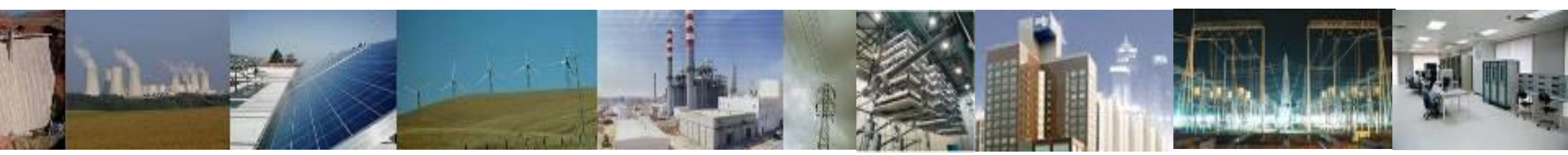
Probabilistic Short-Term Load Forecasting

Yi Wang

Tsinghua University

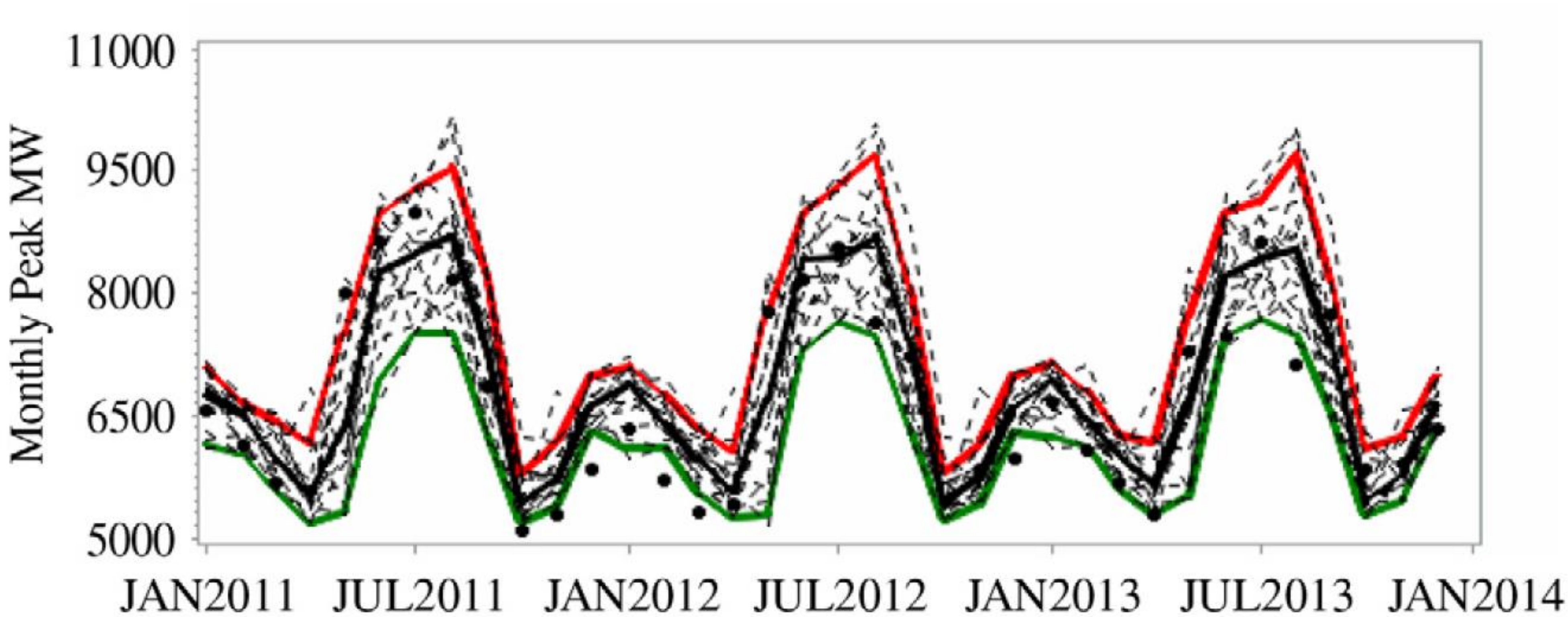


Contents

Backgrounds

What is Probabilistic Load Forecasting?

PLFs can be in the form of quantiles, intervals, or density functions.

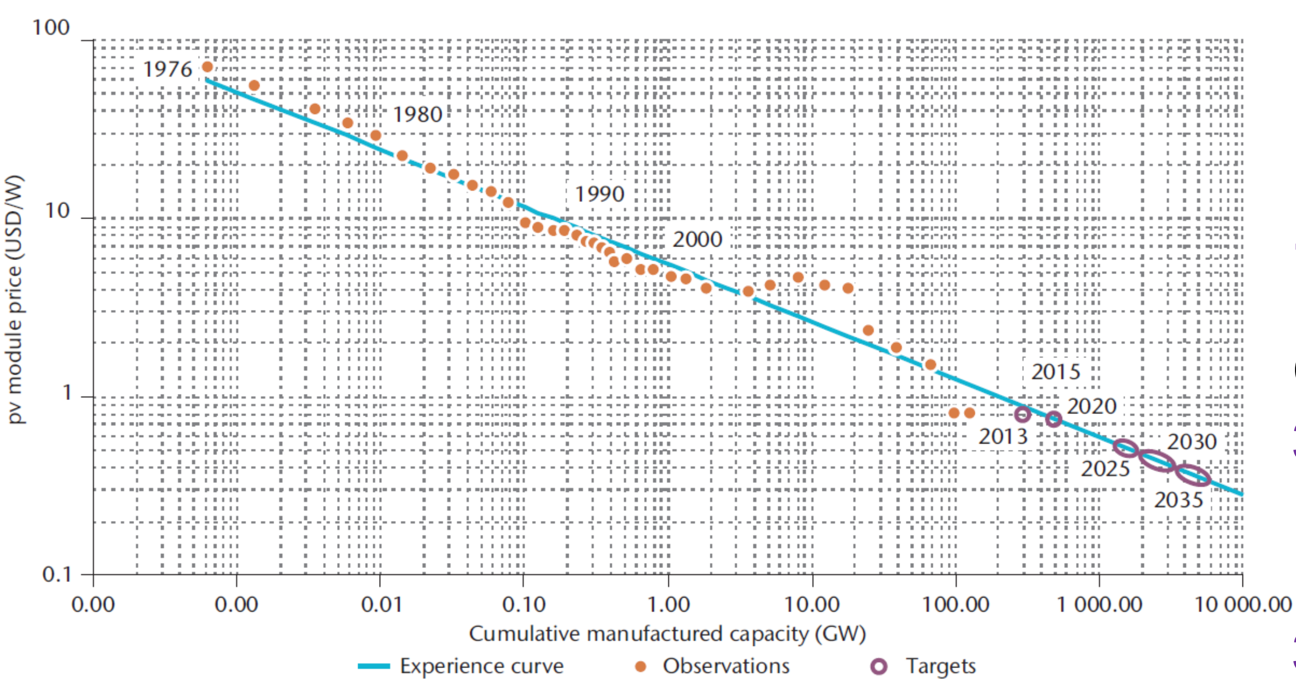


Backgrounds

Why we need probabilistic load forecasting?

- The integration of distributed renewable energy, energy storage, and the implementation of demand response.

- The stochastic mathematical techniques has been applied to power systems operation and planning.



In the year of 2017, China

Distributed PV capacity
19.44GW

Grow rate
360%

Percentage
36.6%

Backgrounds

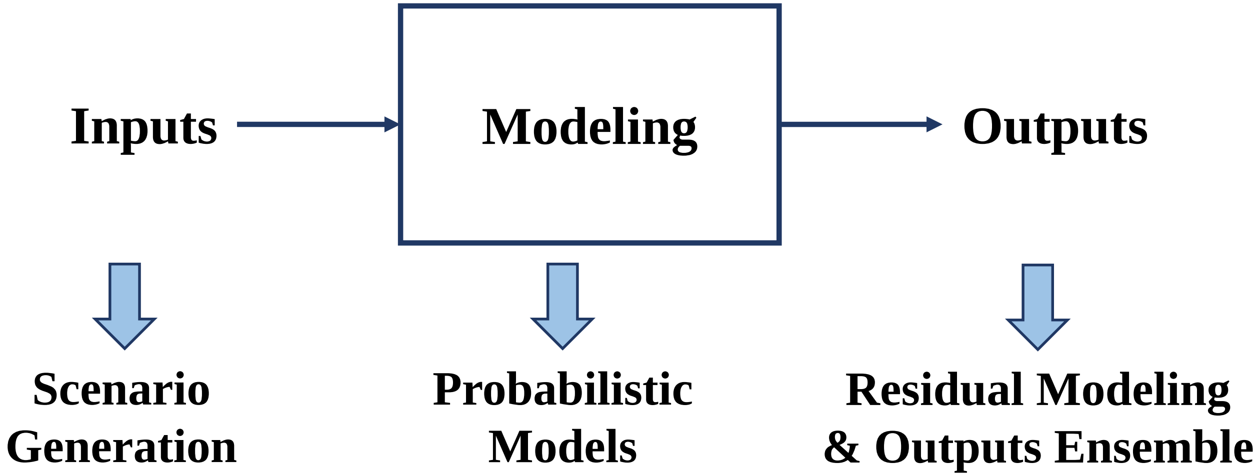
Off-the-shelf regression methods

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 works**

Backgrounds

From point load forecasting to probabilistic forecasting?



Two-stage Bootstrap Sampling
Temperature Scenario Generation

Probabilistic Net Load Forecasting
Pinball Loss Guided LSTM

Combining Probabilistic Forecasts
Conditional Residual Modeling

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.

Two-stage Bootstrap Sampling

Basic Idea

Uncertainty decomposition:

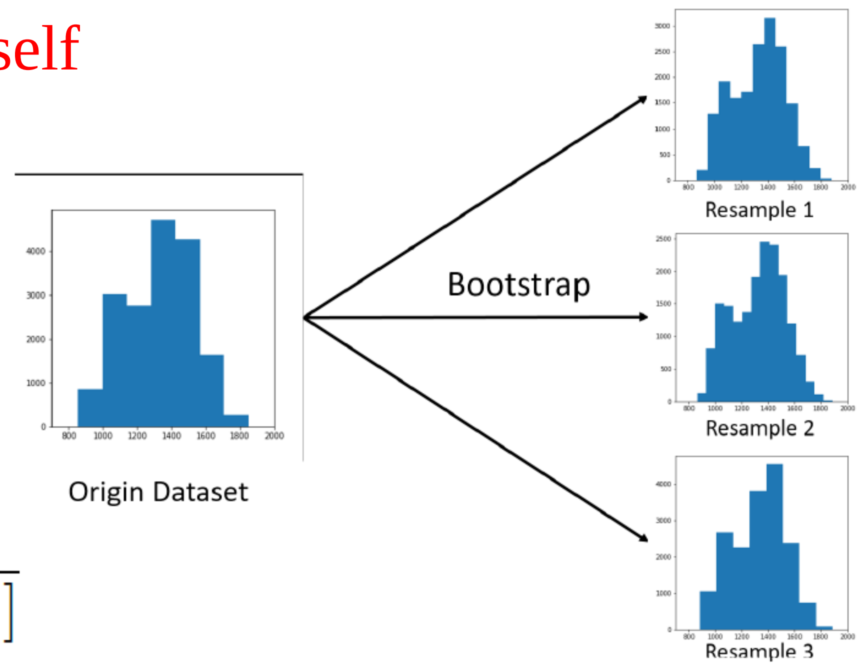
$$Var[Y^*|X=x^*] = Var[Y^* - \hat{m}(x^*)|X=x^*] + Var[\hat{m}(x^*)|X=x^*]$$

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:

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

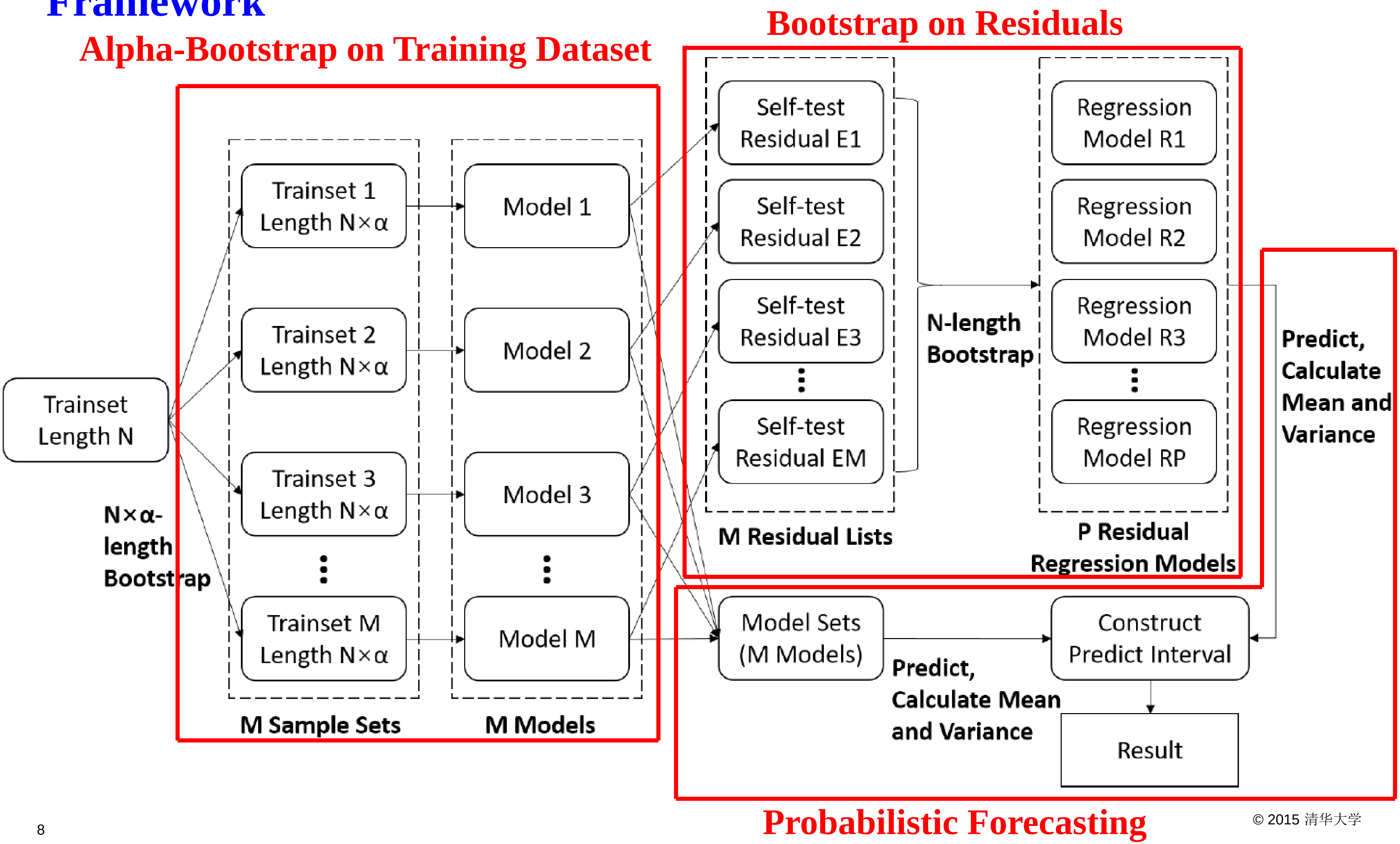


Calculate the quantiles:

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

Two-stage Bootstrap Sampling

Framework



Two-stage Bootstrap Sampling

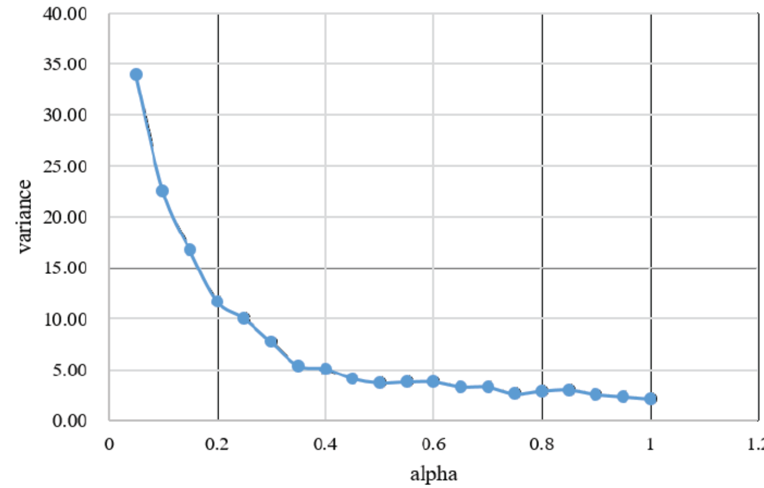
Alpha-Bootstrap on Training Dataset

Resample training dataset:

$$Ts_1, Ts_2, ..., Ts_M$$

Train *M* models:

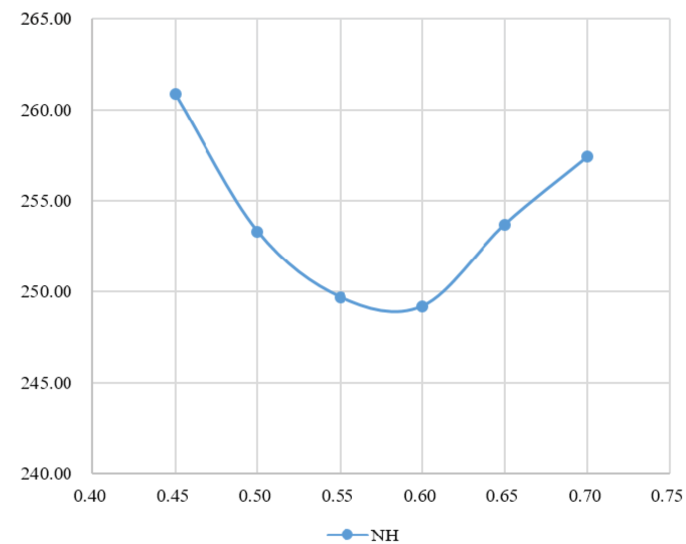
$$m_1, m_2, ..., m_M$$



Variance Estimation:

$$E[\hat{m}(x^*)] = \frac{1}{M}\sum_{i=1}^{M} m_i(x^*)$$

$$Var[\hat{m}(x^*)] = \frac{1}{M-1}\sum_{i=1}^{M} (m(x^*) - E[\hat{m}(x^*)])^2$$



Bootstrap on Residual

Conduct Forecasting:

$$m_1(x_i), m_2(x_i), ..., m_M(x_i)$$

Calculate Error:

$$E_1, E_2, ..., E_M$$

Bootstrap Sampling:

$$T_1, T_2, ..., T_P$$

Build Regression Models:

$$R_1, R_2, ..., R_P$$

Estimate Variance:

$$E[\hat{R}(x^*)] = \frac{1}{P}\sum_{i=1}^{P} R_i(x^*)$$

$$Var[\hat{R}(x^*)] = \frac{1}{P-1}\sum_{i=1}^{P} (R(x^*) - E[\hat{R}(x^*)])^2$$

First Stage:
Strong learners: GBRT, RF
M=200;

Second Stage:
Fast Learners: LR, LSSVM
P=2000;

Two-stage Bootstrap Sampling

Results

QUANTILE REGRESSION RESULT

	Quantile Random Forest					Quantile GBRT				
	MAPE	RMSE	PICP	Pinball	Winkler	MAPE	RMSE	PICP	Pinball	Winkler
NH	3.07%	57.79	0.85	14.85	246.58	3.33%	59.26	0.85	16.37	264.16
RI	3.04%	40.76	0.89	10.45	169.71	3.15%	39.65	0.88	11.15	178.77
SEMASS	3.92%	95.03	0.80	24.77	433.44	4.05%	92.94	0.83	25.39	429.46
CT	3.52%	169.19	0.85	44.53	728.26	3.61%	164.51	0.86	47.01	745.64
ME	2.86%	49.13	0.83	13.54	216.79	2.91%	48.31	0.86	13.92	214.39
NEMASSBOST	3.19%	126.75	0.86	33.15	538.10	3.37%	127.57	0.84	35.70	559.95
VT	3.98%	34.13	0.81	9.14	164.79	4.10%	33.93	0.81	9.51	150.46
WCMASS	3.44%	89.13	0.88	24.32	377.85	3.55%	88.13	0.84	25.53	406.75

FRAMEWORK IN THIS PAPER WITH LSSVM REGRESSION

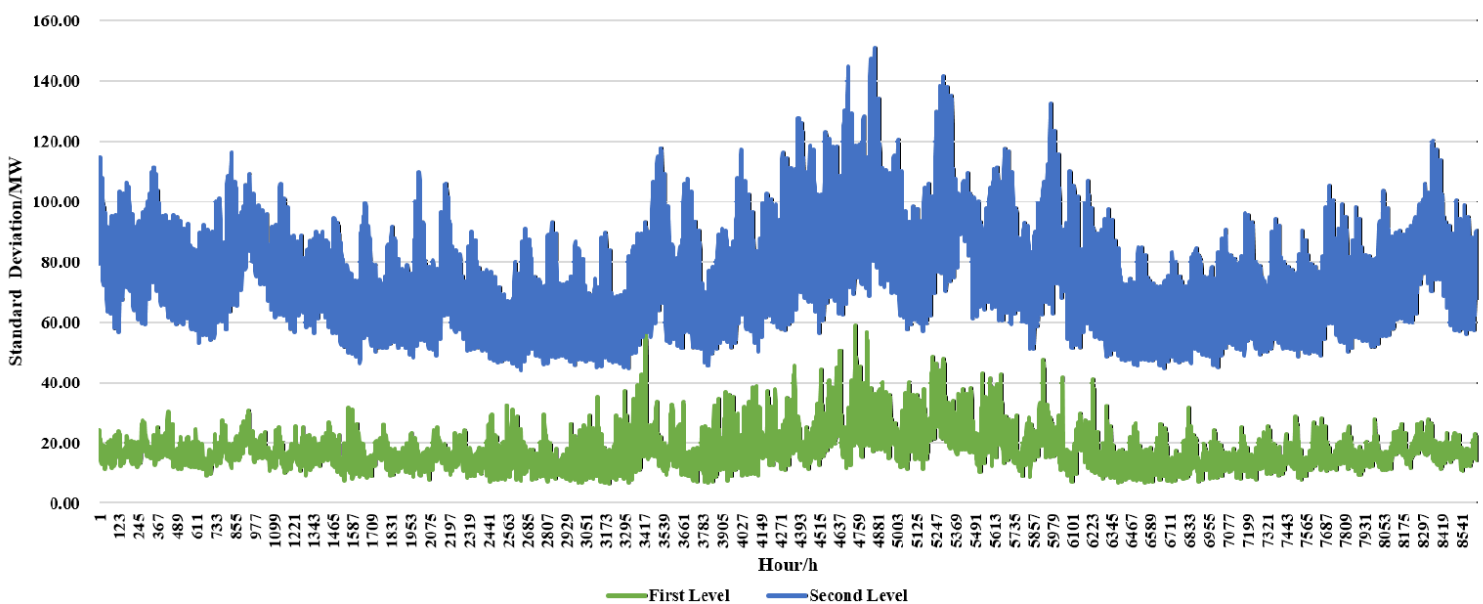
	Random Forest Based					GBRT Based				
	MAPE	RMSE	PICP	Pinball	Winkler	MAPE	RMSE	PICP	Pinball	Winkler
NH	3.21%	57.50	0.88	15.50	246.00	3.21%	57.50	0.88	15.50	246.00
RI	3.13%	39.34	0.90	10.45	165.21	3.18%	39.70	0.90	10.60	166.00
SEMASS	3.96%	91.00	0.87	24.10	387.00	3.99%	91.60	0.85	24.20	394.00
CT	3.60%	165.00	0.88	44.30	702.00	3.52%	163.00	0.90	43.70	691.00
ME	2.83%	47.50	0.90	13.10	200.00	2.86%	47.80	0.90	13.20	201.00
NEMASSBOST	3.24%	124.00	0.88	33.10	519.00	3.27%	125.00	0.88	33.30	521.00
VT	4.01%	33.30	0.83	9.11	155.00	4.10%	33.90	0.83	9.26	157.00
WCMASS	3.40%	86.07	0.90	23.37	372.74	3.46%	86.90	0.89	23.60	377.00

Two-stage Bootstrap Sampling

Results

COMPARE WITH QUANTILE REGRESSION

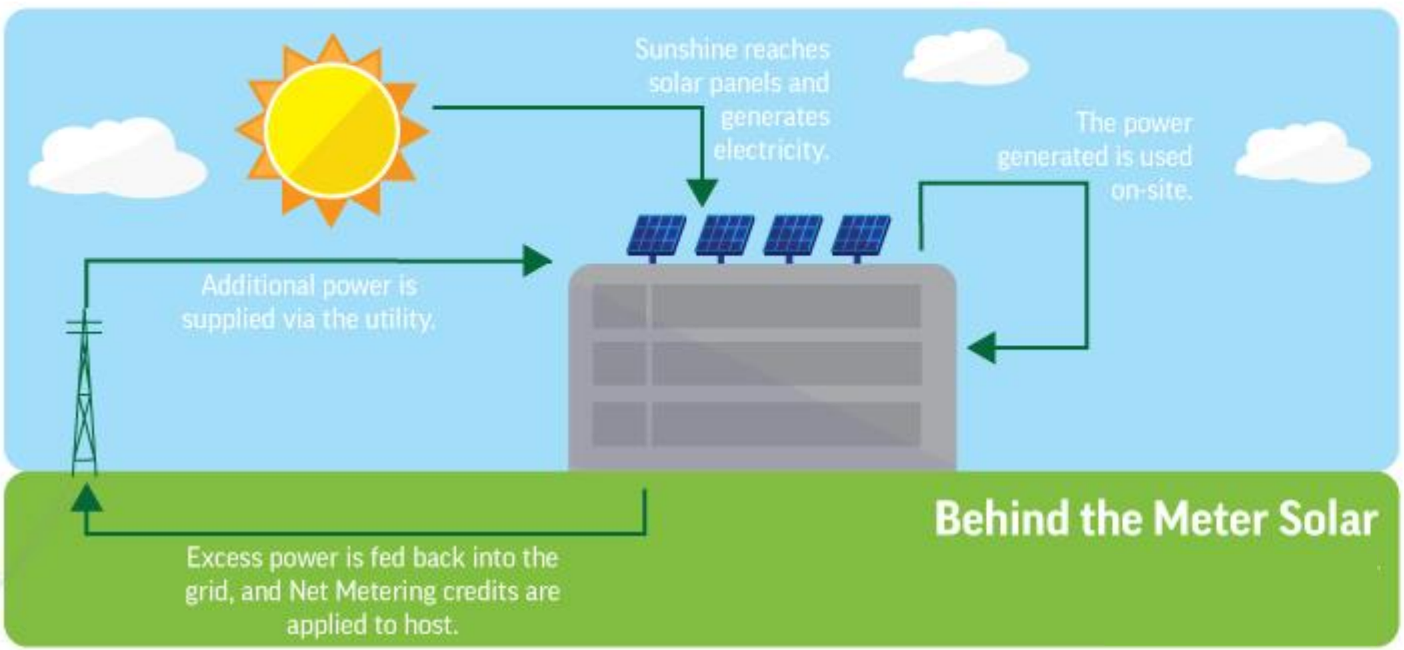
	Random Forest Based					GBRT Based				
	MAPE	RMSE	PICP	Pinball	Winkler	MAPE	RMSE	PICP	Pinball	Winkler
NH	0.14%	-0.51%	3.47%	4.37%	-0.24%	-0.12%	-2.97%	3.54%	-5.31%	-6.87%
RI	0.09%	-3.47%	1.98%	0.01%	-2.65%	0.03%	0.12%	2.61%	-4.94%	-7.14%
SE	0.04%	-4.24%	9.49%	-2.69%	-10.72%	-0.06%	-1.45%	2.84%	-4.67%	-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%



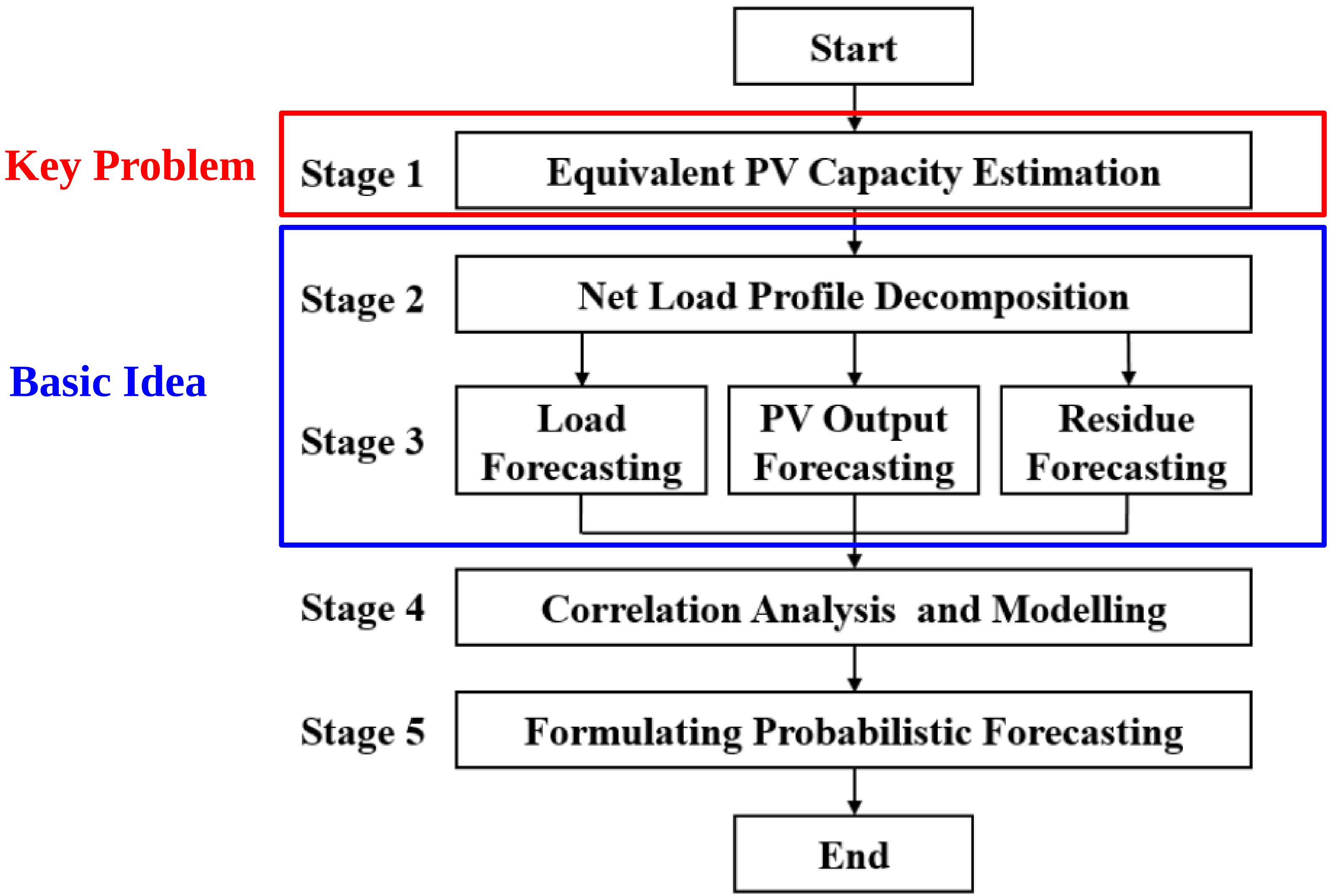
Problem Statement & Basic Idea

Behind-the-meter (BtM) PV are invisible to DSO which poses great challenges to real time situation awareness.

- How to estimate the capacity of BtM PV?
- How to further improve the forecasting accuracy?



Framework


Start
Key Problem
Stage 1
Equivalent PV Capacity Estimation
Basic Idea
Stage 2
Net Load Profile Decomposition
Stage 3
Load Forecasting
PV Output Forecasting
Residue Forecasting
Stage 4
Correlation Analysis and Modelling
Stage 5
Formulating Probabilistic Forecasting
End

PV Capacity Estimation

Initialize the capacity

↓

Estimate PV output

$$P_t \approx C\frac{I_{PV,t}}{1000}\left[1-\mu\left(T_{PV,t}-25\right)\right]$$

↓

Estimate original load

$$L_{E,t}=f_L\left(M_t, W_t, H_t, T_{A,t}\right)$$

↓

Calculate residual

$$Res_t = Net_t - L_{E,t} + P_{E,t}$$

↓

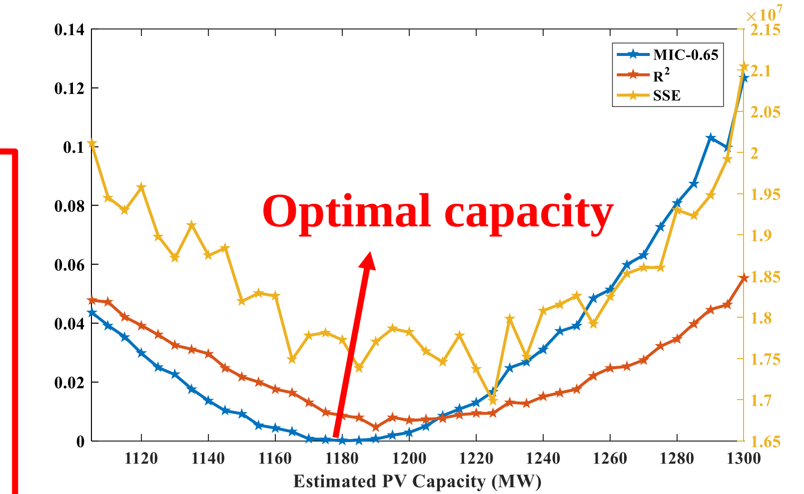
Correlation analysis

$$R = \arg\min_{C_{eq},\beta_{eq},\gamma_{eq}} MIC\left(Res_t, I_{GHI,t}\right)$$

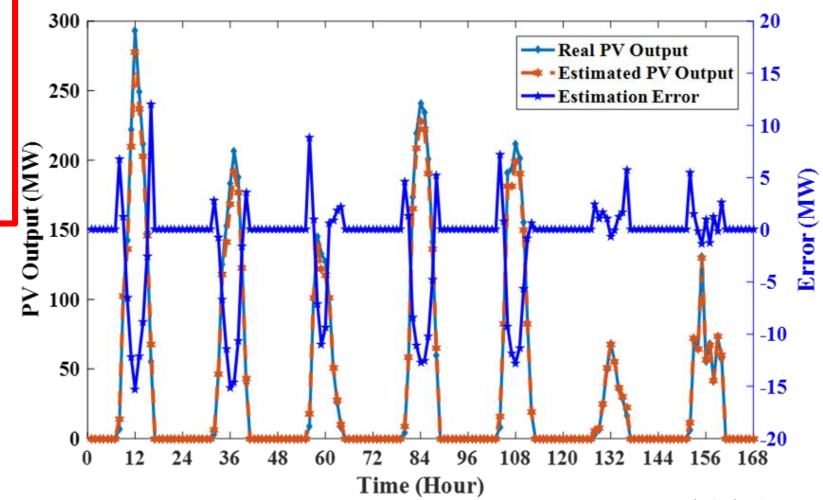
↓

Stop when R does not decrease

Adjust the capacity



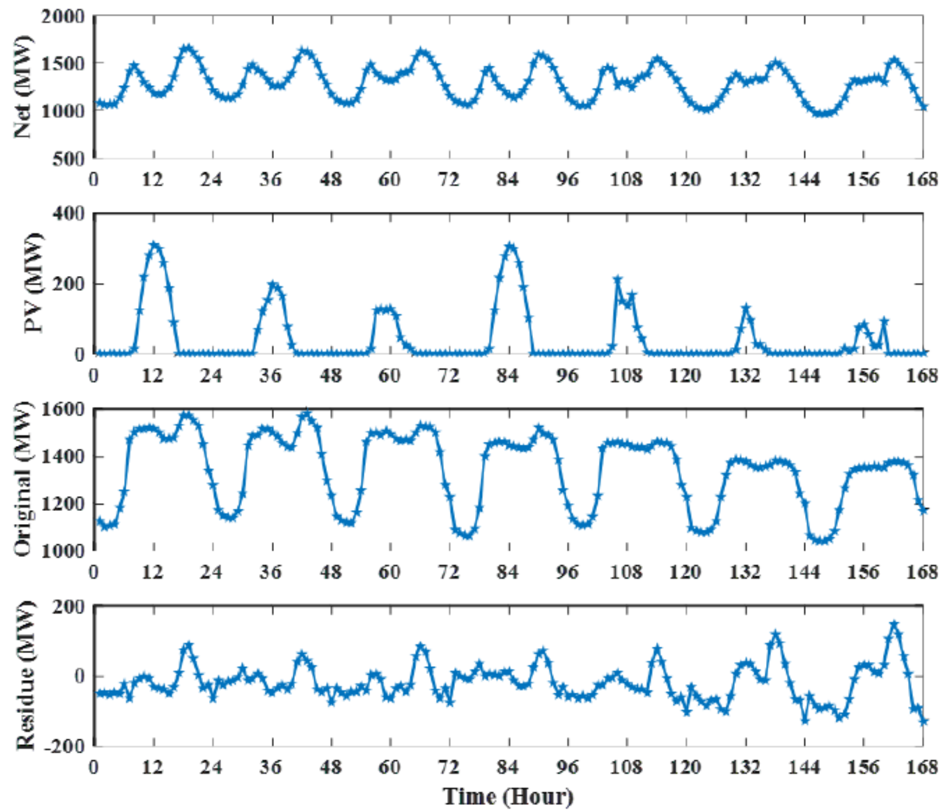
Estimated PV output



Data Simulation

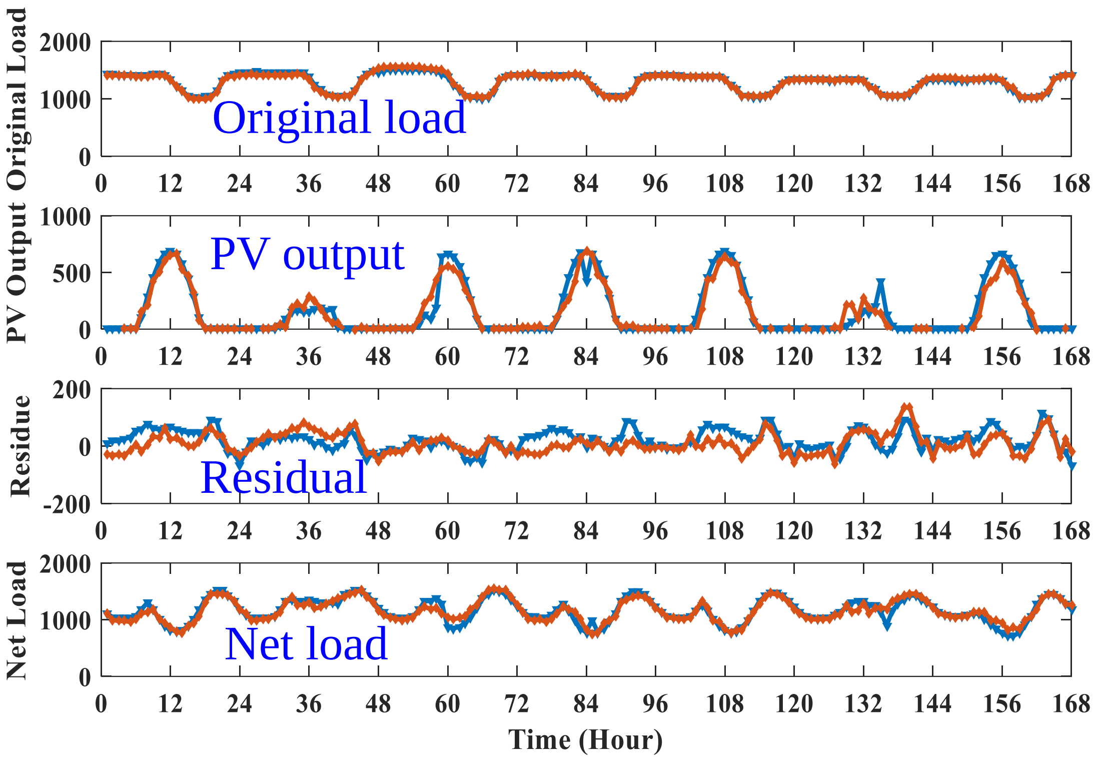
System Advisor Model (SAM)
Developed by NREL

Net load separation

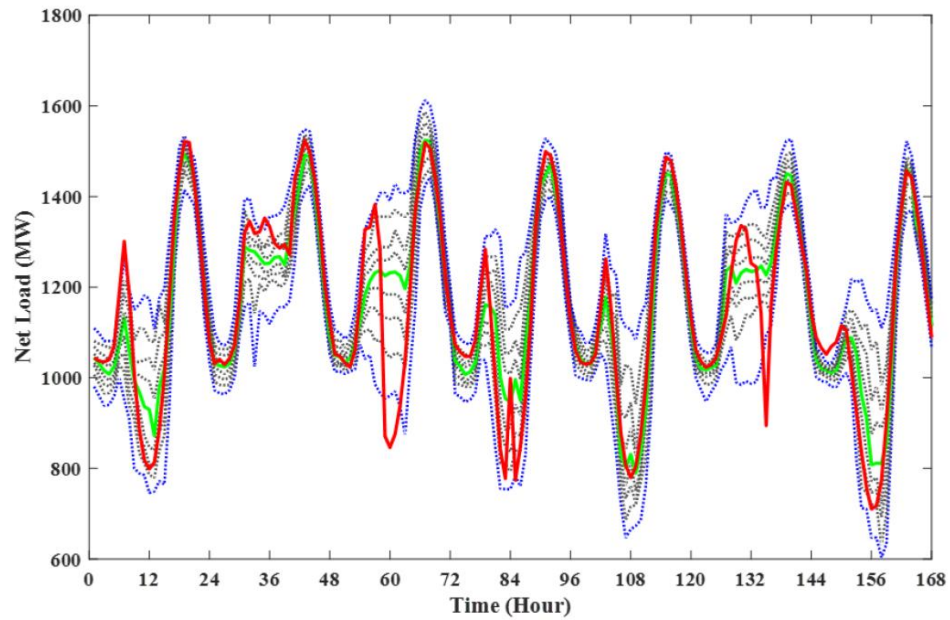


Results

Forecasts for different parts



Probabilistic forecasts：



Probabilistic Net Load Forecasting

Results

Competing methods

	Point Forecasting	Probabilistic Forecasting
Time Series	#1	#4
Considering Temperature	#2	#5
Considering Temperature and Solar Irradiation	#3	#6

Point forecasting

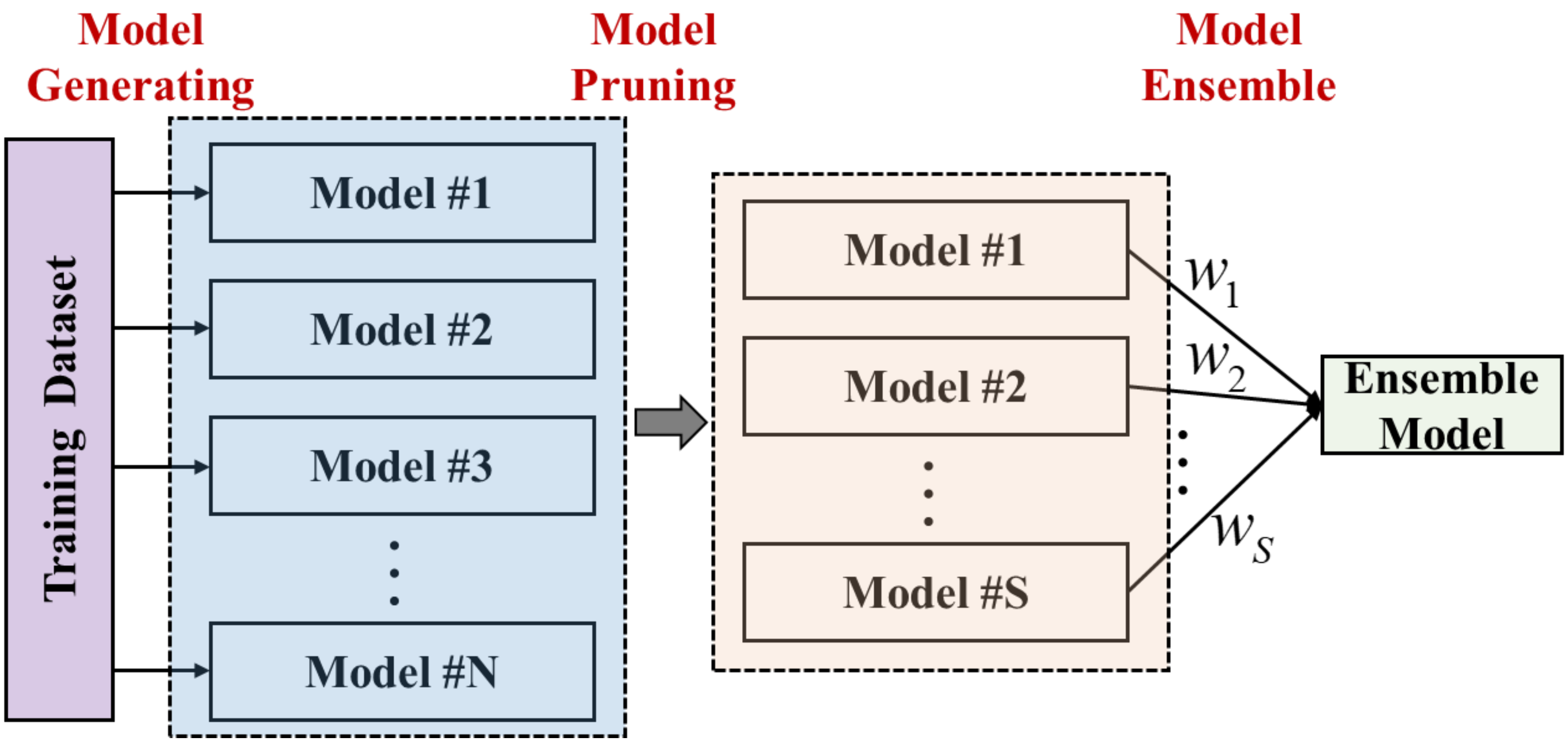
PV Penetration	Proposed Method	Method #1	Method #2	Method #3
0	34.3/2.60	38.3/2.85	40.7/3.06	**34.2/2.59**
5%	**60.1/3.37**	94.6/5.28	101.5/5.47	61.4/3.59
10%	**80.9/4.80**	145.8/8.17	157.5/8.50	83.6/5.23
15%	**109.1/7.28**	221.8/13.1	209.7/12.3	115.0/8.25
20%	**140.8/22.6**	279.1/109.2	267.1/84.1	162.8/43.6

The higher, the better

Probabilistic forecasting

PV Penetration	Proposed Method	Method #4	Method #5	Method #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

Ensemble Learning



QRNN
QGBRT
QRF
…

1) Generate models;
2) Prune models;
3) Combine models

Combining Probabilistic Forecasts

From point forecast to probabilistic forecast

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

$$\begin{aligned} \hat{\omega} = & \arg\min_{\boldsymbol{\omega}} \sum_{t \in T} L_{n,t}\left(\sum_{n=1}^{N} \omega_n f_n(\mathbf{X}_{n,t}, \mathbf{W}_n), y_t\right) \\ s.t. \quad & \sum_{n=1}^{N} \omega_n = 1, \\ & \omega_n \geq 0, \quad \forall n \in \{1, \cdots, N\}. \end{aligned}$$

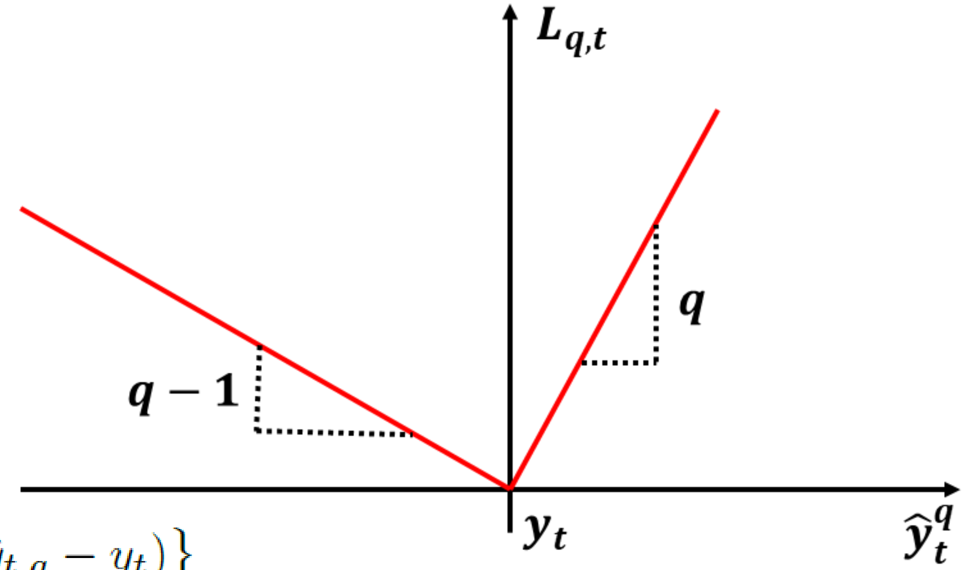
Point Forecasts

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

$$\begin{aligned} \hat{\boldsymbol{\omega}}_q = & \arg\min_{\boldsymbol{\omega}_q} \sum_{t \in T} L_{n,t,q}\left(\sum_{n=1}^{N} \omega_{n,q} f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}), y_t\right) \\ s.t. \quad & \sum_{n=1}^{N} \omega_{n,q} = 1, \\ & \omega_{n,q} \geq 0, \quad \forall n \in \{1, \cdots, N\}. \end{aligned}$$

Quantile Forecasts

Linear Programming Model



$$\hat{y}_{t,q} \approx \sum_{n\in N} \omega_{n,q}\hat{y}_{n,t,q}.$$

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

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

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

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

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

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

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

1) LP problem;
2) Model Selection;

- T1 Individual Model Training
- T2 Individual Model Validation
- T3 Individual Model Test and Ensemble
- T4 Ensemble Model Test

Comparisons
Nine models

1) Naïve Sorting (NS): With each forecasting model producing Q quantiles, a total of $N \times Q$ quantiles can be observed (in some sense) by N forecasting models. By sorting these observations by descending order, a new sequence $\mathbf{S}_t = \{S_{t,j}, j = [1, Q \times N]\}$ can be obtained. And therefore the q-th quantile is estimated as follows:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N}.$$

2) Median Value (MED): The median value of the N q-th quantiles is selected as the final quantile:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N+[N/2]}.$$

3) Simple Averaging (SA): The simple averaging strategy applies equal weights to different methods:

$$w_{n,q} = 1/N.$$

Then, the final combined forecasts are calculated according to Eq. (15).

4) Weighted Averaging (WA): The basic idea of the weighted averaging method is that methods with higher accuracy should be given higher weights:

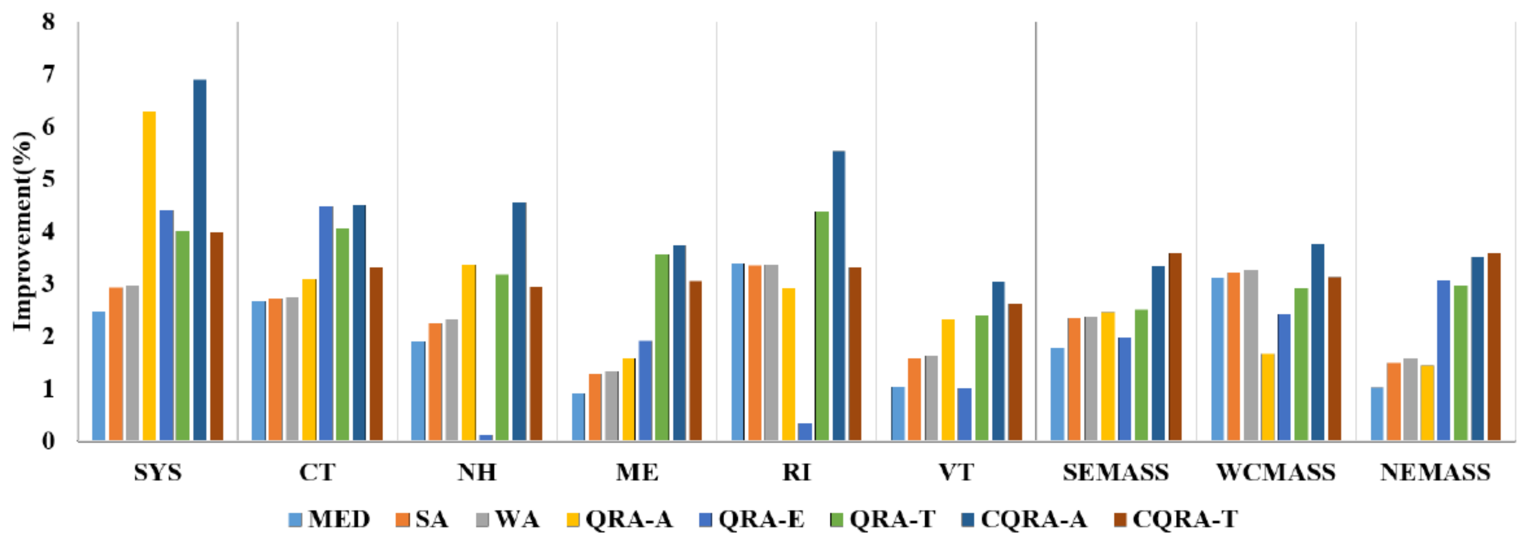
$$w_{n,q} = \frac{\frac{1}{L_{n,q}}}{\sum_{n \in N} \frac{1}{L_{n,q}}}.$$

Constraints / Quantiles	With Constraints	Without Constraints
Averaged Quantiles	5) QRA-E	8) CQRA-E
All Quantiles	6) QRA-A	9) CQRA-A
Targeted Quantiles	7) QRA-T	CQRA-T (Proposed)

Results

PINBALL LOSSES OF THE INDIVIDUAL AND COMBINATION METHODS FOR DIFFERENT ZONES

Methods \ Zones	SYS	CT	NH	ME	RI	VT	SEMASS	WCMASS	NEMASS
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



Combining Probabilistic Forecasts

Results

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

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

Conclusions

Understand where the uncertainties are from

Investigate how the load profile changes

Combine different forecasting methods

Welcome to Tsinghua

References

1. **Yi Wang**, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang*, "Combining Probabilistic Load Forecasts", ***IEEE Trans. Smart Grid***, in press.
2. **Yi Wang**, Qixin Chen, Tao Hong, and Chongqing Kang*, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges", ***IEEE Trans. Smart Grid***, in press.
3. **Yi Wang**, Qixin Chen, Mingyang Sun, and Chongqing Kang* and Qing Xia, "An Ensemble Forecasting Method for the Aggregated Load with Sub Profiles", ***IEEE Trans. Smart Grid***, in press.
4. **Yi Wang**, Ning Zhang, Qixin Chen*, Daniel Kirschen, Pan Li, and Qing Xia, "Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV", ***IEEE Trans. Power Systems***, in press.
5. 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***, in press.
6. Dahua Gan, **Yi Wang**, Ning Zhang*, and Wenjun Zhu, "Enhancing Short Term Probabilistic Residential Load Forecasting with Quantile LSTM", ***The Journal of Engineering***, in press.
7. Mingyang Sun*, **Yi Wang**, Goran Strbac, and Chongqing Kang, "Probabilistic Peak Load Estimation in Smart Cities Using Smart Meter Data", ***IEEE Trans. Industrial Electronics***, in press.

Thanks！