arXiv:2305.04341v1 [stat.ML] 7 May 2023

使用神经网络快速估计广义极值分布参数

Sweta Rai[1],Alexis Hoffman[2], Soumendra Lahiri[3], Douglas W. Nychka[1]Stephan R. Sain[2], 和 Soutir Bandyopadhyay[1] *

[1] 科罗拉多矿业学院应用数学与统计系，戈尔登，科罗拉多州[2] Jupiter Intelligence，博尔德，科罗拉多州[3] 圣路易斯华盛顿大学，密苏里州圣路易斯

摘要

广义极值分布具有重尾行为，因此常用于模拟洪水、干旱、热浪、野火等极端事件。然而，即使对于中等规模的数据集，使用传统最大似然方法估计该分布的参数也可能需要大量计算。为克服这一局限，我们提出一种利用神经网络的高效无似然估计方法。通过广泛的模拟研究，我们证明，所提出的基于神经网络的方法所得的广义极值分布（GEV）参数估计值，其精度与传统最大似然方法相当，但计算速度显著提升。为考虑估计不确定性，我们采用参数自助法，这一方法内嵌于训练好的网络中。最后，我们将该方法应用于第三版社区气候系统模型（CCSM3）在北美洲模拟的1000年年最大气温数据，涵盖三种大气浓度：289 ppm CO_2（前工业时期）、700 ppm CO_2（未来条件）和1400 ppm CO_2, ，并将结果与使用最大似然方法得到的结果进行比较。

关键词：深度神经网络；广义极值分布；参数估计；充分统计量；极端分位数。

*SR、SB 和 DWN 的研究工作部分获得了美国国家科学基金会资助，项目编号为 CMMI-2210840；SL 的研究工作部分获得了美国国家科学基金会资助，项目编号为 CMMI-2210811。通讯作者电子邮箱：sbandyopadhyay@mines.edu。

1 引言

人们普遍认为，地球目前正经历一段气候变化时期，导致天气模式出现极端变化，并持续扰动周边环境。为了为规划提供依据并减轻这些变化带来的风险，深入了解这些极端事件并量化其潜在影响至关重要（参见 Adger（2003）、Power 等（2021））。极值理论（EVT）广泛应用于计量经济学（Bali（2003））、金融学（Broussard 和 Booth（1998））、材料科学（Castillo（2012））、环境科学（Katz（2010）、Martins 和 Stedinger（2000））以及可靠性工程（Smith（1991））等领域，用于估计极端事件发生的概率。有关 EVT 的详细信息，请参阅 Fisher 和 Tippett（1928年）、Hosking 等（1985年）、Smith（1985、1990）、Bali（2003）、Coles 等（2001年）、Gumbel（2004年）、Davison 和 Huser（2015年）以及其中所引文献。在单变量 EVT 框架下，广义极值分布因其灵活性和坚实的理论基础，常用于极端事件建模。其一项重要特征是形状参数，该参数决定了多种尾部行为，从有界分布到重尾分布均可涵盖，详见 Coles 等（2001年）、Gumbel（2004年）、Davison 和 Huser（2015年）、Haan 和 Ferreira（2006）。

通常使用最大似然（ML）或基于矩的方法来估计广义极值分布参数。这些方法都可能较为复杂且耗时，因为形状参数及其他参数的取值范围受到约束，需要进行数值计算。因此，近期拟合广义极值分布的方法开始探索使用精确似然以外的替代方案（参见 Casson & Coles (1999)、El Adlouni & Ouarda (2009)、Opitz 等人 (2018)、Erhardt & Smith (2012)、Castro-Camilo 等人 (2021)）。其中一种方法是采用贝叶斯方法，例如近似贝叶斯计算（ABC），该方法已用于研究极大稳定过程建模中的依赖性（参见 Blum (2010)、Erhardt & Smith (2012)、Erhardt & Sisson (2016)、Sisson 等人 (2018)）。另一方面，近期已有多种深度学习算法被用于极端事件的无似然推断（参见 Creel (2017)、Lenzi 等人 (2021)、Sainsbury-Dale 等人 (2022)、Richards & Huser (2022) 及其中的参考文献）。与经典方法相比，这些研究在准确性方面展现出良好的前景，同时也有望加快整体估计过程。不过，这些方法的应用大多聚焦于利用极大稳定过程对高维空间极值进行建模。

这项工作受到上述近来将近似贝叶斯计算方法和深度学习算法应用于空间极值的研究，以及将神经网络（NN）用于时间序列和空间数据的研究的启发，相关成果见 Chon & Cohen (1997)、Cremanns & Roos (2017)、Gerber & Nychka (2021)、Wikle & Zammit-Mangion (2022) 和 Majumder et al. (2022) 等论文。在这项工作中，我们提出了一种新的估计方法，利用神经网络拟合极端事件的单变量GEV分布。需要指出的是，对任何多变量极值过程建模，通常都需要识别其边缘GEV分布；因此，本文聚焦于计算高效的单变量GEV分布建模。包括 Creel & Kristensen (2013)、Creel (2017) 和 Gerber & Nychka (2021) 在内的多项研究报告了使用信息性统计量而非完整数据集对极值进行建模。通过同时纳入极端分位数和标准四分位数（Q_1, Q_2, 和 Q_3），我们的神经网络能够有效捕捉极端事件建模中的重要尾部行为。将样本分位数作为网络输入，也得到了次序统计量概念的支持。

训练期间，我们输入样本分位数，并应用激活函数生成最优的非线性表示，再据此估计广义极值分布参数。模型输出广义极值分布参数的估计值，这些参数定义了极端事件的分布。为确保模型稳健，我们使用合理参数范围内的模拟值进行训练，并选取了较大的训练样本量。此外，我们还使用验证集监控模型在训练期间的性能。

使用神经网络解决这一问题具有以下优势。

(a) 神经网络架构非常适合推断，能够高效地估计参数。尤其是，网络在训练完成后可以快速评估，与传统方法相比，速度可提升 100–1000 倍。

(b) 为了应对参数估计值的不确定性问题，我们采用了自助法，这一方法得到了以往研究的广泛支持（例如，Cooley 等人（2007年）、Kysel`y（2008年）、Huang 等（2016年）、Lenzi 等人（2021年）、Gamet 和 Jalbert（2022年）、Yadav 等人（2022年）、Sainsbury-Dale 等人（2022年））。具体而言，为了生成置信区间，我们采用了通常计算成本较高的参数自助法。然而，训练好的神经网络能够高效地模拟和评估自助样本，从而快速生成置信区间。

最后，我们使用模型分析了 CCSM3 模型在北美洲三种不同 CO_2 浓度下的模拟运行中的最高气温极值。通过考察年最高气温数据，我们展示了与经典机器学习方法相比，本方法的准确性和优势。神经网络估计器的快速评估速度使我们能够通过参数自助法抽样高效地进行不确定性量化。研究结果表明，我们可以在数秒内生成数百个空间置信区间。

本文其余部分安排如下。第2节概述广义极值分布，并阐述所提出的神经网络模型。第3节展示模拟研究的结果。第4节深入探讨 CCSM3 模拟运行的案例研究。最后，第5节总结研究发现，考察模型的行为及其局限性，并给出结论。

2 方法

本节概述广义极值分布的结构，并介绍我们的模型框架。此外，还介绍作为网络输入所选用的近似统计量，以及模型所采用的网络架构。

2.1 广义极值分布

由 Jenkinson（1955）提出的广义极值分布，具有位置—尺度参数 $(\mu,\ \sigma) \in \mathcal{R} \times (0,\ \infty)$，以及形状参数或尾指数参数 $\xi \in \mathcal{R}$。其累积分布函数（CDF）表示为

$$\mathcal{F}(x) = \begin{cases} \exp\left\{-\left[1+\xi(x-\mu)/\sigma\right]^{-1/\xi}\right\}, & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left[-(x-\mu)/\sigma\right]\right\}, & \text{if } \xi = 0 \end{cases}$$

$\mathcal{F}$ 的定义域由区间确定

$$S_\theta = \{x \in \mathcal{R} : \sigma + \xi(x-\mu) > 0\}, \tag{1}$$

其中 $\theta = (\mu, \sigma, \xi)$。因此，累积分布函数 $\mathcal{F}$ 仅针对落在 S_θ范围内的 x 值定义。在极值分析（EVA）中评估极端事件相关风险时，重现水平是一个关键组成部分。这些水平用于估计在特定时间范围（即由 T表示的重现期）内可能出现的极端分位数的期望值。重现水平的计算公式为 $z_T = \mathcal{F}^{-1}(1 - T^{-1})$，其中 $\mathcal{F}^{-1}$ 表示广义极值分布的逆累积分布函数， T 表示重现期。将 $(1 - T^{-1})$ 分位数用作阈值非常重要，因为它表示在指定重现期内超过该阈值的平均频率。

广义极值分布根据其形状参数 ξ 的符号可呈现三种不同形式：适用于轻尾分布的冈贝尔分布（$\xi = 0$）、适用于重尾分布的弗雷歇分布（$\xi > 0$）以及适用于短尾分布的威布尔分布（$\xi < 0$）。 ξ 的符号还决定了广义极值分布是具有上界（$\xi > 0$）还是下界（$\xi < 0$）。

由于支持集 S_θ 依赖于参数 θ，确定广义极值分布参数最大似然估计（最大似然估计）的渐近性质颇具挑战性。然而，Smith (1985) 和 B¨ucher & Segers (2016年) 已证明，通过限制 $\xi > -1/2$ 的下界，可以保留其渐近性质。因此，在我们的模拟研究中，我们限制 $\xi > -1/2$ 的范围，以确保最大似然估计结果有效。

2.2 近似充分统计量

本文提出了一种间接推断方法，用于估计基础广义极值分布的未知参数$\theta \in \Theta \subset \mathcal{R}^k$。该方法以样本 $y = (y_1, y_2, \ldots, y_n)$为基础，并使用一组最小化的低维统计量作为神经网络的输入。该方法专为推断广义极值分布的重尾行为而设计。它通过利用来自分布下尾和上尾的一组极端分位数来实现这一点，其中包括 Q_1、 Q_2和 Q_3。

虽然在间接推断中使用汇总统计量是一种有用的近似方法，但它未必总能实现最优的渐近效率。为解决这一问题，重要的是仔细选择能够准确捕捉广义极值分布重尾行为、同时尽量降低神经网络输入维度的信息性分位数。为确定最优的极端分位数集合，我们尝试了不同的分位数选择，并观察其对网络行为的影响。选择信息性统计量对于间接推断的成功至关重要，因为它会影响神经网络的计算效率、稳健性和可解释性。因此，我们选择近似充分统计量，以确保计算效率和稳健性。

这一方法受到先前研究的启发（Creel & Kristensen (2013)、Jiang et al. (2017)），这些研究表明，在统计建模中使用信息性统计量取得了令人鼓舞的结果。尽我们所知，这是

首次探索将极端分位数作为神经网络的输入，以估计广义极值分布参数的研究。本研究旨在考察给定的一组分位数能否通过复杂的深度神经网络估计广义极值分布的参数，从而为推断该分布的重尾特性提供一种更高效、更准确的方法。有关神经网络框架的更多细节，请参见第2.3节。

2.3 神经网络框架

我们的估计技术涉及一个深度神经网络，该网络以分位数值作为输入，并返回近似的 $\theta = (\mu, \sigma, \xi)$ 作为因变量输出。令 $\mathcal{P} = \mathcal{P}(\mathbf{y}) \in \mathcal{R}^m$ 表示分位数/百分位数值，它是一个标量向量 $m < n$，其中 m 和 n 分别是 $\mathcal{P}(y)\widehat{\ }$ $\widehat{\ }$ 和 y 的大小。函数 $\mathcal{N}$ 将 $\mathcal{P}$ 映射到 θ，从而得到 $\theta = (\widehat{\mu}, \widehat{\sigma}, \xi)$ 作为相应的估计值

$$\mathcal{N} : \mathcal{R}^m \to \Theta; \text{ where } \mathcal{N}(\mathcal{P}) = \widehat{\theta}.$$

前馈神经网络（FFNN） $\mathcal{N}$ 有 L 层，其中第一层是输入层，最后一层是输出层，其余 $L-2$ 层是隐藏层。设 j^{th} 层的 $\mathcal{N}$ 有 n_j 个神经元， $j = 1$， 2，....., L，并由非线性激活函数 f_j 表征。相邻层之间的交互由递归方程定义为

$$h_j = f_j(b_j + w_j h_{j-1}), \quad j = 1, 2, \ldots, L,$$

其中， h_j 是 j^{th} 层所有神经元的向量输出， b_j 是偏置向量， w_j 是连接（$j-1$）th层与 j^{th}层的权重矩阵， h_{j-1}是（$j-1$）th层的向量输出。该方程描述了网络各层执行的计算：每一层的输入都是前一层的输出，并经过权重矩阵、偏置向量和激活函数变换。因此， $\mathcal{N}$通过后续各层的一系列非线性变换，将输入映射为输出。在训练过程中，会调整权重$\widehat{\ }$和偏置，以最小化给定输入 $\mathcal{P}$对应的预测 θ 与真实 θ 之间的差异。

在构建神经网络算法时，选择合适的性能损失函数来衡量 θ 与$\backsim\theta$ 之间的差异至关重要。衡量神经网络算法性能的一种常用指标是均方误差（MSE）损失，该指标基于选定的一批训练样本计算，表示如下：

$$\text{MSE}(\omega) = \frac{1}{n_B} \sum_{i=1}^{n_B} \|\theta_i - \widehat{\theta}_i(\omega)\|_2^2 = \frac{1}{n_B} \sum_{i=1}^{n_B} \left\{ (\mu_i - \widehat{\mu}_i(\omega))^2 + (\sigma_i - \widehat{\sigma}_i(\omega))^2 + (\xi_i - \widehat{\xi}_i(\omega))^2 \right\}, \quad (2)$$

其中， n_B 是批次大小， ω 表示在给定迭代阶段对网络权重和偏置进行编码的矩阵， $\|\cdot\|_2$ 是 l_2 范数。在训练过程中最小化均方误差损失，对于准确估计 θ至关重要，从而能够更接近真实参数值。

然而，在估计广义极值分布的参数时，必须确保估计参数满足广义极值分布的支撑约束，记为 1。为满足该支撑约束，我们通过添加一个惩罚项来修改均方误差损失函数，该惩罚项用于处理所选批次样本 y_{n_B} 中的任何违规情况。我们将惩罚项$\widehat{}$表示为 $C(y_{n_B}, \theta_{\mathrm{NN}})$：如果$y_{n_B}$中的任何 y_i（ $i = 1$、2、.....、 n_B）违反支撑约束，则返回 1；否则返回 0。

$$C(y_{n_B}, \widehat{\theta}_{NN}) = \begin{cases} 1, & \text{if } y_i \text{ violates domain constraint for any } i = 1, 2, ..., n_B \\ 0, & \text{otherwise.} \end{cases}$$

T因此，我们将问题的正则化损失函数定义如下 :

$$\mathrm{Loss}(\omega) = \mathrm{MSE}(\omega) + \lambda \; C(y_{n_B}, \widehat{\theta}_{\mathrm{NN}}(\omega)), \tag{3}$$

其中 λ $\widehat{}$是用于平衡均方误差与惩罚项的权重， $C(y_{n_B}, \theta_{\mathrm{NN}}(\omega))$.

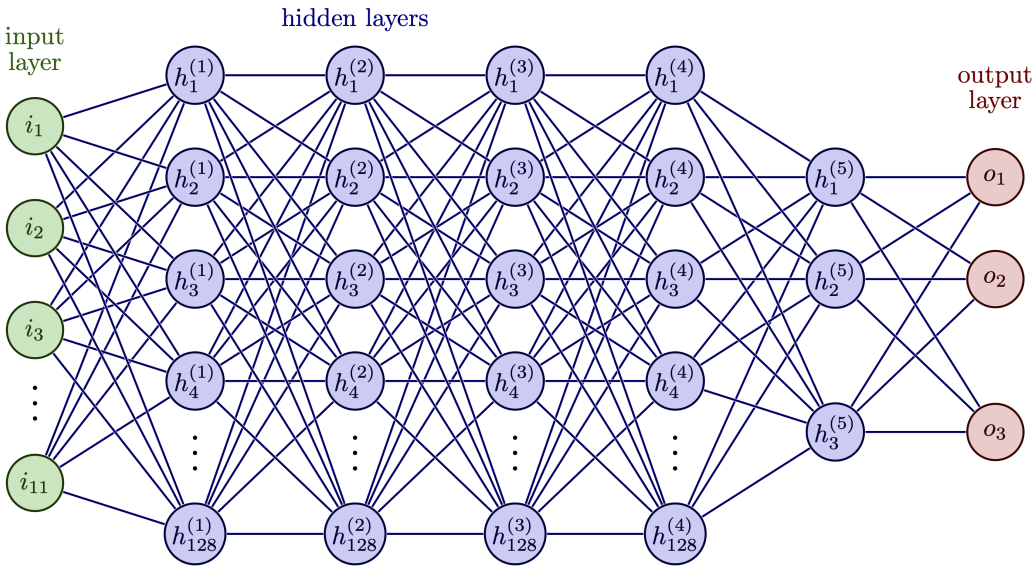

图 1：神经网络模型 $\mathcal{N}$ 的网络架构，以 $\mathcal{P}$；$\{p_1, p_2, \ldots, p_{11}\}$ 作为网络输入，并输出三个标量值。

我们使用合适的优化器来优化该损失函数（例如，本研究使用 RMSprop），其过程包括根据损失函数的梯度迭代更新神经网络的权重。这样，我们便能训练 $\mathcal{N}$ 来估计参数，同时确保其满足广义极值分布的定义域约束。

2.4 网络训练

为训练 $\mathcal{N}$，我们通过模拟广义极值分布在一系列可行参数 θ 下的取值，生成全面的训练数据集。具体而言，我们生成 340，000 种参数配置用于训练和验证，并在 $\mu \in$ (1，50) 的范围内进行均匀抽样，

$\sigma \in (0。1，40)$ 和 $\xi \in (-0.4，0。4)$，其中 μ 和 σ 的取值依据我们对温度、降水和风数据的分析确定。确定参数配置后，我们模拟一个广义极值分布样本，并进一步对样本进行标准化，以提升模型性能。具体而言，我们通过减去样本均值并除以样本四分位距（IQR）进行缩放来实现标准化，这也会重新缩放 μ 和 σ 的取值。值得注意的是，$\xi > -1/2$ 的受限范围与现有文献一致，因为极大似然估计在此范围内有效，且得到既有研究所确立的渐近性质的支持（Smith (1985), B¨ucher & Segers (2016)）。

为了作为神经网络的输入，我们选择一组合适的百分位数，$\mathcal{P}$。为确定合适的百分位数组合，我们考虑生成的广义极值分布样本中从第 0.01^{th} 到第 99.99^{th} 百分位数的取值范围，如下所示：

$$\mathcal{P} = \left\{\mathbf{0.01^{th}},\ \mathbf{0.1^{th}},\ \mathbf{1^{th}},\ \mathbf{10^{th}},\ 25^{th},\ 50^{th},\ 75^{th},\ \mathbf{90^{th}},\ \mathbf{99^{th}},\ \mathbf{99.9^{th}},\ \mathbf{99.99^{th}}\right\}, \tag{4}$$

其中，我们将极端百分位数值加粗以示强调。从这一宽广的取值范围中选择百分位数，可以捕捉广义极值分布样本的完整范围，从而更好地了解该分布的重尾特性。

为了训练神经网络，我们探讨了两种情形：（1）针对训练-验证-测试过程中的参数配置，生成固定样本量为 1000 的广义极值分布样本；（2）生成样本量各异的广义极值分布样本，以更好地模拟现实情况。

（1）针对每种参数配置，模拟一个样本量为 1000 的广义极值分布样本。为优化模拟过程，我们采用了向量化技术，并利用 Google Colab 提供的图形处理器支持。结合这两种方法，我们高效地完成了 340，000 种配置、每种样本量为 1000 的模拟，耗时约 12.6 分钟。为便于训练和验证，我们将 340，000 种参数配置划分为训练集（$N_{train} = 300$，000）和验证集（$N_{valid} = 40$，000），以监测过拟合。此外，我们还定义了测试集（$N_{test} = 10$，000），用于评估模型在上述相同参数范围内的表现。使用较大的训练集可确保优化网络中的权重，从而建立从输入到输出的可靠映射。为防止过拟合，我们在训练过程中采用了提前停止措施，并监测验证损失和学习率。如果验证损失没有改善，就会停止训练。

(2) 为确保我们的研究成果能够泛化至观测数据有限的现实场景，我们考察了用于训练的广义极值分布样本量如何影响研究结果。训练和验证使用相同的 N_{train} 和 N_{valid} 参数配置，但生成广义极值分布样本时，我们选择的样本量范围为 30 至 1000。在训练和验证阶段，共生成 $340,000$ 个参数配置，并为每种广义极值分布样本量随机分配 68,000 个配置。评估时，我们通过固定 μ 为 0，并生成一个 20×20 规则网格 $(\sigma, \xi) \in (0.1, 40) \times (-0.4, 0.4)$，以此创建测试集。随后，针对不同样本量生成广义极值分布样本，每种配置重复 100 次。研究结果表明，将样本量从较小值增加到中等值，可显著提高估计精度。更多详情请参阅第3节。

表 1：前馈神经网络模型摘要

层类型	输出形状	激活函数	参数
稠密层	[-，128]	Relu	1536
稠密	[-，128]	Relu	16512
稠密层	[-，128]	Relu	16512
密集	[-，128]	Relu	16512
全连接	[-，3]	Tanh	387

注：输入层的形状为 [-，11]；输入为 11 个分位数值，输出为估计的广义极值分布参数，形状为 [-，3]。

对于这两种情形，我们都采用一致的网络架构，并使用式 3 所述的损失函数，通过 RMSprop 优化器训练和优化模型。优化器的初始学习率设为 0.001，以便在训练过程中高效、有效地调整模型参数。表 1 全面概述了网络架构，包括每层的输出形状以及所用的激活函数。

我们的神经网络输出层采用自定义激活函数，返回三个标量值，分别对应平移后的 μ、平移后的 σ 和 ξ。我们将 tanh、relu 和 tanh 激活函数组合起来设计该激活函数，以处理这些参数可能的取值范围，并确保估计值准确可靠。为更好地建模重尾分布，我们使用 RMSprop 优化器进行算法优化（参见 Hinton 等人，2012年），每批 64 个样本更新一次权重。我们采用基于式 2 中验证集均方误差的提前停止准则来防止过拟合；当损失未见改善时，即停止训练过程。我们保存第 28th 个 epoch 得到的最佳权重。为了实现该模型，我们选择固定样本情景，因为该模型使用固定的 1000 个广义极值分布样本进行训练，且在准确性方面效率更高。模型完成一个 epoch 需要 260 秒；训练 150 个 epoch，并在第 28th 个 epoch 提前停止，总共需要 2.11 小时。

2.5 样本标准化

在本节中，我们将讨论在训练网络之前对样本进行标准化的重要性。标准化过程有诸多优点，包括提升稳定性、估计效果和性能。不过，我们的主要目标是让网络对不同的尺度和计量单位保持不变，从而使其更为通用，能够适用于更广泛的极端情形。

我们使用样本均值和四分位距对广义极值分布样本进行中心化和尺度变换。设 $y = (y_1, y_2, y_3, \ldots, y_n)$ 是来自 GEV(μ, σ, ξ) 分布的大小为 n 的样本。标准化表示为

$$z = \frac{y - \bar{y}}{\mathrm{IQR}},$$

其中，$\bar{y}$ 是样本均值，IQR 是样本四分位距，而 z 是标准化后的广义极值分布样本。必须注意，将广义极值分布样本标准化并不能保证标准化后的样本仍服从广义极值分布。不过，重新缩放样本可以改变其

位置（μ）和尺度（σ）。 μ 的调整方式如下：

$$\frac{\mu - \bar{y}}{IQR}$$

以及 σ 的调整是

$$\frac{\sigma}{IQR}.$$

为考虑这些变化，我们使用经过变换的百分位数值训练 $\mathcal{N}$，然后将变换反转至原始尺度，以便与真实值进行比较。

通过实施标准化，神经网络可以变得更加稳健，并更好地泛化到降水和风等以不同单位测量的极端事件，从而提高其适用性。标准化通常采用配对方法，例如中位数与四分位距（IQR）配对，或均值与标准差配对。本研究选择了样本均值-四分位距配对，因为其效果优于其他组合。

3 模拟研究

本节展示了模拟研究的结果，研究使用第2.4节所述选定参数范围内生成的 N_{test} 个参数配置组成的测试集，考察神经网络模型 $\mathcal{N}$ 的精度。

3.1 与极大似然方法的比较

为了比较模型性能，我们还计算了测试数据集上 GEV 参数的最大似然估计。最大似然估计使用 R 包 `ismev` （Stephenson（2011））中实现的参数估计方法计算。具体而言， `ismev` 包使用 `optim` 函数进行数值优化，并采用 Nelder-Mead 优化方法计算广义极值分布的最大似然估计（参见 Singer 和 Nelder（2009））。 $\mathcal{N}$ 得到的参数估计值的准确性与最大似然方法的结果一同展示于图2和图3中。

对于我们的模拟研究，情况（1），我们通过运行 $N_{test} = 10{,}000$ 种配置来评估神经网络模型 $\mathcal{N}$ 的性能，每种配置包含 1000 个广义极值分布样本。为比较模型的估计值与最大似然估计，我们使用箱线图来直观呈现不同真实参数值范围内估计值与真实参数值之间的差异。分析结果显示，神经网络估计器的表现与极大似然方法相近，且离群值较少。这些发现表明，神经网络模型 $\mathcal{N}$ 是一种有前景的广义极值分布参数估计方法。

在情况2中，我们使用与前一种情况相同的网络架构和参数配置进行训练和验证。为了评估模型的表现，我们在大小为 20×20 的参数网格（σ，ξ）上生成测试集，其中 $\mu = 0$。对于每种配置，我们针对每个样本量重复生成测试集100次。在图3中，我们比较了神经网络模型和极大似然方法所得参数σ 和 ξ的估计值的均方误差（均方误差）。我们发现，将样本量从72增加到1000，会使神经网络和极大似然方法的均方误差都降低。当样本量为72时，神经网络估计值的均方误差小于极大似然估计值；但当样本量增加到416和1000时，最大似然估计的均方误差更小。

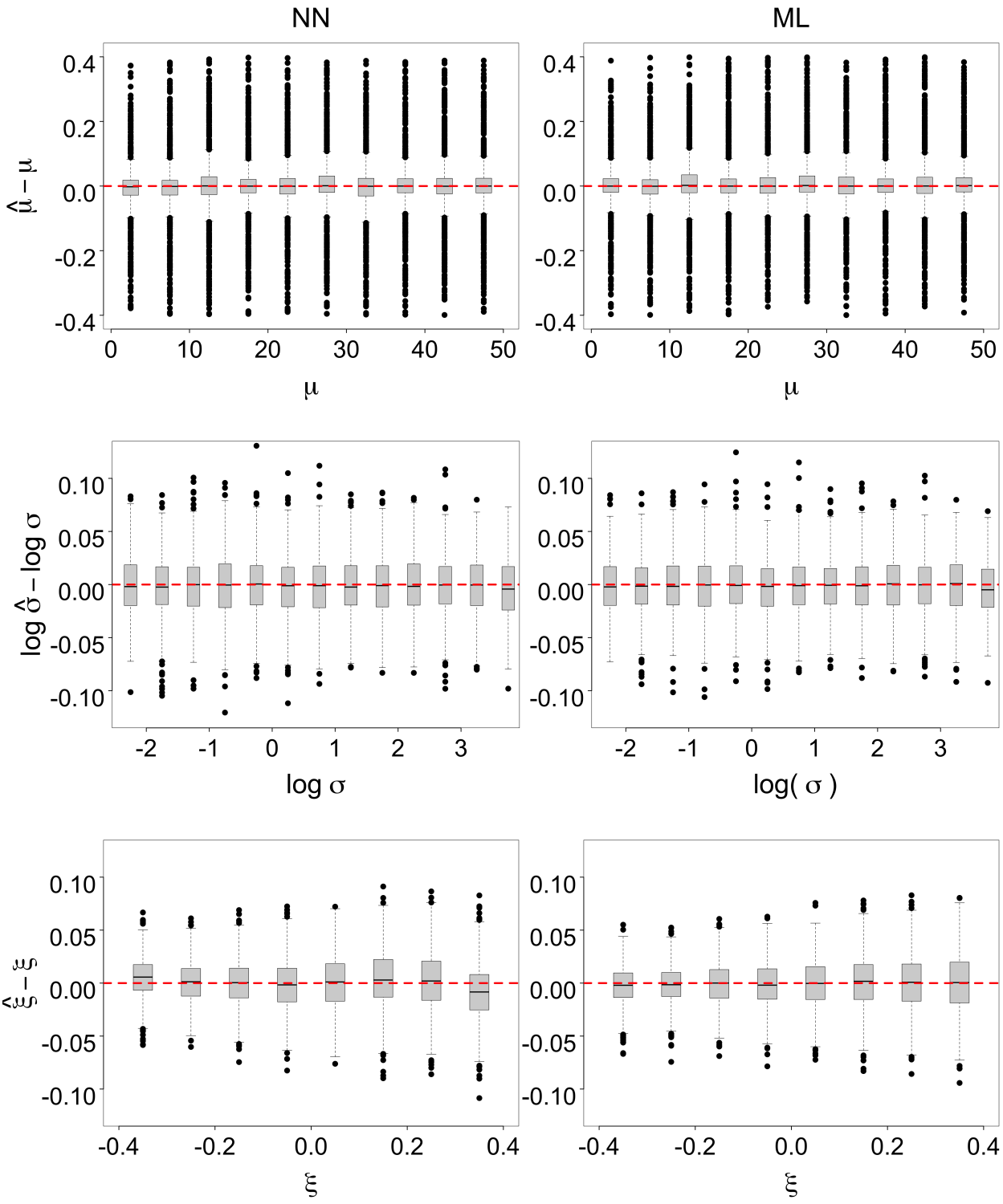

图2：此图总结了我们使用固定的广义极值分布样本（大小为^1000）进行模拟研究的结果。箱线图展示了估计值 $\hat{\mu}$、log $\hat{\sigma}$和 ξ 与其真实值之间的偏差，基于一个包含 10,000 个样本的测试集。横轴表示参数的真实值，纵轴表示估计值与真实值之间的偏差。

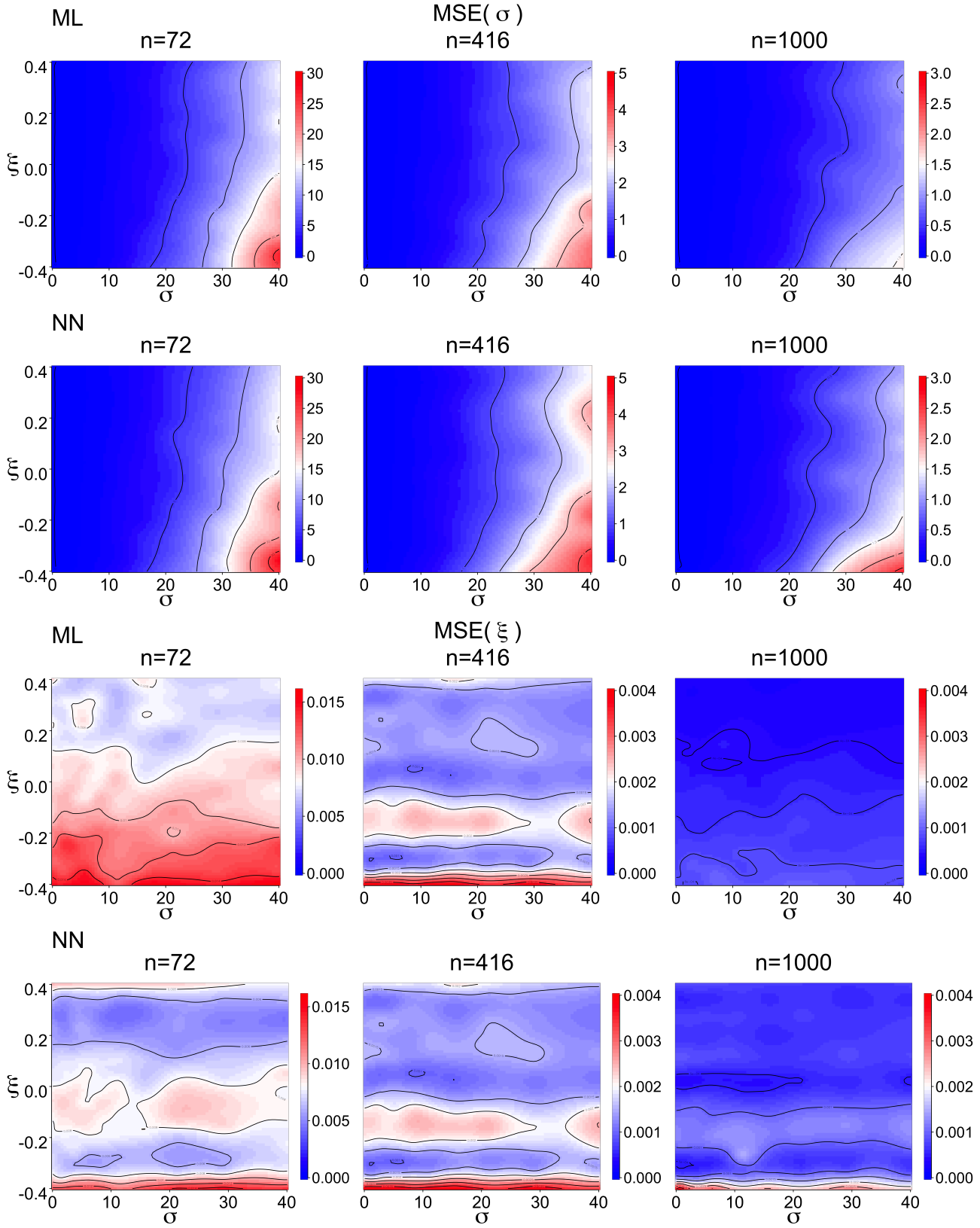

图3：该图基于我们的模拟研究，展示了样本量在72、416和1000之间变化时均方误差的表现。图像图展示了真实值（σ，ξ）下的均方误差，其中横轴为 σ，纵轴为 ξ。11

其均方误差（MSE）低于神经网络估计值。不过，总体而言，在不同样本量下，两种方法的均方误差差异都很小。

总体而言，本研究表明，神经网络模型可以作为传统极大似然方法，在极值建模中的一种可行替代方案。

3.2 自助法

为考虑神经网络所得参数估计值的不确定性，我们采用参数自助法（Efron & Tibshirani（1994年））。我们从原始数据中生成 $B = 900$ 个自助样本，对每个样本拟合神经网络，并计算真实参数的95%置信区间。神经网络训练完成后，自助法不会产生额外的计算成本。我们可以生成B次自助重复抽样，并在几秒内利用神经网络得出相应结果。

为评估基于自助法的置信区间的性能，我们将其与基于似然的置信区间进行比较；后者利用最大似然方法的海森矩阵计算最大似然估计的标准误。我们使用第3.1节所述的 `ismev` 软件包进行计算。我们在测试集上计算基于最大似然的置信区间，并使用在固定的1000个广义极值分布样本上训练的神经网络，在10000个测试集和900次自助重复抽样中获得基于自助法的置信区间。

图4展示了在不同真实参数值下，自助法神经网络置信区间宽度与基于最大似然的置信区间宽度之比。箱线图总结了该比值的分布范围，结果表明，自助法神经网络置信区间略宽于最大似然模型的置信区间，但观察到的差异极小。具体而言，与 σ相比，我们发现 μ 和 ξ 的自助法置信区间宽于最大似然置信区间。自助法得到的置信区间宽度可能大于极大似然方法得到的宽度，因为自助法不假设严格的分布约束。

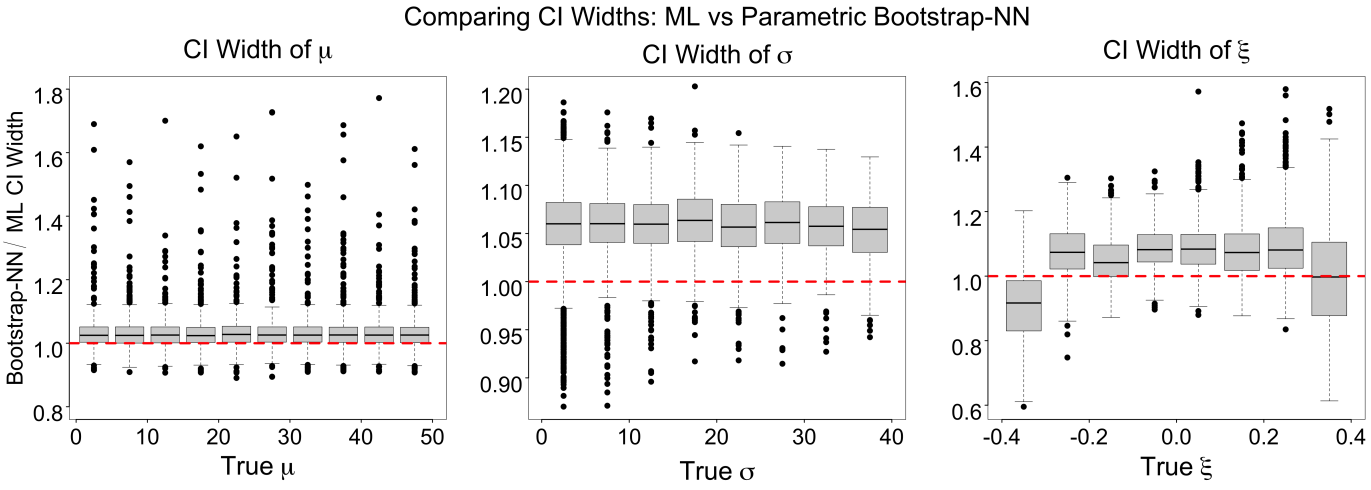

图4：该图比较了在 10,000 个测试集上，自助法神经网络和机器学习方法得到的广义极值分布参数的置信区间宽度。我们使用固定的、经 1000 个样本训练的神经网络，并进行 900 次自助重复抽样，以计算自助法神经网络的输出。横轴表示参数的真实值，纵轴表示自助法神经网络与最大似然方法所得置信区间宽度之比。

3.3 时间比较

神经网络模型 $\mathcal{N}$ 的训练与其他估计方法相比，计算效率更高。该模型在基于云的 Python 平台 Google Colab 上实现，所使用的计算环境配备了 2 个虚拟 CPU 、 32GB 内存，以及以下任一配置：具有 16GB 显存的 P100 GPU，或具有 16GB GPU 显存且系统内存可扩展至 25GB 的 T4 GPU。为便于比较，使用 R 软件包 ismev 计算了最大似然估计。计算在一台配备 2.3 GHz 双核英特尔 i5 处理器和 8GB 内存的笔记本电脑上进行。神经网络模型对 10,000 个测试样本的评估耗时 4 秒，而计算最大似然估计约耗时 10.631 分钟。基于此，我们预计，在扩展到目标模型输出时，速度将显著提升超过 150 倍。

4 案例研究

另一种验证神经网络估计器并测算其运行时间的方法，是复现一项重要气候模型分析的结果。我们分析了第三版社区气候系统模型（CCSM3）千年模拟中的极端温度；该模型是一种广泛用于气候研究的全球气候模型，模拟了不同大气 CO_2 浓度下的情况（Collins 等（2006））。CCSM3 模型完整模拟了大气、陆地、海冰和海洋组成部分；大气和陆地采用 T31 分辨率的网格，海洋和海冰则采用约 1° 分辨率（Huang 等（2016）；Yeager 等（2006））。

我们考虑在北美洲 133 个空间位置开展的 1000年控制模拟，并考察三种 CO_2 浓度情景：前工业时期（289 ppm），以及未来情景中的 700 ppm（全球平均气温升高 3.4℃）和 1400 ppm（全球平均气温升高 6.1℃）（Huang 等（2016））。太阳强迫和气溶胶浓度等关键外部强迫均固定在前工业时期的水平。假设经过预热期后，最终的 1000年模拟处于平稳状态，且不存在气候漂移。根据模型输出，计算每个网格单元每年的日最高气温。

我们使用广义极值分布对空间域内 1000 年的年最高气温进行建模。我们对该域内的每个站点分别拟合广义极值分布，并假设每个站点的 1000 年年最高气温值服从其特定的广义极值分布。为估计每个网格单元的广义极值分布参数，我们使用 $\mathcal{N}$ ，具体方法如前文所述。图 5 展示了神经网络对前工业时期 CCSM3 广义极值分布参数的估计值，以及未来情景下可能发生的变化。结果表明，前工业时期的广义极值分布与 Huang 等（2016）的先前研究结果一致。在对极端温度进行建模时，通常会观察到负形状参数，而我们的模型输出也证实了这一趋势。图 6 进一步展示了神经网络模型与最大似然模型在估计广义极值分布参数方面的性能比较。

最后，采用与第 3.2 节相同的设置，我们可以在约 0.4 秒内，通过 900 次自助法重复抽样计算出一个基于自助法的置信区间。

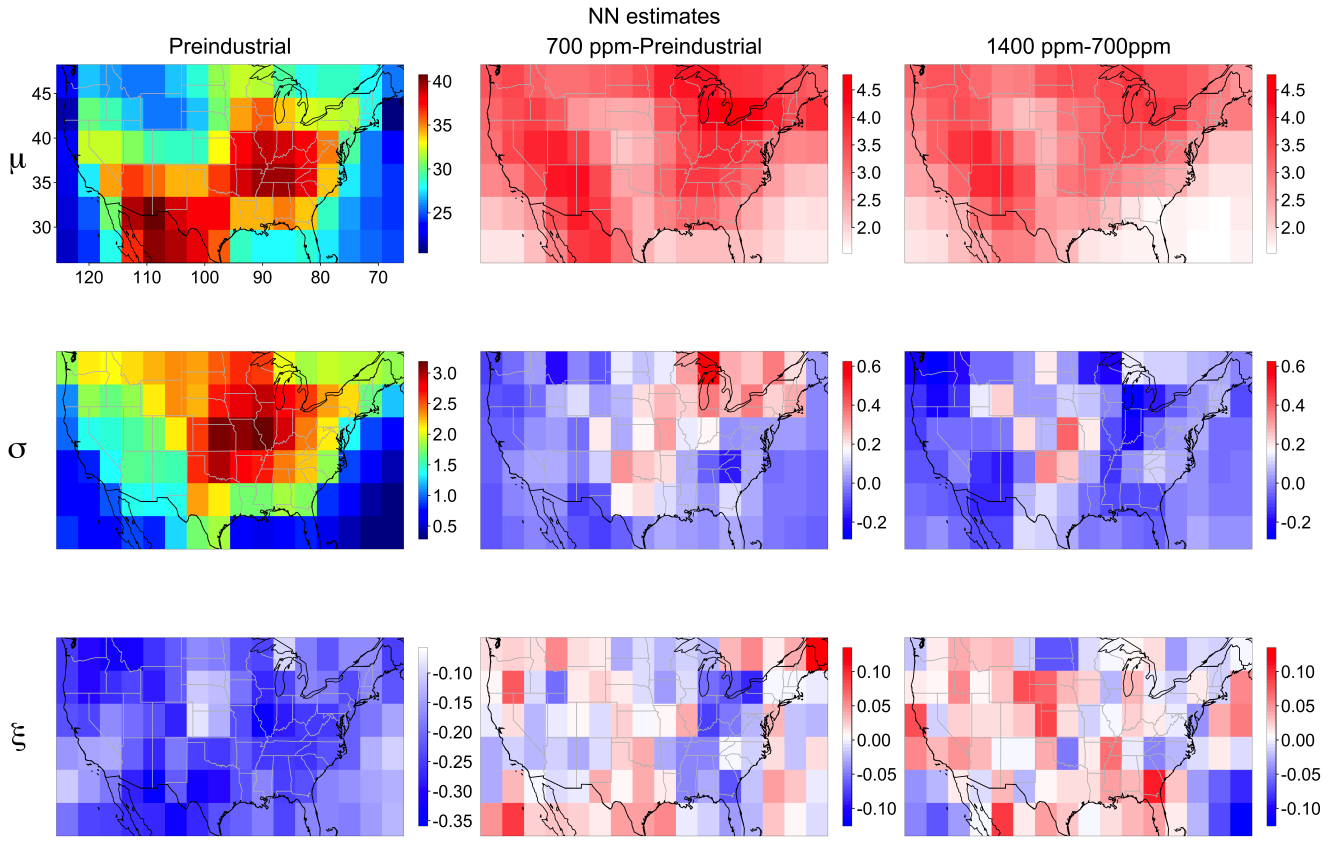

图 5：前工业时期 CCSM3 广义极值分布参数的神经网络估计值，以及未来情景下可能发生的变化。左侧：前工业时期中心的估计值；从 289 ppm CO_2 到 700 ppm CO_2 浓度时参数估计值的预期变化；右侧：从 700 ppm CO_2 到 1400 ppm 时估计值的变化。

5 结论

本研究强调了深度学习算法在无似然推断中的应用进展。结果表明，经过良好训练的神经网络能够以与传统机器学习方法相当的准确度，估计复杂重尾分布（如广义极值分布）的参数。与其他参数相比，形状参数的估计可能比最大似然估计具有更大的变异性；鉴于重尾广义极值分布中形状参数估计的难度，这在预料之中。此外，我们的研究结果表明，计算速度显著提升：处理大型数据集时，与传统机器学习方法相比，模型评估时间缩短至原来的 1/150。使用神经网络让我们能够全面掌控测试样本和训练样本，并处理范围广泛的参数。这使我们能够根据问题的具体要求定制神经网络，并评估其可靠性。

然而，使用神经网络方法时，必须考虑到它的若干局限性。选择适当的超参数是构建神经网络的关键步骤，因为这会显著影响模型的性能和准确性。超参数包括隐藏层的数量、隐藏层中的神经元数量、激活函数的选择、学习率，

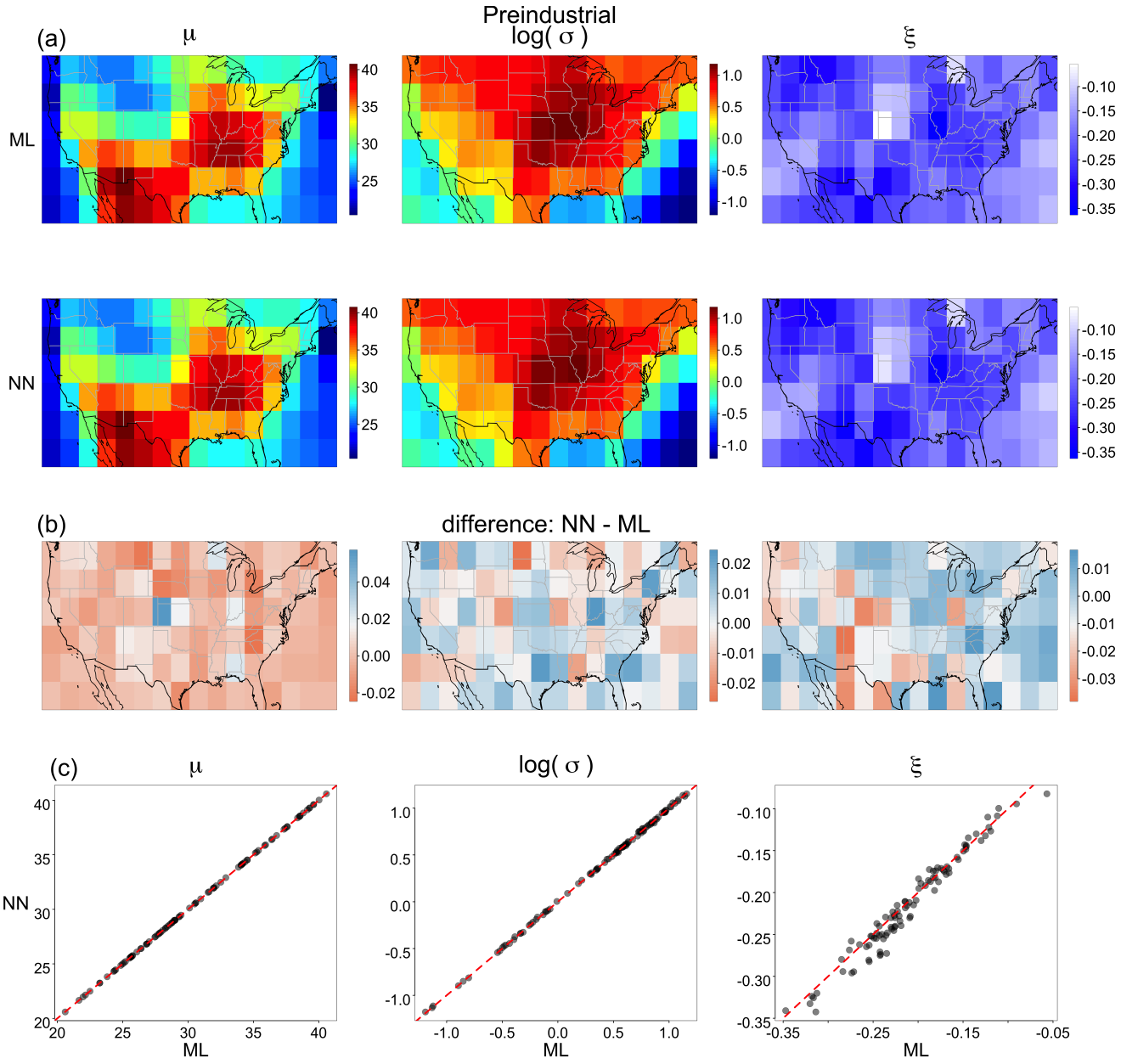

图 6: （a）展示 ML 和神经网络模型在各空间位置上的 CCSM3 广义极值分布参数估计值；（b）展示 ML 和神经网络所得参数估计值之间的差异；（c）通过散点图比较神经网络与 ML 对广义极值分布参数的估计结果。

批次大小必须通过反复试验谨慎选择。由于超参数的搜索空间庞大且复杂，这一过程可能既耗时又困难。此外，优化神经网络模型中每一层的权重可能会产生大量参数，使模型难以处理。此外，为训练集设计时选择合适的参数范围至关重要。还必须谨慎考虑作为网络输入的信息性统计量，因为它们必须提供足够的数据相关信息，才能实现准确估计。

总之，使用所提出的模型时，必须考虑网络架构、超参数的选择以及统计量的选取，以确保结果可靠。我们对广义极值分布参数的估计尚未考虑任何潜在的空间或时间效应。这为未来研究提供了可能性，即探讨将时间依赖结构纳入广义极值分布参数估计，从而提高极值预测的准确性和稳健性。此外，在极值空间建模中进一步应用这一方法，有助于深入了解不同地理位置上极端事件的分布与行为。此外，这种方法也可能有益于其他与重尾分布相关的统计建模方法。

致谢

我们衷心感谢 Whitney Huang 慷慨分享 CCSM3 模型三组千年模拟的 1000 年输出数据，供我们的案例研究使用。

参考文献

Adger, W. N. (2003)，《社会资本、集体行动与适应气候变化》，《经济地理》**79年**(4)，387–404。Bali, T. G. (2003)，《广义极值分布》，《经济学快报》**79年**(3)，423–427。Blum, M. G. (2010)，《近似贝叶斯计算：非参数视角》，《美国统计学会会刊》**105年**(491)，1178–1187。Broussard, J. P. & Booth, G. G. (1998)，《德国股指期货中极值的表现：日内保证金设定的应用》，《欧洲运筹学杂志》**104年**(3)，393–402。B¨ucher, A. & Segers, J. (2016)，《关于广义极值分布的最大似然估计量》，*arXiv* 预印本 *arXiv:1601.05702* 。Casson, E. & Coles, S. (1999)，《极值的空间回归模型》，《极值》1，449–468。Castillo, E. (2012)，《工程中的极值理论》，爱思唯尔。

Castro-Camilo, D., Huser, R. & Rue, H.（2021），‘极端值基于广义极值分布的回归模型的实用策略’，*arXiv*预印本 *arXiv:2106.13110* 。Chon, K. H. & Cohen, R. J.（1997），‘使用人工神经网络进行线性和非线性ARMA模型参数估计’，*IEEE*生物医学工程汇刊 44(3), 168–174。Coles, S., Bawa, J., Trenner, L. & Dorazio, P.（2001），极值统计建模导论，第208卷，Springer。Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Doney, S. C., Hack, J. J., Henderson, T. B. 等（2006），‘第三版社区气候系统模型（CCSM3）’，《气候学报》 **19年**(11), 2122–2143。Cooley, D., Nychka, D. & Naveau, P.（2007），‘极端降水重现水平的贝叶斯空间建模’，《美国统计学会会刊》 **102年**(479), 824–840。Creel, M.（2017），‘用于间接推断的神经网络’，《计量经济学与统计学》 **2年**, 36–49。Creel, M. & Kristensen, D.（2013），‘间接似然推断（修订版）’。Cremanns, K. & Roos, D.（2017），‘深度高斯协方差网络’，*arXiv*预印本 *arXiv:1710.06202* 。Davison, A. C. & Huser, R.（2015），‘极值统计学’，《统计学及其应用年度评论》 **2年**, 203–235。Efron, B. & Tibshirani, R. J.（1994），自助法导论，CRC出版社。El Adlouni, S. & Ouarda, T. B.（2009），‘使用出生-死亡马尔可夫链蒙特卡罗方法联合进行广义极值模型的贝叶斯模型选择和协变量参数估计’，《水资源研究》 **45年**(6)。Erhardt, R. J. & Smith, R. L.（2012），‘空间极值的近似贝叶斯计算’，《计算统计学与数据分析》 **56年**(6), 1468–1481。Erhardt, R. & Sisson, S. A.（2016），‘使用近似贝叶斯计算对极值建模’，《极值建模与风险分析》 第281–306页。Fisher, R. A. & Tippett, L. H. C.（1928年），样本中最大或最小成员频率分布的极限形式，载于 《剑桥哲学学会数学会刊》，第24卷，剑桥大学出版社，第180–190页。Gamet, P. & Jalbert, J.（2022年），‘用于尾部估计的灵活扩展广义帕累托分布’，《环境计量学》 **33年**(6), e2744。Gerber, F. & Nychka, D.（2021年），‘使用神经网络快速估计空间高斯过程模型的协方差参数’，《*Stat*》 **10年**(1), e382。Gumbel, E. J.（2004年），极值统计学，Courier公司。

Haan, L. & Ferreira, A. (2006), 《极值理论：导论》, Vol. 3, Springer. Hinton, G., Srivastava, N. & Swersky, K. (2012), '用于机器学习的神经网络，第 6a 讲：小批量梯度下降概述', 引自 14(8), 2. Hosking, J. R. M., Wallis, J. R. & Wood, E. F. (1985), '基于概率加权矩法估计广义极值分布', *Technometrics* **27**(3), 251–261. Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S. & Moyer, E. J. (2016), '利用广义极值（GEV）分布从千年尺度气候模拟中估计极端温度的变化', 统计气候学、气象学与海洋学进展**2**(1), 79–103. Jenkinson, A. F. (1955), '气象要素年最大值（或最小值）的频率分布', 英国皇家气象学会季刊**81**(348), 158–171. Jiang, B., Wu, T.-y., Zheng, C. & Wong, W. H. (2017), '通过深度神经网络学习用于近似贝叶斯计算的汇总统计量',《中华统计学志》pp. 1595–1618. Katz, R. W. (2010), '气候变化中的极值统计学',《气候变化》**100**(1), 71–76. Kysel`y, J. (2008), '关于使用非参数自助法估计极值模型不确定性的注意事项',《应用气象学与气候学杂志》**47**(12), 3236–3251. Lenzi, A., Bessac, J., Rudi, J. & Stein, M. L. (2021), ' 难处理模型中参数估计的神经网络 ', *arXiv* 预印本 *arXiv:2107.14346* . Majumder, R., Reich, B. J. & Shaby, B. A. (2022), '利用深度学习近似和灵活空间过程对极端径流建模', *arXiv* 预印本 *arXiv:2208.03344* . Martins, E. S. & Stedinger, J. R. (2000), '适用于水文数据的广义最大似然广义极值分位数估计量',《水资源研究》**36**(3), 737–744. Opitz, T., Huser, R., Bakka, H. & Rue, H. (2018), 'INLA迈向极值：用于估计高时空分位数的贝叶斯尾部回归',《极值》**21**, 441–462. Power, S., Lengaigne, M., Capotondi, A., Khodri, M., Vialard, J., Jebri, B., Guilyardi, E., Mc- Gregor, S., Kug, J.-S., Newman, M. et al. (2021), '热带太平洋的年代际气候变率：特征、成因、可预测性与展望',《科学》**374**(6563), eaay9165. Richards, J. & Huser, R. (2022), '用于神经网络极端分位数回归的统一部分可解释框架', *arXiv* 预印本 *arXiv:2208.07581* . Sainsbury-Dale, M., Zammit-Mangion, A. & Huser, R. (2022), '利用置换不变神经网络对难处理模型进行快速最优估计', *arXiv* 预印本 *arXiv:2208.12942*。

Singer, S. & Nelder, J. (2009), ‘Nelder-Mead 算法’, *Scholarpedia* 4(7), 2928. Sisson, S. A., Fan, Y. & Beaumont, M. A. (2018), abc 概述，载于‘《近似贝叶斯计算手册》’, Chapman and Hall/CRC, pp. 3–54. Smith, R. L. (1985), ‘一类非正则情形中的最大似然估计’, *Biometrika* **72**(1), 67–90. Smith, R. L. (1990), ‘极值理论’, 《应用数学手册》**7**(437-471), 18. Smith, R. L. (1991), ‘用于可靠性数据的威布尔回归模型’, 《可靠性工程与系统安全》**34**(1), 55–76. Stephenson, A. (2011), ‘软件包“ismev”’, 在线地址：*http://cran. rproject. org/web/packages/is- mev/ismev. pdf*。访问日期：*9*月。Wikle, C. K. & Zammit-Mangion, A. (2022), ‘空间及时空数据的统计深度学习’, *arXiv* 预印本 *arXiv:2206.02218* . Yadav, R., Huser, R. & Opitz, T. (2022), ‘针对空间相关超阈值数据的灵活贝叶斯分层建模框架’, 《空间统计》p. 100672。