EUROfusion

Bayesian optimization for efficient parameter space coverage with computationally demanding simulations in fusion energy applications

A.E. Järvinen, L. Acerbi, E. Amnell, A. Bharti, R.M. Churchill, G. Clarté, T. Fonghetti, C.S. Furia, T. Fülöp, A. Ho, M. Hoppe, A. Kit, E. Nardon, S. Silburn, and JET Contributors

Workshop on AI for Accelerating Fusion and Plasma Science

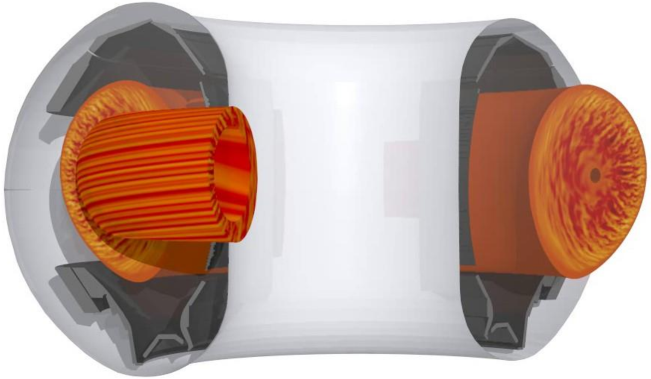
November 29th 2023

VTT FinnFusion PPPL CHALMERS UNIVERSITY OF TECHNOLOGY Aalto University BUSINESS FINLAND Lawrence Livermore National Laboratory CSC EPFL ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE SWISS PLASMA CENTER Research Council of Finland

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

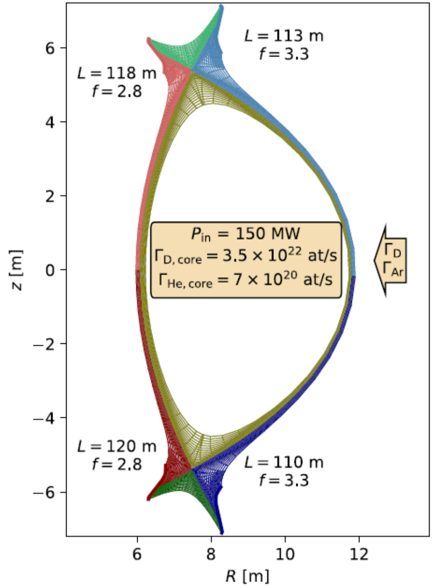
Computationally expensive models with uncertain code parameters are ubiquituos in magnetic confinement fusion energy research

Fusion performance *~ plasma turbulence*



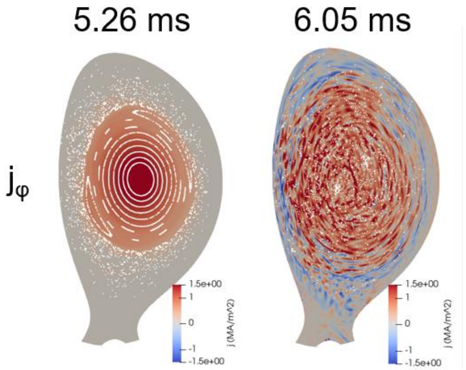
https://genecode.org/
F. Jenko, et al. PoP 2000

Power exhaust and heat load control ~ *plasma-neutral-materials interactions*



L. Aho-Mantila, et al. Nucl. Mat. Ene. 2021
https://doi.org/10.1016/j.nme.2020.100886
SOLPS-ITER: S. Wiesen, et al. J. Nucl. Mat. 2015
https://doi.org/10.1016/j.jnucmat.2014.10.012

Prediction of off-normal events and disruptions ~ *rapidly evolving thermal and relativistic populations & fields*



E. Nardon, et al. Nucl. Fusion 2023
https://doi.org/10.1088/1741-4326/acc417
JOREK: Hoeltzl, et al. Nucl. Fusion 2021
https://doi.org/10.1088/1741-4326/abf99f

Due to the uncertain code parameters, a single forward simulation is rarely sufficient to actually quantify the prediction uncertainty

Common problem statement

Given limited resources of wall clock time or HPC hours, it is not tractable to simply simulate all possible model configurations. Therefore, the question is how to best use the available resources to optimally quantify the uncertainties?

Due to the uncertain code parameters, a single forward simulation is rarely sufficient to actually quantify the prediction uncertainty

Common problem statement

Given limited resources of wall clock time or HPC hours, it is not tractable to simply simulate all possible model configurations. Therefore, the question is how to best use the available resources to optimally quantify the uncertainties?

Forward UQ ~ Uncertainty propagation

Input		Model		Output
Probability distributions of code parameters $\boldsymbol{\theta}$	→	**Computationally expensive model $y = f(\theta, \Phi, \varepsilon)$**	→	**Distributions of model predictions y**

Due to the uncertain code parameters, a single forward simulation is rarely sufficient to actually quantify the prediction uncertainty

Common problem statement

Given limited resources of wall clock time or HPC hours, it is not tractable to simply simulate all possible model configurations. Therefore, the question is how to best use the available resources to optimally quantify the uncertainties?

Forward UQ ~ Uncertainty propagation

Input	Model	Output
Probability distributions of code parameters $\boldsymbol{\theta}$	**Computationally expensive model $y = f(\theta, \Phi, \varepsilon)$**	**Distributions of model predictions y**

Inverse UQ ~ Parameter calibration

Due to the uncertain code parameters, a single forward simulation is rarely sufficient to actually quantify the prediction uncertainty

Common problem statement

Given limited resources of wall clock time or HPC hours, it is not tractable to simply simulate all possible model configurations. Therefore, the question is how to best use the available resources to optimally quantify the uncertainties?

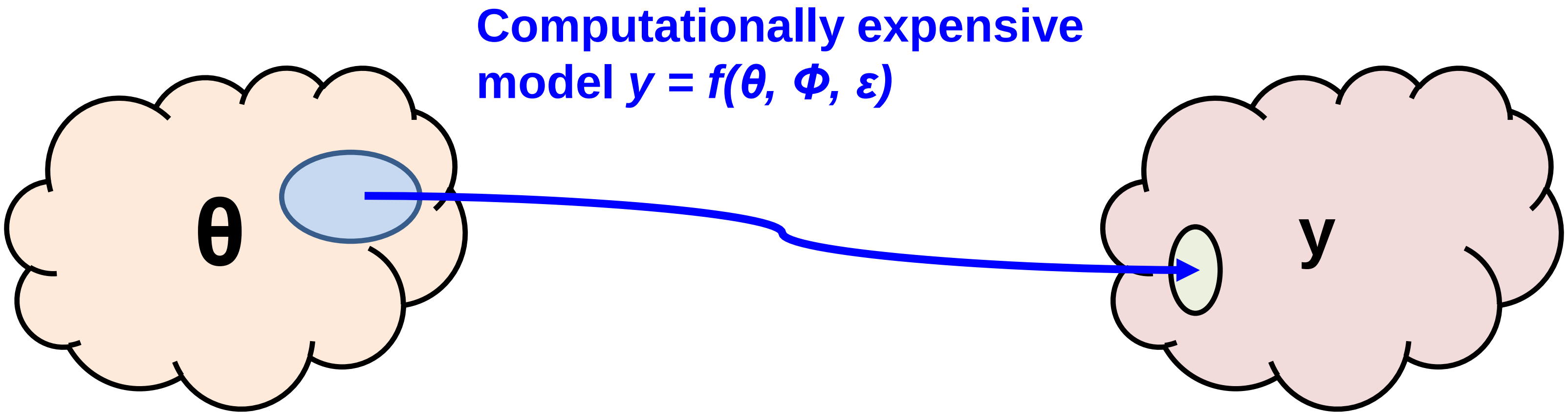
Forward UQ ~ Uncertainty propagation

Input		Model		Output
Probability distributions of code parameters θ	→ / ←	Computationally expensive model $y = f(\theta, \Phi, \varepsilon)$	→ / ←	Distributions of model predictions y

Inverse UQ ~ Parameter calibration

- **Forward UQ is typically needed for addressing the confidence interval for a model prediction: Given the input & code uncertainties, how likely it is that the prediction falls within the tolerance of the system, such as heat load limit on divertor plate?**
- **Inverse UQ is typically needed for model validation: Is the model able to reproduce experimental observations with physically valid input parameters?**

The inverse mapping for θ is defined only implicitly through the forward model

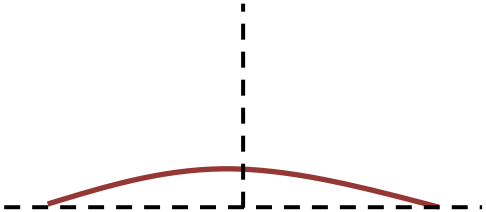


- **With computationally expensive numerical models:**
  - **Given θ, y can be computed with the forward model (potentially including a stochastic term)**
  - **Given y, there is no direct computational model to determine θ**
  - **We will use Bayesian inference to establish probability distributions for θ, given samples of (θ, y)**

Bayesian inference (BI) algorithms provide a principled approach to quantify the uncertainty for the state of the investigated system, given available data

Prior belief

$\pi(\theta)$

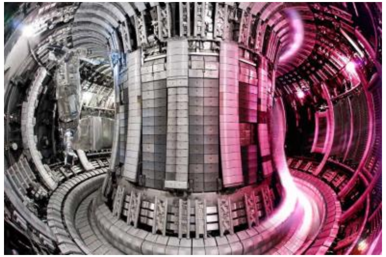


State of high uncertainty

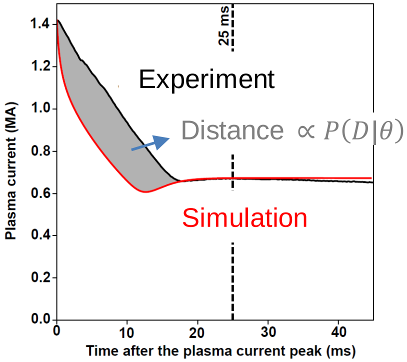
Likelihood

$P(D|\theta)$

Measured data D

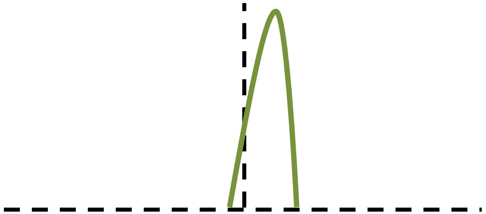


Simulation, input θ



Posterior belief

$\pi(\theta|D) \propto P(D|\theta)\pi(\theta)$



State of low uncertainty

Bayesian inference (BI) algorithms provide a principled approach to quantify the uncertainty for the state of the investigated system, given available data

Likelihood

$P(D|\theta)$

Measured data D

Prior belief

$\pi(\theta)$

Posterior belief

$\pi(\theta|D) \propto P(D|\theta)\pi(\theta)$

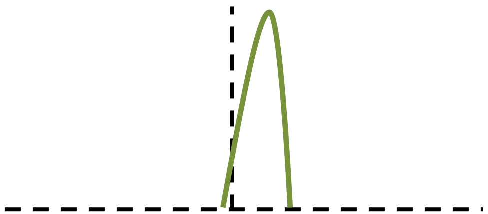
- **Typically the likelihood is not available in closed-form when operating with complicated models**
- **Approximate Bayesian Computation (ABC) or likelihood-free techniques are a way to address these intractable likelihood problems [J.M. Marin, et al. Statistics and Computing 2012 and references therein, https://doi.org/10.1007/s11222-011-9288-2]**
- **With computationally expensive models, data-efficiency is key to complete the Bayesian inference task with acceptable overall computational resources → How to efficiently find the combinations of input parameters that best reproduce the experimental observations?**
- **The forward model is often expensive, not necessarily bijective, does not provide first or second order derivatives easily, and the optimization challenge is expected to be non-convex**

Bayesian optimization is a class of optimization methods focused on finding a global optimum of a forward model within a search space

Bayesian Inference

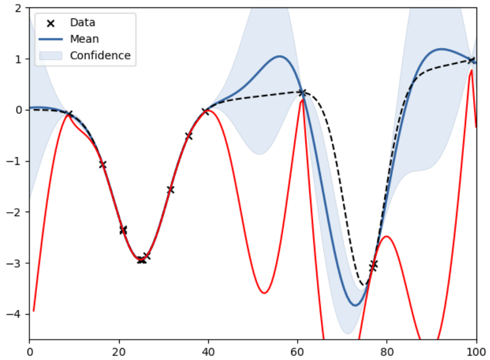
Establish a posterior distribution for uncertain parameters, given evidence *π(θ|D)*

$$\pi(\theta|D) \propto P(D|\theta)\pi(\theta)$$



Bayesian Optimization (BO)

Conduct Bayesian Inference in the space of objective functions to data-efficiently find the global optimum



The ability to optimize expensive ”black-box” functions without access to derivatives makes BO very powerful.

A Bayesian optimization algorithm has two main components

1. **A probabilistic (surrogate) model of the objective function**
2. **An acquisition function**

Steps of a Bayesian optimization algorithm

Apply acquisition function to get the next sample point: x_i

→

Sample the objective function: $y_i = f(x_i)$

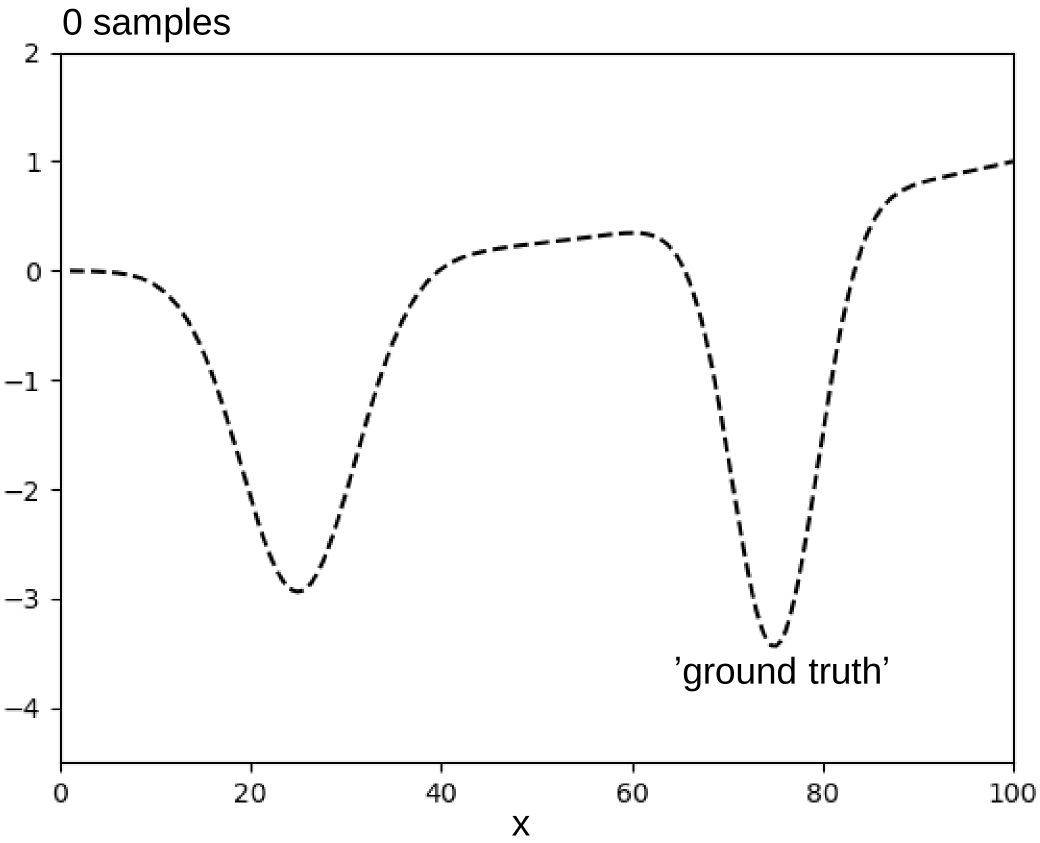
→

Augment the data set: $\{(x_1, y_1),\ldots, (x_i, y_i)\}$

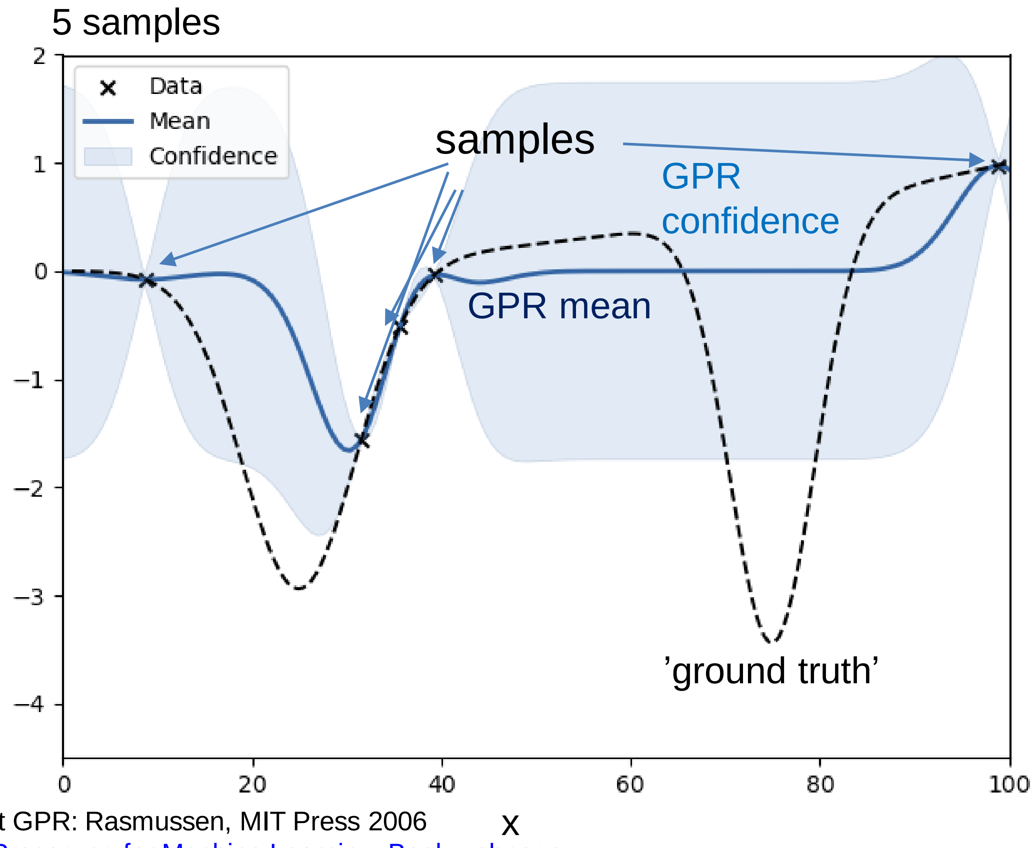
Update the surrogate model

See, e.g. [E. Brochu et al. arXiv:1012.2599 https://arxiv.org/abs//1012.2599]

1D Bayesian optimization example – Find the minimum



1D Bayesian optimization example – Find the minimum



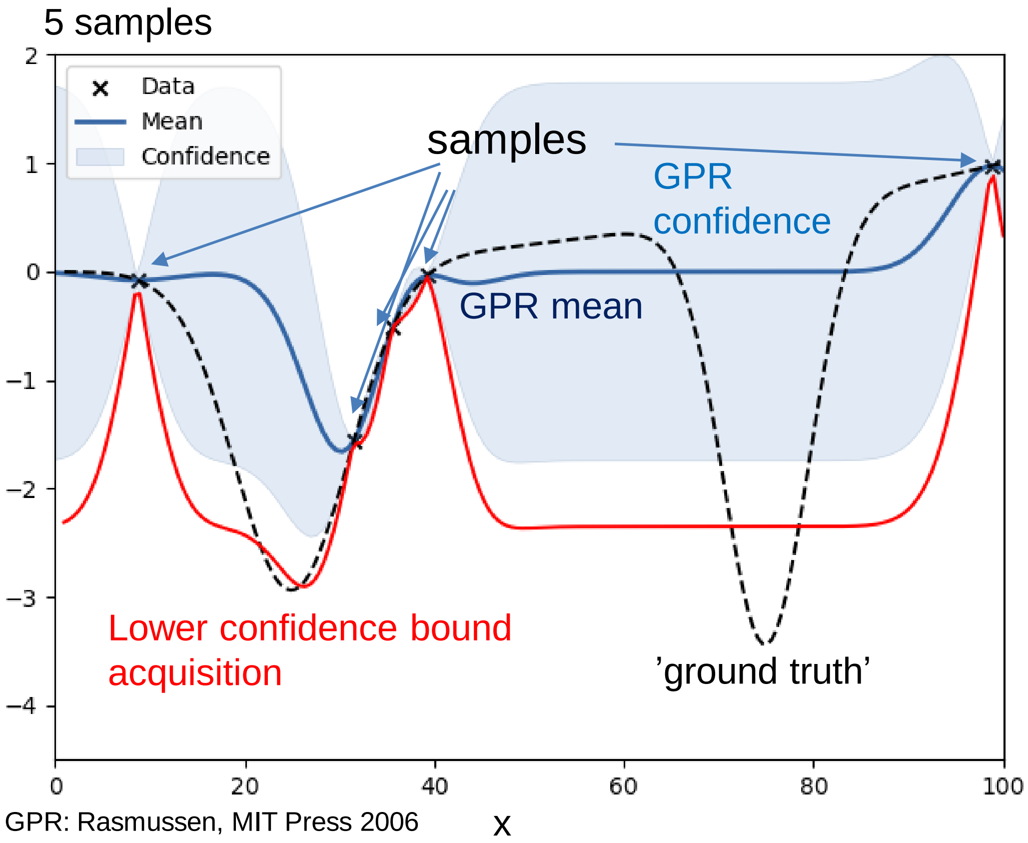
Gaussian process regression as probabilistic surrogate model

- Convenient non-parametric probabilistic regression when the amount of data is limited
- ’Learn’ model hyperparameters from data → gradient based optimization of marginal log-likelihood or through integration over the model hyperparameters (e.g. MCMC)

More about GPR: Rasmussen, MIT Press 2006

[Gaussian Processes for Machine Learning: Book webpage]

1D Bayesian optimization example – Find the minimum



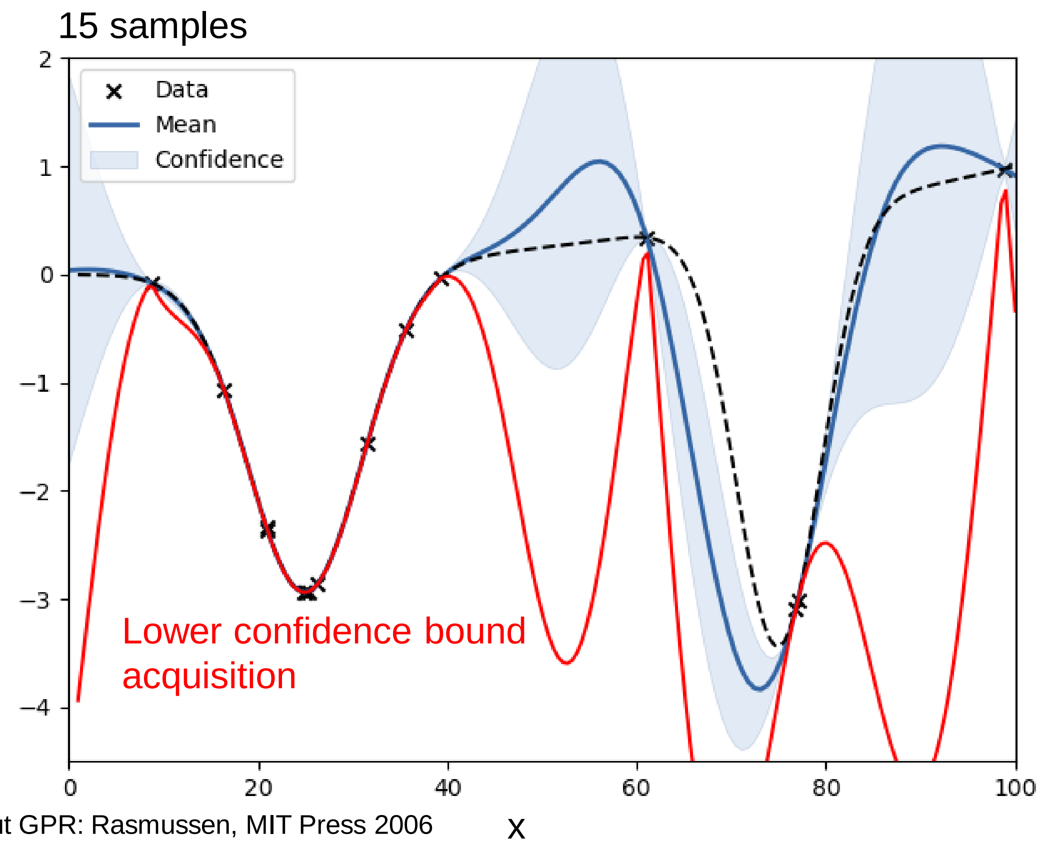
Gaussian process regression as probabilistic surrogate model

- Convenient non-parametric probabilistic regression when the amount of data is limited
- 'Learn' model hyperparameters from data → gradient based optimization of marginal log-likelihood or through integration over the model hyperparameters (e.g. MCMC)

Acquisition function uses the GPR mean and confidence to recommend new query points

More about GPR: Rasmussen, MIT Press 2006
[Gaussian Processes for Machine Learning: Book webpage]

1D Bayesian optimization example – Find the minimum



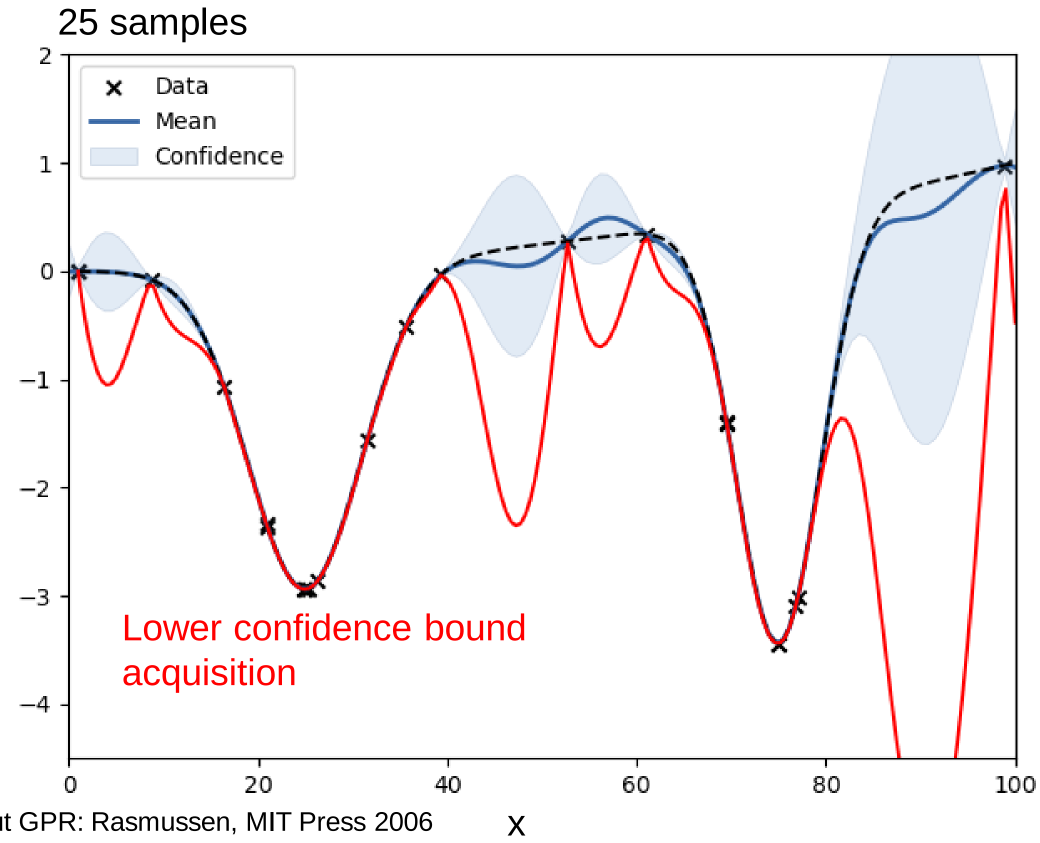
Gaussian process regression as probabilistic surrogate model

- Convenient non-parametric probabilistic regression when the amount of data is limited
- ’Learn’ model hyperparameters from data → gradient based optimization of marginal log-likelihood or through integration over the model hyperparameters (e.g. MCMC)

Acquisition function uses the GPR mean and confidence to recommend new query points

More about GPR: Rasmussen, MIT Press 2006
[Gaussian Processes for Machine Learning: Book webpage]

1D Bayesian optimization example – Find the minimum



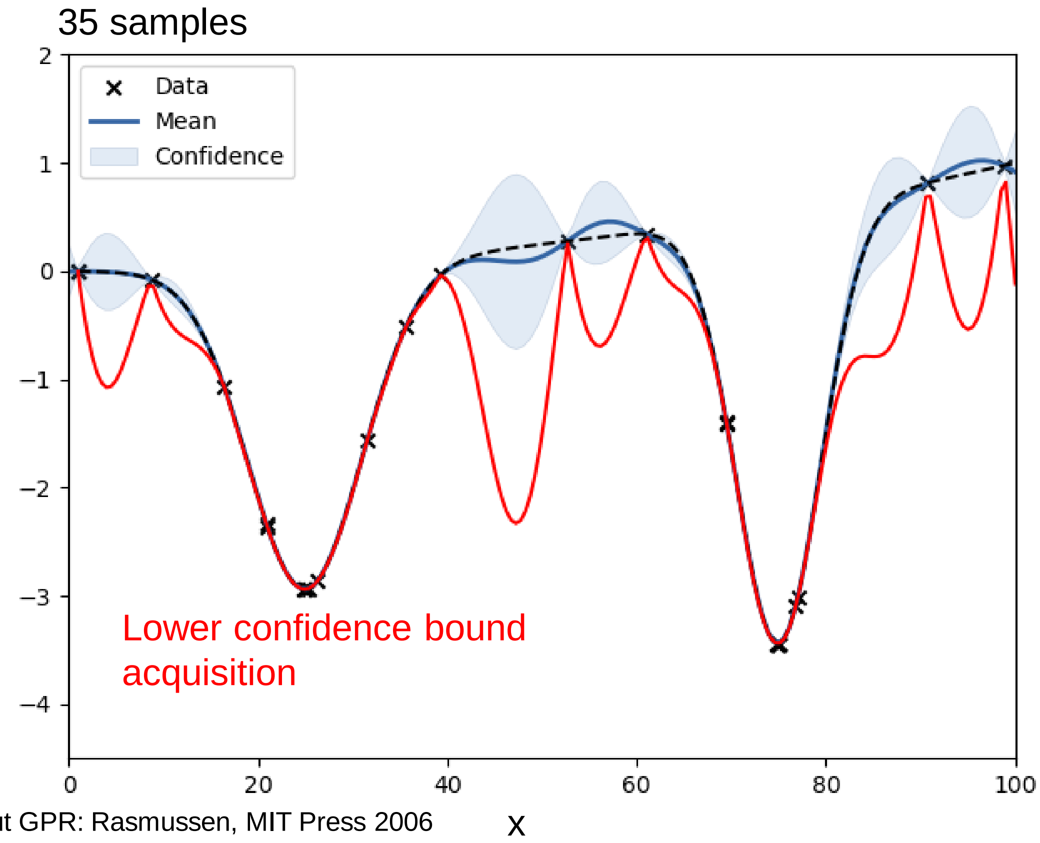
Gaussian process regression as probabilistic surrogate model

- Convenient non-parametric probabilistic regression when the amount of data is limited
- ’Learn’ model hyperparameters from data → gradient based optimization of marginal log-likelihood or through integration over the model hyperparameters (e.g. MCMC)

Acquisition function uses the GPR mean and confidence to recommend new query points

More about GPR: Rasmussen, MIT Press 2006
[Gaussian Processes for Machine Learning: Book webpage]

1D Bayesian optimization example – Find the minimum



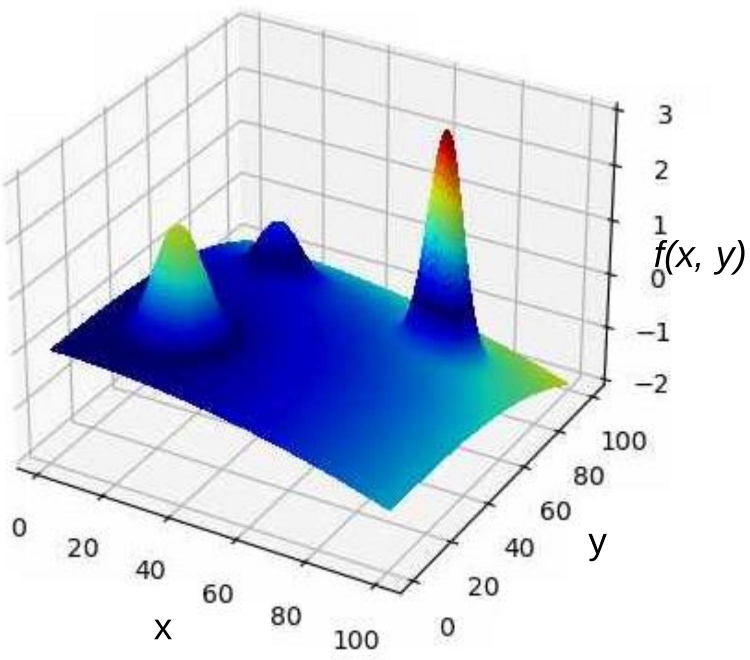
Gaussian process regression as probabilistic surrogate model

- Convenient non-parametric probabilistic regression when the amount of data is limited
- ’Learn’ model hyperparameters from data → gradient based optimization of marginal log-likelihood or through integration over the model hyperparameters (e.g. MCMC)

Acquisition function uses the GPR mean and confidence to recommend new query points

More about GPR: Rasmussen, MIT Press 2006
[Gaussian Processes for Machine Learning: Book webpage]

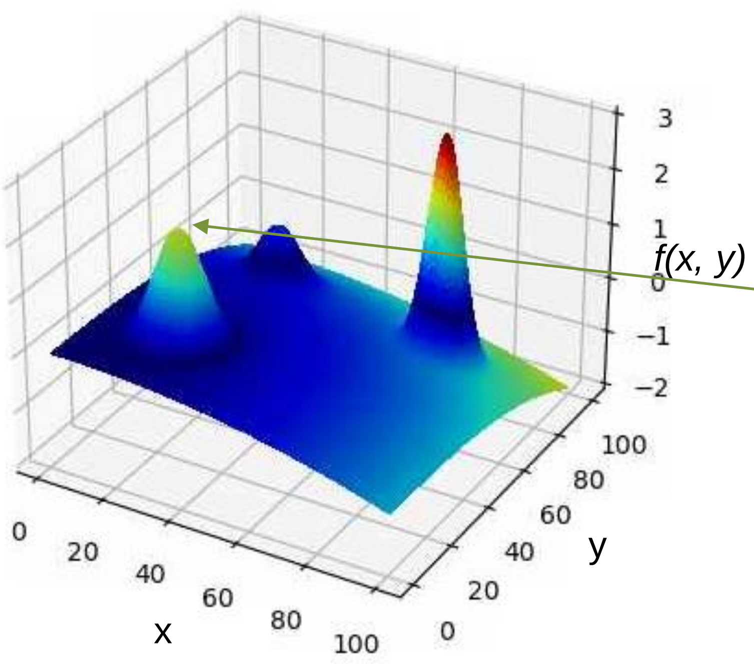
Bayesian optimization in a challenging multimodal landscape



Goal: Find maximum of f(x,y)

- Consider an analytical 2D function with 3 Gaussian peaks and an underlying slope as ground truth

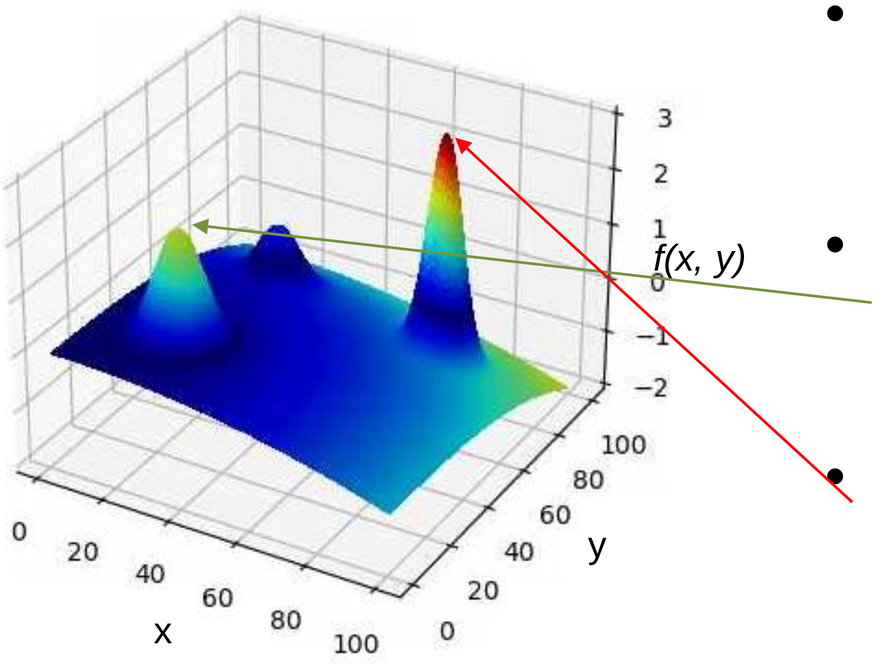
Bayesian optimization in a challenging multimodal landscape



Goal: Find maximum of f(x,y)

- Consider an analytical 2D function with 3 Gaussian peaks and an underlying slope as ground truth
- Following the gradient of the landscape is most likely to lead to the second best peak

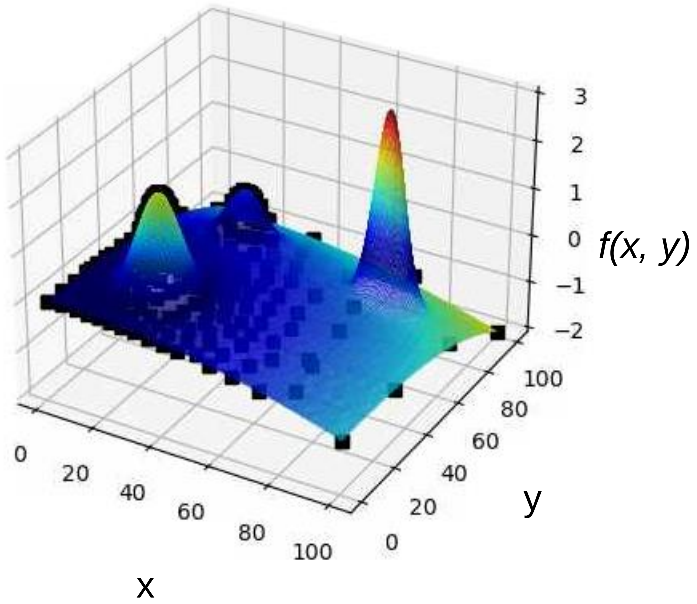
Bayesian optimization in a challenging multimodal landscape

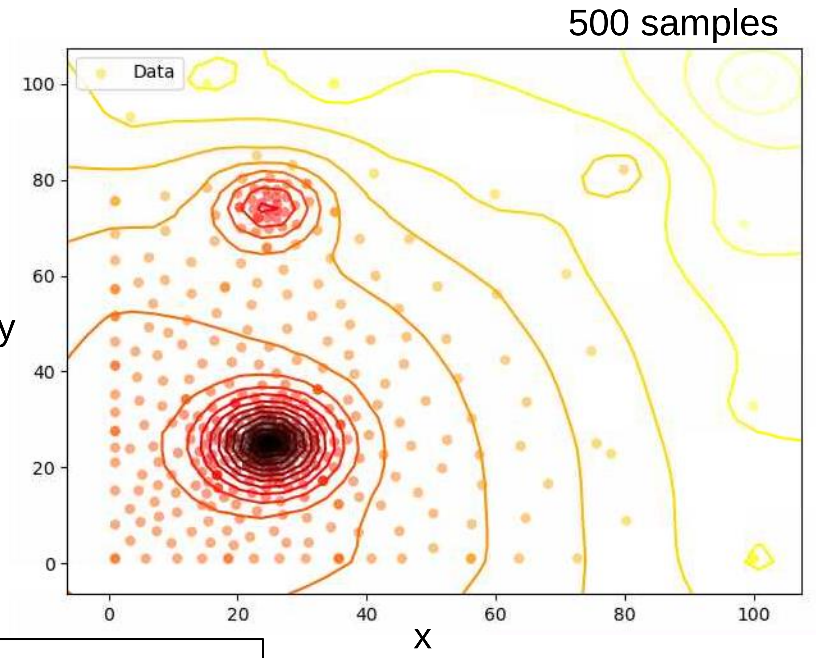


Goal: Find maximum of f(x,y)

- Consider an analytical 2D function with 3 Gaussian peaks and an underlying slope as ground truth
- Following the gradient of the landscape is most likely to lead to the second best peak
- In order to find the global optimum, the algorithm must explore in region that looks not attractive at first

Bayesian optimization in a challenging multimodal landscape

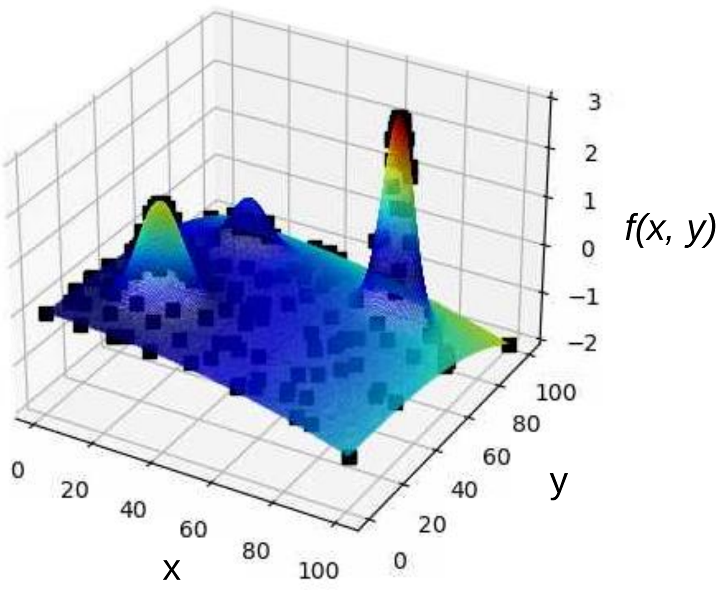


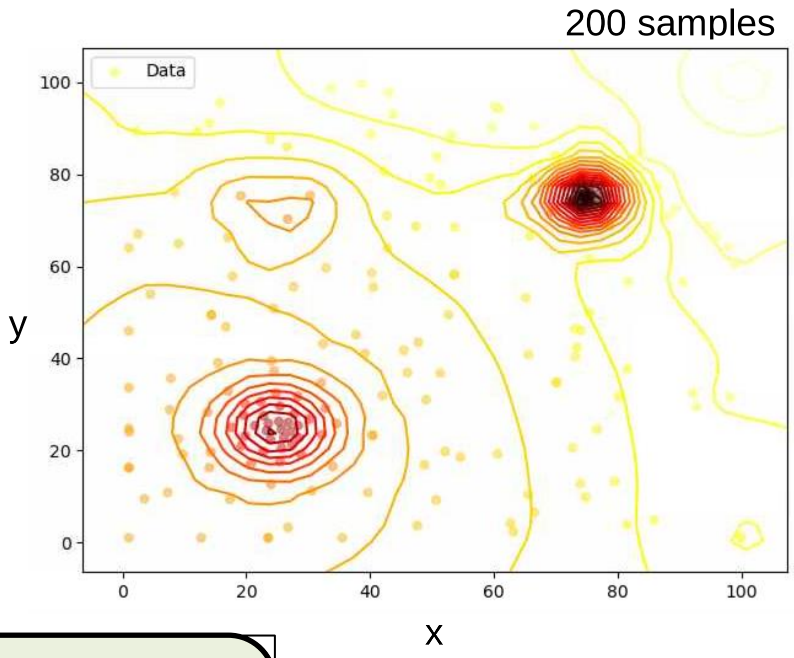


- **With a simple BO setup, the algorithm stops exploration too early and converges to the second best optimum**
- **After 500 samples, no sample near the global optimum**
- **This would be difficult to diagnose in higher dimensionality search space**

- GP with Rational Quadratic kernel
- Upper Confidence Bound Acquisition

Bayesian optimization in a challenging multimodal landscape





- **A trivial way to add exploration is to throw in random samples (not necessarily optimal strategy)**
- **Several ways to encourage exploration exist through designs of the acquisition functions and surrogate models**

- GP with Rational Quadratic kernel
- Upper Confidence Bound Acquisition
  + 50% fully random samples

Batch BO needed for many practical BO application in model validation exercises

- Many numerical models in fusion energy research consume several hours or days of wall clock time per simulation → collecting a few 100 samples could take years
- To populate the search space within an acceptable wall clock time, samples need to be collected in batch parallel to each other → a batch BO approach is needed
- In Batch BO, a batch of search points is queried at once from the acquisition function
- Batch acquisition functions can be categorized as*:
  - Value-estimators: *use a value function g(x) to predict a place-holder value while waiting for the sample*
  - Explorers: *use sequential acquisition for the initial point and fill rest of the batch with exploration*
  - Stochastic: *Draw a stochastic sample rather than just optimum from the acquisition function*
  - Penalizers: *Apply penalty to sample points too close to each other*
  - Mode-finders: *Acquire samples in the modes*
  - Others: *Any other strategy*

* N. Hunt, MSc Thesis, 'Batch Bayesian Optimization', MIT 2020

It is quite common for the simulations to fail to converge with some parameter configurations

Dealing with sample failure is not trivial:

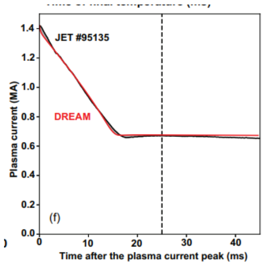
- **Ignoring failed samples is likely to lead to acquisition function recommending sampling the failure region again → inefficient**
- **Representing a failed sample with a place-holder value that is far from optimal introduces a discontinuity in the data that is challenging for the probabilistic surrogate**

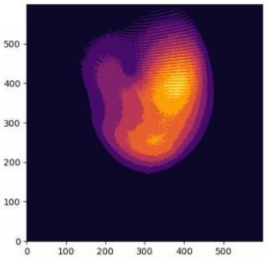
Chakrabarty, et al. IEEE Press 2021 *https://doi.org/10.1109/SMC52423.2021.9658893* **introduced a relatively elegant approach in the paper 'Simulation Failure Robust Bayesian Optimization for Estimating Black-Box Model Parameters', which adds a failure probability classifier**

1. **A failure classifier model to estimate the likelihood of success & failure**
2. **Surrogate model for the objective function (as in standard BO)**
3. **An acquisition function that incorporates the failure probability into the BO framework**

A portfolio of BI and BO tasks pursued in close connection to the EUROfusion Advanced Computing Hub (05) in Finland

DREAM runaway electron simulations





A.E. Järvinen, et al. J. Plasma Phys. 2022
https://doi.org/10.1017/S0022377822001210

Validation of integrared modelling

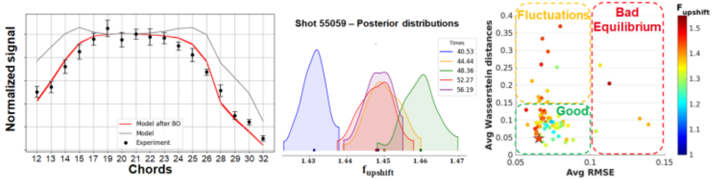
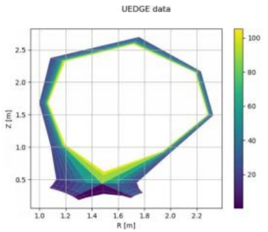
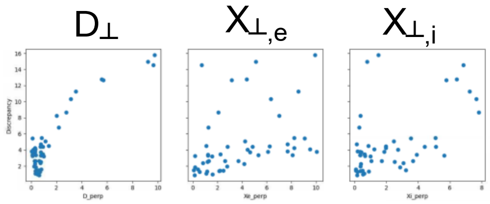


Fig 10. Single time optimization : (a) Forward UQ comparing with default parameters based on posterior distribution showed in (b). (c) Plateau-averaged performance of BO algorithm on the whole hot branch database and resulting fupshift as a color scale

T. Fonghetti, E. Amnell, et al.

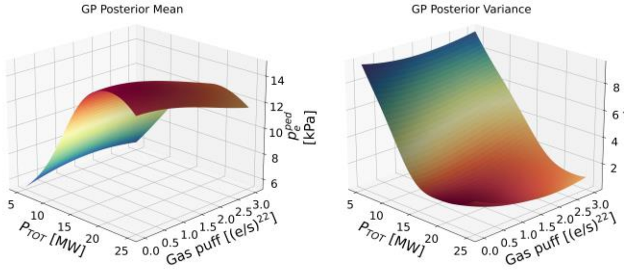
UEDGE SOL model calibration





same UEDGE case as in C.S. Furia, R.M. Churchill 2022 *Plasma Phys. Control. Fusion* **64** 104003
See PiAI seminar October 24, 2022

BO for experiment design



A. Kit,
G. Clarte, et al. ICDDPS-4

Bayesian optimization explored for current quench simulations for a JET plasma with an induced RE beam [Reux PRL 2021]

- Argon massive gas injection 25.4 ms before the current spike
- CQ simulated with DREAM [Hoppe CPC 2021] using fluid model for REs
  - ➢ Instant thermal quench assumed with post TQ T_e as an input parameter
  - ➢ RE seed profile given as input
  - ➢ Amount of injected argon known, but the fraction that is assimilated is an uncertain input parameter
  - ➢ Characteristic wall time (wall conductivity) given as an uncertain input parameter

J. Plasma Phys. (2022), *vol.* 88, 905880612 1
Published by Cambridge University Press

doi:10.1017/S0022377822001210

Bayesian approach for validation of runaway electron simulations

A.E. Järvinen[1,2,†], T. Fülöp[3], E. Hirvijoki[4], M. Hoppe[5], A. Kit[2], J. Åström[2,6] and JET Contributors‡

[1]VTT Technical Research Centre of Finland, FI-02044 VTT, Finland
[2]University of Helsinki, FI-00014 Helsinki, Finland
[3]Chalmers University of Technology, SE-412 96 Göteborg, Sweden
[4]Aalto University, FI-00076 AALTO, Finland
[5]Ecole Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center, CH-1015 Lausanne, Switzerland
[6]CSC-IT Center for Science, FI-02101 Espoo, Finland

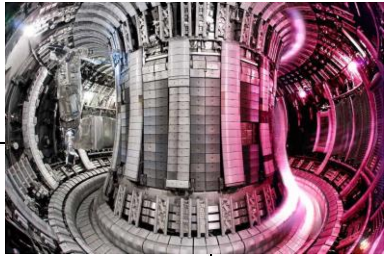
(Received 1 August 2022; revised 11 November 2022; accepted 14 November 2022)

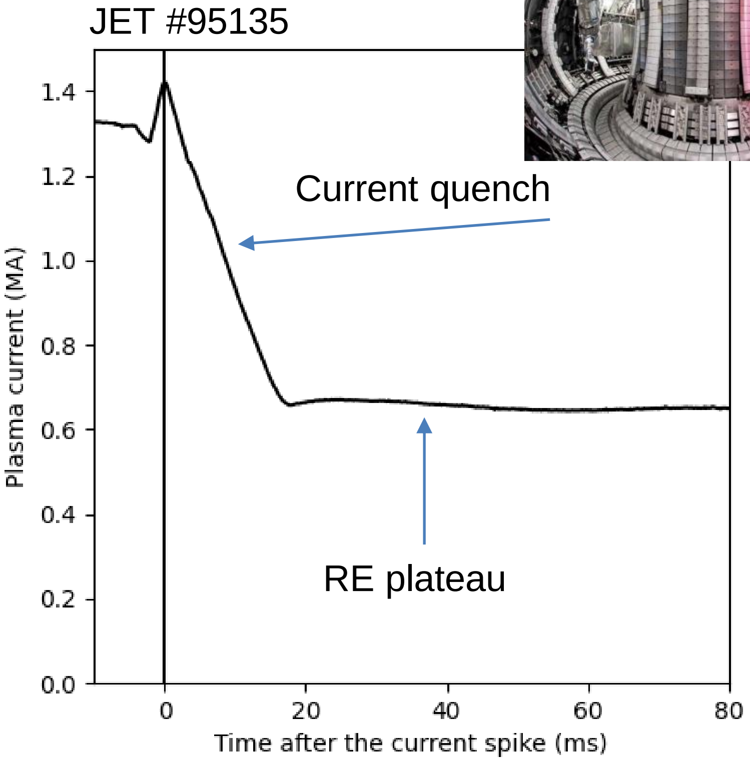
Plasma-terminating disruptions in future fusion reactors may result in conversion of the initial current to a relativistic runaway electron beam. Validated predictive tools are required to optimise the scenarios and mitigation actuators to avoid the excessive damage that can be caused by such events. Many of the simulation tools applied in fusion energy research require the user to specify input parameters that are not constrained by the available experimental information. The conventional approach, where an expert modeller calibrates these input parameters based on domain knowledge, is prone to lead to an intractable validation challenge without systematic uncertainty quantification. Bayesian inference algorithms offer a promising alternative approach that naturally includes uncertainty quantification and is less subject to user bias in choosing the input parameters. The main challenge in using these methods is the computational cost of simulating enough samples to construct the posterior distributions for the uncertain input parameters. This challenge can be overcome by combining probabilistic surrogate modelling, such as Gaussian process regression, with Bayesian optimisation, which can reduce the number of required simulations by several orders of magnitude. Here, we implement this type of Bayesian optimisation framework for a model for analysis of disruption runaway electrons, and explore for simulations of current quench in a JET plasma discharge with an argon induced disruption. We use this proof-of-principle framework to explore the optimum input parameters with uncertainties in optimisation tasks ranging from one to seven

Published in A.E. Järvinen, et al.
Journal of Plasma Physics **88** (2022) 905880612
https://doi.org/10.1017/S0022377822001210

Bayesian optimization explored for current quench simulations for a JET plasma with an induced RE beam [Reux PRL 2021]

- Argon massive gas injection 25.4 ms before the current spike
- CQ simulated with DREAM [Hoppe CPC 2021] using fluid model for REs
  - Instant thermal quench assumed with post TQ T_e as an input parameter
  - RE seed profile given as input
  - Amount of injected argon known, but the fraction that is assimilated is an uncertain input parameter
  - Characteristic wall time (wall conductivity) given as an uncertain input parameter

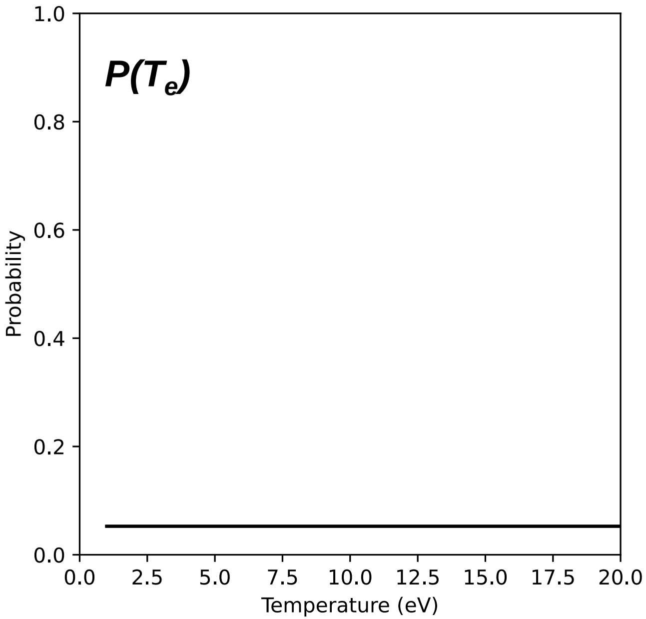




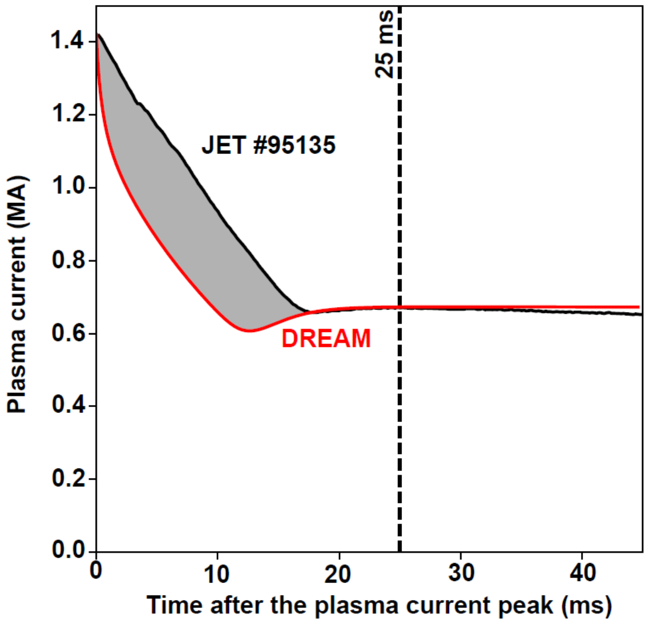
Published in A.E. Järvinen, et al.
Journal of Plasma Physics **88** (2022) 905880612
https://doi.org/10.1017/S0022377822001210

1D Example: Find the post thermal quench T_e that minimizes the discrepancy between the measured and predicted I_p during the current quench

Simple prior $P(T_e)$ = U(1.0, 20)

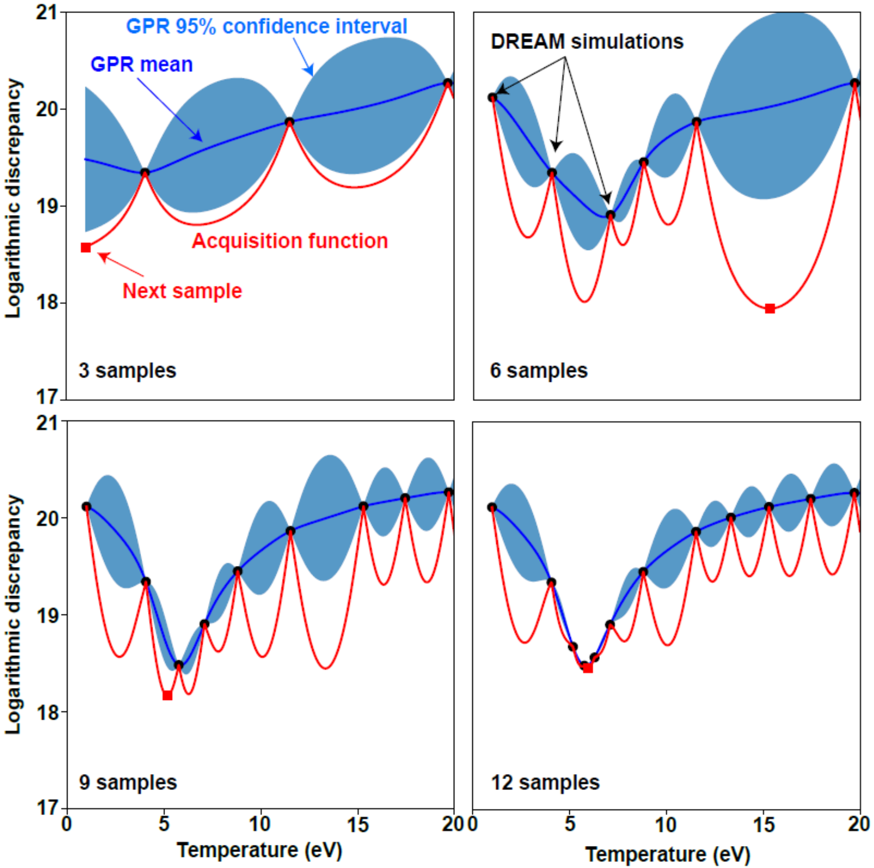


Evaluate the discrepancy as L_1-norm between the measured and predicted I_p



- Bayesian Optimization of Likelihood-Free Inference (BOLFI) method of Engine for Likelihood-Free Inference (ELFI) Python software package is used [Lintusaari, JMLR, 2018 & Gutmann, JMLR, 2016]

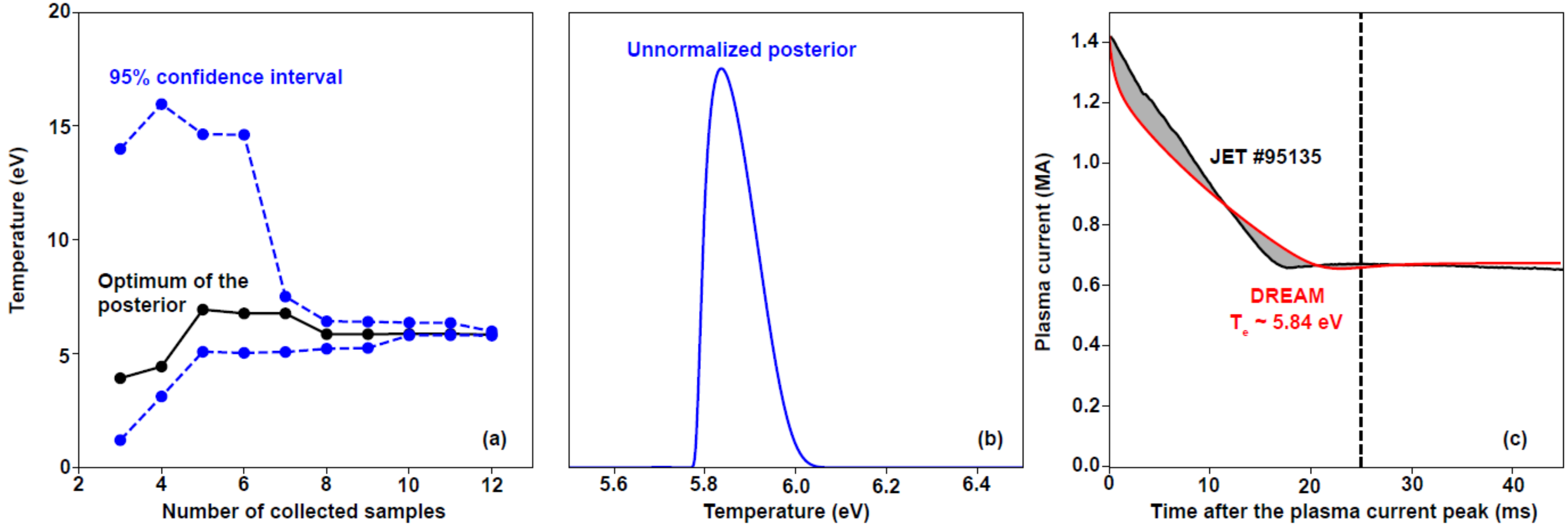
1D proof-of-principle



- Gaussian process provides a sample-efficient probabilistic regression

- Rational quadratic kernel used in the GPR
- Lower confidence bound acquisition function
- f_{AR} = 15 %, τ_{wall} = 5 ms, Radially uniform RE seed distribution

1D proof-of-principle

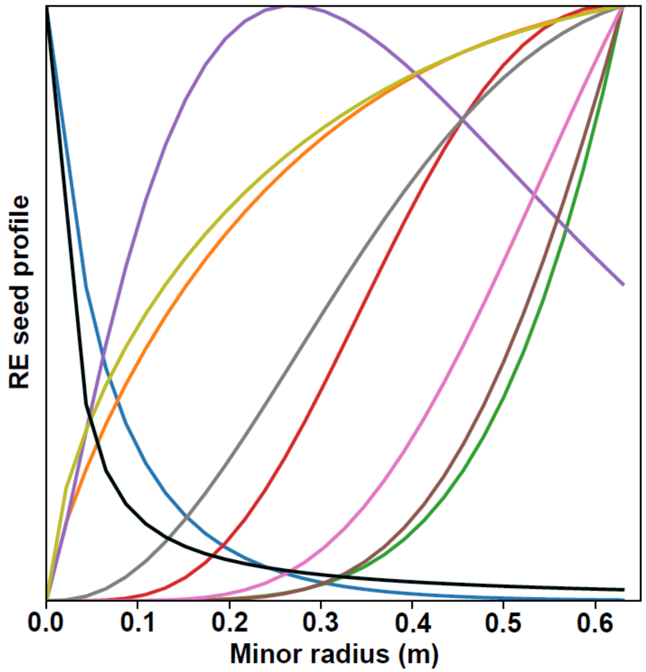


Extending the search space to 5 dimensions

- **Uniform 5D search space**
  - RE seed parameterized as gamma distribution pdf with α, β as inputs

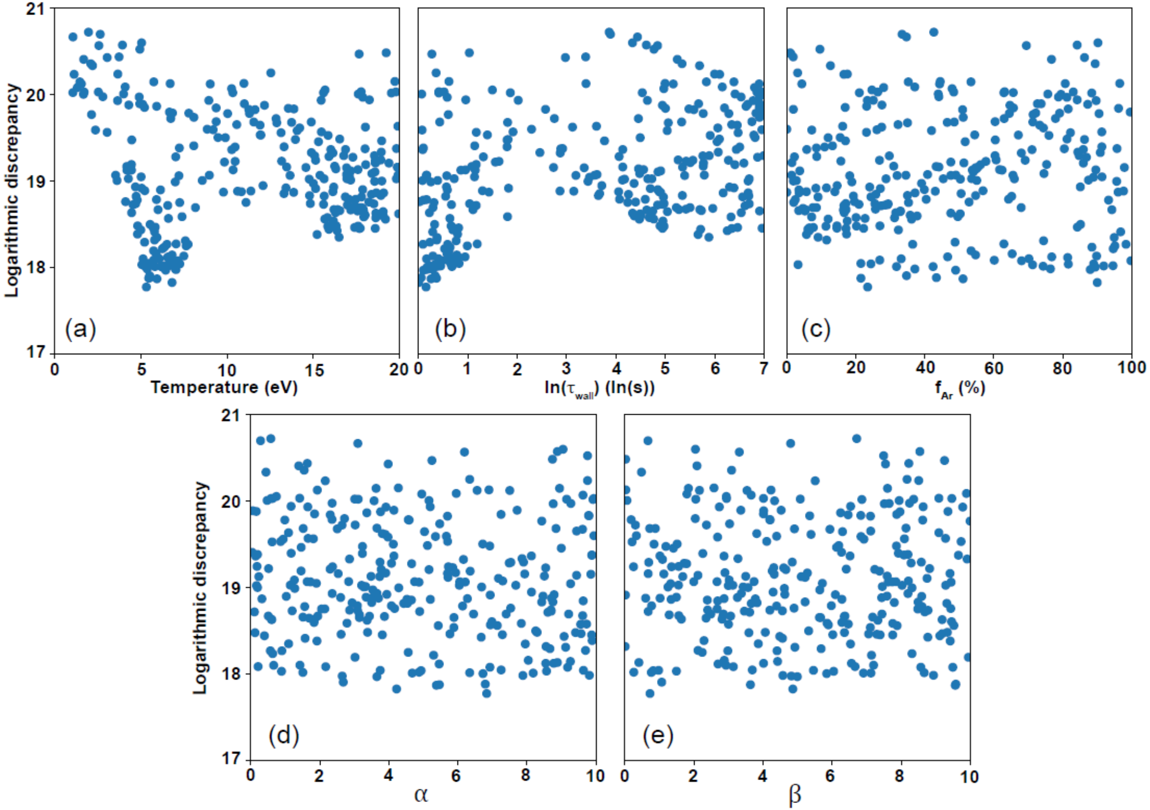
Parameter	Lower bound	Upper bound
T_e (eV)	1.0	20
f_{AR} (%)	0.001	100
$ln(\tau_{wall}$ (ms))	0	7
α, β	0.001	10

10 randomly sampled RE seed profiles



Clear optima are found for T_e and τ_{wall} in the 5D search space

290 samples

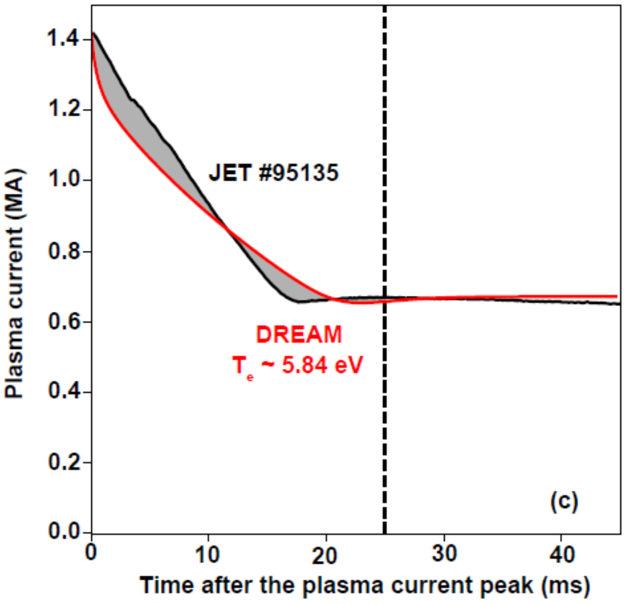


- Global optimum at T ~ 6 eV, τ_{wall} ~ 3 ms
- Local optimum at T ~ 16 eV, τ_{wall} > 50 ms

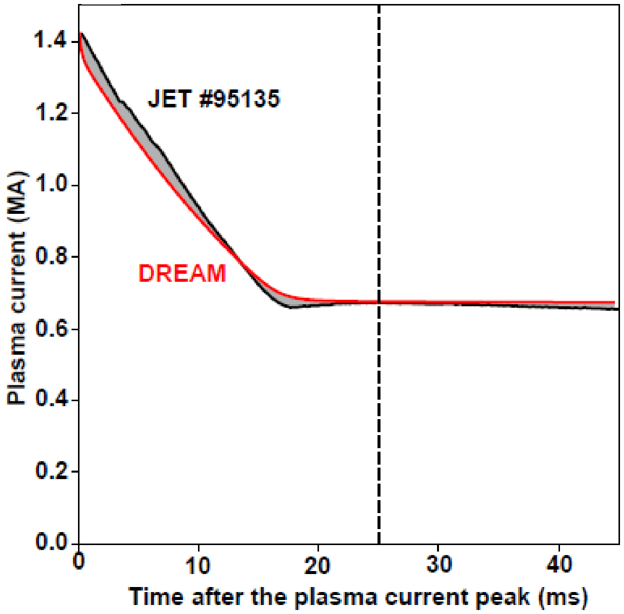
- RandMaxVar stochastic acquisition function used [M. Järvenpää et al. Bayesian Analysis 2019]
- Number of parallel samples is 10

Clearly a better fit to the experimentally measured current evolution is obtained with the 5D optimisation

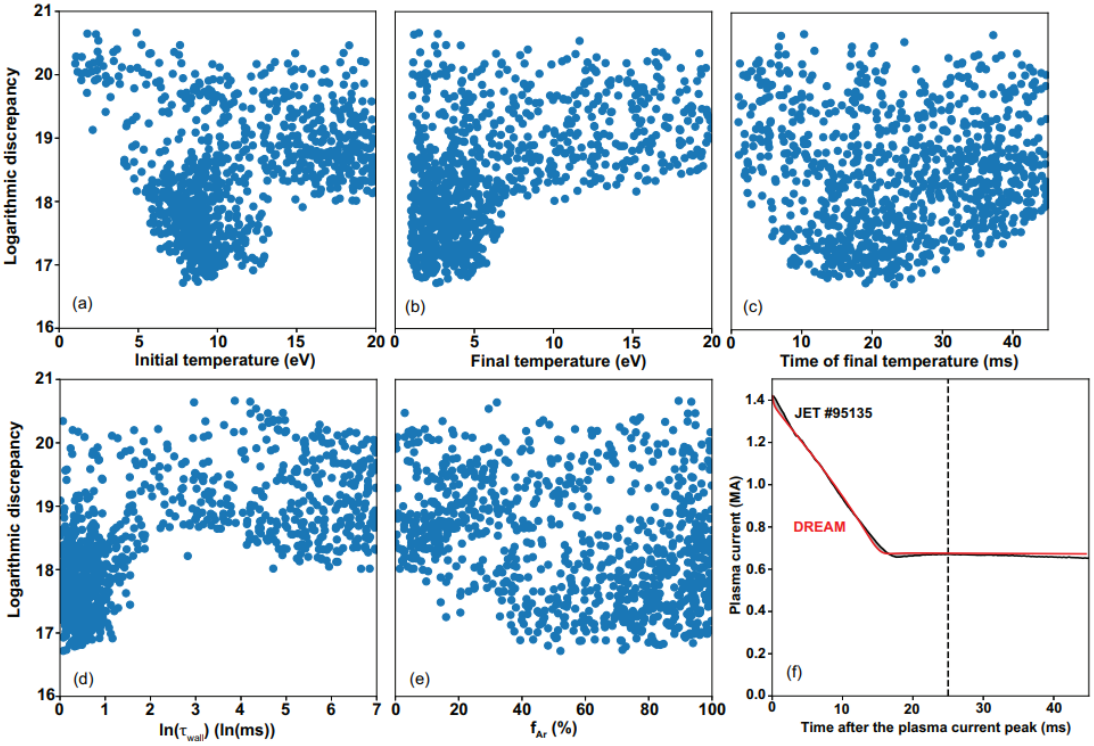
1D search space

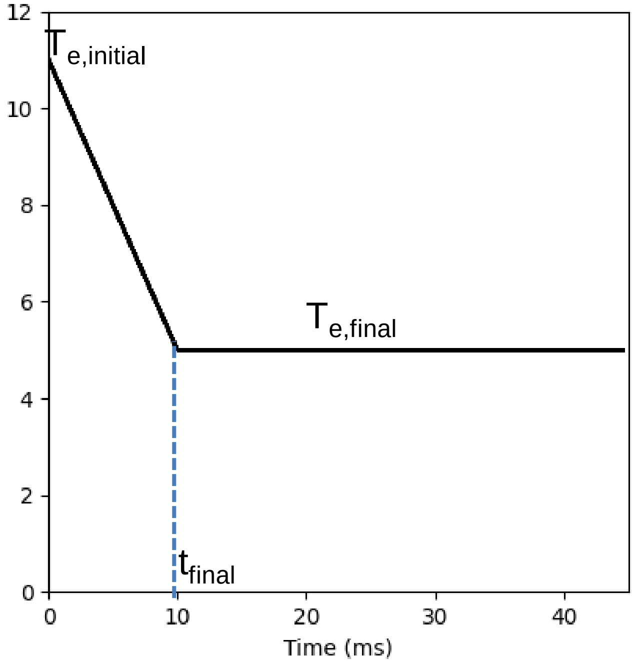


5D search space

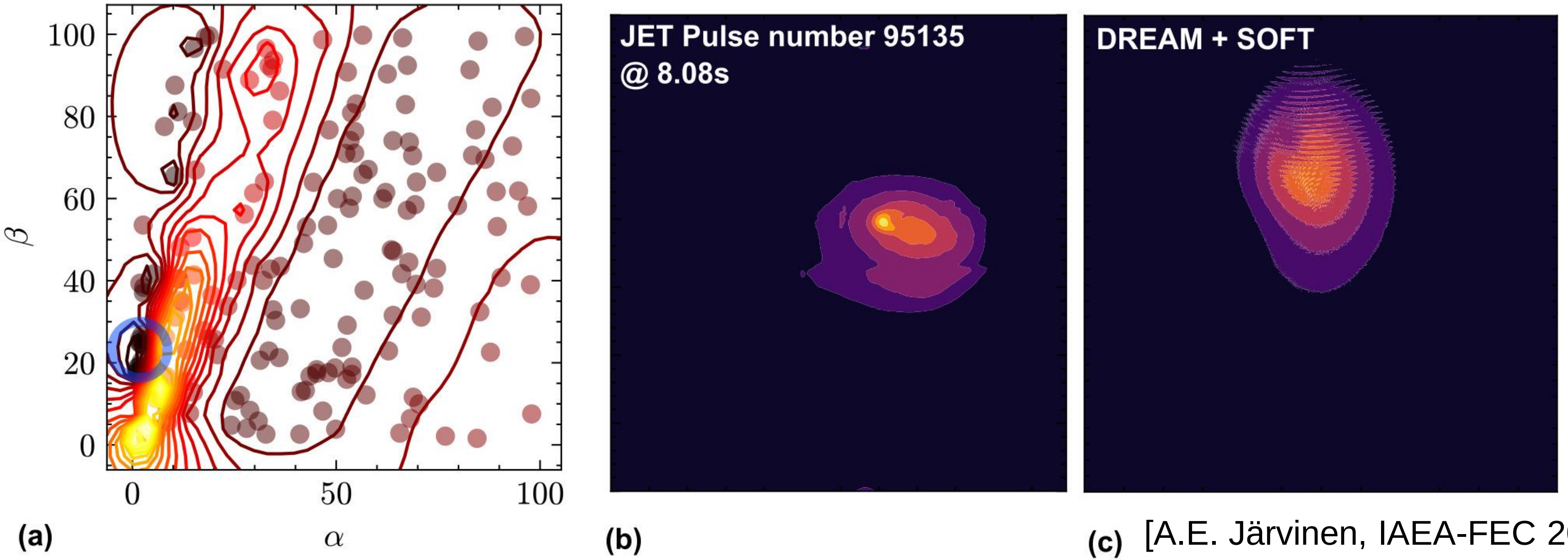


Extending the search space to allow parameterized variation of electron temperature, the algorithm is able to match the evolution of the plasma current quite well





Presently the focus of the project is to use the framework to match the observed synchrotron radiation distribution



[A.E. Järvinen, IAEA-FEC 2023]

- Need to model runaway electron velocity distribution and to generate a synthetic synchrotron emission distribution with an orbit-following code SOFT [Hoppe Nuclear Fusion 2018 https://doi.org/10.1088/1741-4326/aa9abb]
- A forward pass for a single input takes several hours (vs. ~ minute in the previous study with fluid model for REs)

Likelihood-free inference to constrain SOL fluid simulations

IOP Publishing — Plasma Physics and Controlled Fusion

Plasma Phys. Control. Fusion **64** (2022) 104003 (13pp) — https://doi.org/10.1088/1361-6587/ac828d

Normalizing flows for likelihood-free inference with fusion simulations

C S Furia[1,*] and R M Churchill[2]

[1] Rutgers University, New Brunswick, NJ, 08901, United States of America
[2] Princeton Plasma Physics Laboratory, Princeton, NJ, 08540, United States of America

E-mail: chirag.furia@rutgers.edu

Received 18 February 2022, revised 29 June 2022
Accepted for publication 20 July 2022
Published 23 August 2022

CrossMark

Abstract

Fluid-based scrape-off layer transport codes, such as UEDGE, are heavily utilized in tokamak analysis and design, but typically require user-specified anomalous transport coefficients to match experiments. Determining the uniqueness of these parameters and the uncertainties in them to match experiments can provide valuable insights to fusion scientists. We leverage recent work in the area of likelihood-free inference (‘simulation-based inference’) to train a neural network, which enables accurate statistical inference of the anomalous transport coefficients given experimental plasma profile input. UEDGE is treated as a black-box simulator and runs multiple times with anomalous transport coefficients sampled from priors, and the neural network is trained on these simulations to emulate the posterior. The neural network is trained as a normalizing flow model for density estimation, allowing it to accurately represent complicated, high-dimensional distribution functions. With a fixed simulation budget, we compare a single-round procedure to a multi-round approach that guides the training simulations toward a specific target observation. We discuss the future possibilities for use of amortized models, which train on a wide range of simulations and enable fast statistical inference for results during experiments.

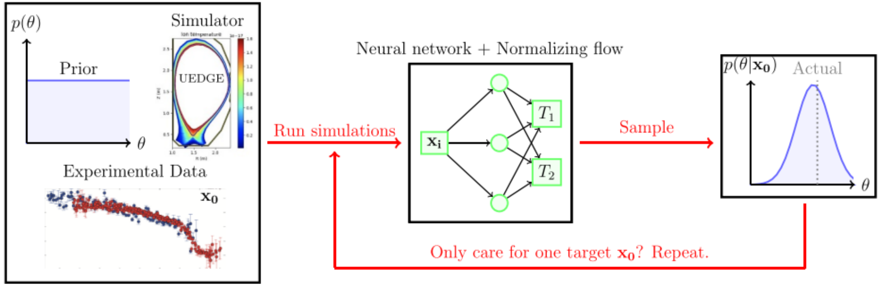
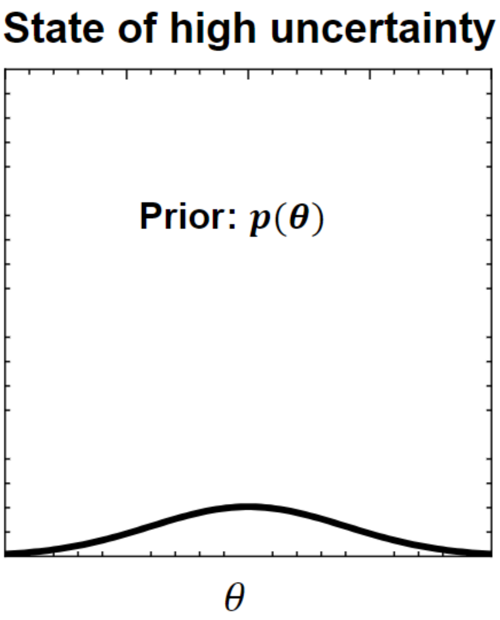


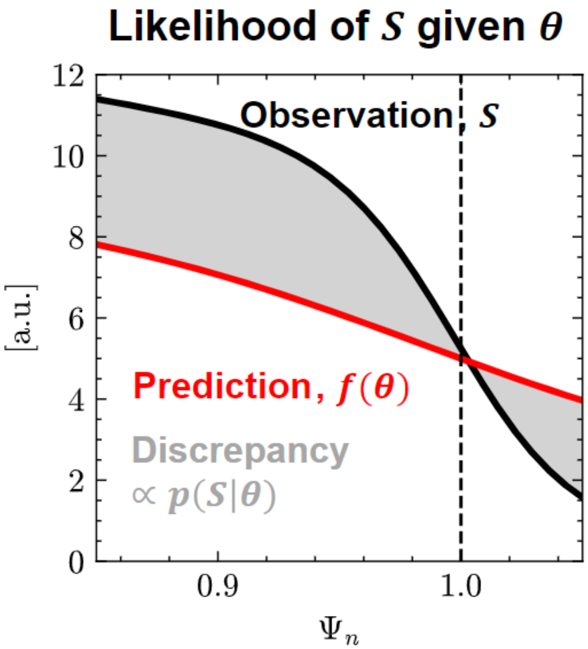
Figure 2. A visualization of (sequential) neural posterior estimation, based on a graphic created by Gonçalves *et al* [5]. For a given prior, simulator, and target observation, N simulations will be run based on samples from the prior. The corresponding $(\theta, \mathbf{x})$ pairs will be used to train a normalizing flow model based on deep neural networks. This flow can then be sampled from for a particular $\mathbf{x}_0$ in order to plot the pdf corresponding to $p(\theta|\mathbf{x}_0)$. For the sake of simplicity, θ is restricted to one dimension in this diagram but in reality will often be multidimensional. If non-amortized SNPE inference is desired, the process of running simulations and training can be repeated by denoting $p(\theta|\mathbf{x}_0)$ as the proposal distribution to be sampled from instead of the prior.

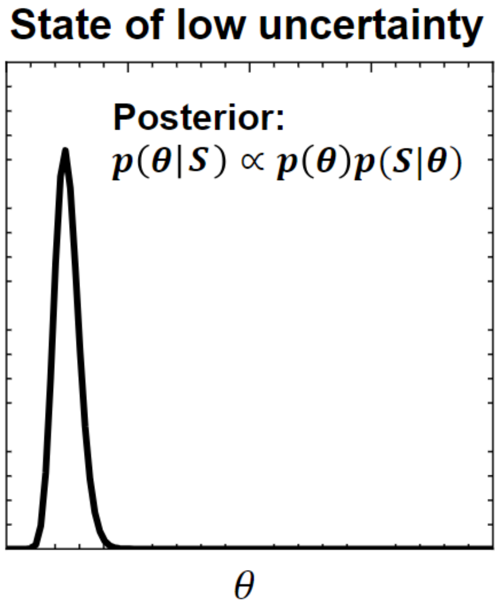
- Use likelihood-free inference to constrain cross-field transport coefficients in UEDGE simulations: $D_\perp$, $Xe_\perp$, $Xi_\perp$ → 3x10 parameters (10 values radially)
- Test the Bayesian optimization framework for the same example UEDGE case

C.S. Furia, R.M. Churchill, Plasma Phys. Control. Fusion **64** (2022) 104003. https://doi.org/10.1088/1361-6587/ac828d

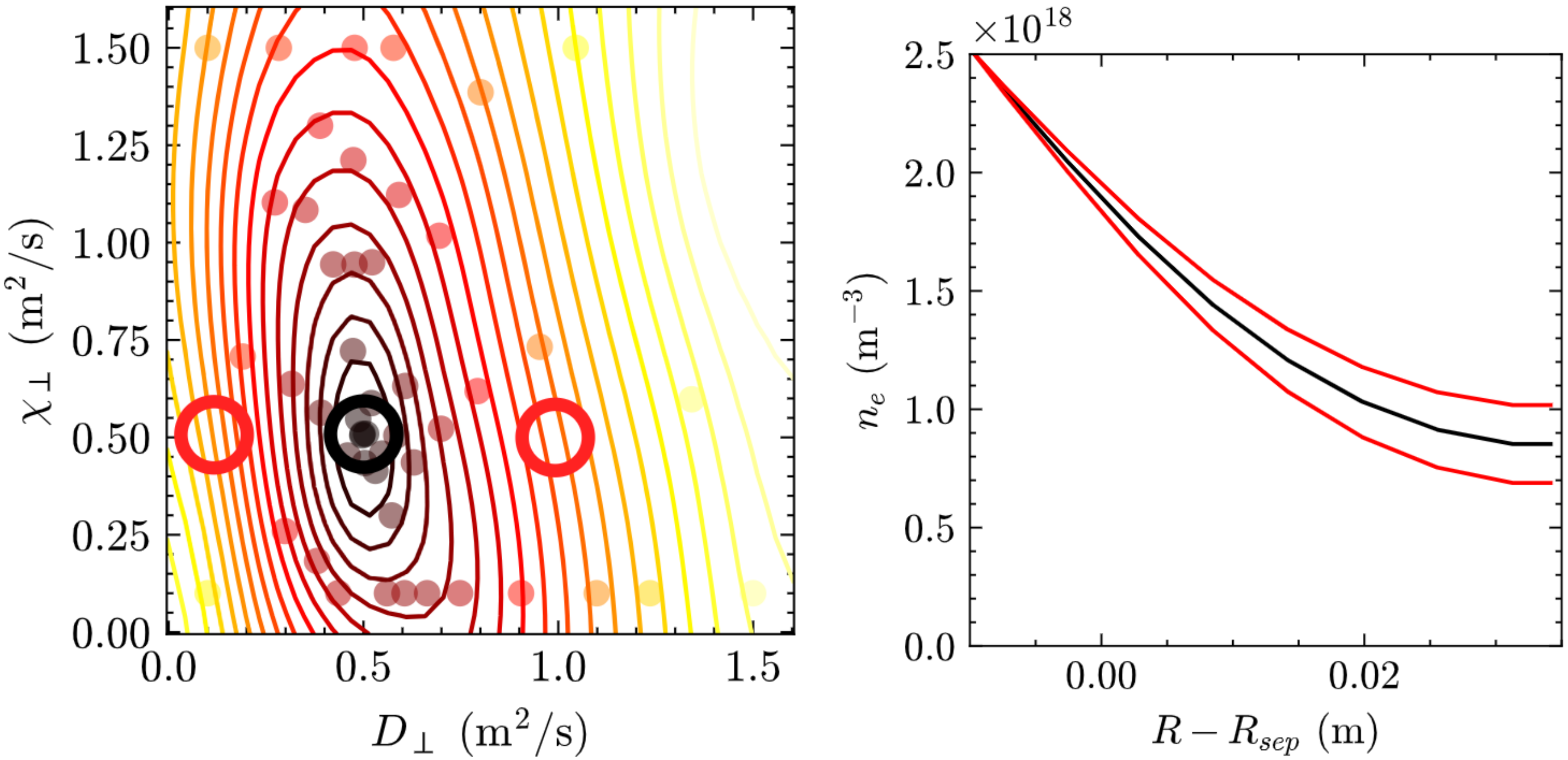
Aim to infer cross-field transport coefficients that reproduce a given target profile





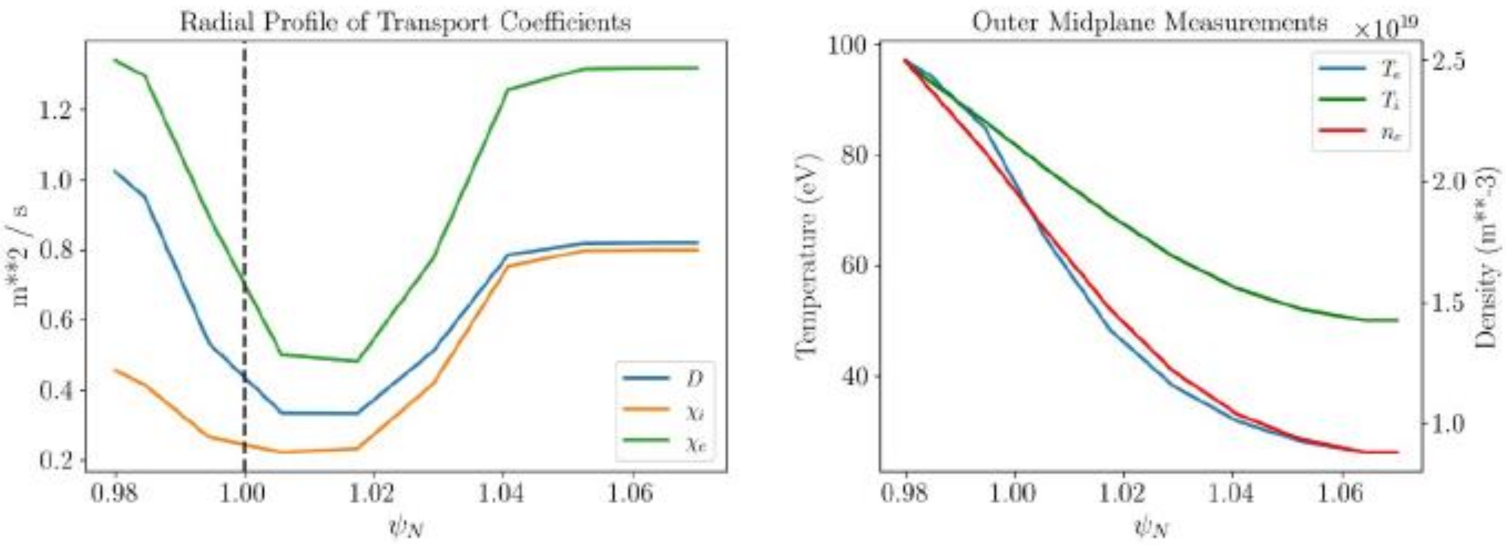


Simple 2D example with D⊥ & X⊥ set as 0.5 m²/s



[A.E. Järvinen, IAEA-FEC 2023]

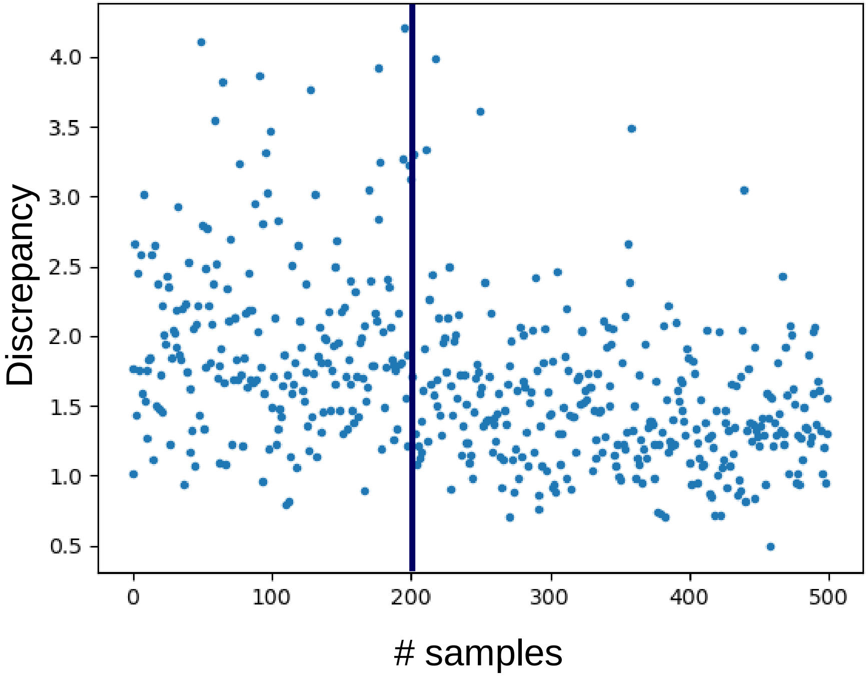
Explore the same (30 dimension) case as Furia and Churchill



C.S. Furia, R.M. Churchill, Plasma Phys. Control. Fusion **64** (2022) 104003.
https://doi.org/10.1088/1361-6587/ac828d

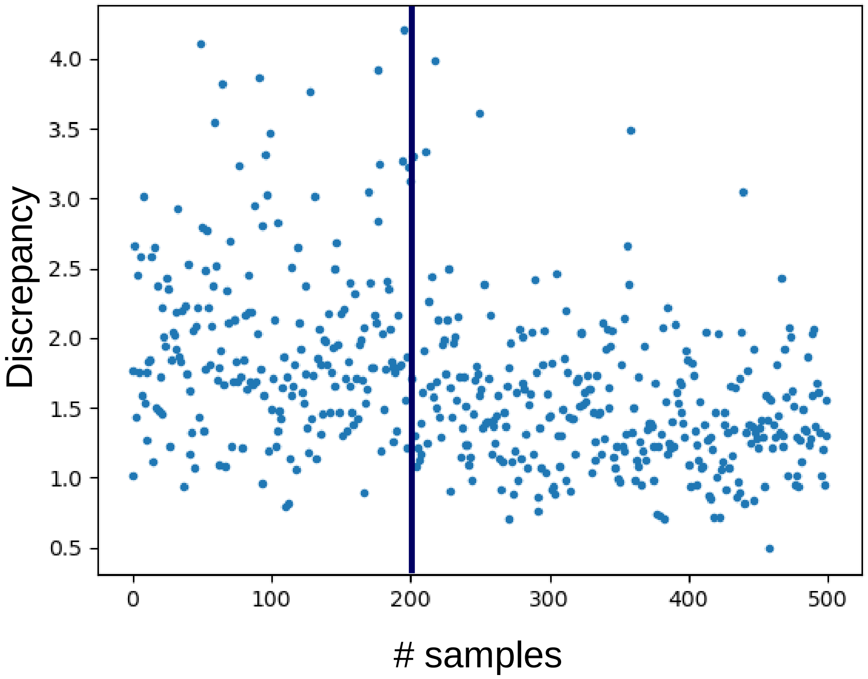
Explore the same (30 dimension) case as Furia and Churchill

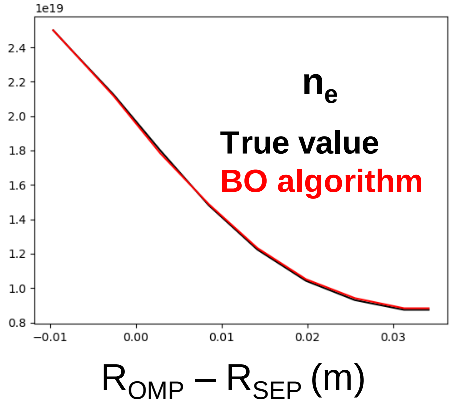
Bayesian optimization started after 200 randomly sampled cases (stochastic acquisition using Thompson sampling)

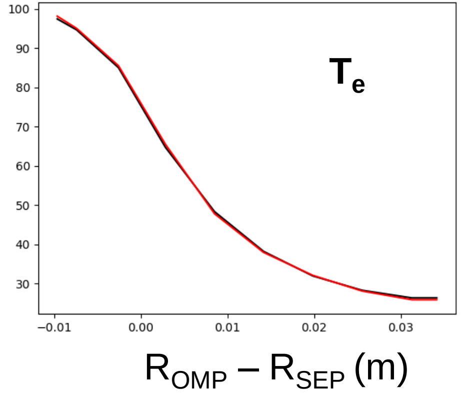


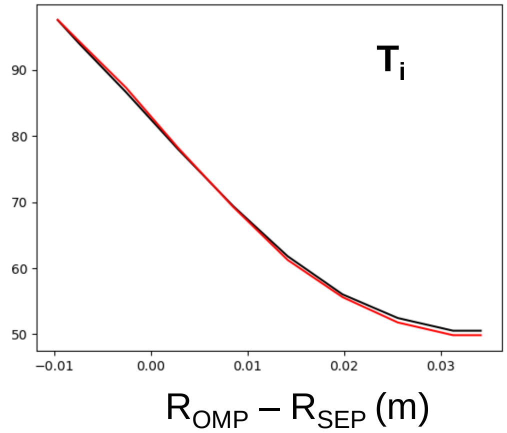
Explore the same (30 Dimension) case as Furia and Churchill

Bayesian optimization started after 200 randomly sampled cases (stochastic acquisition using Thompson sampling)

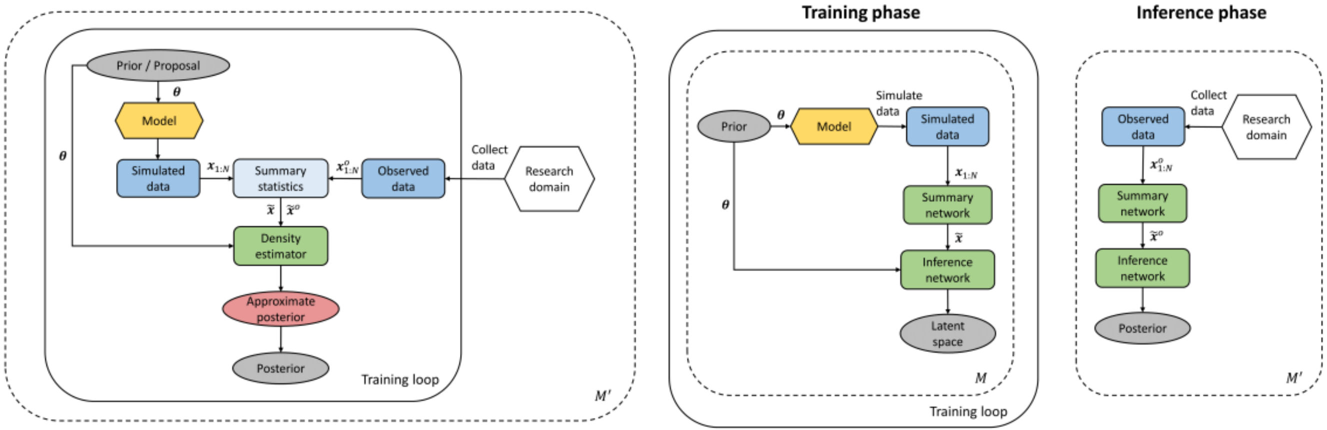








Outlook: From case-based to amortized inference



(a) Case-based inference

(b) Globally amortized inference with BayesFlow

Radev et al. IEEE 2022
https://ieeexplore.ieee.org/document/9298920

- The previous discussion has focused on conducting BO for a single input-simulation-output combination starting from scratch for each BO task – Case-based inference
- In practice, conducting these searches over variety of experiments and configurations, one ends up repeating similar tasks
- A learning algorithm that is able to generalize and use previous experiences to guide searches would be very attractive → called amortized inference – this is much like the human brain is argued to operate [Gershman & Goodman, Cognitive Science 2014 https://api.semanticscholar.org/CorpusID:924780]

Bayesian inference (BI) algorithms provide a principled approach to quantify the uncertainty for the state of the investigated system, given available data

- **These tools hold the potential to significantly reduce the manual work in parameter calibration when validating computationally demanding models for fusion plasmas**
- **With computationally expensive models, data-efficiency is key to reduce the overall CPUh needs → Bayesian optimization (BO) is a way to achieve this → For practical BO for computationally expensive model calibration: batch acquisition, failure handling, and means to appropriately balance exploitation-exploration**
- **A broad portfolio of BI and BO projects are being conducted in close connection the EUROfusion Advanced Computing Hub (05) hosted by the University of Helsinki**
- **Outlook: In future the plan is to investigate the applicability of deep generative models within the BO workflow to estbalish amortized inference capabilities for these tasks – See e.g. Radev et al. IEEE Trans Neural Netw Learn Syst. 2022** BayesFlow: Learning Complex Stochastic Models With Invertible Neural Networks | IEEE Journals & Magazine | IEEE Xplore

Connection to active learning

arXiv:2310.09024v1 [physics.plasm-ph] 13 Oct 2023

Efficient training sets for surrogate models of tokamak turbulence with Active Deep Ensembles

L. Zanisi[1], A. Ho[2,3], J. Barr[4], T. Madula[4], J. Citrin[2,3], S. Pamela[1], J. Buchanan[1], F. Casson[1], V. Gopakumar[1] and JET contributors[5]

[1] United Kingdom Atomic Energy Authority, Culham Centre for Fusion Energy, Abingdon, UK
[2] DIFFER - Dutch Institute for Fundamental Energy Research, Eindhoven, Netherlands
[3]Science and Technology of Nuclear Fusion Group, Eindhoven University of Technology, Eindhoven, Netherlands
[4] UCL's Centre for Doctoral Training in Data Intensive Sciences, University College London
[5] See the author list of "Overview of T and D-T results in JET with ITER-like wall" by CF Maggi et al. to be published in Nuclear Fusion Special Issue: Overview and Summary Papers from the 29th Fusion Energy Conference (London, UK, 16-21 October 2023)

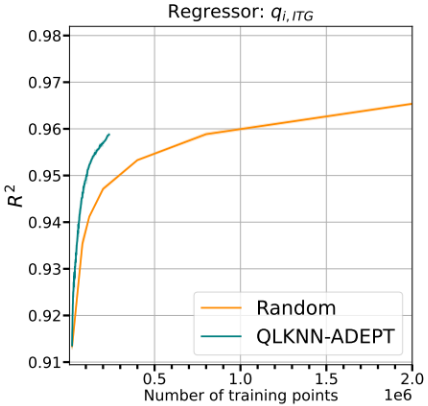
E-mail: lorenzo.zanisi@ukaea.uk

Abstract.
Model-based plasma scenario development lies at the heart of the design and operation of future fusion powerplants. Including turbulent transport in integrated models is essential for delivering a successful roadmap towards operation of ITER and the design of DEMO-class devices. Given the highly iterative nature of integrated models, fast machine-learning-based surrogates of turbulent transport are fundamental to fulfil the pressing need for faster simulations opening up pulse design, optimization, and flight simulator applications. A significant bottleneck is the generation of suitably large training datasets covering a large volume in parameter space, which can be prohibitively expensive to obtain for higher fidelity codes.

In this work, we propose ADEPT (Active Deep Ensembles for Plasma Turbulence), a physics-informed, two-stage Active Learning strategy to ease this challenge. Active Learning queries a given model by means of an acquisition function that identifies regions where additional data would improve the surrogate model. We provide a benchmark study using available data from the literature for the QuaLiKiz quasilinear transport model. We demonstrate quantitatively that the physics-informed nature of the proposed workflow reduces the need to perform simulations in stable regions of the parameter space, resulting in significantly improved data efficiency compared to non-physics informed approaches which consider a regression problem over the whole domain. We show an up to a factor of 20 reduction in training dataset size needed to achieve the same performance as random sampling. We then validate the surrogates on multichannel integrated modelling of ITG-dominated JET scenarios and demonstrate that they recover the performance of QuaLiKiz to better than 10%. This matches

https://arxiv.org/pdf/2310.09024.pdf

- A key challenge in developing machine learning surrogates for computationally expensive models is to establish the training set
- Training set of ~million simulations with a model that costs ~CPUh requires about million CPUh
- In active learning, acquisition function is used to recommend sampling the the full model in parts of the parameter space in which the surrogate model uncertainty is high



- This could be seen BO with the objective function to reduce the uncertainty globally for the surrogate model

Model validation challenge: how to rigorously select the free input parameters for the model to best represent the investigated system

- Usually, a selection of the input parameters are not well constrainted by the available data

- Multidimensional optimization or model calibration is needed to find the most appropriate combination of input parameters → Leading bottleneck in model validation exercises

Conventional validation workflow

Select the experiment(s) and measurements used in the validation exercise

↓

Setup the simulation, grid, and initial conditions to represent the experiment

↓

Multidimensional optimization

Setup the uncertain input parameters

↓

Run the simulation and compare the measured and synthetic signals

Adjust the uncertain input parameters to calibrate the output

↓

Evaluate remaining discrepancy between the measured and synthetic signals

Model validation challenge: how to rigorously select the free input parameters for the model to best represent the investigated system

- **Usually, a selection of the input parameters are not well constrainted by the available data**

- **Multidimensional optimization or model calibration is needed to find the most appropriate combination of input parameters → Leading bottleneck in model validation exercises**

Human is subjective and inefficient in completing this task:

1. **Subjective reasoning for trajectory through the search space**
2. **Poorly quantified uncertainties**
3. **Manual input selection and output processing is inefficient**

An optimization algorithm to conduct this process would be very attractive

Some essential parts of simulating a system with a numerical model

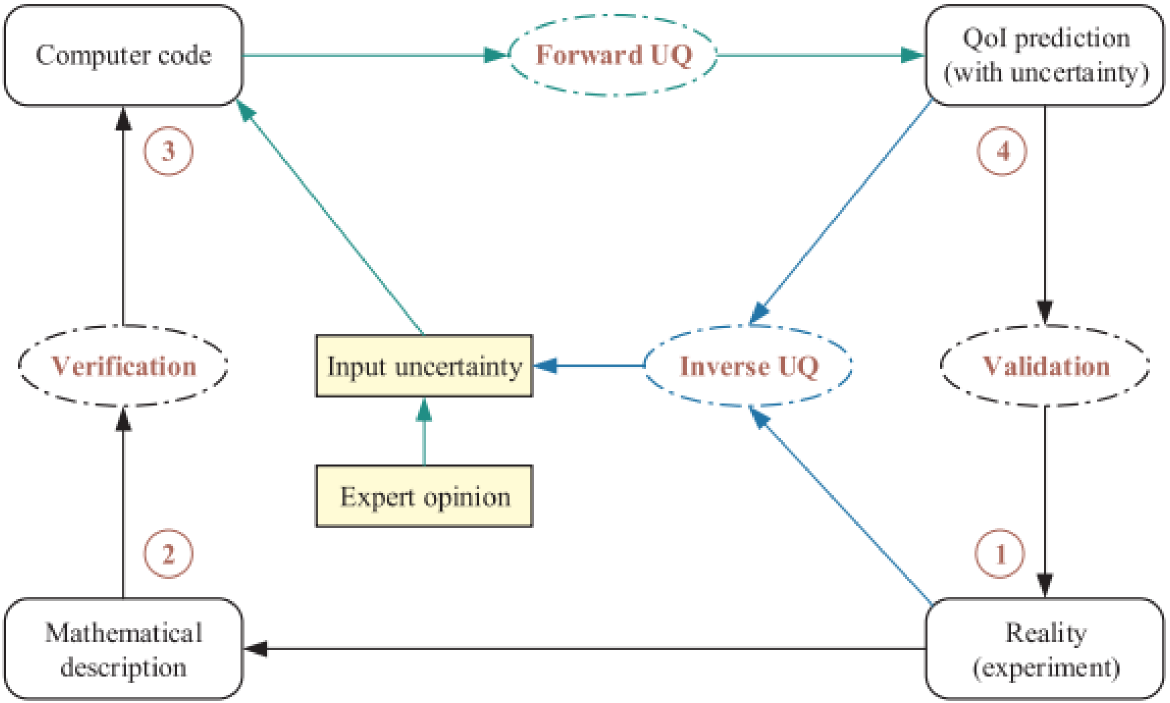


Figure 1 in Wu, et al. Nucl. Eng. Des. 2018
https://doi.org/10.1016/j.nucengdes.2018.06.004