Machine Learning Estimation on the Trace of Inverse Dirac Operator using the Gradient Boosting Decision Tree Regression

Benjamin J. Choi[1], Hiroshi Ohno[1],
Takayuki Sumimoto[2] and Akio Tomiya[3]

[1]Center for Computational Sciences, University of Tsukuba, Japan
[2]FLECT Co., Ltd., Japan
[3]Division of Mathematical Sciences, Tokyo Woman's Christian University, Japan

Lattice 2024 Liverpool,
29 July 2024

Program for Promoting Researches on the Supercomputer Fugaku
Large-scale lattice QCD simulation and development of AI technology

ML PhYs

筑波大学 計算科学研究センター Center for Computational Sciences

Introduction and Motivation

- For Lattice QCD calculations
  - on observables such as cumulants of the chiral order parameter,
  - the trace of operators ($\mathrm{Tr}\ O$) is often necessary.
    - ex) $O = M^{-1}$ (inversed Dirac operator).

- M (Dirac operator): a large sparse matrix on the lattice
  - ➡ Trace estimation by **linear CG solver** for stochastic sources
  - ➡ This needs considerable computational cost.

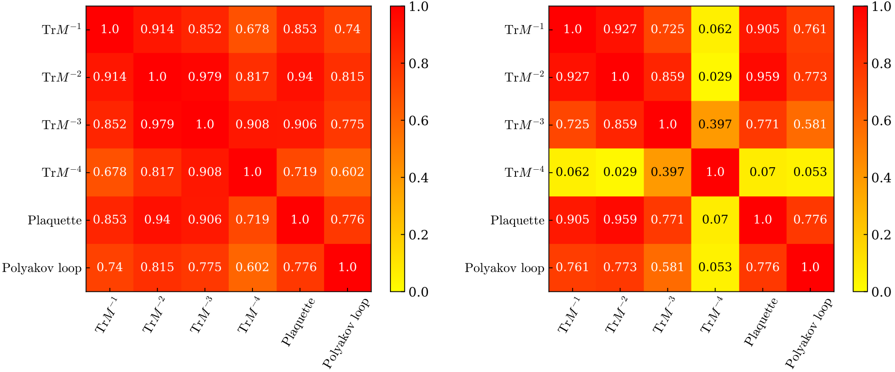
- We present our preliminary result
  - on machine learning estimation of $\mathrm{Tr} M^{-n}$
  - from other observables
    - ex) $\mathrm{Tr} M^{-m}$ where $m < n$, plaquette, Polyakov loop
  - with **gradient boosting decision tree regresson**
  - based on the methodology of Yoon *et al.*, PRD **100** 014504 (2019).

Measurement Information

ID	$L^3 \times T$	β	κ	c_{SW}	N_{conf}
0	$16^3 \times 4$	1.60	0.13575	2.065	5500
1	$16^3 \times 4$	1.60	0.13577	2.065	5500
2	$16^3 \times 4$	1.60	0.13580	2.065	5500
3	$16^3 \times 4$	1.60	0.13582	2.065	5500
4	$16^3 \times 4$	1.60	0.13585	2.065	5500

- Measured by H. Ohno (2017 — 2018)
  - Ohno *et al.* PoS **LATTICE2018** (2018) 174
  - HW: `Oakforest-PACS` system (Boku *et al.* arXiv:1709.08785)
  - SW: `BQCD` program (Haar *et al.* EPJWoC **175** 14011 (LAT2017))
- Iwasaki gauge action, Wilson clover action, $N_f = 4$
- Red row: first order phase transition occurs here.

Correlation Map of Observables



(a) $\kappa = 0.13575$, `ID 0` (heaviest quark) (b) $\kappa = 0.13585$, `ID 4` (lightest quark)

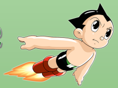
- We evaluate the correlation coefficients for Plaquette, Polyakov loop and $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- Red color: **strong** correlation ➡ machine learning estimation
- Using the correlation between observables, we try ML estimation.

Machine Learning Estimation

Input

$X = \{x_1, x_2, \cdots\}$

Machine



$f(X)$ ex) Gradient Boosting Decision Tree Algorithm

Output

$Y^P \approx Y$

- We adopt machine learning method for the estimation of observables.
  - Yoon *et al.*, PRD **100** 014504 (2019)
- Machine learning estimation

$$X \quad \text{(input)} \quad \rightarrow \quad Y^P \approx Y \quad \text{(output)}$$

  - P: prediction by machine learning
- A machine f determines fit function during the learning sequence

$$f(X) = Y^P \approx Y$$

Machine Learning Estimation of Observables

Input

$$X = \{x_1, x_2, \cdots\}$$

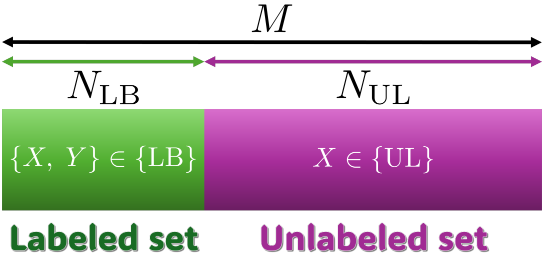
Machine

$f(X)$ ex) Gradient Boosting Decision Tree Algorithm

Output

$$Y^P \approx Y$$

- For ML estimation of $f(X) = Y^P \approx Y$,
  - # of data X: $M = N_{\rm LB} + N_{\rm UL}$
  - # of data Y: $N_{\rm LB}$

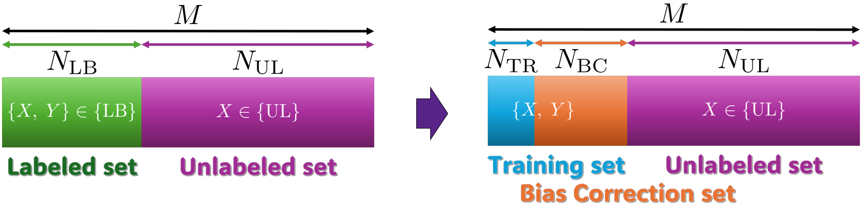


1. Train f to get Y from X on labeled set.
2. Predict Y^P from X on **unlabeled** set:

$$f(X) = Y^P \approx Y\,.$$

- Do we use **all labeled set** for training?

Bias correction in the ML estimation



- In general and in principle,

$$\bar{Y}_{(\mathrm{UL})} = \frac{1}{N_{\mathrm{UL}}} \sum_{i=1}^{N_{\mathrm{UL}}} Y_i^P$$

  is not exact due to the bias.
- Prediction bias in the ML

$$(\mathrm{Bias}) = \langle Y \rangle - \langle Y^P \rangle$$

- We **split** labeled set into training and bias correction set.

1 Training with $\{X, Y\} \in \{\mathrm{TR}\}$

2 Predicting with

$$\bar{Y} = \frac{1}{N_{\mathrm{UL}}} \sum_{i=1}^{N_{\mathrm{UL}}} Y_i^P + \frac{1}{N_{\mathrm{BC}}} \sum_{j=1}^{N_{\mathrm{BC}}} (Y_j - Y_j^P)$$

(Yoon *et al.*, PRD **100** 014504)

Optimal range for ratio of labeled/training set

- To reduce the computational cost ➡ **minimal** sufficient labeled set.
  - In bias correction set, Y as well as Y^P is used to determine $\bar{Y}$.
  - ➡ We should grant **sufficient statistics** to bias correction set.
- ➡ Need to find **optimal range** for ratio of labeled/training set!

- M = # of total data set ($M = N_{\mathrm{LB}} + N_{\mathrm{UL}}$)
- N_{LB} = # of labeled set ($N_{\mathrm{LB}} = N_{\mathrm{TR}} + N_{\mathrm{BC}}$)
- N_{TR} = # of training set

1. Find out minimal $\mathcal{R}_{\mathrm{LB}} \equiv \dfrac{N_{\mathrm{LB}}}{M}$ where $\mathcal{R}_{\mathrm{LB}} = 5, 10, \cdots 50\ \%$
2. Find out maximal $\mathcal{R}_{\mathrm{TR}} \equiv \dfrac{N_{\mathrm{TR}}}{N_{\mathrm{LB}}}$ where $\mathcal{R}_{\mathrm{TR}} = 10, 20, \cdots 90\ \%$

- We also observe
  - $\mathcal{R}_{\mathrm{TR}} = 0\ \%$ to check the labeled set itself,
  - $\mathcal{R}_{\mathrm{TR}} = 100\ \%$ to check the result without the bias correction.

Gradient boosting decision tree regression

- We use LightGBM (Microsoft) via `JuliaAI/MLJ.jl`.
  - boosting stage = 40 ➡ empirically determined with L2-Loss plot.
  - depth of tree = 3, learning rate = 0.1, subsampling = 0.7
    ➡ Same with Yoon *et al.*, PRD **100** 014504 (2019)

- Statistical error estimation: Bootstrap resampling, $N_{\mathrm{BS}} = 10,000$

- Check $\mathcal{P}1$ and $\mathcal{P}2$ as in Yoon *et al.*, PRD **100** 014504 (2019)

1 $\mathcal{P}1$: bias corrected ML prediction

$$\bar{Y}_{\mathcal{P}1} = \frac{1}{N_{\mathrm{UL}}} \sum_{i \in \{\mathrm{UL}\}} Y_i^{\mathrm{P}} + \frac{1}{N_{\mathrm{BC}}} \sum_{j \in \{\mathrm{BC}\}} \left(Y_j - Y_j^{\mathrm{P}} \right) \tag{1}$$

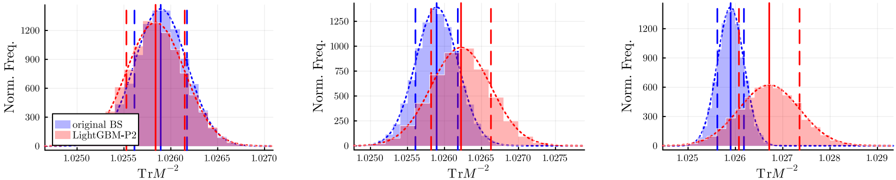
2 $\mathcal{P}2$: weighted average of $\mathcal{P}1$ and direct measured labeled set

$$\bar{Y}_{\mathcal{P}2} = \frac{N_{\mathrm{UL}}}{M} \bar{Y}_{\mathcal{P}1} + \frac{N_{\mathrm{LB}}}{M} \bar{Y}_{(\mathrm{LB})} \tag{2}$$

➡ To improve the statistical precision

Two evaluations of estimation results

- **Evaluation 1**: $\bar{Y}$ (central value) check (blue: original CG, red: ML)



(a) **Score 2** (b) **Score 1** (c) **Score 0**

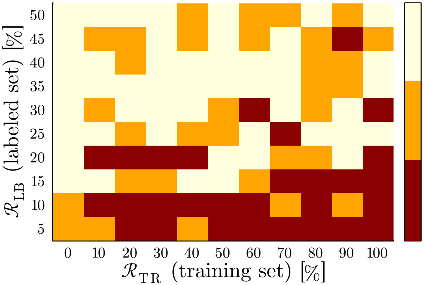
Score	Evaluation criteria ($\mathcal{X} = \mathcal{P}1$ or $\mathcal{P}2$)
2	**Both of** $\bar{Y}_{\mathrm{Orig.}}$ and $\bar{Y}_{\mathcal{X}}$ agree with $1\,\sigma$ level.
1	**Only one** of $\bar{Y}_{\mathrm{Orig.}}$ or $\bar{Y}_{\mathcal{X}}$ comes into the other's $1\,\sigma$ error.
0	$\bar{Y}_{\mathrm{Orig.}}$ and $\bar{Y}_{\mathcal{X}}$ **do not agree with** $1\,\sigma$ level each other.

- **Evaluation 2**: $\sigma_{\mathcal{X}}/\sigma_{\mathrm{Orig.}}$ check ($\sigma_{\mathcal{X}}$: ML error, $\mathcal{X} = \mathcal{P}1$ or $\mathcal{P}2$)

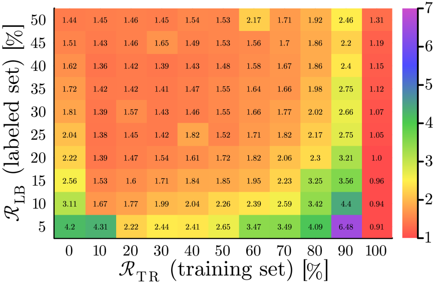
1. If a ML result got Score 2 at the Evaluation 1
2. and gave $\sigma_{\mathcal{X}}/\sigma_{\mathrm{Orig.}} \approx 1$ at the Evaluation 2

➡ ML estimation imitates the original CG result as well as possible.

Example: Plaquette $\to \mathrm{Tr}M^{-3}$ estimation ($\mathcal{P}1$, ID-0)

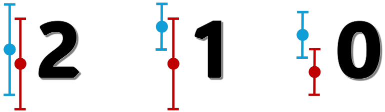


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

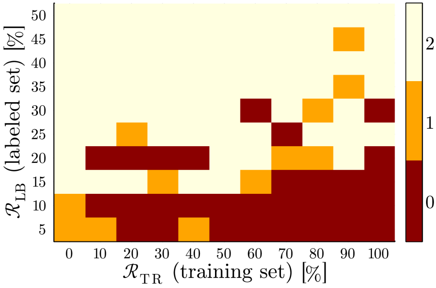
- $\bar{Y}$ score: white, orange, red



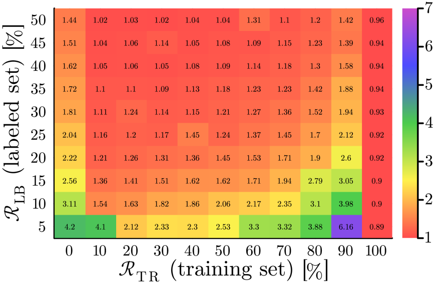
1 **Eval. 1:** central value check
➡ cannot find consistently white region (score 2).

2 **Eval. 2:** $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ check
➡ Roughly $\sigma_{\mathcal{P}1} \approx 1.5\sigma_{\mathrm{Orig.}}$ in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

Example: Plaquette → $\mathrm{Tr}M^{-3}$ estimation ($\mathcal{P}2$, ID-0)

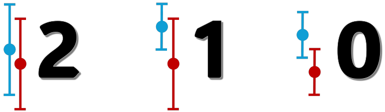


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

- $\bar{Y}$ score: white, orange, red



1 **Eval. 1:** central value check

➡ consistently white region

n $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

2 **Eval. 2:** $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$ check

➡ Roughly $\boldsymbol{\sigma_{\mathcal{P}2} \lesssim 1.1\sigma_{\mathrm{Orig.}}}$

in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

Preliminary results of ML estimation (1)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\boldsymbol{\mathcal{R}}_{\mathbf{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-0`: $\kappa = 0.13575$ (heaviest quark mass)

`ID-0`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 50}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**30**, 50}	{**35**, 40}	{**25**, 40}		
$\mathrm{Tr}M^{-3}$	{**30**, 50}	{**35**, 40}	{**40**, 40}	{**15**, 70}	
$\mathrm{Tr}M^{-4}$	{**45**, 60}	{**45**, 60}	{**40**, 40}	{**35**, 50}	{**25**, 40}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2, 3)$,
  - $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ for plaquette, $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{35}\,\%$ for Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{45}\,\%$

Preliminary results of ML estimation (2)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-2`: $\kappa = 0.13580$ (1st order phase transition)

`ID-2`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**10**, 80}	{**10**, 80}			
$\mathrm{Tr}M^{-2}$	{**10**, 90}	{**10**, 80}	{**10**, 90}		
$\mathrm{Tr}M^{-3}$	{**10**, 90}	{**10**, 70}	{**10**, 80}	{**10**, 90}	
$\mathrm{Tr}M^{-4}$	{**40**, 40}	{**40**, 80}	{**40**, 80}	{**40**, 80}	{**40**, 80}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2, 3)$,
  - $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{10}\,\%$ for plaquette and Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{40}\,\%$

Preliminary results of ML estimation (3)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{30, 50\}$ ➡ 👍 in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-4`: $\kappa = 0.13585$ (lightest quark mass)

`ID-4`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	$\{30, 40\}$	$\{35, 40\}$			
$\mathrm{Tr}M^{-2}$	$\{30, 50\}$	$\{40, 50\}$	$\{30, 40\}$		
$\mathrm{Tr}M^{-3}$	N.A.	N.A.	N.A.	N.A.	
$\mathrm{Tr}M^{-4}$	N.A.	N.A.	N.A.	N.A.	N.A.

- N.A. ➡ Cannot find proper $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\}$.
- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2)$,
  - $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$ for plaquette, $\mathcal{R}_{\mathrm{LB}} \geq 40\,\%$ for Polyakov loop

Summary and to-do list

- We checked the ability of ML estimation $X \to Y$ on
  1. $X =$ Plaquette/Polyakov loop, $Y = \mathrm{Tr} M^{-n}$ $(n = 1, 2, 3, 4)$
  2. $X = \mathrm{Tr} M^{-n}$, $Y = \mathrm{Tr} M^{-m}$ $(n < m)$

  using LightGBM (Microsoft) via `JuliaAI/MLJ.jl`.

- ML estimation works well with heavier quark mass (smaller κ).
  - Especially works well at the 1st order phase transition point.
- ML estimation with $X =$ Plaquette/Polyakov loop
  ➡ works well if labeled set were $\gtrsim$ 30% of total set
- In this preliminary analysis, we found that bias correction works well.

- Expand the analysis with other gauge ensembles.
  - different lattice volume $L^3 \times T$, different lattice spacing a.
- Further investigation and comparison with other linear regression algorithms such as LASSO or RIDGE.

Thank you for your listening!

Backup slides

Preliminary results of ML estimation (Backup, 1)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\boldsymbol{\mathcal{R}}_{\mathbf{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-0`: $\kappa = 0.13575$ (heaviest quark mass)

`ID-0`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 50}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**30**, 50}	{**35**, 40}	{**25**, 40}		
$\mathrm{Tr}M^{-3}$	{**30**, 50}	{**35**, 40}	{**40**, 40}	{**15**, 70}	
$\mathrm{Tr}M^{-4}$	{**45**, 60}	{**45**, 60}	{**40**, 40}	{**35**, 50}	{**25**, 40}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2, 3)$,
  - $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ for plaquette, $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{35}\,\%$ for Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{45}\,\%$

Preliminary results of ML estimation (Backup, 2)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-1`: $\kappa = 0.13577$

`ID-1`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 40}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**20**, 50}	{**35**, 40}	{**15**, 40}		
$\mathrm{Tr}M^{-3}$	{**25**, 50}	{**35**, 40}	{**30**, 60}	{**20**, 80}	
$\mathrm{Tr}M^{-4}$	{**45**, 40}	{**50**, 40}	{**45**, 50}	{**45**, 50}	{**25**, 50}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2, 3)$,
  - $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{30}\,\%$ for plaquette, $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{35}\,\%$ for Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{45}\,\%$

Preliminary results of ML estimation (Backup, 3)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-2`: $\kappa = 0.13580$ (1st phase transition)

`ID-2`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**10**, 80}	{**10**, 80}			
$\mathrm{Tr}M^{-2}$	{**10**, 90}	{**10**, 80}	{**10**, 90}		
$\mathrm{Tr}M^{-3}$	{**10**, 90}	{**10**, 70}	{**10**, 80}	{**10**, 90}	
$\mathrm{Tr}M^{-4}$	{**40**, 40}	{**40**, 80}	{**40**, 80}	{**40**, 80}	{**40**, 80}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2, 3)$,
  - $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{10}\,\%$ for plaquette and Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{40}\,\%$

Preliminary results of ML estimation (Backup, 4)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ ($n = 1, 2, 3$)
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ ($n = 1, 2, 3, 4$)
- $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-3`: $\kappa = 0.13582$

`ID-3`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**10**, 60}	{**20**, 80}			
$\mathrm{Tr}M^{-2}$	{**10**, 90}	{**10**, 60}	{**10**, 80}		
$\mathrm{Tr}M^{-3}$	{**10**, 70}	{**10**, 50}	{**15**, 60}	{**10**, 90}	
$\mathrm{Tr}M^{-4}$	{**40**, 50}	{**40**, 50}	{**40**, 50}	{**40**, 50}	{**40**, 50}

- In the $Y = \mathrm{Tr}M^{-n}$ estimation ($n = 1, 2, 3$),
  - $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{10}\,\%$ for plaquette and Polyakov loop
- Not so good at $Y = \mathrm{Tr}M^{-4}$ estimation: $\mathcal{R}_{\mathrm{LB}} \geq \mathbf{40}\,\%$

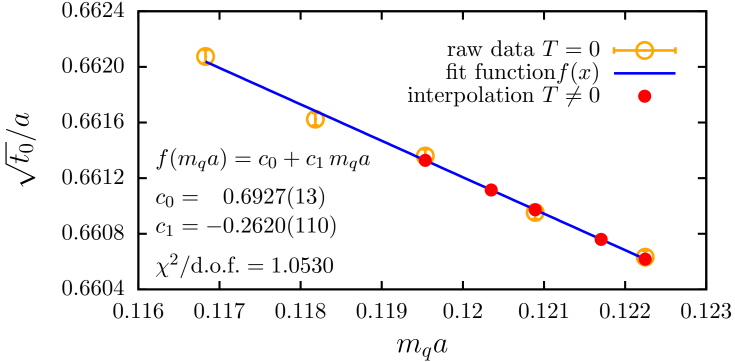
Preliminary results of ML estimation (tBackup, 5)

- For a ML trace estimation $X \to Y$, (X: input, Y: target)
- x-axis: X ➡ Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3)$
- y-axis: Y ➡ $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- $\{\boldsymbol{\mathcal{R}}_{\mathbf{LB}}, \mathcal{R}_{\mathrm{TR}}\} = \{\mathbf{30}, 50\}$ ➡ 👍 in $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ and $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$.
- $\mathcal{P}2$ results of `ID-4`: $\kappa = 0.13585$ (lightest quark mass)

`ID-4`	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 40}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**30**, 50}	{**40**, 50}	{**30**, 40}		
$\mathrm{Tr}M^{-3}$	N.A.	N.A.	N.A.	N.A.	
$\mathrm{Tr}M^{-4}$	N.A.	N.A.	N.A.	N.A.	N.A.

- N.A. ➡ Cannot find proper $\{\mathcal{R}_{\mathrm{LB}}, \mathcal{R}_{\mathrm{TR}}\}$.
- In the $Y = \mathrm{Tr}M^{-n}$ estimation $(n = 1, 2)$,
  - $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{30}\,\%$ for plaquette, $\boldsymbol{\mathcal{R}}_{\mathbf{LB}} \geq \mathbf{40}\,\%$ for Polyakov loop

Fit result of Wilson flow scale parameter at zero T



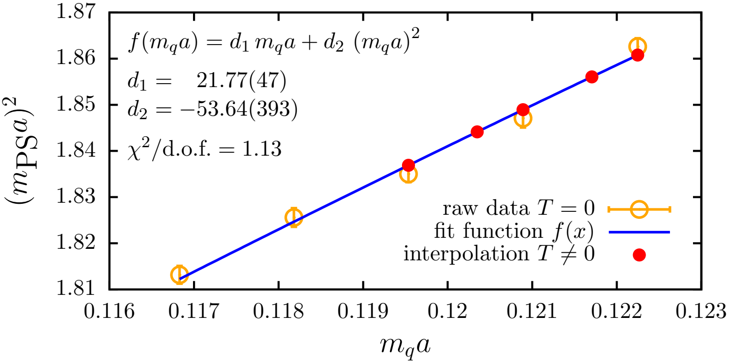
- $a \approx 0.22175$ fm ($\kappa = 0.13575$)
- $a \approx 0.22171$ fm ($\kappa = 0.13577$)
- $a \approx 0.22163$ fm ($\kappa = 0.13580$)
- $a \approx 0.22159$ fm ($\kappa = 0.13582$)
- $a \approx 0.22152$ fm ($\kappa = 0.13585$)

- $V = 16^3 \times 32$, $\beta = 1.60$
- Here, we use

$$\frac{1}{\sqrt{t_0}} = 1.347(30)\ \text{GeV}$$

BMW, JHEP **09** 010 (2012).

Fit result of pseudoscalar meson mass at zero T



- $m_{\mathrm{PS}} \approx 1.214$ GeV ($\kappa = 0.13575$)
- $m_{\mathrm{PS}} \approx 1.213$ GeV ($\kappa = 0.13577$)
- $m_{\mathrm{PS}} \approx 1.211$ GeV ($\kappa = 0.13580$)
- $m_{\mathrm{PS}} \approx 1.209$ GeV ($\kappa = 0.13582$)
- $m_{\mathrm{PS}} \approx 1.207$ GeV ($\kappa = 0.13585$)

- $V = 16^3 \times 32$, $\beta = 1.60$
- In summary, we have

$$m_{\mathrm{PS}} \approx 1.21 \text{ GeV}$$

in these 5 datasets.

Quark mass

- We obtain κ_c (kappa critical) from

$$\kappa_c(g_0^2) = 0.125 + 0.003681192\, g_0^2 + 0.000117\, (g_0^2)^2 + 0.000048\, (g_0^2)^3 - 0.000013\, (g_0^2)^4$$

where $g_0^2 = \frac{2N_c}{\beta}$ is the coupling constant.

- With $\beta = 1.60$, we have $\kappa_c \approx 0.14041$.

- We obtain quark mass $m_q a$ with

$$m_q a = \frac{1}{2}\left(\frac{1}{\kappa} - \frac{1}{\kappa_c}\right)$$

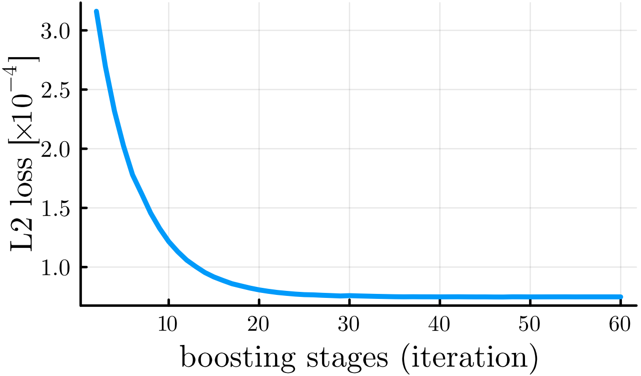
Computational gain

- By the measurement logs of `BQCD` program, we have ($N_{\text{conf}} = 5500$)

ID	κ	Total CPU-time [hour]	All CG [hour]	CG for single $\text{Tr}M^{-n}$ [hour]
0	0.13575	65.91	59.12	3.94
1	0.13577	65.97	59.15	3.94
2	0.13580	65.51	58.66	3.91
3	0.13582	64.44	57.56	3.84
4	0.13585	63.35	56.58	3.77

  - For single gauge conf., CG is called 150 times where $N_{\text{src}} = 10$.
  - For a *single* determination $\text{Tr}\, M^{-n}$, it takes $\approx$ 4 hours.
- For GBDT regression on `Julia` code (with ordinary laptop/desktop),
  - Time for model training $\approx$ 3.17 seconds (99.71% compilation time)
  - Time for $N_{\text{BS}} = 10000$ bootstrap resampling $\lesssim$ 1 minute
  - ➡ negligible comparing with CG time on supercomputer.
- (ex. 1) ML estimation on $\text{Tr}\, M^{-1} \to \text{Tr}\, M^{-2}$ with 30% labeled set,
  - ➡ we save $\approx$ 35 % of CG time.
- (ex. 2) ML estimation on Plaquette $\to \text{Tr}\, M^{-3}$ with 30% labeled set,
  - ➡ we save $\approx$ 70 % of CG time.

Optimal boosting stage

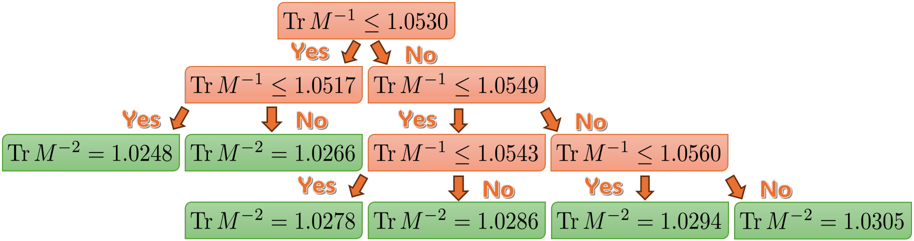


- x-axis: boosting stage (number of iteration)
- y-axis: L2-loss for machine learning $X \to Y$

$$(\text{L2 loss}) \equiv \sum_{i=1}^{N_{\text{TR}}} \left(Y_i - Y_i^P\right)^2$$

- optimal boosting stage ≈ **40**

Visualization of decision tree



- An example with $X = \operatorname{Tr} M^{-1}$, $Y = \operatorname{Tr} M^{-2}$ for $X \to Y$ learning
- Example for a boost stage.
- depth of tree $= 3$
- number of leaf (green cell, $Y = \operatorname{Tr} M^{-2}$) $= 6$

Optimal ratio of labeled/training set (backup)

- With $M = N_{LB} + N_{UL}$ and $N_{LB} = N_{TR} + N_{BC}$, let us assume that
  - $N_{LB} = 0.2M$ — labeled set is 20 % of total set.
  - $N_{TR} = 0.5N_{LB}$ — training set is 50 % of labeled set.
- Then the distribution of TR/BC/UL set is shown as, for example,

TR	UL	UL	UL	UL	BC	UL	UL	UL	UL	TR	UL	UL	UL	UL	BC	$\cdots$

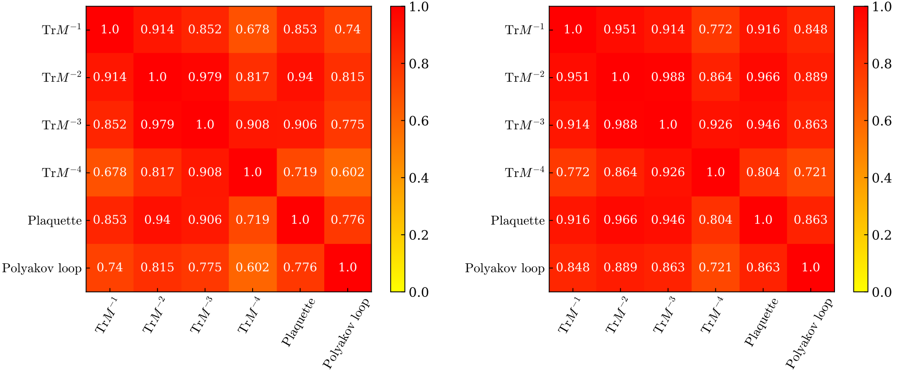
,

where x-axis corresponds to the gauge configuration index.

- To reduce the computational cost ➡ **fewest possible** labeled set.
  - In bias correction set, Y as well as Y^P is used to determine $\bar{Y}$.
  - ➡ We should grant **sufficient statistics** to bias correction set.

- ➡ Need to find **optimal ratio** of labeled/training set!

Correlation Map of Observables (`ID 0`, `ID 1`)

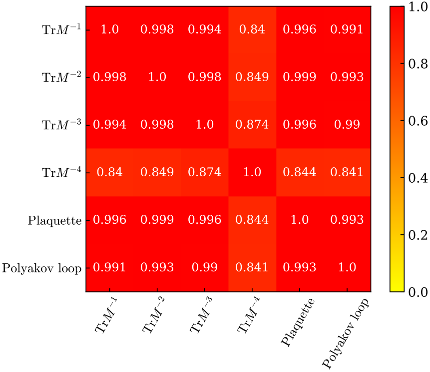


(a) $\kappa = 0.13575$, `ID 0`

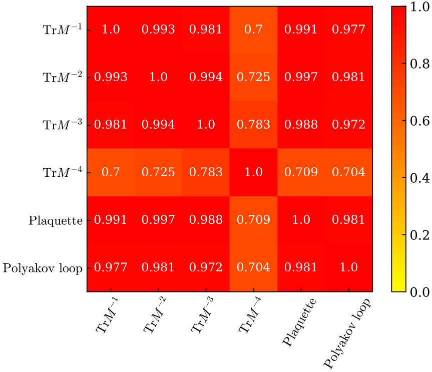
(b) $\kappa = 0.13577$, `ID 1`

- We evaluate the correlation coefficients for Plaquette, Polyakov loop and $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$

Correlation Map of Observables (ID 2, ID 3)



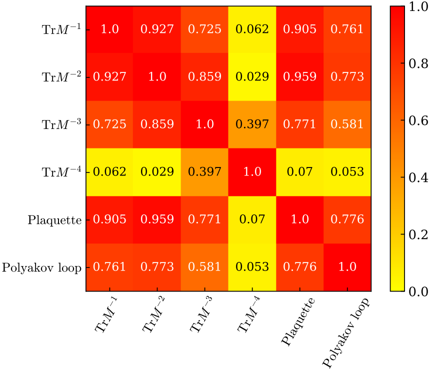
(c) $\kappa = 0.13580$, `ID 2`



(d) $\kappa = 0.13582$, `ID 3`

- We evaluate the correlation coefficients for Plaquette, Polyakov loop and $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$

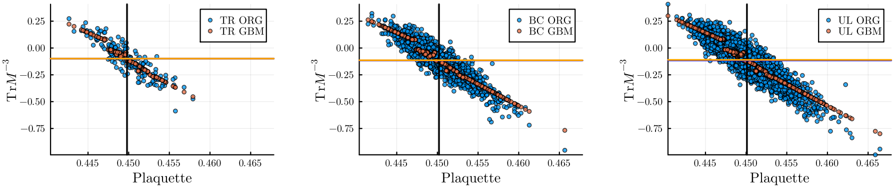
Correlation Map of Observables (ID 4)



(e) $\kappa = 0.13585$, `ID 4`

- We evaluate the correlation coefficients for Plaquette, Polyakov loop and $\mathrm{Tr}M^{-n}$ $(n = 1, 2, 3, 4)$
- We observe weak correlation between $\mathrm{Tr}M^{-4}$ and other observables.

Corr. Scatter of Plaquette → $\mathrm{Tr}M^{-3}$ estimation (ID-0)

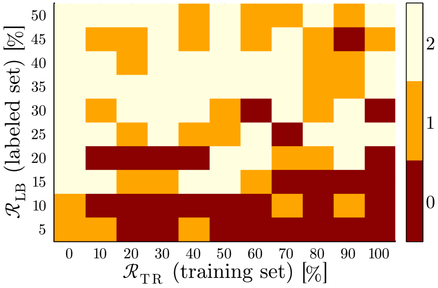


(a) {TR} (b) {BC} (c) {UL}

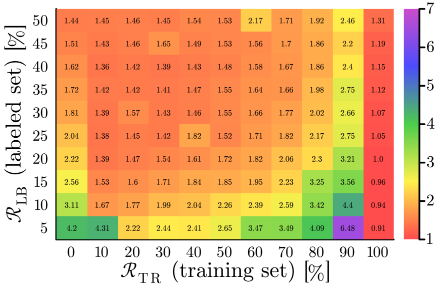
- x-axis: Plaquette
- y-axis: $\mathrm{Tr}\,M^{-3}$
- {TR}: Training set
- {BC}: Bias correction set
- {UL}: Unlabeled set
- $\mathcal{R}_{\mathrm{LB}} = 30\,\%$, $\mathcal{R}_{\mathrm{TR}} = 10\,\%$

- Blue circle: original data
- Orange circle: ML estimation
- Black vertical line:
  ➡ average of plaquette
- Yellow horizontal line:
  ➡ average of $\mathrm{Tr}\,M^{-3}$

Example: Plaquette → TrM^{-3} estimation ($\mathcal{P}1$, ID-0)

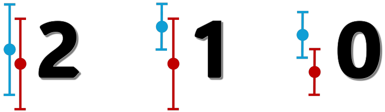


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

- $\bar{Y}$ score: white, orange, red



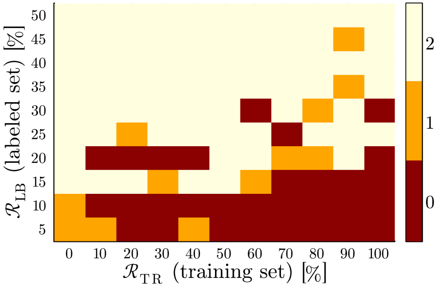
1 **Eval. 1:** central value check

➡ cannot find consistently white region (score 2).

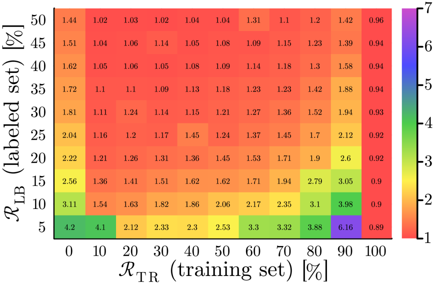
2 **Eval. 2:** $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ check

➡ Roughly $\sigma_{\mathcal{P}1} \approx 1.5\sigma_{\mathrm{Orig.}}$ in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

Example: Plaquette → $\mathrm{Tr} M^{-3}$ estimation ($\mathcal{P}2$, ID-0)

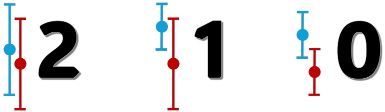


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

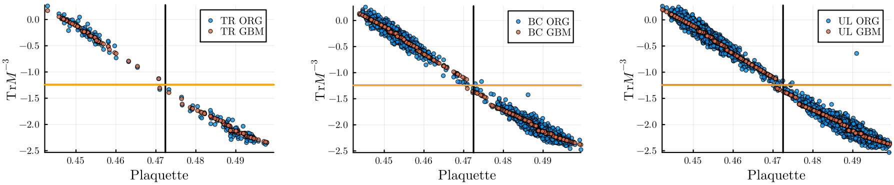
- $\bar{Y}$ score: white, orange, red



1 **Eval. 1:** central value check
➡ consistently white region in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

2 **Eval. 2:** $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$ check
➡ Roughly $\boldsymbol{\sigma_{\mathcal{P}2} \lesssim 1.1\sigma_{\mathrm{Orig.}}}$ in $\mathcal{R}_{\mathrm{LB}} \geq 30\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 50\,\%$

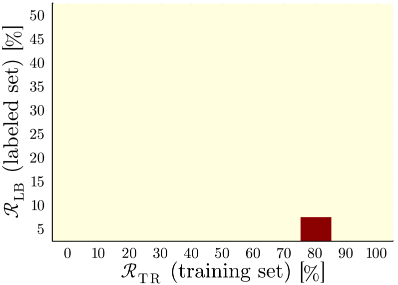
Corr. Scatter of Plaquette → $\mathrm{Tr}M^{-3}$ estimation (ID-2)



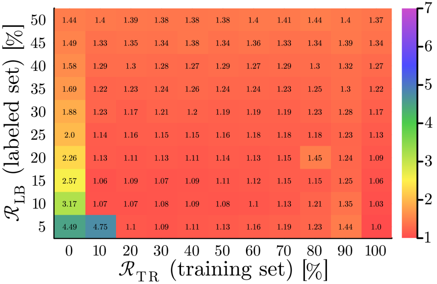
(a) {TR} (b) {BC} (c) {UL}

- x-axis: Plaquette
- y-axis: $\mathrm{Tr}\, M^{-3}$
- {TR}: Training set
- {BC}: Bias correction set
- {UL}: Unlabeled set
- $\mathcal{R}_{\mathrm{LB}} = 30\,\%$, $\mathcal{R}_{\mathrm{TR}} = 10\,\%$
- Blue circle: original data
- Orange circle: ML estimation
- Black vertical line:
  ➡ average of plaquette
- Yellow horizontal line:
  ➡ average of $\mathrm{Tr}\, M^{-3}$

Example: Plaquette $\to \mathrm{Tr}M^{-3}$ estimation ($\mathcal{P}1$, ID-2)

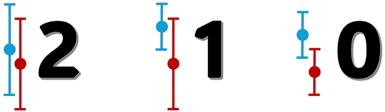


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

- $\bar{Y}$ score: white, orange, red



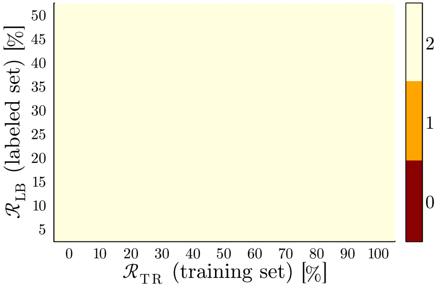
1 **Eval. 1:** central value check

➡ consistently white region in $\mathcal{R}_{\mathrm{LB}} \geq 10\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 90\,\%$

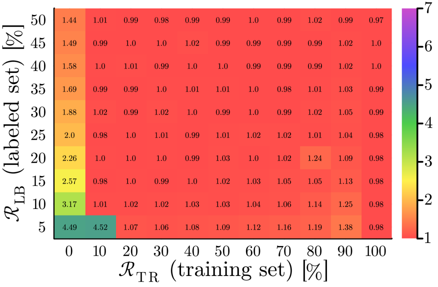
2 **Eval. 2:** $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ check

➡ Roughly $\sigma_{\mathcal{P}1} \approx 1.3\sigma_{\mathrm{Orig.}}$ in $\mathcal{R}_{\mathrm{LB}} \geq 10\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 70\,\%$

Example: Plaquette → $\mathrm{Tr} M^{-3}$ estimation ($\mathcal{P}2$, ID-2)

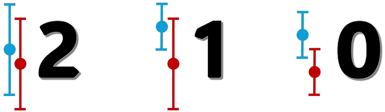


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

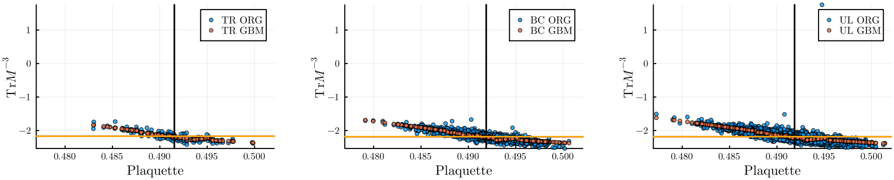
- $\bar{Y}$ score: white, orange, red



1 **Eval. 1:** central value check
➡ consistently white region in $\mathcal{R}_{\mathrm{LB}} \geq 5\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 90\,\%$

2 **Eval. 2:** $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$ check
➡ Roughly $\boldsymbol{\sigma_{\mathcal{P}2} \lesssim \sigma_{\mathrm{Orig.}}}$ in $\mathcal{R}_{\mathrm{LB}} \geq 10\,\%$, $\mathcal{R}_{\mathrm{TR}} \leq 70\,\%$

Corr. Scatter of Plaquette → $\mathrm{Tr} M^{-3}$ estimation (ID-4)

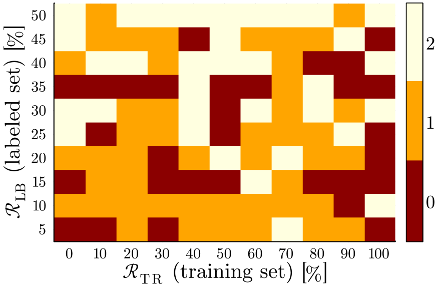


(a) {TR} (b) {BC} (c) {UL}

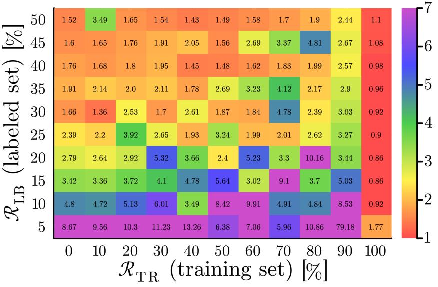
- x-axis: Plaquette
- y-axis: $\mathrm{Tr}\, M^{-3}$
- {TR}: Training set
- {BC}: Bias correction set
- {UL}: Unlabeled set
- $\mathcal{R}_{\mathrm{LB}} = 30\,\%$, $\mathcal{R}_{\mathrm{TR}} = 10\,\%$

- Blue circle: original data
- Orange circle: ML estimation
- Black vertical line:
  ➡ average of plaquette
- Yellow horizontal line:
  ➡ average of $\mathrm{Tr}\, M^{-3}$

Example: Plaquette → $\mathrm{Tr}M^{-3}$ estimation ($\mathcal{P}1$, ID-4)

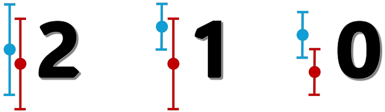


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

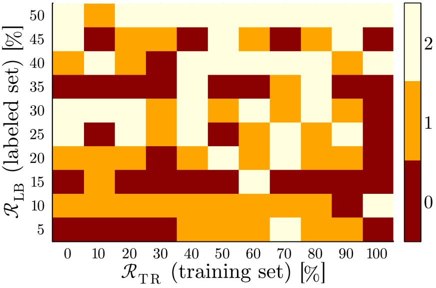
- $\bar{Y}$ score: white, orange, red



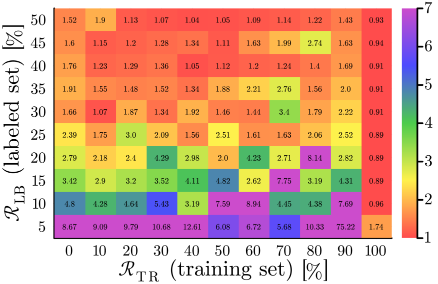
1 **Eval. 1:** central value check
➡ cannot find consistently white region (score 2).

2 **Eval. 2:** $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ check
➡ magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ is randomly distributed.

Example: Plaquette → $\mathrm{Tr}M^{-3}$ estimation ($\mathcal{P}2$, ID-4)

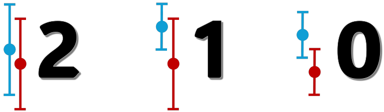


(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

- $\bar{Y}$ score: white, orange, red



1 **Eval. 1:** central value check
➡ cannot find consistently white region (score 2).

2 **Eval. 2:** $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$ check
➡ magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$ is randomly distributed.