iSPEC 2020 2020 IEEE Sustainable Power & Energy Conference

Aggregated load forecasting with fine-grained smart meter data: An ensemble learning approach

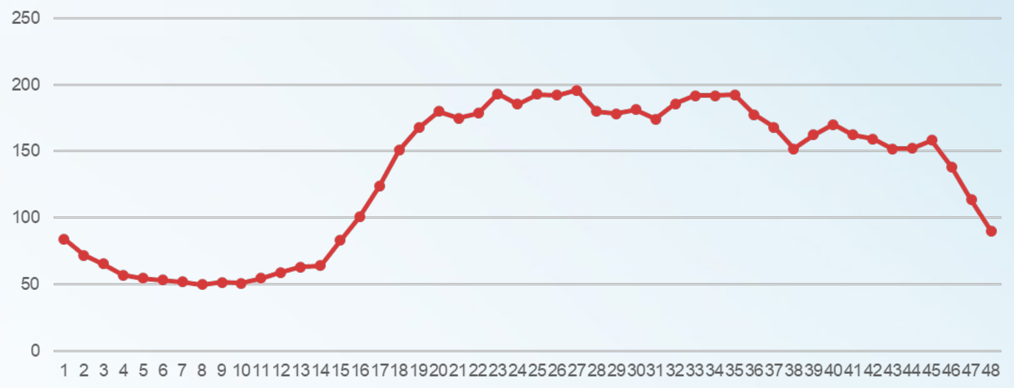
Yi Wang, ETH Zürich

yiwang@eeh.ee.ethz.ch, www.eeyiwang.com

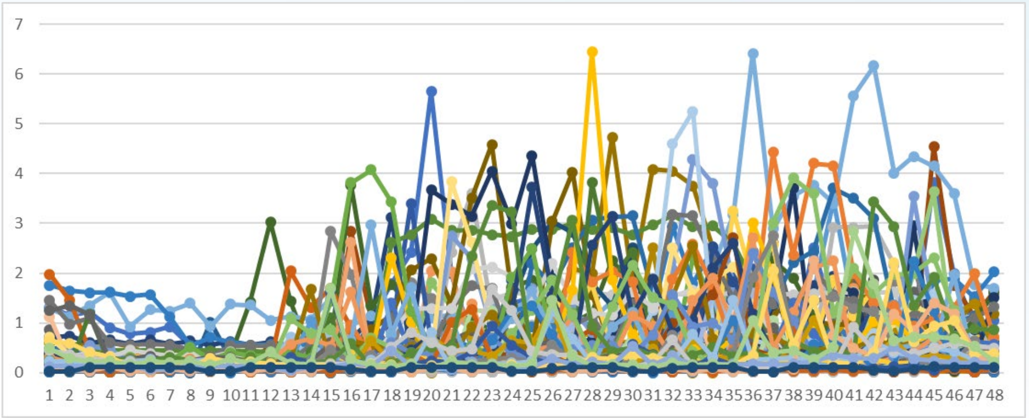
Contents

Introduction

Traditional load forecasting algorithms directly use historical data at the aggregation level.



With the prevalence of smart meters, fine-grained sub profiles reveal more information about the aggregated load and further help improve the forecasting accuracy.

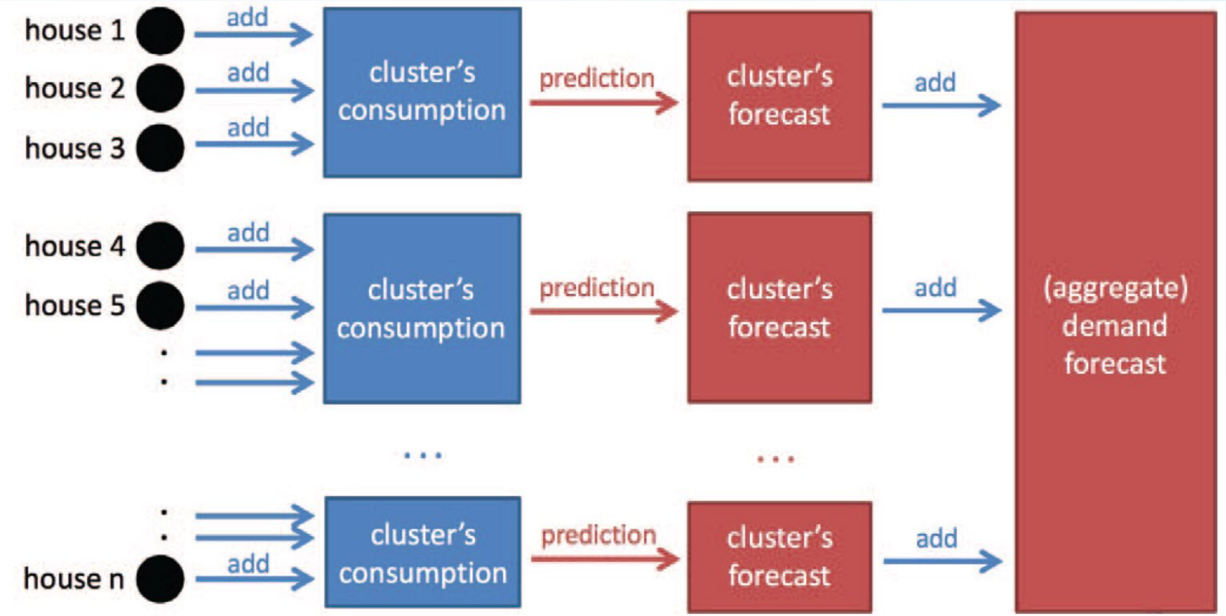


Introduction

Three strategies for aggregated load forecasting (ALF):

1) Top-down; 2) bottom-up; 3) clustering based.

Is it possible to utilize both ensemble techniques and fine-grained subprofiles to further improve the aggregated load forecasting accuracy?

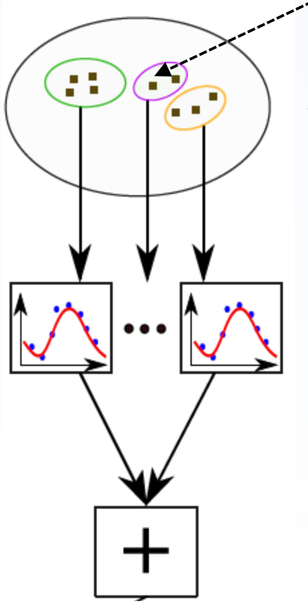


Introduction

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

A three-stage approach for aggregated load forecasting with smart meter data:

Load profile of one consumer



Clustering: divide consumers into different groups

Forecasting: develop forecasting model for each group

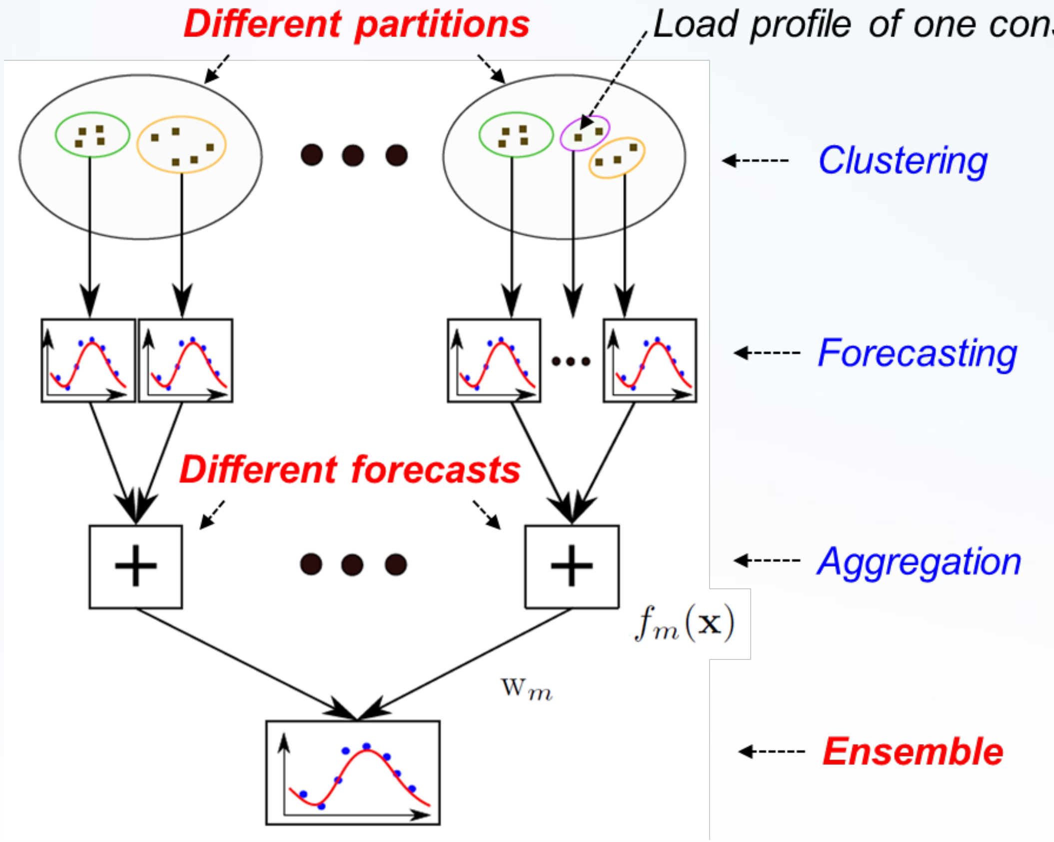
Aggregation: sum forecasts of all groups

Introduction

Go further steps by ensemble learning?

Deterministic ALF

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



Combined model:

$$G(\mathbf{x}) = \sum_{m=1}^{M} w_m f_m(\mathbf{x})$$

Deterministic ALF

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

$$\min_{\mathbf{w}} \quad L(y - G(\mathbf{x}))$$
$$\text{s.t.} \quad \sum_{m=1}^{M} \mathrm{w}_m = 1, \quad \mathrm{w}_m \geq 0, \quad m = 1, ..., M$$

Real load — **The *n*-th predicted load**

$$\hat{\omega} = \arg\min_{\omega} \sum_{t=1}^{T} \frac{1}{T} \frac{|L_{en,t} - \hat{L}_{en,t}|}{L_{en,t}}$$

→ **Minimize MAPE**

$$s.t. \;\; \hat{L}_{en,t} = \sum_{n=1}^{N} \omega_n \hat{L}_{en,n,t}, \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0.$$

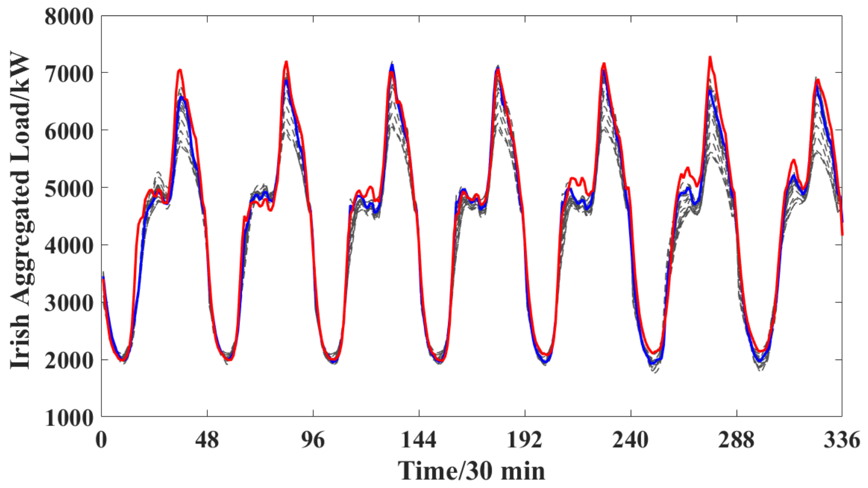
It can be formulated as an LP problem.

To determine the weights for the forecasts

Deterministic ALF

Weights, MAPE, and RMSE of different forecasts with different groups

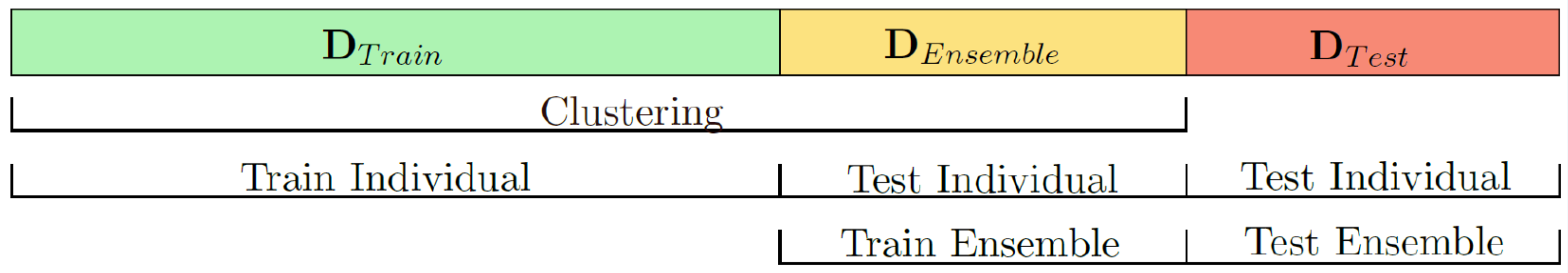
N	1	2	4	8	16	32	64	128	256	...	5237	Ensemble
ω	0.634	0	0	0.271	0	0	0.095	0	0	...	0	/
MAPE	***4.25%***	5.05%	5.29%	4.74%	5.55%	4.66%	4.79%	5.09%	5.59%	...	10.31%	**4.05%**
RMSE	***210.95***	229.73	228.01	217.68	244.9	217.64	227.36	232.61	250.27	...	441.33	**202.88**



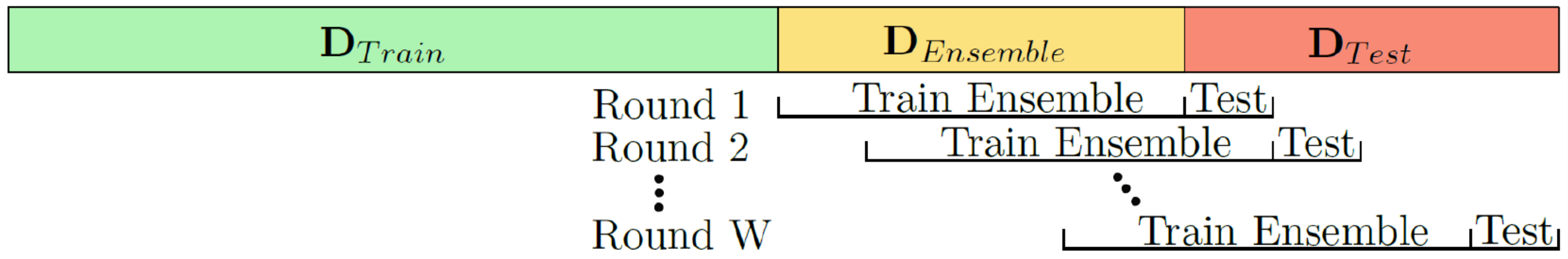
red line: actual load blue line: ensemble forecast
dashed lines: individual forecasts

The MAPE and RMSE of the proposed ensemble method are 4.05% and 202.88 which gain 4.71% and 3.83% improvements, respectively compared with the best individual forecast.

Deterministic ALF



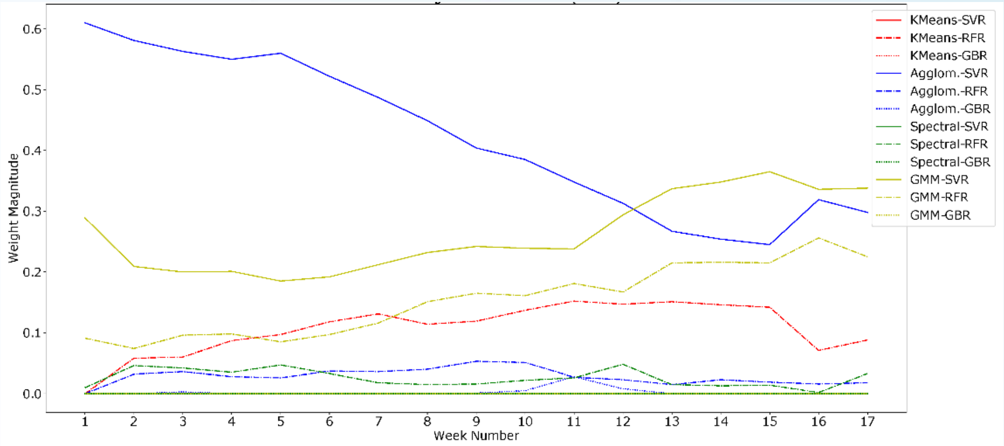
Can we update the weights in a rolling window-based manner?



Deterministic ALF

Benefits of window-based method

Ensemble Method	Error Metrics	Window	Benchmark
COP_{MAPE}	MAPE	2.85%	3.13%
	MAE	106.13	116.66
	RMSE	149.81	166.74
COP_{MSE}	MAPE	2.89%	3.15%
	MAE	107.3	116.8
	RMSE	151.26	166.92

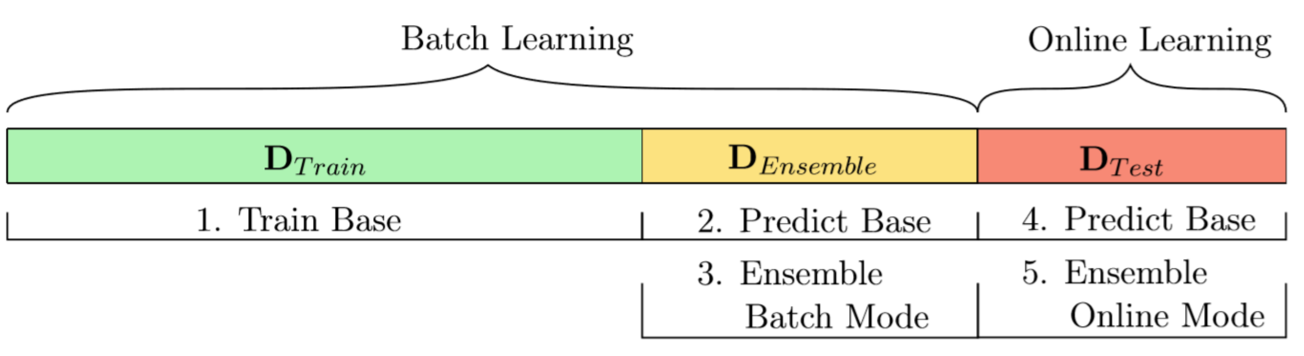


Ensemble weights over 17 weeks of the test set for all individual models.

Deterministic ALF

Combined model:

$$G(\mathbf{x}) = \sum_{m=1}^{M} \mathrm{w}_m f_m(\mathbf{x})$$



Algorithm 1: Online Protocol

input: Initial model weights $\mathbf{w}_1 \in \mathbb{R}^M$, convex loss function ℓ, weight update rule U

for $t = 1, 2, ...$

- Calculate individual predictions $\mathbf{f}_t \in \mathbb{R}^M$
- Predict $\hat{y}_t = \mathbf{w}_t \cdot \mathbf{f}_t$
- Reveal true value $y_t \in \mathbb{R}$
- Calculate loss $\ell(y_t, \hat{y}_t)$
- Update model $\mathbf{w}_{t+1} = U\big(\mathbf{w}_t; \ell(y_t, \hat{y}_t)\big)$

end

Online Convex Optimization (OCO) is a unifying framework for the analysis and design of online algorithms.

Deterministic ALF

- **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

Prevent information loss

Loss ℓ

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

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

Passive: weights do not change every time slot

Deterministic ALF

Update the weights online for a better performance

Errors on test set after online learning

Method	MAPE	SD	MAE	RMSE
SGDR	2.43%	0.025	86.05	122.71
FTRLP	2.23%	0.021	81.09	113.87
OSELM	2.80%	0.029	106.03	155.03
Online Bagging	2.07%	0.021	74.33	106.23
PAR	1.67%	0.015	61.83	86.68
Proposed	1.62%	0.014	59.59	83.21
Best SVR	3.18%	0.032	117.54	171.72
Best RF	2.89%	0.029	108.25	156.84
Best GBRT	3.53%	0.032	127.81	175.78
Batch OPT	2.89%	0.028	107.55	154.88
Window OPT	2.85%	0.028	106.13	149.81

SD: Standard deviation of the absolute percentage error

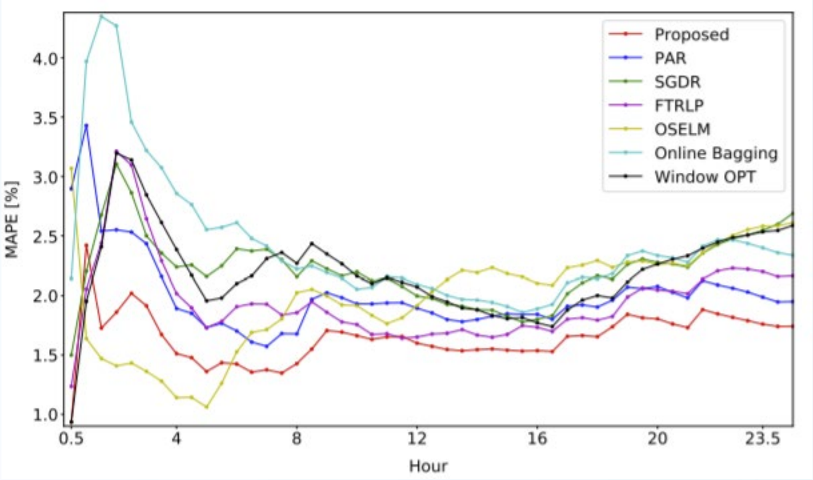
- All ensembles improve their forecasting performance through online learning.
- Nearly all ensembles outperform the benchmarks after online learning.
- The proposed method has the highest accuracy and stability among all examined ensembles.

Deterministic ALF

Update the weights online for a better performance

The hour of break-even for all ensembles

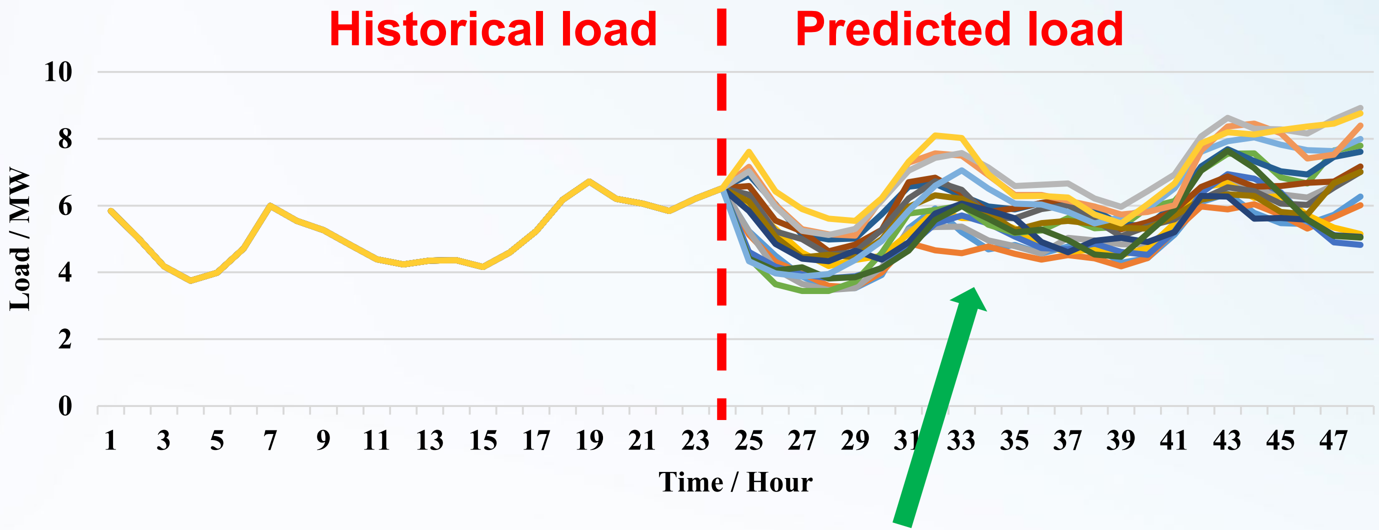
Method	Break-even [hour]			
	MAPE	SD	MAE	RMSE
SGDR	39.5	86.5	41.0	64.0
FTRLP	66.5	87.0	64.0	60.5
PAR	17.5	9.0	19.5	17.5
OSELM	112.0	2.0	2833.5	no
Online Bagging	22.5	4.5	23.0	35.5
Proposed	1.5	2.0	1.5	1.5



MAPE over the course of the first day of online learning

- The proposed method has the earliest break-even after 2 hours for all metrics.
- The other ensembles have the break-even approximately within one or two days.
- An ensemble employing online learning is able to pay off at a relatively early point in time.

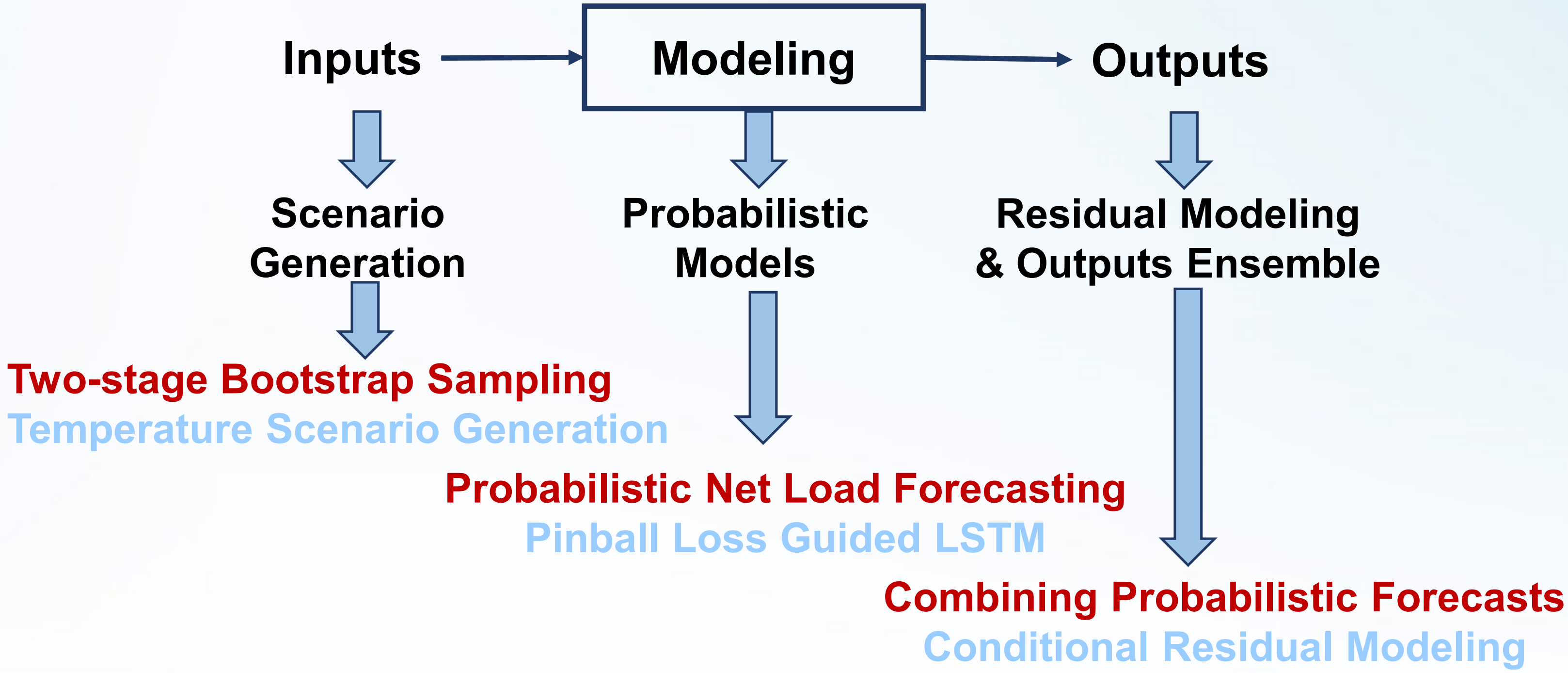
Probabilistic ALF



Compared with deterministic forecasting, probabilistic load forecasts provide comprehensive information about future uncertainties.

Probabilistic ALF

How to obtain probabilistic forecasting?



Probabilistic ALF

Pinball loss (PL) and **Winkler Score (WS)** assess the calibration and sharpness simultaneously.

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

Performance of overall quantiles

$$\mathrm{WS}(L_t, U_t, y_t) = \begin{cases} \delta_t + 2(L_t - y_t)/\alpha & y_t \le L_t \\ \delta_t & L_t < y_t < U_t \\ \delta_t + 2(y_t - U_t)/\alpha & U_t \le y_t \end{cases}$$

Performance of extreme quantiles

Average Coverage Error (ACE) evaluate the reliability of the forecasts.

$$ACE = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{y_i \in [L_i, U_i]\}} - (1-\alpha)$$

Performance of an certain interval

Probabilistic ALF

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

PCA

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

Factor Quantile Regression Averaging

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

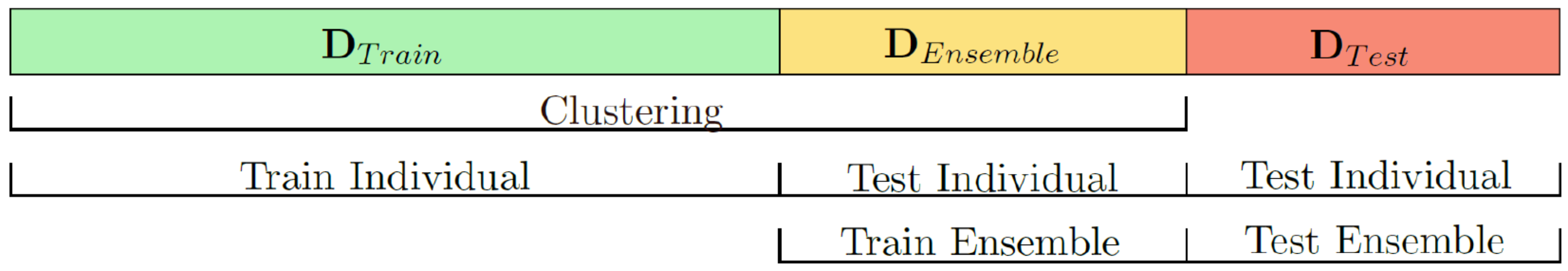
Quantile regression averaging (QRA), a special form of quantile regression, is a kind of model averaging method.

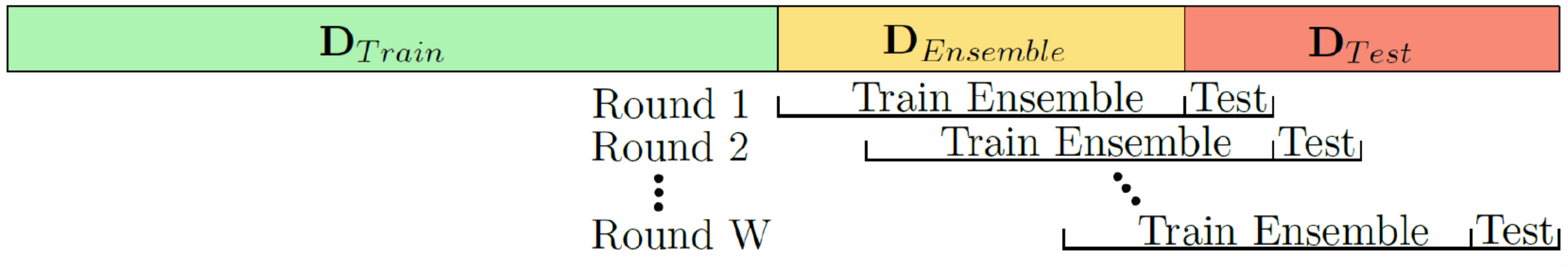
$$\hat{\mathbf{w}}_q = \arg\min_{\mathbf{w}_q} \sum_{i=1}^{n} \rho_q(y_i - \mathbf{x}_i \mathbf{w}_q) + \lambda \|\mathbf{w}_q\|_1$$

LASSO Quantile Regression Averaging

Probabilistic ALF

Similar to deterministic forecasting……





Probabilistic ALF

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

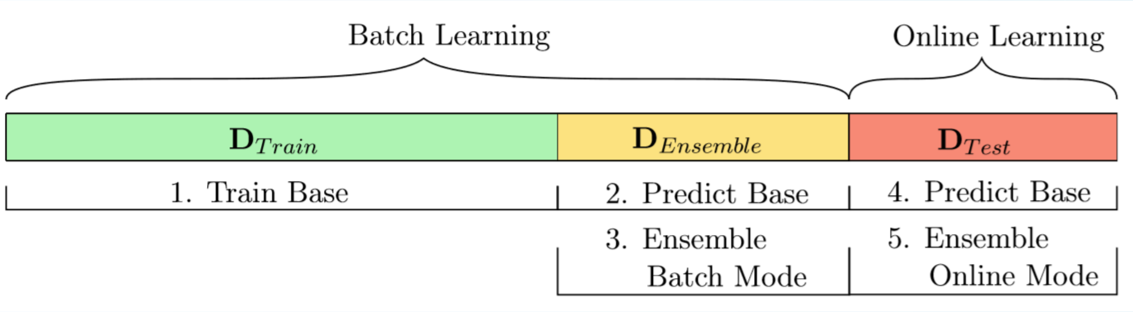
Ensemble Method	Error Metrics	Offline Ensemble	Benchmark 1	Rolling Window-based Ensemble	Benchmark 2
QRA	ACE	-1.73%	-1.85%	-0.56%	-0.92%
	PBL	45.82	50.19	42.28	46.52
	WKS	788.62	846.89	728.13	791.78
FQRA	ACE	-1.80%	-1.85%	**-0.45%**	-0.92%
	PBL	45.82	50.19	**42.26**	46.52
	WKS	787.26	846.89	**727.24**	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.

Probabilistic ALF

Combined model:

$$G(\mathbf{x}) = \sum_{m=1}^{M} \mathrm{w}_m f_m(\mathbf{x})$$



- **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

Prevent information loss

Loss ℓ

Integrate new sample

Probabilistic ALF

- General Formula

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

- L_2-distance :

$$d(\cdot) = \frac{1}{2} \|\cdot\|^2$$

- ε-insensitive quantile loss :

$$\ell_{\varepsilon,q}(\mathbf{w}_q; \mathbf{x}, y) = \begin{cases} q(y - \mathbf{w}_q \cdot \mathbf{x} + \varepsilon(q-1)) & \text{if } y - \mathbf{w}_q \cdot \mathbf{x} > \varepsilon(1-q) \\ 0 & \text{if } -\varepsilon q < y - \mathbf{w}_q \cdot \mathbf{x} < \varepsilon(1-q) \\ (q-1)(y - \mathbf{w}_q \cdot \mathbf{x} + \varepsilon q) & \text{if } y - \mathbf{w}_q \cdot \mathbf{x} < -\varepsilon q \end{cases}$$

- Solving KKT conditions:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t sign(y_t - \mathbf{w}_t \cdot \mathbf{x}_t)\tau_t \mathbf{x}_t \qquad \tau_t = min\left\{C, \frac{\ell_{\varepsilon,q}(y_t, \mathbf{w}_t \cdot \mathbf{x}_t)}{q\,\|\mathbf{x}_t\|_2^2}\right\}$$

Probabilistic ALF

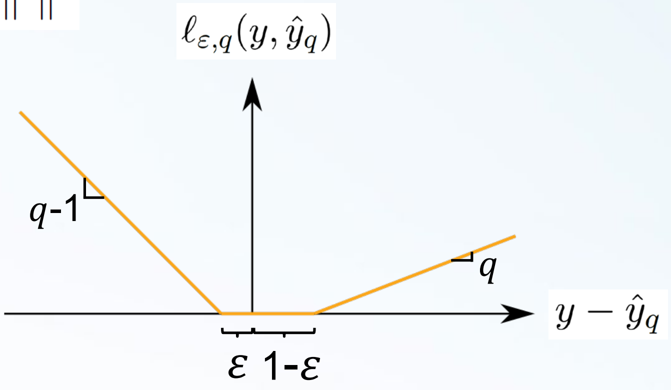
- General Formula

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

- L_2-distance :

$$d(\cdot) = \frac{1}{2}\left\|\cdot\right\|^2$$

- ε-insensitive quantile loss :



- Solving KKT conditions:

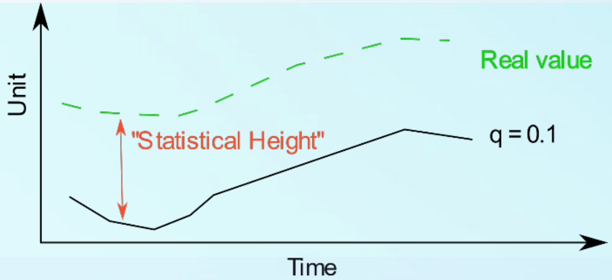
$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t sign(y_t - \mathbf{w}_t \cdot \mathbf{x}_t)\tau_t \mathbf{x}_t \qquad \tau_t = min\left\{C, \frac{\ell_{\varepsilon,q}(y_t, \mathbf{w}_t \cdot \mathbf{x}_t)}{q\left\|\mathbf{x}_t\right\|_2^2}\right\}$$

Probabilistic ALF

Mechanism of Quantile Passive Aggressive Regression

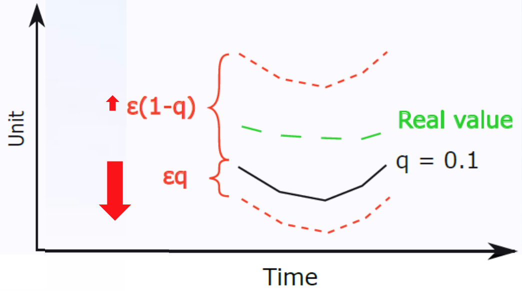
- Extension to probabilistic forecasting: **ε-insensitive loss -> ε-insensitive quantile loss**
- ε-insensitive region: Preserve «quantile height» between y_q and y

- Batch quantile regression
  - Access to whole data sequence
  - «Statistical height» implicitly given



$$\ell_q(y, \hat{y}_q) = \begin{cases} q(y - \hat{y}_q) & \text{if } y \geq \hat{y}_q \\ (q-1)(y - \hat{y}_q) & \text{if } y < \hat{y}_q \end{cases}$$

- Online quantile regression
  - Only access to one sample per round
  - «Statistical height» collapses → Real value
- ε-insensitive quantile: Preserve «statistical height»



Probabilistic ALF

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

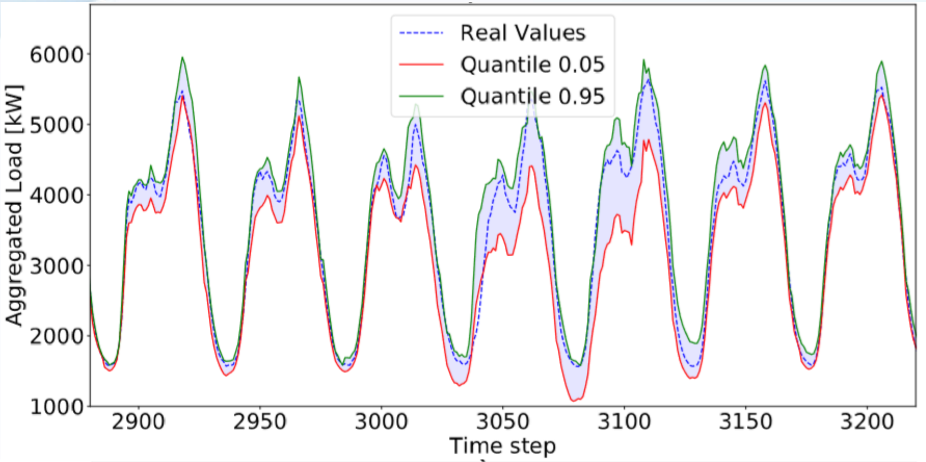
Errors on test set after online learning

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

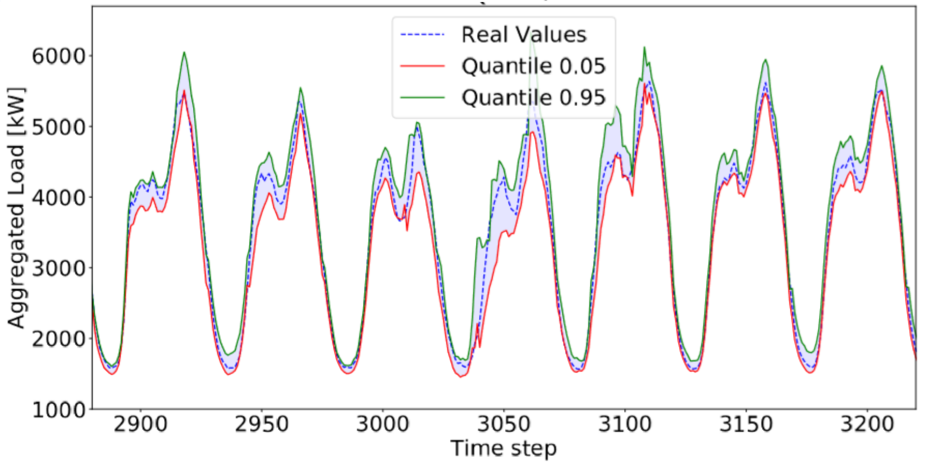
*Window OPT: window-based optimization

- 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.

Probabilistic ALF



QSGD online forecast over one week



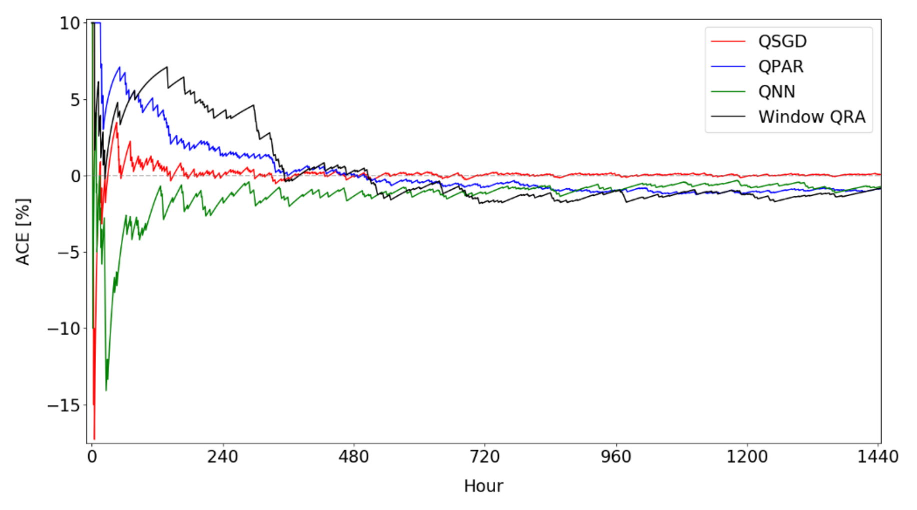
QPAR online forecast over one week

Probabilistic ALF

The performance on Irish load data

The hour of break-even for all ensembles

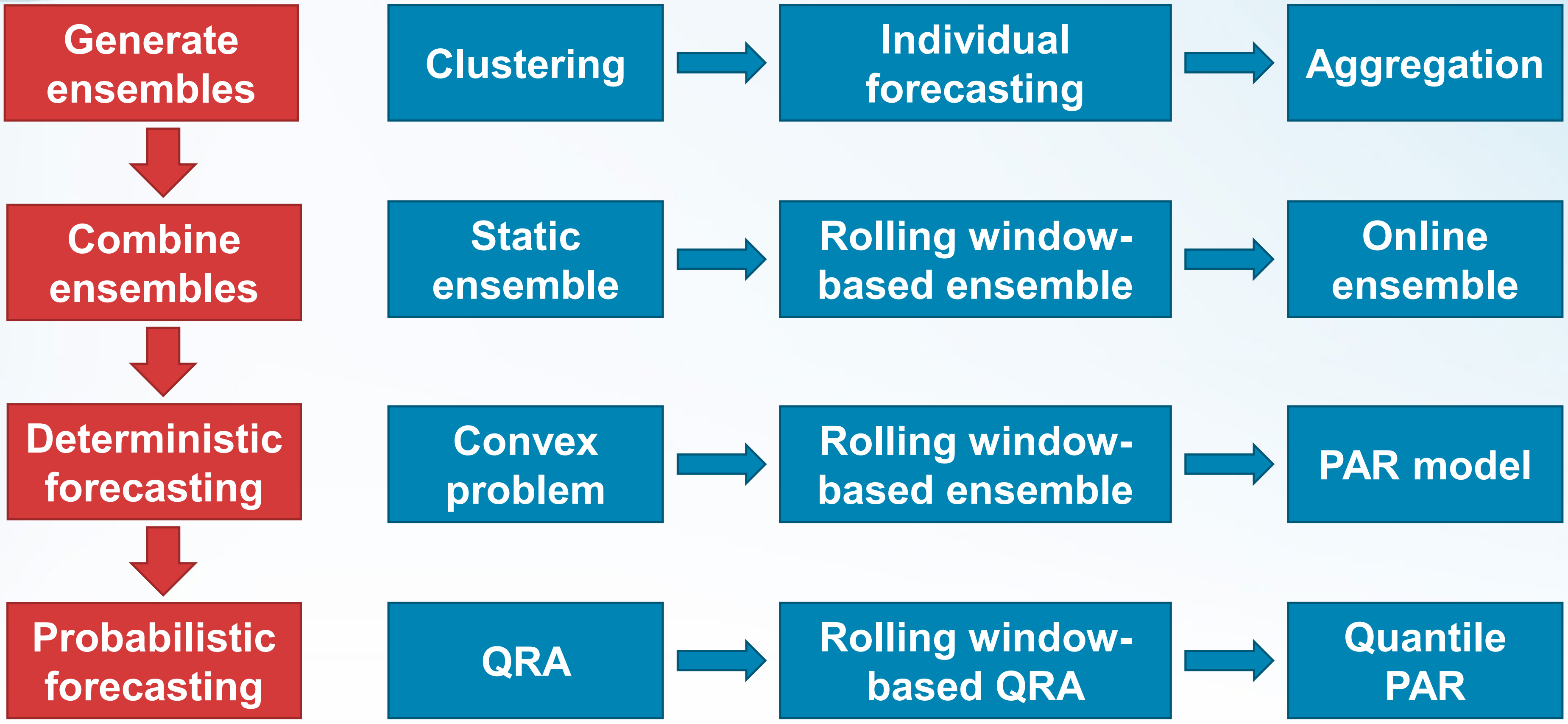
Method	Break-Even ACE	Break-Even PBL	Break-Even WKS
QSGD	508.0 h	35.0 h	307.0 h
QPAR	2810.0 h	138.5 h	253.5 h
QNN	687.0 h	no	no



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.

Conclusions



Conclusions

- High quality point forecasting can be generated by making full use of the fine grained smart meter data;
- On this basis, we can utilize ensemble techniques to further improve the forecasting accuracy;
- Online learning can be a powerful tool in short-term load forecasting by integration new information and the proposed modified PAR model is very suitable in this context, especially as an online ensemble method;
- PAR model can be further extend to quantile PAR model using quantile regression averaging for probabilistic forecasting.

References

1. Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," ***IEEE Transactions on Smart Grid***, 2019, 10(3):3125-3148.
2. Yi Wang, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, "Combining Probabilistic Load Forecasts," ***IEEE Transactions on Smart Grid***, 2019, 10(4):3664-3674.
3. Yi Wang, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, "Combining Probabilistic Load Forecasts," ***IEEE Transactions on Smart Grid***, 2019, 10(4):3664-3674.
4. Yi Wang, Qixin Chen, Mingyang Sun, Chongqing Kang and Qing Xia, "An Ensemble Forecasting Method for the Aggregated Load with Subprofiles," ***IEEE Transactions on Smart Grid***, 2018, 9(4): 3906-3908.
5. Leandro Von Krannichfeldt, Yi Wang*, and Gabriela Hug, "Online Ensemble Learning for Load Forecasting," ***IEEE Transactions on Power Systems***, in press.
6. Tainyi Li, Yi Wang*, and Ning Zhang, "Combining Probability Density Forecasts for Power Electrical Loads," ***IEEE Transactions on Smart Grid***, 2020, 11(2):1679-1690.
7. Yi Wang, Leandro Von Krannichfeldt, Gabriela Hug, "Probabilistic Aggregated Load Forecasting with Fine-grained Smart Meter Data," ***IEEE PowerTech 2021***, submitted.

Thank you for your attention

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