均值函数和协方差函数为$m_{\mathrm{th}}(\boldsymbol{x}) = m_N(\boldsymbol{x}) + m_\delta(\boldsymbol{x})$ 和 $\Sigma_{\mathrm{th}}(\boldsymbol{x}, \boldsymbol{x}') = \bar{c}^2 R_N(\boldsymbol{x}, \boldsymbol{x}';\ \ell) + \bar{c}^2 R_\delta(\boldsymbol{x}, \boldsymbol{x}';\ \ell)$。无论哪种情况，给定一组模型运行结果 $H(\boldsymbol{x}_1^c) \ldots, H(\boldsymbol{x}_{n_c}^c)$，都可以获得新输入 $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m$ 处的后验预测 $\hat{f}^{(N)}(\tilde{\boldsymbol{x}}_1) \ldots, \hat{f}^{(N)}(\tilde{\boldsymbol{x}}_m)$。