

Surrogate Modelling of the Tritium Breeding Ratio

Petr Mánek Graham Van Goffrier

Centre for Doctoral Training in Data Intensive Science
University College London

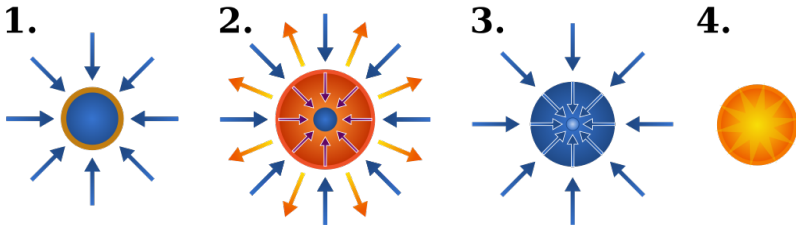
22nd June 2020



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 (^2H) – abundant in naturally-sourced water
 - Tritium (^3H) – extremely rare, but can be produced *in-reactor*



Tritium breeding blankets convert neutron radiation from the fusion plasma into a steady supply of tritium fuel.



However, the tritium breeding ratio (TBR) depends on numerous geometric and material parameters.

TBR evaluation *Paramak* achieves very accurate results by OpenMC Monte Carlo neutronics simulation, 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 designed the Approximate TBR Evaluator (ATE) package to generate training and test datasets from Paramak.

UCL's Hypatia cluster provided the multithreading power for us to produce one million TBR samples, representing 27 days of runtime.

These runs included full evaluations on the 18 continuous and discrete parameters of Paramak, and "slice" evaluations with all discrete parameters frozen.

| | Parameter name | Domain |
|------------|---|--|
| Blanket | Breeder fraction [†] | [0, 1] |
| | Breeder ⁶ Li enrichment fraction | [0, 1] |
| | Breeder material | {Li ₂ TiO ₃ , Li ₄ SiO ₄ } |
| | Breeder packing fraction | [0, 1] |
| | Coolant fraction [†] | [0, 1] |
| | Coolant material | {D ₂ O, H ₂ O, He} |
| | Multiplier fraction [†] | [0, 1] |
| | Multiplier material | {Be, Be ₁₂ Ti} |
| | Multiplier packing fraction | [0, 1] |
| | Structural fraction [†] | [0, 1] |
| First wall | Structural material | {SiC, eurofer} |
| | Thickness | [0, 500] |
| | Armour fraction [‡] | [0, 1] |
| | Coolant fraction [‡] | [0, 1] |
| | Coolant material | {D ₂ O, H ₂ O, He} |
| | Structural fraction [‡] | [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, σ of error, R^2 , R^2_{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:

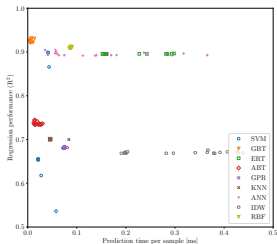
- 1 Hyperparameter tuning (simplified) – Bayesian optimization, discrete features fixed & withheld.
- 2 Hyperparameter tuning – same as #1 but with all features.
- 3 Scaling benchmark – increase training set size.
- 4 Model comparison – train surrogates for practical use.

Experiments 1 & 2: Hyperparameter Tuning

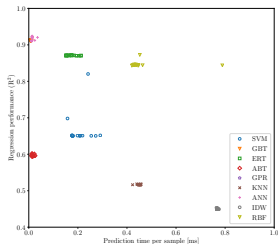
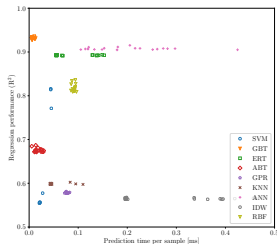
Experiment 1, slice (a)



Experiment 1, slice (b)



Experiment 1, slice (c)

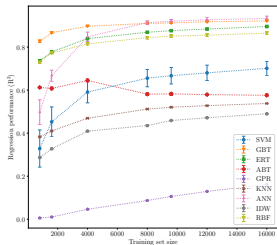


Experiment 2

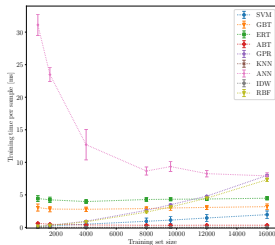
- Plots show $\bar{t}_{\text{pred.}}$ vs. R^2 for 20 best surrogates per family (fastest, most accurate = top left).
- Omitting discrete features yields only a negligible improvement in performance.
- Overall dominated by tree-based surrogates (GBTs, ERTs) and neural networks.

Experiment 3: Scaling Benchmark

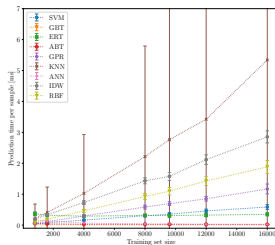
- We observe a hierarchy.
- Best-performing families from the previous experiments also scale the best in $\bar{t}_{\text{pred.}}$.
- More samples: neural networks outperform tree-based models.
- Instance-based surrogates (KNN, IDW) train trivially but have complex lookup.
- Neural networks show inverse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



Regression performance



Training time / sample

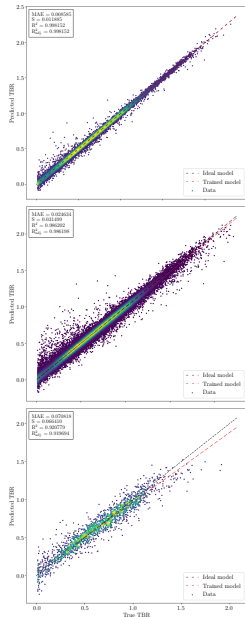


Prediction time / sample

Experiment 4: Model Comparison

- Trained 8 models for practical use.
- Plots show true vs. predicted TBR by Models 1, 2 & 4, coloured by density.
- Best regression performance:
 - Model 1: ANN, 500K samples.
 - $R^2 = 0.998$, $\sigma = 0.013$,
 - $\bar{t}_{\text{pred.}} = 1.124 \mu\text{s}$, $6\,916\,416\times$ faster.
- Fastest prediction:[†]
 - Model 2: ANN, 500K samples.
 - $R^2 = 0.985$, $\sigma = 0.033$,
 - $\bar{t}_{\text{pred.}} = 0.898 \mu\text{s}$, $8\,659\,251\times$ faster.
- Smallest training set:[†]
 - Model 4: GBT, 10K samples.
 - $R^2 = 0.913$, $\sigma = 0.072$,
 - $\bar{t}_{\text{pred.}} = 6.125 \mu\text{s}$, $1\,269\,777\times$ faster.

[†] with acceptable regression performance.

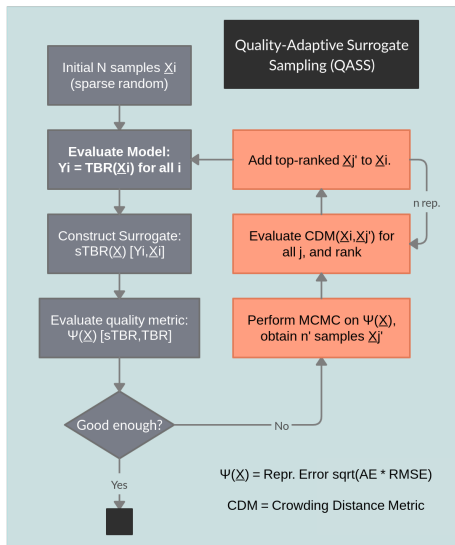


Adaptive Approach

Adaptive sampling takes advantage of surrogate information content *during training* to reduce sample quantity.

We developed a technique which is novel in the literature:

- Construct surrogate quality distribution by nearest-neighbour interpolation.
- Draw candidate samples by quality using MCMC.
- Include samples with high crowding distance.
- Repeat!

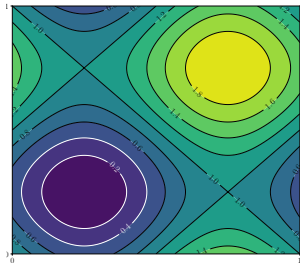


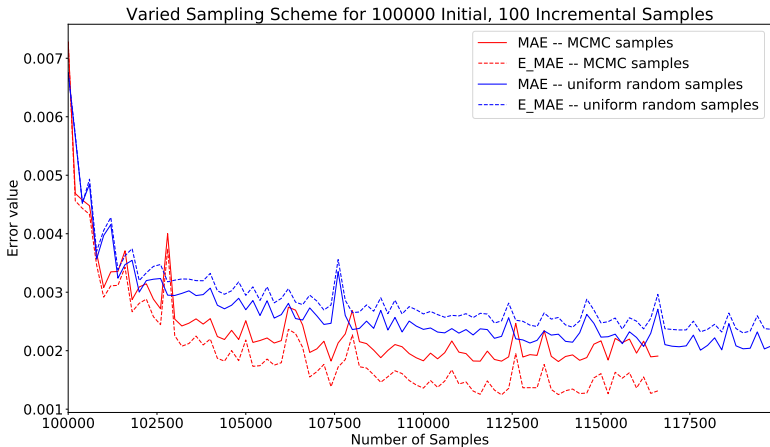
Toy functional TBR theory with wavenumber n , and similar ANN performance to Paramak:

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

Evaluation was performed both on the adaptively-sampled dataset and on a baseline uniform-random dataset.

An baseline scheme with incremental but uniform-random samples was used as a placebo comparison.





Decoupled approach:

- Tuned and compared surrogates from 9 state-of-the-art families.
- Found heuristic: GBTs for $< 10^4$ points and ANNs for $\geq 10^5$ points.
- 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 datapoints, we found surrogates with comparable properties with as little as 10K datapoints.

Adaptive approach:

- TODO
- TODO
- TODO
- TODO

Presented methods portable \rightarrow can be used as cheap approximation of any simulation or black box function.