TECHNISCHE UNIVERSITÄT DARMSTADT

GAUSSIAN PROCESS EMULATORS WITH ACTIVE SUBSPACE LEARNING

with Margarida Companys and Achim Schwenk

Tom Plies

SFB 1245

TECHNISCHE UNIVERSITÄT DARMSTADT

MOTIVATION

- Bayesian LEC inference could be improved through medium-mass constraints
- Full IMSRG calculations are significantly too expensive
- Emulators can be constructed based on Gaussian Processes (GPs)
- Regular GP emulators are too inefficient to approximate IMSRG

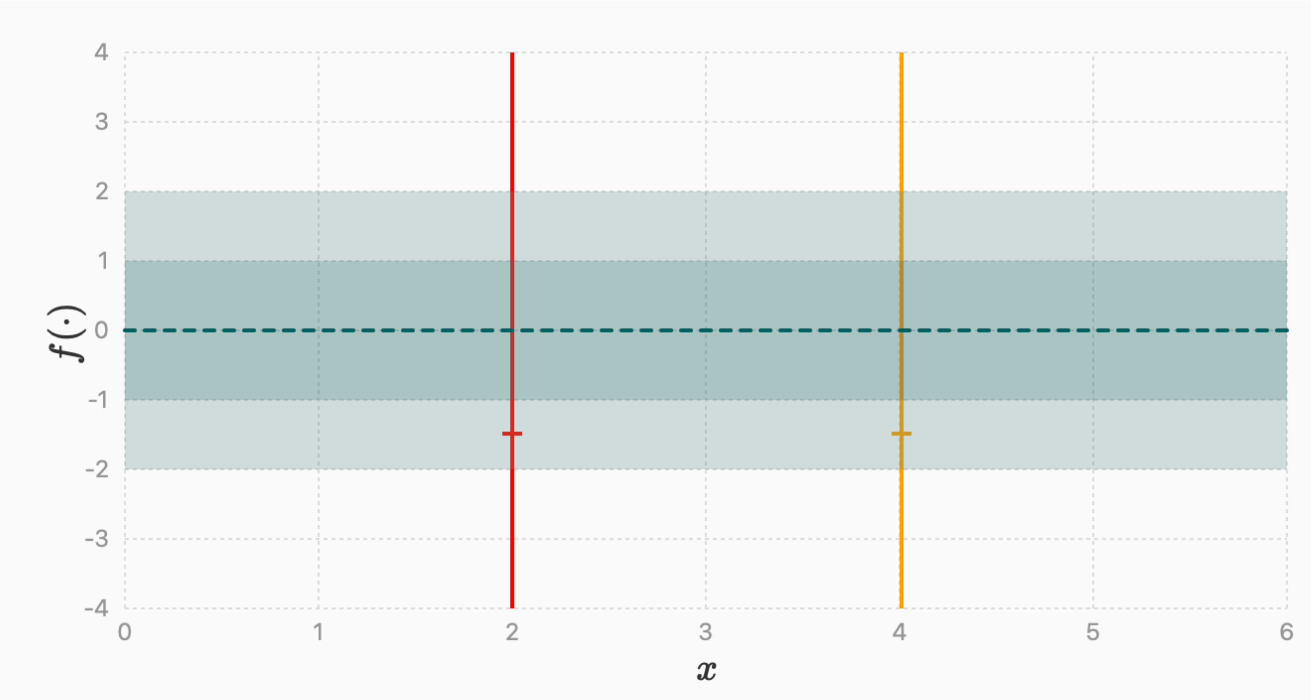
- **Goal: Emulator for nuclear structure observables with Gaussian Processes**

TECHNISCHE UNIVERSITÄT DARMSTADT

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \ldots, f(x_n)) \sim \mathcal{N}\left(m(X), K(X, X')\right)$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* \mid \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg\max_{\theta} \log p(\mathbf{y} \mid \theta)$
- GP effectively an interpolator (for large enough training space)

Unconditioned GP (prior) $\ell = 1$



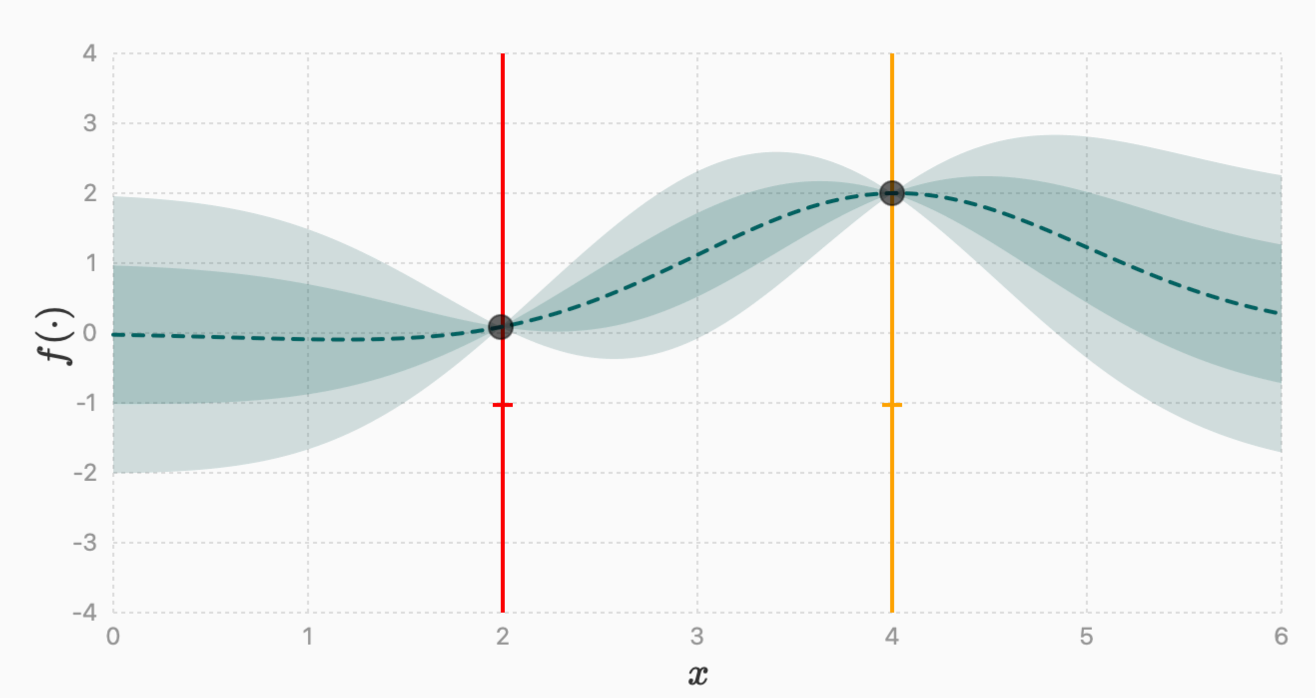
https://www.infinitecuriosity.org/vizgp/

TECHNISCHE UNIVERSITÄT DARMSTADT

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \ldots, f(x_n)) \sim \mathcal{N}\left(m(X), K(X, X')\right)$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* \mid \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg\max_{\theta} \log p(\mathbf{y} \mid \theta)$
- GP effectively an interpolator (for large enough training space)

Conditioned GP (posterior) $\ell = 1$



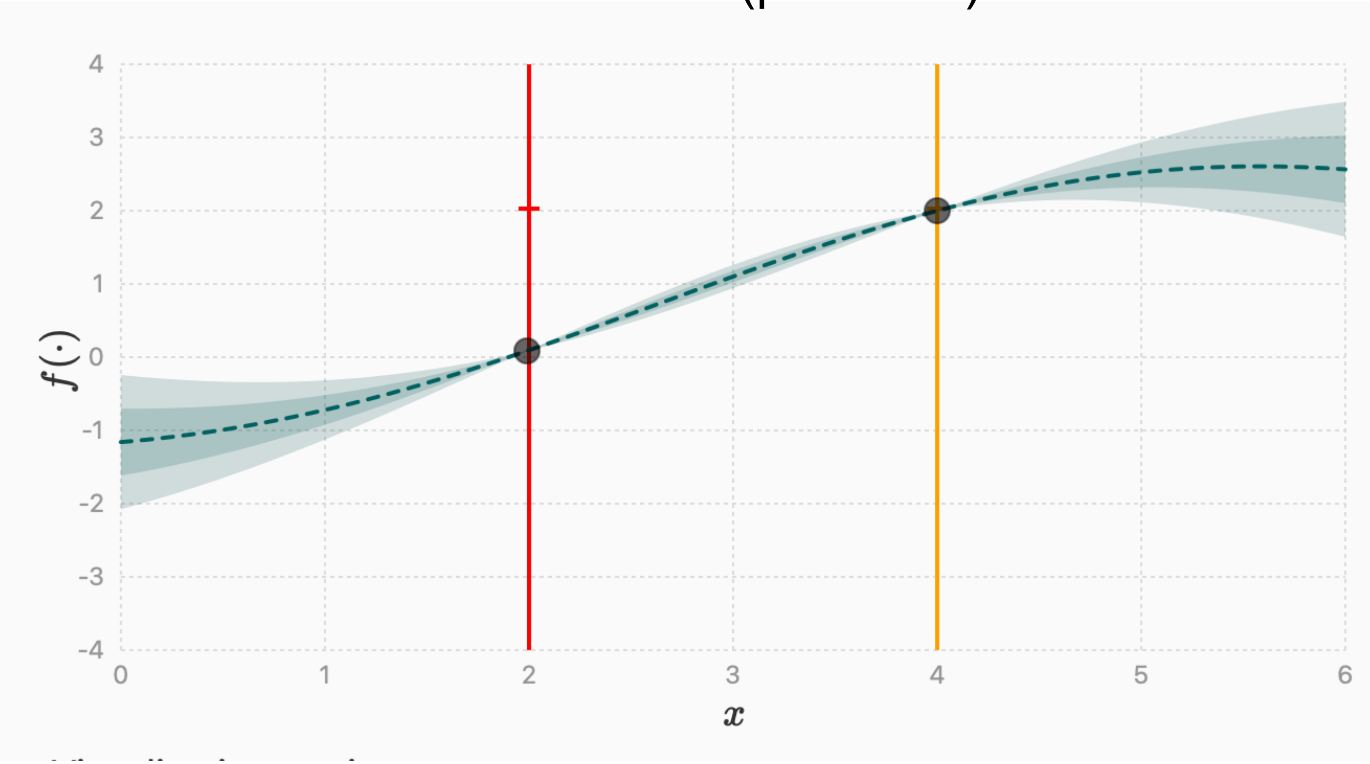
https://www.infinitecuriosity.org/vizgp/

TECHNISCHE UNIVERSITÄT DARMSTADT

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \ldots, f(x_n)) \sim \mathcal{N}\left(m(X), K(X, X')\right)$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* \mid \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg\max_\theta \log p(\mathbf{y} \mid \theta)$
- GP effectively an interpolator (for large enough training space)

Conditioned GP (posterior) $\ell = 3$



https://www.infinitecuriosity.org/vizgp/

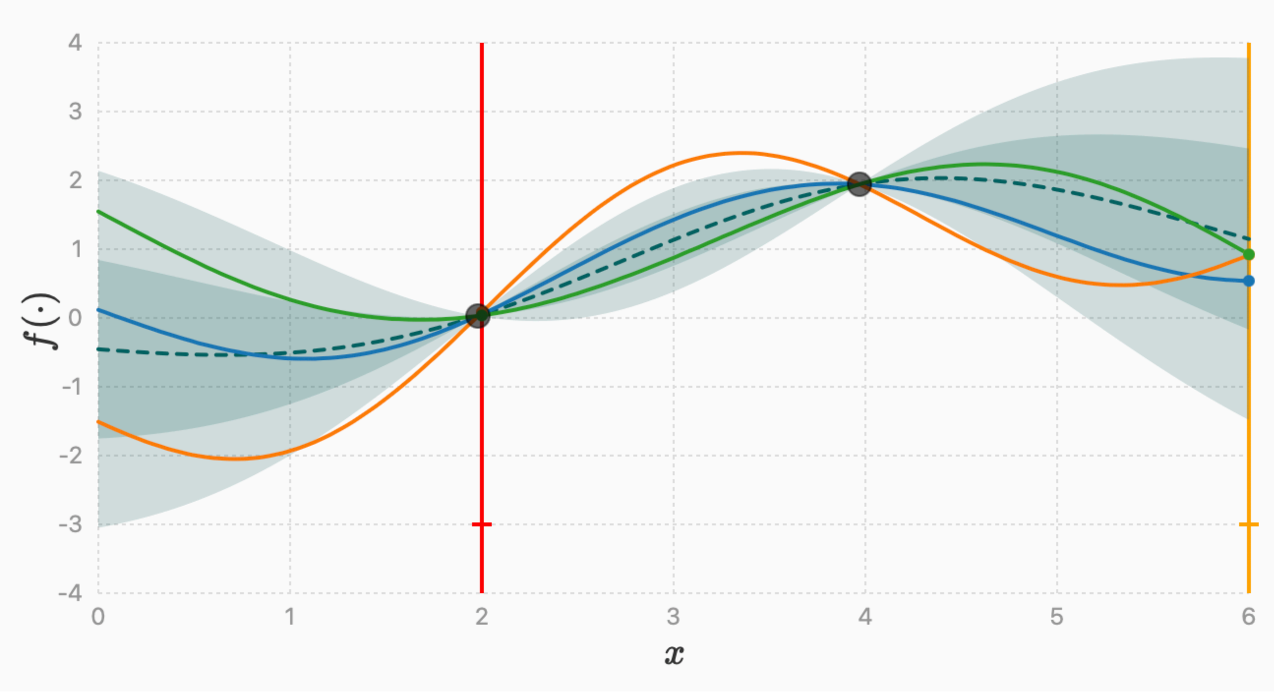
TECHNISCHE UNIVERSITÄT DARMSTADT

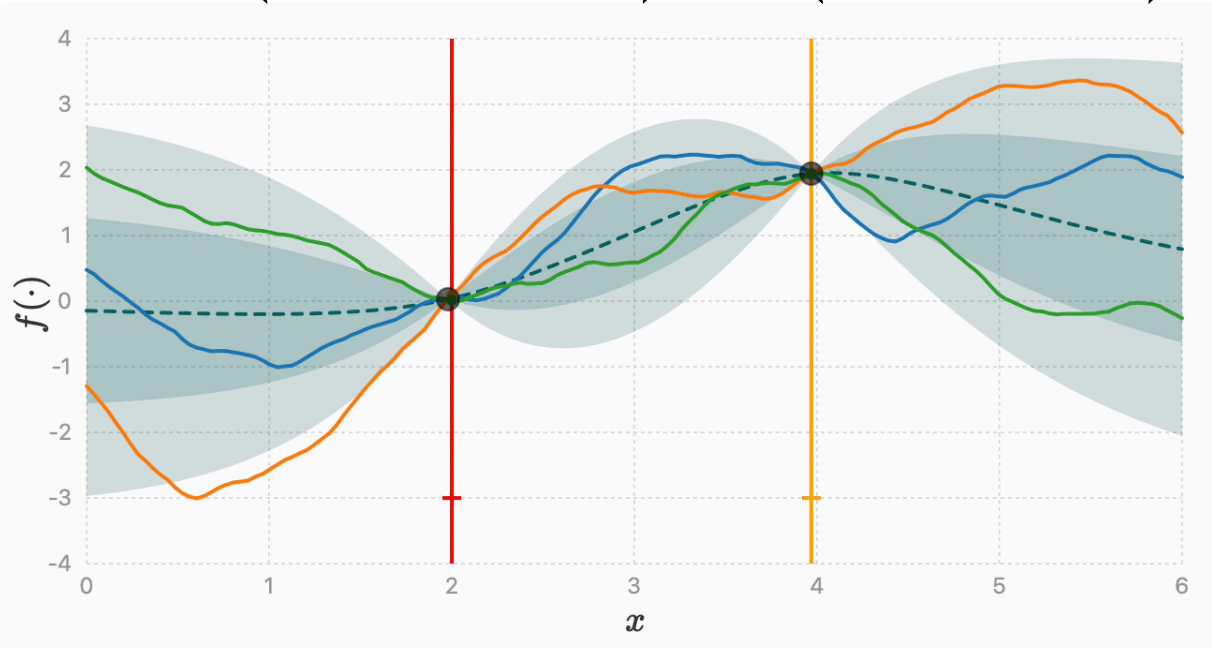
KERNELS

- Squared-exponential kernel vs Matern 3/2

$$K(x,x') = \sigma^2 \exp\left(-(x-x')^T(x-x')/2\ell^2\right)$$

$$K(x,x') = \sigma^2 \left(1 + \frac{\sqrt{3}\,|x-x'|}{\ell}\right) \exp\left(-\frac{\sqrt{3}\,|x-x'|}{\ell}\right)$$





https://www.infinitecuriosity.org/vizgp/

TECHNISCHE UNIVERSITÄT DARMSTADT

MULTIDIMENSIONAL INPUT GAUSSIAN PROCESSES

- Example: Variation of 2 LECs ⟶ Prediction of binding energies

$$K(\boldsymbol{x}, \boldsymbol{x}') = \bar{c}^2 \exp\left(-(\boldsymbol{x} - \boldsymbol{x}')^T L^{-1} (\boldsymbol{x} - \boldsymbol{x}')\right) \qquad \text{with } \boldsymbol{x} = (c_D, c_E)$$

$$L = \begin{pmatrix} \ell_{c_D}^2 & \ell_{c_D,c_E}^2 \\ \ell_{c_E,c_D}^2 & \ell_{c_E}^2 \end{pmatrix}$$

Number of ℓ parameters scales as $\dfrac{n_{\text{LECs}} \times (n_{\text{LECs}} - 1)}{2}$

TECHNISCHE UNIVERSITÄT DARMSTADT

MULTIDIMENSIONAL INPUT GAUSSIAN PROCESSES

- Example: Variation of 2 LECs ⟶ Prediction of binding energies

$$K(\boldsymbol{x}, \boldsymbol{x}') = \bar{c}^2 \exp\left(-(\boldsymbol{x} - \boldsymbol{x}')^T L^{-1} (\boldsymbol{x} - \boldsymbol{x}')\right) \qquad \text{with } \boldsymbol{x} = (c_D, c_E)$$

$$L = \begin{pmatrix} \ell^2_{c_D} & \ell^2_{c_D, c_E} \\ \ell^2_{c_E, c_D} & \ell^2_{c_E} \end{pmatrix} \longrightarrow L' = \begin{pmatrix} \ell^2_A & 0 \\ 0 & \ell^2_B \end{pmatrix}$$

Number of ℓ parameters scales as $\dfrac{n_{\text{LECs}} \times (n_{\text{LECs}} - 1)}{2}$

Goal: Transform parameter space into orthogonal parameter directions

TECHNISCHE UNIVERSITÄT DARMSTADT

ACTIVE SUBSPACE LEARNING

Constantine et al., SIAM J. Sci. Comput. 36 (2014)

- **Goal: Identify directions in LEC space along which the observable is most sensitive**

$$C = \mathbb{E}[\nabla f(x) \nabla f(x)^T]$$

$$C = W \Lambda W^T$$

$$C \approx \tilde{C} = \frac{1}{M} \sum_{j=1}^{M} (\nabla_x f(x_j))(\nabla_x f(x_j))^T$$

- Have to estimate the gradient for each sample

$$\tilde{C} = G G^T \qquad G = \frac{1}{\sqrt{M}} [\nabla_x f_1 \ \ldots \ \nabla_x f_M]$$

SVD: $G = \tilde{W} \sqrt{\tilde{\Lambda}} V^T$

- Divide into active and inactive directions

$$\boldsymbol{x} = W_1 \boldsymbol{y} + W_2 \boldsymbol{z}$$

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \qquad W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}$$

- Transform input into active subspace

$$\boldsymbol{y} = \tilde{W}_1^T \cdot \boldsymbol{x}$$

- Discard inactive directions

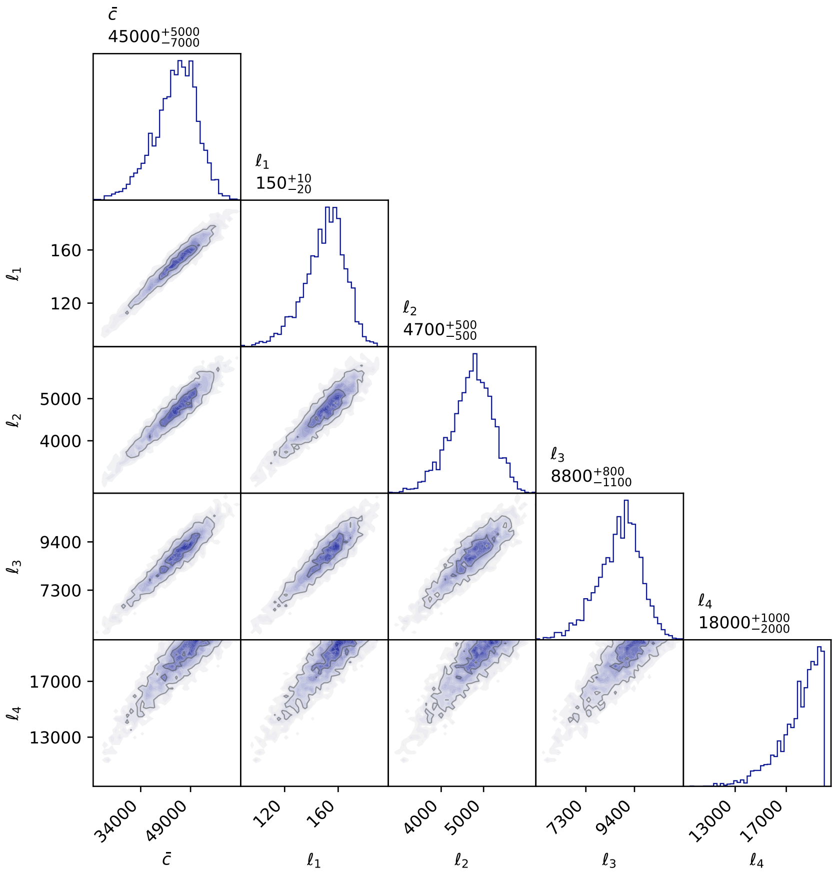
$$\boldsymbol{z} = \tilde{W}_2^T \cdot \boldsymbol{x}$$

TECHNISCHE UNIVERSITÄT DARMSTADT

APPLICATION TO BINDING ENERGIES

Companys et al., arxiv:2510.08362 (2025)

- $^{48}\mathrm{Ca}$ binding energies from Hartree-Fock emulator with 4 LECs : $\boldsymbol{x} = (c_3, c_4, c_D, c_E), f(x) = E_{\mathrm{gs}}$
- Transform data into active subspace: $\boldsymbol{y} = \tilde{W}_1^T \cdot \boldsymbol{x}$
- Sample hyperparameters: $\bar{c}, \ell_1, \ell_2, \ell_3, \ell_4$ with MCMC
- Use maximum a posteriori (MAP) parameter values
- Higher-order directions could potentially be discarded (indicated by large ℓ)

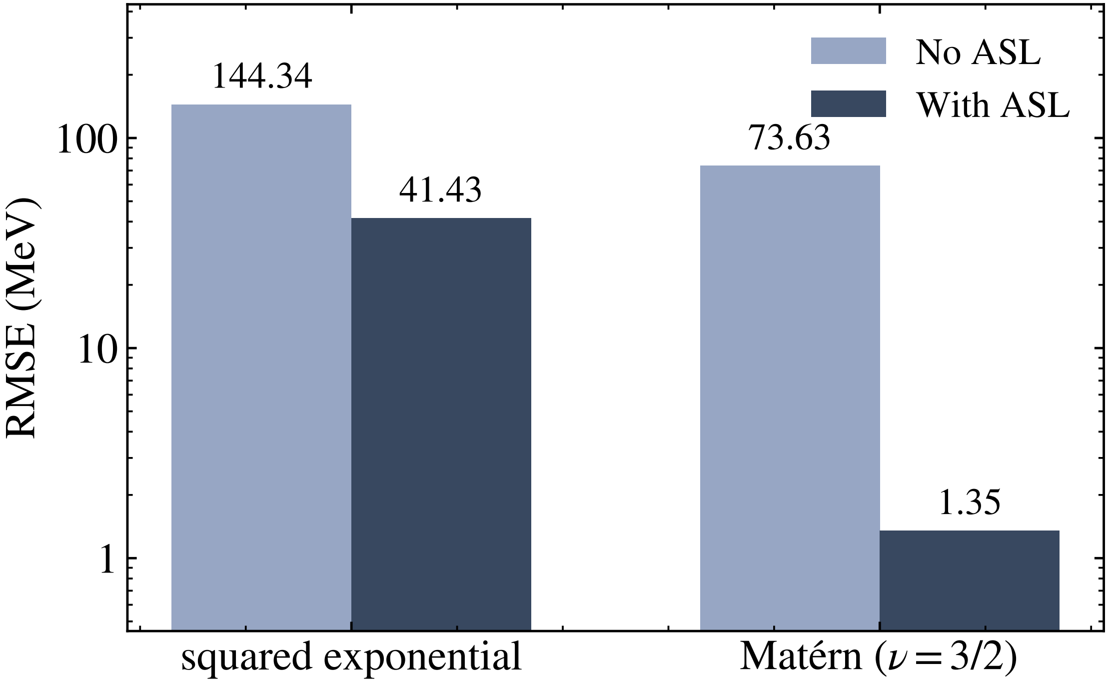


RESULTS

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_{\text{GP,i}})^2}$$

- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\text{HF}}(^{48}\text{Ca}) \sim -300\,\text{MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with ~0.1% deviation
- Poor empirical coverage -> overestimated uncertainties

200 training points



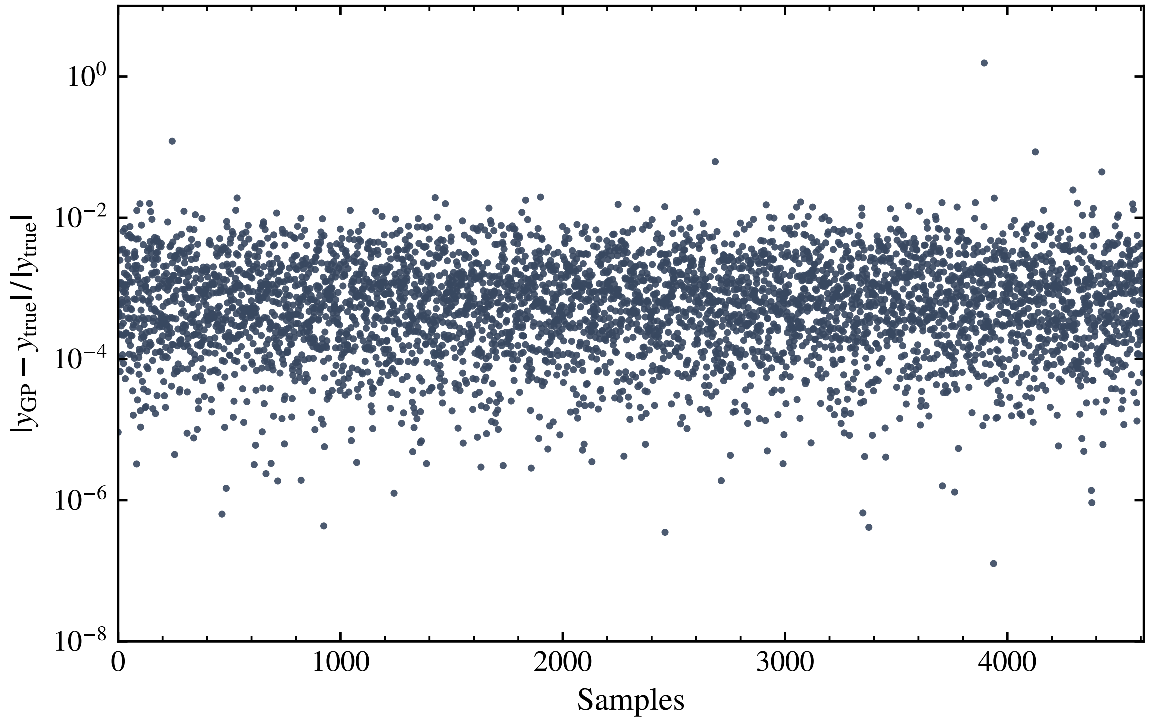
TECHNISCHE UNIVERSITÄT DARMSTADT

RESULTS

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_{\mathrm{GP,i}})^2}$$

- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\mathrm{HF}}(^{48}\mathrm{Ca}) \sim -300\,\mathrm{MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with ~0.1% deviation
- Poor empirical coverage -> overestimated uncertainties

200 training points

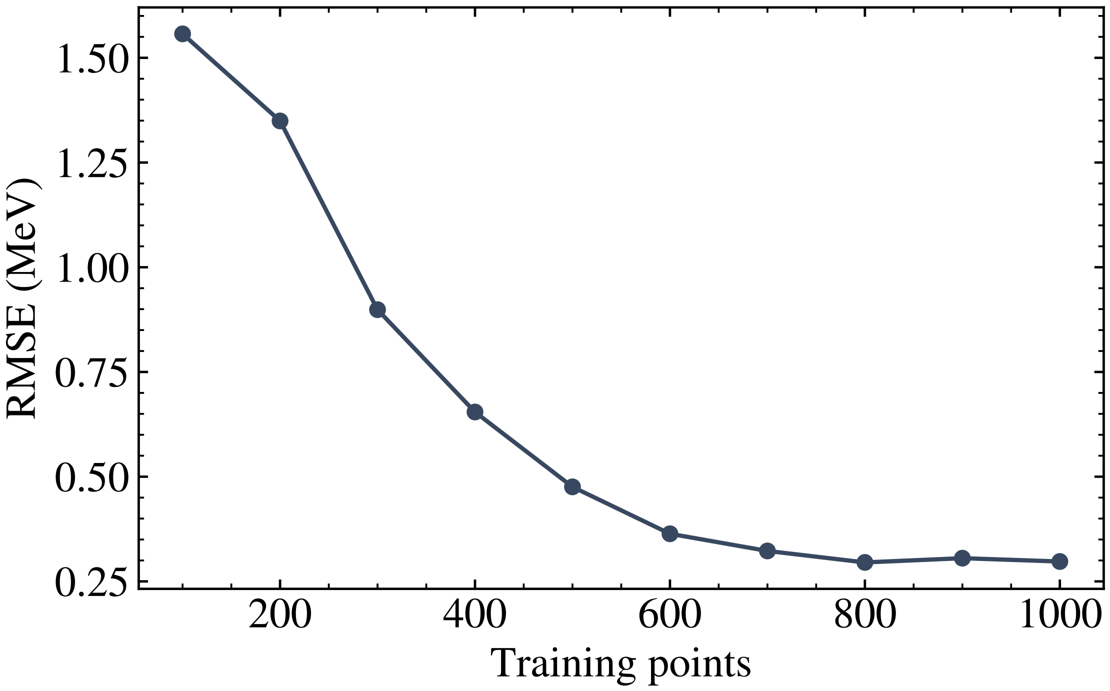


TECHNISCHE UNIVERSITÄT DARMSTADT

RESULTS

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_{\mathrm{GP,i}})^2}$$

- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\mathrm{HF}}(^{48}\mathrm{Ca}) \sim -300\,\mathrm{MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with ~0.1% deviation
- Poor empirical coverage -> overestimated uncertainties



TECHNISCHE UNIVERSITÄT DARMSTADT

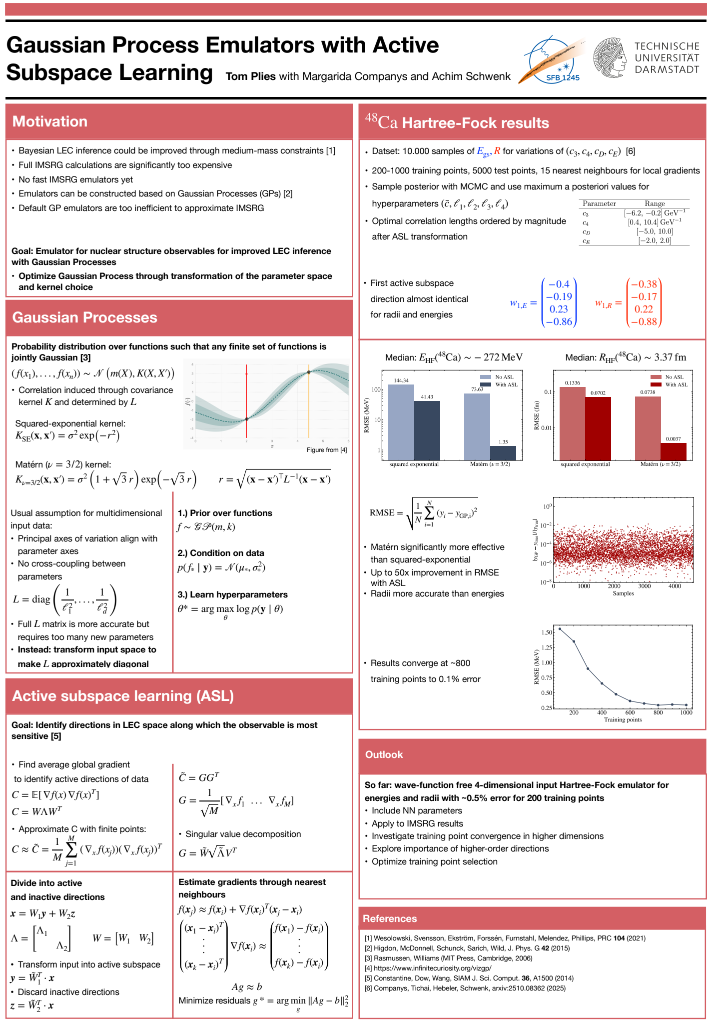
OUTLOOK

- Built wave-function-free multi-input emulator
- Extend to include NN parameters
- Potentially discard inactive directions
- Optimize selection of training points
- **Apply to IMSRG**

TECHNISCHE UNIVERSITÄT DARMSTADT

OUTLOOK

- Built wave-function-free multi-input emulator
- Extend to include NN parameters
- Potentially discard inactive directions
- Optimize selection of training points
- **Apply to IMSRG**



Thank you for your attention

TECHNISCHE UNIVERSITÄT DARMSTADT

LOCAL GRADIENTS

- Assume approximately local linear behavior: $f(\boldsymbol{x}_j) \approx f(\boldsymbol{x}_i) + \nabla f(\boldsymbol{x}_i)^T(\boldsymbol{x}_j - \boldsymbol{x}_i)$

For one point:

$$\begin{pmatrix} x_{j1} - x_{i1} \\ x_{j2} - x_{i2} \end{pmatrix} \cdot \nabla f(\boldsymbol{x_i}) = f(\boldsymbol{x}_j) - f(\boldsymbol{x}_i)$$

$$\nabla f(\boldsymbol{x}_i) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}_i) \\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}_i) \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

For k points:

$$\begin{pmatrix} (\boldsymbol{x}_1 - \boldsymbol{x}_i)^T \\ \vdots \\ (\boldsymbol{x}_k - \boldsymbol{x}_i)^T \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \approx \begin{pmatrix} f(\boldsymbol{x}_1) - f(\boldsymbol{x}_i) \\ \vdots \\ f(\boldsymbol{x}_k) - f(\boldsymbol{x}_i) \end{pmatrix}$$

$$Ag \approx b$$

→ Residual $r = Ag - b$

Minimize residuals $g^* = \arg\min_g \|Ag - b\|_2^2$

- Compute differences in f and $\boldsymbol{x}$ between all **k nearest neighbours** of x_i with $j \in \{1,\ldots,k\}$
- Minimize residuals with least-square fit to find compatible common gradient $\nabla f(\boldsymbol{x_i})$

TECHNISCHE UNIVERSITÄT DARMSTADT

GAUSSIAN PROCESSES

$$f(x) \sim \mathcal{GP}\Big(m(x), k(x, x')\Big), m(x) = \mathbb{E}[f(x)], \qquad k(x, x') = \mathrm{Cov}\Big(f(x), f(x')\Big)\,.$$

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} K + \sigma_n^2 I K_*^\top \\ K_* K_{**} \end{bmatrix} \right),$$

$$\mathbf{f}_* \mid X, \mathbf{y}, X_* \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \mathrm{Cov}(\mathbf{f}_*)\right),$$

$$\bar{\mathbf{f}}_* = \mathbf{m}_* + K_* \left(K + \sigma_n^2 I\right)^{-1} (\mathbf{y} - \mathbf{m}),$$

$$\mathrm{Cov}(\mathbf{f}_*) = K_{**} - K_* \left(K + \sigma_n^2 I\right)^{-1} K_*^\top,$$

$$[K_*]_{ij} = k(x_i, x_{*j}), \quad [K_{**}]_{ij} = k(x_{*i}, x_{*j})\,.$$

SENSITIVITY ANALYSIS

- Identify most sensitive LEC directions
- c_E is most important
- Energies and radii have almost identical dominant directions

$$w_{1,E} = \begin{pmatrix} -0.4 \\ -0.19 \\ 0.23 \\ -0.86 \end{pmatrix} \qquad w_{1,R} = \begin{pmatrix} -0.38 \\ -0.17 \\ 0.22 \\ -0.88 \end{pmatrix}$$

TECHNISCHE UNIVERSITÄT DARMSTADT

RADII RESULTS

- $R(^{48}\mathrm{Ca}) \sim 3.37\,\mathrm{fm}$
- 0.1% deviation at 200 training points