TALENT COURSE 11
LEARNING FROM DATA: BAYESIAN METHODS AND MACHINE LEARNING

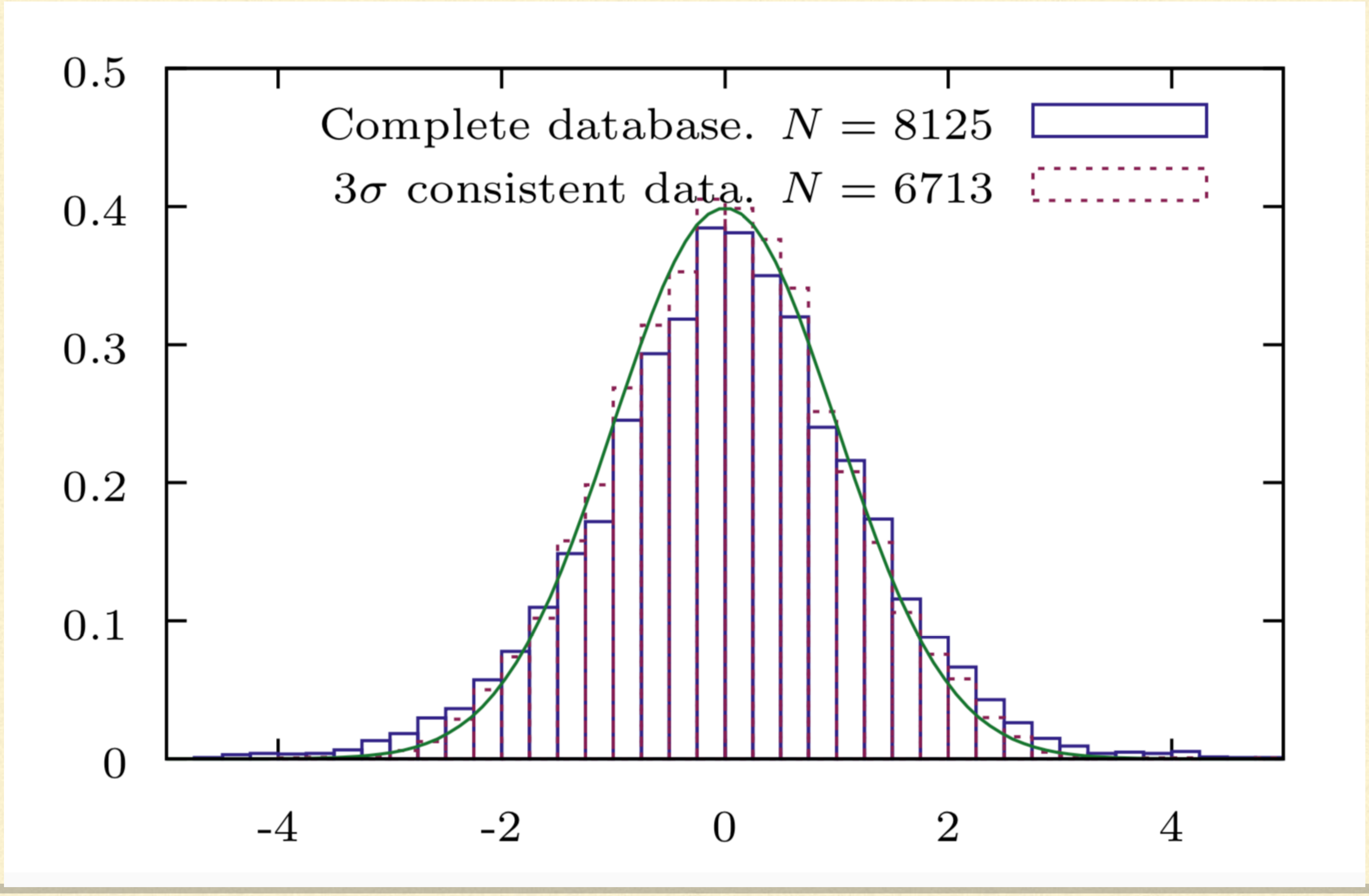
Lecture 21: Model checking I

Daniel Phillips
Ohio University
TU Darmstadt
ExtreMe Matter Institute

TALENT Course 11 is possible thanks to funding from the STFC

QQ plots

Navarro Perez, Ruiz Arriola, Amaro, PRC (2014)



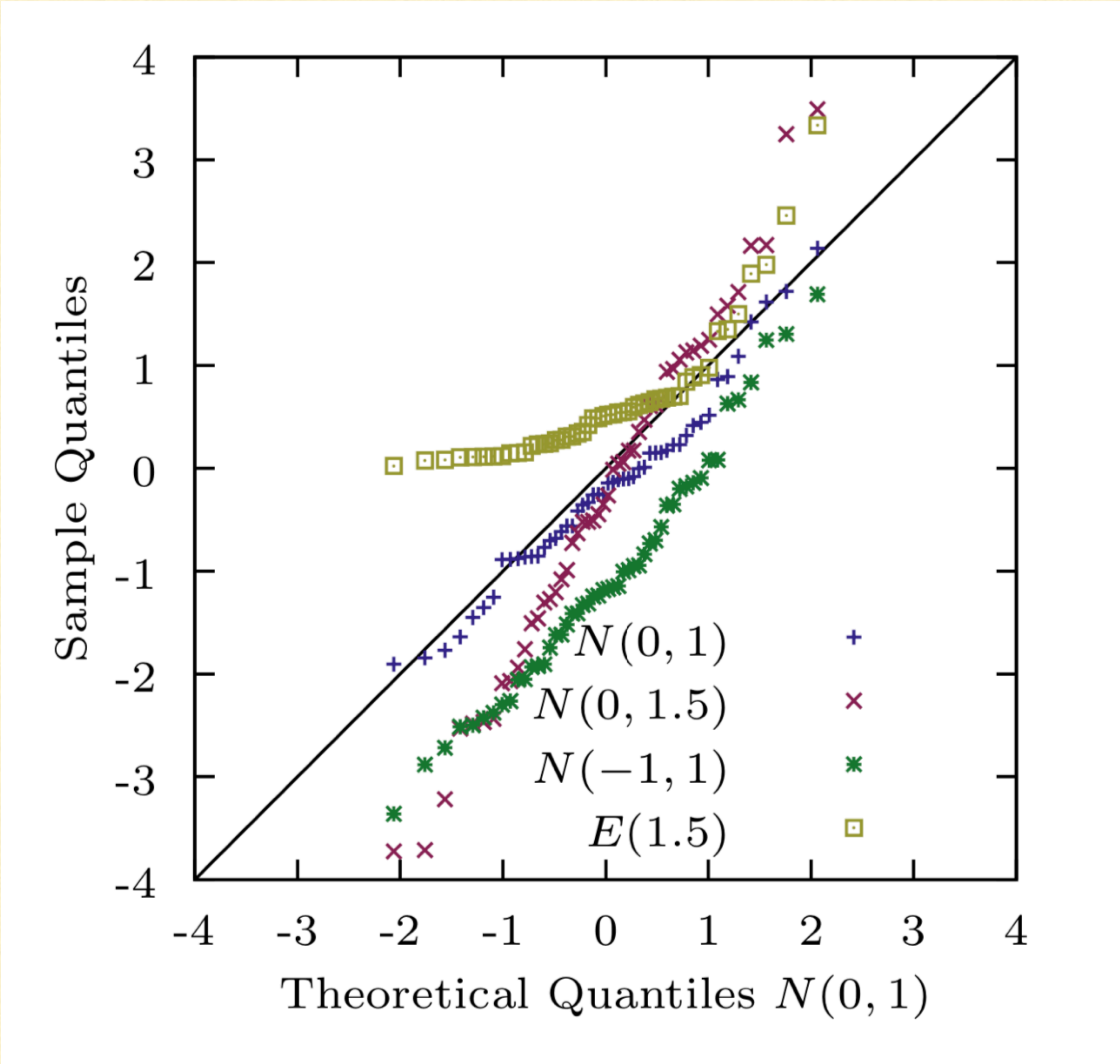
QQ plots

Navarro Perez, Ruiz Arriola, Amaro, PRC (2014)

Arrange data residuals from most negative to most positive

QQ plots

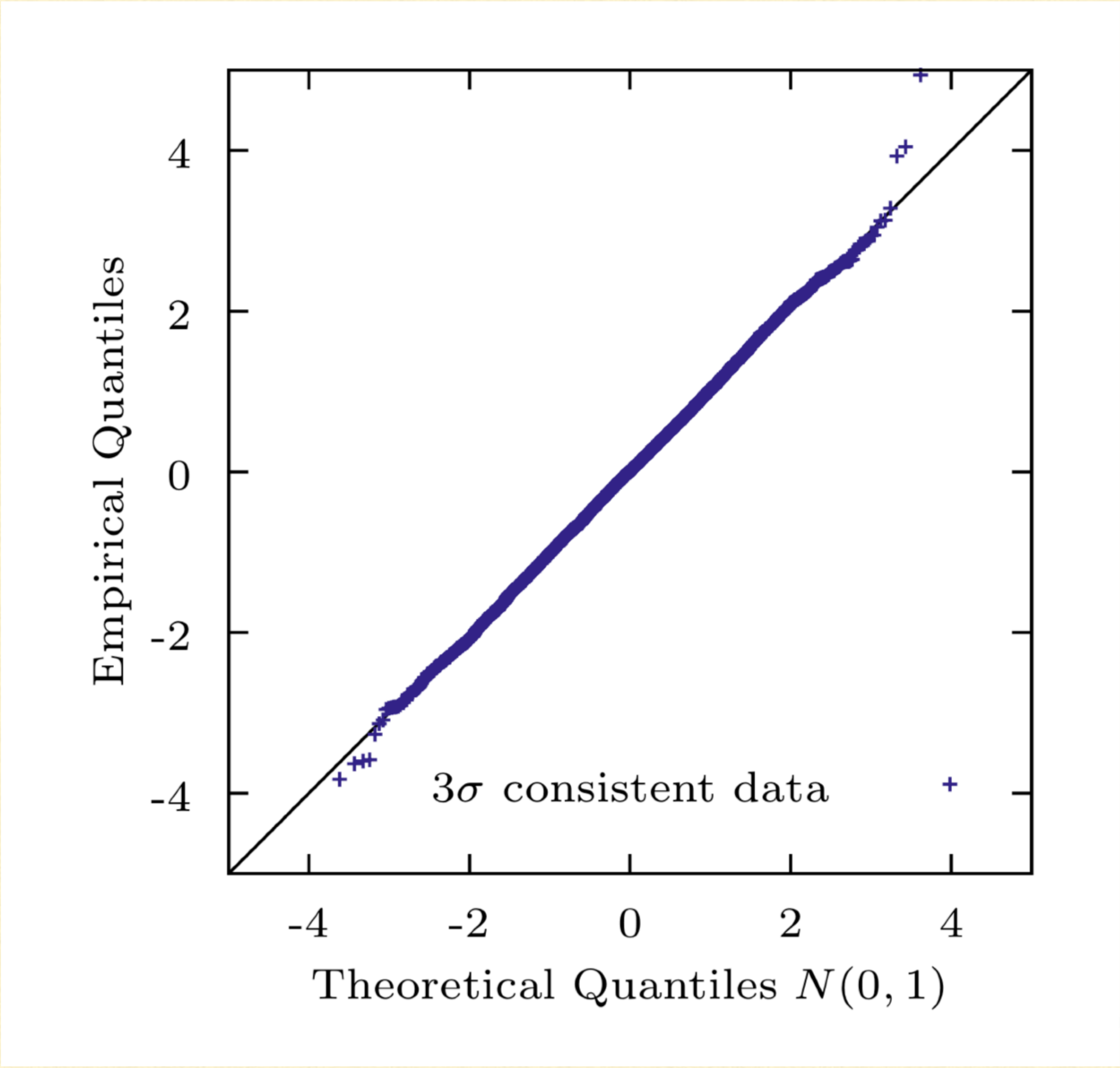
Navarro Perez, Ruiz Arriola, Amaro, PRC (2014)



Arrange data residuals from most negative to most positive

QQ plots

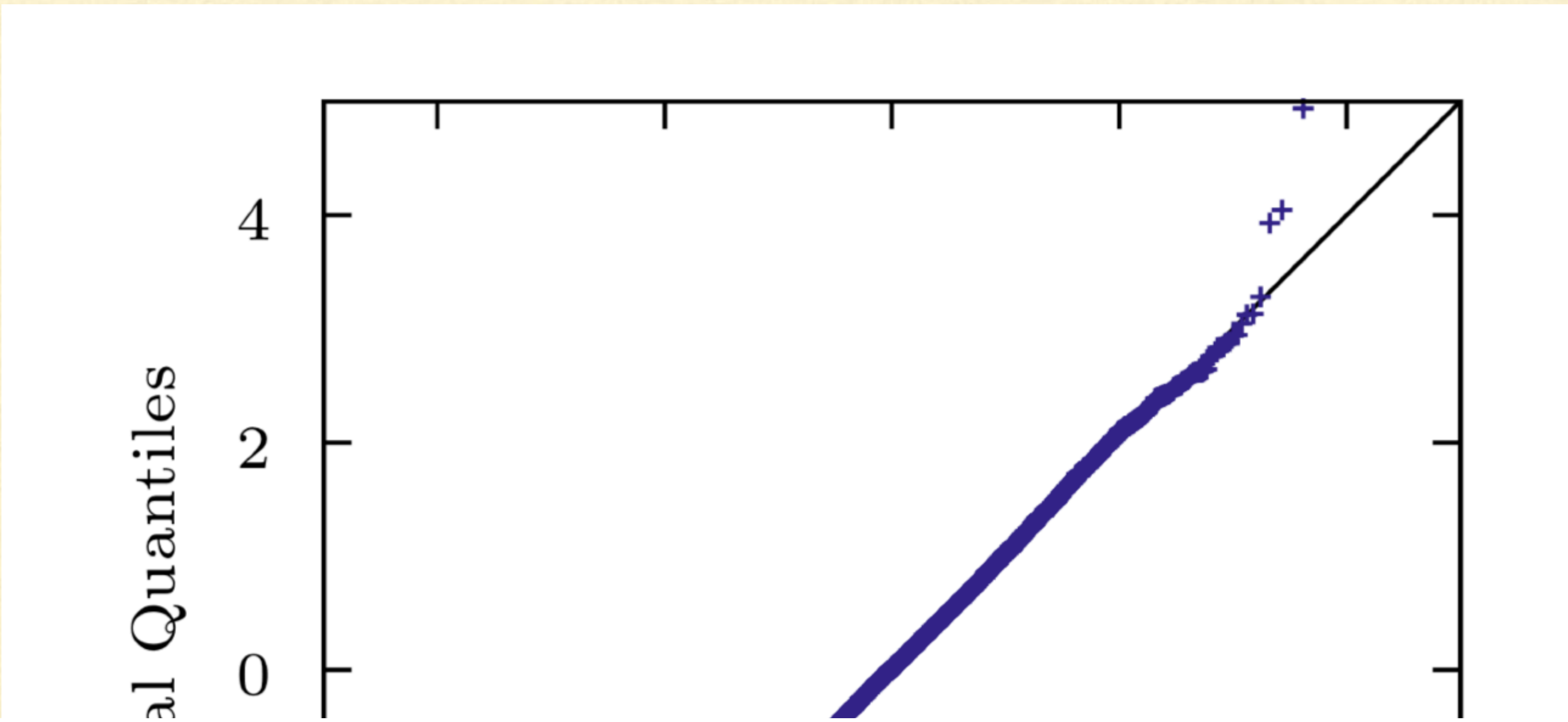
Navarro Perez, Ruiz Arriola, Amaro, PRC (2014)



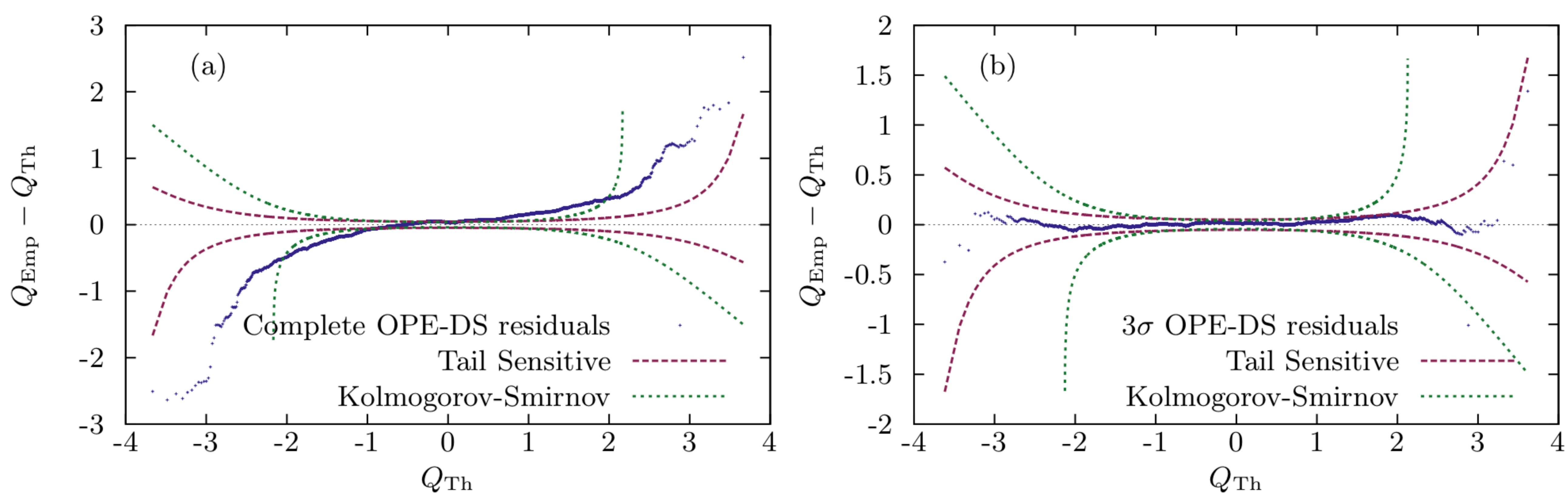
Arrange data residuals from most negative to most positive

QQ plots

Navarro Perez, Ruiz Arriola, Amaro, PRC (2014)



Arrange data residuals from most negative to most positive

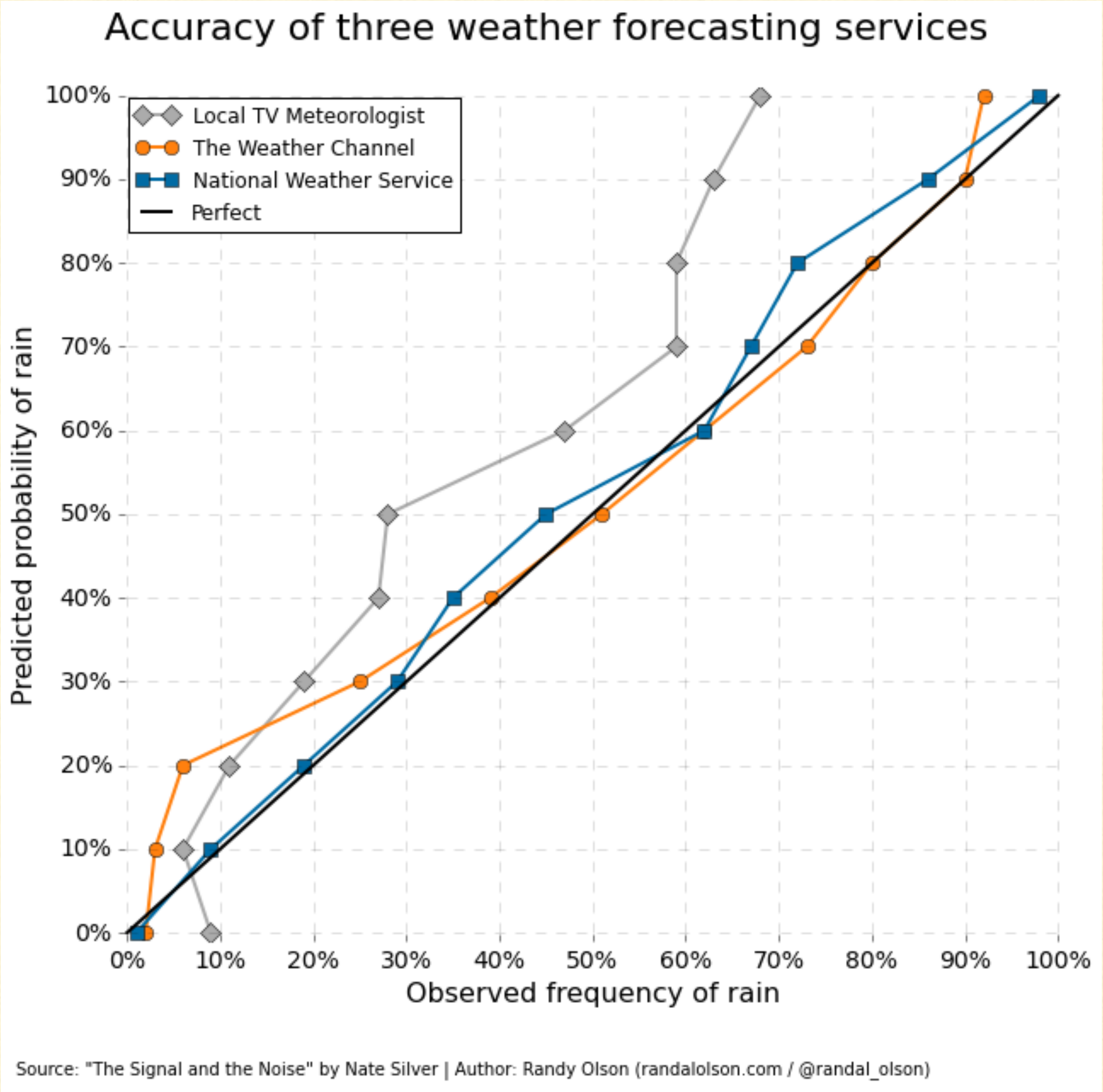


Let’s talk about the weather

Let's talk about the weather

- Consider predictions for percentage probability of rain each day. Bin in 10% intervals.
- Fix a given interval: compute number of days it actually rained, compare.

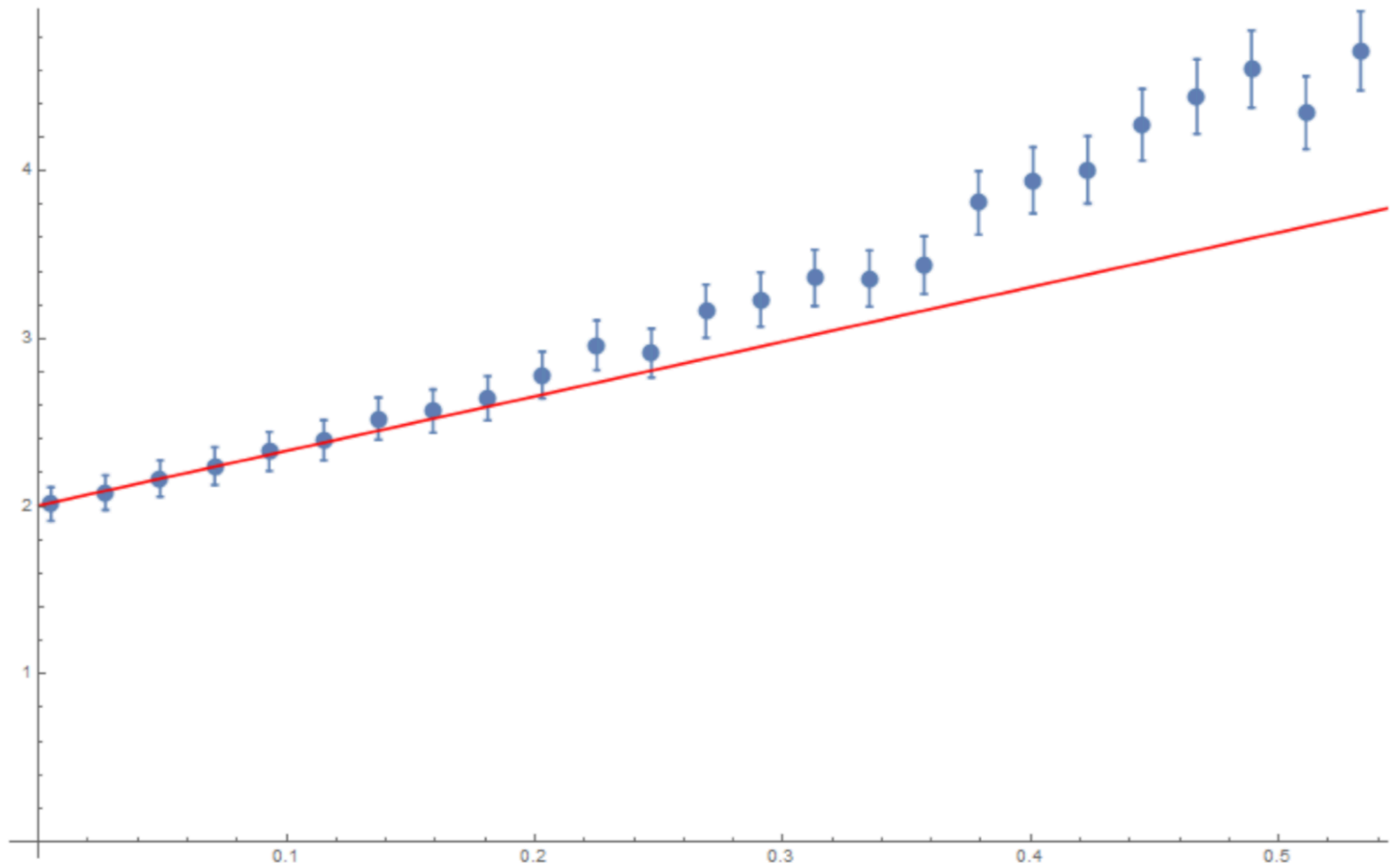
Let's talk about the weather



- Consider predictions for percentage probability of rain each day. Bin in 10% intervals.
- Fix a given interval: compute number of days it actually rained, compare.

Systematic trends in residuals

Furnstahl, DP, Wesolowski, JPG (2014); Billig, HTC thesis (2019)

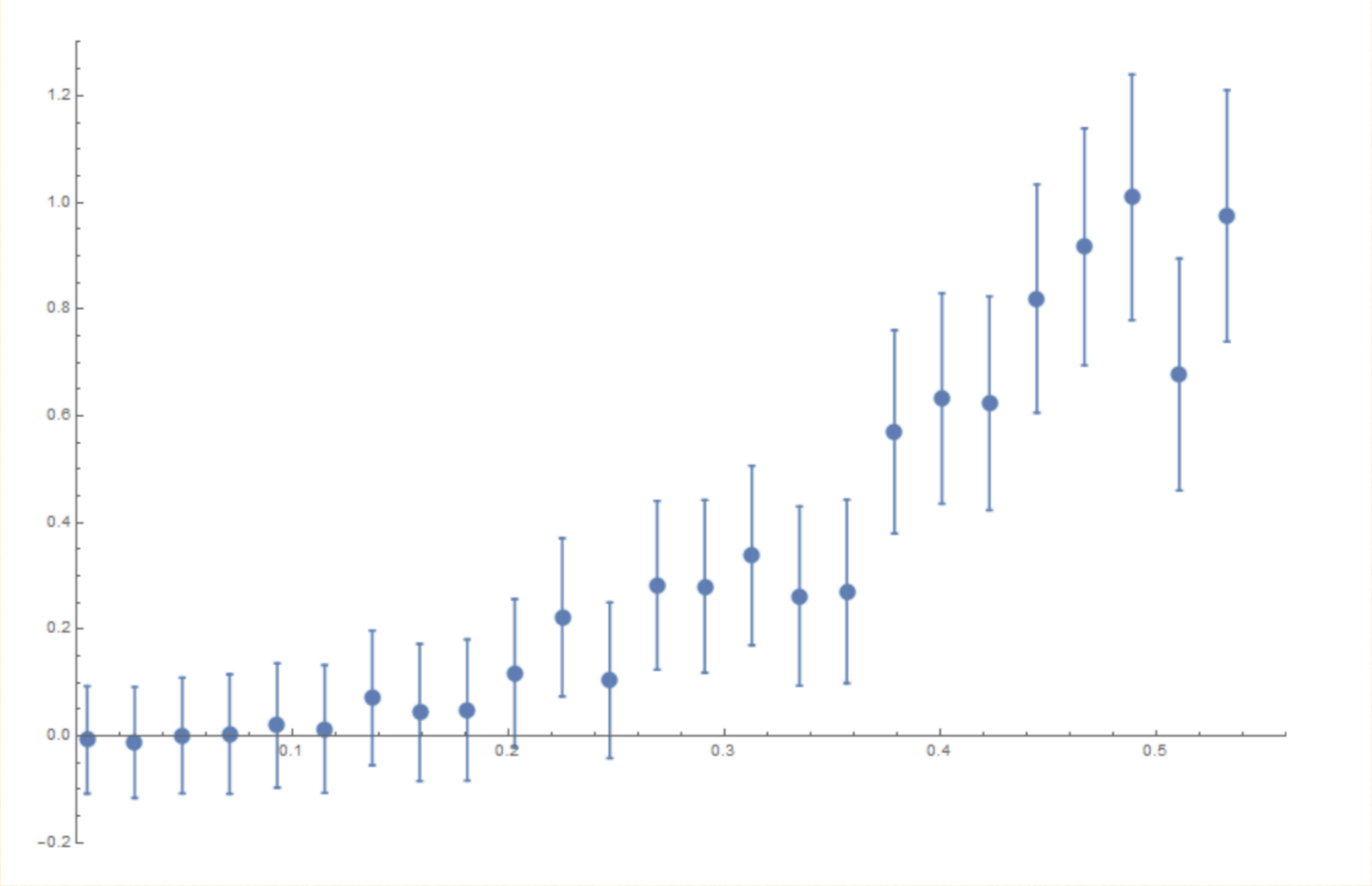


a0	a1
2.00473 ± 0.0834257	3.25548 ± 1.4859

Clear that a linear fit does not produce Gaussian residuals

Systematic trends in residuals

Furnstahl, DP, Wesolowski, JPG (2014); Billig, HTC thesis (2019)



Clear that a linear fit does not produce Gaussian residuals

Behavior of a ChiEFT series

Behavior of a ChiEFT series

- χEFT: encodes low-energy ($p \ll \Lambda_{\chi SB} \simeq 750$ MeV) consequences of QCD's chiral symmetry and the pattern of its breaking
- Expansion in $x \equiv m_\pi/(M_\Delta - M_N) \simeq M_\Delta - M_N/\Lambda_{\chi SB} \simeq 0.4$
- For proton electric polarizability, χEFT$\Rightarrow \alpha_{\mathrm{E1}}^{(p)} = 12.5 - 2.3 + 1.5 = 11.7$

Behavior of a ChiEFT series

- χEFT: encodes low-energy ($p \ll \Lambda_{\chi SB} \simeq 750$ MeV) consequences of QCD's chiral symmetry and the pattern of its breaking
- Expansion in $x \equiv m_\pi/(M_\Delta - M_N) \simeq M_\Delta - M_N/\Lambda_{\chi SB} \simeq 0.4$
- For proton electric polarizability, χEFT$\Rightarrow \alpha_{\rm E1}^{(p)} = 12.5 - 2.3 + 1.5 = 11.7$
- What is the theoretical uncertainty of this result, Δ_2?
- Rewrite as $\alpha_{\rm E1}^{(p)} = \alpha_{\rm LO}[1 + c_1(0.4) + c_2(0.4)^2 + c_3(0.4)^3]$

Behavior of a ChiEFT series

- χEFT: encodes low-energy ($p \ll \Lambda_{\chi SB} \simeq 750$ MeV) consequences of QCD's chiral symmetry and the pattern of its breaking
- Expansion in $x \equiv m_\pi/(M_\Delta - M_N) \simeq M_\Delta - M_N/\Lambda_{\chi SB} \simeq 0.4$
- For proton electric polarizability, χEFT$\Rightarrow \alpha_{\mathrm{E1}}^{(p)} = 12.5 - 2.3 + 1.5 = 11.7$
- What is the theoretical uncertainty of this result, Δ_2?
- Rewrite as $\alpha_{\mathrm{E1}}^{(p)} = \alpha_{\mathrm{LO}}[1 + c_1(0.4) + c_2(0.4)^2 + c_3(0.4)^3]$
- We cannot *know* the result for c_3 before we compute it, but we can use information we have to determine *what values are more or less probable*
- Two different answers, depending on what information we have:
  - What is expectation for c_3 before we know c_0, c_1, c_2?
  - In fact $\{c_n\}=\{1,-0.46,0.75\}$. What then is expectation for c_3?

Behavior of a ChiEFT series

- χEFT: encodes low-energy ($p \ll \Lambda_{\chi SB} \simeq 750$ MeV) consequences of QCD's chiral symmetry and the pattern of its breaking
- Expansion in $x \equiv m_\pi/(M_\Delta - M_N) \simeq M_\Delta - M_N/\Lambda_{\chi SB} \simeq 0.4$
- For proton electric polarizability, χEFT$\Rightarrow \alpha_{\mathrm{E1}}^{(p)} = 12.5 - 2.3 + 1.5 = 11.7$
- What is the theoretical uncertainty of this result, Δ_2?
- Rewrite as $\alpha_{\mathrm{E1}}^{(p)} = \alpha_{\mathrm{LO}}[1 + c_1(0.4) + c_2(0.4)^2 + c_3(0.4)^3]$
- We cannot *know* the result for c_3 before we compute it, but we can use information we have to determine *what values are more or less probable*
- Two different answers, depending on what information we have:
  - What is expectation for c_3 before we know c_0, c_1, c_2?
  - In fact $\{c_n\}=\{1,-0.46,0.75\}$. What then is expectation for c_3?

Griesshammer et al., PPNP (2012)

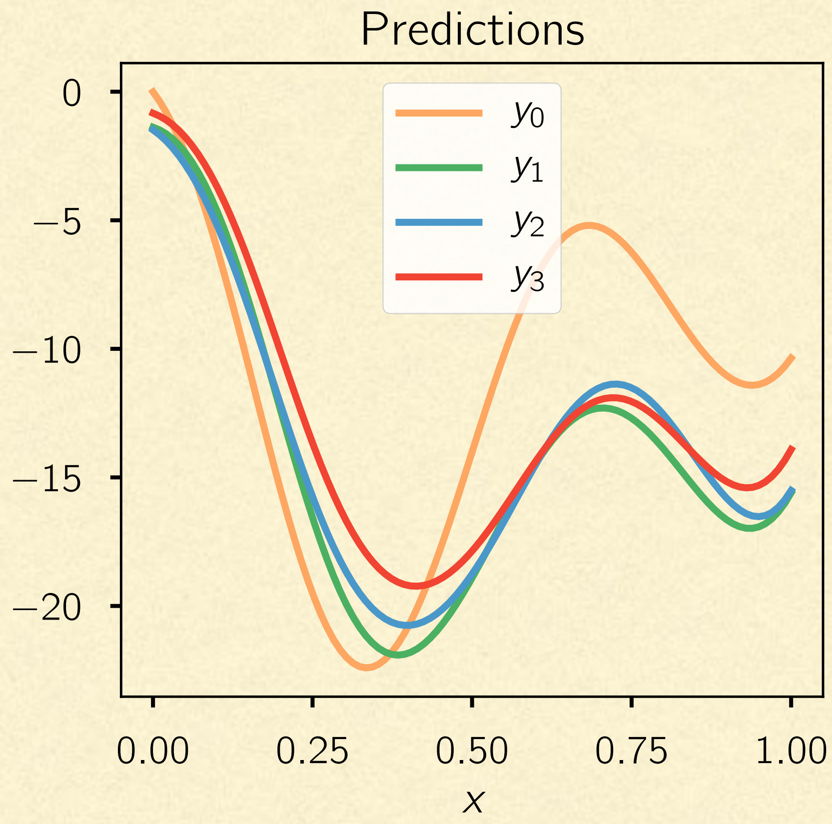
$c_3=1$

An EFT expansion in pictures

- General EFT series for observable to order k: $y = y_{\mathrm{ref}} \sum_{n=0}^{k} c_n Q^n$
- In ChiEFT $Q = \dfrac{(p, m_\pi)}{\Lambda_b}; \quad \Lambda_b \approx 600\ \mathrm{MeV}$

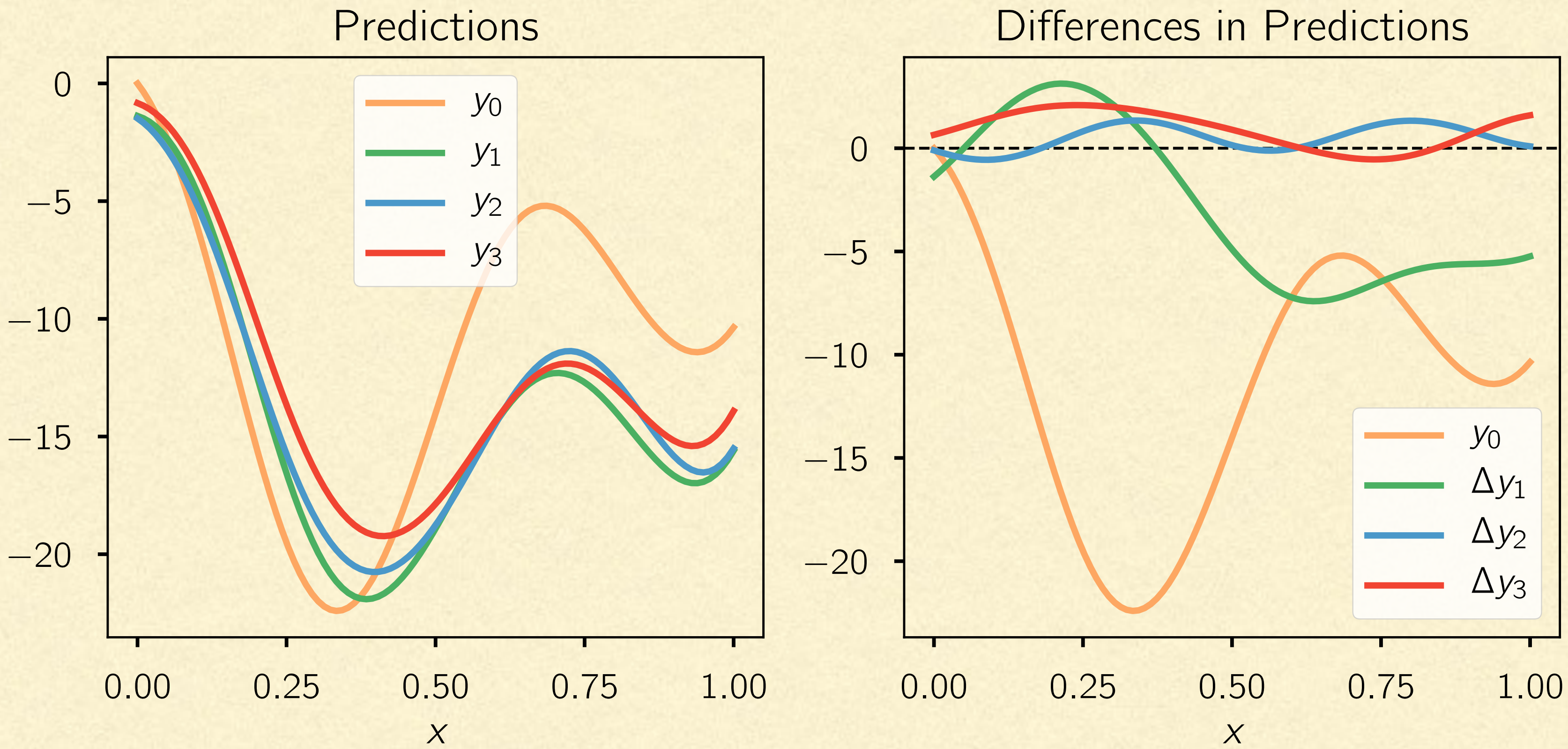
An EFT expansion in pictures

- General EFT series for observable to order k: $y = y_{\rm ref} \sum_{n=0}^{k} c_n Q^n$
- In ChiEFT $Q = \dfrac{(p, m_\pi)}{\Lambda_b}; \quad \Lambda_b \approx 600\ \mathrm{MeV}$



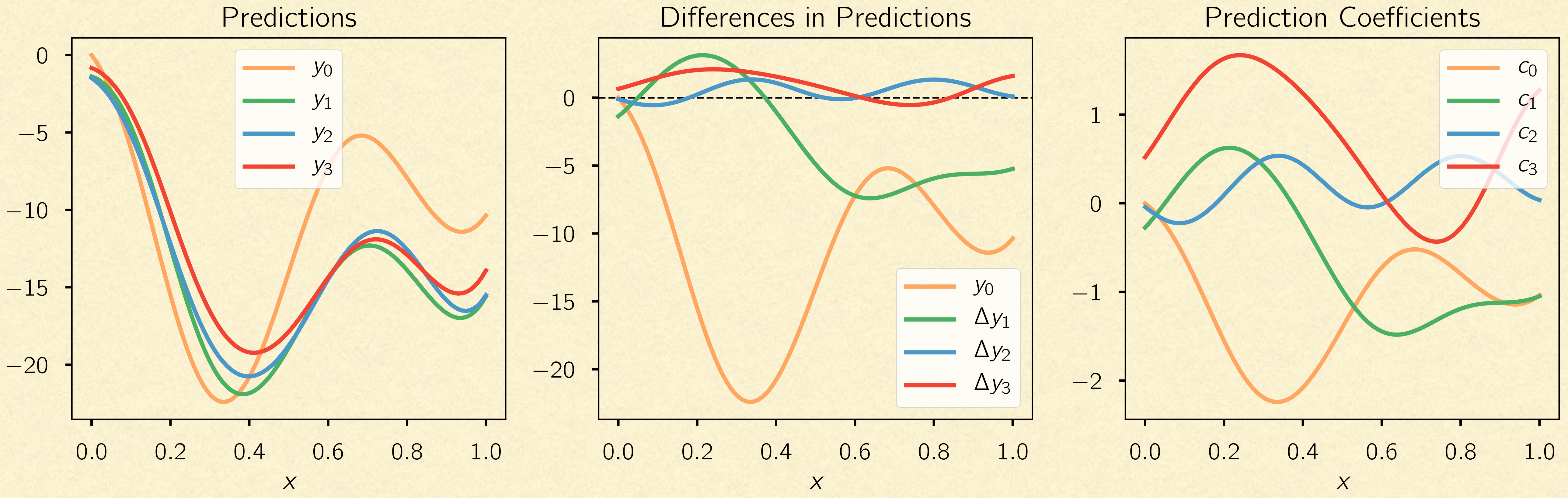
An EFT expansion in pictures

- General EFT series for observable to order k: $y = y_{\rm ref} \sum_{n=0}^{k} c_n Q^n$
- In ChiEFT $Q = \dfrac{(p, m_\pi)}{\Lambda_b}; \quad \Lambda_b \approx 600 \text{ MeV}$



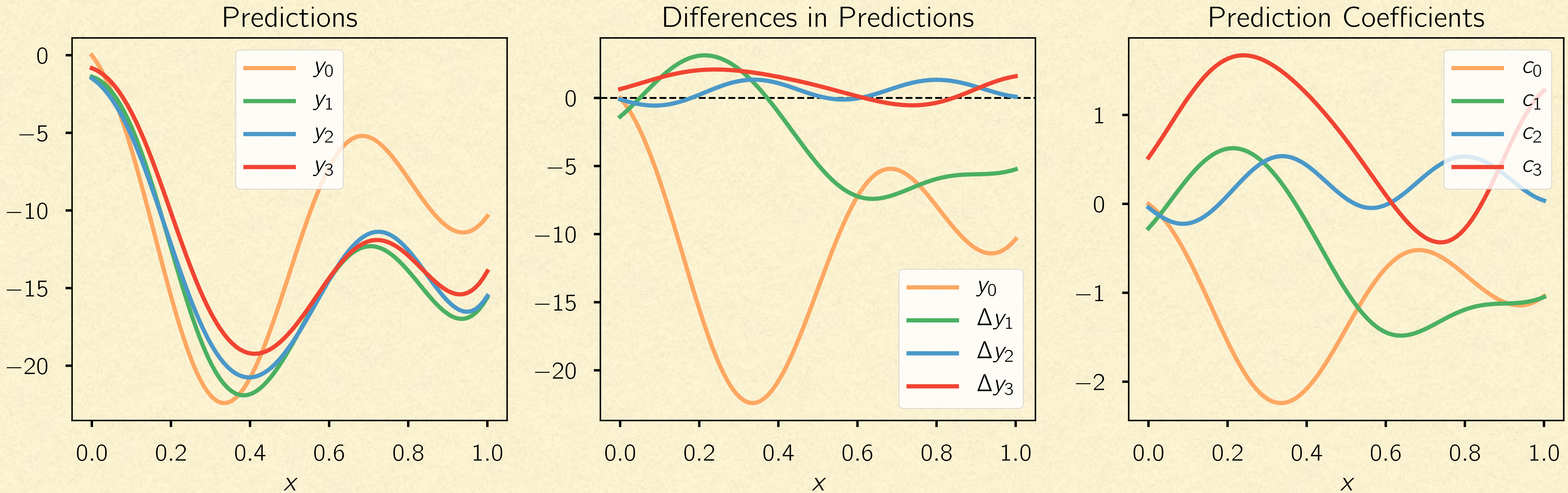
An EFT expansion in pictures

- General EFT series for observable to order k: $y = y_{\mathrm{ref}} \sum_{n=0}^{k} c_n Q^n$
- In ChiEFT $Q = \dfrac{(p, m_\pi)}{\Lambda_b}$; $\Lambda_b \approx 600$ MeV



An EFT expansion in pictures

- General EFT series for observable to order k: $y = y_{\mathrm{ref}} \sum_{n=0}^{k} c_n Q^n$
- In ChiEFT $Q = \dfrac{(p, m_\pi)}{\Lambda_b}$; $\quad \Lambda_b \approx 600\ \mathrm{MeV}$



This is what a healthy observable expansion looks like: bounded coefficients, that do not grow or shrink with order.

Bayes' theorem

Thomas Bayes (1701?-1761)

$$\mathrm{pr}(A\,|\,B, I) = \frac{\mathrm{pr}(B\,|\,A, I)\,\mathrm{pr}(A\,|\,I)}{\mathrm{pr}(B\,|\,I)}$$

Bayes' theorem

Thomas Bayes (1701?-1761)

$$\mathrm{pr}(A \mid B, I) = \frac{\mathrm{pr}(B \mid A, I)\,\mathrm{pr}(A \mid I)}{\mathrm{pr}(B \mid I)}$$

http://www.bayesian-inference.com

Bayes' theorem

Thomas Bayes (1701?-1761)

$$\mathrm{pr}(A \mid B, I) = \frac{\mathrm{pr}(B \mid A, I)\,\mathrm{pr}(A \mid I)}{\mathrm{pr}(B \mid I)}$$

http://www.bayesian-inference.com

$$\mathrm{pr}(\text{model} \mid \text{data}, I) = \frac{\mathrm{pr}(\text{data} \mid \text{model}, I)\,\mathrm{pr}(\text{model} \mid I)}{\mathrm{pr}(\text{data} \mid I)}$$

Bayes' theorem

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A\,|\,B, I) = \frac{\mathrm{pr}(B\,|\,A, I)\,\mathrm{pr}(A\,|\,I)}{\mathrm{pr}(B\,|\,I)}$$

$$\mathrm{pr}(\mathrm{model}\,|\,\mathrm{data}, I) = \frac{\mathrm{pr}(\mathrm{data}\,|\,\mathrm{model}, I)\,\mathrm{pr}(\mathrm{model}\,|\,I)}{\mathrm{pr}(\mathrm{data}\,|\,I)}$$

Posterior

Bayes' theorem

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A\,|\,B, I) = \frac{\mathrm{pr}(B\,|\,A, I)\,\mathrm{pr}(A\,|\,I)}{\mathrm{pr}(B\,|\,I)}$$

Likelihood

$$\mathrm{pr}(\mathrm{model}\,|\,\mathrm{data}, I) = \frac{\mathrm{pr}(\mathrm{data}\,|\,\mathrm{model}, I)\,\mathrm{pr}(\mathrm{model}\,|\,I)}{\mathrm{pr}(\mathrm{data}\,|\,I)}$$

Posterior

Bayes' theorem

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A \mid B, I) = \frac{\mathrm{pr}(B \mid A, I)\,\mathrm{pr}(A \mid I)}{\mathrm{pr}(B \mid I)}$$

Likelihood

Prior

$$\mathrm{pr}(\mathrm{model} \mid \mathrm{data}, I) = \frac{\mathrm{pr}(\mathrm{data} \mid \mathrm{model}, I)\,\mathrm{pr}(\mathrm{model} \mid I)}{\mathrm{pr}(\mathrm{data} \mid I)}$$

Posterior

Bayes' theorem

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A \mid B, I) = \frac{\mathrm{pr}(B \mid A, I)\,\mathrm{pr}(A \mid I)}{\mathrm{pr}(B \mid I)}$$

Likelihood

Prior

$$\mathrm{pr}(\mathrm{model} \mid \mathrm{data}, I) = \frac{\mathrm{pr}(\mathrm{data} \mid \mathrm{model}, I)\,\mathrm{pr}(\mathrm{model} \mid I)}{\mathrm{pr}(\mathrm{data} \mid I)}$$

Posterior

Evidence

Bayes' theorem

Thomas Bayes (1701?-1761)

http://www.bayesian-inference.com

$$\mathrm{pr}(A\,|\,B,I) = \frac{\mathrm{pr}(B\,|\,A,I)\,\mathrm{pr}(A\,|\,I)}{\mathrm{pr}(B\,|\,I)}$$

Likelihood

Prior

$$\mathrm{pr}(\mathrm{model}\,|\,\mathrm{data},I) = \frac{\mathrm{pr}(\mathrm{data}\,|\,\mathrm{model},I)\,\mathrm{pr}(\mathrm{model}\,|\,I)}{\mathrm{pr}(\mathrm{data}\,|\,I)}$$

Posterior

Evidence

Probability as degree of belief cf. frequentist view

Probability for EFT coefficients

Furnstahl, Klco, DP, Wesolowski, PRC, 2015 after Cacciari and Houdeau, JHEP, 2011

Probability for EFT coefficients

Furnstahl, Klco, DP, Wesolowski, PRC, 2015 after Cacciari and Houdeau, JHEP, 2011

- So can we use extracted c_0, c_1, c_2, …, c_k to estimate (in a probabilistic way) c_{k+1}? From there construct $\Delta_k = y_{ref}\, c_{k+1}\, Q^{k+1}$:truncation error

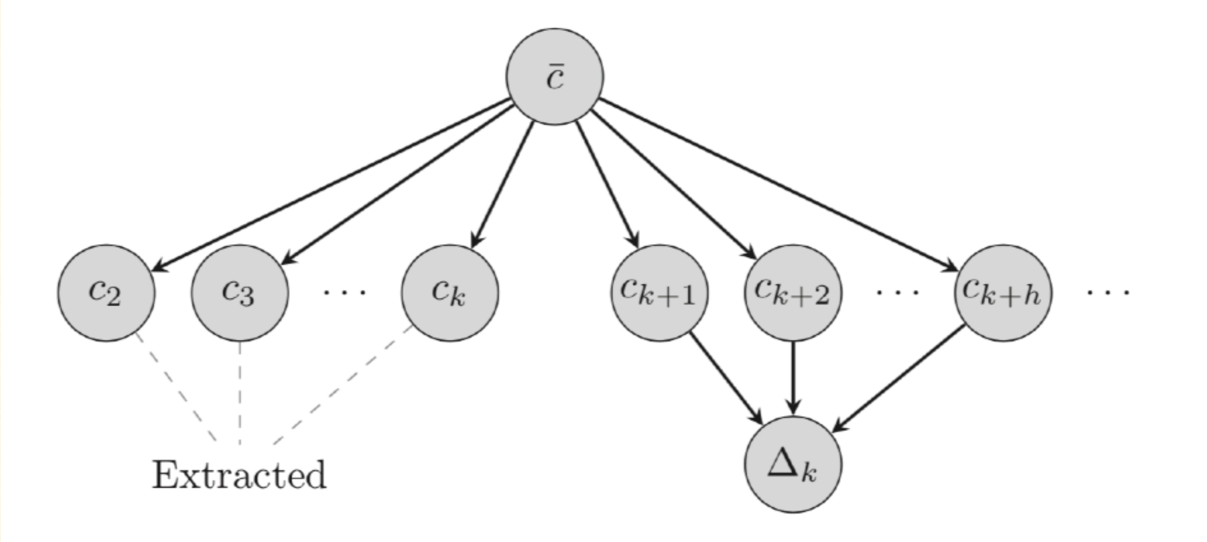
Probability for EFT coefficients

Furnstahl, Klco, DP, Wesolowski, PRC, 2015 after Cacciari and Houdeau, JHEP, 2011

- So can we use extracted c_0, c_1, c_2, …, c_k to estimate (in a probabilistic way) c_{k+1}? From there construct $\Delta_k = y_{ref}\, c_{k+1}\, Q^{k+1}$:truncation error

- Bayesian model:

Parameter cbar sets size of all dimensionless coefficients



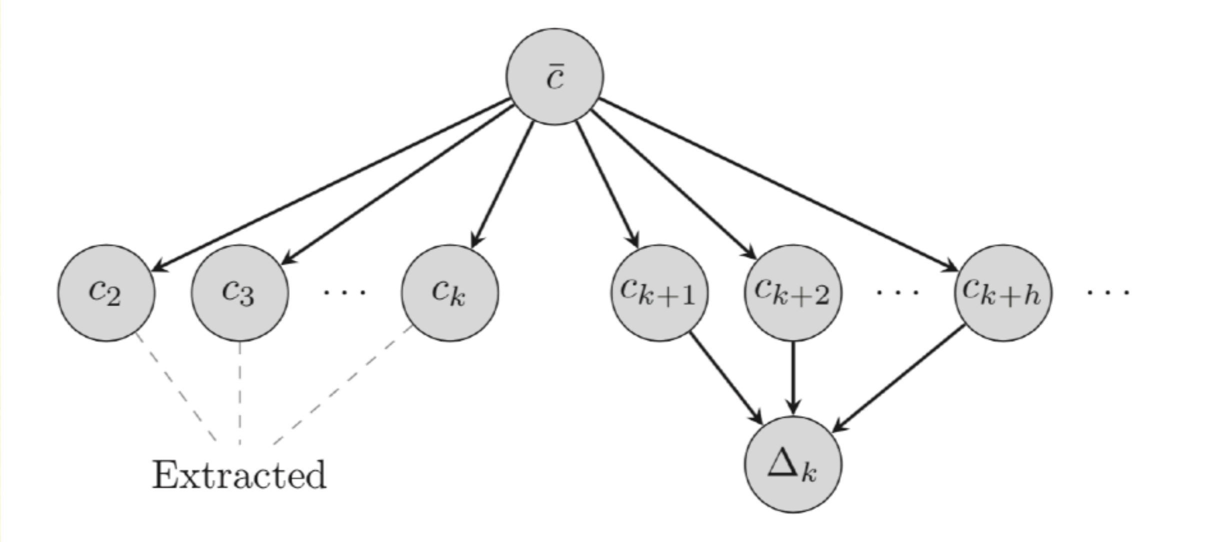
Probability for EFT coefficients

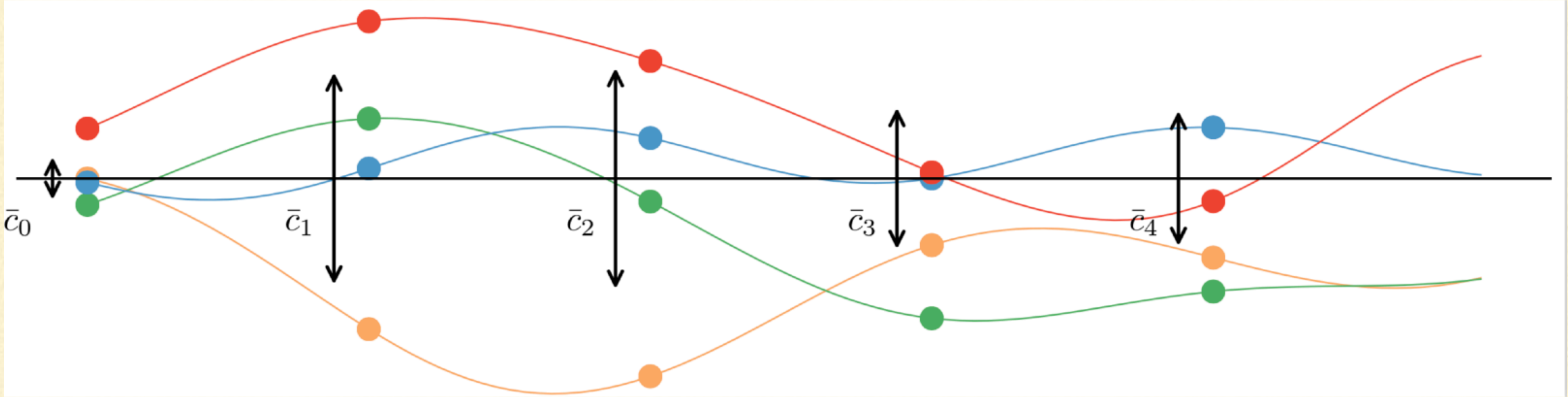
Furnstahl, Klco, DP, Wesolowski, PRC, 2015 after Cacciari and Houdeau, JHEP, 2011

- So can we use extracted c_0, c_1, c_2, …, c_k to estimate (in a probabilistic way) c_{k+1}? From there construct $\Delta_k = y_{ref}\, c_{k+1}\, Q^{k+1}$:truncation error

- Bayesian model:

Parameter cbar sets size of all dimensionless coefficients





First shot: cbar can be different at different kinematic points: "uncorrelated model"

Normal naturalness

Furnstahl, Klco, DP, Wesolowski, PRC, 2015; Melendez, Furnstahl, Wesolowski, PRC, 2017

- c_n's are normally distributed, with mean 0 and standard deviation cbar. that is a) fixed or b) distributed uniformly in its logarithm

$$\mathrm{pr}(c_n|\bar{c}) = \frac{1}{\sqrt{2\pi}\bar{c}} e^{-c_n^2/2\bar{c}^2}; \ \mathrm{pr}(\bar{c}) \propto \frac{1}{\bar{c}}\theta(\bar{c} - \bar{c}_<)\theta(\bar{c}_> - \bar{c})$$

- Marginalization:

$$\mathrm{pr}(c_{k+1}|c_0, c_1, \ldots, c_k) = \int_0^\infty d\bar{c}\, \mathrm{pr}(c_{k+1}|\bar{c})\mathrm{pr}(\bar{c}|c_0, c_1, \ldots c_k)$$

$$= \int_0^\infty \frac{d\bar{c}}{\bar{c}^{k+3}} \exp\left(-\frac{c_{k+1}^2}{2\bar{c}^2}\right) \exp\left(-\frac{(k+1)\langle c^2\rangle}{2\bar{c}^2}\right)$$

- Student's t-distribution results:

$$\mathrm{pr}(c_{k+1}|c_0, c_1, \ldots, c_k) \propto \frac{\Gamma\left(\frac{k+2}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)} \left(\frac{(k+1)\langle c^2\rangle}{(k+1)\langle c^2\rangle + c_{k+1}^2}\right)^{(k+2)/2}$$

- DoB intervals computed using known results for this distribution. Size of error bar set by $<c^2>$, k, Q^{k+1}, and y_{ref}.

NN scattering

Epelbaum, Krebs, Meißner, PRC, 2015

Employ "semi-local" potentials of Epelbaum, Krebs, and Meißner

χEFT: $\mathscr{L}$(N,π)→V$^{(k)}$→δ→σ_{np}

$$\sigma_{np}(E_{\rm lab}) = \sigma_{\rm LO} \sum_{n=0}^{k} c_n(p_{\rm rel}) \left(\frac{p_{\rm rel}}{\Lambda_b}\right)^n$$

$$x = \frac{p_{\rm rel}}{\Lambda_b}$$

NN scattering

Epelbaum, Krebs, Meißner, PRC, 2015

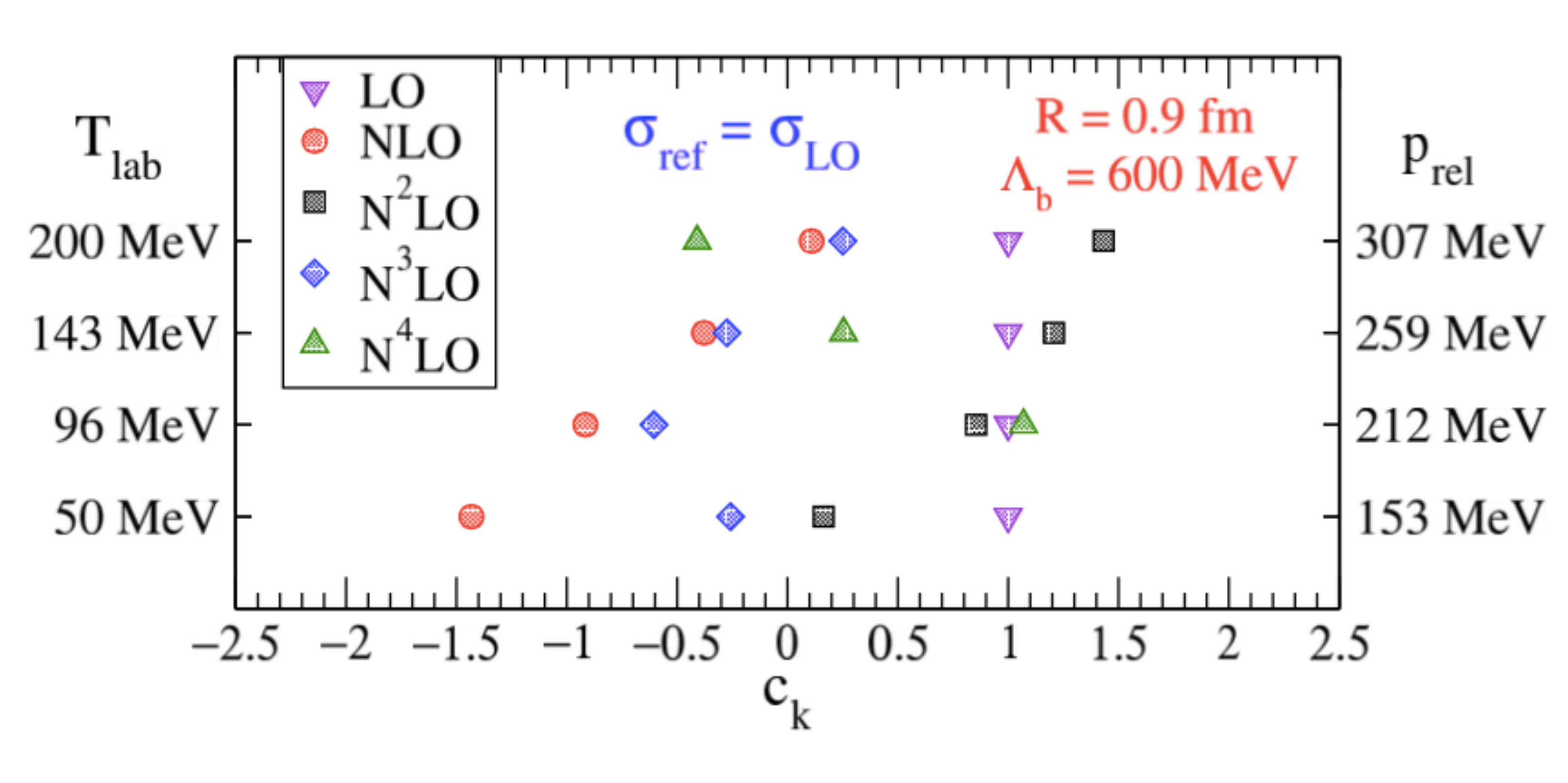
Employ "semi-local" potentials of Epelbaum, Krebs, and Meißner

χEFT: $\mathscr{L}(N,\pi) \rightarrow V^{(k)} \rightarrow \delta \rightarrow \sigma_{np}$

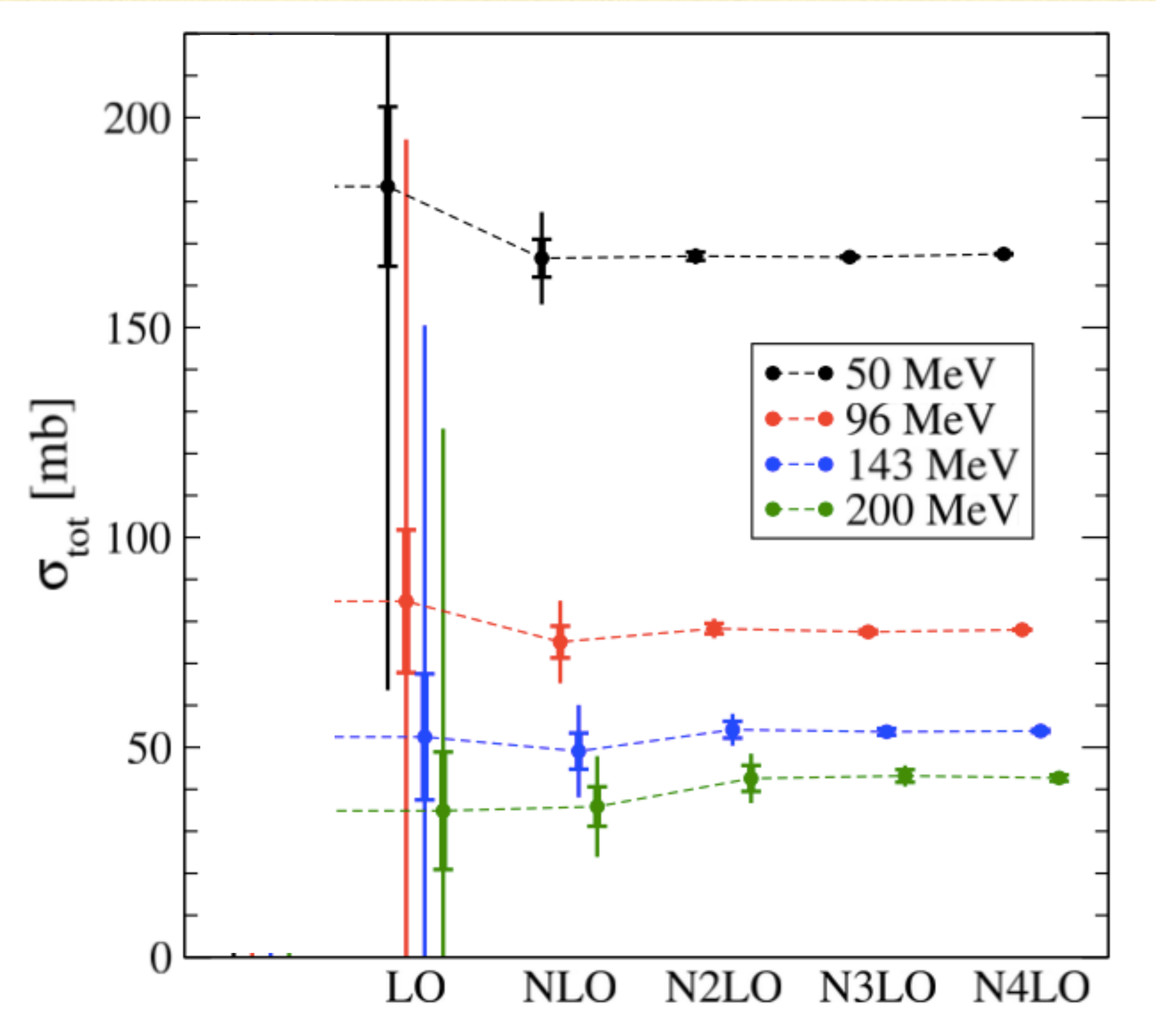
- NN cross section at T_{lab}=50, 96, 143, 200 MeV
- Potential regulated by local function, parameterized by R
- EKM identify Λ_b=600 MeV for smaller R values
- Here: R=0.9 fm data
- Results at LO, NLO, N^{2}LO, N^{3}LO, N^{4}LO (k=0, 2, 3, 4, 5)

$$\sigma_{np}(E_{\text{lab}}) = \sigma_{\text{LO}} \sum_{n=0}^{k} c_n(p_{\text{rel}}) \left(\frac{p_{\text{rel}}}{\Lambda_b}\right)^n$$

$$x = \frac{p_{\text{rel}}}{\Lambda_b}$$

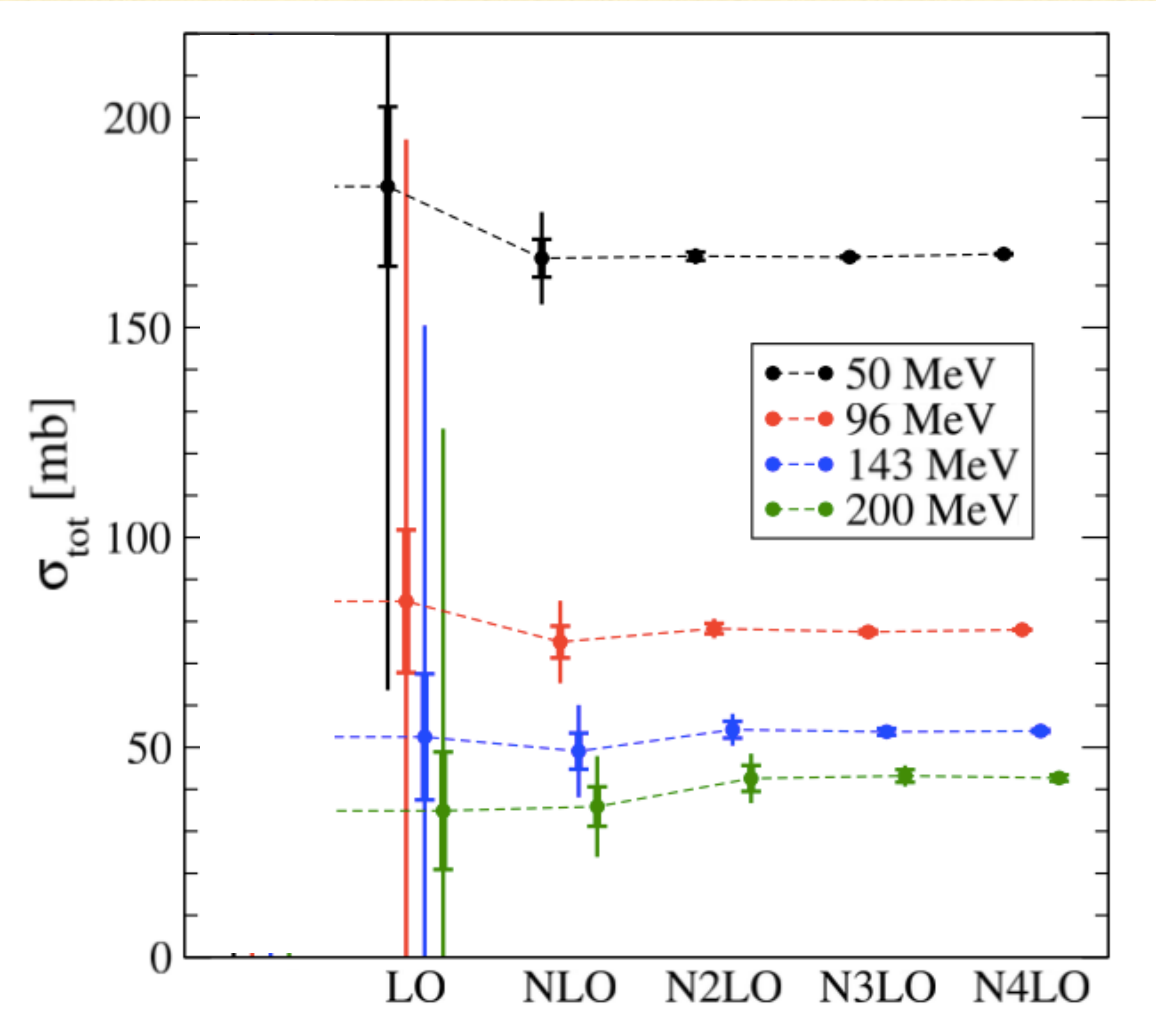


Results

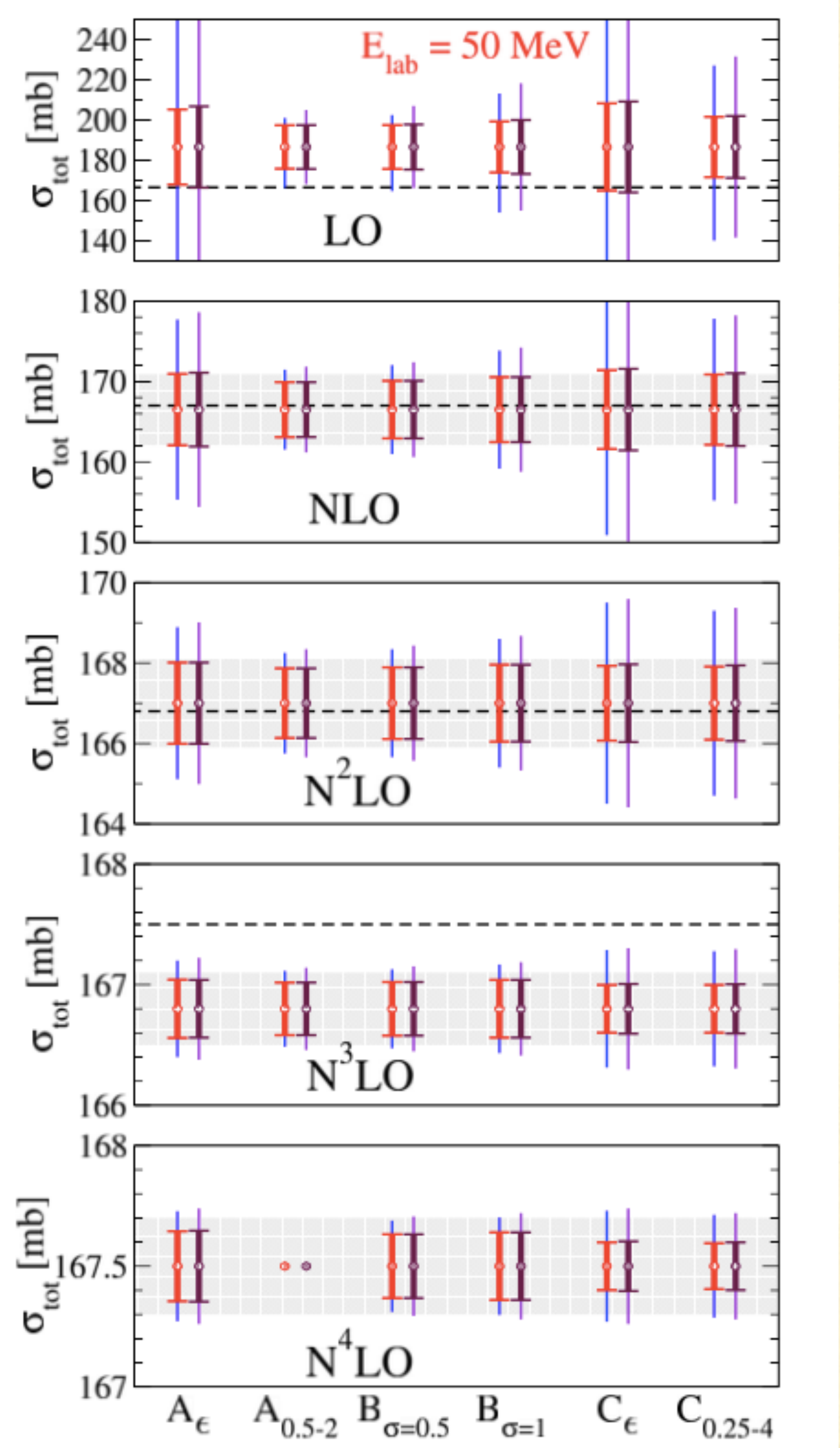


200
150
100
50
0
σ_{tot} [mb]
50 MeV
96 MeV
143 MeV
200 MeV
LO
NLO
N2LO
N3LO
N4LO

Results

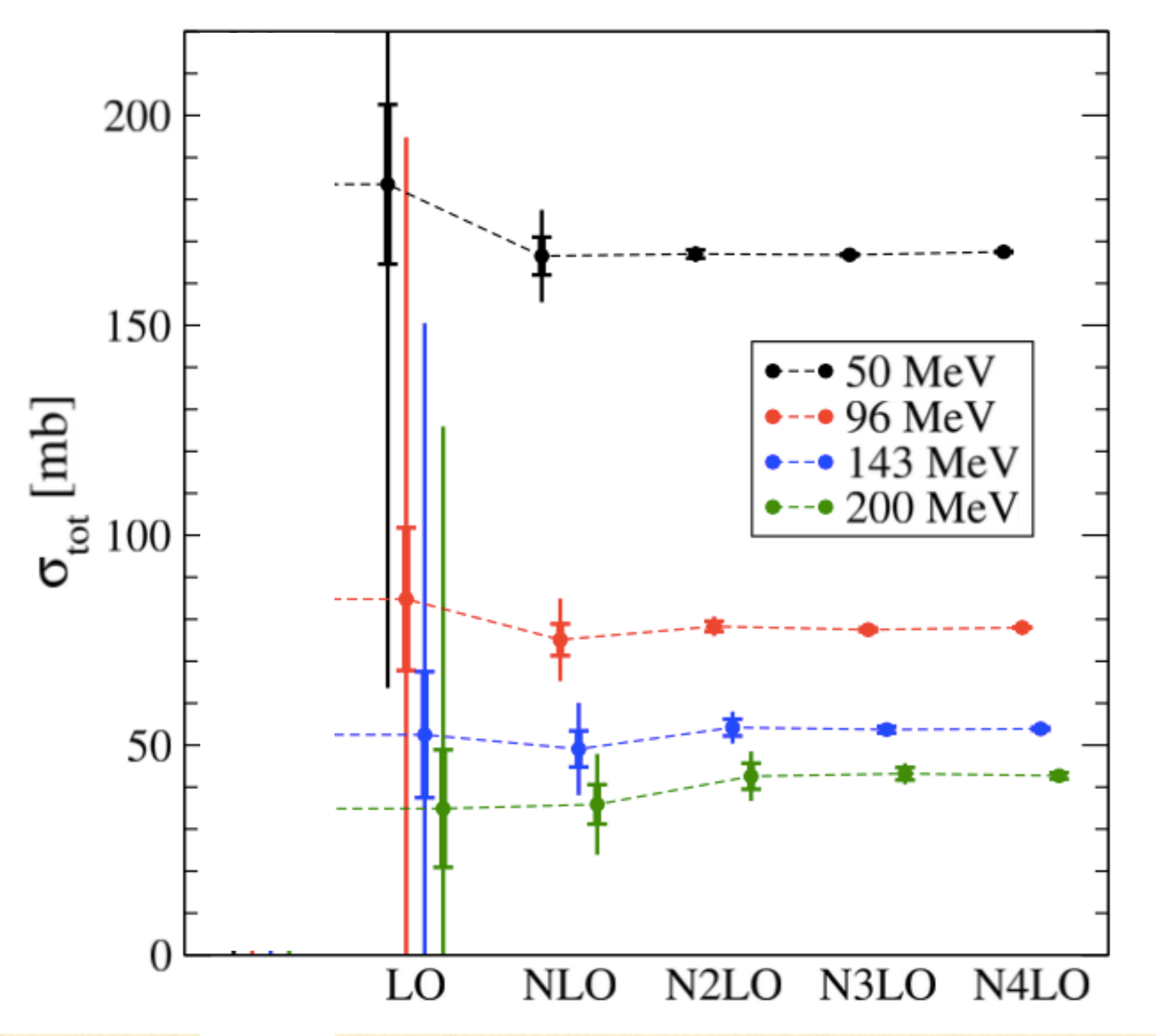


σ_{tot} [mb]
0
50
100
150
200
LO
NLO
N2LO
N3LO
N4LO
50 MeV
96 MeV
143 MeV
200 MeV

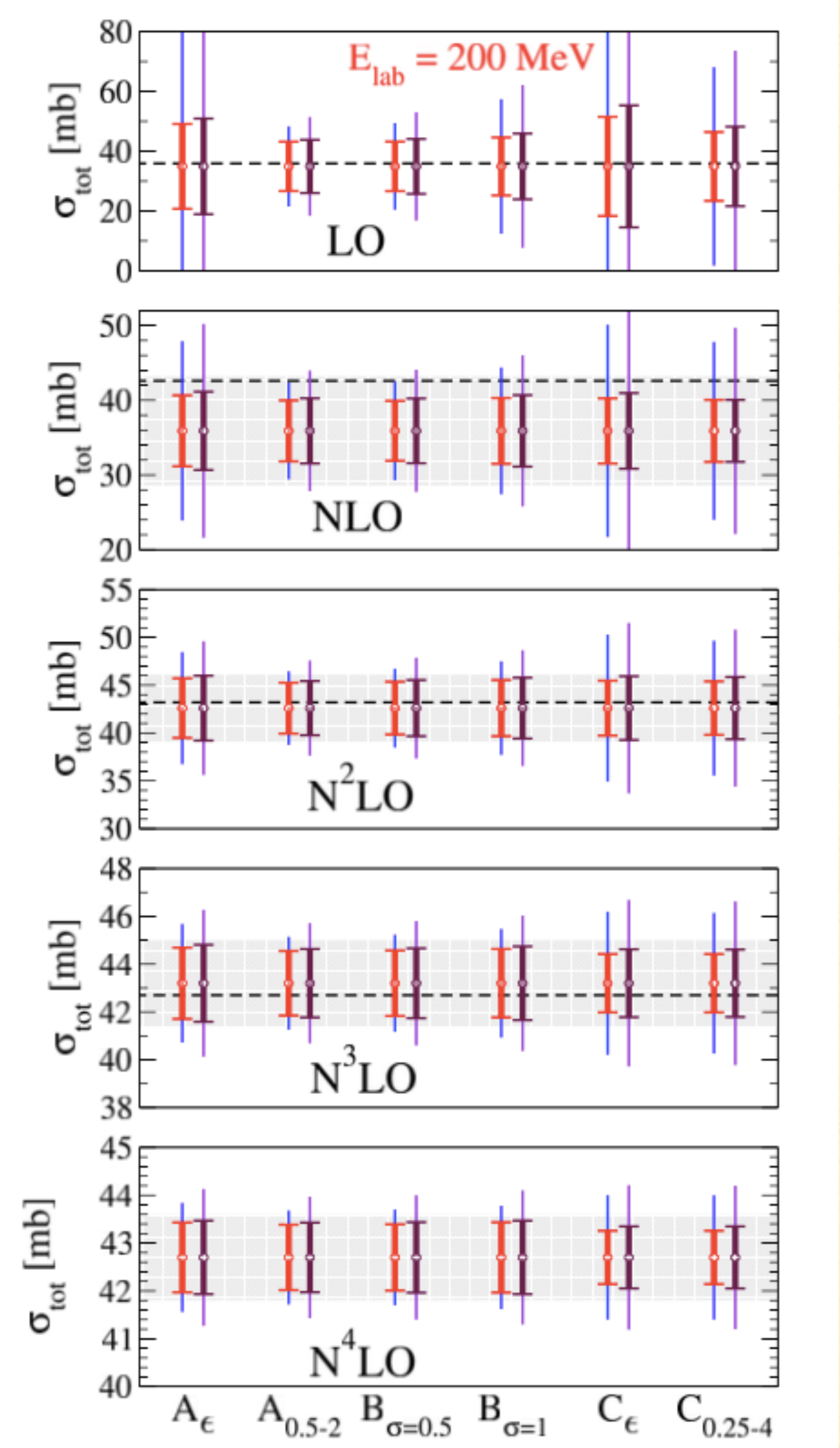


E_{lab} = 50 MeV
σ_{tot} [mb]
LO
NLO
N^2LO
N^3LO
N^4LO
A_ϵ
$A_{0.5-2}$
$B_{\sigma=0.5}$
$B_{\sigma=1}$
C_ϵ
$C_{0.25-4}$

Results



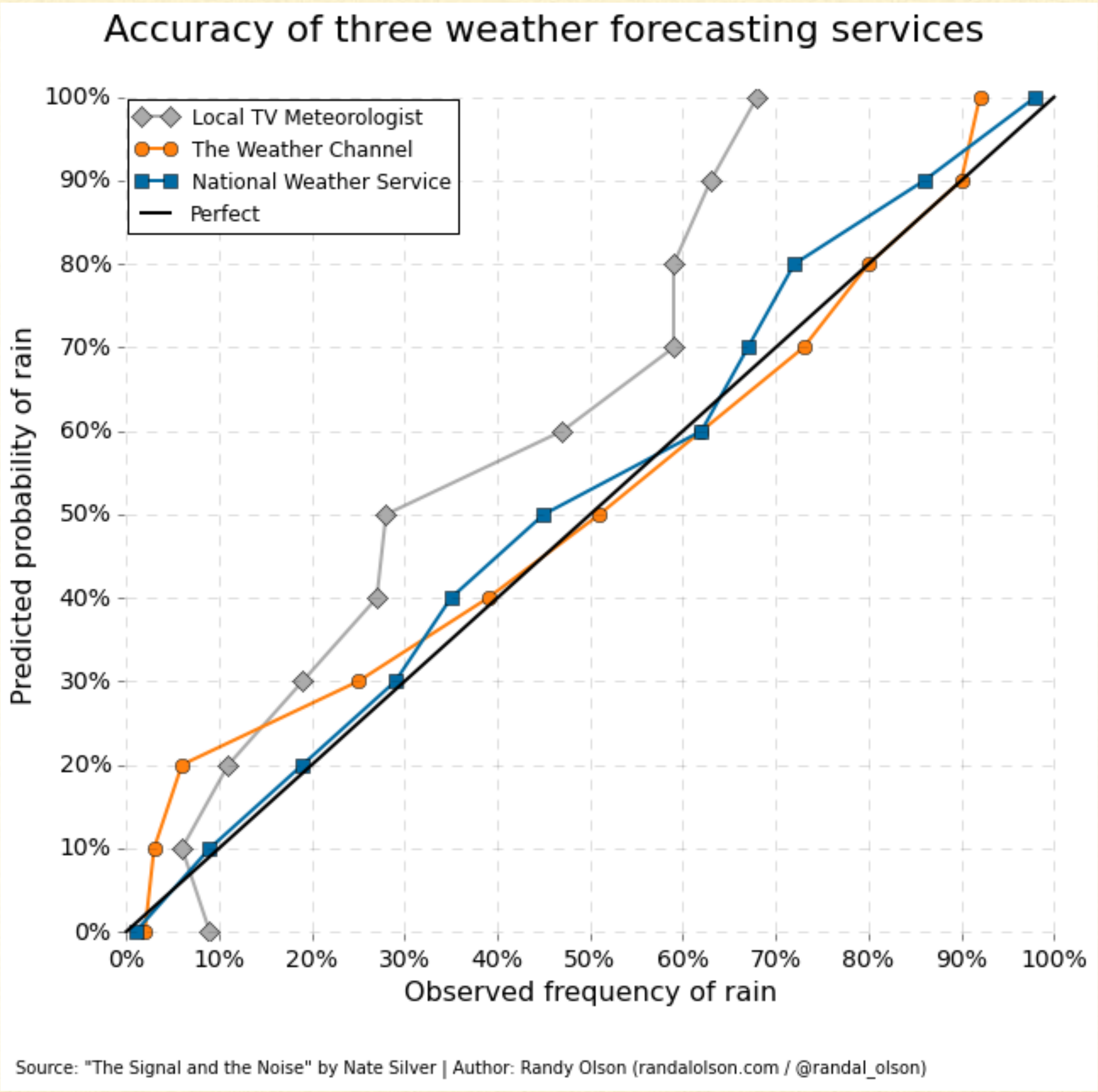
σtot [mb]
200
150
100
50
0
50 MeV
96 MeV
143 MeV
200 MeV
LO
NLO
N2LO
N3LO
N4LO



Elab = 200 MeV
σtot [mb]
LO
NLO
N2LO
N3LO
N4LO
Aϵ
A0.5-2
Bσ=0.5
Bσ=1
Cϵ
C0.25-4

Are these well calibrated?

Are these well calibrated?



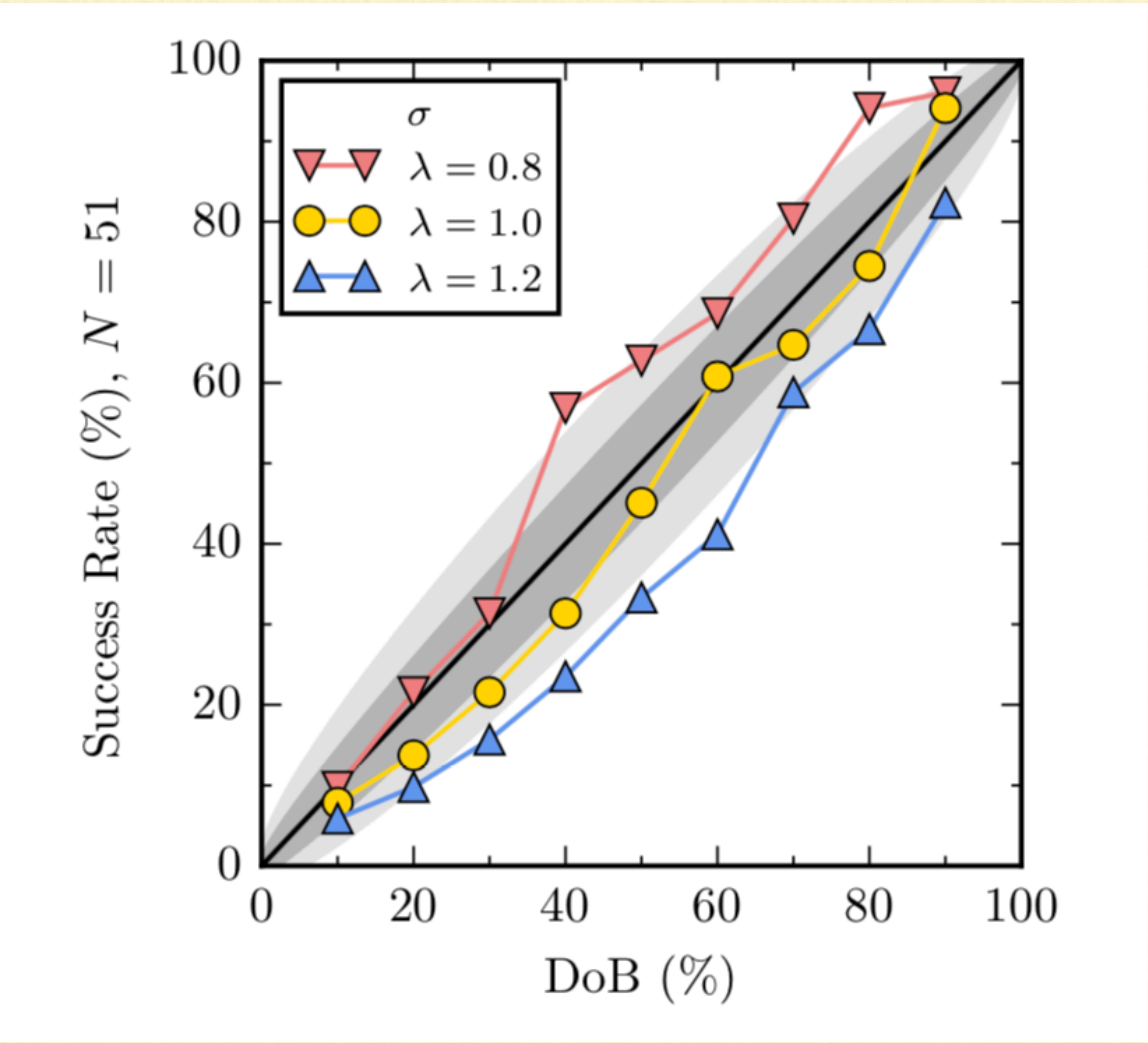
Melendez, Furnstahl, Wesolowski, PRC, 2017
after Furnstahl, Klco, DP, Wesolowski, PRC, 2015
after Bagnaschi, Cacciari, Guffanti, Jenniches, 2015

- Consider predictions at each order, with their error bars, as data and test them to see if the procedure is consistent

- Fix a given DOB interval: compute success ratio, compare

Are these well calibrated?

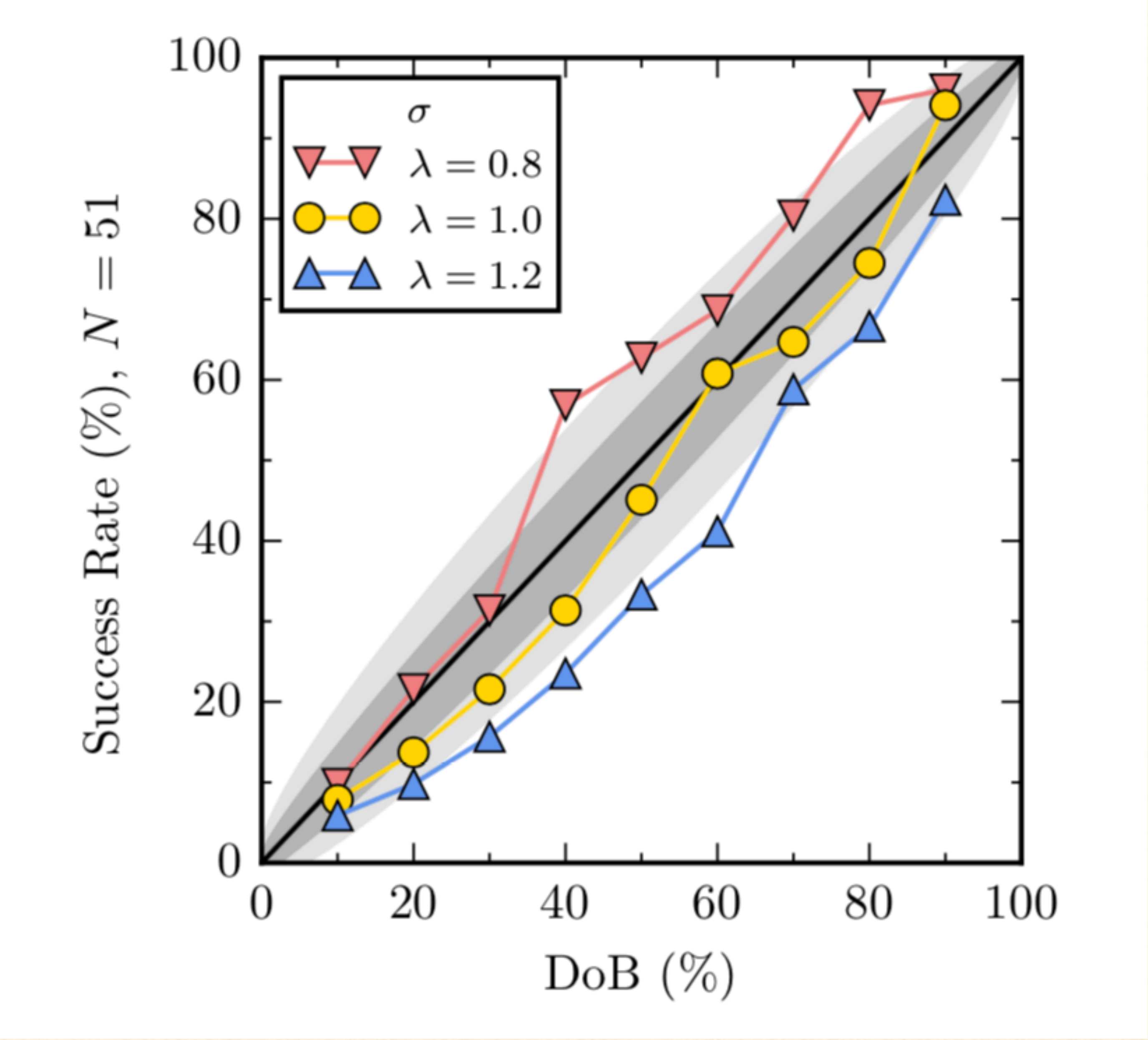
Melendez, Furnstahl, Wesolowski, PRC, 2017
after Furnstahl, Klco, DP, Wesolowski, PRC, 2015
after Bagnaschi, Cacciari, Guffanti, Jenniches, 2015



- Consider predictions at each order, with their error bars, as data and test them to see if the procedure is consistent
- Fix a given DOB interval: compute success ratio, compare
- Look at this for EKM predictions at three different orders and 17 different energies
- Interpret in terms of rescaling of Λ_b by a factor λ

Are these well calibrated?

Melendez, Furnstahl, Wesolowski, PRC, 2017
after Furnstahl, Klco, DP, Wesolowski, PRC, 2015
after Bagnaschi, Cacciari, Guffanti, Jenniches, 2015



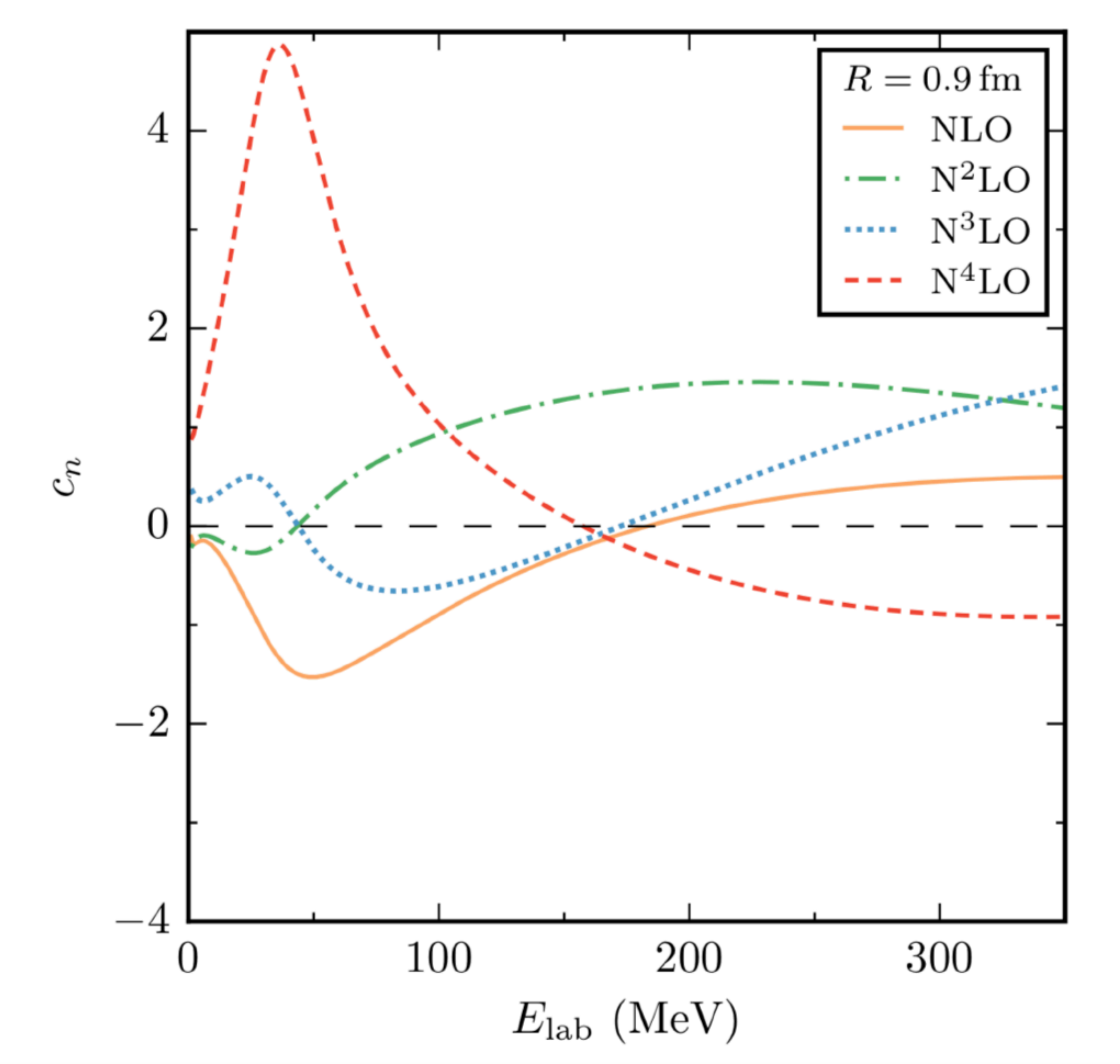
- Consider predictions at each order, with their error bars, as data and test them to see if the procedure is consistent
- Fix a given DOB interval: compute success ratio, compare
- Look at this for EKM predictions at three different orders and 17 different energies
- Interpret in terms of rescaling of Λ_b by a factor λ

No evidence for significant rescaling of Λ_b

Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

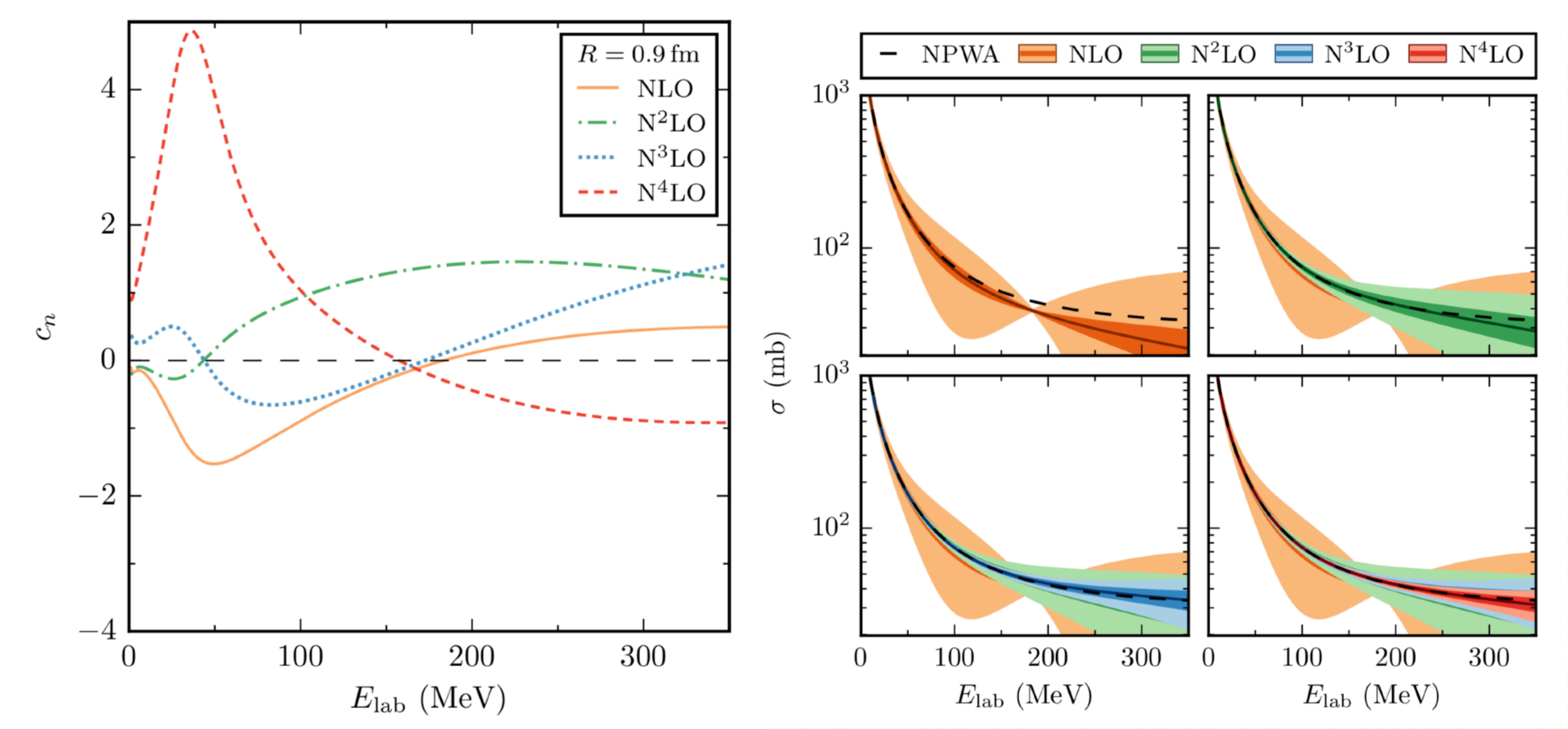
EKM R=0.9 fm potential



Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

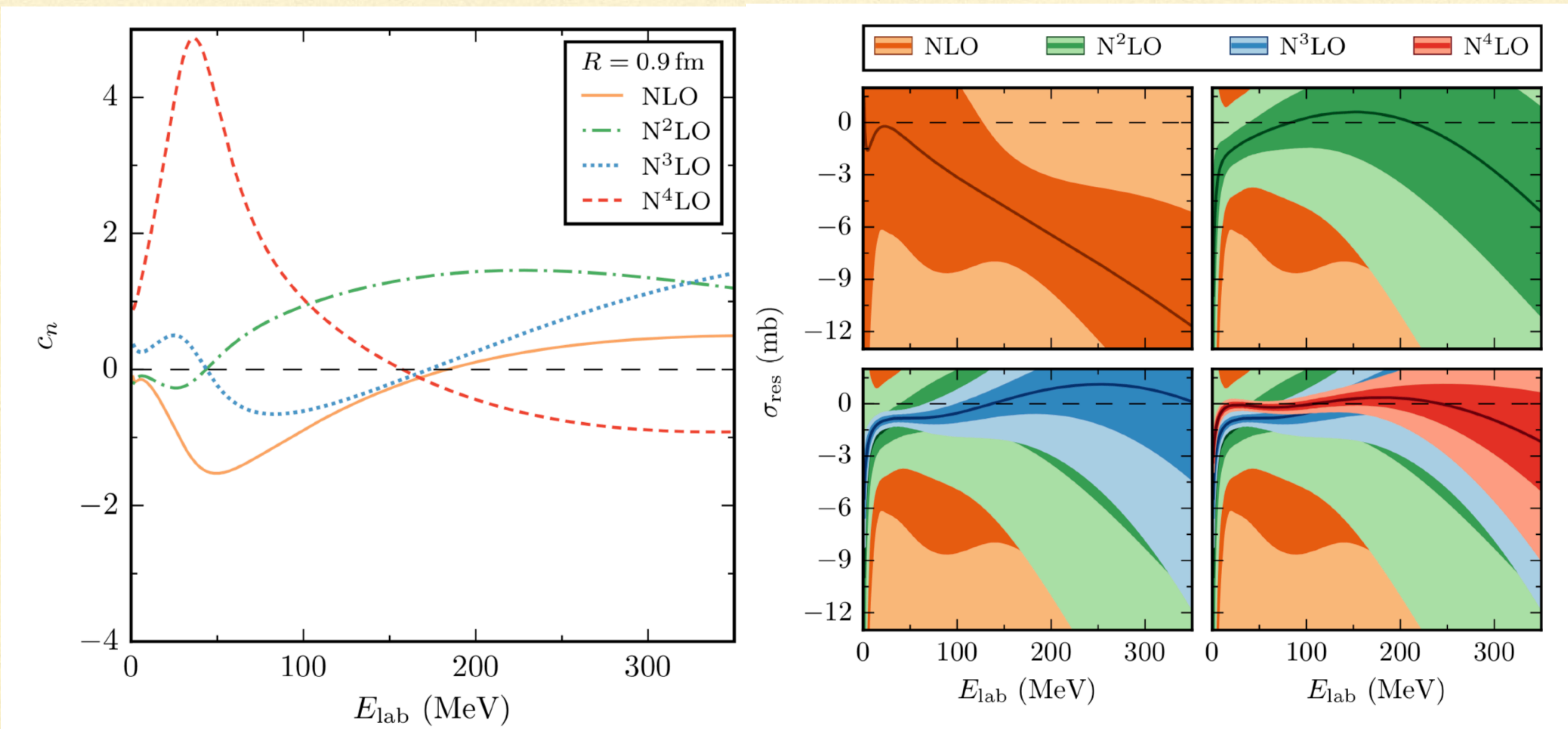
EKM R=0.9 fm potential



Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

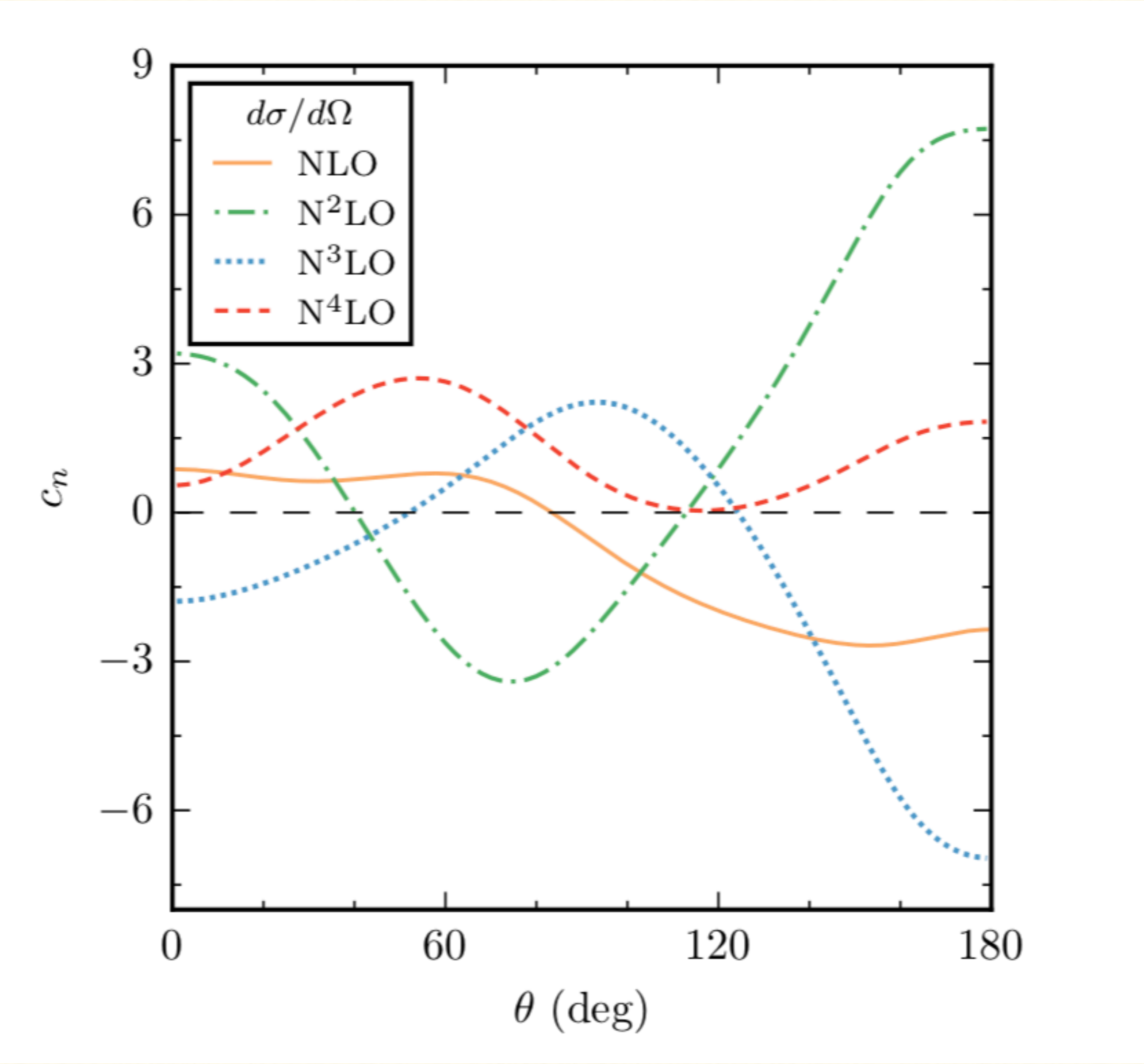
EKM R=0.9 fm potential



Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

EKM R=0.9 fm potential

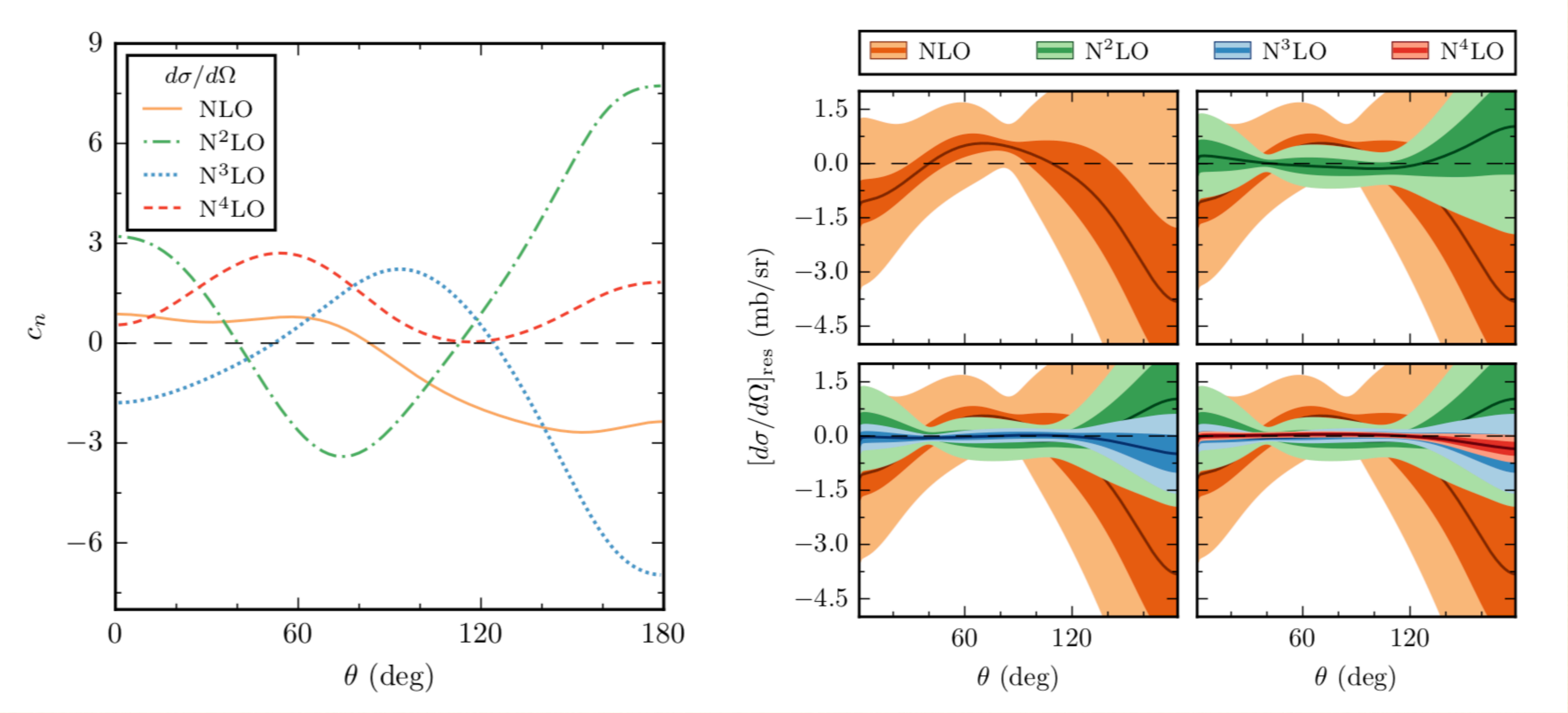


E_{lab}=96 MeV

Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

EKM R=0.9 fm potential

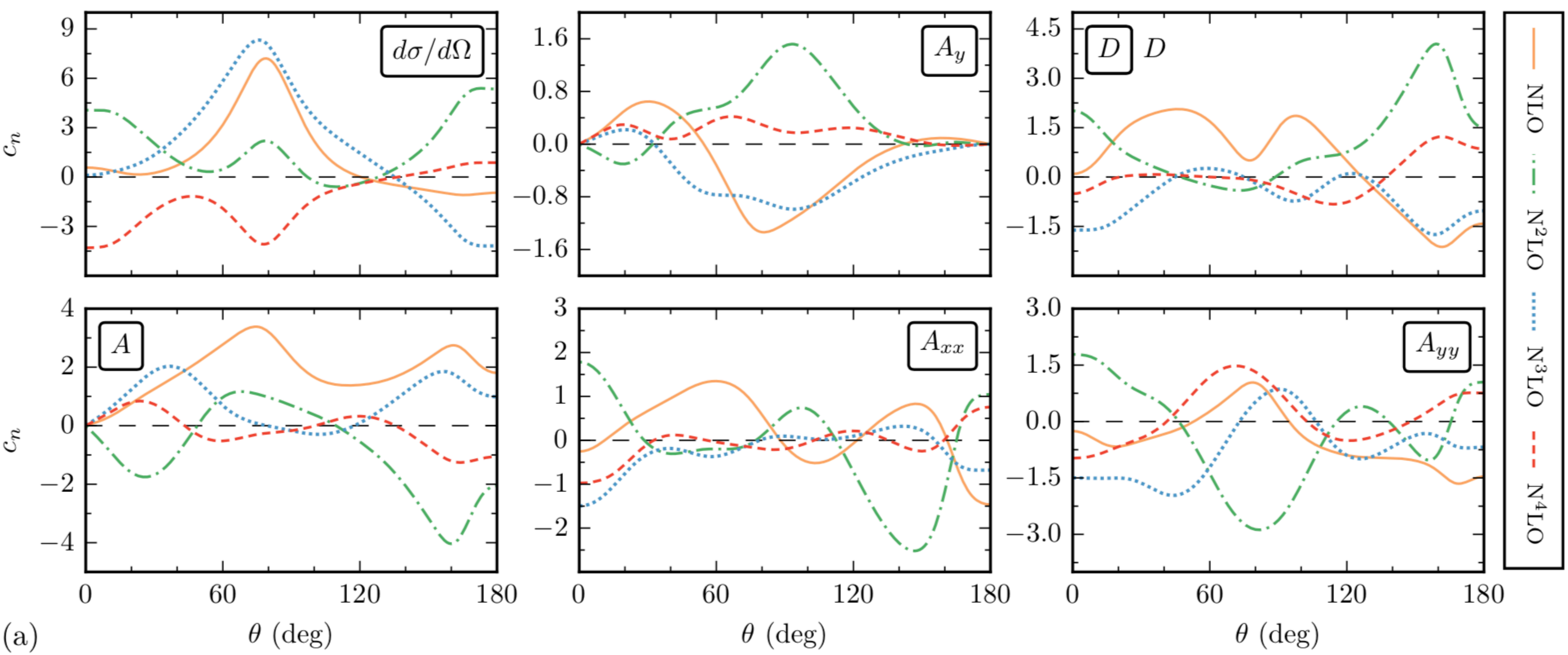


E_{lab}=96 MeV

Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

EKM R=0.9 fm potential

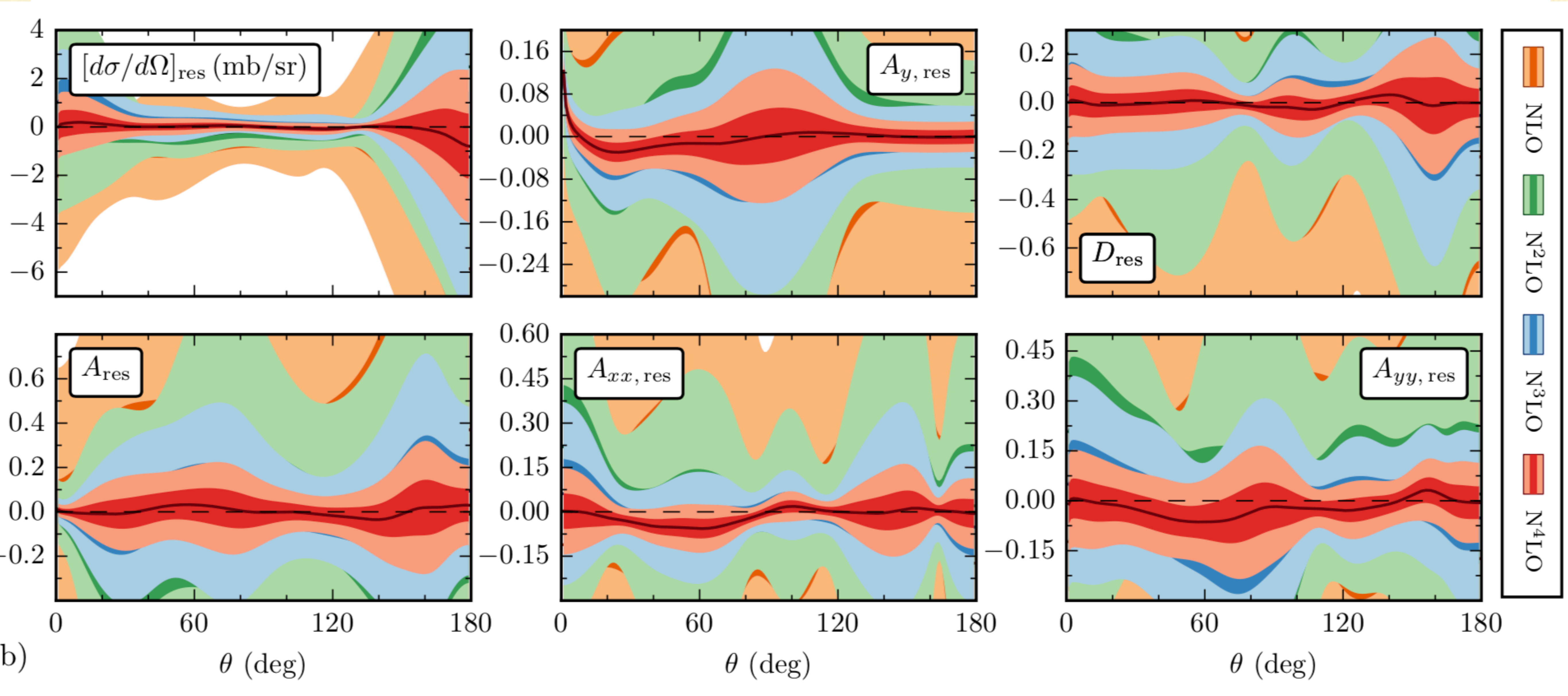


E_{lab}=250 MeV

Error bands for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

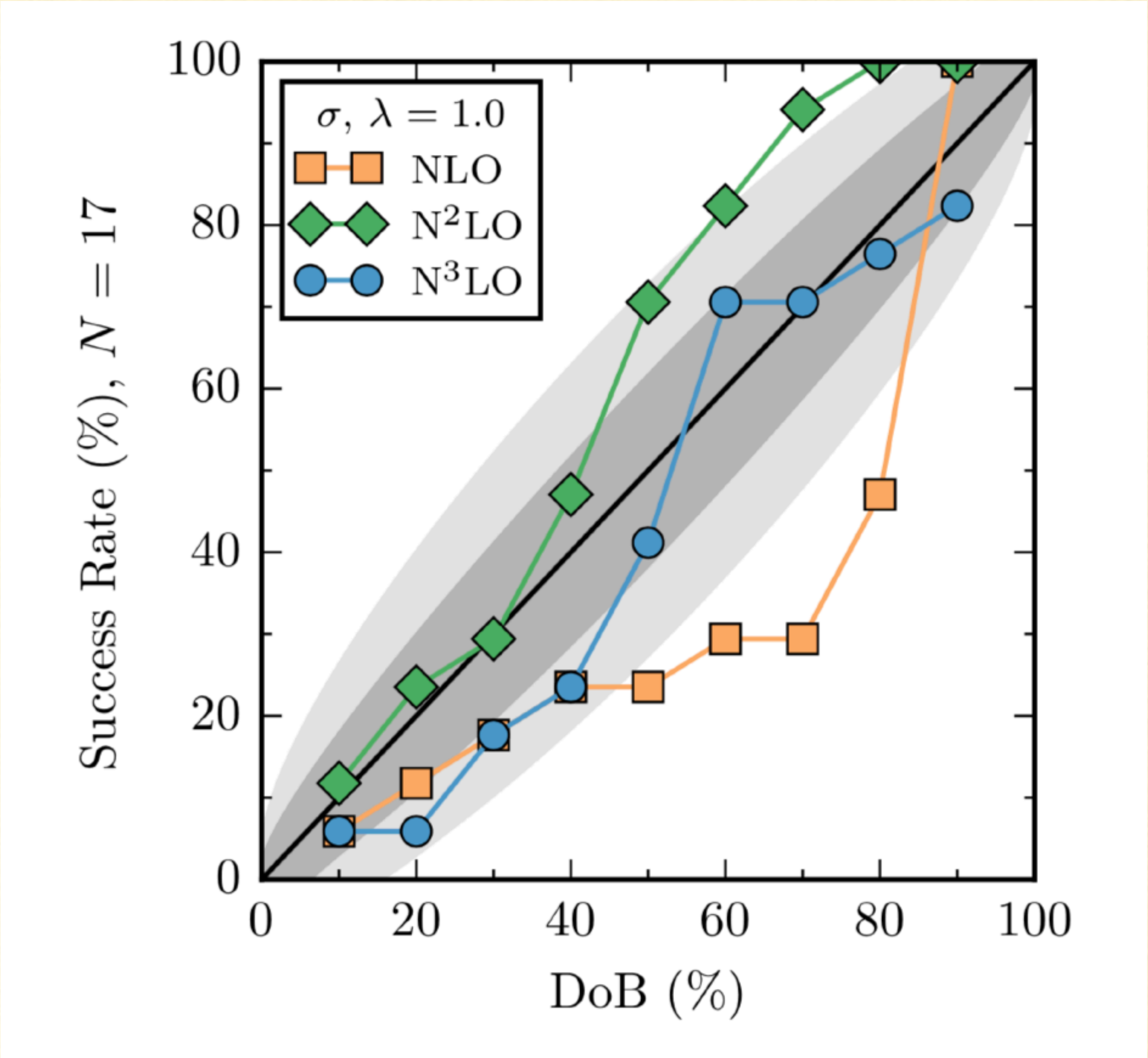
EKM R=0.9 fm potential



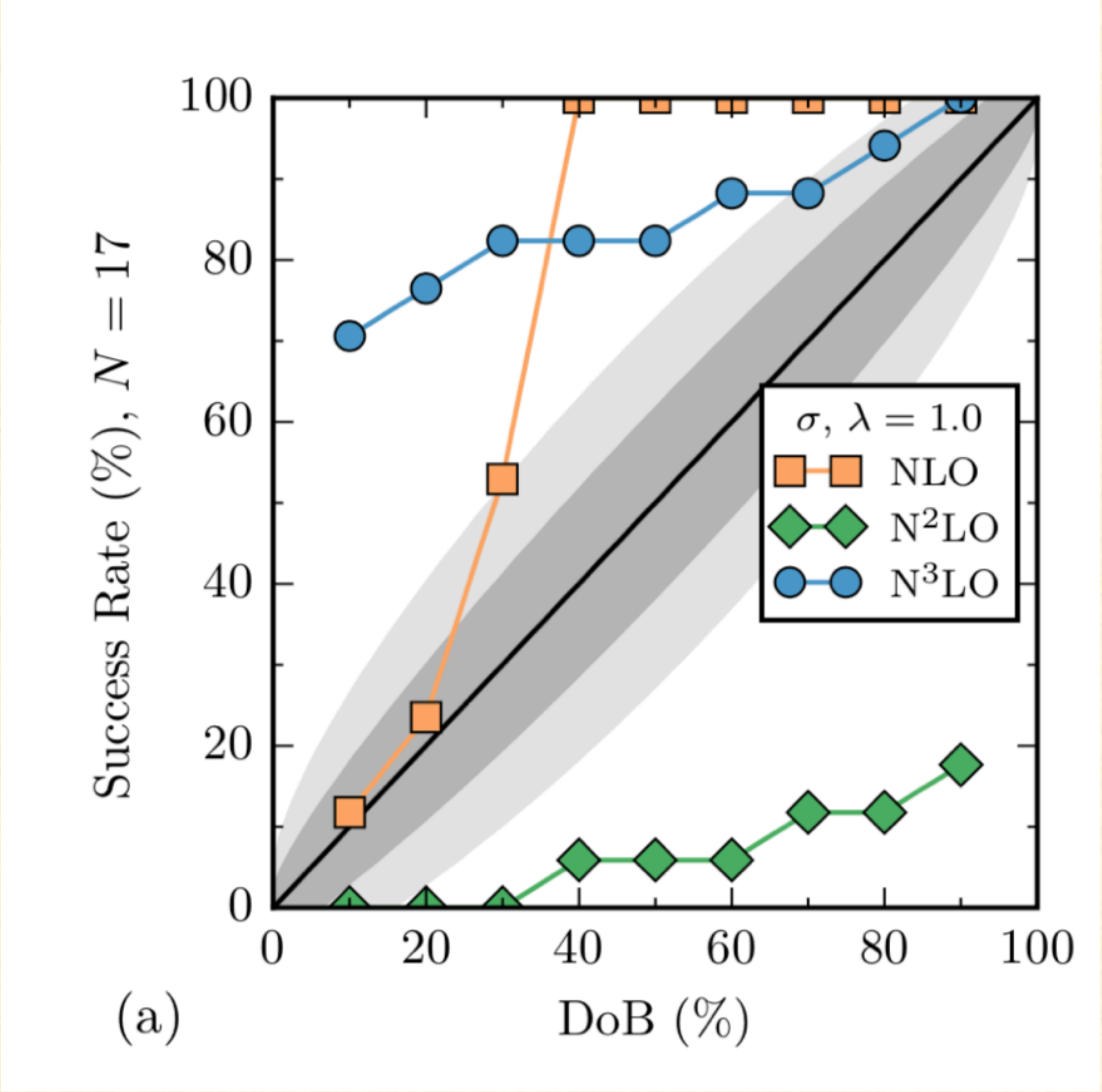
E_{lab}=250 MeV

Physics from calibration plots

R=0.9 fm



R=1.2 fm



- Allows assessment of order-by-order convergence
- Can look at differential cross section and spin observables too