arXiv:2603.01653v1 [stat.AP] 2 Mar 2026

电力网络中天气驱动故障的概率预测：适用于极端与非极端事件的灵活方法

Mateus Maia*、Daniela Castro-Camilo、Jethro Browell，格拉斯哥大学数学与统计学院

2026年3月3日

摘要

电力网络易受天气损害，严重天气事件往往会导致故障和停电。及时预测故障发生情况——从临近预报到提前数日的预测——有助于加强准备、加快响应并缩短停电时长。要在实际运行中发挥作用，此类预测必须量化不确定性，以便进行风险知情的资源分配。我们提出了一种新颖的概率框架，用于预测故障计数，并同时捕捉典型事件和极端事件。非极端故障通过对多个加性分位数回归的估计结果进行线性插值来建模，而极端事件则采用离散广义帕累托分布描述。为纳入天气波动的影响，我们使用集合数值天气预报，从而有助于量化预测中的不确定性。该方法旨在提供未来四日内可靠的故障预测。我们通过数值实验评估该模型，并将其应用于大不列颠两个配电网络的历史故障数据。结果表明，与常规做法及其他建模方法相比，该预测显著改善。2024年10月至2025年3月期间，我们与苏格兰电力能源网络合作开展的从业者试验进一步证明了该预测的运营价值。工程师认为预测足够可靠，可为决策提供依据，并能使网络运营商和电力消费者双方受益。

关键词：电力网络、计数极值理论、广义加性模型、概率预测、分位数回归、天气驱动的故障。

*通讯作者。电子邮箱：mateus.maiamarques@glasgow.ac.uk

1 引言

电力对现代生活至关重要，供应中断可能导致重大的人员与经济影响。因此，确保电力网络的韧性是网络运营商、监管机构和消费者共同的关键优先事项。极端天气事件对暴露的基础设施构成特别威胁，包括架空线路及其支撑电杆和输电塔。韧性的一个关键要素是防备。虽然无法完全防止天气相关损害，但预判极端状况可以采取预防措施，从而增强系统稳健性并支持及时修复。英国近期的研究，例如Cha等人 （2025），评估了风暴 ´Eowyn，以凸显对关键基础设施的跨部门影响，进一步强调预判中断并增强系统韧性的重要性。这凸显了在从数小时到数日的提前时长内预测电力网络中可能发生故障的数量和位置的重要性，从而能够进行主动规划与响应。

尽管已有大量文献研究电力网络对恶劣天气的韧性，但其中大部分聚焦于资产设计与长期网络规划。相比之下，关于短期（提前数日）故障预测的研究仍然有限，目前该领域的商业解决方案也很少——由IBM、GE、DTN、Bellrock和StormGeo等公司提供。现有方法往往仅限于提前一日的预测，且不提供或仅提供有限的不确定性量化。不过，它们确实提供了有力证据，表明天气驱动的故障预测是可行的，而这些方法很可能可以扩展到更长时效的预测以及概率预测。

报道最为全面的故障预测能力来自康涅狄格大学的“停电预测模型”。该模型已被业界使用，可在风暴期间以4公里网格预测美国东北部的故障（Watson 等，2020；Cerrai 等，2019）。它专门针对温带风暴预测停电——停电被定义为任何需要维修队处置的事件。研究表明，天气预报中的误差对总体故障预测误差的贡献很小，误差的主要来源在于天气到故障的建模过程。一项纳入雷暴的扩展研究发表于Alpay等（2020）。中国境内由热带气旋造成的输电塔故障在Huang与Wang （2025）中利用脆弱性曲线和降尺度天气预报进行预测，以应对极端风速预测中的系统性偏差。

其他研究者也为其他地区开发了类似但复杂程度较低的模型，例如芬兰（Brester 等，2020年）。在英国，Tsioumpri 等(2021年)和Wilkinson等(2022年)开发了一种方法论，后来由 DTN 扩展为商业服务。后者揭示了天气预报——具体而言是降水、风速和闪电——与英国电力网络公司的配电网中的故障数量之间的相关性，并提出了用于预测英国电力网络公司配电网故障数量的简单统计模型，该研究针对的是

英国东南部。其他例子包括 Ghaemi 等(2024年)，其基于天气预测变量对输电线路故障进行故障预测；大不列颠的停电也被发现与大尺度天气型有关，这些天气型可提前数周进行有一定技巧评分的预报（Souto 等，2024年）。

本研究的动因是考察苏格兰电力能源网络（SPEN）内 11 个地区的故障分布。在正常运行条件下，控制中心可以利用概率预测工具，更高效地调配资源并主动安排维护，同时考虑典型和罕见的天气情景。故障数量较多的事件在交通信号灯式系统中被归类为红色事件，通常与恶劣天气有关。由于维护团队需要为此做好充分准备，这类事件尤其值得关注，尽管其发生频率相对较低。另一方面，破坏性较小但发生更频繁的琥珀色事件涉及的故障数量较少，其预测结果对于日常资源管理仍然很重要。

图1 左侧面板展示了 SPEN 单个地区十年间每日故障发生情况，呈现了典型和极端故障情景。右侧面板比较了两种建模方法：对于最大 10米风速（$\mathrm{ws10}_{\max}$）超过约 9m/s 的日子，图中显示了在红色事件未发生时发布红色故障带预警的预测概率。朴素模型以叉号表示，并未专门针对极值进行设计，因此往往会为未观测到的红色事件分配相对较高的概率。相比之下，本文提出的方法以圆圈表示，在这一情形下给出的概率估计较低且更稳定，这与其基于混合模型处理极端和非极端行为的方式一致。

在概率预测问题中，分位数回归（QR）最初由Koenker 和 Bassett Jr （1978）提出，现已得到广泛应用，包括能源行业（Zhu和He，2024年；Ruiz-Abell´on 等，2024年；Zhang等，2014）。分位数回归的成功可归因于多种因素：它不仅能描述平均值，还能有效刻画新观测值的分布；对异常值具有稳健性；与传统分布回归模型相比，对分布的假设要求也更弱（Koenker 等，2013）。此外，分位数回归的最新进展，例如加性分位数回归，通过引入加性项提高了灵活性，从而能够描述预测变量与估计分位数之间的非线性效应（Fasiolo 等，2021a；Waldmann 等，2013）。

尽管分位数回归（QR）传统上是针对连续响应变量制定的，但当主要关注点是预测性能而非推断时，也可将其应用于计数数据。这种情形与电力网络的故障预测相关，在此类场景中，准确估计分位数比严格的推断分析更为重要。对于离散响应，点质量的存在以及分位数函数不可微，会给样本分位数及相关推断所依据的标准渐近理论带来挑战（Machado 和 Silva，2005）。为了解决

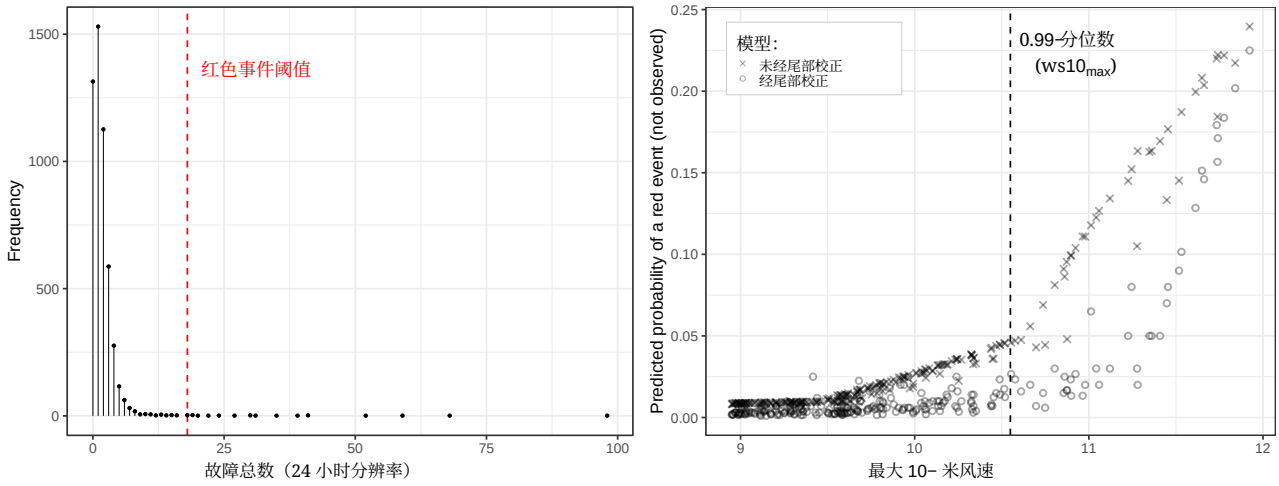


图 1：SPEN 某区十年间每日故障发生情况（左图），竖直虚线表示红色事件阈值。右图显示在非事件日，当$\mathrm{ws10}_{\max}$ 超过 9m/s 时，发布红色预警的估计概率，并对比朴素模型与所提出的模型。

这一问题，已有研究提出了基于响应抖动处理的方法，即向响应变量添加有界连续噪声，以近似连续分布（Machado 和 Silva，2005；Jantre，2023）。然而，引入额外随机性可能导致分位数估计产生偏差，并增加计算成本。当可靠预测是首要目标、正式推断居于次要地位时，直接将分位数回归应用于计数结果仍是一种务实的建模选择。

图1 显示，对故障计数进行建模需要明确处理尾部行为。尽管分位数回归能够生成概率预报，但由于尾部数据稀疏，条件极值的估计受到限制，尤其是在进行协变量条件化时。因此，稳健的尾部预测促使人们将基于分位数的模型与极值理论相结合。以往研究已将建模方法扩展至同时捕捉典型事件和极端事件，并在与能源相关的预测任务中采用基于极值理论的方法纳入条件极端分位数。然而，这些研究主要聚焦于连续响应变量（Gonçalves等，2021；Browell 和 Fasiolo，2021）。例如，Browell 和 Fasiolo（2021）利用加性分位数回归混合对非极端分位数下的电力负荷进行建模，而极端分位数则根据极值理论分布得出；两者均以数值天气预报为条件。

为满足准确预测极值和非极值故障计数的需求，我们提出了*X-flexForecast*，这是一个用于对全范围故障发生情况进行建模的概率框架，并特别关注极值。非极值故障通过对多个加性分位数回归进行线性插值来建模，而极值则采用

离散广义帕累托分布进行捕捉。天气相关的不确定性通过数值天气预报（NWP）纳入模型，从而提高天气驱动预测的可靠性。该框架针对最长四日的提前时长，在预测不确定性与控制室决策的实际需求之间取得平衡。

本文其余部分安排如下。第2 节回顾构成*X-flexForecast*基础的方法学组件，第3节详细介绍所提出的框架，该框架结合集合数值天气预报、加性分位数模型和极值理论，对离散结果进行概率预测。第4 节展示基于真实数据的模拟研究，第5 节将该方法论应用于SPEN 故障数据，第6节则讨论结果、局限性和未来研究方向。

2 背景

本节介绍建模和预报框架的背景。记号与第5节的真实数据应用保持一致。数据由区级故障计数组成$y_{d,\,t}$ 时间为 t，其中 d 为各区的索引。协变量向量 $\mathbf{x}_{d,\,t}$ 包含天气相关变量的汇总统计量，其各组成部分在不同地区之间可能有所不同，因为故障的发生可能由不同的环境因素驱动。我们的目标是将$Y_{d,\,t}$ 视为随机变量，并选择合适的模型来估计条件分布$F(Y_{d,\,t} \mid \mathbf{x}_{d,\,t})$。

2.1 分位数回归

为避免对 $F(Y_{d,\,t} \mid \mathbf{x}_{d,\,t})$ 假设参数形式，并允许其随 $\mathbf{x}_{d,\,t}$ 平滑变化，我们按照 Fasiolo等人 (2021a)，采用广义加性模型进行多重分位数回归。条件分布 $F(Y_{d,\,t} \mid \mathbf{x}_{d,\,t})$ 由多个条件分位数构建

$$q_\alpha(Y_{d,t} \mid \mathbf{x}_{d,t}) = F^{-1}_{Y_{d,t}}(\alpha \mid \mathbf{x}_{d,t}) = g^{-1}\left(\beta_{\alpha,0} + \boldsymbol{\beta}^\top_{\alpha,A}\mathbf{A}_{d,t} + \sum_k f_{\alpha,k}(\boldsymbol{x}^{S_k}_{d,t})\right), \tag{1}$$

在概率水平 $0 < \alpha_1 < \alpha_2 < \cdots < 1$ 下，其中 $g(\cdot)$ 是单调函数。向量$\mathbf{A}_{d,\,t}$ 是一个列向量，包含来自完整协变量向量 $\mathbf{x}_{d,\,t}$的协变量子集，这些协变量以线性形式进入模型，也就是说，不通过基函数展开。这些协变量直接乘以相应的回归系数向量 $\boldsymbol{\beta}_{\alpha,\,A}$，其余协变量则通过平滑函数 $f_{\alpha,\,k}(\cdot)$ 进入模型。

除线性项外，我们还纳入由 k索引的平滑分量，每个平滑分量均定义在协变量向量的一个子向量上。令 $S_k \subseteq \{1, \ldots, p\}$ 表示一个索引集，用于指定$\mathbf{x}_{d,t}$ 的哪些分量进入第 k个平滑项。我们记作$\mathbf{x}^{S_k}_{d,t} = (x^{\{i\}}_{d\,t})_{i \in S_k}$。例如，若 $S_k =$

$\{2,4\}$，则$\mathbf{x}_{d,t}^{S_k} = (x_{d,t}^{\{2\}}, , x_{d,t}^{\{4\}},)$。公式（1）中的加性预测变量包含$\sum_k f_{\alpha,k}(\mathbf{x}_{d,t}^{S_k})$。每个平滑函数$f_{\alpha,k}$ 表示为预先指定的样条基函数的线性组合，$f_{\alpha,k}(\mathbf{x}_{d,t}^{S_k}) = \sum_{j=1}^{K_k} b_{k,j}(\mathbf{x}_{d,t}^{S_k})\gamma_{\alpha,k,j}$，其中$b_{k,j}(\cdot)$是利用$\mathbf{x}_{d,t}^{S_k}$,$\gamma_{\alpha,k,j}$ 计算得到的基函数，为相应的系数，且 K_k 表示基函数的个数。

分位数（或 Pinball）损失的最优平滑方法在 Fasiolo等人（2021a）中有完整描述，这里仅作简要总结。对于概率水平 α，平滑后的损失为 $\tilde{\rho}_\alpha(y-\mu) = (\alpha-1)(y-\mu) + \lambda \log\left(1+e^{\frac{y-\mu}{\lambda\sigma}}\right)$，其中 $\lambda > 0$；当 $\lambda \to 0$ 时，可恢复通常的 Pinball损失。此外，对于异方差的 $y \mid \mathbf{x}$（即 y 的方差取决于 $\mathbf{x}$），最优带宽 $h = \lambda\sigma(\mathbf{x})$ 也取决于 $\mathbf{x}$。在估计回归系数之前，先估计 $\mu(\mathbf{x})$ 和$\sigma(\mathbf{x})$，后者采用广义加性模型方法（GAM；Hastie 和 Tibshirani，1990）。

考虑到 $Y_{d,t}$ 的离散性质，尽管分位数回归传统上是针对连续响应构建的，但也可以扩展以处理离散结果。 Machado 和 Silva (2005) 提出了一种基于 抖动处理的方法，即向计数变量添加均匀随机噪声，从而有效地将其转换为连续变量。这样便可满足一致性和渐近正态性等标准渐近结果，因此当推断模型参数是主要目标时，这种方法尤其有用。

2.2 离散广义帕累托分布

由于极值区域的数据稀疏性，QR 在估计条件极端分位数方面存在局限性；当以外生变量为条件时，这一挑战会进一步加剧（Koenker，2005）。因此，需要采用替代模型来更好地刻画尾部行为。研究分布的右尾，本质上涉及考察超过足够高阈值 u 的超越值。超越值定义为观测值与阈值 u 之间的非负差值。当所研究的响应 Y 为连续变量，并满足最大稳定性假设时（参见例如 Coles，2001，第 4 章），这些超越值渐近地服从广义帕累托分布（GP），其形式为 $F_{\mathrm{GP}}(y \mid Y > u) = 1 - \left(1 + \frac{\xi(y-u)}{\sigma}\right)^{-1/\xi}$ if$\xi \neq 0$，以及 $F_{\mathrm{GP}}(y \mid Y > u) = 1 - \exp\left(-\frac{y-u}{\sigma}\right)$ if $\xi = 0$。其中，$\sigma > 0$ 和 $\xi \in \mathbb{R}$ 分别表示尺度参数和形状参数。显然，上述分布定义于 $y > u$，并且注意到当 $\xi < 0$ 时，$Y \mid Y > u$ 的支撑集为 $\{z \in \mathbb{R} : u < z \leq u - \sigma/\xi\}$。

早期针对阈值超越的回归形式，将广义帕累托分布（GP）引入广义线性和加性建模框架（例如，Davison 和 Smith，1990年；Chavez-Demoulin 和 Davison，2005），使尺度参数和形状参数能够随协变量平滑变化，包括表示时间或空间位置等复杂非平稳效应的协变量，前提是存在足够多的超越值。然而，经典广义帕累托分布假设响应为连续型。为适应离散性，Shimura (2012年）和Hitz 等(2024年）提出了离散广义帕累托分布（DGP），将广义帕累托尾部建模扩展至计数值结果。DGP 具有

此后，该分布已应用于多种离散尾部问题，包括道路交通事故、疾病传播和医院拥堵（Prieto 等，2014；Daouia 等，2023；Ranjbar等，2022）。令$Y_{d,\ t}$表示一个离散随机变量，代表在时间 t于区 d观测到的故障数量，并令$u_{d,\ t}$表示一个较高的整数阈值。我们关注阈值超越，即满足$Y_{d,\ t} \geq u_{d,\ t}$的观测值。为刻画离散数据的尾部行为，我们假设阈值超越的条件分布可通过对平移变量$Y_{d,\ t} - u_{d,\ t}$应用离散广义帕累托（DGP）分布来表示。

具体而言，条件累积分布函数定义为

$$F_{\mathrm{DGP}}(y \mid Y_{d,t} \geq u_{d,t}) = \Pr(Y_{d,t} - u_{d,t} \leq y \mid Y_{d,t} \geq u_{d,t}) = F_{\mathrm{GP}}(y+1) - F_{\mathrm{GP}}(y), \qquad (2)$$

对于整数 $y \geq 0$，其中 $F_{\mathrm{GP}}(\cdot)$ 表示尺度参数为$\sigma > 0$ 、形状参数为 ξ 的广义帕累托分布。

在这一表述下，$Y_{d、\ t} - u_{d、\ t} \mid Y_{d、\ t} \geq u_{d、\ t}$的支撑集为 $\{0, 1, 2, \dots\}$，当 $\xi \geq 0$时；当 $\{0, 1, 2, \dots, \lfloor -\sigma/\xi \rfloor\}$时， $\xi < 0$，其中 $\lfloor \cdot \rfloor$表示向下取整算子。关联概率质量函数的闭式表达式可由公式（2）直接推出，从而能够通过最大似然估计对尾部模型进行精确的基于似然的推断。这一公式自然适用于位置、尺度和形状广义加性模型（GAMLSS）框架，其中尾部分布的尺度参数被建模为天气预测变量的平滑函数 $\mathbf{x}_{d、\ t}$。估计采用 Wood等人（2016）

2.3 天气与故障预测的贝叶斯模型平均

贝叶斯模型平均已扩展至动力学预测模型，见Raftery 等（2005），其中集合预报系统的每个成员都由一个概率分布表示。这些分布通过加权混合进行组合，以生成单一的密度预测，权重则基于各个成员过去的表现。

令 Y表示待预测的未来故障计数， $\mathcal{D} = \{y_1, \dots, y_n\}$表示长度为 n的训练数据中的观测故障计数。考虑一组 $J+1$ 预测模型 $\{M_0, \dots, M_J\}$（在我们的应用中， $J = 50$），每个模型都针对给定的 $\mathcal{D}$，生成 Y 的估计预测分布^。对于模型 M_j，令 f_j 表示根据其估计预测分布得出的点预测（例如，后验平均值或一个或多个特定分位数）。贝叶斯模型平均根据 $p(Y \mid \mathcal{D}) = \sum_{j=0}^{J} p(Y \mid M_j,\ \mathcal{D})\, p(M_j \mid \mathcal{D})$，对各模型的预测分布进行组合，其中 $p(M_j \mid \mathcal{D})$表示给定观测数据时模型 M_j 的后验概率。贝叶斯模型平均的点预测随后作为平均预测分布^ $p(Y \mid \mathcal{D})$的函数得出，而不是对各个点预测 f_j进行加权平均。

^ 此处的故障预报 f_0 由单个高分辨率确定性预报 ^ ^（HRES）成员得出，而 f_1、.....、f_{50}则对应于 50 个成员的欧洲中期

ˆ天气预报中心（ECMWF）集合预报。当 f 包含多个预测分位数时，为简便起见，可以采用分位数平均；不过，其他组合策略（例如Ranjan 和 Gneiting（2010）提出的贝塔变换线性池化）可能会略微改善概率校准。在提前时长很短时（约 1–2 日），集合预报通常略微过度离散：用于表示大气不确定性的初始扰动可能产生大于实际预报误差的离散度（Leutbecher，2009）。模型会校正这种早期的人为离散度，以避免预测不确定性被夸大。随着预报时段延长，集合离散度通常会与观测变率更加一致。

3 方法论

我们提出了*X-flexForecast*，这是一种概率框架，将用于分布主体部分的 QR 与用于尾部的 DGP 模型相结合。这一构造将非参数条件分位数与参数化尾部表示联系起来，从而在 $Y_{d,t}$ 的整个（离散）支撑集上提供一致的预测分布。

3.1 模型设定

ˆ预测估计量 $F_{d,t}(\mathbf{x}_{d,t})$ 由第2.1节中的分位数回归公式和第2.2节中的 DGP 尾部成分定义。基于 QR 的成分用于刻画直至概率水平 $\alpha_{d\,t}^{(T)}$，的条件分布，而 DGP 成分则对阈值超越 $\geq u_{d,t}$ 建模。对于尾部成分，我们不采用完全非平稳的 DGP （Chavez-Demoulin 和 Davison,2005）。若允许尺度参数和形状参数均随时间或空间平滑变化，则每个序列都需要显著更多的超越值才能确保估计稳定，同时也会大幅增加计算复杂度。因此，我们采用简约设定，仅通过尺度参数引入非平稳性，并使用协变量依赖的广义加性模型将其建模为 $\sigma(\mathbf{x}_{d,t}^{T})$。其中，$\mathbf{x}_{d,t}^{T}$ 表示仅纳入尾部模型的天气预测变量子集，而形状参数ξ 则假设为常数。

对于主体部分，我们在较粗的 $q_\alpha(y_{d,\,t} \mid \mathbf{x}_{d,\,t})$网格上估计条件分位数，其中包含数据稀疏的 m 概率水平 $0 < \alpha_1 < \cdots < \alpha_m < 1$，也包括恶劣天气时段。使用经过最优平滑的分位数损失来估计系数，该方法对稀疏性具有稳健性ˆ(见第2.1节)。随后，通过在估计分位数之间进行线性插值，在 F_{QR} 的离散支撑集上构建预测分布$Y_{d,\,t}$ 我们直接对整数响应进行分位数回归，避免使用抖动处理或临时舍入，因为加入随机扰动可能会增加偏差、降低可解释性并带来额外的计算成本（Machado 和 Silva，2005）。

3.2 估计与预测

由于响应$Y_{d,t}$ 是离散的，基于加性分位数回归的分位数$q_{\alpha_{d,t}^{(T)}}$，$(\mathbf{x}_{d,t})$通常不属于$Y_{d,t}$的整数值支撑集。为体现离散性，我们在离散概率水平上估计多个分位数，并通过线性插值延伸至 $\tilde{\alpha}_{d,t}^{(T)}$，$(\mathbf{x}_{d,t}) = \widehat{F}_{QR}\left(\left\lfloor q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_{d,t})\right\rfloor\right)$，从而定义 $y_{d,t} \in \left[0, \left\lfloor q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_{d,t})\right\rfloor\right]$的预测分布。其余支撑集 $y_{d,t} \in \left[\left\lceil q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_{d,t})\right\rceil, +\infty\right)$则由广义帕累托分布建模。

由此得到的预测分布 $\widehat{F}_{d,t}(\mathbf{x}_{d,t})$ 为

$$\begin{cases} \widehat{F}_{QR}(y_{d,t} \mid \mathbf{x}_{d,t}), & \text{if } y \le \left\lfloor q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_t)\right\rfloor, \\ \tilde{\alpha}_{d,t}^{(T)}(\mathbf{x}_{d,t}) + \left(1 - \tilde{\alpha}_{d,t}^{(T)}(\mathbf{x}_{d,t})\right) F_{DGP}\left(y - \left\lceil q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_{d,t})\right\rceil \mid \xi, \sigma(\mathbf{x}_{d,t})\right), & \text{if } y \ge \left\lceil q_{\alpha_{d,t}^{(T)}}(\mathbf{x}_{d,t})\right\rceil, \end{cases} \tag{3}$$

其中 F_{DGP} 是第 2.2 节所述广义帕累托分布的累积分布函数，而 F_{QR} 表示通过第 3.1 节所述线性插值得到的值。

由于广义帕累托分布拟合的是超过$\left\lceil q_{\alpha_{d,t}^{(T)}\text{的超越值}}(\mathbf{x}_{d,t})\right\rceil$，因此 $\alpha_{d,t}^{(T)}$ 的选择至关重要。虽然通常会使用阈值稳定性诊断（例如评估 ξ 在不同阈值下的稳定性），但我们通过选择样本外预测表现最佳的值来实证确定 $\alpha_{d,t}^{(T)}$。所选水平还必须足够高，以确保广义帕累托近似仍然有效（敏感性分析见附录 C，选择详情见附录 D）。

为生成 k 时间范围的提前预测，集合 数值天气预报 输出提供 j 个$\mathbf{x}_d^{(j)}$ 的预测，$t|t+k$ 这些预测在时间 t 可用。将每个成员的输出输入拟合模型，即可得到各成员特定的预测分布 $\widehat{F}(y_{d,t+k} \mid \hat{\mathbf{x}}_d^{(j)}, t|t+k)$，然后通过贝叶斯模型平均进行组合（见第 2.3 节）。为减轻早期的过度离散并提高短时效的可靠性，我们采用加权分位数平均方案，在短预报时效下提高 HRES/控制成员的权重。在提前时长为 0 小时时，HRES/控制成员的权重设为 50，其余五十个成员的权重各为 1。HRES/控制成员的权重随提前时长线性递减，直至 72 小时，此时其权重与其他成员相同。超过 72 小时后，所有成员的权重相等。

4 实现与基准测试

在本节中，我们开展模拟实验，以模拟苏格兰各区发生故障时观察到的数据生成过程。目标是评估所提出的模型能否为典型事件和极端事件提供准确的预测。我们将该模型与适用于此任务的其他方法进行比较，包括加性分位数回归（下文简称$qgam$）模型（Fasiolo 等人，2021a）以及扩展离散广义帕累托模型（EDGP）；Ahmad等人2024年）。$qgam$ 模型被视为

模拟中的基准模型，因为我们的目标之一是证明，所提出的模型在保持数据主体部分采用无分布假设方法的同时，能够改善尾部估计。相比之下，作为 DGP 框架扩展的 EDGP 模型被纳入比较，作为一种竞争方案。该模型通过对 Y 的整个支撑集进行建模，无需预先设定阈值即可处理离散响应和极端事件。不过，与 *X-flexForecast* 相比，它依赖于全局参数分布，可能缺乏捕捉真实数据中变异性的灵活性。

所有结果均使用所引用软件包中提供的实现生成Fasiolo 等人(2021b）和Ahmad 等人(2024），并采用默认参数设置。我们提出的模型在R 软件包 `xflex4cast` [1] 中实现，该模型依赖于 `qgam` (Fasiolo 等人，2021b）和 `GJRM` (Marra 和 Radice，2024软件包，分别使用 qgam 和 DGP 模型，并通过 GAMLSS 框架进行拟合。

4.1 模拟实验

我们的模拟研究旨在构建这样一种设定：数据在一个已知阈值之下和之上分别由两个不同的分布生成，分别称为“主体部分”和“尾部”。这种方法针对每种情况考虑了两个不同的潜在过程。关于连续域中这一通用方法的全面综述见Dey 和 Yan (2016）。将其扩展到离散响应（如Ahmad 等人(2024）所述，随机变量 Y 的分布为：

$$P_{\text{dg-dgp}}(Y=k)=\begin{cases}P_{\text{dg}}(Y=k\mid\boldsymbol{\theta}_B) & \text{for } k\le u-1\\ (1-P_{\text{dg}}(Y\ge u\mid\boldsymbol{\theta}_B))\times P_{\text{dgp}}(Y=k\mid Y\ge u;\boldsymbol{\theta}_T) & \text{for } k\ge u\end{cases} \tag{4}$$

超越阈值定义为 $u = q_{Y_{\text{dg}}}(1-\phi) = \inf\{k\in\mathbb{N}_0 : F_{\text{dg}}(k\mid\boldsymbol{\theta}_B)\ge 1-\phi\}$，其中 $F_{\text{dg}}(\cdot\mid\boldsymbol{\theta}_B)$ 表示离散伽马主体分布的累积分布函数，$q_{Y_{\text{dg}}}(\zeta)$ 表示其 ζ 分位数。由于主体分布是离散的，其累积分布函数仅在整数值处递增，因此阈值不能设在两个整数之间。因此，ϕ 指定目标超越概率，而 u 是主体累积分布函数达到或超过 $1-\phi$ 时的最小整数。在此设定下，主体分量 $P_{\text{dg}}(Y=k\mid\boldsymbol{\theta}_B)$ 定义于 $k\le u-1$，而尾部成分 $P_{\text{dgp}}(Y=k\mid Y\ge u;\boldsymbol{\theta}_T)$ 则刻画 $k\ge u$ 的超越值。这两个分量共同定义 DG–DGP 混合分布 $P_{\text{dg-dgp}}$。参数集合 $\boldsymbol{\theta}_B$ 和 $\boldsymbol{\theta}_T$ 分别对应控制主体分量和尾部成分的分布参数。对于离散伽马主体分布，$\boldsymbol{\theta}_B=(\kappa,\boldsymbol{\beta}_\lambda)$ 由形状参数 κ 和与尺度参数 $\lambda(\mathbf{z})$ 相关的回归系数向量 $\boldsymbol{\beta}_\lambda$ 构成，其中 $\mathbf{z}$ 表示模拟中使用的协变量向量。对于 DGP 尾部，$\boldsymbol{\theta}_T=(\xi,\boldsymbol{\beta}_\sigma)$ 包含形状参数 ξ 和用于定义尺度参数 $\sigma(\mathbf{z})$ 的回归系数向量 $\boldsymbol{\beta}_\sigma$。

[1]可访问：https://github.com/ddoubleblindpeerreview/xflex4cast。

为评估各模型的性能，计算估计分位数 $\hat{q}_\alpha(\mathbf{x}_{d,t})$ 与其真实值之间的均方根误差（RMSE）。计算范围涵盖一系列概率水平 $\alpha = \{0.05, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99, 0.999, 0.9999\}$，其中较高的概率水平专门用于模拟极端事件的发生。对于下文所述的每种模拟设定，实验均重复进行 1000 次；每种设定都反映了苏格兰各区故障发生中可能观察到的情况。为适应可能无法持续观察到重尾行为的情形，形状参数 $\xi \in \boldsymbol{\theta}_T$ 在集合 $\xi \in \{0.0, 0.03\}$ 中变化，以便灵活刻画尾部行为。情景 $\xi < 0$ 未纳入模拟实验，因为在所研究地区的经验数据中并未出现该情景，因此不属于本模拟研究的范围；为简洁起见，故将其省略。

抽样过程首先从观测到的天气预测变量$\mathbf{z}$ 的经验分布中抽取$\mathbf{x}_{d,\,t}^{T_k}$ 在预测变量的一个子集 T_k 中。给定$\mathbf{z}$后，再从方程（所定义的分布中生成一个抽样值4）。参数 $\boldsymbol{\theta}_B$ 和 $\boldsymbol{\theta}_T$ 根据各模拟情景随$\mathbf{z}$ 变化。下面介绍两项模拟实验，以展示在模型设定正确时，该模型捕捉复杂底层过程的灵活性。包含设定错误的模型的其他实验见附录B。对于*X-flexForecast* 方法，分位数水平 $\alpha^{(T)}$ 设为 $1 - \phi$，这与方程（4中定义的真实数据生成过程相对应。关于该参数敏感性分析的进一步讨论，以及在真实 ϕ 未知的实际情形下的合适取值，请参见附录C。

4.1.1 第一种情景：尾部分布中的尺度不变性

在第一个模拟情景中，我们将离散伽马分布的尺度参数定义为随预测变量$\mathbf{z}^{\{1\}}$变化，该分布刻画了主体部分。参数设定为 $\boldsymbol{\theta}_B = \left\{\alpha = 1.5, \gamma(\mathbf{z}^{\{1\}}) = \exp(\beta_0 + \beta_1 \mathbf{z}^{\{1\}})\right\}$。预测变量$\mathbf{z}^{\{1\}}$ 从一个经验分布中抽样，该分布由某个$\mathrm{ws10}_{\max}$ 地区每隔24小时记录一次的数值构建而成。这样可以确保模拟数据反映真实世界预测变量的分布。系数的真实值设为 $\beta_0 = 1$ 和$\beta_1 = 2$。对于模型的尾部，我们假设采用 $P_{\mathrm{dgp}} \sim$广义帕累托分布 $(\xi \in \{0.0, 0.3\},\ \sigma = 2.5)$，分别表示轻尾和重尾行为。最后，阈值以上观测值所占比例设为 $\phi = 0.1$。

对于*qgam* 方法的*X-flexForecast*模型设定及其主体部分，我们使用基维数为 $k = 10$ 的惩罚样条平滑函数对条件分位数进行建模。采用这一方法，是因为离散伽马分布的分位数函数没有闭式表达式，而惩罚样条的灵活性可以弥补这一局限。尾部模型的设定是正确的，仅包含常数项，

这些常数项仍需估计。对于以 κ、σ 和 ξ 为参数的 EDGP 模型，我们遵循 Ah-mad 等人(2024年的建议，对 κ 和 ξ采用常数模型，而 $\sigma(\mathbf{z}^{\{1\}})$ 则通过惩罚样条进行估计，其中 $k = 10$，以确保与 QR方法保持一致。

结果汇总于表1，其中给出了在模拟实验的1000次重复模拟上计算得到的平均均方根误差。为便于阅读，概率水平低于 $\alpha = 0.95$ 的分位数在第二列和第五列中被省略，因为 *X-flexForecast* 与 *qgam* 模型在该范围内得到相同结果。为简洁起见，$\alpha = 0.9$ 分位数也被排除。结果表明，对于大多数主体分位数，EDGP 模型取得了最小的均方根误差值，尽管这些值与其他方法几乎处于同一数量级。对于主体分位数，这很可能是因为 EDGP 受益于其参数化分布假设所隐含的分位数联合估计，而其他方法则独立估计这些分位数。随着与分位数相关的概率水平升高，*X-flexForecast* 模型相对于竞争方法持续取得明显更低的均方根误差值，这反映了其更大灵活性在尾部预测性能上的优势。这表明它能够在分布的主体部分和尾部区域都提供准确估计，并强调其在捕捉整个预测范围方面比其他方法更具适应性。

	$\xi = 0.0$				$\xi = 0.3$			
α	$\bar{q}(\mathbf{X})$	*X-flexForecast*	*qgam*	EDGP	$\bar{q}(\mathbf{X})$	*X-flexForecast*	*qgam*	EDGP
	RMSE				RMSE			
0.05	0.09	0.60 ± 0.32	—	**0.52 ± 0.06**	0.09	0.60 ± 0.34	—	**0.55 ± 0.12**
0.25	1.41	1.23 ± 0.40	—	**0.96 ± 0.05**	1.42	1.22 ± 0.33	—	**0.98 ± 0.42**
0.50	3.98	1.53 ± 0.50	—	**1.01 ± 0.11**	3.99	1.53 ± 0.54	—	**1.04 ± 0.91**
0.75	8.03	1.94 ± 0.66	—	**1.17 ± 0.30**	8.04	1.93 ± 0.68	—	**1.23 ± 1.71**
0.95	14.52	**2.03 ± 1.13**	2.20 ± 0.77	3.26 ± 0.76	14.98	**2.09 ± 1.07**	2.30 ± 0.67	4.07 ± 3.63
0.99	18.55	**2.14 ± 1.18**	3.52 ± 0.63	11.17 ± 2.16	21.53	**2.50 ± 1.10**	3.91 ± 0.94	12.81 ± 5.93
0.999	24.27	**2.69 ± 1.62**	14.18 ± 4.06	22.68 ± 4.26	38.84	**4.36 ± 2.65**	12.34 ± 5.63	25.10 ± 9.77
0.9999	30.13	**3.91 ± 2.81**	60.07 ± 15.27	34.13 ± 6.39	73.33	**11.99 ± 9.72**	40.33 ± 20.18	40.29 ± 14.03

表1：第一种模拟情景下1000次重复模拟的平均均方根误差。*X-flexForecast*、*qgam*和 EDGP 列展示各方法的均方根误差；$\bar{q}(\mathbf{X})$ 列报告水平为 α 时的平均真实分位数。粗体表示最低均方根误差。*qgam* 的主体分位数因与*X-flexForecast*的主体分位数相同而省略。

4.1.2 第二种情景：尾部分布非平稳

在第二种情景中，通过引入 DGP 尺度参数对协变量的依赖来模拟尾部非平稳性，即允许 σ 随额外的天气协变量 $\mathbf{z}^{\{2\}}$ 变化。这种非平稳性完全由协变量驱动，不涉及空间或时间依赖。此设定模拟了与故障计数相关的极端事件

发生时观察到的现实模式。此时，决定 Y分布的预测变量集定义为 $\mathbf{Z} = \{\mathbf{z}^{\{1\}}, \mathbf{z}^{\{2\}}\}$，其中两个协变量均从根据观测数据估计的联合经验分布中抽样。$\mathbf{z}^{\{1\}}$ 仍表示 $\mathbf{ws10}_{\max}$，而 $\mathbf{z}^{\{2\}}$表示同一苏格兰地区在24小时区间内测得的总降水量的第90分位数。鉴于两者存在相关性，使用其联合经验分布对这两个变量进行联合抽样。尾部尺度参数 σ 通过关系式 $\sigma(\mathbf{z}^{\{2\}}) = \exp\left(\beta_0 + \beta_1 \mathbf{z}^{\{2\}}\right)$建模为 $\mathbf{z}^{\{2\}}$ 的函数，其中参数值为 $\beta_0 = -0.05$ 和$\beta_1 = 0.85$。这些参数值是通过对苏格兰某一地区真实24小时数据中的阈值超越值拟合同一 DGP 模型，并采用最大似然估计获得的。此举旨在尽可能贴近第5节所分析数据集的特征。控制离散伽马分布的所有其余参数保持不变。

将 $qgam$ 的模型设定修改为纳入一组大小为$k = 10$ 的基函数，用于 $\mathbf{z}^{\{2\}}$；并对 EGPD 的尺度参数进行了相同的修改。结果汇总于表2。均方根误差（RMSE）的结果与第1节报告的结果一致，表明 X-*flexForecast* 与其他估计主体分位数的方法具有相当的表现。然而，对于较高分位数，X-*flexForecast* 仍展现出更高的预测准确性，这体现在其平均均方根误差值更低。

	$\xi = 0.0$				$\xi = 0.3$			
α	$\bar{q}(\mathbf{X})$	*X-flexForecast*	*qgam*	EDGP	$\bar{q}(\mathbf{X})$	*X-flexForecast*	*qgam*	EDGP
		RMSE				RMSE		
0.05	0.09	**0.60 ± 0.33**	—	0.62 ± 0.08	0.09	0.61 ± 0.36	—	**0.58 ± 0.08**
0.25	1.41	1.21 ± 0.29	—	**0.99 ± 0.07**	1.41	1.22 ± 0.35	—	**0.98 ± 0.08**
0.50	3.99	1.53 ± 0.52	—	**0.95 ± 0.10**	3.98	1.53 ± 0.53	—	**0.98 ± 0.16**
0.75	8.04	1.95 ± 0.64	—	**1.05 ± 0.30**	8.03	1.91 ± 0.65	—	**1.11 ± 0.37**
0.95	13.31	**1.94 ± 1.08**	2.30 ± 0.62	3.55 ± 0.71	13.52	**2.00 ± 1.06**	2.39 ± 0.62	3.67 ± 0.79
0.99	15.34	**2.10 ± 1.09**	4.44 ± 0.78	12.14 ± 2.01	16.63	**2.37 ± 1.13**	4.50 ± 0.86	11.75 ± 2.07
0.999	17.90	**2.72 ± 1.40**	18.11 ± 4.43	24.93 ± 4.00	24.56	**3.89 ± 2.27**	15.13 ± 4.60	22.23 ± 4.03
0.9999	20.55	**4.01 ± 2.68**	64.54 ± 15.42	37.69 ± 6.02	40.41	**9.38 ± 6.06**	51.87 ± 16.12	31.72 ± 5.45

表2：第二种模拟情景下1000次重复模拟的平均均方根误差。X-*flexForecast*、$qgam$和EDGP列显示各方法的均方根误差； $\bar{q}(\mathbf{X})$列报告 α水平下的平均真实分位数。粗体表示最低均方根误差。省略了$qgam$ 的主体分位数，因为它们与X-*flexForecast*的相同。

5 案例研究：苏格兰电力网络的短期故障预测

我们提出这种新型灵活预测方法，旨在为最终用户提供量身定制的短期故障预测。具体而言，该方法致力于为各个

电力配电网络中的地区（组织单元，面积约为 1000 km^2）生成概率故障预报。这些预测最长可提前 4 天，分辨率为 24小时分辨率。该方法明确考虑故障分布中可能出现的重尾行为。其设定符合最终用户的需求：他们需要提前一天至几天调配资源，并实施稳健的风险管理，因此需要准确量化不确定性。

5.1 数据描述

故障的成因有很多，并非只有恶劣天气；然而，故障数量较多的时段通常对应于恶劣天气，包括风暴、雨雪和冰冻以及极端高温。为此，我们将详细的电网数据与高分辨率的观测和预报天气变量相结合，以建立覆盖由SPEN管理的两个许可证区域的十一个区级故障预测模型；这两个区域位于英国两个不同地区，采用日时间分辨率，时间范围为2010年3月至2024年4月。

5.1.1 电网与故障数据

预测给定时段和区内的故障数量是本研究的目标；因此，建模和预测评估都需要故障记录。故障记录是一项监管要求，因此记录得到妥善维护。然而，这些记录并非为预测目的而生成，需要仔细清洗。一个常见问题是，故障是在被发现时记录的，而这未必是其发生的时间。例如，由风暴造成的故障可能数日未被察觉，直到对电网受损部分进行修复时才显现出来。时间戳并不能准确记录故障发生的精确时间，因此不应被过度解读。最一致且可靠的信息记录在电路级（最小的运行单元），尽管更精确的地理位置仅对少数事件可用。我们将电路级故障计数汇总到24小时窗口中，然后在每个区内对这些计数求和（每个区包含许多电路，通常为数百条）。所有建模和预测均在地区级进行：我们分析中的每个观测值都对应一个区—日故障数。其中11个地区属于SPEN的配电网，如图2所示。

5.1.2 天气数据

对于故障建模，使用ERA5再分析数据（Hersbach 等，2025）在针对每个电路和区优化的一组位置上估计天气协变量，并对逐小时场应用双线性插值。所考虑的变量包括10米风速和

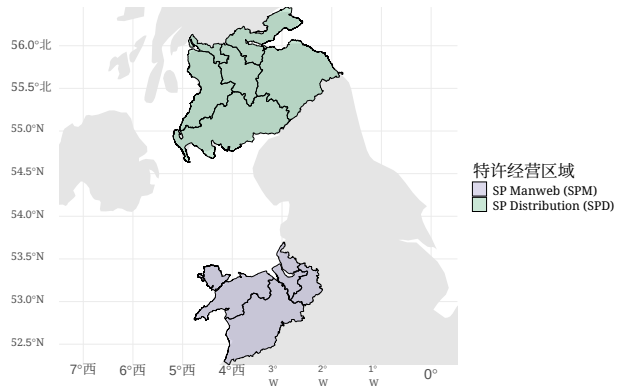


图 2：英国地图，显示 SPEN 运营的两个许可证区域：SP Distribution（SPD）覆盖苏格兰中部和南部（上方，绿色），SP Manweb（SPM）覆盖北威尔士、默西塞德和柴郡（下方，紫色）。许可证区域内标示了区边界——这些空间单元用于汇总故障，以便进行建模和预测。每个区由数百条电路组成。

阵风、2米气温和露点、总降水量、雪深、土壤湿度以及对流有效位能。为与故障计数的日聚合保持一致，计算各区级别的24小时平均值和最大值。这种时间和空间汇总还有助于解释并减轻天气协变量在原始逐小时和空间分辨率下潜在的时空依赖性。

利用这些汇总后的协变量训练模型，随后将拟合后的模型应用于欧洲中期天气预报中心（ECMWF）的高分辨率（HRES）确定性预报和 50 成员集合预报的后处理数值天气预报，从而获得故障预报。对集合成员中的气象变量进行一致的提取和汇总，并使用基于历史数据拟合的线性模型进行偏差订正，最终得到每日分辨率、提前时长最长为四天的 51 成员输入集合。最终的概率故障预报采用第 2.3 节所述的贝叶斯模型平均方法生成。

5.1.3 数据探索与整合

我们在电路级对电网与故障数据进行了探索，但初步建模实验表明，预测单个电路的故障极具挑战性。研究发现，尽管某些电路由于海拔和植被因素更容易发生故障，且总体故障数量更多，但在恶劣天气期间，其故障数量与其他电路相近。恶劣天气期间地区级观测到的大量故障，是由许多电路各自发生少量故障造成的，而非易受影响的电路发生了大量故障。由于较不易受影响的电路故障数据尤为稀疏，因此无法对这些故障的发生情况进行建模。因此，我们旨在预测

利用根据各个电路对应的天气观测位置数据构建的工程化特征（协变量），预测地区级故障。

5.2 结果

上述方法论已应用于一项案例研究，该研究涵盖了SPEN在大不列颠配电网络中的11个地区中的10个，这些网络覆盖苏格兰南部以及英格兰西北部和北威尔士的部分地区（见图2）。由于故障计数较少，结果缺乏信息量，因此本分析未纳入格拉斯哥与克莱德北部。模型选择与评估基于2010年4月至2024年3月的历史故障数据和ERA5天气再分析数据。研究采用包含14折的交叉验证设置，每折覆盖一个完整的4月至次年3月年度周期，确保每折仅包含一个监管年度，即一个完整的冬季，从而与SPEN的业务管理相一致，并减轻各折之间时间依赖可能造成的偏差。评估预测时，使用从欧洲中期天气预报中心获取的历史天气预报：高分辨率预报模型的数据取自业务档案，集合预报系统（EPS）的数据取自TIGGE档案。历史预报覆盖2018年4月至2024年3月，包括一次预报起报（基准时间）——协调世界时午夜，以及从0至96小时的提前时长，并采用可用的最高时间分辨率。

在实际运行中，给定区和时间分辨率下的预测故障数量表示为分类变量 $c_{d,\,t} \in \{R, A, G\}$，分别对应红色、琥珀色和绿色（RAG）故障等级。定义红色—琥珀色和琥珀色—绿色界限的阈值会随区和时间分辨率而变化，其中 $\tau_{d\,t}^{(RA)}{}_{,}$ 和 $\tau_{(AG)\,d,\,t}$ 分别表示各区和时间尺度下的这些转换界限，因为风险管理策略会因地点和时间间隔而异。对于给定观测，使用方程（3）计算 $\widehat{F}_d(\tau_d^{(RA)}{}_{,\,t} \mid \mathbf{x}_{d,\,t})$ 和 $\widehat{F}_d(\tau_d^{(AG)}{}_{,\,t} \mid \mathbf{x}_{d,\,t})$，以确定各类别的概率，随后据此推导各故障等级的最终概率。此外，为了向最终用户传达信息，将类别 $c_{d,\,t}$ 分配给观测 $\mathbf{x}_{d,\,t}$ 时，不仅依据概率最高的类别，还遵循一套启发式规则。具体而言，若绿色类别的概率超过 80%，则分配为绿色类别；若红色类别的概率超过 20%，则分配为红色类别；若前两项条件均不满足，但琥珀色类别的概率高于红色类别，则分配为琥珀色类别。故障等级以及分类预测的分配逻辑，是通过与最终用户进行结构化讨论并开展测试后选定的，以符合其实际运行需求。

在 $qgam$ 模型中，10米风速、表示树木是否长有树叶的指示变量、盛行风向和对流有效位能被视为模型选择期间的候选天气协变量。模型纳入了线性项或三次样条项，每个地区使用10至15个基函数，以确保在捕捉非线性关系方面具备足够的灵活性-

性关系。对于特定地区，除风速之外还纳入树叶、风向和闪电代理变量（对流有效位能）等协变量，这与近期研究结果一致；这些研究报告称，其中一些预测变量与英国的停电之间存在显著关联（Manning 等，2025）。

模型选择依据 k 折交叉验证框架下的样本外（OOS）表现，采用针球损失、锐度、布里尔分数（BS）和 ROC 曲线下面积（AUC）进行评估。基于 OOS 集合计算的 BS 和 AUC，用于选择协变量集合 $\mathbf{x}_{d,\,t}$ 用于拟合 DGP 尾部参数的变量，以及转移概率水平$\alpha_{d\,t}^{(T)}$，，并考虑分类变量 $\mathbf{c}_{d,\,t}$ 对预报可用性的作用。 $qgam$组件中的分位数在以下概率水平进行估计 $\boldsymbol{\alpha} = \{0.05, 0.25, 0.5, \alpha_{d,t}^{(T)}\}$，其中 $\alpha_{d}^{(T)}{}_{,\,t} \in \{0.75, 0.8, 0.9, 0.95\}$。在多个候选设定中，优先选择在各类别上平均而言能最大幅度改善 BS 和 AUC 的配置；当两个标准不一致时，优先采用 BS 作为主要选择度量指标，因为它是严格适当的评分规则，可评估校准度和准确性，而 AUC 主要反映排序性能（Gneiting和Raftery，2007），大体如此。各尾部模型的最终变量选择、 $\alpha_{d\,t}^{(T)}{}_{,}$ 以及相关标准汇总于附录D。

5.2.1 模型评估

通过检验在k折交叉验证过程中基于ERA5再分析生成的后报的可靠性图，验证了故障预测模型的校准。由于故障取整数值且呈右偏，评估校准具有挑战性，因此我们考虑当$\mathbf{x}_{d,\,t}$ 包含处于其最高五分位数的风速时，故障预测的条件可靠性。这些图每个区一张，显示于图3 中的 *X-flexForecast* 模型。24小时分辨率预测表现出良好的校准，尽管一些故障发生不频繁的地区在 $\alpha_{0.25}$分位数上校准较差。这是由于零故障边界导致概率质量集中于零。

为评估将广义帕累托分布模型纳入*X-flexForecast* 框架所带来的模型性能提升，我们计算了各区本方法与$qgam$ 基线之间AUC值的比值。此外，以$qgam$ 作为参考模型计算了BS技巧评分（技巧评分）。

图4 总结了所有地区在24小时分辨率下的结果。左图中，AUC比值大于1表示性能有所提升，小于1则表示与仅使用$qgam$的方法相比性能有所下降。右图中，大于零的值表示 *X-flexForecast* 在预测各类别概率方面有所提升 C，小于零的值则表示性能有所下降。在左右两个面板中，虚线表示参考值，意味着性能没有变化

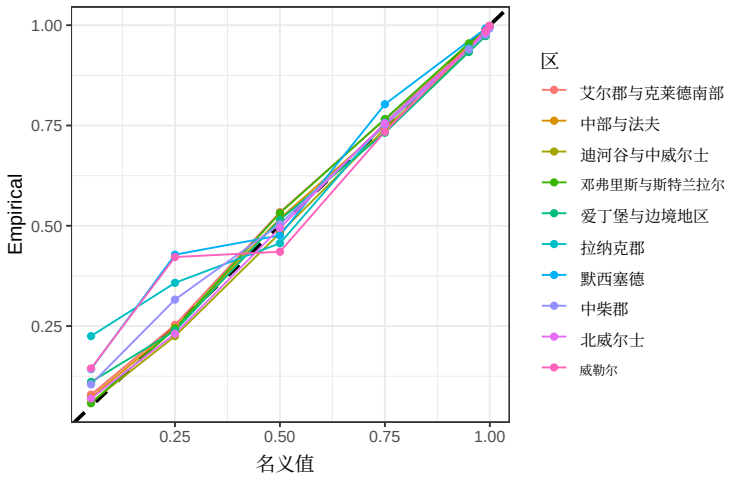


图3：针对风力最大的20%天数，使用 ERA5 进行样本外后报的 X-flexForecast 模型在24小时分辨率预报中的可靠性。

相较于所比较的方法。

分析左图和右图可见，大多数数值都位于参考线上方，这表明将尾部模型与$qgam$ 结合使用，可以改进对故障带的估计。对于某些地区——例如中部与法夫、中柴郡和威勒尔——低于参考线的数值表明红色类别可能表现欠佳。不过，鉴于这些地区红色事件的发生频率较低，将部分红色事件误分类为琥珀色并不一定意味着模型失效，因为估计的正确分类概率仍可能具有参考价值。因此，仅依赖 AUC 并不足够：模型$\widehat{}$生成 $F_{d,\,t}(\mathbf{x}_{d,\,t})$，其评估还需要采用适用于概率预测的评分规则。右图报告了布里尔分数（BS）的技巧评分（SS），其中大多数地区位于参考线右侧，只有迪河谷与中威尔士、拉纳克郡和北威尔士的部分类别略低于参考线。结合左图中的 AUC 比值来看，整体模式仍然支持所提出的方法。因此，图4表明，平均而言，即使在故障带偶尔被误分类的地区，$X\text{-}flexForecast$ 生成的概率预报也比标准$qgam$所得预报更符合观测到的类别。附录D提供了阈值加权连续分级概率评分的补充结果。该评分也能反映完整概率预报的质量，并同时给出微观和宏观 AUC 值。

5.2.2 预测评估

针对2017年4月至2024年3月期间生成了故障再预报，使用历史数值天气预报 以及 $X\text{-}flexForecast$ ，其通过交叉验证训练，训练数据排除了正在进行再预报的时段。针对多模型的每个成员生成故障预报

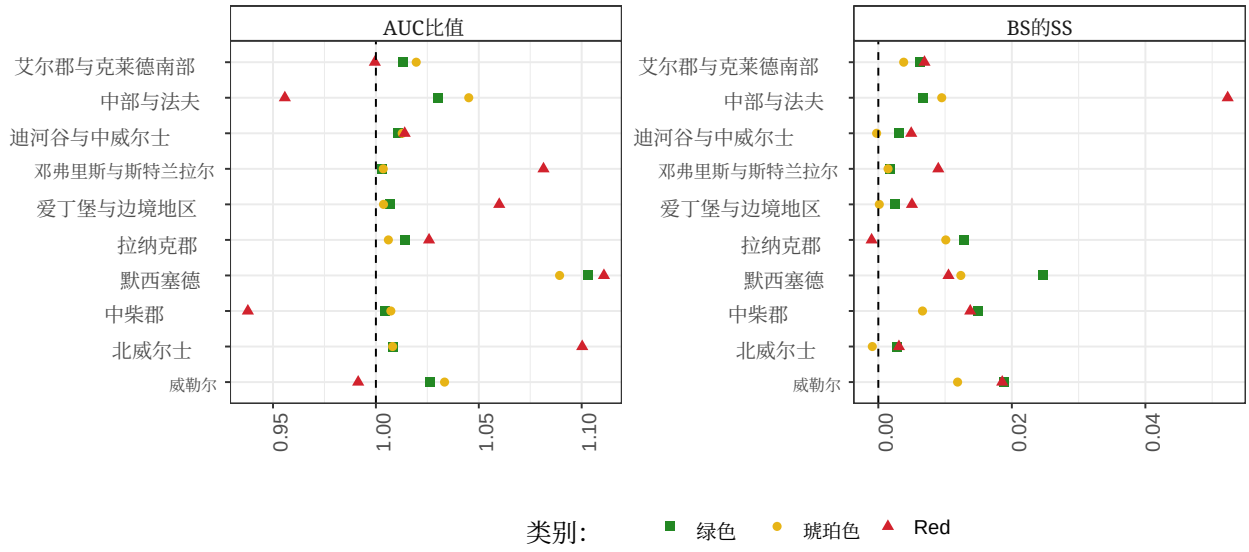


图4：所有地区的24小时分辨率结果。左图：AUC比值 > 1 表明，相较于仅使用$qgam$的方法有所改进。右图：数值 >0 表明，*X-flexForecast* 在预测类别显示概率方面有所改进。虚线表示性能没有变化。

集合，其中包括一个高分辨率预报成员和 50 个 EPS成员，并采用 贝叶斯模型平均（BMA）以及第 3.2 节所述的加权平均方法进行组合。作为比较，还生成了仅使用EPS和仅使用高分辨率预报的再预报。

所得故障预报的校准情况已得到验证，因为这些模型最初以大风日为条件。图5所示的24小时分辨率预报结果证实，大多数地区和预报时效（未显示）均实现了校准。HRES与 EPS 的结合改善了大多数地区的校准情况。然而，对于三个地区而言，这种结合并未解决、也未预期能够解决其较差的校准表现；此前这些地区的故障模型已显示出缺陷。

整体故障预报性能采用七个分位数的Pinball损失进行评估 $\alpha = \{0.05, 0.25, 0.5, 0.75, 0.95, 0.99, 0.999\}$。图6 展示了24小时*X-flexForecast* 预报按预报时效和地区划分的缩放Pinball损失。Pinball评分以各地区在预报时效0时的EPS+HRES值为基准进行缩放。由于地区间存在异质性且事件频率较低，地区间比较受到限制，但结果证实，预报质量会随预报时效延长而小幅下降。EPS+HRES始终优于仅使用EPS和仅使用HRES的方案，尤其是在较短的预报时效（0–48小时）内；而仅使用HRES的方案在大多数地区表现出最高的损失，凸显了单独使用该方案时可靠性较低。

除上述定量评估外，SPEN的工程师还对这些预报的可用性和实用性进行了定性评估。该预报框架已由 Sia 开发为网络应用程序，并于十月起在 SPEN 进行试用

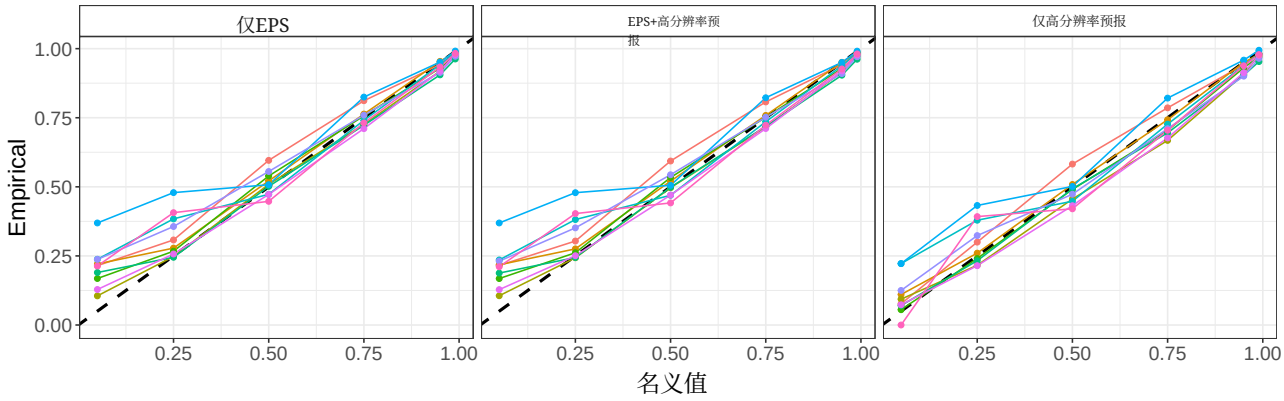


图5：*qgam* 模型的可靠性，预报时效为提前零至四日的样本外再预报，使用欧洲中期天气预报中心的EPS 和/或高分辨率预报确定性预报，针对风力最大的20%天数。各区的颜色同图3。

2024年，直至本文撰写之时。在此期间，工程师报告称，该工具通过提供额外的定量证据，帮助他们在不确定情况下快速决策，从而提升了他们在极端天气事件来临前的态势感知能力和运行准备状态。在多次具体的风暴应对中，该工具改善了电网干预措施的协调和优先级排序，显著提高了运行效率，并缩短了停电持续时间。总体而言，该解决方案的准确性足以为运行决策提供依据；与常规业务相比，它为 SPEN 和电力消费者创造了可衡量的价值。

6 结论

对电力配电网故障进行预测，是保障电力供应韧性的一项关键能力。我们开发了一种故障建模与预测方法，用于量化影响故障预测的两大不确定性来源：恶劣天气导致故障的倾向，以及未来天气条件。此外，*X-flexForecast* 提供的解决方案专门考虑了极端事件的发生以及离散响应。由此得到的概率故障预报经过了定量和定性验证与评估。结果表明，这些预报质量很高，适用于运营决策。

这些结果构成了预报能力的概念验证，展示了未来进行业务化部署的潜力，同时也指出了进一步改进的方向。其中一个方向涉及阈值概率水平的选择 $\alpha_{d\,t}^{(T)}$，$(\mathbf{x}_{d,\,t})$。本研究通过交叉验证确定该参数；不过，也可以借鉴Murphy等人(2024年的方法，并将其调整以适用于 *X-flexForecast* 框架，从而制定自动选择策略，

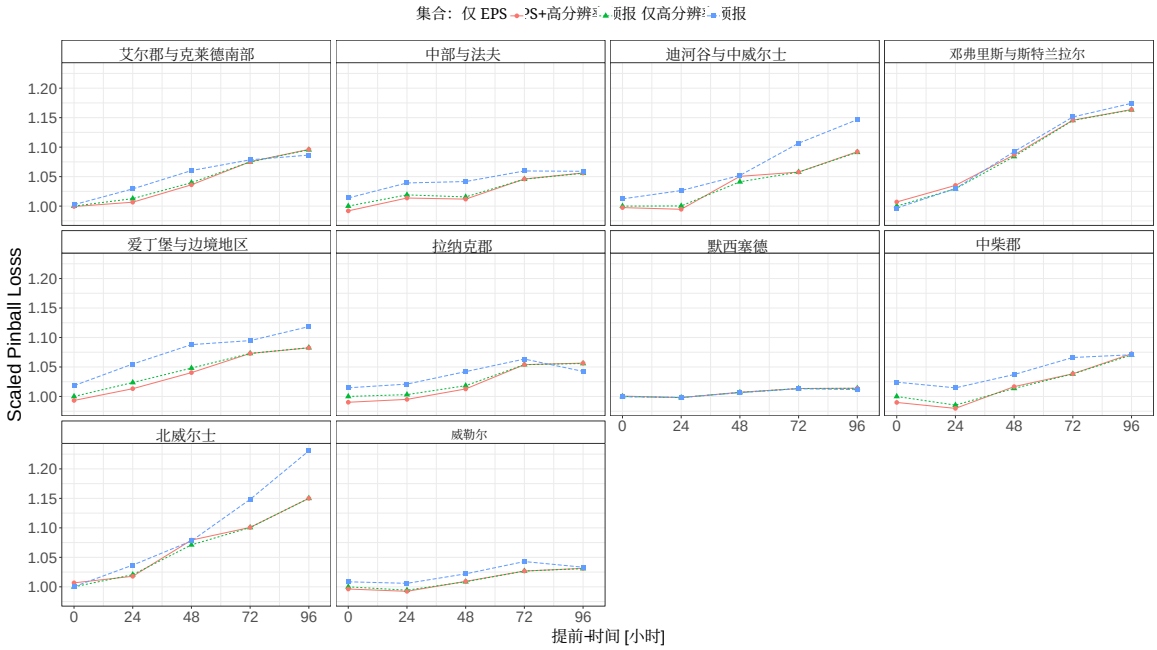


图6：基于提前时长为0至96小时（24小时分辨率）的样本外再预报，使用 EPS 和/或高分辨率预报计算的*X-flexForecast* 模型的缩放弹球损失。各区的评分均以提前时长为0时的 EPS+HRES 预测为基准进行缩放。

以减轻模型估计和部署所需的计算负担，从而提高实施效率。还可以进行其他扩展，但这些扩展并非易事，且超出了本文的范围。其中包括加入目前故障模型尚未使用的协变量（例如时变植被），将方法推广到处于不同气候条件下的网络，扩展多模型集合，使模型适应数值天气预报的变化，探索替代的插值方案，以便根据多个分位数回归构建 F，并优化或替换基于贝叶斯模型平均的预报组合。未来的工作还可在加性预测变量中引入明确的时空项，包括尾部成分（Chavez-Demoulin 和 Davison，2005），以刻画地区之间及时间上的依赖关系，但这会增加模型复杂度。

7 致谢

我们谨向 Abdelouahab Mestassi、Jeremy Joslove、Kitti Kovacs、L´eopoldDuverger-Noinski、Milan Clerc、Miao Zhou、S´ebastien Gerber、Simon Hessami、Simone de Angelis、来自 Sia 的 Wilbur Zhu，以及SPEN 和英国气象局致谢，感谢他们为本研究所作的贡献。我们还感谢使用 ERA5-Land 数据（Hersbach 等人，2025）、ECMWF HRES 和集合预报。本研究是*Predict4Resilience* 项目的一部分，该项目由英国天然气和电力市场办公室及英国研究与创新署管理的战略创新基金资助。

本研究使用的天气数据可从欧洲中期天气预报中心获取：ERA5 数据可从气候数据存储库获取[2]，TIGGE档案中的历史 ECMWF 集合预报[3]，ECMWF业务档案库中的历史 ECMWF确定性/HRES预报[4]。电力网络故障数据的使用受到限制；本研究获许可使用了这些数据。如需获取数据，请联系作者并征得SPEN的许可。第4.1 节所述模拟实验的复现代码可在 GitHub 上获取[5]，并已存档于 Zenodo[6]。

我们没有需要声明的利益冲突。为实现开放获取，作者已将知识共享署名许可（CC BY）授予本次投稿产生的任何作者接受稿版本。

参考文献

Ahmad, T., Gaetan, C. 和Naveau, P.（2024年）一种用于计数数据的扩展广义帕累托回归模型。《统计建模》。

Allen, S.（2024年）加权评分规则：评估概率预测时强调特定结果。《统计软件杂志》110，1–26。

Alpay, B. A., Wanik, D., Watson, P., Cerrai, D., Liang, G. 和 Anagnostou, E.（2020年）雷暴引发停电的动态建模。《预测》*2020*年，第*2*卷，第*151-162*页**2**（2），151–162。

Brester, C., Niska, H., Ciszek, R. 和 Kolehmainen, M.（2020年）利用人工神经网络进行基于天气的电力网络故障预测。载于*2020*年*IEEE*进化计算大会（*CEC*），第1–8页。

Browell, J. 和 Fasiolo, M.（2021）利用条件极值和网格化数值天气预报对区域净负荷进行概率预测。《IEEE智能电网汇刊》12（6），5011–5019。

Cerrai, D.、Wanik, D. W.、Bhuiyan, M. A. E.、Zhang, X.、Yang, J.、Frediani, M. E. 和 Anagnostou, E. N.（2019）通过天气和植被的新表征预测风暴引发的停电。《IEEE开放获取期刊》7，29639–29654。

Cha, Y., White, C. J., Gonzalez, P. L., Wallace, E., Harkin, D., Brett, L., Mattu, K., Smith, R. S., Thom, C., Alix, M. D. 等 (2025)评估自然灾害的级联影响

[2]https://cds.climate.copernicus.eu/[3]
https://apps.ecmwf.int/datasets/data/tigge/[4]
https://www.ecmwf.int/en/forecasts/dataset/operational-archive[5]获取地址：
https://github.com/ddoubleblindpeerreview/xflex4cast。[6]DOI:
10.5281/zenodo.18173632

对关键国家基础设施（关键国家基础设施）的影响，并以苏格兰作为案例研究。 *npj* 《自然灾害》2(1), 108. Chavez-Demoulin, V. 和 Davison, A. C. (2005) 样本极值的广义加性建模。《皇家统计学会会刊》*C*辑：应用统计 54(1), 207 – 222. Coles, S. (2001) 极值统计建模导论。伦敦：施普林格。 Daouia, A.、Stupfler, G. 和 Usseglio-Carleve, A. (2023) 使用离散广义帕累托分布对SARS-CoV-2社区传播进行极值建模。《英国皇家学会开放科学》10(3), 220977. Davison, A. C. 和 Smith, R. L. (1990年) 高阈值超越模型。《皇家统计学会会刊》*B*辑（方法论） 52(3), 393–442. Dey, D. K. 和 Yan, J. (2016) 极值建模与风险分析：方法与应用。CRC出版社。 Fasiolo, M.、Wood, S. N.、Zaffran, M.、Nedellec, R. 和 Goude, Y. (2021a) 快速校准加性分位数回归。 美国统计学会杂志 116(535), 1402– 1412. Fasiolo, M.、Wood, S. N.、Zaffran, M.、Nedellec, R. 和 Goude, Y. (2021b) qgam：在R中进行贝叶斯非参数分位数回归建模。《统计软件杂志》 100(9), 1–31. Ghaemi, A.、Safari, A.、Quteishat, A. 和 Younis, M. A. (2024年) 不同天气条件下输电线路基于堆叠的故障预测研究。《能源》 306, 132572. Gneiting, T. 和 Raftery, A. E. (2007) 严格适当评分规则、预测与估计。美国统计学会杂志 102(477), 359–378. Gneiting, T. and Ranjan, R. (2011年) 使用阈值加权与分位数加权评分规则比较密度预测。《商业与经济统计杂志》29(3), 411–422. Gon˙calves, C.、Cavalcante, L.、Brito, M.、Bessa, R. J. 和 Gama, J. (2021) 风能条件极端分位数预测。《电力系统研究》 190, 106636. Hastie, T. J. 和 Tibshirani, R. J. (1990年) 广义加性模型。伦敦：查普曼与霍尔。

Hersbach, H.、Bell, B.、Berrisford, P.、Biavati, G.、Hor´anyi, A.、Mu˜noz Sabater, J.、Nicolas, J.、Peubey, C.、Radu, R.、Rozum, I.、Schepers, D.、Simmons, A.、Soci, C.、Dee, D. 和 Th´epaut, J.-N.（2025）1940年至今的ERA5单层逐小时时间序列数据。访问日期：2025-04-07。Hitz, A. S.、Davis, R. A. 和 Samorodnitsky, G.（2024）离散极值。数据科学杂志 **22**（4），524–536。Huang, X. 和 Wang, N.（2025）热带气旋期间输电系统风险预测的概率微尺度风场建模。结构安全 第102620页。Jantre, S.（2023）纵向计数数据的贝叶斯分位数回归。统计计算与模拟杂志 **93**（1），103–127。Koenker, R.（2005）分位数回归。计量经济学会专著。英国剑桥：剑桥大学出版社。Koenker, R. 和 Bassett Jr, G.（1978）回归分位数。《计量经济学》：计量经济学会杂志 第33–50页。Koenker, R.、Leorato, S.、Peracchi, F. 等 （2013）分布回归与分位数回归。*CEIS*研究论文 **300**。Leutbecher, M.（2009）集合预报系统的诊断。载于预报与资料同化系统诊断研讨会，第235–266页。Machado, J. A. F. 和 Silva, J. S.（2005）计数分位数。美国统计学会杂志 **100**（472），1226–1237。Manning, C.、Wilkinson, S.、Fowler, H. J. 和 Kendon, E. J.（2025）前期降雨、风向和季节效应可能会放大英国风致停电的风险。《地球与环境通讯》 **6**（1），217。Marra, G. 和 Radice, R.（2024）*GJRM*：广义联合回归建模（*R* 包版本 *0.2-6.7*）。Murphy, C.、Tawn, J. A. 和 Varty, Z.（2024）单变量极值的自动阈值选择及相关推断不确定性。《技术计量学》 第1–10页。Prieto, F.、G´omez-D´eniz, E. 和 Sarabia, J. M.（2014）使用离散广义帕累托分布对道路事故黑点数据进行建模。《事故分析与预防》 **71**，38–49。Raftery, A. E.、Gneiting, T.、Balabdaoui, F. 和 Polakowski, M.（2005）使用贝叶斯模型平均校准预报集合。《月度天气评论》 **133**（5），1155–1174。

Ranjan, R. 和 Gneiting, T.（2010）组合概率预测。英国皇家统计学会会刊*B*辑：统计方法 **72**（1），71–91。Ranjbar, S.、Cantoni, E.、Chavez-Demoulin, V.、Marra, G.、Radice, R. 和 Jaton, K.（2022）季节性病毒与医院拥堵的极值建模：以瑞士一家医院的流感为例。《皇家统计学会会刊》*C*辑：应用统计 **71**（4），884–905。Ruiz-Abell´on, M. C.、Fern´andez-Jim´enez, L. A.、Guillam´on, A. 和 Gabald´on, A.（2024）需求响应中概率预测的应用。应用科学 **14**（21），9716。Shimura, T.（2012）最大吸引域中分布的离散化。极值 **15**，299–317。Souto, L.、Neal, R.、Pope, J. O.、Gonzalez, P. L.、Wilkinson, J. 和 Taylor, P. C.（2024）识别可能导致英国停电的天气模式及其转变。《地球与环境通讯》 **5**（1），49。Tsioumpri, E.、Stephen, B. 和 McArthur, S. D. J.（2021）对监测程度最低的配电网络进行天气相关故障预测。能源 **14**（8），2053。Waldmann, E.、Kneib, T.、Yue, Y. R.、Lang, S. 和 Flexeder, C.（2013）贝叶斯半参数加性分位数回归。《统计建模》 **13**（3），223–252。Watson, P. L.、Cerrai, D.、Koukoula, M.、Wanik, D. W. 和 Anagnostou, E.（2020）具有增长域的天气相关停电模型：结构、性能与普适性。《工程学杂志》 **2020**（10），817–826。Wilkinson, S.、Dunn, S.、Adams, R.、Kirchner-Bossi, N.、Fowler, H. J.、Gonz´alez Ot´alora, S.、Pritchard, D.、Mendes, J.、Palin, E. J. 和 Chan, S. C.（2022）后果预测：预测临近风暴后果的理性框架。气候风险管理 **35**，100412。Wood, S. N.、Pya, N. 和 S¨afken, B.（2016）一般平滑模型的平滑参数与模型选择。美国统计学会杂志 **111**（516），1548–1563。Zhang, Y.、Wang, J. 和 Wang, X.（2014）风力发电概率预测综述。可再生与可持续能源评论 **32**，255–270。Zhu, J. 和 He, Y.（2024）用于风力发电概率预测的大规模多目标进化分位数估计模型。*IEEE*进化计算汇刊 。

附录

本附录介绍真实数据应用中采用的补充模拟实验和详细模型设定。附录 A 首先介绍全文所使用的记号，以便查阅。附录 B 部分考察天气相关协变量设定错误时 *X-flexForecast* 方法的稳健性。尽管存在模型设定错误，*X-flexForecast* 仍能生成连贯的预测分布预测。附录C 进一步介绍 *X-flexForecast* 在概率水平的不同选择下的表现 $\alpha^{(T)}$，该概率水平对应于主体分位数回归组件与尾部之间的阈值。最后，附录 D 概述了用于选择过渡阈值的交叉验证程序 $\alpha_{d\,t}^{(T)}$，以及第 5.2 节中所述 *X-flexForecast* 模型的数据生成过程组件协变量集 5.2，并介绍了用于模型评估的其他指标，包括宏微观AUC和阈值加权连续排序概率评分。

A 记号和评估指标

本附录汇总第 2 节引入的记号，以及第 4 节和第 5 节中使用的评估指标定义。除非另有说明，索引表示区 d 和时间 t。粗体符号表示向量；花体字母表示集合或集合类。对于引用文献中的记号，建议核对原始文献

表A.1：全文使用的记号。

符号	描述
索引，*sets*，和数据	
$d \in \{1, \ldots, D\}$	区索引。
$t \in \{1, \ldots, T\}$	时间索引（日或提前时间戳）。
$\mathcal{D}$	涵盖所有地区和时间的训练数据集。
响应和协变量	
$Y_{d,t} \in \{0, 1, 2, \ldots\}$	表示（d，t）处故障计数的随机变量。
$y_{d,t} \in \{0, 1, 2, \ldots\}$	位于 (d, t) 的观测故障计数。
$\mathbf{x}_{d,t}$	位于 (d, t) 的协变量向量（天气特征）。
$\mathbf{x}_{d,t}^{S_k}$	协变量子集$\mathbf{x}_{d,t}$ 从 S_k 索引集合中提取。
$\mathbf{A}_{d,t}$	是一个包含该协变量子集的列向量
预测分布和分位数	
$F(Y_{d,\,t} \mid \mathbf{x}_{d,\,t})$	给定$\mathbf{x}_{d,\,t}$时$Y_{d,\,t}$ 的条件累积分布函数。

（续见下页）

符号	描述
$q_\alpha(Y_{d,\ t}\mid \mathbf{x}_{d,\ t})$	水平为 $\alpha \in [0,1$ 的条件 α 分位数）。
$\widehat{F}_{QR}$	拟合的加性分位数回归（主体部分）诱导的累积分布函数。
加性分位数回归（主体部分）	
$\beta_{\alpha,0}$	分位数水平 α 处的截距。
$\boldsymbol{\beta}_{\alpha,A}$	系数：$\mathbf{A}_d$ $_{,t}$ 水平 α。
$f_{\alpha,\ k}\ (\mathbf{x}_{d,t}^{S_k})$	平滑分量 k 在水平 α 处。
$b_{k,j}(\cdot),\ K_k$	平滑函数的基函数 $j = 1, \ldots, K_k$ k。
$\gamma_{\alpha,k,j}$	层级 α处的基函数系数 $b_{k,j}$ 。
$g^{-1}(\cdot)$	加性预测变量中使用的逆连接函数。
$\lambda > 0$	平滑针球损失的平滑参数。
$\boldsymbol{\theta}_B$	主体部分（分位数回归/广义加性模型）组件的参数集。
尾部建模：广义帕累托分布（*GP*）和离散广义帕累托分布（*DGP*）	
$u_{d,t}$	用于计算超越值的阈值（$d,\ t$）。
$\alpha_{d,t}^{(T)}$	混合/阈值分位数水平。
$\phi \in (0,\ 1)$	目标超越概率。
$\sigma > 0,\ \xi$	分别来自尾部模型的尺度参数和形状参数。
$\mathbf{x}_{d,t}^{T_k}$	仅使用的天气预测变量子集 σ（$\mathbf{x}_d^T$ $_{,t}$）表示为由 T_k 索引。
$F_{\mathrm{GP}}\ (\cdot)$	连续广义帕累托分布累积分布函数。
$F_{\mathrm{DGP}}(\cdot)$	整数支撑上的离散广义帕累托分布累积分布函数。
$\lfloor\cdot\rfloor,\ \lceil\cdot\rceil$	向下取整和向上取整算子。
$\boldsymbol{\theta}_T$	尾部（离散广义帕累托）分量的参数集。
X-flexForecast	
$\widehat{F}_{d,t}$	来自*X-flexForecast* 在（d,t）处的预测累积分布函数。
$\tilde{\alpha}_{,t}^{(T)\,d}(\mathbf{x}_{d,\ t})$	^对应于 $\lfloor q_{\alpha^{(T)}}$的 F_{QR} 累积分布函数质量 $_{d,t}(\mathbf{x}_{d,\ t})\rfloor$。
HRES	确定性高分辨率预报。
$\hat{f}_j$	由拟合模型 j 得到的估计值。
$\tau_{d,t}^{(AG)}; \tau_{d,t}^{(RA)}$	故障计数 阈值 定义 绿色/琥珀色 and 分别为红色/琥珀色区间。
分位数与损失记号	
$\rho_\alpha(u)$	Pinball损失：$u\big(\alpha - \mathbb{I}\{u < 0\}\big)$。

（下页续）

符号	描述
天气协变量示例（$\mathbf{x}_{d,\ t}$ 的元素）	
$t2m$	2米气温（K）。
tp	总降水量（mm）
$ws10_{\max}$	10米最大风速（m/s）。
$cape_mean_ge100_covar$	平均对流有效位能超过 ≥ 100 J/kg。
tp_{q90}	集合成员降水量的第90百分位数。

表A.2：模拟和案例研究中使用的指标。

度量指标	定义／用途
针球损失（分位数）Loss	对于残差 $u = y - q_\alpha$, $\rho_\alpha(u) = u\big(\alpha - \mathbb{I}\{u < 0\}\big)$。用于训练并评估主体模型的分位数预测。
平滑的针球损失 Loss	采用平滑参数 $\lambda >$ 的可微变体 0；当 $\lambda \to 0$ （参见Fasiolo 等人 (2021a)).
RMSE	点预测的均方根误差（例如，贝叶斯模型平均值）：均方根误差 $= \sqrt{n^{-1}\sum_{i=1}^{n}(\hat{} y_i - y_i)^2}$。
MAE	平均绝对误差：MAE $= n^{-1}\sum_{i=1}^{n} \lceil \hat{} y_i - y_i \rfloor$。
布里尔分数（BS）	对于二元事件 Z，BS $= (\hat{} p - \mathbb{I}\{Z\})^2$。 应用于类别区间的分类，采用一对其余事件（例如琥珀色）。
AUC	用于特定区间 事件分类的 ROC曲线下面积（例如琥珀色与非琥珀色），使用预测事件概率，来自 $\widehat{F}_{d,t}$.
可靠性校准	导出的名义预测区间经验覆盖率 from $\widehat{F}_{d,t}$；用于可靠性汇总。
阈值加权连续排序概率评分	阈值加权连续排序 概率评分：阈值加权连续排序概率评分$(F, y) = \int_0^\infty w(z)\big(F(z) - \mathbb{I}\{y \leq z\}\big)^2 \mathrm{d}z$，权重函数 $w(z) = \mathbb{I}\{z \geq u\}$。强调预测性能在上尾部的表现；根据以下内容的蒙特卡洛样本进行评估 $\widehat{F}_{d,t}$ 使用 `scoringExtra` 实现。

B 第三种情景：设定错误的模型

鉴于条件分位数与可用预测变量之间的真实关系通常未知，评估模型设定错误的影响至关重要。为此，预测变量 Z $=$ (z$^{\{1\}}$, z$^{\{2\}}$) 按照第 4.1.2 节中的设定进行抽样，而响应变量 y 则根据第 4.1.1 节中的模拟设计生成，并确保其与 z$^{\{2\}}$相互独立。通过纳入一个不会影响 y分布的预测变量，这一设置导致模型设定错误。具体而言，在 z$^{\{2\}}$ 上计算维度为 $k = 10$ 的基函数展开，并将其纳入 X-FlexForecast 和 EDGP 的条件分位数及 log (σ(z)) 的参数化中，尽管 z$^{\{2\}}$ 与真实数据生成过程无关。表 B.1 总结了结果，显示基于均方根误差的比较与先前的发现一致：X-FlexForecast 与竞争方法在主体分位数上的精度相近，而 X-FlexForecast 在尾部分位数估计方面仍表现更优。

	$\xi = 0.0$				$\xi = 0.3$			
α	$\bar{q}$(X)	*X-flexForecast*	*qgam*	EDGP	$\bar{q}$(X)	*X-flexForecast*	*qgam*	EDGP
	RMSE				RMSE			
0.05	0.09	**0.61 ± 0.41**	—	0.52 ± 0.06	0.09	0.60 ± 0.35	—	**0.53 ± 0.07**
0.25	1.39	**1.23 ± 0.42**	—	0.96 ± 0.05	1.41	1.22 ± 0.33	—	**0.96 ± 0.14**
0.50	3.94	1.53 ± 0.52	—	**1.03 ± 0.11**	3.97	1.51 ± 0.46	—	**1.04 ± 0.34**
0.75	7.96	1.95 ± 0.65	—	**1.21 ± 0.29**	8.00	1.91 ± 0.63	—	**1.24 ± 0.68**
0.95	14.42	**1.98 ± 1.02**	2.26 ± 0.64	3.24 ± 0.77	14.93	**2.10 ± 0.99**	2.32 ± 0.65	4.19 ± 1.65
0.99	18.45	**2.01 ± 1.02**	4.00 ± 0.84	10.97 ± 2.08	21.48	**2.56 ± 0.99**	4.01 ± 0.89	13.11 ± 3.46
0.999	24.17	**2.60 ± 1.33**	17.23 ± 5.30	22.25 ± 4.09	38.79	**5.01 ± 2.51**	13.07 ± 8.32	26.17 ± 6.56
0.9999	30.03	**3.85 ± 2.38**	69.74 ± 17.36	33.47 ± 6.13	73.28	**14.46 ± 9.53**	42.62 ± 18.77	44.55 ± 8.13

表 B.1：第三个模拟情景下 1000 次重复模拟的平均均方根误差。粗体数值表示三种方法中均方根误差最低的结果。标记为 $\bar{q}$(X) 的列显示了 α 水平下真实分位数的平均值。qgam 的主体分位数数值未列出，因为它们与 *X-flexForecast* 的数值相同。

C *X-flexForecast* 阈值选择的敏感性分析

本附录进一步介绍 *X-flexForecast* 对概率水平 $\alpha^{(T)}$ 的表现，该概率水平决定主体部分分位数回归组件与离散广义帕累托（DGP）尾部之间的转换。在实际预报中，真实的超越概率未知，因此必须选择实际阈值。以下分析考察了在有意选择 $\alpha^{(T)}$的情况下，预测分布及相应尾部参数估计的稳定性。

C.1 估计阈值与数据生成阈值

尽管第 4 节规定 $\alpha^{(T)} = 1 - \phi$ 以使模型与模拟设计相一致，但实际整数阈值为

$$\hat{u}(\boldsymbol{x}_{d,t}) = \left\lceil \hat{q}_{\alpha_{d,t}^{(T)}}(\boldsymbol{x}_{d,t}) \right\rceil,$$

因此与数据生成阈值不同

$$u = \inf\{k \in \mathbb{N}_0 : F_{\mathrm{dg}}(k \mid \boldsymbol{\theta}_B) \geq 1 - \phi\}.$$

造成这种差异的两个结构性特征是：首先，$\hat{q}_{\alpha^{(T)}}(\boldsymbol{x}_{d,\,t})$会受到分位数回归拟合所产生的估计误差影响。其次，分布是离散的：多个相近的概率水平可能对应同一个整数分位数，因而对应同一个实际阈值。因此，$\alpha^{(T)}$ 可以理解为确定了一个可接受阈值的邻域，而非单一的概率水平。

C.2 候选阈值网格

为评估阈值设定错误的敏感性，我们将第4.1.1 节模拟中的所有设置扩展为 $\phi = 0.1$，方法是考虑一组候选转移概率

$$\mathcal{A} = \{(1-\phi) - 0.05, (1-\phi) - 0.04, \ldots, (1-\phi) + 0.04, (1-\phi) + 0.05\},$$

增量为0.01。步长0.01以及十个候选值的使用，为 $\alpha^{(T)}$定义了一个合理的局部网格，体现了标准的超参数调优权衡：在充分探索参数空间的同时避免不必要的密集搜索，其范围和间距可根据分析范围及所需粒度进行调整。对于每个 $\alpha^{(T)} \in \mathcal{A}$，相应阈值$\hat{u}$ 由估计分位数重新计算，尾部模型利用相应超越值重新拟合，并通过估计尾部分位数$\hat{q}_\alpha(\boldsymbol{x})$相对于 $\alpha \in \{0.95, 0.99, 0.999, 0.9999\}$的真实分位数的均方根误差来评估预测准确性。图C.1 总结了所得的缩放均方根误差模式，其中均方根误差按相应真实分位数值的平均值进行缩放。对于中间上分位数（$\alpha = 0.95$，0.99），均方根误差在不同 $\mathcal{A}$下仅有轻微变化，这反映出多个概率水平往往产生相同的整数阈值，且超越值仍处于广义帕累托近似充分适用的范围内。相比之下，对于远尾分位数（$\alpha = 0.999$，0.9999），当 $\alpha^{(T)}$ 设置过高时，均方根误差急剧增大，原因是超越值的$\hat{}$数量减少，从而导致（$\hat{\sigma}$，ξ）的方差增加。略低的 $\alpha^{(T)}$ 值可通过增大超阈值样本量部分抵消这一影响，而过低的阈值则会因非极端观测污染尾部估计而引入显著偏差。

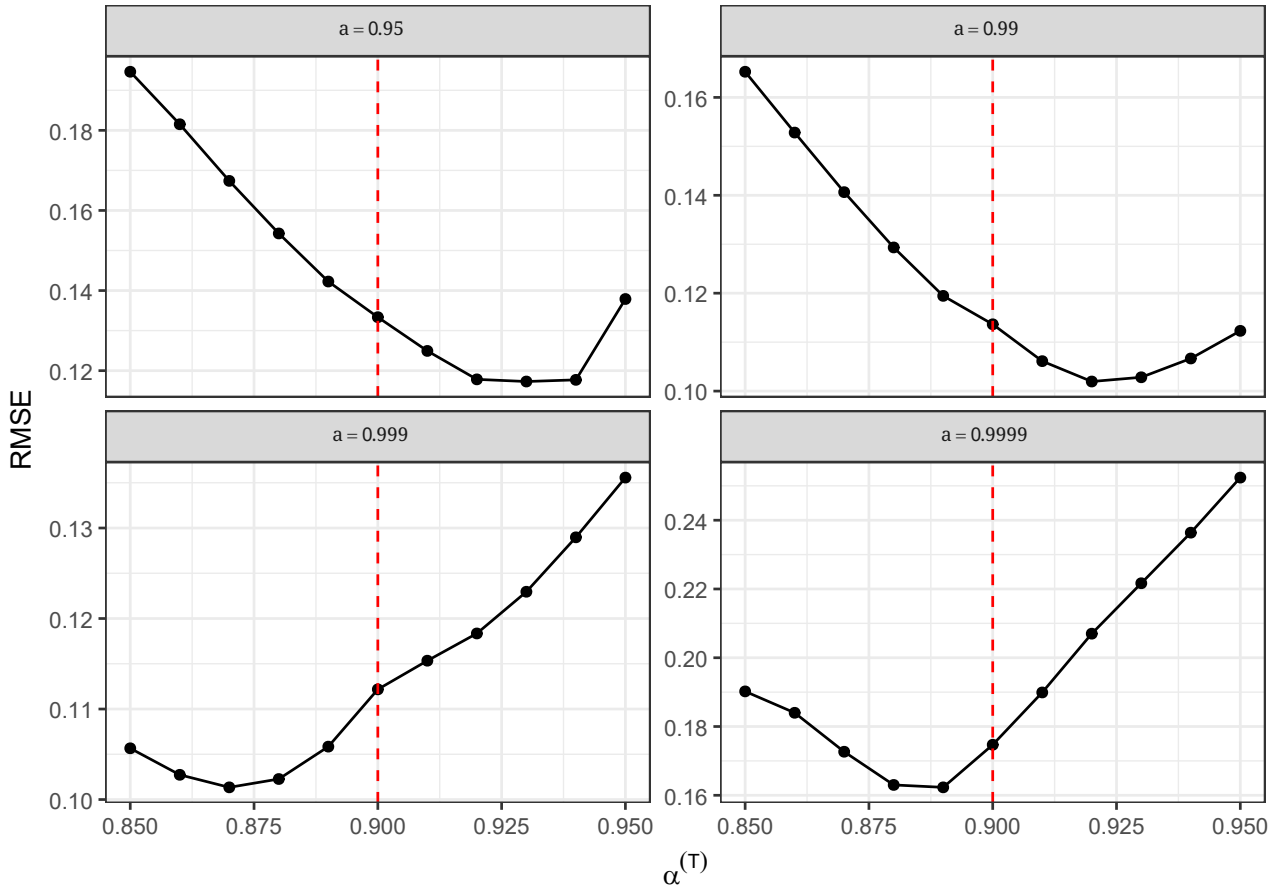
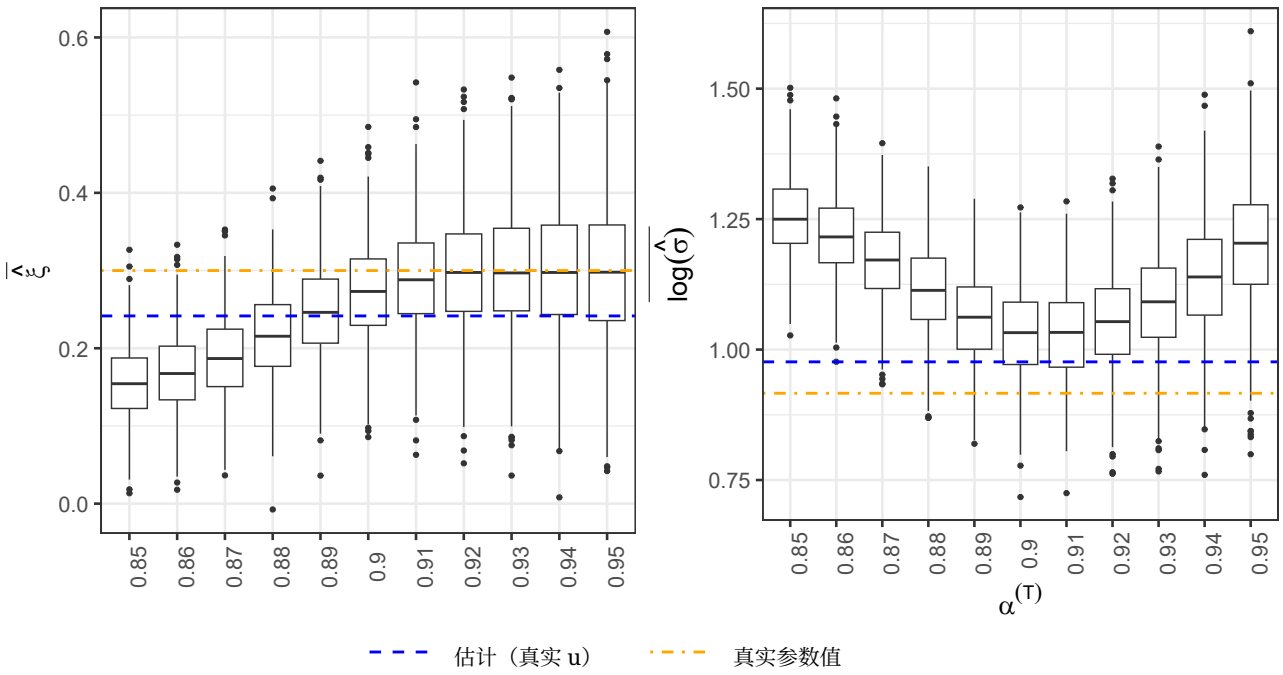


图 C.1：在具有 $\phi = 0.1$ 的模拟设计下，$\alpha \in 0.95$，0。99、0。999，0。9999 的预测尾部分位数在候选阈值水平 $\alpha^{(T)} \in \mathcal{A}$ 下的缩放平均均方根误差；虚线表示该结果。

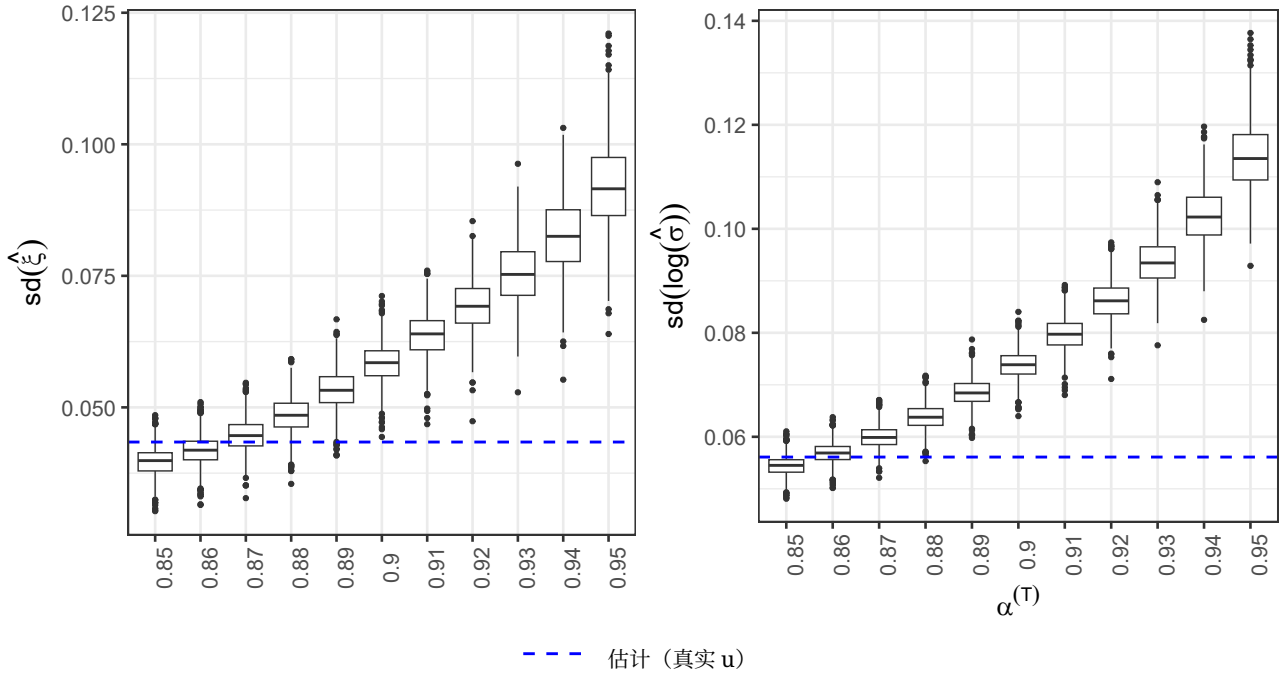
C.3 尾部参数估计

为补充均方根误差结果，我们考察估计尾部参数（$\hat{\xi}$，$\hat{\sigma}$）在每个 $\alpha^{(T)} \in \mathcal{A}$下的平均值和标准差。图C.2中的经验平均值表明$\hat{\xi}$ 在对应于概率水平约为0.89及以上的阈值处保持稳定，而$\hat{\sigma}$ 的平均值仍接近与数据生成阈值相关联的值。图C.3 显示，随着 $\alpha^{(T)}$增大，两个参数的标准差均上升，这与超越集内观测数量的减少相一致。虚线（蓝色）标示使用真实超越阈值u时所得到的估计，而点划线（橙色）则表示真实参数值

在所有实验中，*X-flexForecast* 的预测性能在数据生成值附近邻域内的阈值上保持稳定。将阈值设定得远低于该区域会因将非极端观测纳入尾部模型而引入偏差，而



ˆ 图 C.2：形状参数的平均估计值 ξ 和对数尺度参数 ˆσ 在候选阈值下 $\alpha^{(T)}$ 从 *X-flexForecast* 的尾部模型中获得，该模型由附录 C 中描述的模拟 $\phi = 0.1$得到。



ˆ 图 C.3：估计的形状参数 ξ 和对数尺度参数 ˆσ在各候选阈值下的标准差 $\alpha^{(T)}$ 来自附录 C 中描述的模拟 $\phi = 0.1$。

将阈值设得明显高于该值会减少超越值的数量，并增加估计量方差。在实际相关范围内，该方法能够可靠地估计主体部分和尾部分位数，这与超越模型的经典阈值选择原则一致。

C.4 其他超越概率

为评估方法的普适性，我们针对 $\phi \in \{0.15, 0.20, 0.25\}$重复了分析。所得模式与 $\phi = 0.1$的观察结果相似——详见第C.2节和C.3节——中间阈值下表现稳定，过高的 $\alpha^{(T)}$会导致方差膨胀，过低的 $\alpha^{(T)}$则会产生偏差。这些结果表明，该方法对于一系列合理的 $\alpha^{(T)}$选择具有稳健性。

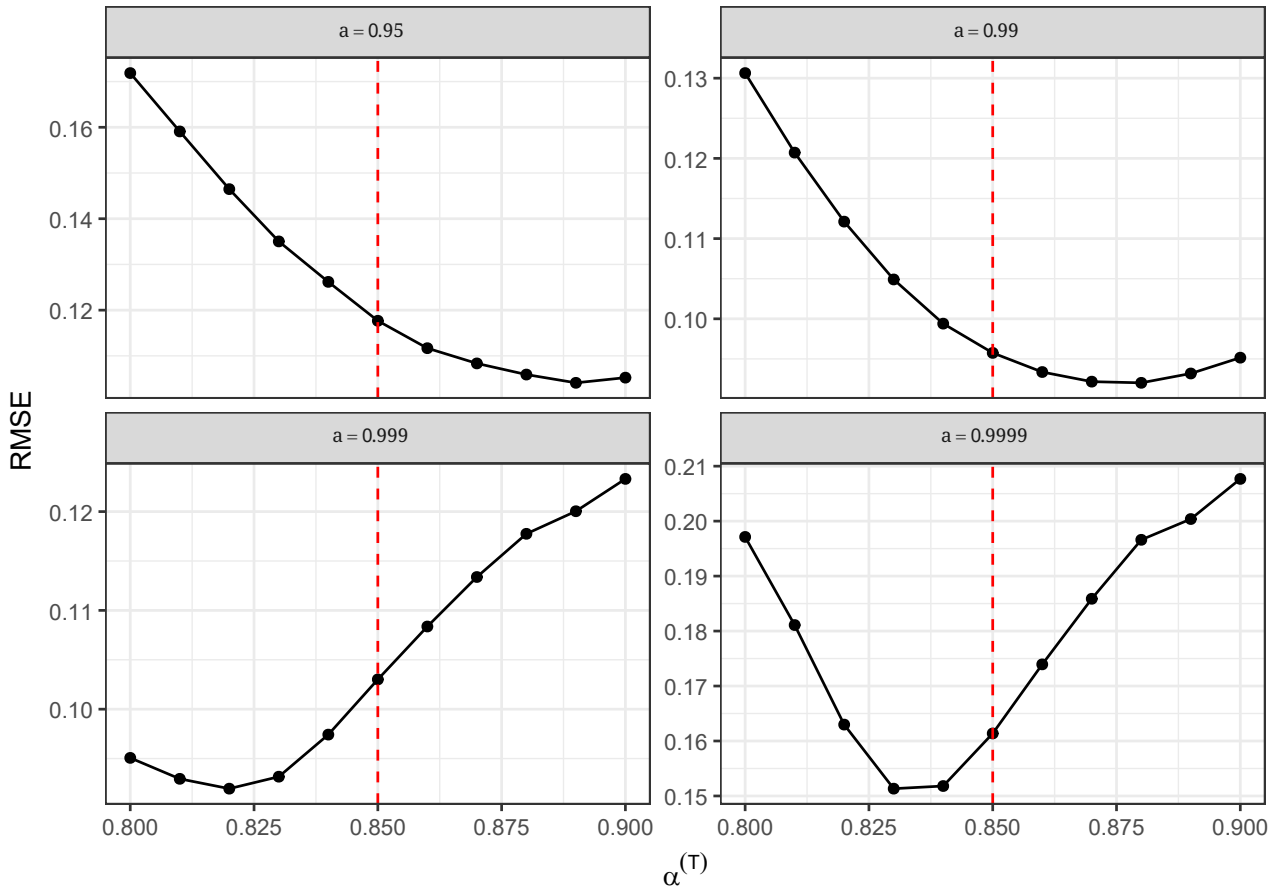
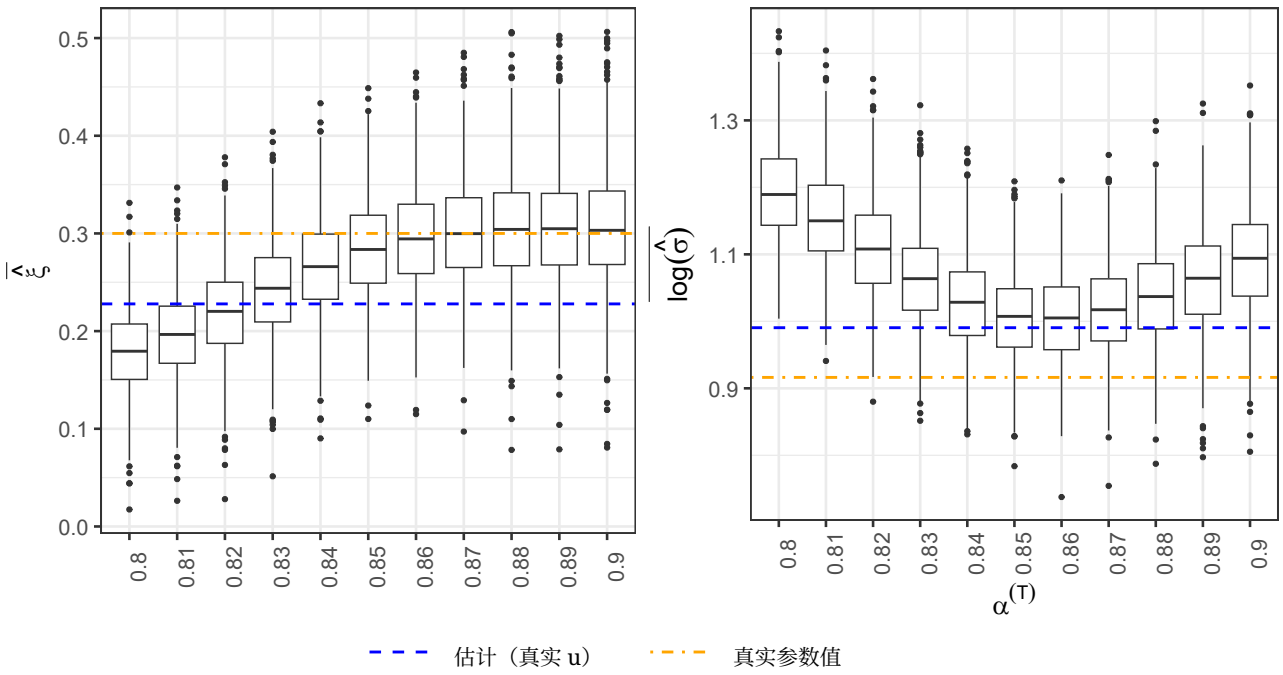
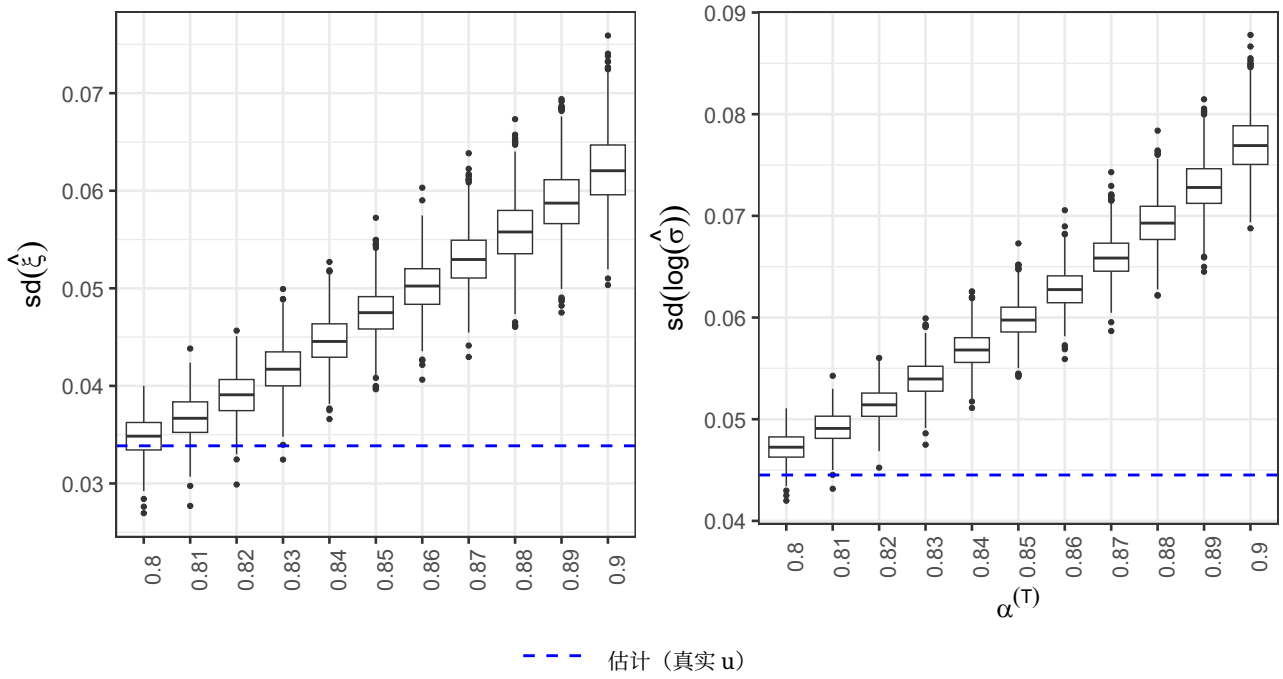


图 C.4：在具有 $\phi = 0.15$ 的模拟设计下，$\alpha \in 0.95$、0.99、0.999、0.9999 的预测尾部分位数均方根误差（RMSE）平均值经缩放后的结果，适用于候选阈值水平 $\alpha^{(T)} \in \mathcal{A}$；虚线表示该模拟设计。



ˆ 图 C.5：形状参数的平均估计值 ξ 和对数尺度参数 $\hat{}\sigma$ 在候选阈值中 $\alpha^{(T)}$ 根据以下尾部模型获得：*X-flexForecast* 该模型源自附录 C 中所述的模拟，$\phi = 0.15$。



ˆ 图 C.6：估计形状参数的标准差 ξ 和对数尺度参数 $\hat{}\sigma$在候选阈值中 $\alpha^{(T)}$ 来自附录 C 中所述的模拟，$\phi = 0.15$。

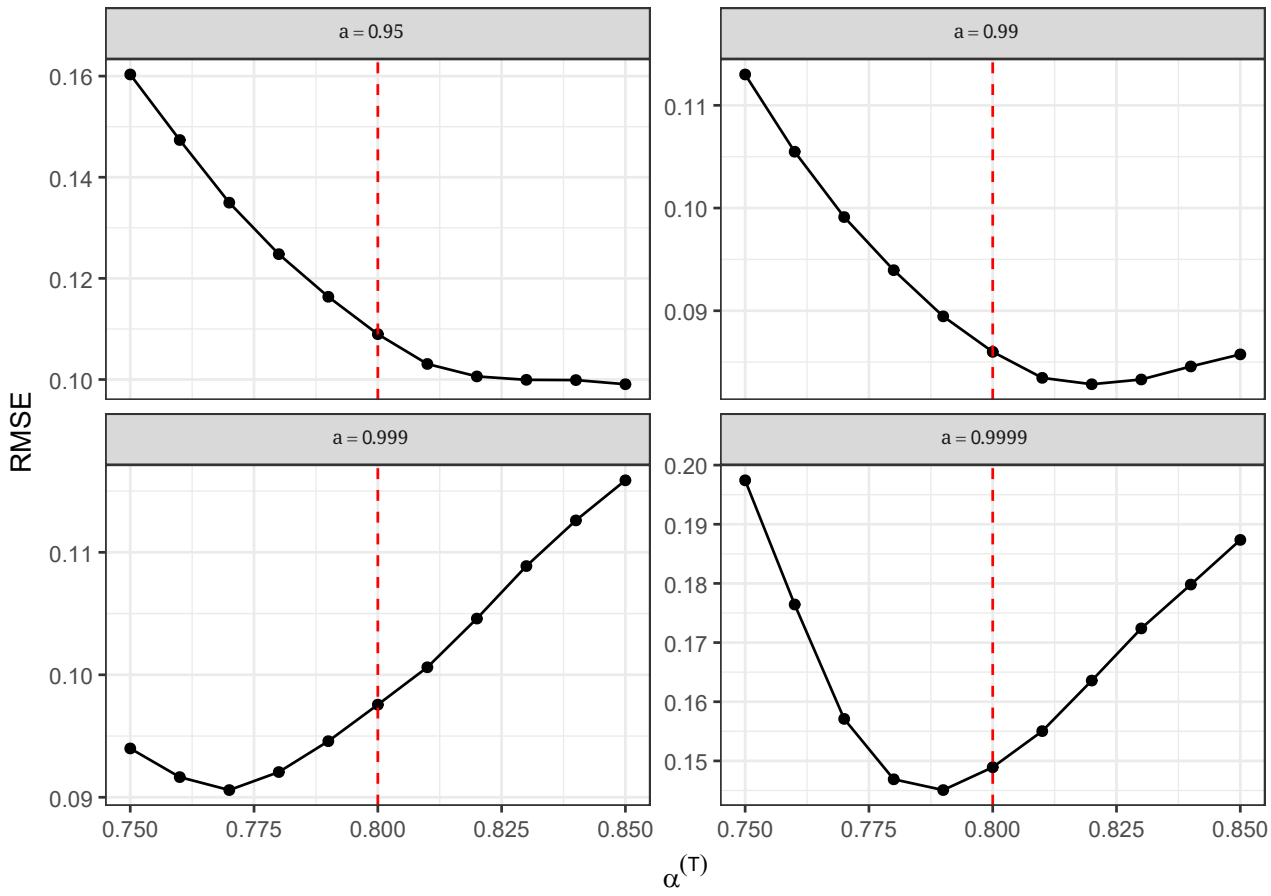
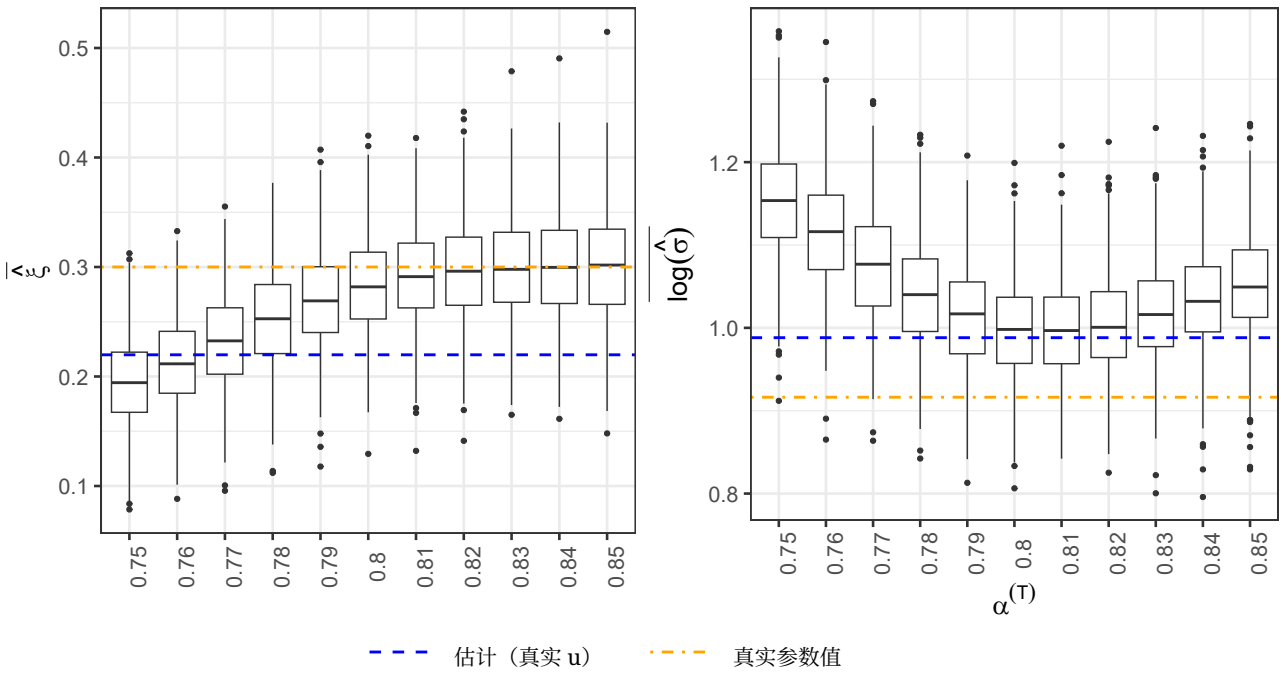
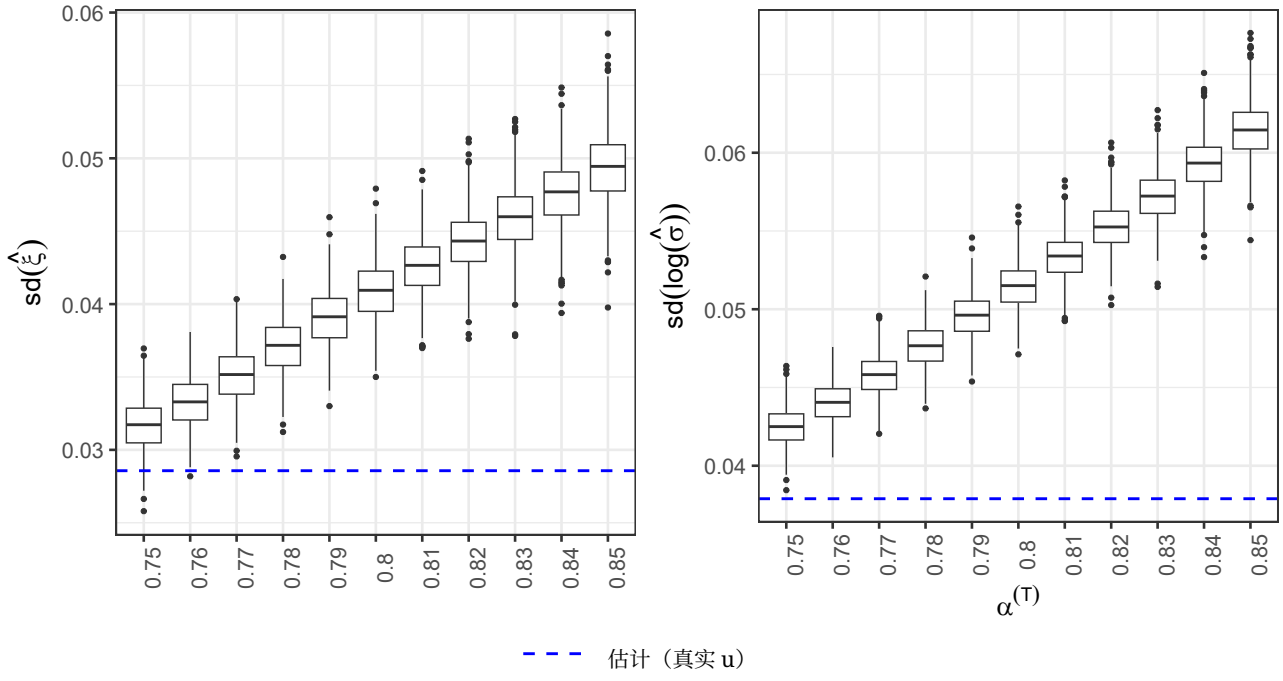


图 C.7：在模拟设计下，针对 $\alpha \in 0.95, 0.99, 0.999, 0.9999$ 的预测尾部分位数在各候选阈值水平 $\alpha^{(T)} \in \mathcal{A}$ 上的缩放平均均方根误差，其中 $\phi = 0.2$ – 由虚线标示。



ˆ 图 C.8：形状参数的平均估计值 ξ 和对数尺度参数 ˆσ 在不同候选阈值下 $\alpha^{(T)}$ 由 *X-flexForecast* 尾部模型得到，该模型基于附录 C 中所述的模拟 $\phi = 0.2$。



ˆ 图 C.9：估计的形状参数的标准差 ξ 和对数尺度参数 ˆσ在不同候选阈值下 $\alpha^{(T)}$ 基于附录 C 中所述的模拟，参数为 $\phi = 0.2$。

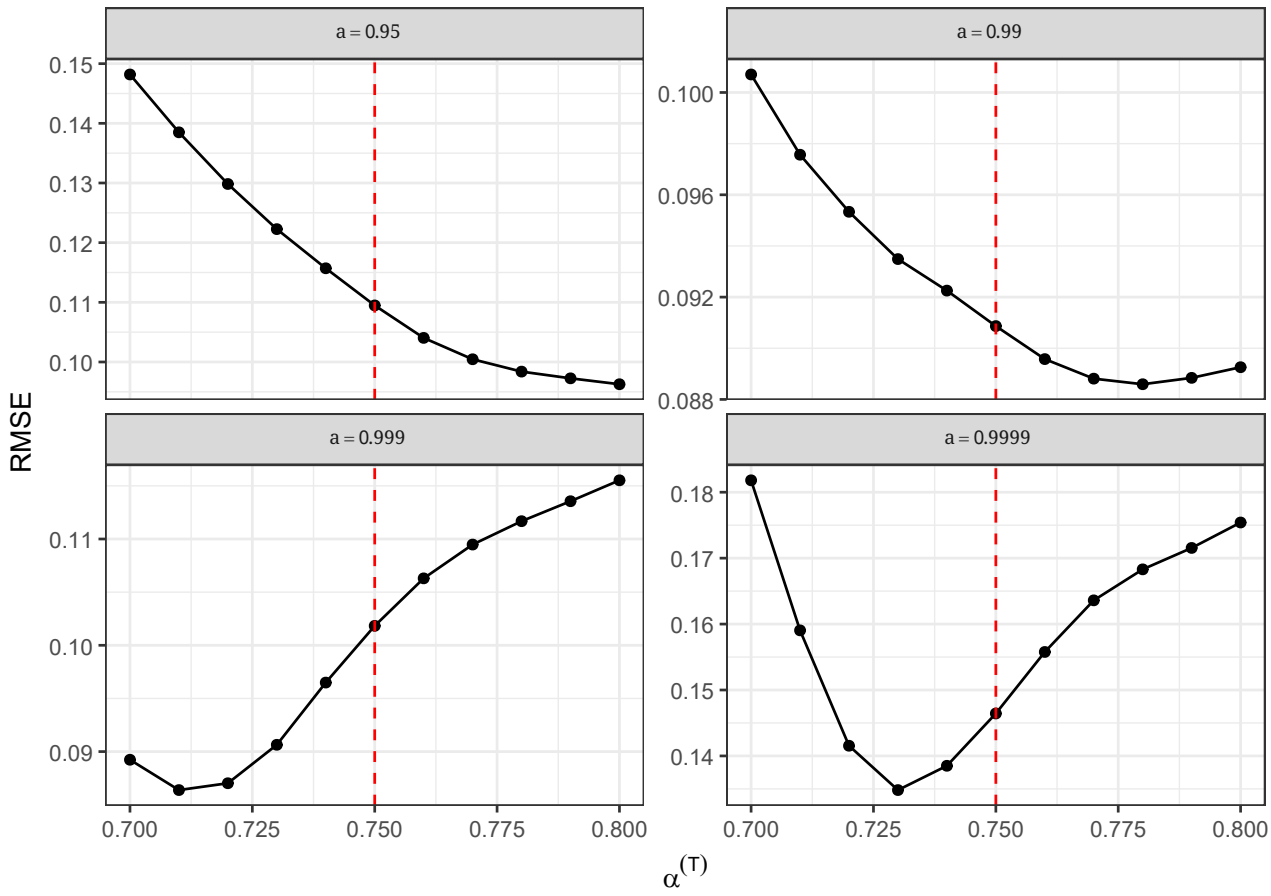
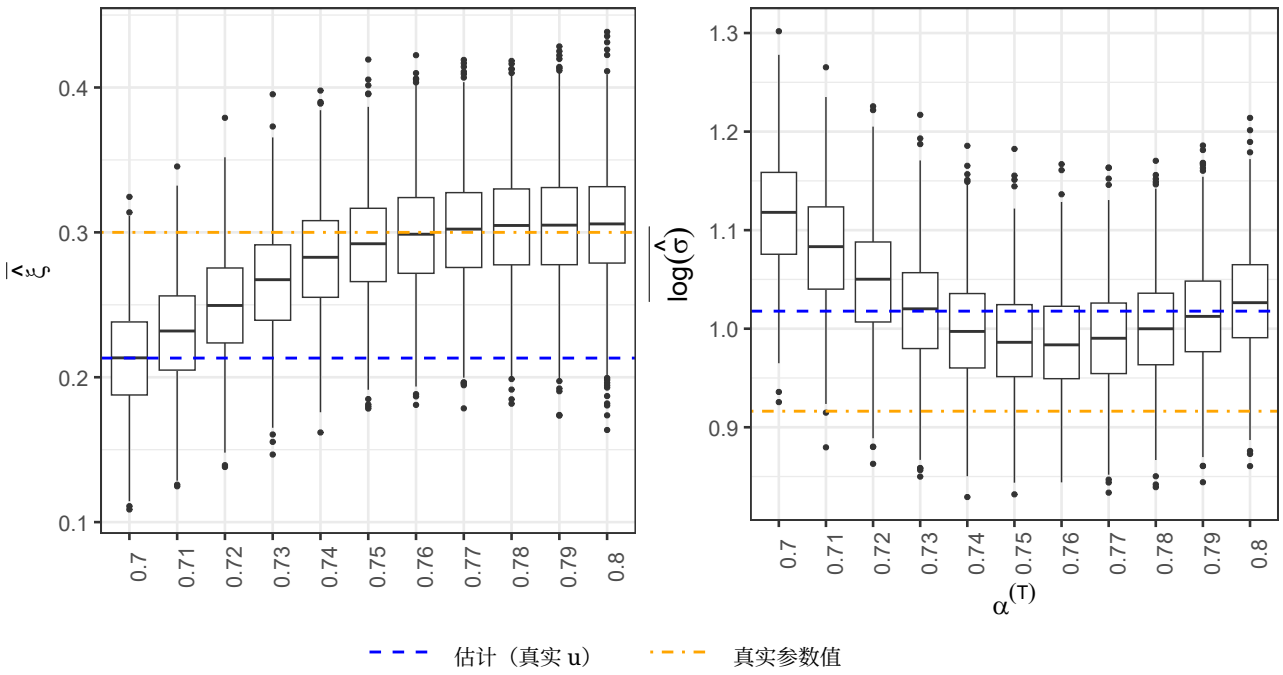
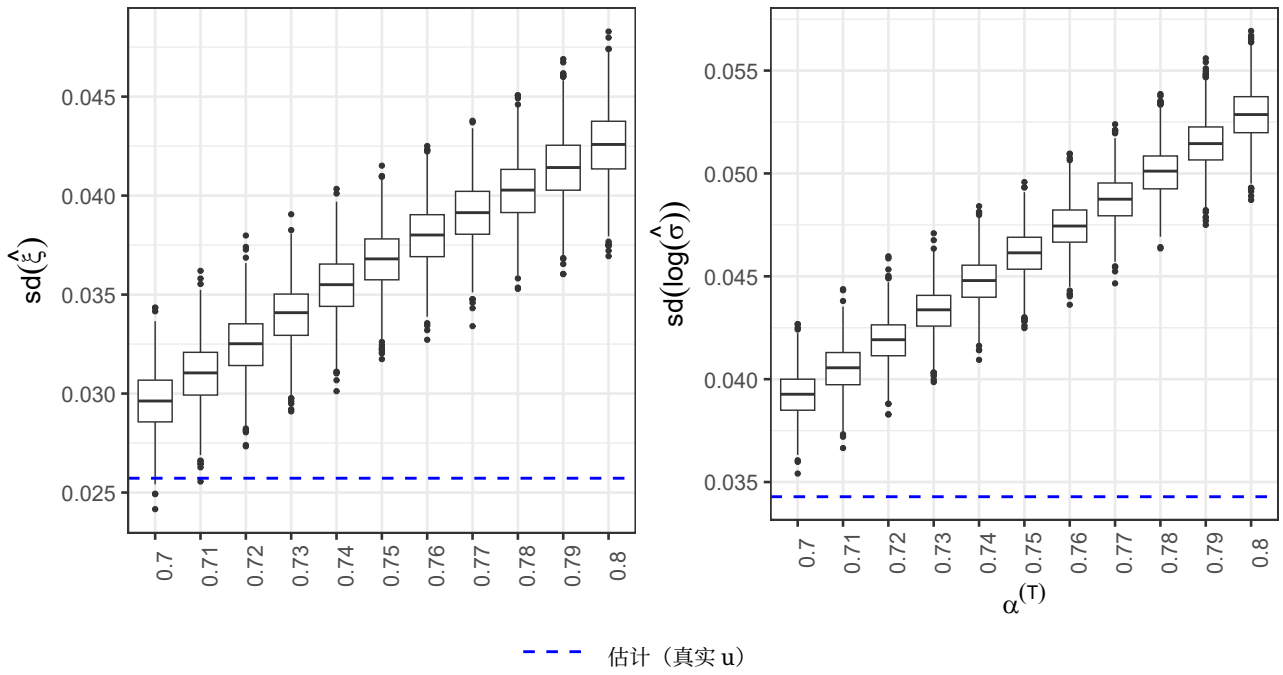


图 C.10：在含有 $\phi = 0.25$ 的模拟设计下，$\alpha \in$0.95，0.99，0.999，0.9999 的预测尾部分位数的缩放平均均方根误差，在候选阈值水平 $\alpha^{(T)} \in \mathcal{A}$ 下的结果——虚线所示。



ˆ 图 C.11：形状参数的平均估计值 ξ 和对数尺度参数 ˆσ 在候选阈值上的结果 $\alpha^{(T)}$ 取自 *X-flexForecast* 尾部模型，该模型基于附录 C 中所述的模拟 $\phi = 0.25$。



ˆ 图 C.12：估计形状参数的标准差 ξ 和对数尺度参数 ˆσ 在候选阈值上的结果 $\alpha^{(T)}$ 来自附录 C 中所述的模拟，参数为 $\phi = 0.25$。

D *X-flexForecast* 模型规格

该*X-flexForecast* 模型（见第2 节所述，并概括于公式（3））需要同时设定转移概率 $\alpha_{d\,t}^{(T)}$，以定义主体–尾部混合，以及协变量集合$\mathbf{x}_{d\,t}^{(DGP)}$，用于对尺度参数 σ（$\mathbf{x}_{d,\,t}^{T}$）进行建模，该尺度参数属于离散广义帕累托（DGP）分布，且针对每个区 d 和时间分辨率 t。利用第5节介绍的数据集，这些量通过14折交叉验证方案加以选择，每一折对应一个完整监管年度（4月至次年3月），从而确保每一折至少包含一个完整冬季。

候选规格基于样本外预测性能进行比较，而非通过优化可微损失函数，因为转移概率 $\alpha_{d\,t}^{(T)}$，无法在形式化学习目标中表示为连续决策变量。相反，$\alpha_{d\,t}^{(T)}$，被视为超参数，并从候选离散网格中选取。对于每个候选阈值，使用交叉验证评分规则评估预测技能，包括布里尔分数（BS）和ROC曲线下面积（AUC），并分别针对绿色、琥珀色和红色类别计算。所选阈值 $\widehat{\alpha}_{d\,t}^{(T)}$，对应于在各类别上使BS与AUC平均改进最大的配置；当两项准则不一致时，优先将BS作为主要选择度量指标，因为它是评估校准与准确性的严格恰当评分规则，而AUC主要反映排序性能（Gneiting和Raftery，2007）。该准则亦用于定义子集$\mathbf{x}_{d,\,t}^{T}$。在此流程下选定的天气预测变量、混合阈值 $\alpha_{d\,t}^{(T)}$，以及DGP分量协变量的最终组合汇总于表D.1。

表 D.　1：各区在时间分辨率 $t = 24$ 下的 DGP 模型规格 d。

d	$\mathbf{x}_{d,t}^{(DGP)}$	$\alpha_{d,t}^{(T)}$
艾尔郡与克莱德南部	$\mathrm{ws10}_{\max} + tp_{q90}$	0.85
中部与法夫	$\mathrm{ws10}_{\max}$	0.80
迪河谷与中威尔士	$\mathrm{ws10}_{\max} + tp_{q90}$	0.90
邓弗里斯与斯特兰拉尔	$\mathrm{ws10}_{\max} + tp_{q90}$	0.90
爱丁堡与边境地区	$\mathrm{ws10}_{\max}$	0.90
拉纳克郡	tp_{q90}	0.90
默西塞德	$t2m_{\mathrm{mean}}$	0.75
中柴郡	tp_{q90} +岬角平均值≥100	0.85
北威尔士	西南$10_{\max} + tp_{q90}$	0.90
威勒尔	$\mathrm{ws10}_{\max} + t2m_{\mathrm{mean}}$	0.80

不过，仍报告 AUC 值作为补充性能度量指标，因为这有助于熟悉该度量指标的业务相关方进行解读（见第 5 节和附录D）。表 D.2报告了各地区和不同方法的微平均和宏平均 AUC

值。

表D.2：$mqgam$ 模型和*X-flexForecast* 模型按区划分的微平均和宏平均AUC。

	qgam		*X-flexForecast*	
区	AUC（宏平均）	AUC（微平均）	AUC（宏平均）	AUC（微平均）
艾尔郡与克莱德南部	0.858	0.996	0.867	0.996
中部与法夫	0.766	0.993	0.812	0.995
迪河谷与中威尔士	0.844	0.992	0.854	0.993
邓弗里斯与斯特兰拉尔	0.857	0.993	0.883	0.993
爱丁堡与边境地区	0.790	0.989	0.829	0.990
拉纳克郡	0.755	0.993	0.767	0.993
默西塞德	0.637	0.986	0.704	0.988
中柴郡	0.839	0.992	0.821	0.992
北威尔士	0.837	0.990	0.869	0.990
威勒尔	0.728	0.988	0.739	0.989

除了使用 BS 评估概率预测外，我们还使用其他适当评分规则评估模型。根据审稿人的建议，我们计算连续排序概率评分（CRPS），尤其是其阈值加权变体（twCRPS；Gneiting 和 Ranjan，2011），该评分通过 `scoringRules`软件包实现（Allen，2024年）。这一选择反映了混合模型的结构：在支撑集的区域 $y \leq \lfloor q_{\alpha_{d\,t}^{(T)}}$，$(\mathbf{x}_t)\rfloor$ 上，多重加性分位数回归模型与 X-flexForecast 的预测一致。因此，未加权的 CRPS 会过多关注预测分布中两个模型相同的部分。为将比较重点放在上尾——即各预测方法存在差异之处——我们使用权重函数为 $w(z) = \mathbb{I}\{z \geq a\}$ 的 twCRPS，其中 $a = \lfloor q_{\alpha_{d\,t}^{(T)}}$，$(\mathbf{x}_t)\rfloor$ 且 $b = \infty$，从而使评估能够隔离尾部行为的差异。由于预测分布 $F_{d,t}(\mathbf{x}_{d,t})$ 并无闭式表达式，twCRPS 通过 `twcrps sample` 计算；该方法基于从各模型下的 $F_{d,t}(\mathbf{x}_{d,t})$ 抽取的 Monte Carlo 样本估计评分。为确保数值稳定性，我们使用 $n_{\text{samples}} = 100{,}000$ 个样本。更多方法细节见 Allen（2024年）。图 D.1 展示了 twCRPS 技巧评分的经验密度。该分布的大部分质量位于零参考线右侧，表明在大多数情况下，X-flexForecast 模型的 twCRPS 值较低，因此预测技能高于标准的多重 qgam 方法。我们注意到，在少数地区且总共不到四天的情况下，基于 twCRPS 的比较出现了不稳定现象。

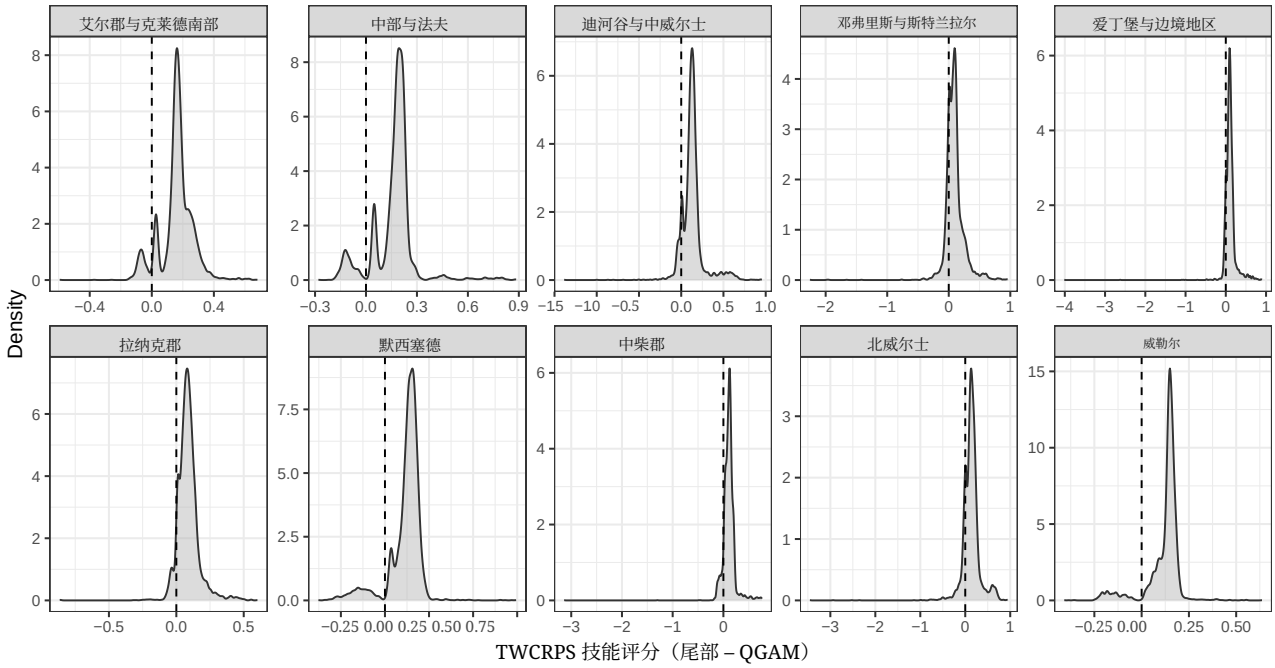


图 D.1：以$qgam$ 作为参考模型时，十个地区的 twCRPS 技能评分经验密度。垂直虚线表示零，意味着两种方法的表现相同。在每个地区，分布的大部分质量都位于正值一侧，表明相较于标准的多重$qgam$ 模型，X-$flexForecast$ 尾部模型在大多数预报实例中取得了更低（即更优）的阈值加权连续秩概率得分。

在这些情况下，twCRPS 使用的阈值高于$qgam$预测分布的有效支撑集。由于$qgam$ 模型的上尾可能相对较短，其经验累积分布函数在加权区域内较早达到 1，导致 twCRPS 降为零，并使由此得到的技巧评分异常偏大或不稳定。这种现象反映的是$qgam$ 尾部的结构性截断，而非预测性能的真正提升。为使图表更清晰，这些情况未纳入相应图表，但在底层分析中均已充分计入。鉴于此类情况罕见且影响有限，这一现象不会影响总体结论；此处报告仅为完整起见。