Two slides on Gaussian Processes

- Non-parametric, probabilistic model for a function
- Suppose we already know f at $x_1, x_2, x_3, \ldots, x_n$.
- Specify how f(y) is correlated with $f(x_1), f(x_2), \ldots..$; don't specify underlying functional form.
- But value of f(y) is not deterministic: it's given by a (Gaussian) probability distribution. Prior is: $\pi(f(x)\,|\,I) \sim \mathcal{N}(0, k(x,y))$
- Correlation decreases as points get further away from each other.
- Specify correlation matrix of f at x and y, e.g.:

$$k(x,y) = \bar{c}^2 \exp\left(-\frac{(x-y)^2}{2\ell^2}\right)$$

Statistical model choices

- Two parameters $\bar{c}^2$ and ℓ: $\mathrm{pr}(\bar{c}^2\,|\,I) \sim \chi^{-2}(\nu_0, \tau_0^2)$; $\mathrm{pr}(\ell\,|\,I)$ uniform

Bayesian updating for GPs

- $\pi(f(x)\,|\,I) \sim \mathcal{N}(0, k(x, y))$
- Then updated by training data to give p(f(x)|f(x_1),…,f(x_n),I)
- $p(f(x)\,|\,f(x_1), I) \sim \mathcal{N}(\mu, K')$
- $\mu = k(x, x_1)k(x_1, x_1)^{-1}f(x_1)$
- $K' = k(x, x) - k(x, x_1)k(x_1, x_1)^{-1}k(x_1, x)$
- k: “prior kernel”; K’: “conditional variance”

GPs for EFT truncation errors

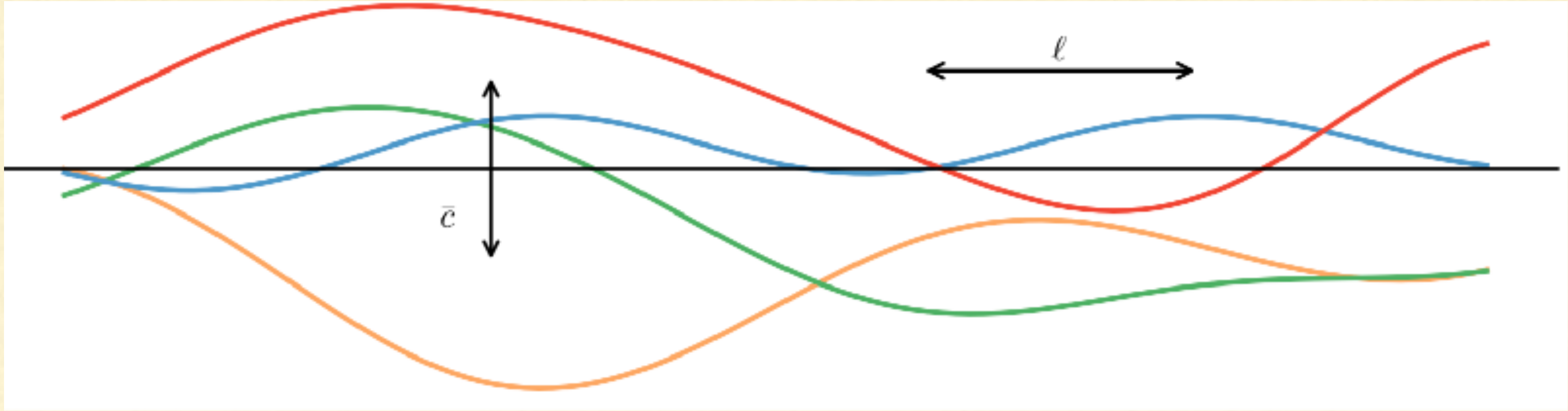
Melendez, Wesolowski, Furnstahl, DP, Pratola, PRC (2019)

$$y = y_{\rm ref} \sum_{n=0}^{k} c_n(p/m_\pi) Q^n$$

Function c_n is not a constant.
But the c_n's at different values of p aren't independent random variables either

Our hypothesis:

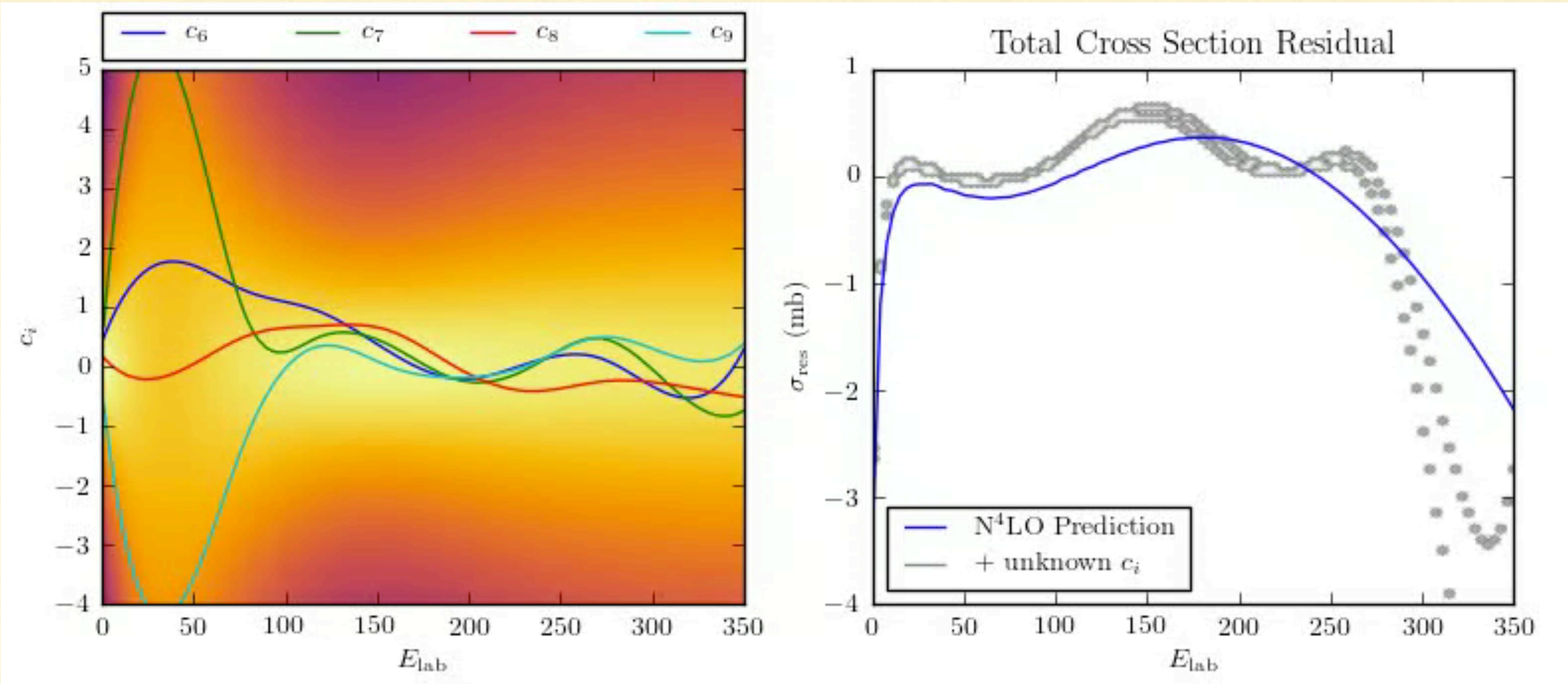
EFT coefficients at different orders can be modeled as independent draws from a *Gaussian Process* with a stationary kernel



- Gaussian distribution at each point
- With correlation structure parameterized by a single $\bar{c}^2$ and ℓ at all orders

Inferring the next coefficient(s)

$$\pi(c_n(x)) \sim \mathcal{N}\left[0, \bar{c}^2 \exp\left(-\frac{(x-y)^2}{2\ell^2}\right)\right]$$



GP "model" for χEFT coefficients, trained on c_2 -c_5: predict distribution of

$$\Delta\sigma(E) = \sigma_{\mathrm{ref}}[c_6(E)Q^6 + c_7(E)Q^7 + c_8(E)Q^8 + c_9(E)Q^9 + c_{10}(E)Q^{10}]$$

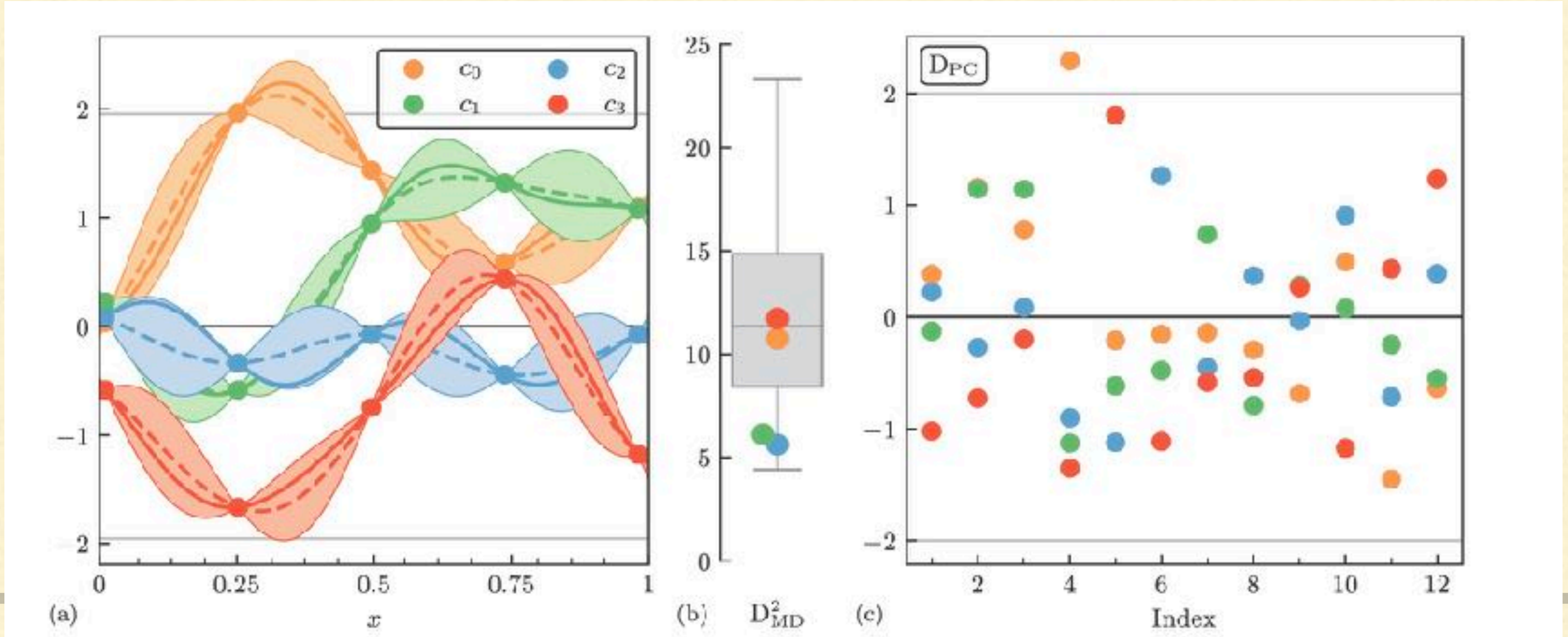
Diagnostics

Melendez, Furnstahl, DP, Pratola, Wesolowski, in preparation
after Bastos & O'Hagan, Technometrics, 2009

- Assess performance of fitted GP on "validation data".
- Errors are correlated, so can't just add up number of sigmas. "Consistency plot" does not account for correlations.
- Define Mahalanobis distance, which does account for correlations

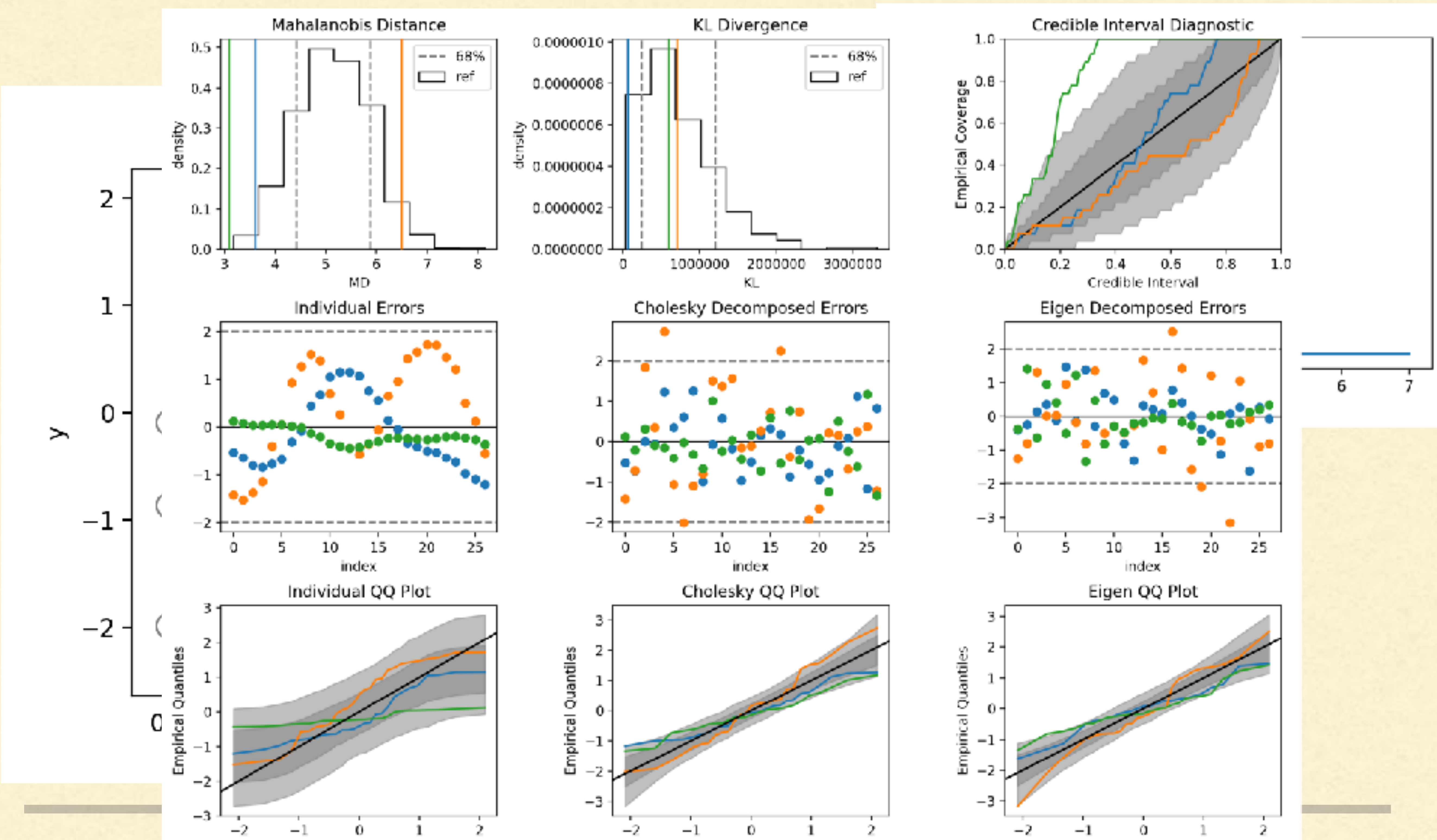
$$\mathrm{MD} \equiv (y(x_i) - \mu)^T [k(x_i, x_j)]^{-1} (y(x_j) - \mu)$$

- Write $k(x_i,x_j)=G^T G$ with G from PC decomposition; then form PC errors

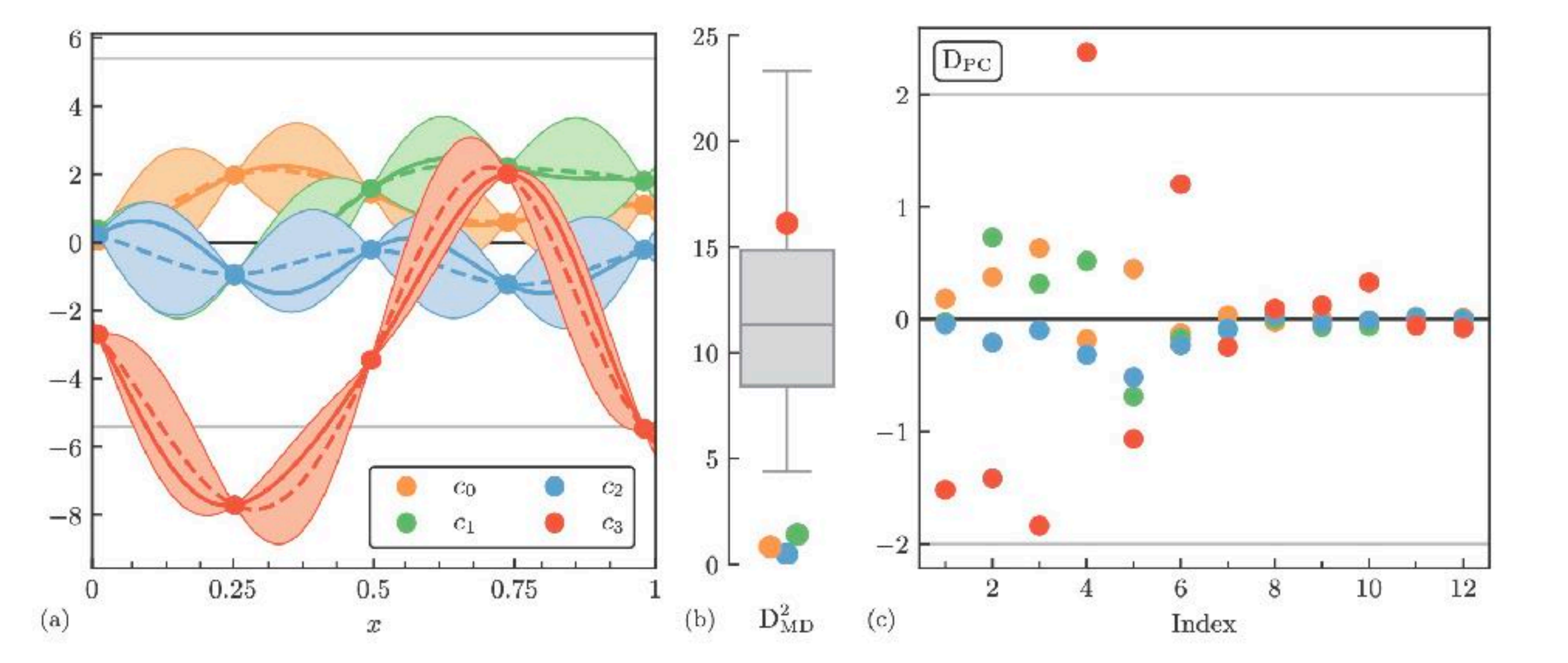


What can go wrong 1: different cbar's

- Try to fit a single GP to data generated using different variances



Misspecified GP



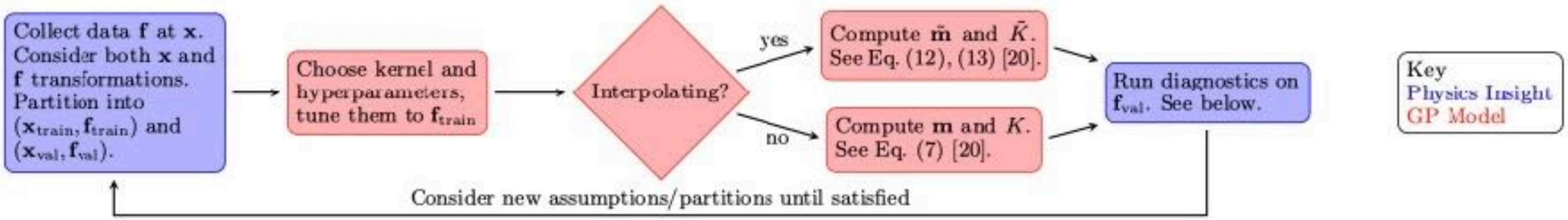
Q=0.3 rather than "true" Q=0.5: so size of coefficient c_n grows as $(1.67)^n$

Also leads to misestimated length scale, $l_{est} < l_{true}$

Model checking

https://github.com/buqeye/gsum

Melendez et al. (2019), Millican et al. (2024), Bastos & O'Hagan (2009)



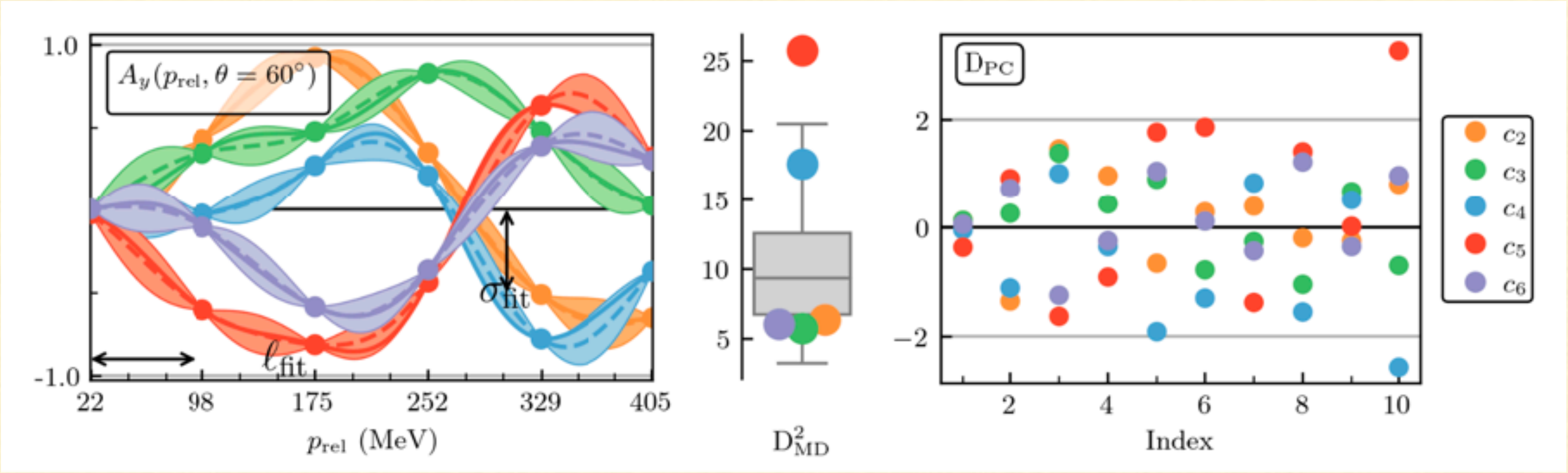
Diagnostic	Formula	Motivation	Success	Failure
Visualize the function	—	Does $\mathbf{f}_{\rm val}$ look like a draw from a GP? What kind of GP?	$\mathbf{f}_{\rm val}$ "looks similar" to draws from a GP	$\mathbf{f}_{\rm val}$ "stands out" compared to GP draws
Mahalanobis Distance $\mathrm{D}^2_{\rm MD}$	$(\mathbf{f}_{\rm val} - \mathbf{m})^\intercal K^{-1}(\mathbf{f}_{\rm val} - \mathbf{m})$	Can we *quantify* how much the $\mathbf{f}_{\rm val}$ looks like a GP?	$\mathrm{D}^2_{\rm MD}$ follows its theoretical distribution (χ^2_M)	$\mathrm{D}^2_{\rm MD}$ lies too far away from the expected value of M
Pivoted Cholesky $\mathbf{D}_{\rm PC}$	$G^{-1}(\mathbf{f}_{\rm val} - \mathbf{m})$	Can we understand why $\mathrm{D}^2_{\rm MD}$ is failing?	At each index, points follow standard Gaussian	Many cases (see below)
Credible Interval $\mathrm{D}_{\rm CI}(P)$ for $P \in [0,1]$	$\frac{1}{M}\sum_{i=1}^{M} \mathbf{1}[\mathbf{f}_{{\rm val},i} \in \mathrm{CI}_i(P)]$	Do $100P\%$ credible intervals capture data roughly $100P\%$ of the time?	Plot $\mathrm{D}_{\rm CI}(P)$ for $P \in [0,1]$; the curve should be within errors of $\mathrm{D}_{\rm CI}(P) = P$,	$\mathrm{D}_{\rm CI}(P)$ is far from $100P\%$, particularly for large $100P\%$ (e.g., 68% and 95%).

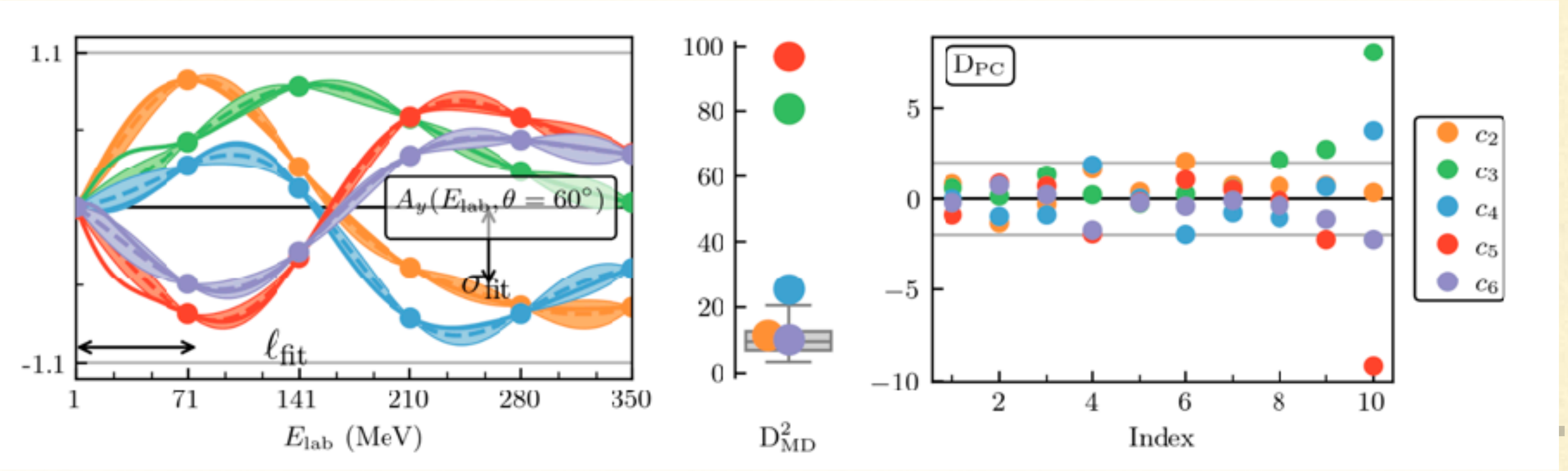
Variance	Length Scale	Observed Pattern in $\mathbf{D}_{\rm PC}$
$\sigma_{\rm est} = \sigma_{\rm true}$	$\ell_{\rm est} = \ell_{\rm true}$	Points are distributed as a standard Gaussian, with no pattern across index (e.g., only $\approx 5\%$ of points outside 2σ lines).
$\sigma_{\rm est} = \sigma_{\rm true}$	$\ell_{\rm est} > \ell_{\rm true}$	Points look well distributed at small index but expand to a too-large range at high index.
$\sigma_{\rm est} = \sigma_{\rm true}$	$\ell_{\rm est} < \ell_{\rm true}$	Points look well distributed at small index but shrink to a too-small range at high index.
$\sigma_{\rm est} > \sigma_{\rm true}$	$\ell_{\rm est} = \ell_{\rm true}$	Points are distributed in a too-small range at all indices.
$\sigma_{\rm est} < \sigma_{\rm true}$	$\ell_{\rm est} = \ell_{\rm true}$	Points are distributed in a too-large range at all indices.

NN physics choice I: energy/momentum?

Millican, Furnstahl, Melendez, Phillips, Pratola, PRC (2024)

- Is it $c_n(p)$ or $c_n(E)$ that has a single length scale? Diagnostics!

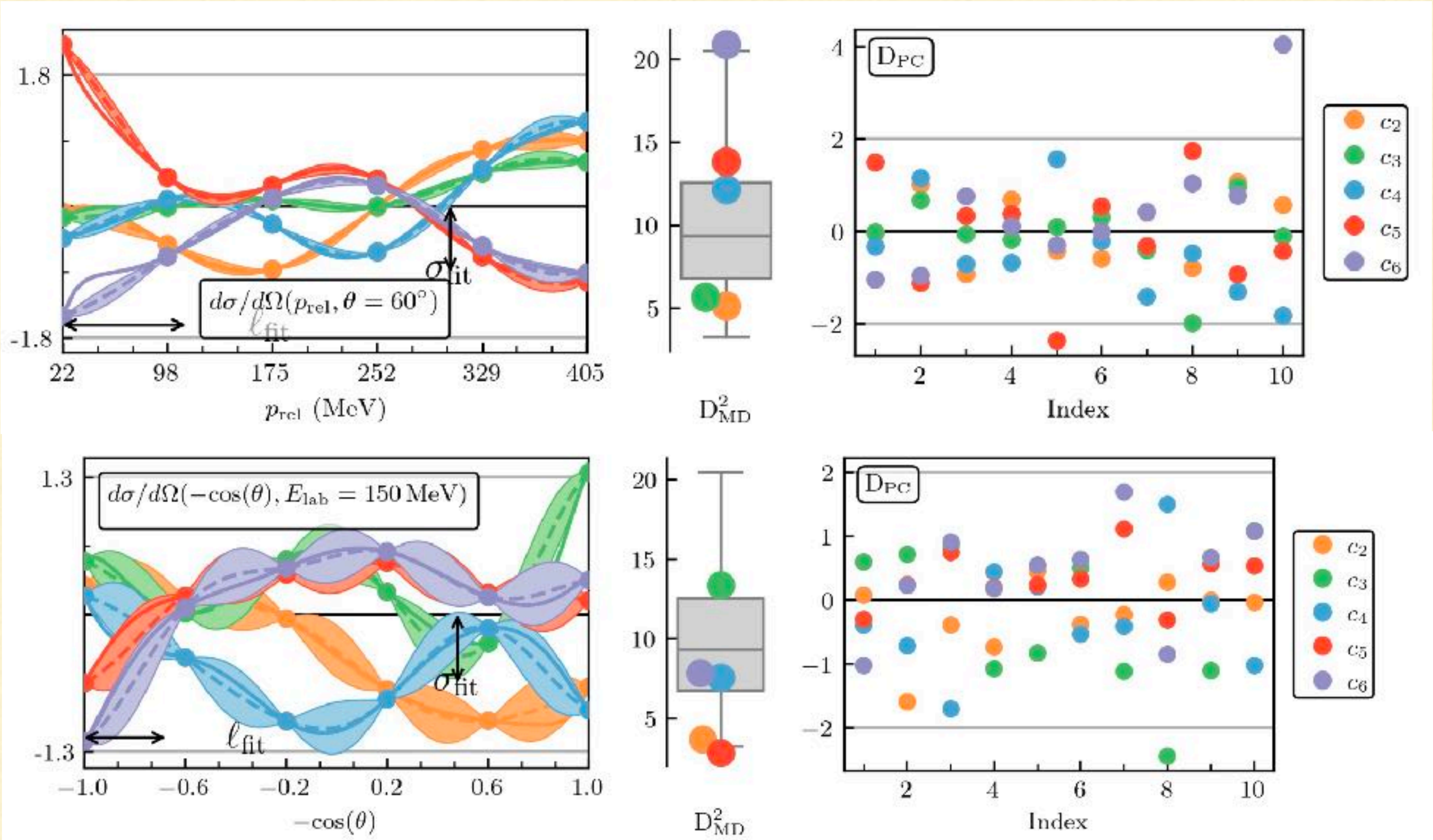




What does success look like?

Millican, Furnstahl, Melendez, DP, Pratola, PRC (2024)

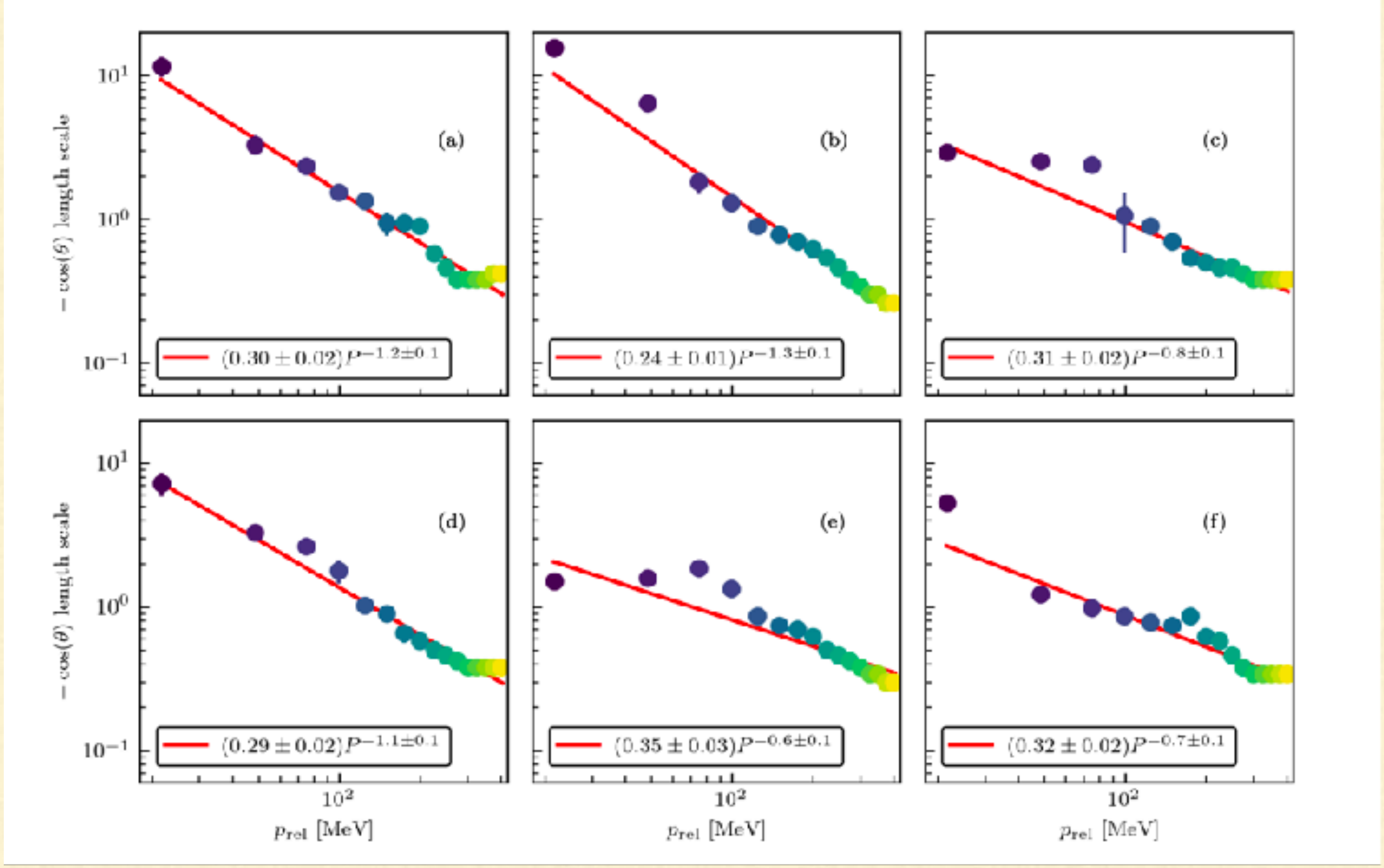
SMS 500 MeV, choice of expansion parameter Q=Q_{sum}



The GP is not 2D stationary in (p_{rel},cos(θ_{cm}))

Millican, Furnstahl, Melendez, DP, arXiv (2025)

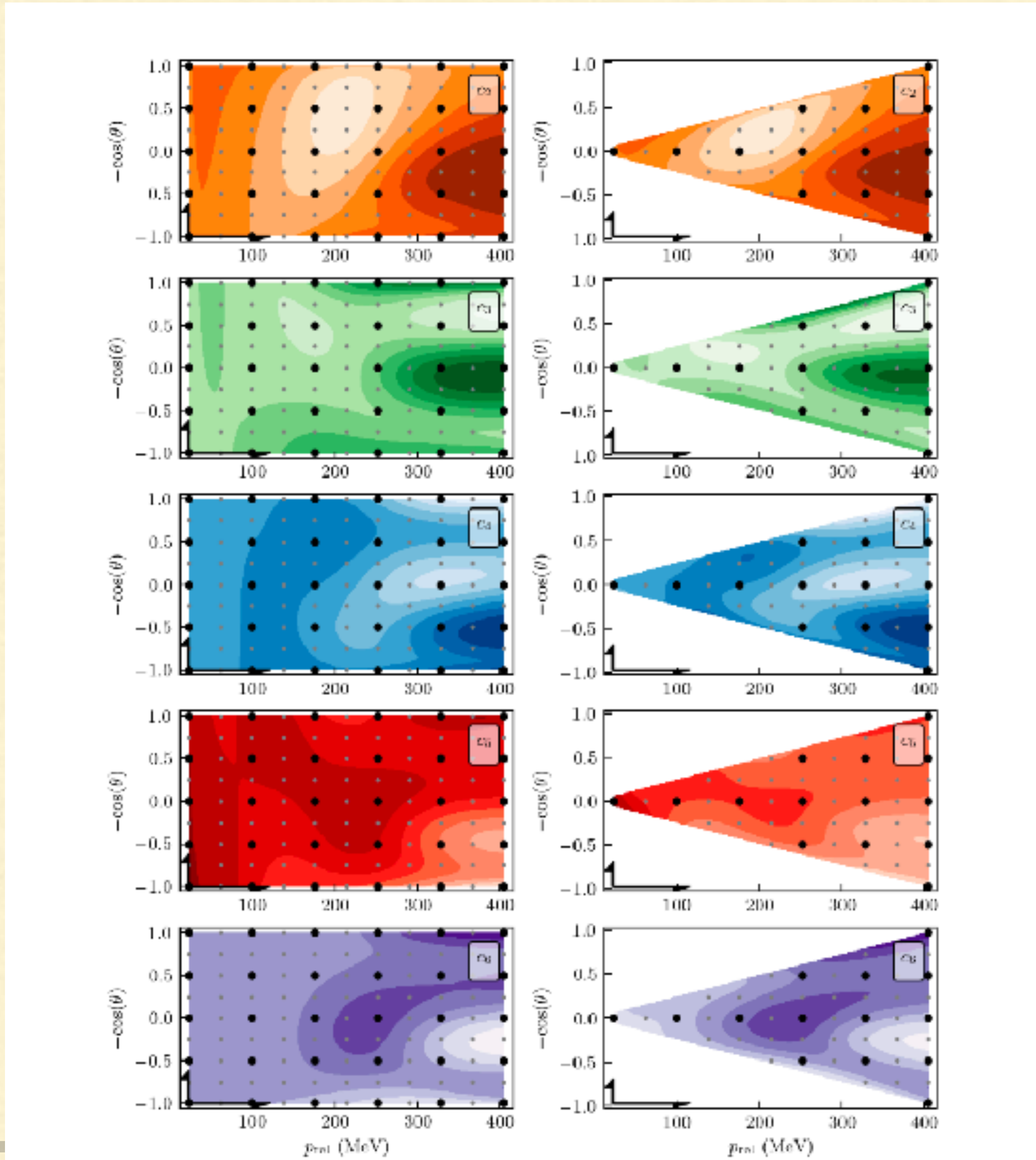
SMS 500 MeV



- $\ell_\theta \sim 1/p$ to within uncertainties
- "Warp" input space to account for 1/p effect

$$\ell_\theta(p) = \ell_\theta \left(\frac{405\ \text{MeV}}{p} \right)$$

Unwarped vs warped coefficients



Coefficients vary faster with angle at high momentum than they do at low momentum

Redefine input space: $\theta \rightarrow \theta / p_{rel}$

Can also just put in length scale $\sim 1/p_{rel}$

Unwarped or warped? Diagnostics!

$d\sigma/d\Omega$ from EMN 500 MeV; input space: (p_{rel},-cos(θ)); $Q = Q_{\text{sum}} \equiv \dfrac{p_{\text{rel}} + m_\pi}{\Lambda_b + m_\pi}$