SAND2024-01951C

ALLEVIATING THE COST OF BAYESIAN CALIBRATION WITH EMBEDDED MODEL ERROR

(USING GAUSSIAN PROCESS SURROGATES AND MULTILEVEL MCMC)

Sandia National Laboratories

Arun Hegde
(ahegde@sandia.gov)
Cosmin Safta
Habib Najm

The Ohio State University

Elan Weiss
Wolfgang Windl

OVERVIEW

- Motivating example: Bayesian calibration for *interatomic potentials*

- Review two model error frameworks
  - Additive model error
  - **Embedded model error**

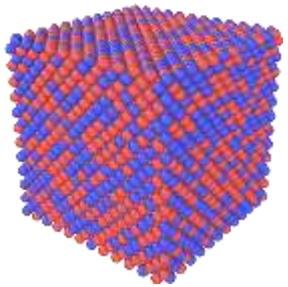
- Computational strategy for Gaussian embedded model error
  - **Likelihood approximation**
  - Gaussian Process (GP) surrogates with Gaussian inputs
    - **Sparse GPs** & **multilevel MCMC** for inference

- Case study: calibrating an EAM potential with embedded model error

MOTIVATING EXAMPLE: CALIBRATING INTERATOMIC POTENTIALS

Design of advanced energy storage devices often requires materials simulations at the *atomistic* level.

For practical scale systems, density functional theory (DFT) based simulations are untenable, motivating the use of empirical methods involving **interatomic potentials**.

$f_\theta($  $) \longrightarrow$ **energy of the atomic configuration**

Given their empirical nature, **effective calibration of such models** requires careful consideration of the **model error***.

*Hegde, A., Weiss, E., Windl, W., Najm, H., & Safta, C. "Bayesian calibration of interatomic potentials for binary alloys." *Computational Materials Science* 214 (2022): 111660.

MOTIVATING EXAMPLE: CALIBRATING INTERATOMIC POTENTIALS

$\theta \in \mathbb{R}^5$ → **Interatomic potential model** → LAMMPS → $\mathcal{M}(x_i;\theta)$

molecular dynamics simulation software*

5-parameter RAMPAGE* Embedded Atom Method (EAM) potential.

$$E_i = \frac{1}{2}\sum_{j\neq i} V_{\alpha\beta}(r_{ij}) + F_\alpha\left(\sum_{j\neq i}\rho_{\alpha\beta}(r_{ij})\right)$$

$$V_{AB}(r) = D\left(e^{-2\alpha(r-r_{eq})} - 2e^{-\alpha(r-r_{eq})}\right)$$

$$\rho_{BA}(r) = r^6\left(e^{-S_A r} + 2^9 e^{-2S_A r}\right)$$

$$\rho_{AB}(r) = r^6\left(e^{-S_B r} + 2^9 e^{-2S_B r}\right)$$

Model output: **102** quantities of interest (QoIs) total.

17 compositions ranging from 3% Au to 97% Au

For each composition, 6 physical properties:

- lattice parameter
- mixing enthalpy
- elastic constants C11, C12, C44
- bulk modulus

*L. Ward, A. Agrawal, K.M. Flores, and W. Windl. Rapid production of accurate embedded-atom method potentials for metal alloys. (2012). arXiv:cond-mat.mtrl-sci/1209.0619

Indexing the 102 different **QOIs:**

finite composition space $x \in \mathcal{X}$

physical properties $i \in \{\mathrm{lat}, \mathrm{mix}, \mathrm{C}_{11}, \mathrm{C}_{12}, \mathrm{C}_{44}, \mathrm{bulk}\}$

BAYESIAN CALIBRATION WITH ADDITIVE MODEL ERROR

$$\boxed{y(x_i)} = \boxed{\mathcal{M}(x_i;\theta)} + \boxed{\epsilon(x_i)}$$

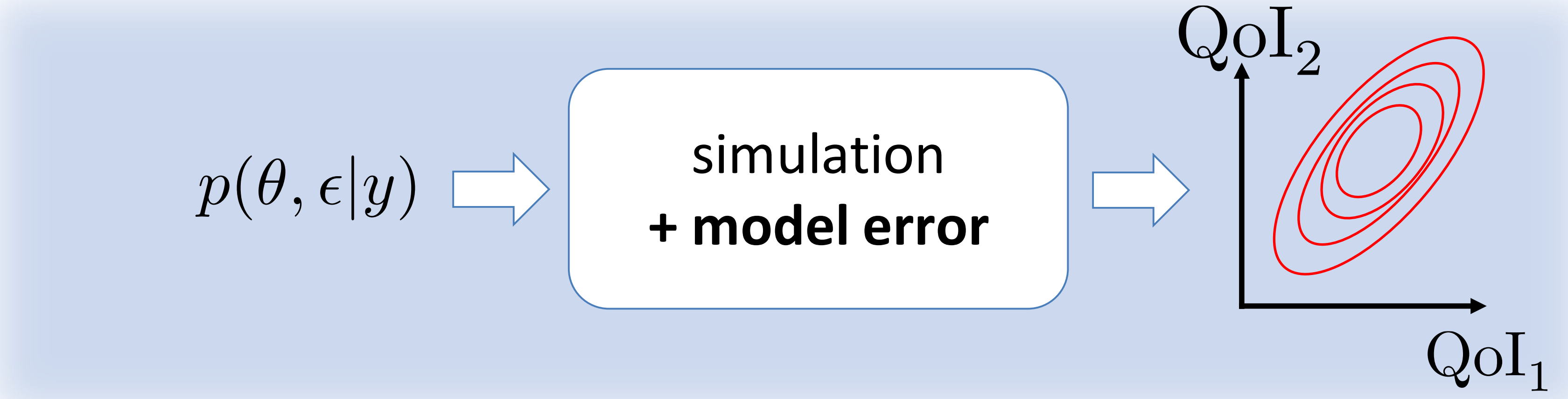
DFT data   simulation   model error

$$\boxed{p(\theta,\epsilon|y) = \frac{p(y|\theta,\epsilon)p(\theta,\epsilon)}{p(y)}}$$

Principal objective is to jointly infer the unknown parameters and the model error given DFT data.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464.

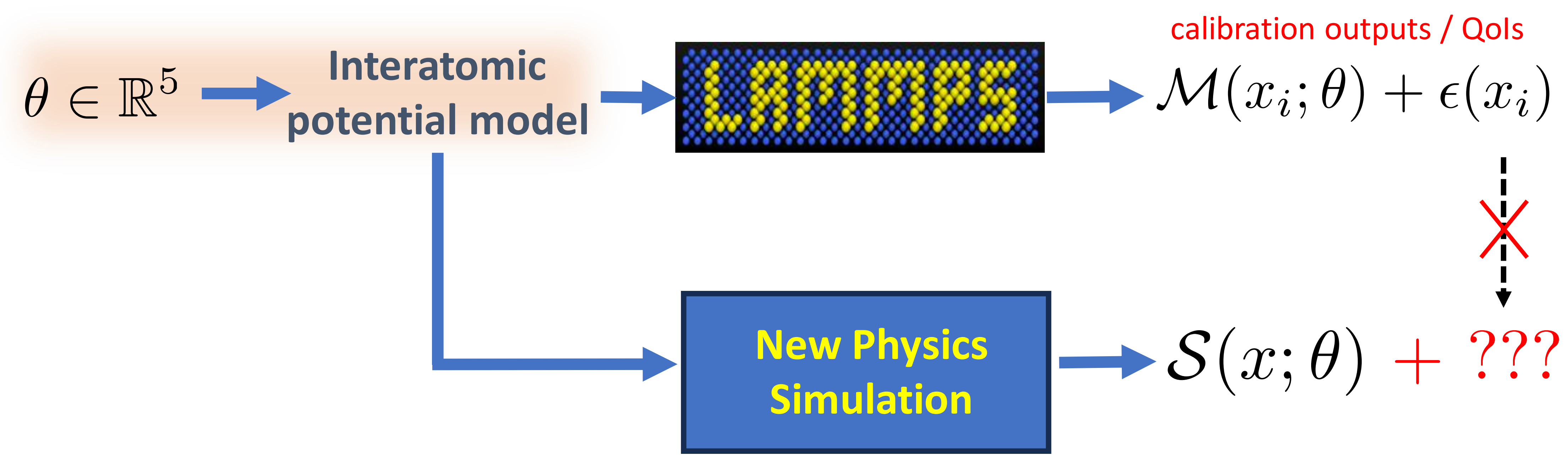
Predictions **w/ model error**



ective is to jointly
known parameters
del error given DFT

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464.

TRANSFERABILITY TO NEW PREDICTION DOMAINS



An additive error model **does not always** transfer to new prediction domains, particularly when the new prediction QoIs differ from the calibration QoIs

TRANSFERABILITY TO NEW PREDICTION DOMAINS

calibration outputs / QoIs

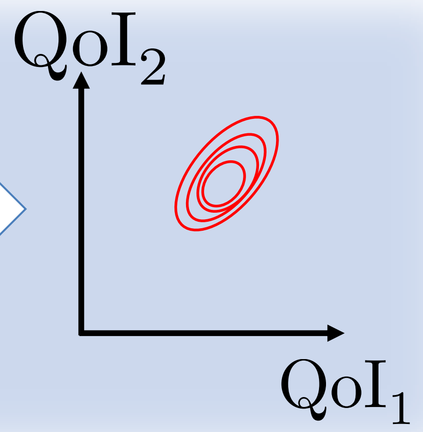
$\theta \in \mathbb{R}$ … $\epsilon(x_i)$

???

Predictions w/o model error

(transferable to the new setting / simulation)

$p(\theta|y)$ ⇒ simulation ~~+ model error~~ ⇒



An additive error model **does not always** transfer to new prediction domains, particularly when the new prediction QoIs differ from the calibration QoIs

EMBEDDED MODEL ERROR: MULTIVARIATE GAUSSIAN

Simplified form of embedded model error --
model error added to the model parameters

$\lambda = \theta + \epsilon$ → Interatomic potential model → LAMMPS → $\mathcal{M}(x_i; \lambda)$

$$\boxed{\epsilon \sim \mathcal{N}(0, S)}$$

Interatomic potential model → New Physics Simulation → $\mathcal{S}(x; \lambda)$

Inference target: $p(\theta, S|y)$

- Full uncertainty representation – model error entirely transferable to new prediction domains
- Predictions constrained by the model – predictive distributions supported only by values realizable by the model.
    - predictions with uncertainties satisfy the model constraints (e.g., conservation laws), but **also inherit the biases of the model.**

ABC-BASED LIKELIHOOD APPROXIMATION

ABC = **A**pproximate **B**ayesian **C**omputation

Challenge: the corresponding likelihood does not have a closed form.

Workaround: use an ABC-based approximate likelihood strategy*

$$p_{\mathrm{ABC}}(y|x,\theta,S) = \frac{1}{\nu\sqrt{2\pi}} \prod_{x,i}^{N} \exp\left(- \frac{\overbrace{(\mu(x_i) - y(x_i))^2}^{\text{nominal error}} + \overbrace{(\sigma(x_i) - \gamma|\mu(x_i) - y(x_i)|)^2}^{\text{error coverage via uncertainty}}}{2\nu^2} \right)$$

$$\mu(x_i) = \mathbb{E}_{\lambda\sim\mathcal{N}(\theta,S)}[\mathcal{M}(x_i;\lambda)]$$

$$\sigma(x_i) = \sqrt{\mathbb{E}_{\lambda\sim\mathcal{N}(\theta,S)}[(\mathcal{M}(x_i;\lambda) - \mu(x_i))^2]}$$

*Sargsyan, K., H. N. Najm, and R. Ghanem. "On the statistical calibration of physical models." *International Journal of Chemical Kinetics* 47.4 (2015): 246-276.

ABC-BASED LIKELIHOOD APPROXIMATION

ABC = **A**pproximate **B**ayesian **C**omputation

Challenge: the corresponding likelihood does not have a closed form.

Workaround: use an ABC-based approximate likelihood strategy*

$$p_{\mathrm{ABC}}(y|x,\theta,S) = \frac{1}{\nu\sqrt{2\pi}} \prod_{x,i}^{N} \exp\left(-\frac{\overbrace{(\mu(x_i)-y(x_i))}^{\text{nominal error}}{}^2 + \overbrace{(\sigma(x_i)-\gamma|\mu(x_i)-y(x_i)|)}^{\text{error coverage via uncertainty}}{}^2}{2\nu^2}\right)$$

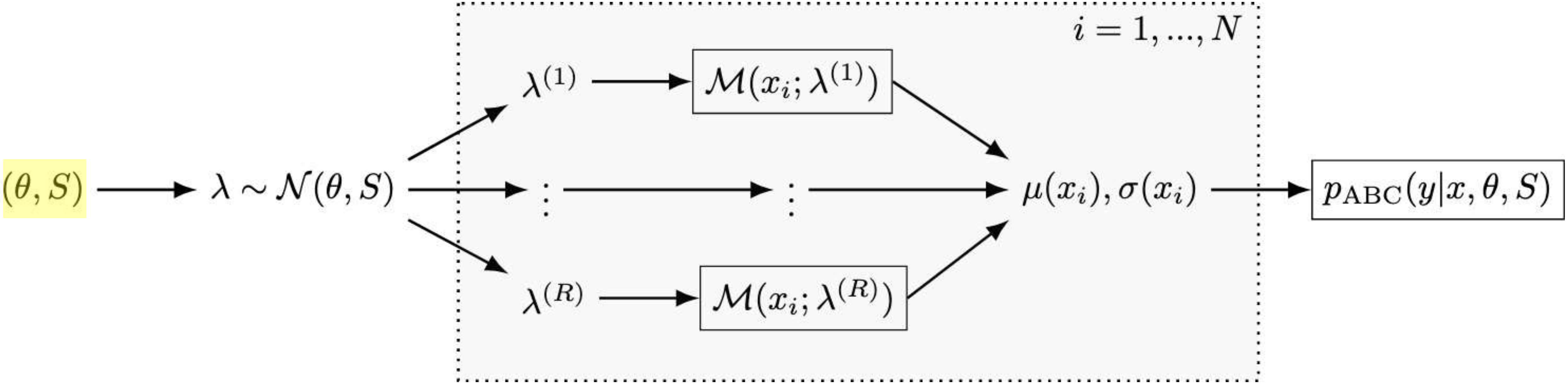
Expensive! Many model evaluations (samples) required per likelihood evaluation.

$$\mu(x_i) = \mathbb{E}_{\lambda\sim\mathcal{N}(\theta,S)}[\mathcal{M}(x_i;\lambda)]$$

$$\sigma(x_i) = \sqrt{\mathbb{E}_{\lambda\sim\mathcal{N}(\theta,S)}[(\mathcal{M}(x_i;\lambda)-\mu(x_i))^2]}$$

*Sargsyan, K., H. N. Najm, and R. Ghanem. "On the statistical calibration of physical models." *International Journal of Chemical Kinetics* 47.4 (2015): 246-276.

COST OF A SINGLE ABC-LIKELIHOOD EVALUATION



In the following slides, we will construct a surrogate encompassing this boxed region.

Computational expense of the simulation often necessitates the use of cheaper-to-evaluate surrogate models of the θ-to-$\mathcal{M}(x_i;\theta)$ map.

When the surrogate models are Gaussian processes (GPs) with Gaussian/RBF kernels,

$$k(\theta,\theta';\nu) = \sigma_k^2 \exp\left(-\frac{1}{2}(\theta-\theta')^T\Gamma^{-1}(\theta-\theta')\right)$$

The GP prediction equations at a new parameter value are Gaussian with mean:

training data: $\{(\theta^j, q^j)\}_{j=1}^M$

$$\mathrm{m}_\star(\theta) = m(\theta) + K(\theta,\Theta)\beta \text{ with } \beta = (K(\Theta,\Theta)+\sigma_n^2 I)^{-1}(q - m(\Theta))$$

σ_n^2 is from the GP likelihood (additive noise)

and variance:

$$\sigma_\star^2(\theta) = \sigma_k^2 - K(\theta,\Theta)(K(\Theta,\Theta)+\sigma_n^2 I)^{-1}K(\Theta,\theta)$$

GP SURROGATES WITH GAUSSIAN INPUTS

Surrogate model: $\cancel{\mathcal{M}(x_i;\lambda)} \Rightarrow \mathcal{N}\left(m_\star(\lambda), \sigma_\star^2(\lambda)\right)$

with embedding: $\lambda \sim \mathcal{N}(\theta, S)$

induces a pushforward density on the surrogate output

If the embedded model error is Gaussian, then we can analytically* compute the quantities that appeared in the ABC-likelihood,

$$m_{x_i}(\theta, S) = \mu(x_i) = \beta^T \ell$$

$$v_{x_i}(\theta, S) = \sigma^2(x_i) = \sigma_k^2 + \sigma_n^2 - \mathrm{trace}((K(\Theta,\Theta) + \sigma_n^2 I)^{-1} L) + \mathrm{trace}\left((L - \ell\ell^T)\beta\beta^T\right)$$

where $\ell \in \mathbb{R}^M$ and $L \in \mathbb{R}^{M\times M}$ require $\mathcal{O}(M)$ and $\mathcal{O}(M^2)$ computations (with respect to the surrogate model training data).

*Candela, J. Q., Girard, A., Larsen, J., & Rasmussen, C. E. (2003, April). Propagation of uncertainty in Bayesian kernel models-application to multiple-step ahead forecasting. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* (Vol. 2, pp. II-701). IEEE.

REDUCTION THROUGH SPARSE GPS AND MULTI-LEVEL MCMC

Challenge: each ABC-likelihood evaluation requires N surrogate evaluation, where each surrogate costs $\mathcal{O}(M^2)$ to evaluate.

Strategy: Build reduced-cost surrogate models via inducing point (pseudo data) approximations to GPs.

- Evaluation cost $\mathcal{O}(J^2)$ where J is the number of inducing points.

increasing fidelity,
increasing evaluation cost

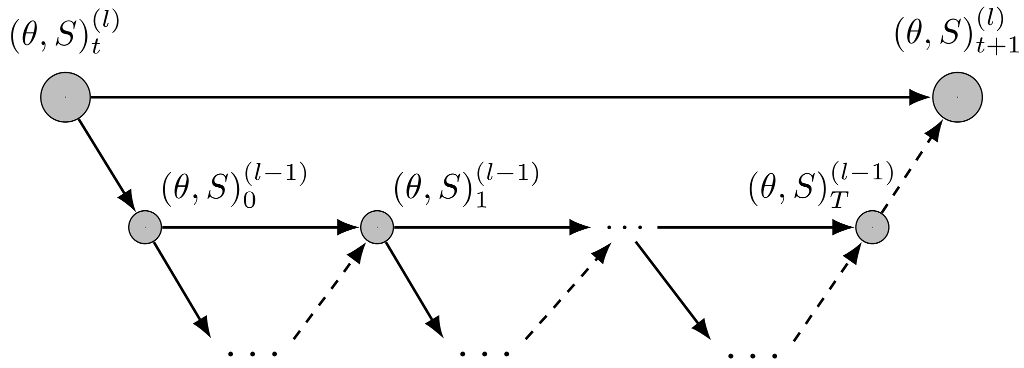
$J_{\text{fine}} = M$

$J_{\text{mid}} < J_{\text{fine}}$

$J_{\text{coarse}} < J_{\text{mid}}$

leads to likelihoods of varying fidelity

use multi-level MCMC algorithms for inference, e.g., multi-level delayed-acceptance (MLDA)*



*Lykkegaard, M. B., Dodwell, T. J., Fox, C., Mingas, G., & Scheichl, R. (2023). Multilevel delayed acceptance MCMC. *SIAM/ASA Journal on Uncertainty Quantification*, *11*(1), 1-30.

CASE STUDY: CALIBRATING POTENTIALS WITH EMBEDDED ERROR

Case 1: Embedding model error into the 5 parameters *individually*

v.s.

Case 2: Embedding into all 5 parameters *simultaneously*

(includes all correlations between model parameters)

Case 1 Setup

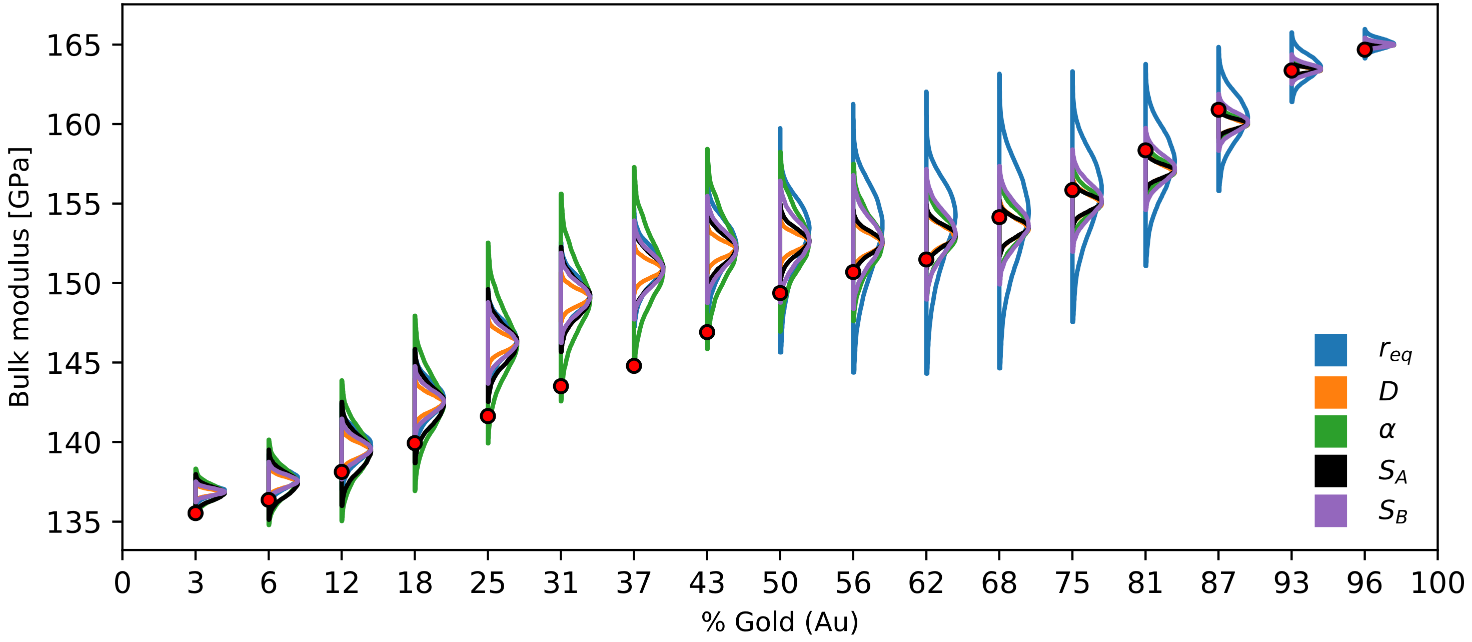
- model parameters + univariate embedding
- 102 QoIs
- ABC-Likelihood w/ GP surrogates
- Adaptive MCMC
  - resulted in low autocorrelations, effective sample size > 2000

Case 2 Setup

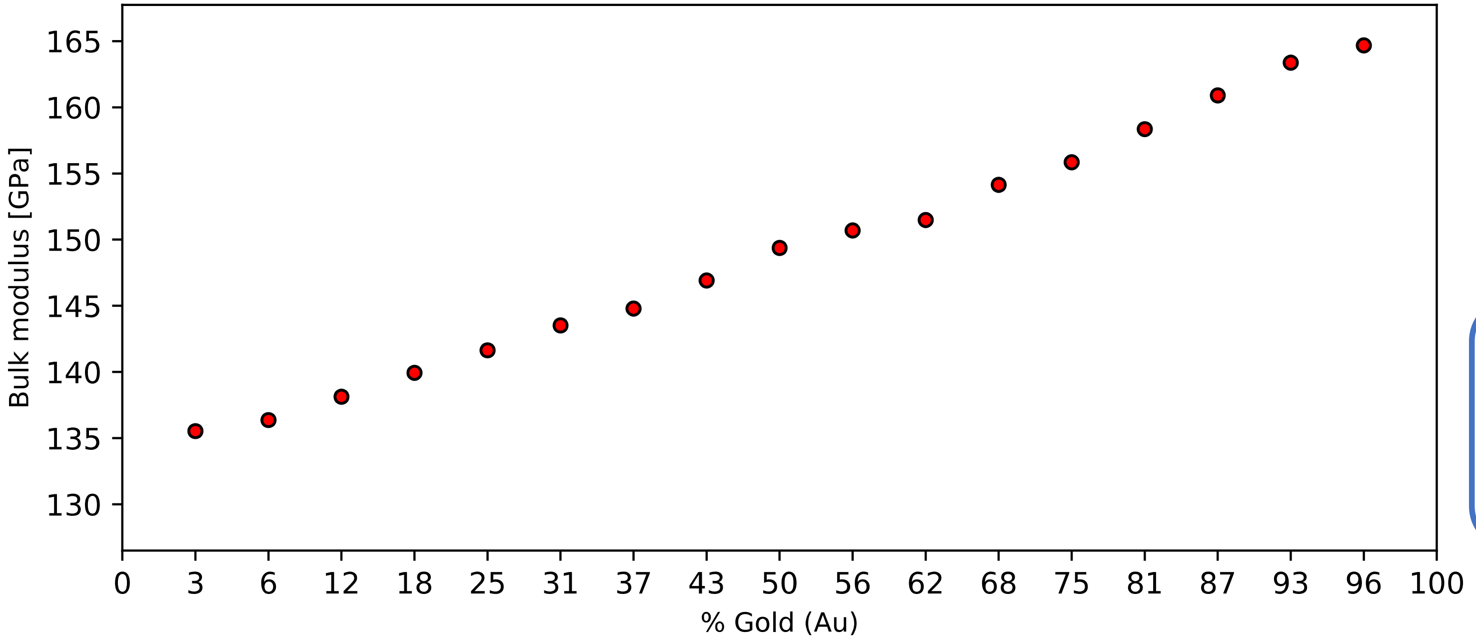
- model parameters + multivariate embedding
- 102 QoIs
- ABC-Likelihood w/ GP surrogates
- Adaptive MCMC: very high autocorrelations, even after > 1 week of sampling

Sparse GPs + Multilevel MCMC (MLDA)
  - 3 levels (w/ 500 and 1000 inducing points)
  - Very low autocorrelations, effective sample size > 2000

Individual embedding



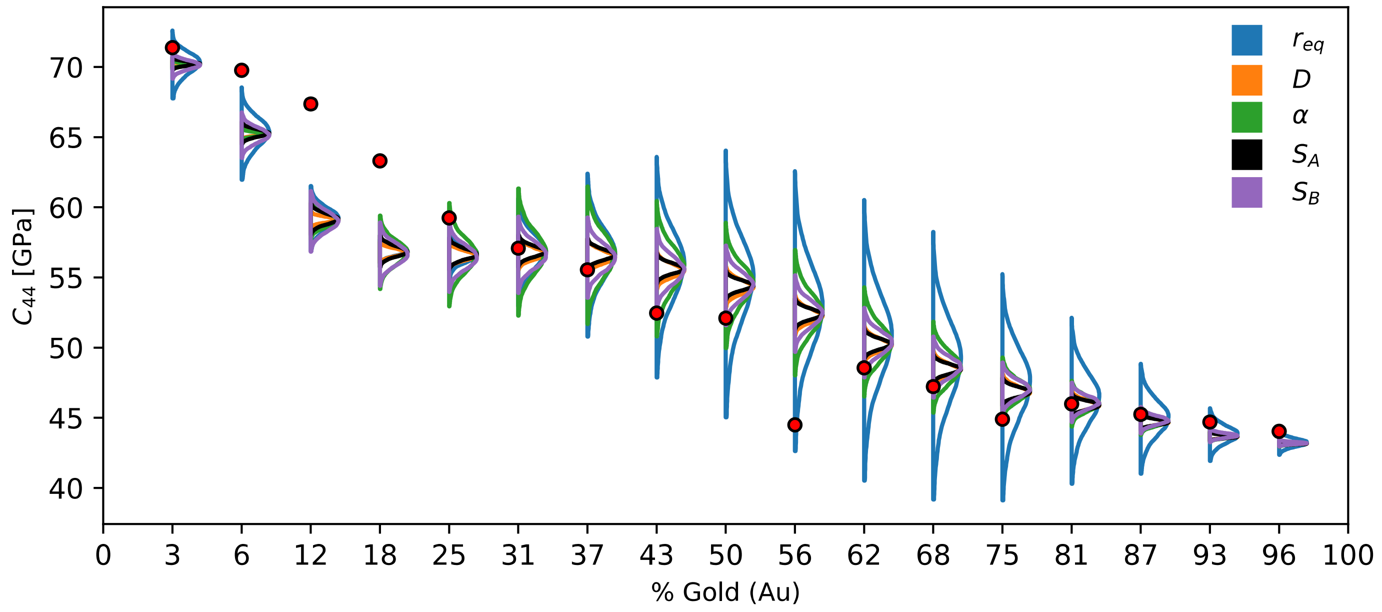
Full embedding



- Embedding-dependent predictive capabilities
- Composition-dependent trends
- Full embedding generally provides better coverage of the data

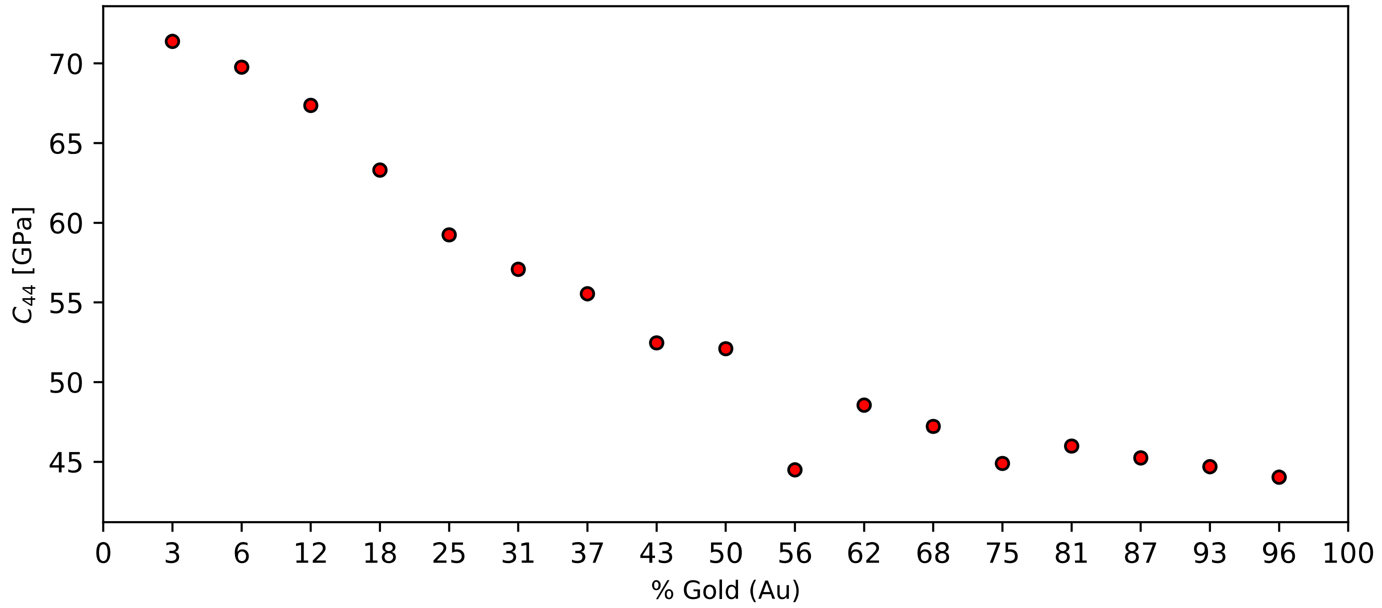
In comparison with the additive model error, this uncertainty is fully transferable.

Individual embedding



- Embedding-dependent predictive capabilities
- Composition-dependent trends

Full embedding



Even with the full embedding, the model struggles to predict the DFT data at low compositions gold – reflects limitations of the underlying potential model.

SUMMARY

- UQ & calibration for interatomic potential models

- Embedded model error (Gaussian embedding)
  - ABC-likelihoods & GP surrogates
  - Further reduction through sparse GPs & multilevel MCMC

- Case study: predictive utility of embedded model error
  - Full uncertainty representation – new predictions account for the calibrated model error

    ➡ Depicted uncertainties are fully transferable to new prediction settings
  - Predictions constrained by the model – predictions with uncertainties satisfy the model constraints, and in doing so also inherit the biases of the model.

    ➡ Embedding-dependence and composition-dependence in predictions

    ➡ Inability to predict certain QoIs within uncertainty
  - Combined, these findings provide new and useful diagnostic information that was not present in our prior work.

ACKNOWLEDGEMENTS

Arun Hegde, **Cosmin Safta**, and **Habib Najm** are supported by the U.S. Department of Energy, Office of Science:

- Advanced Scientific Computing Research (ASCR)
- Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute
- National Energy Research Scientific Computing Center (NERSC)

Elan Weiss and **Wolfgang Windl** are supported by:

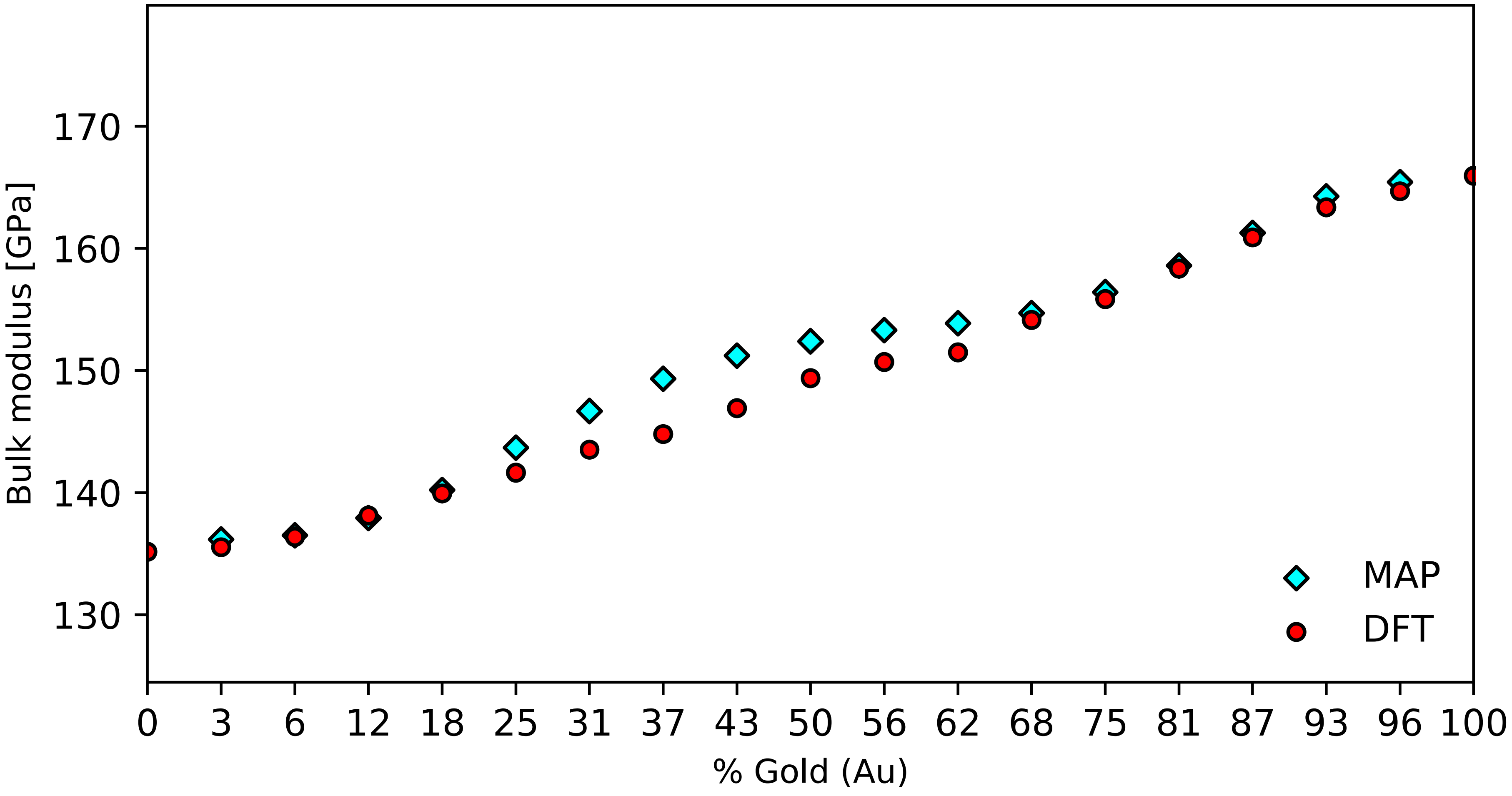
- Air Force Office of Scientific Research (AFOSR) under Contract No.FA9550-19-1-0378
- Ohio Supercomputer Center, Project No. PAS0072

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the presentation do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

APPENDIX

ILLUSTRATION: PREDICTIONS WITH AND WITHOUT MODEL ERROR

With (left, grey) and without (right, **blue**) a simple form of additive error (iid Gaussian)



Notably,

- **predictions w/o model error** under-represent uncertainty.
- **predictions with model error** provide coverage of the DFT data.

In *new* prediction settings – where additive model error cannot be transferred -- the uncertainty will be considerably under-represented.

*Hegde, A., Weiss, E., Windl, W., Najm, H., & Safta, C. "Bayesian calibration of interatomic potentials for binary alloys." *Computational Materials Science* 214 (2022): 111660.

GP SURROGATES WITH GAUSSIAN INPUTS

Surrogate model: $\cancel{\mathcal{M}(x_i;\lambda)} \Rightarrow \mathcal{N}\left(m_\star(\lambda), \sigma_\star^2(\lambda)\right)$

If the embedded model error is Gaussian, then we can analytically* compute the quantities that appeared in the ABC-likelihood

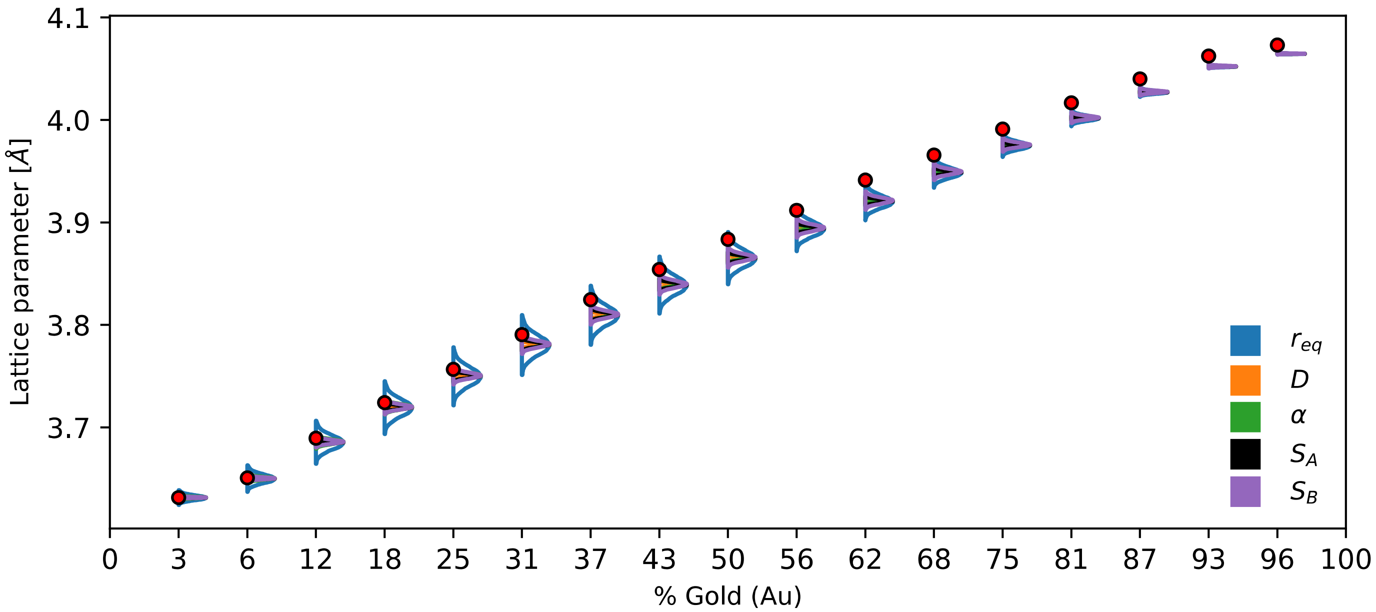
$$m_{x_i}(\theta, S) = \mu(x_i) = \beta^T \ell$$

$$v_{x_i}(\theta, S) = \sigma^2(x_i) = \sigma_k^2 + \sigma_n^2 - \mathrm{trace}((K(\Theta,\Theta) + \sigma_n^2 I)^{-1} L) + \mathrm{trace}\left((L - \ell\ell^T)\beta\beta^T\right)$$

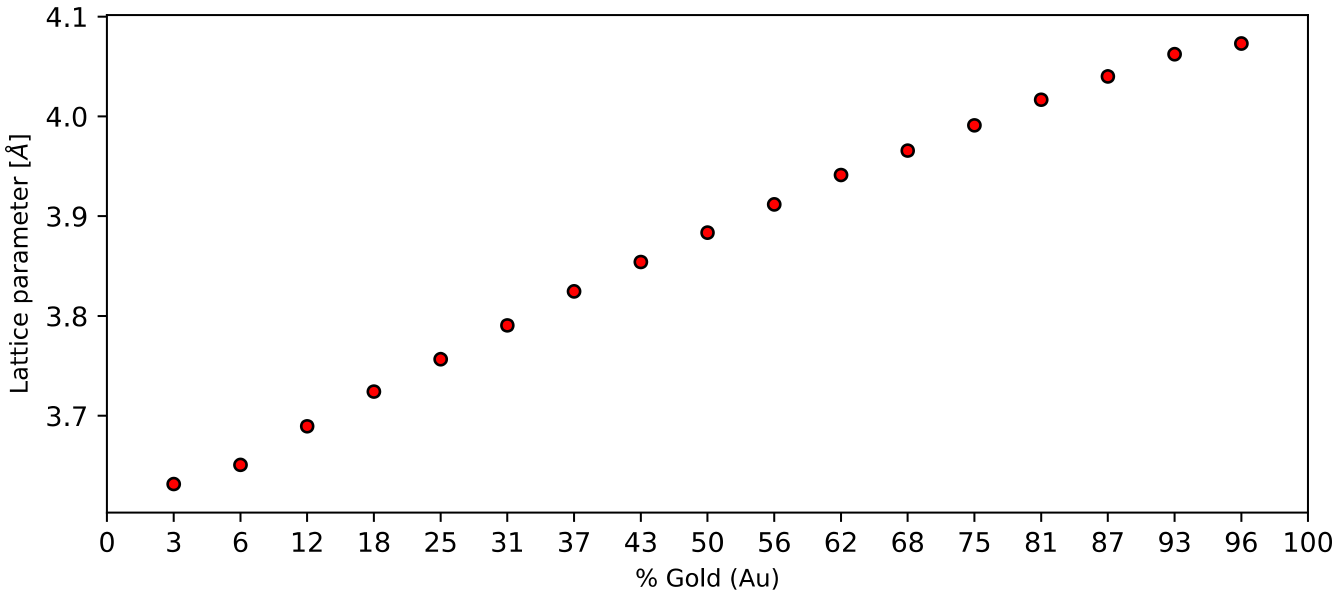
$$\ell_i = \sigma_k^2 \frac{1}{\sqrt{\det(\Gamma_{\mathcal{A}}^{-1} S + I)}} \exp\left(-\frac{1}{2}((\lambda_{\mathcal{A}^c} - \theta^i_{\mathcal{A}^c})^T \Gamma_{\mathcal{A}^c}^{-1}(\lambda_{\mathcal{A}^c} - \theta^i_{\mathcal{A}^c})\right) \exp\left(-\frac{1}{2}(\theta - \theta^i_{\mathcal{A}})^T (S + \Lambda_{\mathcal{A}})^{-1}(\theta - \theta^i_{\mathcal{A}})\right)$$

$$L_{ij} = k(\lambda|_{\lambda_{\mathcal{A}}=\theta}, \theta^i) k(\lambda|_{\lambda_{\mathcal{A}}=\theta}, \theta^j) \cdot \frac{1}{\sqrt{\det(2\Gamma_{\mathcal{A}}^{-1} S + I)}} \exp\left(2\left(\theta - \frac{1}{2}(\theta^i_{\mathcal{A}} + \theta^j_{\mathcal{A}})\right)^T \Gamma_{\mathcal{A}}^{-1}(2\Gamma_{\mathcal{A}}^{-1} + S^{-1})^{-1}\Gamma_{\mathcal{A}}^{-1}\left(\theta - \frac{1}{2}(\theta^i_{\mathcal{A}} + \theta^j_{\mathcal{A}})\right)\right).$$

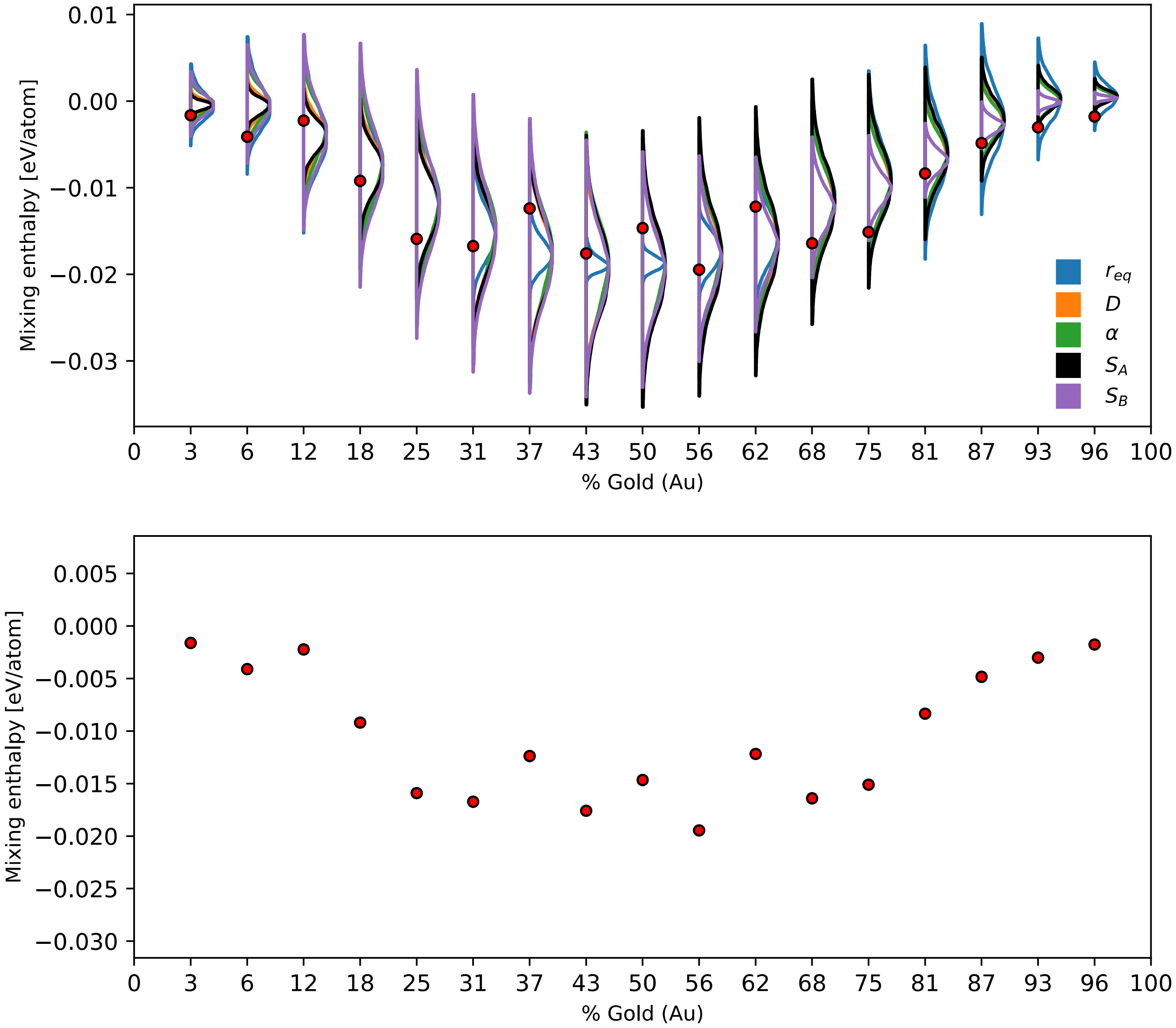
*Candela, J. Q., Girard, A., Larsen, J., & Rasmussen, C. E. (2003, April). Propagation of uncertainty in Bayesian kernel models-application to multiple-step ahead forecasting. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* (Vol. 2, pp. II-701). IEEE.



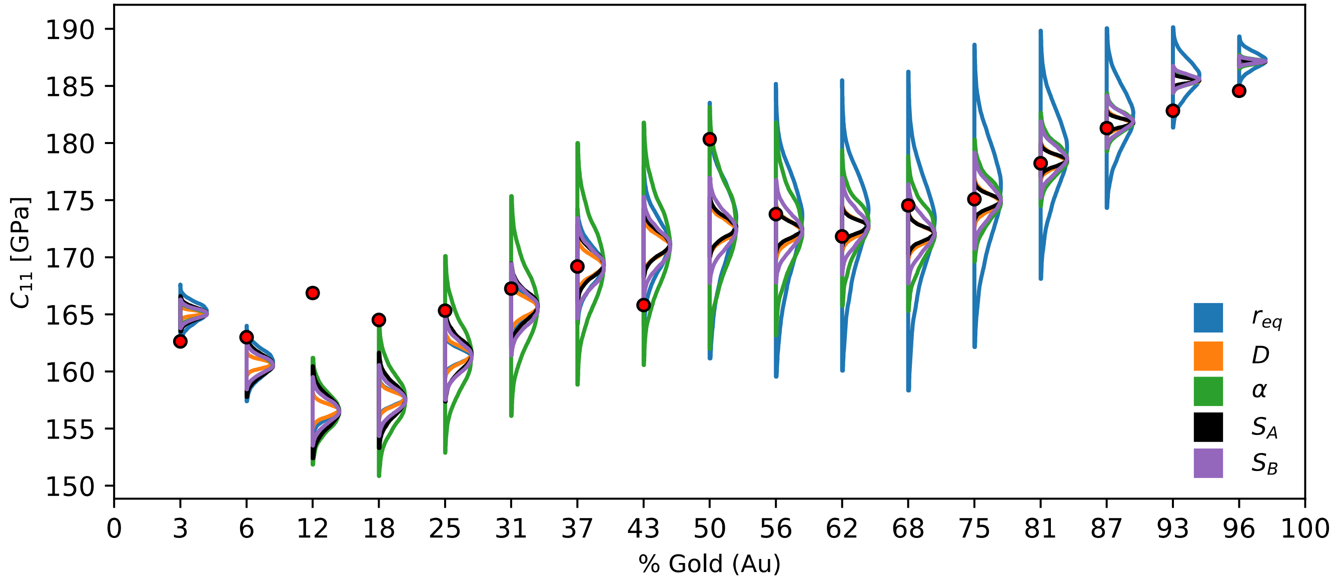
4.1
4.0
3.9
3.8
3.7
Lattice parameter [Å]
0 3 6 12 18 25 31 37 43 50 56 62 68 75 81 87 93 96 100
% Gold (Au)
r_{eq}
D
α
S_A
S_B



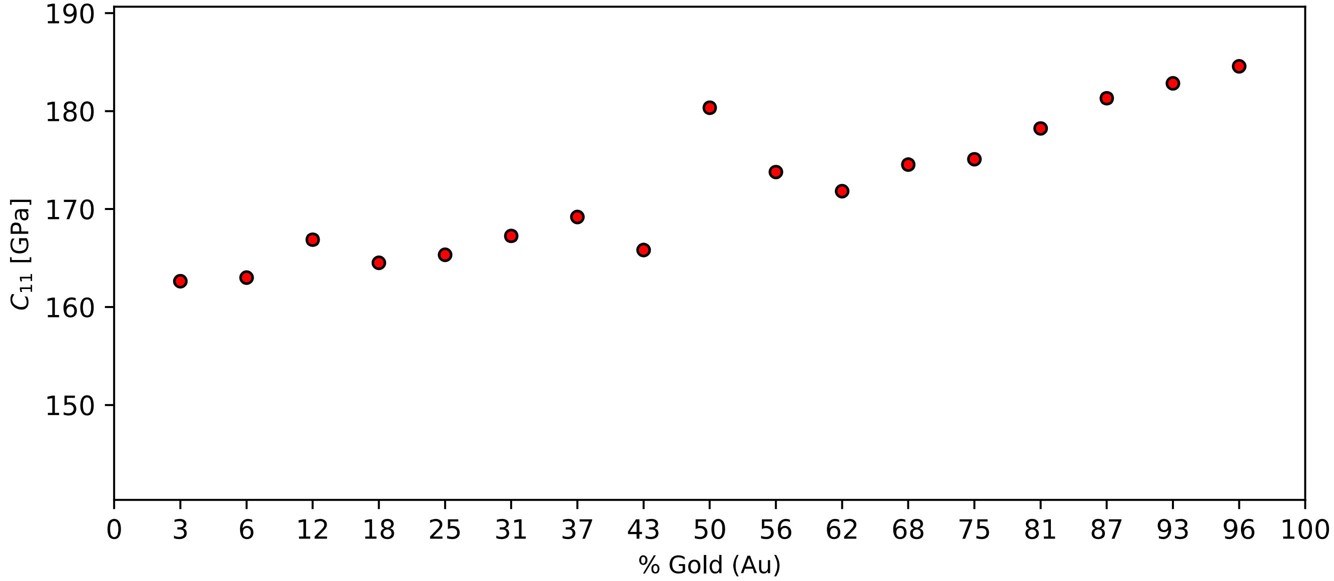
4.1
4.0
3.9
3.8
3.7
Lattice parameter [Å]
0 3 6 12 18 25 31 37 43 50 56 62 68 75 81 87 93 96 100
% Gold (Au)



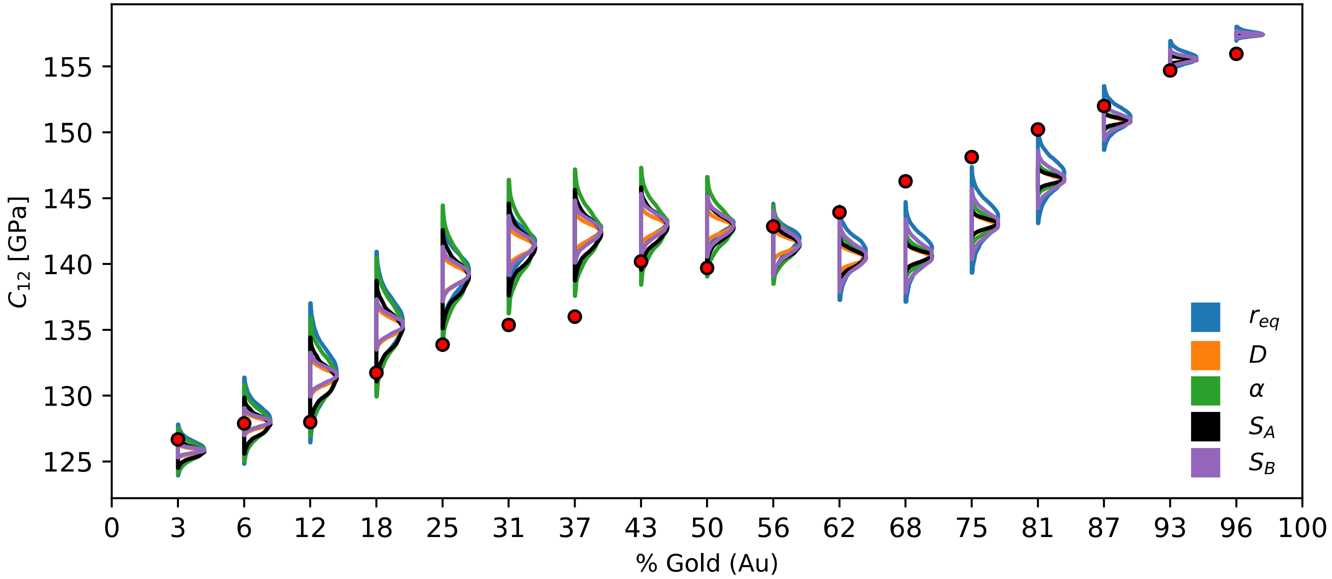
0.01
0.00
−0.01
−0.02
−0.03
Mixing enthalpy [eV/atom]
0
3
6
12
18
25
31
37
43
50
56
62
68
75
81
87
93
96
100
% Gold (Au)
r_{eq}
D
α
S_A
S_B
0.005
0.000
−0.005
−0.010
−0.015
−0.020
−0.025
−0.030
Mixing enthalpy [eV/atom]
% Gold (Au)



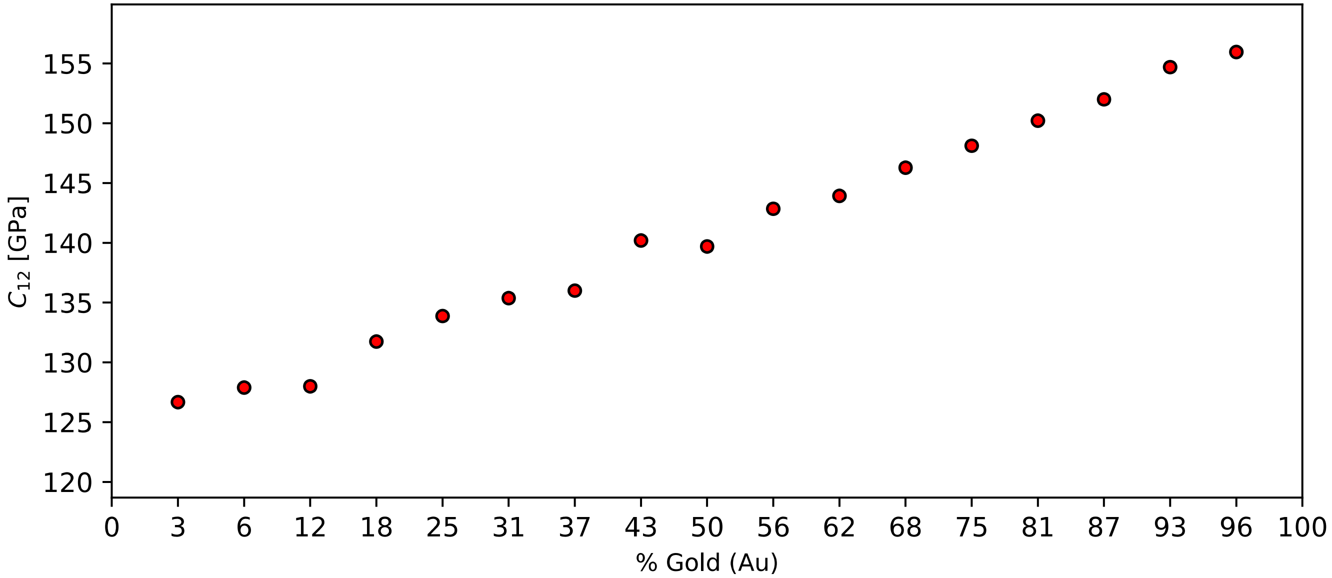
190
185
180
175
170
165
160
155
150
C_{11} [GPa]
0 3 6 12 18 25 31 37 43 50 56 62 68 75 81 87 93 96 100
% Gold (Au)
r_{eq}
D
α
S_A
S_B



190
180
170
160
150
C_{11} [GPa]
0 3 6 12 18 25 31 37 43 50 56 62 68 75 81 87 93 96 100
% Gold (Au)



C_{12} [GPa]
% Gold (Au)
r_{eq}
D
α
S_A
S_B



C_{12} [GPa]
% Gold (Au)