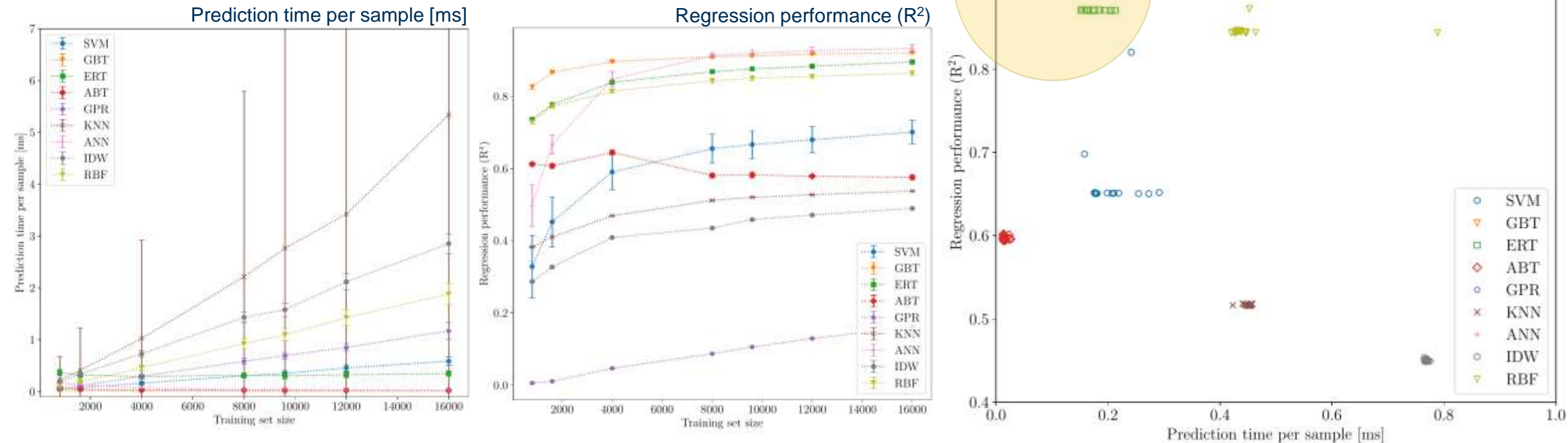


Surrogate modelling of TBR: Methodology

- A decoupled approach:
 - Generated 900K samples from OpenMC simulation (using simplified spherical model).
 - Trained & reviewed a diverse set of surrogate families in a 5-fold cross-validation setting.
- Metrics of interest:
 - Regression performance (as coefficient of determination, R^2),
 - Prediction time per sample (measured as wall time).
- How many datapoints are needed for learning a sufficiently accurate surrogate?



Surrogate modelling of TBR: Results

- 2 clear winners: decision trees (gradient boosted, extremely randomised) and dense neural networks.
- Both are fast: observed relative speedups of order 10^6 w.r.t. multi-threaded TBR MC (on 40 CPUs)
- A proposed heuristic depending on amount of training data available:
 - Small dataset (or has yet to be obtained) \rightarrow DTs provide satisfactory approximation early on in training with as little as $\sim 10K$ points.
 - Data available in large quantities ($\sim 100K+$ points) \rightarrow DNNs overall scale better in R^2 , offer constant prediction time.

