# Surrogate Modelling of the Tritium Breeding Ratio

Petr Mánek    Graham Van Goffrier

Centre for Doctoral Training in Data Intensive Science
University College London

8th October 2020

UCL CDT DIS

UK Atomic Energy Authority

# Project Background

Nuclear fusion – the energy of the future!

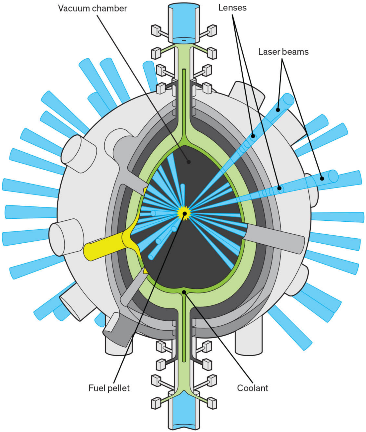



Illustration by Chris Philpot, courtesy of IEEE Spectrum.

- Designing next-generation Inertial Confinement Fusion (ICF) facility.
- Search for optimal reactor design.
- Fueling – important for viability.
- Require fuel of 2 varieties:
  - Deuterium $^2$H – abundant in naturally-sourced water.
  - Tritium $^3$H – extremely rare.
- Modern reactors can generate tritium during operation.

# Problem Description

Tritium breeding blankets convert neutron radiation to $^3$H fuel:

$$ {}^1_0\mathrm{n} + {}^6_3\mathrm{Li} \to {}^3_1\mathrm{H} + {}^4_2\mathrm{He} \qquad {}^1_0\mathrm{n} + {}^7_3\mathrm{Li} \to {}^3_1\mathrm{H} + {}^4_2\mathrm{He} + {}^1_0\mathrm{n} $$

$^3$H balance described by Tritium Breeding Ratio (TBR) $= \dfrac{\text{fuel bred}}{\text{fuel consumed}}$

- Depends on numerous geometric and material parameters.
- Evaluated by *Paramak* – OpenMC neutronics simulation.
- Slow . . . we want to consider as many reactor designs as possible!

## Our Challenge:

Produce a *fast* TBR surrogate that strongly approximates Paramak.

# Data Generation

Produced datasets by sampling Paramak outputs over its 7 discrete and 11 continuous input parameters at random.

Deployed at UCL's Hypatia cluster:

- Created 1M points.
- 27 days of runtime.

2 classes of runs:

- All parameters free.
- Discrete fixed, continuous free.

| | Parameter name | Domain |
|---|---|---|
| Blanket | Breeder fraction$^{\dagger}$ | $[0, 1]$ |
| | Breeder $^{6}\mathrm{Li}$ enrichment fraction | $[0, 1]$ |
| | Breeder material | $\{Li_2TiO_3, Li_4SiO_4\}$ |
| | Breeder packing fraction | $[0, 1]$ |
| | Coolant fraction$^{\dagger}$ | $[0, 1]$ |
| | Coolant material | $\{D_2O, H_2O, He\}$ |
| | Multiplier fraction$^{\dagger}$ | $[0, 1]$ |
| | Multiplier material | $\{Be, Be_{12}Ti\}$ |
| | Multiplier packing fraction | $[0, 1]$ |
| | Structural fraction$^{\dagger}$ | $[0, 1]$ |
| | Structural material | $\{$SiC, eurofer$\}$ |
| | Thickness | $[0, 500]$ |
| First wall | Armour fraction$^{\ddagger}$ | $[0, 1]$ |
| | Coolant fraction$^{\ddagger}$ | $[0, 1]$ |
| | Coolant material | $\{D_2O, H_2O, He\}$ |
| | Structural fraction$^{\ddagger}$ | $[0, 1]$ |
| | Structural material | $\{$SiC, eurofer$\}$ |
| | Thickness | $[0, 20]$ |

Groups marked$^{\dagger\ddagger}$ required to sum to 1.

Conventional regression task – search for a cheap surrogate $\hat{f}(x)$ that minimizes dissimilarity with an expensive function $f(x)$:

- Regression performance: mean absolute error, $\sigma$ of error, $R^2$, $R^2_{\text{adj.}}$
- Computational complexity: training & prediction time / sample

Decoupled training approach:

1. Collect training dataset at random.
2. Use data to train a surrogate.

Adaptive training approach:

1. Collect initial training dataset at random.
2. Use data to train a surrogate.
3. Collect more data in regions where surrogate performed poorly.
4. Repeat steps 2 and 3.

# Decoupled Approach

# Outline

Compared 9 state-of-the-art surrogate families:

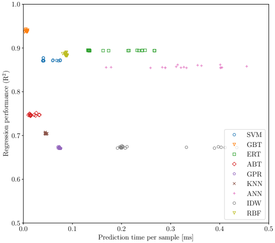
- Support vector machines,
- Gradient boosted trees,
- Extremely randomized trees,
- AdaBoosted decision trees,
- Gaussian process regression,
- $k$ nearest neighbors,
- Artificial neural networks (MLP),
- Inverse distance weighting,
- Radial basis functions.

Performed 4 experiments:

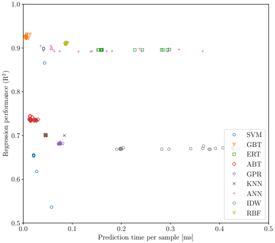
1. Hyperparameter tuning (simplified) – Bayesian optimization, discrete features fixed & withheld.
2. Hyperparameter tuning – same as #1 but with all features.
3. Scaling benchmark – increase training set size.
4. Model comparison – train surrogates for practical use.

# Experiments 1 & 2: Hyperparameter Tuning

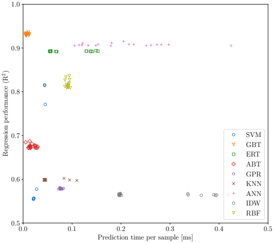
Experiment 1, slice (a)

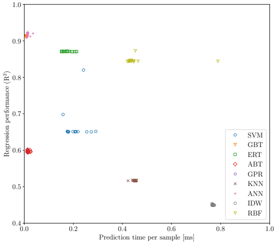



Experiment 1, slice (b)




Experiment 1, slice (c)



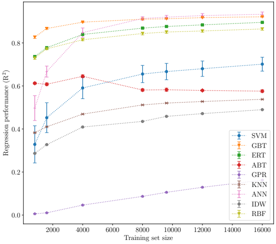




Experiment 2

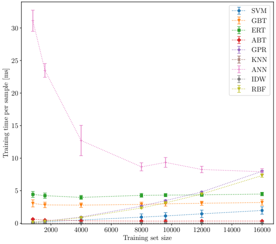
- Showing $\bar{t}_{\text{pred.}}$ vs. $R^2$ for the 20 best surrogates per family (top left $\Leftrightarrow$ fastest, most accurate).
- Omit discrete features $\rightarrow$ negligible performance improvement.
- Dominated by trees (GBTs, ERTs) and neural networks.

# Experiment 3: Scaling Benchmark

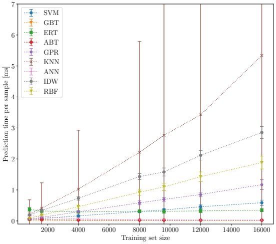
- We observe a hierarchy.
- Trees and neural networks scale the best in $\bar{t}_{\text{pred.}}$.
- Maximizing training set size, neural networks dominate.
- Instance-based surrogates (KNN, IDW) train trivially but have slow lookup.
- Neural networks show inverse scaling due to parallelization.




Regression performance




Training time / sample



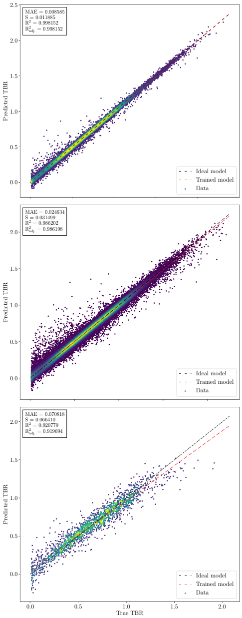

Prediction time / sample

# Experiment 4: Model Comparison

- Trained 8 models for practical use.
- Plots show true vs. predicted TBR by Models 1, 2 & 4, coloured by density.

- Model 1 – best regression performance:
  - ANN (4-layer MLP), 500K samples.
  - $R^2 = 0.998$, $\sigma = 0.013$,
  - $\bar{t}_{\text{pred.}} = 1.124\,\mu\text{s}$, speedup $6\,916\,416\times$
- Model 2 – fastest prediction:†
  - ANN (2-layer MLP), 500K samples.
  - $R^2 = 0.985$, $\sigma = 0.033$,
  - $\bar{t}_{\text{pred.}} = 0.898\,\mu\text{s}$, speedup $8\,659\,251\times$
- Model 4 – smallest training set:†
  - GBT, 10K samples.
  - $R^2 = 0.913$, $\sigma = 0.072$,
  - $\bar{t}_{\text{pred.}} = 6.125\,\mu\text{s}$, speedup $1\,269\,777\times$

† with acceptable regression performance.



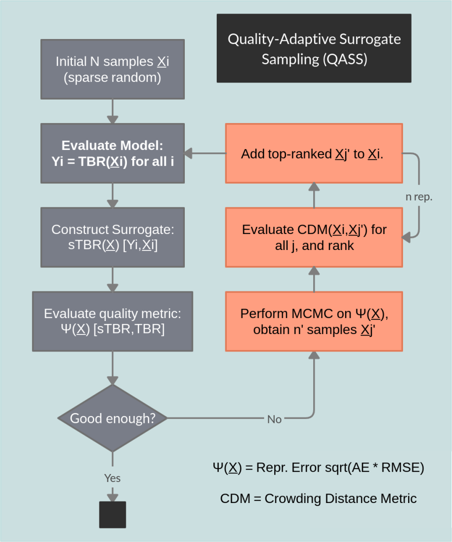
# Adaptive Approach

# Adaptive Sampling: Theory

How can we take advantage of surrogate information content *during training* to reduce sample quantity?

We developed a new technique:

1. Construct surrogate quality distribution by nearest-neighbour interpolation.
2. Draw candidate samples by quality using MCMC.
3. Include samples with greatest separation from neighbours.
4. Repeat!



# Application on Toy Theory

Toy functional TBR theory with wavenumber $n$, and qualitatively comparable ANN performance to Paramak:

$$\mathrm{TBR}_{\mathrm{toy}} = \frac{1}{|C|} \sum_{i \in C} \left[1 + \sin(2\pi n(x_i - 1/2))\right]$$

(where $C$ enumerates all continuous variables)

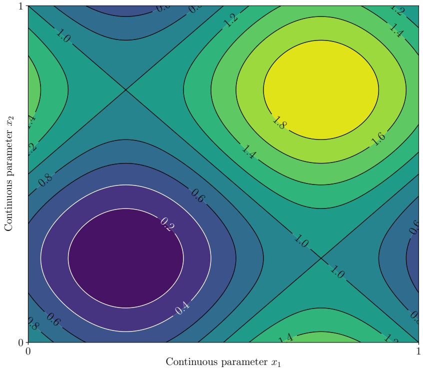
Evaluation set:

- Adaptive samples
- Generated during runtime

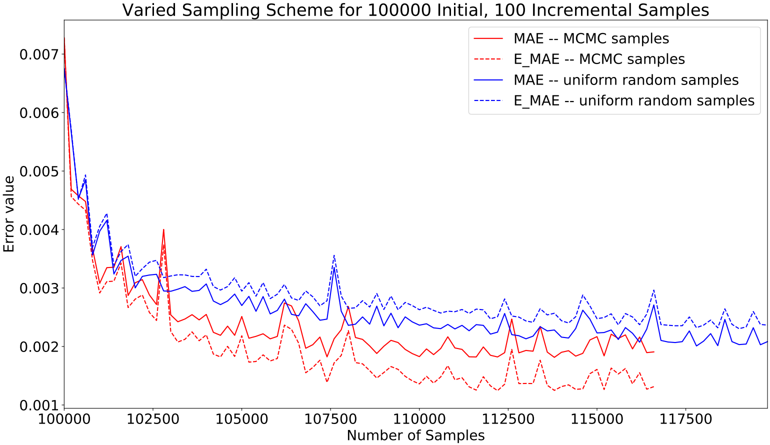
Validation set:

- Uniform random samples
- Generated independently

Placebo comparison – incremental uniform-random samples, no MCMC.



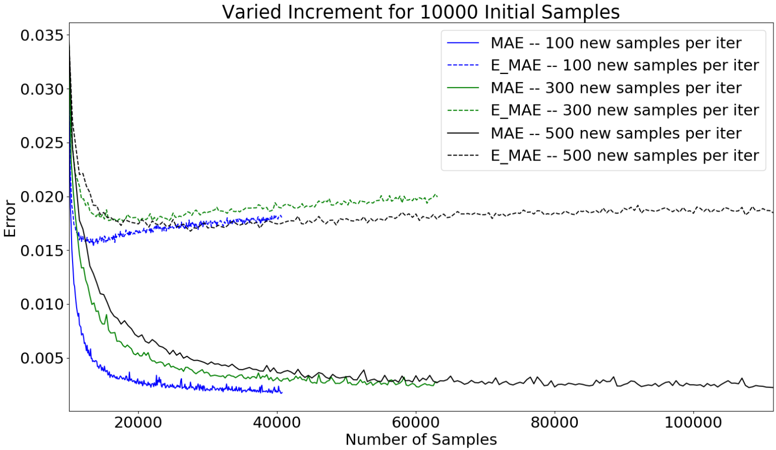
# Adaptive Sampling: Results




60% decrease in MAE for validation set (dashed)

Equivalently, 6% decrease in samples needed for same accuracy

# Adaptive Sampling: Results




Fewer incremented samples can lead to better accuracy!

But depends on initial samples, specific model – further study needed.

# Conclusion

Decoupled approach:

- Tuned and compared surrogates from 9 state-of-the-art families.
- Found heuristic: GBTs for $< 10^4$ samples, ANNs for $\geqslant 10^5$ samples.
- Fastest found surrogate predicts TBR with standard deviation of error $0.033$ in $0.898\,\mu s$, which is $8 \cdot 10^6 \times$ faster than Paramak.
- While this used 500K samples, we found surrogates with comparable properties with as little as 10K samples.

Adaptive approach (on toy theory):

- New theoretical approach QASS developed, based on MCMC.
- 60% decrease in evaluation MAE demonstrated.
- 6% decrease in expensive TBR samples needed.
- Strong potential for further reduction via hyperparameter tuning.

All presented methods portable $\longrightarrow$ can be used as cheap approximation of any simulation or black box function.

# Thank you for listening!

Petr Mánek ............................. petr.manek.19@ucl.ac.uk
Graham Van Goffrier ........... graham.vangoffrier.19@ucl.ac.uk

Further reading:

- Single page abstract (available online).
- Journal article, currently in internal pre-submission review (available online): *Fast Regression of the Tritium Breeding Ratio in Fusion Reactors*.
- Industry group project final report (available online).
- All models, plots, training data, source code and technical documentation. `https://github.com/ucl-tbr-group-project`