Combining Probabilistic Load Forecasts

Yi Wang

ETH Zurich

www.eeyiwang.com

August 8, 2019, Atlanta

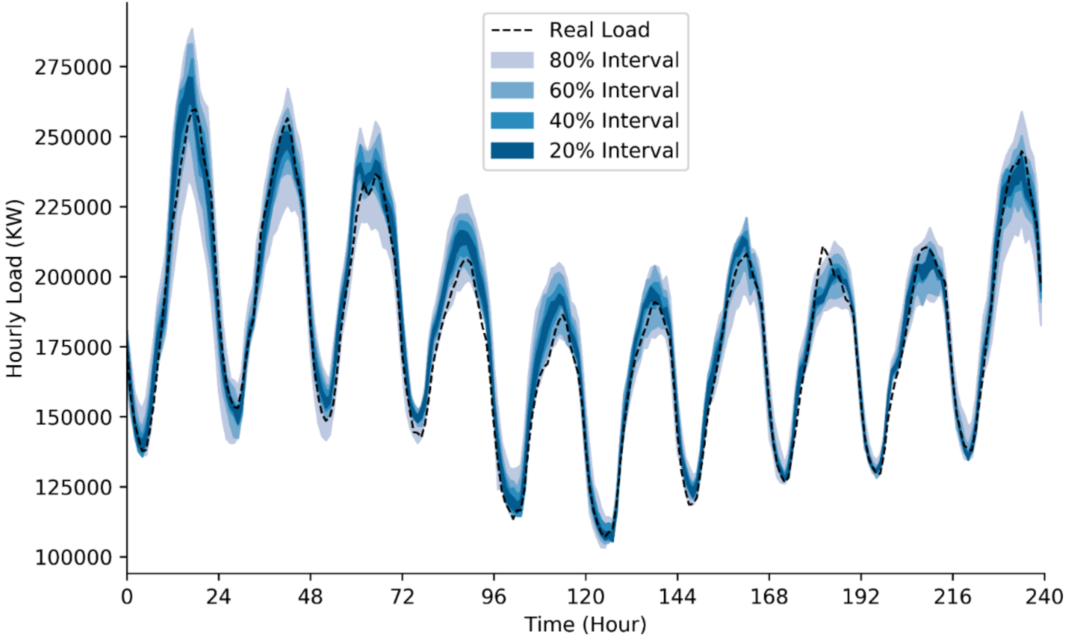
IEEE PES Power & Energy Society®

IEEE

Outline

- Background
- Problem formulation
- Combining quantile forecasts
- Combining density forecasts
- Conclusions

Backgrounds



Probabilistic load forecasts (PLF) provide comprehensive information about future uncertainties.

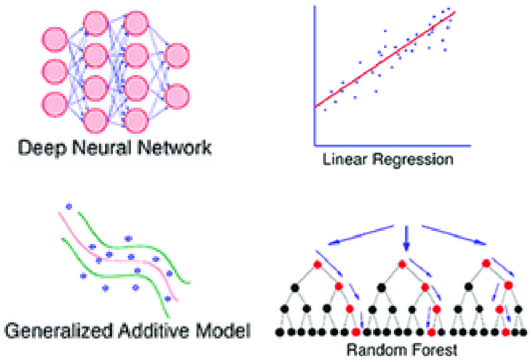
PLFs can be in the form of quantiles, intervals, or density functions.

Backgrounds

*In statistics and machine learning, **ensemble methods** use **multiple learning algorithms** to obtain **better predictive performance** than could be obtained from any of the constituent learning algorithms alone [1].*

Western Phrase: Two heads are better than one;

Chinese Saying: Three vice-generals are equal to one Zhuge Liang



- ❖ Which method is the best?
- ❖ Is it possible to combine these methods?

[1] https://en.wikipedia.org/wiki/Ensemble_learning

Backgrounds

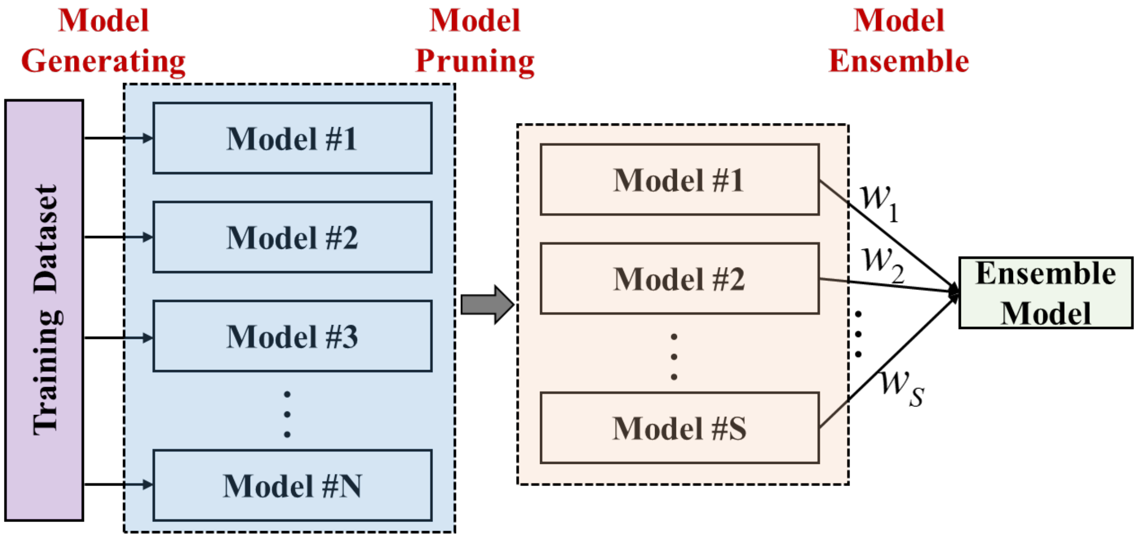
Various ensemble methods have been studied to combine multiple point forecasts. However, **combining probabilistic load forecasts is a rarely touched area**.

Combine point forecasts	Combine probabilistic forecasts
One dimension	High dimension
RMSE, MAPE	Reliability, sharpness, calibration
Analytical solution	???

Contributions of our work:

- ***New problem***: *Extend the ensemble method to the PLF area;*
- ***Elegant formulation***: *Formulate the combining problem into an LP or QP model.*

Problem formulation



1) How to generate multiple PLF models?

2) Among the *N* forecasting models, how many and which methods should be selected for the final ensemble formation process?

3) How much weight should be given to each method for the optimal combination?

Problem formulation

Loss function

Combined forecast

Determine the weights

$$\min_{\omega_n} \quad TL = \sum_{t=1}^{T} L\left(\sum_{n=1}^{N} \omega_n F_{n,t}, y_t\right)$$

Real load

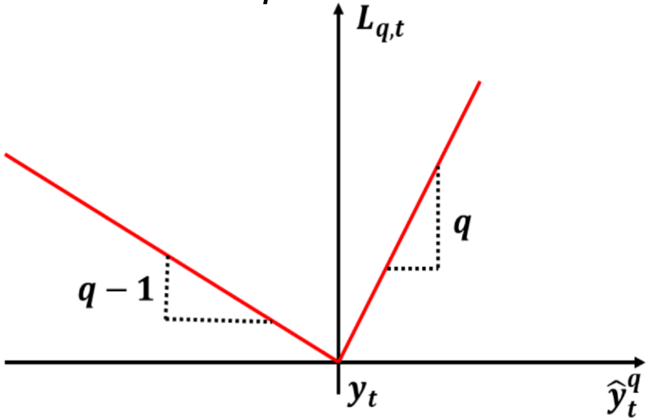
$$\text{s.t.} \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0$$

Summation and non-negative constraints

A deep investigation of the loss function is the key to formulate and solve the optimization problem.

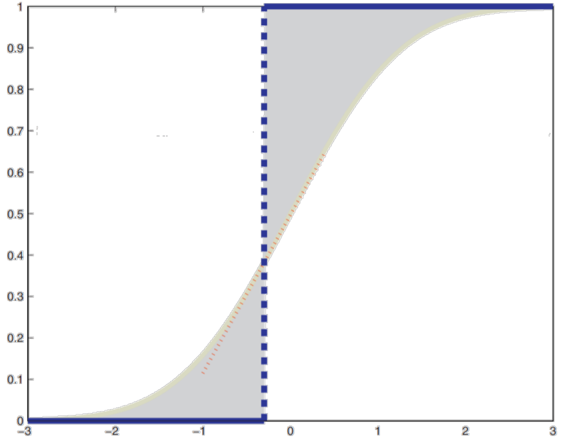
Problem formulation

Pinball loss *and* ***CRPS*** *assess the calibration and sharpness simultaneously, thus balancing the statistical consistency between the distributional forecasts and the observations and the concentration of the predictive distributions*



Pinball loss

$$\mathrm{PL}\left(\hat{y}_{t,q}, y_t\right) = \begin{cases} \left(y_t - \hat{y}_{t,q}\right) q & \hat{y}_{t,q} \le y_t \\ \left(\hat{y}_{t,q} - y_t\right)(1-q) & \hat{y}_{t,q} > y_t \end{cases}$$



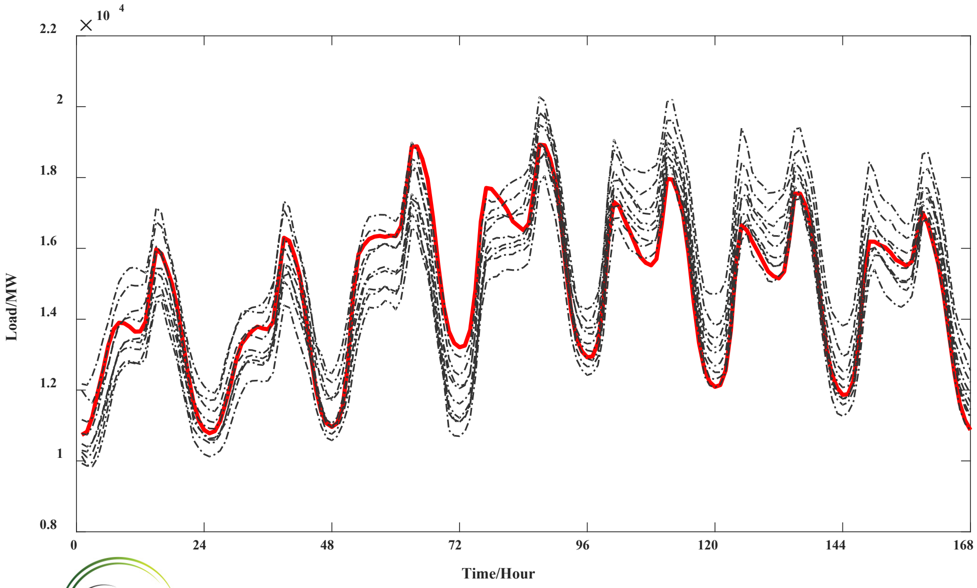
Continuous ranked probability score (CRPS)

$$\mathrm{CRPS}\left(F_t, y_t\right) = \int_{-\infty}^{\infty} \left(F_t(z) - 1\left(z - y_t\right)\right)^2 dz$$

Combining quantile forecasts

$$\mathrm{PL}\left(\hat{y}_{t,q}, y_t\right) = \begin{cases} \left(y_t - \hat{y}_{t,q}\right) q & \hat{y}_{t,q} \le y_t \\ \left(\hat{y}_{t,q} - y_t\right)\left(1-q\right) & \hat{y}_{t,q} > y_t \end{cases}$$

$$\hat{y}_{q,t} \approx \sum_{n=1}^{N} \omega_{n,q} \hat{y}_{n,q,t}$$



$$\min_{\omega_{n,q}} \quad TL = \sum_{t=1}^{T} \mathrm{PL}\left(\sum_{n=1}^{N} \omega_{n,q} \boxed{\hat{y}_{n,q,t}}, y_t\right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \omega_{n,q} = 1, \quad \omega_{n,q} \ge 0$$

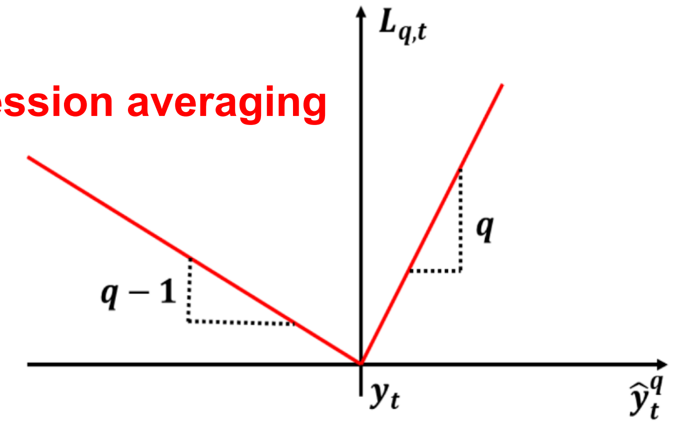
Combining quantile forecasts

Constrained quantile regression averaging

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

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

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



Auxiliary decision variables $v_{t,q} = \max\left\{q\left(y_t - \hat{y}_{t,q}\right), (1-q)\left(\hat{y}_{t,q} - y_t\right)\right\}$

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

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

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

- **LP problem**
- **Model selection**

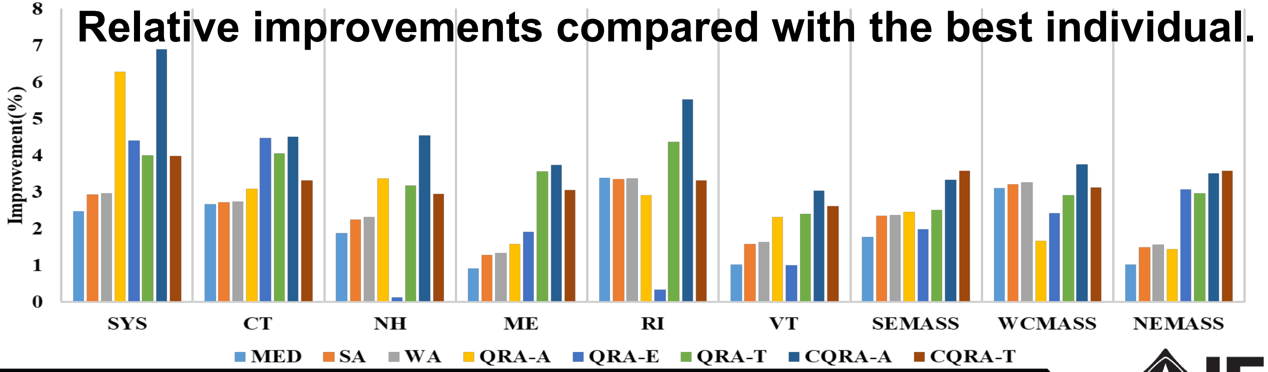
Combining quantile forecasts

We try to combine 13 individual methods and test the combined forecasts in ISO-NE dataset.

Pinball losses of different combining methods.

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

Relative improvements compared with the best individual.



Combining quantile forecasts

Models that are selected for different quantiles for total load (SYS).

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

Combining quantile forecasts

Models that are selected for the 90-th quantile for different zones.

Zones / Models	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

Combining density forecasts

The applications of the CRPS have been hampered by a lack of readily computable solutions to the integral:

$$\mathrm{CRPS}\left(F_t, y_t\right) = \int_{-\infty}^{\infty} \left(F_t\left(z\right) - 1\left(z - y_t\right)\right)^2 dz$$

This drawback is overcome by [1]:

$$\mathrm{CRPS}\left(F_t, y_t\right) = E\left|Y - y\right| - \frac{1}{2} E\left|Y - Y'\right|$$

Let's consider a simple case: **Gaussian Mixture Distribution**

[1] L. Baringhaus and C. Franz, "On a new multivariate two-sample test," Journal of Multivariate Analysis, vol. 88, no. 1, pp. 190–206, 2004.

Combining density forecasts

Two lemmas for Gaussian model: $$\mathrm{CRPS}(F_t, y_t) = E|Y - y| - \frac{1}{2}E|Y - Y'|$$

Lemma 1: *The expectation of an absolute value of a finite mixture distribution is the* ***weighted sum of the corresponding expectations*** *of absolute values of the components of the finite mixture distribution.*

If $X_1, X_2, \cdots X_N$ *are the* N *components of the finite mixture distribution* X *with weights* $\omega_1, \omega_2, \cdots \omega_N$ *, then* $E|X| = \sum_{n=1}^{N} \omega_n E|X_n|$

Combining density forecasts

Two lemmas for Gaussian model: $\mathrm{CRPS}(F_t, y_t) = E|Y-y| - \frac{1}{2}E|Y-Y'|$

Lemma 2: *If X and Y are independent random variables that are finite mixtures of normal distributions, then their sum is also a finite mixture of normal distributions. i.e., if*

$$f_X(x) = \sum_{l=1}^{L} \omega_l \phi(x \mid \mu_l, \sigma_l),\ f_Y(y) = \sum_{m=1}^{M} \omega_m \phi(y \mid \mu_m, \sigma_m)$$

$$\omega_l \geq 0,\quad \omega_m \geq 0,\quad \sum_{l=1}^{L} \omega_l = 1,\quad \sum_{m=1}^{M} \omega_m = 1$$

where $\phi(\cdot \mid \mu, \sigma)$ *is the PDF of normal distribution* $N(\mu, \sigma)$ *, then the PDF of* $Z=X+Y$ *is:*

$$f_Z(z) = \sum_{l=1}^{L} \sum_{m=1}^{M} \omega_l \omega_m \phi(x \mid \mu_l + \mu_m, \sqrt{\sigma_l^2 + \sigma_m^2})$$

Combining density forecasts

If individual density forecasts are **Gaussian distributed** $f_n(x)$, the combined forecast follows Gaussian mixture distribution:

$$f_X(x) = \sum_{n=1}^{N} \omega_n f_n(x)$$

Then, the CRPS can be calculated as:

$$\mathrm{CRPS}(F, y) = \sum_{n=1}^{N} \omega_n E\left|Y_n - y\right| - \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{2} \omega_i \omega_j E\left|Y_i - Y'_j\right|$$

The expectation of the absolute value of a normal distribution $N(\mu, \sigma)$ can be calculated as follows:

$$E\left|X\right| = \int_{-\infty}^{\infty} \left|x\right| f(x)dx = \int_{-\infty}^{0} -xf(x)dx + \int_{0}^{\infty} \left|x\right| f(x)dx$$

$$= \sigma\sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2}{2\sigma^2}} + \mu[2\Phi(\frac{\mu}{\sigma}) - 1]$$

Combining density forecasts

Thus, we have:

$$\mathrm{CRPS}(F,y)=\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i,j}\omega_i\omega_j+\sum_{n=1}^{N}\beta_n\omega_n$$

where

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

$$\beta_n=\sqrt{\frac{2}{\pi}}\sigma_n\exp\left(-\frac{(\mu_n-y)^2}{2\sigma_n^2}\right)+(\mu_n-y)\left[2\Phi\left(\frac{(\mu_n-y)}{\sigma_n}\right)-1\right]$$

Finally, we have:

$$\min_{\omega}\quad \omega^T Q\omega+c^T\omega$$

$$\text{s.t.}\quad \mathbf{1}^T\omega=1\quad \omega\geq 0$$

QP problem!

Is this convex?

Combining density forecasts

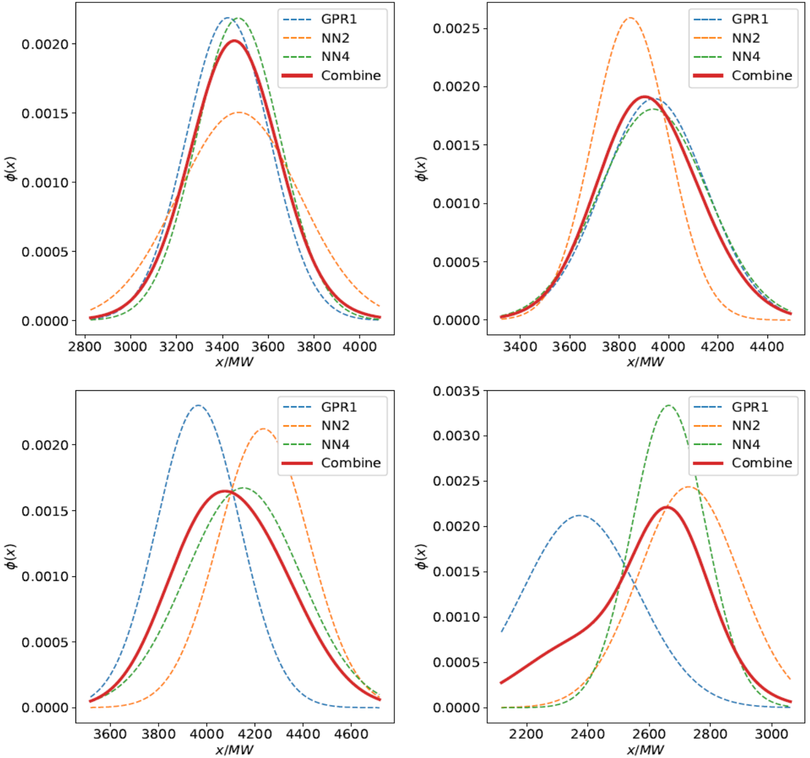
Performances of combined models

Metric \ Method	Simple Average	MAPE Based	CRPS Based
MAPE / %	5.48	5.09	5.11
MAE / MW	214.41	198.99	199.43
CRPS / MW	151.19	141.58	141.07

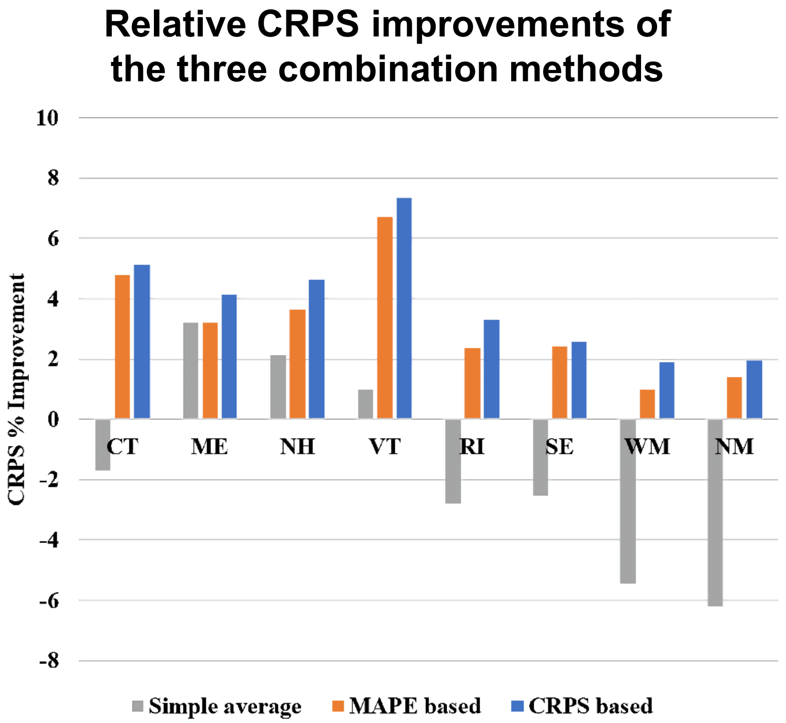
CRPS of the Best Individual Model and Combined Models

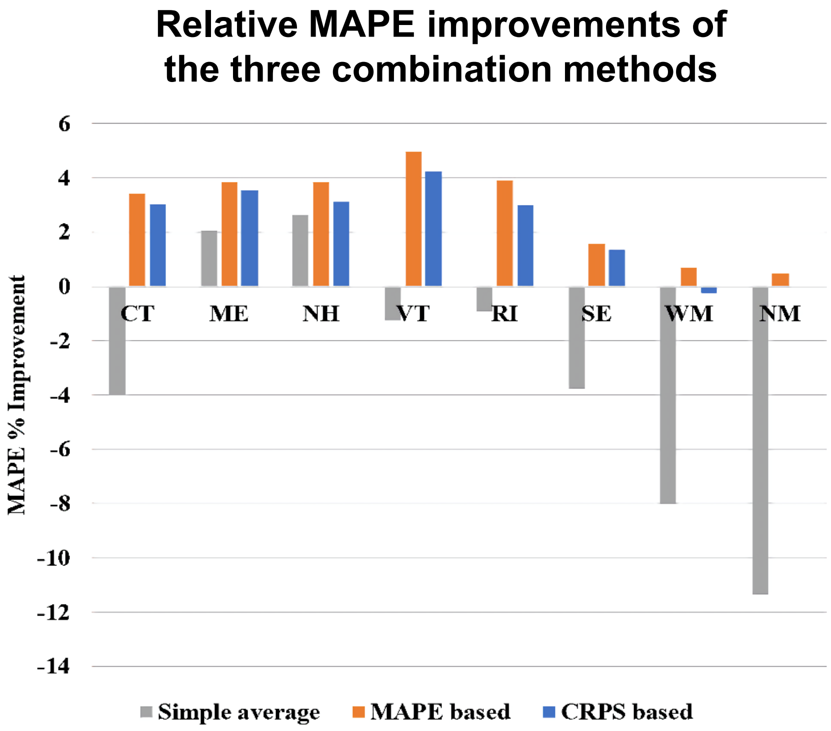
Region \ Method	Best Individual	Simple Average	MAPE Based	CRPS Based
CT	148.7	151.19	141.58	141.07
ME	33.8	32.71	32.71	32.40
NH	43.0	42.08	41.43	41.01
VT	21.2	20.99	19.78	19.64
RI	31.4	32.28	30.66	30.36
SE	61.4	62.95	59.91	59.83
WM	71.4	75.28	70.68	70.04
NM	103.7	110.14	102.25	101.66

PDFs of predictions of four typical days given by the base models and their combination



Combining density forecasts





Combining density forecasts

Weights of the base models in the CRPS-guided model

	GPR1	GPR2	GPR3	GPR4	GPR5	NN1	NN2	NN3	NN4	NN5	XGB1	XGB2	XGB3	XGB4	XGB5
CT	0.33	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.47	0.00	0.00	0.00	0.00	0.00	0.00
ME	0.00	0.04	0.00	0.11	0.22	0.14	0.09	0.03	0.00	0.21	0.16	0.00	0.00	0.00	0.00
NH	0.15	0.01	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.07	0.09	0.18	0.14	0.06	0.27
VT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.12	0.07	0.12	0.14	0.13	0.23
RI	0.00	0.00	0.01	0.00	0.00	0.00	0.09	0.00	0.00	0.02	0.34	0.00	0.13	0.24	0.17
SEMASS	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.37	0.01	0.34
WCMASS	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.09	0.09	0.00	0.38	0.13	0.22
NEMASSBOST	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.04	0.00	0.31	0.33	0.15

Weights of the base models in the MAPE-guided model

	GPR1	GPR2	GPR3	GPR4	GPR5	NN1	NN2	NN3	NN4	NN5	XGB1	XGB2	XGB3	XGB4	XGB5
CT	0.26	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00
ME	0.00	0.12	0.00	0.00	0.40	0.00	0.32	0.02	0.00	0.06	0.07	0.00	0.00	0.00	0.00
NH	0.08	0.16	0.01	0.00	0.00	0.11	0.00	0.03	0.00	0.02	0.22	0.00	0.00	0.10	0.27
VT	0.00	0.00	0.04	0.00	0.00	0.02	0.00	0.00	0.15	0.05	0.05	0.13	0.29	0.00	0.26
RI	0.00	0.02	0.03	0.00	0.08	0.00	0.10	0.04	0.00	0.00	0.30	0.00	0.31	0.13	0.00
SEMASS	0.00	0.08	0.01	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.10	0.00	0.48	0.00	0.29
WCMASS	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.12	0.20	0.00	0.60	0.00	0.00
NEMASSBOST	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.54	0.43	0.00

Conclusions

- ➢ Two combining approaches are proposed for multiple probabilistic load forecasts in the form of quantiles or density;
- ➢ The combining model is formulated as an LP or QP problem that can be easily solved;
- ➢ The final combined model can further improve the forecasting performance compared with the best individual model;
- ➢ These methods are tested on several open load datasets.

- ❖ Prove the convexity of the QP model?

$$\min_{\omega} \quad \omega^T Q\omega + c^T \omega \quad \text{s.t.} \;\; \mathbf{1}^T \omega = 1 \;\; \omega \geq 0$$

- ❖ Apply the combining methods to renewable energy and price forecasting?

Thanks for your attention.

Yi Wang

ETH Zurich

www.eeyiwang.com

August 8, 2019, Atlanta

IEEE PES Power & Energy Society®

IEEE