Operator Learning for spectral reconstruction in lattice QCD

Based on arXiv:2607.03311

Alessandro De Santis

July 29, 2026

B.S. = **Backup Slides**
clickable link

Introduction: spectral functions in Lattice QCD

- Spectral functions are key observables in hadron physics and are related to Euclidean correlators

$$C(an) = \int_0^\infty \mathrm{d}\omega\, e^{-\omega a n} \rho(\omega)$$

- In finite-volume lattice QCD spectral functions are distributions and **must be smeared**

$$\rho(\omega) = \sum_{n=1}^{\infty} c_n \delta\big(\omega - \omega_n\big)$$

$$\rho[K] = \int_0^\infty \mathrm{d}\omega\, K(\omega)\rho(\omega)$$

Introduction: spectral functions in Lattice QCD

- Spectral functions are key observables in hadron physics and are related to Euclidean correlators

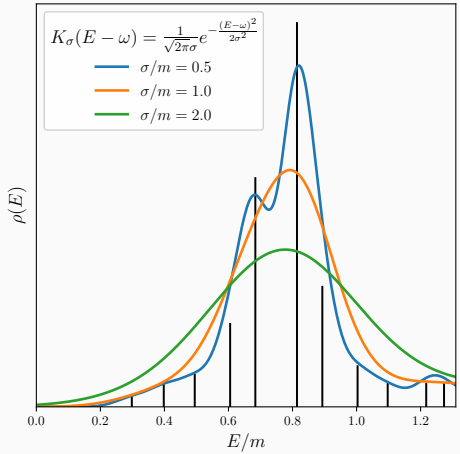
$$C(an) = \int_0^\infty \mathrm{d}\omega\, e^{-\omega a n} \rho(\omega)$$

- In finite-volume lattice QCD spectral functions are distributions and **must be smeared**

$$\rho(\omega) = \sum_{n=1}^{\infty} c_n \delta(\omega - \omega_n)$$

$$\rho[K] = \int_0^\infty \mathrm{d}\omega\, K(\omega)\rho(\omega)$$

smearing with Gaussian kernel of width σ



Introduction: methods

- The Hansen-Lupo-Tantalo (HLT) method is designed to extract smeared spectral functions from Euclidean correlators with **controlled systematic and statistical errors** and in a **model-independent** way

$$\rho[K] = \lim_{N\to\infty} \sum_{n=1}^{N} g_n \cdot C(an)$$

$$A[g] = \int_0^\infty \mathrm{d}\omega \left| K(\omega) - \sum_{n=1}^{N} g_n e^{-\omega a n} \right|^2$$

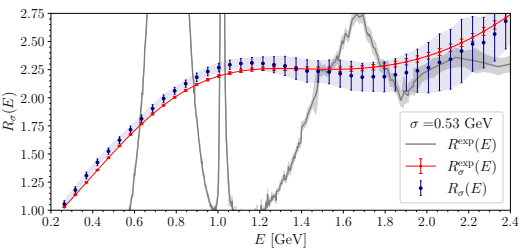
Introduction: methods

- The Hansen-Lupo-Tantalo (HLT) method is designed to extract smeared spectral functions from Euclidean correlators with **controlled systematic and statistical errors** and in a **model-independent** way

$$\rho[K] = \lim_{N\to\infty} \sum_{n=1}^{N} g_n \cdot C(an)$$

$$A[g] = \int_0^\infty \mathrm{d}\omega \left| K(\omega) - \sum_{n=1}^{N} g_n e^{-\omega a n} \right|^2$$

smeared R-ratio PRL 130 (2023) 24, 241901

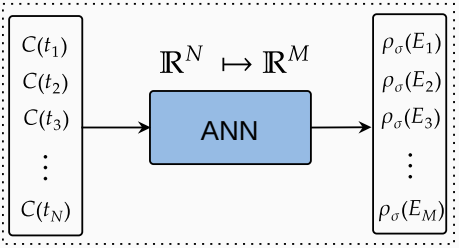


- F. Margari's talk for update on $R(E)$
- A. De Santis's talk for $\bar{B}_s \to X_{\bar{s}c}\ell\bar{\nu}$
- S. Hashimoto's talk for a list of talks related to spectral reconstruction
- We're open to disclose correlators and algorithms for benchmarks, talk to us!

Introduction: methods

What about machine learning?

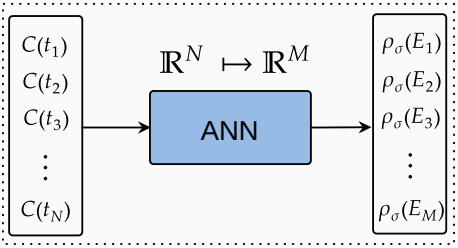
- In EPJC 84 (2024) 1, 32 together with M. Buzzicotti and N. Tantalo I devised a **model-independent** machine-learning method with a solid **estimate of systematic errors**
- 😊 Same performance as HLT
- 😡 **re-training** for different σ and energies



Introduction: methods

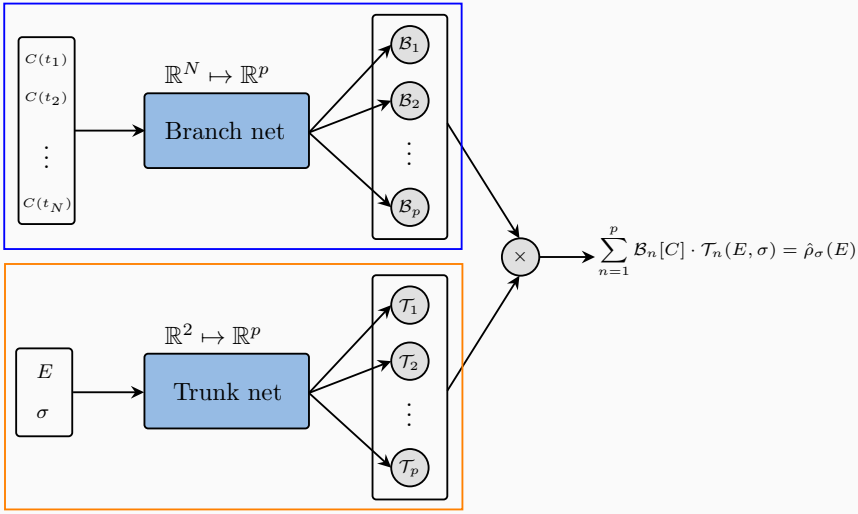
What about machine learning?

- In EPJC 84 (2024) 1, 32 together with M. Buzzicotti and N. Tantalo I devised a **model-independent** machine-learning method with a solid **estimate of systematic errors**
- 😊 Same performance as HLT
- 😡 **re-training** for different σ and energies



- In **Operator Learning** the network learns a **map between function spaces**
- based on Deep Operator Network (DeepONet) introduced in Nature MI 3, 218–229 (2021)

DeepONet architecure



Comparison between methods

HLT

$$\hat{\rho}_\sigma(E) = \sum_{n=1}^{N} C(an) \cdot g_n(E, \sigma)$$

- N is limited by the number of time slices
- linear in the correlator
- basis of coefficients given by exponentials
- $g_n(E, \sigma) = \mathcal{O}(10^N) \implies$ explicit regulator

DeepONet

$$\hat{\rho}_\sigma(E) = \sum_{n=1}^{p} \mathcal{B}_n[C] \cdot \mathcal{T}_n(E, \sigma)$$

- p can be arbitrarily large
- highly non-linear in the correlator
- basis of coefficients given by architecture
- $\mathcal{T}_n(E, \sigma) = \mathcal{O}(1) \implies$ implicit regulator

Comparison between methods

HLT

$$\hat{\rho}_\sigma(E) = \sum_{n=1}^{N} C(an) \cdot g_n(E, \sigma)$$

- N is limited by the number of time slices
- linear in the correlator
- basis of coefficients given by exponentials
- $g_n(E, \sigma) = \mathcal{O}(10^N) \implies$ explicit regulator

DeepONet

$$\hat{\rho}_\sigma(E) = \sum_{n=1}^{p} \mathcal{B}_n[C] \cdot \mathcal{T}_n(E, \sigma)$$

- p can be arbitrarily large
- highly non-linear in the correlator
- basis of coefficients given by architecture
- $\mathcal{T}_n(E, \sigma) = \mathcal{O}(1) \implies$ implicit regulator

valididation and benchmark in a controlled setting

Training set incorporating prior physical knowledge

- **benchmark model**: equivalent of R-ratio for the two-dimensional O(3) non-linear σ-model
- the model is **exactly integrable**

$$\rho^{\mathrm{O}(3)}(E) = \theta(E - 2m)\rho_2^{\mathrm{O}(3)}(E) + \sum_{n=1}^{\infty} \rho_{2+2n}^{\mathrm{O}(3)}(E)$$

Training set incorporating prior physical knowledge

- **benchmark model**: equivalent of R-ratio for the two-dimensional O(3) non-linear σ-model
- the model is **exactly integrable**

$$\rho^{\mathrm{O}(3)}(E) = \theta(E-2m)\rho_2^{\mathrm{O}(3)}(E) + \sum_{n=1}^{\infty} \rho_{2+2n}^{\mathrm{O}(3)}(E)$$

- Gaussian Processes (GP) generalize the concept of probability to space of functions

$$\rho(E) \sim \mathcal{GP}\left(\rho_2^{\mathrm{O}(3)}(E), \mathcal{K}(E,E')\right)$$

- with **covariance kernel**

$$\mathcal{K}(E,E') = \sigma_{\mathrm{GP}}^2 \exp\left(-\frac{(E-E')^2}{2\ell^2}\right)$$

Training set incorporating prior physical knowledge

- **benchmark model**: equivalent of R-ratio for the two-dimensional O(3) non-linear σ-model
- the model is **exactly integrable**

$$\rho^{\mathrm{O}(3)}(E) = \theta(E-2m)\rho_2^{\mathrm{O}(3)}(E) + \sum_{n=1}^{\infty} \rho_{2+2n}^{\mathrm{O}(3)}(E)$$

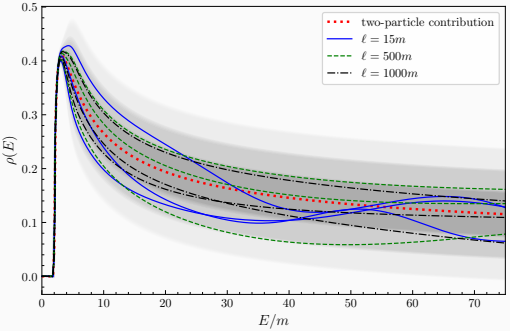
- Gaussian Processes (GP) generalize the concept of probability to space of functions

$$\rho(E) \sim \mathcal{GP}\left(\rho_2^{\mathrm{O}(3)}(E), \mathcal{K}(E,E')\right)$$

- with **covariance kernel**

$$\mathcal{K}(E,E') = \sigma_{\mathrm{GP}}^2 \exp\left(-\frac{(E-E')^2}{2\ell^2}\right)$$

- correlation length ℓ determines the smoothness
- σ_{GP}^2 determines the GP variance



Training set incorporating prior physical knowledge

- **benchmark model**: equivalent of R-ratio for the two-dimensional O(3) non-linear σ-model
- the model is **exactly integrable**

$$\rho^{\mathrm{O}(3)}(E) = \theta(E-2m)\rho_2^{\mathrm{O}(3)}(E) + \sum_{n=1}^{\infty} \rho_{2+2n}^{\mathrm{O}(3)}(E)$$

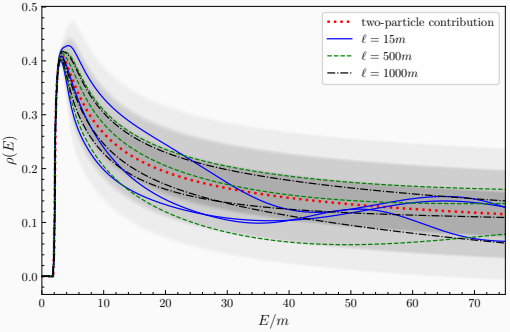
- Gaussian Processes (GP) generalize the concept of probability to space of functions

$$\rho(E) \sim \mathcal{GP}\left(\rho_2^{\mathrm{O}(3)}(E), \mathcal{K}(E,E')\right)$$

- with **covariance kernel**

$$\mathcal{K}(E,E') = \sigma_{\mathrm{GP}}^2 \exp\left(-\frac{(E-E')^2}{2\ell^2}\right)$$

- correlation length ℓ determines the smoothness
- σ_{GP}^2 determines the GP variance



- $\ell \to 0$ and $\sigma_{\mathrm{GP}} \to \infty$: **model independence** (but not in this work)

Training set incorporating prior physical knowledge

- Extract N_ρ random functions $\rho(E)$ from the GP

  input $C(t) = \int_0^\infty \mathrm{d}\omega\, e^{-\omega t}\rho(\omega) + \delta^{\mathrm{noise}}(t)$

  output $\rho_\sigma^{\mathrm{true}}(E) = \int_0^\infty \mathrm{d}\omega\, K_\sigma(E-\omega)\rho(\omega)$

- for fixed grid of times Ω_t and random (E, σ)

Training set incorporating prior physical knowledge

- Extract N_ρ random functions $\rho(E)$ from the GP

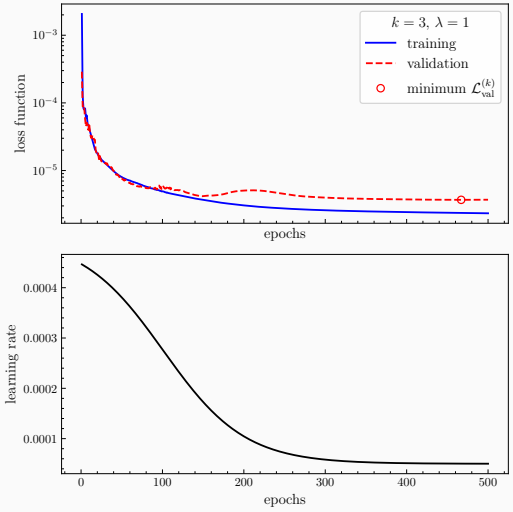
  input $C(t) = \int_0^\infty \mathrm{d}\omega\, e^{-\omega t}\rho(\omega) + \delta^{\mathrm{noise}}(t)$

  output $\rho_\sigma^{\mathrm{true}}(E) = \int_0^\infty \mathrm{d}\omega\, K_\sigma(E-\omega)\rho(\omega)$

- for fixed grid of times Ω_t and random (E, σ)

- **Supervised** training using MSE

$$\mathcal{L} = \frac{1}{N_\rho}\sum_{n=1}^{N_\rho}\left(\rho_{\sigma,n}^{\mathrm{output}}(E) - \rho_{\sigma,n}^{\mathrm{true}}(E)\right)^2$$



Ensembles of neural networks

			Branch Net		Trunk Net			$\lambda = 1$	
k	$N_\rho \times 10^{-5}$	p	layers	neurons	layers	neurons	params	$\mathcal{L}_{\text{val}}^{(k)} \times 10^6$	$w^{(k)}$
1	1	16	2	64	2	64	45216	3.8	0.072
2	1	32	2	64	2	64	47296	3.8	0.072
3	1	64	2	64	2	64	51456	3.7	0.073
4	1	128	2	64	2	64	59776	4.0	0.071
5	1	64	2	128	2	128	119168	4.8	0.065
6	1	64	2	256	2	256	303744	4.7	0.066
7	2	64	2	64	2	64	51456	3.3	0.076
8	4	64	2	64	2	64	51456	2.6	0.081
9	8	64	2	64	2	64	51456	2.3	0.083
10	1	16	3	64	3	64	53536	4.6	0.066
11	1	32	3	64	3	64	55616	4.3	0.069
12	1	64	3	64	3	64	59776	4.5	0.067
13	1	128	3	64	3	64	68096	4.2	0.069
14	1	64	1	64	1	64	43136	9.5	0.041
15	1	128	1	64	1	64	51456	12.1	0.031

- **Key idea**: combine different predictions to estimate the neural network error

Systematic errors and closure test

- central value from weighted average

$$\hat{\rho}_\sigma(E) = \sum_{k=1}^{N_{\text{nets}}} w^{(k)}\, \hat{\rho}_\sigma^{(k)}(E)$$

- systematic error from weighted spread

$$\Delta_\sigma^{\text{net}}(E) = \sqrt{\sum_{k=1}^{N_{\text{nets}}} w^{(k)} \left(\hat{\rho}_\sigma^{(k)}(E) - \hat{\rho}_\sigma(E)\right)^2}$$

- total reconstruction error

$$\Delta_\sigma^{\text{rec}}(E) = \sqrt{\left(\Delta_\sigma^{\text{stat}}(E)\right)^2 + \left(\Delta_\sigma^{\text{net}}(E)\right)^2}$$

Systematic errors and closure test

- central value from weighted average

$$\hat{\rho}_\sigma(E) = \sum_{k=1}^{N_{\text{nets}}} w^{(k)}\, \hat{\rho}_\sigma^{(k)}(E)$$

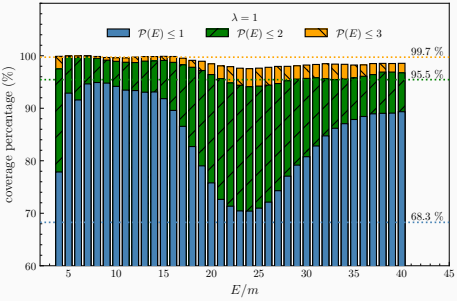
- systematic error from weighted spread

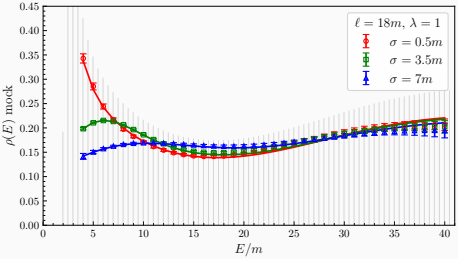
$$\Delta_\sigma^{\text{net}}(E) = \sqrt{\sum_{k=1}^{N_{\text{nets}}} w^{(k)} \left(\hat{\rho}_\sigma^{(k)}(E) - \hat{\rho}_\sigma(E)\right)^2}$$

- total reconstruction error

$$\Delta_\sigma^{\text{rec}}(E) = \sqrt{\left(\Delta_\sigma^{\text{stat}}(E)\right)^2 + \left(\Delta_\sigma^{\text{net}}(E)\right)^2}$$

- **closure test** with 1000 unseen mock data





Benchmark on two-dimensional O(3) non-linear σ-model

Correlators from JHEP 07 (2022) 034

ID	$(L/a) \times (T/a)$	β	am	mL	mT
A1	640×320	1.63	0.0447967(66)	29	14
A2	1280×640	1.72	0.0257692(35)	33	17
A4	2880×1440	1.85	0.0112591(21)	32	16
B1	5760×1440	1.85	0.0112601(75)	65	16
B2	2880×2880	1.85	0.0112463(82)	32	32

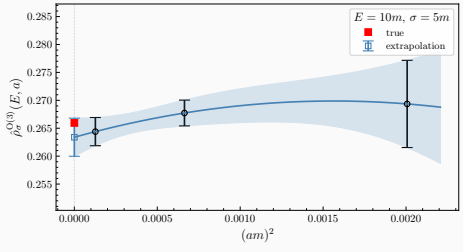
- A1 and A2 correlators interpolated at Ω_t (B.S.)
- B1 and B2 for finite-size effects (B.S.)
- A1, A2 and A4 for **continuum limit**

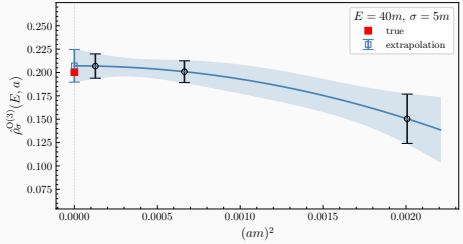
Benchmark on two-dimensional O(3) non-linear σ-model

Correlators from JHEP 07 (2022) 034

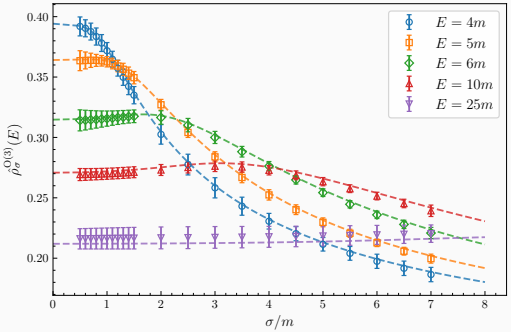
ID	$(L/a) \times (T/a)$	β	am	mL	mT
A1	640×320	1.63	0.0447967(66)	29	14
A2	1280×640	1.72	0.0257692(35)	33	17
A4	2880×1440	1.85	0.0112591(21)	32	16
B1	5760×1440	1.85	0.0112601(75)	65	16
B2	2880×2880	1.85	0.0112463(82)	32	32

- A1 and A2 correlators interpolated at Ω_t (B.S.)
- B1 and B2 for finite-size effects (B.S.)
- A1, A2 and A4 for **continuum limit**



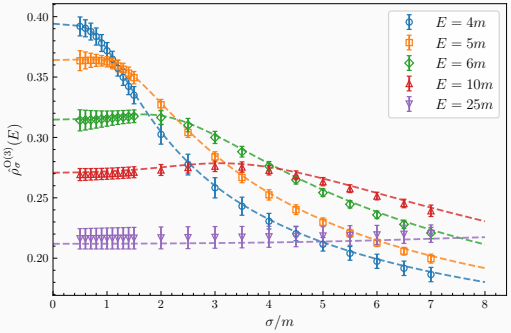


Removal of the smearing parameter



- dependence on σ is steep near resonances and multi-particle thresholds

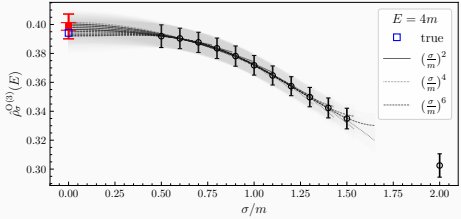
Removal of the smearing parameter

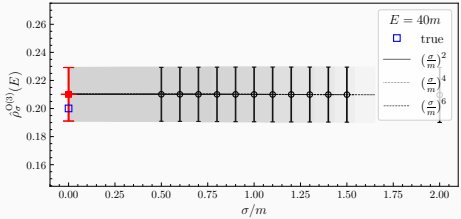


- dependence on σ is steep near resonances and multi-particle thresholds

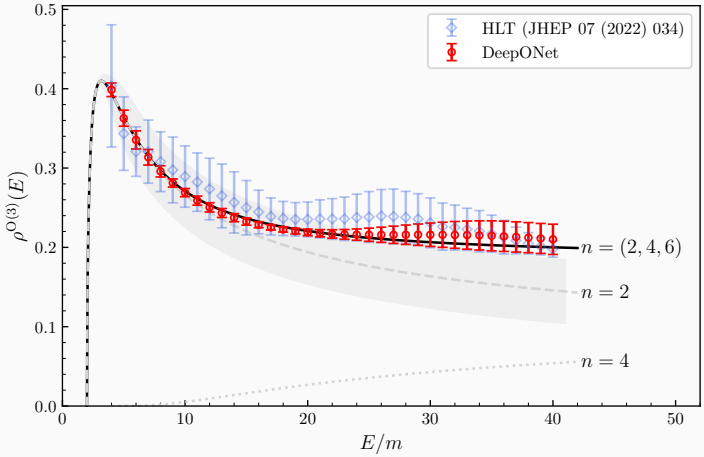
- **only after** infinite-volume limit,

$$\rho(E) = \lim_{\sigma \to 0} \rho_\sigma(E)$$





Final results and comparison with HLT



- perfect agreement with the theoretical result including two-,four- and six-particle states
- Much more precise than HLT $\rightarrow$ **dependence on prior information to be assessed**

Outlook: the R-ratio

The smeared R-ratio

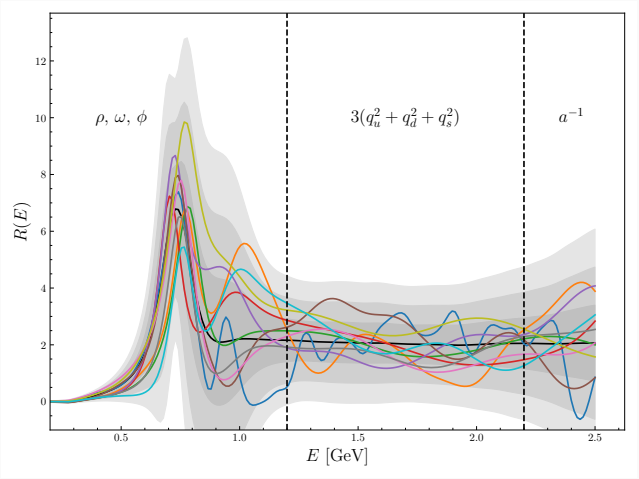
- Phenomenologically important

$$R(E) = \frac{\sigma(e^+e^- \to X)}{\sigma(e^+e^- \to \mu^+\mu^-)}$$

- Associated with $C(t) = \langle V_i(t) V_i(0) \rangle$

- physics-informed training set with different levels of smoothness

- this is **not a family of parametric functions!**



The smeared R-ratio

- Simulations at physical point
- Two discretizations for $V_\mu(x)$ with the same continuum limit
- Gaussian Negative Log Likelihood loss to predict the error

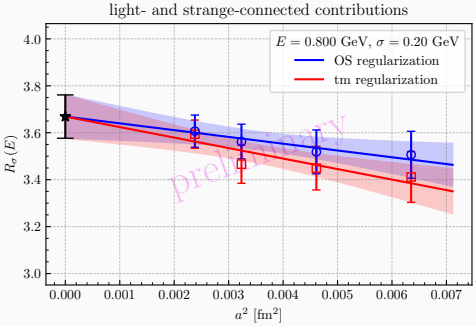
$$\mathcal{L} \simeq \frac{1}{2}\left[\frac{(\rho_\sigma^{\text{pred}}(E) - \rho_\sigma^{\text{true}}(E))^2}{\Delta^2} + \log \Delta^2\right]$$

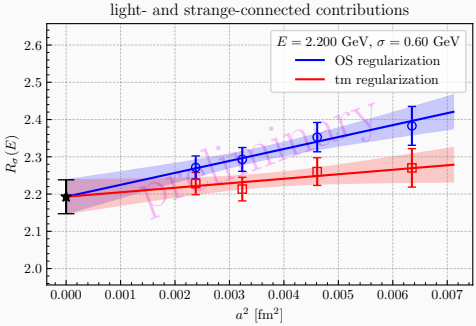
ETMC ensembles ($M_\pi \sim M_\pi^{\text{phys}}$)

ID	$L \times T$	a [fm]	L [fm]
B48	48×96	0.080	3.82
B64	64×128	0.080	5.09
B96	96×192	0.080	7.63
C80	80×160	0.068	5.46
D96	96×192	0.057	5.46
E112	112×224	0.049	5.48

M. Garofalo and A. Evangelista's talks

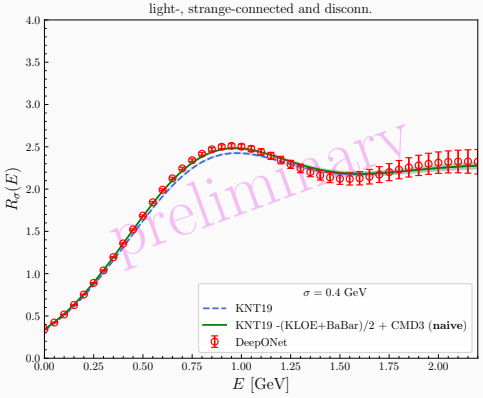
The smeared R-ratio: continuum limits

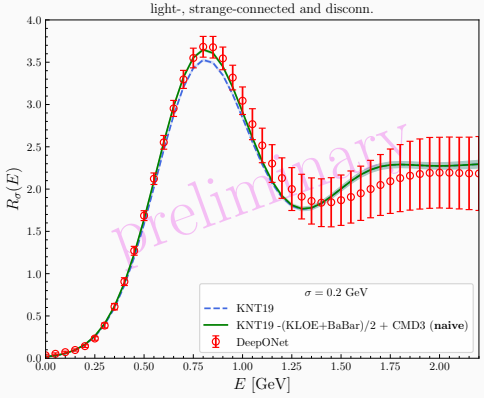




- **Independent trainings** and correlators from **independent Monte Carlo simulations**
- Excellent controll over continuum limits

The smeared R-ratio: confront experiments





- permille precision for $\sigma \geq 0.4$ GeV
- More precise than HLT (see F. Margari's talk on Friday)
- Large discrepancy from KNT19 but below one when replacing BaBar/KLOE with CMD3
- **no free lunch**: errors increase consistently when σ is reduced

Conclusions

- The operation of extracting **smeared spectral functions** from Euclidean correlators can be successfully rephrased in the **Operator Learning** framework
- In practice this can be done using **DeepONets** and the systematic errors can be reliably estimated using and ensemble of networks
- The method is **benchmarked** by reconstructing the spectral function in the two-dimensional O(3) non-linear σ-model with final agreement with the theoretical result
- Preliminary studies reveals that the smeared R-ratio can be calcualted with unprecedented precision

Conclusions

- The operation of extracting **smeared spectral functions** from Euclidean correlators can be successfully rephrased in the **Operator Learning** framework
- In practice this can be done using **DeepONets** and the systematic errors can be reliably estimated using and ensemble of networks
- The method is **benchmarked** by reconstructing the spectral function in the two-dimensional O(3) non-linear σ-model with final agreement with the theoretical result
- Preliminary studies reveals that the smeared R-ratio can be calcualted with unprecedented precision

Thank you for the attention!

Backup Slides (B.S.)

Finite size effects

- Finite size effects for smeared quantities are exponentially suppressed and expect to be, in general, small
- I checked for residual FSEs

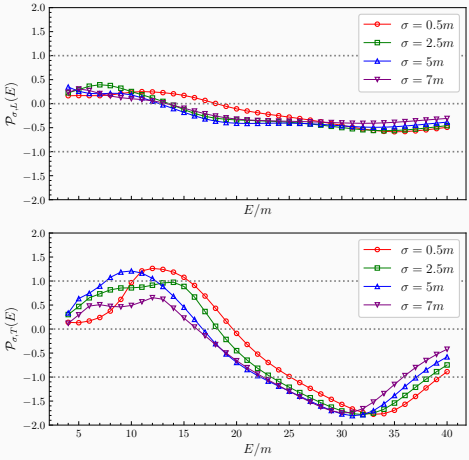
$$\mathcal{P}_{\sigma,L}(E) = \frac{\hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{B1}] - \hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{A4}]}{\sqrt{\left(\Delta_\sigma^{\mathrm{rec}}(E)[\mathrm{B1}]\right)^2 + \left(\Delta_\sigma^{\mathrm{rec}}(E)[\mathrm{A4}]\right)^2}}$$

$$\mathcal{P}_{\sigma,T}(E) = \frac{\hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{B2}] - \hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{A4}]}{\sqrt{\left(\Delta_\sigma^{\mathrm{rec}}(E)[\mathrm{B2}]\right)^2 + \left(\Delta_\sigma^{\mathrm{rec}}(E)[\mathrm{A4}]\right)^2}}$$

- Treated as systematic errors

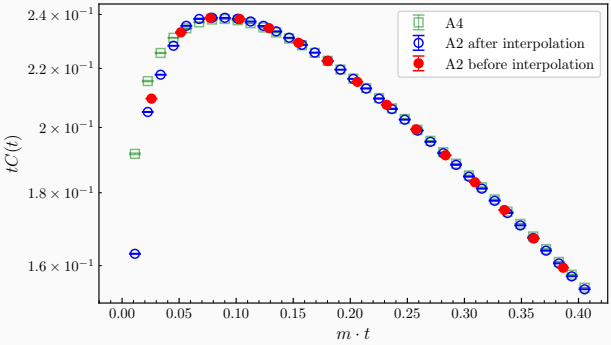
$$\Delta_\sigma^L(E) = \left|\hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{B1}] - \hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{A4}]\right| \mathrm{erf}\left(\frac{\mathcal{P}_{\sigma,L}(E)}{\sqrt{2}}\right)$$

$$\Delta_\sigma^T(E) = \left|\hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{B2}] - \hat{\rho}_\sigma^{\mathrm{O}(3)}(E)[\mathrm{A4}]\right| \mathrm{erf}\left(\frac{\mathcal{P}_{\sigma,T}(E)}{\sqrt{2}}\right)$$



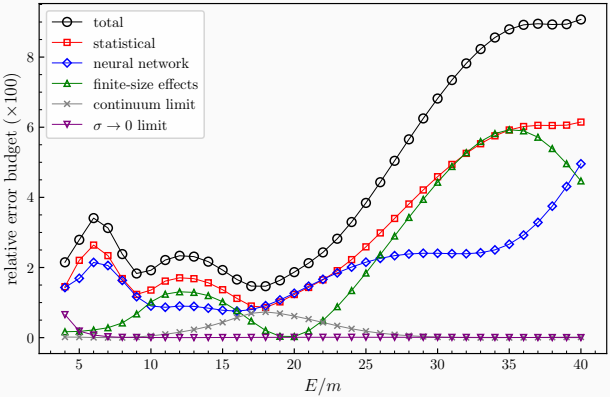
Correlator interpolation

- Interpolation is needed to match the grid of times used during training
- Cubic-spline ansatz to get a smooth interpolation
- Discretizations effects are well present in the data
- The interpolated correlator does not contain more physical information than the original one!

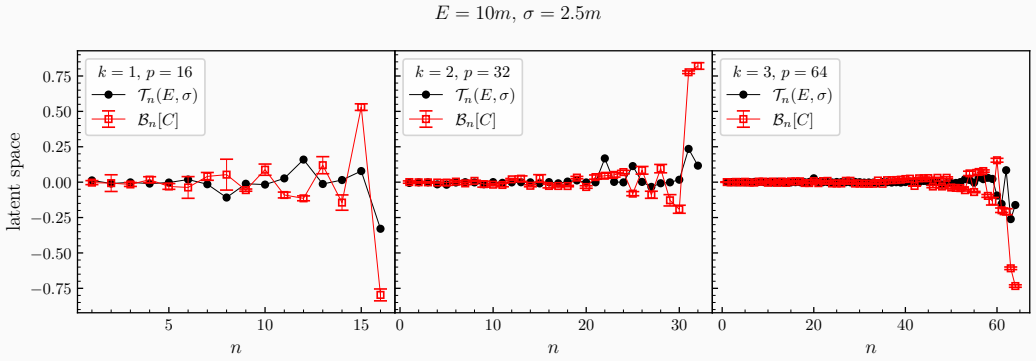


Error budget

- Continuum limit $\sigma \to 0$ limits with small systematics
- All the other systematics are equally important
- Increase of the error at large energies, in accordance to more severe reconstruction problem

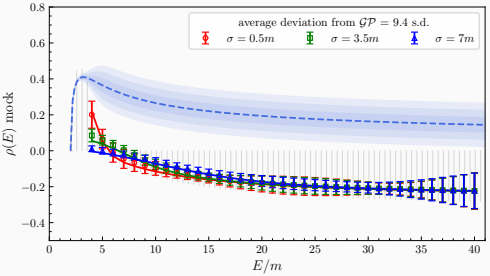


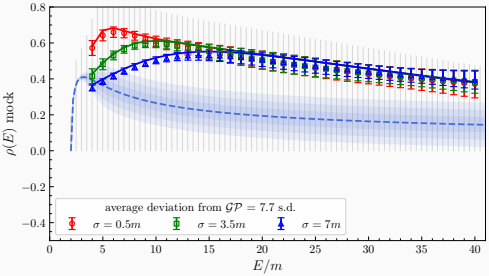
Latent spaces



- Typical order of magnitude of latent representations is 1

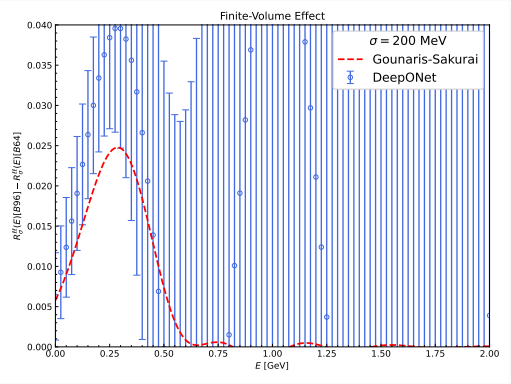
Generalization to out-of-distribution data

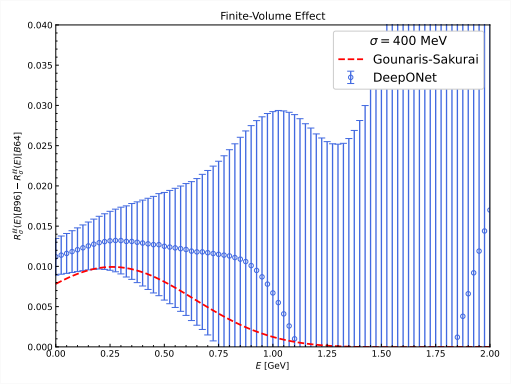




- Generalization to very out-of-distribution functions and even negative ones
- Learning properties of the operator independent of the training data!

FVE: DeepONet vs Model





- Excellent agreement in Finite-volume Effects between DeepONet and prediction from Gounaris-Sakurai model