Supported by:

2024 IEEE Sustainable Power and Energy Conference (iSPEC 2024)

End-to-End Forecasting towards Economic Operation of Microgrid using Derivative-free Learning

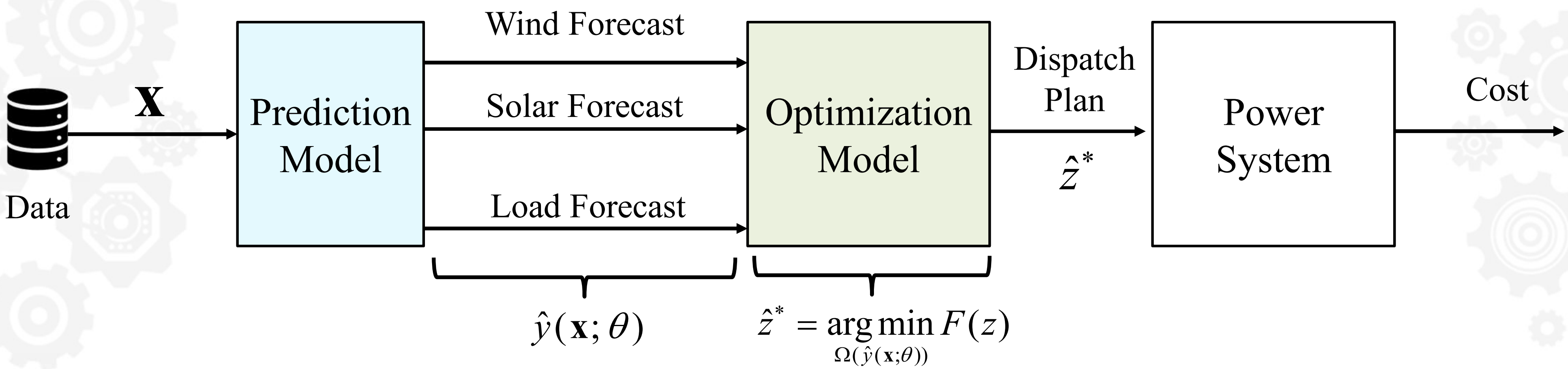
Chuanqing Pu

College of Smart Energy, Shanghai Jiao Tong University

sashabanks@sjtu.edu.cn

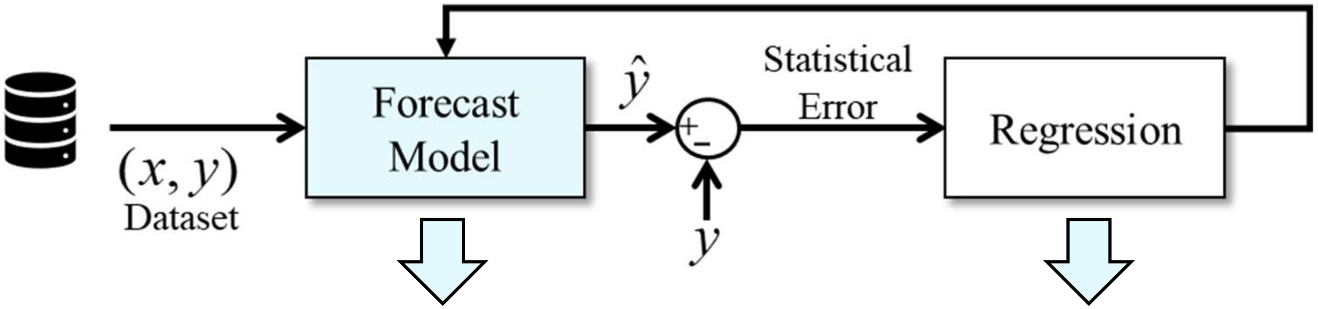
Predict then Optimize

- Step1: Predict uncertain parameters
- Step2: Optimize the decision variables



1. Training a Forecasting Model

- Statistical Loss-based Training Methods



Neural networks
Regression trees
GBDTs
....

$$\min_{\theta} L_{MSE}(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

2. Building a Optimization Model for MG

- Objective Function

$$\min_{\boldsymbol{z}^{(\mathrm{i})},\boldsymbol{z}^{(\mathrm{ii})}} F(\hat{\boldsymbol{y}}, \boldsymbol{z}^{(\mathrm{i})}, \boldsymbol{z}^{(\mathrm{ii})}) = \sum_{t=0}^{23} f(\hat{y}_t, z_t^{(\mathrm{i})}, z_t^{(\mathrm{ii})})$$

- Constraints

$$f(\hat{y}_t, z_t^{(\mathrm{i})}, z_t^{(\mathrm{ii})}) = \pi_t P_{\mathrm{G},t} + \omega_{\mathrm{B}} (P_{\mathrm{ESdch},t} - P_{\mathrm{ESch},t}) + \omega_{\mathrm{DG}} (P_{\mathrm{DG},t} + \Delta P_{\mathrm{DG},t}) + \omega_{\mathrm{C}} L_{\mathrm{C},t}$$

Forecasts $\hat{y}_t = \left\{\hat{P}_{\mathrm{w.max,t}}, \hat{P}_{\mathrm{pv.max,t}}, \hat{L}_t\right\}$

Decision (Stage I) $z_t^{(\mathrm{i})} = \{P_{\mathrm{ESdch},t}, P_{\mathrm{ESch},t}, P_{\mathrm{DG},t}\}$

Decision (Stage II) $z_t^{(\mathrm{ii})} = \{P_{\mathrm{w},t}, P_{\mathrm{pv},t}, \Delta P_{\mathrm{DG},t}, P_{\mathrm{G},t}, L_{\mathrm{C},t}\}$

Renewable Energy

$$0 \leqslant P_{\mathrm{w},t} \leqslant \hat{P}_{\mathrm{w.max,t}}$$
$$0 \leqslant P_{\mathrm{pv},t} \leqslant \hat{P}_{\mathrm{pv.max,t}}$$

Energy Storages

$$0 \leq P_{\mathrm{ESch},t}, P_{\mathrm{ESdch},t} \leq P_{\mathrm{ES.max}}$$
$$E_{\mathrm{ES.min}} \leq E_{\mathrm{ES},t} \leq E_{\mathrm{ES.max}}$$
$$E_{\mathrm{ES},t+1} = E_{\mathrm{ES},t} + \eta_{\mathrm{ch}} P_{\mathrm{ESch},t} - \frac{1}{\eta_{\mathrm{dch}}} P_{\mathrm{ESdch},t}$$

Diesel Generator

$$0 \leq P_{\mathrm{DG},t} \leq \alpha P_{\mathrm{DG.max}}$$
$$-k\delta_{\mathrm{rup}} P_{\mathrm{DG.\,max}} \leq \Delta P_{\mathrm{DG},t} \leq k\delta_{\mathrm{rup}} P_{\mathrm{DG.\,max}}$$
$$P_{\mathrm{DG},t+1} - P_{\mathrm{DG},t} \leq \delta_{\mathrm{rup}} P_{\mathrm{DG.\,max}}$$
$$P_{\mathrm{DG},t} - P_{\mathrm{DG},t+1} \leq \delta_{\mathrm{rdn}} P_{\mathrm{DG.\,max}}$$
$$0 \leq P_{\mathrm{G},t} \leq P_{\mathrm{G.max}}$$

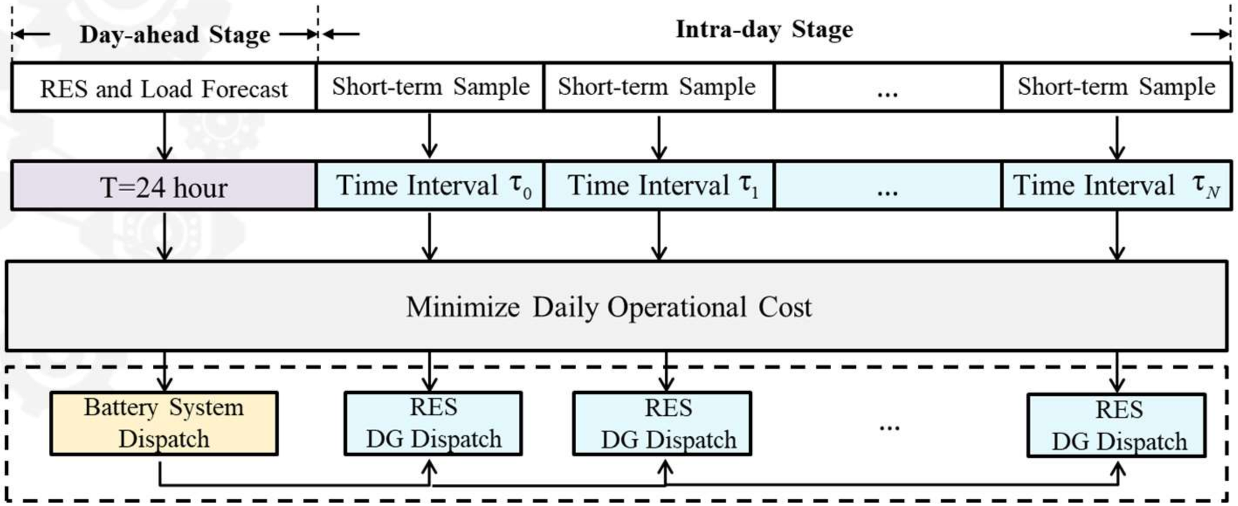
Power Balance

$$P_{\mathrm{w},t} + P_{\mathrm{pv},t} + P_{\mathrm{ESdch},t} - P_{\mathrm{ESch},t} + P_{\mathrm{DG},t} + \Delta P_{\mathrm{DG},t} + P_{\mathrm{G},t} = \hat{L}_t - L_{\mathrm{C},t}$$

2. Building a Optimization Model for MG

Two Stage Operational Model

- Step1: Optimize the energy storage dispatch plan
- Step2: Optimize the RES and DG power



Mathematical Formulation

$$\boldsymbol{z}^{*(\mathrm{i})} = \underset{\boldsymbol{z}^{(\mathrm{i})}, \boldsymbol{z}^{(\mathrm{ii})}}{\mathrm{armin}} F(\hat{\boldsymbol{y}}, \boldsymbol{z}^{(\mathrm{i})}, \boldsymbol{z}^{(\mathrm{ii})})$$

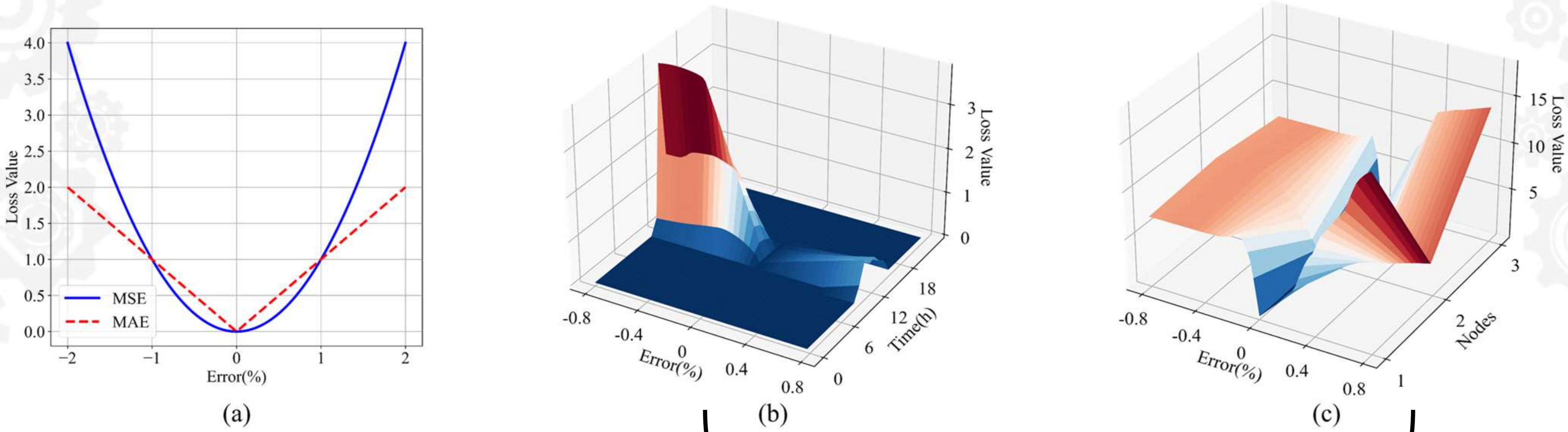
$$z_t^{(\mathrm{ii})} = \underset{z_t^{(\mathrm{ii})}}{\mathrm{argmin}} f(y_t, z_t^{*(\mathrm{i})}, z_t^{(\mathrm{ii})}) \quad \forall t = 0, \ldots, 23$$

$$F(\boldsymbol{y}, \boldsymbol{z}^{*(i)}(\hat{\boldsymbol{y}}), \boldsymbol{z}^{*(ii)})$$

Actual Operational Cost

Why End-to-End Learning? The Key Motivation

- Accurate forecasts do not necessarily lead to better scheduling benefits (lower operating costs)
- Statistical loss functions such as MSE and MAE **do not perfectly quantify the value** of prediction results for scheduling decisions
- The impact of prediction errors from devices at different times, in different directions, and at different nodes **is not equalized**

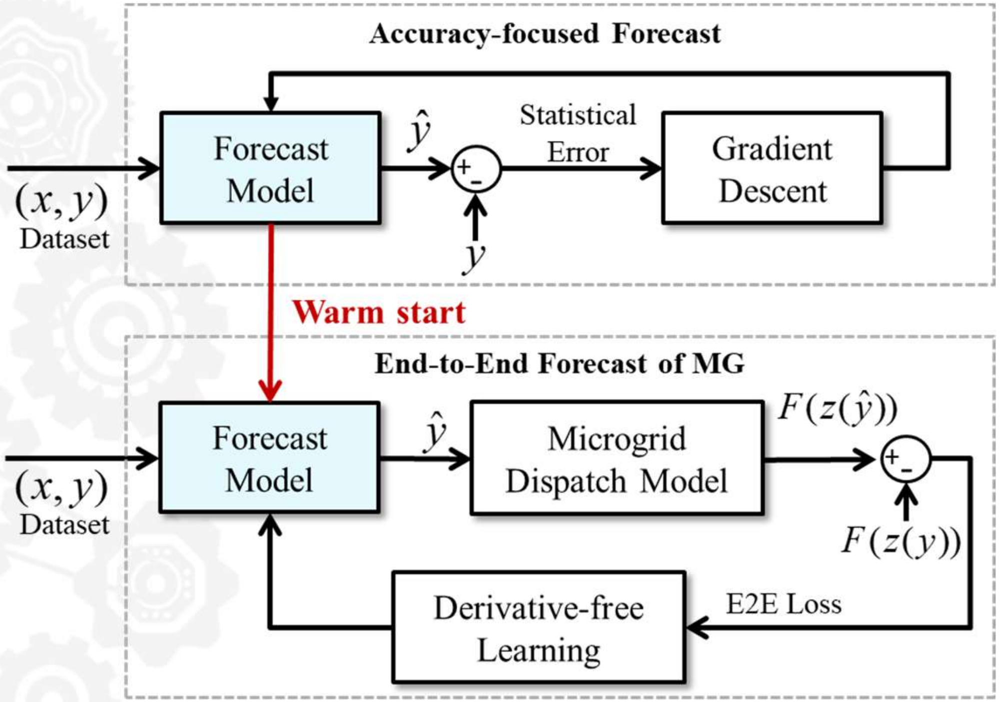


MSE Loss

Asymmetric Loss

What is End-to-End Learning?

iSPEC 2024 KUCHING



- Define the decision loss as:

$$L_{\text{decision}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = F(\hat{z}^{\text{DA}*}(\hat{\boldsymbol{y}}), \hat{z}^{\text{RT}*}(\hat{z}^{\text{DA}*})) - F(z^{\text{DA}*}(\boldsymbol{y}), z^{\text{RT}*}(z^{\text{DA}*}))$$

- where

$$z^{\text{DA}*}(\boldsymbol{y}) = \underset{\boldsymbol{z}^{\text{DA}}, \boldsymbol{z}^{\text{RT}} \in \Omega(\boldsymbol{y})}{\text{argmax}} \quad F(z^{\text{DA}}, z^{\text{RT}})$$

- End-to-End (E2E) Learning

$$\theta^* = \underset{\theta}{\text{argmin}} \mathbb{E}_{(x_i, y_i) \sim \mathbb{D}} L_{\text{decision}}(\boldsymbol{y}, \hat{\boldsymbol{y}})$$

- **Minimize the downstream decision loss directly**

Derivative-free End-to-End Learning

- Using a neural network (NN) as a prediction model, we consider backpropagation:

$$\frac{\partial L_{\text{decision}}}{\partial \theta} = \frac{\partial L_{\text{decision}}}{\partial F} \cdot \frac{\partial F}{\partial \hat{z}^{\mathrm{RT}*}} \cdot \boxed{\frac{\partial \hat{z}^{\mathrm{RT}*}}{\partial \hat{z}^{\mathrm{DA}*}} \cdot \frac{\partial \hat{z}^{\mathrm{DA}*}}{\partial \hat{y}}} \cdot \frac{\partial \hat{y}}{\partial \theta}$$

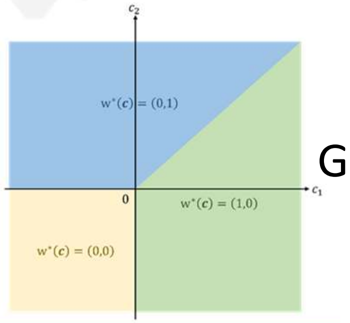
✓ Equal to 1

✓ Cost Function

?

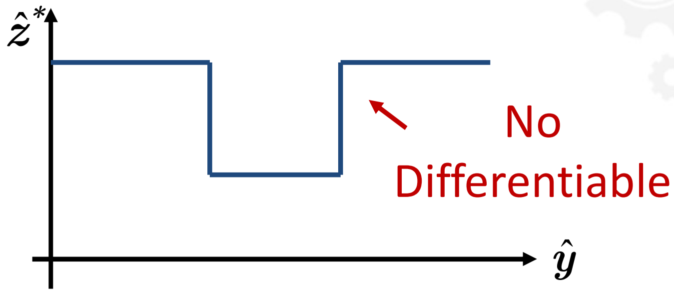
✓ Computational Graph in NN

- Linear Programming (LP):



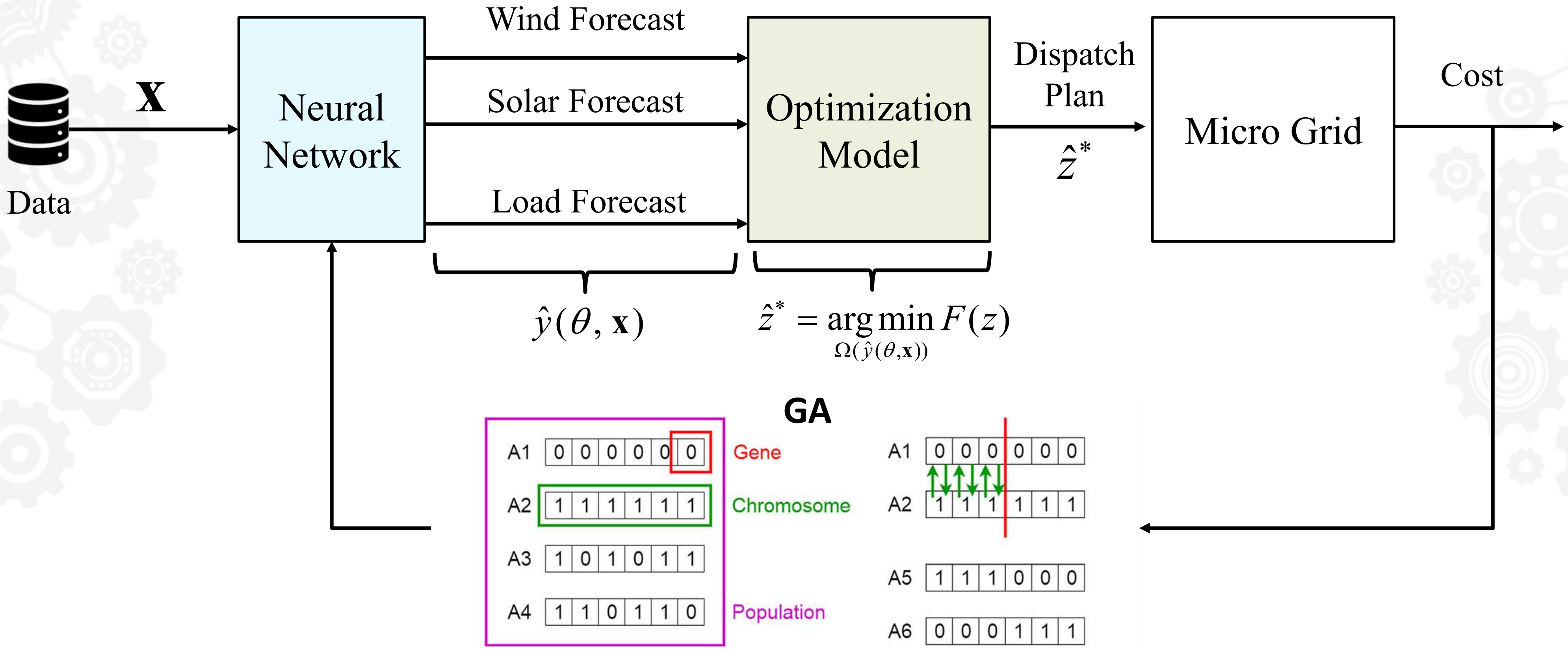
Gradient= **0**

- Mixed Integer Programming (MIP):



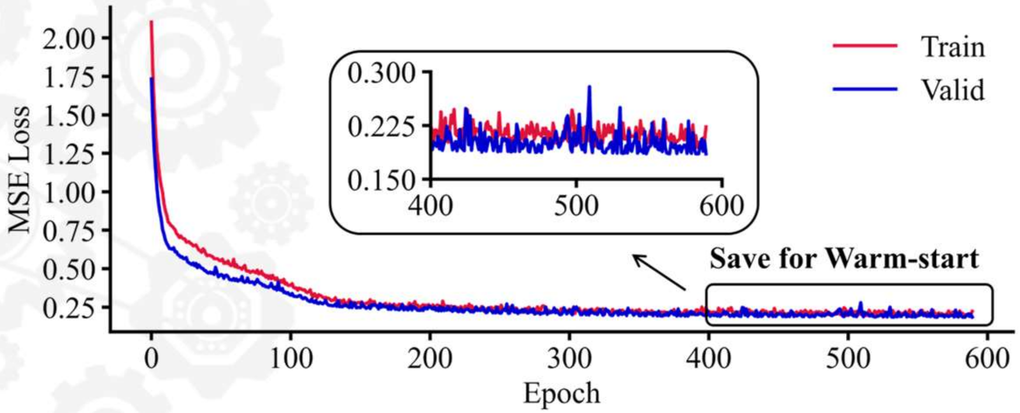
Derivative-free End-to-End Learning

- Use Genetic Algorithm (GA) to optimize NN parameters directly from the operating costs



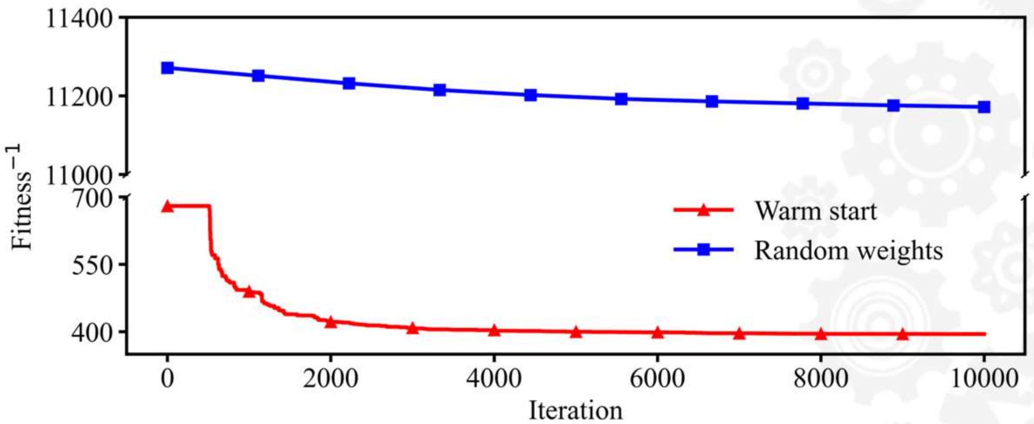
Case Study

1. Pre-train



Loss curve of pre-trained accuracy-focused model.

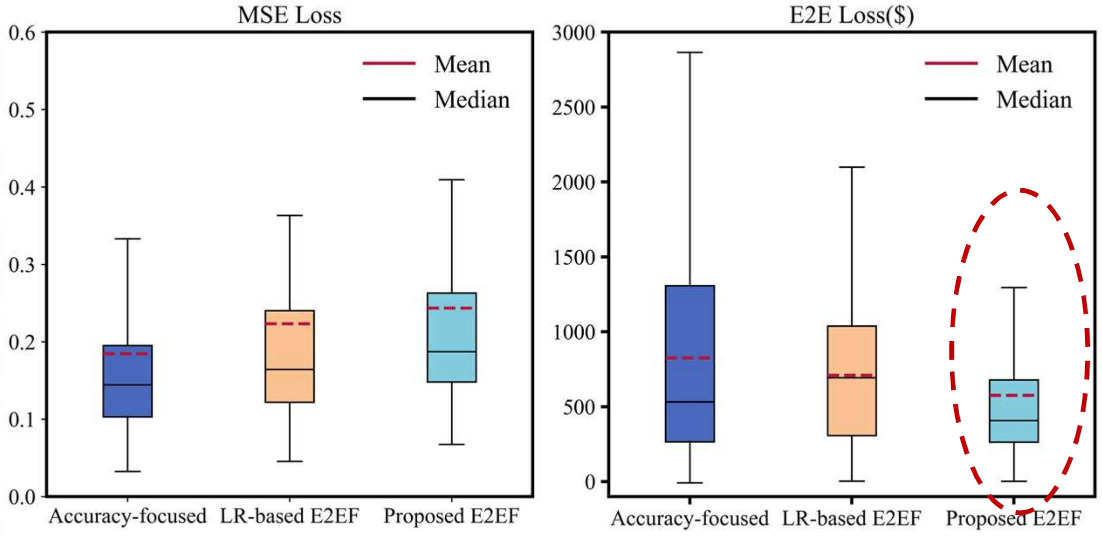
2. E2E Learning



Fitness curve of warm-start GA-DNN compared with random weights start.

Case Study

- Higher MSE Loss, but **lower costs**



Distribution of MSE and E2E Loss of existing method in common test dataset.

TABLE I
COMPARISON OF EXISTING FORECAST METHOD IN TEST DATASET

	Accuracy-focused	LR-based E2EF	Proposed E2EF
MSE	0.185	0.223	0.244
MAE	0.292	0.337	0.357
Operational Cost($)	8049.75	7934.34	7800.04
E2E Loss($)	824.77	709.36	575.07
Cost Save(%)	/	1.43	3.10

Conclusion

iSPEC 2024 KUCHING

- A genetic algorithm (GA) employed as a derivative-free technique for end-to-end learning

  To averts the negative impacts of steep gradients that commonly impede E2E optimization

- A pre-trained accuracy-focused model for warm-start.

  To ensures GA improves upon the benchmark set by the well-trained accuracy-focused model, minimizing the inefficiencies of heuristic approaches.

- Case studies reveal that the proposed E2EF methodology markedly reduces the operational costs of MG compared to conventional accuracy-focused forecast model, while also outperforming existing linear E2E approaches.

Thank you for listening !

Chuanqing Pu
College of Smart Energy, Shanghai Jiao Tong University
sashabanks@sjtu.edu.cn