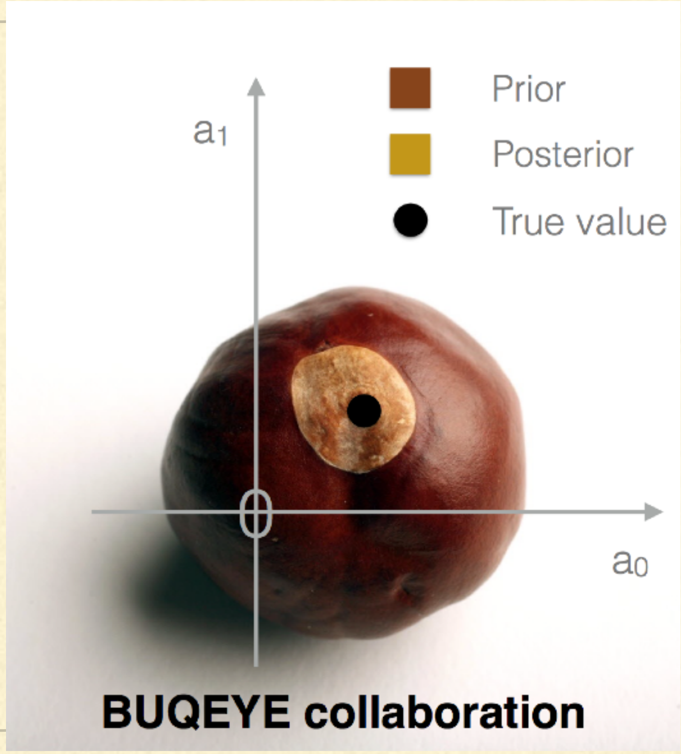
TALENT Course 11
Learning from Data: Bayesian methods and machine learning



Lecture 23: Model checking II

Daniel Phillips
Ohio University
TU Darmstadt
ExtreMe Matter Institute

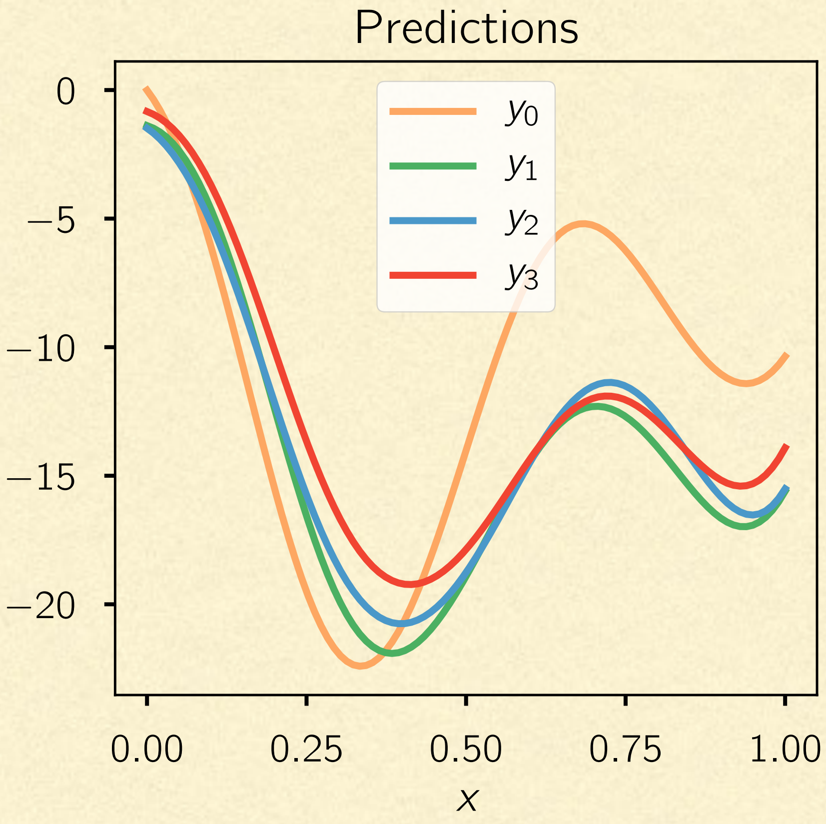
TALENT Course 11 is possible thanks to funding from the STFC

An EFT expansion in pictures

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

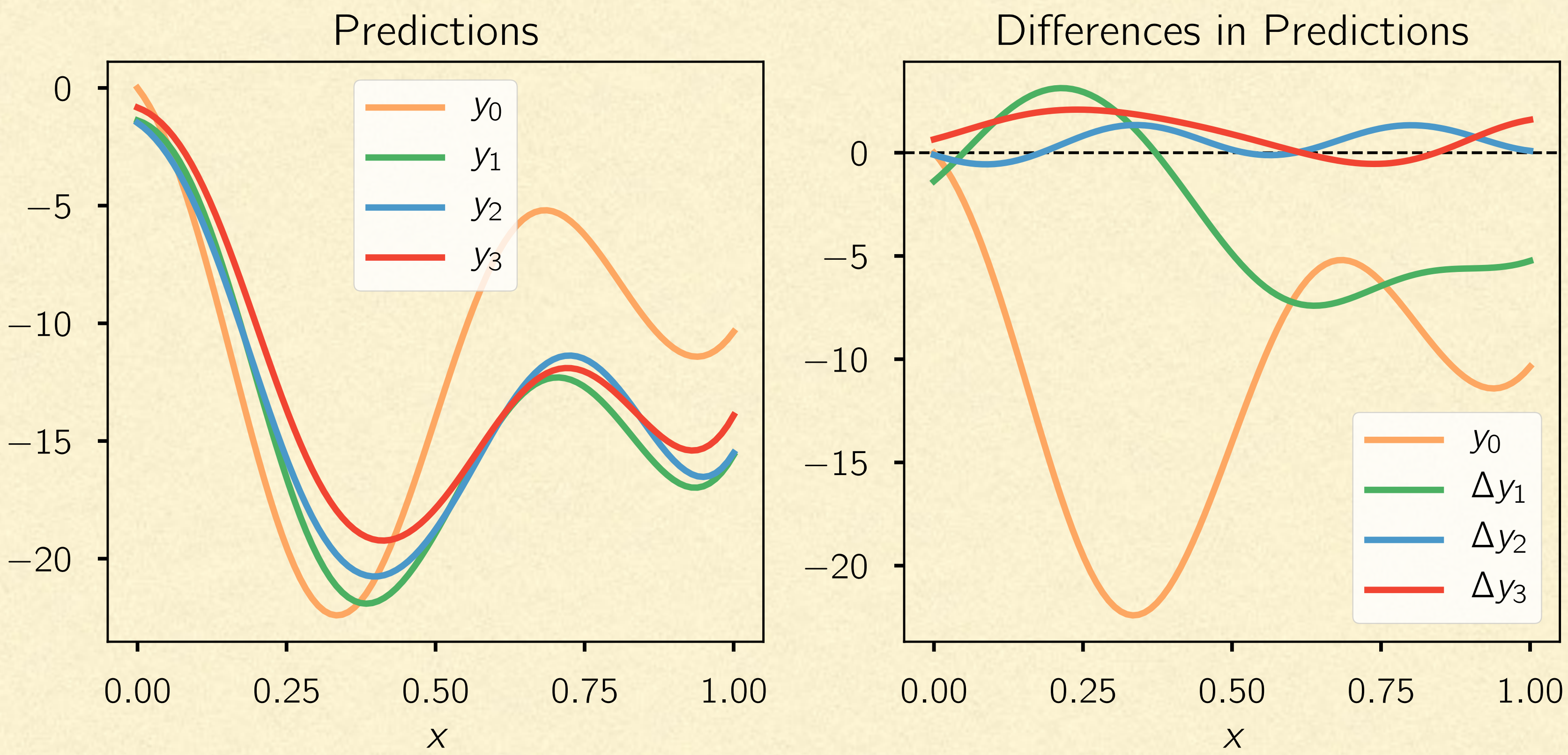
An EFT expansion in pictures

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



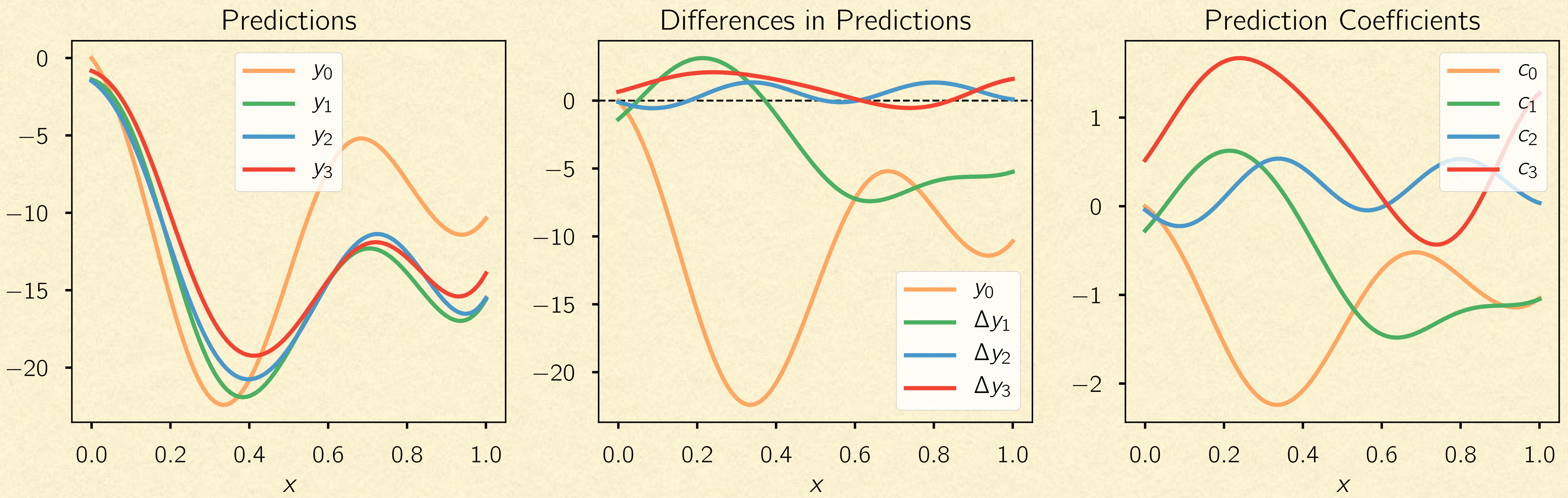
An EFT expansion in pictures

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



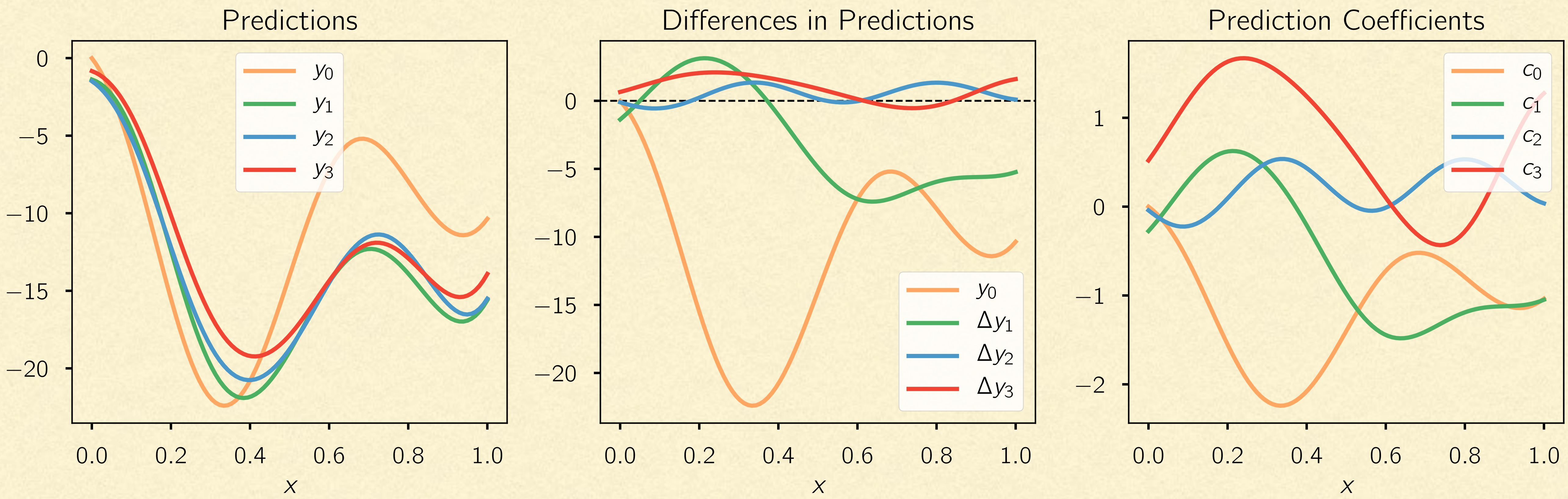
An EFT expansion in pictures

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



An EFT expansion in pictures

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



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

Bayesian EFT Parameter Estimation

$$X(p) = X_0(p) \sum_{i=0}^{k} c_i(\{a_j\}) x^i$$

Bayesian EFT Parameter Estimation

- General EFT series for observable to order k: $X(p) = X_0(p) \sum_{i=0}^{k} c_i(\{a_j\}) x^i$
- Truncation error=$X_0(p)\ c_{k+1}\ x^{k+1}$ + ...
- Posterior for parameters **a**={a_0,...,a_k} by marginalizing:

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\mathrm{max}}) = \int dc_{k+1} \ldots dc_{k_{\mathrm{max}}} \mathrm{pr}(\mathbf{a}, c_{k+1}, \ldots, c_{k_{\mathrm{max}}}|D,k,k_{\mathrm{max}})$$

and Bayesing $= \int dc_{k+1} \ldots dc_{k_{\mathrm{max}}} \dfrac{\mathrm{pr}(D|\mathbf{a}, c_{k+1}, \ldots, c_{k_{\mathrm{max}}}, k, k_{\mathrm{max}})\mathrm{pr}(\mathbf{a}|\bar{a}_{\mathrm{fix}}) \prod_{j=k+1}^{k_{\mathrm{max}}} \mathrm{pr}(c_j|\bar{c}_{\mathrm{fix}})}{\mathrm{pr}(D|k,k_{\mathrm{max}})}$

Bayesian EFT Parameter Estimation

- General EFT series for observable to order k: $X(p) = X_0(p) \sum_{i=0}^{k} c_i(\{a_j\}) x^i$
- Truncation error=$X_0(p)\ c_{k+1}\ x^{k+1}$ + ….
- Posterior for parameters **a**={a_0,…,a_k} by marginalizing:

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\max}) = \int dc_{k+1}\ldots dc_{k_{\max}} \mathrm{pr}(\mathbf{a},c_{k+1},\ldots,c_{k_{\max}}|D,k,k_{\max})$$

and Bayesing $= \int dc_{k+1}\ldots dc_{k_{\max}} \dfrac{\mathrm{pr}(D|\mathbf{a},c_{k+1},\ldots,c_{k_{\max}},k,k_{\max})\mathrm{pr}(\mathbf{a}|\bar{a}_{\mathrm{fix}})\prod_{j=k+1}^{k_{\max}}\mathrm{pr}(c_j|\bar{c}_{\mathrm{fix}})}{\mathrm{pr}(D|k,k_{\max})}$

- Marginalization over c's produces revised correlation matrix in standard likelihood, accounts for uncertainties (and correlation structure) induced by omitted terms

$$\mathrm{pr}(\mathbf{a}\,|\,D) \propto \exp\left(-\frac{1}{2}\mathbf{r}^T(\mathbf{\Sigma}_{\mathrm{exp}}+\mathbf{\Sigma}_{\mathrm{th}})^{-1}\mathbf{r}\right)\exp\left(-\frac{\mathbf{a}^2}{2\bar{a}^2}\right) \qquad \mathbf{r} \equiv \mathbf{y}_{\mathrm{exp}} - \mathbf{y}_{\mathrm{th}}$$

Bayesian EFT Parameter Estimation

- General EFT series for observable to order k: $X(p) = X_0(p) \sum_{i=0}^{k} c_i(\{a_j\}) x^i$
- Truncation error=$X_0(p)\ c_{k+1}\ x^{k+1}$ + ….
- Posterior for parameters **a**={a_0,…,a_k} by marginalizing:

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\max}) = \int dc_{k+1}\dots dc_{k_{\max}}\mathrm{pr}(\mathbf{a},c_{k+1},\dots,c_{k_{\max}}|D,k,k_{\max})$$

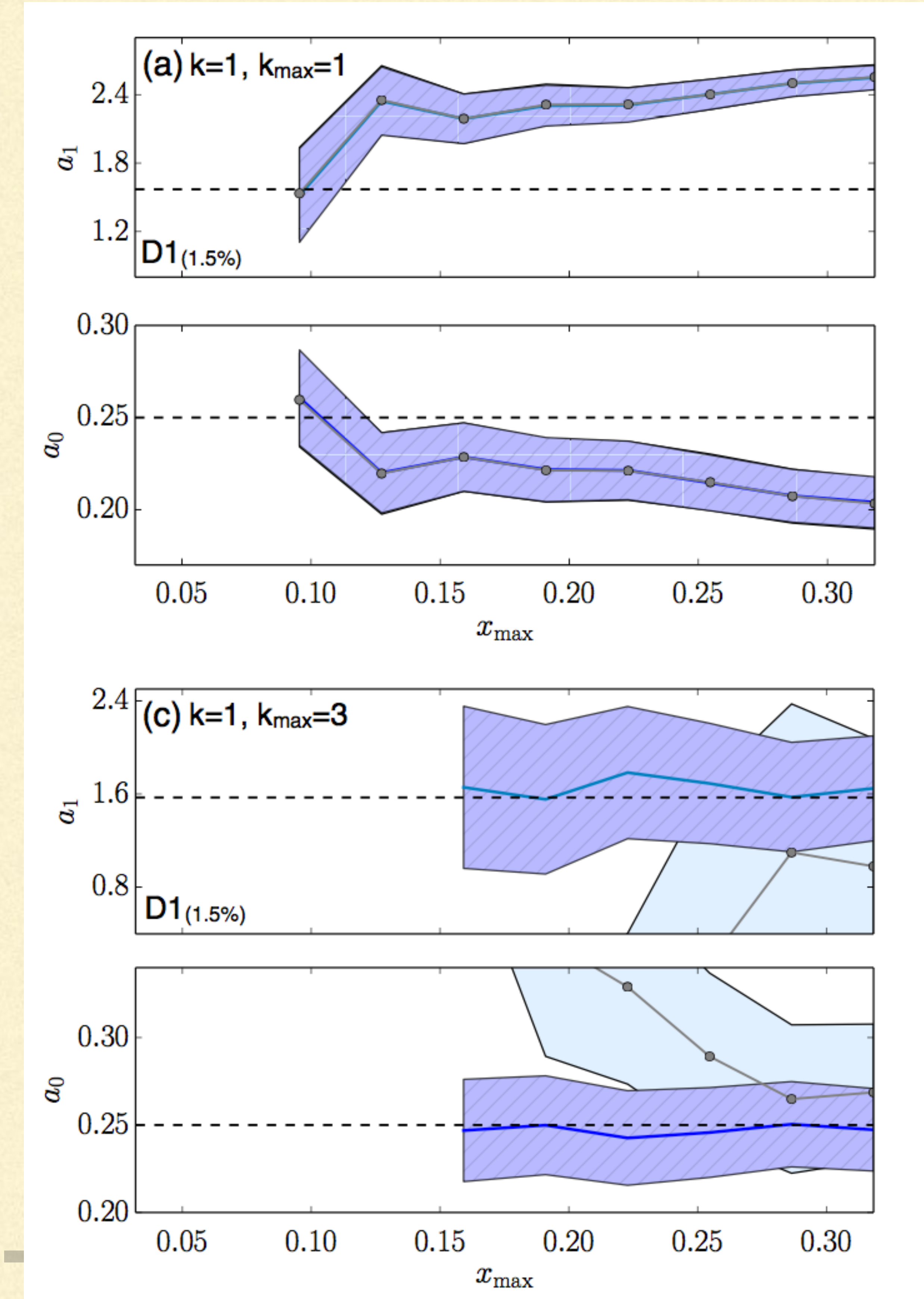
and Bayesing $= \int dc_{k+1}\dots dc_{k_{\max}} \dfrac{\mathrm{pr}(D|\mathbf{a},c_{k+1},\dots,c_{k_{\max}},k,k_{\max})\mathrm{pr}(\mathbf{a}|\bar{a}_{\mathrm{fix}})\prod_{j=k+1}^{k_{\max}}\mathrm{pr}(c_j|\bar{c}_{\mathrm{fix}})}{\mathrm{pr}(D|k,k_{\max})}$

- Marginalization over c's produces revised correlation matrix in standard likelihood, accounts for uncertainties (and correlation structure) induced by omitted terms

$$\mathrm{pr}(\mathbf{a}\,|\,D) \propto \exp\left(-\frac{1}{2}\mathbf{r}^T(\mathbf{\Sigma}_{\mathrm{exp}}+\mathbf{\Sigma}_{\mathrm{th}})^{-1}\mathbf{r}\right)\exp\left(-\frac{\mathbf{a}^2}{2\bar{a}^2}\right) \qquad \mathbf{r}\equiv\mathbf{y}_{\mathrm{exp}}-\mathbf{y}_{\mathrm{th}}$$

- Normal distribution for LECs. Here we take a fixed $\bar{a}$, could also marginalize over it.

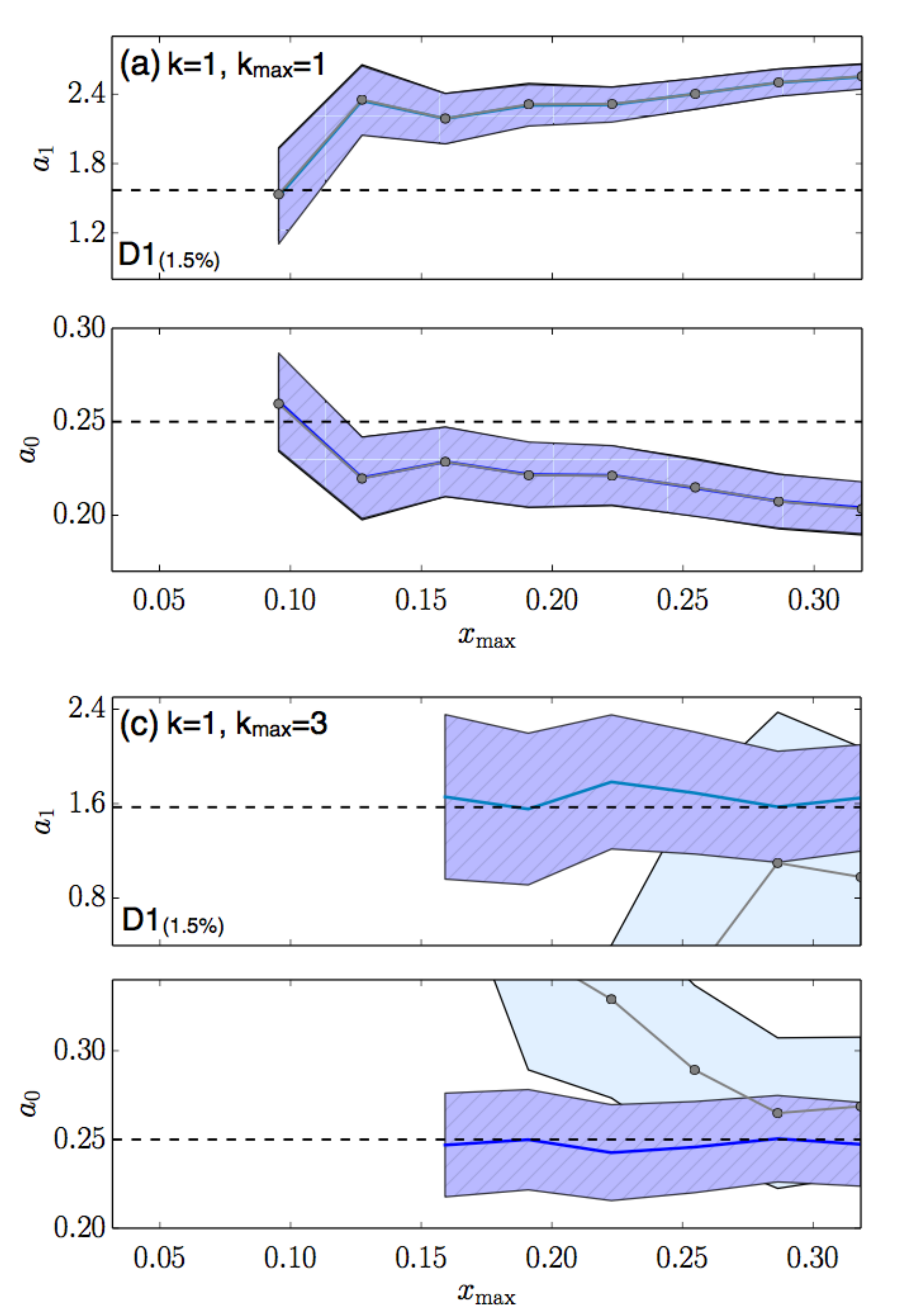
Truncation Errors Help



Wesolowski et al., JPG (2016)

$$(\boldsymbol{\Sigma}_{\text{th,corr}})_{ij} = (\mathbf{y}_{\text{ref}})_i(\mathbf{y}_{\text{ref}})_j\bar{c}^2 \sum_{n=k+1}^{k_{\text{max}}} x_i^n x_j^n$$

Truncation Errors Help

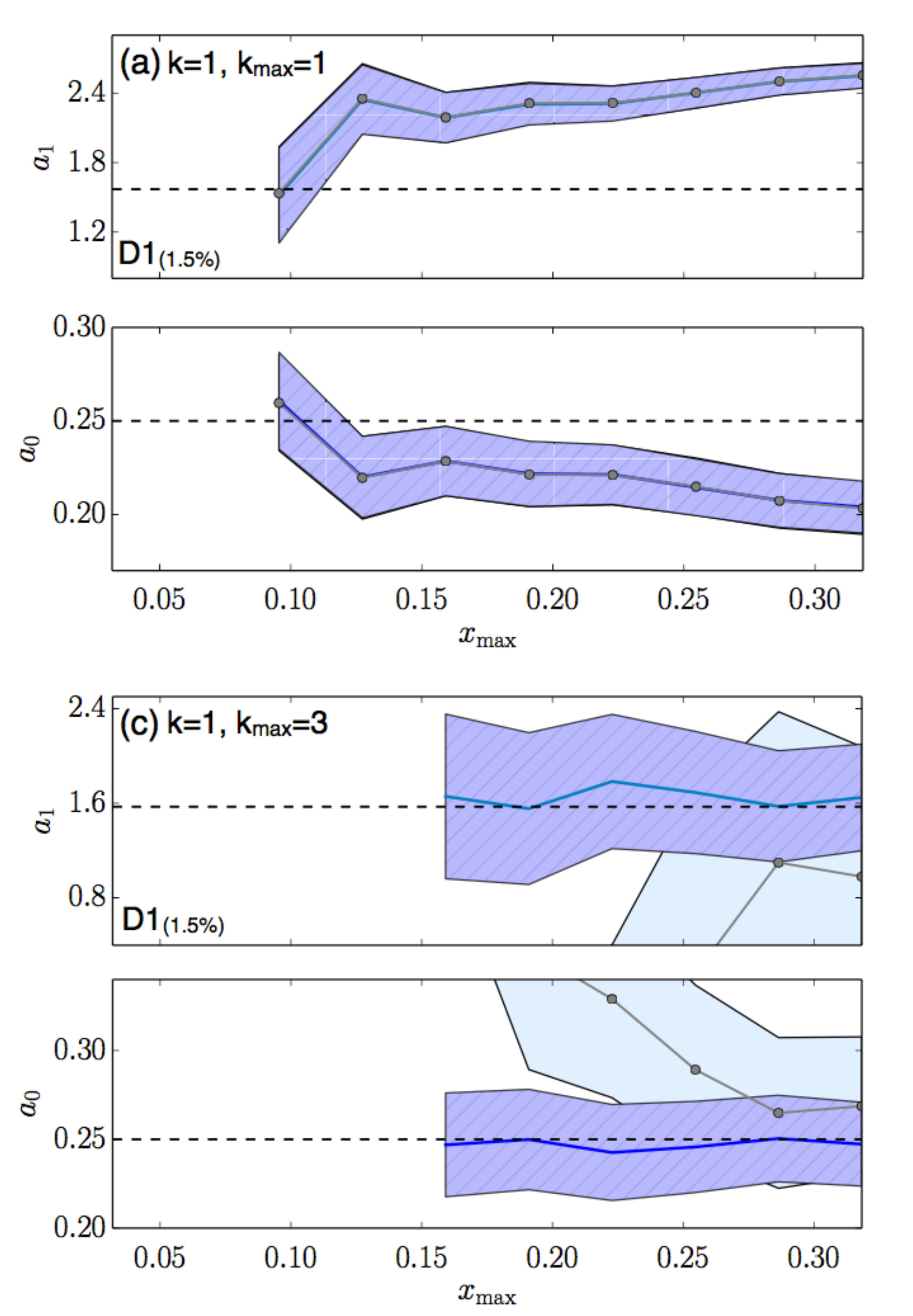


Wesolowski et al., JPG (2016)

$$(\Sigma_{\text{th,corr}})_{ij} = (\mathbf{y}_{\text{ref}})_i(\mathbf{y}_{\text{ref}})_j\bar{c}^2 \sum_{n=k+1}^{k_{\max}} x_i^n x_j^n$$

- Plot results for c_0 and c_1 for fit to data at $x \in [0, x_{max}]$ and look for stability as a function of x_{max}

Truncation Errors Help

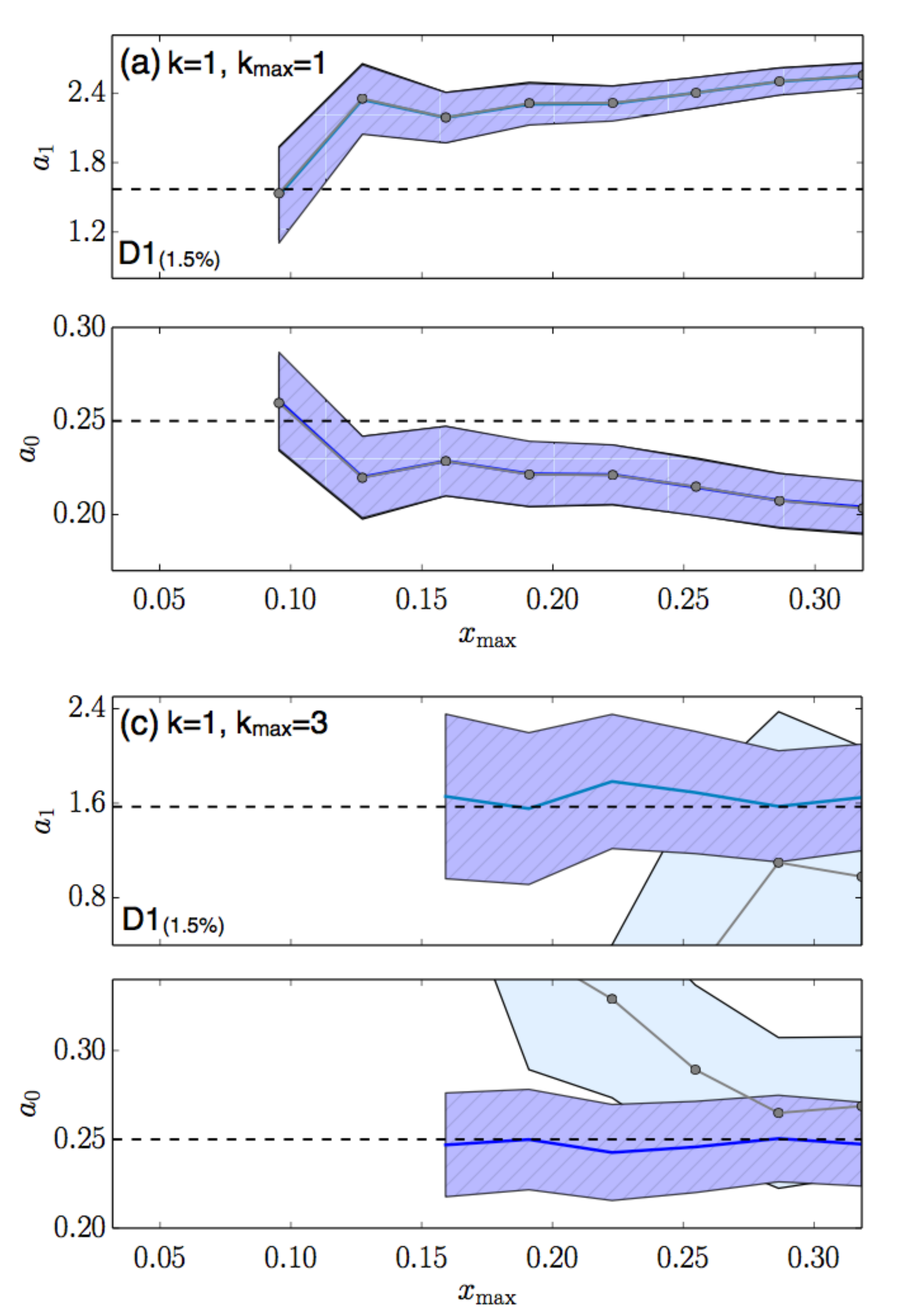


Wesolowski et al., JPG (2016)

$$(\Sigma_{\text{th,corr}})_{ij} = (\mathbf{y}_{\text{ref}})_i(\mathbf{y}_{\text{ref}})_j\bar{c}^2 \sum_{n=k+1}^{k_{\max}} x_i^n x_j^n$$

- Plot results for c_0 and c_1 for fit to data at $x \in [0,x_{max}]$ and look for stability as a function of x_{max}
- Lack of stability for $k=k_{max}=1$. Underfitting.

Truncation Errors Help



Wesolowski et al., JPG (2016)

$$(\Sigma_{\text{th,corr}})_{ij} = (\mathbf{y}_{\text{ref}})_i(\mathbf{y}_{\text{ref}})_j\bar{c}^2 \sum_{n=k+1}^{k_{\max}} x_i^n x_j^n$$

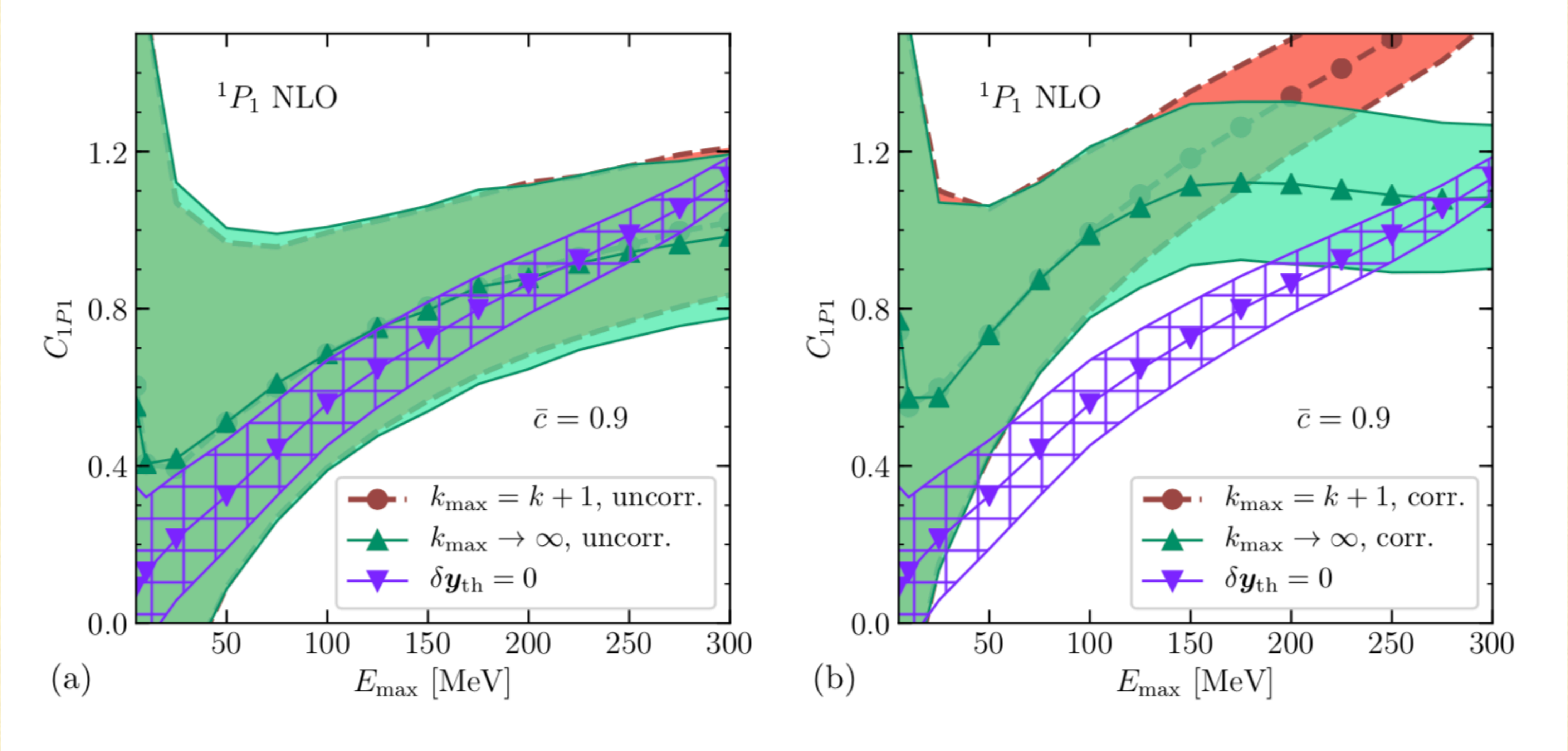
- Plot results for c_0 and c_1 for fit to data at $x \in [0,x_{max}]$ and look for stability as a function of x_{max}
- Lack of stability for $k=k_{max}=1$. Underfitting.
- $k=1$, $k_{max}=3$: once truncation errors are incorporated correct result is achieved earlier, and with smaller error bars, than for uniform-prior (least-squares) result

E_{max} plots for NN

Wesolowski, Furnstahl, Melendez, DP, JPG (2019)

Epelbaum, Krebs, Meißner, EPJA 51, 53 (2015)

- EKM semi-local potential with R=0.9 fm

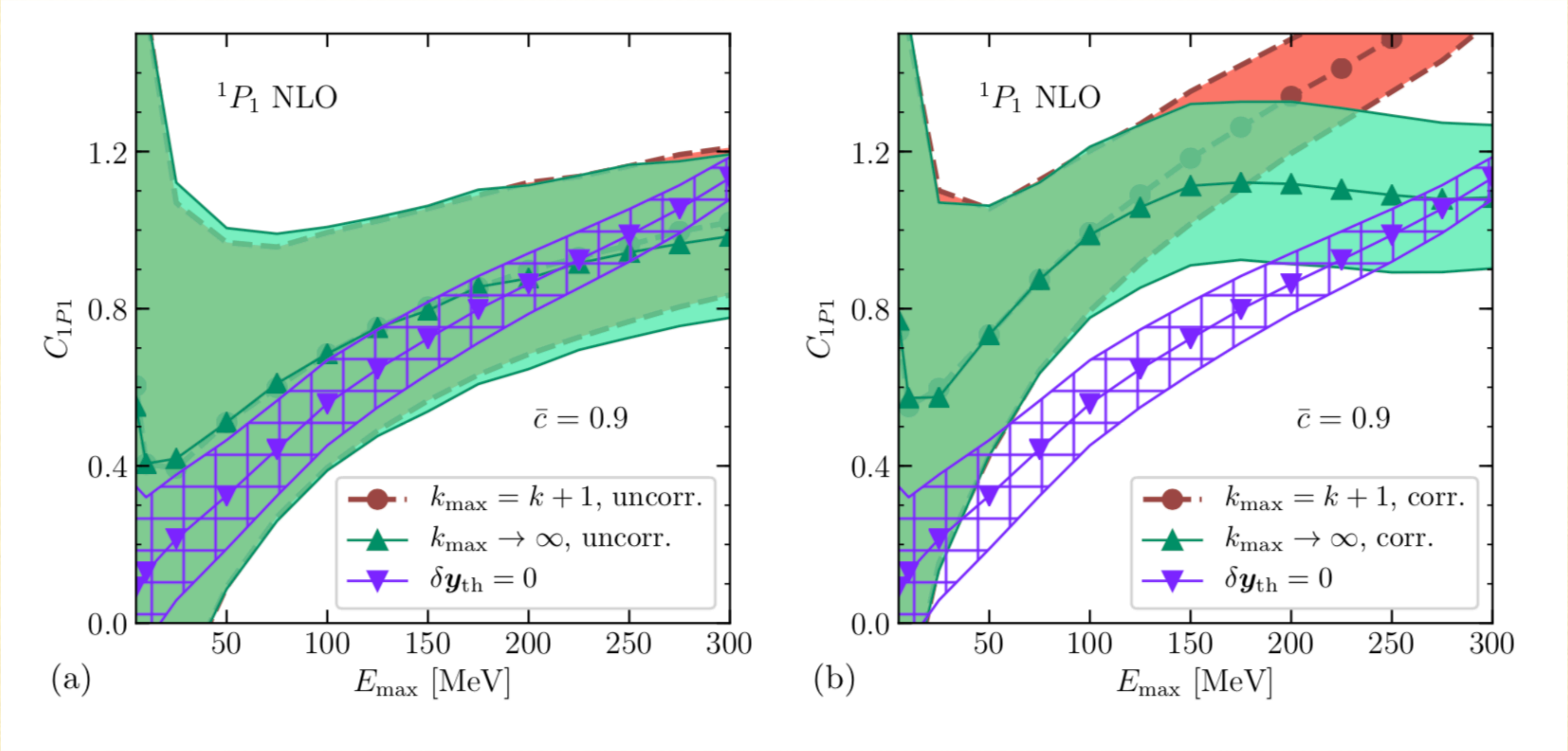


E_{max} plots for NN

Wesolowski, Furnstahl, Melendez, DP, JPG (2019)

Epelbaum, Krebs, Meißner, EPJA 51, 53 (2015)

- EKM semi-local potential with R=0.9 fm



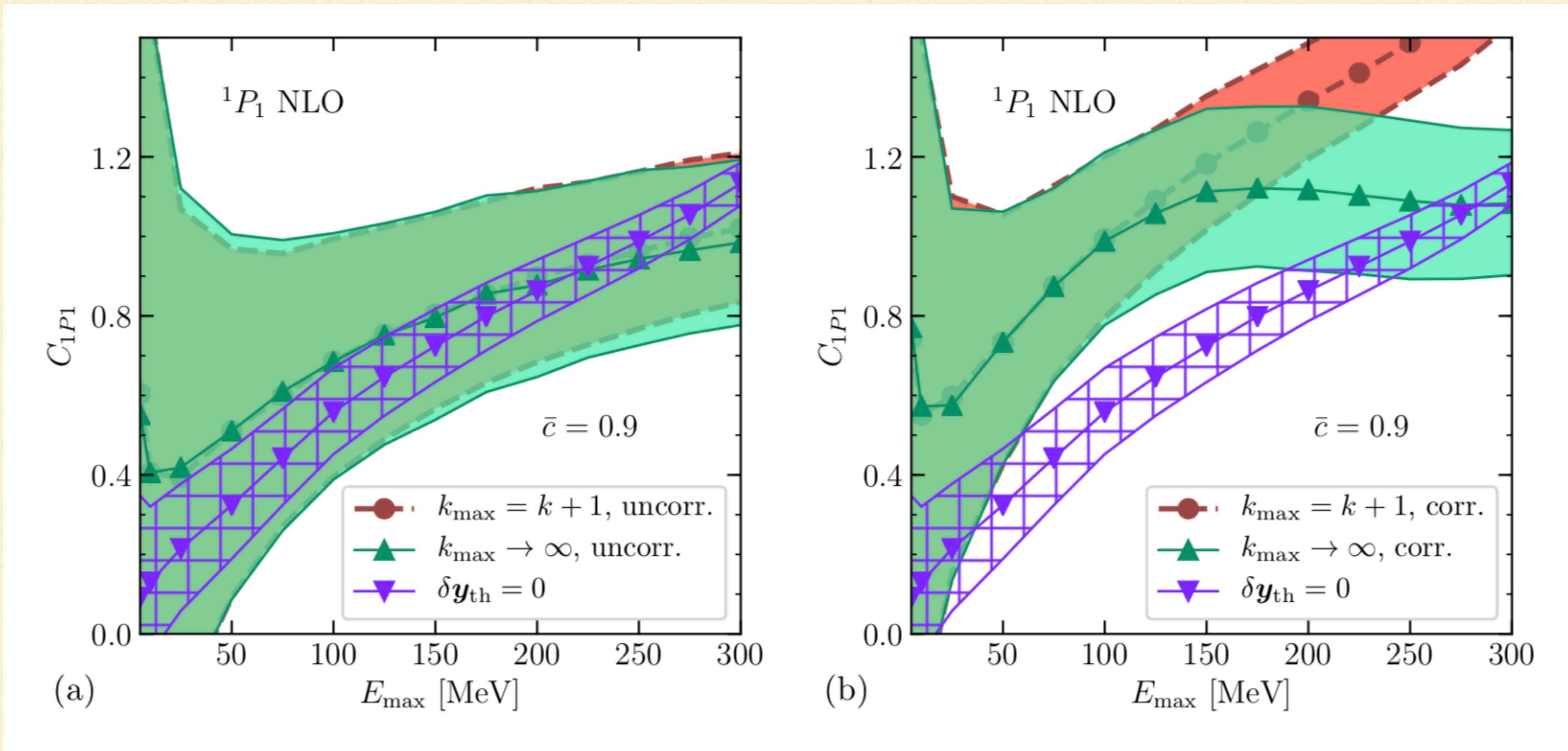
$$(\boldsymbol{\Sigma}_{\rm th,uncorr})_{ij} = (\mathbf{y}_{\rm ref})^2 \bar{c}^2 \delta_{ij} \sum_{n=k+1}^{k_{\rm max}} x_i^{2n}$$

E_{max} plots for NN

Wesolowski, Furnstahl, Melendez, DP, JPG (2019)

Epelbaum, Krebs, Meißner, EPJA 51, 53 (2015)

- EKM semi-local potential with R=0.9 fm



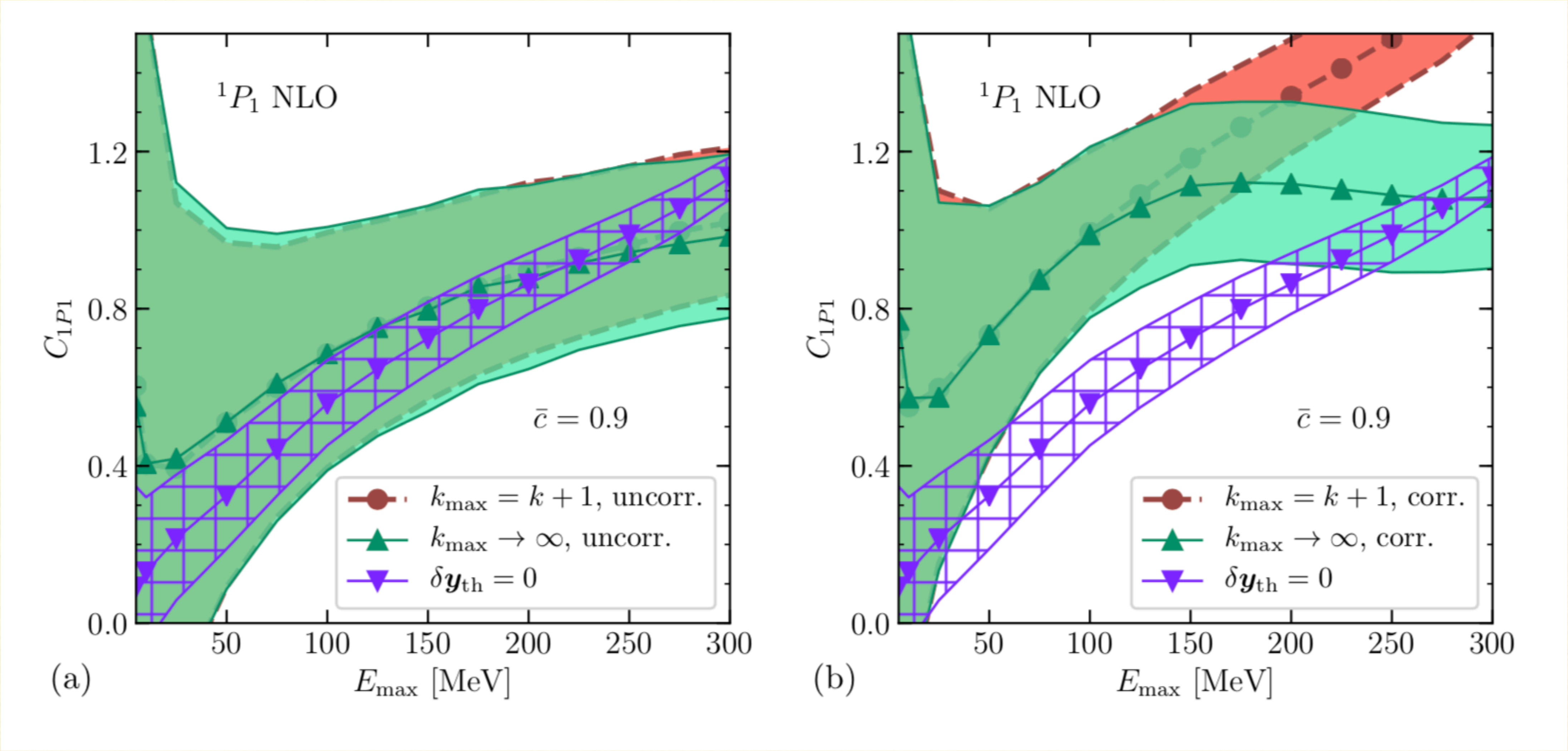
$$(\mathbf{\Sigma}_{\mathrm{th,uncorr}})_{ij} = (\mathbf{y}_{\mathrm{ref}})^2 \bar{c}^2 \delta_{ij} \sum_{n=k+1}^{k_{\mathrm{max}}} x_i^{2n} \qquad (\mathbf{\Sigma}_{\mathrm{th,corr}})_{ij} = (\mathbf{y}_{\mathrm{ref}})_i (\mathbf{y}_{\mathrm{ref}})_j \bar{c}^2 \sum_{n=k+1}^{k_{\mathrm{max}}} x_i^n x_j^n$$

E_{max} plots for NN

Wesolowski, Furnstahl, Melendez, DP, JPG (2019)

Epelbaum, Krebs, Meißner, EPJA 51, 53 (2015)

- EKM semi-local potential with R=0.9 fm



$$(\boldsymbol{\Sigma}_{\rm th,uncorr})_{ij} = (\mathbf{y}_{\rm ref})^2 \bar{c}^2 \delta_{ij} \sum_{n=k+1}^{k_{\rm max}} x_i^{2n} \qquad (\boldsymbol{\Sigma}_{\rm th,corr})_{ij} = (\mathbf{y}_{\rm ref})_i (\mathbf{y}_{\rm ref})_j \bar{c}^2 \sum_{n=k+1}^{k_{\rm max}} x_i^n x_j^n$$

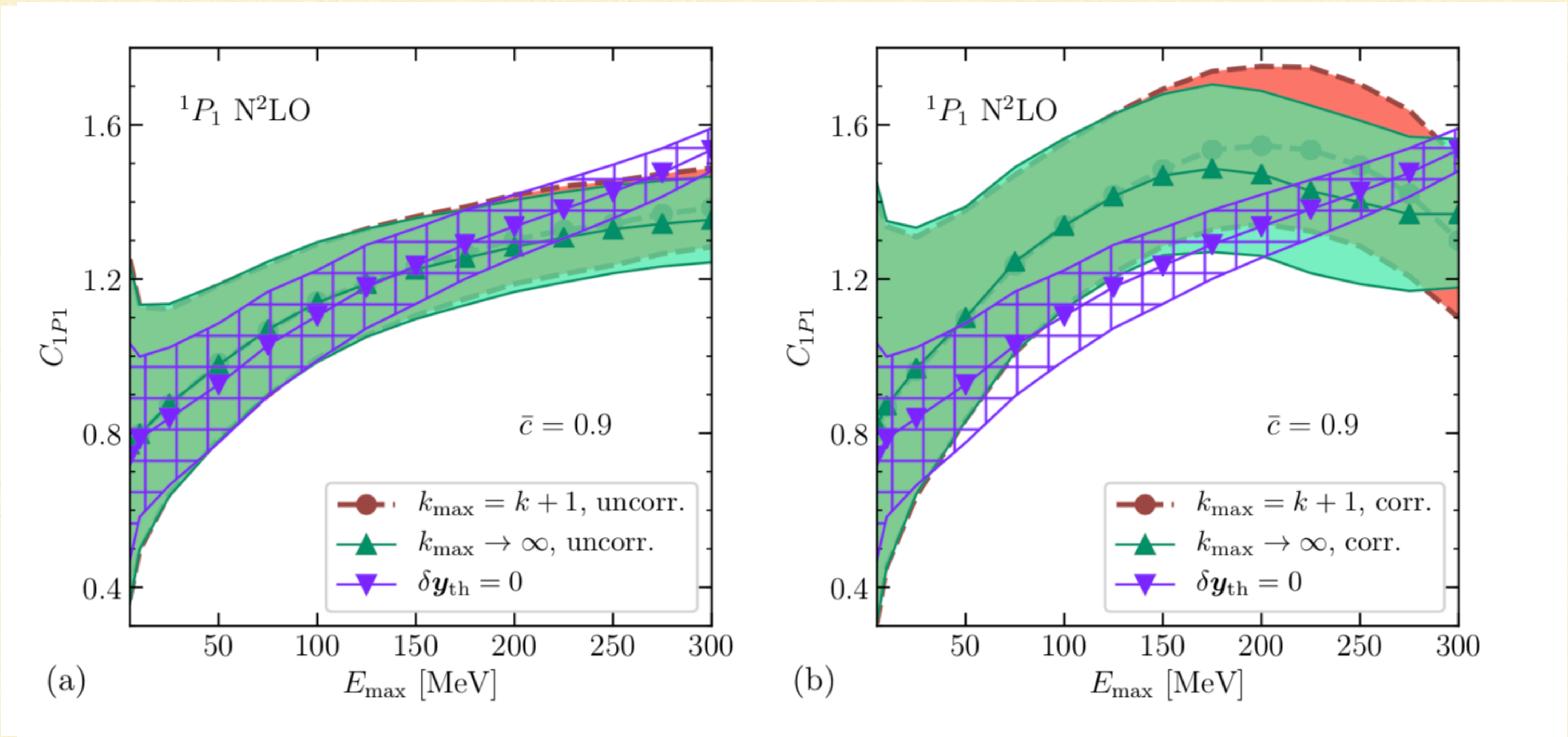
- Can resum truncation error to all orders (under assumptions about its correlation across orders): tests validity of FOTA

E_{max} plots for NN

Wesolowski, Furnstahl, Melendez, DP, JPG (2019)

Epelbaum, Krebs, Meißner, EPJA 51, 53 (2015)

- EKM semi-local potential with R=0.9 fm

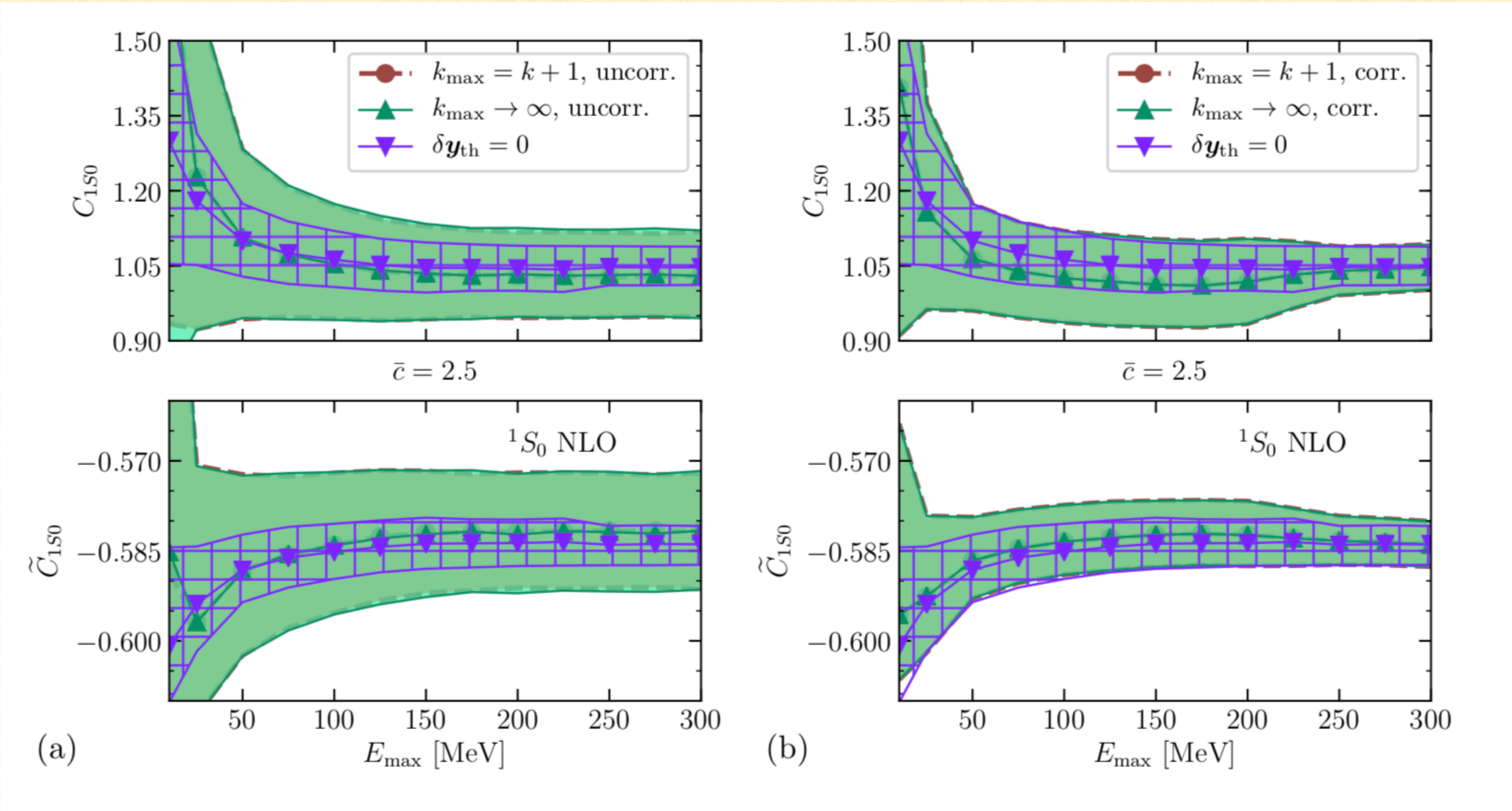


$$(\boldsymbol{\Sigma}_{\rm th,uncorr})_{ij} = (\mathbf{y}_{\rm ref})^2 \bar{c}^2 \delta_{ij} \sum_{n=k+1}^{k_{\max}} x_i^{2n} \qquad (\boldsymbol{\Sigma}_{\rm th,corr})_{ij} = (\mathbf{y}_{\rm ref})_i (\mathbf{y}_{\rm ref})_j \bar{c}^2 \sum_{n=k+1}^{k_{\max}} x_i^n x_j^n$$

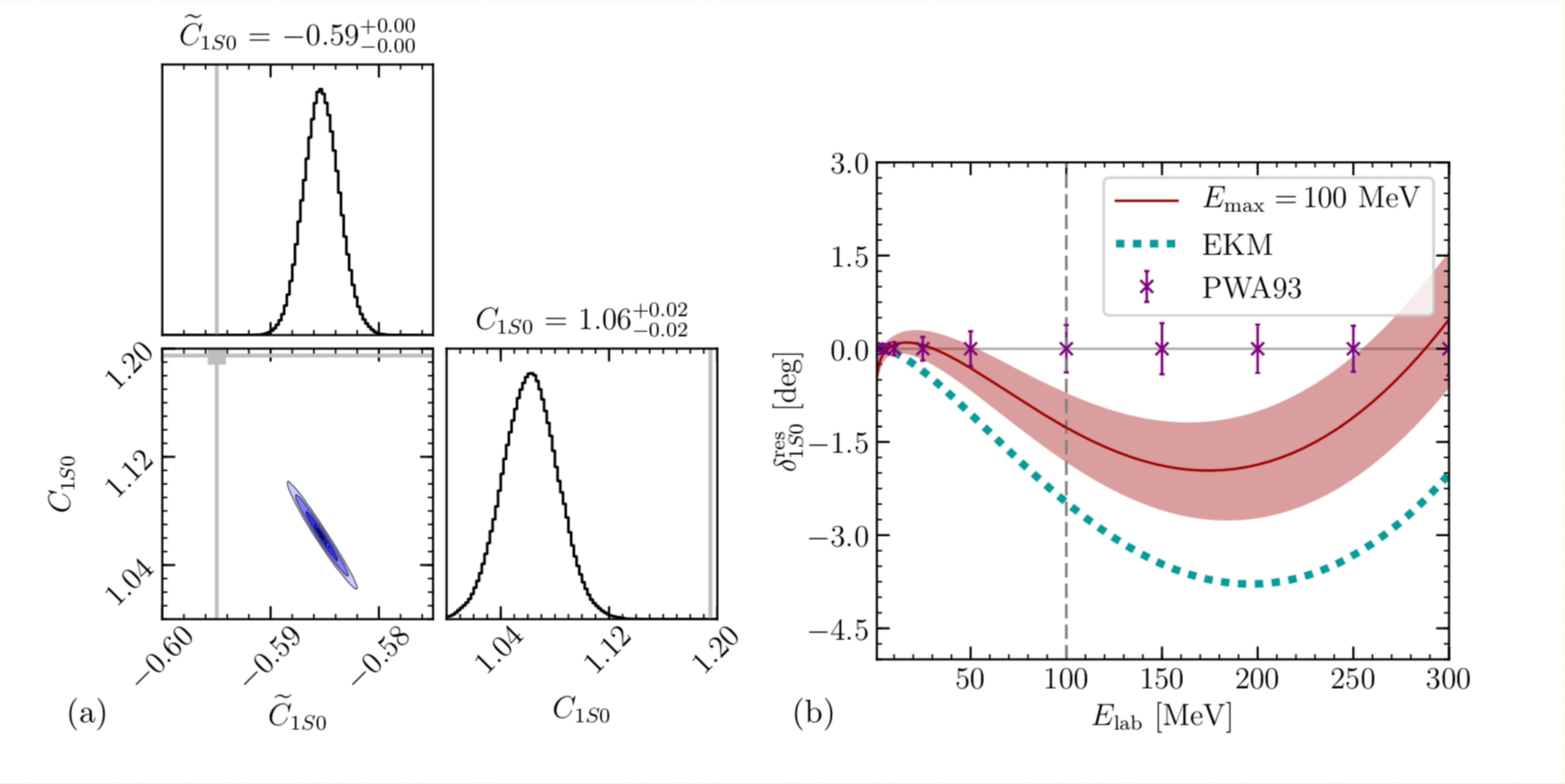
- Can resum truncation error to all orders (under assumptions about its correlation across orders): tests validity of FOTA

E_{max} plot in the $^1S_0$ at NLO

NLO potential in $^1S_0$ channel $V^{(2)}_{1S0}(p',p) = \frac{4\pi}{f_\pi^2}\left(\tilde{C}^{np}_{1S0} + C_{1S0}\frac{p^2+p'^2}{\Lambda_b^2}\right)$

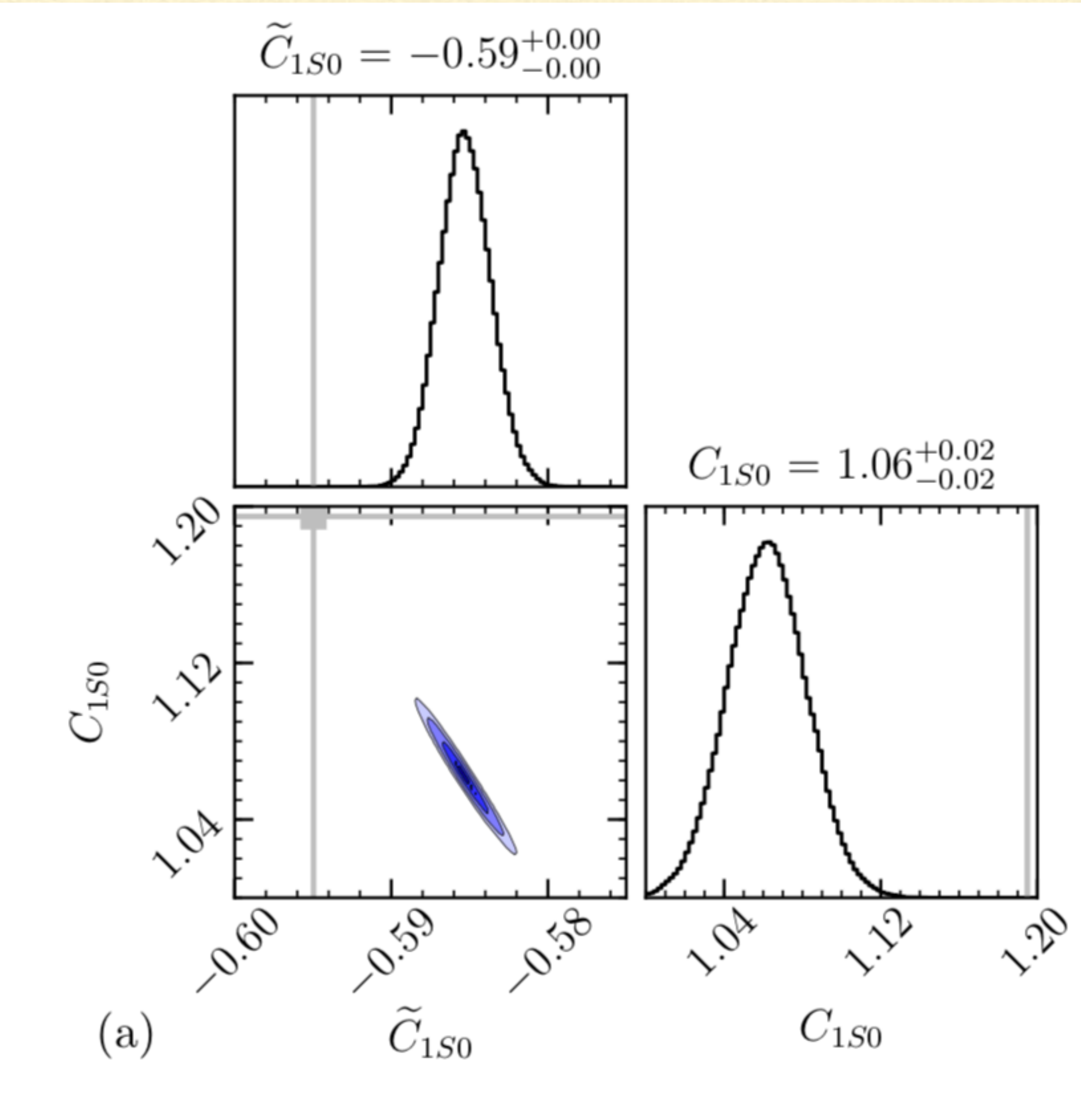


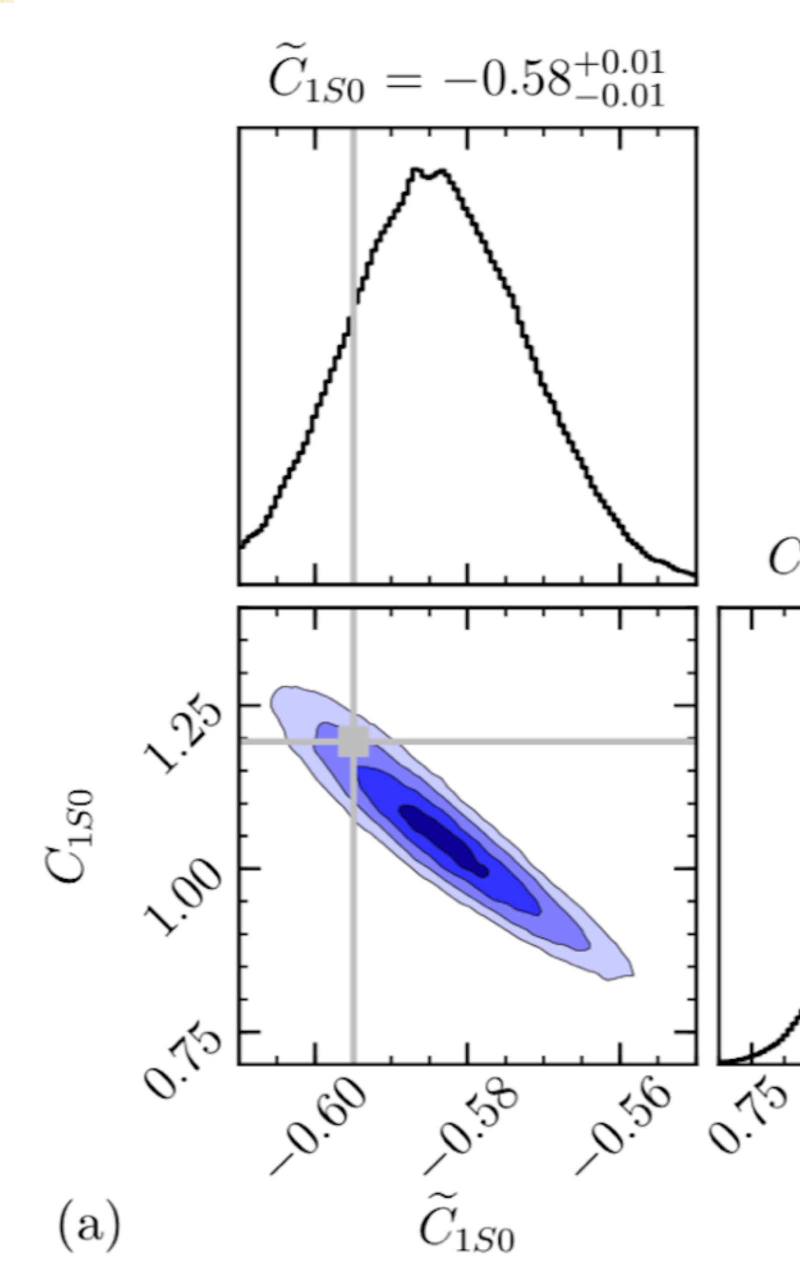
Changes in Parameter Estimates



Including truncation errors changes central values and (esp.) errors

Changes in Parameter Estimates



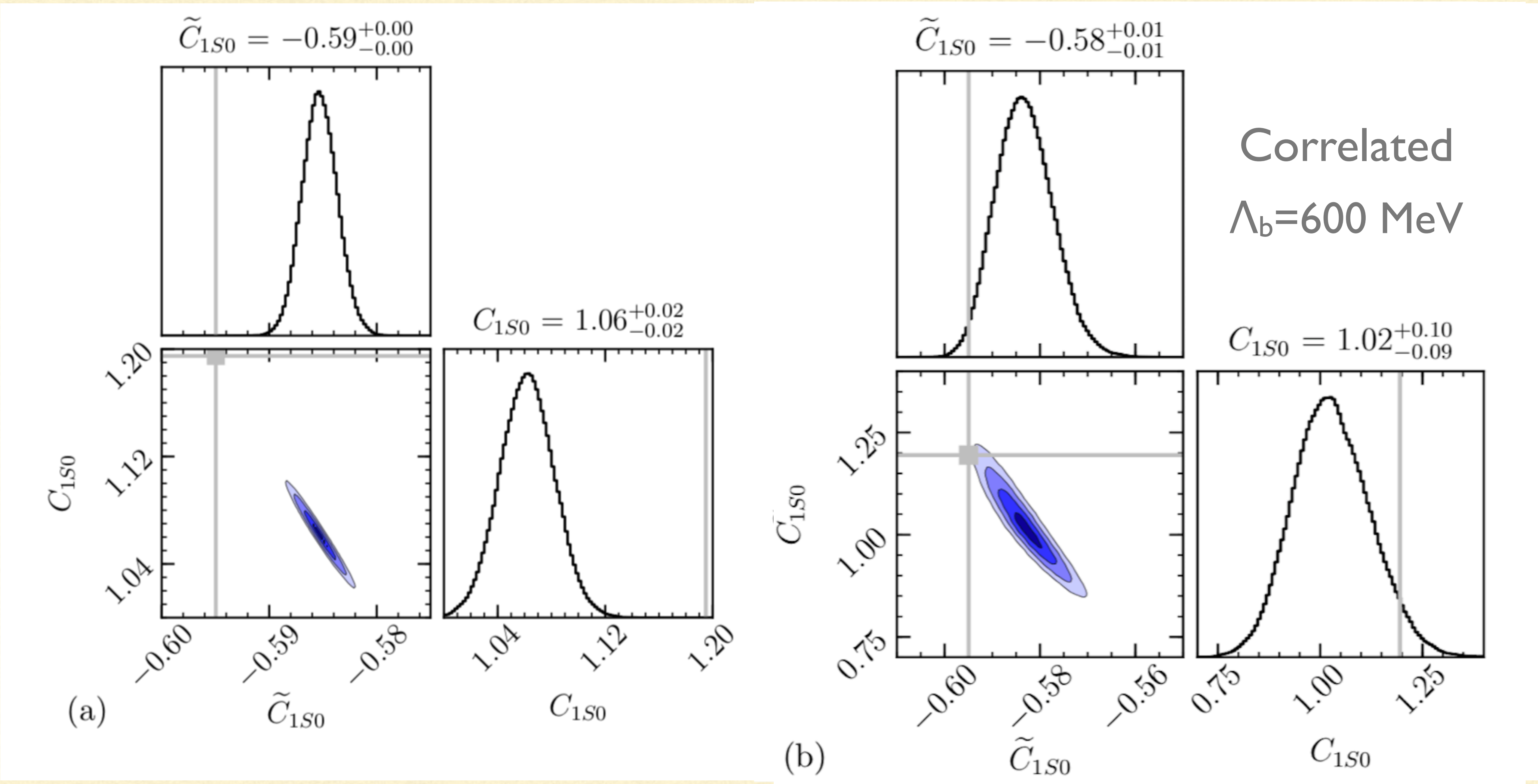


Uncorrelated

Λ_b=600 MeV

Including truncation errors changes central values and (esp.) errors

Changes in Parameter Estimates



Including truncation errors changes central values and (esp.) errors

NN scattering

Epelbaum, Krebs, Meißner, PRC, 2015

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

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

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

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

NN scattering

Epelbaum, Krebs, Meißner, PRC, 2015

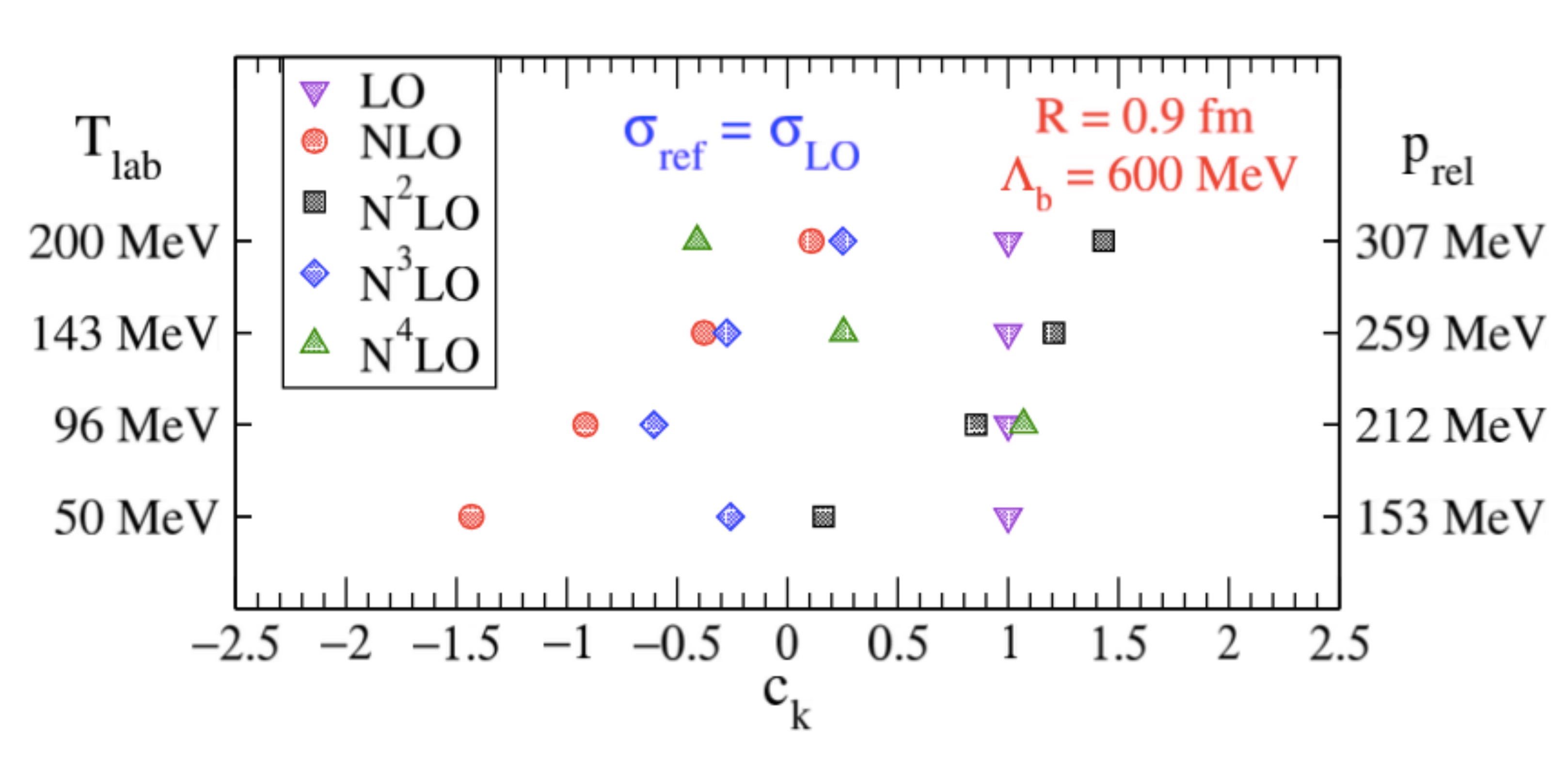
Employ "semi-local" potentials of Epelbaum, Krebs, and Meißner

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

- 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, N²LO, N³LO, N⁴LO (k=0, 2, 3, 4, 5)

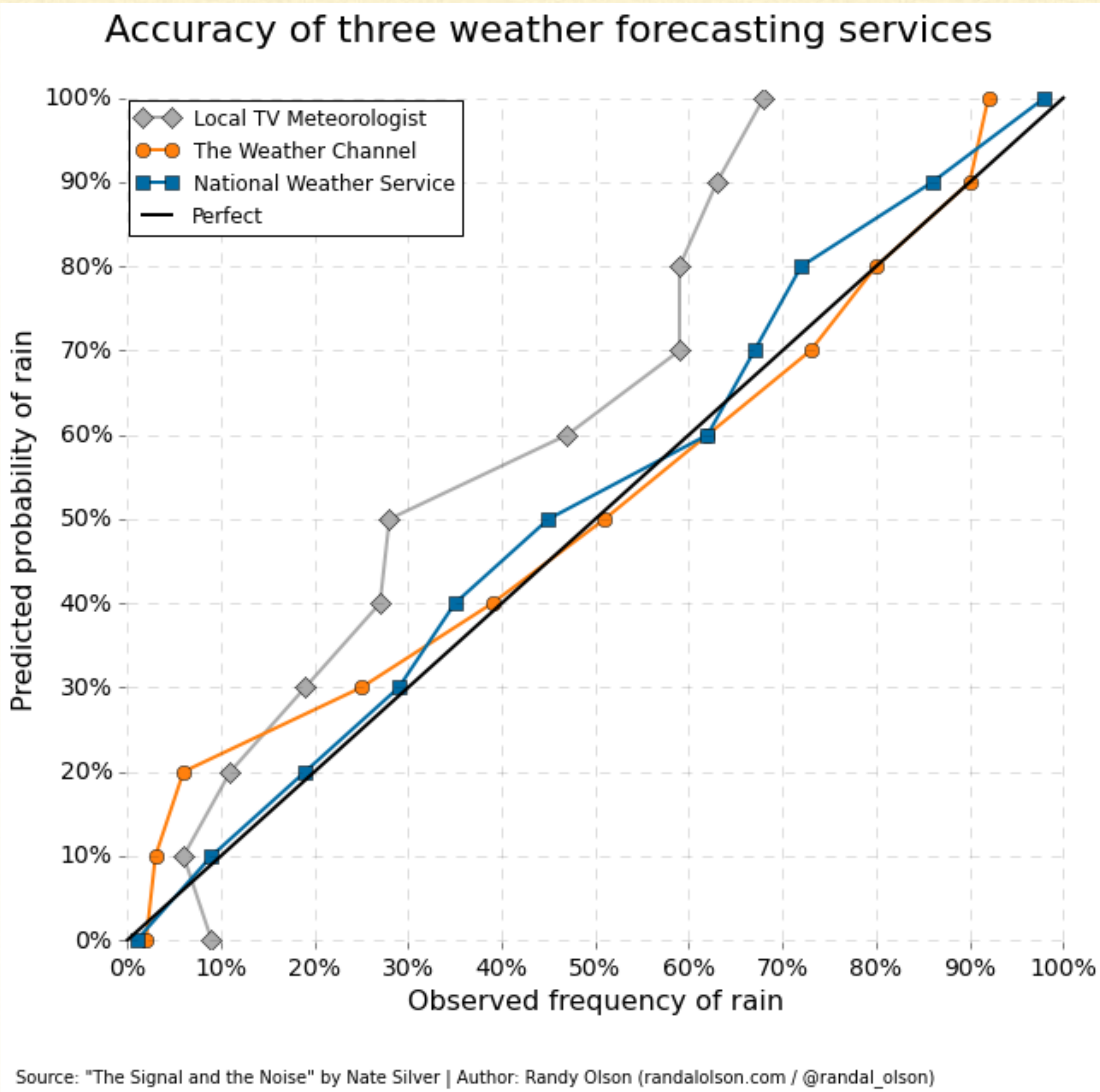
$$\sigma_{np}(E_{\rm lab}) = \sigma_{\rm LO} \sum_{n=0}^{k} c_n(p_{\rm rel}) \left(\frac{p_{\rm rel}}{\Lambda_b}\right)^n$$

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



EFT truncation error model checking

EFT truncation error model checking

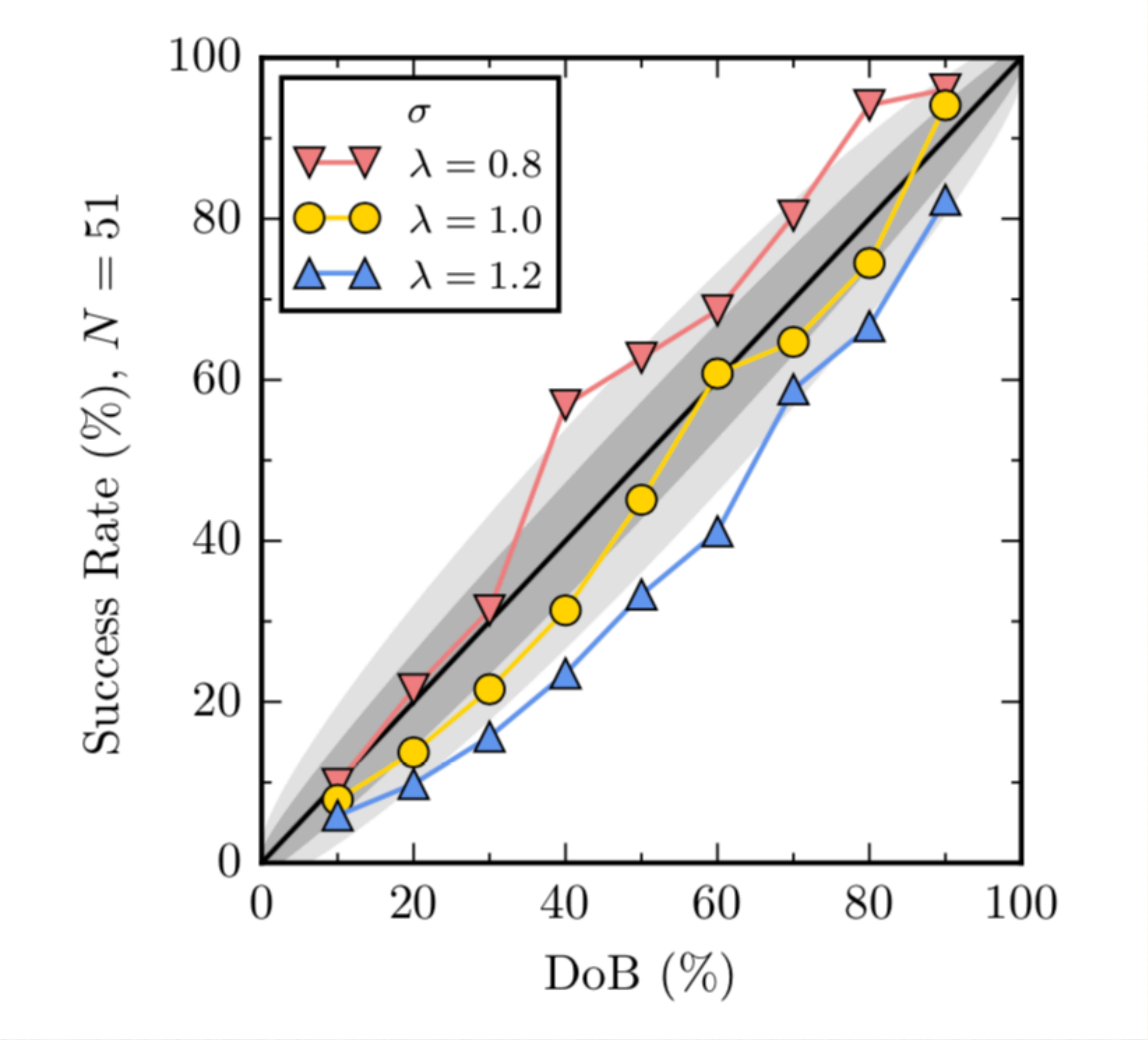


Melendez, Furnstahl, Wesolowski, PRC, 2017
after Furnstahl, Klco, DP, Wesolowski, PRC, 2015
after Bagnaschi, Cacciari, Guffanti, Jenniches, 2015

- 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

EFT truncation error model checking

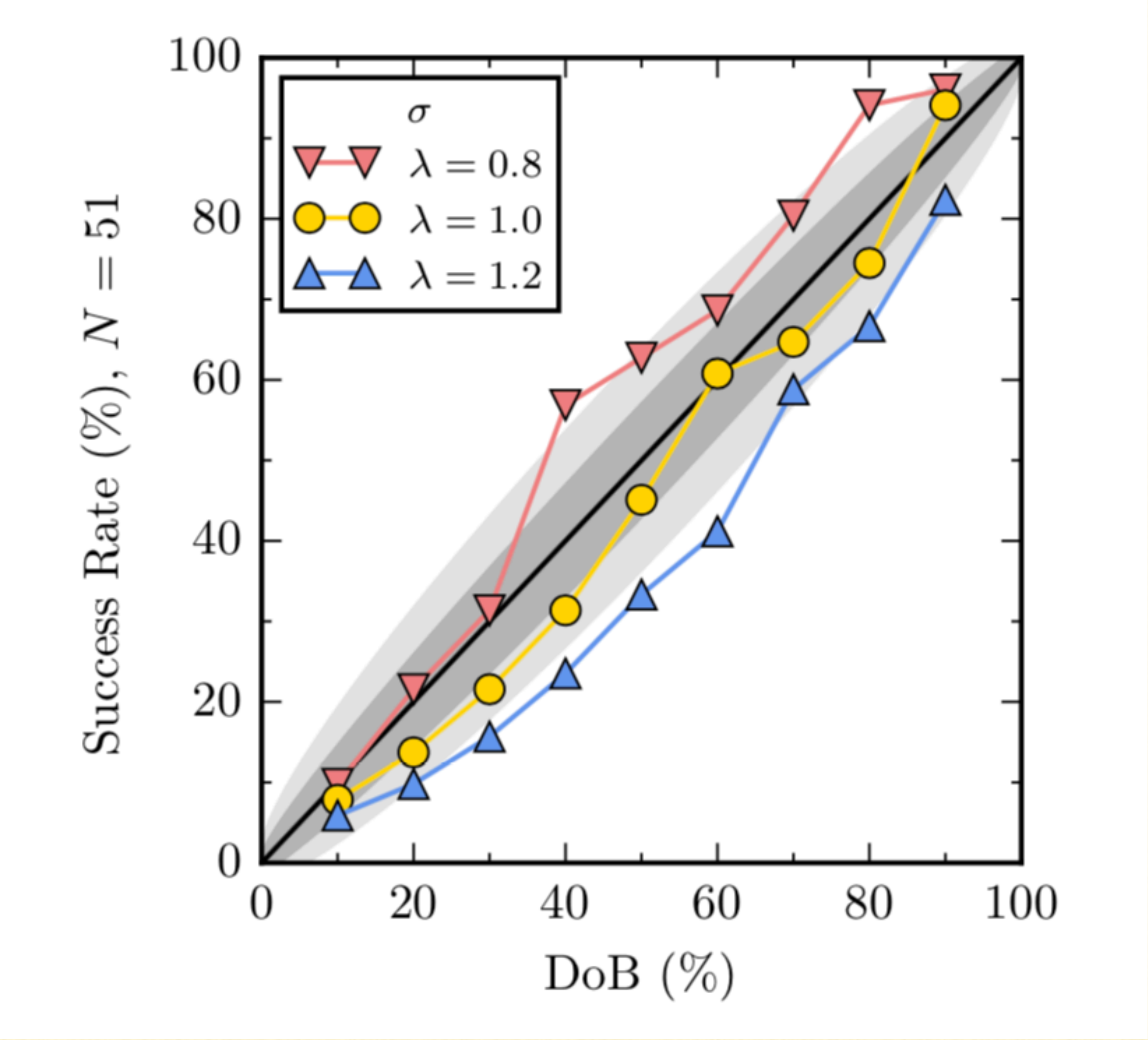
Melendez, Furnstahl, Wesolowski, PRC, 2017
after Furnstahl, Klco, DP, Wesolowski, PRC, 2015
after Bagnaschi, Cacciari, Guffanti, Jenniches, 2015



- 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

EFT truncation error model checking

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

Breakdown-scale Inference

- Λ_b determines the size of the c_n's. Choose it too big, and they'll be too big. Choose it too small, they'll be too small. And progressively so as one moves to higher and higher order.

- We have a theory for $\mathrm{pr}(c_n|c_0, c_1, \ldots, c_k)$: now use Bayes' theorem to see how (im)probable are the c_n's that dimensionful EFT coefficients (b_n's) produce for a given Λ_b.

Breakdown-scale Inference

- Λ_b determines the size of the c_n's. Choose it too big, and they'll be too big. Choose it too small, they'll be too small. And progressively so as one moves to higher and higher order.

- We have a theory for pr(c_n|c_0, c_1, …, c_k): now use Bayes' theorem to see how (im)probable are the c_n's that dimensionful EFT coefficients (b_n's) produce for a given Λ_b.

At one energy:

$$\mathrm{pr}(\Lambda_b|b_2,\ldots,b_k) \propto \frac{1}{\Lambda_b}\left(\frac{\Lambda_b^{k+2}}{(k+1)\langle b^2\rangle}\right)^{\frac{k-1}{2}}$$

(NLO: k=2, NNLO: k=3, N^3LO: k=4, etc.)

Breakdown-scale Inference

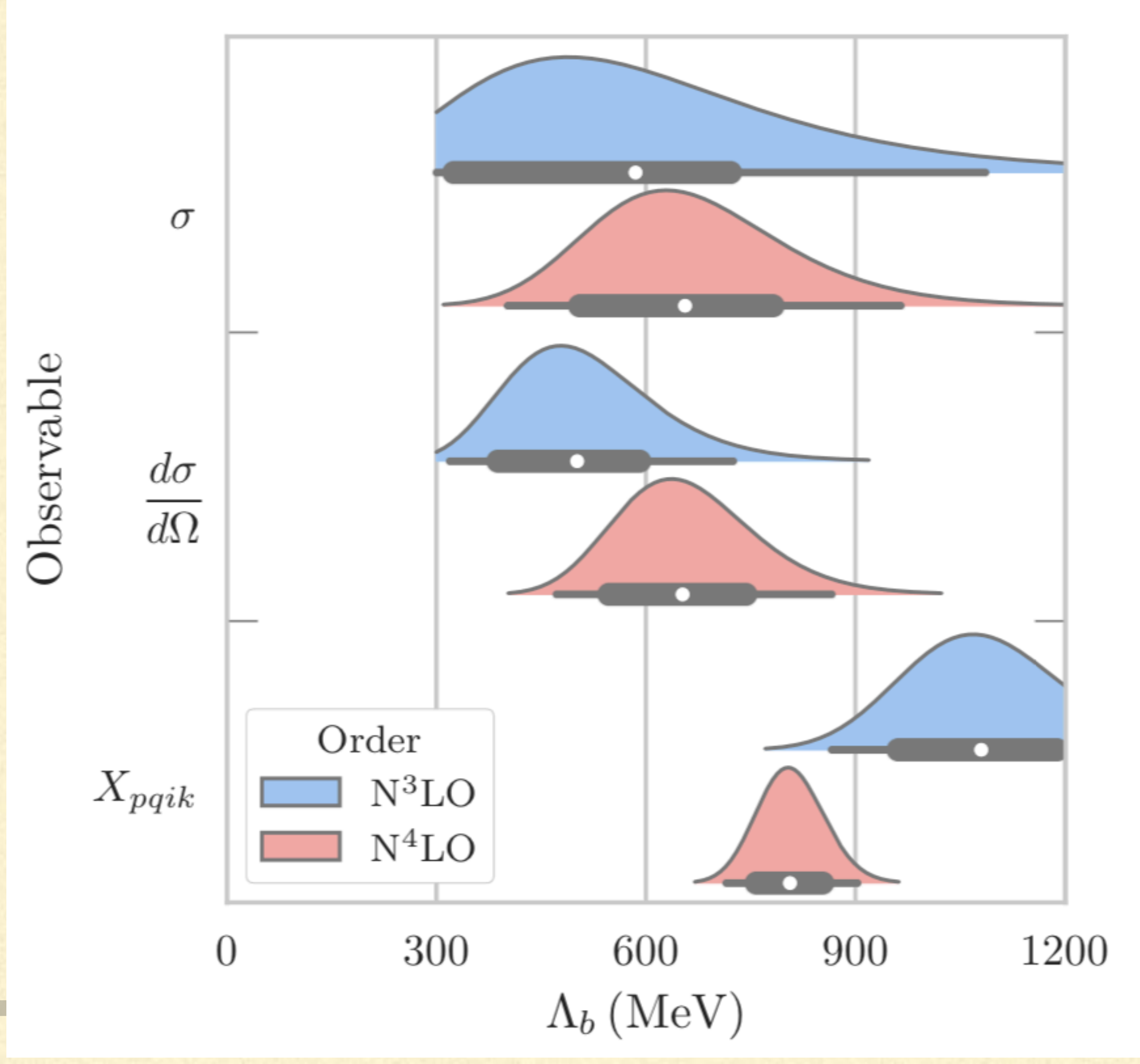
- Λ_b determines the size of the c_n's. Choose it too big, and they'll be too big. Choose it too small, they'll be too small. And progressively so as one moves to higher and higher order.
- We have a theory for pr(c_n|c_0, c_1, …, c_k): now use Bayes' theorem to see how (im)probable are the c_n's that dimensionful EFT coefficients (b_n's) produce for a given Λ_b.

At one energy:

$$\mathrm{pr}(\Lambda_b|b_2,\ldots,b_k) \propto \frac{1}{\Lambda_b}\left(\frac{\Lambda_b^{k+2}}{(k+1)\langle b^2\rangle}\right)^{\frac{k-1}{2}}$$

(NLO: k=2, NNLO: k=3, N^3LO: k=4, etc.)

Using 5 energies (and 2 angles):



Breakdown-scale Inference

- Λ_b determines the size of the c_n's. Choose it too big, and they'll be too big. Choose it too small, they'll be too small. And progressively so as one moves to higher and higher order.

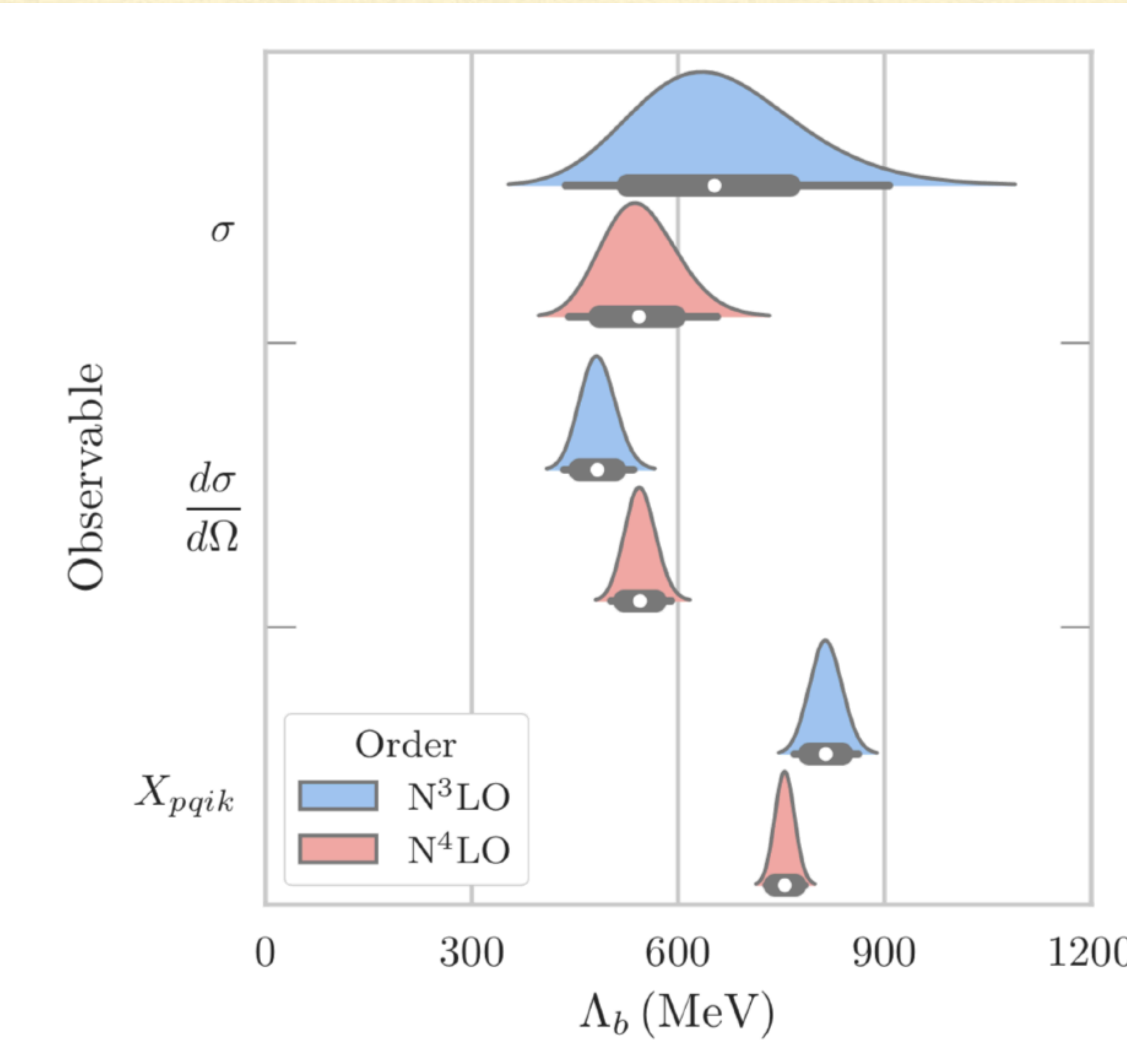
- We have a theory for pr($c_n|c_0, c_1, \ldots, c_k$): now use Bayes' theorem to see how (im)probable are the c_n's that dimensionful EFT coefficients (b_n's) produce for a given Λ_b.

At one energy:

$$\mathrm{pr}(\Lambda_b|b_2,\ldots,b_k) \propto \frac{1}{\Lambda_b}\left(\frac{\Lambda_b^{k+2}}{(k+1)\langle b^2\rangle}\right)^{\frac{k-1}{2}}$$

(NLO: k=2, NNLO: k=3, N^3LO: k=4, etc.)

Using 17 energies (and 7 angles):



Breakdown-scale Inference

- Λ_b determines the size of the c_n's. Choose it too big, and they'll be too big. Choose it too small, they'll be too small. And progressively so as one moves to higher and higher order.

- We have a theory for pr(c_n|c_0, c_1, ..., c_k): now use Bayes' theorem to see how (im)probable are the c_n's that dimensionful EFT coefficients (b_n's) produce for a given Λ_b.

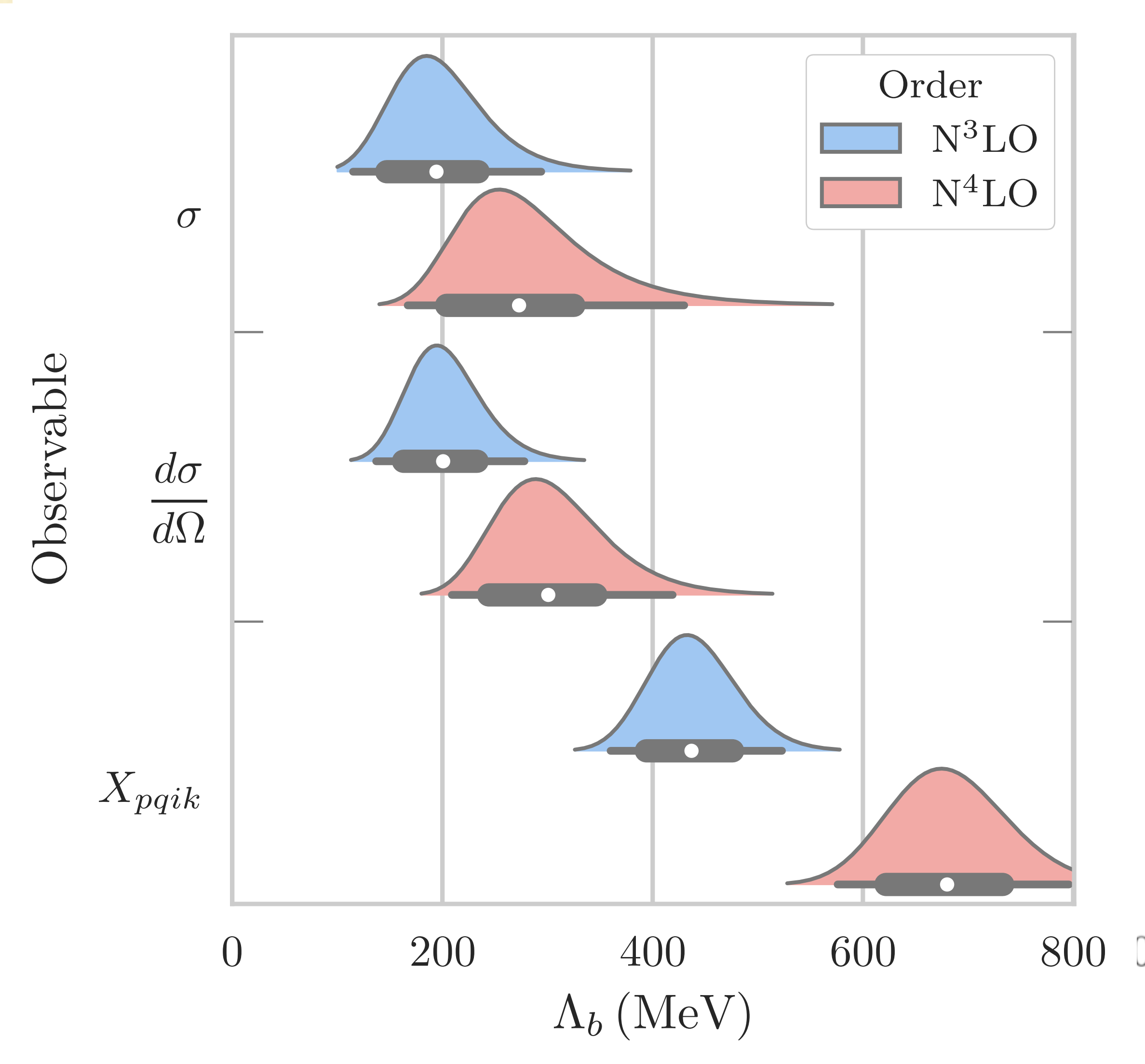
At one energy:

$$\mathrm{pr}(\Lambda_b|b_2,\ldots,b_k) \propto \frac{1}{\Lambda_b}\left(\frac{\Lambda_b^{k+2}}{(k+1)\langle b^2\rangle}\right)^{\frac{k-1}{2}}$$

(NLO: k=2, NNLO: k=3, N³LO: k=4, etc.)

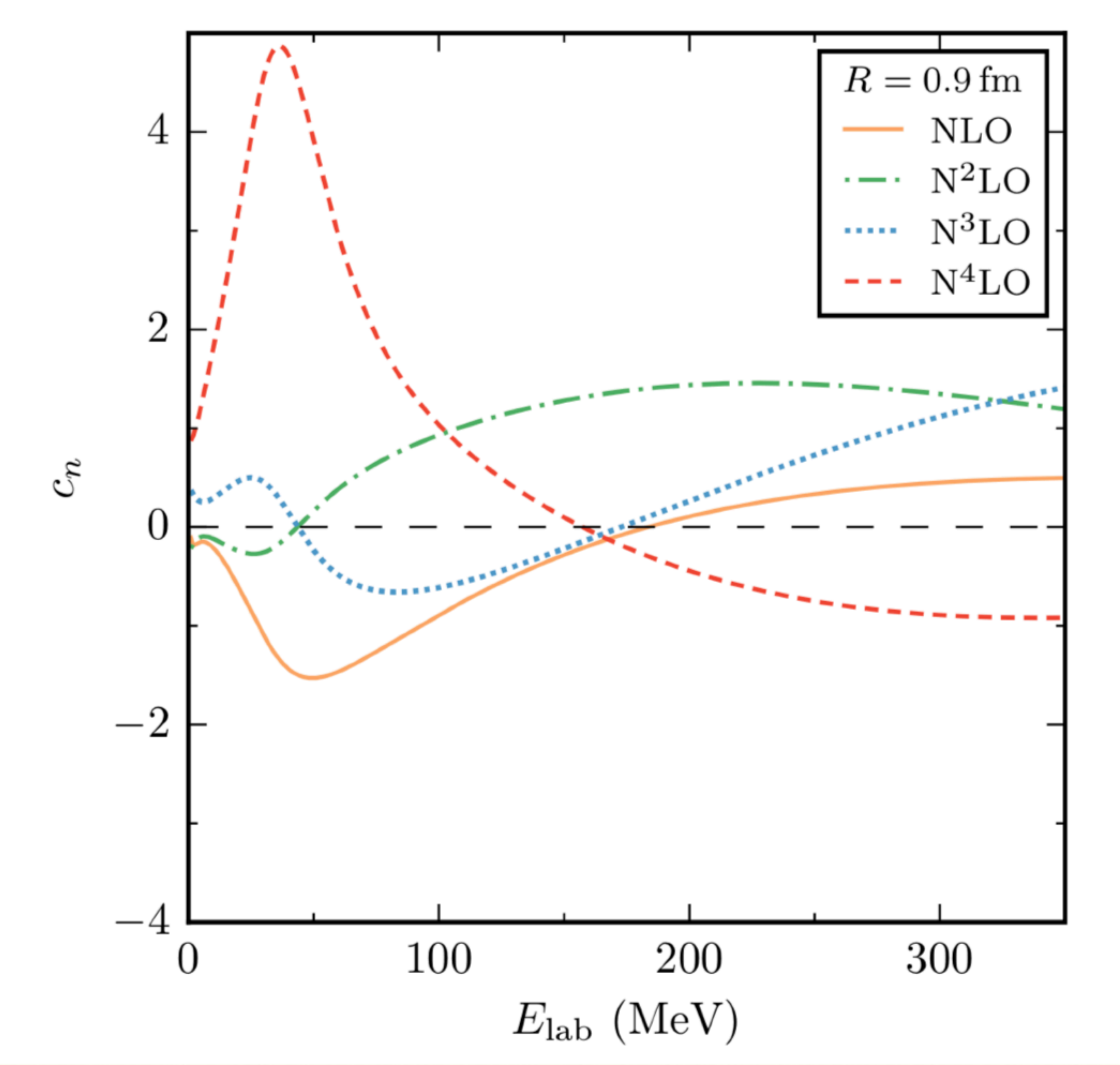
Using 17 energies (and 7 angles):

R=1.2 fm



Functional Data

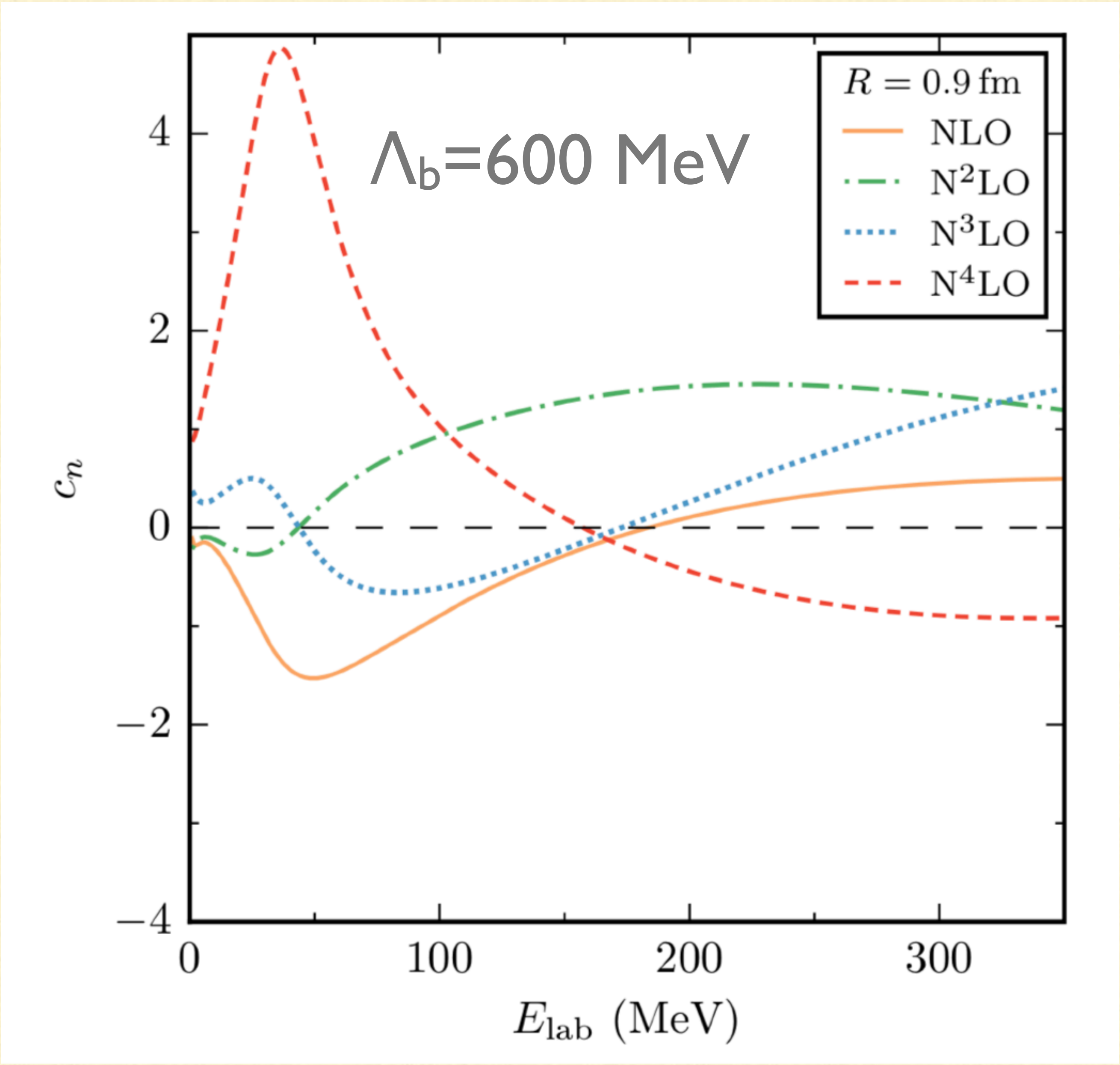
- Truncation-error analysis assumed we have 119 independent data points
- But we don't
- We have a function for each observable at each order
- Can we understand the properties of these functions, and so do Λ_b inference and compute success ratios rigorously?



$$\sigma(E) = \sigma_0(E)\left[1 + c_2(E)x^2 + c_3(E)x^3 + c_4(E)x^4 + c_5(E)x^5\right]$$

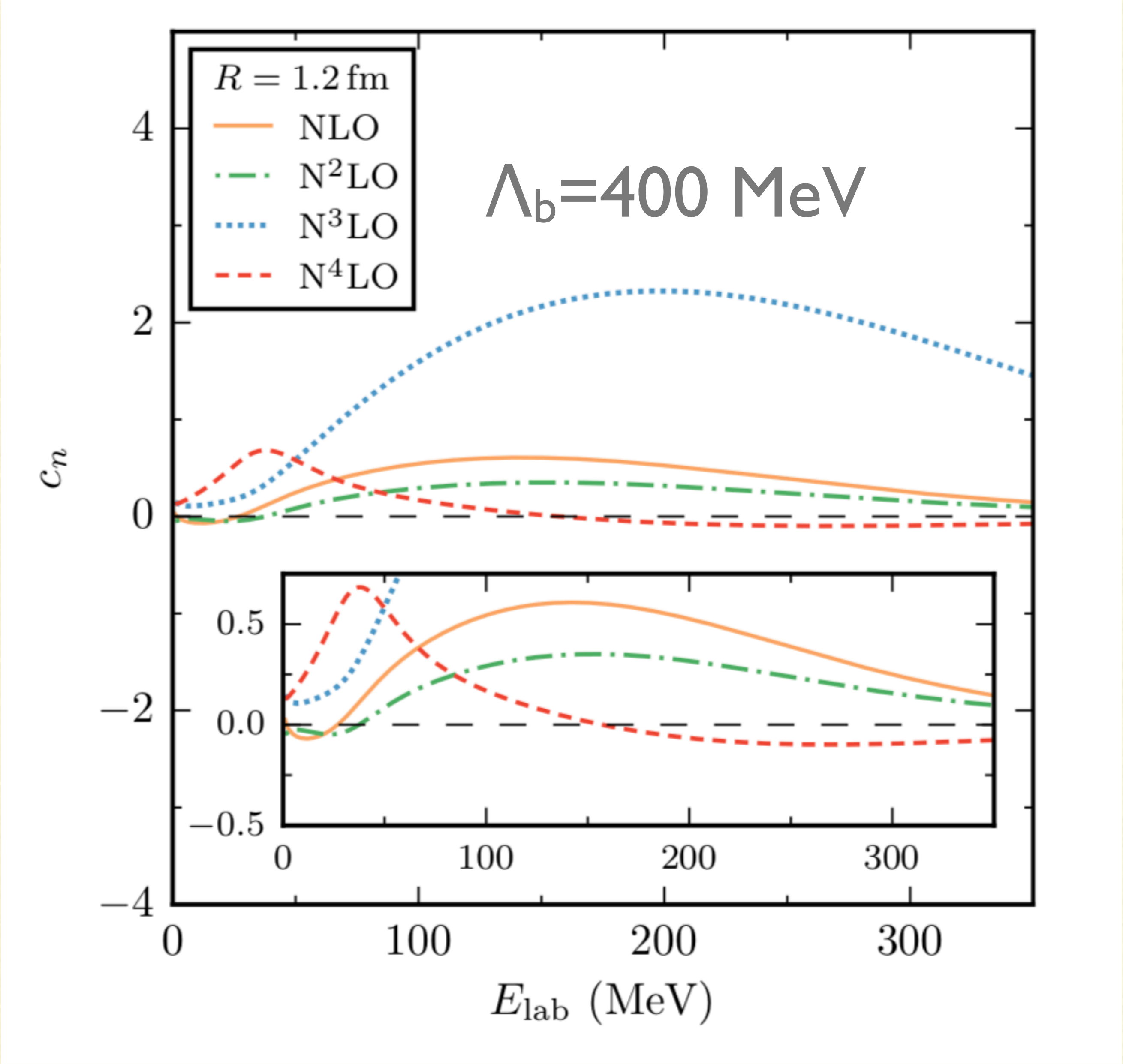
Observations and Questions

- c_n's do not grow or shrink with n: good Λ_b choice
- Bounded functions, mostly between -2 and 2
- Each "takes a turn" at being largest
- Not oscillating quickly in this energy range



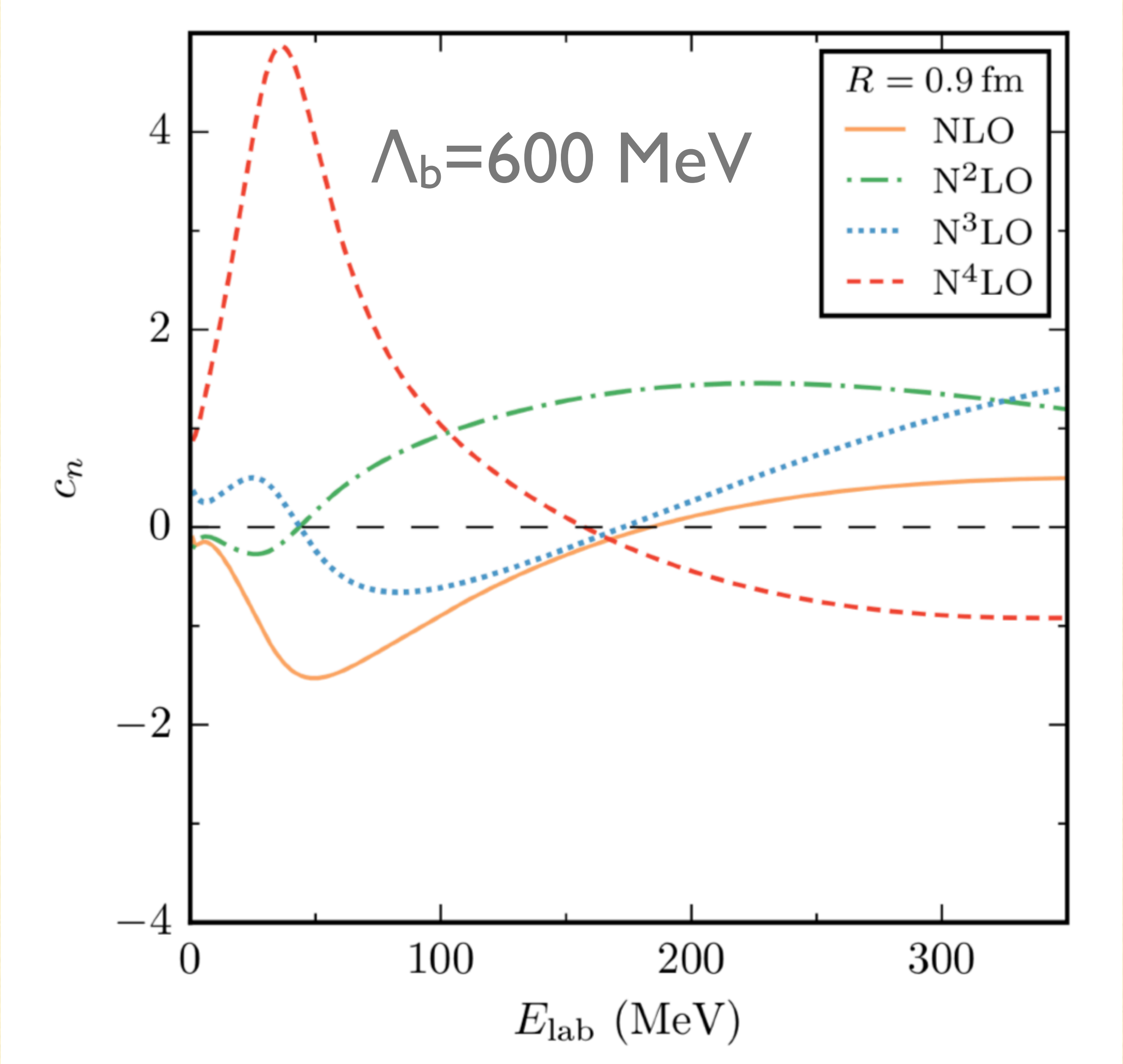
Observations and Questions

- c_n's do not grow or shrink with n: good Λ_b choice
- Bounded functions, mostly between -2 and 2
- Each "takes a turn" at being largest
- Not oscillating quickly in this energy range



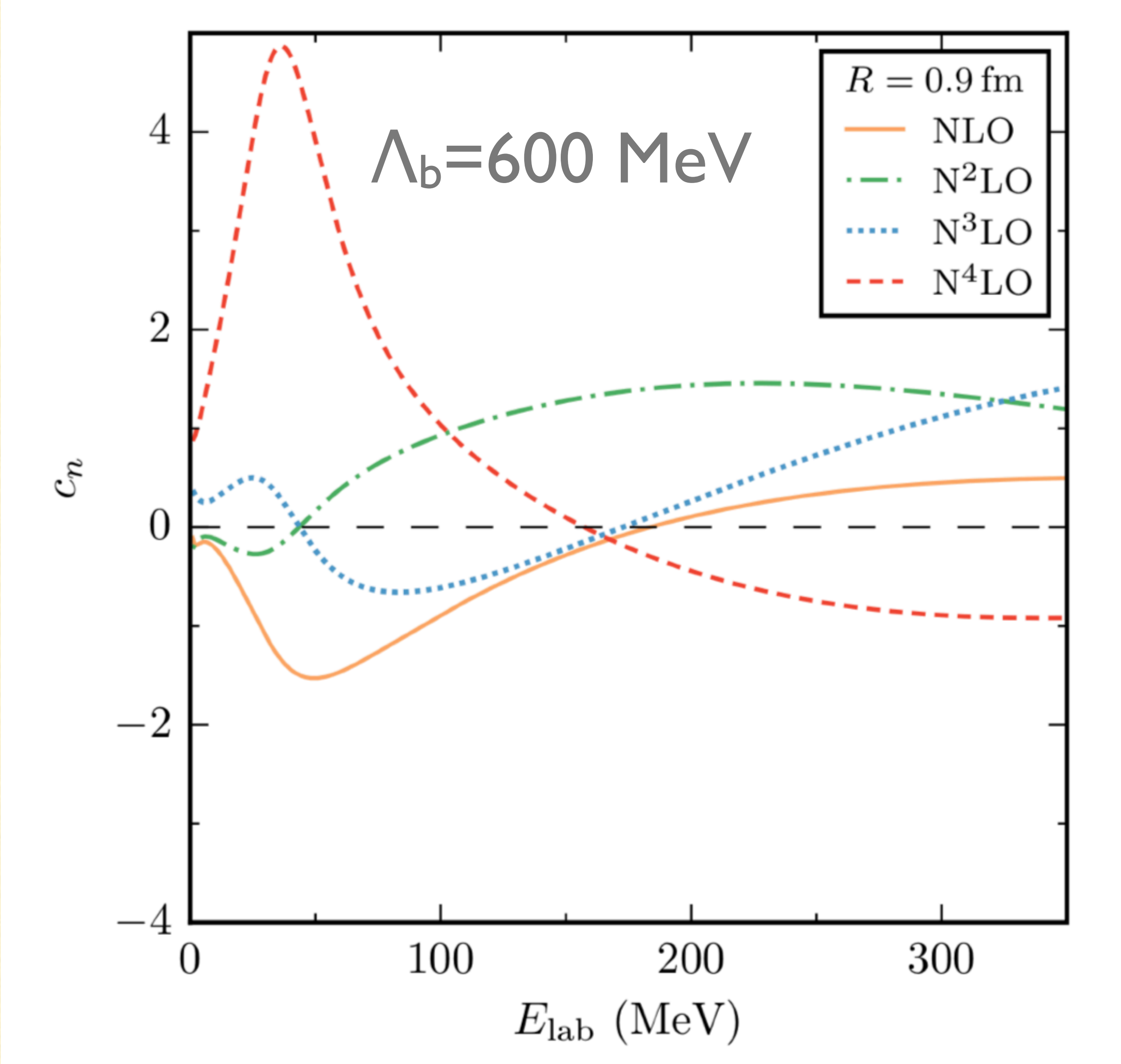
Observations and Questions

- c_n's do not grow or shrink with n: good Λ_b choice
- Bounded functions, mostly between -2 and 2
- Each "takes a turn" at being largest
- Not oscillating quickly in this energy range



Observations and Questions

- c_n's do not grow or shrink with n: good Λ_b choice
- Bounded functions, mostly between -2 and 2
- Each "takes a turn" at being largest
- Not oscillating quickly in this energy range

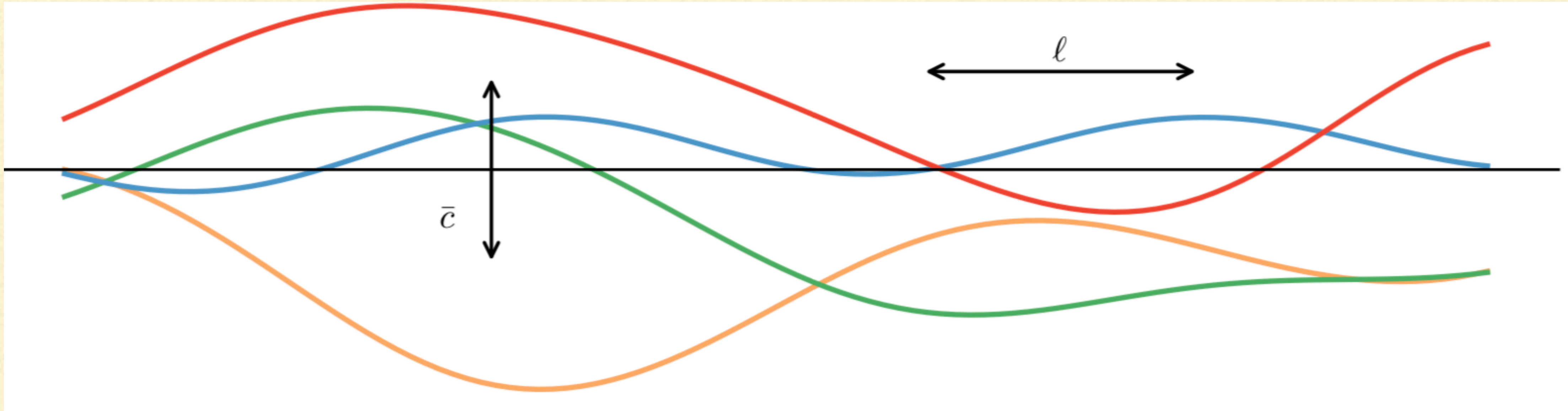


Physics questions:

- Do curves all fluctuate around zero with some common variance?
- What is the correlation length? Is it different at each order?

Hypothesis: the c_n's are a GP

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

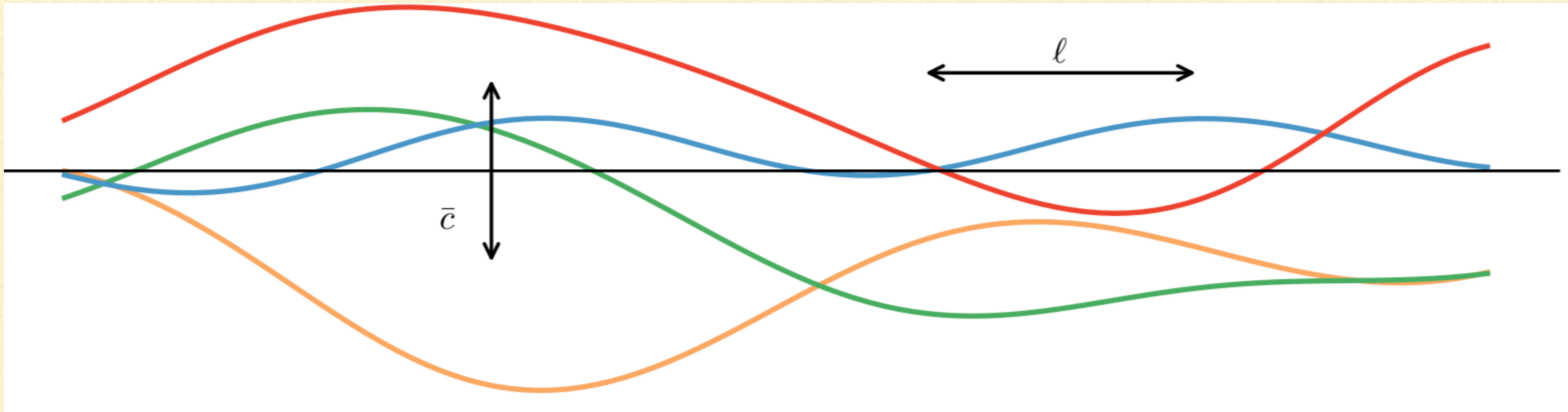


GP describes "functional data" as

- Gaussian distribution at each point.
- With specified correlation structure.

Hypothesis: the c_n's are a GP

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



GP describes "functional data" as

- Gaussian distribution at each point.
- With specified correlation structure.

Our hypothesis:

EFT coefficients at different orders can be modeled as:

- independent but identical realizations of one Gaussian Process;
- with a correlation structure; here we use a "squared exponential" (Gaussian) kernel, but we test it

Gaussian Process model

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

Gaussian Process model

- Specify correlation matrix of c_n at x and y, e.g.:

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

Gaussian Process model

- Specify correlation matrix of c_n at x and y, e.g.:

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

- Assumptions?

Gaussian Process model

- Specify correlation matrix of c_n at x and y, e.g.:

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

- Assumptions?
  - Mean of zero

Gaussian Process model

- Specify correlation matrix of c_n at x and y, e.g.:

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

- Assumptions?
  - Mean of zero
  - Stationary

Gaussian Process model

- Specify correlation matrix of c_n at x and y, e.g.:

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

- Assumptions?
  - Mean of zero
  - Stationary
  - Squared exponential correlation structure

Gaussian Process model

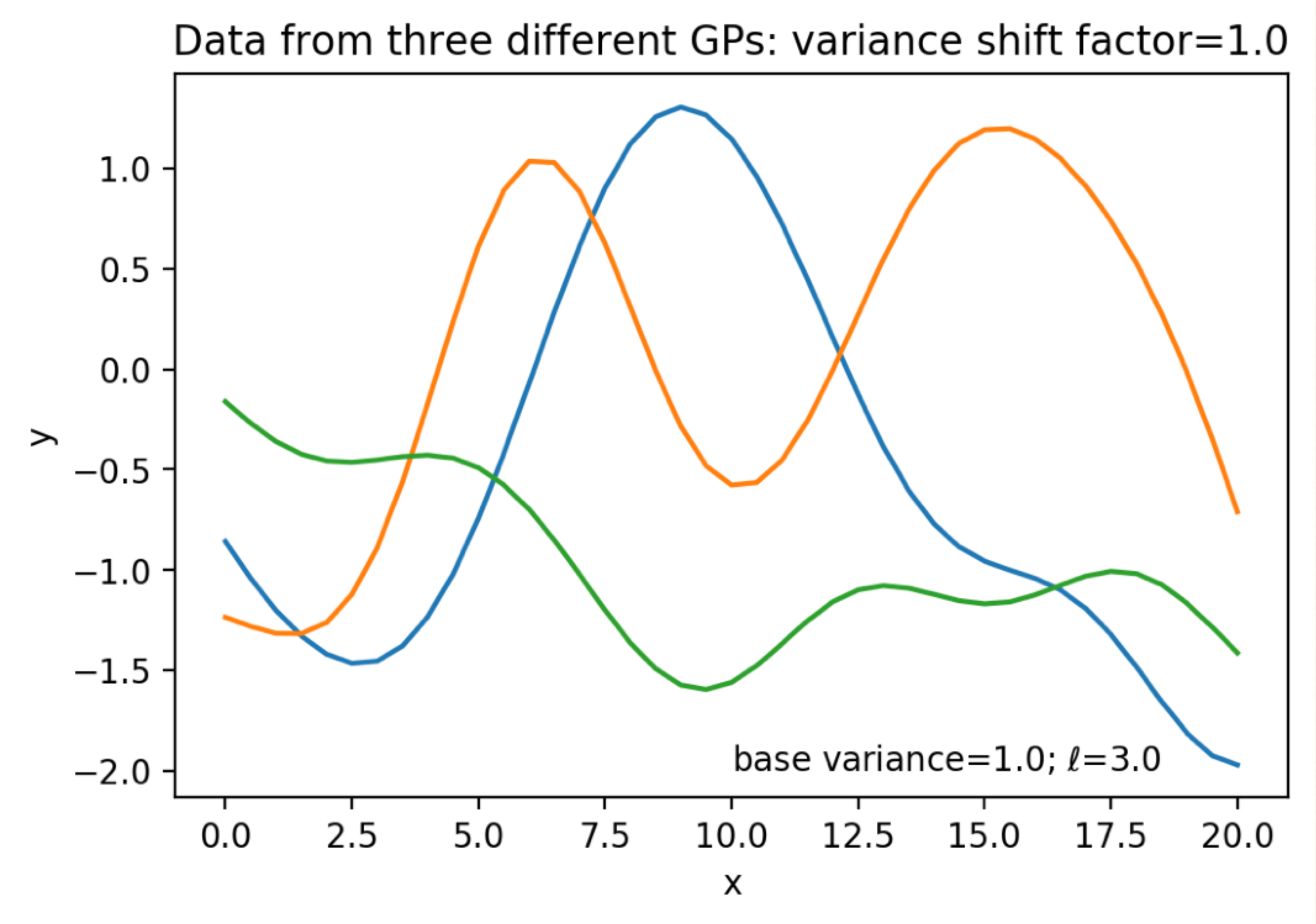
- Specify correlation matrix of c_n at x and y, e.g.:

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

- Assumptions?
  - Mean of zero
  - Stationary
  - Squared exponential correlation structure
- Need to estimate two parameters: cbar (variance) and ℓ (correlation length)

Learning cbar and ℓ: Toy Example

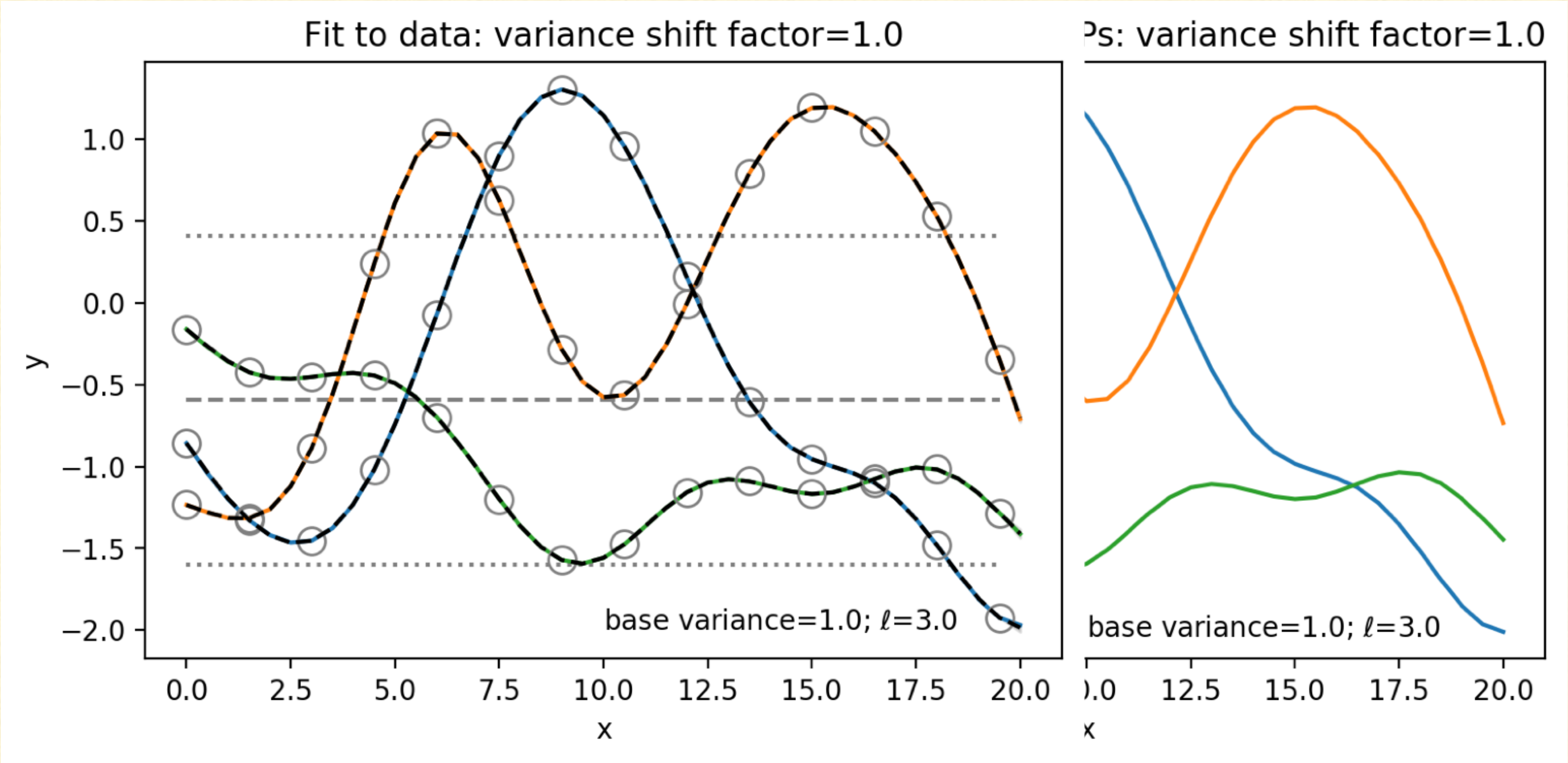
- Generate three data sets from an underlying GP with mean 0, variance 1 and length scale 3.



- Fit one GP to "training data" from all three data sets ("three realizations"): estimate μ, cbar, and ℓ.

Learning cbar and ℓ: Toy Example

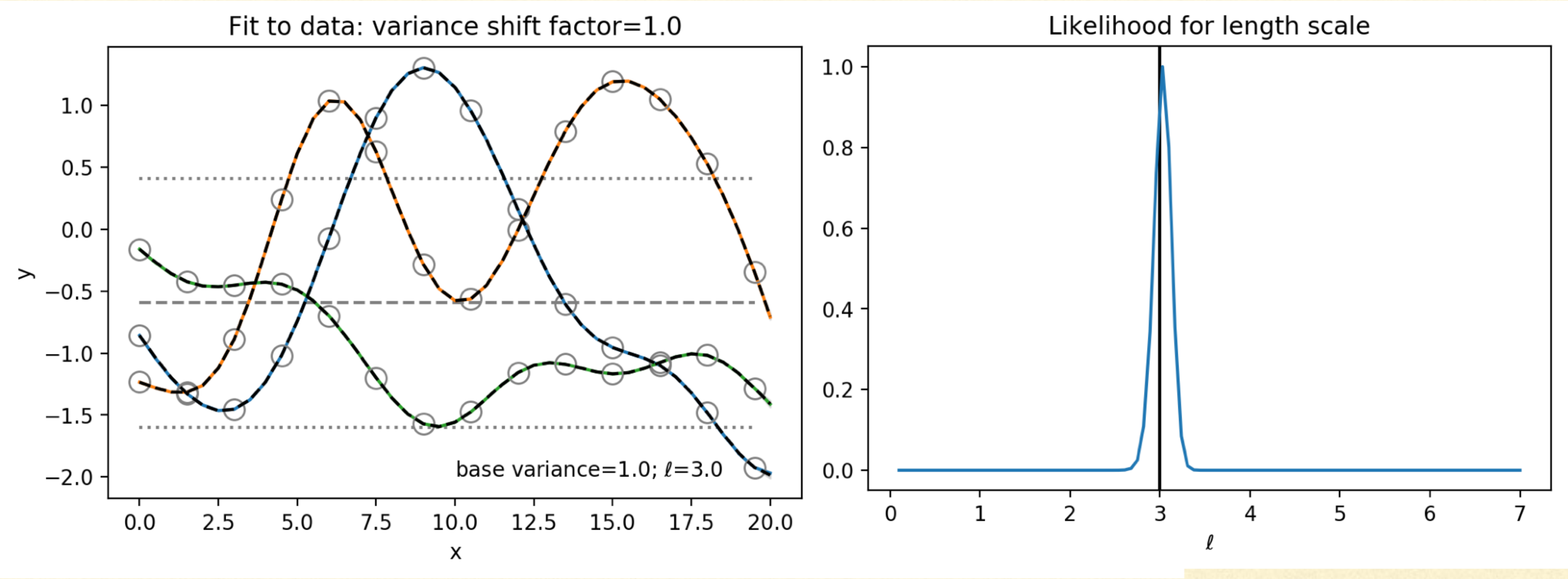
- Generate three data sets from an underlying GP with mean 0, variance 1 and length scale 3.



- Fit one GP to “training data” from all three data sets (“three realizations”): estimate μ, cbar, and ℓ.

Learning cbar and ℓ: Toy Example

- Generate three data sets from an underlying GP with mean 0, variance 1 and length scale 3.



- Fit one GP to "training data" from all three data sets ("three realizations"): estimate μ, cbar, and ℓ.

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

  ⇒equal variance at all locations

  ⇒correlations depend only on x-y

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

  ⇒equal variance at all locations

  ⇒correlations depend only on x-y

- Poor estimate of cbar

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

  ⇒equal variance at all locations

  ⇒correlations depend only on x-y

- Poor estimate of cbar

  ⇒credible intervals too wide or too narrow

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

  ⇒equal variance at all locations

  ⇒correlations depend only on x-y

- Poor estimate of cbar

  ⇒credible intervals too wide or too narrow

- Poor estimate of ℓ

What can go wrong

Bastos & O'Hagan, Technometrics, 2009

- Assumption of GP

  ⇒data no more than 2-3 σ from mean

  ⇒data equally likely above or below

- Assumption of stationarity

  ⇒equal variance at all locations

  ⇒correlations depend only on x-y

- Poor estimate of cbar

  ⇒credible intervals too wide or too narrow

- Poor estimate of ℓ

  ⇒credible intervals too wide or too narrow in vicinity of training points

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

- Assess performance of fitted GP on "testing data" set.

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

- Assess performance of fitted GP on "testing data" set.
- Errors are correlated, so can't just add up number of sigmas and/or compute uncorrelated χ^2
- But plotting individual errors can still be useful

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

- Assess performance of fitted GP on "testing data" set.
- Errors are correlated, so can't just add up number of sigmas and/or compute uncorrelated χ^2
- But plotting individual errors can still be useful
- "Consistency plot" has to account for correlations.

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

- Assess performance of fitted GP on "testing data" set.
- Errors are correlated, so can't just add up number of sigmas and/or compute uncorrelated χ^2
- But plotting individual errors can still be useful
- "Consistency plot" has to account for correlations.
- Define Mahalanobis distance, which does account for correlations

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

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

- Assess performance of fitted GP on "testing data" set.
- Errors are correlated, so can't just add up number of sigmas and/or compute uncorrelated χ^2
- But plotting individual errors can still be useful
- "Consistency plot" has to account for correlations.
- Define Mahalanobis distance, which does account for correlations
$$\mathrm{MD} \equiv (y(x_i) - \mu)^T [k(y(x_i), y(x_j))]^{-1} (y(x_j) - \mu)$$
- Write $k(y(x_i),y(x_j))=G\,G^T$ with G from PC or SV decomposition then form "Pivoted Cholesky" or "Eigen" errors from $G(y(x_i)-\mu)$

GP diagnostics sumnmary

Bastos & O'Hagan, Technometrics, 2009

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

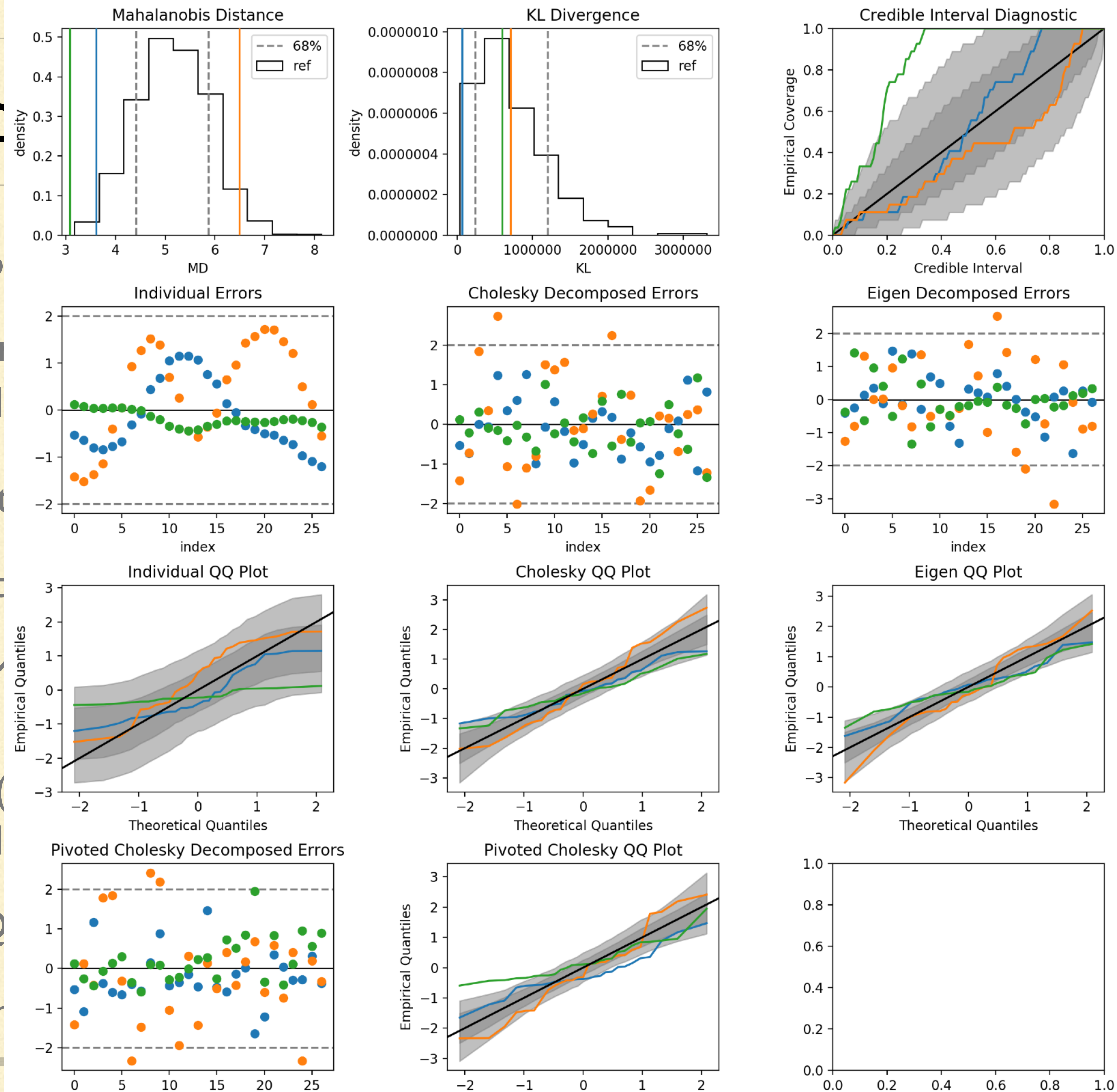
- Assess performance of fitted GP on "testing data" set.
- Errors are correlated, so can't just add up number of sigmas and/or compute uncorrelated χ^2
- But plotting individual errors can still be useful
- "Consistency plot" has to account for correlations.
- Define Mahalanobis distance, which does account for correlations
$$\mathrm{MD} \equiv (y(x_i) - \mu)^T [k(y(x_i), y(x_j))]^{-1} (y(x_j) - \mu)$$
- Write $k(y(x_i),y(x_j))=G\,G^T$ with G from PC or SV decomposition then form "Pivoted Cholesky" or "Eigen" errors from $G(y(x_i)-\mu)$
- Can QQ plot any of these individual errors
- Although in the end structure as a function of index may be more revealing

GP c

2009

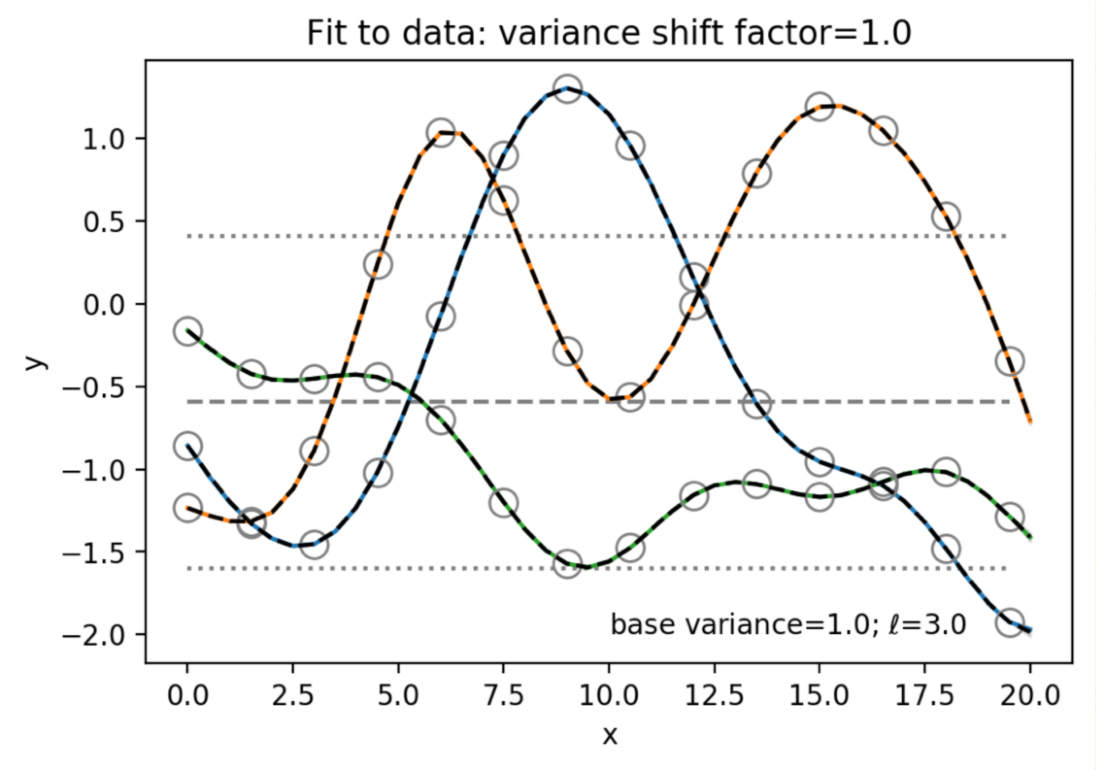
(2019)

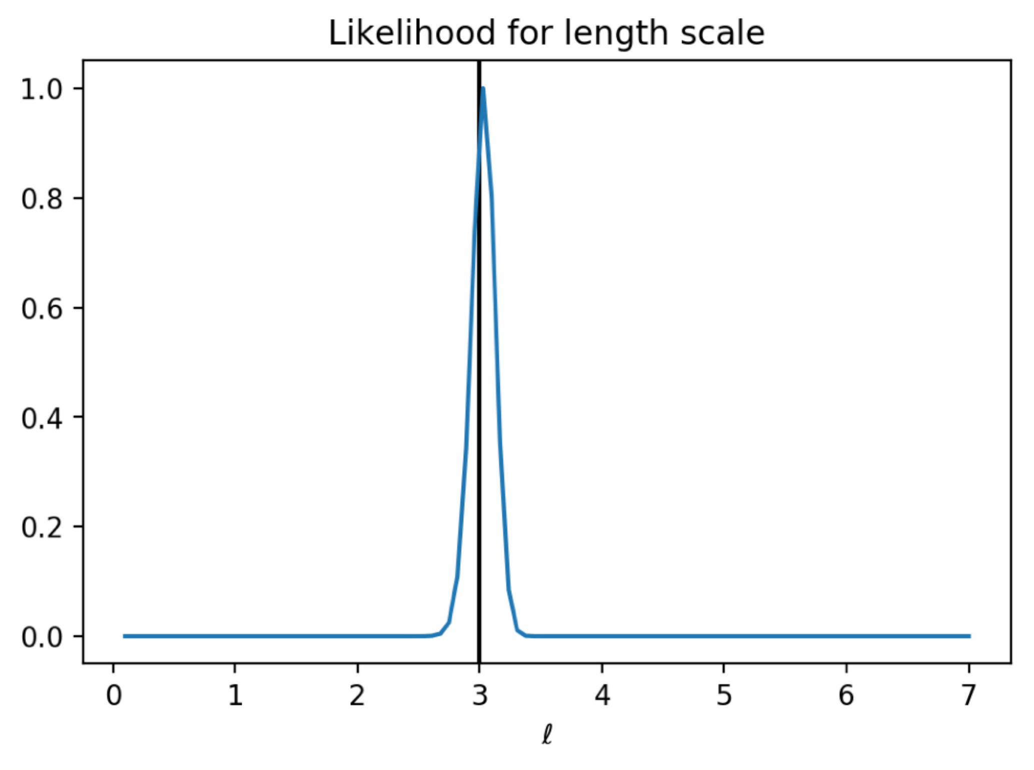
- Assess p
- Errors ar ute uncorrel
- But plott
- “Consist
- Define M
- Write k(m “Pivoted
- Can QQ
- Althoug ng



Model checking a GP

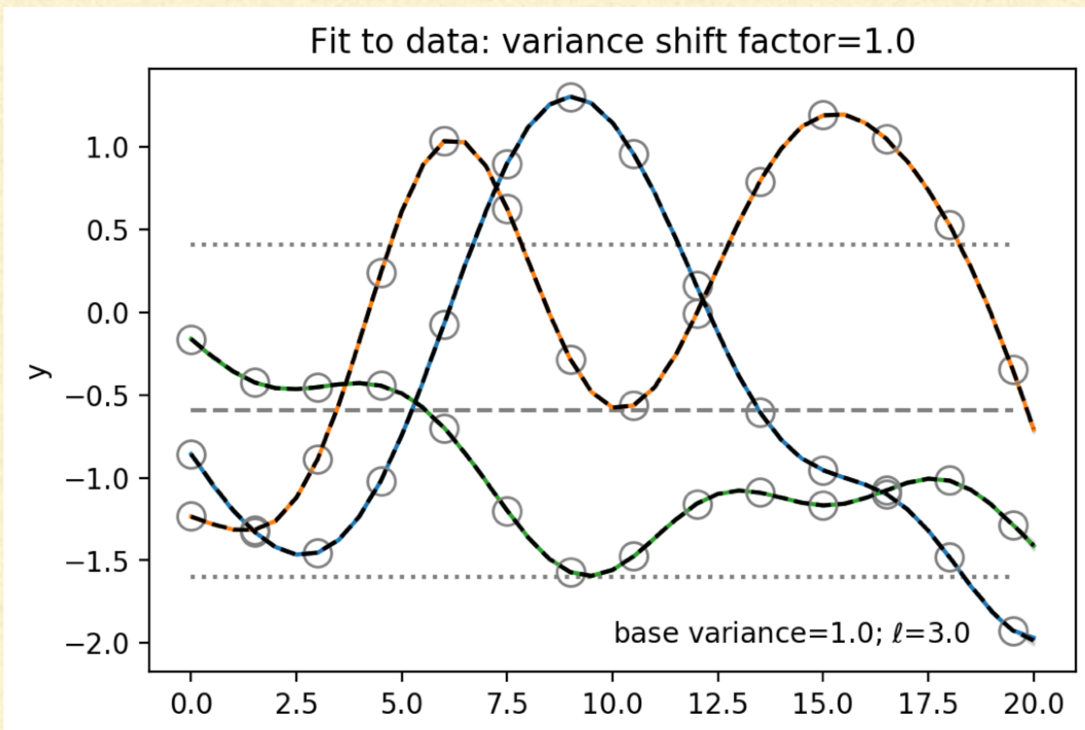
- Fit one GP to "training data" from all three data sets ("three realizations"): estimate μ, cbar, and ℓ.

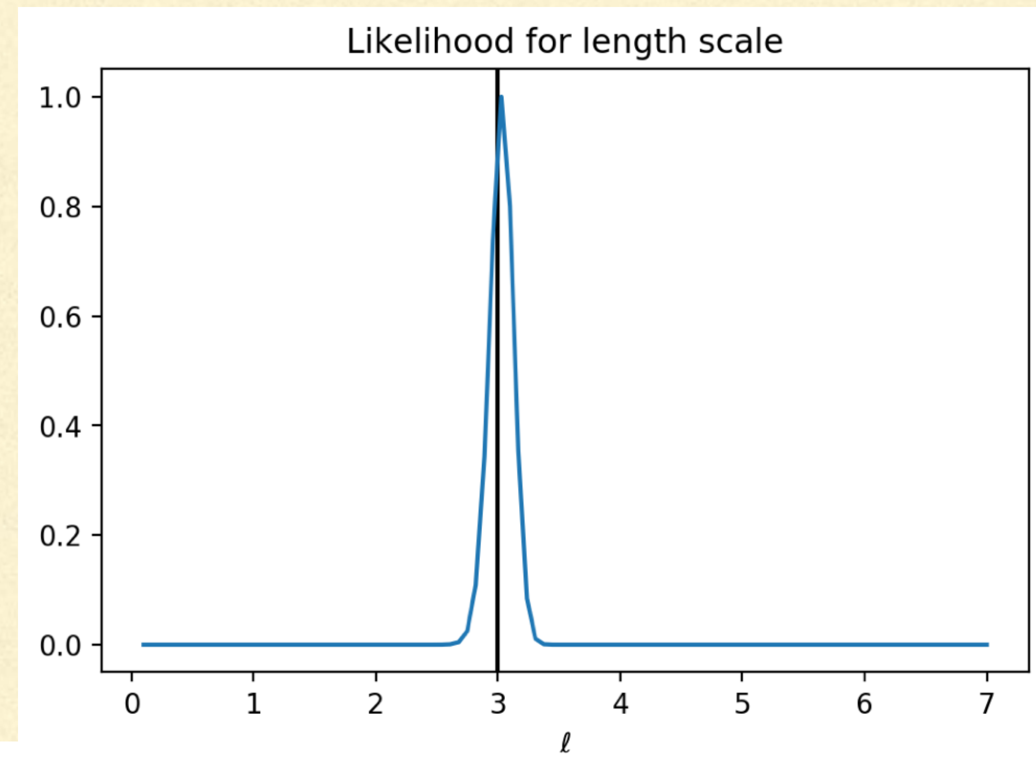


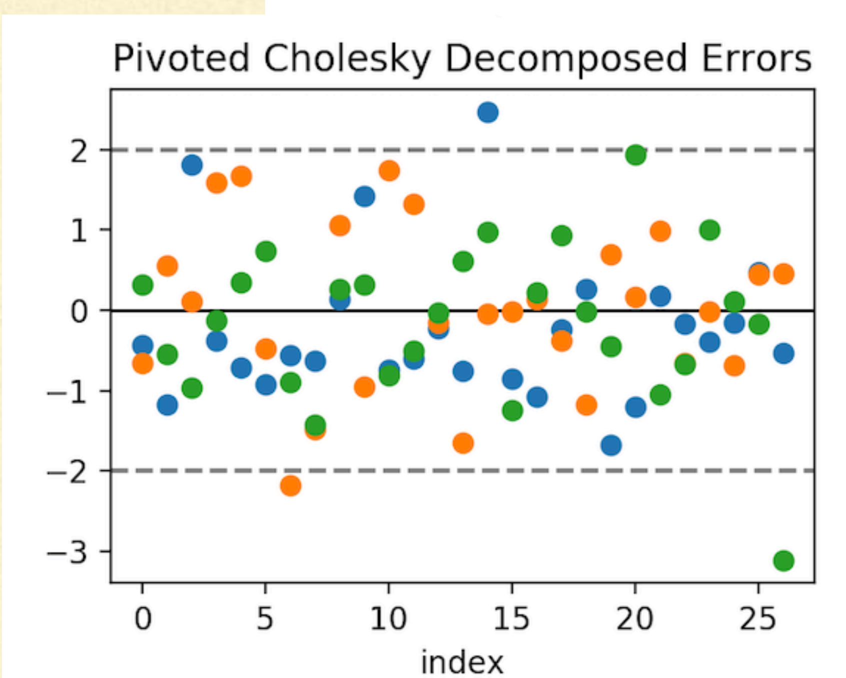


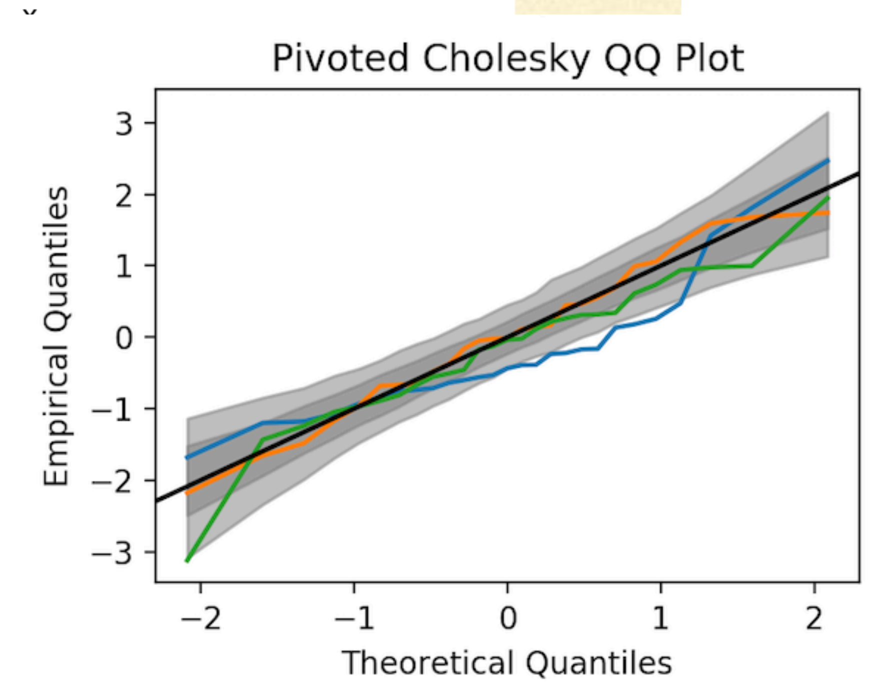
Model checking a GP

- Fit one GP to "training data" from all three data sets ("three realizations"): estimate μ, cbar, and ℓ.



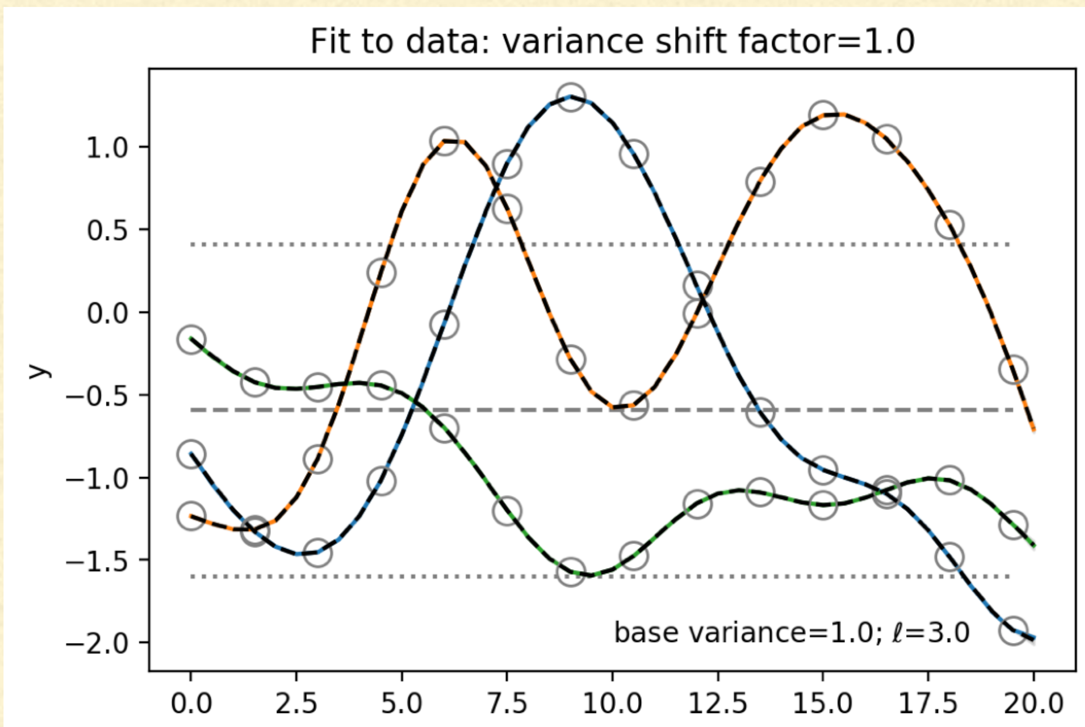


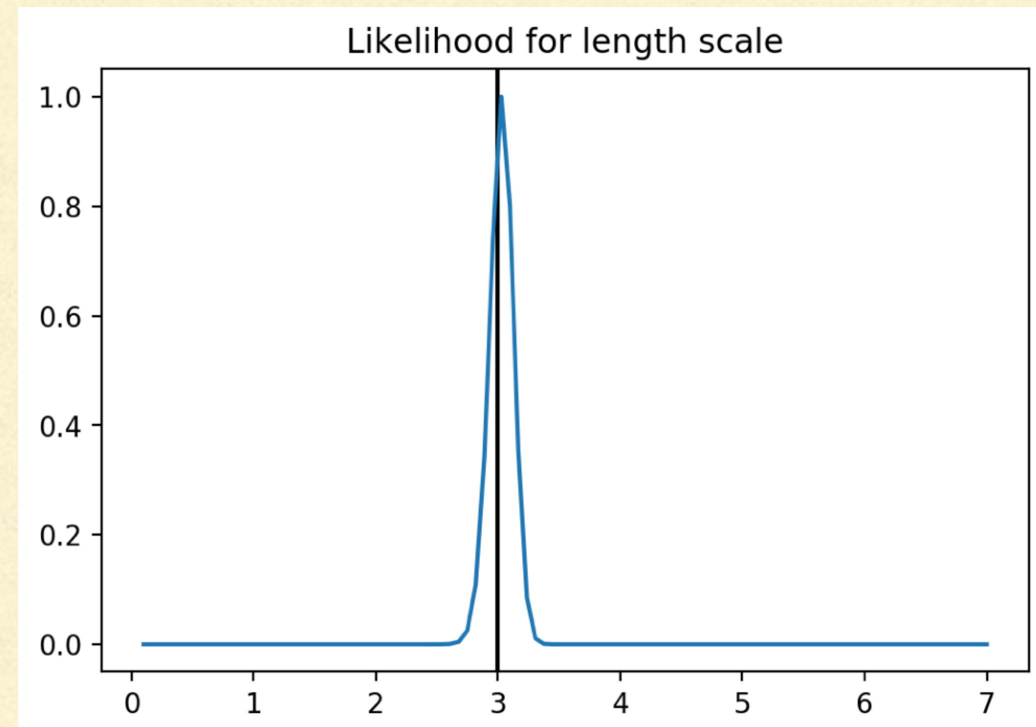


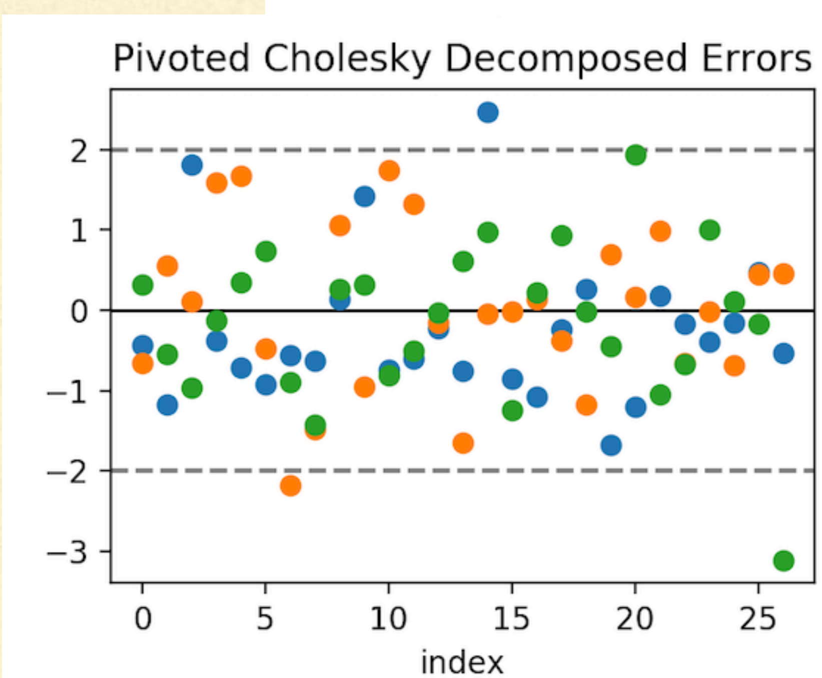


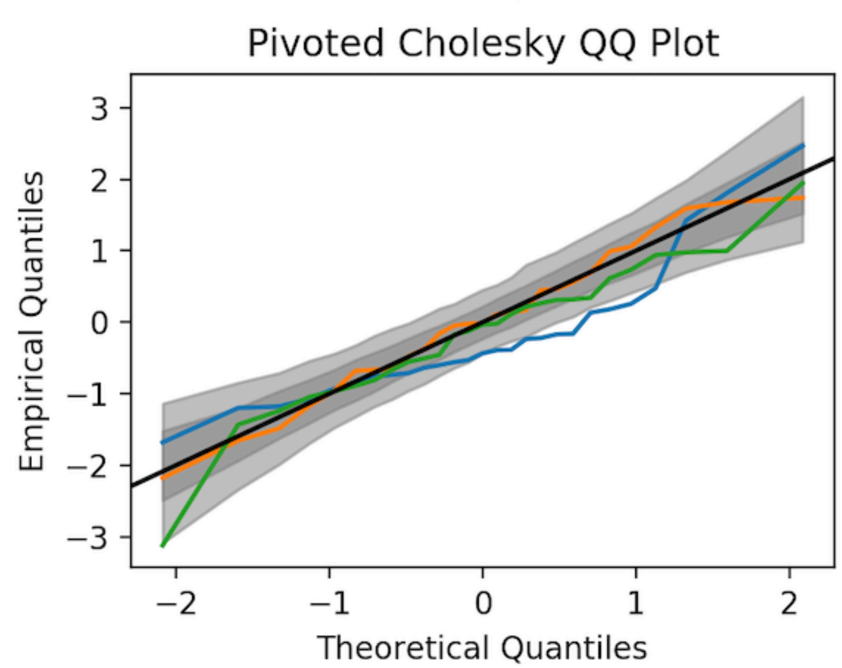
Model checking a GP

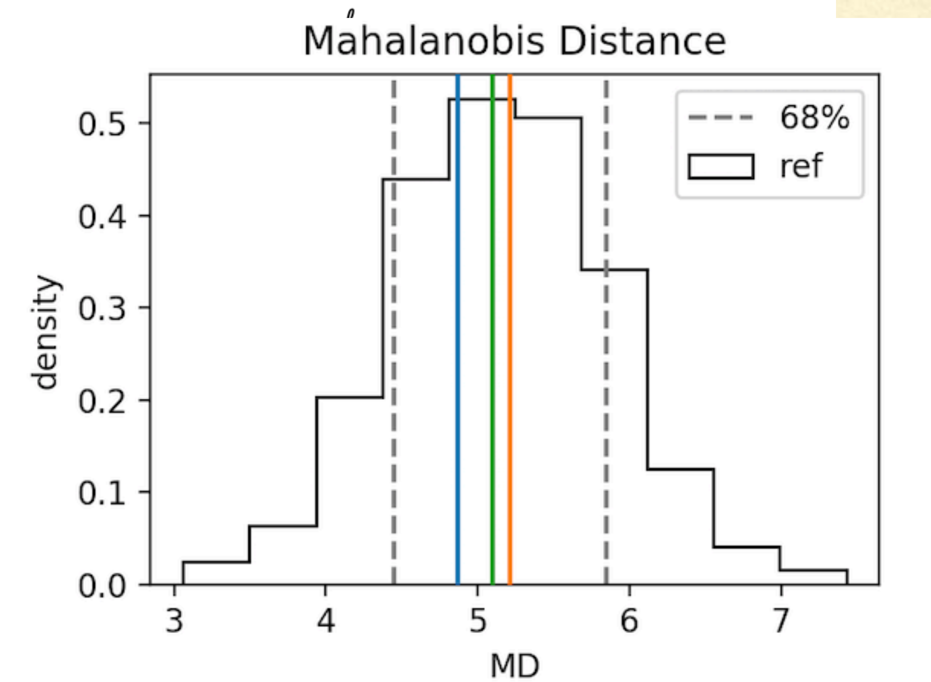
- Fit one GP to “training data” from all three data sets (“three realizations”): estimate μ, cbar, and ℓ.









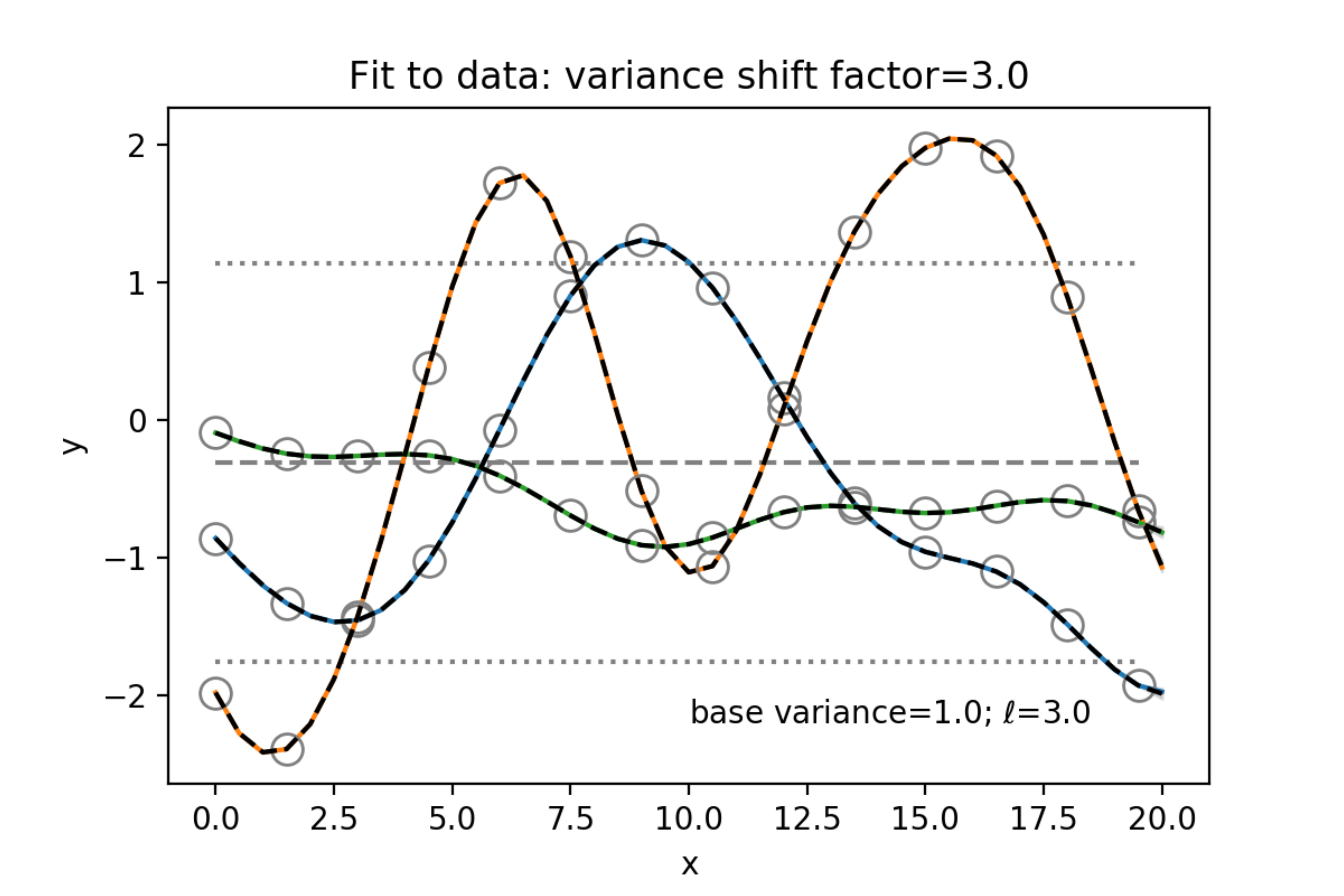


What can go wrong 1: different cbar's

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

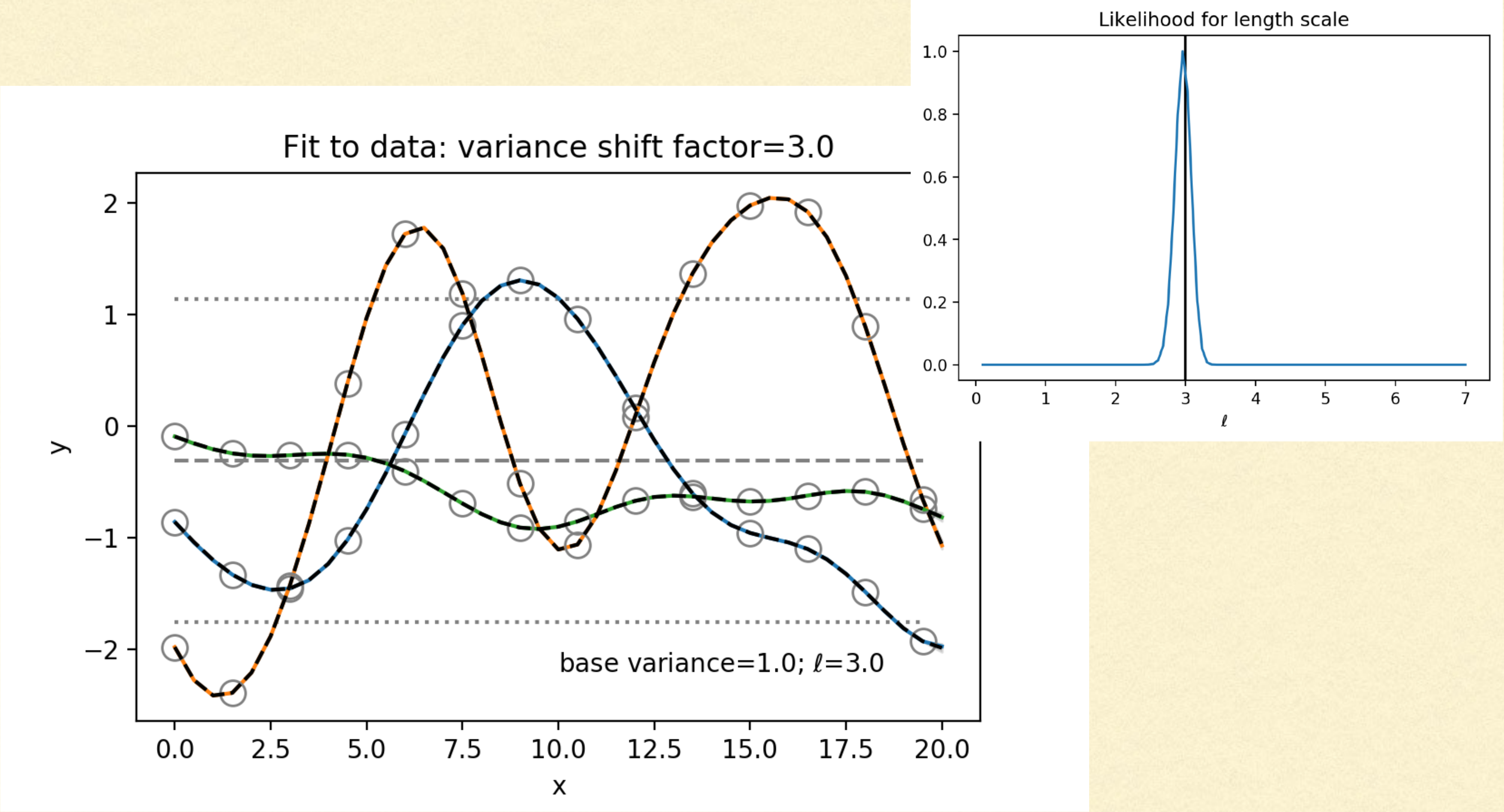
What can go wrong 1: different cbar's

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



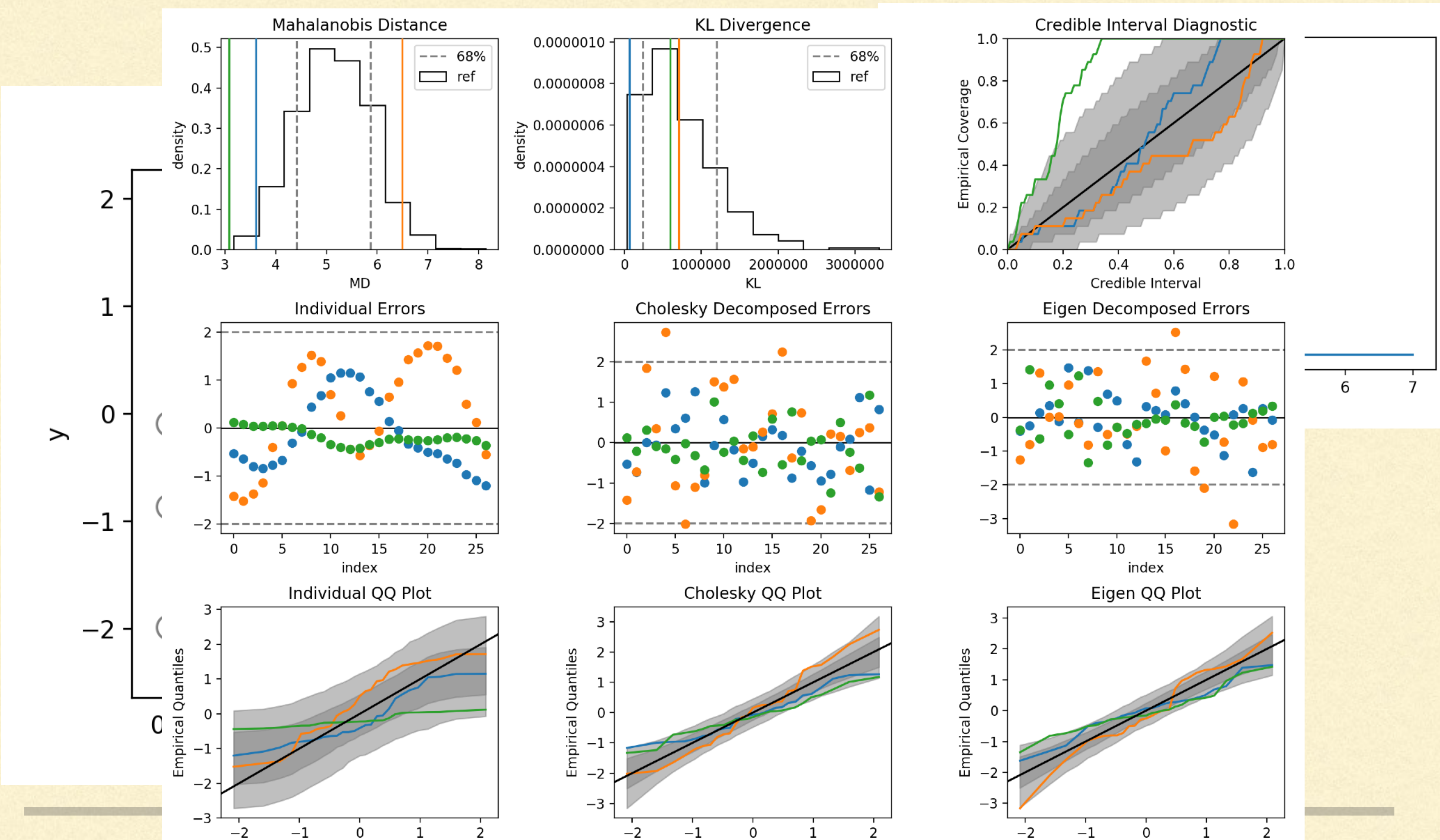
What can go wrong 1: different cbar's

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



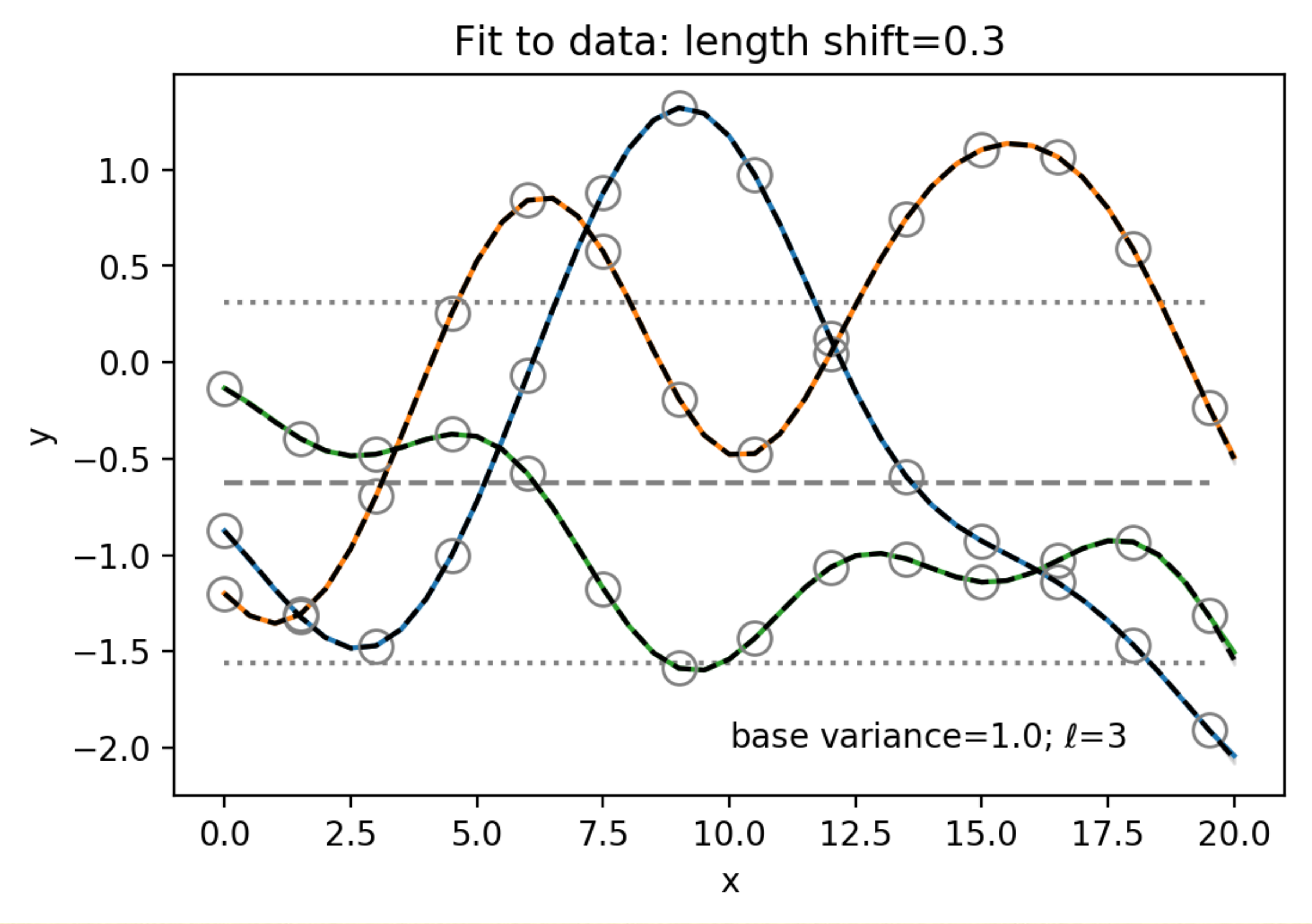
What can go wrong 1: different cbar's

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



What can go wrong II: different ℓs

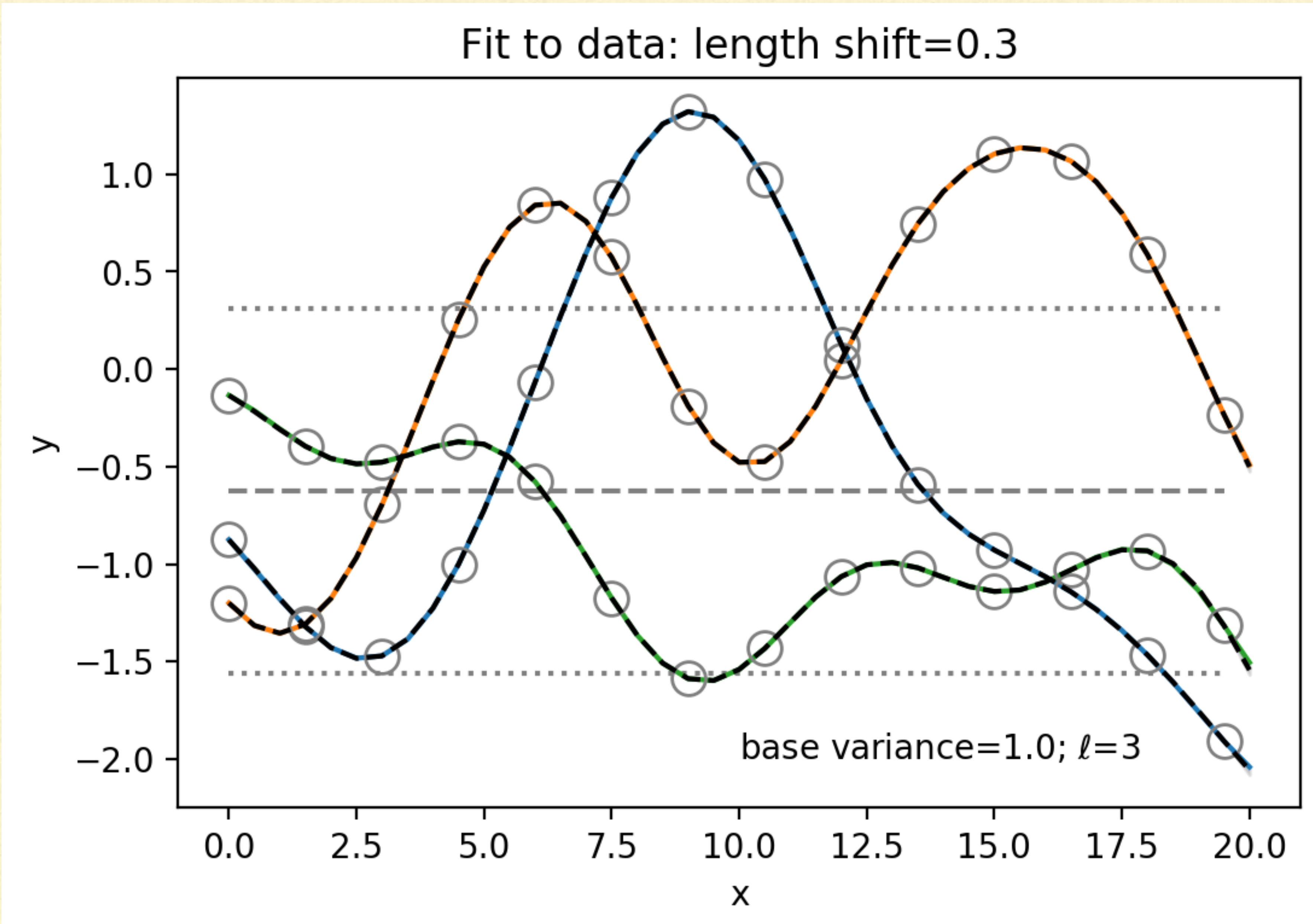
- Try to fit single GP to data with different ℓs



What can go wrong II: different ℓs

- Try to fit single GP to data with different ℓs

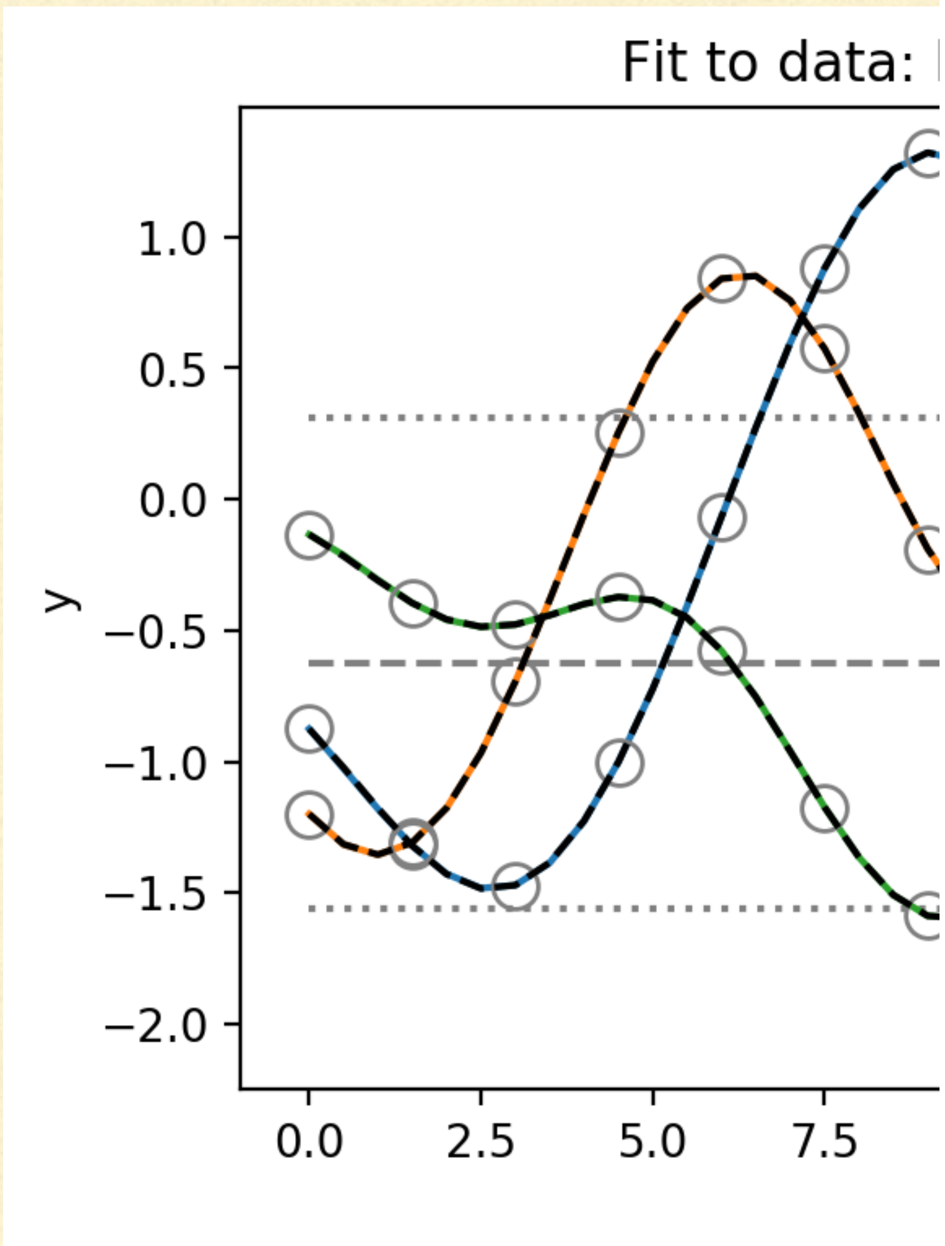
ℓ=3.1

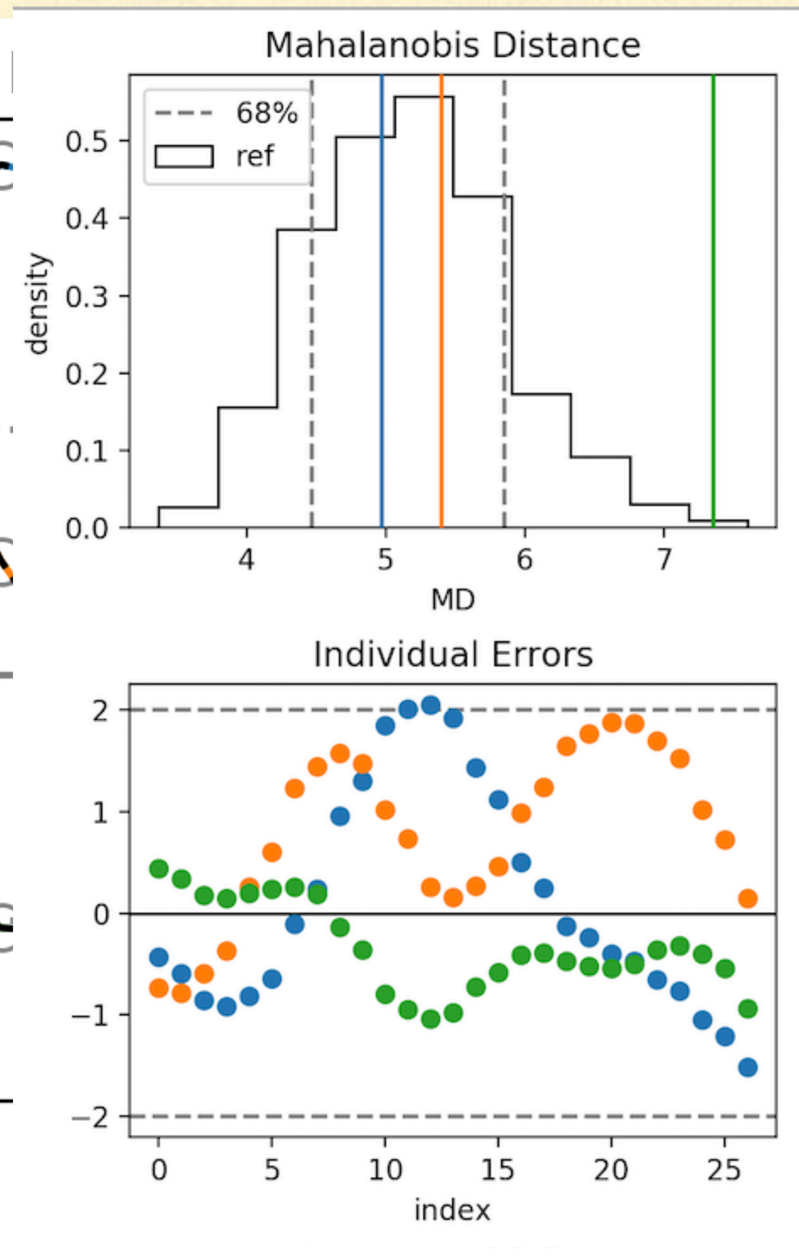


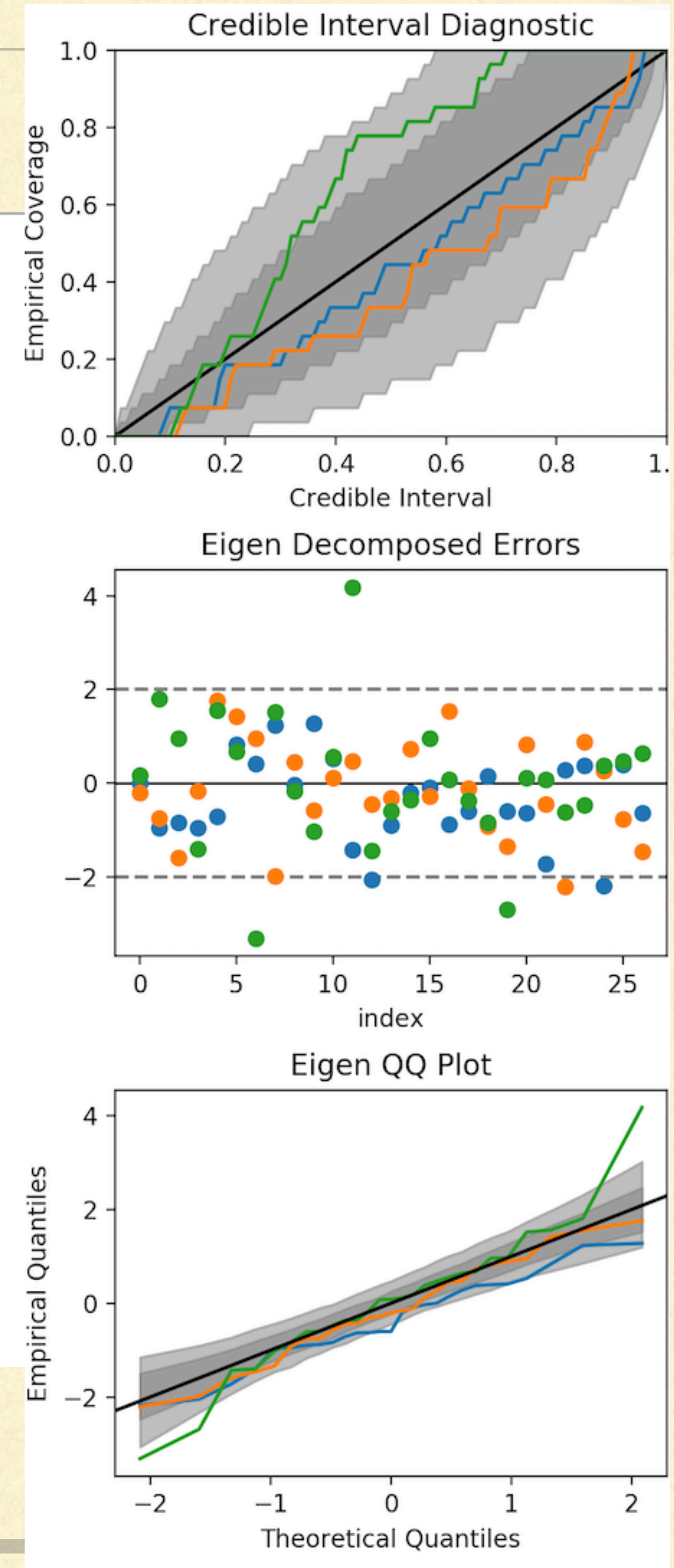
What can go wrong II: different ℓs

- Try to fit single GP to data with different ℓs

ℓ=3.1

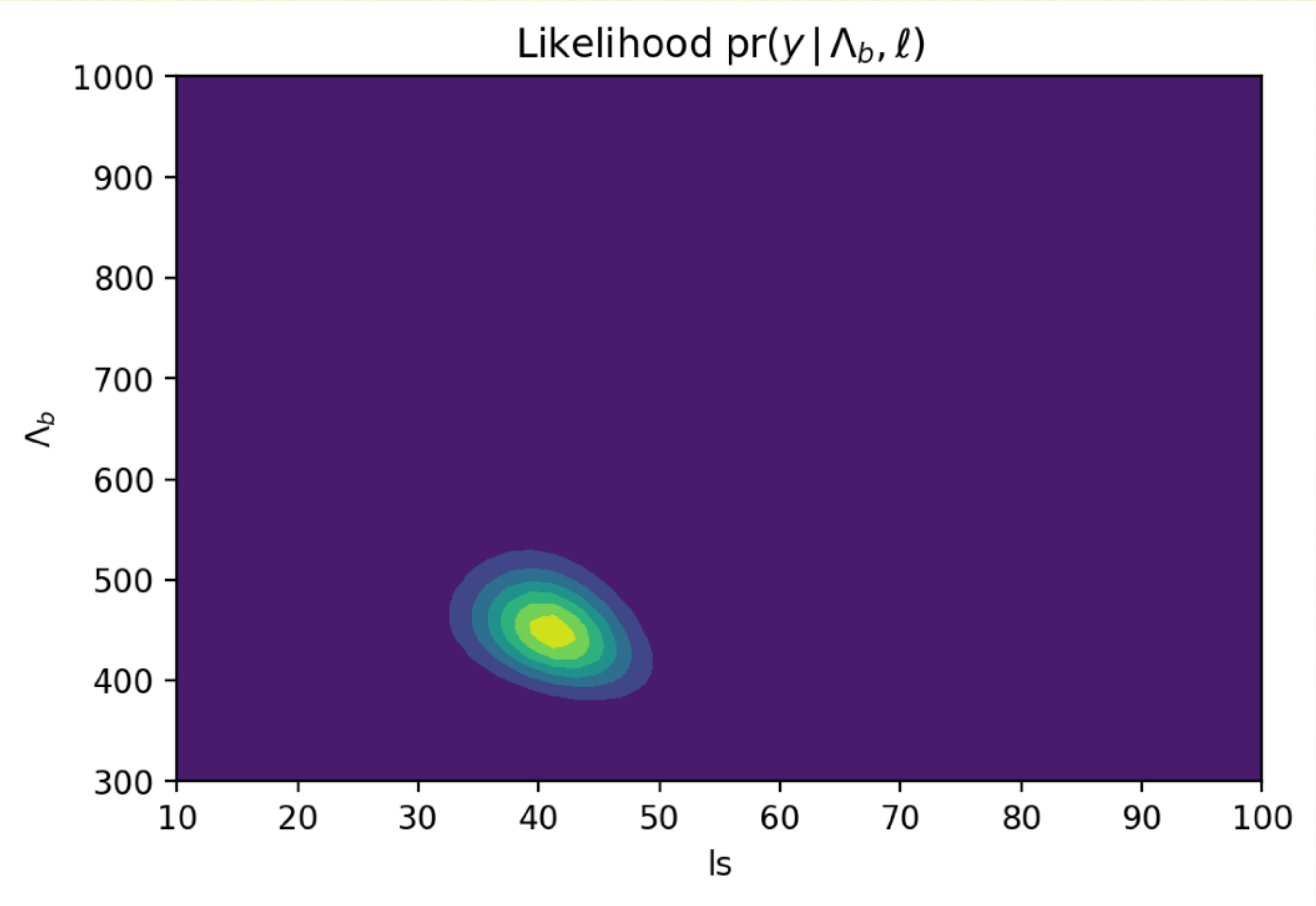






Λ_b and ℓ inference

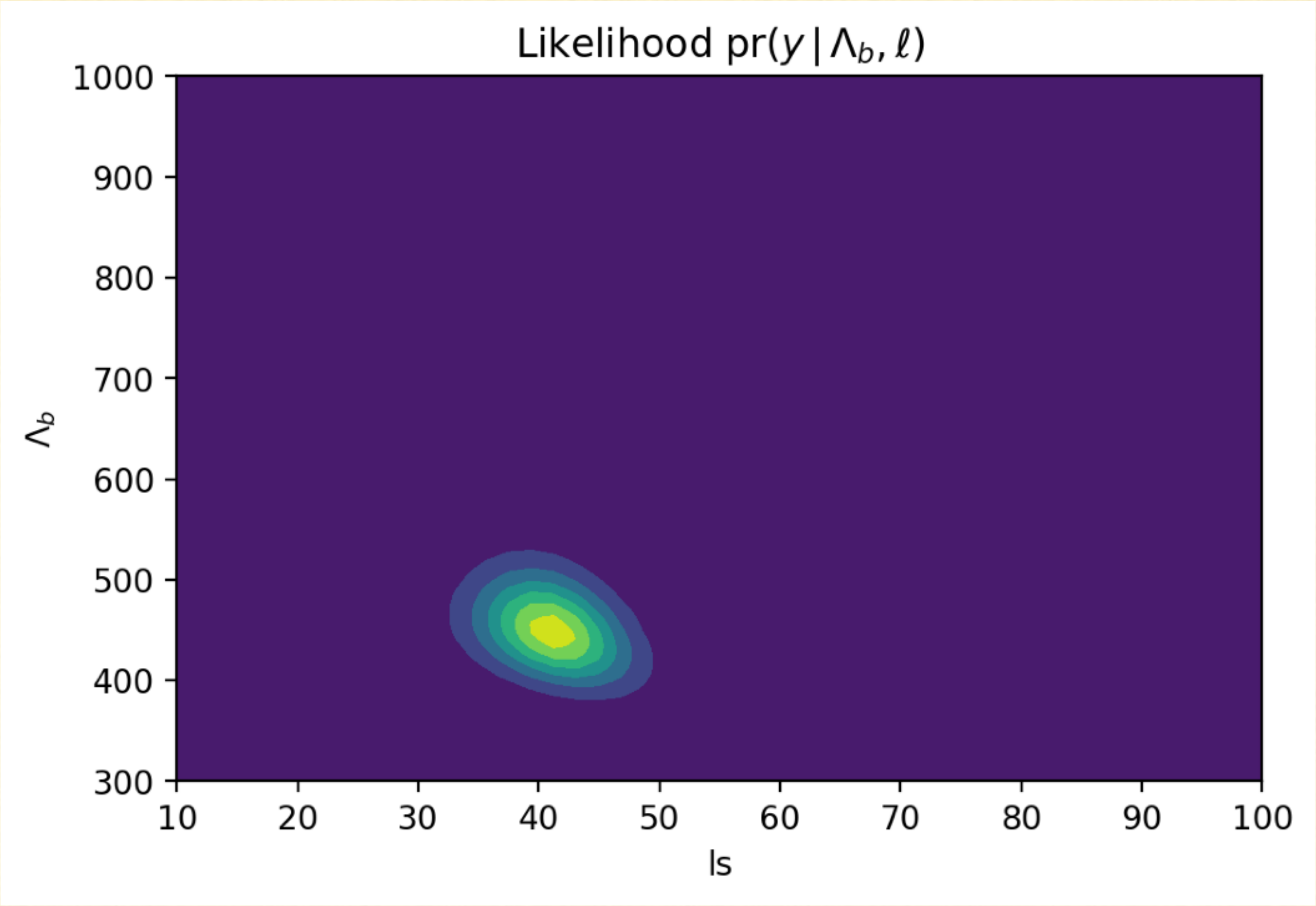
PRELIMINARY



RKE potential; Λ=500 MeV;
Total cross section only

Λ_b and ℓ inference

PRELIMINARY

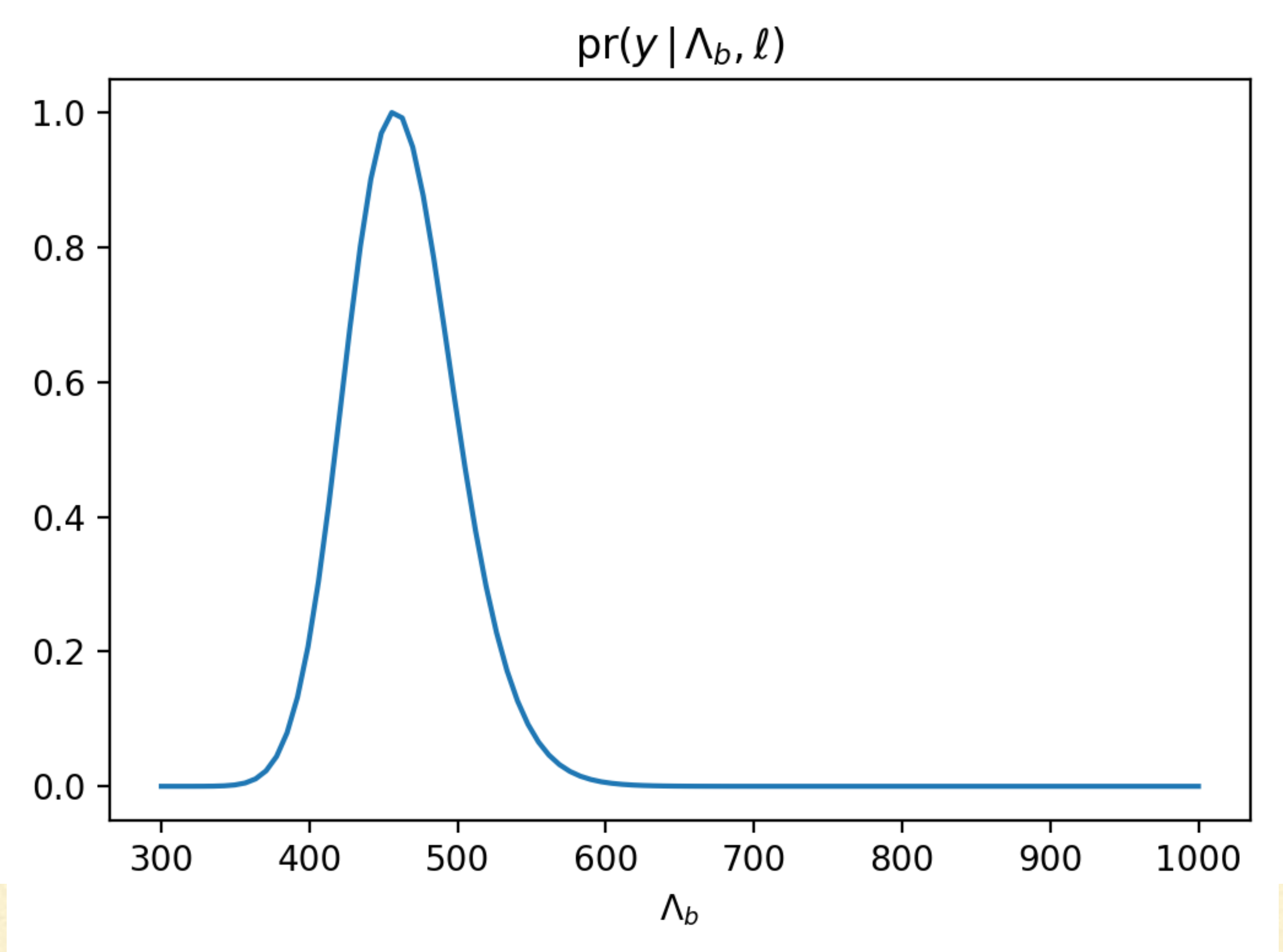


RKE potential; Λ=500 MeV;
Total cross section only

$\ell \approx 40$ MeV

Λ_b and ℓ inference

PRELIMINARY

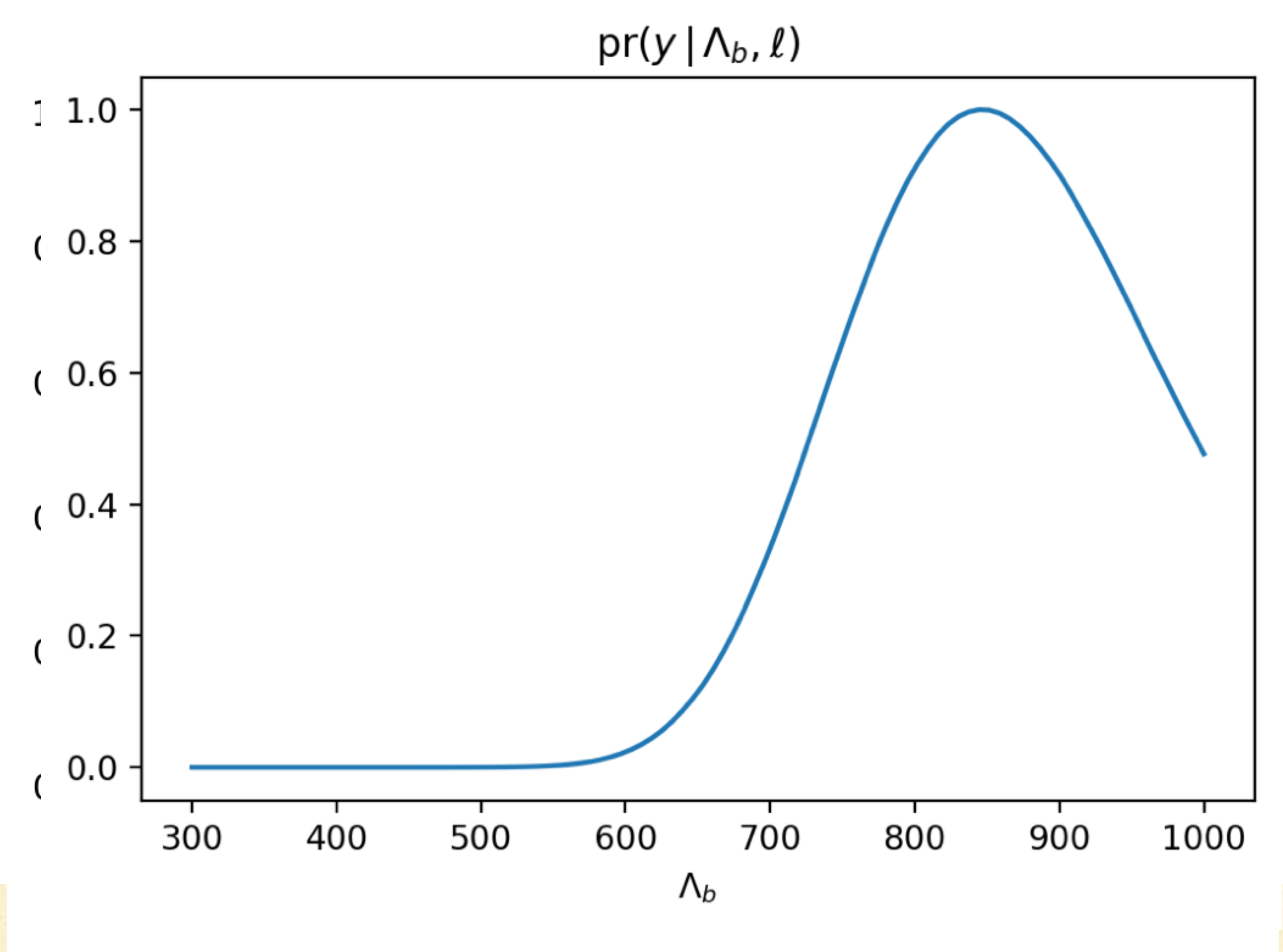


RKE potential; Λ=500 MeV;
Total cross section only

$\ell \approxeq 40$ MeV

Λ_b and ℓ inference

PRELIMINARY



RKE potential; Λ=500 MeV;

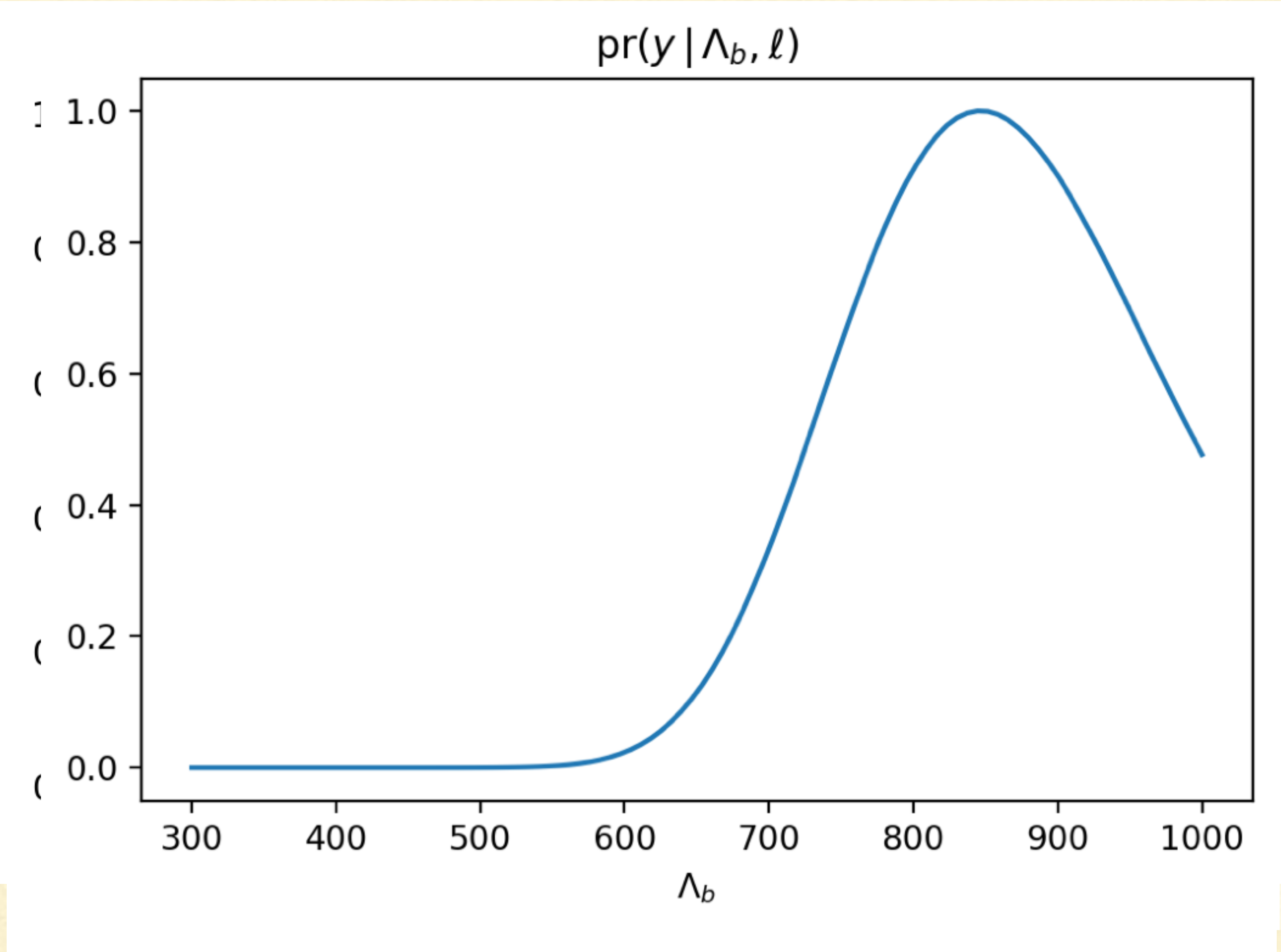
Total cross section only

Restricted domain

$\ell \approx 116$ MeV

Λ_b and ℓ inference

PRELIMINARY



RKE potential; Λ=500 MeV;

Total cross section only

Restricted domain

$\ell \approx 116$ MeV

To do:

- Model non-stationarity
- Play with “switch over” of expansion parameter at $p \approx m_\pi$
- Include other observables in analysis