Using AI for Efficient Statistical Inference of Lattice Correlators Across Mass Parameters

Aida El-Khadra, Andrew Lytle, Jiayu Shen, **Octavio Vega**

Department of Physics, University of Illinois, Urbana-Champaign

Background

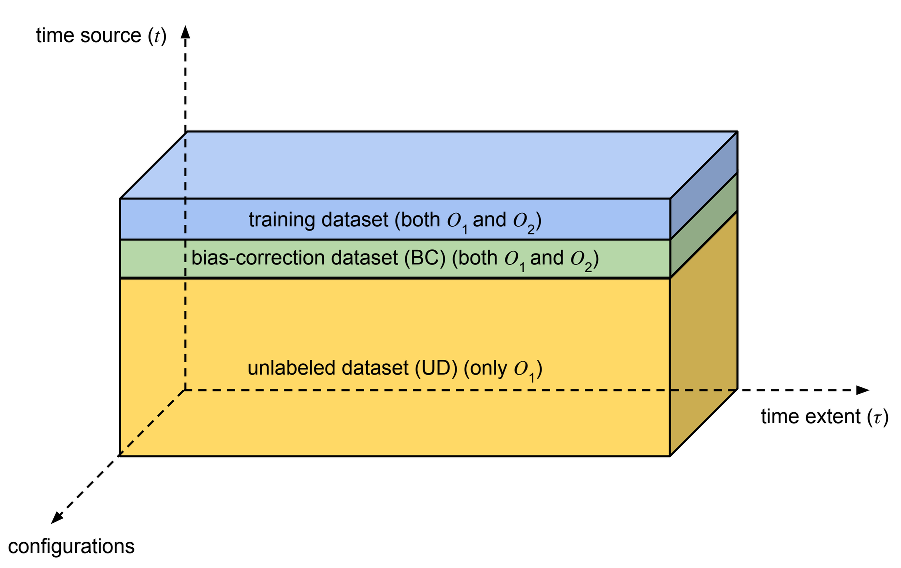
Lattice QCD is notorious for its computational expense. Modern simulations require large compute allocations on institutional clusters to handle the high number of Dirac operator inversions used to construct correlation functions. Machine learning (ML) techniques that can reduce the cost of correlator generations would therefore be beneficial for accelerating lattice field theory calculations. We apply supervised learning to infer two-point lattice correlation functions at different target masses. Our work proposes a new method for separating data into training and bias correction subsets for efficient uncertainty estimation. We also benchmark our ML models against a simple ratio method.

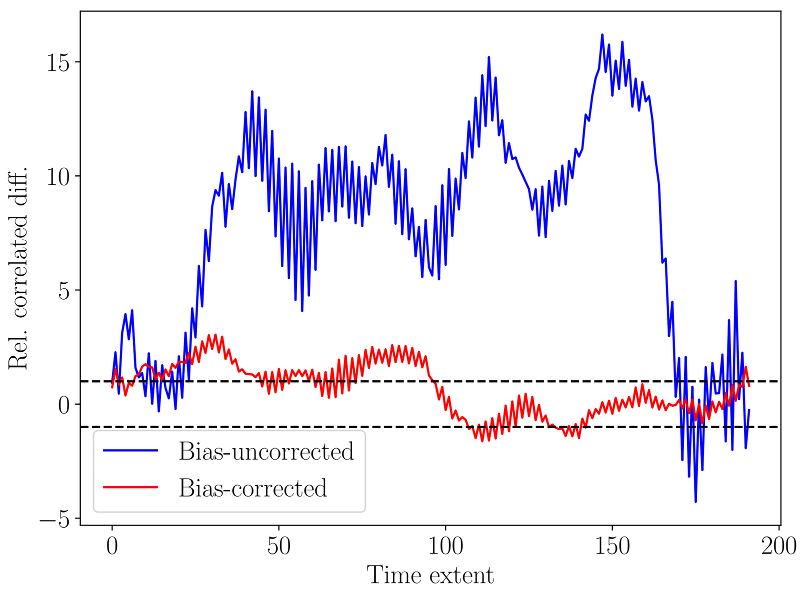
Our Method

Adapted from the ML estimator introduced in [1], our method uses the source times as indices to separate data into subsets for training, bias-correction, and prediction, thereby preserving the configuration axis to give unbiased configuration-wise predictions. The formula describing our procedure (for time extent τ and configuration i) is

$$\overline{O_i}(\tau) = \langle O_i(\tau)^{\text{pred}} \rangle_{\text{UD}} + \langle O_i(\tau) - O_i^{\text{pred}}(\tau) \rangle_{\text{BC}}. \tag{1}$$

A depiction of the data allocation is shown on the left below, and the right plot depicts the effect of bias-correction on the correlated difference of predicted correlation functions:





Key differences and benefits of our method compared to [1] and [2]:

1. Gives observable estimates at individual configurations, so uncertainties can be estimated without intensive bootstrapping or repetitive training.
2. All training source times are "seen" by our ML models simultaneously, and network weights are shared across source times for a greater chance to capture their correlations.

Dataset Information

Our dataset consists of meson two-point functions computed over a single gauge ensemble generated by the MILC collaboration [3], the parameters of which are tabulated below:

a [fm]	β	$N_s^3 \times N_t$	$N_{\text{src}} \times N_{\text{confs}}$	m_l/m_s	$(am_l)_{\text{sea}}$	$(am_s)_{\text{sea}}$	$(am_c)_{\text{sea}}$
0.042	7.00	$64^3 \times 192$	24×1028	1/5	0.00316	0.0158	0.188

The data was generated as a part of the FNAL-MILC collaboration's D- and B-meson semileptonic decay program [4]. The valence quark masses used are shown in the following table:

$(am_l)_{\text{valence}}$	$(am_s)_{\text{valence}}$	$(am_h)_{\text{valence}}$
0.00311	0.01555	{0.164, 0.1827, 0.365, 0.548, 0.731, 0.767 }

Correlation Functions

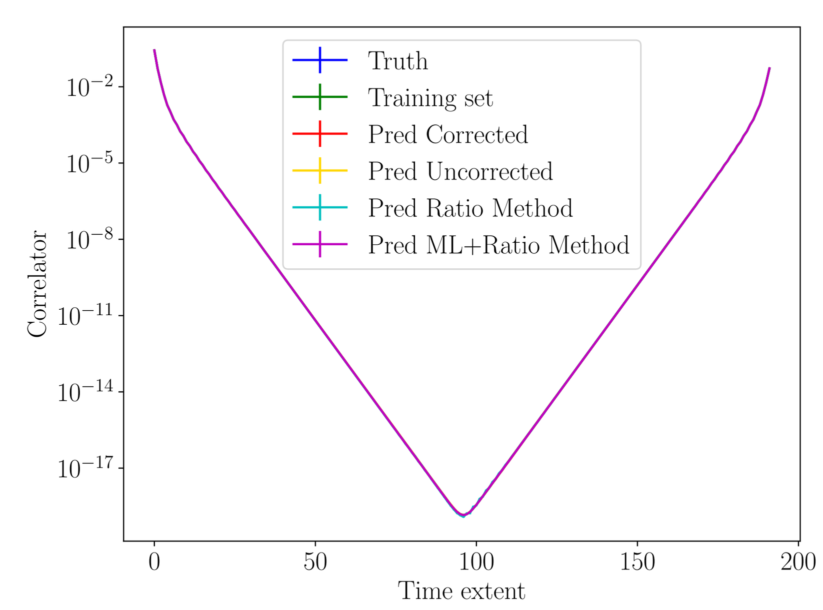
The correlation function representing the creation of an $H_{(s)}$ meson at source time t_0 and destruction at $t_0 + t$ is constructed as

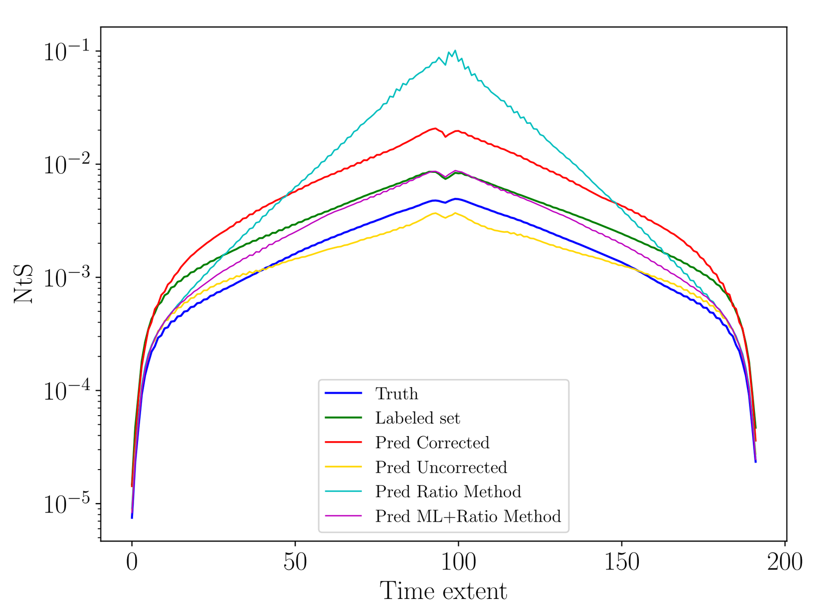
$$C^{2pt}_{H_{(s)}}(t) = \sum_{\mathbf{x},\mathbf{y}} \langle H_{(s)}(t+t_0, \mathbf{x}) H^\dagger_{(s)}(t_0, \mathbf{y}) \rangle \tag{2}$$

where $H_{(s)}$ is a local staggered interpolating operator with spin-taste structure $\gamma_5 \otimes \gamma_5$ that creates and destroys mesons with pseudoscalar quantum numbers at rest.
The correlation functions can be expressed as a spectral decomposition involving all energy eigenstates $|n\rangle$ that couple to the interpolating operator:

$$C^{2pt}_{H_{(s)}}(t) = \sum_{n=0} (-1)^{n(t+1)} \frac{|\langle 0|H_s^\dagger|n\rangle|^2}{2E_n} \left(e^{-E_n t} + e^{-E_n (T-t)} \right) \tag{3}$$

An example of predicted correlators and their noise-to-signal ratios are plotted below:





Important spectroscopic information and leptonic decay rates can be extracted by fitting the correlator data to the form (3) to determine the few lowest energy eigenvalues, E_n, and the overlap amplitudes $|\langle 0|H_s^\dagger|n\rangle|^2$. Obtaining accurate energies and amplitudes from inferred correlators thus forms a stringent test and benchmark for our inference methods.

Numerical Fit Results

We fit our correlator data to the spectral decomposition given in (3), using the `gvar` [5], `lsqfit` [6], and `corrfitter` [7] packages. The resulting values of amplitudes a_n and energy splitting dE_n for the ground and first excited states from bias-corrected ML predictions compared with results for truth-level data are seen below:

	Fit Parameters					
Method	a_0	dE_0	a_1	dE_1	χ^2/d.o.f.	Q
TRUTH	0.053791 (43)	0.398392 (50)	0.0689 (10)	0.1838 (21)	1.07	0.30
MLP	0.053825 (76)	0.398386 (78)	0.0712 (22)	0.1883 (44)	0.98	0.54
CNN	0.053906 (68)	0.398458 (75)	0.0740 (18)	0.1938 (36)	1.10	0.25
Transformer	0.053824 (84)	0.398377 (85)	0.0713 (23)	0.1885 (47)	0.95	0.61
Decision Tree	0.05382 (10)	0.39835 (10)	0.0693 (38)	0.1854 (72)	1.09	0.27

Ratio Method as a Benchmark

For two correlated observables O_1 and O_2, given O_1 computed on all configurations and O_2 computed on *some* configurations, we can estimate the value of O_2 with a lower uncertainty:

$$\langle O_2 \rangle_{\text{HP}} = \langle O_1 \rangle_{\text{HP}} \frac{\langle O_2 \rangle_{\text{LP}}}{\langle O_1 \rangle_{\text{LP}}}. \tag{4}$$

$\langle O \rangle_{\text{HP}} := (1/|\mathcal{S}_{\text{HP}}|) \Sigma_{i \in \mathcal{S}_{\text{HP}}} O(i)$ is the high-precision (HP) sum over a large subset $\mathcal{S}_{\text{HP}}$, and similarly for $\langle O \rangle_{\text{LP}}$ with $|\mathcal{S}_{\text{LP}}| < |\mathcal{S}_{\text{HP}}|$. We also use a "boosted" ratio method:

$$\langle O_2 \rangle_{\text{HP, boosted}} := \langle O_1 \rangle^\alpha_{\text{HP}} \frac{\langle O_2 \rangle_{\text{LP}}}{\langle O_1 \rangle^\alpha_{\text{LP}}}, \tag{5}$$

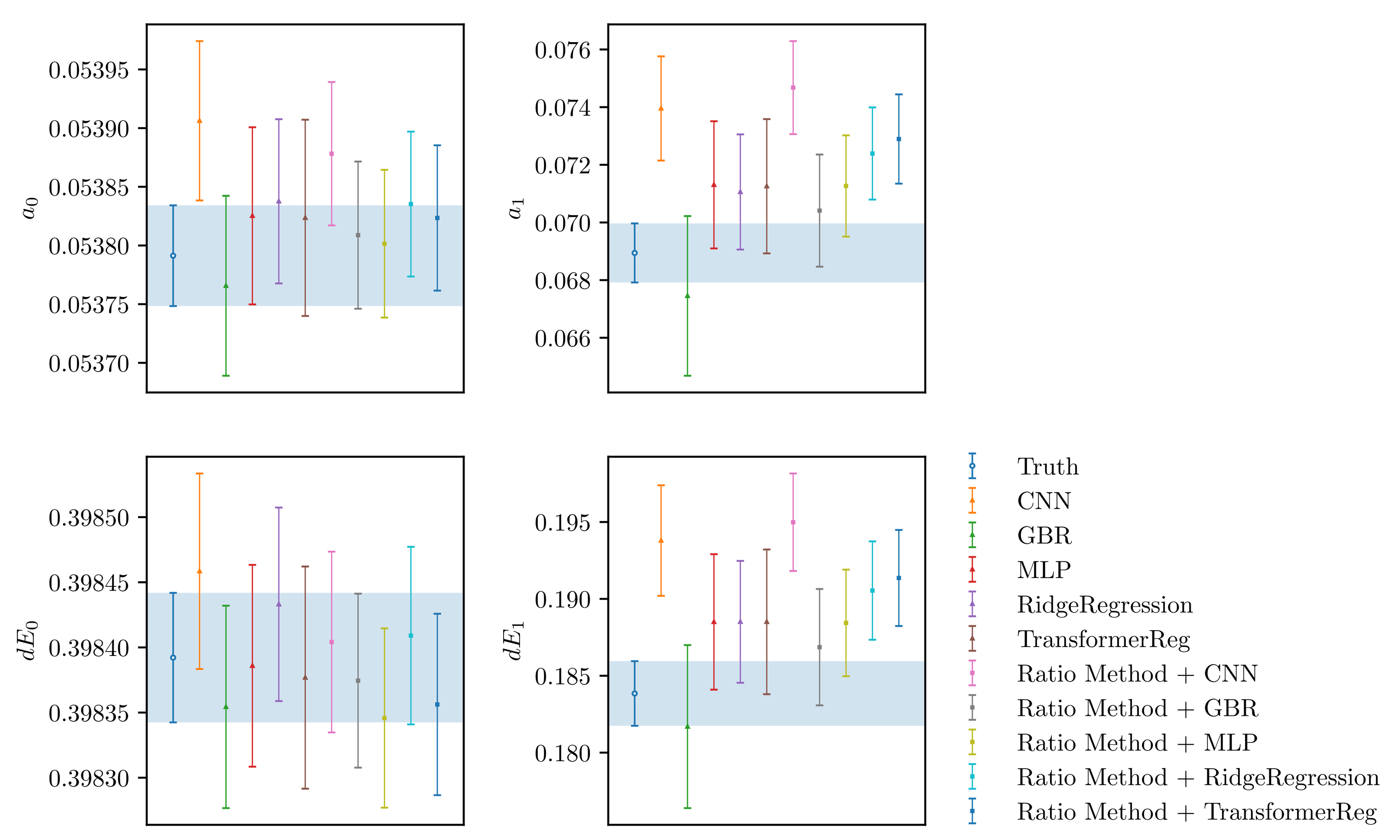
where α is a constant tailored to minimize the uncertainty of $\langle O_2 \rangle_{\text{HP, boosted}}$.
The (boosted) ratio method combined with ML ("(b)RM+ML") is formulated as

$$\langle O_2 \rangle_{\text{HP, (b)RM+ML}} := \langle O_1 \rangle^\alpha_{\text{HP}} \frac{\langle O_2^{\text{pred}} \rangle_{\text{LP}}}{\langle O_1^{\text{pred}} \rangle^\alpha_{\text{LP}}}. \tag{6}$$

We tabulate a comparison of the different ratio methods' performances below:

	Fit Parameters					
Method	a_0	dE_0	a_1	dE_1	χ^2/d.o.f.	Q
TRUTH	0.053791 (43)	0.398392 (50)	0.0689 (10)	0.1838 (21)	1.07	0.30
RM	0.05398(12)	0.39855(15)	0.0747(22)	0.1959(47)	0.91	0.73
bRM	0.053821(62)	0.398351(70)	0.0729(15)	0.1913(31)	1.15	0.14
RM + MLP	0.053811(64)	0.398353(70)	0.0723(17)	0.1903(34)	1.13	0.18
bRM + MLP	0.053802(63)	0.398346(69)	0.0713(18)	0.1884(35)	1.16	0.14

The first two amplitude and energy fit parameters are plotted below and compared for different bias-corrected ML models with ratio method combined with ML as well as with truth data.



References

[1] B. Yoon, T. Bhattacharya and R. Gupta, *Machine Learning Estimators for Lattice QCD Observables*, Phys. Rev. **D100** (2019) 014504 [1807.05971].

[2] J. Kim, G. Pederiva and A. Shindler, *Machine learning mapping of lattice correlated data*, Phys. Lett. **B856** (2024) 138894 [2402.07450].

[3] MILC collaboration, A. Bazavov *et al.*, *Lattice QCD Ensembles with Four Flavors of Highly Improved Staggered Quarks*, Phys. Rev. **D87** (2013) 054505 [1212.4768].

[4] Fermilab Lattice and MILC collaborations, A. Bazavov *et al.*, *D-meson semileptonic decays to pseudoscalars from four-flavor lattice QCD*, Phys. Rev. **D107** (2023) 094516 [2212.12648].

[5] G. Peter Lepage, *gvar* (Version 13.1), 2024.

[6] G. Peter Lepage, *lsqfit* (Version 13.2.2), 2024.

[7] G. Peter Lepage, *corrfitter* (Version 8.2), 2024.

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN