arXiv:2603.09428v1 [stat.AP] 10 Mar 2026

使用分层分解先验的贝叶斯物种分布模型

Luisa Ferrari
摩德纳与雷焦艾米利亚大学经济系

Massimo Ventrucci
博洛尼亚大学统计科学系

Alex Laini
都灵大学生命科学与系统生物学系

摘要

理解环境、空间和时间过程在塑造物种分布中的相对贡献，是生态学的核心目标。贝叶斯物种分布模型（SDM）为此提供了灵活的框架，但方差分量的先验设定仍颇具挑战。

为解决这一问题，我们将层级分解（HD）先验框架应用于潜在高斯物种分布模型，从而能够直接、透明地控制方差分解。HD方法将方差重新参数化为总方差和一组易于解释的比例，并通过一棵分解树来组织这些比例；该树既反映模型架构，也体现具有生态学意义的效应分组。

我们探讨了一种适用于物种分布模型的默认树设计原则，并介绍了逐步实施该方法的实用工作流程。我们使用NOAA东北渔业科学中心秋季底层拖网调查中39种底层鱼类的存在-缺失数据来展示该框架。结果表明，该方法的预测性能与既有先验相当，同时显著提升了方差归因和先验敏感性分析的可解释性与透明度。

关键词：直观先验设定；潜在高斯模型；P样条；方差分解

1 引言

物种分布模型（SDM）广泛应用于生态学中，用于模拟物种出现情况。这类模型假设，在感兴趣区域和给定时间段内，物种出现的分布是 3 种不同因素共同作用的结果（Warton 等人，2015；Ovaskainen 等人，2017）：非生物因素，即环境特征；生物因素，即种内与种间相互作用（如捕食、竞争和互利共生）；以及其他随机过程（如生态漂变或环境随机性）。尽管目前尚未建立能够实现这一目标的完全稳健的统计框架，厘清这些因素的相对贡献仍然是自然保护的一项关键目标。

形式上，物种分布模型将响应变量 Y 作为考量对象，用以表征物种出现情况，并以有无记录、个体计数或生物量百分比估计值表示。观测的地点和时间（$\boldsymbol{Z} = [Z_1, Z_2]$，$T$）通常是已知的，同时还会获得一组环境协变量（$\boldsymbol{X} = [X_1, \ldots, X_P]$），因为假定它们与物种出现情况相关。

假设 Y 服从指数族分布

其线性预测因子为 η；一般的物种分布模型可写为：

$$\eta = \mu + \sum_{j=1}^{J} f_j(\boldsymbol{X}) + \sum_{l=1}^{L} f_l(\boldsymbol{Z}, T). \tag{1}$$

该模型类别能够为每个协变量 X_1，……，X_P（例如 $f_1(X_1) + \cdots + f_P(X_P) + f_{\boldsymbol{Z}}(\boldsymbol{Z}) + f_T(T)$）设置主效应，还能表示不同环境变量之间的交互作用、时空交互作用、额外的非结构化噪声等。表示非生物因素和生物因素效应的 $f_j(\cdot)$ 和 $f_l(\cdot)$ 函数，可通过采用贝叶斯潜在高斯模型（潜在高斯模型）框架（Rue 等人，2009年；Hrafnkelsson，2023年）来方便地近似：

$$f_j(\boldsymbol{X}) = \boldsymbol{D}_j^{\mathsf{T}}(\boldsymbol{X})\boldsymbol{u}_j, \qquad \boldsymbol{u}_j \mid \sigma^2_{A,j} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{A,j}\boldsymbol{Q}^-_{A,j}\right), \quad j = 1, \ldots, J \tag{2a}$$

$$f_l(\boldsymbol{Z}, T) = \boldsymbol{G}_l^{\mathsf{T}}(\boldsymbol{Z}, T)\boldsymbol{v}_l, \qquad \boldsymbol{v}_l \mid \sigma^2_{B,l} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{B,l}\boldsymbol{Q}^-_{B,l}\right), \quad l = 1, \ldots, L \tag{2b}$$

方程 (2a)-(2b) 指定了两组结构化加性效应。每个非生物组分 $f_j(\boldsymbol{X})$、$j = 1$、……、J 都通过基 $\boldsymbol{D}_j(\boldsymbol{X})$ 和系数 $\boldsymbol{u}_j$ 表示；这些系数服从方差为 $\sigma^2_{A,\,j}$ 且精度矩阵为 $\boldsymbol{Q}_{A,\,j}$ 的零均值高斯先验。类似地，每个生物组分由基 $\boldsymbol{G}_l(\boldsymbol{Z}, T)$ 和系数 $\boldsymbol{v}_l$ 表示，其方差为 $\sigma^2_{B,\,l}$，精度矩阵为 $\boldsymbol{Q}_{B,\,l}$。方差参数控制非生物效应 (A) 和生物效应 (B) 的贡献。在贝叶斯框架中，模型设定还需为尺度参数 $\sigma^2_{A,\,j}$、$\sigma^2_{B,\,l}$ 指定先验：传统上，$\sigma^2_{A,\,j}$ 参数被固定为较大的值，以便系数自由变化；而 $\sigma^2_{B,\,l}$ 则被赋予相互独立的弥散先验，以反映先验信息的缺乏。

最近，Fuglstad 等人(2020年) 提出了一种创新方法，称为层级分解（HD）先验，用于进行尺度参数的先验设定。这种新方法可以为方差设定联合先验，并且能够轻松地融入现有的

关于模型效应相对重要性的先验知识，同时还能对模型的灵活性进行正则化。Hem 等（2021）和 Marques 等（2023）的研究发现，HD先验与其他最先进的替代方法相比具有竞争力。HD先验在具有结构化数据和/或可利用专家知识的领域尤具价值。在设计相对标准的情形下（例如多层模型或时空模型），可以构建 HD先验以体现模型结构，并以简约原则为指导，优先选择更简单的模型，除非数据为增加复杂度提供了有力证据。或者，HD先验的设计也可以遵循弱先验信息原则，旨在尽量减少模型结构本身所隐含假设之外的额外假设。HD框架的应用示例包括疾病制图（Franco-Villoria 等，2022；Riebler 等，2016）、人口学（Fuglstad 等（2020）中的新生儿死亡率）和林业（Marques 等，2023）。HD先验也非常适用于专家知识通常以方差比例形式表达的情境，包括疾病制图（Wakefield，2007）、基因组学（Holand 等，2013）和生态学（Peres-Neto 等，2006）；不过，目前直接 elicitation 仅应用于植物育种数据（Hem 等，2021）。

HD先验为解决贝叶斯物种分布模型中的先验征询问题提供了一种切实可行的方案。首先，由于这类模型通常将非生物效应和生物效应分别建模，因此可以直接构建分解树（Ovaskainen等，2017）。其次，生态学家通常掌握有关方差分解的专家知识（Pettit，1990），这些知识可以直接编码到模型的先验设定中。第三，限制模型的灵活性是一个重要方面，而采用 HD 先验这类有原则的先验，能更轻松地实现这一点。

本文探讨了在尚未探索的物种分布模型背景下实施 HD先验。具体而言，我们概述了在物种分布模型中实施 HD 方法所需的工作流程，

该流程包括树设计和先验设定。此外，我们强调了初步标准化过程的重要性，它对于正确解释基于 HD 的重新参数化是必要的。

本文通过一个真实案例研究来说明研究方案，该案例使用了Hui 等人(2023年提供的数据集，报告了北大西洋 39 种鱼类物种的出现/缺失数据，数据来自 NOAA-NEFSC 调查（NOAA东北渔业科学中心，2024 年）。该应用展示了相较于更传统方法，采用 HD先验的优势。此外，该数据集所需模型的复杂性，为突出并讨论实践中指定 HD 先验时可能出现的一些关键挑战提供了机会。

本文其余部分安排如下：第 2 节介绍引导性示例，呈现数据集和所选模型；第 3 节简要回顾 HD 方法；第 4 节介绍将 HD先验应用于物种分布模型的方法；第 5 节报告案例研究的结果；第 6 节通过讨论作结。

2 引导性示例

我们使用 NOAA-NEFSC 秋季底拖网调查数据库的一部分数据集（NOAA东北渔业科学中心，2024 年），该数据集经Hui 等人(2023年)处理和研究，并在 https://github.com/fhui28/CBFM 上公开发布（最后访问时间：2026 年 1 月）。核心研究问题是量化不同模型组成部分的方差贡献，即环境因素以及残差空间和时间贡献。

该数据集包含 $N = 5892$ 项观测，这些观测是在2000年至2019年间每年秋季于美国东北大陆架采集的；每个地点仅调查过一次。以下物种的出现情况： $N_{\text{species}} =$

39种不同底栖鱼类以有无响应的形式进行记录 Y。数据的可视化概览见补充材料。

为每次观测收集了以下环境协变量：表层温度 (X_1)、底层温度 (X_2)、表层盐度(X_3)、底层盐度 (X_4)、深度 (X_5)。此外，还报告了一个二元协变量，用于表示调查船 (X_6)的类型，并将其作为额外的环境条件处理，因此协变量总数为 $P = 6$ 。空间位置采用UTM坐标系报告，其中 $\boldsymbol{Z} = [Z_1, Z_2]$ 表示纬度和经度，而 T 协变量用于表示年份，因为数据仅在秋季收集。

在建模方面，将每个物种的出现情况视为边际结果，并分别使用单独但相同的物种分布模型进行拟合。这里使用的物种分布模型借鉴了Hui 等人 (2023年)提出的联合模型。该模型基于探索性分析，假设空间效应、时间效应与连续环境协变量的效应彼此可分离。为简洁起见，省略下标，每个响应 Y 假设服从参数为 $g^{-1}(\eta)$ 的伯努利分布，其中 $g(\cdot)$ 为logit链接函数，线性预测因子 η 定义如下：

$$\eta = \mu + \sum_{p=1}^{5} f_p(X_p) + f_6(X_6) + f_S(\boldsymbol{Z}) + f_T(T) \tag{3}$$

该模型随后被指定为潜在高斯模型。具体而言，对于五个连续协变量 $p = 1$, ……, 5， $f_p(X_p)$采用P样条（Lang和Brezger（2004），Eilers和Marx（1996））进行建模，以纳入可能的平滑非线性效应：特别地，基D$_p(X_p)$由20个B样条组成，而精度矩阵Q$_A$，$_p$ 被指定为二阶内在高斯马尔可夫随机场（IGMRF，Rue and Held（2005））。$f_6(X_6)$是虚拟变量协变量上的两水平独立同分布效应； $f_T(T)$被设定为一阶随机游走（即一阶I GMRF）。 $f_S(\boldsymbol{Z})$使用一个

ICAR 模型（即一阶IGMRF 先验），作用于空间域底层的离散网格上，其系数通过二维B样条基变换到连续空间，以诱导平滑性（更多细节见补充材料）。因此，该模型隐式包含以下方差参数： $\sigma_1^2, \ldots, \sigma_6^2, \sigma_S^2, \sigma_T^2$。

3 分层分解先验框架

HD 先验框架（Fuglstad 等人，2020年）通过两步程序，以直观的方式纳入关于模型效应相对重要性的先验知识。为便于说明，考虑方程（1）–（2）中给出的潜在高斯模型。HD 方法的第一步是将 $\sigma_{A,j}^2$、$\sigma_{B,l}^2$ 重新参数化为一组更直观的参数。具体而言，新参数包括一个总方差：

$$V = \sum_j \sigma_{A,j}^2 + \sum_l \sigma_{B,l}^2 \tag{4}$$

以及一组以 ω 字母表示的方差比例。这些比例通过树设计构造：递归地将总方差 V划分为若干组方差参数，直到每个分支都只包含一个尺度参数。在每次划分中，将各分支中尺度参数之和除以其父节点中的尺度参数之和，得到相应比例。二元划分得到标量比例，而多分支划分则产生单纯形。例如，在物种分布模型中，用户通常能直观判断非生物效应所解释的方差份额。这可以通过在非生物效应与生物效应之间进行初始二元划分来表示，从而定义比例：

$$\omega_A = \frac{\sum_j \sigma_{A,j}^2}{\sum_j \sigma_{A,j}^2 + \sum_l \sigma_{B,l}^2}. \tag{5}$$

该流程的第二步是为新参数 V和 ωs 指定先验分布及其超参数，使之能够反映先验信念：由此得到原始方差参数的联合先验分布，从而有效整合专家知识，并符合贝叶斯原理。HD 先验可以通过 `makemyprior` 由 Hem 等人 (2024 年) 开发的 R 包，以便于用户使用的方式进行设计。

关于比例参数，Fuglstad 等人 (2020年) 考虑了两种备选的先验规范。如果对于某次分割的各个分支没有先验偏好，则使用狄利克雷先验来表达对总方差如何分配的无知。相反，尤其对于二元分裂，用户可以使用惩罚复杂度（PC）先验，表达对某个分支的先验偏好。Simpson 等人 Simpson 等人 (2017) 提出的 PC 框架提供了一种有原则的方法，用于定义向基准模型（即偏好的参数值）收缩的先验。更准确地说，PC先验通过为模型与其基准版本之间基于 KLD 的距离指定一个率参数为 λ 的指数分布来定义。根据 Hem 等人 (2024 年)， $PC_b(\cdot)$ 表示一种 PC先验，其中 b 是基准模型下的参数值；对于方差参数，通常使用 PC_0 ，以促使其向基准模型收缩，即方差为 0。率参数 λ 控制收缩程度，并通过尾概率陈述进行校准。PC先验应用广泛，尤其适用于高斯方差参数，例如方程（4）中的 PC_0，此时 $\sqrt{V}$可化为对 $\sqrt{V}$的指数先验。

—

相比之下，推导 PC_0 比例参数的先验需要更多工作，因为其形式取决于涉及的具体效应，也可能取决于低层比例参数，这在计算上效率低下（Fuglstad 等人，2020年）。然而，在许多情况下，潜在高斯模型的方差比例 PC_0 先验 ω被发现具有以下Fuglstad 等人，2020年；Franco-Villoria 等人，2022年）简化形式：

$$\pi(\omega) = \frac{\lambda \exp(-\lambda\sqrt{\omega})}{2\sqrt{\omega}[1 - \exp(-\lambda)]} \qquad 0 < \omega < 1 \tag{6}$$

这对应于 $\sqrt{\omega}$上的指数分布，截断于 $\sqrt{\omega}=1$。这个简化解不依赖于效应的基或精度矩阵的选择，也不依赖于分解树较低层级分割处的参数。我们在补充材料中证明，方程（6）是物种分布模型中大多数实际相关情形下正确的 PC_0 先验。我们给出的证明基于对所涉及矩阵秩的一个条件，该条件是否成立取决于所考虑的效应类型。

HD 方法的主要优势在于，它将难以解释的方差尺度参数化转换为比例尺度参数化。通过先验设定，这一转换能够直接纳入有关不同效应相对贡献的先验信息，使模型假设更加透明，并让生态学家及其他非专业用户能够直接依据直观的效应重要性进行推理，而不必理解抽象的方差分量。然而，该方法依赖于比例参数准确反映各效应在分支内相对于父节点总方差所占的方差份额。只有当原始效应经过正确缩放，使其对应的方差参数与实际方差贡献相匹配时，这一条件才成立；传统的效应设定并不总能满足这一要求。

为了解决这个问题，Ferrari 和 Ventrucci(2025) 提出了一种标准化流程，适用于潜在高斯模型中的所有效应。标准化可确保每个效应的方差参数（例如 σ_j^2）准确反映相关效应（例如 $f_j(\cdot)$ 对总方差的贡献。因此，这一步对于解读 HD 先验至关重要，因为它能确保从分解树导出的比例参数有意义地表示各效应的相对重要性。

4 物种分布模型的 HD先验

4.1 物种分布模型效应的标准化

要将 HD先验正确应用于物种分布模型，必须确保所有效应（无论是非生物效应还是生物效应）都已正确标准化，使参数 $\sigma^2_{A,\ j}$ 和 $\sigma^2_{B,\ j}$ 分别对应其效应实际的方差贡献。根据Ferrari 和 Ventrucci (2025)，这一要求可形式化定义如下：

$$\sigma^2_{A,j} = Var_{\boldsymbol{X},\boldsymbol{u}_j}\left[f_j(\boldsymbol{X}) \mid \sigma^2_{A,j}\right], \qquad j = 1, \ldots, J, \tag{7a}$$

$$\sigma^2_{B,l} = Var_{\boldsymbol{Z},T,\boldsymbol{v}_l}\left[f_l(\boldsymbol{Z}, T) \mid \sigma^2_{B,l}\right], \qquad l = 1, \ldots, L. \tag{7b}$$

如（7）所示，在此框架中，效应的方差贡献定义为该效应相对于所涉及协变量的变异性。因此， $\boldsymbol{X}$、 $\boldsymbol{Z}$和 T （高斯系数$\mathbf{u}_j$、$\mathbf{v}_l$除外）必须视为具有已知概率分布的随机变量，才能对（7）中的量进行形式化定义。重要的是，对协变量所作的分布假设不一定要旨在反映其真实的数据生成过程；相反，这些假设应当用于将用户对效应方差贡献的理解转化为精确的数学定义。请注意，只需指定与给定模型效应相关的协变量的联合分布即可。未进入同一效应的协变量之间的依赖结构不会影响标准化程序，因此可以忽略；为方便起见，可以假定未进入同一效应的协变量组彼此独立。我们建议将均匀性作为对协变量和时空域进行分布假设时的默认选择。这一建议兼顾了数学上的便利性和可解释性。具体而言，均匀性为用户带来了解释上的优势：选择均匀分布时，方差

相对于协变量而言，只是支撑集内所有地点测得的趋势（即效应的实现值）的方差；与必须考虑每个点上不同的概率密度相比，这对用户而言更为直观。

能够保证（7）的标准化程序通过两步程序修改效应的基和精度矩阵，其中两步都明确依赖于对协变量分布的假设。第一步是对效应施加适当的约束以确保其均值为零（若尚未满足这一条件）：

$$\mathbb{E}_{\boldsymbol{X}}[f_j(\boldsymbol{X})] = 0, \qquad j = 1, \ldots, J, \tag{8a}$$

$$\mathbb{E}_{\boldsymbol{Z},T}[f_l(\boldsymbol{Z}, T)] = 0, \qquad l = 1, \ldots, L. \tag{8b}$$

可以通过适当修改效应的精度矩阵或基函数来满足这一要求。第二步是将每个效应乘以适当的常数，该常数称为参考标准差（Sørbye 和 Rue, 2014）：

$$\widetilde{f}_j(\boldsymbol{X}) = \frac{f_j(\boldsymbol{X})}{\sqrt{\mathrm{Var}_{\boldsymbol{X},\boldsymbol{u}_j}\left[f_j(\boldsymbol{X}) \mid \sigma^2_{A,j} = 1\right]}}, \qquad j = 1, \ldots, J, \tag{9a}$$

$$\widetilde{f}_l(\boldsymbol{Z}, T) = \frac{f_l(\boldsymbol{Z}, T)}{\sqrt{\mathrm{Var}_{\boldsymbol{Z},T,\boldsymbol{v}_l}\left[f_l(\boldsymbol{Z}, T) \mid \sigma^2_{B,l} = 1\right]}}, \qquad l = 1, \ldots, L. \tag{9b}$$

请注意，方程（9）分母中的方差分别是在以$\sigma^2_{A,j} = 1$和$\sigma^2_{B,l} = 1$为条件时计算的。这样缩放可确保$\widetilde{}$所有效应$f_j(\boldsymbol{X})$和$f_l(\boldsymbol{Z}, T)$都以统一的尺度表示。对于许多效应（包括常用的积分高斯马尔可夫随机场），以参考标准差定义的缩放常数可以使用 GitHub 上提供的`scaleGMRF`软件包轻松计算：`https://github.com/LFerrariIt/scaleGMRF`（Ferrari 和 Ventrucci, 2025）。

4.1.1 将标准化应用于 NOAA-NEFSC 案例研究

考虑方程（3）中的模型。在这种情况下，标准化只需选择边缘分布$\prod_{p=1}^{P}\pi$（x_p）π（$\boldsymbol{z}$）π（t），因为模型中不存在交互项。首先考虑环境协变量，我们假设每个协变量均服从均匀分布。为确定连续协变量的均匀分布范围，我们采用经验范围，但先排除被认为具有高杠杆值的点（共 3 个），因为这些点可能会严重扭曲结果。所选范围与现有知识相符，能够反映美国东北大陆架所研究环境因素的变化情况（NOAA东北渔业科学中心，2024 年）。分类变量 X_6 和 T也均假设服从均匀分布。对于空间坐标 $\boldsymbol{Z}$，我们同样假设其在一个多边形上服从均匀分布，该多边形涵盖样本中的所有地点，并使用凹包算法构建（更多详情和图示请参见补充材料）。

一旦指定了分布假设，就可以对所有效应执行标准化过程。然而，处理P样条会带来额外的挑战，因为二阶IGMRF（传统P样条设定中采用）的方差参数只衡量效应偏离其零空间的程度，而其零空间即为线性趋势。因此，在这种情况下，$\sigma_1^2,\ldots,\sigma_5^2$参数只衡量效应偏离其线性趋势的程度，而不是量化其对应效应对方差的总体贡献。为此，Ferrari 和 Ventrucci(2025)提出将P样条效应重新设定为两个独立的组成部分：一个线性效应，以及一个受到约束、不会干扰线性效应的P样条效应。因此，每个 $f_p(X_p)$都重新定义为：

$$f_p(X_p) = f_{Lp}(X_p) + f_{Np}(X_p) \tag{10}$$

其中 $f_{Lp}(X_p)$是一个线性效应，其对应的方差参数为 σ^2_{Lp}；而 $f_{Np}(X_p)$是一个P样条效应（通过适当的约束去除线性趋势），其对应参数为σ^2_{Np}。有关这一替代设定的更多细节，请参见Ferrari 和 Ventrucci(2025)。

最后，已对所有 $f_{Lp}(X_p)$ 和$f_{Np}(X_p)$ 分别进行标准化，其中 $p = 1$、. . . 、 5 ； 此外，还对 $f_6(X_6)$ $f_{\boldsymbol{Z}}(\boldsymbol{Z})$ $f_T(T)$ 进行了标准化。

4.2 树设计

对物种分布模型中的所有效应进行适当标准化后，我们可以考虑应用HD框架，首先进行树设计。根据 Fuglstad 等人(2020年) 的观点，“必须选择合适的树结构，以便进行所需的比较”：这意味着生成的比例参数应衡量用户能够直接直观理解的各组效应之间的相对重要性。因此，树设计最好针对具体应用，并由专家主导。然而，先验信息往往与当前案例研究并不直接相关，而是来自一般领域知识与建模原则（例如简洁性）的结合。因此，我们为公式（2）所示的通用物种分布模型提出一棵默认分解树，如图1所示。该树旨在提供生态学家普遍能够直观理解的参数，无论具体应用场景如何。此默认树设计基于物种分布模型背后的理论、建模原则以及 HD文献中的相关考量。

图中树的第一次分裂1 将总方差划分为 SDM理论所认定的两个主要变异来源：非生物/环境效应的贡献；以及额外时空变异的贡献，后者可能代表生物相互作用和随机过程。在第二层，前一步产生的两个子节点各自进一步分裂。在这两种情况下，主效应（单个协变量的函数）都与交互项（两个

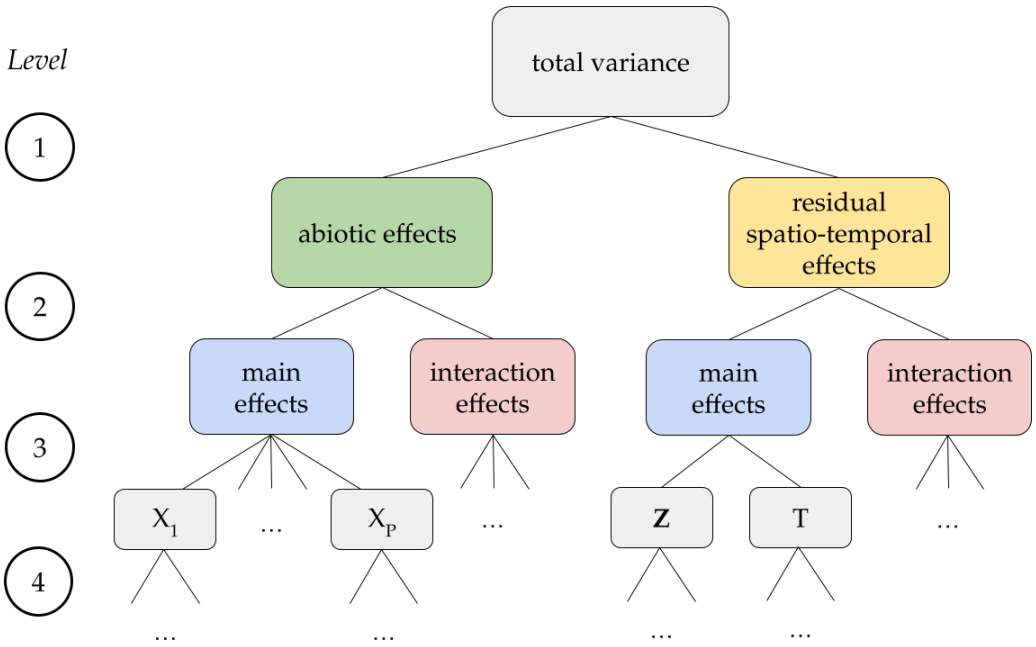


图1：物种分布模型的默认分解树。每个节点代表相应效应的所有方差参数之和 σ^2 。

协变量）；请注意， $\boldsymbol{Z}$ 被视为单个协变量。如果模型中不包含交互作用，则该层可直接从树中剪除。第三层涉及潜在的多分支划分，即根据各效应作为其函数的协变量来划分效应。第一次分割用于将不同环境协变量的效应划分到多个分支，而第二次分割则将空间效应与时间效应分开。来自第2层任一分支的交互项，可能需要在第3层进行额外的分割。

所提出的树能够很好地适应大多数物种分布模型应用。此后，只要用户掌握更为精细的信息，便可随意调整该树。如果图1 中的所有灰色节点都只包含一个方差参数，那么该树即已完整。然而，某些协变量可能通过多个效应进入模型。在这种情况下，图1 中的方案应通过额外的分割加以扩展，直至所有子节点均为单元素集合。我们建议使用第四层分裂，通过连续的二叉分支按灵活性递增来分离效应，因为这样可以简化先验设定。例如，一次第4层分割可以区分某一环境协变量效应的线性成分与非线性成分。

4.2.1 NOAA-NEFSC 案例研究的树设计

案例研究中的模型共有 13 个需要进行先验设定的方差参数： σ^2_{L1}， σ^2_{N1}，……， σ^2_{L5}， σ^2_{N5}， σ^2_{6}， σ^2_{S}， σ^2_{T}。该应用的树设计如图2所示。默认树的第2层经过剪枝；在第4层，则针对全部 5 个连续环境协变量，在其线性贡献与非线性贡献之间添加二元分裂。按父节点划分子节点后，我们得到以下新参数：

$$V = \left(\sum_{p=1}^{5} \sigma_{Lp}^2 + \sigma_{Np}^2\right) + \sigma_6^2 + \sigma_S^2 + \sigma_T^2 \tag{11a}$$

$$\omega_A = \frac{\left(\sum_{p=1}^{5} \sigma_{Lp}^2 + \sigma_{Np}^2\right) + \sigma_6^2}{V} \tag{11b}$$

$$\boldsymbol{\omega}_X = \left[\frac{\sigma_{L1}^2 + \sigma_{N1}^2}{\omega_A \cdot V}, \ldots, \frac{\sigma_{L5}^2 + \sigma_{N5}^2}{\omega_A \cdot V}, \frac{\sigma_6^2}{\omega_A \cdot V}\right] \tag{11c}$$

$$\omega_S = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_T^2} \tag{11d}$$

$$\omega_{Np} = \frac{\sigma_{Np}^2}{\sigma_{Lp}^2 + \sigma_{Np}^2}, \qquad p = 1, \ldots, 5. \tag{11e}$$

4.3 先验设定

根据图1，我们讨论适用于新的 HD 参 V 数和 ωs 的先验选择。尽管原始 HD 框架以自底向上方式构建联合先验，但大多数实际实现都假设不同分割之间的比例相互独立。因此，我们采用分割特定参数之间的先验独立性；这一选择在简化先验设定与计算的同时，也符合常见做法。

总方差 对于 V，当似然为非高斯时，HD文献建议采用 PC先验；对于高斯似然，Fuglstad 等人(2020年) 则主张采用尺度不变的杰弗里斯先验。在物种分布模型的背景下，这两种选择都颇具吸引力，因为 PC_0 先验能够为高度灵活的模型提供正则化，而杰弗里斯先验则无需进行超参数调优。

第1层 第1层会生成一个比例参数 ω_A，来自公式（5）。可以很容易地用均匀分布来表达无知

，也可以使用尺度不变的杰弗里斯先验。如果先验信息表明情况并非如此，则可以通过适当的贝塔分布，或采用具有所需基准模型和反映用户不确定性的浓度参数的PC先验来引入这些信息 Hem 等人 （2024 年）。

第3层 第3层的分割根据父节点所代表的协变量，将其效应划分为多个分支。Fuglstad 等（2020）建议使用对称狄利克雷分布 Dir($q, \ldots, q$)，其中超参数 q 控制稀疏度，可对其进行调整，以反映关于各分支间方差划分的假设。Hem 等（2024）建议使用单个比例 ω_p 的边际先验来设定 q ，使得 $P(\text{logit}(1/4) < \text{logit}(\omega_p) - \text{logit}(1/P) < \text{logit}(3/4)) = 0.5$。另外，Zhang 等（2022）的研究推导出，在选择 $q < 1/2$ 时，R2D2 先验诱导的系数先验具有理想的收缩行为。

第2层和第4层 第2层和第4层会产生类似的二元划分，根据效应的灵活程度将其分开（例如，主效应与交互效应，或线性效应与非线性效应）。遵循 Fuglstad 等人 (2020年的指导原则，一种自然的选择是将比例参数解释为较灵活分支的比例贡献，并为其指定 PC_0先验，这与模型简约性原则相符。在大多数情况下，所得的 PC_0 先验会简化为方程（6）给出的简化形式。即使在无法进行这种简化的情况下，出于实际考虑，我们仍建议采用相同的函数形式，因为它具有显著的计算优势，尽管此时它不再对应于精确的 PC 先验。

4.3.1 NOAA-NEFSC 案例研究的先验设定

参数 ω_A 表示非生物因素相对于生物因素所解释的方差比例。参数 ω_S 量化了空间效应相对于时间效应的相对贡献。由于缺乏关于这些参数的先验信息，我们为两个参数指定宽泛的均匀先验。这一选择不预先偏好任何允许的取值，而是将所有方差划分视为同等合理；这意味着预期方差会均等划分，同时也容许存在较大的不确定性。另一方面， $\boldsymbol{\omega}_X$ 用于衡量6个环境协变量各自的比例贡献，而我们能够对这些量提出先验假设。具体而言，我们认为深度 可能发挥重要作用；表层温度 与底层温度之间，以及表层盐度 与底层盐度 之间的相关性，可能导致每对协变量中只有一个被选中；最后，我们对调查船可能发挥的作用没有先验知识。我们通过设定 $\boldsymbol{\omega}_X \sim \text{Dir}(q)$，其中 $q = 0.5$，来量化这些先验信念；这会使协变量效应适度收缩。

第4层的5个分割均将5个环境协变量的线性贡献与非线性贡献分离开来。对于其中每一个 ω_{Np} 参数，我们选择一个非常平坦的PC_0 先验，以便先验上对非线性贡献几乎不施加收缩。根据补充材料中的证明，我们知道在这种情况下PC_0 先验具有其简化的函数形式，即公式（6）。此处通过采用非线性效应的降秩表示（具体使用20个B样条基函数）来保证这一简化。为设定超参数，我们考虑一个尾概率陈述： $P(\omega < U) = \alpha$。Franco-Villoria 等人 (2022年)指出必须遵守以下条件： $\alpha \geq \sqrt{U}$ 以得到 $\lambda > 0$。因此，中位数至多为0.25，这可通过 $\lambda \to 0$得到。我们选择 $\lambda = 0.1$，从而使中位数（$\alpha = 0.5$）为0.238，这表明先验上由协变量效应贡献的方差中仅约1/4来自

其非线性组成部分。先验设定通过假定 V服从尺度不变的杰弗里斯先验而得以完成。

5 结果

5.1 性能评估

首先将所提出的 HD 先验的性能与更传统的选择进行比较，具体而言：对所有σ^2 参数采用独立的逆伽马分布(1,5e-5)，这是 INLA 中的默认设定（IG）；对所有$_0(\lambda)$ 参数采用独立的 PC σ^2 先验，其超参数 λ满足$P(\sigma > 3) = 0.05$，如Fuglstad 等人(2020年)（PC）。

模型在 INLA（Rue 等人，2009年）中拟合，该软件提供所有模型参数的联合后验样本以及许多其他有用的输出。数据集被划分为 $\boldsymbol{y} = [\boldsymbol{y}_{\text{train}}^T, \boldsymbol{y}_{\text{test}}^T]^T$ 其中训练集$\boldsymbol{y}_{\text{train}}$ 包含截至2015年的所有观测值，而测试集 $\boldsymbol{y}_{\text{test}}$ 包含截至2020年的其余观测值，共计1028个观测值，约占数据集总量的17%。模型在训练集上拟合，并使用Hui 等人(2023年)（对数似然、布里尔分数、Tjur R^2）所用的相同标准评估测试集上的预测表现，同时采用更易于解释的准确率指标。所有指标均基于测试集中所有实例的点估计 $\hat{}p_i = \text{logistic}(E[\eta_i|\boldsymbol{y}_{\text{train}}, \boldsymbol{x}_i, \boldsymbol{z}_i, t_i])$ i 。图3 展示了数据集中全部39个鱼类物种在这4项指标上的分布。

首先，我们可以注意到，由于数据集中观测值数量庞大，先验设定对预测性能的影响并不显著。尽管如此，结果仍存在一些差异：HD 先验与广受欢迎的 PC 先验方案相比颇具竞争力，并且优于逆伽马分布，尤其是在准确度和布里尔分数方面优势明显。因此，这些结果表明，HD先验为此前已验证的先验规范提供了一种可信的替代方案，在

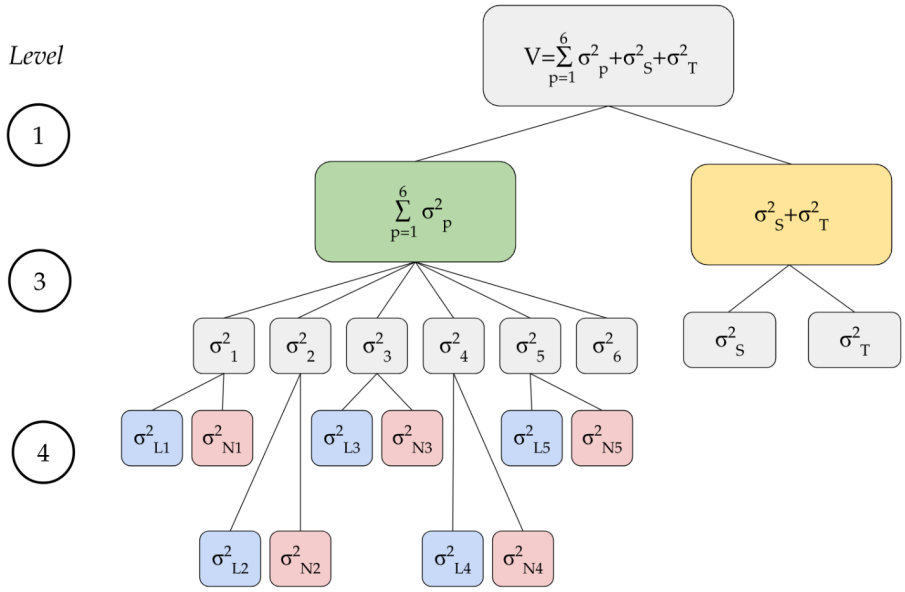


图2：基于第 4.2 节默认提议的案例研究分解树。

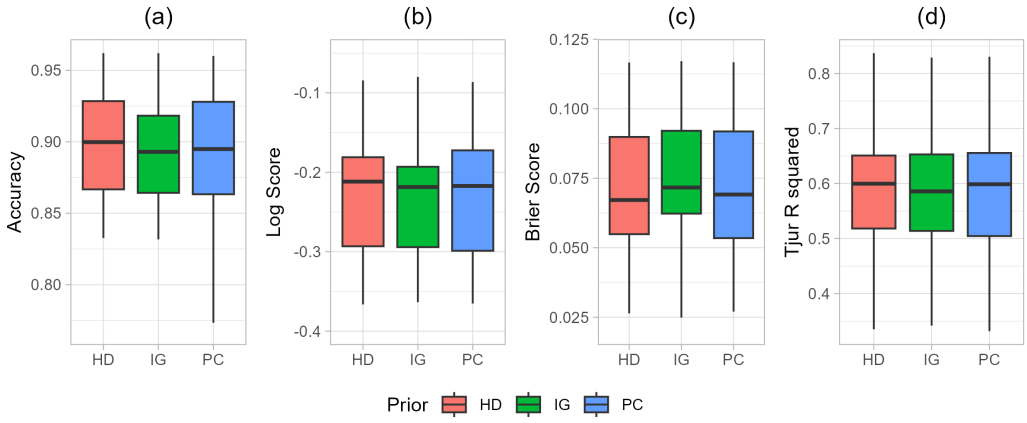


图3：比较4种不同先验设定在测试集上进行预测的性能。

文献，至少在本例的背景下是如此。

5.2 方差分解结果

使用 HD 先验，我们现在可以估计不同效应对所有物种出现的方差贡献；这也是生态学家的一个核心研究问题。我们针对每个物种使用完整数据集拟合模型，然后利用 R-INLA 软件包的 `inla.posterior.sample()` 函数获得模型参数的后验样本（Rue 等，2009）。我们使用相对于协变量上所选均匀分布计算的有限总体方差，来估计各效应的方差贡献： $s^2_{Lp} = Var_{X_p}[f_{Lp}(X_p)|\beta_p]$， $p = 1$，.. ., 5； $s^2_{Np} = Var_{X_p}[f_{Np}(X_p)|\boldsymbol{u}_p]$， $p = 1$，. . ., 5； $s^2_6 = Var_{X_6}[f_6(X_6)|\beta_6]$； $s^2_S = Var_{\boldsymbol{Z}}[f_S(\boldsymbol{Z})|\boldsymbol{v}_S]$； $s^2_T = Var_T[f_T(T)|\boldsymbol{v}_T]$。

令 $s^2_p = s^2_{Lp} + s^2_{Np}$， $p = 1$，.....，5，随后计算各模型项占总方差的比例：

$$\boldsymbol{\phi} = \frac{1}{\sum_{p=1}^{6} s^2_p + s^2_S + s^2_T} \left[s^2_1, \ldots, s^2_6, s^2_S, s^2_T\right]. \tag{12}$$

图 4 报告了 $\boldsymbol{\phi}$ 中所有 39 个物种对应条目的后验均值。

我们的分析表明，影响出现情况变异的主要因素为空间效应、深度效应以及底层温度效应。这与先前的研究结果一致（Hui 等人，2023年；Kleisner 等人，2016），因为所选物种均为底栖物种，而深度是影响栖息地适宜性的一个公认因素，通常可作为多种环境梯度的替代指标，例如压力、光照可用性和底质类型。底层温度也是决定底栖物种分布的关键因素，反映了生理限制和代谢偏好。

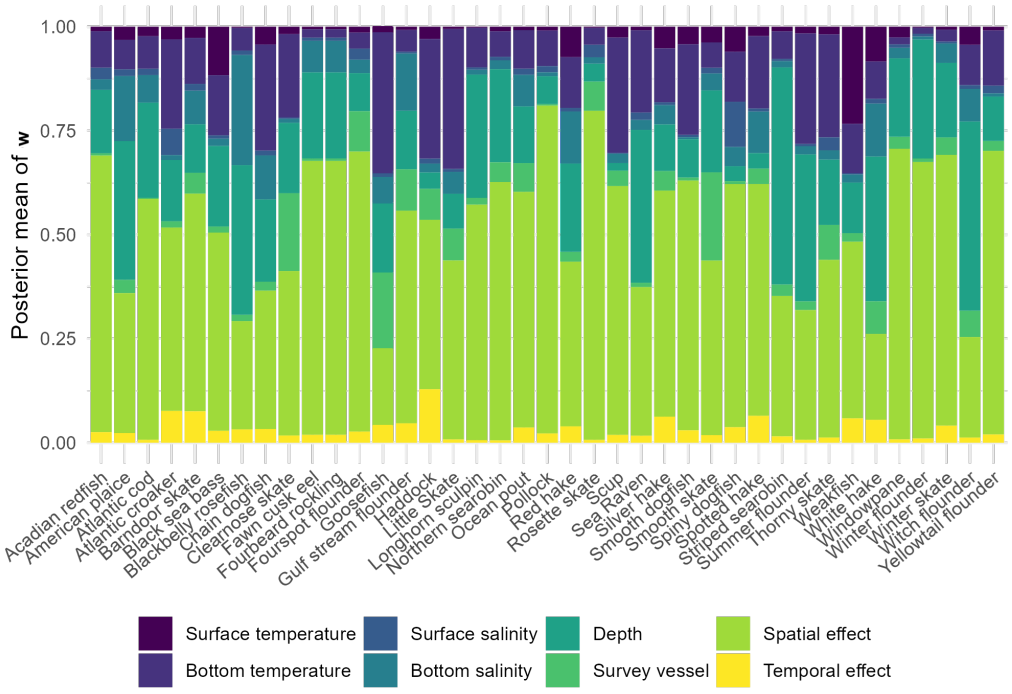


图4：39种不同物种的 ϕ 各条目的后验均值。

5.3 先验敏感性分析

最后，我们探讨与传统先验规范相比，HD 方法如何使先验敏感性分析过程更加透明、直观。在标准方法中，超参数选择对模型行为的影响往往是间接的，难以解释，因此用户很难将实质性的先验信念转化为具体的建模决策。

相比之下，HD框架允许从业者直接控制先验的特定方面，例如施加于方差分量的收缩程度。这种可解释性有助于开展先验敏感性分析，使人们能够更直接地评估模型对特定先验假设的稳健性或敏感性。这里，我们通过拟合同一模型来说明这一点：采用相同的 HD 先验结构，只改变先验的超参数 q，该先验针对 $\boldsymbol{\omega}_X$，它控制环境协变量比例方差贡献所受的收缩程度。图 5 展示了$q = [1, 0.5, 1/6]$ 物种安康鱼（最常见的物种之一）的方差分解估计。随着 q 减小，先验会越来越倾向于稀疏的方差分配，逐步将较小的分量收缩至零。图6中不同环境协变量的估计趋势也证实了这一点：其中表层盐度的趋势逐渐收缩至更接近 0。观察到的行为既可预测又易于解释，凸显了 HD 方法能够更直观地控制先验的影响，从而有助于以严谨的方式开展先验敏感性分析，并提高贝叶斯建模的透明度。

6 讨论

本文将 HD 先验框架引入物种分布模型领域，旨在提高其可解释性和透明度。

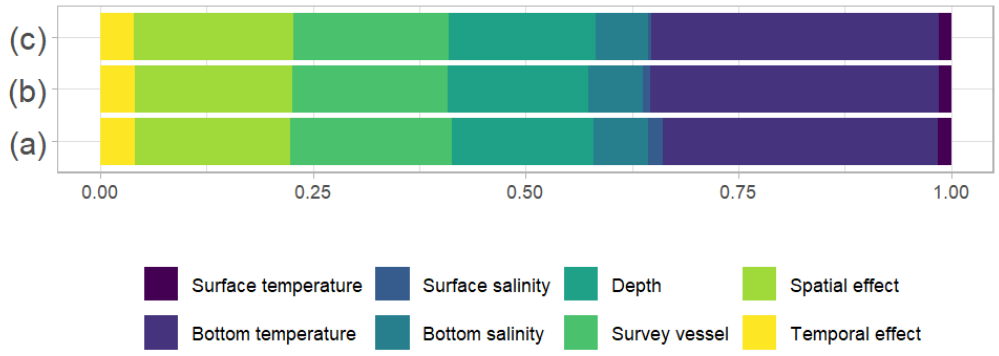
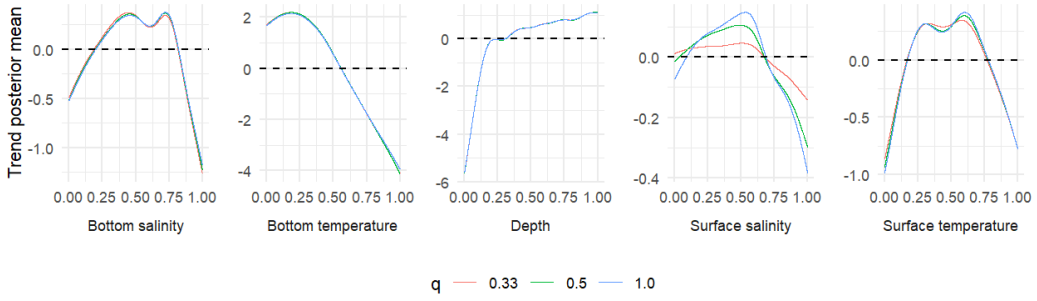


图5：安康鱼的 ϕ 各项后验均值： (a) $q = 1$, (b)$q = 0.5$, (c) $q = 1/6$。



5个环境协变量（采用归一化尺度）在 q不同取值下 $f_p(X_p)$ $p = 1,\ldots,5$的后验均值。

通过明确的方差划分视角进行先验设定。通过将方差分量重新参数化为总方差和若干组比例，该框架为表达关于环境协变量以及物种分布模型中空间效应和时间效应相对贡献的先验信念，提供了一种有原则的方法。我们首先提出了一种默认的HD树，将SDM理论、一般建模原则以及以往HD应用的见解结合起来进行设计。该默认结构具有足够的通用性，能够适用于广泛的物种分布模型情境，同时也足够灵活，可在专家知识可用时将其纳入其中。我们还针对由此得到的重新参数化提供了先验设定的实用指南。在这方面，方差比例的简化闭式惩罚复杂度先验大幅降低了实现复杂度。展望未来，这一默认策略可以扩展到具有类似模型结构的其他应用领域，例如疾病制图和环境质量评估。

我们将该方法应用于 NOAA-NEFSC数据集，展示了 HD先验在复杂建模场景中的用法。首先，我们说明了如何对每个模型组成部分进行标准化，从而对相关方差参数作出一致的解释。随后，我们重点介绍了默认树的灵活性，并展示了如何对其进行剪枝或扩展，以适应特定研究的特点。在先验设定阶段，我们提出对归因于 P样条效应中非线性组成部分的方差比例使用 PC_0 先验。这一选择值得进一步探究，尤其是要评估其能否作为一种通用机制，在保留足够灵活性的同时鼓励模型简约性。就预测性能而言，HD先验的表现与常用替代方案相当。重要的是，HD框架能够进行透明且有针对性的先验敏感性分析，而传统设定中缺乏这一点。总的来说，HD先验为物种分布模型文献增添了有价值的新方法，使方差在环境、空间和时间效应之间的分配方式具有更高的可控性和清晰度。

跨环境、空间和时间效应的先验。

这项研究另一个有趣之处，在于提出了一种量化方差贡献的替代方法。我们不再采用基于协变量经验分布的传统方法（Ovaskainen 和 Abrego，2020年），而是考虑使用用户指定的协变量分布假设来量化后验量。经验分布可以视为这一框架中的一种特殊情况，但它缺乏稳健性，也无法在不同案例研究之间进行有意义的比较。相反，采用与样本无关的选择（如均匀分布），能够提高方差贡献对用户的可解释性，并且适用于不同案例研究。进一步探索这一替代表述方式，或许是未来研究的一个有趣方向。

最后需要指出的是，我们在案例研究分析中采用的堆叠建模方法具有局限性，也未能反映当代群落生态学理论；该理论强调生物相互作用在物种出现中的作用。目前，该领域越来越多地采用联合物种分布模型（JSDMs；（Warton 等，2015），这类模型会明确考虑物种间的残差相关性。因此，未来研究的一个重要方向是考察 HD 先验在联合物种分布模型中的适用性。

数据可用性 支撑本研究发现的数据可在 GitHub 存储库 https://github.com/fhui28/CBFM 中公开获取。

资助信息 作者获得了欧盟通过欧盟下一代计划，在“PNRR—任务4‘教育与研究’—组成部分C2，投资1.1‘国家研究计划及国家重大项目基金（PRIN）’”计划下提供的资助。该项目由意大利大学与研究部（MUR）资助，项目名称为：“面向元群落的元条形码：迈向一种

群落生态学的遗传学方法（META2）”，项目代码：2022PA3BS2（CUP E53D23007580006），大学与研究部（MUR）第1015号资助法令，日期为2023年7月7日。

参考文献

Eilers, P. H. 和 Marx, B. D.（1996）。使用 B样条和惩罚项进行灵活平滑。《统计科学》，11(2):89–121。

Ferrari, L. 和 Ventrucci, M.（2025）。一种将基于方差分解的先验纳入潜在高斯模型的标准化程序。斯堪的纳维亚统计学杂志，第 1–31 页。

Franco-Villoria, M.、Ventrucci, M. 和 Rue, H.（2022）。时空疾病制图模型中的方差分解。《医学研究中的统计方法》，31(8):1566–1578。

Fuglstad, G.-A.、Hem, I. G.、Knight, A.、Rue, H. 和 Riebler, A.（2020）。方差参数的直观联合先验。《贝叶斯分析》，15(4):1109 – 1137。

Gombin, J.、Vaidyanathan, R. 和 Agafonkin, V.（2020年）。concaveman：一种速度极快的二维凹包算法。R 包。

Hem, I.、Fuglstad, G.-A. 和 Riebler, A.（2024 年）。makemyprior：直观地构建 R语言中方差参数的联合先验。《统计软件期刊》，110(3)：1–39。

Hem, I. G.、Selle, M.L.、Gorjanc, G.、Fuglstad, G.-A. 和 Riebler, A.（2021）。通过直观地纳入专家知识，对加性变异和非加性变异进行稳健建模。《遗传学》，217(3)。

Holand, A. M., Steinsland, I., Martino, S., 和 Jensen, H.（2013）。动物模型与积分嵌套拉普拉斯近似。《*G3*：基因、基因组、遗传学》，3(8):1241–1251。

Hrafnkelsson, B.（2023年）。使用贝叶斯潜在高斯模型进行统计建模：地球物理学与环境科学中的应用。Springer Cham。

Hui, F. K., Warton, D. I., Foster, S. D., 和 Haak, C. R.（2023年）。时空联合物种分布建模：一种基函数方法。生态学与进化方法，14(8):2150–2164。

Kleisner, K. M., Fogarty, M. J., McGee, S., Barnett, A., Fratantoni, P., Greene, J., Hare, J.A., Lucey, S. M., McGuire, C., Odell, J., 等（2016）。次区域气候速度对海洋物种群落分布和空间范围的影响。*PLOS ONE*，11(2):e0149220。

Knorr-Held, L.（2000）。疾病风险中不可分离时空变异的贝叶斯建模。*Statistics in Medicine*，19(17-18):2555–2567。

Lang, S. 和 Brezger, A.（2004）。贝叶斯P样条。计算与图形统计学杂志，13(1):183–212。

Marques, I., Wiemann, P. F. 和 Kneib, T.（2023）。采用固定样地设计的森林清查数据方差分解多层模型。农业、生物与环境统计学杂志，28(4):706–725。

NOAA东北渔业科学中心（2024 年）。NEFSC秋季底拖网调查。访问日期：2024年8月30日。

Ovaskainen, O. 和 Abrego, N.（2020年）。联合物种分布建模：*R*语言应用。剑桥大学出版社。

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., 和 Abrego, N.（2017）。如何从群落数据中获得更多信息？一个概念框架及其作为模型和软件的实现。《生态学快报》，20(5):561–576。

Peres-Neto, P. R., Legendre, P., Dray, S., 和 Borcard, D.（2006）。物种数据矩阵的变异分解：各部分的估计与比较。《生态学》，87(10):2614–2625。Pettit, L.（1990）。正态分布的条件预测序位。《皇家统计学会杂志：*B*辑（方法论）》，52(1):175–184。Riebler, A., Sørbye, S. H., Simpson, D., 和 Rue, H.（2016）。一种直观的贝叶斯空间模型，用于疾病制图并考虑尺度调整。《医学研究中的统计方法》，25(4):1145–1165。Rue, H. 和 Held, L.（2005）。《高斯马尔可夫随机场：理论与应用》。CRC出版社。Rue, H., Martino, S., 和 Chopin, N.（2009）。利用积分嵌套拉普拉斯近似对潜在高斯模型进行近似贝叶斯推断。《皇家统计学会杂志：*B*辑（统计方法论）》，71(2):319–392。Simpson, D., Rue, H., Riebler, A., Martins, T. G., 和 Sørbye, S. H.（2017）。惩罚模型组件复杂度：构建先验的一种严谨实用方法。《统计科学》，32(1):1–28。Sørbye, S. H. 和 Rue, H.（2014）。空间建模中内在高斯马尔可夫随机场先验的尺度调整。《空间统计学》，8:39–51。Wakefield, J.（2007）。计数数据的疾病制图与空间回归。《生物统计学》，8(2):158–183。Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., 和 Hui, F. K. C.（2015）。变量众多：群落生态学中的联合建模。《生态学与进化趋势》，30(12):766–779。

Zhang, Y. D., Naughton, B. P., Bondell, H. D., 和 Reich, B. J.（2022年）。基于模型拟合先验的贝叶斯回归：r2-d2 收缩先验。美国统计学会会刊，117(538):862–874。

补充材料

用于采用分层分解先验的贝叶斯物种分布模型

S1 空间效应设定

考虑第 2 节中介绍的 NOAA-NEFSC数据集。图 7 展示了该数据的时空概况。

样本中的空间位置均位于北大西洋的一个特定区域，即美国东北大陆架。如图8 (a)所示，该区域的采样并不均匀，有些区域的采样比其他区域更为充分。我们考虑采用均匀分布，使感兴趣区域内各处具有相同的重要性，从而能够更直观地定义可归因于空间分量的方差贡献。具体而言，我们将感兴趣区域定义为一个包含样本中所有地点的多边形。该多边形旨在覆盖所有观测到的空间位置，并使用凹包算法确定（Gombin等人，2020年）：最终形状在图8 (a)中以绿色阴影区域表示。随后，在多边形区域内生成大量均匀分布的点，以快速近似计算该分布下所需的量。

一旦明确定义了 $\boldsymbol{Z}$ 的分布，便可相应地指定并缩放空间效应。这里，我们选择二维P样条（Lang和Brezger (2004）来拟合连续空间坐标上的变异性。在该数据集中，地点之间的距离差异很大，因此可以采用多种分辨率来拟合空间趋势。这里，我们选择捕捉较大尺度的空间格局，创建一个由50x50 km网格单元构成的网格，并在每个网格单元中放置一个二维B样条。二维B样条基可以通过克罗内克

积由两个单变量B样条基构成（本例中，两个基函数之间的距离均为50 km）。图8（b）展示了网格以及其中一个基函数在感兴趣多边形上的取值。对于精度矩阵，我们可以直接使用基于整个矩形网格构造的一阶内蕴高斯马尔可夫随机场。不过，只有在确实要拟合整个矩形曲面时，这种方法才合适。然而在本例中，感兴趣的区域仅为不规则多边形。因此，我们决定移除在整个多边形区域内取值均为零的所有基函数。图$K_S = 267$所示的保留基函数对应的网格单元（ $\boldsymbol{W}$ ）以红色标出。从这些红色网格或格网上，我们可以得到邻接矩阵，该矩阵 $\boldsymbol{W}$ 用于构造精度矩阵$\boldsymbol{Q}_S$，它对应于图8（c）中标出的不规则格网上的一阶内蕴高斯马尔可夫随机场：

$$\boldsymbol{Q}_S = \mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}.$$

因此，该模型可以定义为：

$$f_S(\boldsymbol{Z}) = \boldsymbol{B}^T(\boldsymbol{Z})\boldsymbol{v}_S,$$
$$\boldsymbol{v}_S|\sigma_S^2 \sim N\left(\mathbf{0}, \sigma_S^2 \frac{\boldsymbol{Q}_S^*}{C_S}\right)$$

约束条件为 $\boldsymbol{S}_S^T\boldsymbol{v}_S = \mathbf{0}$，其中 $\boldsymbol{B}(\cdot,\cdot)$是按上文详述方式构建的二元B样条， $\boldsymbol{S}_S$ 是 $\boldsymbol{Q}_S$的零空间。

然而，约束 $\boldsymbol{S}_S^T\boldsymbol{v}_S = 0$并不意味着 $E_{\boldsymbol{Z}}[f_S(\boldsymbol{Z})|\boldsymbol{v}_S] = 0$而仅意味着 $\sum_{k=1}^{K_S} v_{S,k} = 0$。这意味着该效应受到一项约束，而这项约束无法直接在空间坐标的尺度上解释，因此不便使用，并且会与截距参数产生可识别性问题μ。遵循 Ferrari 和 Ventrucci(2025)，我们将精度矩阵替换为一个修改后的版本 $\widetilde{\boldsymbol{Q}}_S$ ，其零空间为 $\widetilde{\boldsymbol{S}}_S$ ，使得 $\widetilde{\boldsymbol{S}}_S^T\boldsymbol{v}_S = 0 \implies$

$E_{\boldsymbol{Z}}[f_S(\boldsymbol{Z})] = 0$，这样更方便。

S2 比例的 PC_0 先验简化形式的证明

设 X_0 和 X_1 为两个一般的协变量，它们也可以是多元的，也可以表示时空协变量。设 N_0 为满足 $\pi_{X_0}(x) > 0$的取值集合 $x \in \mathcal{X}_0$ 的基数，即 $\mathcal{X}_0$ 是$\mathcal{X}_0$ 的支撑集，并设 N_1 对 X_1也具有相同定义。若 N_0 或 N_1 为无限集，则对概率分布进行离散化，使这些量仍然很大但有限。设 $f(X_0, X_1; \omega)$ 表示分解树中二元分裂的父节点内效应的加权和，即：

$$f(X_0, X_1; \omega) = \sqrt{1-\omega} f_0(X_0) + \sqrt{\omega} f_1(X_1) \qquad \omega \in [0, 1]$$

$$\begin{aligned} f_0(X_0) &= \boldsymbol{D}_0^T(X_0)\boldsymbol{u}_0 \qquad \underset{K_0 \times 1}{\boldsymbol{u}_0} \sim N(\boldsymbol{0}, \boldsymbol{Q}_0^*) \\ f_1(X_1) &= \boldsymbol{D}_1^T(X_1)\boldsymbol{u}_1 \qquad \underset{K_1 \times 1}{\boldsymbol{u}_1} \sim N(\boldsymbol{0}, \boldsymbol{Q}_1^*). \end{aligned}$$

设 $\boldsymbol{D}_0(\cdot)$ 是由 $K_0 \leq N_0$ 个函数组成的基，$\boldsymbol{D}_1(\cdot)$ 是由 $K_1 \leq N_1$个函数组成的基。设 $\boldsymbol{x}_0, \boldsymbol{x}_1$ 为维数为 $N \leq N_0 \times N_1$ 的行向量，使得它们的 i^{th} 个元素构成唯一的一对 $x_{i0} \in \mathcal{X}_0, x_{i1} \in \mathcal{X}_1, \pi_{X_0,X_1}(x_{i0}, x_{i1}) > 0$：根据设计，$x_{i0}, x_{i1},\ i = 1,..., N$这些配对涵盖了所有可能观测到的协变量实现。因此，$\boldsymbol{D}_0(\boldsymbol{x}_0)$ 和 $\boldsymbol{D}_1(\boldsymbol{x}_1)$ 分别成为维数为 $K_0 \times N$和 $K_1 \times N$的矩阵。

$\boldsymbol{\Sigma}(\omega)$ 的 $f(\boldsymbol{x}_0, \boldsymbol{x}_1; \omega)$ 协方差矩阵等于：

$$\begin{gathered} \boldsymbol{\Sigma}(\omega) = (1-\omega)\boldsymbol{\Sigma}_0 + \omega\boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_0 = \boldsymbol{D}_0^T(\boldsymbol{x}_0)\boldsymbol{Q}_0^*\boldsymbol{D}_0(\boldsymbol{x}_0); \qquad \boldsymbol{\Sigma}_1 = \boldsymbol{D}_1^T(\boldsymbol{x}_1)\boldsymbol{Q}_1^*\boldsymbol{D}_1(\boldsymbol{x}_1). \end{gathered}$$

可以得到 $\boldsymbol{\Sigma}(\omega)$（下文称为 $R(\omega)$）的秩上界：秩$(\boldsymbol{\Sigma}_0)$+秩$(\boldsymbol{\Sigma}_1)$。注意到 $R(0)$ =秩$(\boldsymbol{\Sigma}_0)$，且 $R(1)$ =秩$(\boldsymbol{\Sigma}_1)$，这些量的上界为：

$$R(0) \leq \min[N_0, K_0] \qquad R(1) \leq \min[N_1, K_1] \tag{13}$$

我们可以得到 $R(\omega)$ 的上界为：

$$R(\omega) \leq \min[N_0, K_0] + \min[N_1, K_1]. \tag{14}$$

由于该上界可能小于 N，因此 $\pi(\boldsymbol{y};\ \omega)$ 的概率密度 $f(\boldsymbol{x}_0,\ \boldsymbol{x}_1;\ \omega)$ 可以使用非正常形式（Rue and Held，2005）表示；当$R(\omega) = N$时，该形式可简化为经典形式：

$$\pi(\boldsymbol{y};\omega) = \frac{1}{\sqrt{(2\pi)^{R(\omega)} \cdot |\boldsymbol{\Sigma}(\omega)|^*}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T \boldsymbol{\Sigma}^*(\omega)\boldsymbol{y}\right). \tag{15}$$

注意，$|\cdot|^*$ 表示广义行列式（即非零特征值的乘积）。

推导 ω 的 PC先验的第一步，是计算 $f(\boldsymbol{x}_0,\ \boldsymbol{x}_1;\ \omega)$ 与 $f(\boldsymbol{x}_0,\ \boldsymbol{x}_1;\ \omega_0)$ 之间的 KL散度，其中 ω_0 是选定的基准模型。在此背景下，我们仅考虑 $\omega_0 = 0$ 的情况；注意，$\omega_0 = 1$ 的情况与之等价。在将 KL散度计算为 ω 的函数后，假设距离 $d(\omega;\ \omega_0) = \sqrt{2 \cdot KLD[\pi(\boldsymbol{y};\ \omega)||\pi(\boldsymbol{y};\ \omega_0)]}$ 服从（截断）指数分布，并求解 ω，即可得到 PC先验的函数形式。

使用公式（15）中的 f（$\boldsymbol{x}_0,\ \boldsymbol{x}_1;\ \omega$），KL散度-

基于距离 $d(\omega;\omega_0)$ 简化为：

$$d(\omega;\omega_0)=\sqrt{\operatorname{tr}\left[\boldsymbol{\Sigma}^*(\omega_0)\boldsymbol{\Sigma}(\omega)\right]-R(\omega)-\log\frac{|\boldsymbol{\Sigma}(\omega)|}{|\boldsymbol{\Sigma}(\omega_0)|}+\left[R(\omega_0)-R(\omega)\right]\log(2\pi)}. \tag{16}$$

两个多元高斯分布之间基于 KLD 的距离可通过以下步骤简化为公式（16）：

$$\begin{aligned}
d(\omega;\omega_0)&=\sqrt{2\cdot KLD[\pi(\boldsymbol{y};\omega)||\pi(\boldsymbol{y};\omega_0)]}\\
&=\sqrt{2\int\log\left[\frac{\sqrt{(2\pi)^{R(\omega_0)}\cdot|\boldsymbol{\Sigma}(\omega_0)|}\exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^*(\omega)\boldsymbol{y}\right)}{\sqrt{(2\pi)^{R(\omega)}\cdot|\boldsymbol{\Sigma}(\omega)|}\exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}^*(\omega_0)\boldsymbol{y}\right)}\right]\pi(\boldsymbol{y};\omega)\mathrm{d}\boldsymbol{y}}\\
&=\sqrt{\left[R(\omega_0)-R(\omega)\right]\log(2\pi)+\log\frac{|\boldsymbol{\Sigma}(\omega_0)|}{|\boldsymbol{\Sigma}(\omega)|}+\int\boldsymbol{y}^T[\boldsymbol{\Sigma}^*(\omega_0)-\boldsymbol{\Sigma}^*(\omega)]\boldsymbol{y}\cdot\pi(\boldsymbol{y};\omega)\mathrm{d}\boldsymbol{y}}\\
&=\sqrt{\left[R(\omega_0)-R(\omega)\right]\log(2\pi)+\log\frac{|\boldsymbol{\Sigma}(\omega_0)|}{|\boldsymbol{\Sigma}(\omega)|}+E_{\pi(\boldsymbol{y};\omega)}\left[\boldsymbol{y}^T[\boldsymbol{\Sigma}^*(\omega_0)-\boldsymbol{\Sigma}^*(\omega)]\boldsymbol{y}\cdot\pi(\boldsymbol{y};\omega)\right]}
\end{aligned}$$

回顾二次型期望的公式，可得：

$$\begin{aligned}
E_{\pi(\boldsymbol{y};\omega)}\left[\boldsymbol{y}^T[\boldsymbol{\Sigma}^*(\omega_0)-\boldsymbol{\Sigma}^*(\omega)]\boldsymbol{y}\cdot\pi(\boldsymbol{y};\omega)\right]&=\operatorname{tr}\{[\boldsymbol{\Sigma}^*(\omega_0)-\boldsymbol{\Sigma}^*(\omega)]\boldsymbol{\Sigma}(\omega)\}\\
&=\operatorname{tr}[\boldsymbol{\Sigma}^*(\omega_0)\boldsymbol{\Sigma}(\omega)]-R(\omega)
\end{aligned}$$

方程（16）仍然是一个复杂的公式，并未给出 ω的先验的简单函数形式，因为它直接依赖于所涉及效应的所选基和精度矩阵。

秩和条件是检查 $\boldsymbol{\Sigma}_0$ 和 $\boldsymbol{\Sigma}_1$ 的非零特征值个数之和，或等价地，其秩之和是否小于或等于

维数 N。

$$R(0) + R(1) \leq N. \tag{17}$$

这个不等式会带来许多推论。首先，假设 $\mathbf{\Sigma}_0$ 和 $\mathbf{\Sigma}_1$ 都不可能是零矩阵，则它意味着二者都是奇异的。更重要的是，它保证了 $\mathbf{\Sigma}(\omega)$ 可以改写为一种更便于处理的形式。

设 $\boldsymbol{e}_0$、1、...、$\boldsymbol{e}_{0,\ R(0)}$ 是与非零特征值 λ_0、1、...、$\lambda_{0,\ R(0)}$ 相关联的 $R(0)$ 个特征向量。设 $\boldsymbol{e}_1$、1、...、$\boldsymbol{e}_{1,\ R(1)}$ 是与非零特征值 λ_1、1、...、$\lambda_{1,\ R(1)}$ 相关联的 $R(1)$ 个特征向量。如果秩之和条件（17）成立，则 $\mathbf{\Sigma}(\omega)$ 可以用 $\mathbf{\Lambda}_0$ =diag（λ_0、1、...、$\lambda_{0,\ R(0)}$、0....、0）、 $\mathbf{\Lambda}_1$ =diag（0、...、0、 λ_1、1、...、$\lambda_{1,\ R(1)}$）以及一个共同特征基 $\boldsymbol{V} = [\boldsymbol{e}_{0,1}, ..., \boldsymbol{e}_{0,R(0)}, \mathbf{0}, ..., \mathbf{0}, \boldsymbol{e}_1, ..., \boldsymbol{e}_{R(1)}]$ 表示为：

$$\mathbf{\Sigma}(\omega) = \boldsymbol{V}[(1-\omega)\mathbf{\Lambda}_0 + \omega\mathbf{\Lambda}_1]\boldsymbol{V}^T.$$

重写 $\mathbf{\Sigma}(\omega)$ 可确保距离函数 (16) 能够进一步简化。

在共同特征基 $\boldsymbol{V}$ 介于协方差矩阵 $\mathbf{\Sigma}_0$ 与 $\mathbf{\Sigma}_1$之间的假设下，可以按照 Franco-Villoria 等人 (2022年) 中给出的证明来简化该距离。首先，注意到 $\mathbf{\Sigma}(\omega)$ 及其广义逆可以改写为：

$$\begin{aligned}
\mathbf{\Sigma}(\omega) &= \boldsymbol{V}\left[(1-\omega)\mathbf{\Lambda}_0 + \omega\mathbf{\Lambda}_1\right]\boldsymbol{V}^T \\
\mathbf{\Sigma}^*(\omega) &= \boldsymbol{V}\left[(1-\omega)\mathbf{\Lambda}_0 + \omega\mathbf{\Lambda}_1\right]^*\boldsymbol{V}^T
\end{aligned}$$

然后可以看出，方程（16）的第一项可化简为

变为仅依赖于 $\mathbf{\Lambda}_0$ 和 $\mathbf{\Lambda}_1$中特征值的函数:

$$\begin{aligned}\operatorname{tr}\left[\mathbf{\Sigma}^*(\omega_0)\mathbf{\Sigma}(\omega)\right] &= \operatorname{tr}[\boldsymbol{V}\left[(1-\omega_0)\mathbf{\Lambda}_0+\omega_0\mathbf{\Lambda}_1\right]^*\boldsymbol{V}^T\boldsymbol{V}\left[(1-\omega)\mathbf{\Lambda}_0+\omega\mathbf{\Lambda}_1\right]\boldsymbol{V}^T] \\ &= \operatorname{tr}[\left[(1-\omega_0)\mathbf{\Lambda}_0+\omega_0\mathbf{\Lambda}_1\right]^*\left[(1-\omega)\mathbf{\Lambda}_0+\omega\mathbf{\Lambda}_1\right]] \\ &= \sum_{n:\lambda_{0,n}+\lambda_{1,n}>0}\frac{(1-\omega)\lambda_{0,n}+\omega\lambda_{1,n}}{(1-\omega_0)\lambda_{0,n}+\omega_0\lambda_{1,n}}\end{aligned}$$

最后，在共同特征基假设下，方程（16）的第二项也可以简单地用特征值表示为:

$$\begin{aligned}\log\frac{|\mathbf{\Sigma}(\omega)|}{|\mathbf{\Sigma}(\omega_0)|} &= \log\frac{|(1-\omega)\mathbf{\Lambda}_0+\omega\mathbf{\Lambda}_1|}{|(1-\omega_0)\mathbf{\Lambda}_0+\omega_0\mathbf{\Lambda}_1|} \\ &= \sum_{n:\lambda_{0,n}+\lambda_{1,n}>0}\log\frac{(1-\omega)\lambda_{0,n}+\omega\lambda_{1,n}}{(1-\omega_0)\lambda_{0,n}+\omega_0\lambda_{1,n}}\end{aligned}$$

此时，距离函数还可以进一步简化，因为在秩和条件下，对于每个 $n\in[1,N]$，$\lambda_{0,\,n}$ 和$\lambda_{1,\,n}$ 之间至少有一个特征值为 0。这将得到以下简化形式:

$$\frac{(1-\omega)\lambda_{0,n}+\omega\lambda_{1,n}}{(1-\omega_0)\lambda_{0,n}+\omega_0\lambda_{1,n}}=\begin{cases}\frac{1-\omega}{1-\omega_0}, & \lambda_{0,n}>0,\lambda_{1,n}=0\\ \frac{\omega}{\omega_0}, & \lambda_{0,n}=0,\lambda_{1,n}>0\end{cases}$$

距离函数的最终形式如下:

$$d(\omega;\omega_0)=\sqrt{\begin{aligned}&R(0)\frac{1-\omega}{1-\omega_0}+R(1)\frac{\omega}{\omega_0}-R(0)\log\frac{1-\omega}{1-\omega_0}-R(1)\log\frac{\omega}{\omega_0}\\&-R(\omega)+[R(\omega_0)-R(\omega)]\log(2\pi)\end{aligned}}$$

该函数不包含 $\mathbf{\Sigma}_0$ 和 $\mathbf{\Sigma}_1$ 的特征值，但仍取决于 $R(\omega)$，因此计算效率较低。不过，注意到对于 $\omega_0=0$，距离不是有限值，因此必须将其作为极限来计算。

首先考虑 $d(\omega;\omega_0)^2\omega_0$:

$$\begin{aligned} d(\omega;\omega_0)^2\omega_0 = {} & \omega_0 R(0)\frac{1-\omega}{1-\omega_0} - \omega_0 R(0)\log\frac{1-\omega}{1-\omega_0} \\ & + R(1)\omega - \omega_0 R(1)\log\frac{\omega}{\omega_0} \\ & - \omega_0 R(\omega) + \omega_0[R(\omega_0) - R(\omega)]\log(2\pi) \end{aligned}$$

计算 $\omega_0 \to 0$ 的极限，可得:

$$\lim_{\omega_0\to 0} d(\omega;\omega_0)^2\omega_0 = R(1)\omega$$

然而，对于 $\omega_0 = 0$，该距离并非有限值，因此必须改为计算 $\omega_0 = 0$ (Franco-Villoria 等人，2022年）处的极限。可以证明:

$$\lim_{\omega_0\to 0} d(\omega;\omega_0) = R(1)\sqrt{\omega} \tag{18}$$

其中， $R(1)$ 是相对于 ω 的常数。对距离指定指数分布（在 $d(1)$ 处截断，因为 ω 的上界为 1），使用变量替换公式可得 PC 先验的最终形式:

$$\pi(\omega) = \frac{\widetilde{\delta}R(1)\exp(-\widetilde{\delta}R(1)\sqrt{\omega})}{2\sqrt{\omega}[1-\exp(-\widetilde{\delta}R(1))]}.$$

消除了对常数 $R(1)$ 的依赖，因为其中包含模型参数，无法与 $\widetilde{\ }$ $\widetilde{\ }$指数分布的超参数 δ识别。记 $\delta = \delta R(1)$，可得公式（6），这证明了在满足秩和条件时，PC_0 先验的简化形式必然成立。

该 δ 超参数可以使用概率陈述来选择，例如$P(\omega < U) = \alpha$。然而，为了获得有效的概率分布（即 $\delta > 0$），Franco-Villoria 等人(2022年指出，必须满足以下条件:

必须遵守：

$$\alpha \geq \sqrt{U}. \tag{19}$$

因此，中位数至多可为0.25，该值可通过 $\delta \to 0$得到。

因为秩和条件保证PC先验具有更简单的解，推导该先验时宜先检验这一条件。若使用公式（13）的上界未能满足该条件，则下一步应推导实际的 $R(0)$与 $R(1)$，并再次检验该条件。

当 $R(0)$与 $R(1)$可解析获得时，可以验证秩和假设，但实践中这种情况很少见。因此，可改用它们的上界（公式（13））来检验该假设；这些上界通常对用户是显式可得的。我们在此讨论三种情形，其中PC_0 先验（针对 ω ）是合理选择，并可能源于为物种分布模型设计的分解树（图1）。

线性效应与非线性效应。设 $X = X_0 = X_1$ 为单变量。因此，$N_0 = N_1 = N$ ，因为这两个协变量完全相依。设 $f_0(X)$为线性效应（在对 X进行标准化之后），$f_1(X)$为具有有限个 K_1 系数的非线性效应，并在设计上被约束为零截距和线性趋势。$R(0)$与 $R(1)$的上界为： $R(0) \leq 1$, $R(1) \leq K_1$ 。因此，只要$K_1 < N$，它们的和$R(0) + R(1) \leq K_1 + 1$ 便满足公式（17）的条件。这一情形可能由图1中该树第4层的分割产生。

线性主效应与交互效应。 令 $X_0 = X_1 = [X_A, X_B]$，其中 X_A， X_B 是两个相互独立的单变量协变量。同样， $N_0 = N_1 =$

N和 $N = N_A \cdot N_B$。令 $f_0(X_A, X_B)$和 $f_1(X_A, X_B)$定义如下:

$$f_0(X_A, X_B) = \sqrt{1-\phi}\widetilde{X_A}u_A + \sqrt{\phi}\widetilde{X_B}u_B; \qquad f_1(X_A, X_B) = \widetilde{X_A}\widetilde{X_B}u_1$$

$\widetilde{\ }$ $\widetilde{\ }$其中 X_A、X_B是 X_A的标准化版本 X_B。在这种情况下，R（1）$\leq$1. R（0）取决于 ϕ 的值，但仍可使用方程（14）的相同公式求得其上界：R（0）$\leq$ 最小值$[N_A, 1]$ + 最小值$[N_B, 1]$ $\Longrightarrow R$（0）$\leq$ 2.因此，只要满足 3 $\leq N_A \cdot N_B$，秩和条件就得到满足；当 X_A、X_B均为连续协变量时，该条件始终成立。此情形涵盖了图1中树的左分支在第3层的分割。

有限维主效应与克罗内克积交互效应。考虑上述情形，现在将 $f_0(X_A$、$X_B)$ 定义为 X_A 和 X_B 的两个有限维效应的加权和，其系数分别为 K_A 和 K_B，且 $K_A \leq N_A$、$K_B \leq N_B$:

$$f_0(X_A, X_B) = \sqrt{1-\phi}\boldsymbol{D}_A^T(X_A)\boldsymbol{u}_A + \sqrt{\phi}\boldsymbol{D}_B^T(X_B)\boldsymbol{u}_B$$
$$\underset{K_A \times 1}{\boldsymbol{u}_A} \sim N(\boldsymbol{0}, \boldsymbol{Q}_A^*); \qquad \underset{K_B \times 1}{\boldsymbol{u}_B} \sim N(\boldsymbol{0}, \boldsymbol{Q}_B^*).$$

令 $f_1(X_A, X_B) = \boldsymbol{D}^T(X_A, X_B)\boldsymbol{u}$表示一种交互效应，它由 $f_0(X_A, X_B)$中主效应的基与精度矩阵的克罗内克积构造而成，即 IV 由 Knorr-Held (2000) 为不可分时空效应定义的交互效应类型。

$$\boldsymbol{D}(X_A, X_B) = \boldsymbol{D}_A(X_A) \otimes \boldsymbol{D}_B(X_B); \qquad \underset{K_A \cdot K_B \times 1}{\boldsymbol{u}} \sim N(\boldsymbol{0}, [\underset{K_A \times K_A}{\boldsymbol{Q}_A} \otimes \underset{K_B \times K_B}{\boldsymbol{Q}_B}]^*).$$

.....的上界 R（0）和 R（1）容易求得为 R（0）$\leq K_A + K_B$ R（1）$\leq K_A \cdot K_B$。因此，只要满足以下条件，秩和条件就成立:
$K_A + K_B + K_A \cdot K_B \leq N_A \cdot N_B$。如果 X_A 和 X_B 是连续协变量，则该条件总是成立。在许多其他情况下，该条件也成立，例如X_A 和 X_B 为离散变量时，包括Franco-

Villoria等人（2022年）。后一种情况涵盖了主要空间效应和时间效应与其非加性交互作用之间可能的划分（第3层，图1右侧分支）。

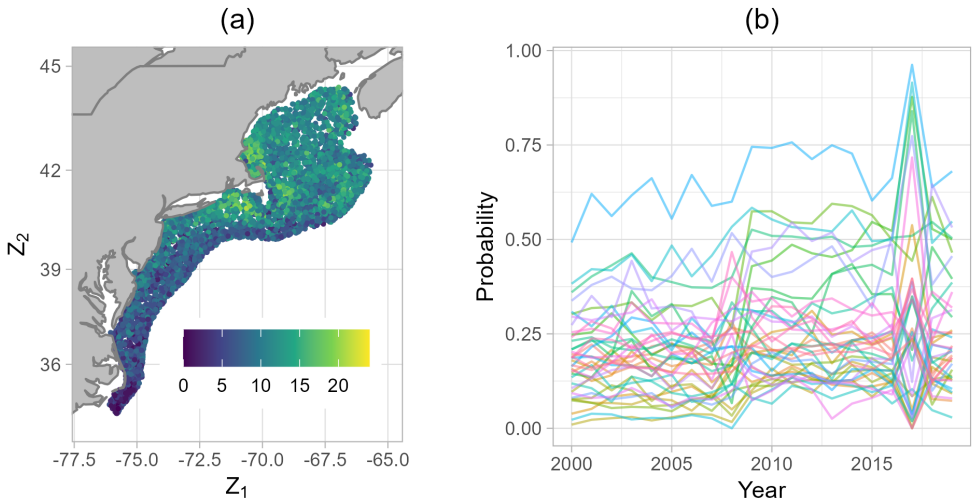


图7：案例研究数据集概览：(a) 各地点检测到的不同物种数量；(b) 39 个物种各自的出现观测比例随时间的变化。

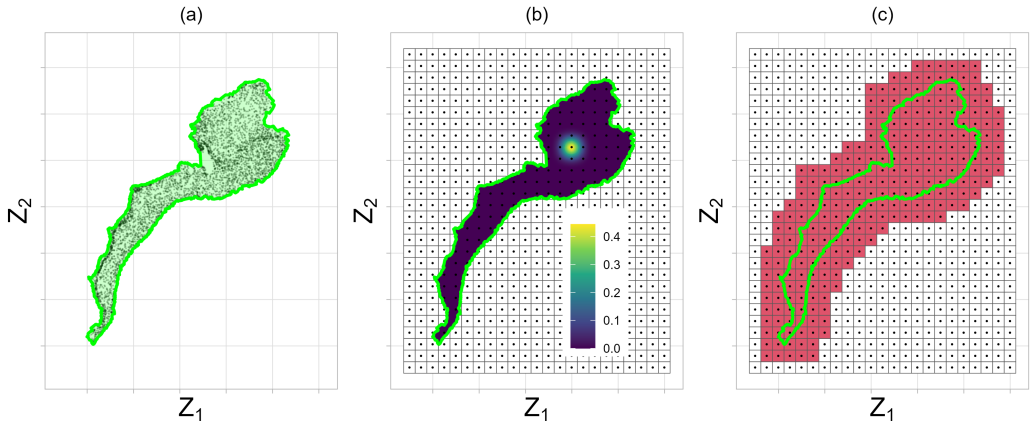


图8：（a）在 Z_1，Z_2 空间坐标系中以黑点表示的采样位置，以及包含所有数据点、以绿色描边并填充的所选凹多边形。（b）50×50千米网格及其网格单元质心（黑点），并展示感兴趣区域内以某个网格单元为中心的一个B样条基函数的取值。（c）红色网格单元表示相应的基函数（以其质心为中心）在实线绿色边界所示多边形上取非零值。

.