2021.11.4

概率预测

主要关注内容：概率预测的实现及评估方法，希望对概率预测有些概念。具体领域包括**需求预测**（特别希望看到intermittent demand）以及**电力预测**（有GEFCcom2014、GEFCom2017比赛的支持）。前者为主。

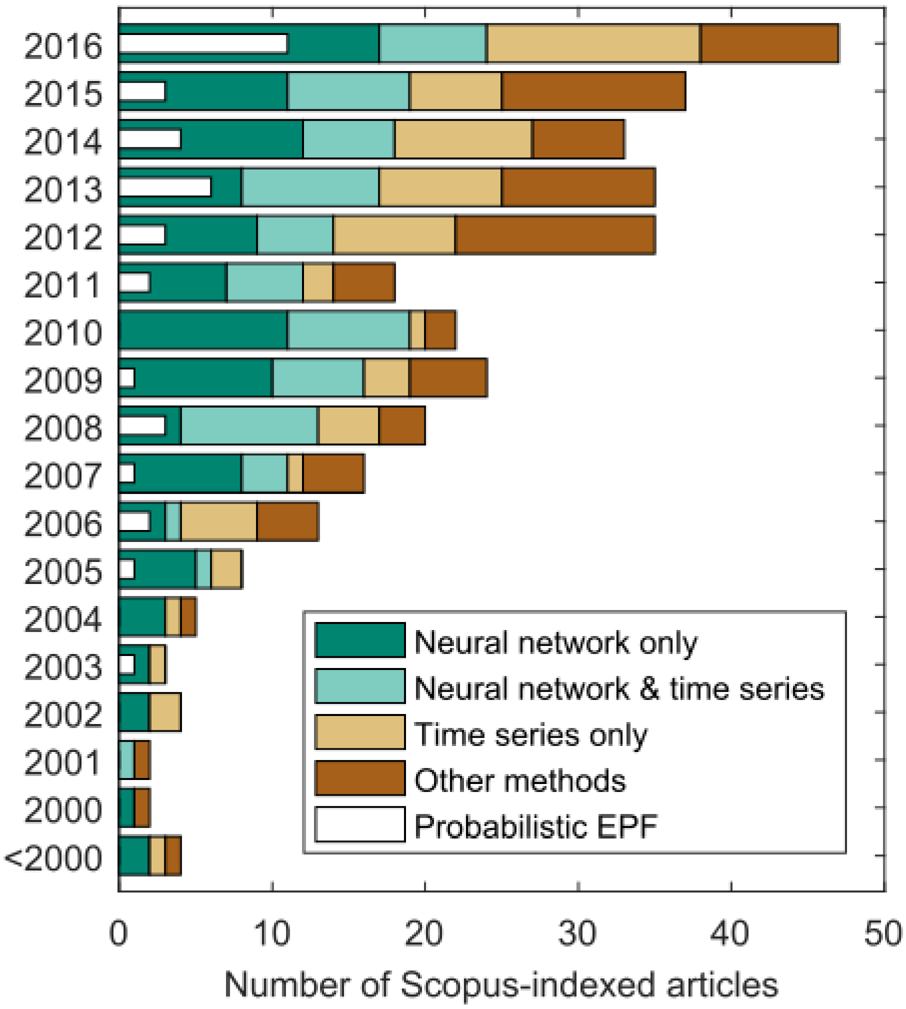
综述：*Recent advances in electricity price forecasting: A review of probabilistic forecasting*

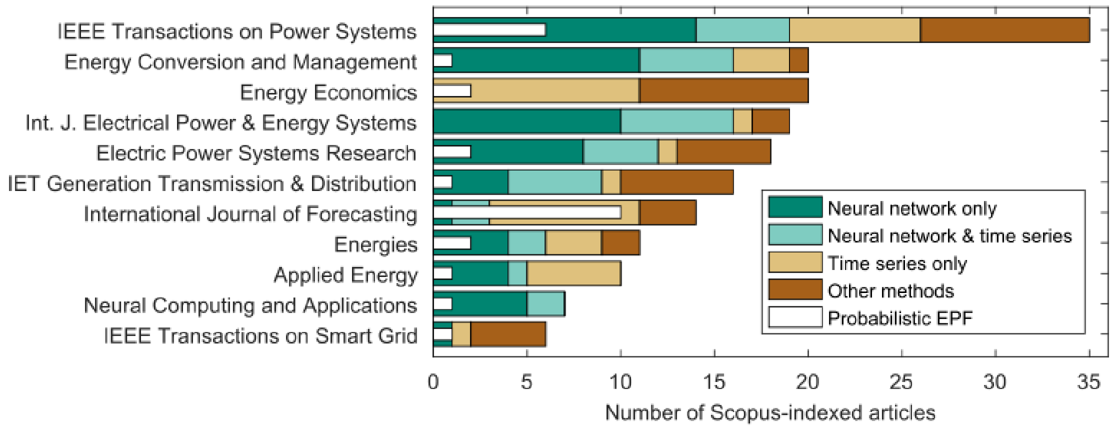
(Jakub Nowotarski, Rafał Weron, 2018) (Renewable and Sustainable Energy Reviews)

本文是综述文章，从电价预测与GEFCcom2014结果展开。

研究情况

在电价预测领域，研究逐年增长，概率预测也基本如此（但比重不大，最高也就20%）。该领域预测方法以神经网络为主。而主要期刊的方法分布也呈现出学科背景的差异：偏工程的期刊更重视网络模型的使用，而偏经济的期刊则以时间序列统计方法为主；此外，IJF最关注概率预测问题。





GEFCom2014比赛的代表方法

- 第一名：对不同任务使用3种不同模型，分别是：**广义加性模型（GAM）&分位数回归**；**分位数回归平均组合**（自回归、回归、GAM、随机森林、梯度提升等13模型组合）；**核分位数Lasso回归。**
- 第二名：以**分位数回归平均**（QRA）为主，分为**点预测**（两个结构相同、样本不同自回归）、**预过滤**、**分位数回归建模**（结合之前自回归与外生变量）和**后处理**（平滑分位数曲线以避免分位数交叉）四步。
- 第三名：**一个仅有5个sigmoid激活函数的MLP**（单隐藏层5个神经元），假设误差服从正态分布。
- 第四名：一个**带正则项的分位数回归，交叉验证选择变量**，乘子交替方向法(ADMM)求解损失函数最小化。

（这个时期，还是统计方法更能占优，与对Kaggle比赛评估的论文 Kaggle forecasting competitions: An overlooked learning

opportunity (Bojer & Meldgaard, 2020）结果相近）

概率预测概述

作者认为概率预测方法可分为四类：

- **历史模拟或经验预测区间**：举例为 Value at Risk （VaR）的历史模拟，独立于模型，计算误差经验分布的样本分位数。
- **基于分布的概率预测**：对于以正态误差驱动的模型（ARIMA等），密度预测可以设置为近似误差密度的高斯分布。（个人感觉之前所看概率预测组合中所用的单一方法基本都属于此，比如AR、VAR、GARCH类、基于经济理论的一般均衡模型（统计本质还是线性回归）。）
- **Bootstrap预测区间**：先估计模型的参数与残差样本，之后使用参数与残差不断生成伪数据，用伪数据拟合模型并预测，得到预测区间。（常见于神经网络中，应该是先做了点预测再考虑误差）
- **分位数回归平均**：直接估计分位数与自变量的关系，不借助分布。

评价与指标

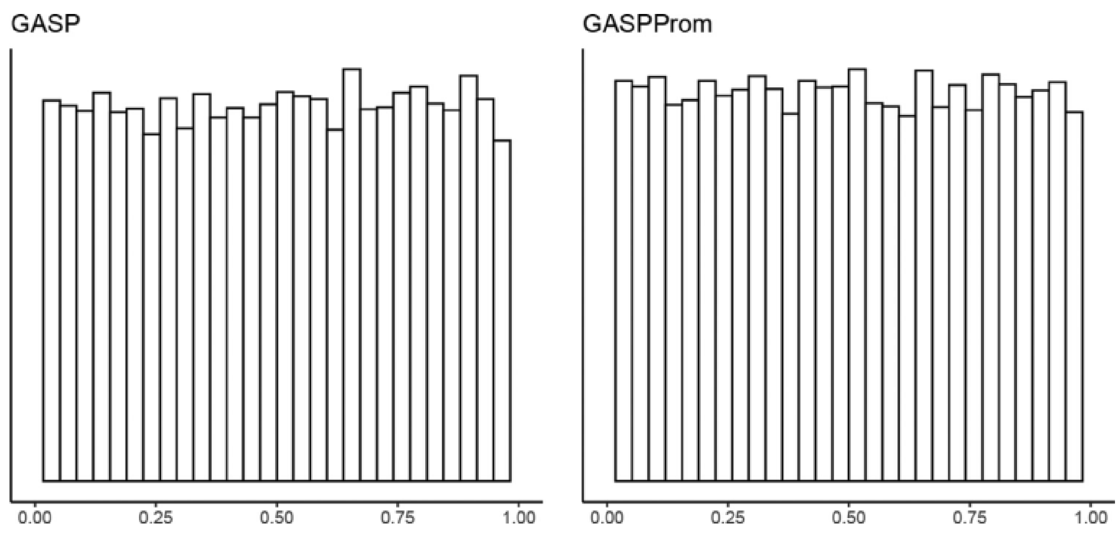
（上一周分享应该也提到了这个内容）

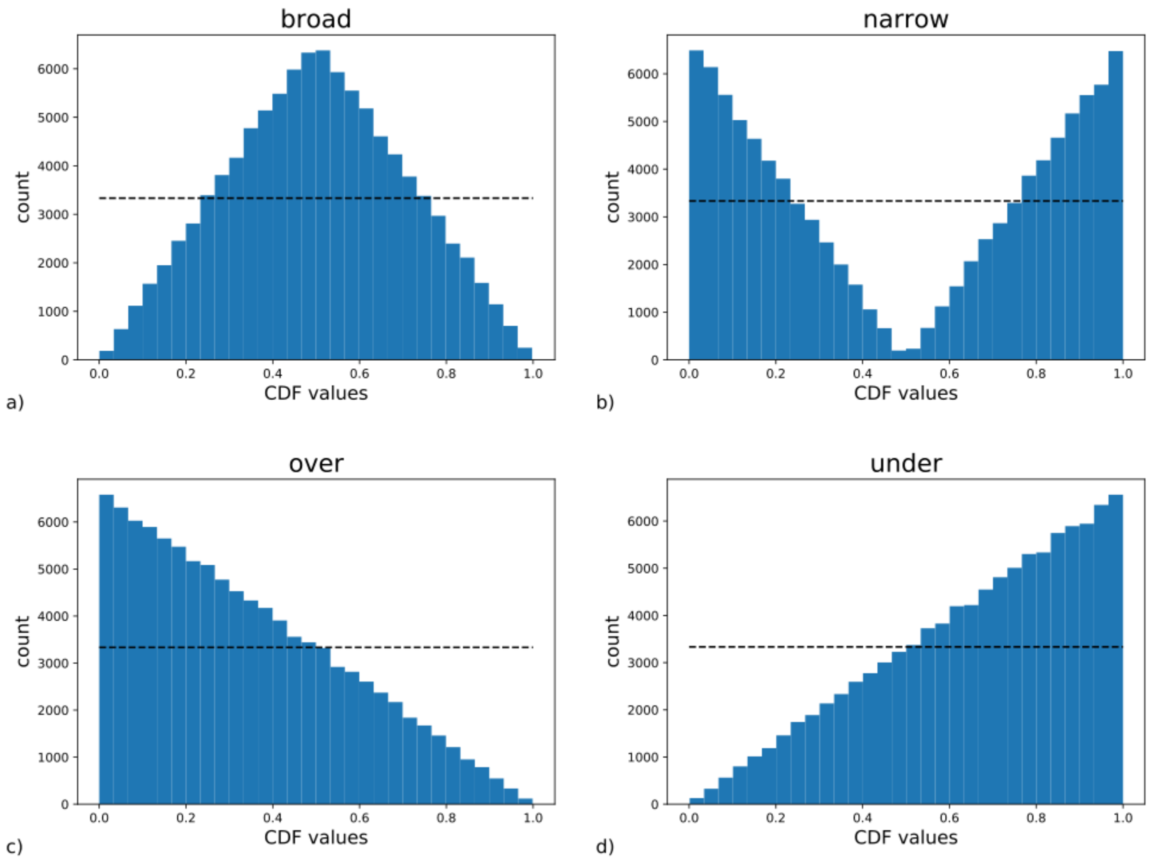
Gneting et al. (2014) 指出概率预测的目标是 “**maximize the sharpness** of the predictive distributions, **subject to reliability**” 。

- **Reliability**：文献中又称 calibration 或 unbiasedness，指预测与观测的一致性（如，90%预测区间覆盖90%观测值）。
- **Sharpness**：衡量预测区间有多紧密的覆盖真实数据，希望预测宽度尽量窄。（预测的浓度）

（或许可以和效度与信度类比？）

Reliability的检验：**Kupiec test**（非条件覆盖检验，构造预测值是否击中真实区间的示性量，再进行似然比检验）； **Christoffersen test**（条件覆盖检验，在前者基础上引入条件、独立性检验）；也可用拟合优度检验；使用**概率积分变换（PIT）进行可视化检验**，要求最后出现的直方图符合一个 [0,1] 均匀分布。（最后一个还很常用）





Sharpness评价：与评分规则（损失函数？）挂钩，之前提到的对数分数属于此。

Pinball Loss：分段线性损失，也是分位数回归中要最小化的函数。

$$Pinball(\hat{Q}_{P_t}(q), P_t, q) = \begin{cases} (1-q)(\hat{Q}_{P_t}(q) - P_t), & \text{for } P_t < \hat{Q}_{P_t}(q), \\ q(P_t - \hat{Q}_{P_t}(q)), & \text{for } P_t \geq \hat{Q}_{P_t}(q), \end{cases}$$

Winkler score：一个可以同时评估Reliability和区间宽度的指标：

$$Winkler_t = \begin{cases} \delta_t, & \text{for } P_t \in [\hat{L}_t, \hat{U}_t], \\ \delta_t + \frac{2}{\alpha}(\hat{L}_t - P_t), & \text{for } P_t < \hat{L}_t, \\ \delta_t + \frac{2}{\alpha}(P_t - \hat{U}_t), & \text{for } P_t > \hat{U}_t, \end{cases}$$

Continuous Ranked Probability Score (CRPS)：为解决对数得分不够稳健的问题，使用此评分。积分改成求和就是离散化的指标DRPS。

$$CRPS(\hat{F}_{P_t}, P_t) = \int_{-\infty}^{\infty} (\hat{F}_{P_t}(x) - \mathbb{1}_{\{P_t \leq x\}})^2 dx,$$

以上指标可以作为评价标准，也可以作为训练中优化的函数。

上次讲论文中有提到结论 **Statistical methods is enough, emphasise sharpness is not recommended** (Mitchell and Wallis，2011)； 问题何在?

需求概率预测

Forecasting the intermittent demand for slow-moving inventories: A modelling approach

(Snyder et al., 2012) (International Journal of Forecasting)

假设了两个间断需求模型：

1. 时间连续，交易零星：每个时期内交易次数服从泊松分布，考虑交易规模服从对数序列，则每个时期的综合需求量呈**负二项分布**；
2. 时间离散，被分成大致相等几段： **hurdle shifted Poisson process** （原始的泊松分布向正值平移1，然后概率为 1-p 取 0，p 取正值）

此外还假设了分布均值随时间变化的情况，分为3类：

1. Static：恒定不变
2. Damped dynamic：平稳过程，受均值与真实值的一期滞后影响
3. Undamped dynamic：非平稳过程

Table 2
Recurrence relationships for the mean.

Relationship	Recurrence relationship	Restrictions
Static	$\mu_t = \mu_{t-1}$	
Damped dynamic	$\mu_t = (1-\phi-\alpha)\mu + \phi\mu_{t-1} + \alpha y_{t-1}$	$\mu > 0, \phi > 0, \alpha > 0$ $\phi + \alpha < 1$
Undamped dynamic	$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}$	$\delta > 0, \alpha > 0$ $\delta + \alpha = 1$

对hurdle shifted Poisson process ，其正值概率 p 也可以按如上的方式时变。以上模型使用极大似然估计，并在预测中前向模拟以获得预测序列的分布。

此外作者提到了一个基于Croston的概率假设 (Hyndman et al, 2008) ，其认为时间间隔服从负二项分布，正需求服从shifted poisson 分布，两个分布参数预测时用简单指数平滑更新（需要注意：向前预测时会出现长期收敛到定值的问题！）

概率预测的评价指标为预测似然分数(PLS)与DRPS（CRPS的离散版本）。

作者认为此方法对库存管理是有用的。

Intermittent Demand Forecasting with Deep Renewal Processes

(Türkmen et al., 2019) （Arxiv文章，貌似没见刊，而且正文仅4页，不够完善；但思路可以参考，因为 intermittent demand 概率预测实在不多）

时间间隔服从负二项分布（参数与上一时刻的时间间隔估计有关），需求大小服从泊松分布（参数与上一时刻的需求大小估计有关），而在在Croston中估计是由指数平滑完成的。据此，其将上述模型改为：

时间间隔服从一非负离散分布（负二项分布）或连续分布（伽马分布），需求大小服从一非负离散分布（负二项分布），而分布所需的参数等性质则依赖于LSTM生成。

$$Q_i \sim \mathcal{NN}\left(g_q(\mathbf{h}_i)\right) \qquad M_i \sim \mathcal{NN}\left(g_m(\mathbf{h}_i)\right) \qquad \mathbf{h}_i = LSTM_{\Theta}(\mathbf{h}_{i-1}, Q_i, M_i)$$

DeepAR: Probabilistic forecasting with autoregressive recurrent networks

(Salinas et al., 2020) (International Journal of Forecasting)

提出**DeepAR**解决零售业预测问题（可能是M5选手参考的文章之一？）

作者认为贡献在于：（1）提出了用于概率预测的RNN，结合计数数据负二项似然；（2）表明基于现代深度学习的方法可以有效地解决概率预测问题。而作者认为的方法优势有4点：

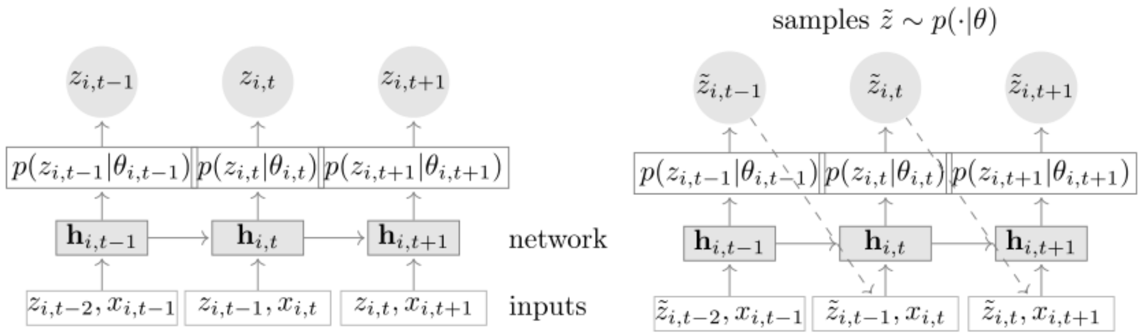
- 捕捉季节性、相关性所需人工干预少
- 以蒙特卡洛方式采样预测，可计算预测范围内任何子范围的一致估计
- 通过类似的项目可以预测较短时间序列
- 不假设高斯噪声，可以选择其它统计分布

预测方法：

目标是估计时间序列未来的条件分布（给定过去的序列与外生变量）

$$P(\mathbf{z}_{i,t_0:T} | \mathbf{z}_{i,1:t_0-1}, \mathbf{x}_{i,1:T})$$

训练与预测时基于如下的自回归循环神经网络。左侧是训练部分，接受前一时刻的输出以及数据信息，训练概率分布，生成预测；右侧是预测部分，每次预测时将预测结果代入下一时刻进行迭代，生成一条预测序列，再重复多次；实践中保持所有训练样本的条件和预测范围的总长度T和相对长度都是固定的。（这里有一个小问题：训练时上一时刻序列值是真值，但预测时只能由上一时刻预测值产生之后的预测，造成一个脱节的问题；但是作者认为这对本文的结果影响不大。）



文中考虑了正态与负二项两种分布，但是其它分布也可以使用。分布的参数与h挂钩，通过一个连接函数与h的线性表示相连（类似于广义线性回归）。优化的损失函数为对数似然函数。

学习中不需要细致的超参调整，即可用于不同数据集，而且不需要大量的数据集（文中认为几百条序列就可以应用）。

A score-driven model of short-term demand forecasting for retail distribution centers

(Hoeltgebaum et al., 2021) (Journal of Retailing)

所用方法为一个**广义自回归分数模型**（GAS）；针对的数据为有一定汇总的销售数据，故未考虑间断性，但是考虑了数据的偏态，文中使用的是对数正态分布。

模型的结构如下所示，这里假定均值时变而方差不变。（方差时变的模型作者也尝试了，但估计结果一般。）方程通过ARMA的平稳可逆条件即可保证模型的平稳性。

$$\log q_t = \mu_t + \epsilon_t$$

$$\mu_t = \omega + \sum_{i=1}^{p} \phi_i \mu_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

$$\text{where } \epsilon_t \sim \text{NID}(0, \sigma^2).$$

$$p(q_t|\mathcal{F}_{t-1}) = \frac{1}{q_t\sigma\sqrt{2\pi}}e^{-\frac{(\ln(q_t)-\mu_t)^2}{2\sigma^2}}, \quad t = 1, \ldots, T$$

此外，通过引入连接函数的方式，可以将外生变量纳入模型中：

$$E(q_t|\mathcal{F}_{t-1}) = e^{\mu_t + \gamma' \Lambda_t + \frac{\sigma^2}{2}}$$

所考虑的外生变量包括星期、假日、价格、促销行为哑变量等。

预测时也是用蒙特卡洛法不断采样去得到分布。、

实证：使用巴西一家大型零售企业的数据，总体销售层次；与线性回归、ETS、神经网络进行比较，点预测占优，而且reliability和sharpness较好。

Demand Forecasting of Individual Probability Density Functions with Machine Learning

(Wick et al., 2021) (Operations Research Forum)

主要的方法是 **Cyclic Boosting**，一个可解释的机器学习方法。

文中使用负二项分布，对两个参数（均值与方差）分别估计。均值解释成各因子的乘积，而特征又循环的由g生成。通过迭代得到估计。另一个参数通过优化负对数似然得到。

$$\hat{\mu}_i = c \cdot \prod_{j=1}^{p} f_j^k \quad \text{with } k = \{x_{j,i} \in b_j^k\}$$

$$g_{j,t}^k = \frac{\sum_{x_{j,i}\in b_j^k} y_i}{\sum_{x_{j,i}\in b_j^k} \hat{\mu}_{i,\tau}} \quad \text{where } f_{j,t}^k = \prod_{s=1}^{t} g_{j,s}^k$$

为了避免“时间混杂”（Temporal Confounding）的问题，**不直接使用任何自身的滞后项进行建模**，以保证可解释性（与M5第二名的特征选择类似）；而使用指数平滑来捕获误差修正项，以使用序列自身的信息优化结果。

使用了M5食物的100条序列进行实证，发现由于零膨胀等问题，负二项分布在分布两端的预测不佳，但比泊松模型强很多。

其它领域概率预测

Density forecasting of daily electricity demand with ARMA-GARCH, CAViaR, and CARE econometric models

(Can Bikcoraa, Lennart Verheijenb, Siep Weilanda, 2018)（Sustainable Energy, Grids and Networks)

电力预测，使用模型：

- ARIMAX-GARCH：交叉验证来筛选模型，估计方法尝试了迭代最小二乘、非线性最小二乘和拟极大似然，使用蒙特卡洛前向传播方法估计密度（根据模型生成大量前向预测）。
- 分位数回归：法1是CAViaR，对过去的测量值和分位数本身的过去值进行分位数回归；法2是CARE，是法1的期望分位数回归

Probabilistic Forecasting with Spline Quantile Function RNNs

(Gasthaus et al., 2019) the 22nd International Conference on Artificial Intelligence and Statistics(会议论文)

使用方法：单调回归样条对条件分位数函数进行概率建模；样条的形状由神经网络（RNN)参数化；神经网络的参数是通过最小化连续排序概率得分来学习的(CRPS)。

在模拟和电力、交通、wiki、域名四个数据集进行了实验。（但预测结果的评估还是用的点预测指标）

Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression

(He et al., 2021) (Neurocomputing) 风电功率预测

使用方法：**三次样条插值+支持向量回归**，利用得到的分位数使用核密度估计；样条插值目的是对特异点进行调整以使模型预测结果更优。（疑问：这会不会影响概率预测对极端值的把握，从而使概率预测对极端情况更不重视）

考虑了点预测与区间预测的评价；使用预测区间归一化平均宽度(PINAW)和预测区间覆盖概率(PICP)来度量区间预测；为进一步衡量，将二者组合成CWC指标（注意：在第一篇综述中明确指出这个指标有问题！）

个人想法

还是想以intermittent demand为主进行概率组合预测。而individual model或许可以如下尝试：

历史模拟或经验预测区间：可以用WSS与VZ等基于历史数据的Bootstrap方法进行概率预测生成。

基于分布的概率预测：

- 基于负二项/tweedie/零膨胀等分布的模型，想尝试DeepAR（M5的头部模型主要是LGBM和DeepAR，后者点预测第三名、不确定预测第四名都有应用）；
- 还可以尝试 *Forecasting the intermittent demand for slow-moving inventories: A modelling approach* 一文的方法；
- 考虑需求间隔&需求规模（Croston系列）或者需求发生概率&需求规模（TSB类）构造两个概率，结合基本模型的指数平滑更新方式进行概率预测，而基本更新方式（分布参数的指数平滑）也可能涉及SBA、SBJ、mSBA、mTSB等......

分位数回归平均：感觉可以直接用相关外生与滞后变量进行分位数回归，得到分位数估计。可以参考GEFCom2014比赛的头部模型等。（2017比赛的相关内容还没关注）

同时，对于概率预测组合，可能有以下新情况：

- 不同分布类型的概率进行组合——会对组合的推导产生影响吗
- 参数模型与非参数模型进行组合——使用非参结果的“经验分布”，或者干脆组合方式进行调整