“I do not know”: Quantifying Uncertainty in Neural Network Based Approaches for Non-Intrusive Load Monitoring

Vibhuti Bansal*
Bharati Vidyapeeth’s College of Engineering, India

Rohit Khoiwal*
Rajasthan Technical University
India

Hetvi Shastri*
UMass Amherst
USA

Haikoo Khandor
IIT Gandhinagar
India

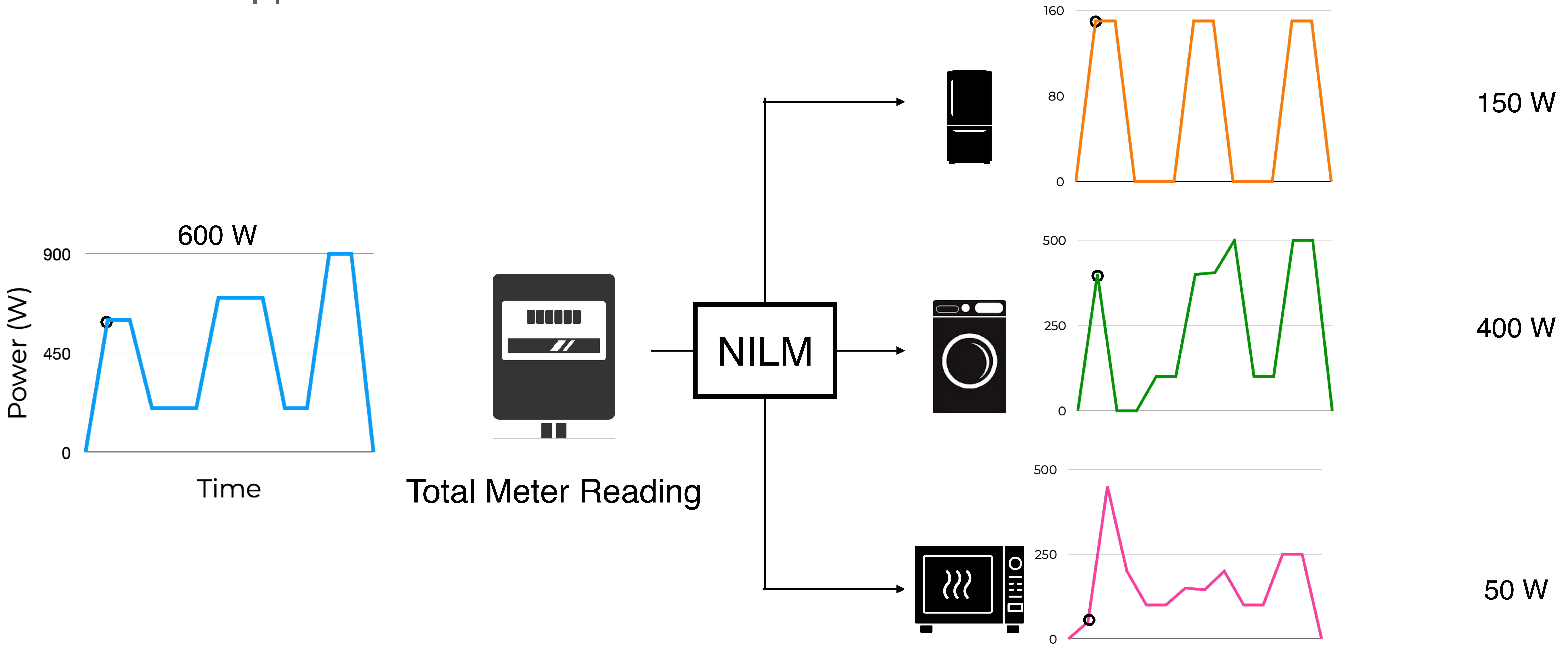
Nipun Batra
IIT Gandhinagar
India

Buildsys 2022, Boston, MA

* refers to equal contribution

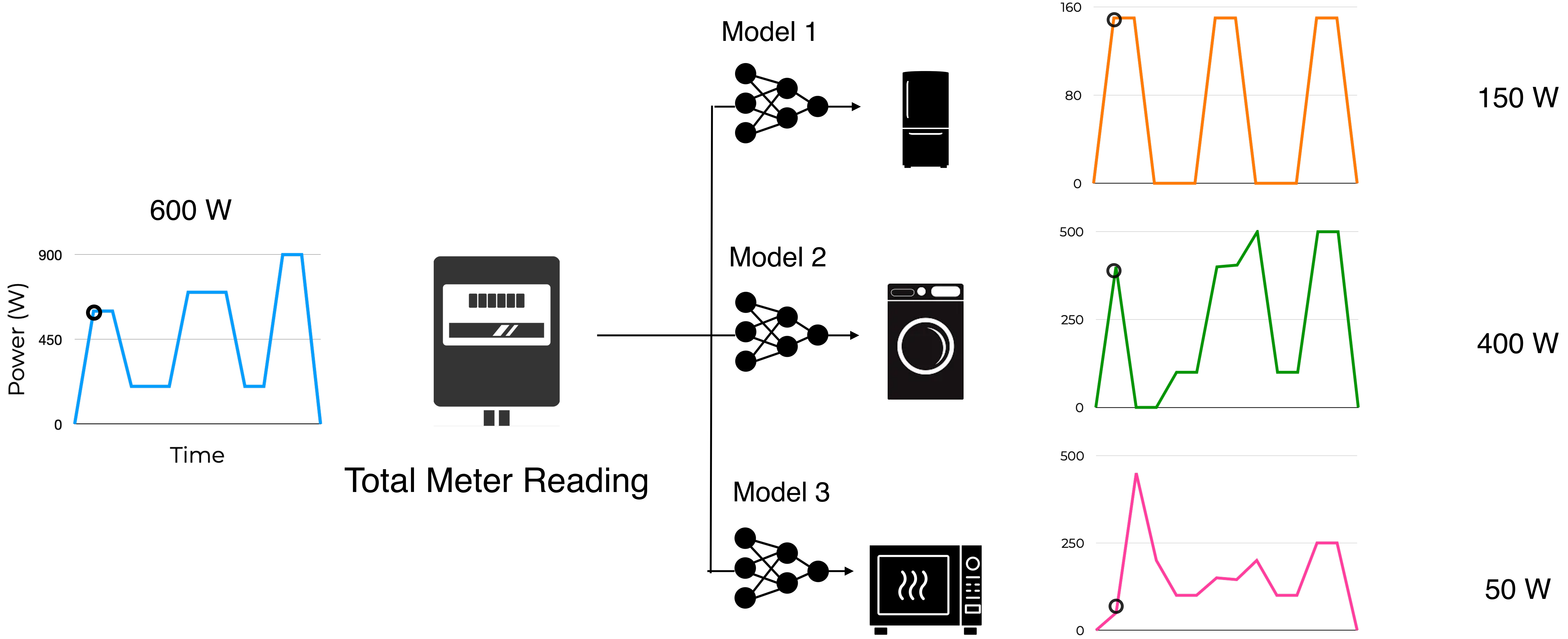
Non Intrusive Load Monitoring (NILM)

- Non intrusive load monitoring is a task of disaggregating mains reading into its constituent appliances.

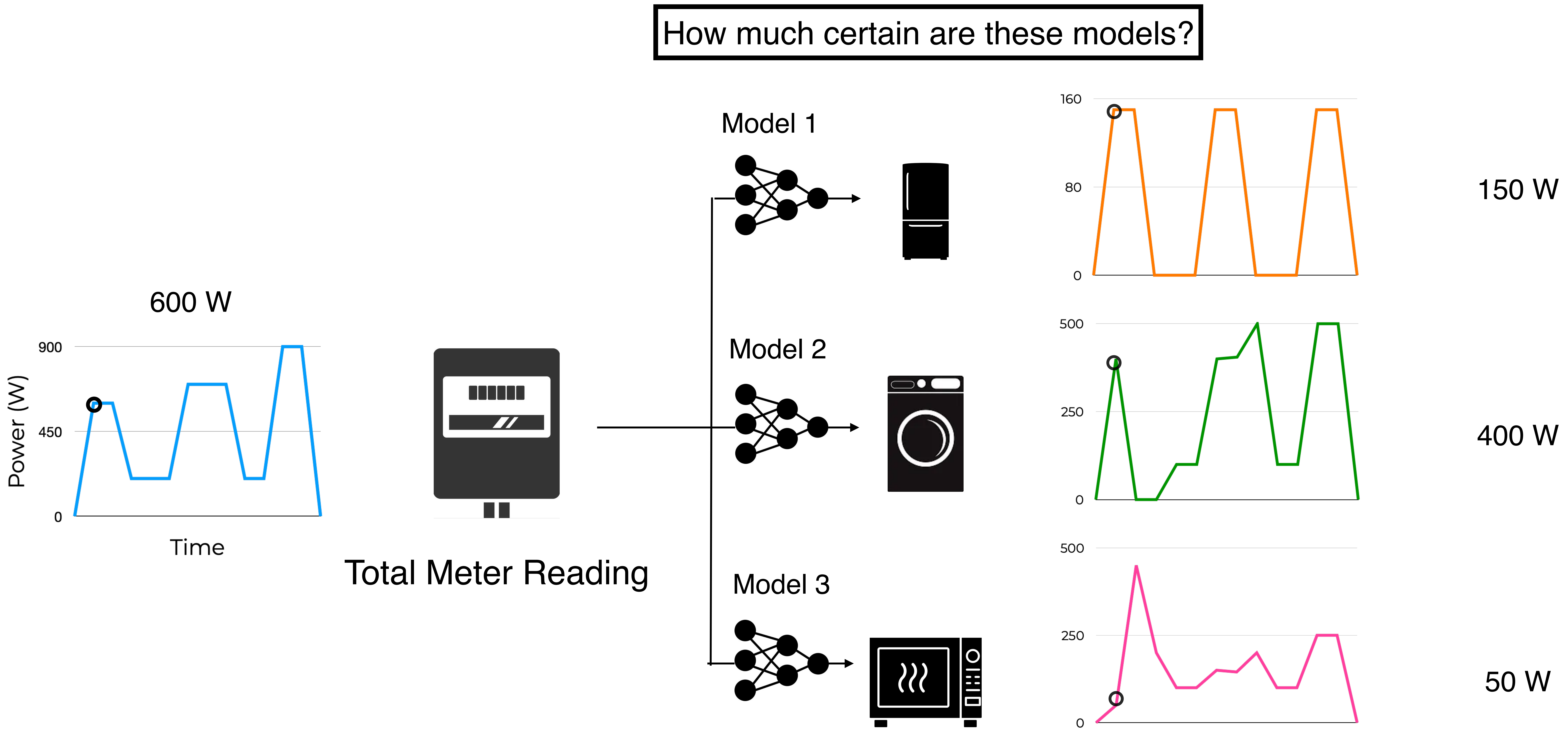


Motivation

- Normally, conventional neural networks for NILM give provide a point estimate for the appliance reading.

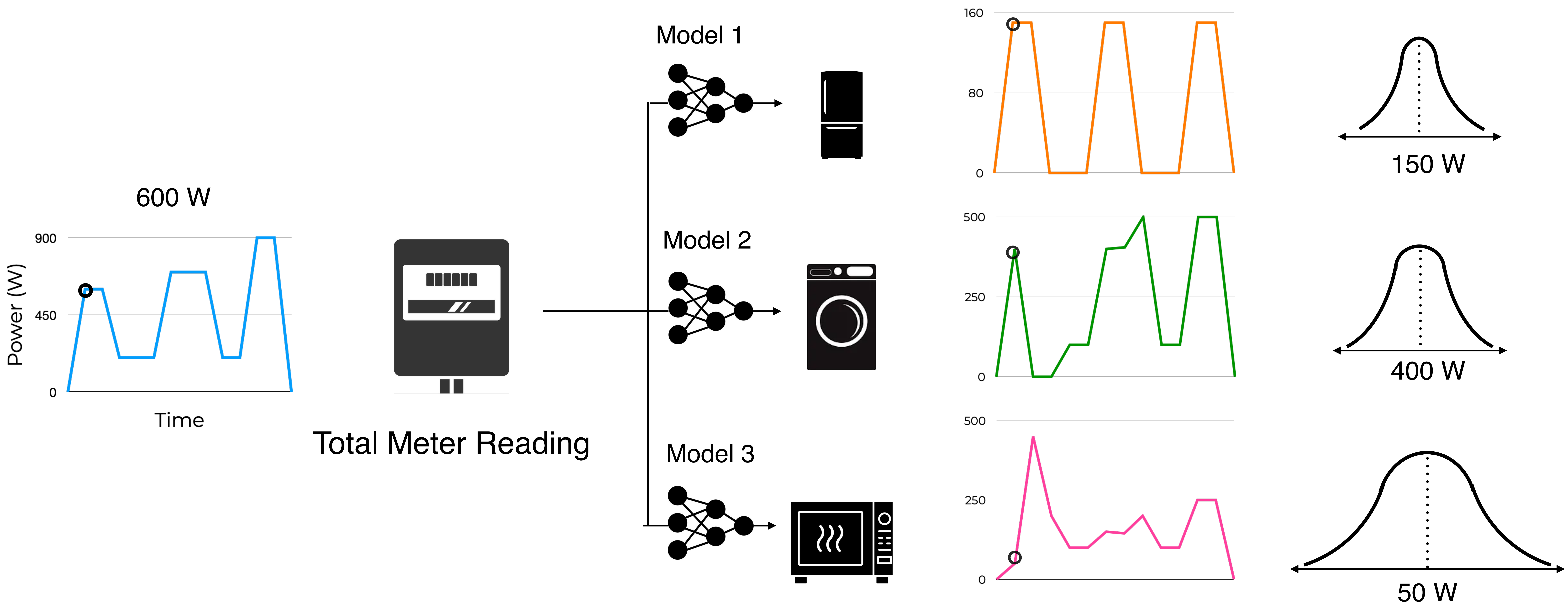


Motivation


How much certain are these models?
600 W
900
450
0
Power (W)
Time
Total Meter Reading
Model 1
Model 2
Model 3
160
80
0
150 W
500
250
0
400 W
500
250
0
50 W

Problem Statement

- To estimate the value of uncertainty along with model prediction.



Aleatoric (Data) Uncertainty

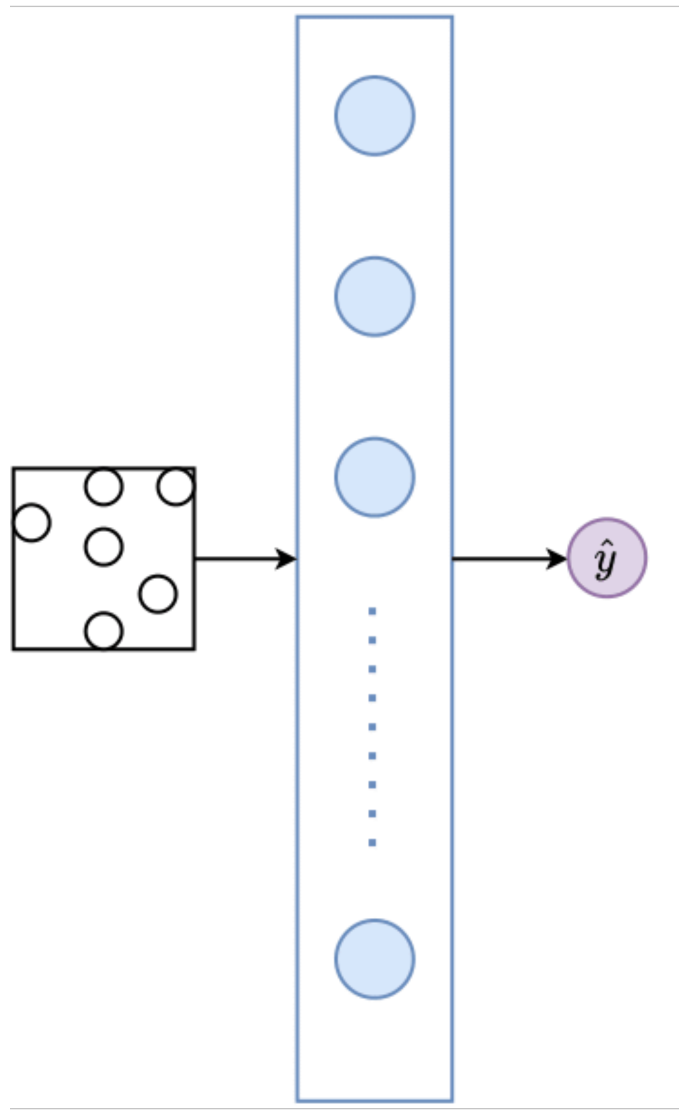
- Uncertainty caused due missing data, imprecise tools used to capture data.

Aleatoric (Data) Uncertainty

- Uncertainty caused due missing data, imprecise tools used to capture data.

Homoskedastic Regression

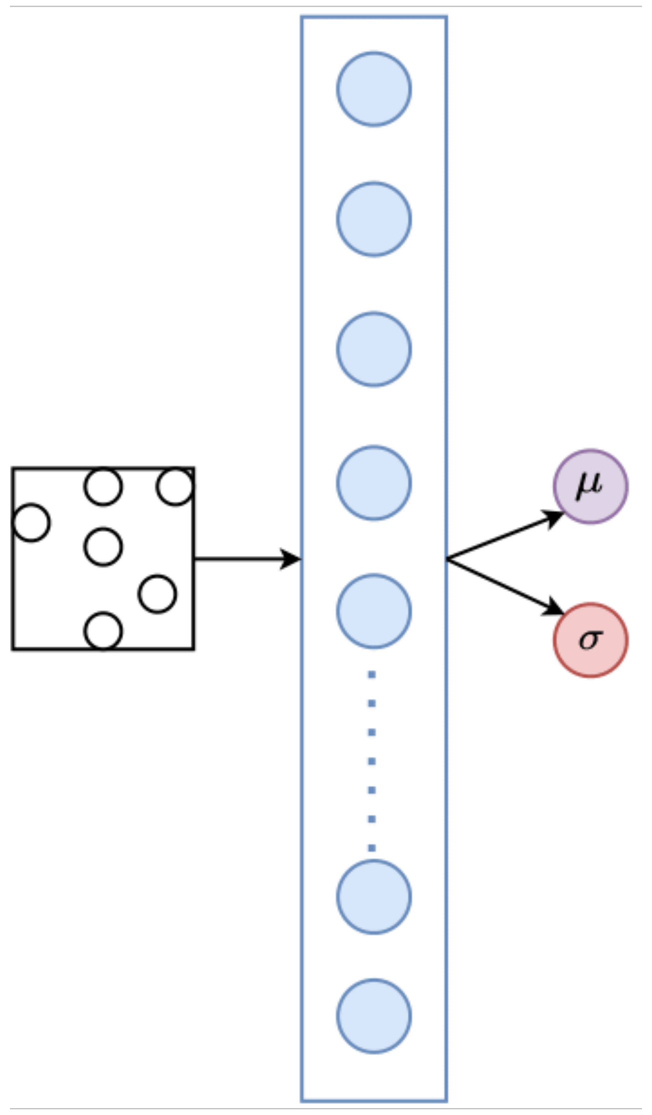
- Estimates constant variance distribution across entire dataset.



MLP

Heteroskedastic Regression

- Estimates of different variance for different datapoint.



Heteroskedastic MLP

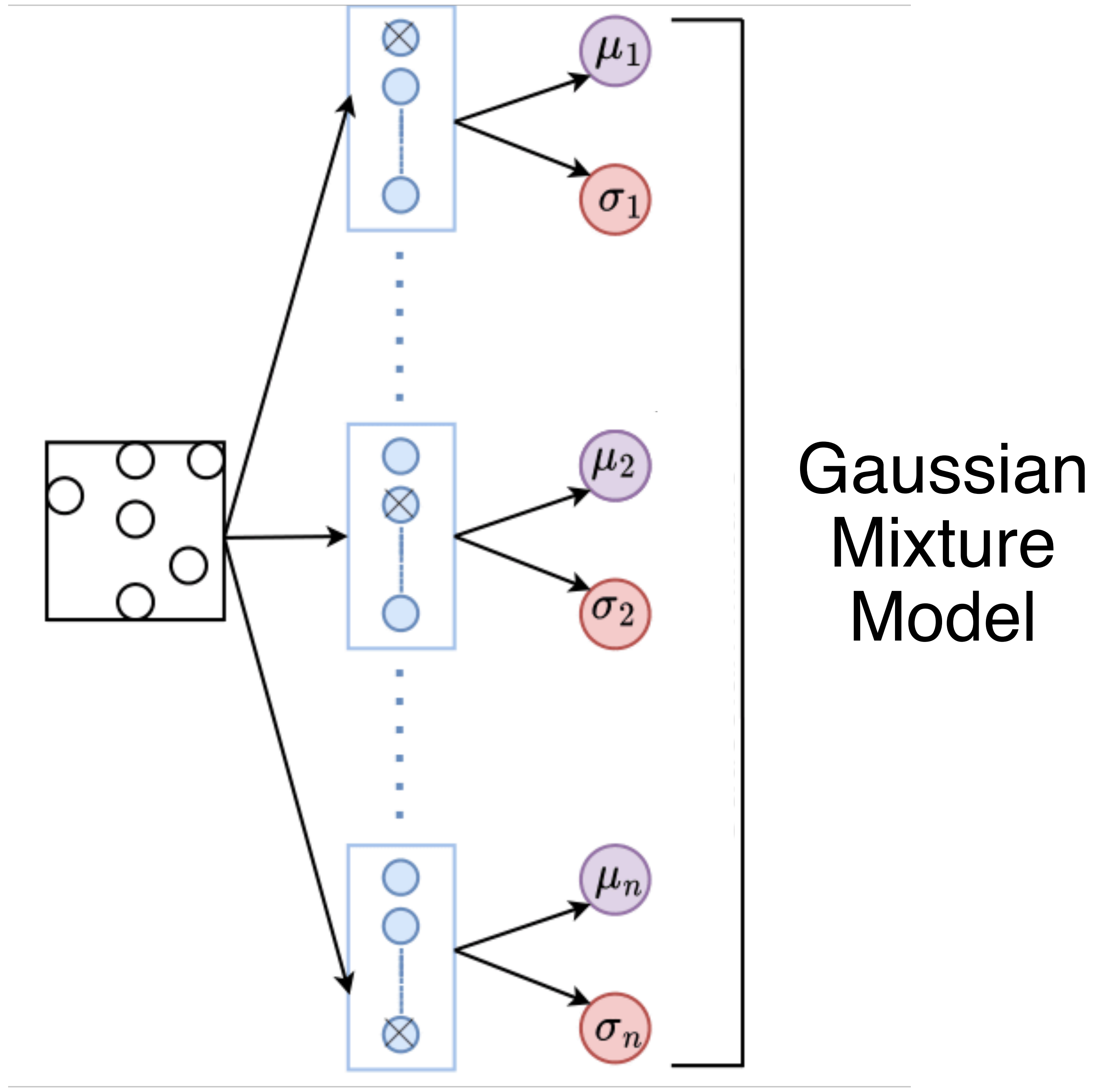
Epistemic (Model) Uncertainty

- Uncertainty caused due to limited amount of data and knowledge.

Epistemic (Model) Uncertainty

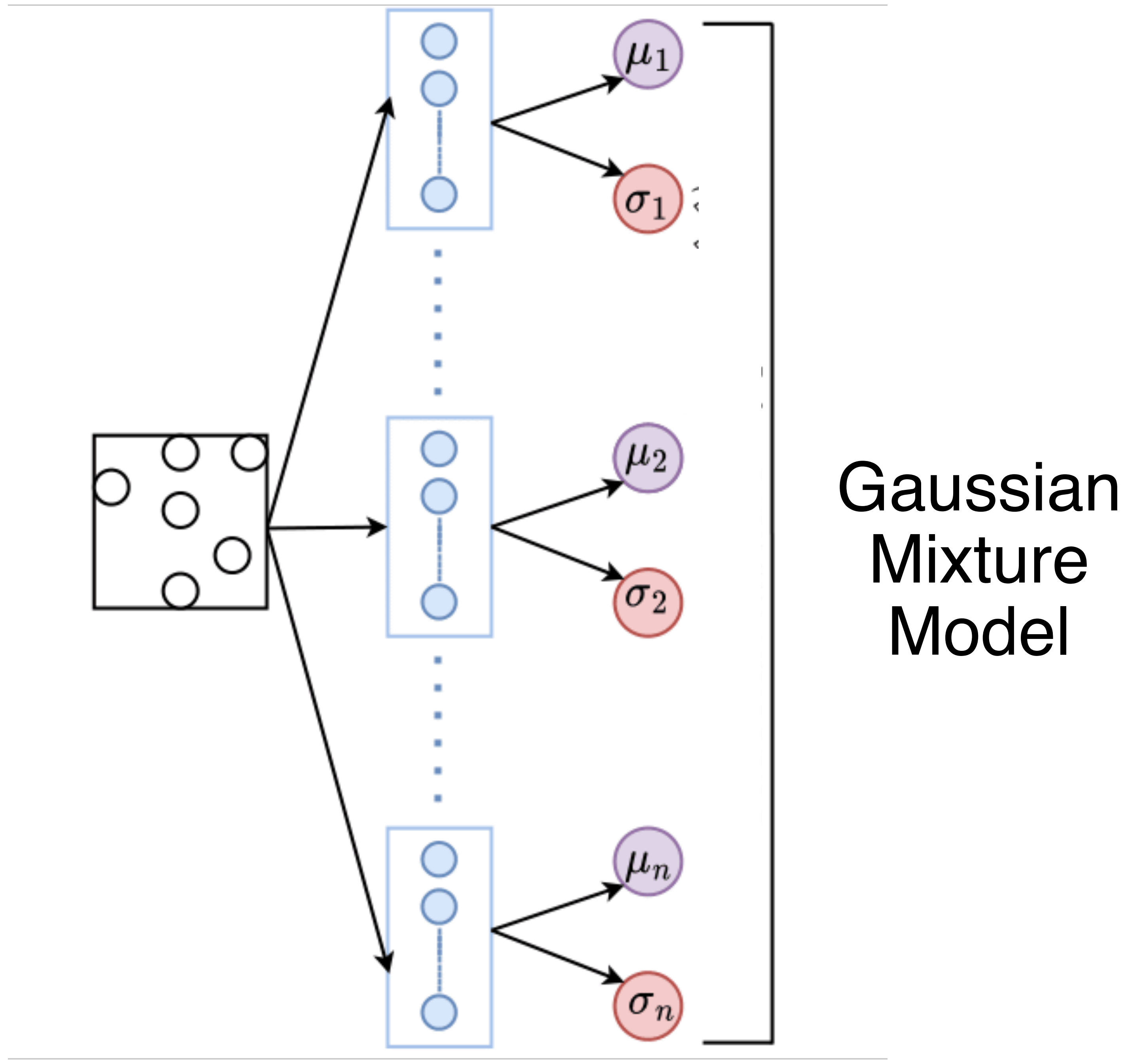
- Uncertainty caused due to limited amount of data and knowledge.

Monte Carlo dropout (MC)



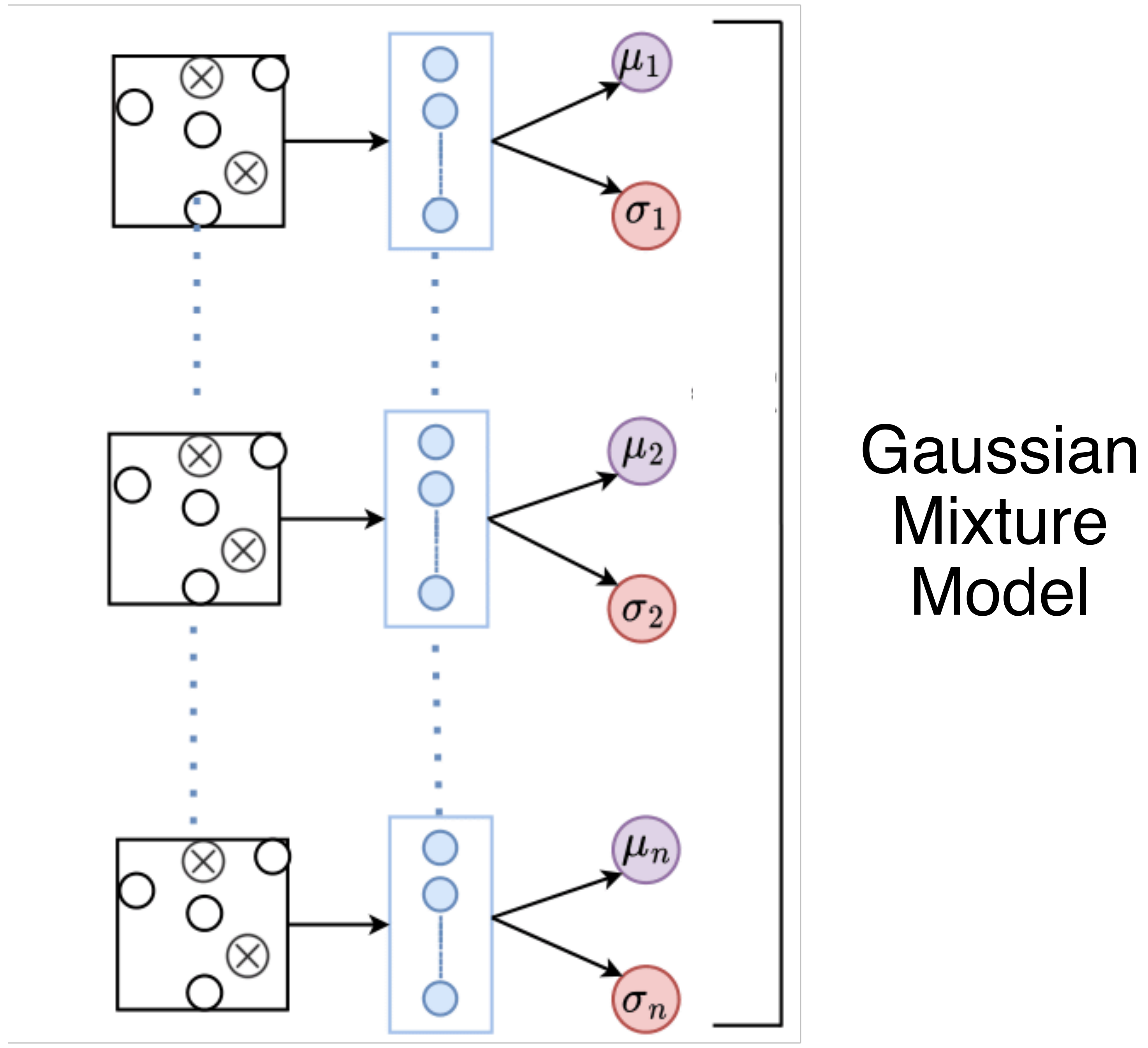
MC dropout on Heteroskedastic Model

Deep Ensemble (DE)



Deep Ensemble on Heteroskedastic Model

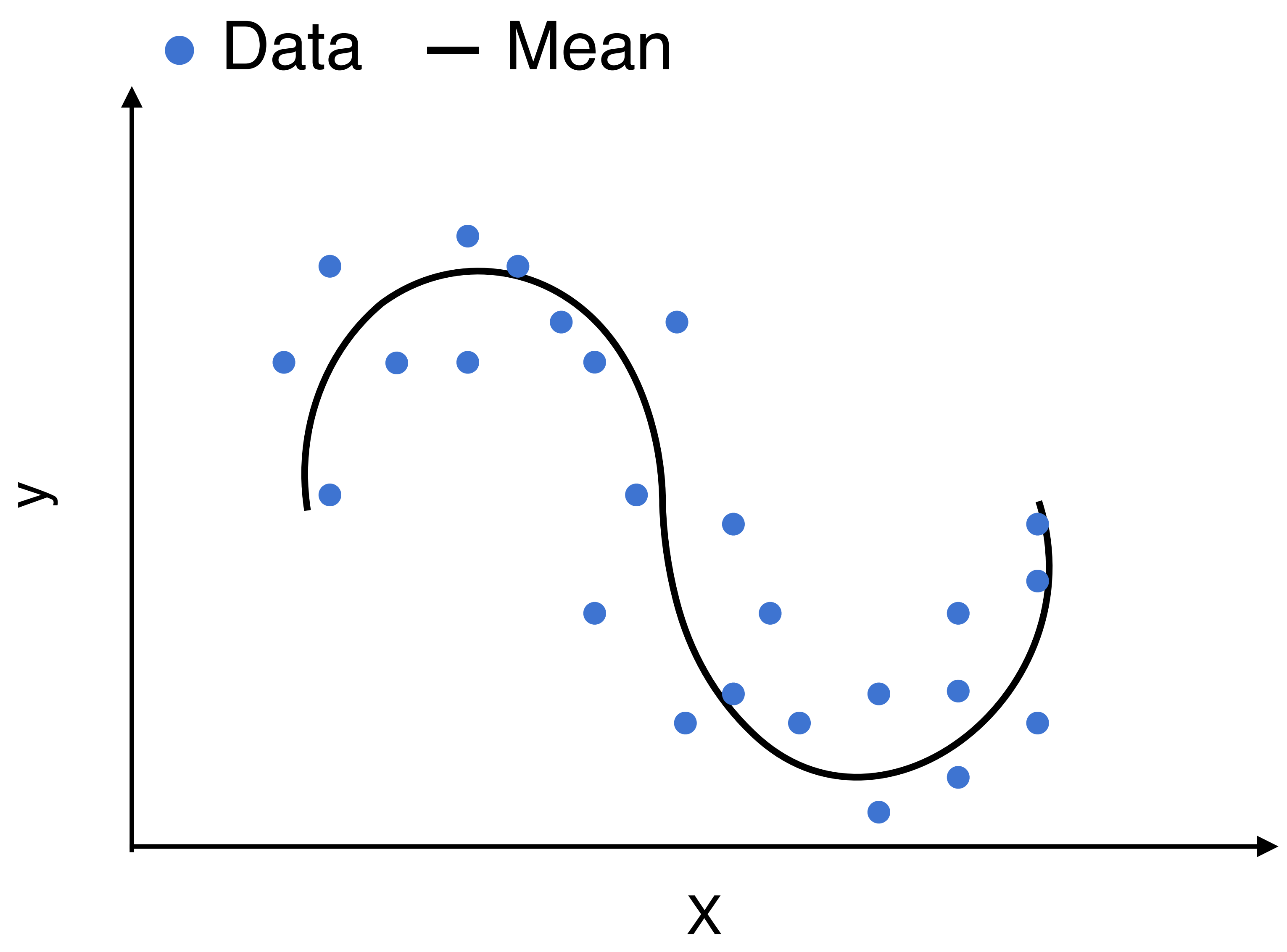
Bootstrap (BS)



Bootstrap on Heteroskedastic Model

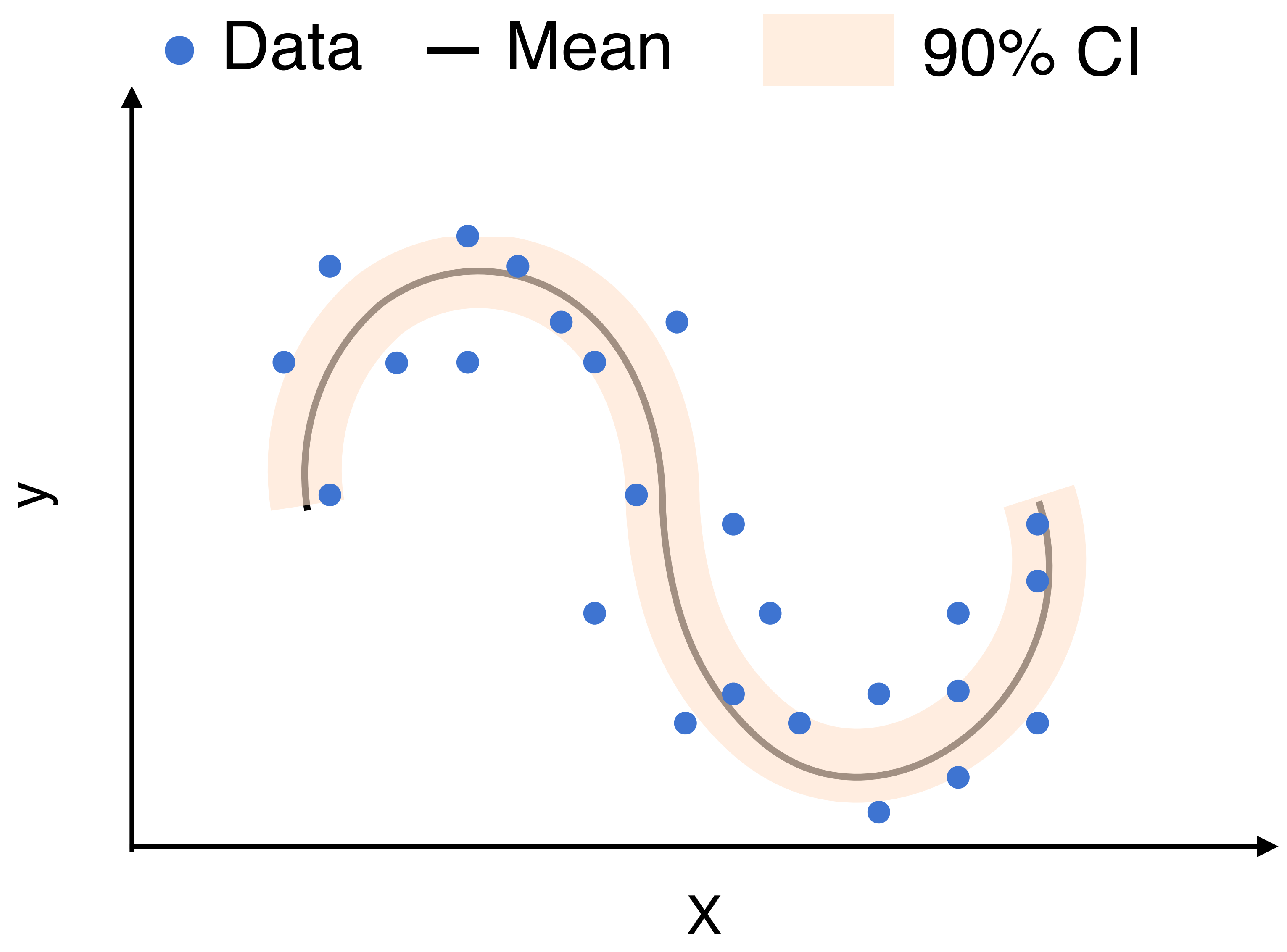
Quantifying predictive uncertainty

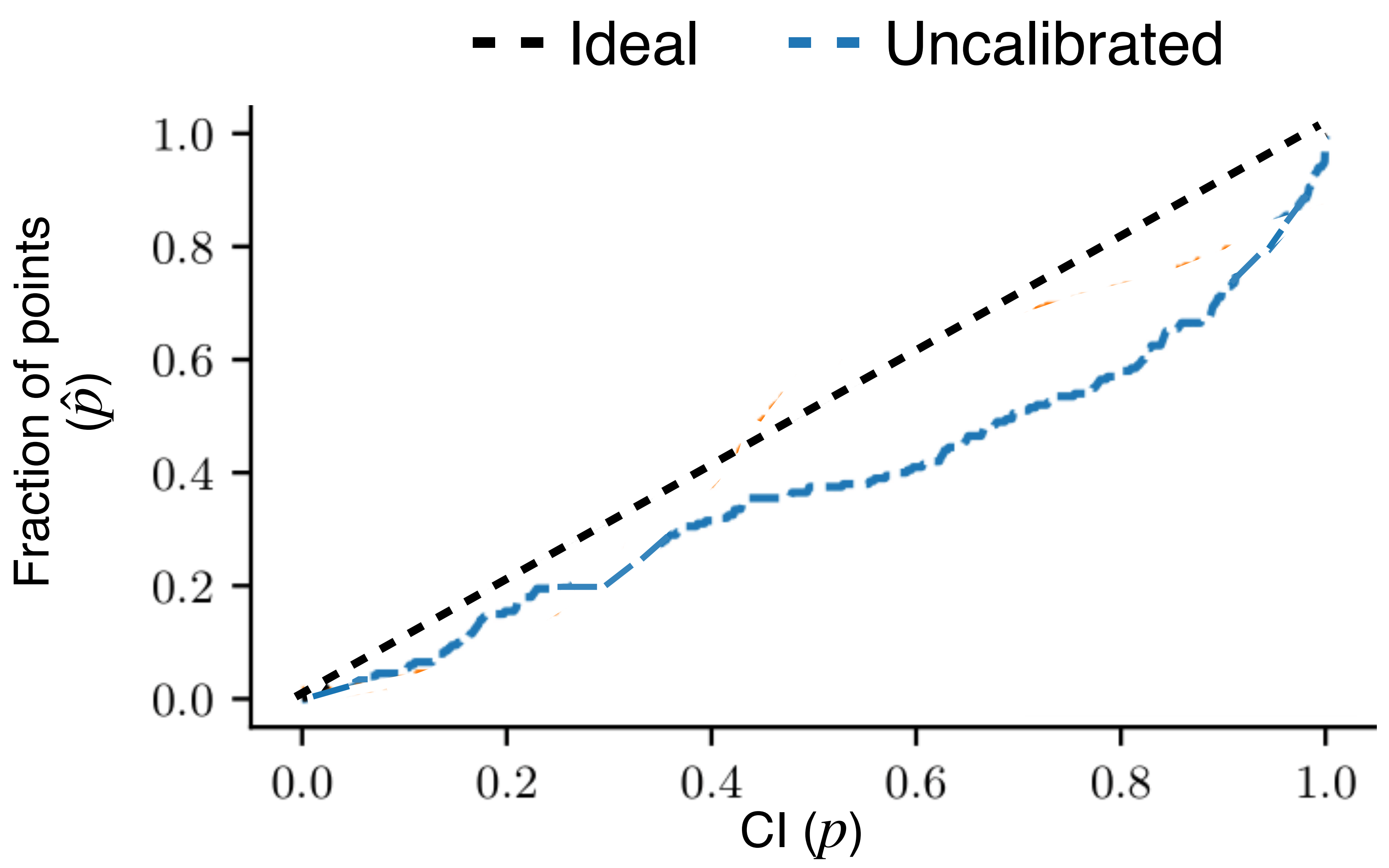
- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.



Quantifying predictive uncertainty

- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.



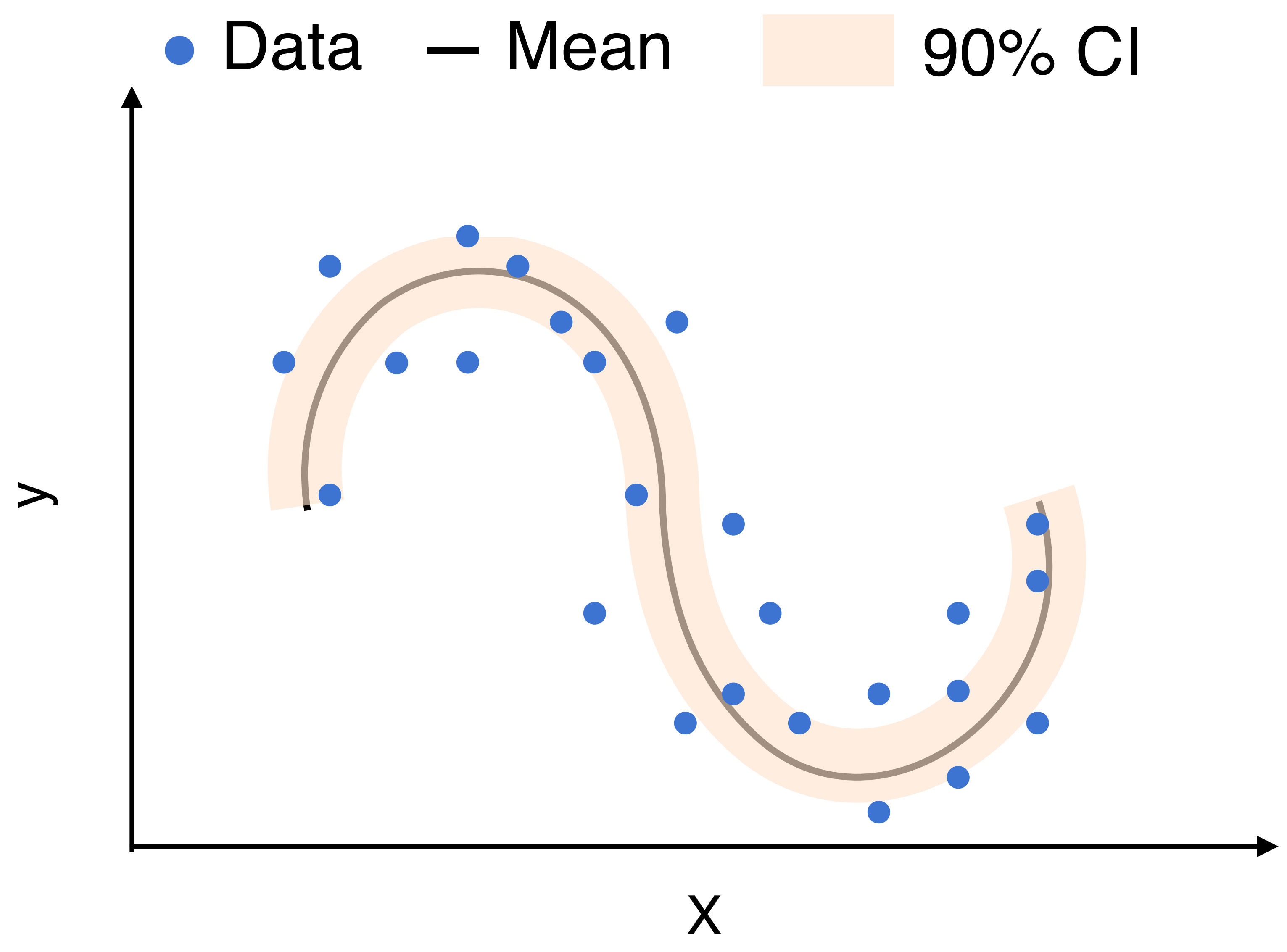


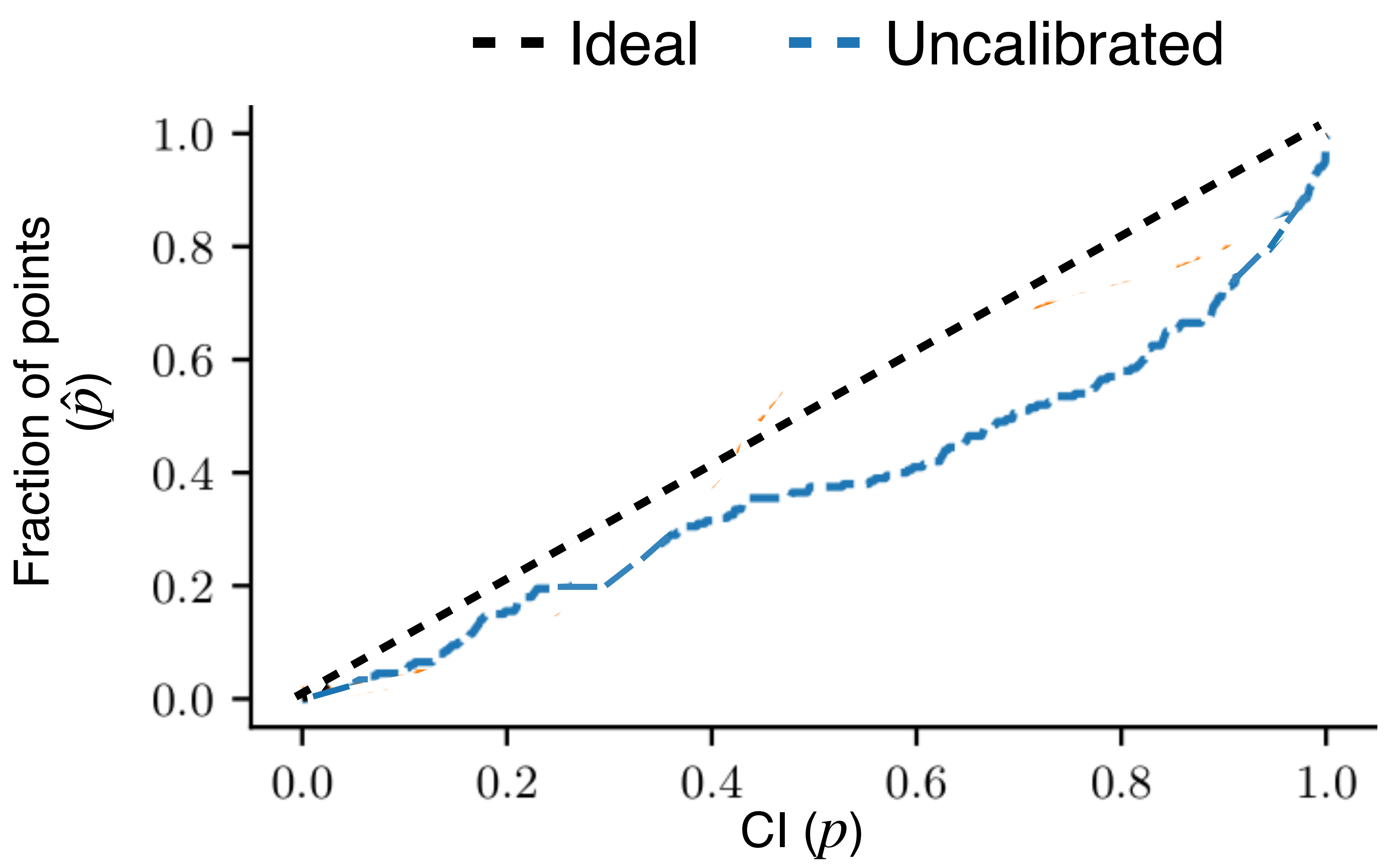
Reliability Plot

Quantifying predictive uncertainty

- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.

$$\text{Expected Calibration Error} \quad (\text{ECE}) = \frac{\sum_{p=0.1}^{P} |\hat{p}(p) - p|}{P}$$

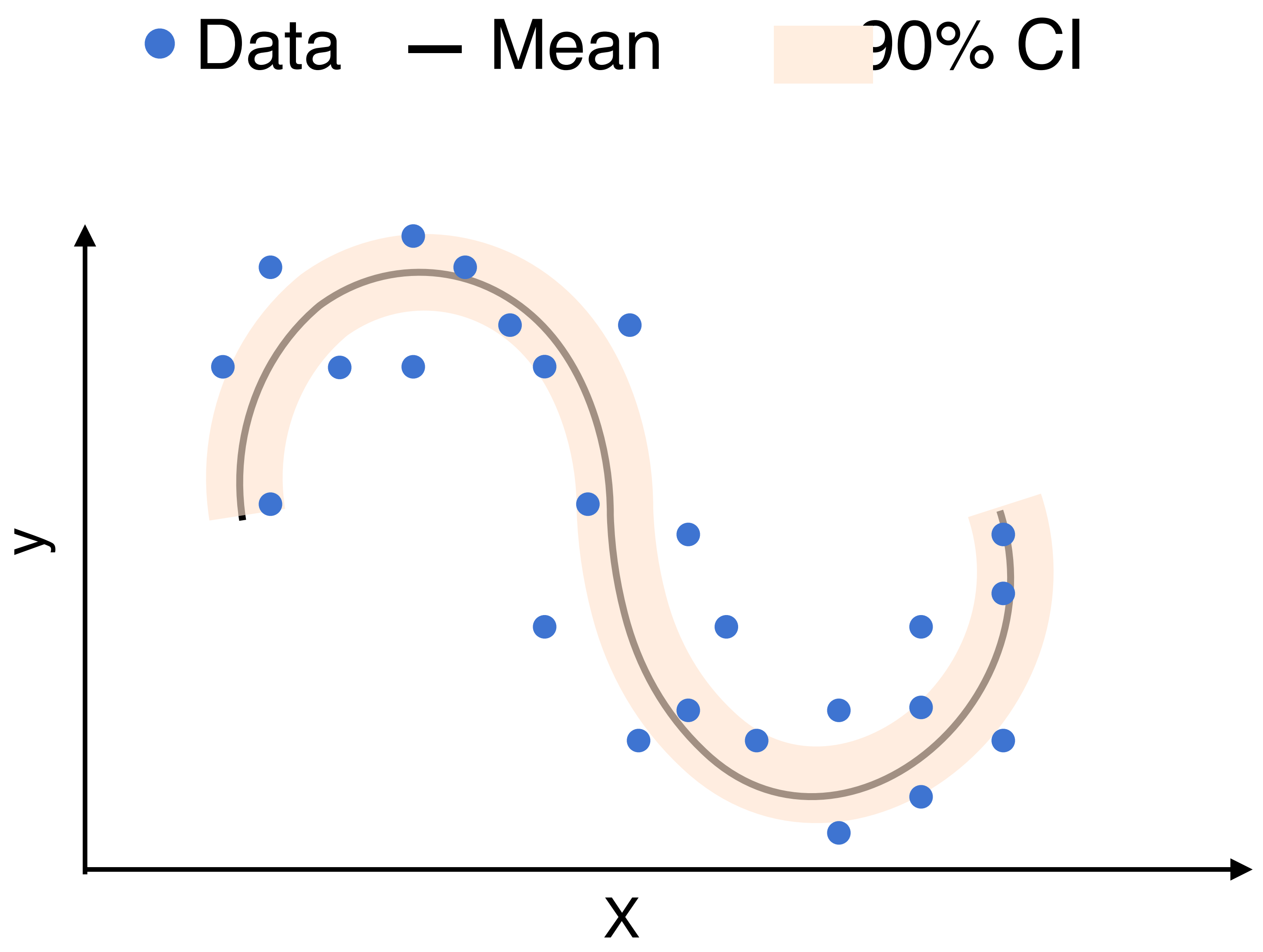




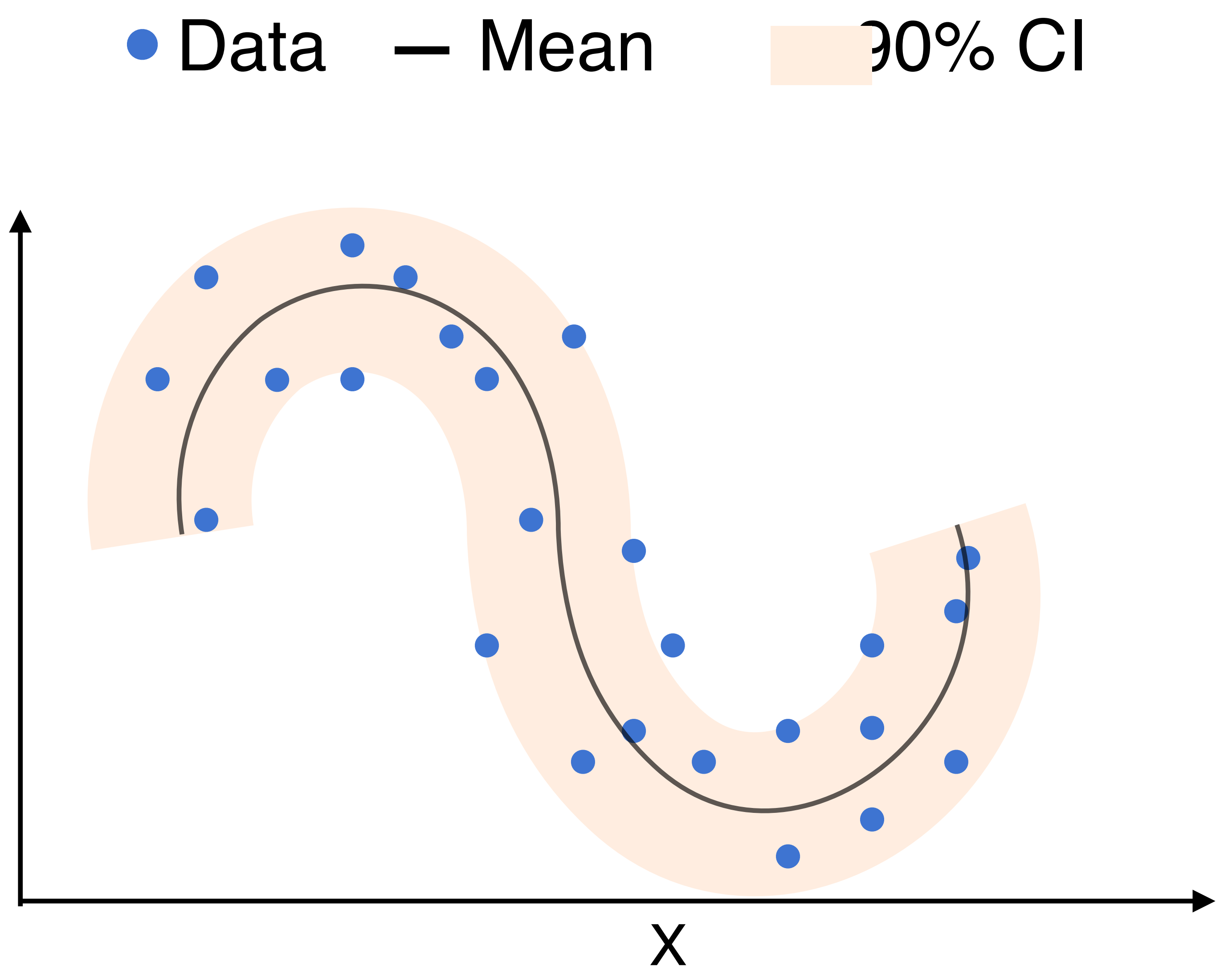
Reliability Plot

Model Recalibration

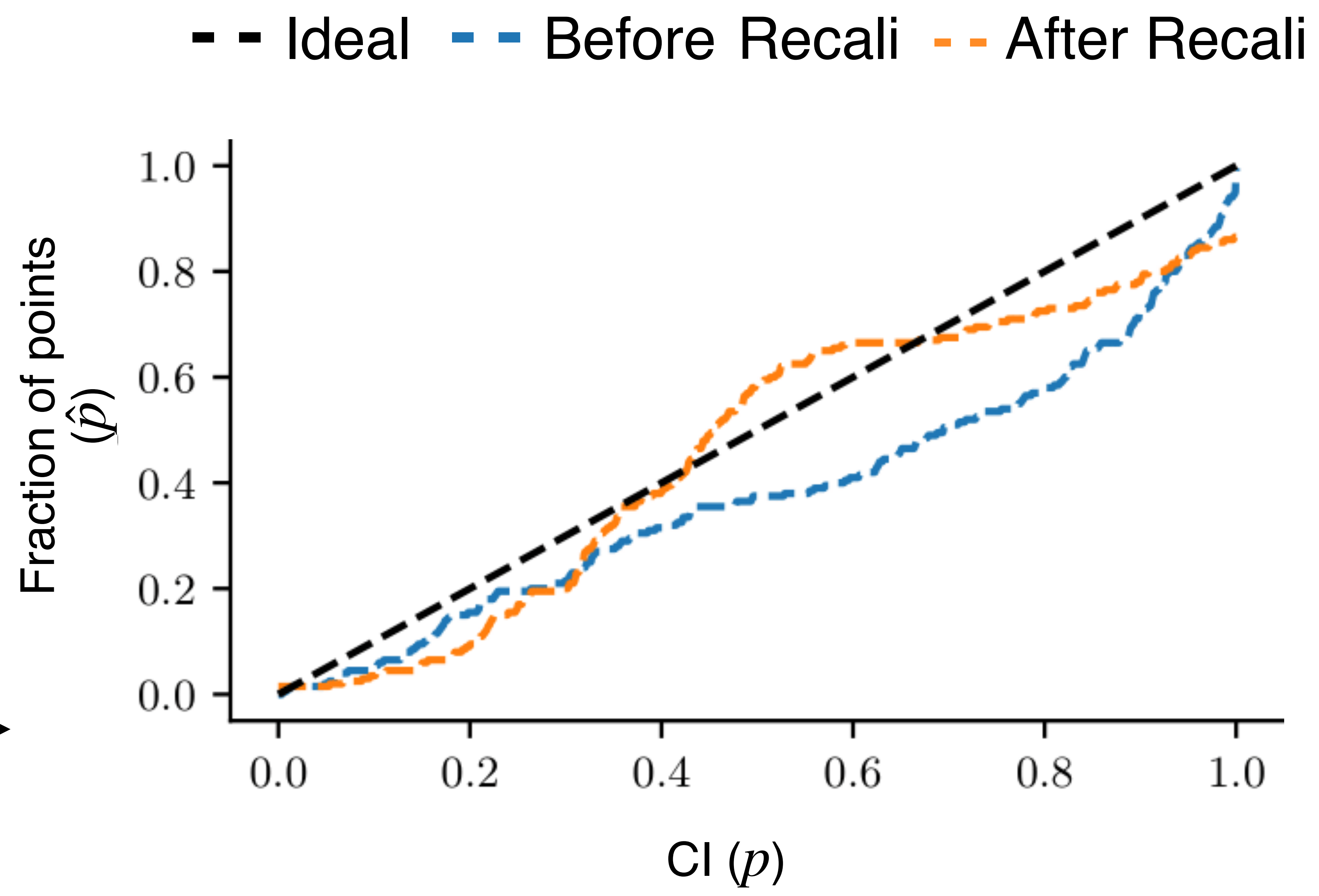
- Increase or decrease the band width of the confidence interval to cover same number empirically found points as confidence interval.



(a). Before Recalibration



(b). After Recalibration

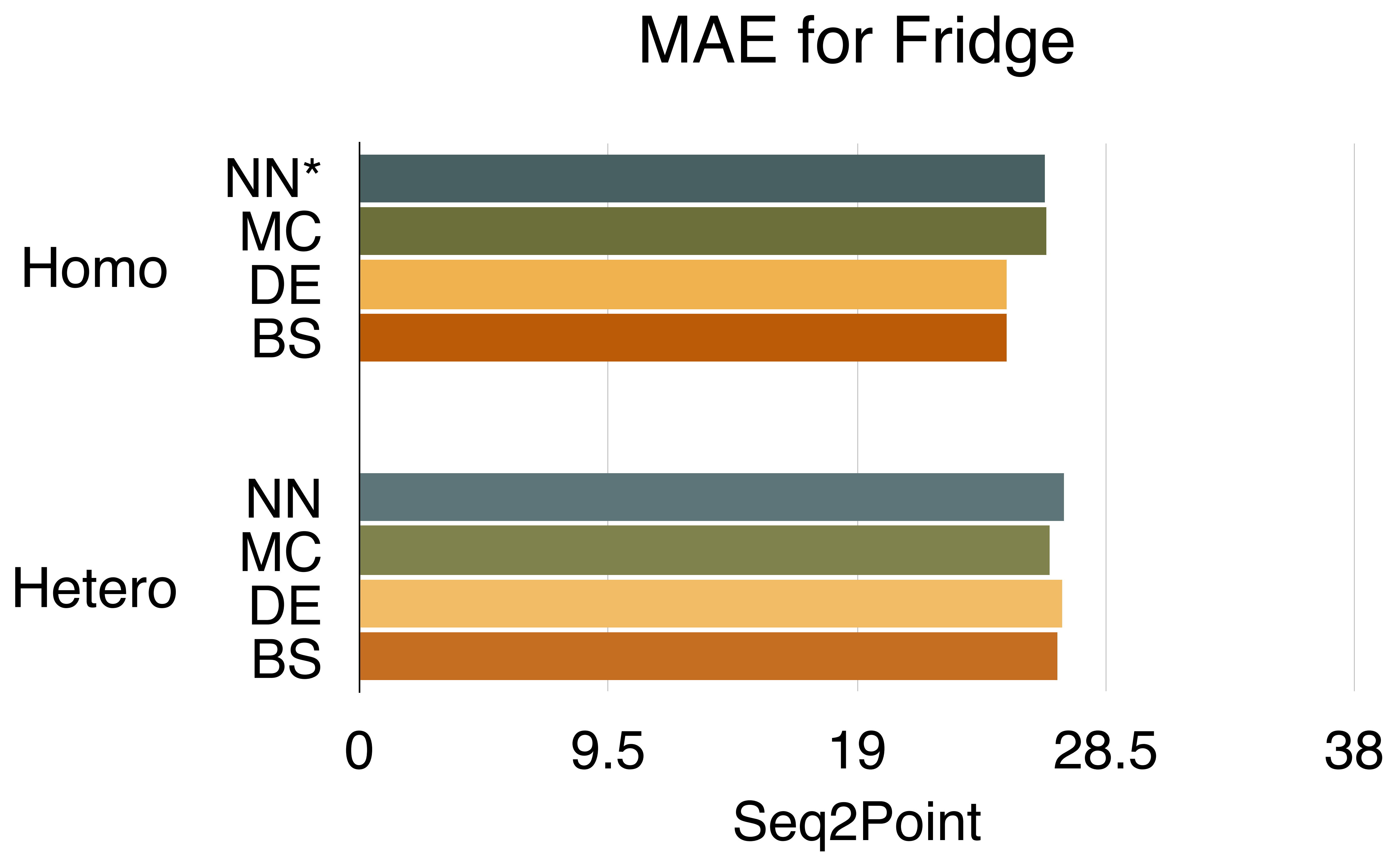


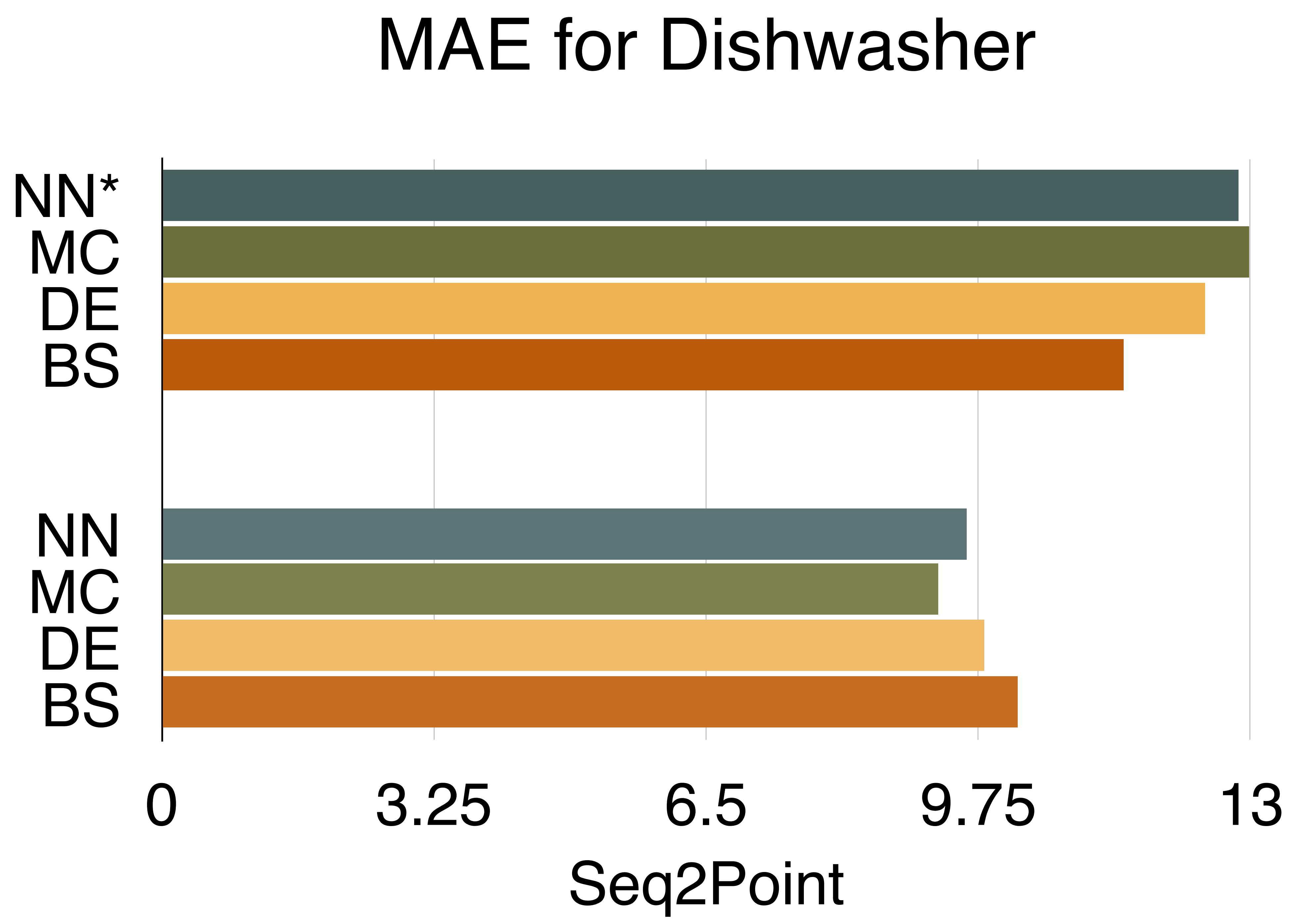
Evaluation

Dataset	ReDD
Appliances	Fridge, Dishwasher, Microwave
Metric	Mean absolute error (MAE), Expected calibration Error (ECE)
Neural Network Models	Seq2point, Bilstm with Attention

Do NNs with uncertainty achieve comparable error on conventional metrics to the baselines?

- Comparable or better performance on MAE while additionally incorporating the notion of uncertainty.

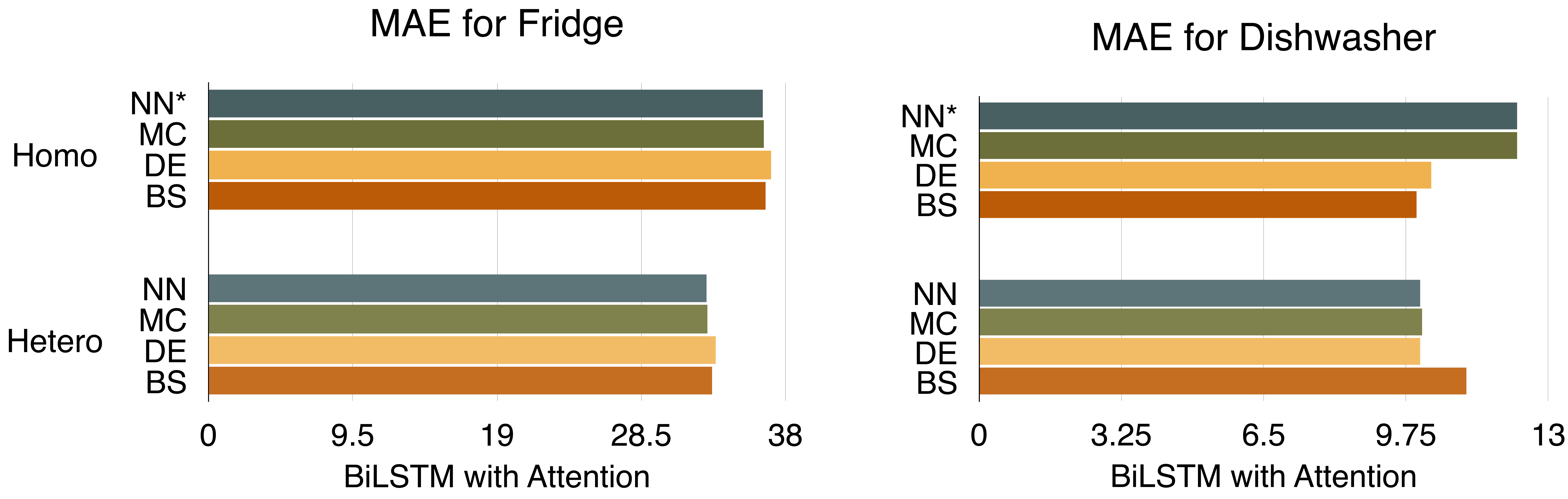




* refers to baseline

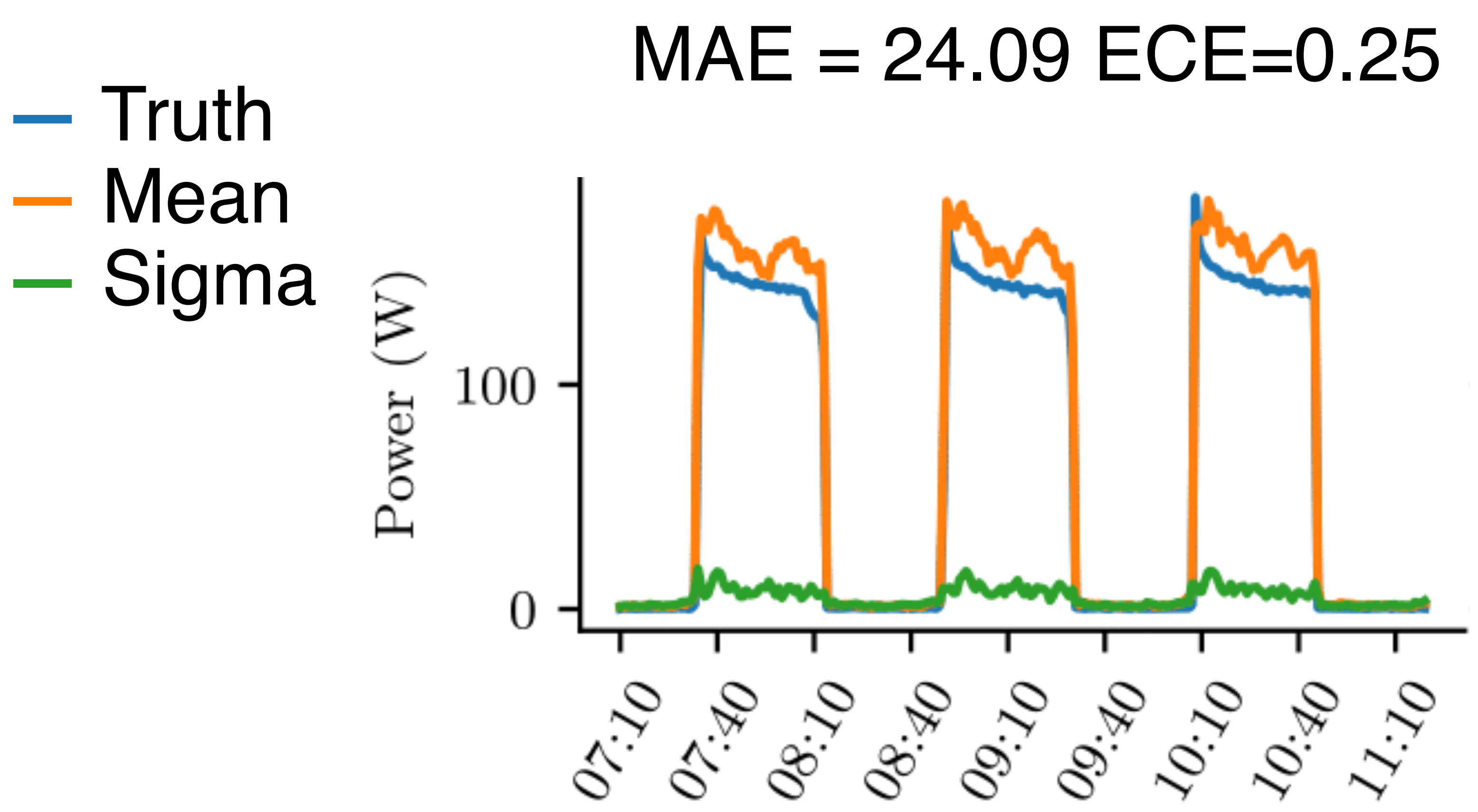
Do NNs with uncertainty achieve comparable error on conventional metrics to the baselines?

- Comparable or better performance on MAE while additionally incorporating the notion of uncertainty.

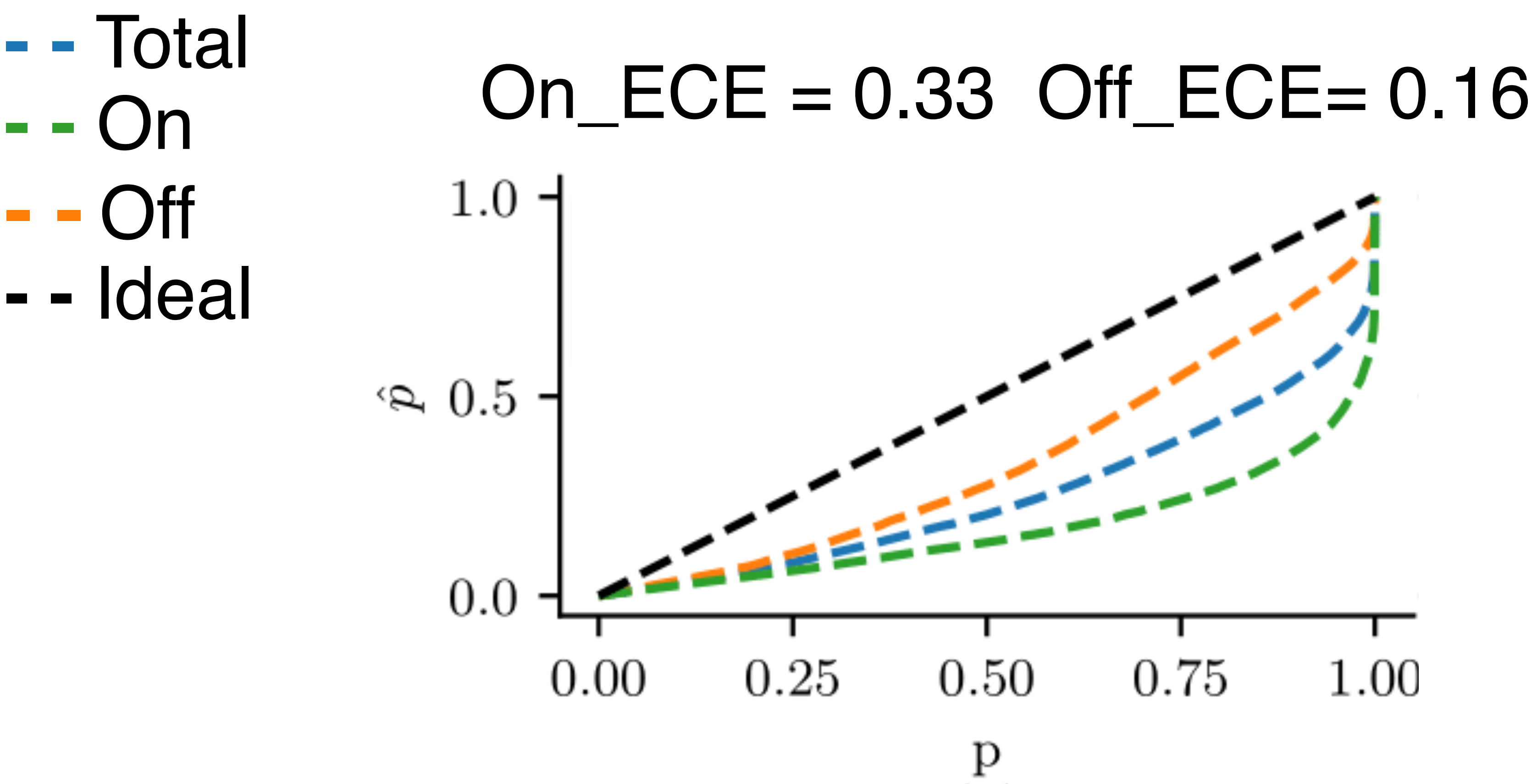


* refers to baseline

Do we need to do analysis of statewise (On and Off) calibration?

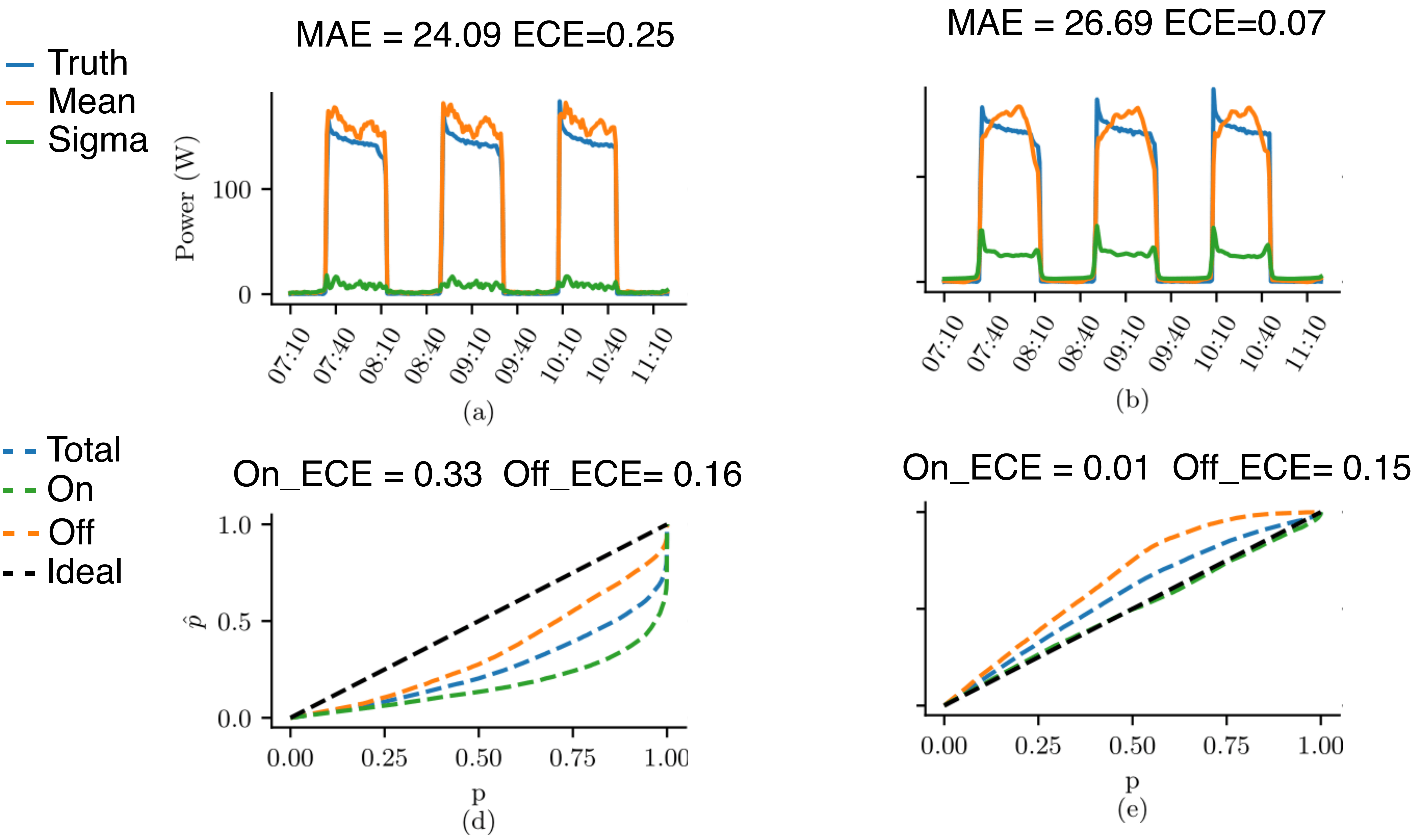


(a)

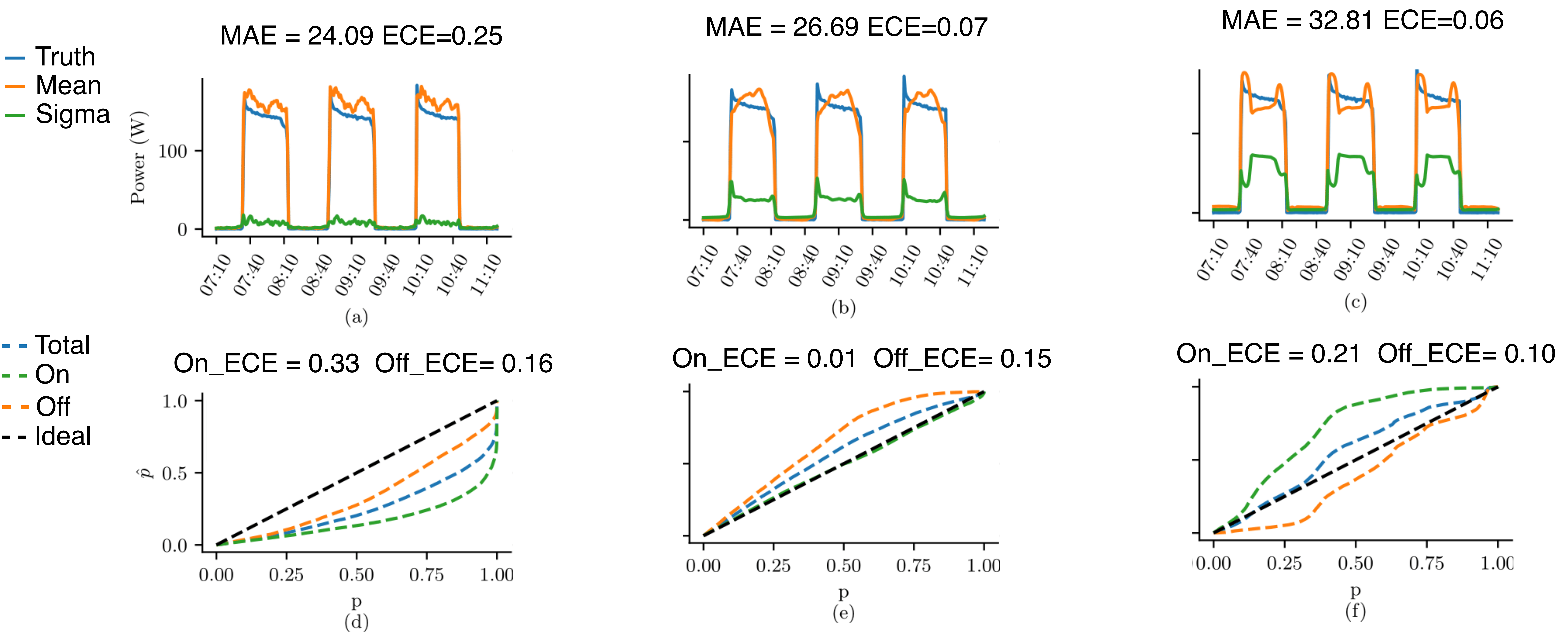


(d)

Do we need to do analysis of statewise (On and Off) calibration?

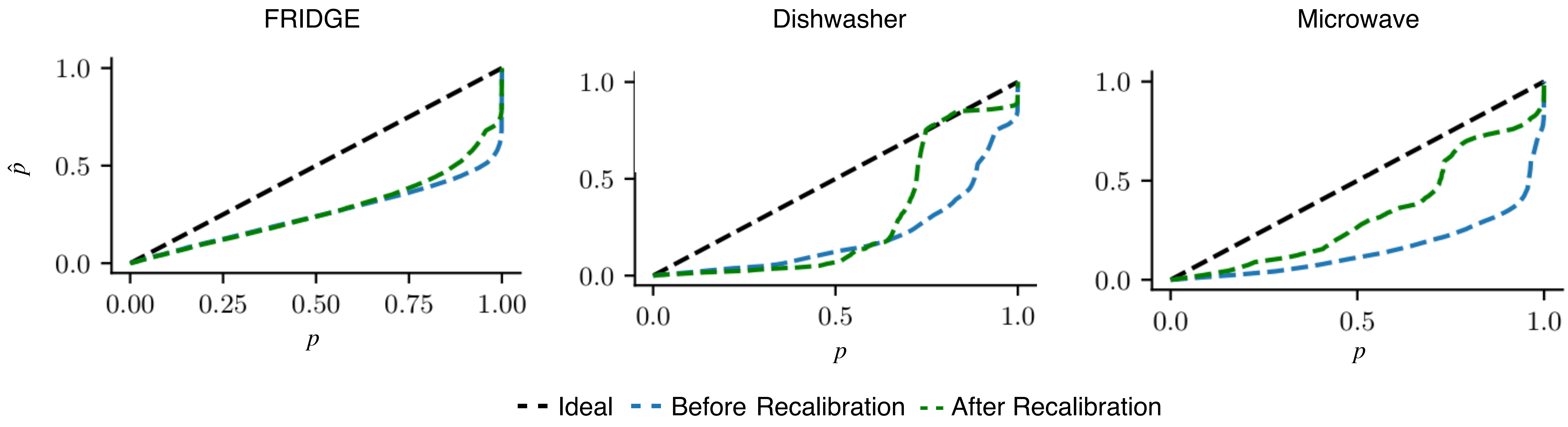


Do we need to do analysis of statewise (On and Off) calibration?



Can recalibration improve model uncertainty?

- After recalibration the reliability plot approaches more towards the ideal line.



Limitations and Future Work

Limitations and Future Work

- Studied performance over one dataset, we plan to expand over more datasets.

Limitations and Future Work

- Studied performance over one dataset, we plan to expand over more datasets.
- In future we plan to add Markov chain Monte Carlo (MCMC) methods and variational inference methods.

Limitations and Future Work

- Studied performance over one dataset, we plan to expand over more datasets.
- In future we plan to add Markov chain Monte Carlo (MCMC) methods and variational inference methods.
- We plan to study our techniques on out of distribution (O.O.D) settings.

Summary

Summary

- We implemented three ways to get uncertainty from neural network for NILM.

Summary

- We implemented three ways to get uncertainty from neural network for NILM.
- We found that new models incorporating uncertainty performed well on conventional metrics.

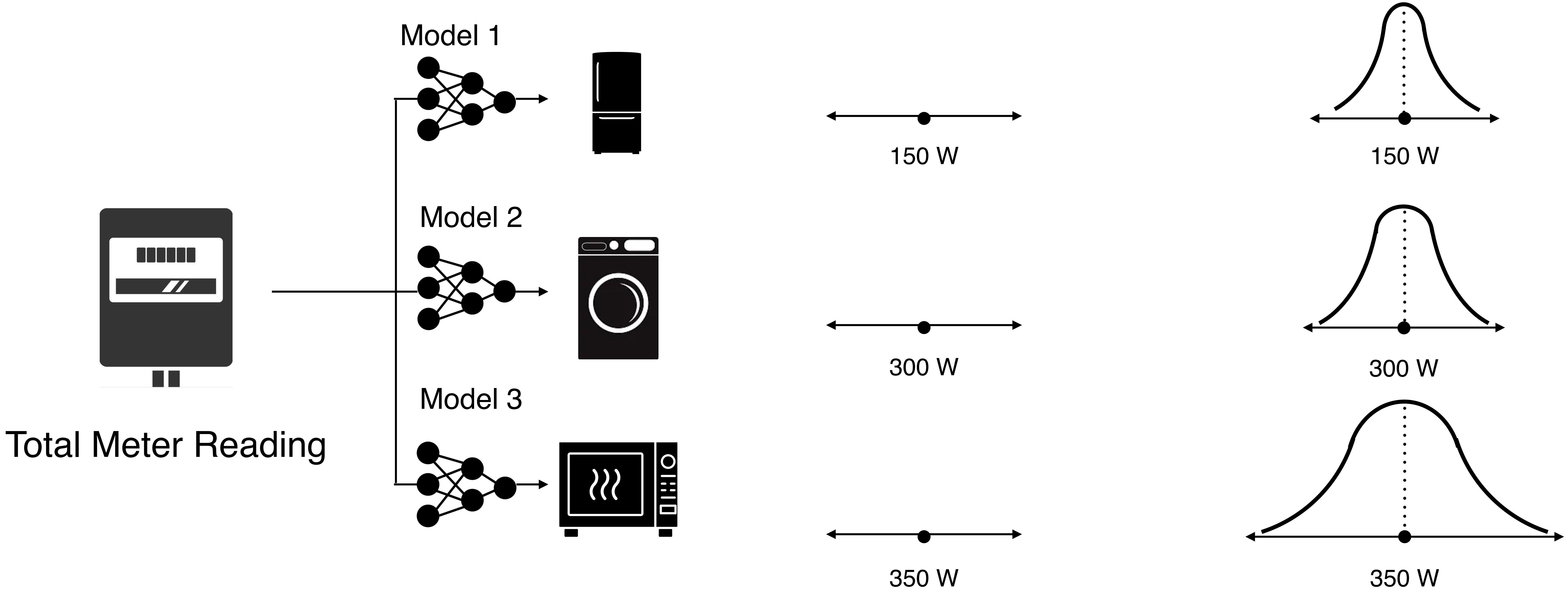
Summary

- We implemented three ways to get uncertainty from neural network for NILM.
- We found that new models incorporating uncertainty performed well on conventional metrics.
- We showed the need of studying calibration of different states of an appliance separately.

Backup Slides

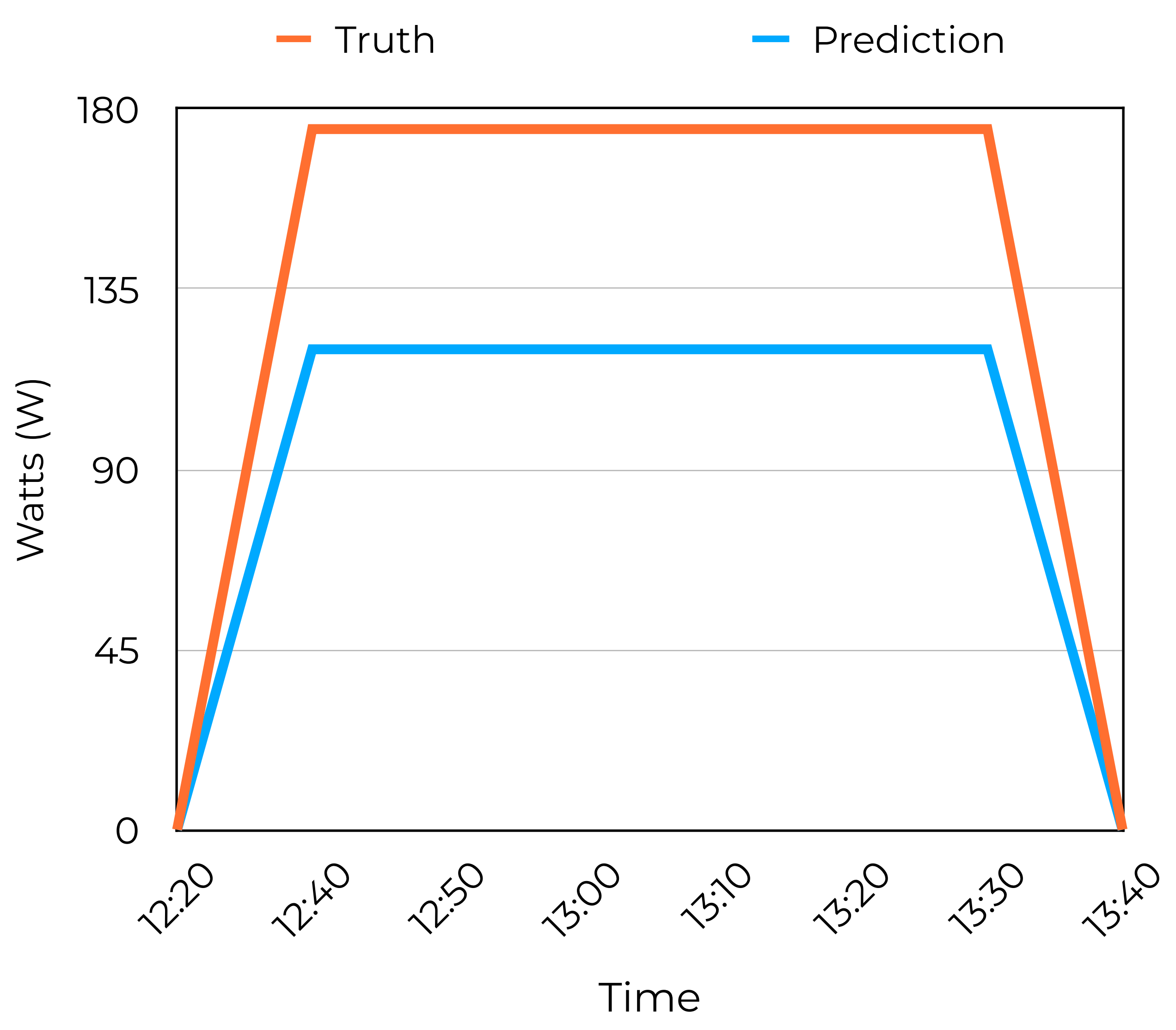
Motivation

- If value of uncertainty is high then model is unsure and decision maker can consider predictive uncertainty.

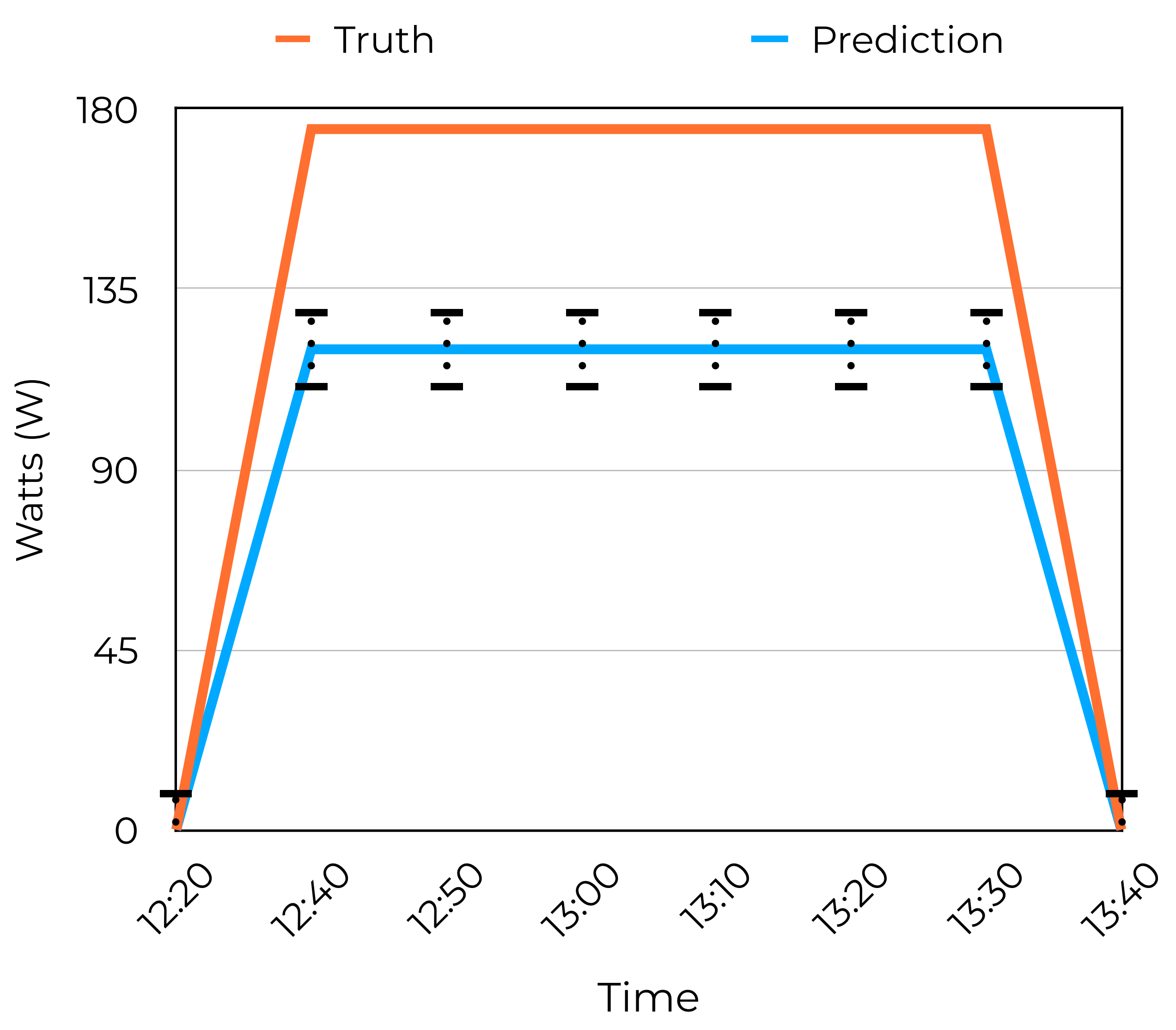


Motivation

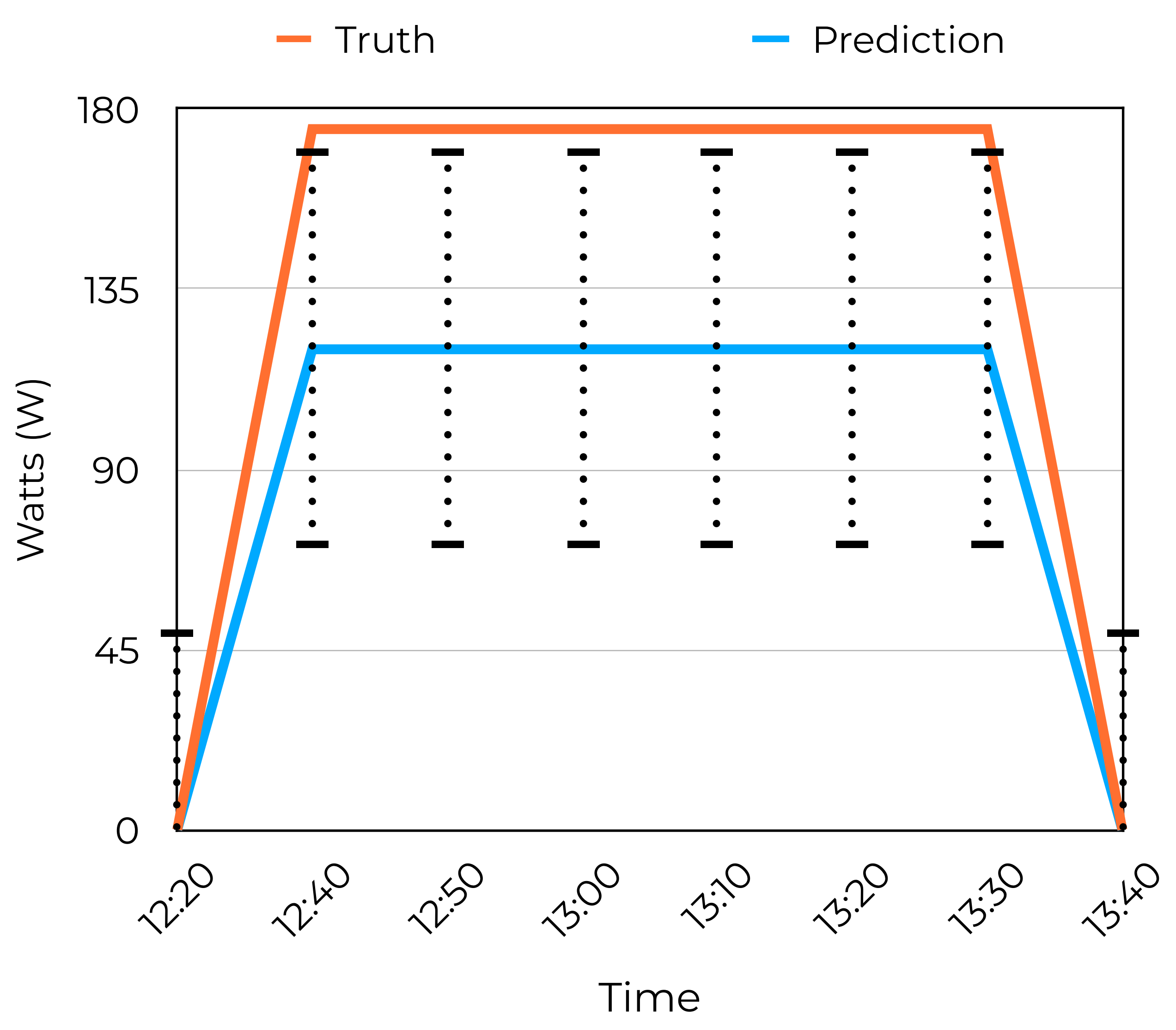
- If value of uncertainty is high then model is unsure and decision maker can consider predictive uncertainty before deciding.



Model capturing no uncertainty



Model capturing low uncertainty



Model capturing high uncertainty

Contribution

- *We have implemented a total of 14 such model variants over the state-of-the-art NNs for NILM.*

- *Proposed "re-calibration" method to improve the uncertainty quantification for our models*

- *Provided qualitative understanding of predictive uncertainty from NILM perspective.*

Loss details

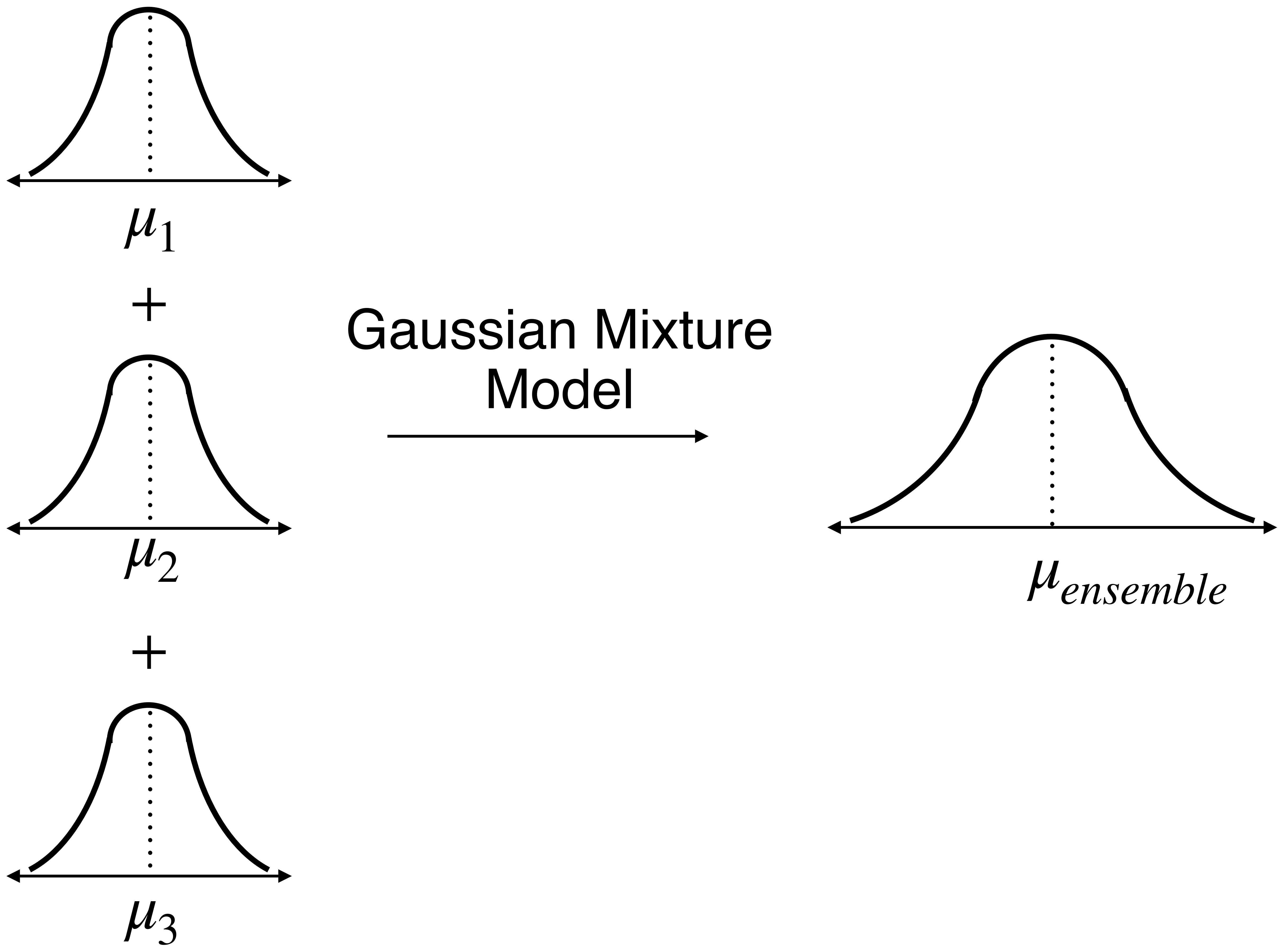
Homoskedastic Regression

- Loss $= \dfrac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}$

Heteroskedastic Regression

- $\text{Loss} = \dfrac{\sum_{i=1}^{N} \left(-\frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(y_i - \hat{y}_i\right)^2\right)}{N}$

Prediction from an ensemble of NNs



Neural Networks for NILM

1. Seq2Point

    - Seq2Point maps the sequence of mains power to a point appliance power prediction.

2. BiLSTM with Attention Mechanism

    - BiLSTM considers the future and past aggregate reading data while training and testing.
    - Attention mechanism provide more focus to some part of the sequence window.

Prediction from an ensemble of NNs

Homoskedastic NN

- μ_i :- Prediction after forward pass in NN.

$$\mu_{\text{ensemble}} = \frac{\sum_{i=1}^{N} \mu_i}{N}$$

$$\sigma_{\text{ensemble}} = \sqrt{\frac{\sum_{i=1}^{N} (\mu_i - \mu_{\text{ensemble}})^2}{N}}$$

Heteroskedastic NN

- μ_i, σ_i :- Prediction after forward pass in NN.

$$\mu_{\text{ensemble}} = \frac{\sum_{i=1}^{N} \mu_i}{N}$$

$$\sigma_{\text{ensemble}} = \sqrt{\frac{\sum_{i=1}^{N} (\sigma_i^2 + \mu_i^2)}{N} - \mu^2_{\text{ensemble}}}$$

Types of Uncertainty

Aleatoric Uncertainty

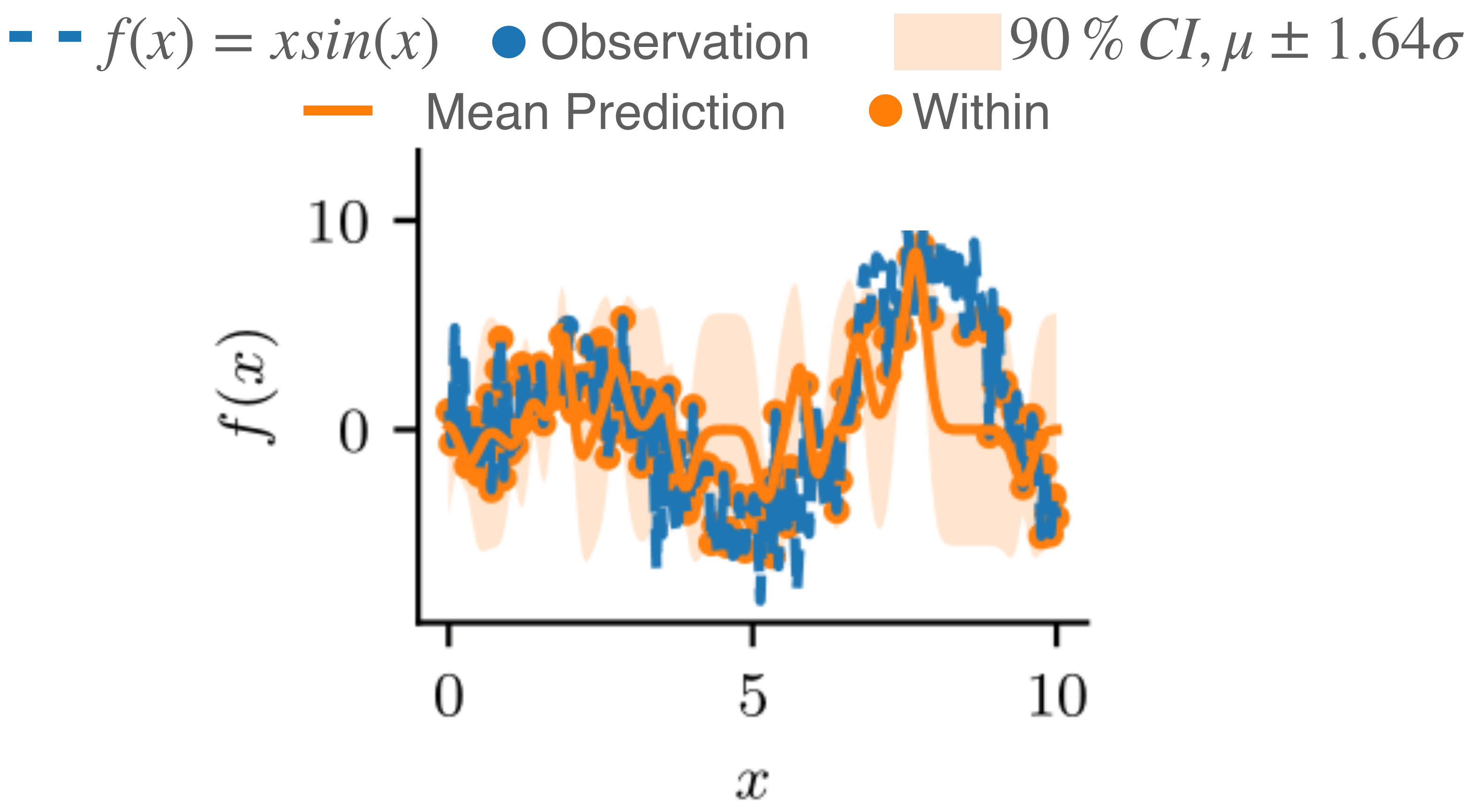
- This type of uncertainty deals with data uncertainty.

Epistemic Uncertainty

- This type of uncertainty deals with data uncertainty of parameters.

Quantifying predictive uncertainty

- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.



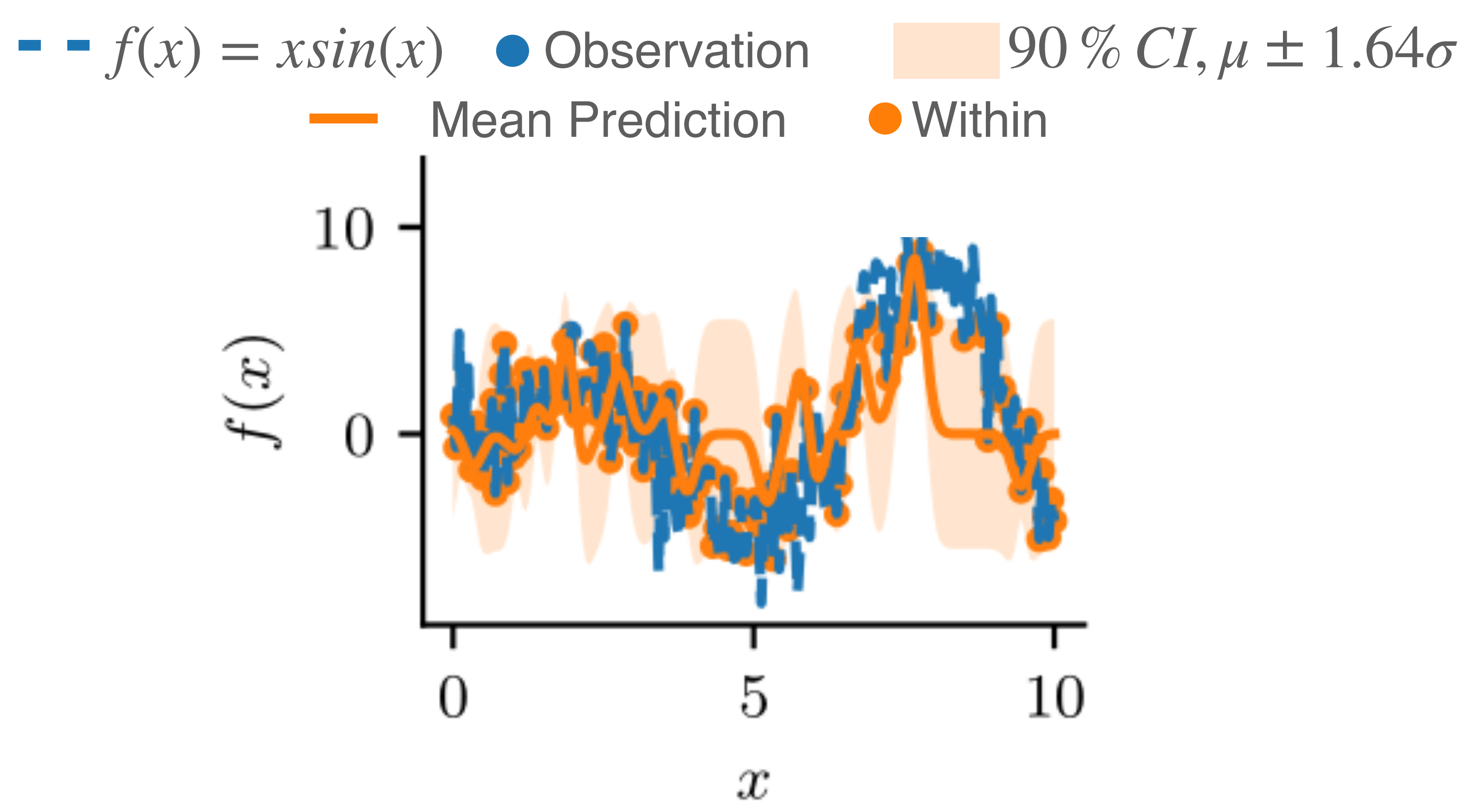
(a)

% of points within CI:-
72.5%

Quantifying predictive uncertainty

- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.

$$\text{Expected Calibration Error} \quad \text{(ECE)} = \frac{\sum_{p=1}^{P} |\hat{p}(p) - p|}{P}$$



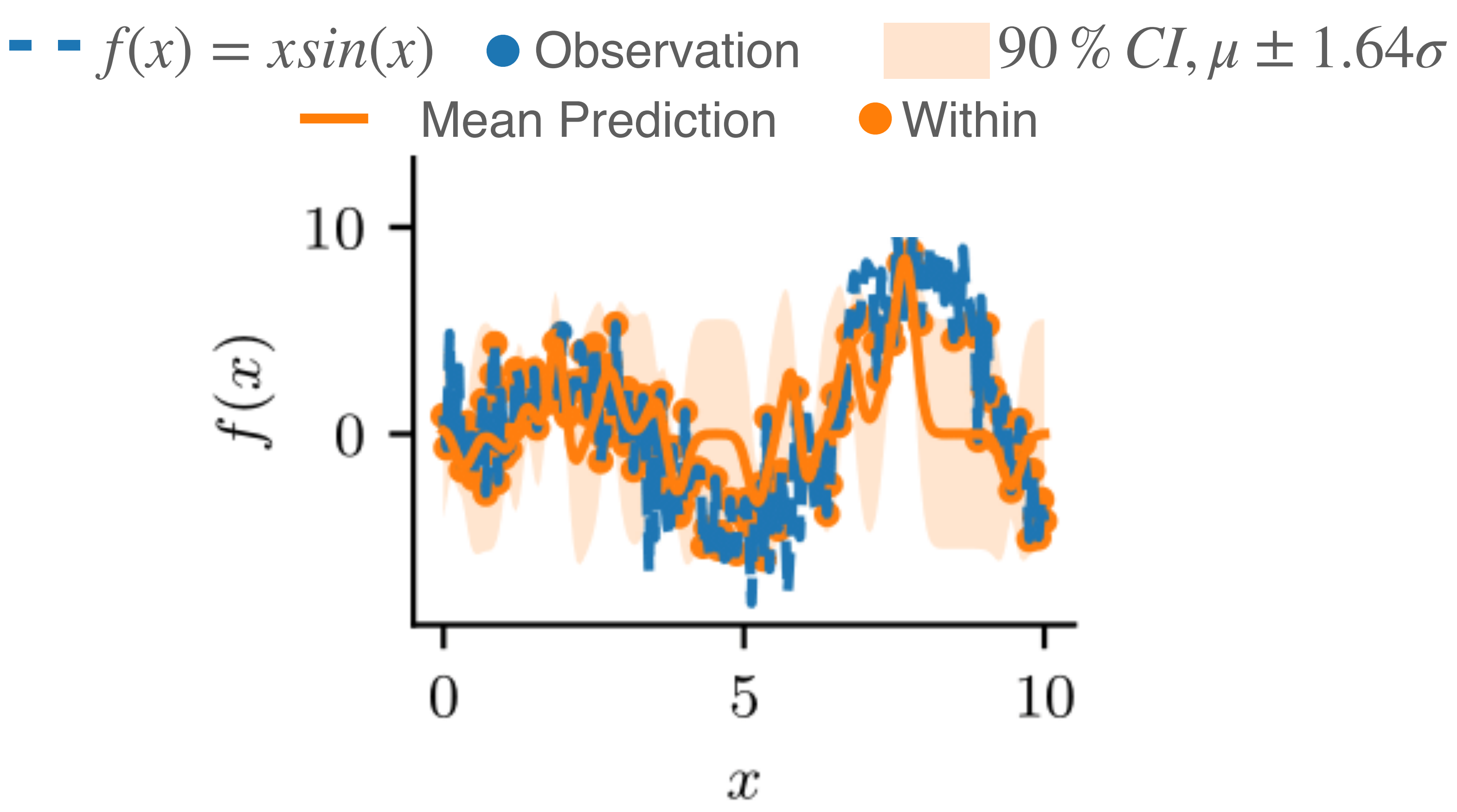
(a)

% of points within CI:- 72.5%

Quantifying predictive uncertainty

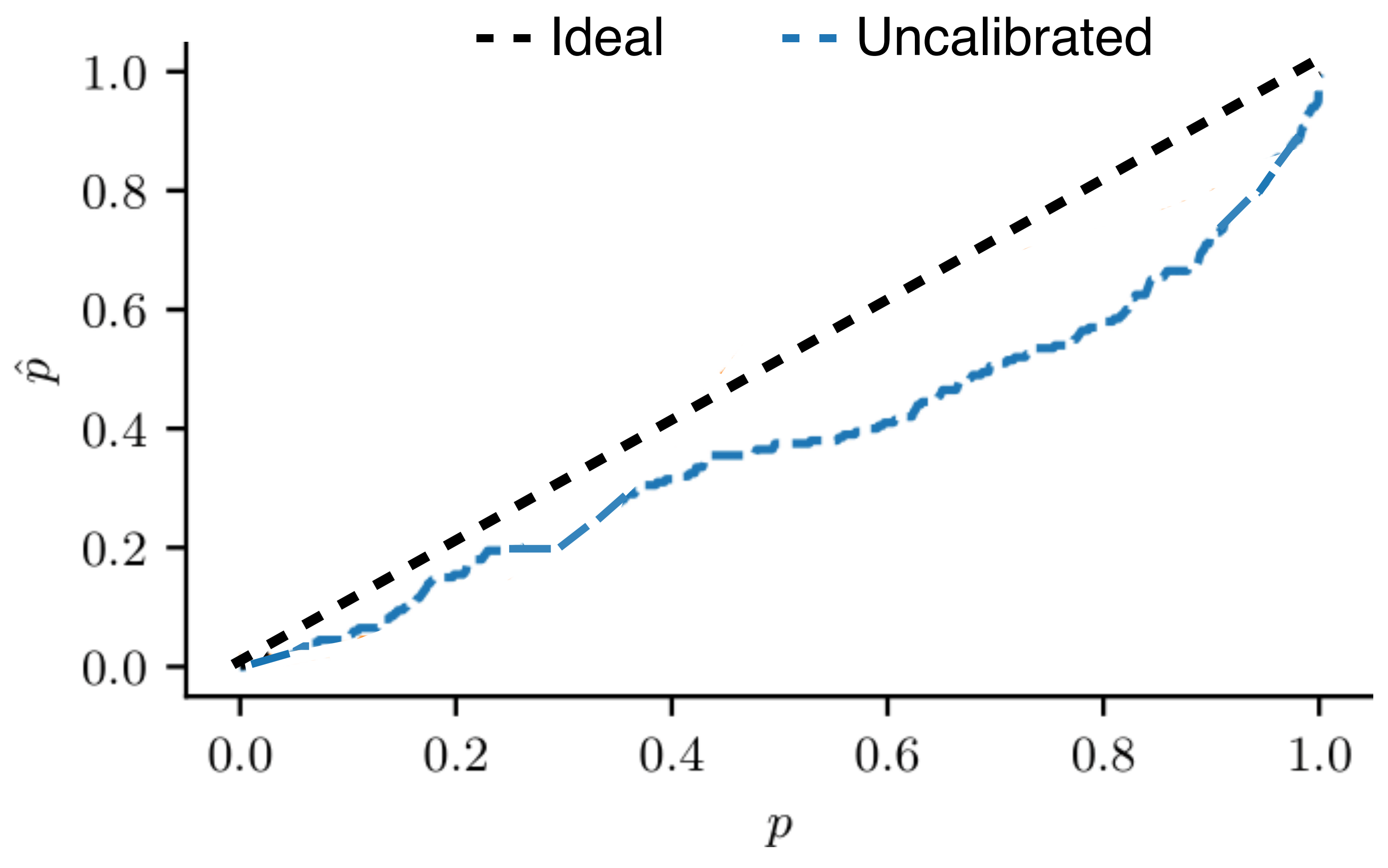
- The chosen confidence interval as p and empirically found fraction of points $\hat{p}$ within the band should be ideally equal.

$$\text{Expected Calibration Error} \quad (\text{ECE}) = \frac{\sum_{p=1}^{P} |\hat{p}(p) - p|}{P}$$



(a)

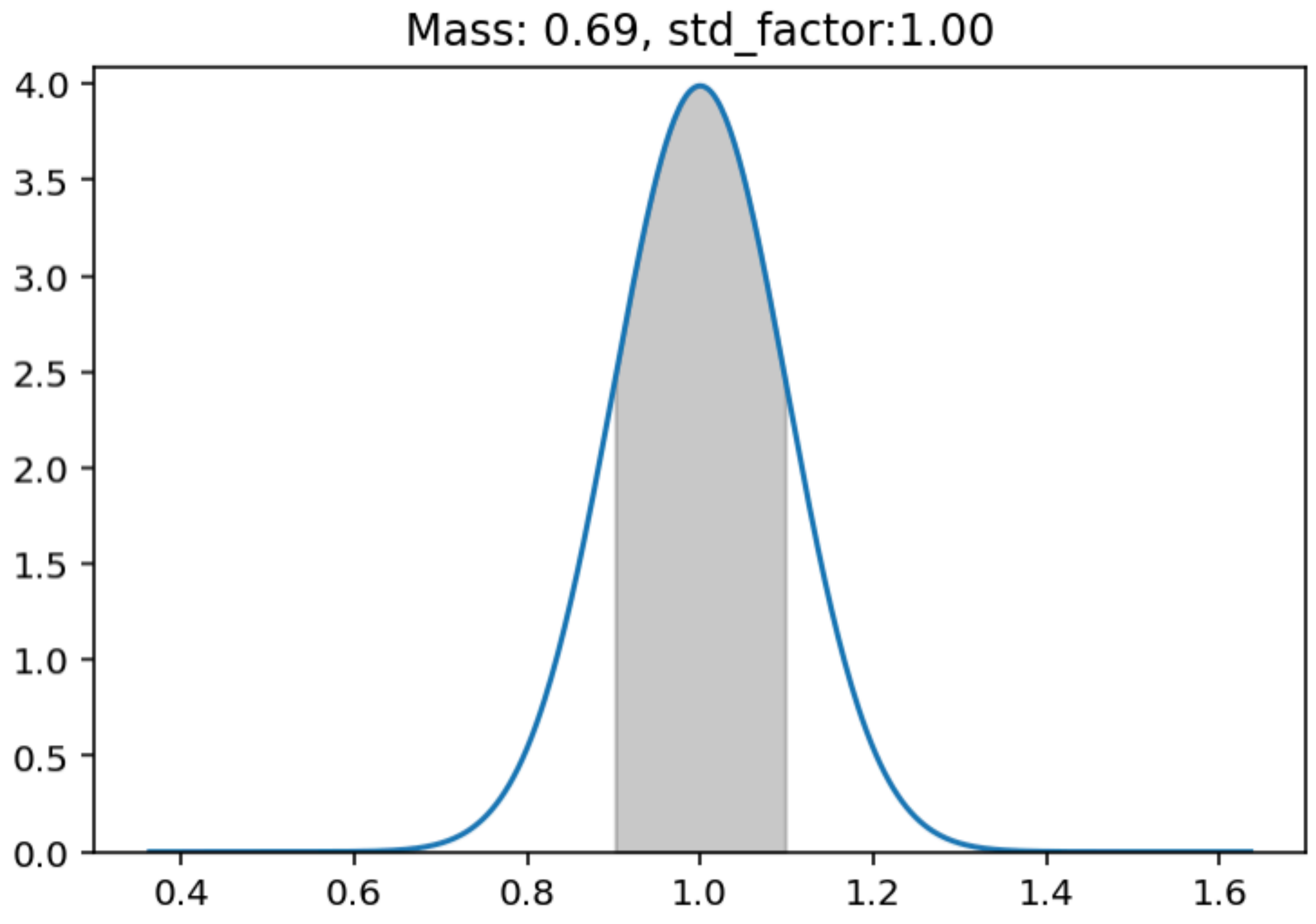
% of points within CI:-
72.5%



Reliability Plot

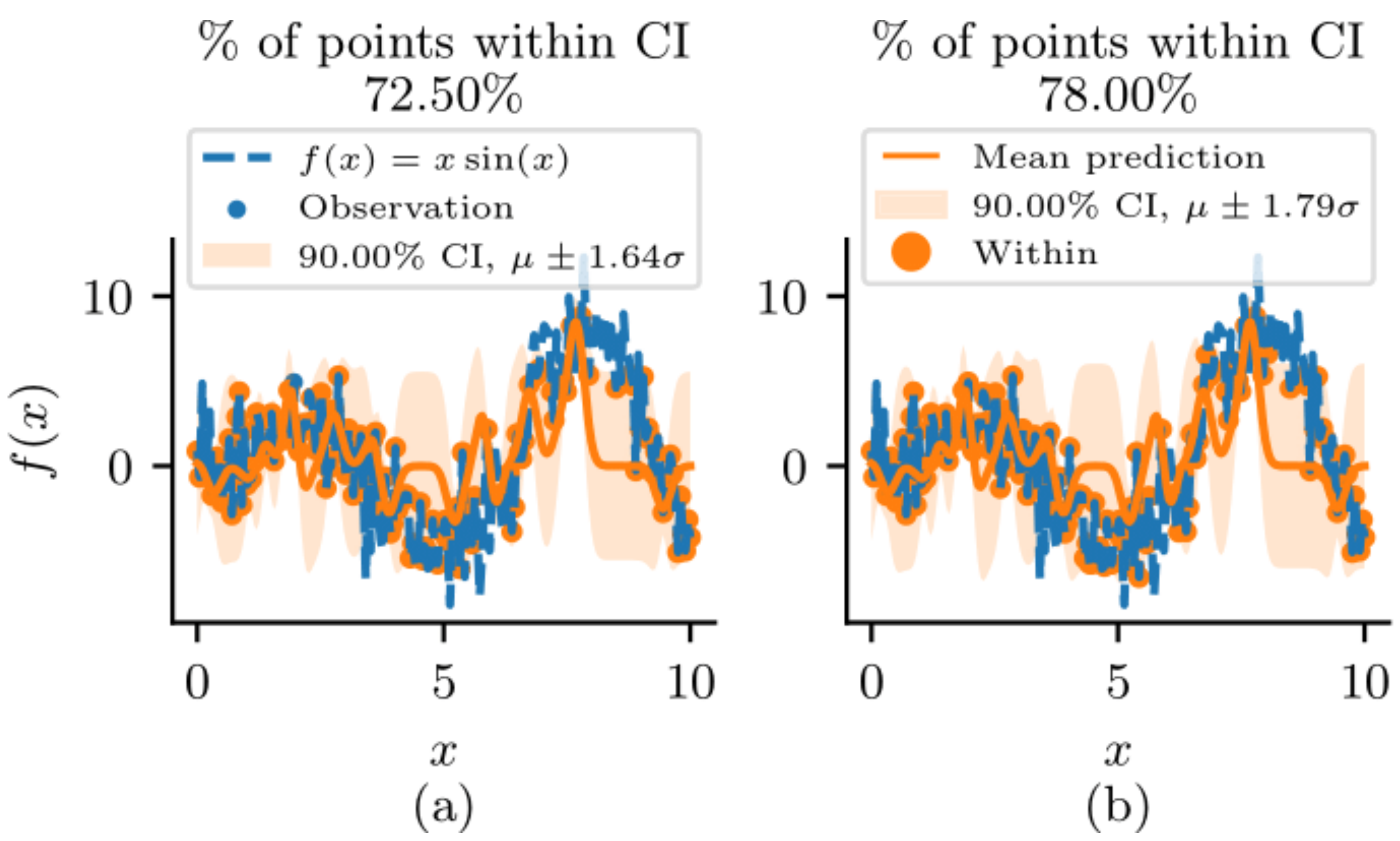
How we found Standard deviation factor from CI?

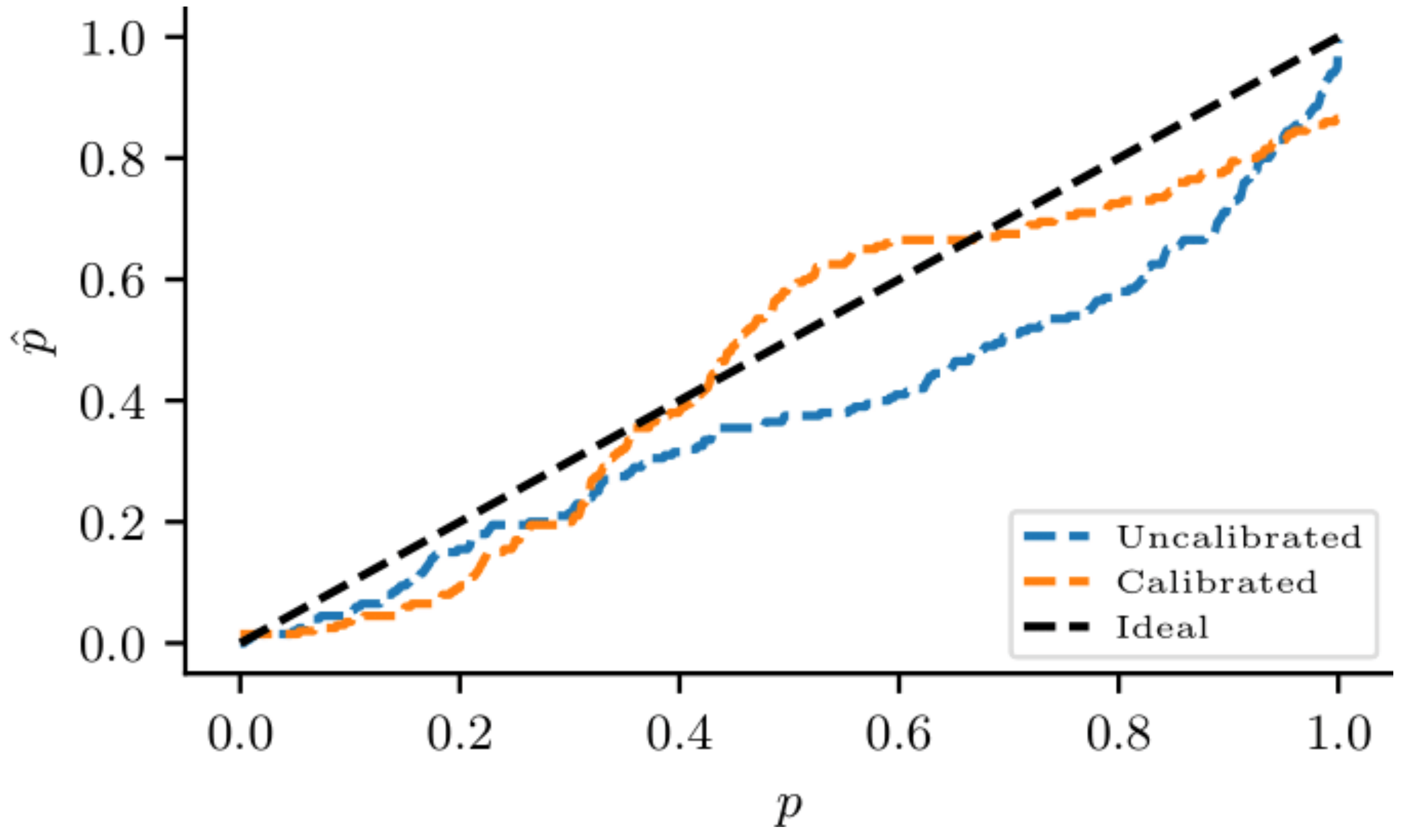
- $Pr(\mu - z\sigma \leq X \leq \mu + z\sigma) = 2cdf(z) - 1 = mass$
- $z = cdf^{-1}((mass + 1)/2)$



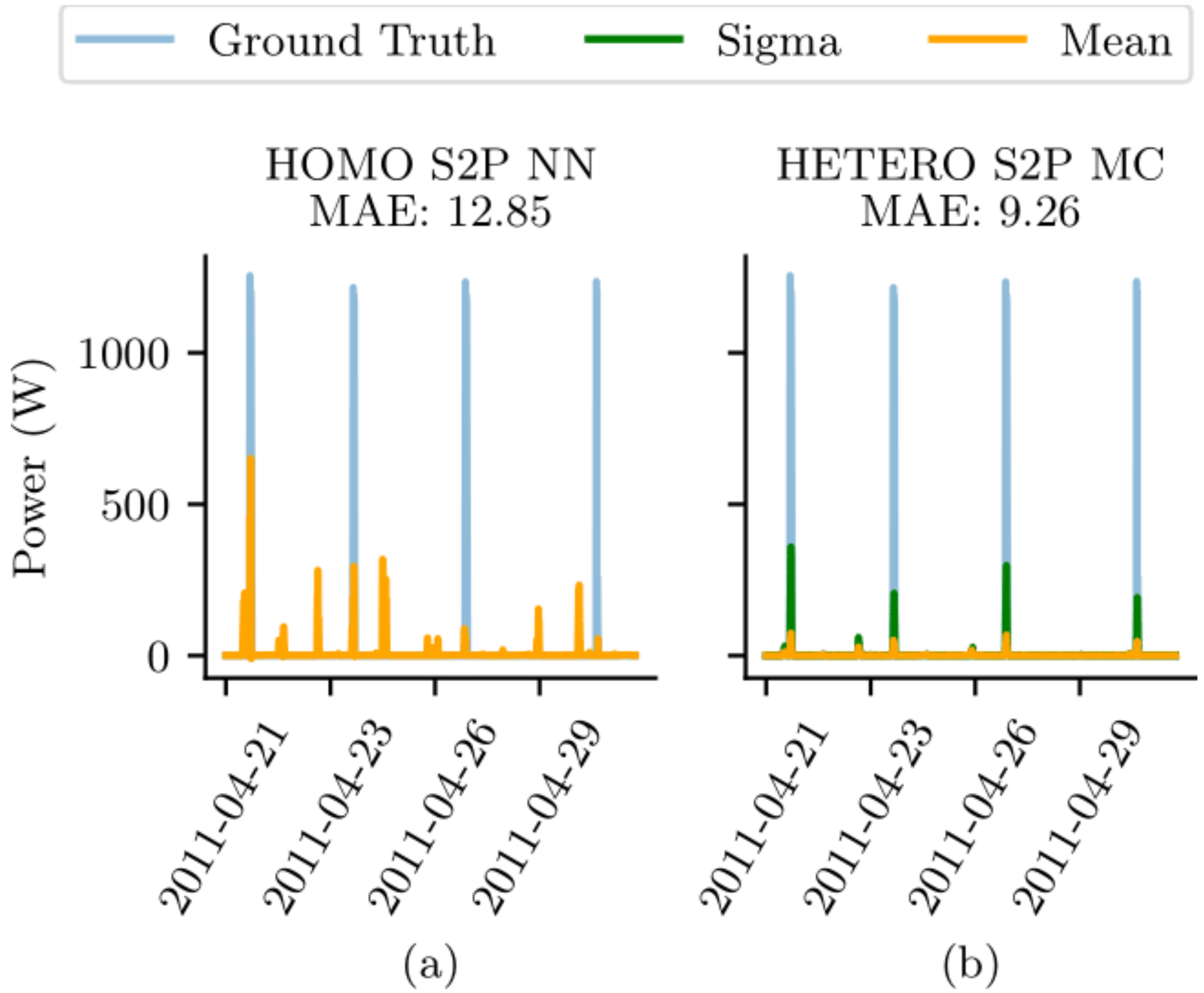
Model Recalibration

- Model recalibration methods tries to increase or decrease the band width of the confidence interval to cover same number empirically found points as confidence interval.





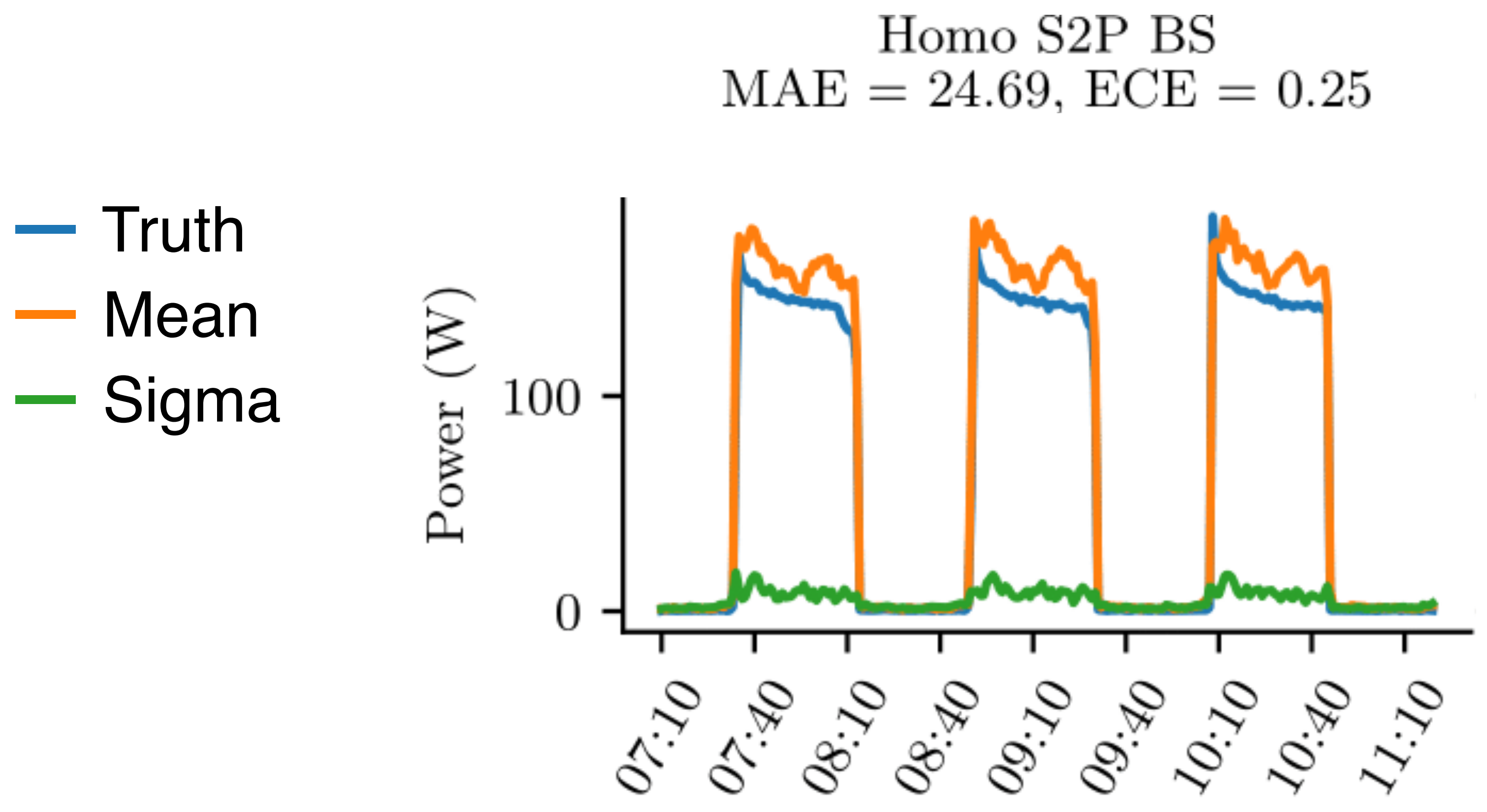
- Higher uncertainty when appliance changes state as model likely to be uncertain during transition and more confident once it observes more sample.



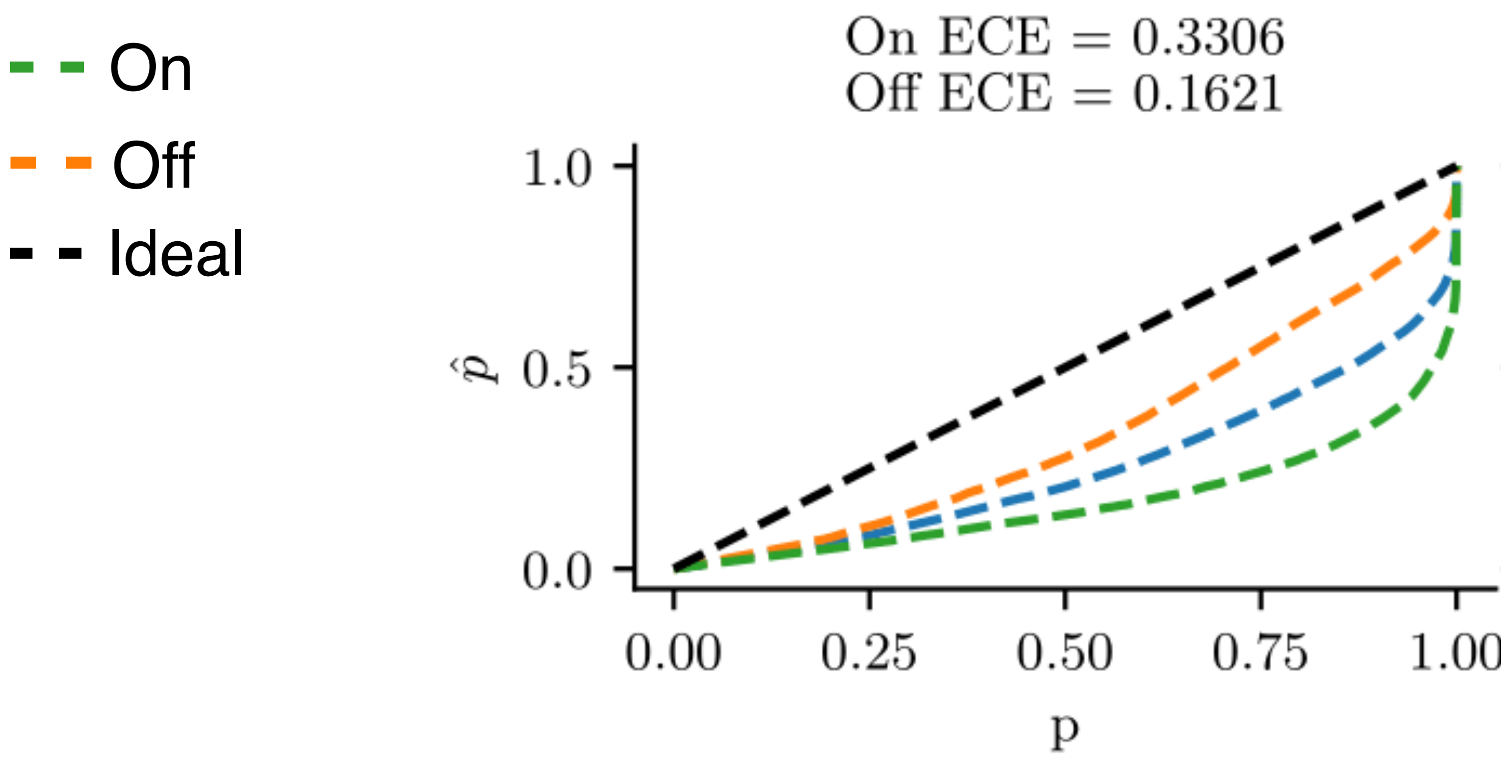
Dishwasher

* refers to baseline

- Different appliance states can have highly varying calibration curves.
- Can achieve an overall low calibration if the individual states calibration errors cancel out each other.

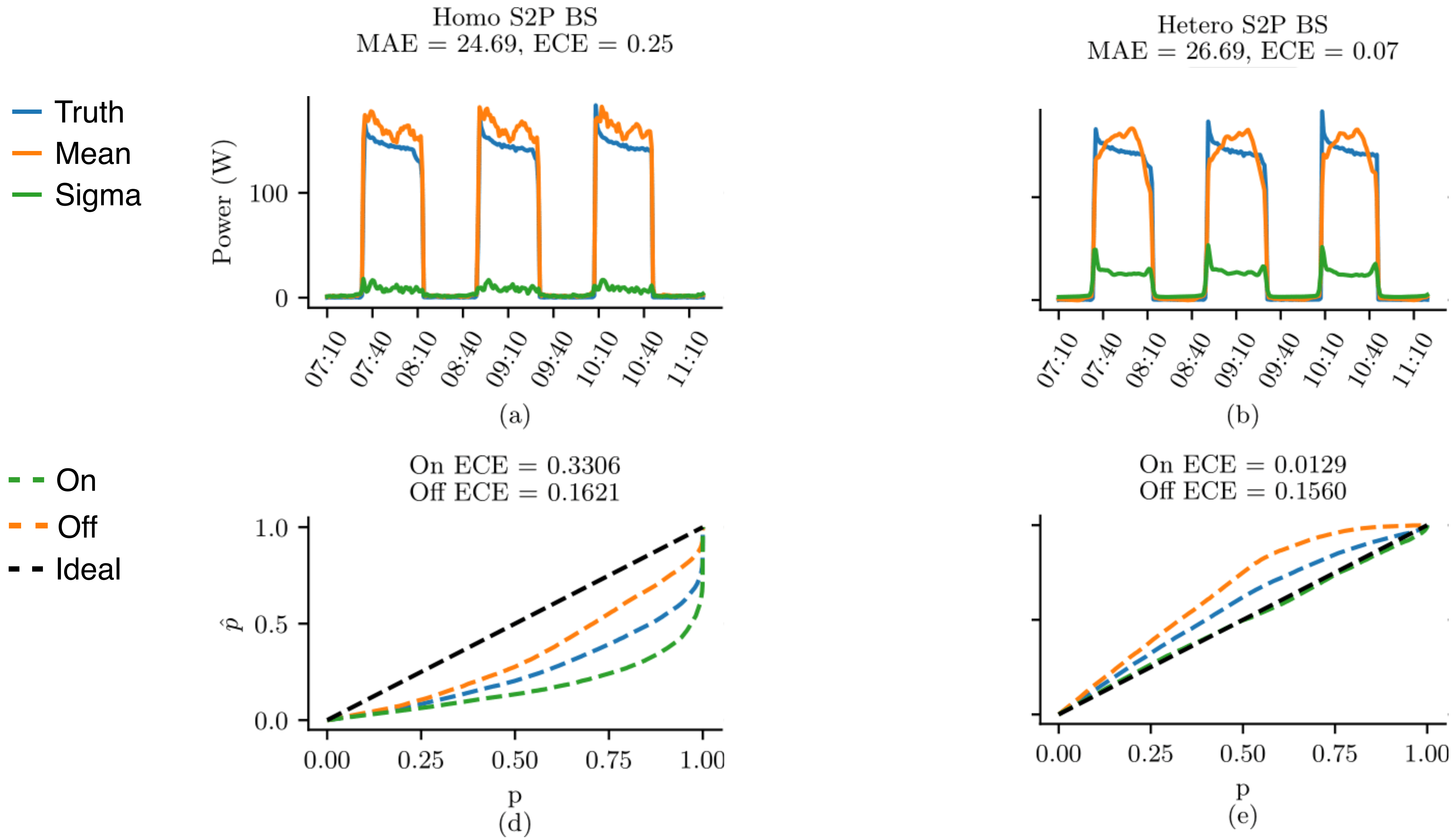


(a)

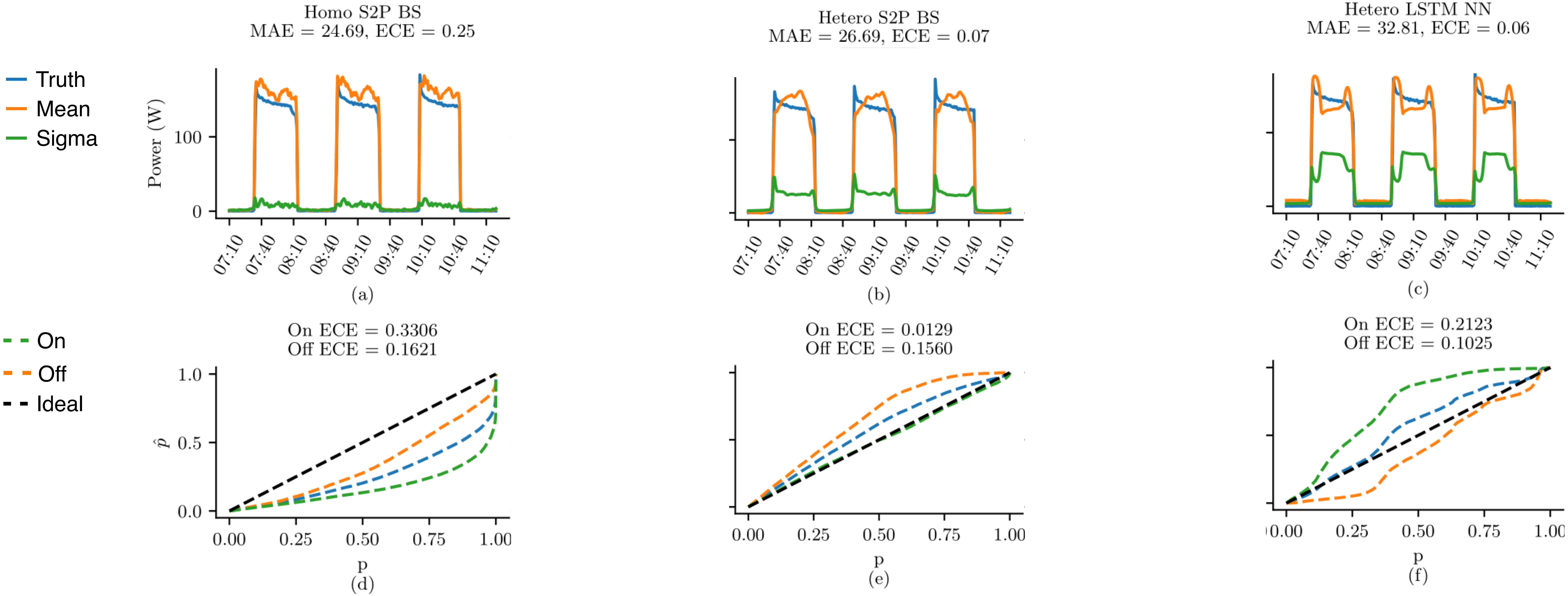


(d)

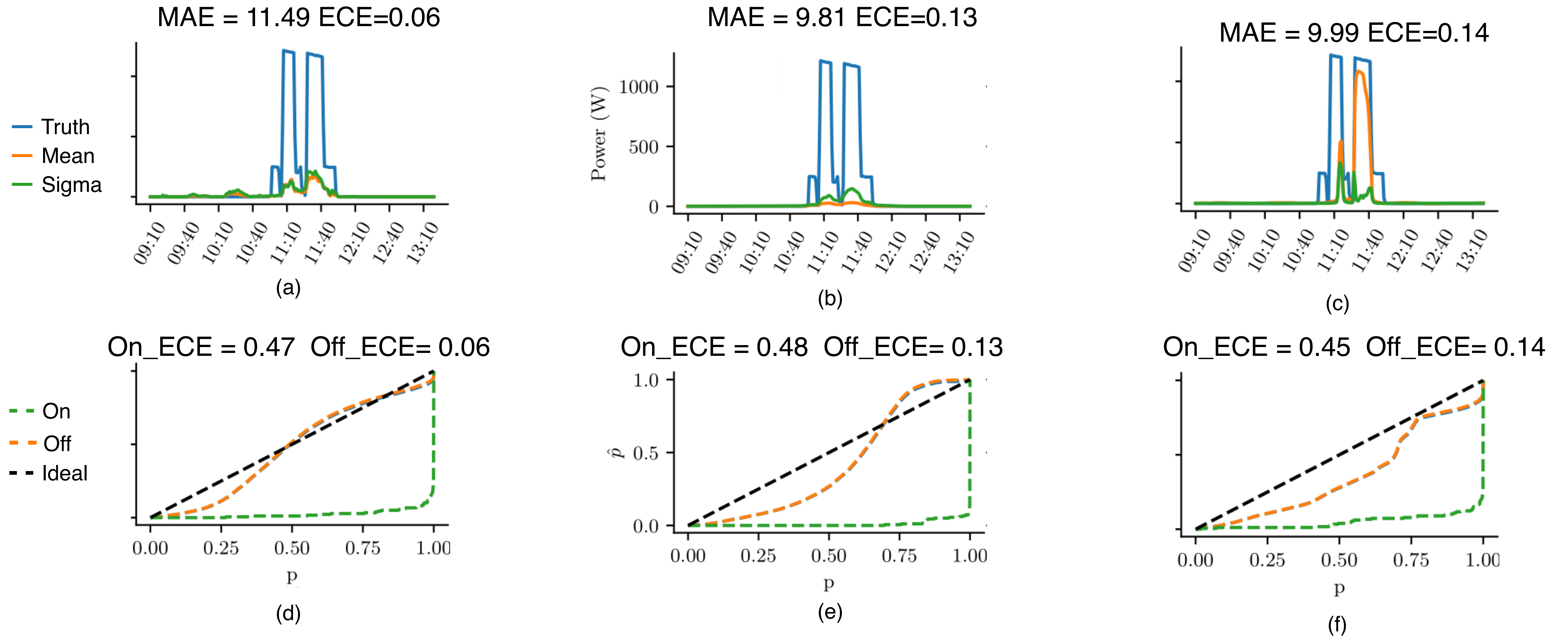
- Different appliance states can have highly varying calibration curves.
- Can achieve an overall low calibration if the individual states calibration errors cancel out each other.



- Different appliance states can have highly varying calibration curves.
- Can achieve an overall low calibration if the individual states calibration errors cancel out each other.

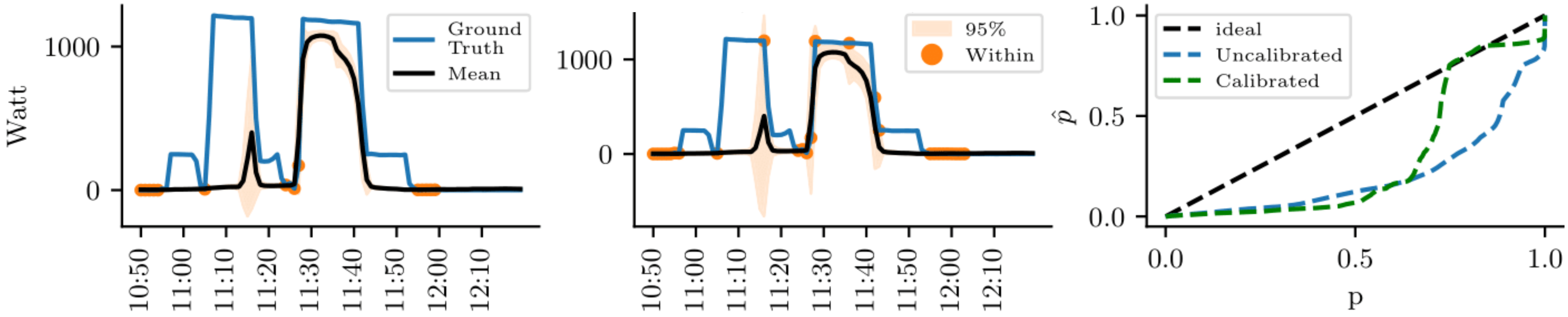


- Good ECE may hide the imbalance between the different states and so recommended to see state-wise ECE.
- Difficult to estimate uncertainty accurately for sparse appliances.



- Improvement in quantifying model uncertainty of dishwasher and microwave for both “ON” and “OFF” state.
- Still chances of improvement in quantifying model uncertainty in “ON” state.

Dishwasher

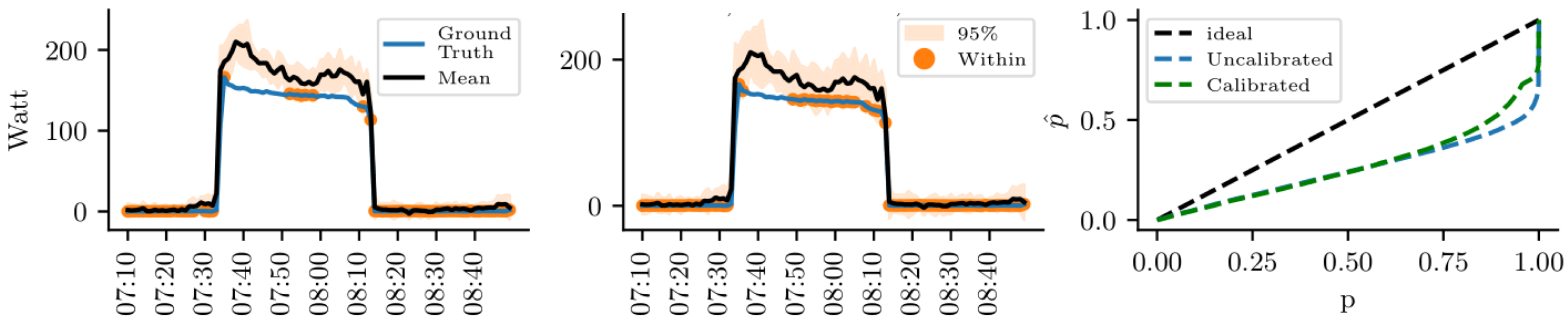


BiLSTM with Attention, Before: 14% , After: 26%

Can recalibration improve model uncertainty?

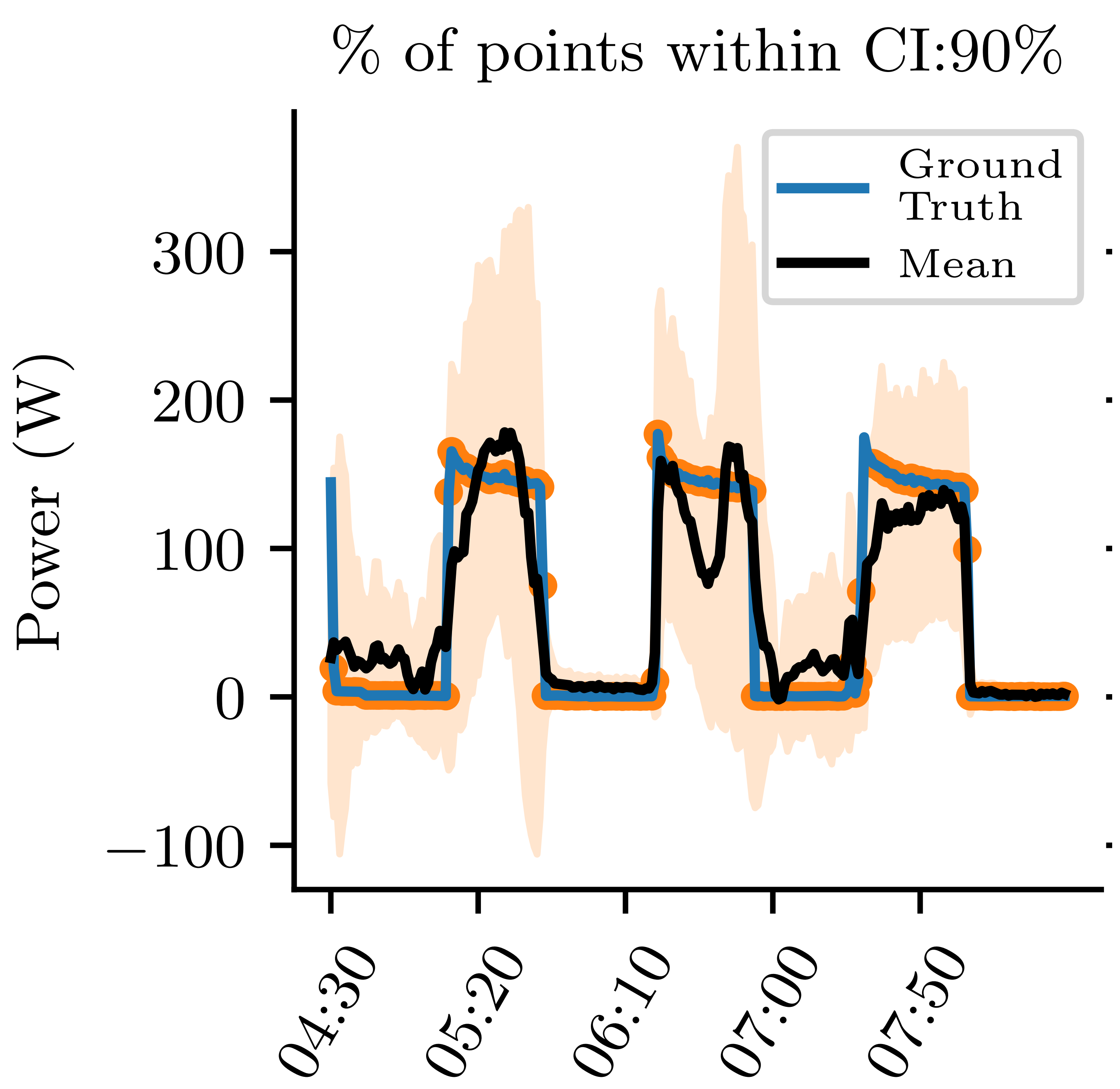
- Improvement from 65% points in 95% confidence interval to 80% points which can be seen even in reliability diagram.

Fridge

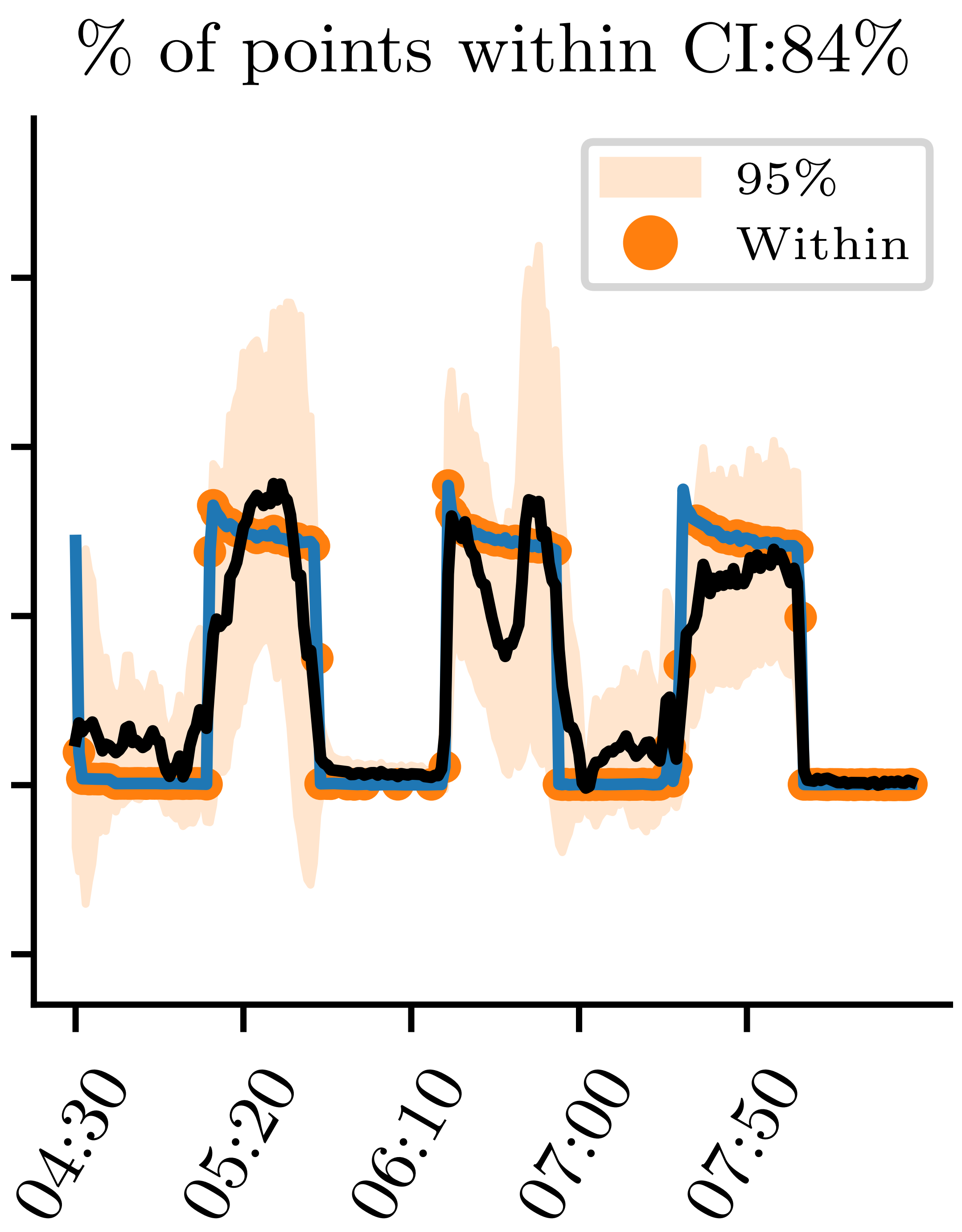


Seq2point HOMO MC, Before: 65%, After 80%

- Good recalibration will require similar characteristic between calibration and test dataset.



(a). Before Recalibration



(b). After Recalibration