SAND2022-8981C

Sandia National Laboratories

Bayesian calibration for summary statistics with applications to a cluster dynamics model

P. Robbe, L. Bonnet, T. Casey, K. Sargsyan, H. Najm (SNL)
C. Matthews, M. Cooper, D. Andersson (LANL)

July 18, 2022

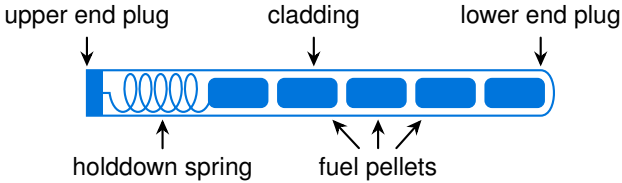
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

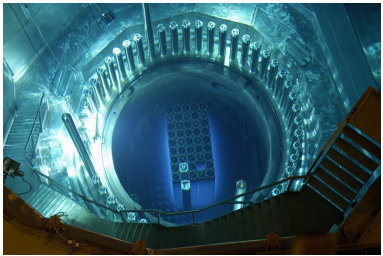
Acknowledgements

This material is based upon work supported by the U. S. Department of Energy, Office of Nuclear Energy and Office of Science, Office of Advanced Scientific Computing Research (ASCR) through the Scientific Discovery through Advanced Computing (SciDAC) project on Simulation of Fission Gas

UO_2 nuclear fuel rods

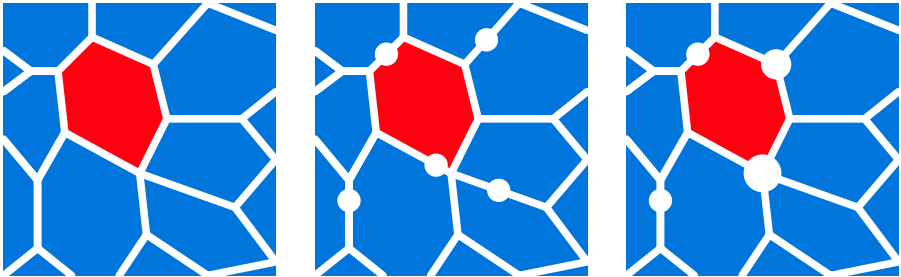
- Nuclear fuel assembly consists of an array of fuel rods filled with fuel pellets





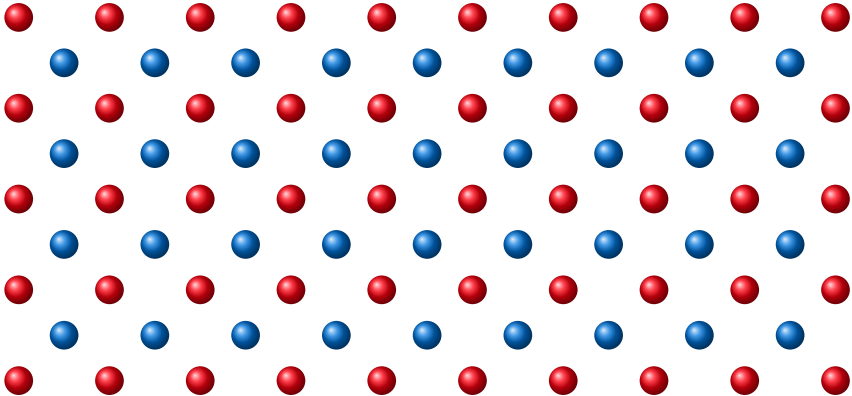
Fission gas release

- Nuclear fuel pellets have a microstructure composed of grains
- Fission gas release happens in three distinct stages: [Andersson et al., 2014]
  1. Nucleation and growth of intra-granular bubbles
  2. Diffusion to the grain boundaries
  3. Growth and interconnection of gas bubbles at grain boundaries



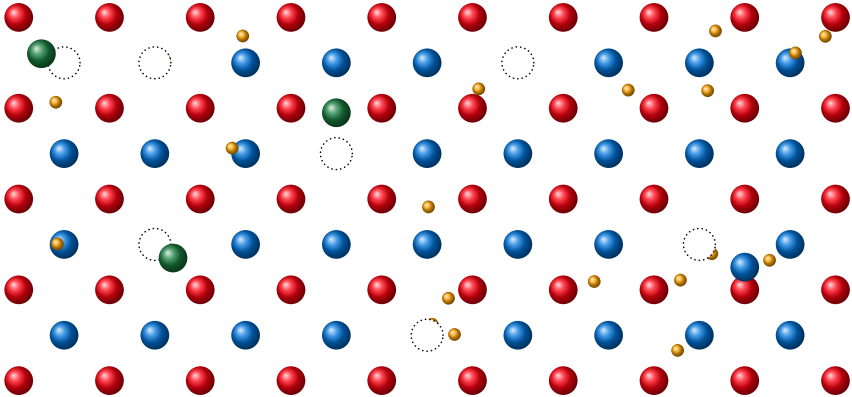
Nucleation of intra-granular bubbles

- Uranium ( ) and oxygen ( ) form a lattice



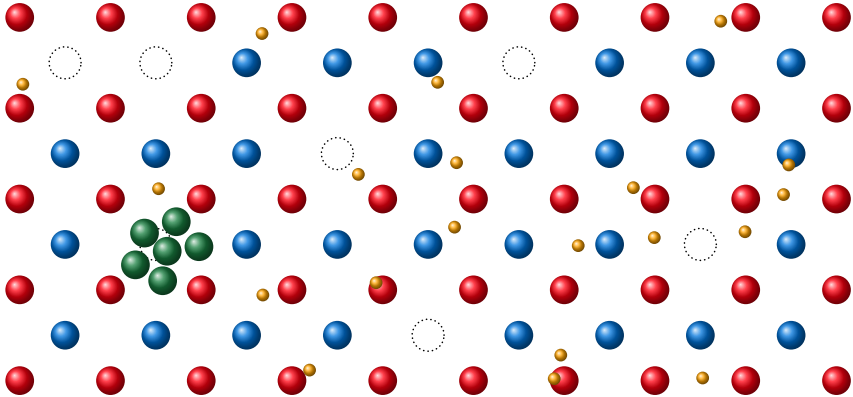
Nucleation of intra-granular bubbles

- Uranium () and oxygen () form a lattice
- Free neutrons () initialize the fission process, creating byproducts such as xenon ()



Nucleation of intra-granular bubbles

- Uranium (●) and oxygen (●) form a lattice
- Free neutrons (●) initialize the fission process, creating byproducts such as xenon (●)
- Xenon gas bubbles get trapped inside the vacancies (○) and migrate to the grain boundary

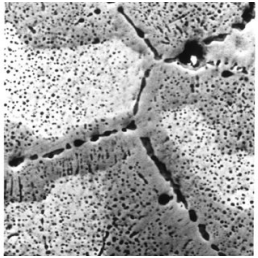


Mechanistic model for intra-granular fission gas evolution

- Evolution equation for the concentration C_n of each cluster with n Xe atoms

$$\frac{\partial C_n}{\partial t} = \underbrace{\dot{F} y_n}_{\substack{\text{production} \\ \text{of new Xe} \\ \text{clusters}}} + \underbrace{D_n(T) \nabla C_n}_{\substack{\text{diffusion} \\ \text{of Xe} \\ \text{clusters}}} - \underbrace{Q(C_n)}_{\substack{\text{interaction} \\ \text{between} \\ \text{clusters}}}$$

- The diffusion coefficient $D_n(T)$ is computed by
  - Density Functional Theory (DFT) under thermal equilibrium [Perriot et al., 2019]
  - The `Centipede` cluster dynamics code under irradiation [Matthews et al., 2020]

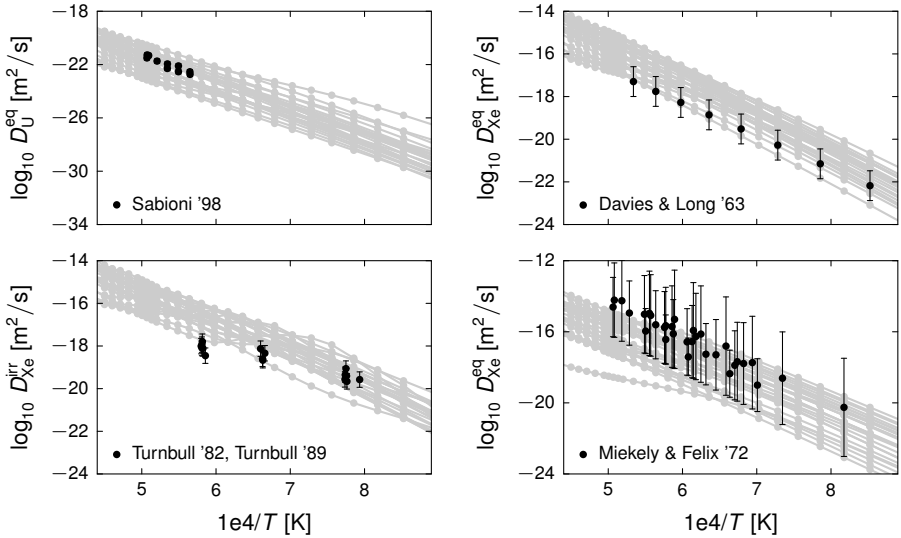


We monitor the following quantities:

- U diffusivity under thermal equilibrium conditions ($D_{\mathrm{U}}^{\mathrm{eq}}$)
- Xe diffusivity under thermal equilibrium conditions ($D_{\mathrm{Xe}}^{\mathrm{eq}}$)
- Xe diffusivity under irradiation conditions ($D_{\mathrm{Xe}}^{\mathrm{irr}}$)
- O non-stoichiometry $UO_{2\pm x}$

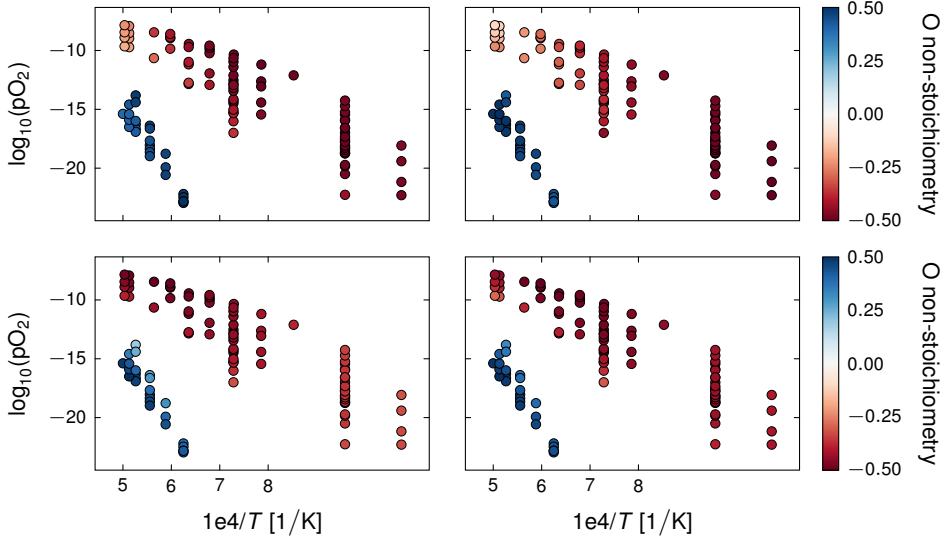
Measurements of the diffusivity

- `Centipede` predictions of $D_{\mathrm{U}}^{\mathrm{eq}}$, $D_{\mathrm{Xe}}^{\mathrm{eq}}$ and $D_{\mathrm{Xe}}^{\mathrm{irr}}$ with measurements taken from the literature



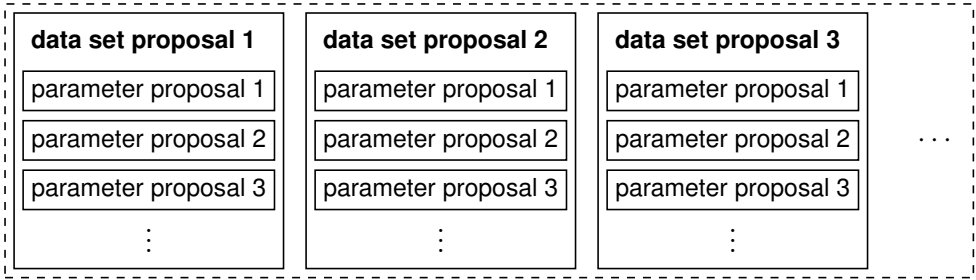
Measurements of the stoichiometry

- `Centipede` predictions of O non-stoichiometry $UO_{2\pm x}$



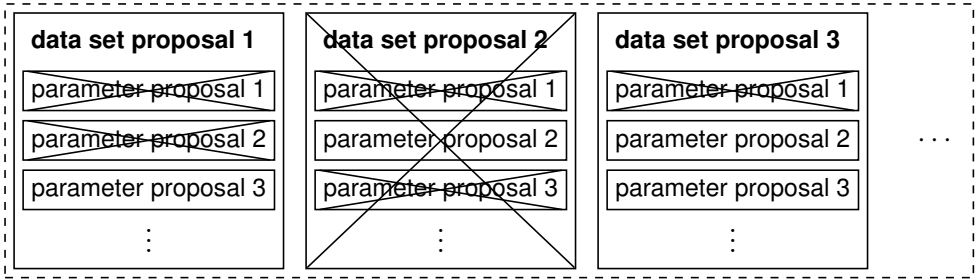
Objectives

- Our goal is to perform Bayesian calibration of the `Centipede` cluster dynamics code, taking into account that only summary statistics of the data (i.e., a measurement value with an associated measurement error) are available
- Data-free inference (DFI) involves an outer iteration in the data space and an inner iteration in the parameter space [Berry et al., 2012, Chowdhary and Najm, 2016]



Objectives

- Our goal is to perform Bayesian calibration of the `Centipede` cluster dynamics code, taking into account that only summary statistics of the data (i.e., a measurement value with an associated measurement error) are available
- Data-free inference (DFI) involves an outer iteration in the data space and an inner iteration in the parameter space [Berry et al., 2012, Chowdhary and Najm, 2016]



- We want to develop a computationally tractable DFI approach that can deal with multiple experimental data sets of different size

Bayesian calibration

- Given a data set

$$\mathcal{D}_d = \{\boldsymbol{x}_d^{(n)}, y_d^{(n)}, s_d^{(n)}\}_{n=1}^{N_d}$$

  and a prior $p(\boldsymbol{\nu})$, we want to compute the posterior $p(\boldsymbol{\nu}|\mathcal{D}_d)$
- Prior and posterior are connected through the likelihood function (Bayes' theorem)

$$p(\boldsymbol{\nu}|\mathcal{D}_d) \propto \mathcal{L}_d(\boldsymbol{\nu})p(\boldsymbol{\nu})$$

- Assume model outputs are available as

$$z_d^{(n)} := \underbrace{f_d(\boldsymbol{x}_d^{(n)}, \boldsymbol{\nu})}_{\text{model}} + \underbrace{\sigma_d^{(n)}\eta}_{\text{additive noise}}$$

  where $\eta \sim \mathcal{N}(0, 1)$ is a standard normal random variable
- This yields a log likelihood of the form

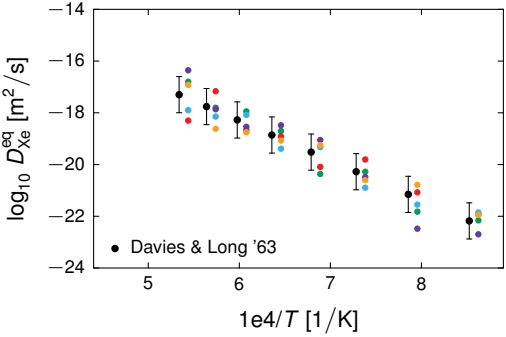
$$\log \mathcal{L}_d(\boldsymbol{\nu}) = -\frac{1}{2}\sum_{n=1}^{N_d}\left[\log(2\pi{\sigma_d^{(n)}}^2) + \frac{\left(z_d^{(n)} - f_d(\boldsymbol{x}_d^{(n)}, \boldsymbol{\nu})\right)^2}{{\sigma_d^{(n)}}^2}\right]$$

  and samples of the posterior can be obtained using an accept-reject scheme

Bayesian calibration

- DFI approach defines consistent data sets using a metric based on maximum entropy
- We will formulate an alternative metric for consistency that avoids the outer loop over the data space altogether
- Define a collection of K_d synthetic data sets $\mathcal{Z}_d^{(k)} := \{z_d^{(n,k)}\}_{n=1}^{N_d}$, $k = 1, 2, \ldots, K_d$, where

$$z_d^{(n,k)} \sim \mathcal{N}\left(y_d^{(n)}, \beta_d s_d^{(n)^2}\right)$$



Consistent data sets

- Each synthetic data set represents an opinion on the posterior $p(\boldsymbol{\nu}|\mathcal{D}_d)$ that can be combined using logarithmic pooling

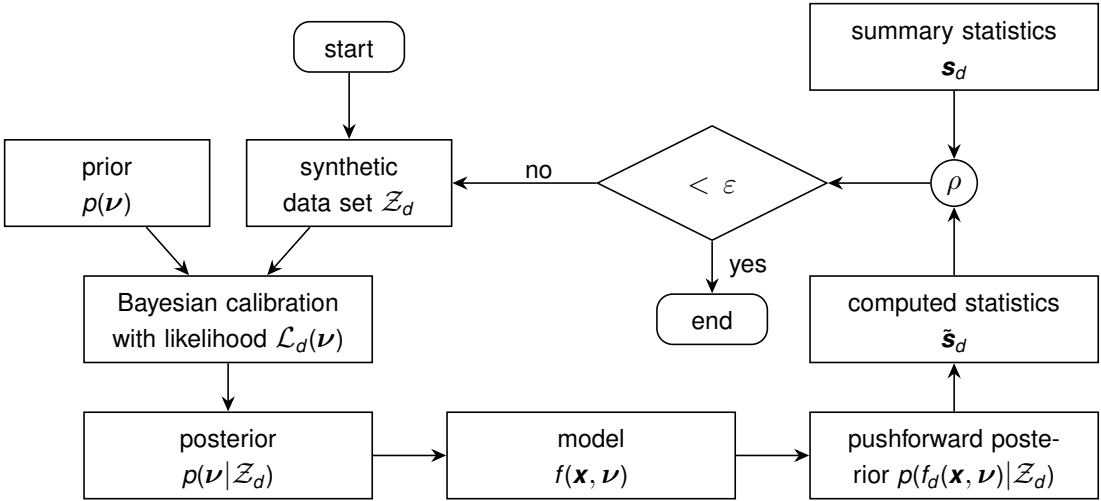
$$\log \mathcal{L}_d(\boldsymbol{\nu}) := -\frac{1}{K_d}\sum_{k=1}^{K_d} \log \mathcal{L}_d^{(k)}(\boldsymbol{\nu})$$

- The pooled posterior can be propagated through the model to obtain the pushforward posterior $p(f_d(\boldsymbol{x}, \boldsymbol{\nu})|\mathcal{Z}_d)$
- A data set is consistent if summary statistics $\tilde{\boldsymbol{s}}_d = \{\tilde{s}_d^{(n)}\}_{n=1}^{N_d}$ extracted from the pushforward posterior agree with the reported summary statistics $\boldsymbol{s}_d = \{s_d^{(n)}\}_{n=1}^{N_d}$, i.e.,

$$\rho(\boldsymbol{s}_d, \tilde{\boldsymbol{s}}_d) < \varepsilon$$

where ρ is a distance metric, and the condition can be satisfied by choosing an appropriate data scaling parameter β_d

Framework for obtaining consistent data sets



Full likelihood construction

- A final likelihood can be constructed by combining the information from different data sources

$$\log \mathcal{L}(\boldsymbol{\nu}) := \sum_{d=1}^{D} \alpha_d \log \mathcal{L}_d(\boldsymbol{\nu})$$

  where $\boldsymbol{\alpha} = (\alpha_d)_{d=1}^{D}$ are positive weights that allow us to express our confidence in the respective experimental data sets
- To account for the different number of measurement locations N_d in each data set, choose

$$\alpha_d := \frac{D N_d^{-1}}{\sum_{d=1}^{D} N_d^{-1}}$$

d	experimental data set	measured quantity	N_d	T0 [K]	Hf_p02
1	Sabioni	$D_{\mathrm{U}}^{\mathrm{eq}}$	10	1973	5.10
2	Davies & Long	$D_{\mathrm{Xe}}^{\mathrm{eq}}$	8	1973	5.10
3	Turnbull	$D_{\mathrm{Xe}}^{\mathrm{irr}}$	16	1973	5.10
4	Miekely & Felix	$D_{\mathrm{Xe}}^{\mathrm{eq}}$	32	1973	6.11
5	stoichiometry	$UO_{2\pm x}$	104	-	-

Polynomial chaos expansion surrogates

- Using the `Centipede` predictions $f_d(\boldsymbol{x}_d^{(n)}, \boldsymbol{\nu})$ directly in the log-likelihood is too expensive
- We propose to replace the model predictions by a polynomial surrogate model, i.e.,

$$f_d(\boldsymbol{x}_d^{(n)}, \boldsymbol{\nu}) \approx \mathcal{P}_d^{(n)}(\boldsymbol{\nu}) := \sum_{\boldsymbol{u} \in \mathcal{I}} c_{\boldsymbol{u}} \phi_{\boldsymbol{u}}(\boldsymbol{\xi}(\boldsymbol{\nu}))$$

where the multi-index $\boldsymbol{u}$ is part of the index set $\mathcal{I} \subseteq \mathbb{N}_0^d$, $c_{\boldsymbol{u}}$ is an unknown coefficient, and $\phi_{\boldsymbol{u}}$ is a multivariate orthogonal polynomial expressed in terms of the i.i.d. random variables $\boldsymbol{\xi} := (\xi_1, \xi_2, \ldots, \xi_d)$

$$\phi_{\boldsymbol{u}}(\boldsymbol{\xi}) := \prod_{j=1}^{s} \phi_{u_j}(\xi_j)$$

- We use weighted iterative Bayesian compressive sensing [Sargsyan et al., 2014] to compute the index set $\mathcal{I}$ and the coefficients $c_{\boldsymbol{u}}$, $\boldsymbol{u} \in \mathcal{I}$

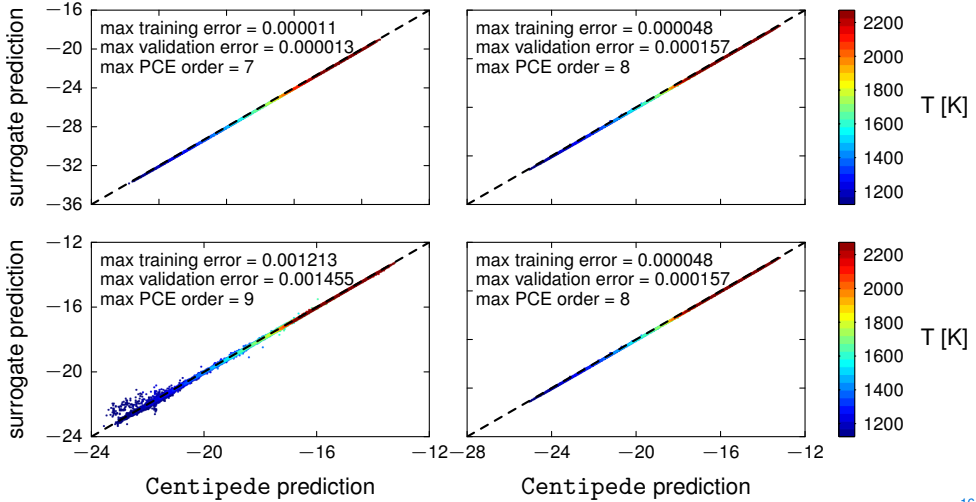
Numerical experiments

- `Centipede` is implemented within the phase-field code `MARMOT` [Tonks et al., 2012] using the `MOOSE` framework [Permann et al., 2016]
- `MOOSE` provides easy access to the C++ Finite Element code `libMesh` [Kirk et al., 2006] and the scalable nonlinear PDE solver `PETSc` [Balay et al., 2022]
- Polynomial surrogates are constructed using `UQTk` [Debusschere, 2017]
- DFI approach implemented in C++ and linked to `UQTk` libraries
- Used global sensitivity analysis (GSA) to identify a subset of 9 important parameters (initially 183 parameters)
- Additionally, two operating conditions
  - `T0`: temperature at which UO_2 is perfectly stoichiometric
  - `Hf_pO2`: the temperature of the oxygen partial pressure

  are allowed to vary separately for each data set (for a total of $7 - 2 + 4 \times 2 =$ 15 parameters to calibrate)

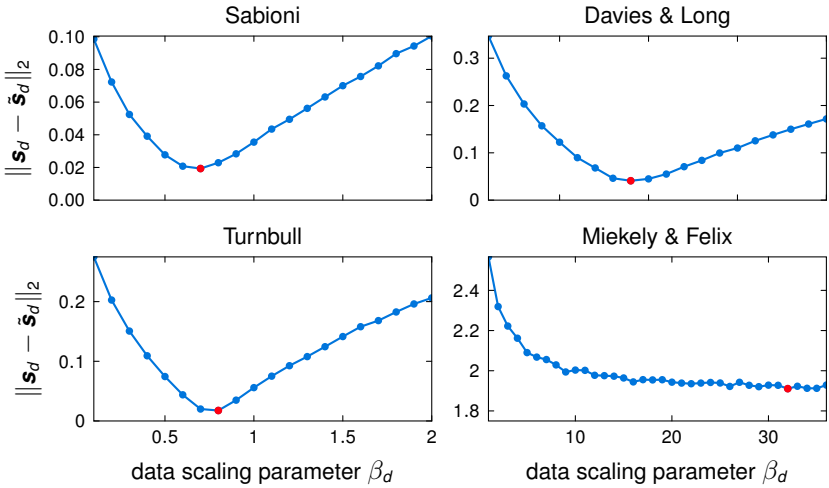
PCE surrogate model construction

- PCE surrogates constructed using 10,000 training points and 1,000 validation points
- PCE basis constructed using BCS shows good surrogate accuracy



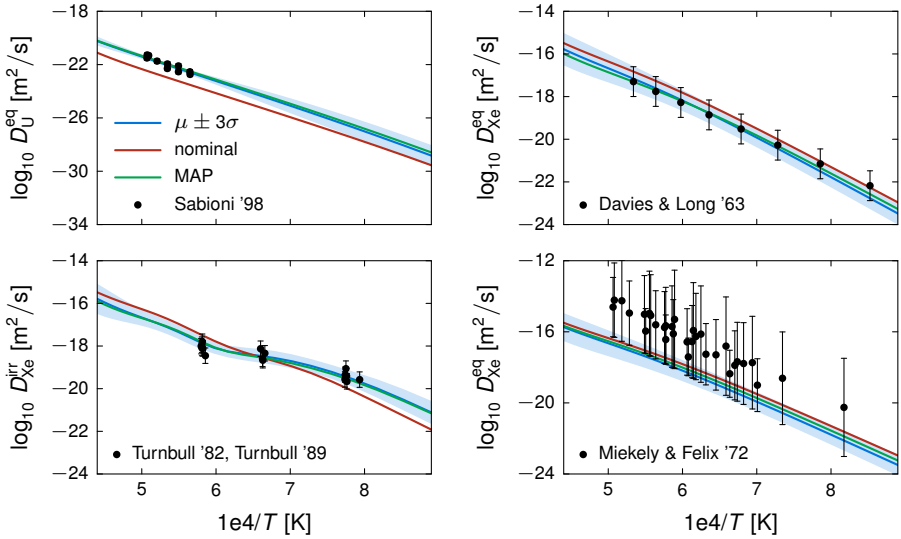
Generating consistent data sets

- Performed a grid search over the standard deviation scaling factors β_d, $d = 1, 2, \ldots, D$ to find consistent data sets where the distance metric $\rho(\boldsymbol{s}_d, \tilde{\boldsymbol{s}}_d) := \left\| \boldsymbol{s}_d - \tilde{\boldsymbol{s}}_d \right\|_2$



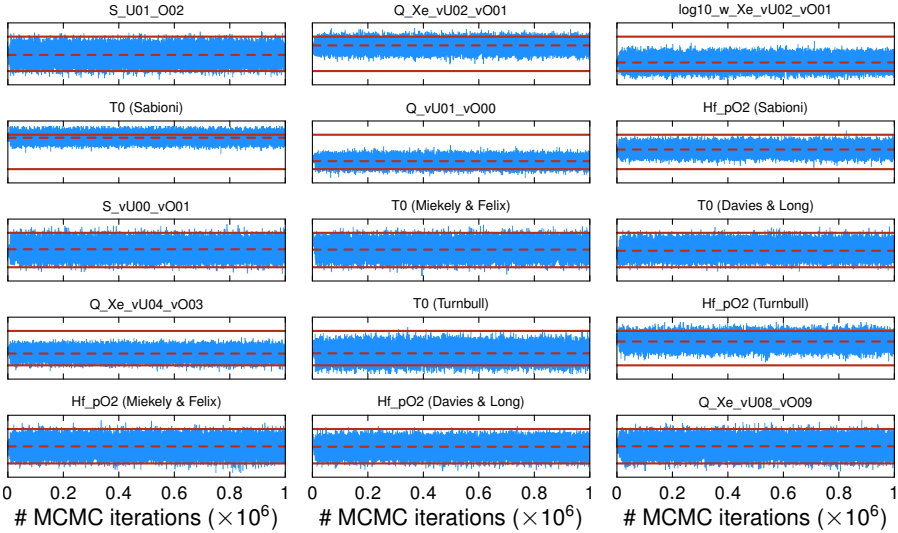
Pushforward posterior

- Calibrated pushforward posterior diffusivities show good match to the data



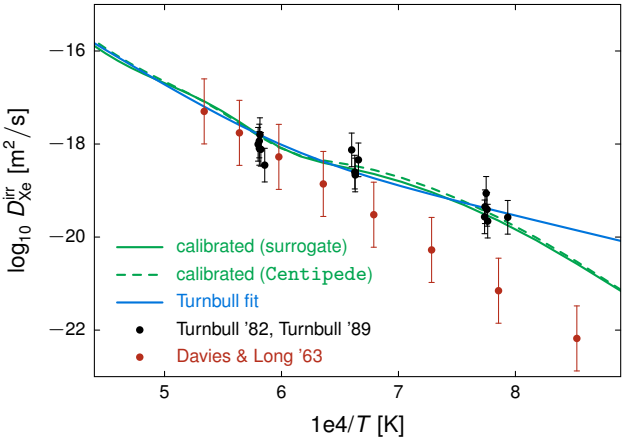
MCMC chains

- Good mixing of MCMC chains



Turnbull correlation

- Comparison of the pushforward MAP prediction for Xe irradiation computed with surrogate model and actual `Centipede` code shows good agreement
- Comparison to Turnbull fit (best available experimental fit from literature)



Conclusion and future work

- We devised a Bayesian inference strategy to estimate the parameters in the cluster dynamics code `Centipede`, where the reported data on UO_2 diffusivity and stoichiometry is available as mean values with an associated measurement error
- Our calibration strategy involves synthetic data sets that are consistent in the sense that the statistics computed from the pushforward posterior agree with the reported summary statistics
- We used our Bayesian inference strategy to calibrate 9 `Centipede` parameters against 5 different experimental data sets with 208 measurements in total
- The predicted pushforward posterior is in good agreement with the experimental data
- In future efforts, we will work on improving the surrogate accuracy to be able to include more parameters in the study (original model consists of 183 parameters)

Conclusion and future work

- We devised a Bayesian inference strategy to estimate the parameters in the cluster dynamics code `Centipede`, where the reported data on UO_2 diffusivity and stoichiometry is available as mean values with an associated measurement error
- Our calibration strategy involves synthetic data sets that are consistent in the sense that the statistics computed from the pushforward posterior agree with the reported summary statistics
- We used our Bayesian inference strategy to calibrate 9 `Centipede` parameters against 5 different experimental data sets with 208 measurements in total
- The predicted pushforward posterior is in good agreement with the experimental data
- In future efforts, we will work on improving the surrogate accuracy to be able to include more parameters in the study (original model consists of 183 parameters)

Thank you

References

- Andersson, D., Garcia, P., Liu, X.-Y., Pastore, G., Tonks, M., Millett, P., Dorado, B., Gaston, D., Andrs, D., Williamson, R., Martineau, R., Uberuaga, B., and Stanek, C. (2014). Atomistic modeling of intrinsic and radiation-enhanced fission gas (Xe) diffusion in $UO_{2\pm x}$: Implications for nuclear fuel performance modeling. *J. Nucl. Mater.*, 451:225–242.
- Balay, S., Abhyankar, S., Adams, M. F., Benson, S., Brown, J., Brune, P., Buschelman, K., Constantinescu, E. M., Dalcin, L., Dener, A., Eijkhout, V., Gropp, W. D., Hapla, V., Isaac, T., Jolivet, P., Karpeev, D., Kaushik, D., Knepley, M. G., Kong, F., Kruger, S., May, D. A., McInnes, L. C., Mills, R. T., Mitchell, L., Munson, T., Roman, J. E., Rupp, K., Sanan, P., Sarich, J., Smith, B. F., Zampini, S., Zhang, H., Zhang, H., and Zhang, J. (2022). PETSc Web page.
- Berry, R. D., Najm, H. N., Debusschere, B. J., Marzouk, Y. M., and Adalsteinsson, H. (2012). Data-free inference of the joint distribution of uncertain model parameters. *Journal of Computational Physics*, 231(5):2180–2198.
- Chowdhary, K. and Najm, H. N. (2016). Data free inference with processed data products. *Statistics and Computing*, 26(1):149–169.
- Debusschere, B. (2017). Uncertainty quantification toolkit. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
- Kirk, B. S., Peterson, J. W., Stogner, R. H., and Carey, G. F. (2006). libmesh: a c++ library for parallel adaptive mesh refinement/coarsening simulations. *Engineering with Computers*, 22(3):237–254.

References (cont.)

- Matthews, C., Perriot, R., Cooper, M., Stanek, C. R., and Andersson, D. A. (2020). Cluster dynamics simulation of xenon diffusion during irradiation in UO_2. *Journal of Nuclear Materials*, 540:152326.
- Permann, C. J., Tonks, M. R., Fromm, B., and Gaston, D. R. (2016). Order parameter re-mapping algorithm for 3d phase field model of grain growth using fem. *Comp. Mater. Sci.*, 115:18–25.
- Perriot, R., Matthews, C., Cooper, M. W., Uberuaga, B. P., Stanek, C. R., and Andersson, D. A. (2019). Atomistic modeling of out-of-pile xenon diffusion by vacancy clusters in UO_2. *J. Nucl. Mater.*, 520:96–109.
- Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., and Thornton, P. (2014). Dimensionality reduction for complex models via Bayesian compressive sensing. *International Journal for Uncertainty Quantification*, 4(1).
- Tonks, M. R., Gaston, D., Millett, P. C., Andrs, D., and Talbot, P. (2012). An object-oriented finite element framework for multiphysics phase field simulations. *Computational Materials Science*, 51(1):20–29.