

# Surrogate Modelling of the Tritium Breeding Ratio

Petr Mánek   Graham Van Goffrier

Centre for Doctoral Training in Data Intensive Science  
University College London

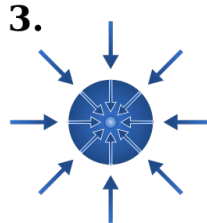
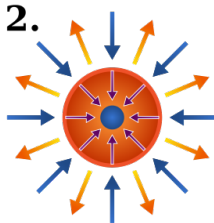
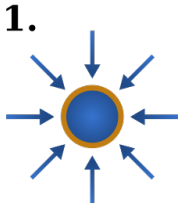
22nd June 2020



UK Atomic  
Energy  
Authority

## Nuclear fusion – the energy of the future!

- Must produce and contain an extremely hot and dense plasma
  - Magnetic Confinement Fusion (MCF): toroidal circulation
  - Inertial Confinement Fusion (ICF): spherical compression
- Modern designs require enriched Hydrogen fuel of two varieties:
  - Deuterium ( $^2\text{H}$ ) – abundant in naturally-sourced water.
  - Tritium ( $^3\text{H}$ ) – extremely rare, but can be produced *in-reactor*.



Tritium breeding blankets convert neutron radiation to tritium fuel:



Tritium breeding ratio (TBR) = fuel bred / fuel consumed

- Depends on numerous geometric and material parameters.
- Evaluated precisely by OpenMC neutronics simulation *Paramak*, but is computationally expensive.

## Our Challenge:

Produce a fast TBR function that strongly approximates Paramak, making use of the latest in surrogate modelling techniques.

We produced training and test datasets by uniform random sampling over the 7 discrete and 11 continuous parameters of Paramak.

Paramak was deployed on UCL's Hypatia cluster:

- Generated 1M samples.
- 27 days of runtime.

2 classes of runs:

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

Groups of fractions marked<sup>††</sup> are required to sum to 1.

	Parameter name	Domain
Blanket	Breeder fraction <sup>†</sup>	[0, 1]
	Breeder <sup>6</sup> Li enrichment fraction	[0, 1]
	Breeder material	{Li <sub>2</sub> TiO <sub>3</sub> , Li <sub>4</sub> SiO <sub>4</sub> }
	Breeder packing fraction	[0, 1]
	Coolant fraction <sup>†</sup>	[0, 1]
	Coolant material	{D <sub>2</sub> O, H <sub>2</sub> O, He}
	Multiplier fraction <sup>†</sup>	[0, 1]
	Multiplier material	{Be, Be <sub>12</sub> Ti}
	Multiplier packing fraction	[0, 1]
	Structural fraction <sup>†</sup>	[0, 1]
	Structural material	{SiC, eurofer}
First wall	Thickness	[0, 500]
	Armour fraction <sup>†</sup>	[0, 1]
	Coolant fraction <sup>†</sup>	[0, 1]
	Coolant material	{D <sub>2</sub> O, H <sub>2</sub> O, He}
	Structural fraction <sup>†</sup>	[0, 1]
	Structural material	{SiC, eurofer}
	Thickness	[0, 20]

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

2 approaches for surrogate training:

- 1 Decoupled – trains models from previously sampled datapoints.
- 2 Adaptive – repeats sampling & model training, increases sampling density in low-performance regions.

## Decoupled Approach

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

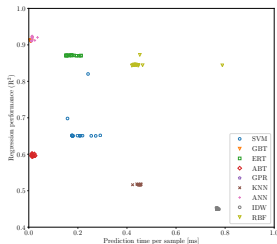
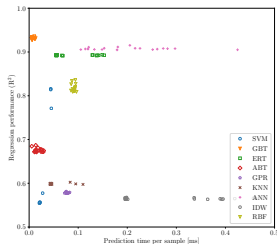
Experiment 1, slice (a)



Experiment 1, slice (b)



Experiment 1, slice (c)



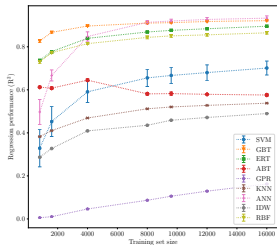
Experiment 2

- Plots show  $\bar{t}_{\text{pred.}}$  vs.  $R^2$  for 20 best surrogates per family (top left  $\Leftrightarrow$  fastest, most accurate).
- Omitting discrete features yields only a negligible improvement in performance.
- Overall dominated by tree-based surrogates (GBTs, ERTs) and neural networks.

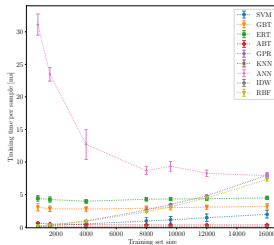


# Experiment 3: Scaling Benchmark

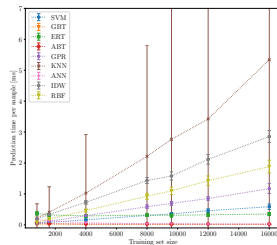
- We observe a hierarchy.
- Best-performing families from the previous experiments also scale the best in  $\bar{t}_{\text{pred.}}$ .
- More samples: neural networks outperform tree-based models.
- Instance-based surrogates (KNN, IDW) train trivially but have complex lookup.
- Neural networks show inverse scaling due to parallelization.



Regression performance



Training time / sample

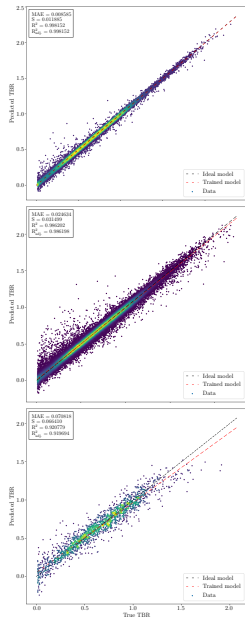


Prediction time / sample

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

<sup>†</sup> with acceptable regression performance.

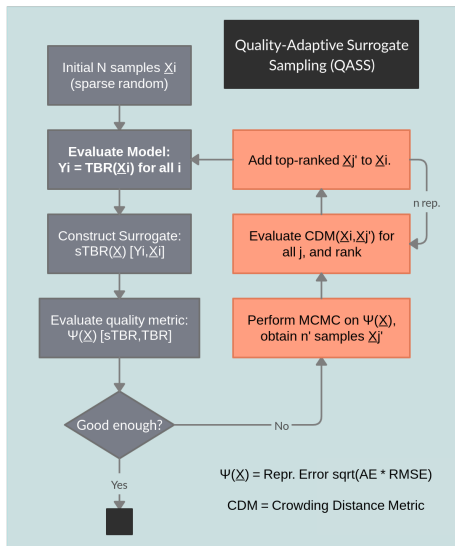


## Adaptive Approach

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!



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

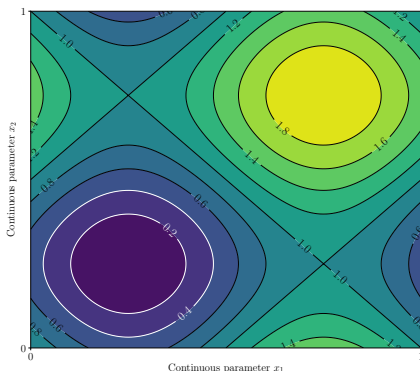
$$\text{TBR} = \frac{1}{|C|} \sum_{i \in C} [1 + \sin(2\pi n(x_i - 1/2))]$$

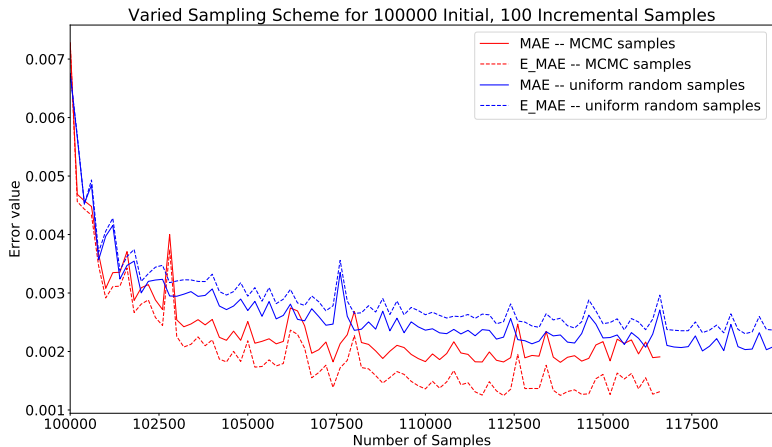
(where  $C$  is the set of variable indices)

2 evaluation sets:

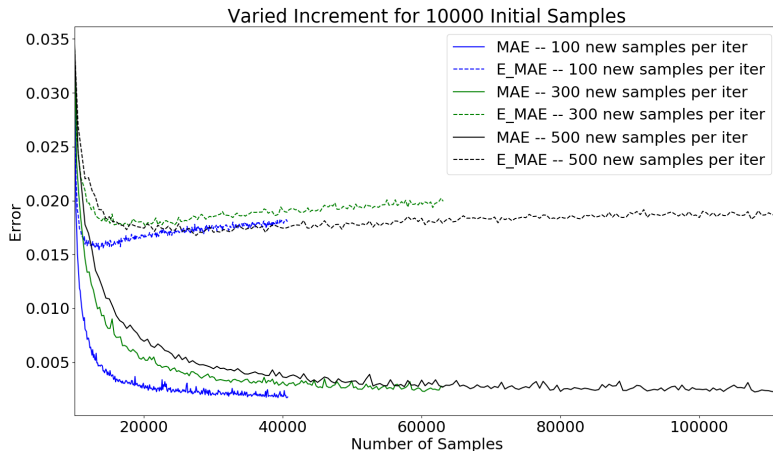
- Adaptively-sampled dataset
- Independent random dataset

Placebo comparison – a baseline scheme without MCMC, incremental uniform-random samples.





60% decrease in MAE for independent evaluation (dashed)  
Equivalently, 6% decrease in samples needed for same accuracy



Fewer incremented samples can lead to better accuracy!

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

## Decoupled approach:

- Tuned and compared surrogates from 9 state-of-the-art families.
- Found heuristic: GBTs for  $< 10^4$  samples, ANNs for  $\geq 10^5$  samples.
- Fastest found surrogate predicts TBR with standard deviation of error 0.033 in  $0.898 \mu\text{s}$ , 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  $\rightarrow$  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>