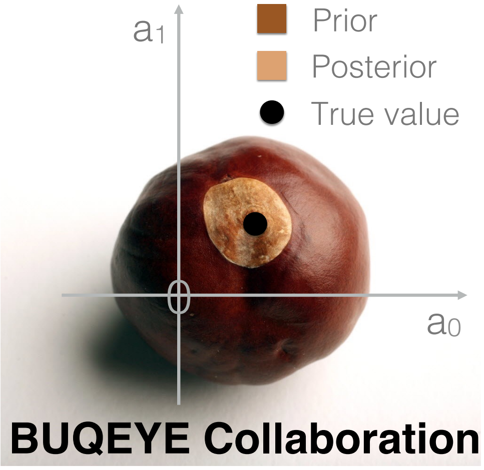
Bayesian methods for effective field theories

Sarah Wesolowski
The Ohio State University
INT Bayes Workshop 2016



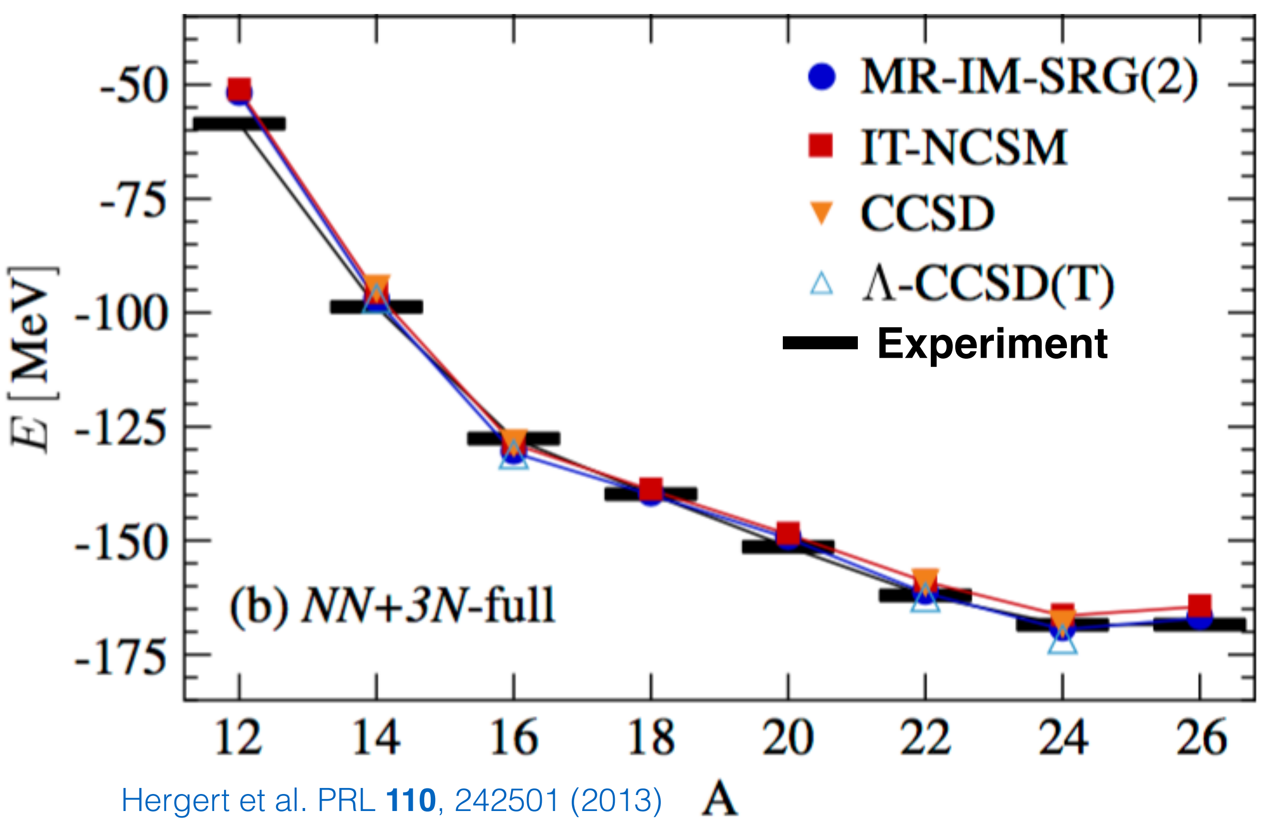
- Dick Furnstahl (OSU)
- Daniel Phillips (OU)
- Arbin Thapilaya (OU)
- Natalie Klco (OU -> UW)
- Harald Griesshammer (GWU)

Outline

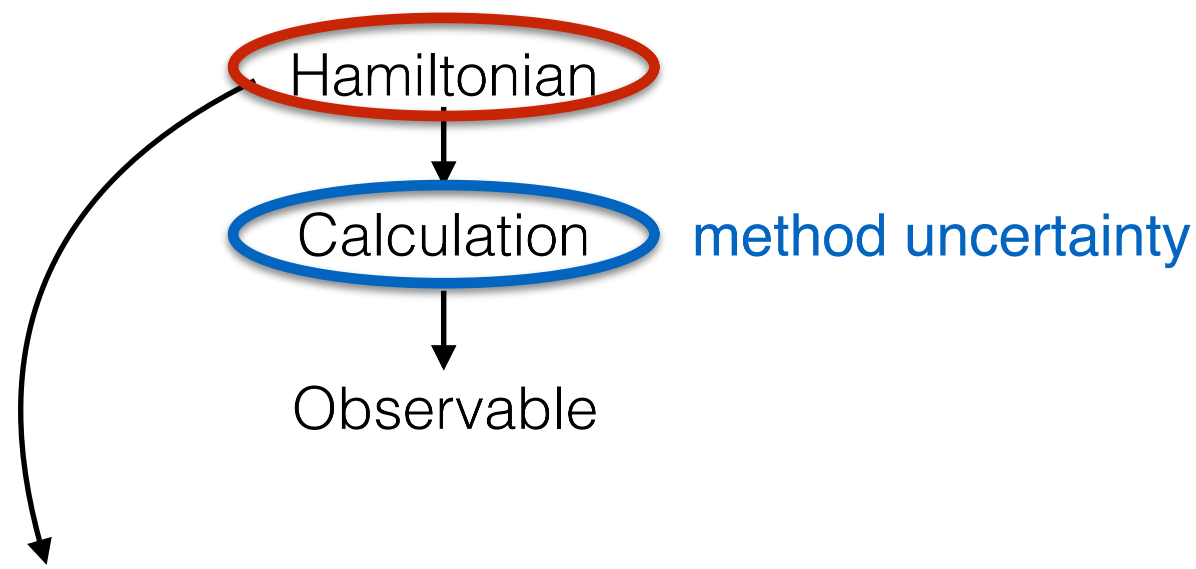
- Introduction to effective field theories
- The parameter estimation problem
- Diagnostics for parameter estimation
- Model problems to the chiral EFT problem
- Conclusions and questions

Era of precision observable calculations

Ground-state energies for even oxygen isotopes



Hergert et al. PRL **110**, 242501 (2013)

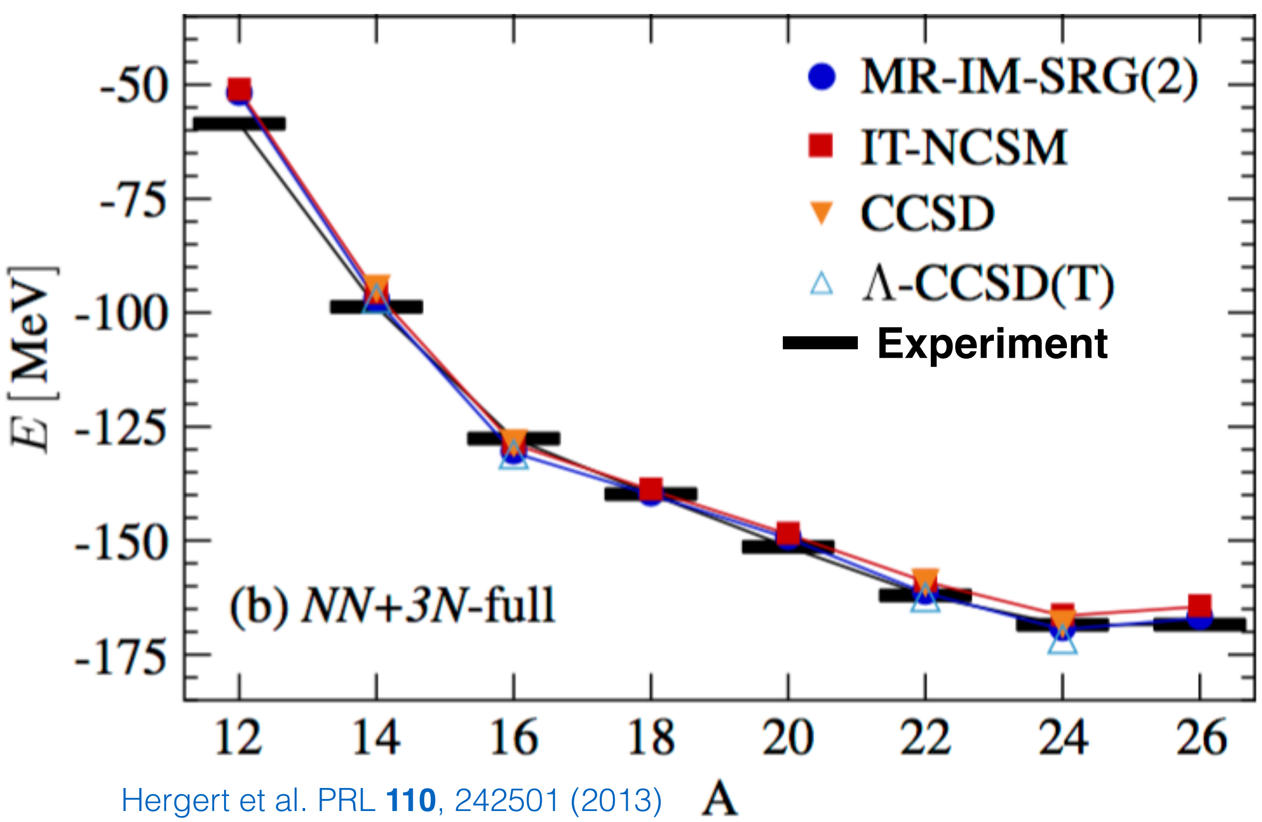


Where are the theory errors?

Uncertainty quantification for modern nuclear calculations

Era of precision observable calculations

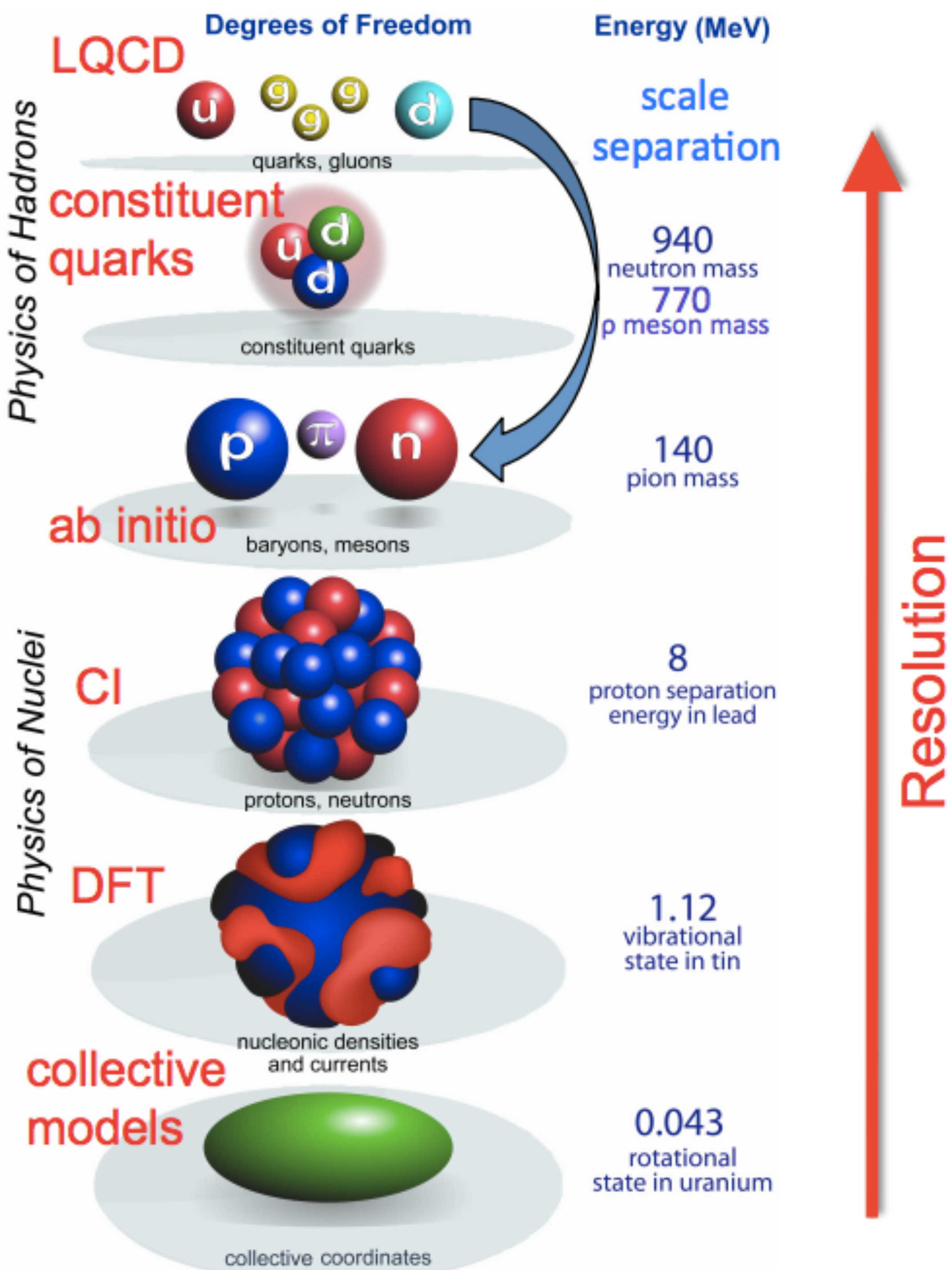
Ground-state energies for even oxygen isotopes



Hergert et al. PRL **110**, 242501 (2013)

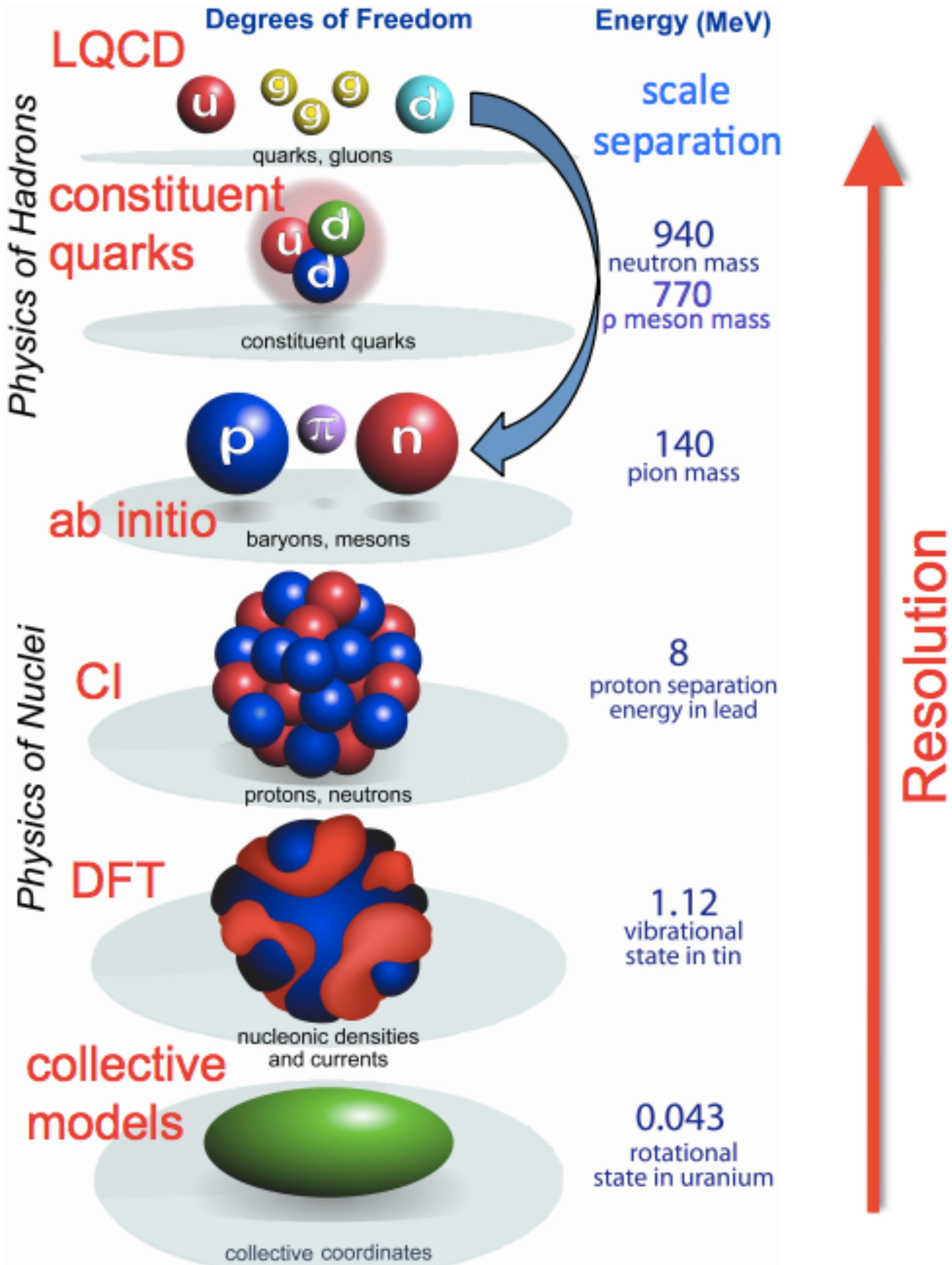
- Include all sources of uncertainty
- Theory uncertainty from input interaction.
- Current efforts on interaction uncertainty:
  - Navarro Pérez *et al.*, JPG **42**, 034013 (2015)
  - Epelbaum *et al.*, EPJA **51**, 5 (2015)
  - Carlsson *et al.*, PRX **6**, 011019 (2016)
- Focus on Bayesian approach:
  - SW *et al.*, JPG **43** 074001(2016)
  - RJF *et al.*, PRC **92**, 024005 (2015)
  - RJF *et al.*, JPG **42**, 034028 (2015)

Effective field theories and you



- Relevant degrees of freedom depend on resolution
- Construct most general theory consistent with symmetries
- “Model-independent” predictions
- Scale separation leads to expansion parameter
- “Natural-sized” low-energy constants (LECs): controlled expansion

Effective field theories and you



- Relevant degrees of freedom depend on resolution
- Construct most general theory consistent with symmetries
- "Model-independent" predictions
- Scale separation leads to expansion parameter
- "Natural-sized" low-energy constants (LECs): controlled expansion

Can estimate theoretical uncertainties!

Parameter estimation problem (schematic)

Fit coefficients (LECs) of this function

$$V(x) \sim V_0 \sum_{n=0}^{k} a_n x^n$$

$$x = \frac{p}{\Lambda_b}$$

Small expansion parameter x

Parameter estimation problem (schematic)

These are O(1)

Fit coefficients (LECs) of this function

$$V(x) \sim V_0 \sum_{n=0}^{k} a_n x^n$$

$$x = \frac{p}{\Lambda_b}$$

Small expansion parameter x

Solution method: e.g., Lippman-Schwinger

Observable: match to experiment

$$\eta(x)$$

Parameter estimation problem (schematic)

Uncertainty from fitting to data

$$V(x) \sim V_0 \sum_{n=0}^{k} a_n x^n + V_0 \sum_{n=k+1}^{k_{\max}} a_n x^n$$

All higher powers we left out of calculation

Solution method:
e.g., Lippman-Schwinger

Numerical errors in calculation method

$$\eta(x)$$

Parameter estimation problem (schematic)

Order to which we calculate

$$V(x) \sim V_0 \sum_{n=0}^{k} a_n x^n + V_0 \sum_{n=k+1}^{k_{\max}} a_n x^n$$

Account for corrections above k with this: marginalization!

↓

Solution method:
e.g., Lippman-Schwinger

↓

$$\eta(x)$$

Issues to address

Uncertainty budget

- a_n's from fit to data
- truncation error
- calculation method

$$V(x;\mathbf{a},k) = V_0 \sum_{n=0}^{k} a_n x^n + \Delta_k(x)$$

$$x = \frac{p}{\Lambda_b}$$

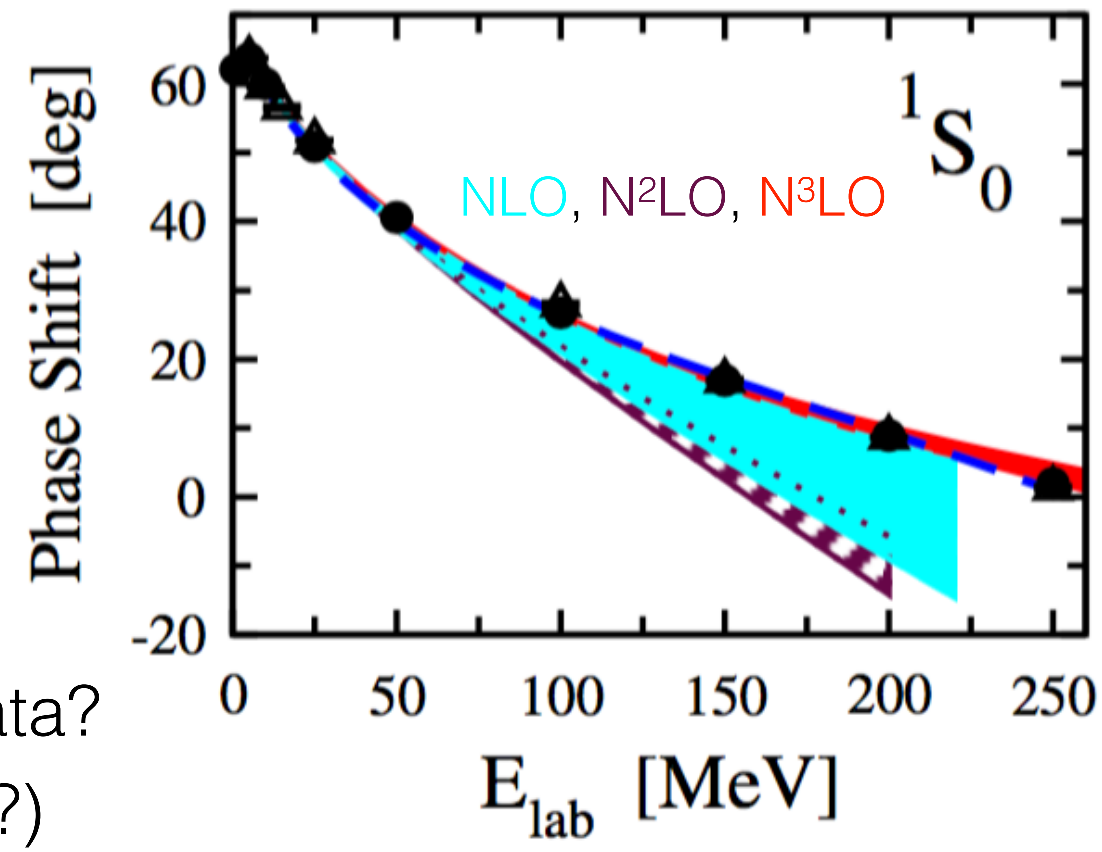
What could possibly go wrong?

- Form of V at higher order may not be known
- V_0 and Λ_b may not be properly identified
- Can have unnaturally large a_n's
- Regulator artifacts (won't go into detail here)

What was done before

- Vary piece of calculation:
  Cutoff regulator Λ
  - Convergence with order unclear
  - Underestimates uncertainty
  - No statistical interpretation

Fits of $V(x; \mathbf{a}, k)$ to one type of $\eta(x)$



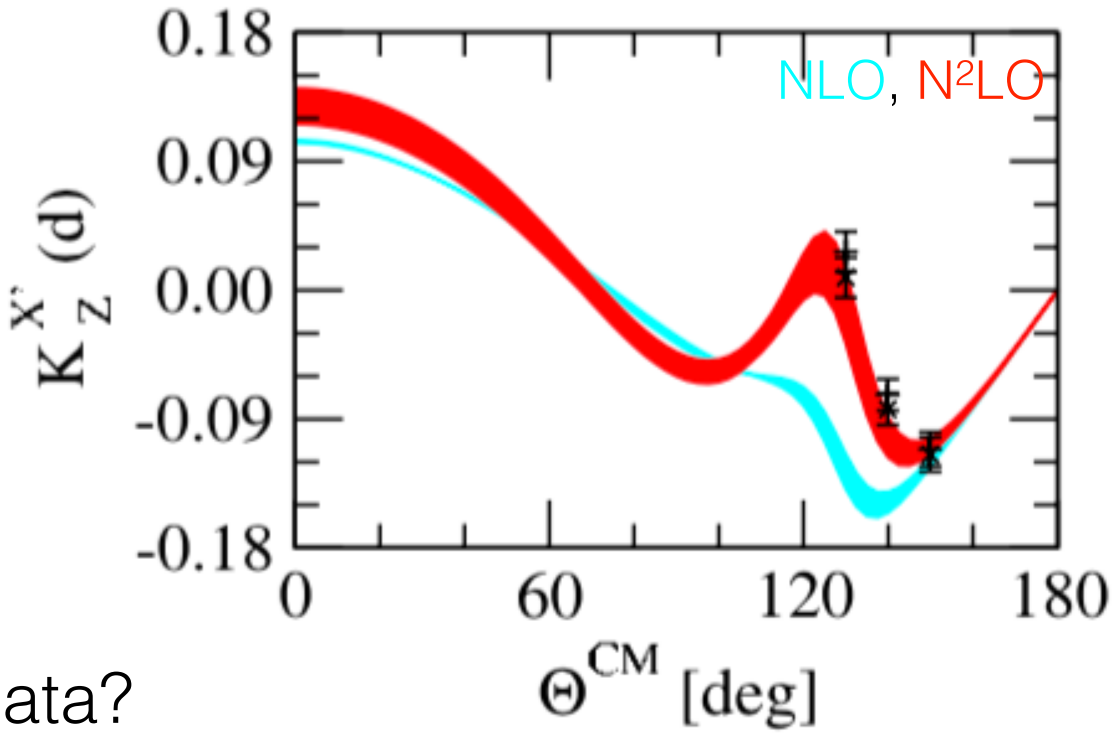
Epelbaum et al., Rev. Mod. Phys. **81** 1773–825

- Issues with past analyses:
  - What is the LEC uncertainty from fit to data?
  - Optimization procedures opaque (priors?)
  - Underfitting avoided by limiting range of data

What was done before

- Vary piece of calculation:
  Cutoff regulator Λ
  - Convergence with order unclear
  - Underestimates uncertainty
  - No statistical interpretation

Use $V(x; \mathbf{a}, k)$ to predict a $\eta(x)$ not in fit

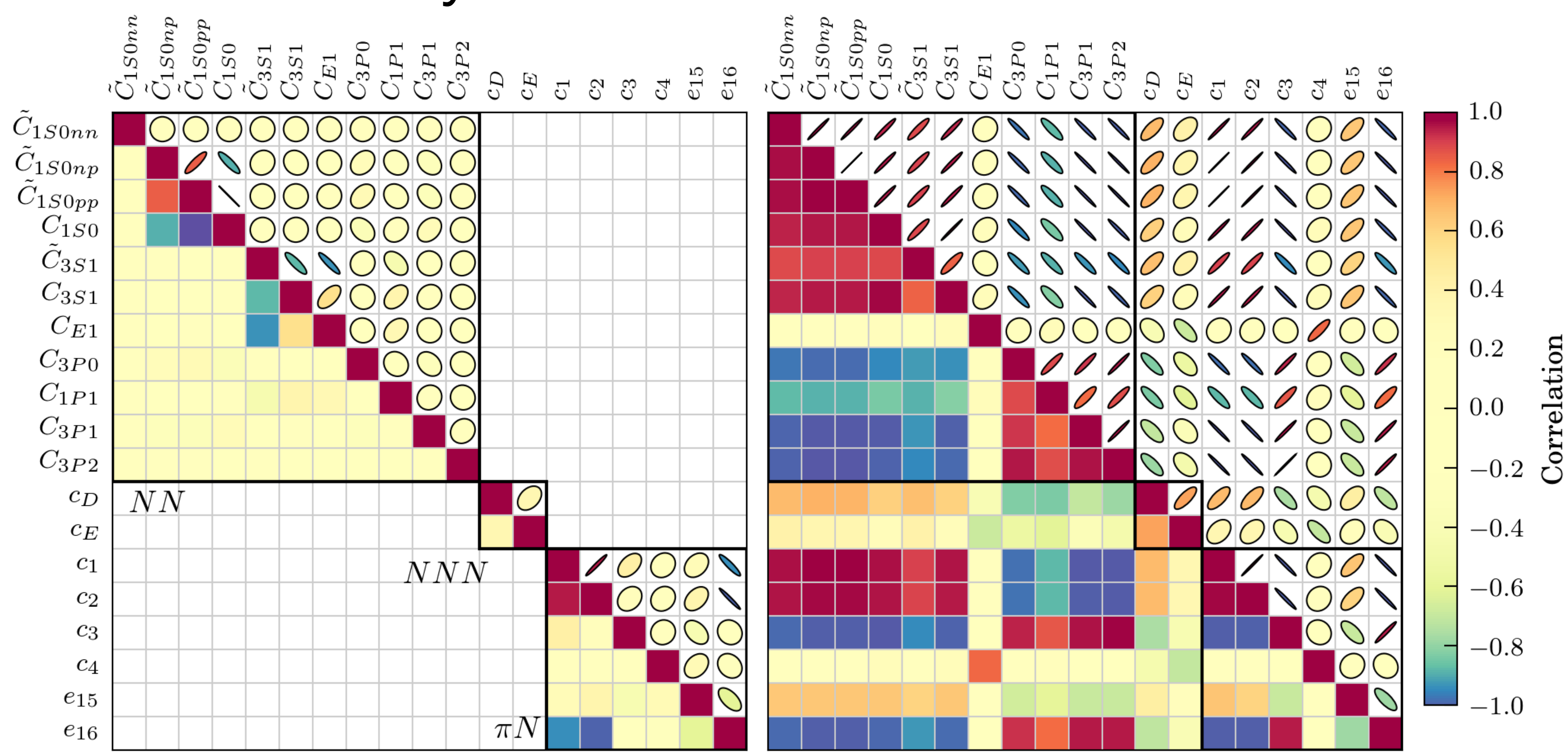


- Issues with past analyses:
  - What is the LEC uncertainty from fit to data?
  - Optimization procedures opaque (priors?)
  - Underfitting avoided by limiting range of data

Epelbaum et al., Rev. Mod. Phys. **81** 1773–825
One proton-to-deuteron polarization transfer coefficient

More modern approaches: uncertainty from data

- Covariance analysis of LECs **a** from EFT fit to data



Carlsson *et al.*, *Phys*. Rev. X **6**, 011019 (2016)

- Are they really Gaussian?
- Inclusion of truncation error in fits? (Birge factors)
- Simultaneous or separate fits to all types of data?

More modern approaches: truncation uncertainty

$$\eta(x;k) = \eta_0 \sum_{n=0}^{k} c_n x^n + \Delta_k(x)$$

Given these order-by-order calculations…

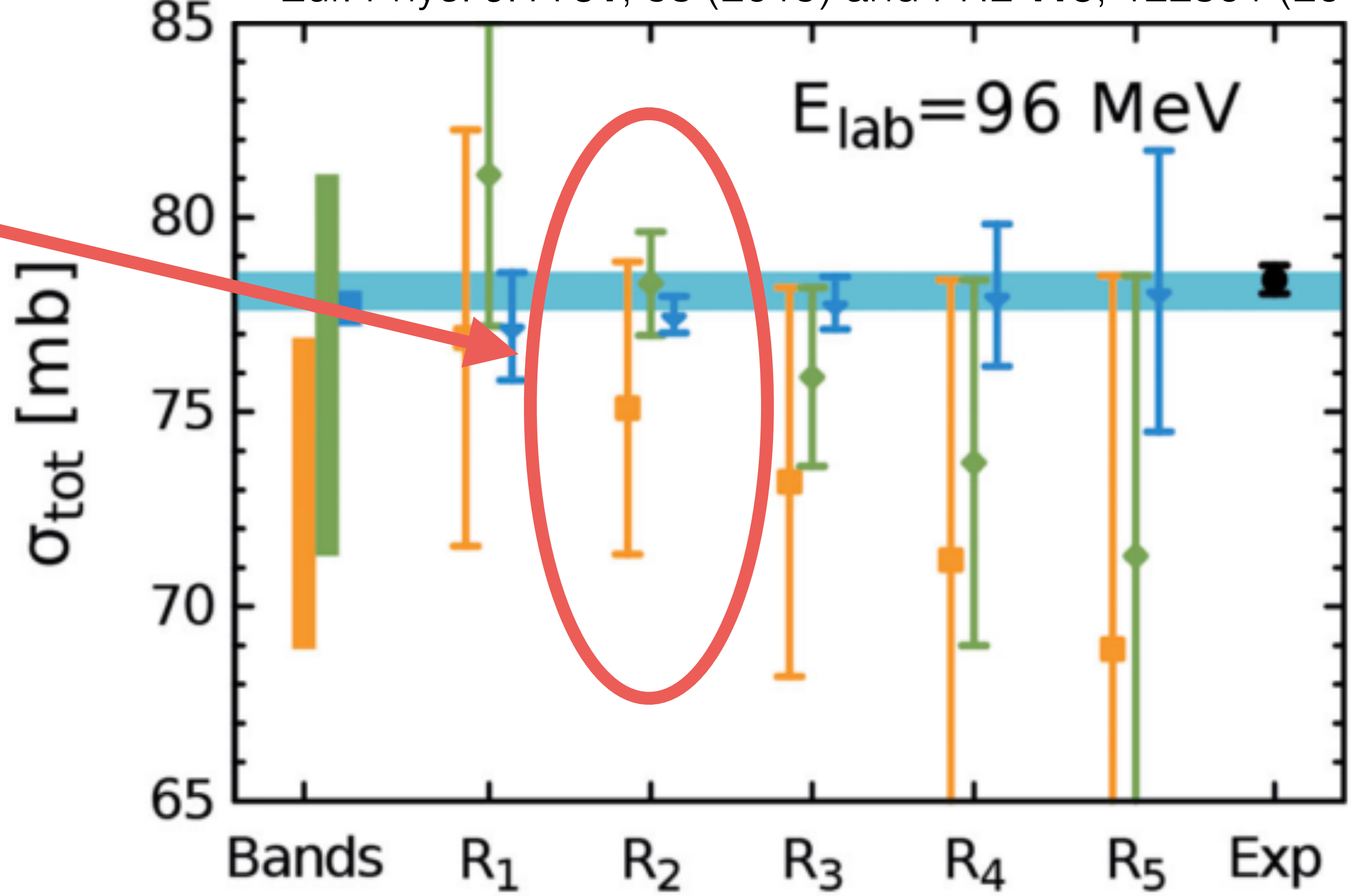
What is this?

Estimate next-order correction

$$c_{k+1} = \max\{c_0,\, c_1, \ldots, c_k\}$$

$$\Delta_k(x) \approx \eta_0\, c_{k+1}\, x^{k+1}$$

Epelbaum, *et al.*,
Eur. Phys. J. A **51**, 53 (2015) and PRL **115**, 122301 (2015)



Procedure for error estimates based on order-by-order convergence

More modern approaches: truncation uncertainty

$$\eta(x;k) = \eta_0 \sum_{n=0}^{k} c_n x^n + \Delta_k(x)$$

Given these order-by-order calculations…

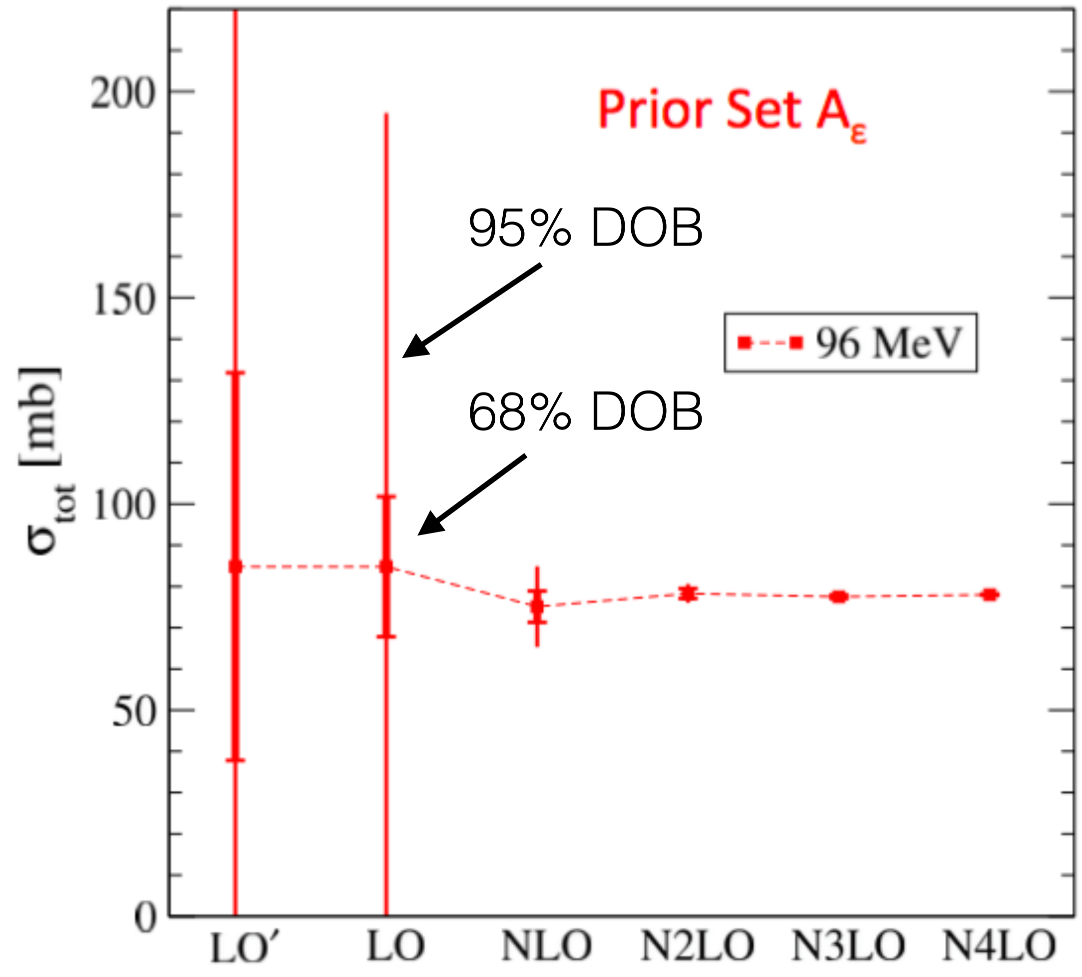
What is this?

Estimate next-order correction

$$c_{k+1} = \max\{c_0,\, c_1, \ldots, c_k\}$$

$$\Delta_k(x) \approx \eta_0\, c_{k+1}\, x^{k+1}$$

Furnstahl, Phillips, SW, Klco,
Phys. Rev. C 92 024005 (2015)



Bayesian derivation gives statistical interpretation and full pdf

See talk of H. Griesshammer from last week for applications to other EFTs!

Set up Bayesian approach

Goal: estimate coefficient posterior

$$\mathrm{pr}(\mathbf{a} | D, k, k_{\mathrm{max}}, I)$$

I: any other information

Vector of LECs
$\{a_0, a_1, \ldots a_k\}$

Data

k : truncation order
k_{max} : higher-order effects

Known prior information:

$\mathrm{pr}(\mathbf{a}|\bar{a})$ Naturalness prior

$\mathrm{pr}(\bar{a})$ Prior for naturalness parameter

Set up Bayesian approach

Goal: estimate coefficient posterior

"Integrate in" higher-order coefficients $\{a_{k+1}, \ldots a_{kmax}\}$

$$\mathrm{pr}(\mathbf{a}|D, k, k_{\max}) = \int d\mathbf{a}_{\mathrm{marg}}\, \mathrm{pr}(\mathbf{a}, \mathbf{a}_{\mathrm{marg}}|D, k, k_{\max})$$

Bayes theorem

$$\propto \int d\mathbf{a}_{\mathrm{marg}}\; \mathrm{pr}(D|\mathbf{a}, \mathbf{a}_{\mathrm{marg}}, k, k_{\max})\; \mathrm{pr}(\mathbf{a}, \mathbf{a}_{\mathrm{marg}}|k, k_{\max})$$

Set up Bayesian approach

Goal: estimate coefficient posterior

"Integrate in" higher-order coefficients $\{a_{k+1}, \ldots a_{kmax}\}$

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\max}) = \int d\mathbf{a}_{\mathrm{marg}}\,\mathrm{pr}(\mathbf{a},\mathbf{a}_{\mathrm{marg}}|D,k,k_{\max})$$

Bayes theorem

$$\propto \int d\mathbf{a}_{\mathrm{marg}}\;\mathrm{pr}(D|\mathbf{a},\mathbf{a}_{\mathrm{marg}},k,k_{\max})\;\mathrm{pr}(\mathbf{a},\mathbf{a}_{\mathrm{marg}}|k,k_{\max})$$

Integrate in naturalness parameter

$$\propto \int d\mathbf{a}_{\mathrm{marg}}\,d\bar{a}\;\mathrm{pr}(D|\mathbf{a},\mathbf{a}_{\mathrm{marg}},k,k_{\max})\;\mathrm{pr}(\mathbf{a},\mathbf{a}_{\mathrm{marg}}|\bar{a},k,k_{\max})\;\mathrm{pr}(\bar{a}|k,k_{\max})$$

Diagnostic framework

Set up

Guidance

Parameter estimation

Validation

Predictions

“Bayesian parameter estimation for effective field theories”
SW, Klco, Furnstahl, Phillips, Thapilaya
J.Phys. G **43**, 074001 (2016)

Diagnostic framework

- Set up
- Guidance
- Parameter estimation
- Validation
- Predictions

- Specify model to be fit: $V(x) \sim V_0 \sum_{n=0}^{k} a_n x^n$
- Prior information
  - naturalness, symmetries, etc.
  - functional form, hyperparameters

$$\mathrm{pr}(\mathbf{a}|\bar{a}) \sim e^{-\mathbf{a}^2/2\bar{a}^2}$$

$$\mathrm{pr}(\bar{a}) = \begin{cases} \frac{1}{\bar{a}\ln(\bar{a}_>/\bar{a}_<)} & \text{when} \quad \bar{a} \in [\bar{a}_<, \bar{a}_>] \\ 0 & \text{otherwise} \end{cases}$$

- Likelihood to be used

$$\mathrm{pr}(D|\mathbf{a}, k, k_{\max}) \sim e^{-\chi^2/2}$$

Diagnostic framework

- Set up
- Guidance
- Parameter estimation
- Validation
- Predictions

Real-life example:

$$V_{\text{contact}}(\vec{p},\vec{p}\,') = \widetilde{C}_{(1S0)} + C_{(1S0)}(p^2 + p'^2) + D^1_{(1S0)} p^2 p'^2 + D^2_{(1S0)}(p^4 + p'^4) + \ldots$$

$$\mathcal{C}_n = \text{scaled } \{\widetilde{C}_{(1S0)}, C_{(1S0)}, D^1_{(1S0)}, D^2_{(1S0)}, \ldots\}$$

$$\text{scaled} \equiv \text{correct } f_\pi,\ 4\pi,\ \Lambda_b \text{ factored out}$$

$$\text{pr}(\mathcal{C}_n|\bar{a}, I) \propto e^{-\mathcal{C}_n^2/2\bar{a}^2}$$

Added complication: we don't know Λ_b a priori

Diagnostic framework

Set up

Guidance

Parameter estimation

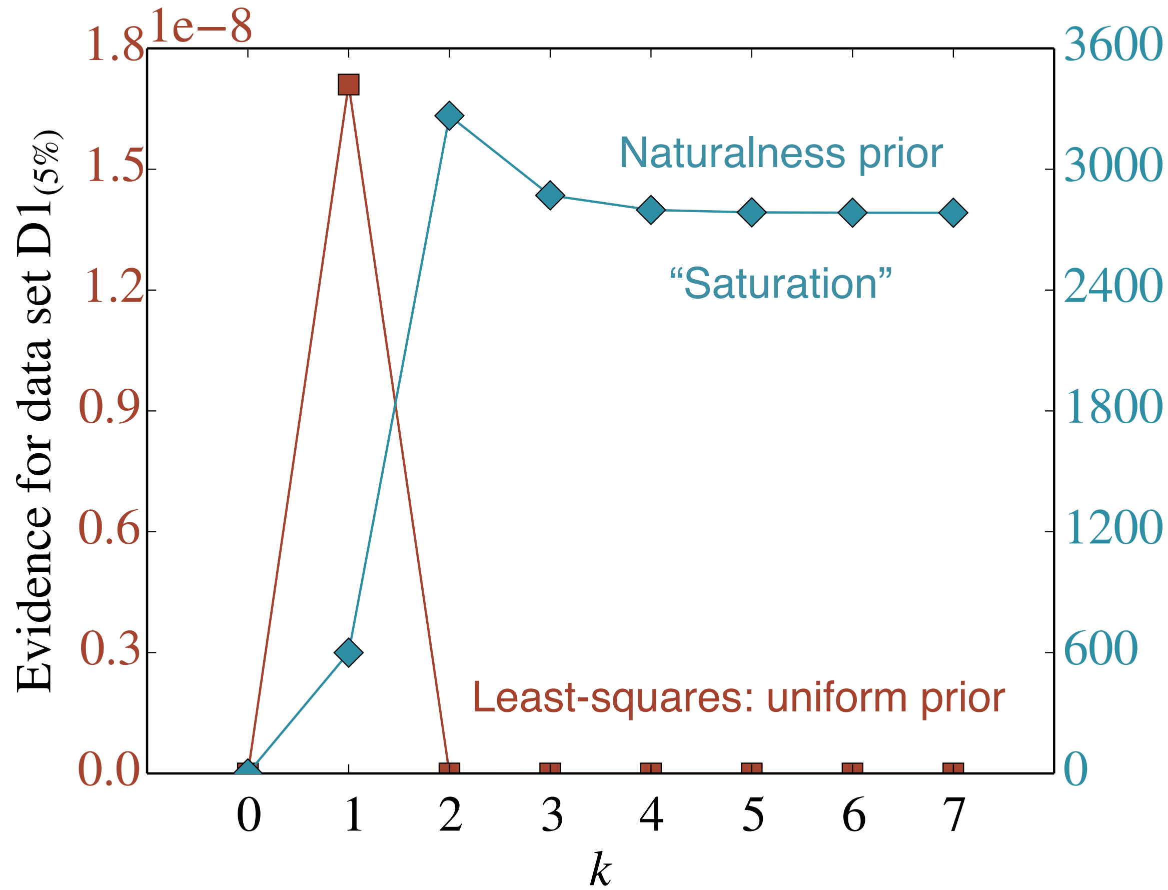
Validation

Predictions

Information from available data D

Calculate Bayes factors for different orders:

$$\frac{\mathrm{pr}(M_k|D)}{\mathrm{pr}(M_{k+1}|D)} = \frac{\int d\mathbf{a}_1 \; \mathrm{pr}(D|\mathbf{a}_1, M_k) \; \mathrm{pr}(\mathbf{a}_1|M_k)}{\int d\mathbf{a}_2 \; \mathrm{pr}(D|\mathbf{a}_2, M_{k+1}) \; \mathrm{pr}(\mathbf{a}_2|M_{k+1})}$$



Diagnostic framework

Set up

Guidance

Parameter estimation

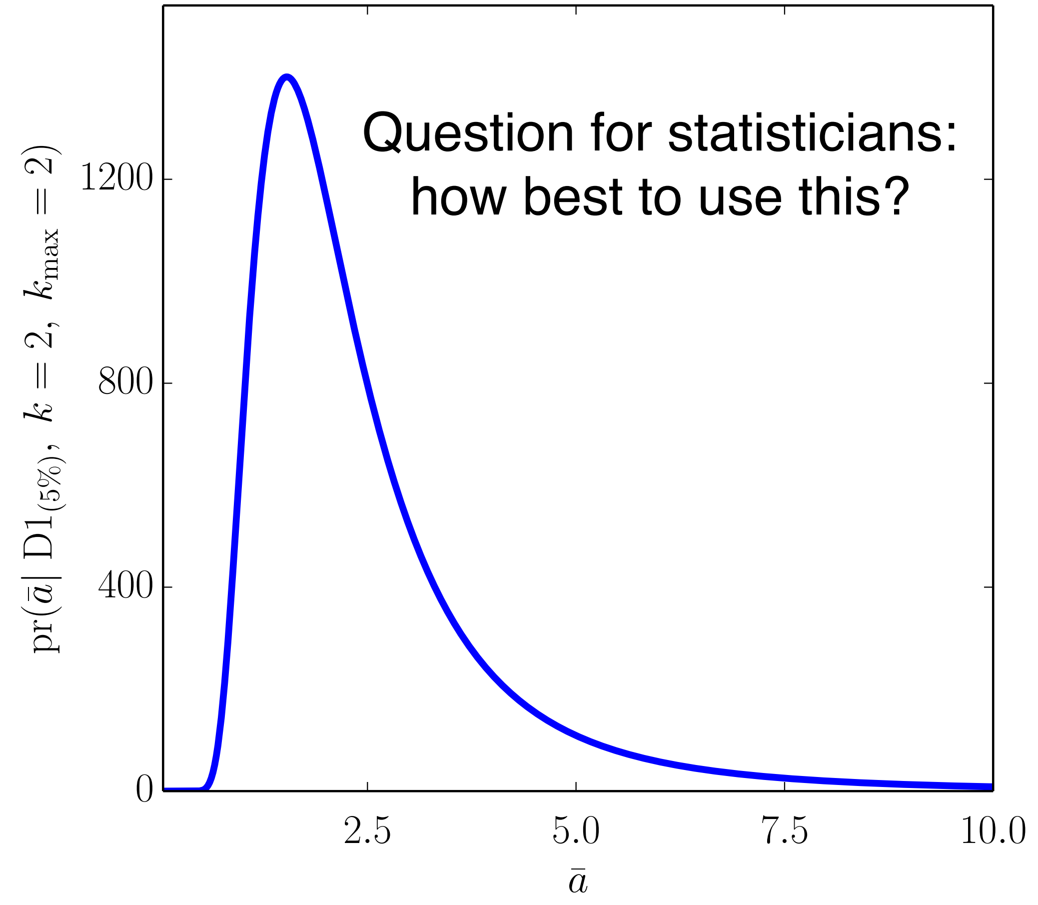
Validation

Predictions

Information from available data D

Calculate posterior for naturalness hyperparameter:

$$\mathrm{pr}(\bar{a}|D,I) = \int d\mathbf{a}\ \mathrm{pr}(D|\mathbf{a},I)\ \mathrm{pr}(\mathbf{a}|\bar{a},I)\ \mathrm{pr}(\bar{a}|I)$$



Diagnostic framework

- Set up
- Guidance
- Parameter estimation
- Validation
- Predictions

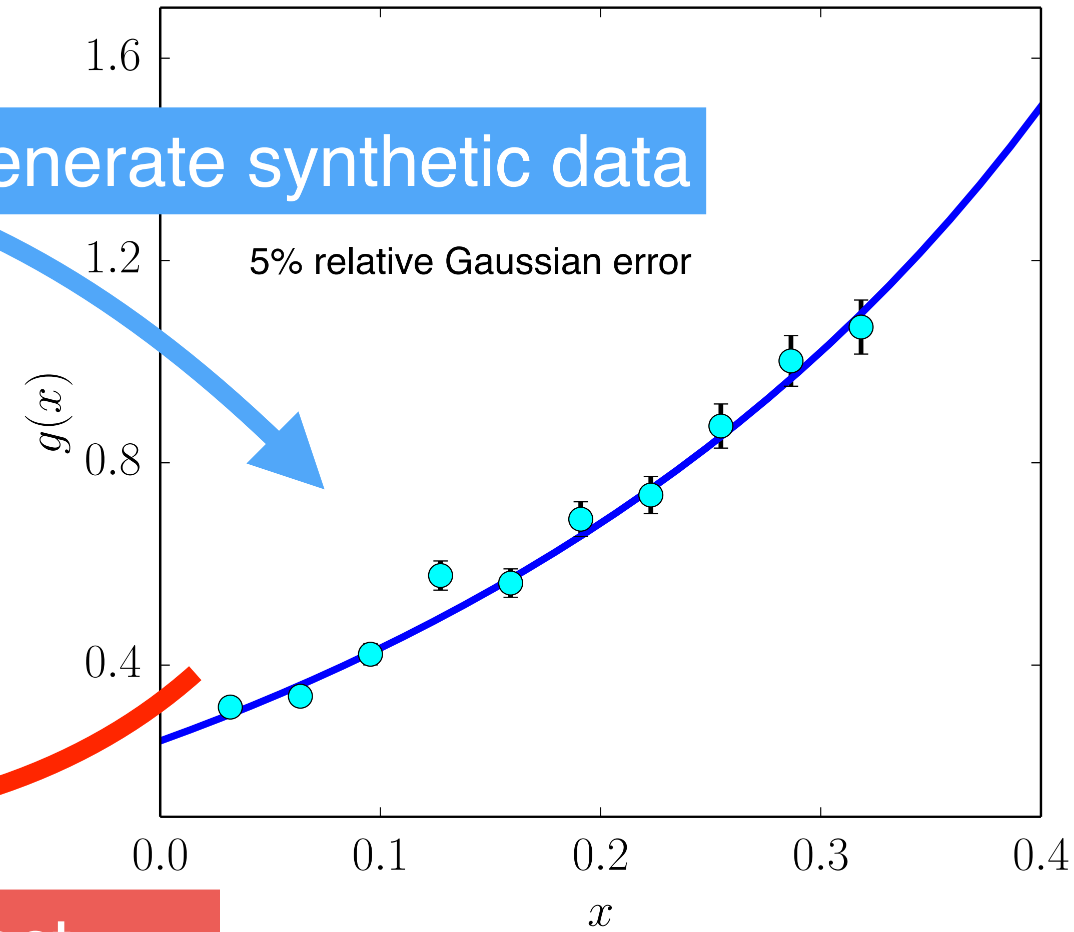
Detour to a model problem to explore generic features in EFT fitting…

Detour: parameter estimation in linear case

Consider a model observable [Schindler/Phillips (2009)]:

"Real World": $g(x) = \left(\frac{1}{2} + \tan\left(\frac{\pi x}{2}\right)\right)^2$

Generate synthetic data



Expansion: $g(x) = 0.25 + 1.57x + 2.47x^2 + \ldots$

Model "EFT": $g(x) = \sum_{n=0}^{k} a_n x^n$

Try to extract coefficients of model

Parameter estimation question:
Are estimates consistent with underlying expansion?

Detour: parameter estimation in linear case

Posterior | Likelihood | Prior

$$\mathrm{pr}(\mathbf{a}|D, k, k_{\max}) \propto \mathrm{pr}(D|\mathbf{a}, k, k_{\max}) \times \mathrm{pr}(\mathbf{a}|k, k_{\max})$$

Quantity of interest

Controlled by data
~ exp[-χ^2/2]

Prior knowledge

Detour: parameter estimation in linear case

Posterior — Likelihood — Prior

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\mathrm{max}}) \propto \mathrm{pr}(D|\mathbf{a},k,k_{\mathrm{max}}) \times \mathrm{pr}(\mathbf{a}|k,k_{\mathrm{max}})$$

Quantity of interest — Controlled by data ~ exp[-χ^2/2] — Prior knowledge

Compare results for different prior assumptions, take $k_{max} = k$

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times \mathbf{1}$$

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times e^{-\mathbf{a}^2/2\bar{a}^2}$$

Detour: parameter estimation in linear case

Posterior		Likelihood		Prior
$\mathrm{pr}(\mathbf{a}\|D,k,k_{\max})$	$\propto$	$\mathrm{pr}(D\|\mathbf{a},k,k_{\max})$	$\times$	$\mathrm{pr}(\mathbf{a}\|k,k_{\max})$
Quantity of interest		Controlled by data ~ $\exp[-\chi^2/2]$		Prior knowledge

Compare results for different prior assumptions, take $k_{max} = k$

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times \mathbf{1}$$

$k_{\max}$	χ^2/dof	a_0	a_1	a_2
true		0.25	1.57	2.47
1	2.2	0.20±0.01	2.6±0.1	
2	1.6	0.25±0.02	1.6±0.4	3.3±1.3
3	1.9	0.27±0.04	1.0±1	8.1±8.0
4	2.0	0.33±0.07	−1.9±3	45±30
5	1.4	0.57±0.3	−15±7	280±100

- Highly unstable for k > 2
- Errors large and unstable
- χ^2/dof looks good!

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times e^{-\mathbf{a}^2/2\bar{a}^2}$$

$k_{\max}$	a_0	a_1	a_2
true	0.25	1.57	2.47
1	0.20±0.01	2.6±0.1	
2	0.25±0.02	1.6±0.4	3.1±1
3	0.25±0.02	1.7±0.5	3.0±2
4	0.25±0.02	1.7±0.5	3.0±2
5	0.25±0.02	1.7±0.5	3.0±2

- Stable with increasing k
- Errors are reasonable
- "Saturates" at k = 2

Detour: parameter estimation in linear case

Posterior — Likelihood — Prior

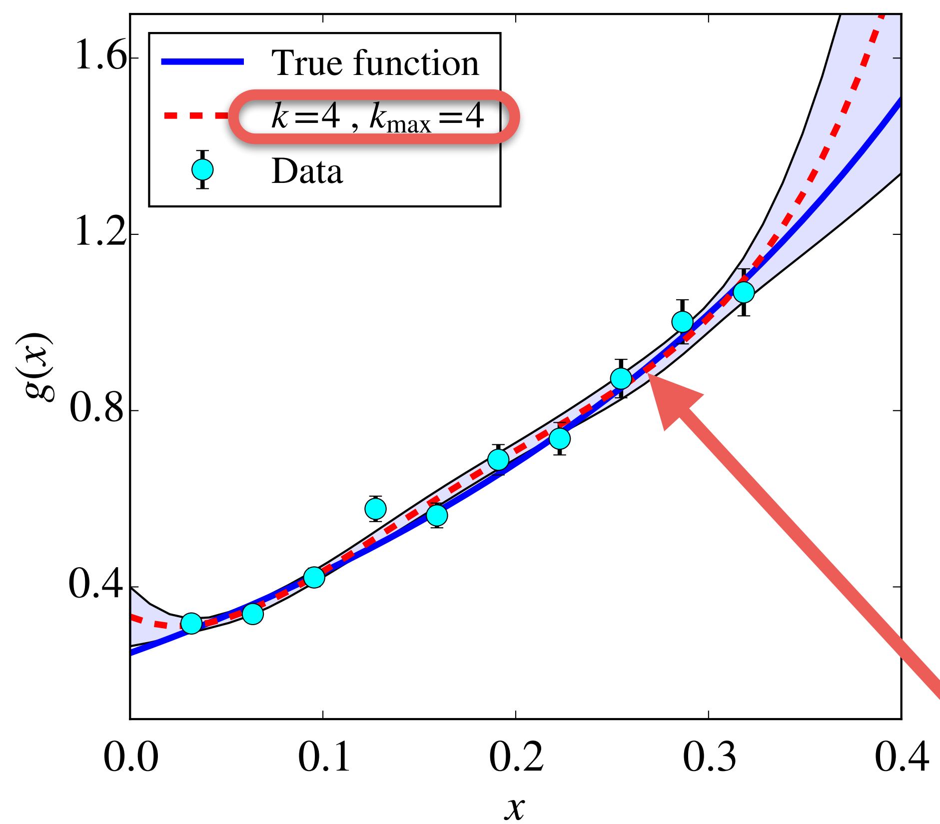
$$\mathrm{pr}(\mathbf{a}|D,k,k_{\max}) \propto \mathrm{pr}(D|\mathbf{a},k,k_{\max}) \times \mathrm{pr}(\mathbf{a}|k,k_{\max})$$

Quantity of interest

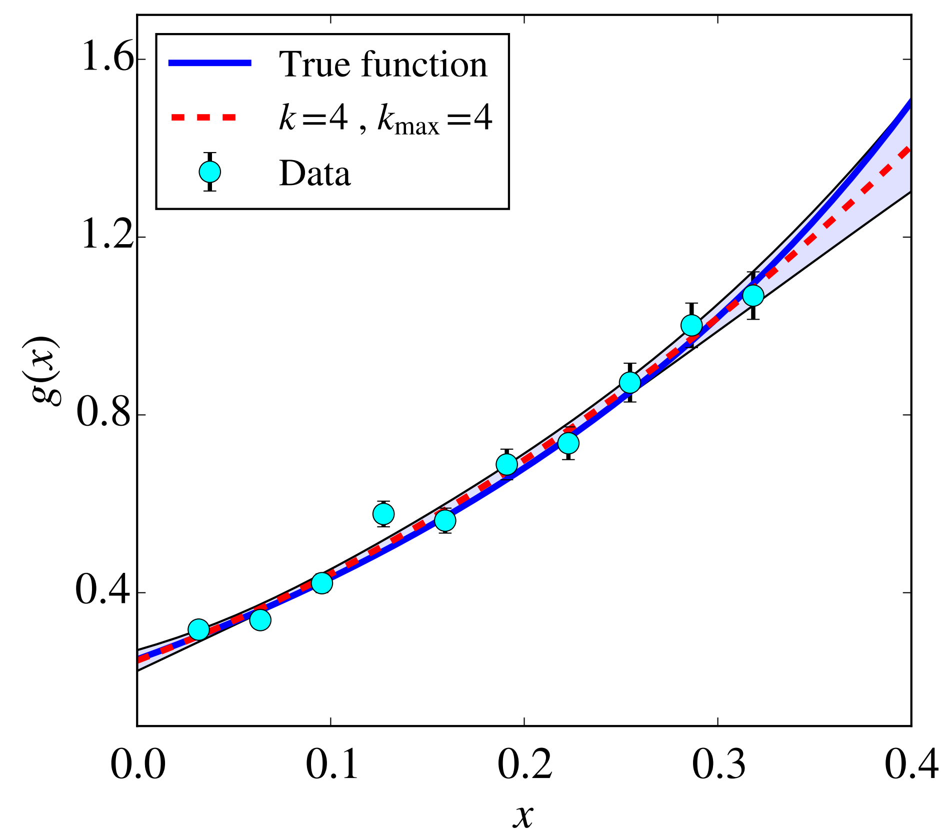
Controlled by data ~ exp[-χ^2/2]

Prior knowledge

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times \mathbf{1}$$



$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times e^{-\mathbf{a}^2/2\bar{a}^2}$$



"Overfitting" - fine-tuned to reproduce data in fit range

Detour: parameter estimation in linear case

Posterior $\quad$ Likelihood $\quad$ Prior

$$\mathrm{pr}(\mathbf{a}|D,k,k_{\max}) \propto \mathrm{pr}(D|\mathbf{a},k,k_{\max}) \times \mathrm{pr}(\mathbf{a}|k,k_{\max})$$

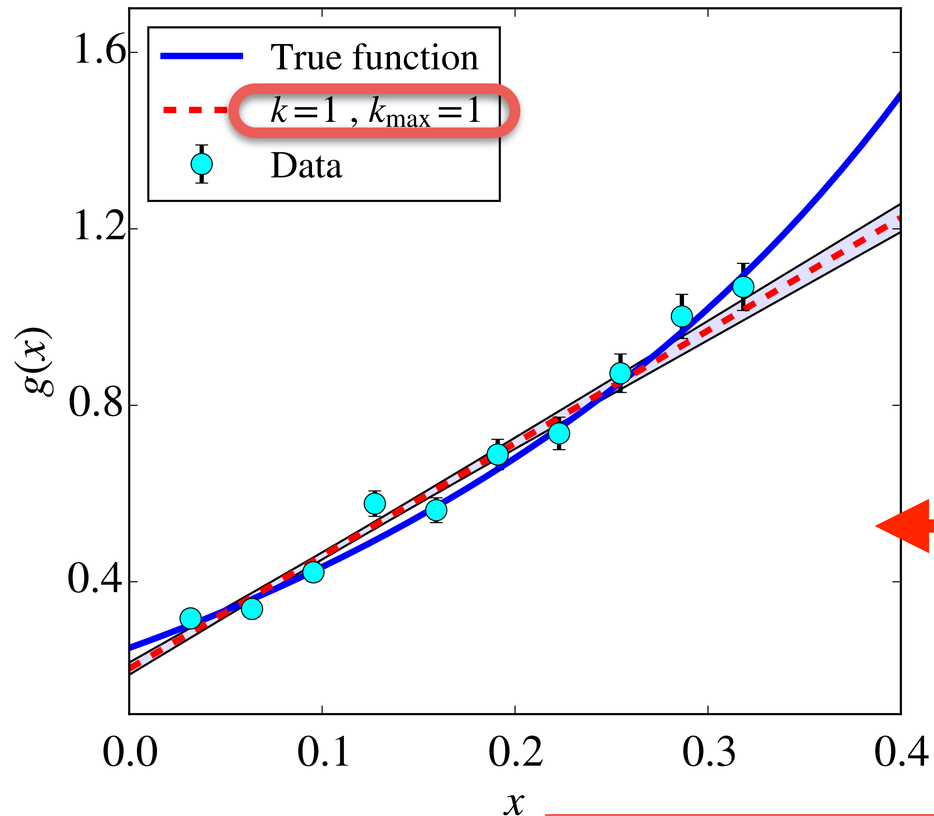
Quantity of interest

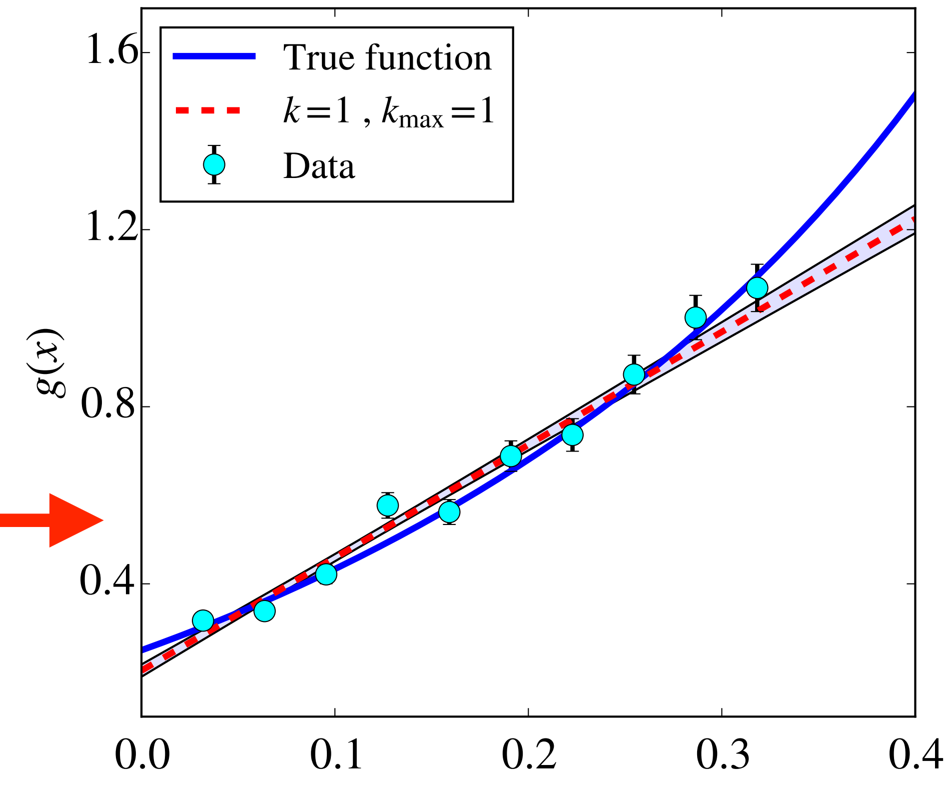
Controlled by data
~ exp[-χ^2/2]

Prior knowledge

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times \mathbf{1}$$

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times e^{-\mathbf{a}^2/2\bar{a}^2}$$



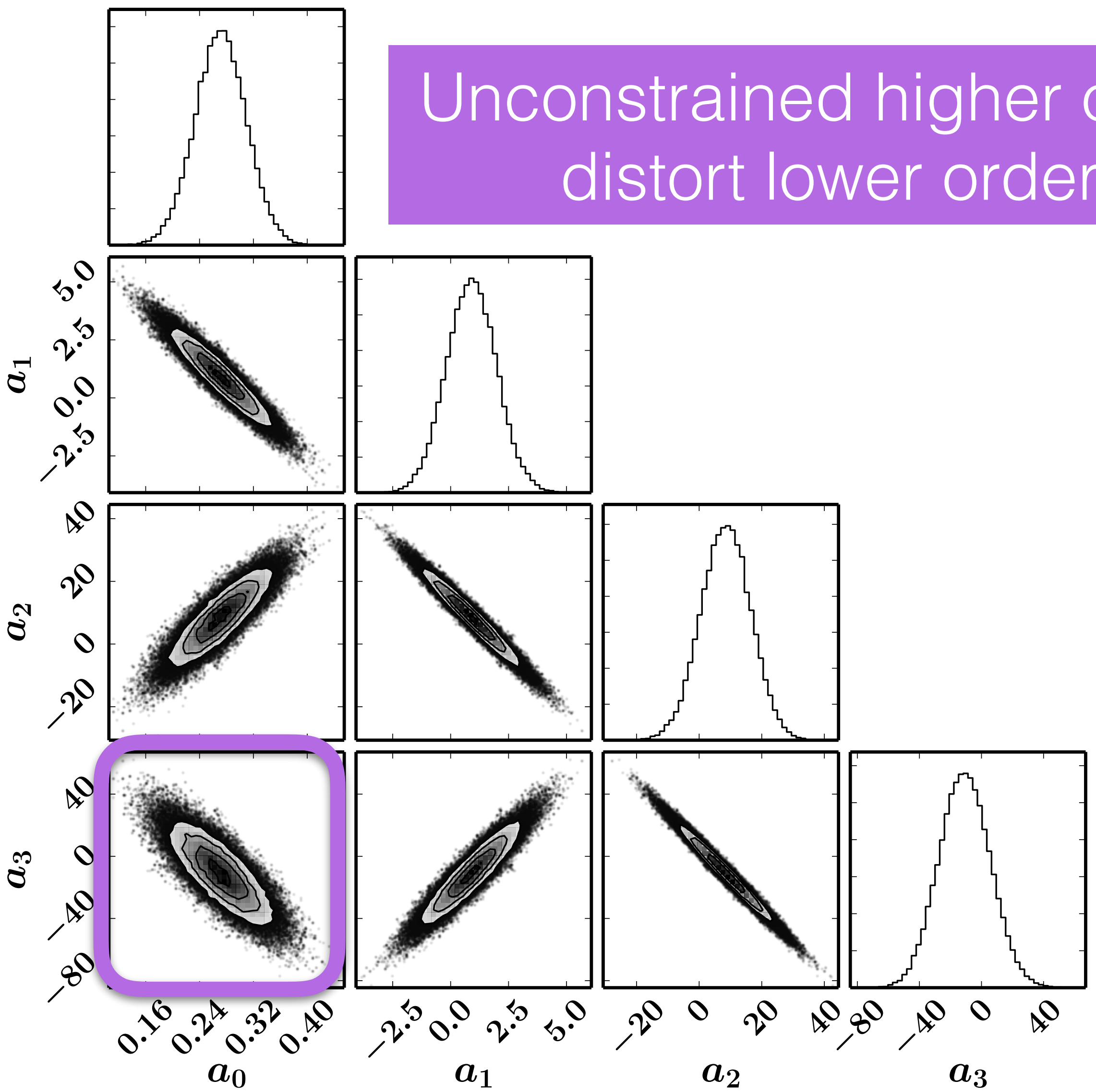


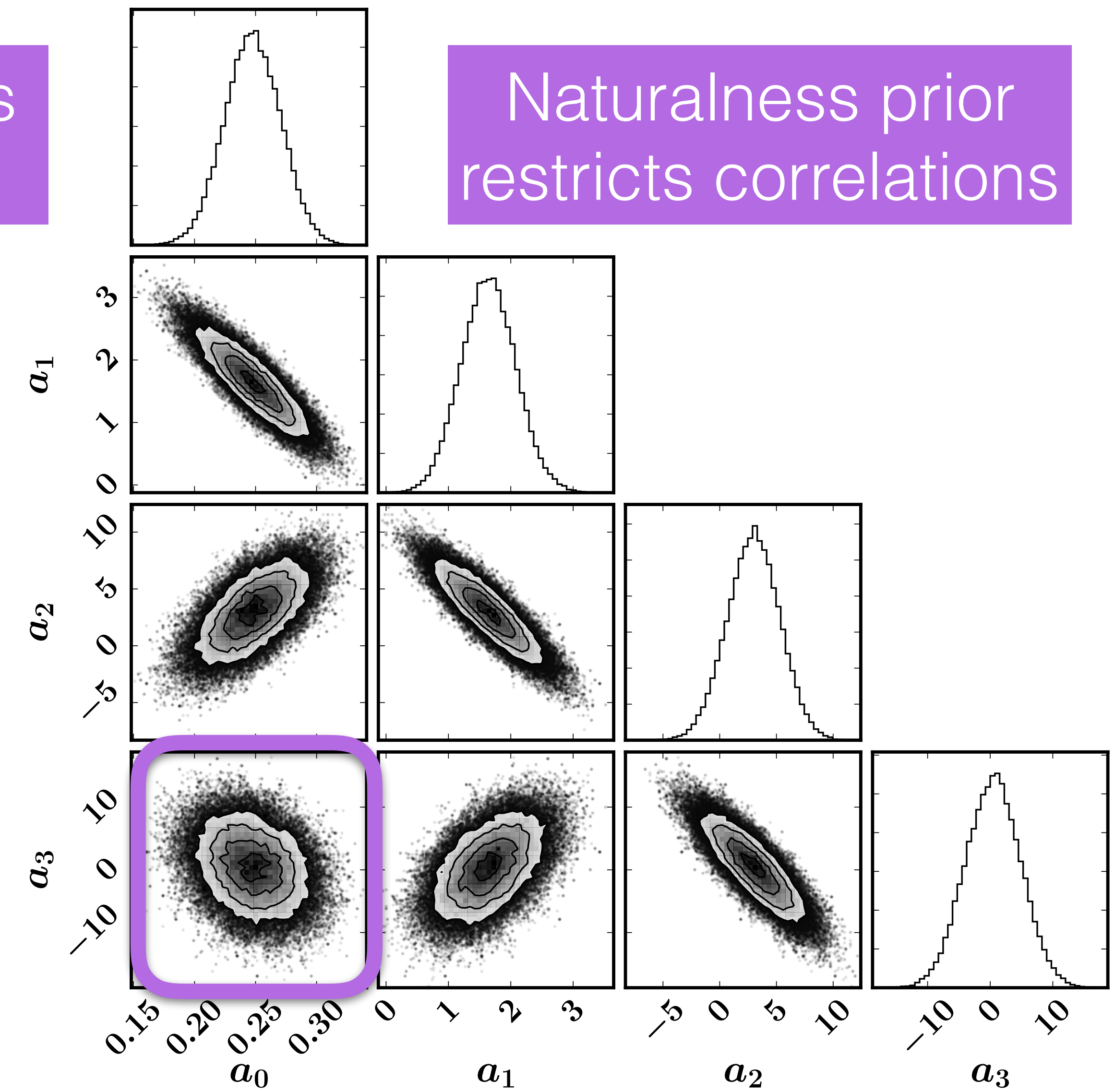
"Underfitting" - model too simplistic for data
Naturalness prior makes no difference

Detour: parameter estimation in linear case

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times \mathbf{1}$$

$$\mathrm{pr}(\mathbf{a}|D,k) \propto e^{-\chi^2/2} \times e^{-\mathbf{a}^2/2\bar{a}^2}$$



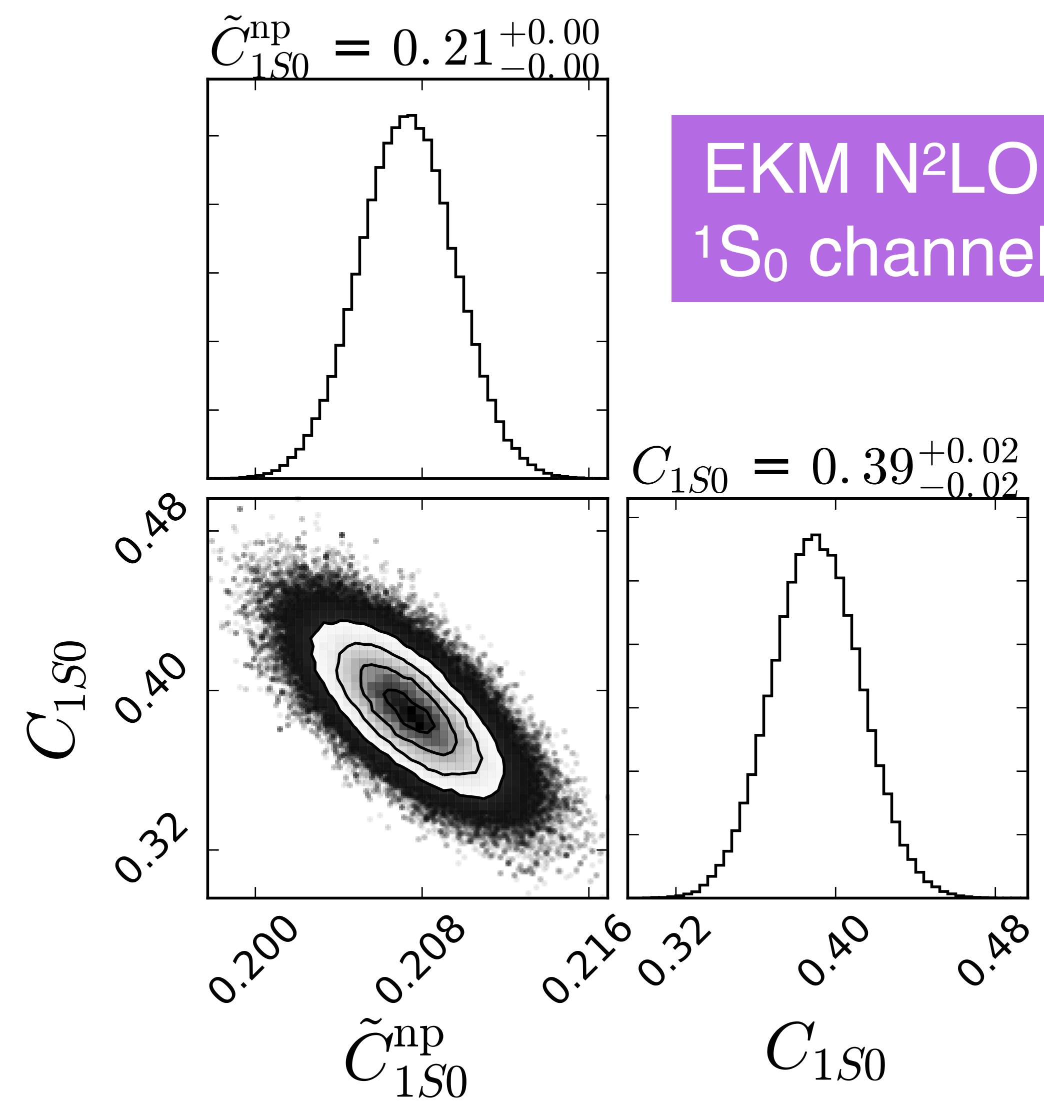


Diagnostic framework

- Set up
- Guidance
- **Parameter estimation**
- Validation
- Predictions

What about real life?

Estimate posterior and associated diagnostics

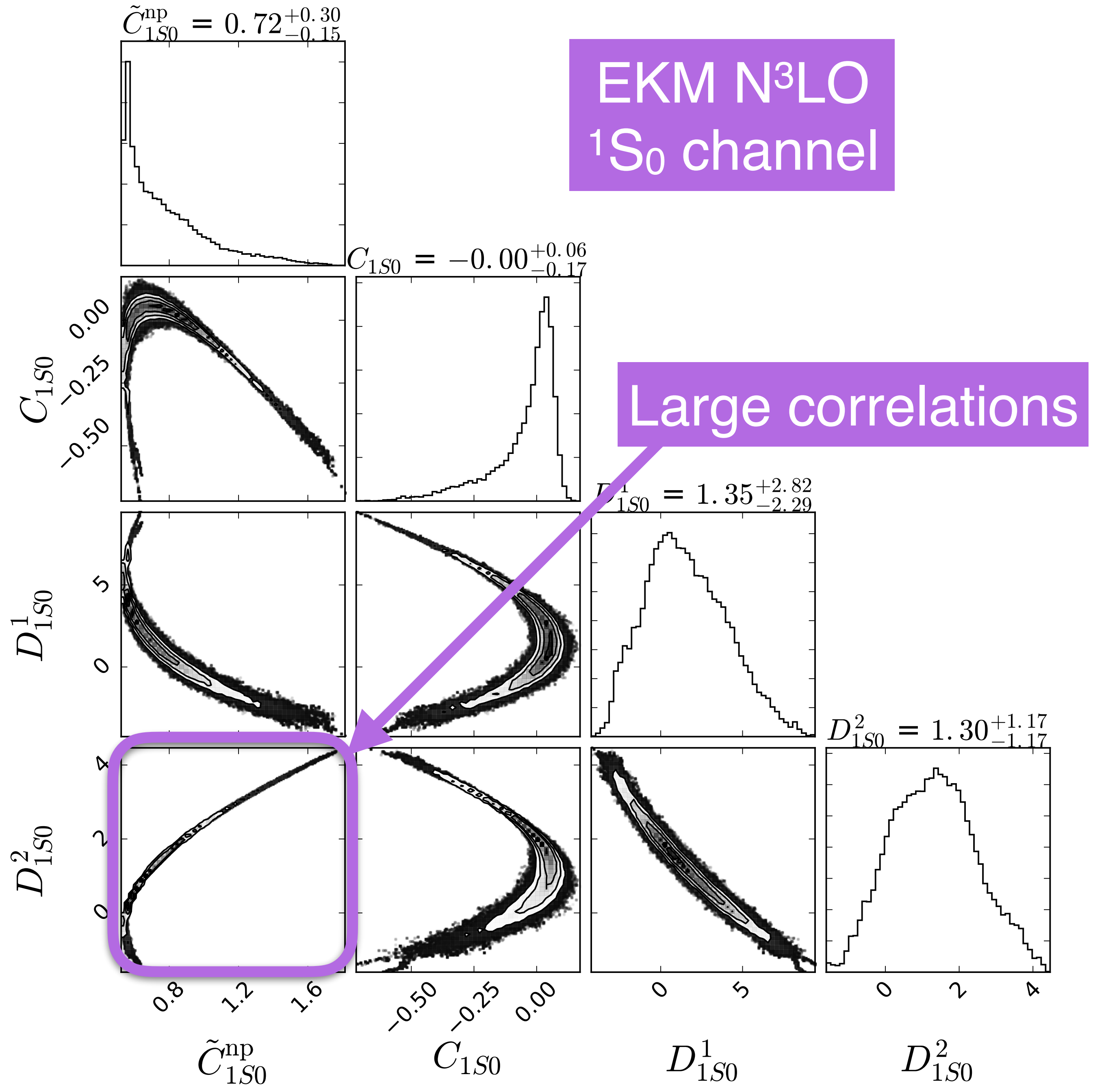


EKM N^2LO
$^1S_0$ channel

Data used: synthetic phase shifts at N^2LO

Diagnostic framework

- Set up
- Guidance
- Parameter estimation
- Validation
- Predictions



Data used: synthetic phase shifts at N^{3}LO

Why are the correlations so large?

N^3LO s-wave coefficients on-shell

- E.g., single linear combination of $D^1_{(1S0)}$ and $D^2_{(1S0)}$:

$$D^1_{(1S0)} p^2 p'^2 + D^2_{(1S0)} (p^4 + p'^4) = \frac{1}{4}\left(D^1_{(1S0)} + 2D^2_{(1S0)}\right)(p^2 + p'^2)^2 - \frac{1}{4}\left(D^1_{(1S0)} - 2D^2_{(1S0)}\right)(p^2 - p'^2)^2$$

N^3LO NN contacts: p^4

- All s-wave LECs from EKM fits to Nijmegen phase shifts:

Combination	$^1S_0$ N^{3}LO	$^1S_0$ N^{4}LO	$^3S_1$ N^{3}LO	$^3S_1$ N^{4}LO
D^1	−1.59	−5.50	−7.13	−6.18
D^2	2.65	4.18	5.64	4.70
$\frac{1}{4}(D^1 + 2D^2)$	0.93	0.71	1.04	0.80
$\frac{1}{4}(-D^1 + 2D^2)$	1.72	3.47	4.60	3.89

Why are the correlations so large?

N^{3}LO s-wave coefficients on-shell

- E.g., single linear combination of $D^1_{(1S0)}$ and $D^2_{(1S0)}$:

$$D^1_{(1S0)}p^2p'^2 + D^2_{(1S0)}(p^4 + p'^4) = \frac{1}{4}\left(D^1_{(1S0)} + 2D^2_{(1S0)}\right)(p^2 + p'^2)^2 - \frac{1}{4}\left(D^1_{(1S0)} - 2D^2_{(1S0)}\right)(p^2 - p'^2)^2$$

N^{3}LO NN contacts: p^4

- All s-wave LECs from EKM fits to Nijmegen phase shifts:

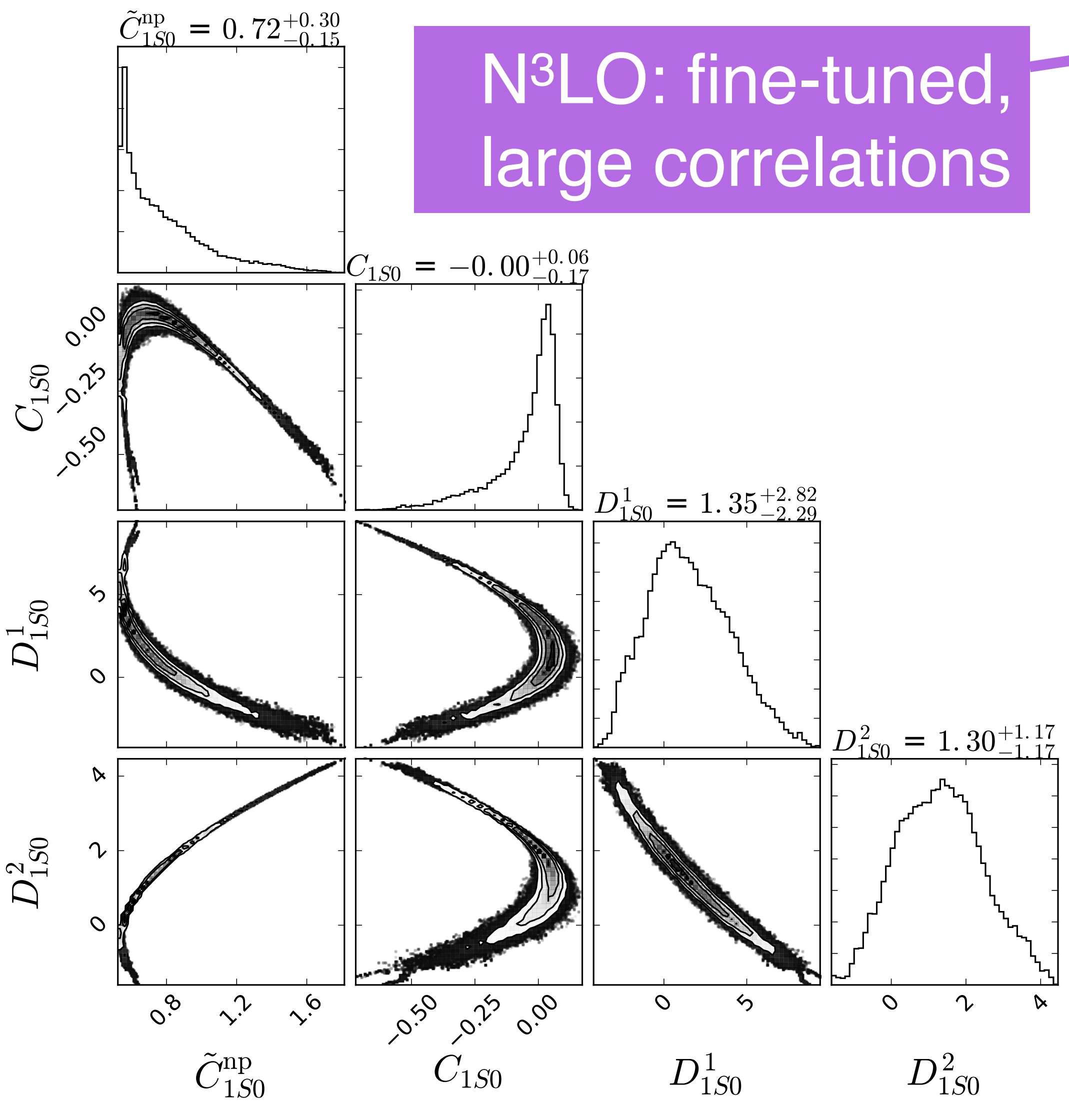
Combination	$^1S_0$ N^{3}LO	$^1S_0$ N^{4}LO	$^3S_1$ N^{3}LO	$^3S_1$ N^{4}LO
D^1	−1.59	−5.50	−7.13	−6.18
D^2	2.65	4.18	5.64	4.70
$\frac{1}{4}(D^1 + 2D^2)$	0.93	0.71	1.04	0.80
$\frac{1}{4}(-D^1 + 2D^2)$	1.72	3.47	4.60	3.89

enhancement of x4 could indicate overfitting: explore further!

Why are the correlations so large?

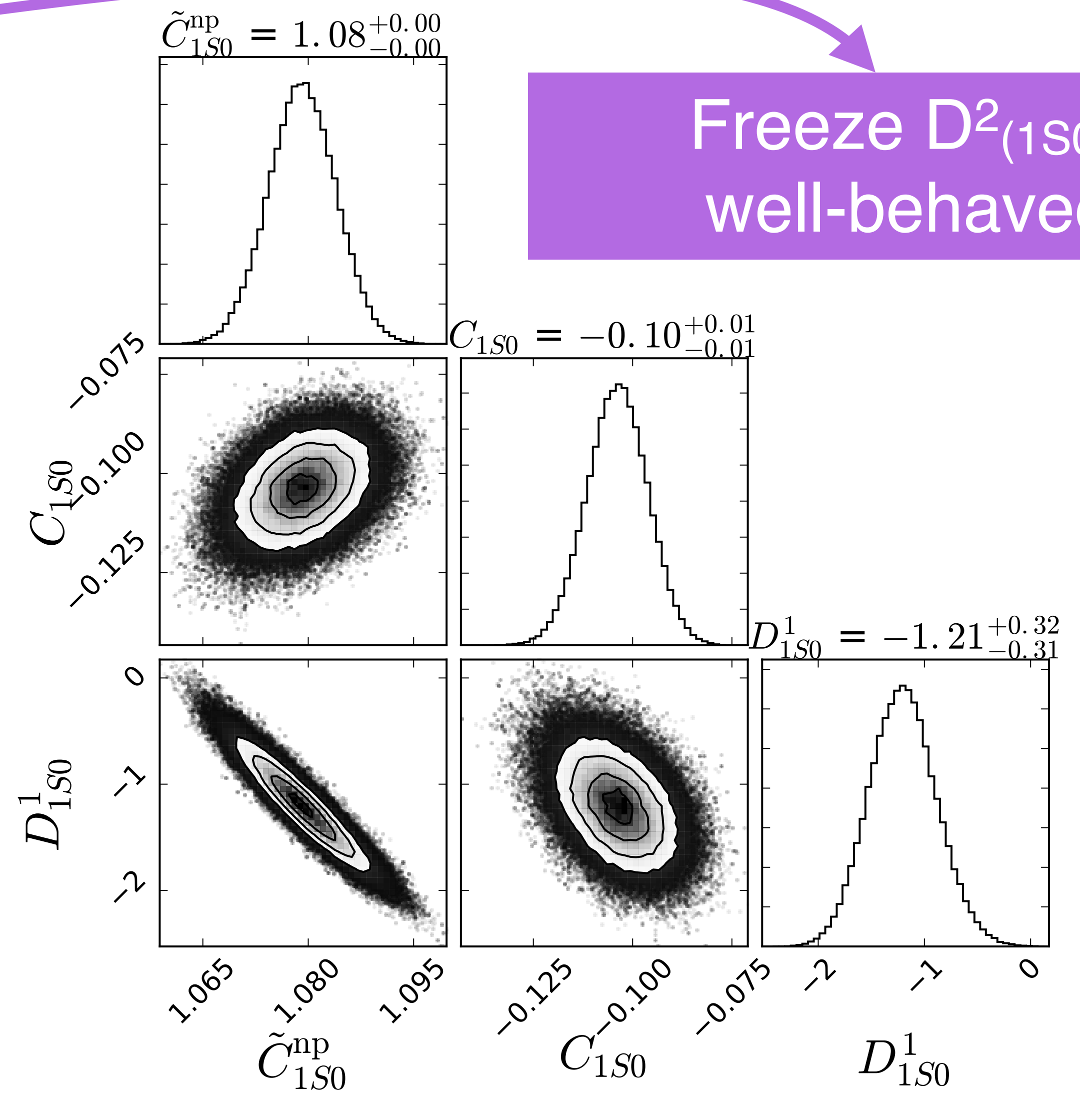
N³LO s-wave coefficients on-shell

- Is there overfitting once we go to N³LO in the s-waves?



N³LO: fine-tuned, large correlations

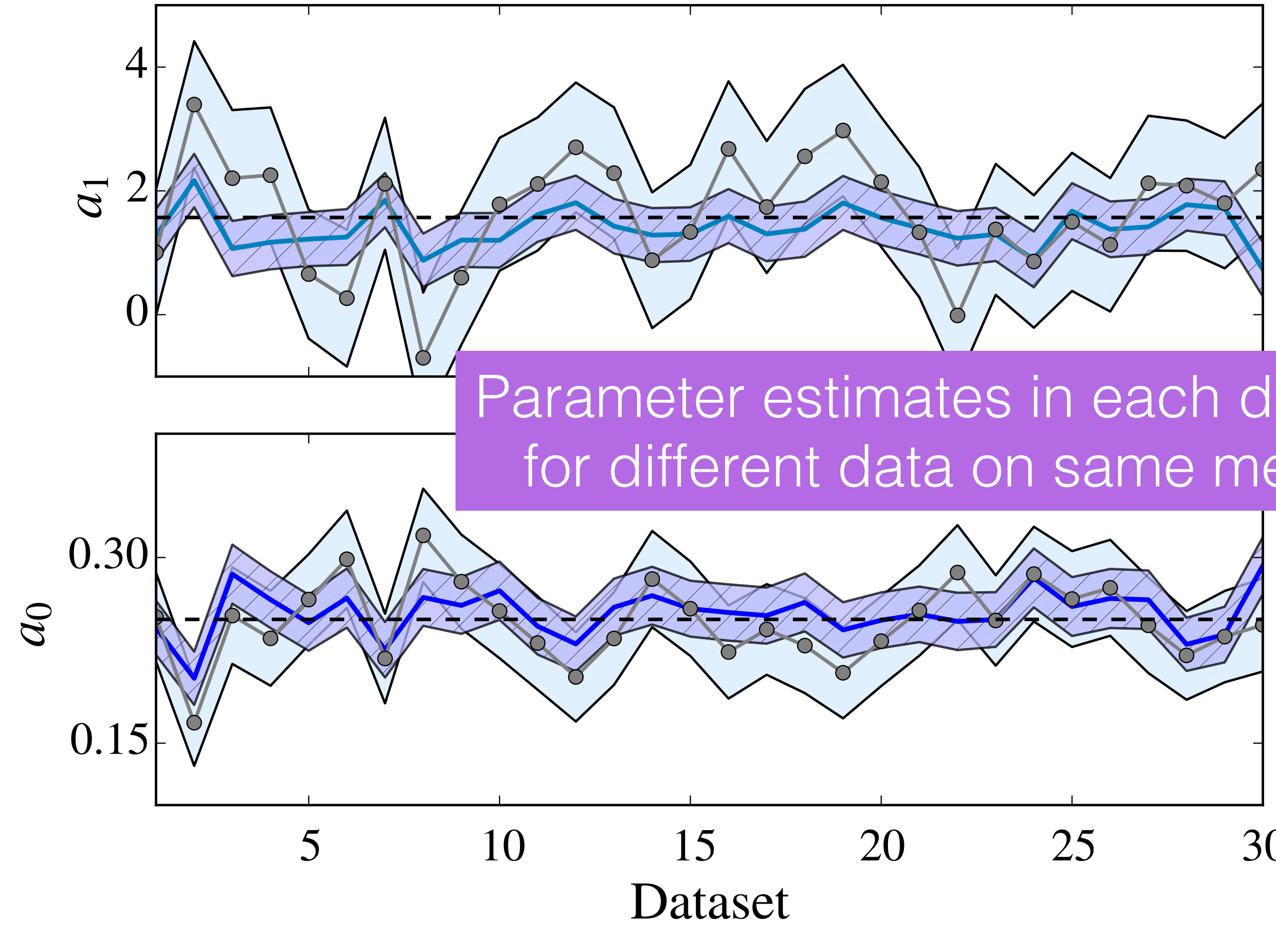
Freeze $D^2_{(1S0)}$, well-behaved



Diagnostic framework

- Set up
- Guidance
- Parameter estimation
- Validation
- Predictions

- Fluctuations between subdivided data sets
- Accumulate data, likelihood-prior competition
- Check power-law behavior (Lepage plots)



Parameter estimates in each dimension for different data on same mesh in x

Diagnostic framework

Set up

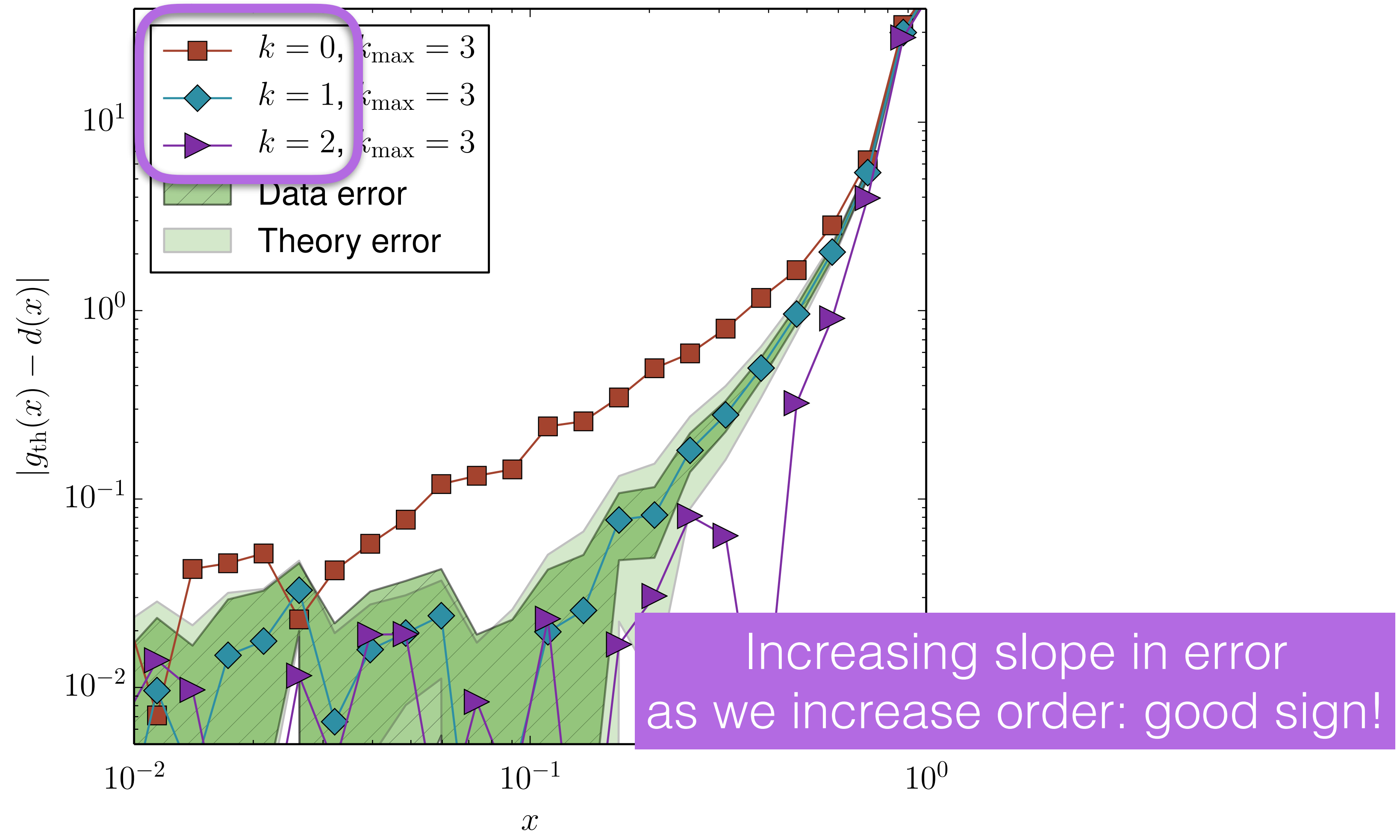
Guidance

Parameter estimation

Validation

Predictions

- Fluctuations between subdivided data sets
- Accumulate data, likelihood-prior competition
- Check power-law behavior (Lepage plots)



Diagnostic framework

Set up

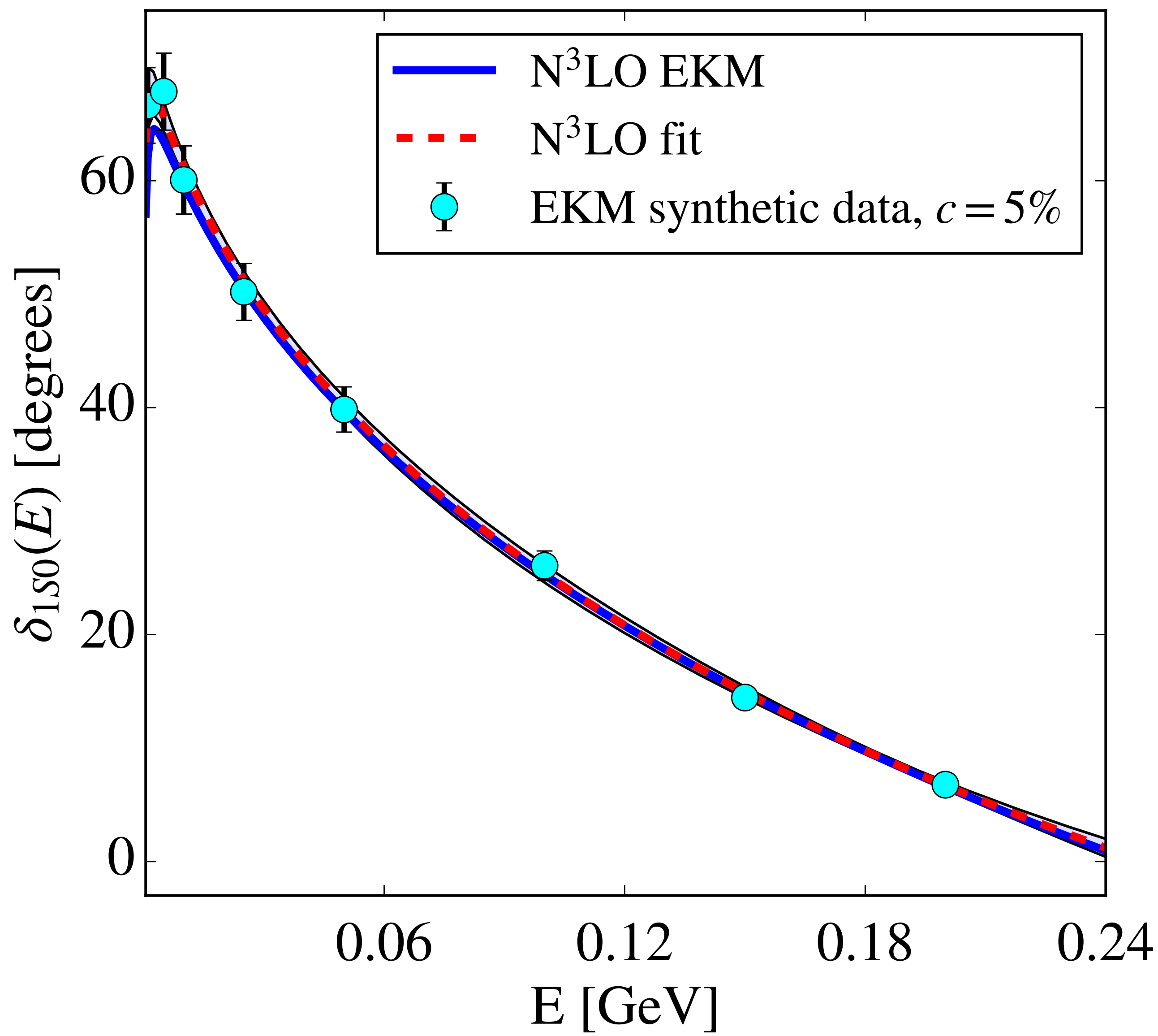
Guidance

Parameter estimation

Validation

Predictions

Propagate errors to predictions!



Also can calculate truncation uncertainty

Summary

- Recently completed diagnostic framework for generic EFTs
  - S. Wesolowski et al., "Bayesian parameter estimation for effective field theories", J. Phys G 43, 074001 (2016)
- Parameter estimation for EKM NN interactions in progress
  - Explore fitting to phase shifts vs. full cross sections
  - Nearing computational wall for evidence integrals
  - Exploring possible redundancy in s-waves

Next steps

- Find Bayesian interpretations for current EFT fitting methods
  - Naive least-squares is never used by practitioners- what are assumptions/priors?
- Parallelization of MCMC calculations
- Nested sampling?
- Mixture models as alternative to evidence

Questions

- Alternatives to Bayes factors?
- Relationship to common practices in EFTs

$$\chi^2 = \sum_{i=0}^{N} \left(\frac{\eta(x_i; \mathbf{a}, k) - d_i}{\sigma_i^2 + (C_x x^{k+1})^2} \right)^2$$

Fit C_x so $\chi 2$/dof = 1: Birge factor?

- What if form of $V(x; \mathbf{a}, k) = V_0 \sum_{n=0}^{k} a_n x^n$ is not known at higher k?
- How could we determine Λ_b?
- Model selection in pionless EFT sandbox.
- Where would GP emulators be useful?
- Should we be orthogonalizing the posterior?

Backup: Prototype EFT

$$V(x) = V_0 \sum_{n=0}^{k} \mathcal{A}_n(x)\ x^n \quad \text{with} \quad x = \frac{p}{\Lambda_b}$$

Natural-sized coefficients

Small expansion parameter x

$$\mathcal{A}_n(x) = a_n(\mu) + f_n(x, \mu)$$

$a_n, \ f_n \sim 1$ when $\mu \sim \Lambda_b$

$f_n(x, \mu)$ encodes IR physics at order n

Backup: Marginalize higher-order corrections

Marginalization over higher-order effects

$$k = 0,\ \ k_{\mathrm{max}} = 3$$

$$g(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

$$\mathrm{pr}(\mathbf{a}|D,I) = \int d\mathbf{a}_{\mathrm{marg}}\ \mathrm{pr}(\mathbf{a},\mathbf{a}_{\mathrm{marg}}|D,I)$$

$$\mathrm{pr}(\mathbf{a}|D,I) \propto \int d\mathbf{a}_{\mathrm{marg}}\ \mathrm{pr}(D|\mathbf{a},\mathbf{a}_{\mathrm{marg}},I)\mathrm{pr}(\mathbf{a},\mathbf{a}_{\mathrm{marg}}|I)$$

Correlated higher-order errors, see [arXiv:hep-ph/0101051]
and [arXiv:1407.0657]