ETH zürich

# 概率性负荷预测研究与展望

王 毅

苏黎世联邦理工学院

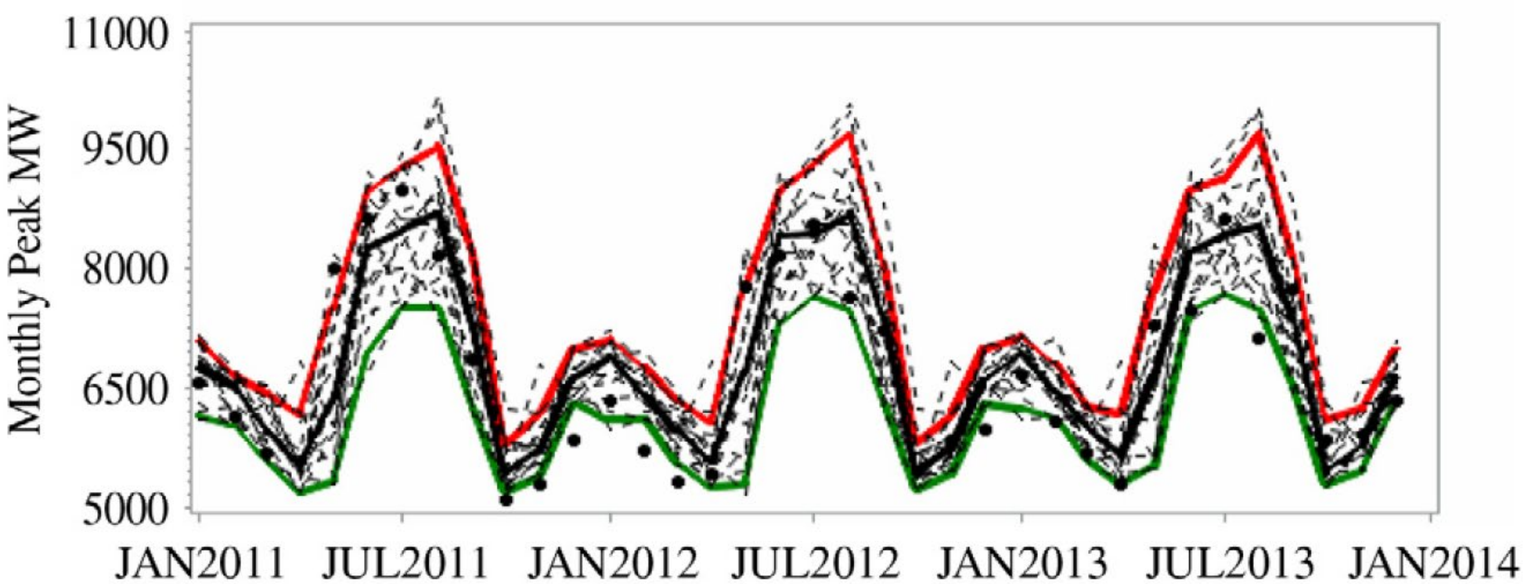
2020年12月13日

# 背景介绍

ETHzürich

## 什么是概率预测?

**概率预测一般通过一系列分位数、区间、概率密度等方式表现。**




$$Var[Y^*|X=x^*] = Var[Y^* - \hat{m}(x^*)|X=x^*] + Var[\hat{m}(x^*)|X=x^*]$$

**Predictions are ineluctably vitiated by errors, originating from noise in the explanatory variables (e.g. due to the chaotic nature of weather conditions) as well as model misspecifications.**

## Off-the-shelf regression methods

Y=f(X)

- Linear regression
- ANN (Artificial Neural Network)
- SVM (Support Vector Machine)
- GBRT (Gradient Boosting Regression Tree)
- RF (Random Forest)

- **Quantile regression**
- **Gaussian Process regression**

- **Hate tedious mathematic derivation**
- **Put more emphasis on how load forecasting method works**

# 概率性负荷预测研究进展

ETHzürich

传统点预测可以抽象为“输入-模型-输出” 三部分，从传统点预测拓展到概率预测也需要从这三部分着手：

输入 → 模型 → 输出

| | 输入 | 模型 | 输出 |
|---|---|---|---|
| 点预测到概率预测的拓展 | 1.1输入的特征选择问题 | 1.2构建分位数回归等概率模型 | 1.3对预测误差进行建模 |
| 以集成预测为核心的建模 | 2.1点预测集成 | 2.2分位数集成 | 2.3概率密度集成 |
| 挖掘预测新对象 | 3.1单一用户预测<br>3.4集成负荷预测 | 3.2峰值预测<br>3.5场景预测 | 3.3净负荷预测<br>3.6在线预测 |

# 点预测到概率预测的拓展

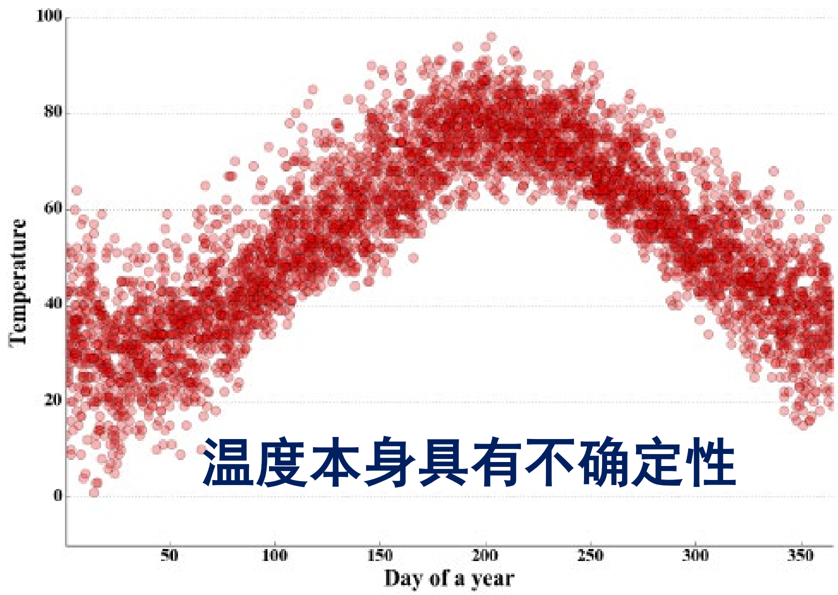
ETH zürich

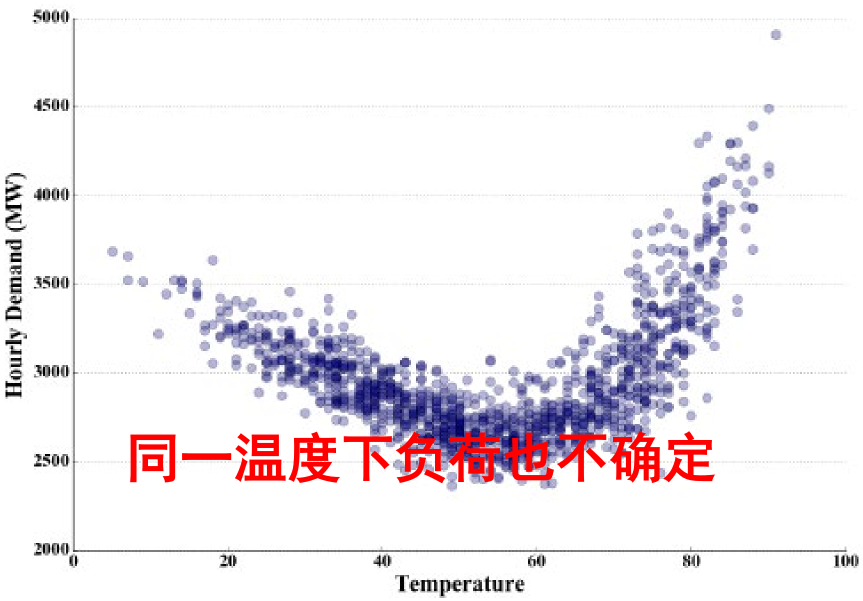
## ◆ 研究进展1.1：分位数回归模型

背景：负荷和温度具有较强的相关性，传统预测方法在引入温度变量时没有充分考虑未来温度的不确定性？

难点：如何挖掘温度等输入数据的不确定？如何描述模型带来的不确定性？如何将这两种不确定性同时考虑，更好的构建待预测负荷的概率模型？

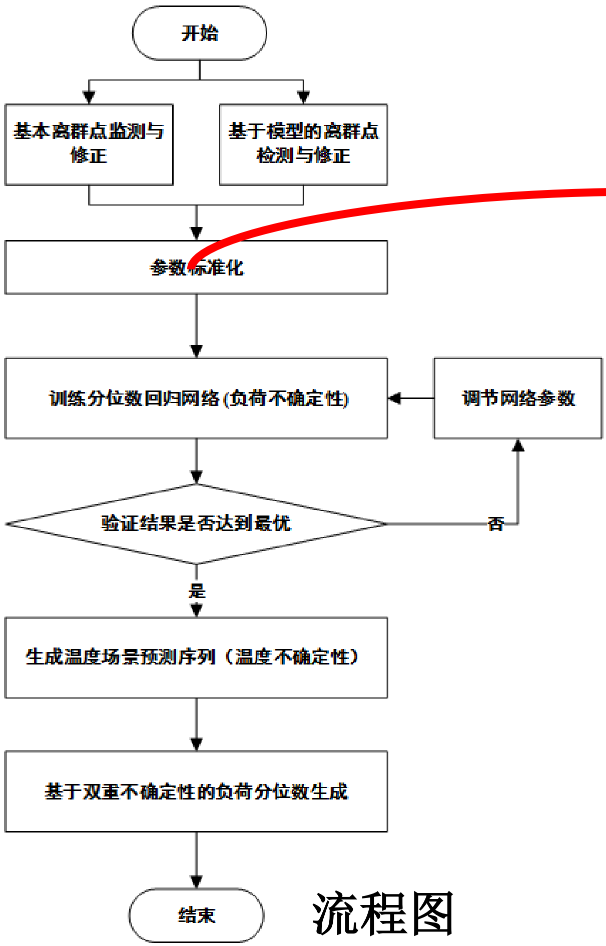
ISO New England 2004-2015小时级数据



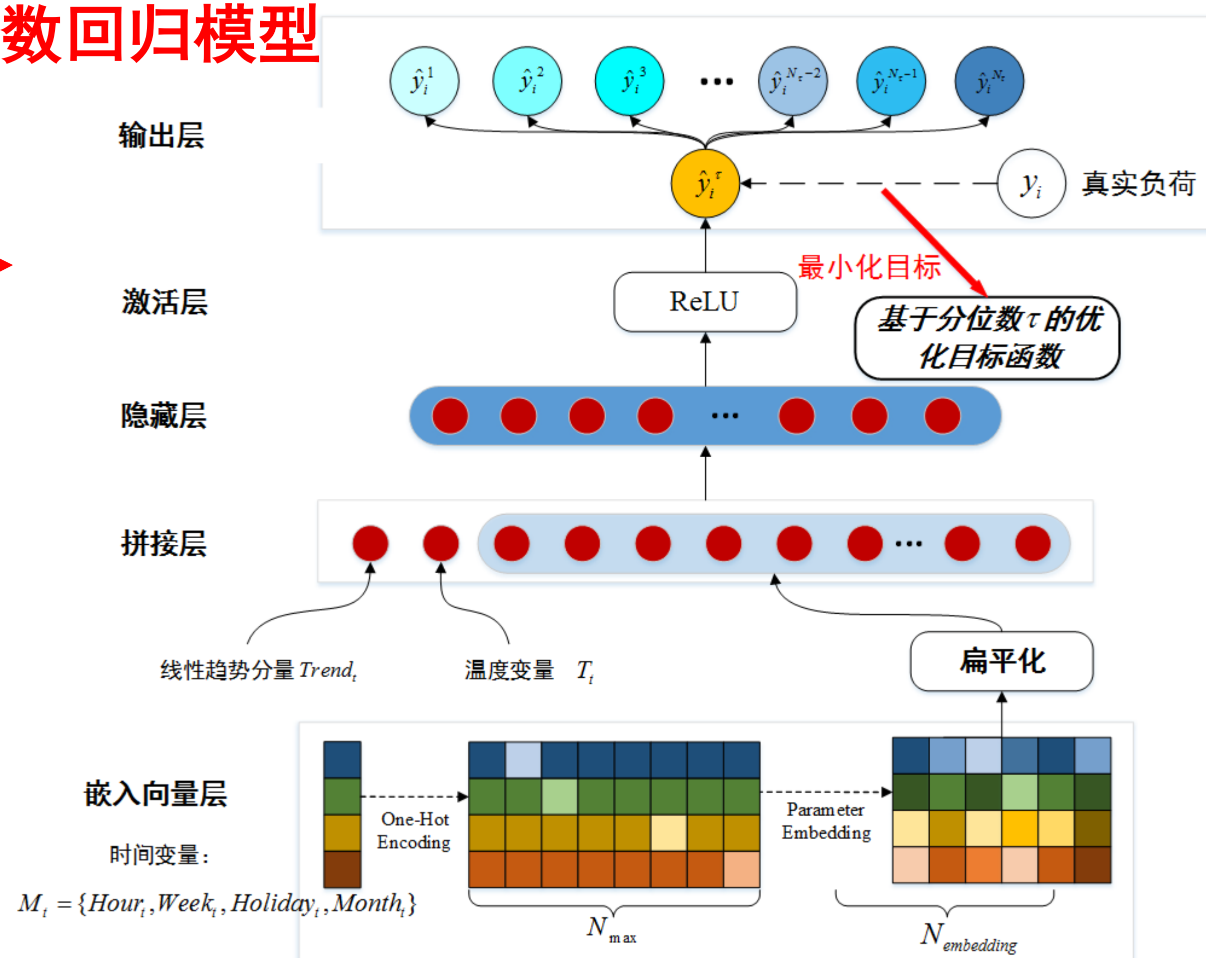



# 点预测到概率预测的拓展

## ◆ 研究进展1. 1：分位数回归模型




流程图



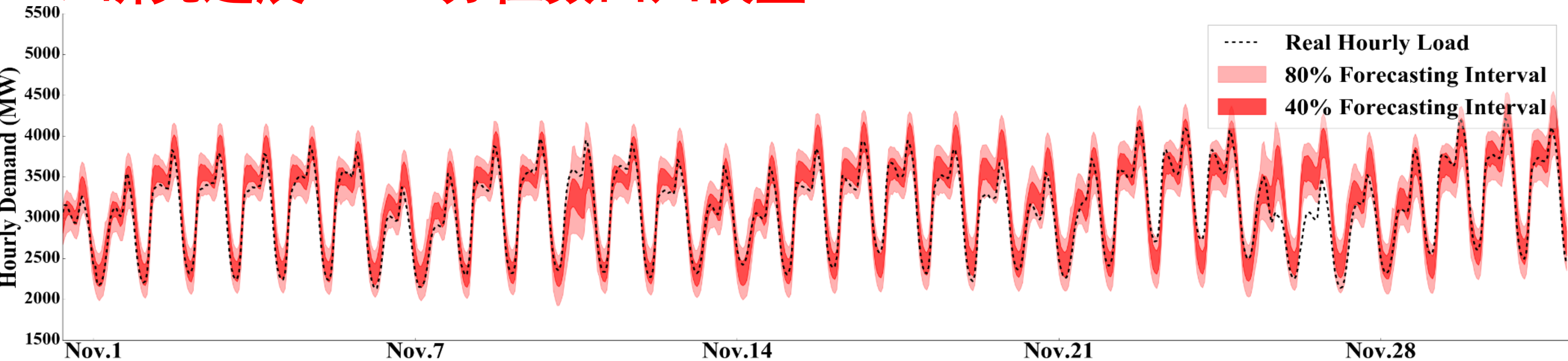

分位数回归网络示意图

方案：生成多个温度场景描述温度不确定性；构建分位数回归模型描述负荷不确定性；最后将所有场景重新整合。

# 点预测到概率预测的拓展

ETHzürich

## ◆ 研究进展1.1：分位数回归模型

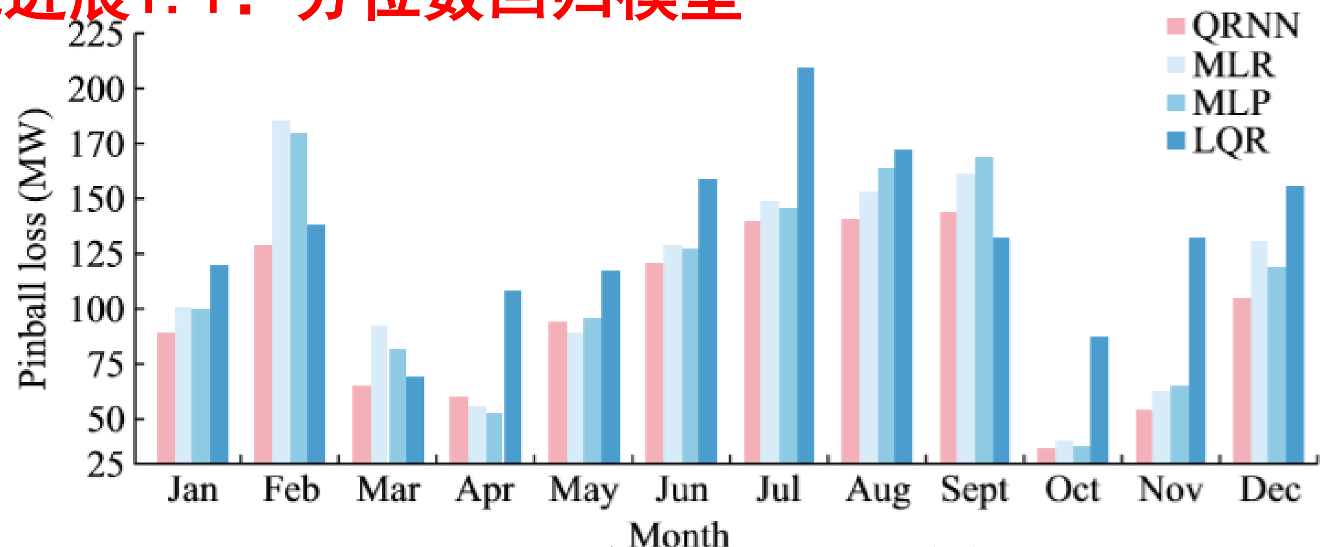



| Zone | QRNN | MLR | MLP | LQR | MaxRI |
|---|---|---|---|---|---|
| CT | **104.9** | 111.8 | 110.8 | 133.1 | 21.2% |
| SEMA | **52.0** | 56.9 | 54.3 | 62.3 | 16.5% |
| NEMA | **75.7** | 84.4 | 81.2 | 96.9 | 21.9% |
| WCMA | **51.6** | 56.8 | 55.6 | 69.2 | 25.4% |
| VT | 15.3 | 14.8 | **14.7** | 19.6 | 21.9% |
| NH | **31.1** | 33.8 | 33.1 | 37.8 | 17.7% |
| RI | **25.9** | 28.1 | 27.3 | 31.7 | 18.3% |
| ME | **22.1** | 25.8 | 25.1 | 27.1 | 18.5% |

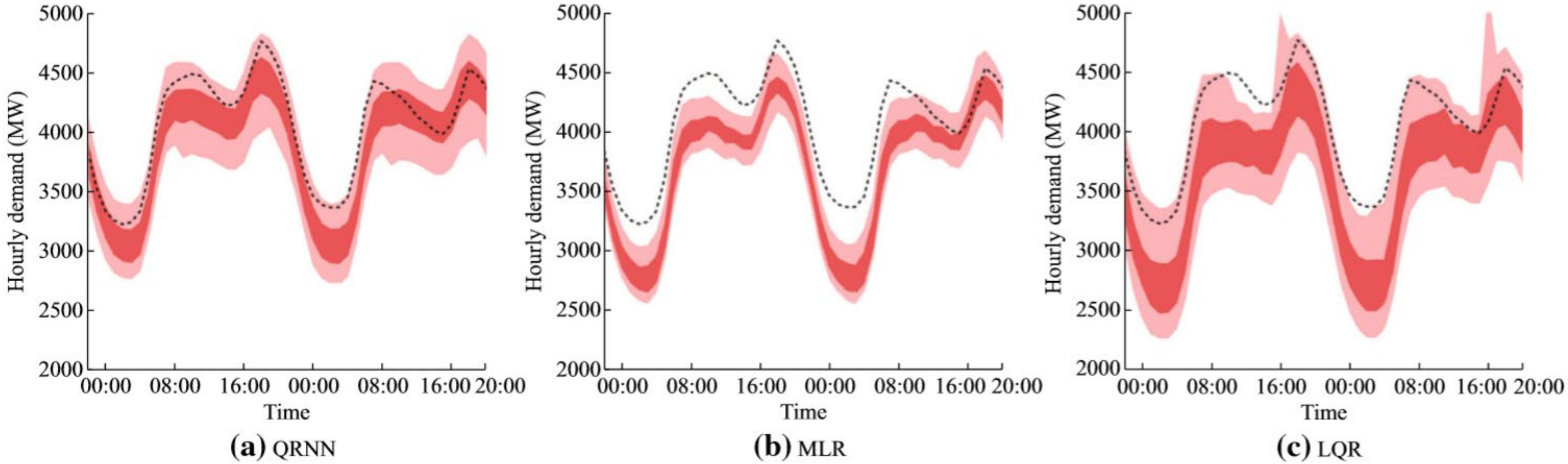
**2015年11月预测结果（已知2014年及以前数据）**

**Pinball Loss结果，越低越好**
(MLR, MLP, LQR均为对比用基本模型)
(MaxRI代表最大相对提升度)

◆ 研究进展1.1：分位数回归模型




不同月份的**Pinball loss**对比




不同不同方法的预测结果对比

## ◆ 研究进展1.2：特征选择问题

**目前已经有较多基于点预测的特征选择方法，LASSO就是其中之一。**

**一个直观的问题是：LASSO能否也应用与概率预测呢？**

这是一个常用且好用的预测模型……

$$\hat{y}_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) + \underbrace{\sum_{d=1}^{N_D} f(\widetilde{T}_{t,d}) + \sum_{h=1}^{N_H} f(T_{t-h})}_{\text{recency effects}}$$

然而，模型的输入变量实在太多……

$$\begin{aligned} N_F =& 1+11+6+23+23\times 6 \\ &+ (3+3\times 11+3\times 23)(1+N_D+N_H) \\ =& 284+105(N_D+N_H) \end{aligned}$$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{N_T} l(r_t) + \lambda \|\boldsymbol{\beta}\|_1; \quad r_t = y_t - \boldsymbol{\beta}^T \mathbf{X}_t$$

$$\hat{\boldsymbol{\beta}}_q = \arg\min_{\boldsymbol{\beta}_q} \sum_{t=1}^{N_T} \rho_q(r_{q,t}) + \lambda_q \|\boldsymbol{\beta}_q\|_1; \; r_{q,t} = y_t - \boldsymbol{\beta}_q^T \mathbf{X}_t$$

# 点预测到概率预测的拓展

ETHzürich

## ◆ 研究进展1.2：特征选择问题

$$\hat{\boldsymbol{\beta}}_q = \arg\min_{\boldsymbol{\beta}_q} \sum_{t=1}^{N_T} \rho_q(r_{q,t}) + \lambda_q \|\boldsymbol{\beta}_q\|_1;\ r_{q,t} = y_t - \boldsymbol{\beta}_q^T \mathbf{X}_t$$

求取拉格朗日函数

$$\begin{aligned} L_\gamma(\mathbf{r}, \boldsymbol{\beta}, \mathbf{u}) = &\rho_q(\mathbf{r}) + \lambda \|\boldsymbol{\beta}\|_1 + \mathbf{u}^T(\mathbf{y} - \boldsymbol{\beta}^T \mathbf{X} - \mathbf{r}) \\ &+ \frac{\gamma}{2}\|\mathbf{y} - \boldsymbol{\beta}^T \mathbf{X} - \mathbf{r}\|_2^2 \end{aligned}$$

基于ADMM的迭代

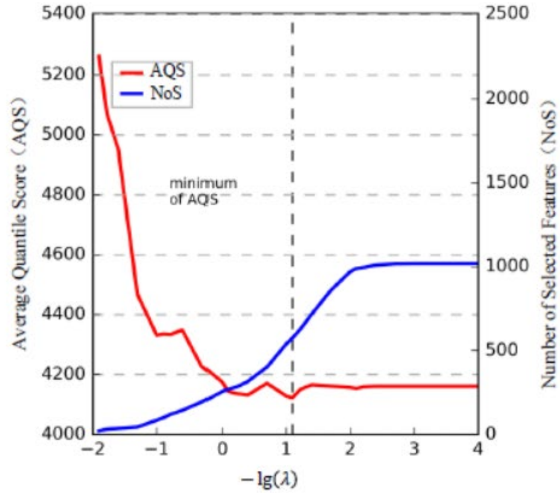
$$\begin{aligned} \boldsymbol{\beta}^{k+1} &:= \arg\min_{\boldsymbol{\beta}} L_\gamma(\boldsymbol{\beta}, \mathbf{r}^k, \mathbf{u}^k) \\ \mathbf{r}^{k+1} &:= \arg\min_{\mathbf{r}} L_\gamma(\boldsymbol{\beta}^{k+1}, \mathbf{r}, \mathbf{u}^k) \\ \mathbf{u}^{k+1} &:= \mathbf{u}^k + \gamma(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{k+1} - \mathbf{r}^{k+1}) \end{aligned}$$

$$\begin{aligned} \mathbf{r}^{k+1} := &\arg\min_{\mathbf{r}} \rho_q(\mathbf{r}) + \mathbf{u}^{(k)T}\mathbf{s} + \frac{\gamma}{2}\|\mathbf{s}\|_2^2 \\ &\arg\min_{\mathbf{r}} \rho_q(\mathbf{r}) + \frac{\gamma}{2}\|\mathbf{s} + (1/\gamma)\mathbf{u}^k\|_2^2 - \frac{1}{2\gamma}\|\mathbf{u}^k\|_2^2 \\ &\arg\min_{\mathbf{r}} \rho_q(\mathbf{r}) + \frac{\gamma}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{k+1} - \mathbf{r} + \frac{1}{\gamma}\mathbf{u}^k\|_2^2 \end{aligned}$$

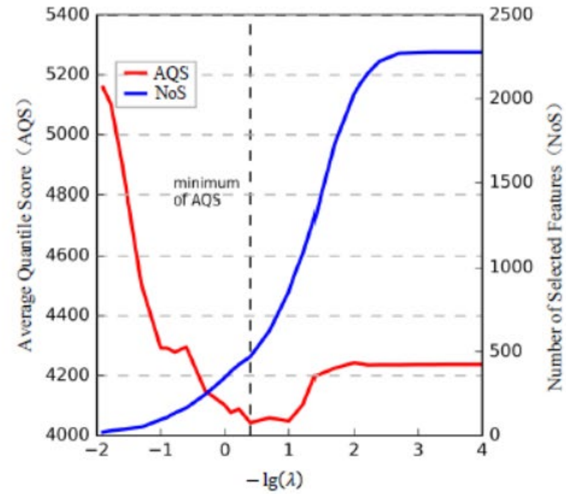
# 点预测到概率预测的拓展

ETH zürich

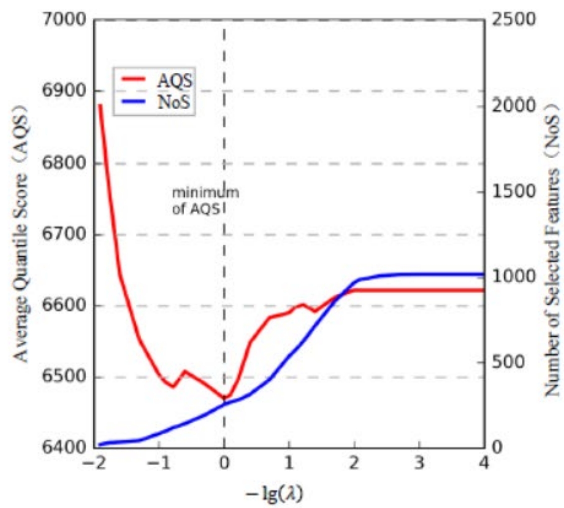
## ◆ 研究进展1.2：特征选择问题



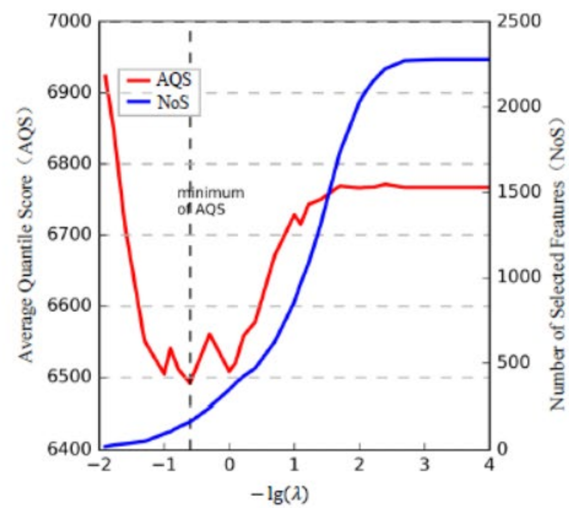

(a) D3-H4 model




(b) D7-H12 model




(a) D3-H4 model



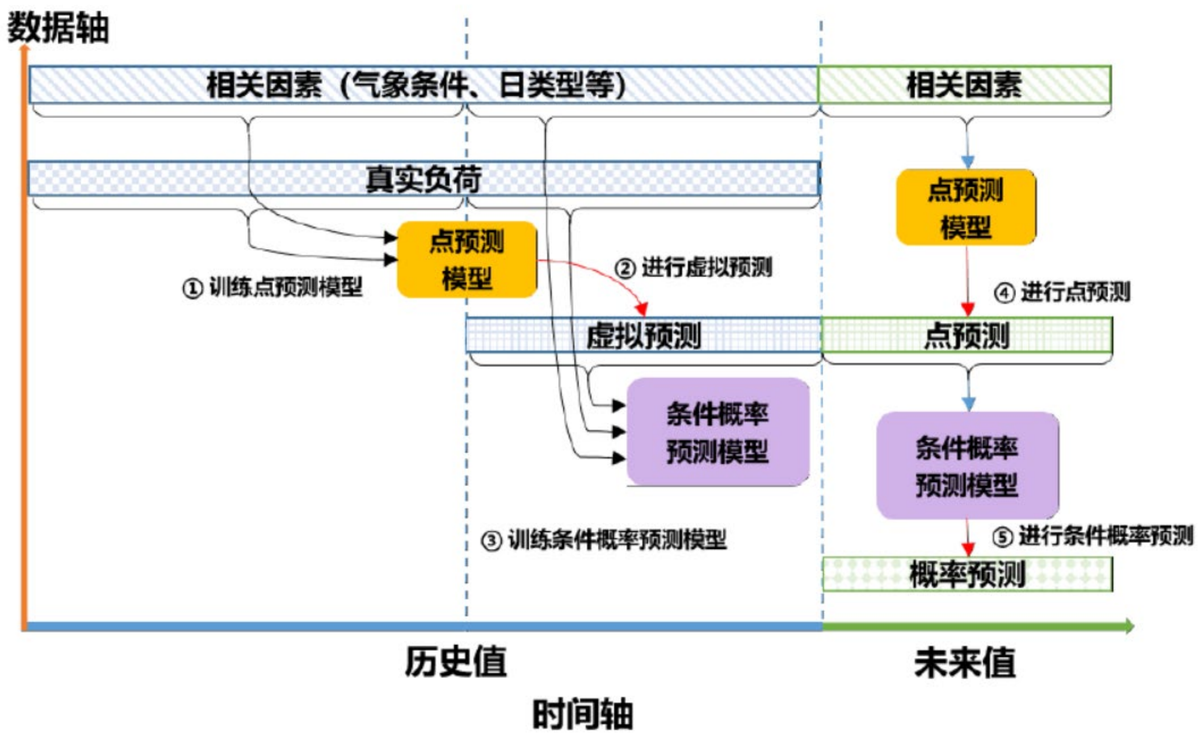

(b) D7-H12 model

# 点预测到概率预测的拓展

ETH zürich

## ◆ 研究进展1.3：条件残差建模

背景：目前已经有很多成熟的点预测方法，概率预测能够更好的利用点预测结果呢？
难点：对点预测的残差进行建模能够得到概率预测结果，但是预测残差不同时段不同，不同负荷水平不同，受到很多条件的影响，不方便直接统计建模。



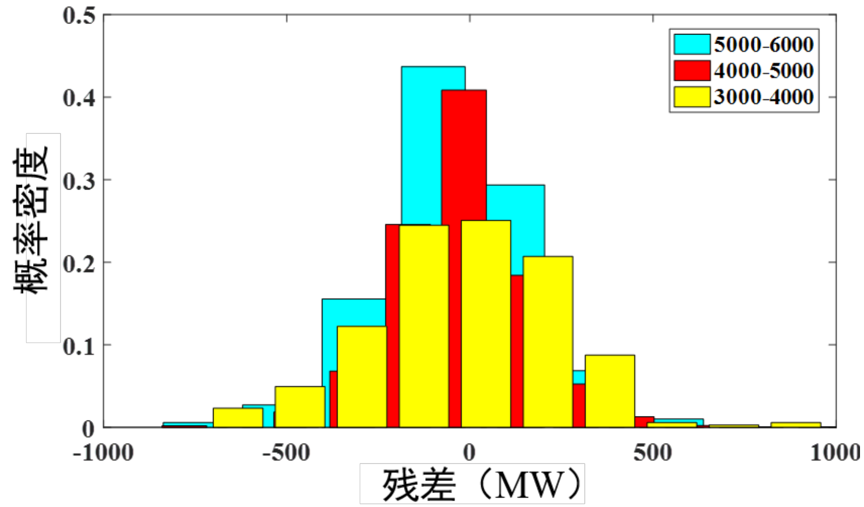

$$\epsilon_{t,q} = g_q(\mathbf{W}_q, [\mathbf{X}_t, \boxed{\hat{y}_t}])$$

把点预测结果作为分位数回归模型的输入，残差作为拟合对象，构建残差相对于输入和预测值的条件概率预测模型

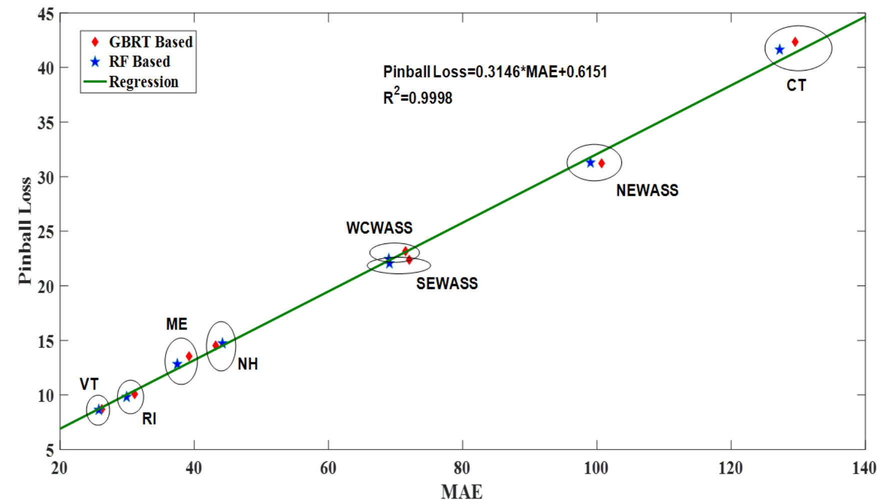
# 点预测到概率预测的拓展

ETH zürich

## ◆ 研究进展1.3：条件残差建模




预测残差分布




概率预测误差和点预测误差之间的线性关系

效果：一方面能够有效提过预测精度，相对于直接分位数回归能提高10%左右！
另一方面，可以发现点预测精度很大程度上决定了概率预测结果！

| | | CT | ME | NH | RI | VT | NE | SE | WC | SYS |
|---|---|---|---|---|---|---|---|---|---|---|
| 直接预测 | Pinball | 47.02 | 13.92 | 16.34 | 11.15 | 9.51 | 35.75 | 25.40 | 25.52 | 164.74 |
| | WS | 749.17 | 213.70 | 263.15 | 179.10 | 150.56 | 561.60 | 429.26 | 405.81 | 2610.44 |
| 条件残差建模 | MAPE | 3.78 | 2.95 | 3.36 | 3.28 | 4.21 | 3.55 | 4.17 | 3.68 | 3.16 |
| | RMSE | 171.82 | 49.06 | 59.74 | 40.60 | 34.81 | 133.21 | 94.54 | 91.85 | 582.09 |
| | MAE | 127.25 | 37.44 | 44.22 | 29.95 | 25.79 | 99.01 | 68.99 | 69.06 | 433.44 |
| | Pinball | **41.61** | **12.82** | **14.69** | **9.79** | **8.66** | **31.27** | **22.47** | **22.06** | **137.85** |
| | WS | **666.62** | **201.03** | **237.47** | **155.31** | **147.04** | **506.61** | **370.21** | **357.49** | **2261.74** |
| 改进(%) | Pinball | 11.50 | 7.86 | 10.11 | 12.23 | 8.95 | 12.55 | 11.52 | 13.55 | 16.32 |
| | WS | 11.02 | 5.93 | 9.76 | 13.28 | 2.34 | 9.79 | 13.76 | 11.91 | 13.36 |

# 点预测到概率预测的拓展

## ◆ 研究进展2. 1：点预测的概率集成

**Basic Idea**

**Uncertainty decomposition:**

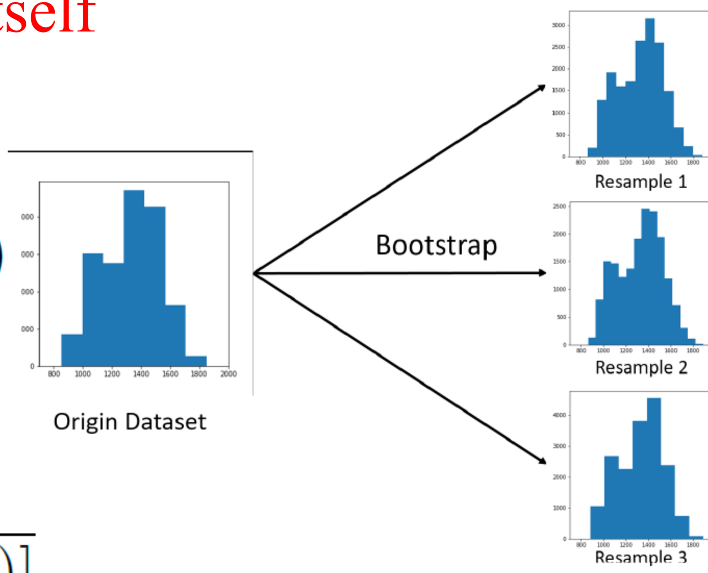
$$Var[Y^*|X=x^*] = Var[Y^* - \hat{m}(x^*)|X=x^*] + Var[\hat{m}(x^*)|X=x^*]$$

1) The possible errors that Y* fall beside the point forecast

2) The uncertainty of the model m(x*) itself

**According to the central limit theory:**

$$\frac{Y^* - E[\hat{Y}^*]}{\sqrt{Var[Y^* - \hat{m}(x^*)] + Var[\hat{m}(x^*)]}} \sim N(0,1)$$



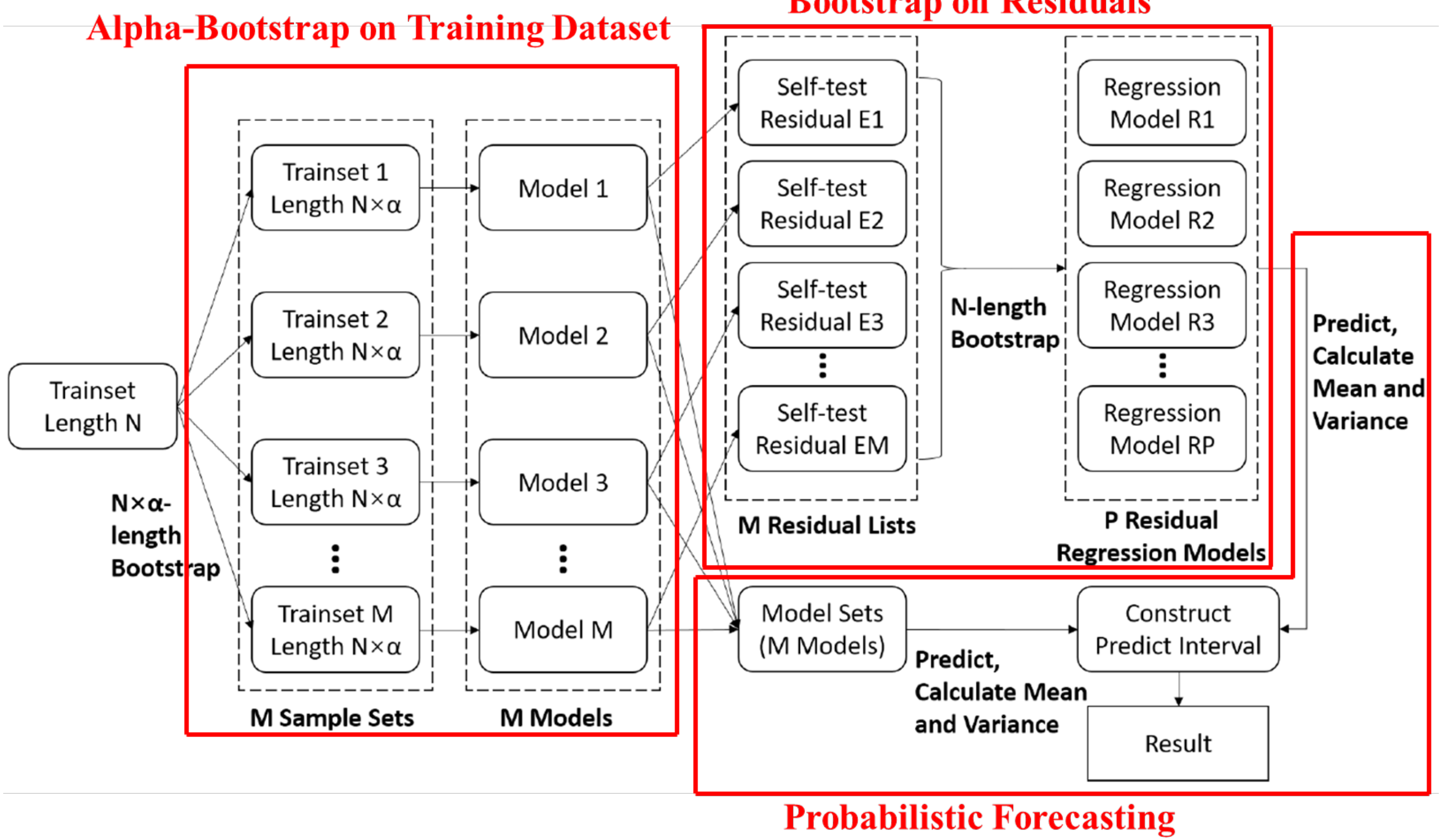

**Calculate the quantiles:**

$$\overline{\hat{Y}^*} \pm z_{1-\beta/2}\sqrt{Var[Y^* - \hat{m}(x^*)] + Var[\hat{m}(x^*)]}$$

# 点预测到概率预测的拓展

## ◆ 研究进展2. 1：点预测的概率集成

### Framework



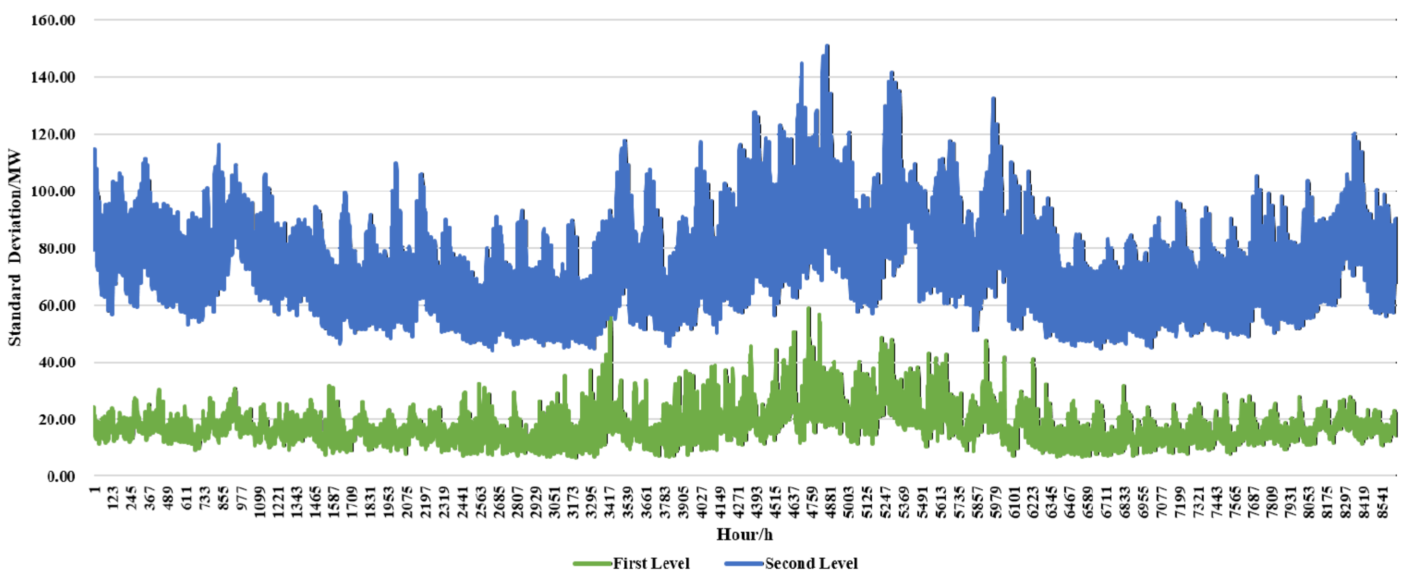
# 点预测到概率预测的拓展

ETHzürich

## ◆ 研究进展2.1：点预测的概率集成

Compare with Quantile regression

| | Random Forest Based | | | | | GBRT Based | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MAPE | RMSE | PICP | Pinball | Winkler | MAPE | RMSE | PICP | Pinball | Winkler |
| NH | 0.14% | -0.51% | 3.47% | 4.37% | -0.24% | -0.12% | -2.97% | 3.54% | -5.31% | -6.87% |
| RI | 0.09% | -3.47% | 1.98% | 0.01% | -2.65% | 0.03% | 0.12% | 2.61% | -4.94% | -7.14% |
| SE | 0.04% | -4.24% | 9.49% | -2.69% | -10.72% | -0.06% | -1.45% | 2.84% | -4.67% | -8.26% |
| CT | 0.08% | -2.47% | 2.55% | -0.51% | -3.61% | -0.09% | -0.92% | 4.63% | -7.04% | -7.33% |
| ME | -0.03% | -3.31% | 8.56% | -3.26% | -7.75% | -0.05% | -1.06% | 5.45% | -5.19% | -6.25% |
| NE | 0.05% | -2.17% | 2.48% | -0.14% | -3.55% | -0.10% | -2.02% | 4.16% | -6.72% | -6.96% |
| VT | 0.03% | -2.44% | 2.85% | -0.34% | -5.94% | 0.00% | -0.10% | 2.12% | -2.64% | 4.34% |
| WC | -0.04% | -3.44% | 1.71% | -3.90% | -1.35% | -0.09% | -1.39% | 6.44% | -7.57% | -7.31% |
| AVER | 0.04% | -2.76% | 4.14% | -0.81% | -4.47% | -0.06% | -1.22% | 3.97% | -5.51% | -5.72% |



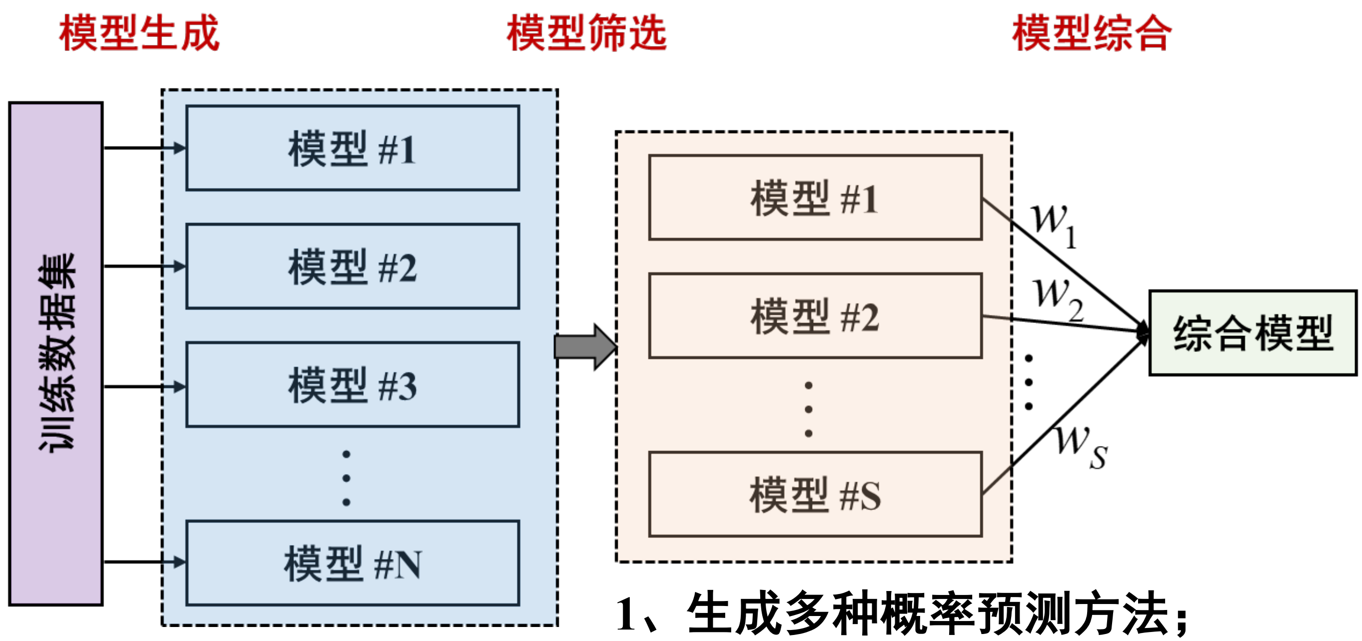
# 点预测到概率预测的拓展

ETH zürich

## ◆ 研究进展2.2：分位数预测集成

**背景：** 目前以后较多研究将点预测结果“集成”起来进一步提升点预测精度。那么多个概率预测模型怎么“集成”进一步提升概率预测精度呢？

**难点：** 点预测集成是一个“一维”问题，而概率预测结果是“高维”的，如何有效且高效整合概率预测结果是个难点。




1、生成多种概率预测方法；
2、去掉冗余或无用的预测模型；
3、实现多种模型的最优综合。

# 点预测到概率预测的拓展

ETH zürich

## ◆ 研究进展2. 2：分位数预测集成

## 从点预测综合到概率预测综合

$$f_e(\mathbf{X}_{n,t},\boldsymbol{\omega}) = \sum_{n=1}^{N}\omega_n f_n(\mathbf{X}_{n,t},\mathbf{W}_n).$$

$$\hat{\omega} = \arg\min_{\boldsymbol{\omega}} \sum_{t\in T} L_{n,t}\left(\sum_{n=1}^{N}\omega_n f_n(\mathbf{X}_{n,t},\mathbf{W}_n), y_t\right)$$

$$s.t. \quad \sum_{n=1}^{N}\omega_n = 1,$$

$$\omega_n \geq 0, \quad \forall n\in\{1,\cdots,N\}.$$

点预测

$$f_{e,q}(\mathbf{X}_{n,t},\boldsymbol{\omega}_q) = \sum_{n=1}^{N}\omega_{n,q} f_{n,q}(\mathbf{X}_{n,t},\mathbf{W}_{n,q}).$$

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} L_{n,t,q}\left(\sum_{n=1}^{N}\omega_{n,q} f_{n,q}(\mathbf{X}_{n,t},\mathbf{W}_{n,q}), y_t\right)$$

$$s.t. \quad \sum_{n=1}^{N}\omega_{n,q} = 1,$$

$$\omega_{n,q} \geq 0, \quad \forall n\in\{1,\cdots,N\}.$$

分位数预测

以分位数损失函数（Pinball Loss）为目标函数，构建优化模型确定不同分位数结果的权重，能够将其装化成为线性规划问题！

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} L_{t,q}(\hat{y}_{t,q}, y_t)$$

$$= \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T}\sum_{q\in Q} \max\left\{q(y_t-\hat{y}_{t,q}), (1-q)(\hat{y}_{t,q}-y_t)\right\}$$

$$s.t. \quad \hat{y}_{t,q} = \sum_{n\in N}\omega_{n,q}\hat{y}_{n,t,q}, \quad \sum_{n\in N}\omega_{n,q} = 1, \quad \omega_n \geq 0 \quad \forall n.$$

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} v_{t,q}$$

$$s.t. \quad \hat{y}_{t,q} = \sum_{n\in N}\omega_{n,q}\hat{y}_{n,t,q}, \quad \sum_{n\in N}\omega_{n,q} = 1, \quad \omega_{n,q} \geq 0 \quad \forall n.$$

$$v_{t,q} \geq q(y_t-\hat{y}_{t,q}), \quad v_{t,q} \geq (1-q)(\hat{y}_{t,q}-y_t)$$

$$\{v_{t,q}-q(y_t-\hat{y}_{t,q})\}\{v_{t,q}-(1-q)(\hat{y}_{t,q}-y_t)\} = 0.$$

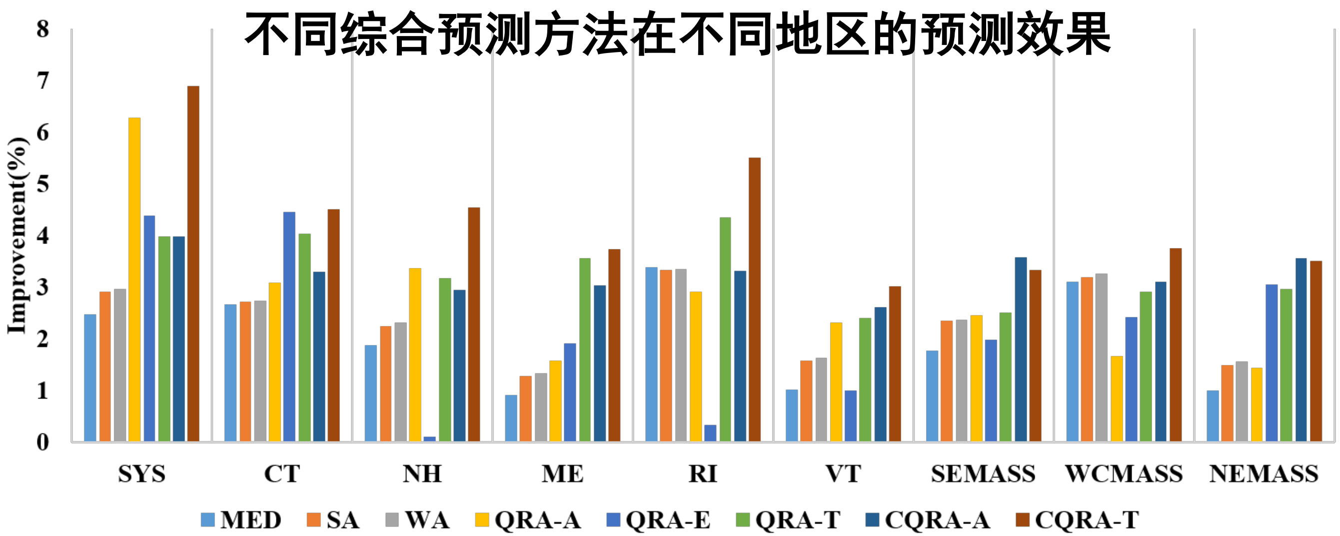
模型线性化、模型选择

# 点预测到概率预测的拓展

ETHzürich

## ◆ 研究进展2.2：分位数预测集成

| 区域 / 方法 | SYS | CT | NH | ME | RI | VT | SE | WC | NE |
|---|---|---|---|---|---|---|---|---|---|
| BI | 288.563 | 81.478 | 27.216 | 18.146 | 21.756 | 12.426 | 42.307 | 41.939 | 63.685 |
| NS | 327.569 | 95.058 | 31.586 | 19.003 | 25.738 | 13.247 | 48.817 | 47.041 | 71.873 |
| MED | 281.607 | 79.359 | 26.713 | 17.981 | 21.044 | 12.300 | 41.570 | 40.676 | 63.048 |
| SA | 280.375 | 79.322 | 26.618 | 17.916 | 21.053 | 12.233 | 41.336 | 40.638 | 62.752 |
| WA | 280.266 | 79.306 | 26.600 | 17.908 | 21.049 | 12.227 | 41.329 | 40.616 | 62.706 |
| QRA-E | 276.417 | 77.995 | 27.184 | 17.806 | 21.683 | 12.303 | 41.484 | 40.949 | 61.793 |
| QRA-A | 271.519 | 79.037 | 26.330 | 17.864 | 21.140 | 12.145 | 41.295 | 41.252 | 62.783 |
| QRA-T | 277.487 | 78.313 | 26.380 | 17.523 | 20.847 | 12.135 | 41.271 | 40.752 | 61.849 |
| CQRA-E | 356.527 | 100.925 | 33.829 | 22.767 | 26.540 | 15.616 | 51.765 | 51.544 | 79.131 |
| CQRA-A | 277.510 | 78.870 | 26.437 | 17.610 | 21.059 | 12.109 | **40.847** | 40.672 | **61.491** |
| CQRA-T | **269.953** | **77.961** | **26.034** | **17.492** | **20.619** | **12.061** | 40.941 | **40.422** | 61.524 |



# 点预测到概率预测的拓展

ETHzürich

## ◆ 研究进展2.2：分位数预测集成

### 对于总负荷（SYS）预测不同模型在不同分位数下的权重

| Models \ Quantiles | 10-th | 20-th | 30-th | 40-th | 50-th | 60-th | 70-th | 80-th | 90-th |
|---|---|---|---|---|---|---|---|---|---|
| #1 | 0 | 0 | 0 | 0.128 | 0.123 | 0 | 0.015 | 0 | 0.102 |
| #2 | 0 | 0 | 0 | 0.177 | 0.022 | 0.236 | 0.154 | 0.004 | 0 |
| #3 | 0.036 | 0 | 0 | 0.041 | 0.255 | 0 | 0.123 | 0.302 | 0 |
| #4 | 0.385 | 0.444 | 0.281 | 0 | 0 | 0.030 | 0 | 0 | 0.068 |
| #5 | 0.165 | 0 | 0 | 0.200 | 0.298 | 0.339 | 0.092 | 0 | 0.134 |
| #6 | 0.037 | 0.093 | 0.537 | 0.264 | 0 | 0 | 0.000 | 0.251 | 0 |
| #7 | 0 | 0.131 | 0 | 0.071 | 0 | 0 | 0.265 | 0.051 | 0.218 |
| #8 | 0 | 0.207 | 0.152 | 0 | 0.158 | 0.003 | 0.350 | 0.133 | 0 |
| #9 | 0.377 | 0.047 | 0.030 | 0.117 | 0.143 | 0.392 | 0 | 0.206 | 0.333 |
| #10 | 0 | 0.078 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| #11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.052 | 0.145 |
| #12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| #13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### 对于不同区域负荷预测不同模型在90分位数下的权重

| Models \ Zones | SYS | CT | NH | ME | RI | VT | SEMASS | WCMASS | NEMASS |
|---|---|---|---|---|---|---|---|---|---|
| #1 | 0.102 | 0.144 | 0.231 | 0.015 | 0.001 | 0.355 | 0 | 0 | 0.196 |
| #2 | 0 | 0 | 0 | 0.082 | 0.074 | 0.146 | 0.071 | 0 | 0 |
| #3 | 0 | 0 | 0.031 | 0 | 0 | 0.079 | 0 | 0.196 | 0 |
| #4 | 0.068 | 0 | 0.089 | 0.349 | 0 | 0 | 0.038 | 0 | 0 |
| #5 | 0.134 | 0 | 0 | 0 | 0.272 | 0 | 0.199 | 0.318 | 0.199 |
| #6 | 0 | 0 | 0.283 | 0.231 | 0.226 | 0.096 | 0 | 0 | 0.136 |
| #7 | 0.218 | 0 | 0.058 | 0.058 | 0 | 0.082 | 0.166 | 0.218 | 0.049 |
| #8 | 0 | 0.129 | 0.308 | 0.079 | 0.197 | 0 | 0.173 | 0.076 | 0.087 |
| #9 | 0.333 | 0.341 | 0 | 0.185 | 0.021 | 0.243 | 0.290 | 0.192 | 0.333 |
| #10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| #11 | 0.145 | 0.267 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| #12 | 0 | 0 | 0 | 0 | 0.210 | 0 | 0 | 0 | 0 |
| #13 | 0 | 0.119 | 0 | 0 | 0 | 0 | 0.062 | 0 | 0 |

# 点预测到概率预测的拓展

ETH zürich

## ◆ 研究进展2.3：概率密度预测集成

还是一样的套路：

$$\min \quad TL = \sum_{t=1}^{T} L\left(\sum_{i=1}^{N} \omega_i f_i(x_t), y_t\right)$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \omega_i = 1, \quad \omega_i \geq 0$$

然而概率密度的损失函数太复杂！

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(z) - \mathbb{1}(z-y))^2 dz$$

$$\mathrm{CRPS}(F, y) = \mathrm{E}_F|Y - y| - \frac{1}{2}\mathrm{E}_F|Y - Y'|$$

幸好，我们发现了一个优美的性质……

$$f_X(x) = \sum_{i=1}^{N} \omega_i f_i(x)$$

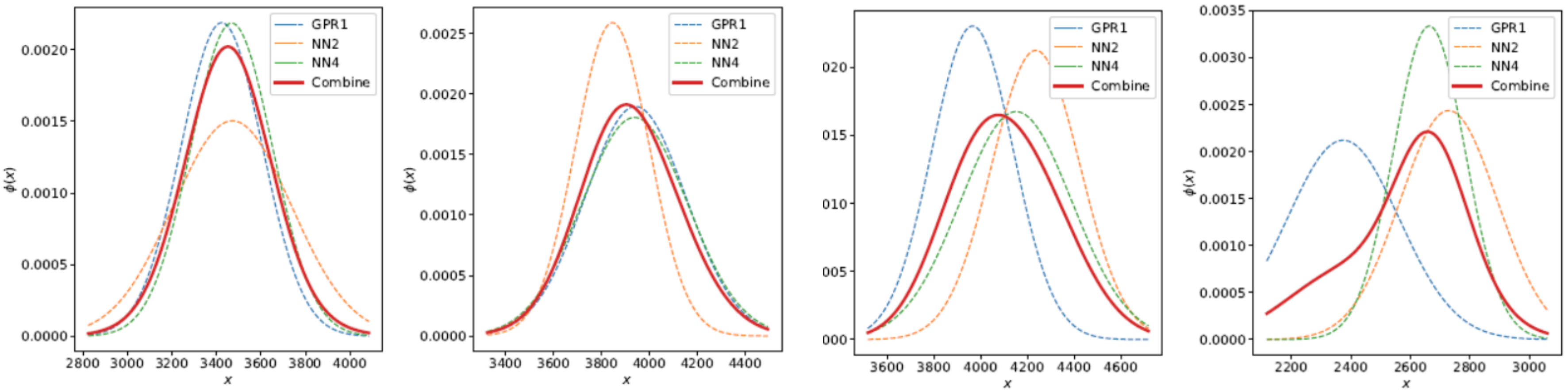
$$\mathrm{CRPS}(F, y) = \sum_{i=1}^{n}\sum_{k=1}^{n} \alpha_{i,k}\omega_i\omega_k + \sum_{i=1}^{n} \beta_i\omega_i$$

$$\alpha_{i,k} = -\frac{1}{\sqrt{2\pi}}\sqrt{\sigma_i^2+\sigma_k^2}\exp\left(-\frac{(\mu_i-\mu_k)^2}{2(\sigma_i^2+\sigma_k^2)}\right) - \frac{\mu_i-\mu_k}{2}\left[2\Phi\left(\frac{(\mu_i-\mu_k)}{\sqrt{\sigma_i^2+\sigma_k^2}}\right)-1\right]$$

$$\beta_i = \sqrt{\frac{2}{\pi}}\sigma_i\exp\left(-\frac{(\mu_i-y)^2}{2\sigma_i^2}\right) + (\mu_i-y)\left[2\Phi\left(\frac{\mu_i-y}{\sigma_i}\right)-1\right]$$

# 点预测到概率预测的拓展

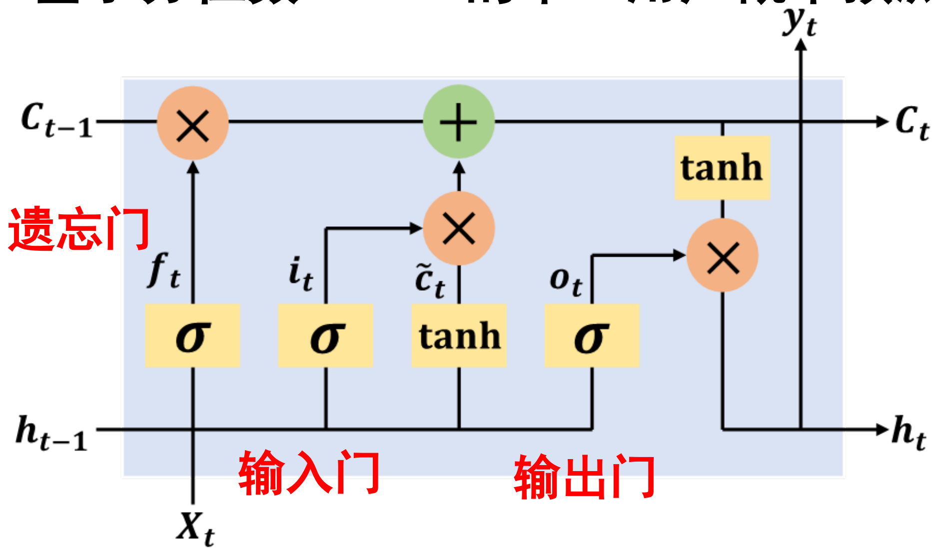
## ◆ 研究进展2. 3：概率密度预测集成



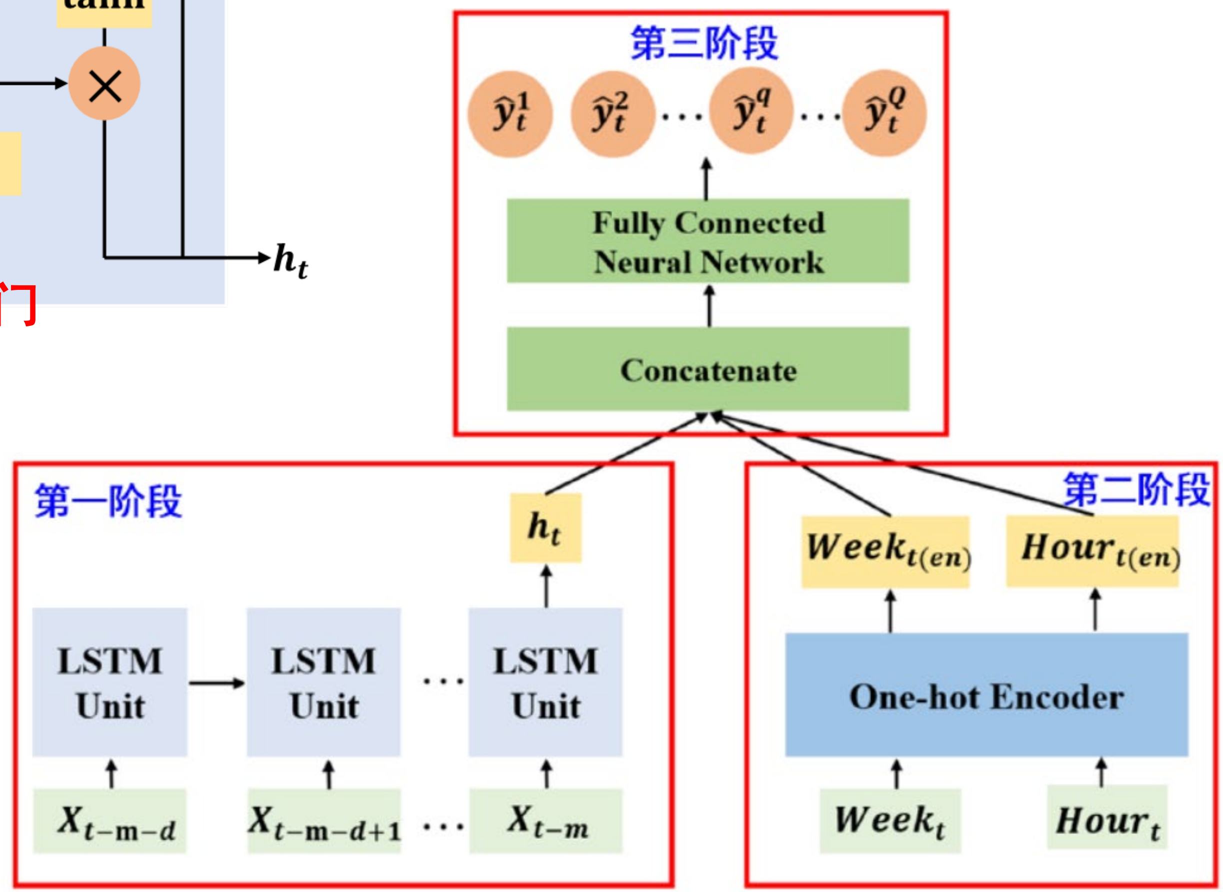

| Method / Region | Best Individual | Simple Average | MAPE Based | CRPS Based |
|---|---|---|---|---|
| CT | 5.27 | 5.48 | 5.09 | 5.11 |
| ME | 3.38 | 3.31 | 3.25 | 3.26 |
| NH | 4.17 | 4.06 | 4.01 | 4.04 |
| VT | 4.03 | 4.08 | 3.83 | 3.86 |
| RI | 4.34 | 4.38 | 4.17 | 4.21 |
| SEMASS | 4.49 | 4.66 | 4.42 | 4.43 |
| WCMASS | 4.36 | 4.71 | 4.33 | 4.37 |
| NEMASSBOST | 4.40 | 4.90 | 4.38 | 4.40 |

## ◆ 研究进展3.1：单一用户概率预测

### 基于分位数LSTM 的单一用户概率预测




$h$和$c$用于保存短期信息和长期信息




1、分位数LSTM 是LSTM和分位数损失函数的结合；
2、利用独热编码器对日类型等进行建模；
3、损失函数为所有的分位数损失。

# ◆ 研究进展3.1：单一用户概率预测

## 分位数损失的修正，以便于训练

$$H(y_t,\hat{y}_t^q)=\begin{cases}\dfrac{(\hat{y}_t^q-y_t)^2}{2\epsilon} & 0\le|\hat{y}_t^q-y_t|\le\epsilon\\ |\hat{y}_t^q-y_t|-\dfrac{\epsilon}{2} & |\hat{y}_t^q-y_t|>\epsilon,\end{cases} \Rightarrow L_{q,t}(y_t,\hat{y}_t^q)=\begin{cases}(1-q)H(y_t,\hat{y}_t^q) & \hat{y}_t^q\ge y_t\\ qH(y_t,\hat{y}_t^q) & \hat{y}_t^q<y_t.\end{cases}$$

不同方法对100 个居民用户预测的平均分位数损失(kW)

| | **QLSTM** | QRNN | I_QRNN | QGBRT | I_QGBRT | LSTM+E | I_LSTM+E |
|---|---|---|---|---|---|---|---|
| 30 分钟 | **0.0837** | 0.0867 | 3.46% | 0.0886 | 5.50% | 0.0905 | 7.52% |
| 1 小时 | **0.0963** | 0.0990 | 2.76% | 0.1030 | 6.48% | 0.1020 | 5.62% |
| 2 小时 | **0.1018** | 0.1040 | 2.18% | 0.1077 | 5.50% | 0.1061 | 4.13% |
| 4 小时 | **0.1031** | 0.1054 | 2.19% | 0.1090 | 5.40% | 0.1077 | 4.27% |

不同方法对100 个工商业用户预测的平均分位数损失(kW)

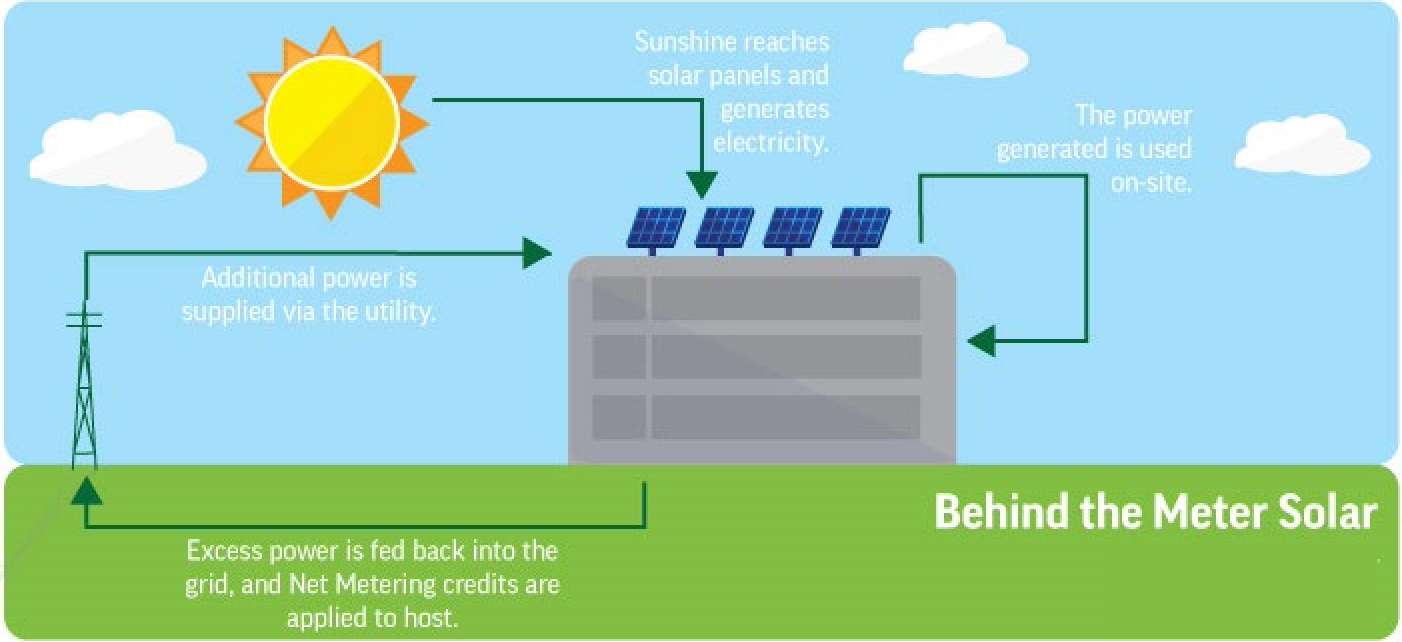
| | **QLSTM** | QRNN | I_QRNN | QGBRT | I_QGBRT | LSTM+E | I_LSTM+E |
|---|---|---|---|---|---|---|---|
| 30 分钟 | **0.1213** | 0.1275 | 4.89% | 0.1461 | 16.98% | 0.1391 | 12.81% |
| 1 小时 | **0.1552** | 0.1613 | 3.79% | 0.1975 | 21.43% | 0.1775 | 12.56% |
| 2 小时 | **0.1805** | 0.1883 | 4.16% | 0.2381 | 24.21% | 0.2081 | 13.29% |
| 4 小时 | **0.1982** | 0.2114 | 6.27% | 0.2671 | 25.80% | 0.2252 | 12.01% |

# 挖掘预测新对象

## ◆ 研究进展3. 2：“净负荷”概率预测

分布式光伏并网成为未来我国乃至全世界可再生能源发展的重要趋势，为了更好消纳分布式可再生能源，保障电力系统安全稳定，需要对分布式可再生能源进行“实时感知”。

分布式不可见光伏的辨识是用电行为结构解析的重要部分，是传统非侵入式辨识的重要补充。




用电与气象等多元数据融合

数据模型与物理模型融合

**不可见分布式光伏辨识方法特点：**

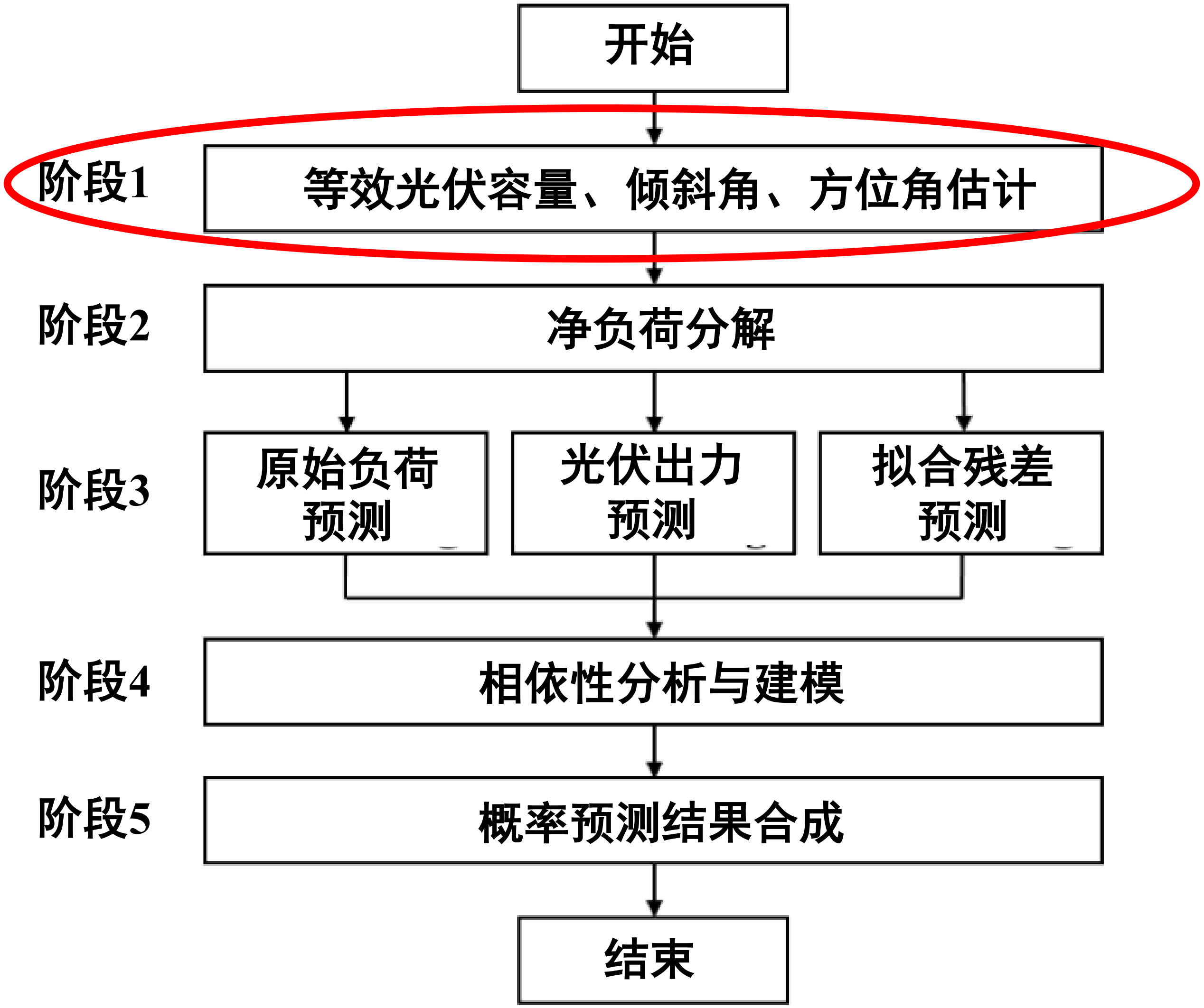
- ➢ 融合多元数据
- ➢ 融合物理模型，引入等效的概念
- ➢ 用预测结果侧面反映辨识效果

# 挖掘预测新对象

ETHzürich

## ◆ 研究进展3.2： “净负荷”概率预测

**含不可见光伏的广义负荷结构解析方案：**



# 挖掘预测新对象

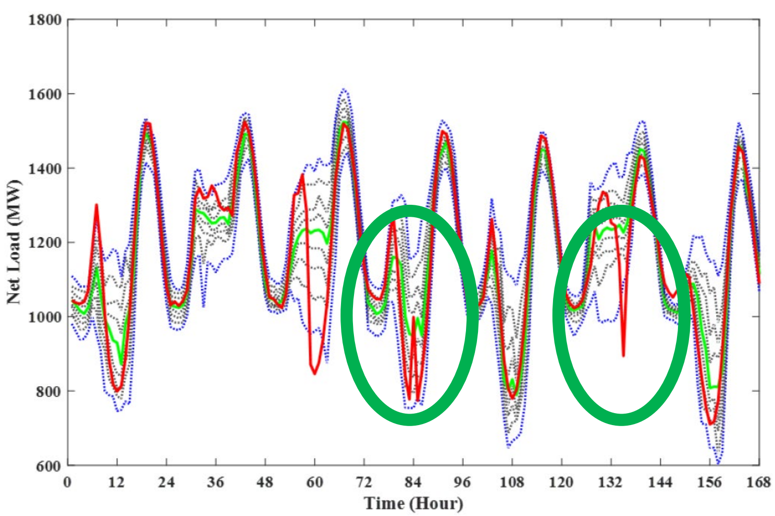
## ◆ 研究进展3.2：“净负荷”概率预测

**随着分布式光伏的增加，“检测-分离-预测-合成”的预测策略效果越明显**

**点预测效果(MW)**

| 光伏渗透率 | 本章方法 | 方法 1 | 方法 2 | 方法 3 |
|---|---|---|---|---|
| 0 | 34.3/2.60 | 38.3/2.85 | 40.7/3.06 | **34.2/2.59** |
| 5% | **60.1/3.37** | 94.6/5.28 | 101.5/5.47 | 61.4/3.59 |
| 10% | **80.9/4.80** | 145.8/8.17 | 157.5/8.50 | 83.6/5.23 |
| 15% | **109.1/7.28** | 221.8/13.1 | 209.7/12.3 | 115.0/8.25 |
| 20% | **140.8/22.6** | 279.1/109.2 | 267.1/84.1 | 162.8/43.6 |

**净负荷概率预测结果**




**概率预测效果(MW)**

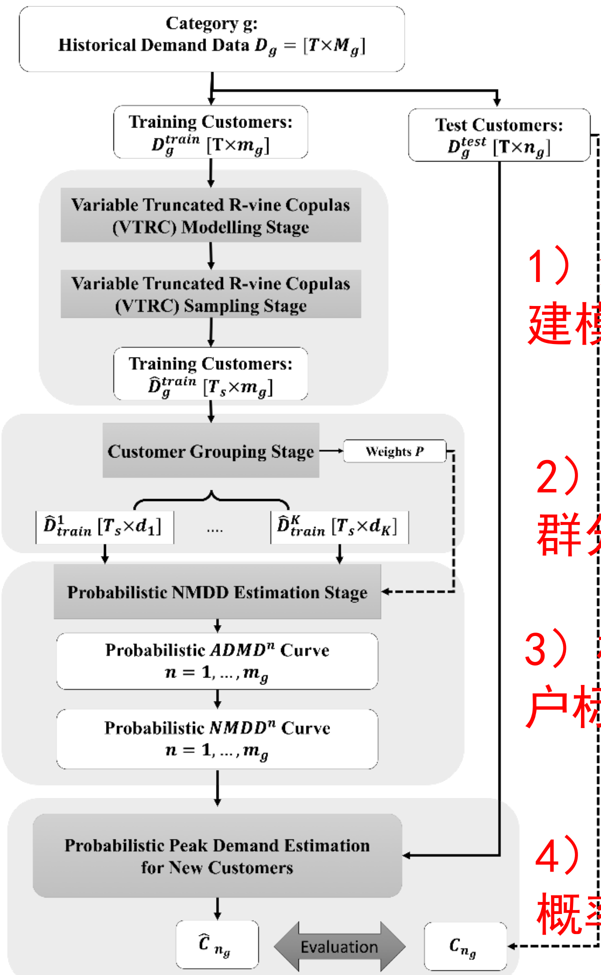
| 光伏渗透率 | 本章方法 | 方法 4 | 方法 5 | 方法 6 |
|---|---|---|---|---|
| 0 | 34.2 | 42.1 | 38.8 | **34.0** |
| 5% | **43.4** | 60.1 | 58.1 | 45.7 |
| 10% | **55.9** | 82.7 | 80.5 | 63.2 |
| 15% | **69.2** | 108.7 | 107.5 | 80.3 |
| 20% | **82.5** | 135.2 | 133.7 | 97.7 |

ETHzürich

## ◆ 研究进展3.3：峰值负荷概率预测

背景：如果新建一个小区并为其安装配电变压器，那其容量应该如何选取呢？能否通过估计未来用户数量及类型，进而估计未来负荷峰值呢？

难点：没有历史数据无法做“预测”，如何将充分利用其他以后多元用户能电表数据？




1）负荷相关性建模与采样

2）用户聚类组群分析

3）构建多元用户标准曲线

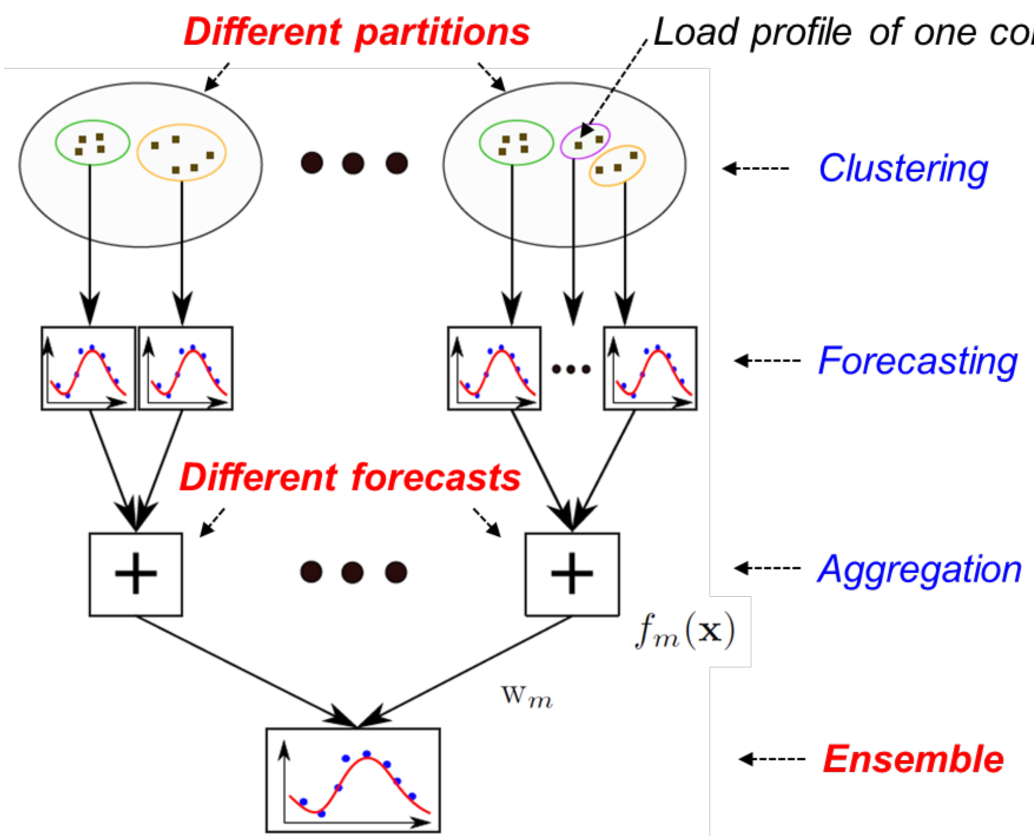
4）新用户峰值概率预测

方案：不同类型（富裕程度、社会阶层、职业状态等）的用户用电峰值特性不同，可以利用该特性估计未来待规划区域的负荷峰值，从而确定变电站等设备容量。

## ◆ 研究进展3. 4：集成负荷预测

**Primary idea:** instead of treating the aggregated load as a whole, partitioning consumers into several groups and making predictions might help improve load forecasting.

If there are different partitions of consumers, we can obtain different load forecasts.




$$\mathbf{x}_i = [1, \hat{y}_{1,i}, \ldots, \hat{y}_{K,i}] \qquad i \in [1, \ldots, n]$$

$$\hat{\mathbf{w}}_q = \arg\min_{\mathbf{w}_q} \sum_{i=1}^{n} \rho_q(y_i - \mathbf{x}_i \mathbf{w}_q)$$

Quantile regression averaging (QRA), a special form of quantile regression, is a kind of model averaging method.

# 挖掘预测新对象

ETHzürich

## ◆ 研究进展3.4：集成负荷预测

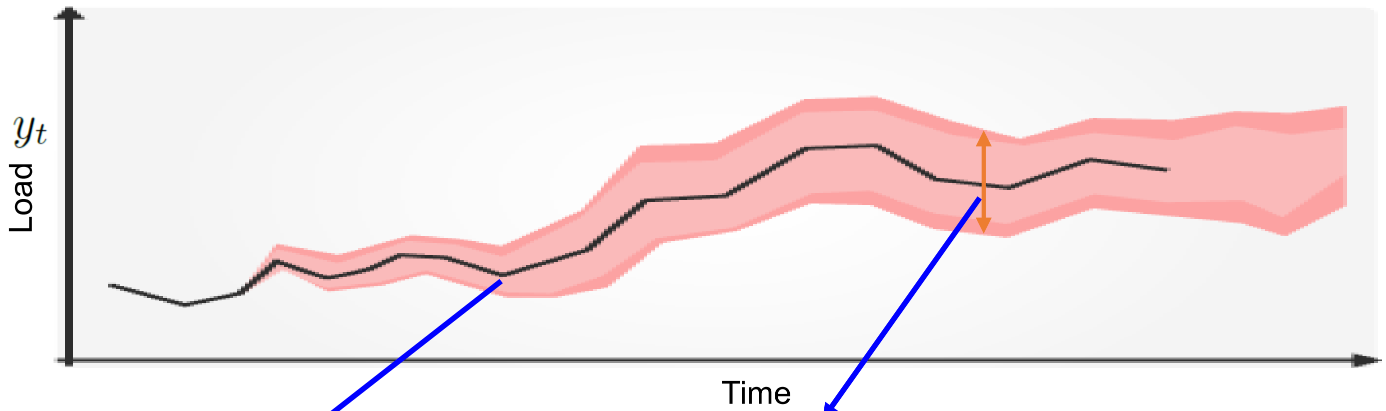
Error metric comparison for all ensemble methods with a Prediction Interval of 90%.

| Ensemble Method | Error Metrics | Offline Ensemble | Benchmark 1 | Rolling Window-based Ensemble | Benchmark 2 |
|---|---|---|---|---|---|
| QRA | ACE | -1.73% | -1.85% | -0.56% | -0.92% |
| | PBL | 45.82 | 50.19 | 42.28 | 46.52 |
| | WKS | 788.62 | 846.89 | 728.13 | 791.78 |
| FQRA | ACE | -1.80% | -1.85% | **-0.45%** | -0.92% |
| | PBL | 45.82 | 50.19 | **42.26** | 46.52 |
| | WKS | 787.26 | 846.89 | **727.24** | 791.77 |
| LQRA | ACE | **-1.71%** | -1.83% | -0.63% | -0.98% |
| | PBL | **45.84** | 50.2 | 42.26 | 46.53 |
| | WKS | **785.77** | 845.7 | 724.74 | 791.55 |

- The two naive benchmarks are obtained by directly forecasting the total loads without dimension reduction and clustering.
- Benchmark 2 updates the weights in a rolling window-based approach, while Benchmark 1 does not.

## ◆ 研究进展3. 5：场景预测

From the perspective of forecasting, the electrical load $y_t$ contains two parts:




deterministic part:

$$\hat{y}_t = f(\mathbf{X}_t)$$

uncertain part:

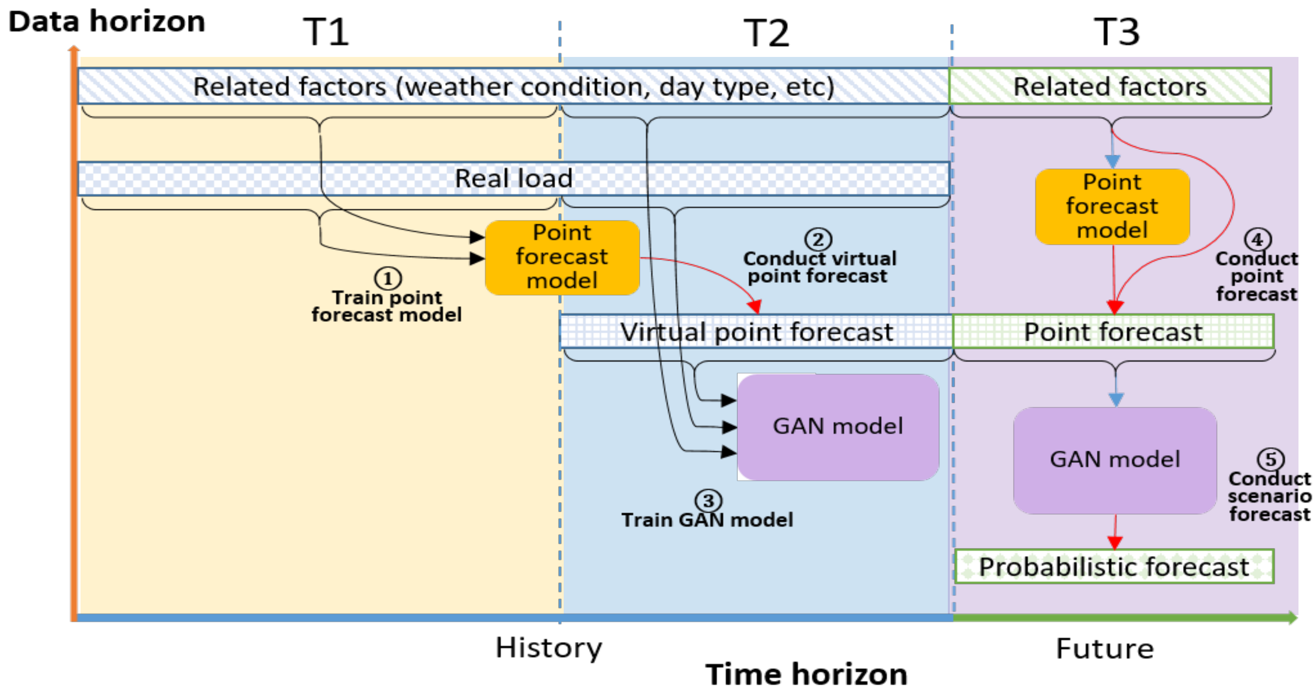
$$\epsilon_t = y_t - \hat{y}_t$$

| Train point forecast model $f(\cdot)$ | Train residual generation model $G(\cdot)$ | Test both models $f(\mathbf{X}_t) + G(\mathbf{C}_t, z)$ |
|---|---|---|
| **T1** | **T2** | **T3** |

# 挖掘预测新对象

## ◆ 研究进展3. 5：场景预测

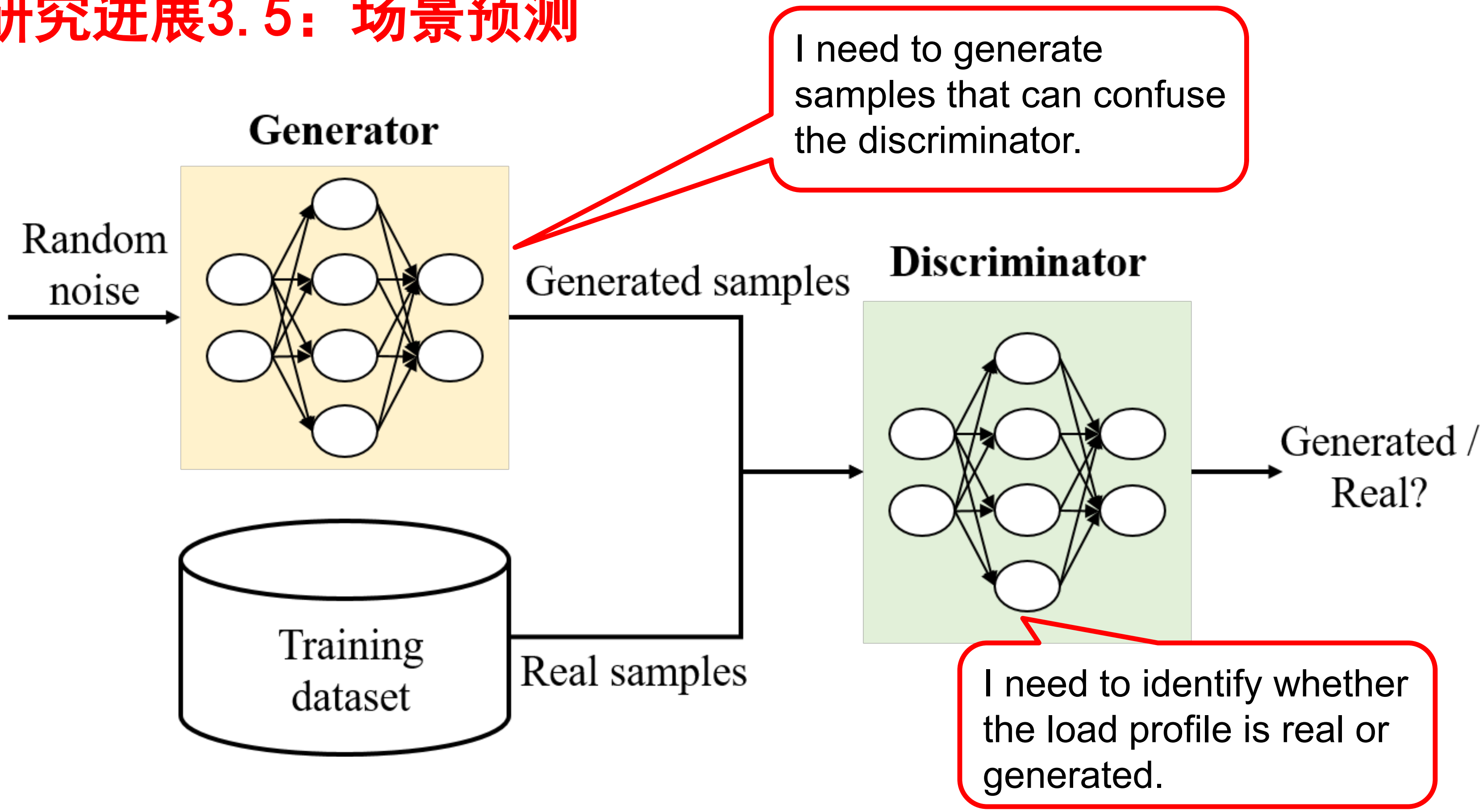
### 基本研究框架




| | | |
|---|---|---|
| ➢ **Train** $f(\mathbf{X}_t)$ using point forecasting models such as MLP/SVR/RF/GBRT | ➢ **Test** $f(\mathbf{X}_t)$<br>➢ **Calculate** $\epsilon_t$<br>➢ **Train** $G(\mathbf{C}_t, z)$ | ➢ **Test** $f(\mathbf{X}_t)$<br>➢ **Test** $G(\mathbf{C}_t, z)$<br>➢ **Final forecasts** $f(\mathbf{X}_t) + G(\mathbf{C}_t, z)$ |

**Formulate the generation model $G(\cdot)$ using GAN !**

## ◆ 研究进展3.5：场景预测




The **adversarial game** between these two neural networks can be presented as a min-max optimization model:

$$\max_{\theta_D} \mathbb{E}_{S_r}[\log(D(s_r;\theta_D))]+\mathbb{E}_Z[\log(1-D(G(z;\theta_G);\theta_D))]$$

# ◆ 研究进展3. 5：场景预测

## Performance w.r.t. Uncertainty

| Point Forecasts | Uncertainty Modeling | PL | WS(a=0.2) | WS(a=0.1) |
|---|---|---|---|---|
| AVE | Proposed | **12.38** | **180.1** | 262.14 |
| | CWGAN | 13.7 | 191.49 | 261.59 |
| | QRF | 14.23 | 189.64 | **231.84** |
| | QGBRT | 14.05 | 190.66 | 243.66 |
| SVR | Proposed | **12.55** | **182.04** | 259.84 |
| | CWGAN | 12.78 | 190.91 | 281.9 |
| | QRF | 14.5 | 194.66 | **240.76** |
| | QGBRT | 14.75 | 201.88 | 255.55 |
| RF | Proposed | **12.91** | **183.32** | 260.07 |
| | CWGAN | 13.1 | 187.58 | 263.7 |
| | QRF | 14.44 | 194.06 | 242.99 |
| | QGBRT | 13.94 | 186.66 | **233.99** |
| GBRT | Proposed | **12.24** | **172.15** | 236.34 |
| | CWGAN | 13.11 | 180.22 | 235.96 |
| | QRF | 14.06 | 183.18 | **223.12** |
| | QGBRT | 14.36 | 189.39 | 236.59 |

Comparison among different PLF methods

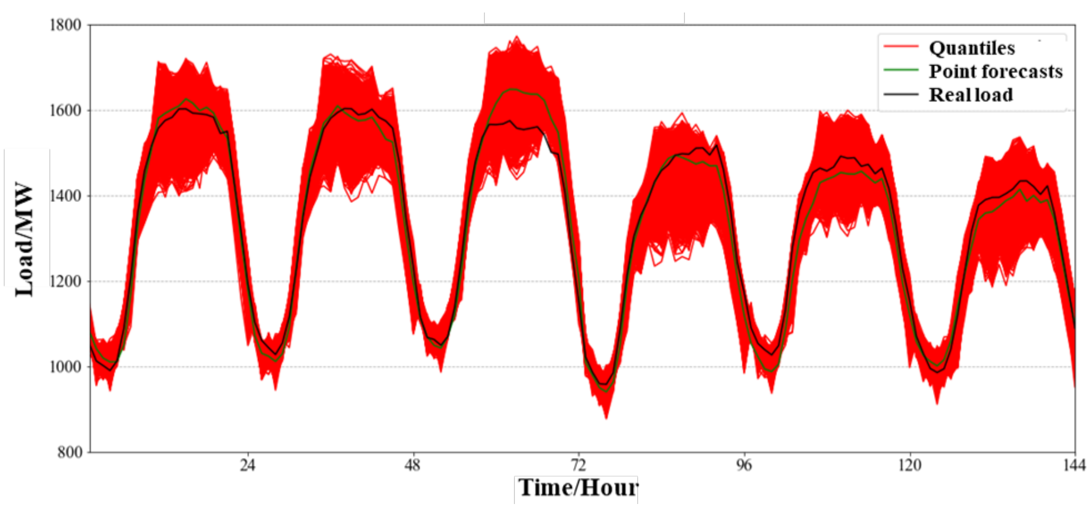
- For different point forecasts, our proposed CWGAN-GP model outperforms the CWGAN model, QRF, and QGBRT in terms of **PL and WS ( a= 0.2)**.
- However, QRF instead of the CWGAN-GP model performs better in terms of **WS ( a= 0.1).**

# 挖掘预测新对象

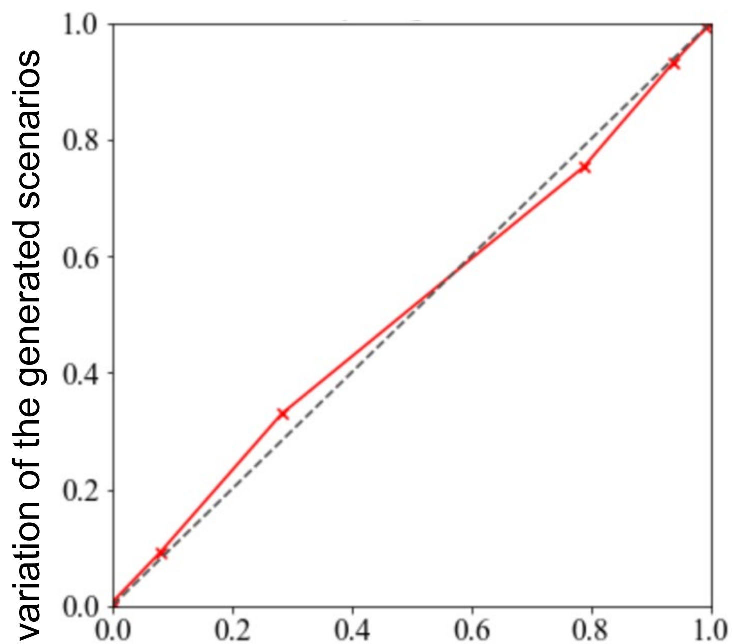
ETHzürich

## ◆ 研究进展3. 5：场景预测

## Performance w.r.t. Variation




10,000 generated scenarios of six selected days




The Q-Q plot

- ➢ The distribution of the variation of the generated scenarios is very similar to that of the real load profiles.
- ➢ Thus, **the generated scenarios can well represent the variations of the load profiles**.

## ◆ 研究进展3. 6：在线预测

- **General formula**

$$\mathbf{w}_{t+1} = \arg\min_{\mathbf{w}} \left[\ d(\mathbf{w}, \mathbf{w}_t) + \eta_t \ell(y_t, \mathbf{w} \cdot \mathbf{x}_t)\ \right]$$

Distance $d$ ⟷ Loss $\ell$

Prevent information loss ⟷ Integrate new sample

**Passive Aggressive Regression**

$$\mathbf{w}_{t+1} = \arg\min_{\mathbf{w}} \left[ \|\mathbf{w} - \mathbf{w}_t\|_1 + \ell_\varepsilon(y_t, \mathbf{w} \cdot \mathbf{f}_t) + \lambda \|\mathbf{w}\|_1 \right]$$

**Aggressive:** weights change if losses are big enough

$$\ell_\varepsilon(y_t, \mathbf{w} \cdot \mathbf{f}) = \begin{cases} 0 & \text{if } |y - \mathbf{w} \cdot \mathbf{f}| \le \varepsilon \\ |y - \mathbf{w} \cdot \mathbf{f}| & \text{otherwise} \end{cases}$$

**Passive:** weights do not change every time slot

# 挖掘预测新对象

ETH zürich

## ◆ 研究进展3.6：在线预测

The performance on Irish load data

Errors on test set after batch learning

| Method | ACE | PBL | WKS |
|---|---|---|---|
| QSGD | -0.92% | 51.60 | 722.43 |
| QPAR | 2.23% | 47.61 | 1075.02 |
| QNN | -2.55% | 54.94 | 776.86 |
| Batch QRA | -5.25% | 44.55 | 734.64 |
| Window QRA | -1.90% | 40.30 | 659.94 |

*QSGD: Quantile Stochastic Gradient Descent
*QPAR: Quantile Passive Aggressive Regression
*QNN: Quantile Neural Network

Errors on test set after online learning

| Method | ACE | PBL | WKS |
|---|---|---|---|
| QSGD | -0.02% | 30.04 | 527.94 |
| QPAR | -1.69% | 29.47 | 484.59 |
| QNN | -0.64% | 56.10 | 930.23 |
| Batch QRA | -5.25% | 44.55 | 734.64 |
| Window QRA | -1.90% | 40.30 | 659.94 |

*Window OPT: window-based optimization

- All ensembles outperform the benchmarks after online learning except QNN
- The proposed method has the highest accuracy regarding pinball loss and winkler score
- A substantial performance improvement can be achieved by ensembles incorporating online learning.

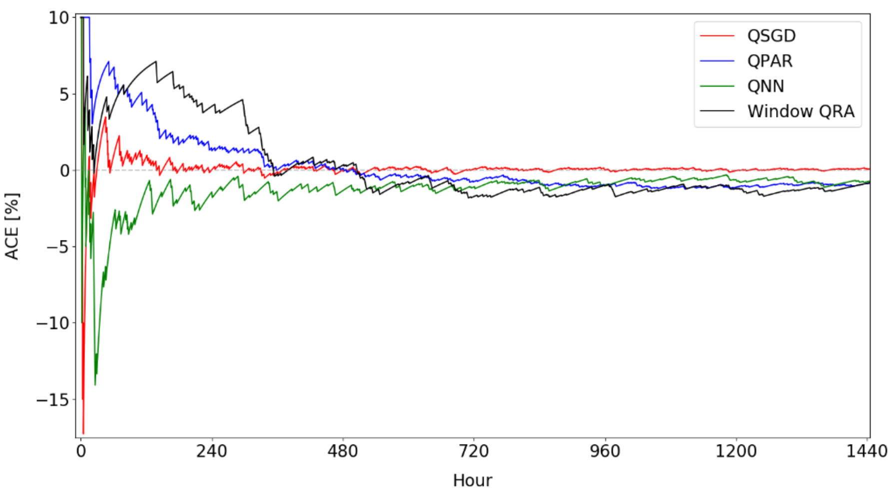
# 挖掘预测新对象

## ◆ 研究进展3. 6：在线预测

### The performance on Irish load data

The hour of break-even for all ensembles

| Method | Break-Even ACE | Break-Even PBL | Break-Even WKS |
|---|---|---|---|
| QSGD | 508.0 h | 35.0 h | 307.0 h |
| QPAR | 2810.0 h | 138.5 h | 253.5 h |
| QNN | 687.0 h | no | no |




ACE over the course of the first two months of online learning

- The proposed QPAR has earliest WKS break-even
- QSGD has earliest Break-even for ACE and PBL
- Online learning enables to outperform batch approach within a month.

# 概率性负荷预测研究展望

ETHzürich

| ◆什么是好的预测论文？ | ◆未来我们还可以做什么？ |
| --- | --- |
| ➢ 挖掘新的负荷预测问题； | ➢ 分层概率性负荷预测 |
| ➢ 避免排列组合式的论文； | ➢ 在线概率性负荷预测 |
| ➢ 充分结合物理背景； | ➢ 大数据下的概率性负荷预测 |
| ➢ 回答“为什么我的方法好？” | ➢ 面向成本的概率性负荷预测 |
| ➢ 全面、公开、综合的算例分析； | ➢ 保护隐私的概率性负荷预测 |

# 做个广告~

ETH zürich

IEEE.org | IEEE *Xplore* Digital Library | IEEE Standards | IEEE Spectrum | More Sites | Login | Create Account

IEEEDataPort™ DATASETS | COMPETITIONS | SUBMIT A DATASET | SEARCH | IEEE

Competition

Active Competition

## DAY-AHEAD ELECTRICITY DEMAND FORECASTING: POST-COVID PARADIGM

You must login to view this form.


| | |
|---|---|
| End Date: | 05/31/2021 |
| Citation Author(s): | Mostafa Farrokhabadi |
| Submitted by: | Mostafa Farrokhabadi |
| Last updated: | Thu, 12/10/2020 - 14:58 |
| DOI: | 10.21227/67vy-bs34 |
| License: | Creative Commons Attribution |

786 Views

| | |
|---|---|
| Categories: | Artificial Intelligence<br>COVID-19<br>Power and Energy |
| Keywords: | Load Forecasting, artificial intelligence, covid-19 |

ACCESS DATASET | CITE | SHARE/EMBED | Enter a Submission

### ABSTRACT

The ongoing COVID-19 related shutdowns have had a profound impact on the electric demand profiles worldwide, as governments put strict mitigation and/or suppression measures in place. The global electrical demand plummeted around the planet in March, April, and May 2020, with countries such as Spain and Italy experiencing more than 20% decrease in their usual electric consumption. In view of such massive electric demand changes, electricity network operators are facing unprecedented challenges in scheduling energy resources, as energy forecasting systems struggle to provide an accurate demand prediction. In fact, power systems operational reliability highly depends on an accurate projection of the future demand and scheduling an appropriate mixture of generation resources accordingly. Particularly, day-ahead forecasts are critical in managing market operation uncertainty. Thus, recent changes expose operators to technical and financial risks, further reinforcing the adverse economic impacts of the pandemic.

### DATASET FILES

You must be an approved participant in this data competition to access dataset files. To request access you must first login.

Login

### QUESTIONS?

# 参考文献

ETHzürich


**1.** **Yi Wang**, Leandro Von Krannichfeldt, Gabriela Hug, "Probabilistic Aggregated Load Forecasting with Fine-grained Smart Meter Data," *IEEE PowerTech 2021*, submitted.

2. Jialun Zhang, **Yi Wang***, Francois Bouffard, and Gabriela Hug, "Cost-Oriented Load Forecasting," *IEEE Trans. Power Systems*, submitted.

3. Jean-Francois Toubeau, Thomas Morstyn, Jeremie Bottieau, Kedi Zheng, Dimitra Apostolopoulou, Zacharie De Greve, **Yi Wang***, and Francois Vallee, "Capturing Spatio-Temporal Dependencies in the Probabilistic Forecasting of Distribution Locational Marginal Price," *IEEE Trans. Smart Grid*, submitted.

4. Leandro Von Krannichfeldt, **Yi Wang***, and Gabriela Hug, "Online Ensemble Learning for Load Forecasting," *IEEE Trans. Power Systems*, in press.

5. Jiawei Zhang, **Yi Wang**, Mingyang Sun, and Ning Zhang, "Two-Stage Bootstrap Sampling for Probabilistic Load Forecasting," *IEEE Trans. Engineering Management*, in press.

6. Tao Hong, Pierre Pinson, **Yi Wang**, Rafal Weron, Dazhi Yang, and Hamidreza Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access Journal of Power and Energy*, 2020, 7:376-388.

7. Shu Zhang, **Yi Wang***, Yutian Zhang, Dan Wang, and Ning Zhang, "Load Probability Density Forecasting by Transforming and Combining Quantile Forecasts," *Applied Energy*, 2020, 277:115600.

**8.** **Yi Wang**, Gabriela Hug, Zijie Liu, and Ning Zhang, "Modeling Load Forecast Uncertainty Using Generative Adversarial Networks," *Electric Power Systems Research*, 2020, 189:106732.

9. Tainyi Li, **Yi Wang**, and Ning Zhang, "Combining Probability Density Forecasts for Power Electrical Loads," *IEEE Trans. Smart Grid*, 2020, 11(2):1679-1690.

# 参考文献


10. Mingyang Sun, Tingqi Zhang, Yi Wang, Goran Strbac, and Chongqing Kang, “Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting,” *IEEE Trans. Power Systems*, 2020, 35(1):188-201.
11. **Yi Wang**, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, “Combining Probabilistic Load Forecasts,” *IEEE Trans. Smart Grid*, 2019, 10(4):3664-3674.
12. **Yi Wang**, Dahua Gan, Ning Zhang, Le Xie, and Chongqing Kang, “Feature Selection for Probabilistic Load Forecasting via Sparse Penalized Quantile Regression,” *Journal of Modern Power Systems and Clean Energy*, 2019, 7(5):1200-1209.
13. Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Trans. Smart Grid*, 2019, 10(3):3125-3148.
14. **Yi Wang**, Dahua Gan, Mingyang Sun, Ning Zhang, and Chongqing Kang, “Probabilistic Individual Load Forecasting Using Pinball Loss Guided LSTM,” *Applied Energy*, 2019, 235: 10-20.
15. Mingyang Sun, **Yi Wang**, Goran Strbac, and Chongqing Kang, “Probabilistic Peak Load Estimation in Smart Cities Using Smart Meter Data,” *IEEE Trans. Industrial Electronics*, 2019, 66(2): 1608-1618.
16. **Yi Wang**, Qixin Chen, Ning Zhang, and Yishen Wang, “Conditional Residual Modeling for Probabilistic Load Forecasting,” *IEEE Trans. Power Systems*, 2018, 33(6): 7327-7330.
17. **Yi Wang**, Qixin Chen, Mingyang Sun, Chongqing Kang and Qing Xia, “An Ensemble Forecasting Method for the Aggregated Load with Subprofiles,” *IEEE Trans. Smart Grid*, 2018, 9(4): 3906-3908.
18. Dahua Gan, **Yi Wang**, Shuo Yang, and Chongqing Kang, “Embedding Based Quantile Regression Neural Network for Probabilistic Load Forecasting,” *Journal of Modern Power Systems and Clean Energy*, 2018, 6(2): 1-11.

ETH zürich

# Thank you for your attention

Yi Wang | yiwang@eeh.ee.ethz.ch | www.eeyiwang.com