

从分布到分位数神经基础模型：以电力价格预测为例

Alessandro Brusaferri, Danial Ramin, and Andrea Ballarino

摘要—尽管神经网络在多时间尺度的概率预测中实现了高预测准确性，但对于预报员来说，理解导致特征条件输出的基本机制仍然是一个重大挑战。

在这项工作中，我们通过引入分位数神经基模型来进一步解决这一关键问题，该模型将分位数广义加性模型的可解释性原则融入端到端神经网络训练框架中。

为此，我们利用了共享基分解和权重因子分解的方法，补充了位置、尺度和形状的神经模型，同时避免了任何参数分布假设。

我们在次日电价预测上验证了我们的方法，实现了与分布回归和分位数回归神经网络相当的预测性能，同时通过从输入特征到输出预测的时间范围内的非线性映射学习提供了对模型行为的有价值的见解。

Index Terms—神经网络，广义加性模型，分位数回归，分布回归，时间序列，概率预测，电价

I. 介绍

概率电力价格预测 (PEPF) 在日前电力市场中的挑战构成了一个具有重要实际意义的关键研究领域。准确地生成预测并可可靠地量化不确定性对于众多市场参与者至关重要，包括公用事业公司、零售商、聚合商和主要消费者 [1]。电价表现出显著的特性，如高波动性和突然波动，这些由复杂因素驱动，包括分布式能源需求模式、生产成本以及气象条件 [2]。可再生能源技术日益整合进一步复杂化了这一环境，尽管它们在减少全球排放方面发挥了关键作用 [3]。在过去几十年中，开发了一系列方法来捕捉和表示概率电力价格预测中的内在不确定性。这些方法从基本的预测区间技术到先进的方法如离散条件分位数估计和全面分布模型 [4]。为了详细了解这些发展，读者可参考近期的综述文章 [5],[6]。

目前，该领域正日益关注基于神经网络的方法论，这些方法论利用其在条件空间内的自适应映射能力，并

从不断扩大的计算资源和先进的工具中获益 [6]。在此范式下，参数化复杂函数形式（参见例如，[7]）的分布深度神经网络 (D-DNNs) 以及分位数回归深度神经网络 (QR-DNN)（参见例如，[8]）在建模价格分布模式的复杂特征方面显示出巨大的潜力。这些特征包括明显的异方差性、重尾行为和不对称性。

然而，这种增强的建模能力引入了关于模型可解释性的挑战。这些架构的基本不透明性限制了对预测生成过程的理解，可能在关键决策环境中创建障碍。最重要的是，这些模型将输入特征转换为密度参数并在预测范围内输出分位数的机制仍然很大程度上无法让从业者理解。各种事后模型不可知的方法，包括基于 LIME 的代理建模和 Shapley 值方法的扩展，已经被开发出来通过局部模型近似和特性贡献分析提供可解释性 [9]。然而，实现对这些模型执行的复杂内部计算的可靠全局解释仍然从根本上具有挑战性 [10]。

因此，一个替代的研究方向集中在将透明性直接纳入神经网络设计中。此类别中的一个显著例子是由 [11] 提出的 Neural Additive Model (NAM) 框架，并随后在 [12] 中应用于预测问题。建立在广义加性模型 (GAM) 结构的基础上，NAM 使用专门神经网络的线性组合，每个网络处理单独特征，从而揭示底层输入输出关系。除了 PEPF 应用之外，[13] 的近期工作已将 NAM 从最初的点预测重点扩展到全面的分布回归，通过适应更广泛的 GAMLSS 框架 [14]，形成了位置、尺度和形状神经加性模型 (NAMLSS)。在此工作的基础上，在 [15] 中进一步开发并部署了 NAMLSS 模型以应用于 PEPF 任务，通过特征形状函数的基分解 [16]，其中采用了一个独特的密集映射来学习一组由可训练线性投影组合而成的共享基。虽然实现了与参数化灵活 Johnson's SU 密度的 D-DNNs 相媲美的概率预测性能，但这种架构（标记为 NBMLSS）通过在预测步骤中学习从非线性特

Alessandro Brusaferri (corresponding author), Danial Ramin, and Andrea Ballarino are with CNR, Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, via A. Corti 12, Milan, Italy (e-mail: alessandro.brusaferri@cnr.it).

征级别到分布参数的映射提供了更多的模型行为洞察力。然而，NBMLSS 模型仍然受到依赖于参数假设的限制，就像基线 D-DNNs 一样。

在这项工作中，我们进一步探索了神经加性模型框架在概率预测方面的应用，特别关注于具有挑战性的日前电力价格预测任务。为此，我们提出了一种非参数分位数回归方法作为 NBMLSS 架构的对应物，该方法借鉴了分位数广义加性模型 [17],[18] 的设计原则，并旨在增强在更广泛的 QR-DNN 框架中的可解释性。我们将开发的方法在此之后标记为 QNBM，即分位数神经基础模型。我们使用来自德国和比利时日前电力市场的公开数据集与 D-DNN、QR-DNN 和 NBMLSS 模型进行了实验比较，这些数据集涵盖了近期时间段和多样的市场特征。

II. 方法

我们集中研究一类多步概率预测模型，旨在识别整个预测范围 $h = 1, \dots, H$ （例如接下来的 24 小时）内的目标分布分位数集合 $\gamma \in \Gamma$ （如十分位、百分位等）。然而，我们强调所提出的方案可以很容易地应用于使用不同模型来估计不同分位数或进行特定阶段推断的子案例中。在此背景下，神经网络通常被实现为参数化函数，将输入变量集映射到预测范围内的一次传递中的预测值：

$$p(y_{t+1} \dots y_{t+h} \mid y_{t-k:t}, z_{t-k:t}, x_{t+h}) = f_{\Theta}(y_{t-k:t}, z_{t-k:t}, x_{t+h})$$

其中 Θ 概括了模型的参数， k 是每个输入序列观测历史中涉及的最大滞后。在 PEPF 应用中，右侧条件变量集通常包括直到当前日期为止的目标价格 $y_{t-k:t}$ 的过去值以及进一步的外生特征。外生集合可以包含过去几天的观测值 $z_{t-k:t}$ （例如，之前的电力需求）和预测变量 x_{t+h} ，如电力负荷预测、可再生能源发电预测以及常量特征（例如，星期几编码等）。遵循深度神经网络 (DNN) 模型在 [7] 中为 PEPF 提出的前馈类，条件集合随后被构造成大小为 n_f 的展平张量，定义为此处的 $\mathbf{x}_d = [x_{d,1}, \dots, x_{d,n_f}]$ 。为了保持一致性，在预测范围内的每一天 d 的目标价格值，例如接下来的 24 小时，标记为 $\mathbf{y}_d = [y_d^1, \dots, y_d^H]$ 。

A. 从 NBMLSS 到分位数神经基础模型

形式上，分位数神经基础模型 (QNBM) 的结构如下：

$$z_k(x_{d,i}) = \mathbf{a} \left[\sum_{j=1}^{n_u} \omega_{j,k}^{(2)} \mathbf{a} \left[\omega_j^{(1)} x_{d,i} \right] + \omega_{0,k}^{(2)} \right], k = 1, \dots, n_z \quad (1)$$

$$f_i(x_{d,i}) = \sum_{k=1}^{n_z} W_{(i,k)} z_k(x_{d,i}), i = 1, \dots, n_f \quad (2)$$

$$\hat{q}_h^\gamma(\mathbf{x}_d) = \beta_h^\gamma + \sum_{i=1}^{n_f} V_{(h,\gamma,i)} f_i(x_{d,i}), h = 1, \dots, H, \gamma \in \Gamma \quad (3)$$

显然，该设计遵循 NBMLSS 模型的基础扩展框架，但用直接映射到条件输出分位数的方式替换了原本用于将链接函数映射到分布参数的最后一层。

详细来说，方程 1 定义了基数为 $n_z \in \mathbb{Z}^+$ 的一组中的 k 次共享基函数， $\omega_j^{(1)} \in \mathbb{R}^{n_u}$ 和 $\omega_{j,k}^{(2)} \in \mathbb{R}^{n_u \times n_z}$ 分别表示共享神经网络第一层和第二层的可训练权重，而 $\mathbf{a}[\cdot]$ 表示应用于整个网络的非线性激活函数。尽管最初提出的神经加性模型采用了指数激活函数来捕捉高度非线性和锯齿状的特征映射 [11]，但最近的研究表明，在实践中标准 ReLU 激活通常能提供更优的表现 [19]。此外，通过在训练过程中随机省略单元，引入 Dropout 层以促进基函数之间的去相关性。

随后的方程通过从基础输出 $z_k(x_{d,i})$ 导出的可训练线性投影来建模逐步目标条件分布分位数 $\gamma \in \Gamma$ ，针对每个输入特征 $x_{d,i}$ 进行评估。具体来说，矩阵 $W \in \mathbb{R}^{n_z \times n_f}$ 用于将共享的基础函数聚合为特定于特征的形状函数 $f_i(\cdot)$ ，而张量 $V \in \mathbb{R}^{H \times |\Gamma| \times n_f}$ 结合这些形状函数的输出以生成逐步分位数估计值 $\hat{q}_h^\gamma(\mathbf{x}_d)$ 。 $\beta_h^\gamma \in \mathbb{R}$ 表示可训练的偏置项。

尽管方程 1 – 3 中的直接表述不仅需要存储完整的权重张量 W 和 V ，而且还需在反向传播过程中计算并保留它们的梯度。随着特征数量和预测范围内的分位数维度的快速增长，这导致了训练期间显著的内存消耗和计算开销。为了解决这些问题，我们通过矩阵分解引入可训练的低秩近似。具体来说，对于矩阵 $M \in \mathbb{R}^{m \times n}$ ，我们将其近似为：

$$M \approx AB^\top, \text{ with: } A \in \mathbb{R}^{m \times r}, B \in \mathbb{R}^{n \times r}, r \ll m, n \quad (4)$$

W 和 V 都可以这样分解，这取决于应用的需求。张量 V 被转换为 $\tilde{V} \in \mathbb{R}^{H \times |\Gamma| \times n_f}$ 并随后以适当的形状进行

重塑。秩参数 r 控制压缩/精度权衡。除了计算优势之外，该分解旨在强制识别更为紧凑且可能更稳健的表示形式 [20]。

整个架构通过最小化在每个阶段 $\hat{q}_h^\gamma(\mathbf{x}_d)$ 预测的条件分位数集 $\gamma \in \Gamma$ 上的平均多阶段 Pinball 损失来进行端到端训练（例如，通过 Adam 优化器），计算公式为：

$$\sum_d \sum_h \sum_\gamma (y_d^h - \hat{q}_h^\gamma(\mathbf{x}_d)) \gamma 1\{y_d^h > \hat{q}_h^\gamma(\mathbf{x}_d)\} + (\hat{q}_h^\gamma(\mathbf{x}_d) - y_d^h)(1 - \gamma) 1\{y_d^h \leq \hat{q}_h^\gamma(\mathbf{x}_d)\} \quad (5)$$

其中数据集 $\mathcal{D}_n \equiv \{(\mathbf{x}_d, \mathbf{y}_d)\}_{d=1}^n$ 是通过使用滑动窗口方法分割条件和外生时间序列，并随后将结果片段合并为统一的输入-输出对生成的。为了防止过拟合，采用了基于验证的提前停止策略，即如果在指定数量的纪元内没有观察到改进，则停止训练。此外，我们在常见模型重新校准之外集成了可逆实例归一化（RevIN）层（参见 [21]），旨在缓解潜在的分布偏移。进一步的实现细节（如超参数调整等）将在下一节专门讨论实验评估中进行探讨。

III. 实验与结果

为了评估 QNBM 模型，我们利用了最近由 [22] 整理并进一步扩展到 [15] 的公开数据集。这些数据集包括来自 ENTSO-E 透明度平台的样本，涵盖多个欧洲电力市场。在本研究中，我们将重点放在德国和比利时地区，以应对多样的市场条件。其他电力市场的探索留待未来的研究。数据集涵盖了截至 2024 年 9 月 30 日的时间范围，使得能够在近期预测环境特征下的高波动性条件下进行评估。每个数据集包括负荷和可再生能源（风能和太阳能）发电的每小时预测作为外生输入。此外，还使用循环正弦和余弦变换对时间序列年龄和星期几等时间特征进行了编码。

实验设置遵循 NBMLSS 研究 [15] 中的协议以保持一致性。输入-输出对使用标准滑动窗口方法生成，对于每一天 d ，条件向量 $\mathbf{x}_d$ 包含来自前 1 天 d 、前 2 天 d 和前 7 天 d 的历史价格值，以及相应的逐小时日前外生变量。

样本外测试使用了从 2023 年 10 月 1 日至 2024 年 9 月 30 日的数据。初始训练和验证集（在重新校准之前）包括了从 2019 年 1 月 1 日起到测试期开始前的所有样本。超参数调整使用的是每个测试集之前的连续一年数据，采用 Optuna[23] 进行交叉验证，并基于 20%

表 I: 调整后的超参数

	QR-深度神经网络		QNBM	
	DE	BE	DE	BE
n_u	640	128	64	64
l_r	1e-4	5e-4	5e-4	5e-4
d_r	0.1	0.1	0.1	0.3

的随机子样本进行提前停止。为了管理不同实验设置下的计算需求，在测试阶段模型每周重新校准一次，并在交叉验证过程中使用四个顺序折叠进行再训练。训练使用 Adam 优化器，最多进行 800 个 epoch，并且在连续 20 个 epoch 内没有改进时触发提前停止机制，批次大小为 128 的滑动窗口，结构化为 32 样本的小块。

作为基线，除了 NBMLSS 架构和来自 [15] 的分布神经网络（两者都映射 Johnson’s SU 密度，后者在此后标记为 J-DNN）之外，我们还开发了一个分位数回归神经网络（QR-DNN），遵循 [8]。QR-DNN 和 QNBM 都被配置成在一个单独的传递中提供每个预测步骤的条件百分位数（即 $|\Gamma| = 99$ ）。所有模型均使用 TensorFlow Probability 库 [24] 实现，通过开发自定义层和损失函数实现。为了与 J-DNN 保持一致，QR-DNN 被结构化为包含两个隐藏层，采用 ReLU 激活函数并结合 dropout 正则化。每个层的单元数、dropout 率和学习率分别使用离散集 [64, 128, 512, 640, 768]、[0, 0.1, 0.3, 0.5] 和 [1e-3, 5e-4, 1e-4, 5e-5] 进行网格搜索调优。为了与 NBMLSS 配置保持一致，QNBM 模型的共享特征网络中的单元数在减小范围内 $n_u, n_z = [32, 64, 128]$ 内进行调整。本研究中的实验中，秩参数设置为 $r = 16$ 。不同市场选择的超参数显示于表 I 中。QR-DNN 和 QNBM 模型都获得了与其各自分布设置相一致的配置。然后，每个模型从不同的随机初始化开始重新校准五次。预测结果使用标准分位数平均法进行聚合。为了保持一致性，我们采用了来自 [15] 的结果用于 J-DNN 和 NBMLSS 模型，包括 RevIn 和 vincentization 组件。

测试集上的概率预测结果见表 II。连续排名概率得分（CRPS）使用 99 个百分位数计算的平均 Pinball 损失来近似。我们报告了 5% 显著性水平下 Kupiec 无条件覆盖检验对于 50% 和 90% 预测区间的检验结果，以及预测区间覆盖率概率（PICP）。图 1 展示了多元 Diebold-Mariano (DM) 检验的结果，比较了模型预测的 CRPS 损失范数。此外，百分位数的整体覆盖性能在图 2 – 3 中进行了说明。QR-DNN 基线实现了与分布模型相当

表 II: 测试集结果

德	PICP _{50%}	PICP _{90%}	PICP _{98%}	MAE	CRPS
J-DNN	54.3(13)	91.4(17)	97.8(23)	10.499	3.809
QR-DNN	40.6(1)	83.1(2)	95.7(13)	10.629	3.858
NBMLSS	53.8(18)	91.3(24)	97.9(23)	10.230	3.728
QNBM	52.2(18)	91.1(23)	98.1(24)	10.411	3.789
是	PICP _{50%}	PICP _{90%}	PICP _{98%}	MAE	CRPS
J-DNN	48.5(20)	89.0(21)	97.0(19)	13.431	4.847
QR-DNN	40.7(0)	84.4(4)	96.4(15)	13.432	4.863
NBMLSS	48.7(21)	87.9(19)	97.1(20)	12.758	4.644
QNBM	47.3(18)	88.9(19)	97.9(23)	12.826	4.653

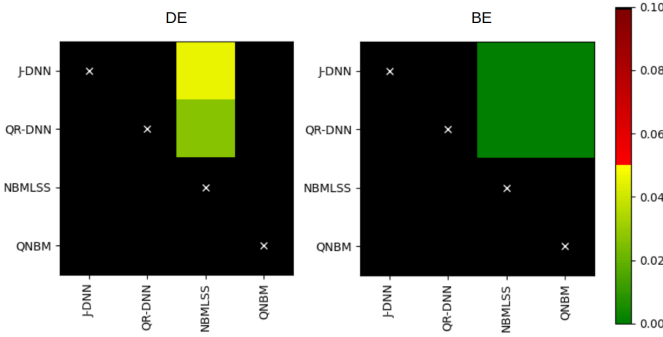


图 1: DM 测试的测试集 CRPS 分数

的 CRPS 和 MAE 指标，但在中心分位数上的预测区间表现出更高的置信度，特别是在 50% 和 98% 水平处覆盖率差异明显。这种行为可能归因于潜在的分位数过拟合问题，这是 QR-DNN 模型易发生的。有必要进行进一步的研究，例如探索额外的正则化技术。QNBM 模型在得分和校准方面与 NBMLSS 架构表现出一致的表现。此外，我们观察到 QR-DNN 和 QNBM 模型都没有显示出分布在零附近和负价格结算处分布模型所见的巨大尾部峰值。DE 测试集上预测分位数样本的结果见图 4。这种行为可能归因于它们通过非参数设置对极端位置分位数的放置提供了更精细的控制，而不是由分布假设强加的更为受限的形式。

除了其概率预测能力外，QNBM 架构的主要目标类似于 NBMLSS 模型，即为用户提供每个输入变量如何影响预测分布特征的可解释描述，从而揭示驱动模型输出的关键因素。图 5 – 6 说明了在 DE 市场上，跨整个时间段对外生变量影响预测分位数 $\gamma = 0.05$ 和 $\gamma = 0.95$ 提取的形状函数。提取遵循应用于 NBMLSS 研究最终测试样本的掩码外生子集方案，确保一致性。不同的颜色表示不同的集成组件，每个组件都来自于单

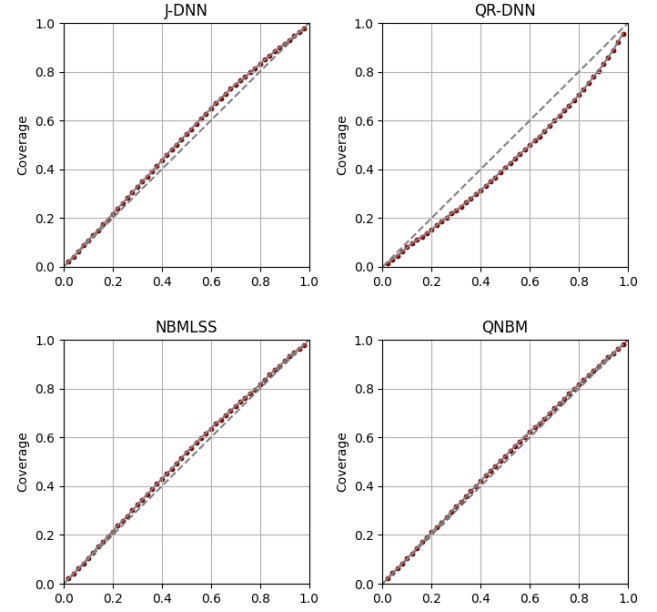


图 2: DE 校准图

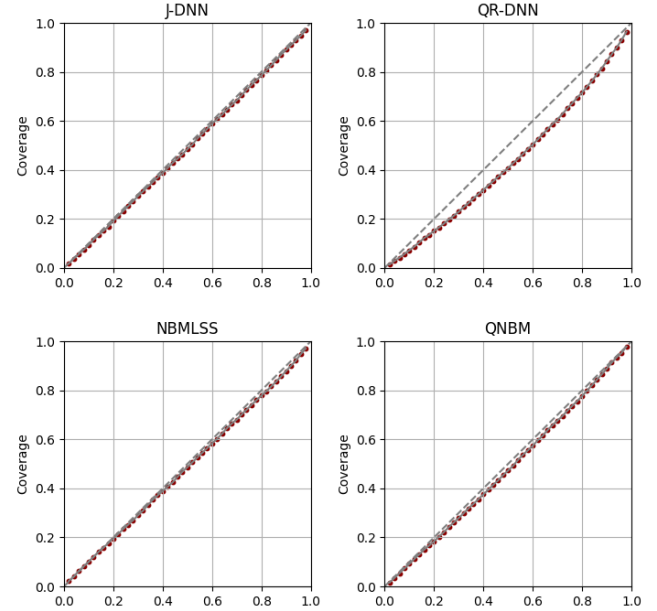


图 3: 贝校准图

独的重新校准运行。显然，学习管道在多次执行中一致收敛到了类似的形状函数，显示出与在 NBMLSS 模型（见 [15]）中观察到的关系相当的结果。这些包括可再生能源发电预测的影响减弱以及负荷特征在高峰时段相比清晨和傍晚具有更陡峭的效果。然而，QNBM 架构继承了特征之间潜在的近似共轭性（参见例如 [25], [26]）以及现代机器学习管道中常见的总体欠指定问题（参见例如 [27], [28], [29] 及其参考文献）。正如 [15] 也

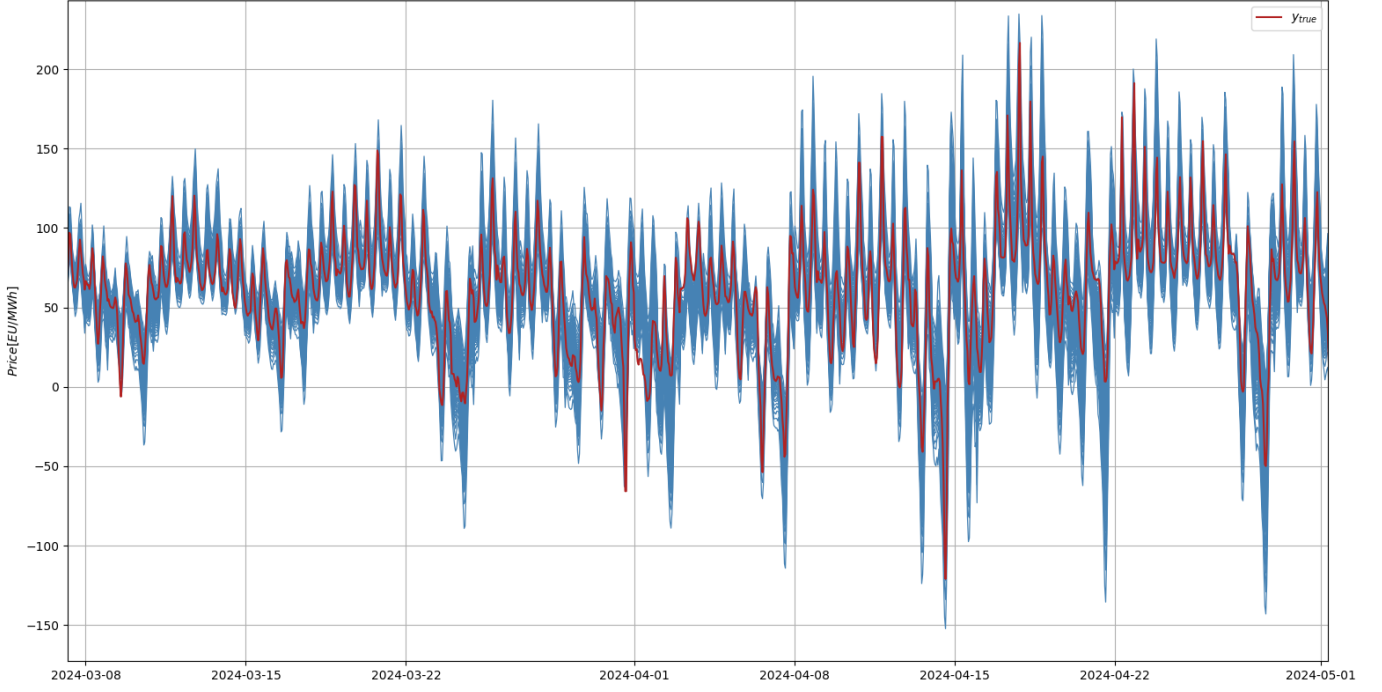


图 4: DE 市场测试集上预测分布的百分位数样本

指出的那样，形状函数之间的非线性交互作用在文献中尚未被彻底研究。因此，我们呼应 [29] 的建议，有效的特征选择应该包含领域专家的意见。此外，与其依赖单一解决方案，采用候选模型集合目前代表了最务实的折衷方案。这种考虑在预测环境中尤为重要，在这些环境中，条件变量内的冗余信息比例外更常见——例如，在清晨时段价格值之间的相关性。在此框架内，神经加法模型如 NBMLSS 和 QNBM 可以补充 D-DNNs 和 QR-DNNs 的灵活性，提供在整个输入域中特征效应更为透明的解释，并帮助用户在关键的模型开发阶段。

IV. 结论与下一步发展

在这项工作中，我们介绍了分位数神经基础模型 (QNBM)，通过显式的形状函数表示提供了量化回归神经网络 (QR-DNN) 的一种更可解释的替代方案，类似于位置、尺度和形状神经基础模型 (NBMLSS) 在分布神经网络 (D-DNN) 中所起的作用。通过避免参数化的分布假设，QNBM 能够更加灵活地表示特定条件分位数。相比之下，NBMLSS 以参数化方式建模整个条件分布，当选择的参数族不正确时，这可能会限制其捕捉复杂分位数行为的能力。然而，从参数级到分位数级 (例如，百分位) 映射的转换引入了计算挑战，特别是在增加条件维度和多时段设置的情况下。为了解决这个问题，我们在基础函数和形状函数投影中引入矩阵分解，

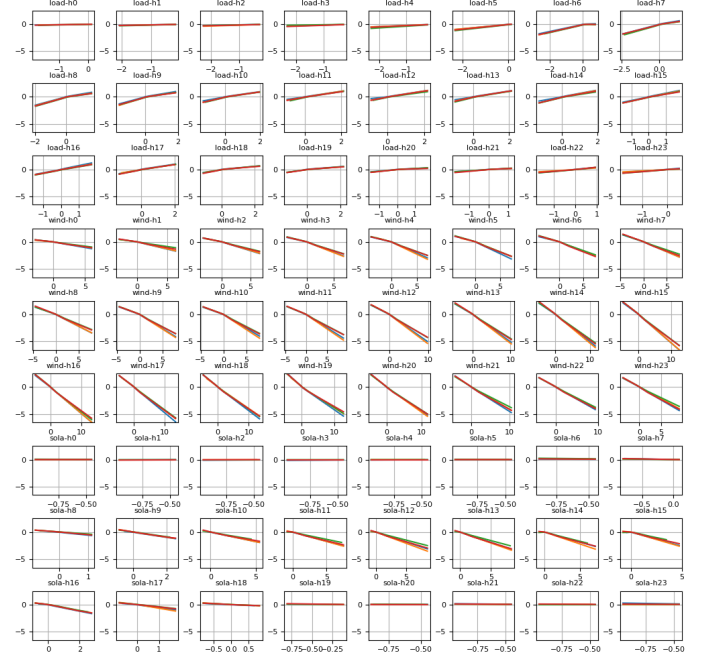


图 5: 预测分位数 $\gamma = 0.05$ 在 DE 重新校准 (5 个集成组件) 的形状函数

超越共享的神经网络骨干结构。所提出的方法应用于德国和比利时电力市场的日前电价预测这一具有挑战性的任务，涵盖近期特征表现为高波动性的时期。实验设置遵循 NBMLSS 研究以确保一致性和可比性。QNBM

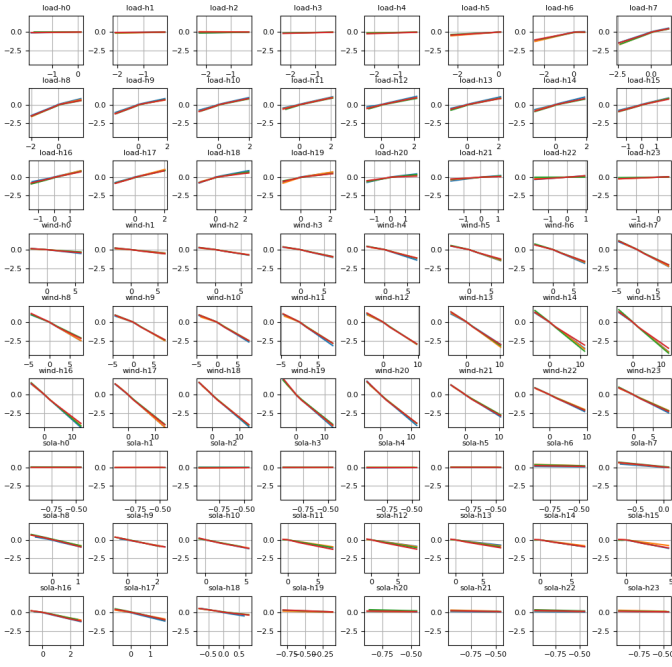


图 6: 预测分位数 $\gamma = 0.95$ 在 DE 重新校准 (5 个集成组件) 的形状函数

模型展示了与 NBMLSS 架构相当的表现, 同时缓解了分布公式中观察到的过度负尖峰问题。总体而言, 通过 D-DNN/NBMLSS 和 QR-DNN/QNBM 模型分别示例化的分布建模和非参数建模方法代表了互补的选择。它们的实际部署或集成应根据应用的具体需求进行指导, 在可解释性、灵活性以及底层数据特征之间取得平衡。

几个未来研究的方向仍然开放。我们计划深入调查共曲线问题, 例如通过利用 Rashomon 集合近似技术以及探索旨在控制多变量共曲线的自动特征选择方法。此外, 我们打算探索进一步正则化策略和架构改进的结合, 并将我们的研究扩展到更广泛的预测应用范围。

参考文献

- [1] A. Mashlakov, T. Kuronen, L. Lensu, A. Kaarna, and S. Honkapuro, "Assessing the performance of deep learning models for multivariate probabilistic energy forecasting," *Applied Energy*, vol. 285, p. 116405, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261920317748>
- [2] A. Ciarreta, B. Martinez, and S. Nasirov, "Forecasting electricity prices using bid data," *International Journal of Forecasting*, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207022000711>
- [3] S. Madadkhani and S. Ikonnikova, "Toward high-resolution projection of electricity prices: A machine learning approach to quantifying the effects of high fuel and co2 prices," *Energy Economics*, vol. 129, p. 107241, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0140988323007399>

- [4] A. Brusaferri, M. Matteucci, P. Portolani, and A. Vitali, "Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices," *Applied Energy*, vol. 250, pp. 1158–1175, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261919309237>
- [5] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1548–1568, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364032117308808>
- [6] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark," *Applied Energy*, vol. 293, p. 116983, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261921004529>
- [7] G. Marcjasz, M. Narajewski, R. Weron, and F. Nov, "Distributional neural networks for electricity price forecasting," *Energy Economics*, vol. 125, p. 106843, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0140988323003419>
- [8] A. Brusaferri, A. Ballarino, L. Grossi, and F. Laurini, "Online conformalized neural networks ensembles for probabilistic forecasting of day-ahead electricity prices," *Applied Energy*, vol. 398, p. 126412, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261925011420>
- [9] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Electricity price forecasting on the day-ahead market using machine learning," *Applied Energy*, vol. 313, p. 118752, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261922002057>
- [10] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, pp. 206–215, 05 2019.
- [11] R. Agarwal, L. Melnick, N. Frosst, X. Zhang, B. Lengerich, R. Caruana, and G. E. Hinton, "Neural additive models: interpretable machine learning with neural nets," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, ser. NIPS '21. Red Hook, NY, USA: Curran Associates Inc., 2024.
- [12] W. Jo and D. Kim, "Neural additive time-series models: Explainable deep learning for multivariate time-series prediction," *Expert Systems with Applications*, vol. 228, p. 120307, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423008096>
- [13] A. Thielmann, R.-M. Kruse, T. Kneib, and B. Säfken, "Neural additive models for location scale and shape: A framework for interpretable neural regression beyond the mean," in *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Dasgupta, S. Mandt, and Y. Li, Eds., vol. 238. PMLR, 02–04 May 2024, pp. 1783–1791. [Online]. Available: <https://proceedings.mlr.press/v238/frederik-thielmann24a.html>
- [14] S. Hirsch, J. Berrisch, and F. Ziel, "Online distributional regression," 2024. [Online]. Available: <https://arxiv.org/abs/2407.08750>
- [15] A. Brusaferri, D. Ramin, and A. Ballarino, "Nbmlss: probabilistic forecasting of electricity prices via neural basis models for location scale and shape," 2024. [Online]. Available: <https://arxiv.org/abs/2411.13921>

- [16] F. Radenovic, A. Dubey, and D. Mahajan, “Neural basis models for interpretability,” in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS ’22. Red Hook, NY, USA: Curran Associates Inc., 2024.
- [17] P. Gaillard, Y. Goude, and R. Nedellec, “Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 1038–1050, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207015001545>
- [18] M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude, “qgam: Bayesian nonparametric quantile regression modeling in r,” *Journal of Statistical Software*, vol. 100, no. 9, p. 1 – 31, 2021. [Online]. Available: <https://www.jstatsoft.org/index.php/jss/article/view/v100i09>
- [19] K. Bouchiat, A. Immer, H. Yèche, G. Rätsch, and V. Fortuin, “Improving neural additive models with bayesian principles,” 2024. [Online]. Available: <https://arxiv.org/abs/2305.16905>
- [20] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” 2021. [Online]. Available: <https://arxiv.org/abs/2106.09685>
- [21] T. Kim, J. Kim, Y. Tae, C. Park, J.-H. Choi, and J. Choo, “Reversible instance normalization for accurate time-series forecasting against distribution shift,” in *International Conference on Learning Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=cGDAkQo1C0p>
- [22] K. Aliyon and J. Ritvanen, “Deep learning-based electricity price forecasting: Findings on price predictability and european electricity markets,” *Energy*, vol. 308, p. 132877, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544224026513>
- [23] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” 2019. [Online]. Available: <https://arxiv.org/abs/1907.10902>
- [24] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, “Tensorflow distributions,” *arXiv preprint arXiv:1711.10604*, 2017. [Online]. Available: <https://arxiv.org/abs/1711.10604>
- [25] C. Zhong, Z. Chen, J. Liu, M. Seltzer, and C. Rudin, “Exploring and interacting with the set of good sparse generalized additive models,” in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS ’23. Red Hook, NY, USA: Curran Associates Inc., 2024.
- [26] L. Kovács, “Feature selection algorithms in generalized additive models under concavity,” *Comput. Stat.*, vol. 39, no. 2, p. 461 – 493, Nov. 2022. [Online]. Available: <https://doi.org/10.1007/s00180-022-01292-7>
- [27] A. D’Amour, K. Heller, D. Moldovan, B. Adlam, B. Alipanahi, A. Beutel, C. Chen, J. Deaton, J. Eisenstein, M. D. Hoffman, F. Hormozdiari, N. Houlsby, S. Hou, G. Jerfel, A. Karthikesalingam, M. Lucic, Y. Ma, C. McLean, D. Mincu, A. Mitani, A. Montanari, Z. Nado, V. Natarajan, C. Nielson, T. F. Osborne, R. Raman, K. Ramasamy, R. Sayres, J. Schrouff, M. Seneviratne, S. Sequeira, H. Suresh, V. Veitch, M. Vladymyrov, X. Wang, K. Webster, S. Yadlowsky, T. Yun, X. Zhai, and D. Sculley, “Underspecification presents challenges for credibility in modern machine learning,” *J. Mach. Learn. Res.*, vol. 23, Jan. 2022.
- [28] J. Siems, K. Ditschuneit, W. Ripken, A. Lindborg, M. Schambach, J. S. Otterbach, and M. Genzel, “Curve your enthusiasm: concavity regularization in differentiable generalized additive models,” in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS ’23. Curran Associates Inc., 2024.
- [29] X. Zhang, J. Martinelli, and S. John, “Challenges in interpretability of additive models,” *IJCAI 2024 Workshop on Explainable Artificial Intelligence*, 2024.