arXiv:2302.03148v1 [stat.AP] 6 Feb 2023

基于潜在高斯模型和非平稳随机偏微分方程结构的高分辨率全球降水降尺度

张家晨 应用与计算数学及统计系，圣母大学（美国）和
Matthew Bonas 应用与计算数学及统计系，圣母大学（美国）和
Diogo Bolster 土木与环境工程系及地球科学系，圣母大学（美国）和
Geir-Arne Fuglstad 数学科学系，挪威科技大学（ NTNU ），挪威特隆赫姆和 Stefano Castruccio*应用与计算数学及统计系，圣母大学（美国）

2023年2月8日

*scastruc@nd.edu

摘要

获取高空间分辨率的降水数据地图，可以为利益相关者在城市尺度评估水资源的可持续获取提供重要依据。在极高空间分辨率下，对降水这一非平稳稀疏过程进行制图，需要在有地面观测站的位置对全球数据集进行插值，并采用能够捕捉复杂非高斯全球时空依赖结构的统计模型。在本研究中，我们提出了一种新方法：通过带有缓冲区的局部变形随机偏微分方程（SPDE），捕捉潜在高斯过程的空间依赖，从而允许陆地和海洋具有不同的空间结构。SPDE 的有限体积近似结合积分嵌套拉普拉斯近似，可为数千万条观测数据实现可行的贝叶斯推断。模拟研究表明，与平稳模型和无缓冲区替代方案相比，所提出方法具有更强的可预测性。随后，我们利用该方法生成了覆盖美国的高分辨率日降水量模拟结果。

关键词：潜在高斯分布模型；降水；随机偏微分方程；积分嵌套拉普拉斯近似

1 引言

准确的高分辨率降水数据对于有效预测和管理水资源至关重要（Clark 等，2015）。对驱动降水的物理过程进行建模方面取得了显著进展，使全球气候模型能够进行更真实的模拟，从而提供更可靠的预测。然而，现代气候模型的复杂度极高，这意味着计算和存储成本高昂，限制了全球气候模拟所能达到的空间分辨率。因此，由于粗分辨率无法充分呈现降水模式，也无法捕捉所关注物理过程的尺度，模拟结果与观测数据之间存在显著的不确定性和偏差（Wood 等，2021）。这可能导致对特定位置的影响被高估或低估，或对事件发生时间的判断有误；例如，这可能决定某地是否发生洪涝（Sapountzis 等，2021）。因此，精细化全球预测，并以高空间尺度绘制降雨发生概率和降水强度图，对于科学研究具有重要意义，能够为洪涝韧性影响评估模型和用于干旱预测的农业模型提供依据。

原则上，可以将粗分辨率全球数据集作为区域天气模型（例如天气研究与预报模型（WRF，Skamarock 等人（2019））的边界条件，从而生成高分辨率降水。这种动力降尺度方法（Sain 等，2011）的优点是能够生成物理上自洽的高分辨率空间场，但同时也需要付出相当大的计算和存储资源成本，

此外，模型设置还需要专业知识，只有少数研究中心、大学或企业能够负担这些投入。一种成本更低的解决方案，是建立全球数据与地面观测之间的经验关系，并在有地面数据的位置拟合该关系。假设这一关系在未观测位置至少近似成立，就可以通过校正全球数据集生成高分辨率地图。这种统计降尺度方法（Berrocal 等人，2010）速度快、计算成本低，并且在地球科学文献中长期以来已被证明卓有成效。要使这种方法奏效，全球数据和地面数据必须位于相同位置；但这并非理所当然，因为全球数据是以大范围区域的平均值定义的。因此，有必要使用空间统计模型，在地面观测的相同位置对全球模拟值进行插值，并评估这些估计值的不确定性。

全球空间数据需要构建专门的模型，其理论属性与欧几里得空间上的空间过程存在显著差异。事实上，Gneiting（2013）指出，若采用应用最广泛的协方差模型 Mat´ern 函数，要构造一个以大圆距离为度量的有效球面过程，就必须对参数空间施加严格限制。过去二十年间，针对全球数据的新建模方法不断涌现。其中，Jun 和 Stein（2007，2008）提出将球面嵌入三维空间，采用 Mat´ern 模型并应用偏导数，以获得更大的灵活性。这类模型不仅能够刻画各向同性行为，还能刻画轴对称性，即跨纬度的非平稳行为——

变化（Jones，1963）。Jun（2011）将这一方法推广到多元全局过程。针对网格数据，Castruccio 和 Stein（2013）提出了一类快速且灵活的轴对称模型谱类。随后，该方法被推广至非参数谱估计（Castruccio 和 Genton，2014）、三维变量（Castruccio 和 Genton，2016）、不同的陆地/海洋行为（Castruccio and Guinness，2017），以及多元过程（Edwards et al.，2019）。在理论方面，通过基分解确定各向同性过程在高维球面上的性质，已取得重大进展，参见 Arafat 等（2020）和 Porcu 等（2020）。关于这一主题的两篇近期综述，可参见 Jeong 等（2017）和 Porcu 等（2018）。

Lindgren 等人（2011）的开创性工作提出了一种新颖而不同的视角：将一类 Matˊern 模型与具有马尔可夫性质的扩散-反应随机偏微分方程（SPDE）的解联系起来，并采用有限体积进行推断。就全球模型而言，这种方法的关键洞见在于，平面上的原始 SPDE 可以直接适用于球面，同时还具有无需边界条件的额外优势。尽管 SPDE 在其最初的表述中产生的是平稳模型，但通过引入空间变化系数，研究者已经提出了非平稳扩展方案。从嵌套随机偏微分方程（Bolin 和 Lindgren，2011）到带有物理屏障的模型（Bakka 等人，2019），人们已经提出了多种替代方案。最近，Fuglstad 等人（2015）；Fuglstad 和 Castruccio（2020）通过允许 SPDE 具有局部变形，扩展了这一方法

即通过空间变化的标量场和向量场实现。所提出的方法展现出良好的应用前景，但迄今为止仅限于高斯情形；鉴于非高斯数据建模所面临的挑战，以及这些模型所带来的计算开销，将其推广到非高斯数据绝非易事。

本文提出一种基于随机偏微分方程的非高斯、非平稳全球时空模型，该模型采用局部变形，并在陆地与海洋之间设置缓冲区，以刻画空间依赖的突变。模型通过潜在高斯模型对非高斯性进行建模，即假设非高斯边缘行为在给定条件下于各位置之间相互独立，再通过具有高斯结构的潜在过程捕捉空间依赖。借助以下方法，仍可对超大规模数据集进行推断：1）由随机偏微分方程的有限体积解产生的潜在高斯模型的稀疏精度矩阵；2）利用积分嵌套拉普拉斯近似（INLA，Rue 等人（2009））快速近似后验计算所需的高维积分。该模型尤其适用于全球日降水量等高度非高斯数据，因此我们将其用于：1）拟合全球再分析数据；2）提供与地面观测位置相同位置处的插值数据；3）结合地面观测和插值数据进行降水降尺度；从而4）生成降水高分辨率地图。

本文结构安排如下。第2节介绍本文所用的数据。第3节详细阐述潜在高斯模型的方法论，具体包括时间分量和空间分量。第4节说明如何进行推断，以及

稀疏性和数值近似如何减轻计算负担。第5节通过数值分析评估后验一致性，并考察所提模型相较于更简单的替代模型在可预测性方面的提升。第6节将所提模型应用于降水数据，并展示其能够为美国大陆提供日降水量的高分辨率地图。第7节以讨论作结。为确保可复现性，本文末尾提供了代码和数据所在代码库的信息。

2 数据描述

我们重点研究由美国国家航空航天局全球建模与同化办公室（GMAO）制作的现代时期回顾性研究与应用分析第2版（MERRA-2，Gelaro 等人（2017））中的全球日降水量数据。MERRA-2 是一种再分析数据产品，整合了来自卫星仪器的观测数据，被认为是对地球系统状态的最佳表征之一。该数据采用规则网格，网格分辨率分别为经度和纬度方向的 $0.625^\circ \times 0.5^\circ$，共包含 $n = 207,936$ 个位置。我们重点关注有连续记录的最近年份，即 2021 年，并使用每日最大降雨率（MRR，单位为 $\mathrm{kg/m^2\cdot s}$）。为将 MRR 转换为降水量，我们将其除以水密度 1,000（$\mathrm{kg/m^3}$），再乘以 1,000 将单位换算为毫米，同时乘以 $86{,}400s$，以得到日降水量。我们假设每个位置的 MRR 持续一整天，这

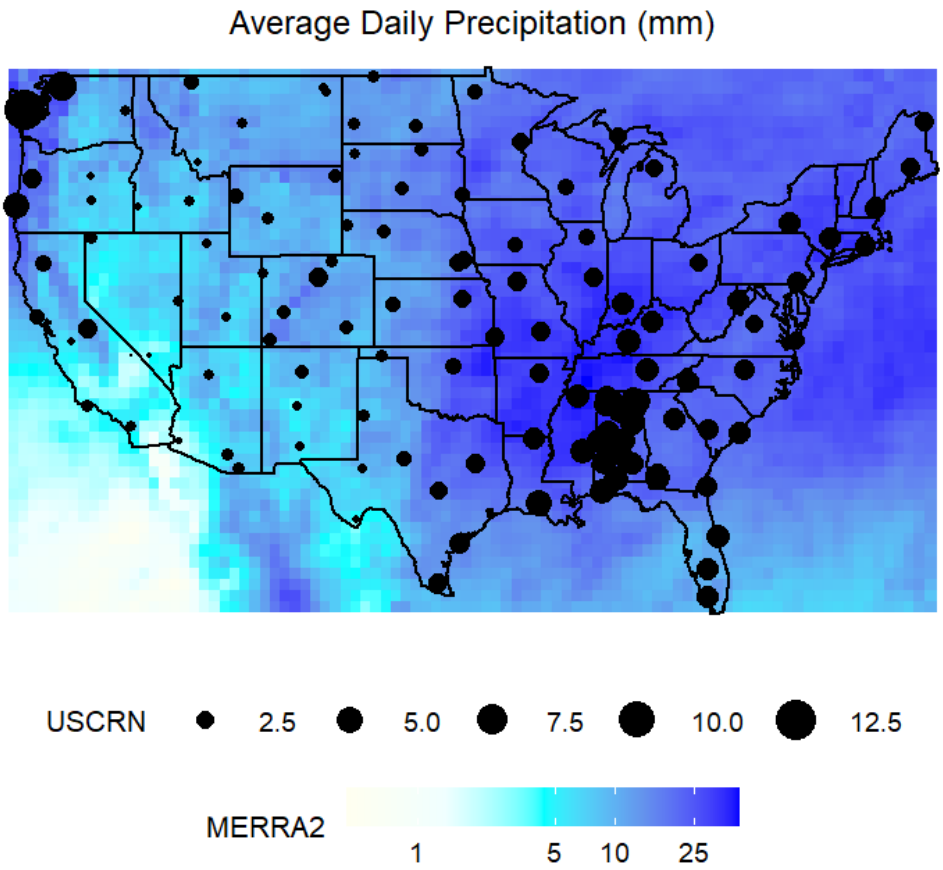

图1：2021年1月 1st日至2021年12月 31th日期间，各 USCRN 站点和 MERRA-2 网格点的日平均降水量（单位：毫米）。

会导致一定程度的高估，这一点从图1中两种不同的图例刻度可以清楚看出。第6节中的降尺度方法将通过对（插值后的）MERRA-2 与 USCRN 进行线性变换来解决这一问题。

对于地面观测数据，我们采用美国地表气候参考网络（USCRN，美国国家海洋和大气管理局（2022））的adata产品，其中包含来自美国大陆各地气候监-

测站的连续记录。USCRN监测站实时记录总降水量，单位为毫米（mm），记录间隔为5分钟。数据由Geonor T-200B降水计采集，其最大容量为600 mm。该降水计使用一个降水收集桶，桶的周围设有风雪防护罩，并通过加热防止寒冷地区发生冰积聚。连接到收集装置上的三根金属丝会以与桶重相关的频率振动，随后将这些振动频率转换为仪表深度（单位为mm）。在本研究中，我们采用来自131个不同监测站的数据，这些数据经过后处理，转换为日分辨率，并形成从 1^{st}年1月到 31th 2021年12月的连续记录。图1展示了USCRN传感器的位置，以及2021年日总降水量的平均值。作为比较，同一图中还展示了同期MERRA-2网格点的日平均降水量。从图中可以清楚看出，日平均降水量最高的地区位于美国西北部和东南部，而较干燥的地区则从加利福尼亚州东部边界一直延伸至密西西比河。

3 方法论

3.1 潜在高斯模型

我们提出一种时空潜在高斯模型（Rue等，2009），其定义适用于球面上的任意空间点 $\mathbf{s} \in \mathbb{S}^2$ 以及时间 $t = 1, 2, \ldots$：

$$Y(\mathbf{s}, t) \mid \mu(\mathbf{s}, t), \boldsymbol{\theta}_{\mathrm{MRG}} \sim h(\mu(\mathbf{s}, t), \boldsymbol{\theta}_{\mathrm{MRG}}), \tag{1a}$$

$$g(\mu(\mathbf{s}, t)) = \sum_{p=1}^{P} \beta_p f_p(\mathbf{s}) + f^{\mathrm{time}}(\mathbf{s}, t) + \epsilon(\mathbf{s}), \tag{1b}$$

$$f^{\mathrm{time}}(\mathbf{s}, t) = \sum_{k=1}^{K} \left\{ \zeta_k(\mathbf{s}) \sin\left(\frac{2\pi k t}{\delta}\right) + \zeta_k'(\mathbf{s}) \cos\left(\frac{2\pi k t}{\delta}\right) \right\}, \tag{1c}$$

其中，$h(\cdot)$ 表示在给定潜在场和超参数的条件下，$Y(\cdot)$ 的边缘分布，且属于指数族，其均值为 $\mu(\mathbf{s}, t)$，其结构由潜在高斯过程通过连接函数 $g(\cdot)$ 决定。边缘参数 $\boldsymbol{\theta}_{\mathrm{MRG}}$ 刻画一阶以上的矩，也可以为空。如果边缘分布为高斯分布，则有 $Y(\mathbf{s}, t) \sim \mathcal{N}(\mu(\mathbf{s}, t), \boldsymbol{\theta}_{\mathrm{MRG}})$，连接函数 $g(\cdot)$ 就是恒等函数（Dunn 和 Smyth，2018）。例如，若边缘分布为伯努利分布，则有 $Y(\mathbf{s}, t) \sim \mathcal{B}(\mu(\mathbf{s}, t))$，并可选择 logit 函数作为连接函数（Dunn 和 Smyth，2018）。我们假设，潜在空间中的变换后均值 $g(\mu(\mathbf{s}, t))$ 由位置特定的时间效应 $f^{\mathrm{time}}(\mathbf{s}, t)$、$p = 1$、……、$P$位置特定协变量 $f_p(\mathbf{s})$ 以及空间误差 $\epsilon(\mathbf{s})$ 建模。时间效应 $f^{\mathrm{time}}(\mathbf{s}, t)$ 由 K 个谐波描述，其参数为 $\boldsymbol{\zeta}(\mathbf{s}) =$

$(\zeta_1(\mathbf{s}) \ldots, \zeta_K(\mathbf{s}))^\top$ 以及 $\boldsymbol{\zeta}'(\mathbf{s}) = (\zeta_1'(\mathbf{s}) \ldots, \zeta_K'(\mathbf{s}))^\top$。假设我们在 $\mathbf{s}1, \ldots, \mathbf{s}_n$ 处观测到一个样本，则方程 (1c) 中的时间参数总数为 $\boldsymbol{\theta}_{\text{time}} = \{\boldsymbol{\theta}_{\text{time}}(\mathbf{s}_1) \ldots, \boldsymbol{\theta}_{\text{time}}(\mathbf{s}_n)\}$，其中 $\boldsymbol{\theta}_{\text{time}}(\mathbf{s}_i) = \{\boldsymbol{\zeta}(\mathbf{s}_i)\ \boldsymbol{\zeta}'(\mathbf{s}_i)\}$，共计 $2Kn$ 个参数。周期 $\delta \in \{365, 366\}$ 取决于所考虑的是闰年还是平年。我们假设，空间随机效应 $\epsilon(\mathbf{s})$ 是一个均值为零、时间上独立的高斯随机场的一次实现，其协方差函数取决于一些参数 $\boldsymbol{\theta}_{\text{space}}$，这些参数将在下一节中说明。

3.2 空间相关结构

$\epsilon(\mathbf{s})$ 空间依赖的最简单模型是平稳且各向同性的，也就是说，它们假设这种依赖是 $\|\mathbf{s}_1 - \mathbf{s}_2\|$ 的函数。其中，最受欢迎的选择之一无疑是 Mat´ern 模型，其两个位置$\mathbf{s}_1$,$\mathbf{s}_2$ 之间的相关性定义如下（Stein，1999）

$$\text{Corr}(\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2)) = C(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|\mathbf{s}_1 - \mathbf{s}_2\|}{\rho} \right)^\nu K_\nu \left(\frac{\|\mathbf{s}_1 - \mathbf{s}_2\|}{\rho} \right),$$

其中，K_ν 是第二类修正贝塞尔函数，具有光滑度参数$\nu > 0$（即控制均方可微性的度）和范围参数$\rho > 0$。如果要对大型数据集进行推断，就无法存储包含所有位置之间协方差的矩阵，似然评估也可能在计算上极具挑战，甚至无法实现。与直接使用协方差矩阵不同，

过去十年中一种流行的解决方案是，将具有 Mat´ern 协方差的高斯分布过程识别为以下分数阶反应扩散随机偏微分方程的（唯一）平稳解（Whittle，1954）：

$$\left(\frac{1}{\rho^2} - \Delta\right)^{\nu/2+1/2} \epsilon(\mathbf{s}) = \mathcal{W}(\mathbf{s}),\ \mathbf{s} \in \mathbb{R}^2, \tag{2}$$

其中， Δ 是拉普拉斯算子， $\mathcal{W}$(s)是空间高斯白噪声。Lindgren 等人（2011年）利用当（2）中的 ν 为整数时连续马尔可夫过程与离散高斯马尔可夫随机场（GMRF）之间的“显式联系”，证明了如果所有位置都排列在二维格点上，就可以通过对稀疏精度矩阵应用卷积来近似 GMRF 的协方差结构。此外，不在格点上的位置也可以通过在定义域上进行三角剖分来插值和近似。最终，这意味着 Mat´ern 协方差可以用稀疏精度矩阵近似，从而能够更快速、可行地推断空间结构 ϵ(·)。在本研究中，我们采用了一个定义在球面上的类似 SPDE，其定义为

$$\left(\frac{1}{\rho^2} - \Delta_{\mathbb{S}^2}\right)^{\nu/2+1/2} \epsilon(\mathbf{s}) = \mathcal{W}(\mathbf{s}),\ \mathbf{s} \in \mathbb{S}^2, \tag{3}$$

其中， $\Delta_{\mathbb{S}^2}$ 是拉普拉斯算子。

上述随机偏微分方程方法具有明显的计算优势，但其公式仅适用于平稳的各向同性过程（Rue等，2009）。随机偏微分方程

算子则可以推广，以纳入非平稳构造，同时仍能得到稀疏精度矩阵。本文采用最初由 Fuglstad 等人（2019）提出的空间变化的随机偏微分方程，但也有研究提出了其他处理空间变化参数的方法（Rue等，2009）以及嵌套随机偏微分方程（Bolin 和 Lindgren，2011）。我们假设球面上的一个位置具有极坐标 $\mathbf{s} = (L,\ l)$，其中 L 是纬度，l 是经度。我们引入两个项：一个向量场 $\mathbf{v}(\cdot) = (v_1(\cdot)\ v_2(\cdot))^\top$ 以及一个取正值的标量场 $\rho(\cdot)$。随后，我们将逆变形张量定义为：

$$\mathbf{G}(\mathbf{s})^{-1} = \rho(\mathbf{s})^2 \frac{\mathbf{I}_2 + \mathbf{v}(\mathbf{s})\mathbf{v}(\mathbf{s})^\top}{\sqrt{1 + \|\mathbf{v}(\mathbf{s})\|^2}}.$$

可以证明，使用上述定义的空间变化度量张量，沿$\mathbf{v(s)}$方向的距离会缩放为 $1/(\rho\mathbf{(s)}(1 + \|\mathbf{v(s)}\|^2)^{\frac{1}{4}}$。在$\mathbf{v(s)}$的正交方向上，距离会缩放为 $(1 + \|\mathbf{v(s)}\|^2)^{\frac{1}{4}}/\rho\mathbf{(s)}$。因此，向量场 $\mathbf{v}(\cdot)$指定了每个位置处局部各向异性效应的方向，而 $\rho(\cdot)$ 表示其强度。在指定度量张量 $\mathbf{G(s)}$后，可以证明，对随机偏微分方程 (3) 进行适当的变量替换可得（Fuglstad and Castruccio，2020）：

$$[|\mathbf{G}(\mathbf{s})|^{\frac{1}{2}} - \nabla \cdot |\mathbf{G}(\mathbf{s})|^{\frac{1}{2}} \mathbf{G}(\mathbf{s})^{-1} \nabla] \epsilon(\mathbf{s}) = |\mathbf{G}(\mathbf{s})|^{\frac{1}{4}} \mathcal{W}(\mathbf{s}),\ \mathbf{s} \in \mathbb{S}^2. \tag{4}$$

3.3 球谐函数

向量场**v**(·)和标量场 ρ(·) 均可分别通过球矢量谐函数和球谐函数等基函数分解来指定。然而，全球模型需要采用更灵活的方法，不仅要考虑缓慢变化的非平稳性，还要考虑由陆地和海洋等大型地理描述符所决定的突变（Castruccio and Guinness，2017）。为了通过随机偏微分方程构建有效模型，同时仍能考虑突变，我们采用 Bakka 等人（2019）提出的缓冲方法。具体而言，我们沿海岸线设置缓冲区，并使用一个单独的参数描述缓冲区内依赖强度的乘性下降 $d \in [0,1]$，因此，我们针对陆地和海洋定义域分别提出球谐函数分解：

$$\log\{\rho^j(\mathbf{s})\} = \sum_{l=0}^{\mathcal{L}} \sum_{m=-l}^{l} \alpha_{ml}^j Y_l^m(\mathbf{s}),$$

其中，α_{ml}^j 是实值系数，Y_l^m(**s**) 是度为l 、阶为 m 的拉普拉斯球谐函数，$j = \{$land, ocean$\}$ 指定了**s** 所在位置的地理描述。类似地，向量场**v**(·) 可描述为：

$$\mathbf{v}^j(\mathbf{s}) = \sum_{l=1}^{\mathcal{L}} \sum_{m=-l}^{l} \{E_{lm}^{(1,j)} \nabla Y_m^l(\mathbf{s}) + E_{lm}^{(2,j)} \hat{\mathbf{r}}(\mathbf{s}) \nabla \times Y_m^l(\mathbf{s})\},$$

其中 $\hat{}\mathbf{r}$ 是正径向方向上的单位向量，$E_{lm}^{(1,j)}$ 和 $E_{lm}^{(2,j)}$ 是

实系数， $\mathcal{L}$是基函数中的最高阶数。此外，为了考虑微尺度变异，我们假设陆地和海洋上的过程都具有块金效应τ_j^2。总之，该模型的空间参数为 $\boldsymbol{\theta}_{\text{space}} = \Big\{d,\ \Big\{\tau_j^2,\ j \in \{\text{陆地、海洋}\}\Big\},\ \Big\{\alpha_{ml}^j,\ E_{lm}^{(1,j)},\ E_{lm}^{(2,j)},\ m = -l,\ \ldots\ldots,\ l;\ l = 1,\ \ldots\ldots,\ \mathcal{L},\ j \in \{\text{陆地、海洋}\}\Big\}\Big\}$，总计包含 $6(\mathcal{L}^2 + 2\mathcal{L}) + 3$ 参数。

我们为所有参数采用先验相互独立的标准正态分布作为先验；对于受限为正值的参数，则进行对数变换。相同的设置也应用于模拟研究和应用中使用的参数。

4 推断

我们提出一种逐步推断方法，以逐步降低参数空间的整体维度。首先，我们在每个位置独立估计 $\boldsymbol{\theta}_{\text{time}}$，然后在给定时间参数的条件下估计 $\boldsymbol{\theta}_{\text{space}}$。Edwards 等人（2020年）表明，逐步方法可实现渐近一致推断；Castruccio 和 Guinness（2017年）则表明，对于我们在此研究中使用的这类规模较大但仍有限的数据集，不确定性和偏差传播的影响较小。

4.1 步骤1：时间结构

第一步是在不考虑空间效应和协变量效应的情况下，针对每个位置独立进行推断。我们将方程（1）重新定义如下：

$$\begin{aligned} Y(\mathbf{s},t) &\sim h(\mu(\mathbf{s},t),\boldsymbol{\theta}_{\mathrm{MRG}}), \\ g(\mu(\mathbf{s},t)) &= \textstyle\sum_{p=1}^{P}\beta_p f_p(\mathbf{s}) + \sum_{k=1}^{K}\left\{\zeta_k(\mathbf{s})\sin\left(\frac{2\pi kt}{\delta}\right) + \zeta_k'(\mathbf{s})\cos\left(\frac{2\pi kt}{\delta}\right)\right\}. \end{aligned} \tag{5}$$

时间参数向量 $\boldsymbol{\theta}_{\mathrm{time}}$ 和线性参数 $\beta_1,\ldots,\beta_p$ 采用最小二乘法进行估计，并在后续推断步骤中将这些参数视为固定值。$\hat{}\,\hat{}\,\hat{}$ 得到 $\boldsymbol{\theta}_{\mathrm{time}},\beta_1,\ldots,\beta_p$ 后，以它们为条件，可以估计空间过程 $\epsilon(\mathbf{s})$ 的空间参数 $\boldsymbol{\theta}_{\mathrm{space}}$。

4.2 步骤2：空间协方差结构

我们在球面上定义一组三角形 $T_1,\ldots,T_{n_T}$，并采用有限体积法对式（4）中的随机偏微分方程进行离散化。我们将逆矩阵张量重新定义为 $\mathbf{G}(\mathbf{s})^{-1} = \rho(\mathbf{s})^2\mathbf{H}(\mathbf{s})$，其中 $|\mathbf{H}(\mathbf{s})| = 1$；随后在全局网格上生成的三角形 T_i 上对其积分，并求取该随机偏微分方程的分片常数解。对于所有三角形 T_i，有如下分布相等关系：

$$\left[\int_{T_i}\frac{1}{\rho(\mathbf{s})^2} - \nabla\cdot\mathbf{H}(\mathbf{s})\nabla\right]\epsilon(\mathbf{s})\mathrm{d}V \overset{d}{=} \int_{T_i}\frac{1}{\rho(\mathbf{s})}\mathcal{W}(\mathbf{s})\mathrm{d}V. \tag{6}$$

此处 $\nabla\cdot$是 散度算子， ∇是梯度算子，而$\mathbf{H}(\cdot)$是一个 2×2分段 ise

连续可微的扩散张量，dV 是三角形上的曲面测度。这使得我们能够将随机偏微分方程转化为一组线性方程，用于描述一个假定在每个三角形上均为常量的高斯向量。

与 Bertolazzi 和 Manzini（2007年）；Fuglstad 和 Castruccio（2020）类似，令 $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ 表示三角形中心处的值向量，则可计算出以下 $n \times n$ 矩阵 $\mathbf{A_H}$，以描述离散近似：

$$\left(\sum_{j=1}^{3} \int_{\sigma_{i,j}} (\mathbf{H}(\mathbf{s}) \nabla \epsilon(\mathbf{s}))^\top n_{i,j} \mathrm{d}\mathbf{s} \right)_{i=1}^{n} \approx \mathbf{A_H} \boldsymbol{\epsilon}.$$

这里，$\sigma_{i,j}$ 表示三角形 T_i 的三条边。然后，我们将其与一个 $n \times n$对角矩阵 $\mathbf{D}$结合，其中 $d_{ii} = |T_i|/\rho(x_i)^2$, ，从而得到：

$$\left(\int_{T_i} \frac{\epsilon(\mathbf{s})}{\rho(\mathbf{s})^2} \mathrm{d}\mathbf{s} - \sum_{j=1}^{3} \int_{\sigma_{i,j}} (\mathbf{H}(\mathbf{s}) \nabla u(s))^T \mathbf{n}_{i,j} \mathrm{d}\mathbf{s} \right)_{i=1}^{n} \approx (\mathbf{D} - \mathbf{A_H}) \boldsymbol{\epsilon}.$$

通过这一近似，方程（6）中表达的分布相等现在可以表示为：

$$(\mathbf{D} - \mathbf{A_H}) \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{L}),$$

其中，$\mathbf{L}$ 是一个元素为 $l_{ii} = |T_i|/\rho(x)_i^2$ 的 $n \times n$ 对角矩阵。这意味着$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$，而$\mathbf{Q}$ 是定义如下的稀疏精度矩阵：

$$\mathbf{Q} = (\mathbf{D} - \mathbf{A_H})^\top \mathbf{L}^{-1} (\mathbf{D} - \mathbf{A_H}).$$

因此，有限体积法能够确保得到稀疏精度矩阵，从而减轻处理大型全球数据的计算负担，并提高非平稳模型在推断过程中的计算速度。

4.3 潜在高斯模型的推断

为了对潜在高斯模型进行推断，本文采用嵌套拉普拉斯近似（INLA，Rue 等人（2009））。这是一种替代传统马尔可夫链蒙特卡罗（MCMC）的贝叶斯推断方法，可进一步减轻计算负担。INLA 是一种确定性方法，可快速近似高维积分，并利用可表示为潜在高斯马尔可夫随机场的模型所具有的计算特性。因此，本研究采用 INLA 方法进行推断。在所提出的潜在高斯模型结构下，观测数据向量记为 $\boldsymbol{Y} = (Y(\mathbf{s}_1) \ldots, Y(\mathbf{s}_n))^\top$，其对应位置为 $\mathbf{s}_i$，可由超参数向量 $\boldsymbol{\theta}_{\text{space}}$ 描述。为简便起见，本节中我们始终用 $\boldsymbol{\theta}$ 表示超参数向量 $\boldsymbol{\theta}_{\text{space}}$。若以潜在空间场$\mathbf{X}$为条件，则观测数据在边际上相互独立，其似然为:

$$\pi(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(Y(\mathbf{s}_i)|X(\mathbf{s}_i), \boldsymbol{\theta}),$$

其中 $\boldsymbol{X} = (X(\mathbf{s}_1) \ldots, X(\mathbf{s}_n))^\top$ 是均值为零的高斯场，并通过随机偏微分方程方法建模，其精度矩阵为$\mathbf{Q}(\boldsymbol{\theta})$。因此，潜在效应与

超参数可以写为:

$$\begin{aligned}\pi(\mathbf{X},\boldsymbol{\theta}|\mathbf{Y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{X}|\boldsymbol{\theta})\prod_{i=1}^{n}\pi(Y(\mathbf{s}_i)|X(\mathbf{s}_i),\boldsymbol{\theta}))\\ &\propto \pi(\boldsymbol{\theta})|Q(\boldsymbol{\theta})|^{1/2}\exp\{-\tfrac{1}{2}\mathbf{X}^{\top}Q(\boldsymbol{\theta})\mathbf{X}\}\prod_{i=1}^{n}\pi(Y(\mathbf{s}_i)|X(\mathbf{s}_i),\boldsymbol{\theta}),\end{aligned}$$

其中 $|\boldsymbol{Q}(\boldsymbol{\theta})|$ 是精度矩阵的行列式。主要目标是近似后验边际分布 $\pi(X(\mathbf{s}_i)|\mathbf{Y})$、 $\pi(\boldsymbol{\theta}|\mathbf{Y})$和 $\pi(\theta_j|\mathbf{Y})$。我们感兴趣的边际后验分布可写为:

$$\begin{aligned}\pi(X(\mathbf{s}_i)|\boldsymbol{Y}) &= \int \pi(X(\mathbf{s}_i)|\boldsymbol{\theta},\boldsymbol{Y})\pi(\boldsymbol{\theta}|\boldsymbol{Y})\mathrm{d}\boldsymbol{\theta}\\ \pi(\theta_j|\boldsymbol{Y}) &= \int \pi(\boldsymbol{\theta}|\boldsymbol{Y})\mathrm{d}\theta_{-j}.\end{aligned}$$

INLA 方法的核心思想是利用上述形式构造嵌套近似。通过近似 $\pi(\boldsymbol{\theta}|\mathbf{Y})$ 和 $\pi(X(\mathbf{s}_i)|\boldsymbol{\theta},\mathbf{Y})$，并使用数值积分消去 $\boldsymbol{\theta}$，即可计算潜在场 $\pi(X(\mathbf{s}_i)|\mathbf{Y})$ 的边际分布。换句话说，潜在参数的后验边际分布可通过以下方式得到:

$$\tilde{\pi}(X(\mathbf{s}_i)|\boldsymbol{Y}) = \sum_k \tilde{\pi}(X(\mathbf{s}_i)|\boldsymbol{\theta}_k,\boldsymbol{y})\times\tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{Y})\times\Delta_k,$$

where Δ_k 是网格中与超参数向量 $\boldsymbol{\theta}_k$ 相关联的权重 .

5 模拟研究

在本节中，我们用 NS-LS 表示所提出的、包含陆海效应的非平稳潜在高斯模型 (4)，用 NS 表示不含陆海效应的非平稳模型。我们还考虑平稳 SPDE 模型 (3)，并用 S-LS 表示包含陆海效应的模型，用 S 表示不含陆海效应的模型。在第 5.1 节中，我们从 NS-LS 的高斯边缘分布中进行模拟，以数值评估超参数及其所产生的协方差矩阵的后验一致性。在第 5.2 节和第 5.3 节中，我们分别从采用恒等链接和 logit 链接的高斯边缘分布与伯努利边缘分布中进行模拟，以评估 NS-LS 相对于 NS、S-LS 和 S 的插值（克里金法）性能。

由于本文的核心贡献在于模型的空间分量，因此本节中我们将假设这是一个不含协变量的纯空间过程。换言之，模型（1）可简化为

$$Y(\mathbf{s}) \sim h(\mu(\mathbf{s}), \boldsymbol{\theta}_{\mathrm{MRG}}), \tag{7a}$$

$$g(\mu(\mathbf{s})) = \epsilon(\mathbf{s}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\mathrm{space}})). \tag{7b}$$

在高斯情形下，我们也有 $\boldsymbol{\theta}_{\mathrm{MRG}} = \sigma^2 = 0.05$；而在伯努利情形下，没有定义边缘参数，因此 $\boldsymbol{\theta}_{\mathrm{MRG}} = \emptyset$。

对于每次模拟，我们在单位球面上抽取 $n = 2{,}000$ 个数据点，然后从均值为 1、标准差为 0.5 的正态分布中抽取 $\boldsymbol{\theta}_{\text{space}}$ 的参数，并假定这些参数固定（在其他样本或分布下也观察到了类似结果）。每次模拟都包含由所得协方差矩阵 $\boldsymbol{\Sigma}(\boldsymbol{\theta}_{\text{space}})$ 生成的 $n_r = 100$ 个重复样本。我们从具有 $\mathcal{L} = 1$ 的 NS-LS 模型中模拟数据，因此总共有 $6(\mathcal{L}^2 + \mathcal{L}) + 3 = 21$ 超参数。我们进行 $n_s = 100$ 次独立模拟，并从汇总性能及其不确定性两方面报告结果。

5.1 高斯情形下的后验一致性

为了通过数值方法评估后验一致性，在每次模拟中，我们考虑逐渐增加的重复样本数 $n_r = 10, \ldots, 100$。推断时假设采用相同的模型（7），并使用包含 $n_T = 2,000$ 个三角形的网格。对于不同的 n_r 水平，我们获得超参数的后验分布，并将其与真实值进行比较。随着 n_r 增加，超参数的后验分布趋近于真实参数 $\boldsymbol{\theta}_{\text{space}}$，由此可在经验上验证后验一致性。

表1：所有高斯情形模拟中，真实超参数与后验分布之间的 MSE 中位数（四分位距） n_s。

n_r	20	40	60	80	100
MSE 中位数（四分位距）	0.32 (0.13)	0.25 (0.07)	0.14 (0.05)	0.05 (0.05)	0.01 (0.007)

图2展示了在重复样本数 n_r 不断增加的情况下，两个超参数的所有 n_s 个后验分布的函数型箱线图（Sun 和 Genton，2011年）。显然，后验均值与真实参数值保持一致，并且后验

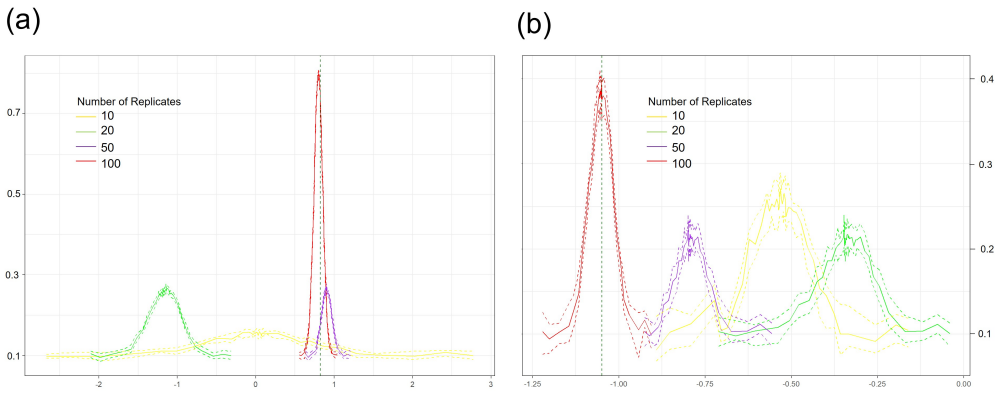

图 2：在 n_s 次模拟中，针对不同重复样本数n_r，两个超参数的后验分布的函数箱线图（Sun 和 Genton，2011年）：(a) α_{11}^2 和 (b) $E_{10}^{(2,2)}$ 。竖直虚线表示超参数的真实值。

标准差随着重复样本数的增加而减小。虽然这里展示的是 NS-LS 的结果，但在其他所有模型（NS、S-LS 和 S）中也观察到了类似的模式。表1展示了从 NS-LS 模型估计得到的超参数后验均值与真实值在所有超参数及所有 $n_s = 100$ 次模拟中的均方误差中位数和四分位距（IQR）。均方误差中位数随着重复样本数的增加而减小。

为对所有超参数进行统一比较——其数量很快就会变得难以处理（例如，使用 $\mathcal{L} = 4$ 时会有 6($4^2 + 4$) 个 $+ 3 = 123$ 超参数）——我们还比较超参数所隐含的协方差矩阵与真实协方差矩阵。我们通过 Kullback-Leibler散度（KLD）评估协方差之间的差异；对于均值为 $\boldsymbol{\mu}_0$ 、 $\boldsymbol{\mu}_1$ 阶的高斯分布，

而协方差矩阵 $\boldsymbol{\Sigma}_0$ 和 $\boldsymbol{\Sigma}_1$ 可简化为:

$$\frac{1}{2}\left(\mathrm{tr}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0) - n + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) + \ln\left(\frac{\det\boldsymbol{\Sigma}_1}{\det\boldsymbol{\Sigma}_0}\right)\right).$$

ˆ 在我们的例子中， $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = \mathbf{0}$、 $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\text{space}})$ 和 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\text{space}})$ ，因此 KLD 衡量真实协方差与估计协方差之间的距离。图3展示了 NS-LS（面板 (a)）和 S（面板 (b) ）的结果，其中给出了随着实现值数量 n_r 增加，在所有 $n_s = 100$ 次模拟中 KLD 的函数型箱线图（Sun 和 Genton，2011年）。函数型箱线图用于展示 50% 中心区域的包络（粉色区域）、中位数曲线（黑线）和最大非离群包络（外侧蓝线）。与参数估计的情况相同，我们观察到，即使训练集中的重复样本数量相对较少，估计的协方差也会逐渐趋近真实协方差。具体而言，重复样本达到 40 个后，估计的协方差几乎与真实协方差无法区分。

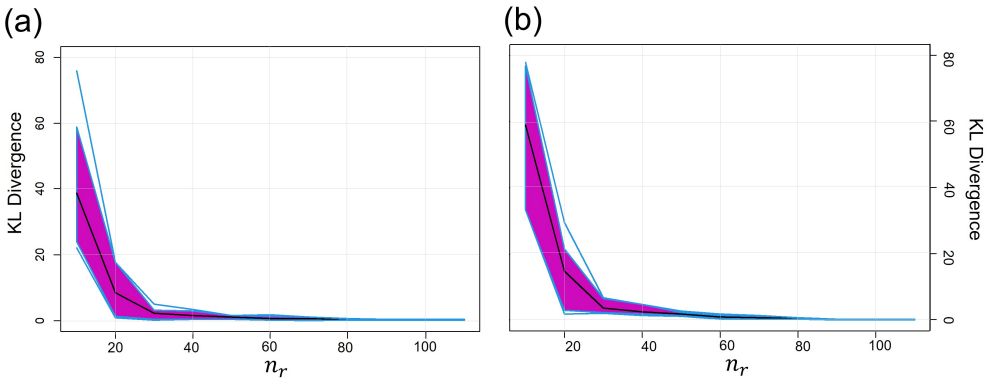

图3：在 $n_s = 100$ 次模拟中，依据 (a) NS-LS 和 (b) S-LS 得到的真实协方差矩阵与估计协方差矩阵之间 KLD 的函数型箱线图。

5.2 高斯情形下的插值性能

为了评估插值性能，我们对所有四种模型的超参数进行推断，并使用这些参数在指定位置进行插值。我们考虑两种情形：（1）将所有 n 数据点用于训练集，并在相同位置进行插值；（2）将 92% 的 n位置纳入训练集，其余 8% 留出用于交叉验证。测试位置位于图 S1 所示的三个选定区域内。插值性能通过均方误差（MSE）衡量。

两种情形的结果均列于表 2。显而易见，在所有位置情形 (1) 和交叉验证设置 (2) 下，NS-LS 模型的均方误差均为四种模型中最小的。具体而言，与 S-LS 模型相比，NS-LS 模型在所有位置上的 MSE 中位数降低了 14.6%。与 NS 模型和 S 模型相比，NS-LS 模型的均方误差也分别降低了 10.4% 和 16.7%，改进相当显著。这些结果清楚地表明，将陆海效应和缓冲区构建与 NS 模型结合使用，能够带来显著改进。

表 2：各模型插值性能比较。前两列显示高斯情形下所有 $n_s = 100$ 次模拟的 MSE中位数（四分位距），分别对应（1）所有位置和（2）交叉验证。后两列显示伯努利情形下相同两种情形的曲线下面积（AUC）中位数（四分位距）。

模型	位置	NS-LS	S-LS	NS	S
高斯分布	所有位置	90.12 (8.17)	105.47 (9.94)	100.55 (11.25)	108.25 (11.23)
	交叉验证	9.11 (1.04)	21.88 (1.18)	19.35 (1.62)	21.39 (1.59)
《伯努利》	所有位置	0.824 (0.048)	0.769 (0.074)	0.782 (0.051)	0.753 (0.050)
	交叉验证	0.707 (0.072)	0.676 (0.081)	0.672 (0.079)	0.641 (0.079)

5.3 伯努利情形下的插值性能

现在，我们评估伯努利分布采用 logit 链接时的可预测性。与第 5.2 节一样，我们既评估将所有位置用作训练集的情形，也评估采用与之前相同测试位置的交叉验证情形。图 4 展示了在所有位置和验证位置上，以 S 模型为参照，NS-LS 与 S-LS 的受试者工作特征曲线（ROC）在所有 $n_s = 100$ 次模拟中的平均差异。NS 模型的 ROC 与 S-LS 模型的 ROC 在视觉上无法区分，因此未展示该模型的相关结果。两种情形下的 ROC 差异均表明，NS-LS 模型整体上优于平稳 S 模型（因为 ROC 差异始终为正），也整体上优于 S-LS 模型，尤其是在曲线的中段。正如预期，在交叉验证情形（面板 (b)）下，NS-LS 的改进幅度更大，因为模型在未观测位置上的增益更为明显。

为了全面评估所有可能阈值选择下的表现，我们考察所有模型的受试者工作特征曲线（ROC）的曲线下面积（AUC），并在表 2 中报告结果。在最佳情况下，即预测完美时，若在所有阈值选择下真阳性率均为 100%，则 AUC 应等于 1；在最差情况下，即随机猜测时，AUC 应为 0.5。在此情形下，AUC 越接近 1，预测表现就越好。如表 2 所示，NS-LS 在两种情况下均优于其他所有模型。更具体地说，在所有位置上，NS-LS 相较于 S-LS、NS 和 S 模型分别提升了 7.2%、5.3% 和 9.4%。

这些结果与第 5.2 节中的结果一致，因为在 NS 模型中加入陆海效应和缓冲区构建，确实能够提升模型表现。

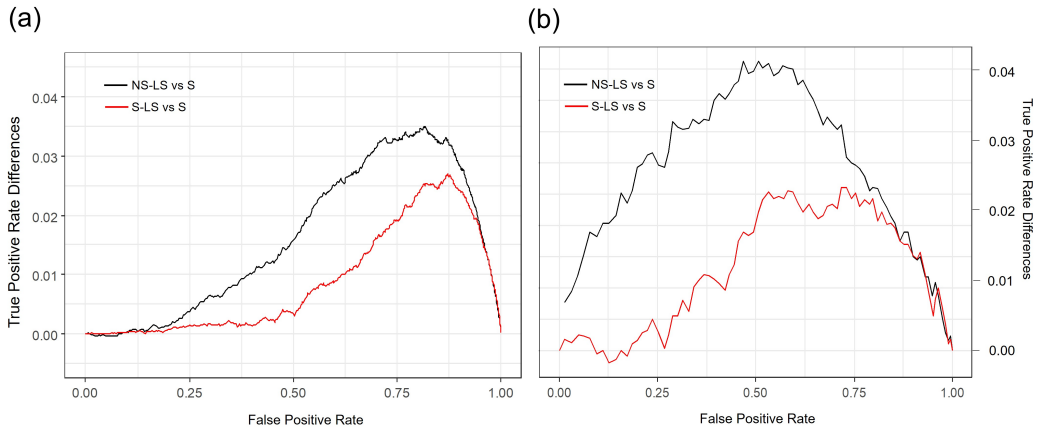

图 4：在 (a) 所有位置和 (b) 交叉验证中，NS-LS 与 S 模型（黑线）以及 S-LS 与 S 模型（红线）的 ROC曲线在全部 $n_s = 100$ 次模拟中的平均差异。NS 的 ROC曲线与 S-LS 模型的 ROC曲线在视觉上无法区分，因此未展示与该模型相关的结果。

6 应用

本节使用第2节详述的数据，以及第3节介绍的非平稳SPDE潜在高斯模型，来估计全球降雨事件发生的概率和降水强度。在第6.1节中，我们讨论对全球 MERRA-2 数据集的拟合，以及利用地面 USCRN 降水测量校正插值后的 MERRA-2 数据的降尺度方法。在第6.2节中，我们提供评估指标，以评估模型性能。

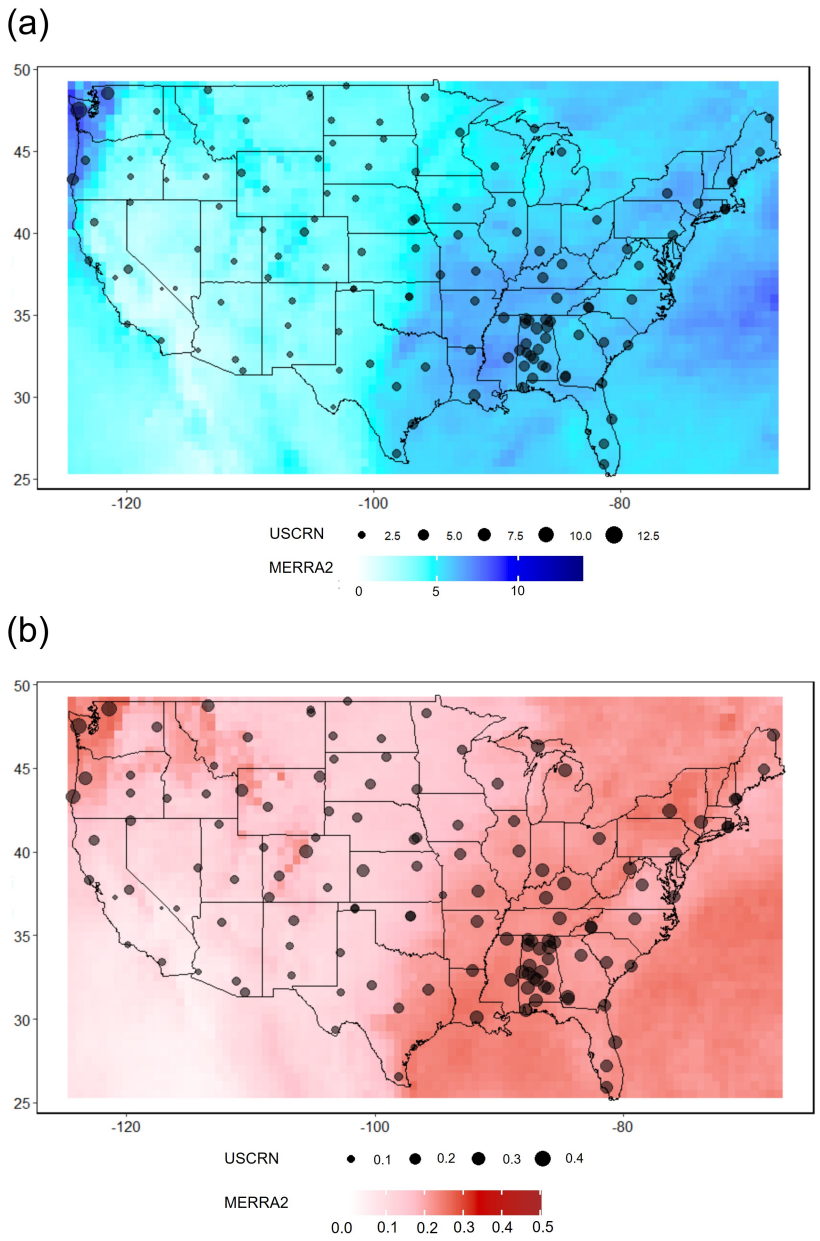

图 5：平均（a）日降水量和（b）降水概率。根据非平稳全球 SPDE 模型 (4)，在与地面观测相同的地点对全球数据集进行插值，拟合线性模型 (9)，并利用所得关系生成降尺度地图。

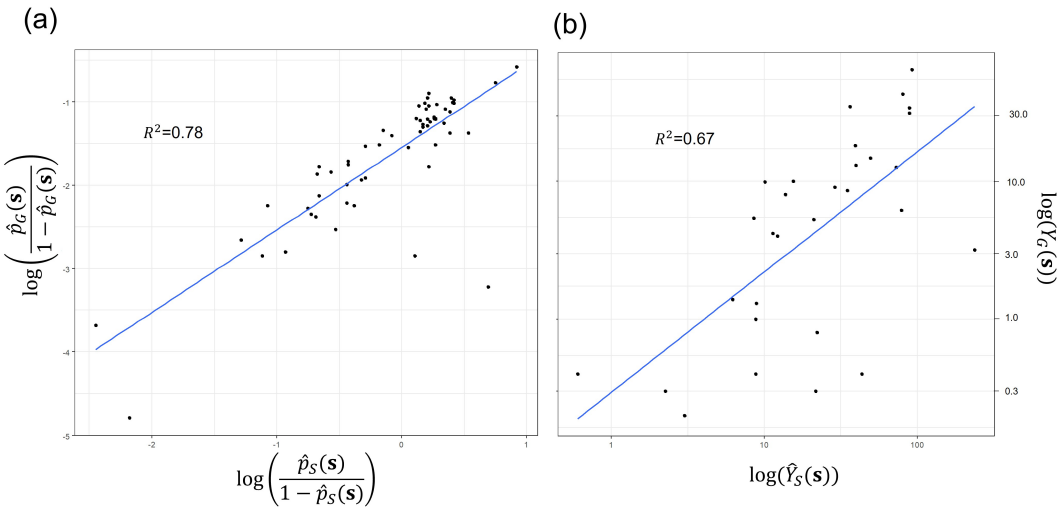

图6：2021年2月 1st，使用(a)方程(9a)和(b)方程(9b)中所述的降尺度模型拟合得到的曲线。

6.1 全球降水建模与降尺度

我们首先聚焦于MERRA-2数据，并考虑两组全球数据：1）二元降雨发生事件；2）发生降雨时的实际降雨强度。接着，我们使用 $\mathcal{L}=1$，将非平稳偏微分方程(4)与潜在高斯模型(1)拟合；对于降雨发生事件，采用伯努利边缘分布和logit链接函数 $g(\cdot)$；对于降雨强度，则采用Gamma分布和负逆链接函数。关于边缘分布选择的验证见补充材料，其中图S3展示了456个样本位置处的降水直方图（经度和纬度分辨率为 $18.75^\circ \times 15^\circ$），并绘出了估计的Gamma密度。样本位置在空间上较为稀疏，以减轻任何空间影响。

在这两种情况下，均不假设存在额外的协变量，并假设时间分量采用 $K=2$ 个谐波

，因为图 S2 中的模型选择结果表明这是最优选择。形式上，模型 (1) 现在具体化为以下两个模型:

$$\log\left(\frac{\mu(\mathbf{s},t)}{1-\mu(\mathbf{s},t)}\right) = f^{\text{time}}(\mathbf{s},t) + \epsilon(\mathbf{s}), \quad \text{precipitation probability} \tag{8a}$$

$$-\mu(\mathbf{s},t)^{-1} = f^{\text{time}}(\mathbf{s},t) + \epsilon(\mathbf{s}). \quad \text{precipitation intensity} \tag{8b}$$

直方图显示，降水强度服从 Gamma分布，其形状参数为 0.826，尺度参数为 0.184。推断基于包含 $n_T = 2$、340 个三角形的全局三角剖分进行，其中 1、134 个位于感兴趣区域（美国本土连续部分），其余 1、206 个覆盖世界其他地区。

我们得到超参数的后验分布，并据此预测 USCRN 的 131 个地面观测位置处的降水概率和降水强度，见图1。随后通过线性回归将这些预测调整（降尺度）至点分辨率。由于我们对每个时间点分别进行降尺度，为简便起见，以下省略时间依赖性，并分别用 $Y_G(\mathbf{s})$ 和 $Y_S(\mathbf{s})$ 表示 USCRN 和 MERRA2 的降水强度（G=地面观测，S=模拟），用 $p_G(\mathbf{s})$ 和 $p_S(\mathbf{s})$ 表示降水发生概率。我们还分别用 $Y_S(\mathbf{s})$ 和 $\hat{}p_S(\mathbf{s})$ 表示根据所提出的 SPDE 模型估计的强度和发生概率。最后，我们估计

通过针对每个位置独立拟合潜在高斯模型（1）作为时间序列模型，估计 USCRN 数据的降水发生概率，即假设不存在空间依赖，并将该估计记为 $\hat{}p_G(\mathbf{s})$。我们进一步假设 USCRN 与 MERRA2 的降水发生概率和降水强度之间存在线性关系：

$$\log\left(\frac{\hat{}p_G(\mathbf{s})}{1-\hat{}p_G(\mathbf{s})}\right) = \beta_0^{(O)} + \beta_1^{(O)}\log\left(\frac{\hat{}p_S(\mathbf{s})}{1-\hat{}p_S(\mathbf{s})}\right) + \xi_O(\mathbf{s}),\ \text{降水概率 (9a)}$$

$$\log\left(Y_G(\mathbf{s})\right) = \beta_0^{(I)} + \beta_1^{(I)}\log\left(\hat{}Y_S(\mathbf{s})\right) + \xi_I(\mathbf{s}),\ \text{降水强度 (9b)}$$

其中 $\xi_j(\mathbf{s}) \sim \mathcal{N}(0,\ \sigma_j^2)$ $j \in \{O, I\}$ 在空间上独立同分布。图 S4 中残差的变异函数函数型箱线图（每条曲线代表一个不同的时间点）支持误差具有空间独立性的假设。随后使用普通最小二乘法估计降水强度的降尺度参数 $\beta_0^{(I)}$ 和 $\beta_1^{(I)}$ 。

6.2 结果与评估

根据上述模型得到的降水发生概率和降水强度的降尺度结果分别显示在图5(a)和(b)中，深色气泡表示 USCRN 数据的平均值。美国的预测图显示，西雅图附近的日降水量和降水强度较高，而拉斯维加斯附近的数值最低；总体而言，模型预测与美国各地的地面观测值相

符。为评估模型性能，我们分别计算降水发生概率和降水强度的均方根误差（RMSE）。强度和降水发生概率的 RMSE 分别为 2.01 mm 和 0.14 mm。为了评估 SPDE 模型平滑处理所带来的价值，我们还使用式（9）中的线性模型进行降尺度，但假设不拟合空间模型，即不将 MERRA-2 数据插值到 USCRN 站点的位置。相反，我们采用原始分辨率的 MERRA-2 数据，并将每个 USCRN 站点所在网格单元中的数值赋予该站点。换言之，我们将 $p_S(\mathbf{s}\ t)$ 和 $Y_S(\mathbf{s}\ t)$ 作为协变量。对于降水强度和降水发生概率，该模型的 RMSE 分别为 82.74 mm 和 0.28 mm。因此，所提出的随机偏微分方程方法显著缩小了 MERRA-2 与 USCRN 之间的差异，使降水强度和降水发生概率的 RMSE 分别降低了 97.6% 和 50%。图6展示了 2021 年 2 月 1st 使用式（9a）和（9b）中的降尺度模型拟合得到的直线。两个线性模型的 R^2 分别为 0.78 和 0.67，对应降水概率和降水强度。

我们还通过交叉验证评估模型的不确定性。首先，移除一个地面观测位置的数据，并使用其余观测数据拟合模型。接着，利用模型超参数的估计后验分布，构造被移除位置处降水发生概率或降水强度后验均值的 95% 可信区间。然后，我们重复

相同的流程，对 USCRN 中全部 131 个位置逐一进行验证。最后，我们统计这 131 个 95 % 可信区间中有多少覆盖了真实值。对于降水，93.1%（122/131）的 95% 可信区间覆盖了真实值；对于降雨概率，91.6%（120/131）的 95% 可信区间覆盖了真实值。

7 结论与讨论

在这项工作中，我们提出了一种新颖的非平稳时空SPDE模型，能够对全球数据集中的降水发生概率和概率强度进行平滑。随后将这一插值数据集与地面观测相结合，生成高分辨率（经降尺度）的降水图，从而能够以比原始模拟数据（即原生分辨率下的全球数据）更高的准确度，预测未采样位置的地面观测结果。原则上，可以将MERRA-2用作边界条件，驱动采用WRF等模型的区域模拟，以获得同样具有高空间分辨率的降水图；此外，这种方法还能生成符合物理定律的预测结果。然而，这种动力降尺度方法的实施要复杂得多，因为它需要大量的计算与存储资源，以及相当丰富的专业知识才能正确配置WRF。因此，我们提出的统计降尺度 方法速度快得多，也更易于实施，无需专门的计算资源。通过地面观测调整模拟结果的方法，也可以视为一种偏差

校正方法，即一种校正模拟结果的方法（例如，参见 Yuan 等（2019）；Kim 等（2015）和 Ho 等（2012）；Hawkins 等（2013）的综述）。地球科学领域的大量文献将偏差校正作为调整边缘分布一阶矩（Hemer 等，2012；Chen 等，2012），并可能调整二阶矩（Teutschbein 和 Seibert，2012；Li 等，2019）的一种手段；然而，与近期其他研究（Piani 和 Haerter，2012；Vrac 和 Friederichs，2014）类似，这种方法也可用于调整非高斯特征。

由于采用了用于生成空间模型的随机偏微分方程有限体积近似，所提出的统计模型能够扩展到未来空间分辨率更高的再分析数据产品。如果能将温度和湿度等其他物理变量作为协变量纳入考虑，就可以生成更逼真的降尺度模式。正如本研究所提出的，协变量既可以作为式（1b）中的潜在高斯模型纳入，也可以作为额外输入加入决定随机偏微分方程模型形变的标量场或向量场。具体实现时，可以假设协变量产生线性贡献，也可以借助神经网络实现非线性贡献（Hu et al., 2022）。原则上，可以对多个变量进行联合建模。不过，这会大幅增加方法上的难度和计算开销，因为快速、灵活、多变量且非高斯的全球模型目前仍是一个活跃的研究领域（Genton and Kleiber, 2015）。

致谢

本研究得到美国国家科学基金会 DMS 2014166 号资助。

参考文献

Ahmed Arafat、Pablo Gregori 和 Emilio Porcu。球面上连续各向同性正定核在原点处的舍恩伯格系数与曲率。《统计与概率快报》，156：108618，2020年。ISSN 0167-7152。doi：https://doi.org/10.1016/j. spl.2019.108618。

Haakon Bakka、Jarno Vanhatalo、Janine B. Illian、Daniel Simpson 和 H˚avard Rue。具有物理屏障的非平稳高斯分布模型。《空间统计》，29：268–288，2019。ISSN 2211-6753。doi：https://doi.org/10.1016/j.spasta.2019.01.002。URL https://www. sciencedirect.com/science/article/pii/S221167531830099X。

V.J. Berrocal、A.E. Gelfand 和 D.M. Holland。用于数值模型输出的时空降尺度器。《农业、生物与环境科学杂志》，15：176–197，2010年。

Enrico Bertolazzi 和 Gianmarco Manzini。关于二维各向异性扩散问题中单元中心有限体积近似的顶点重构。《应用科学中的数学模型与方法》，17(01):1–32，2007年。doi: 10.1142/S0218202507001814。URL https://doi.org/10.1142/S0218202507001814。

David Bolin 和 Finn Lindgren。由嵌套随机偏微分方程生成的空间模型及其在全球臭氧制图中的应用。《应用统计年鉴》，5(1):523 – 550，2011年。doi: 10.1214/10-AOAS383。URL https://doi.org/10.1214/10-AOAS383。

Stefano Castruccio 和 Marc G. Genton。《超越轴对称性：一类改进的全球数据模型》。*Stat*，3(1)：48–55，2014。

Stefano Castruccio 和 Marc G. Genton。《利用统计模型压缩气候集合：一种用于全球三维时空温度的算法》。*Technometrics*，58(3)：319–328，2016。

Stefano Castruccio 和 Joseph Guinness。《一种将大尺度地理描述符纳入全球过程的演化谱方法》。《皇家统计学会杂志》。*C* 辑（应用统计），66(2)：329–344，2017。ISSN 00359254，14679876。URL http://www.jstor.org/stable/44682577。

Stefano Castruccio 和 Michael L. Stein。《气候集合的全球时空模型》。《应用统计年鉴》，7(3)：1593 – 1611，2013。doi：10.1214/13-AOAS656。URL https://doi.org/10.1214/13-AOAS656。

Lian Chen、S. C. Pryor 和 Dongliang Li。《评估政府间气候变化专门委员会 AR5气候模型在模拟和预测中国风速方面的表现》。《地球物理研究杂志：大气》，117:D24102，2012 年。doi:

10.1029/2012JD017533。网址：https://agupubs.onlinelibrary.wiley.com/doi/abs/ 10.1029/2012JD017533。

Martyn P Clark、Ying Fan、David M Lawrence、Jennifer C Adam、Diogo Bolster、David J Gochis、Richard P Hooper、Mukesh Kumar、L Ruby Leung、D Scott Mackay 等。改进地球系统模型中水文过程的表示。《水资源研究》，51(8):5929–5956，2015年。

Peter K. Dunn 和 Gordon K. Smyth。《广义线性模型及其 R 语言示例》。施普林格纽约，纽约，第一版，2018年。

M. Edwards、S. Castruccio 和 D. Hammerling。用于气候集合的多变量全球时空随机生成器。《农业、生物与环境科学杂志》，24(3):464–483，2019年。

M. Edwards、S. Castruccio 和 D. Hammerling。边际参数化时空模型与逐步极大似然估计。《计算统计与数据分析》，151:107018，2020年。

Geir-Arne Fuglstad 和 Stefano Castruccio。使用非平稳全球时空SPDE模型压缩气候模拟。《应用统计年鉴》，14(2):542 – 559，2020年。doi: 10.1214/20-AOAS1340。URL https://doi.org/ 20-AOAS1340。

Geir-Arne F Uglstad、Finn Lindgren、Daniel Simpson 和 H˚avard Rue。探索一类新的

具有变化的局部各向异性的非平稳空间高斯随机场。《中华统计学志》，25(1):115–133，2015。ISSN 10170405，19968507。网址：http://www.jstor.org/stable/24311007。

Geir-Arne Fuglstad、Daniel Simpson、Finn Lindgren 和 H˚avard Rue。构建能够惩罚高斯随机场复杂度的先验。《美国统计协会杂志》，114(525):445–452，2019。doi: 10.1080/01621459.2017.1415907。URL https://doi.org/10.1080/01621459.2017.1415907。

Ronald Gelaro、Will McCarty、Max J. Su´arez、Ricardo Todling、Andrea Molod、Lawrence Takacs、Cynthia A. Randles、Anton Darmenov、Michael G. Bosilovich、Rolf Reichle、Krzysztof Wargan、Lawrence Coy、Richard Cullather、Clara Draper、Santha Akella、Vir- ginie Buchard、Austin Conaty、Arlindo M. da Silva、Wei Gu、Gi-Kong Kim、Randal Koster、Robert Lucchesi、Dagmar Merkova、Jo n Eric Nielsen、Gary Partyka、Steven Paw- son、William Putman、Michele Rienecker、Siegfried D. Schubert、Meta Sienkiewicz 和 Bin Zhao。现代回顾性分析与研究应用（MERRA-2）第 2 版。《气候杂志》，30(14):5419–5454，2017 年 06 月。ISSN 0894-8755。doi: 10.1175/JCLI-D-16-0758.1。URL https://doi.org/10.1175/JCLI-D-16-0758.1。

Marc G. Genton 和 William Kleiber。多元地理统计学的交叉协方差函数。*Statistical Science*，30(2)：147–163，2015 年。doi：10.1214/14-STS487。

Tilmann Gneiting. 球面上的严格与非严格正定函数。《伯努利》，

19(4):1327 – 1349，2013 年。doi：10.3150/12-BEJSP06。网址：
https://doi.org/10.3150/ 12-BEJSP06。

Ed Hawkins、Thomas M. Osborne、Chun Kit Ho 和 Andrew J. Challinor。用于作物建模的气候预测校准与偏差校正：欧洲地区的理想化案例研究。《农业与森林气象学》，170:19–31，2013年。ISSN 0168-1923。doi：
https://doi.org/10.1016/j.agrformet.2012.04.007。利用气候模型集合进行农业预测。

Mark Hemer、Kathleen McInnes 和 Roshanka Ranasinghe。用于澳大利亚东南部动力波浪模型的气候及变率偏差调整：对气候模型推导的风进行处理。《海洋动力学》，62(1):87–104，2012年。ISSN 16167341。URL
http://search.ebscohost.com.proxy.library.nd.edu/login.aspx? direct=true&db=aph&AN=70162084&site=ehost-live。

Chun Kit Ho、David B. Stephenson、Matthew Collins、Christopher A. T. Ferro 和 Simon J. Brown。校准策略：气候变化预测中不确定性的额外来源。《美国气象学会通报》，93(1):21–26，2012年01月。doi：10.1175/2011BAMS3110.1。

Wenjing Hu、Geir-Arne Fuglstad 和 Stefano Castruccio。采用基于神经网络的形变的全球海表温度随机局部扩散模型。*Stat*，11(1):e431，2022。doi：
https://doi.org/10.1002/sta4.431。

Jaehong Jeong、Mikyoung Jun 和 Marc G. Genton。用于全球空间统计的球面过程模型。*Statistical Science*，32(4):501 – 513，2017。doi：10.1214/17-STS620。URL：https://doi.org/10.1214/17-STS620。

R.H. Jones. 球面上的随机过程。*Annals of Mathematical Statistics*，34:213– 218，1963年。

M. Jun. 地球上多元过程的非平稳互协方差模型。*Scandinavian Journal of Statistics*，38(4):726–747，2011年。

Mikyoung Jun 和 Michael L Stein. 一种在球面上构造时空协方差函数的方法。*Technometrics*，49(4):468–479，2007年。doi: 10.1198/004017007000000155。URL https://doi.org/10.1198/004017007000000155。

Mikyoung Jun 和 Michael L. Stein. 全球数据的非平稳协方差模型。《应用统计年鉴》，2(4):1271 – 1289，2008年。doi: 10.1214/08-AOAS183。URL https://doi.org/10.1214/08-AOAS183。

Kue Bum Kim、Hyun-Han Kwon 和 Dawei Han。考虑观测数据背后分布参数不确定性的区域气候模型模拟偏差校正方法。*Journal of Hydrology*，530:568–579，2015。ISSN 0022-1694。doi：https://doi.org/10.1016/j.jhydrol.2015.10.015。URL http://www.sciencedirect.com/ science/article/pii/S002216941500774X。

Delei Li，Jianlong Feng、Zhenhua Xu、Baoshu Yin、Hongyuan Shi 和 Jifeng Qi。统计 ti-

对 CORDEX东亚模拟风速进行偏差校正。《地球与空间科学》，6(2):200–211，2019年。doi：10.1029/2018EA000493。URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EA000493。

Finn Lindgren、H˚avard Rue 和 Johan Lindstr¨om。《高斯场与高斯马尔可夫随机场之间的显式联系：随机偏微分方程方法》。《皇家统计学会会刊：*B*辑（统计方法学）》，73 (4):423–498，2011年。doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x。URL https: //rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x。

美国国家海洋和大气管理局。美国气候参考网络。https://www.ncei.noaa.gov/access/crn/，2022年。最后访问日期：2022/10/25。

C. Piani 和 J. O. Haerter。气候模型中温度与降水联结函数的二维偏差校正。《地球物理研究快报》，39:L20401，2012年。doi: 10.1029/2012GL053839。

Emilio Porcu、Alfredo Alegria 和 Reinhard Furrer。对随时间演变且在全球空间范围内相互依赖的数据进行建模。《国际统计评论》，86(2):344–377，2018。Emilio Porcu、Rachid Senoussi、Enner Mendoza 和 Moreno Bevilacqua。球面上非平稳协方差函数的约化问题与形变方法。《电子统计学杂志》，14(1):890 – 916，2020。doi: 10.1214/19-EJS1670。URL https://doi.org/10.1214/19-EJS1670。

H˚avard Rue、Sara Martino 和 Nicolas Chopin。通过使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。《皇家统计学会会刊：*B*辑（统计方法学）》，71(2):319–392，2009。doi：https:// doi.org/10.1111/j.1467-9868.2008.00700.x。URL https://rss.onlinelibrary.wiley. com/doi/abs/10.1111/j.1467-9868.2008.00700.x。

Stephan R. Sain、Doug Nychka 和 Linda Mearns。函数方差分析与区域气候实验：动态降尺度的统计分析。环境计量学，22(6):700– 711，2011年。

M Sapountzis、AKastridis、AP Kazamias、A Karagiannidis、P Nikopoulos 和 KLagouvar- dos。卫星降水数据在无水文站流域山洪水文分析中的应用及不确定性。*Glob. Nest J*，23:388–399，2021年。

W. C. Skamarock、J. B. Klemp、J. Dudhia、D. O. Gill、Z. Liu、J. Berner、J. G. Wang、Powers、M. G. Duda、D. M. Barker 和 Huang X.-Y。《高级研究WRF 4.0版说明》。技术报告，美国国家大气研究中心技术说明 NCAR/TN-556+STR，2019年。M.L. Stein。《空间数据插值：克里金法的一些理论》。施普林格，纽约，1999年。

Ying Sun 和 Marc G. Genton。《函数箱线图》。计算与图形统计学杂志，20(2):316–334，2011年。doi: 10.1198/jcgs.2011.09224。URL https: // doi.org/10.1198/jcgs.2011.09224。

Claudia Teutschbein 和 Jan Seibert。区域气候模式模拟的偏差校正 mu-

用于水文气候变化影响研究：不同方法的综述与评估。*Journal of Hydrology*, 456-457:12–29，2012。ISSN 0022-1694。doi: https://doi.org/10.1016/j.jhydrol.2012.05.052。URL http://www.sciencedirect.com/ science/article/pii/S0022169412004556。

Mathieu Vrac 和 Petra Friederichs。多变量——变量间、空间和时间——偏差校正。《气候杂志》，28(1):218–237，2014 年 12 月。ISSN 0894-8755。doi: 10.1175/JCLI-D-14-00059.1。URL https://doi.org/10.1175/JCLI-D-14-00059.1。

P. Whittle。平面上的平稳过程。*Biometrika*，41(3-4):434–449，1954 年 12 月。ISSN 0006-3444。doi: 10.1093/biomet/41.3-4.434。URL https://doi.org/10.1093/biomet/41.3-4.434。

Raul R Wood、Flavio Lehner、Angeline G Pendergrass 和 Sarah Schlunegger。多个全球气候模型大型集合中跨时间尺度的降水变率变化。环境研究快报，16(8):084022，2021 年。

Qifen Yuan、Thordis L. Thorarinsdottir、Stein Beldring、Wai Kwok Wong、Shaochun Huang 和 Chong-Yu Xu。针对日平均气温未来预估的高分辨率网格偏差校正与随机降尺度新方法。应用气象学与气候学杂志，58(12):2617–2632，2019年12月。doi: 10.1175/JAMC-D-19-0086.1。URL https://doi.org/10.1175/JAMC-D-19-0086.1。