NEA
NUCLEAR ENERGY AGENCY

OECD NEA NSC/WPRS/EGMUP

Benchmark on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering

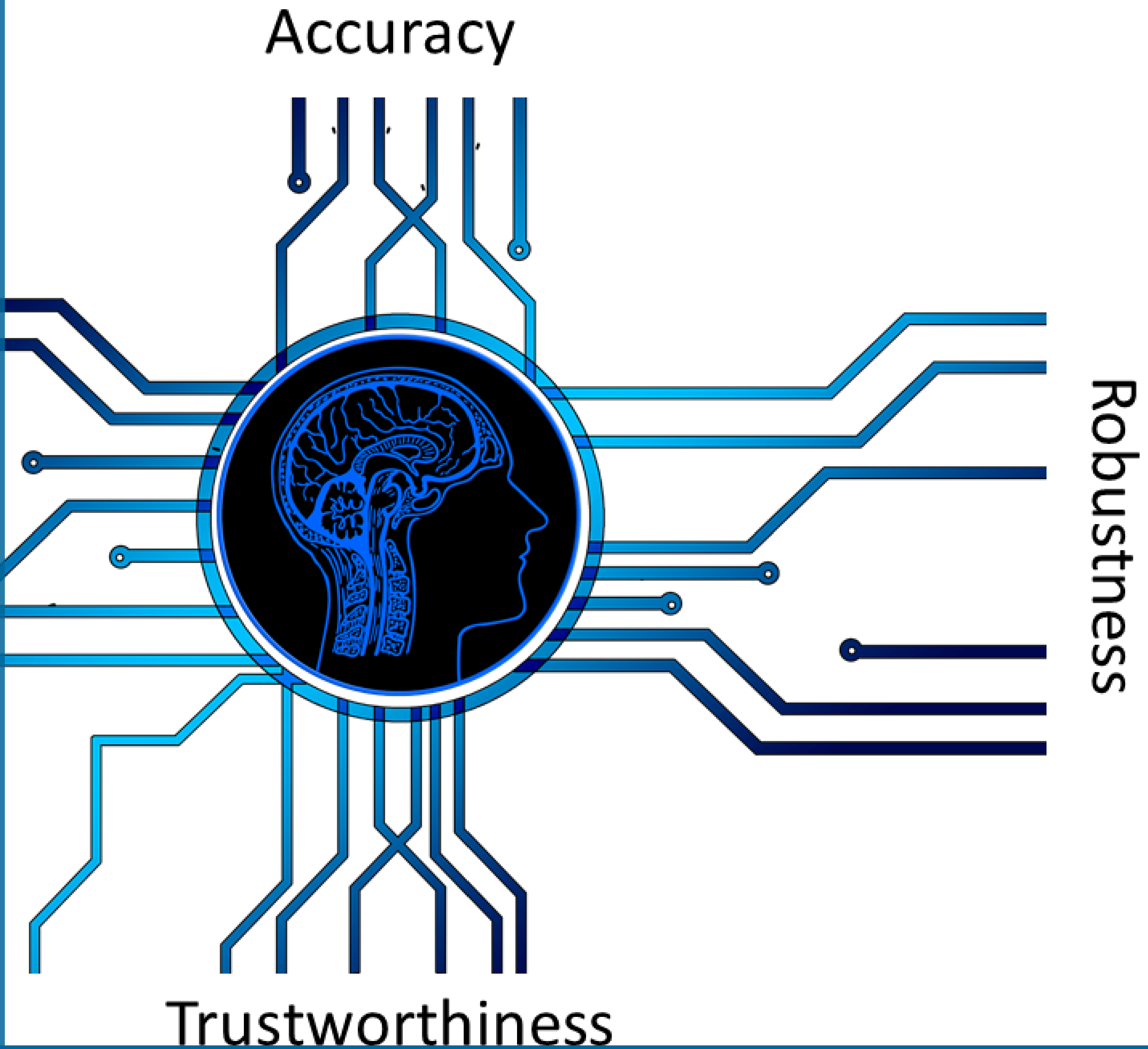
Critical Heat Flux (CHF) Benchmark Exercise - Phase I Results

J.-M. Le Corre
(Westinghouse, SWD)

G. Delipei, X. Wu
(North Carolina State University, USA)

X. Zhao
(University of Tennessee, USA)

NEA Secretariat: O. Buss

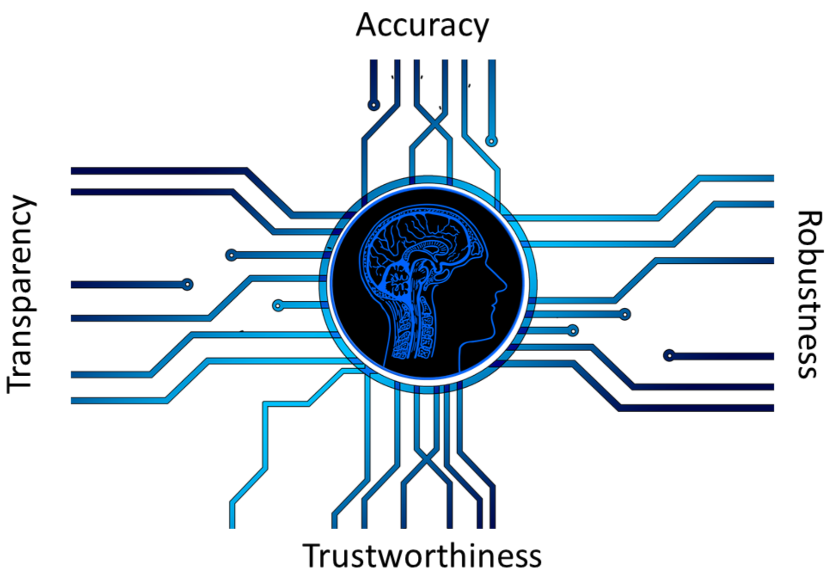


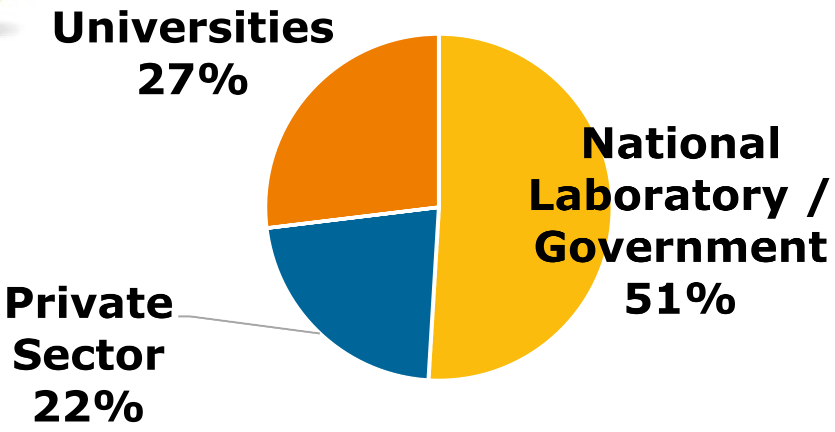
OECD NEA Task Force on AI&ML for Scientific Computing in Nuclear Engineering

Task Force Coordinators: X. Wu and G. Delipei, North Carolina State University (NCSU)

- **Targets:**
  - Build communities of practice to exchange know-how on AI and ML applications
  - Support the development and performance assessment of ML methods
  - Leverage the insights gained from the benchmarks to distill lessons learned and to **provide guidelines for future AI & ML applications** in scientific computing in nuclear engineering.
- **Developed first international AI/ML benchmark for Reactor Core Simulations: Critical Heat Flux (CHF) AI&ML Benchmark**
  - Leverage the NRC CHF database to train data-driven ML models ➔ Test improvement of CHF predictions
    - 47 submitted solutions from 30 institutions, Blind tests for interpolation and extrapolation performance
- **Future:** Benchmarks on **interpretation, classification, and prediction of time series data from research reactors**

NEW **November 2025: Start of benchmark based on PUR-1 data**

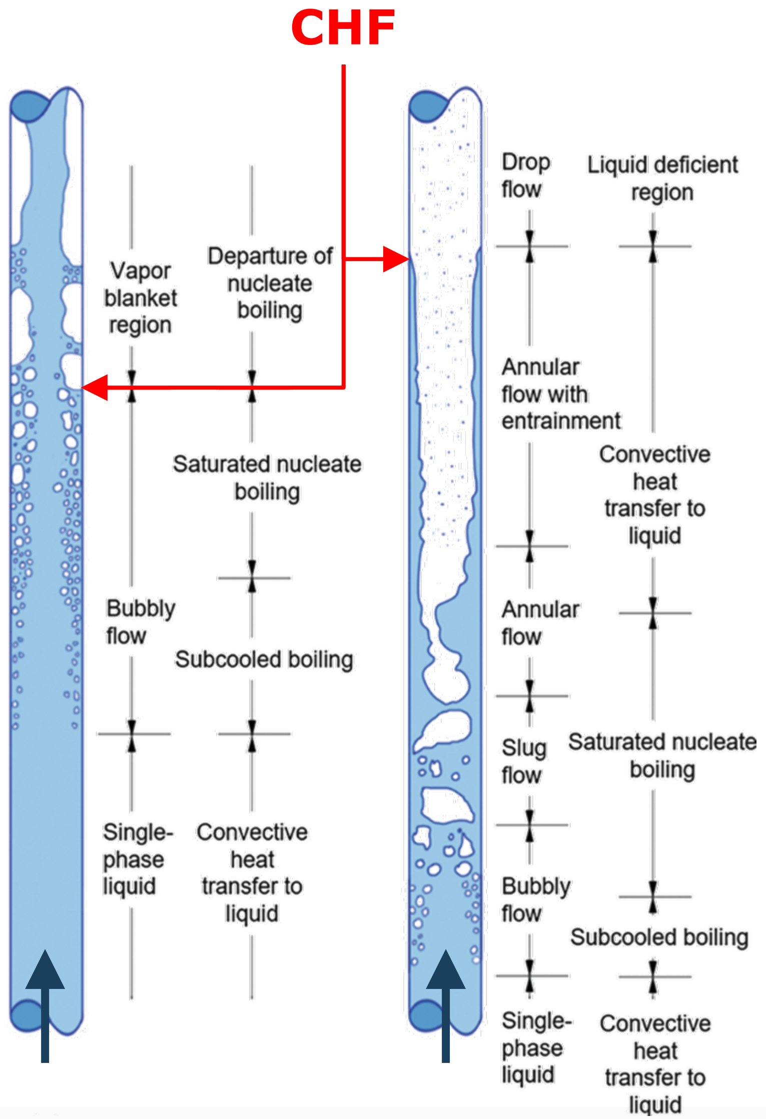




http://oe.cd/wprs-ai

Critical Heat Flux (CHF) in Convective Flow Boiling

- CHF defines a critical wall heat transfer transition
- Beyond CHF a significant increase in wall temperature is expected:
  - Accelerating wall oxidation
  - Potentially fuel rod failure
- Key safety and design criterion in nuclear reactor operation
  - Require costly test data
  - Predicted with (mostly) empirical models



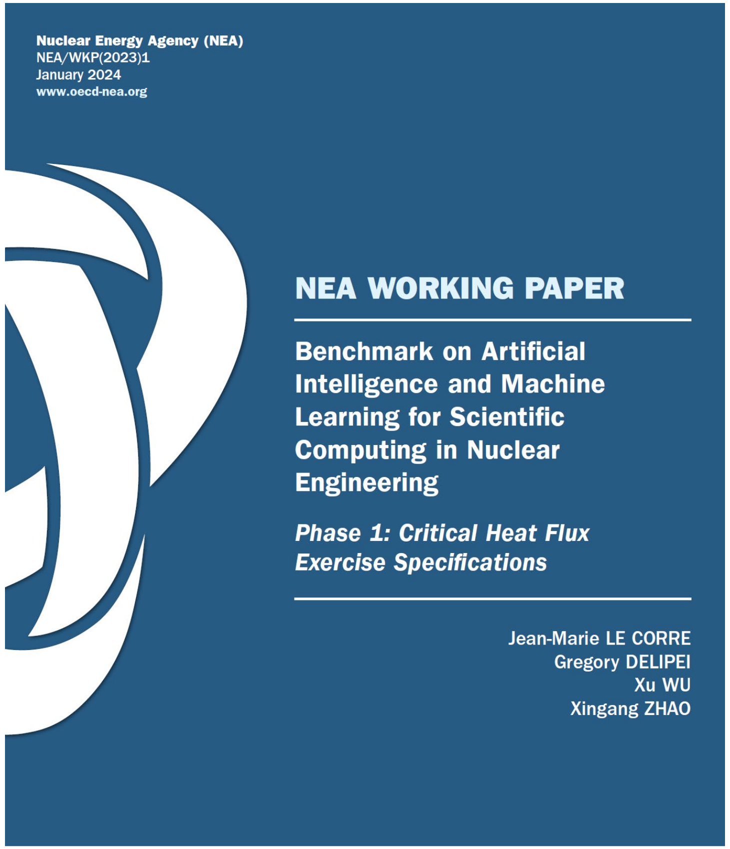
Liu, Y., Dinh, N. (2018). Flow Boiling in Tubes.
In: *Handbook of Thermal Science and Engineering*. Springer, Cham.

Critical Heat Flux Exercises

Coordinators: Dr. Jean-Marie Le Corre (Westinghouse Electric Sweden, benchmark lead), Dr. Xingang Zhao (UTK), Dr. Xu Wu and Dr. Gregory Delipei (NCSU)

- **Phase 1:**
  - **Regression & Feature Analysis**
  - Analyses
    - Statistical analysis and comparison to blind data
    - Extrapolation vs. interpolation performance
  - Schedule
    - Kickoff in November 2023
    - Submission dateline, August 2024
    - Preliminary results, November 2024

- **Phase 2:**
  - **Uncertainty Quantification**
  - Transfer learning: investigate if the ML models have predictive power for unseen geometries and conditions
  - Fuel bundle benchmark
  - To be started in 2025


Nuclear Energy Agency (NEA)
NEA/WKP(2023)1
January 2024
www.oecd-nea.org

NEA WORKING PAPER

Benchmark on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering

Phase 1: Critical Heat Flux Exercise Specifications

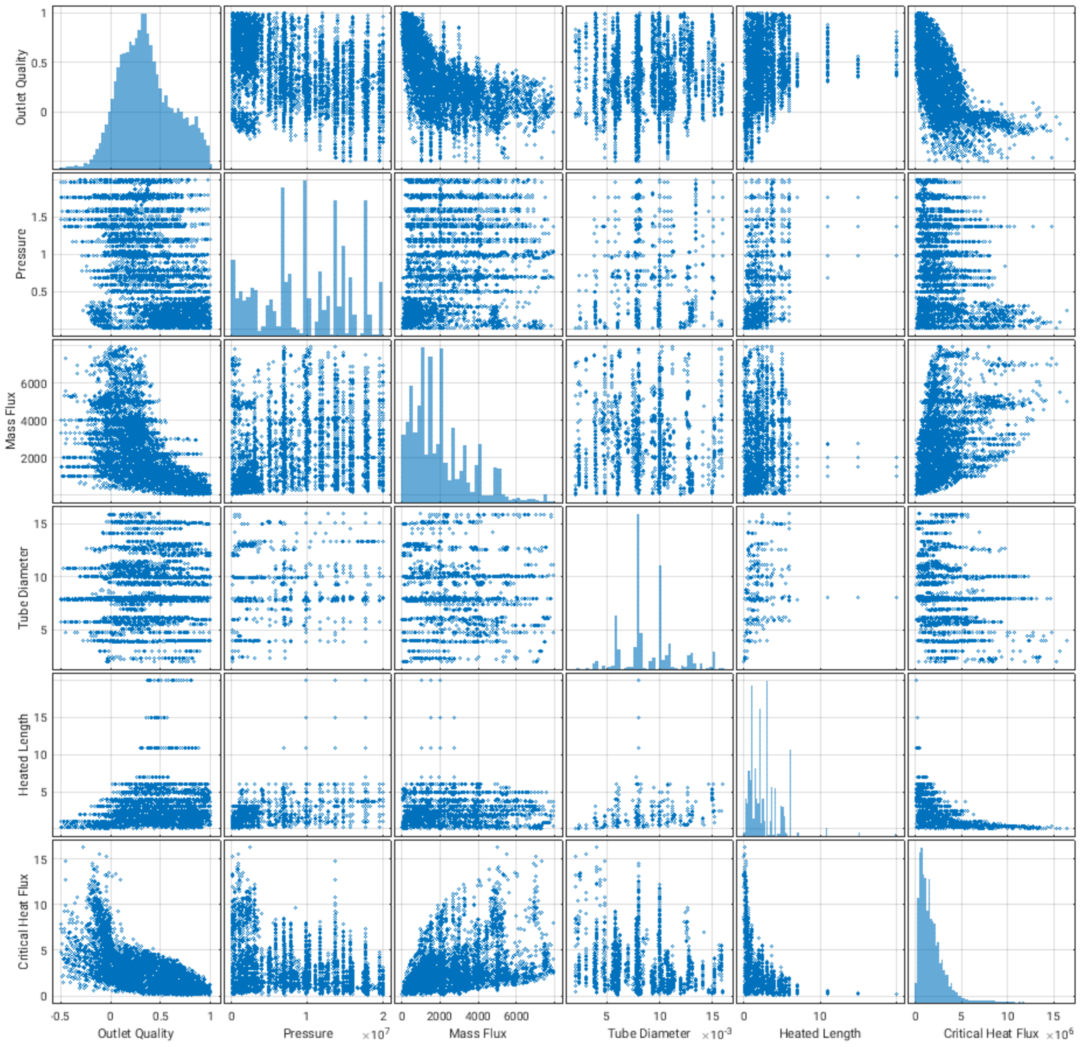
Jean-Marie LE CORRE
Gregory DELIPEI
Xu WU
Xingang ZHAO

Public Database – Overview

- Based on 2006 CHF lookup table data made available by US NRC
  https://www.nrc.gov/reading-rm/doc-collections/nuregs/knowledge/km0011/index.html

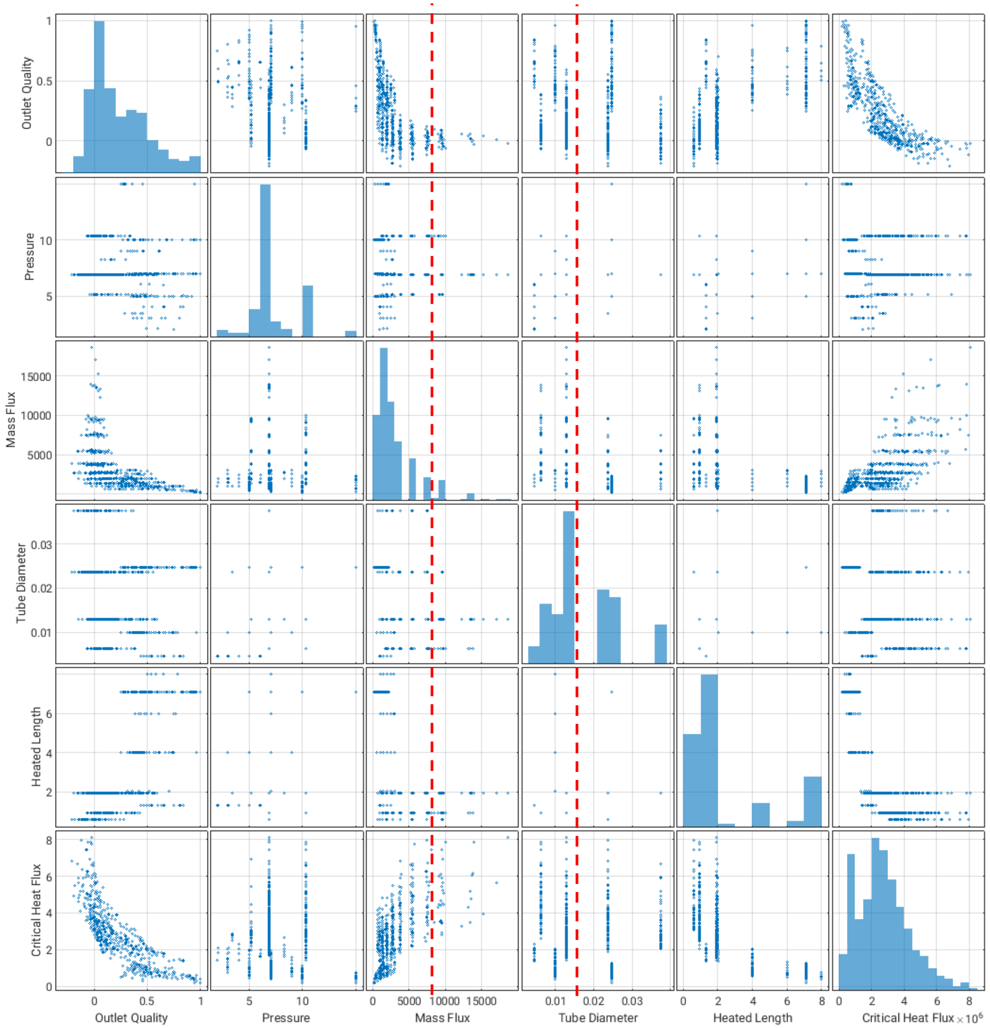
- Total of 24579 CHF data points
  - 59 sources, collected over 60 years
  - Vertical tubes, steady-state, uniformly heated
    → Outlet CHF only (0-D problem)

Variable	CHF [kW/m²]	P [kPa]	G [kg/m²/s]	X [-]	D [m]	L [m]
Min value	50	100	8.2	-0.497	0.002	0.05
Max value	16339.3	20000	7694	0.999	0.016	20



Blind Database – Overview

- 560 data points from 4 datasets
- About 250 data points outside NRC database range
  - D > 16 mm (219)
  - G > 8000 kg/m²/s (34)
- 1 interpolation and 2 extrapolation regions

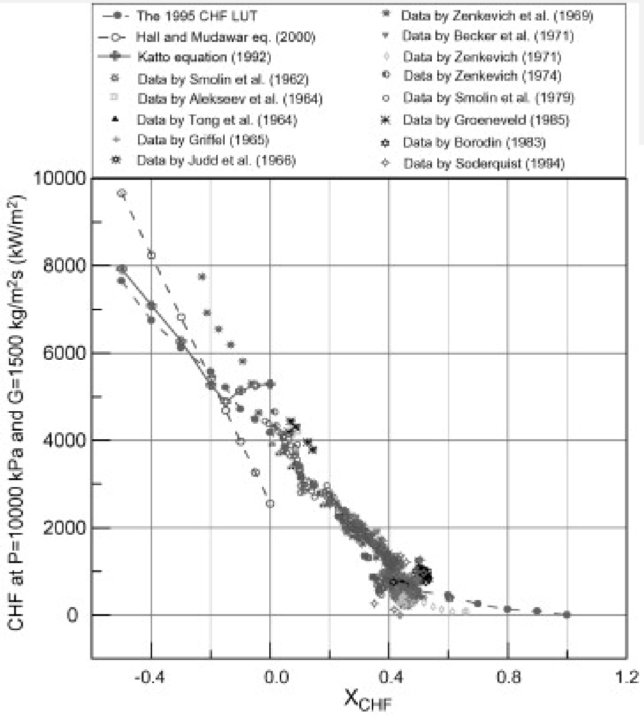


Variable	CHF [kW/m²]	*P* [kPa]	*G* [kg/m²/s]	*X* [-]	*D* [m]	*L* [m]
Min value	204	2040	179	-0.209	0.0046	0.61
Max value	8101.9	15000	18580	0.998	0.0375	8

Extrapolation

Slice Datasets – Overview

- Keep everything constant except one parameter



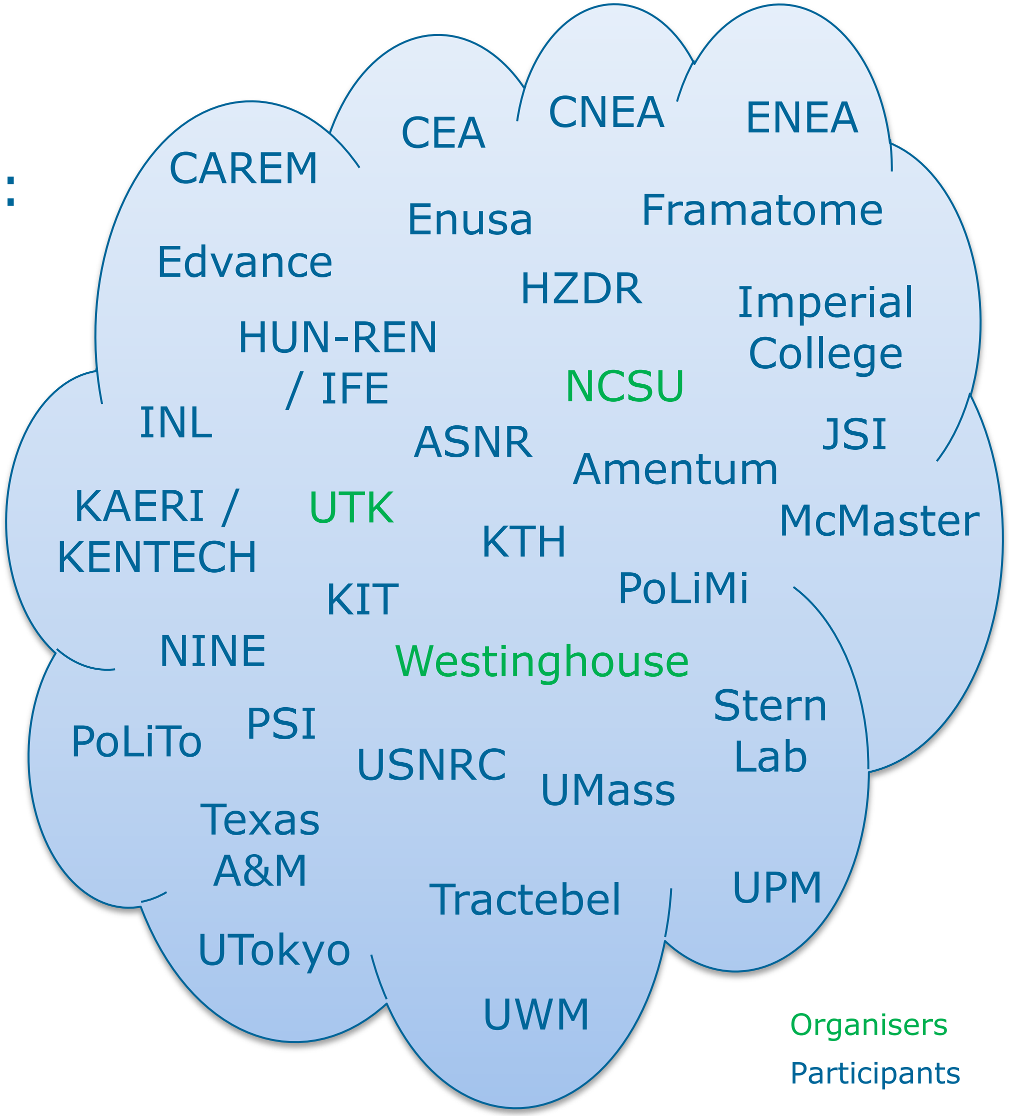
- Example of varying quality
- For each parameter (*D*, *L*, *P*, *G*, *X*)
  - Two regions are selected (10 datasets)
  - Arbitrary 15 equidistant values
- Allow to easily assess:
  - Correct physical behavior
  - Overfitting

Slice #	D [mm]	L [m]	P [kPa]	G [kg/m^2/s]	X [-]	Data file
1	**0 - 16**	6.000	14701	998.5	0.391	`Slice_01.csv`
2	**0 - 16**	6.000	9807	1003.3	0.529	`Slice_02.csv`
3	8.01	**0 - 20**	9806	1000.0	0.587	`Slice_03.csv`
4	8.11	**0 - 20**	2009	752.2	0.756	`Slice_04.csv`
5	8.00	0.998	**0 - 20000**	2006.0	0.140	`Slice_05.csv`
6	13.40	3.658	**0 - 20000**	20040.2	0.378	`Slice_06.csv`
7	8.00	1.570	12750	**0 - 8000**	0.144	`Slice_07.csv`
8	10.00	4.966	16000	**0 - 8000**	0.343	`Slice_08.csv`
9	8.14	1.943	9831	1519.5	**-0.5 - 1.0**	`Slice_09.csv`
10	8.00	0.997	17650	2002.7	**-0.5 - 1.0**	`Slice_10.csv`

Phase 1 Results Submission – Overview

- 48 models submitted from 31 institutions
- Various ML algorithms used, broadly categorized as:
  - Neural network based (30)
  - Tree based (17)
  - Gaussian Process (1)

  + 2 Reference models (from benchmark specifications)

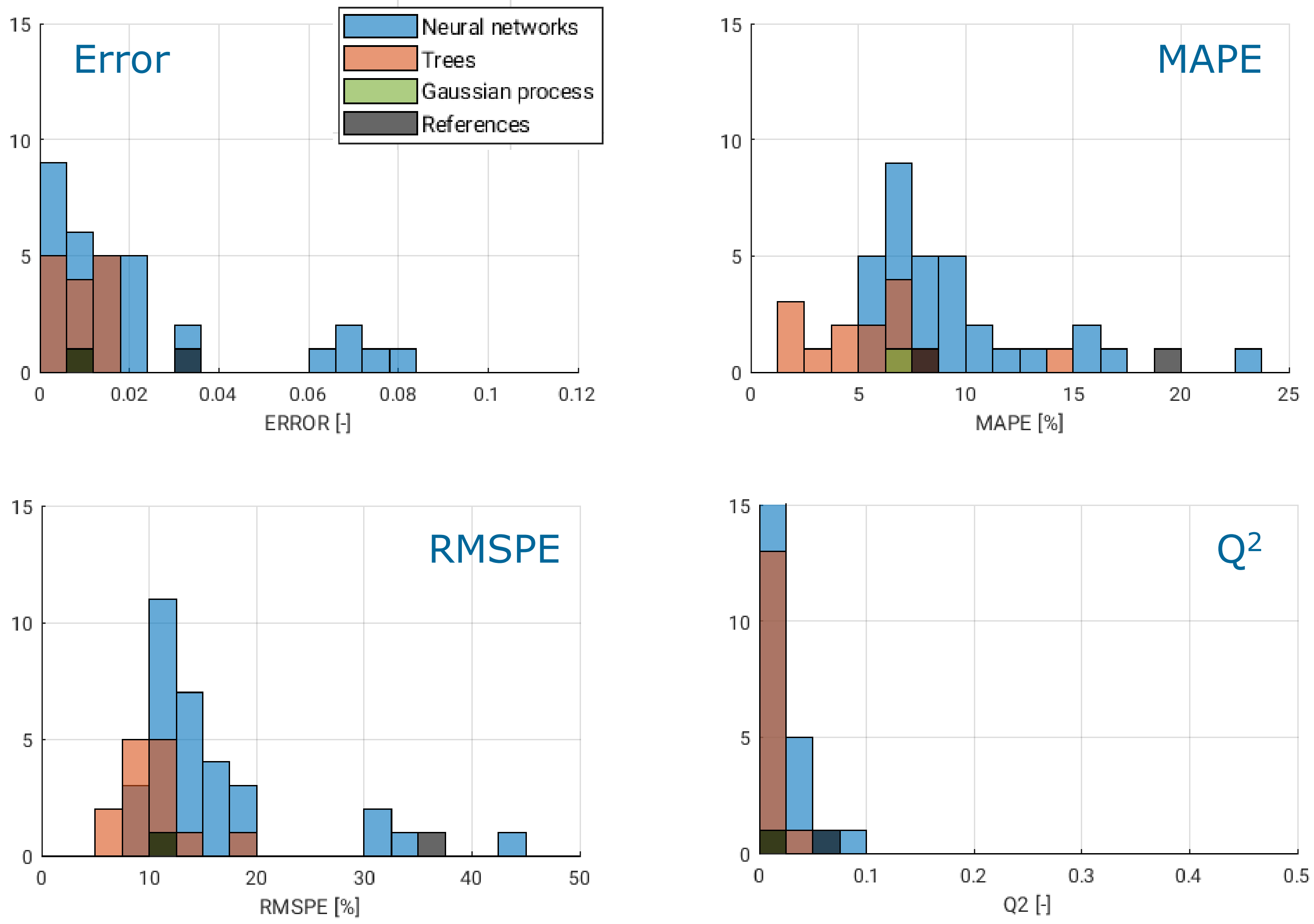


Dimensionality Analysis

- Optional task but highly recommended
- Not considered by most participants
- Most models: $CHF = f(D, P, G, X, L)$
- Limited use of non-dimensional parameters
- Limited use of data driven methods
- Did not provide conclusive results

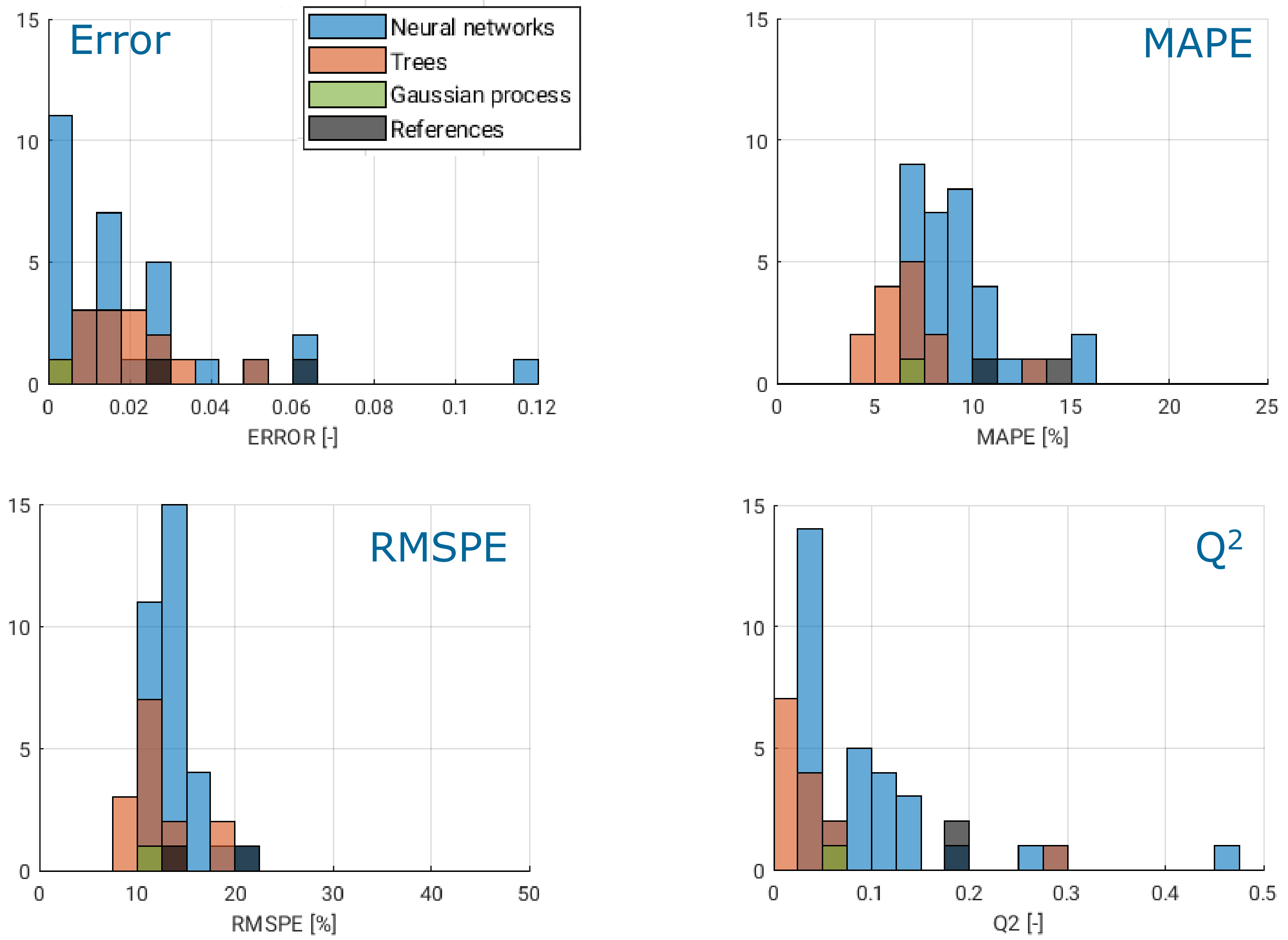
Public Database – Summary of Models Performance

- Color-coded P/M CHF performance comparison
- Up to 3 to 5 times lower metrics as compared to reference CHF LUT
- Tree-based models generally seems to perform better



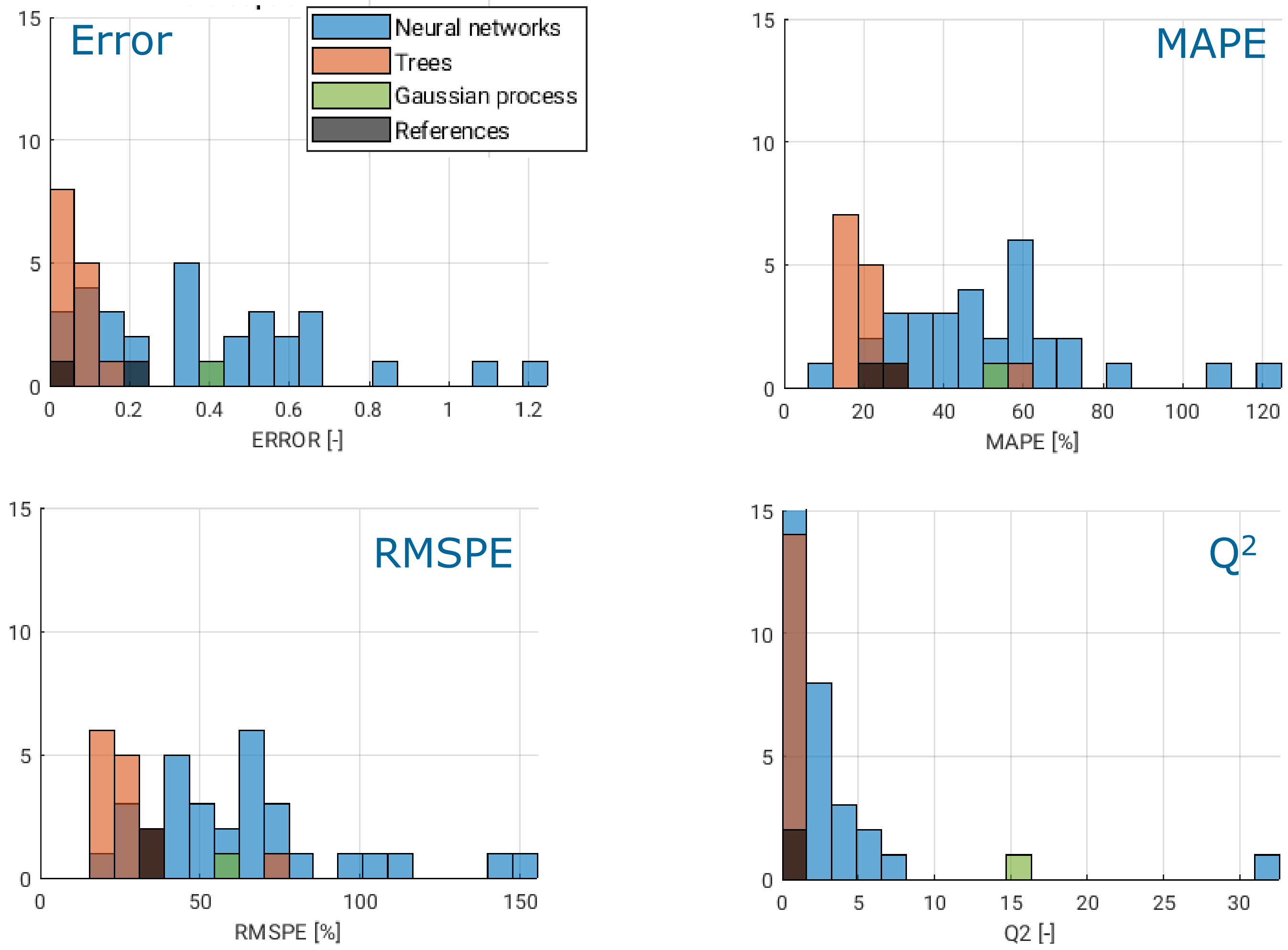
Blind Database – Summary of Models Performance (1)

- Interpolation region
  - Up to 2 to 3 times lower metrics as compared to LUT



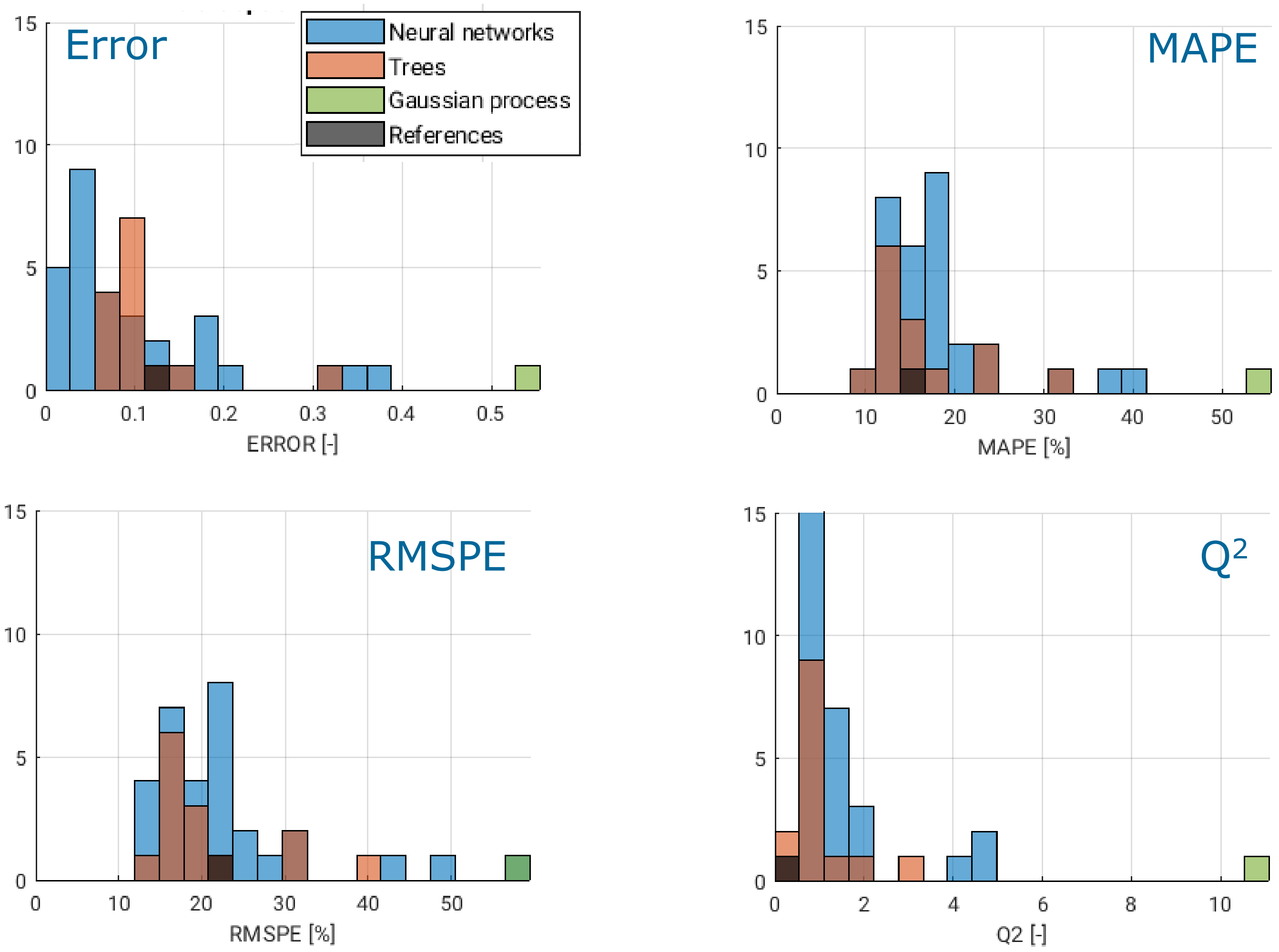
Blind Database – Summary of Models Performance (2)

- Extrapolation region #1
  - D > 16 mm
  - Up to 2 times lower metrics as compared to reference CHF LUT



Blind Database – Summary of Models Performance (3)

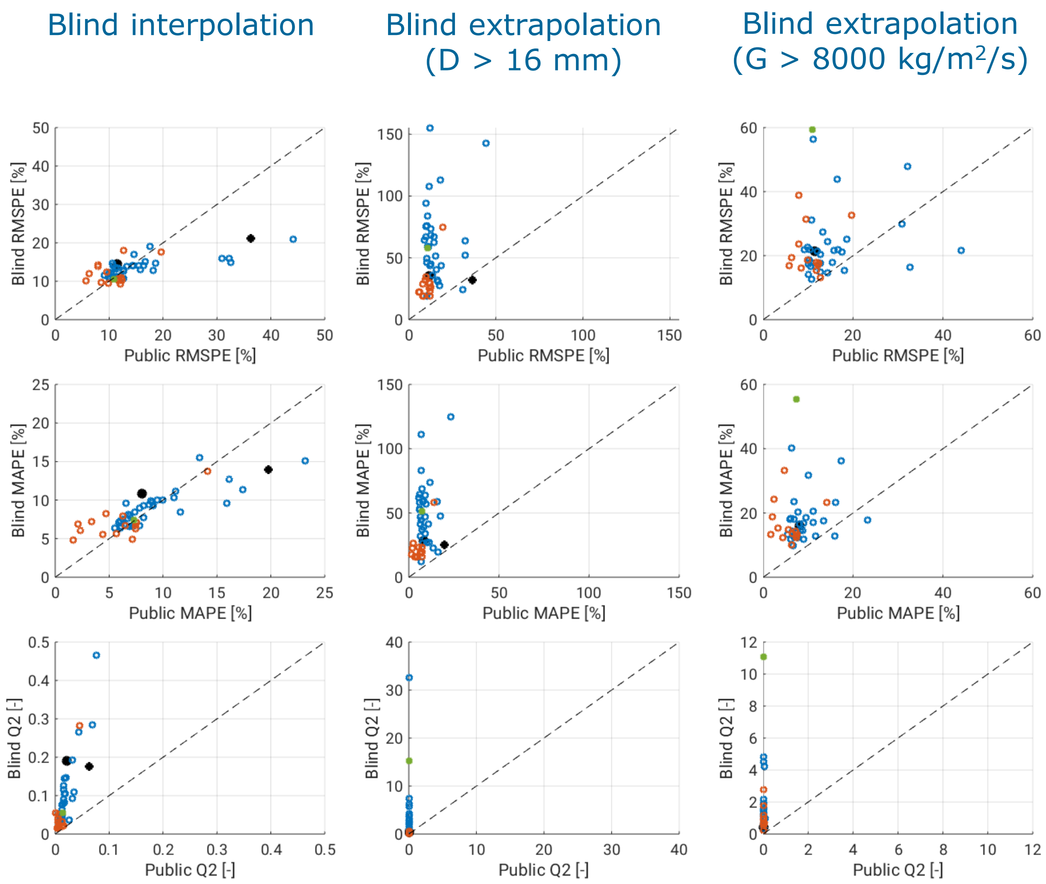
- Extrapolation region #2
  - G > 8000 kg/m²/s
  - No possible comparison with reference CHF LUT



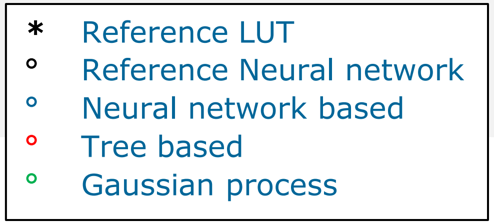
Statistical Metrics – Blind vs Public Databases

* Reference LUT
o Reference Neural network
o Neural network based
o Tree based
o Gaussian process

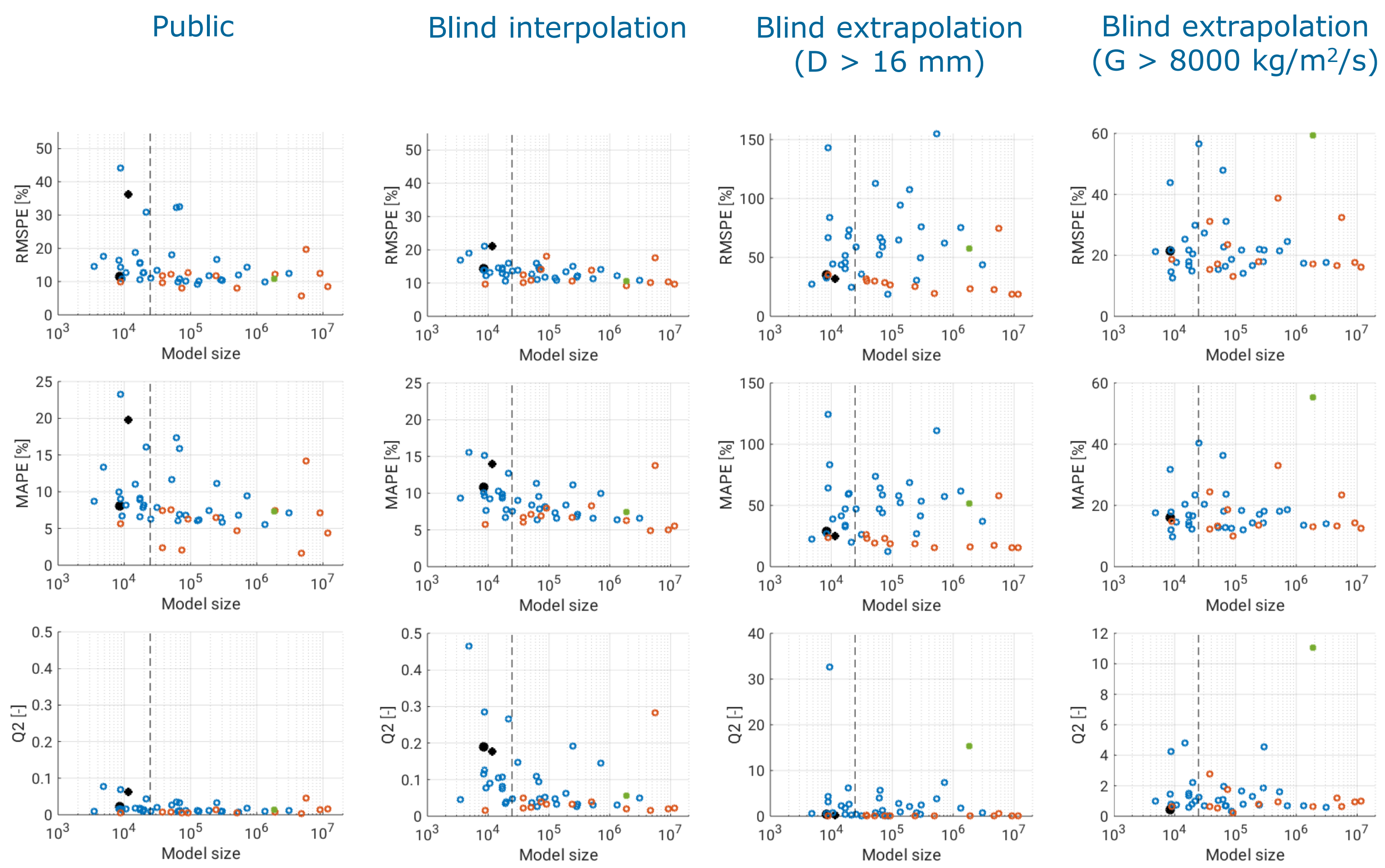
- ***How are the models performing when applied to the blind database, as compared to the public database?***
- Blind interpolation
  - Good behavior
  - General lower discrepancies with tree-based models
  - Improvements for models with large discrepancies
  - Degradation in Q^2 error (due to ≠ range of data)
- Extrapolation
  - Expected significant degradation
  - Higher degradation when extrapolating to large *D* (NN)
  - Tree-based models seems to extrapolate better



Statistical Metrics vs Model Size



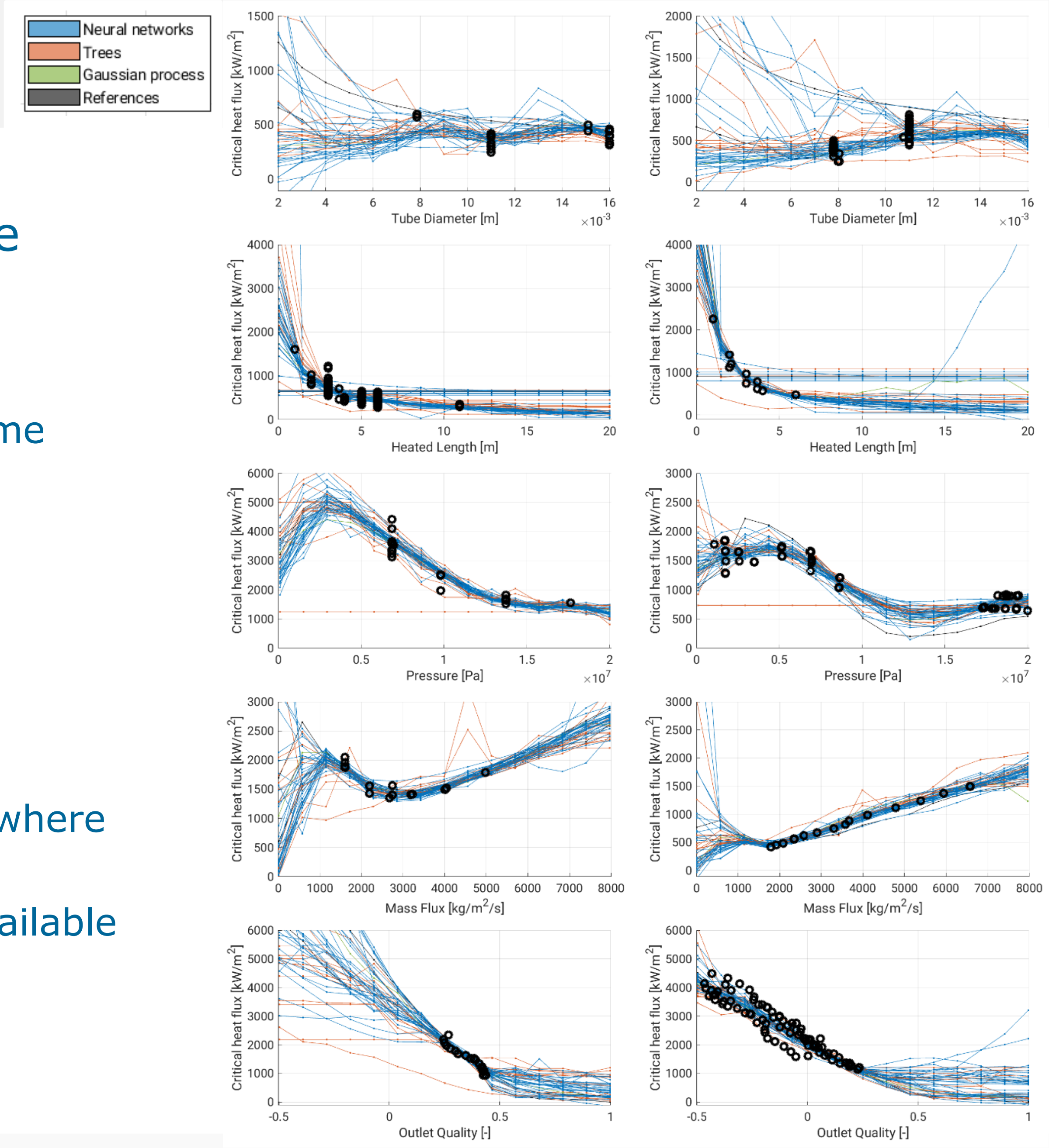
- ***How are the models performing with respect to their size?***
  - Note that most model sizes are larger than the training database (24579)…
  - … but regularization can help
- Public & blind interpolation
  - Performance increases with model size
  - However, smaller models with good performance may be of higher interest
- Blind extrapolation
  - Expected significant degradation
  - Tree-based models seems to extrapolate better, including for large size



Slice Datasets Results

- All submitted results plotted for each slice
  - Along with corresponding exprimental points (**o**)
  - Some models are independant of *L*
  - Strange constant results (with *L*, *P* and *G*) for some models → see Gitlab issue board

- Trends
  - No obvious over-fitting for most models
  - Larger variability wih *D*
  - Mostly consistent behavior with *G* and *X* (and *L*) where data is available
  - Large (epistemic) variability where data is not available (incl. within data range)



Initial Plans for Phase 2

Coordinators: Dr. Jean-Marie Le Corre (Westinghouse Electric Sweden, benchmark lead), Dr. Xingang Zhao (UTK), Dr. Xu Wu and Dr. Gregory Delipei (NCSU)

- **Phase 2:**
  - Benchmark specifications to be issued by end of 2025
  - New participants are welcome to join
  - Several CHF databases to be provided in specifications
    - Annulus
    - Tube with non-uniform axial power
    - Rod bundle (later)
    - Potentially: additional blind datasets
  - Three main tasks
    - Advanced uncertainty quantification
    - Transfer learning to more complex geometries
    - Model interpretability and explainability

Thank you for your attention

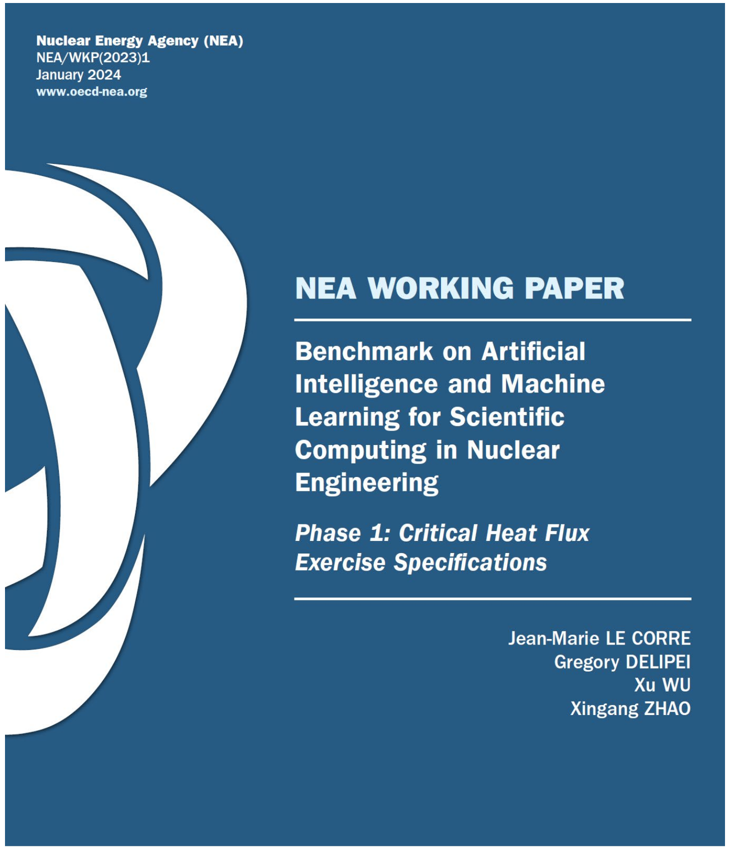
Please contact wprs@oecd-nea.org if you have any questions or comments regarding this presentation.

Backup slides

Critical Heat Flux Exercises (2)

Coordinators: Dr. Jean-Marie Le Corre (Westinghouse Electric Sweden, benchmark lead), Dr. Xingang Zhao (UTK), Dr. Xu Wu and Dr. Gregory Delipei (NCSU)

- **Phase 1:**
  - Benchmark specifications issued in January 2024
  - Three databases provided in specifications
    - Public database (24579 points)
    - Blind database (560 points)
    - Slice datasets (10*16 points)
  - Three main tasks
    - Dimensionality analysis
    - Selection and training of regression models
    - Blind and slice applications
  - Benchmark results documentation
    - WPRS benchmark meeting (July 2025)
    - NEA working paper (2025) – on-going
    - NURETH-21 paper & presentation (Sept. 2025)
    - Journal paper


Nuclear Energy Agency (NEA)
NEA/WKP(2023)1
January 2024
www.oecd-nea.org

NEA WORKING PAPER

Benchmark on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering

Phase 1: Critical Heat Flux Exercise Specifications

Jean-Marie LE CORRE
Gregory DELIPEI
Xu WU
Xingang ZHAO

OECD BETTER POLICIES FOR BETTER LIVES
NEA NUCLEAR ENERGY AGENCY

List of Submitted Models

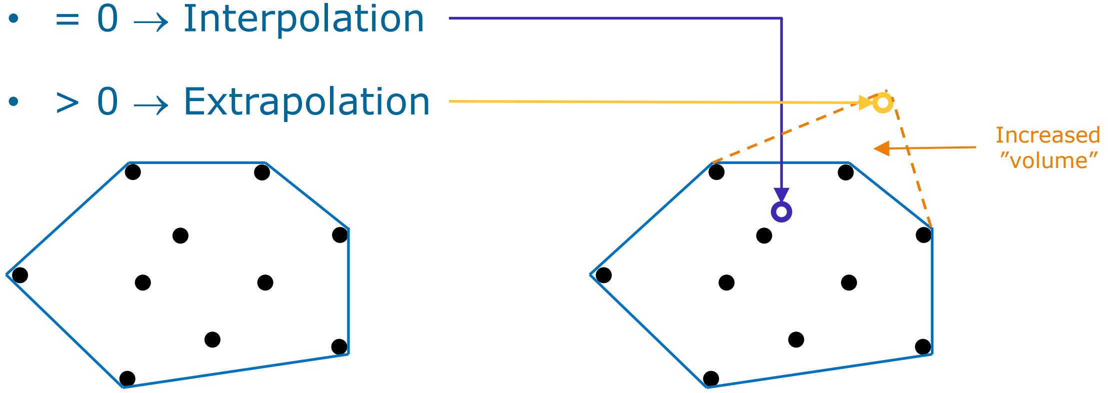
Reference models	(2)
Neural network based	(32)
Tree based	(15)
Gaussian Process	(1)

Geometrical parameters		**Saturated fluid properties**		
D	Diameter	k_l	Liquid thermal conductivity	
L	Length	Lo	Laplace length	$\left(\sigma/9.8\cdot(\rho_f-\rho_g)\right)^{0.5}$
S	Heated surface	σ	Surface tension	
		ρ_f	Liquid density	
Fluid parameters		ρ_g	Vapor density	
P	Pressure	μ_f	Liquid viscosity	
G	Mass flux	μ_g	Vapor viscosity	
X	Equilibrium outlet quality			
H_{out}	Outlet enthalpy	**Non-dimensional numbers**		
T_{in}	Inlet temperature	D^*	Diameter	$D/8$ mm
CHF	Critical Heat Flux	Re_f	Liquid Reynolds number	$G\cdot D/\mu_f$
		Re_g	Vapor Reynolds number	$G\cdot D/\mu_g$
Acronyms		We	Weber number	$G^2\cdot D/(\sigma\cdot\rho_f)$
MAPE	Mean Absolute Percentage Error	N_{rho}	Density ratio	ρ_f/ρ_g
PCs	Principal components	Ah	Ahmad number	$We^{1.333}\cdot Re_f^{-0.133}\cdot Re_g^{-0.2}$
RMSPE	Root Mean Square Percentage Error	Bo	Boiling number	$CHF/(G\cdot h_{fg})$
STD	Standard deviation			

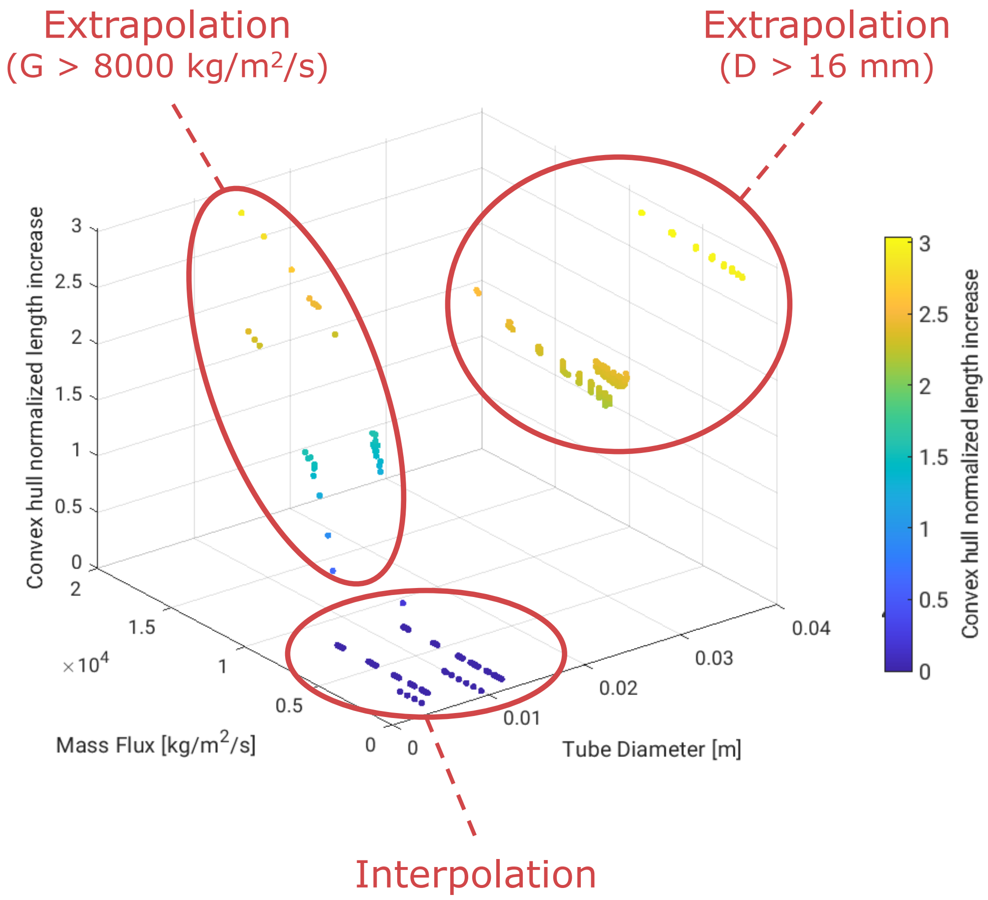
		ID	Algorithm	Inputs	Output
CHF look up table	1	LUT	Look up table	D, P, G, X	CHF
Westinghouse / Uppsala University	2	NN	Neural network	D, L, P, G, X	CHF
CAREM-CNEA	3	CAREM	Neural network	D, L, P, G, X	CHF
CEA Saclay	4	CEA	Residual neural network	D, L, P, G, X	CHF
Comisión Nacional de Energía Atómica	5	CNEA1	Neural network *2	D^*, G, X, Re_f, We, k_l, N_{rho}	CHF
	6	CNEA2	XGBoost	D^*, G, X, Re_f, We, k_l, Lo	$CHF/\sigma^{0.5}$
Edvance	7	EDVANCE	Neural network	D, P, G, X, H_{out}, S	CHF
ENEA	8	ENEA	Neural network	D, L, P, G, X	CHF
Enusa	9	ENUSA	Neural network	D, L, P, G, X	CHF
Framatome	10	FRAMATOME1	Bayesian additive regression trees	6 PCs	log(CHF)
	11	FRAMATOME2	Neural network	6 PCs	CHF
	12	FRAMATOME3	Neural network	6 PCs	CHF
HUN-REN Center for Energy Research / Institute for Energy Technology (IFE)	13	HUNREN&IFE1	Neural network	D, L, P, G, X	CHF
	14	HUNREN&IFE2	XGBoost	D, L, P, G, X	CHF
Helmholtz-Zentrum Dresden-Rossendorf	15	HZDR1	XGBoost	D, L, P, G, Tin	CHF
	16	HZDR2	XGBoost	D, L, P, G, X	CHF
Imperial College London	17	IMPERIAL	Deep gaussian process	D, L, P, G, X	CHF
Idaho National Laboratory	18	INL1	Random forest (1 tree)	D, L, P, G, X	CHF
	19	INL2	Random forest (2 trees)	D, L, P, G, X	CHF
	20	INL3	Kolmogorov–Arnold network	D, L, P, G, X	CHF
Autorité de sûreté nucléaire et de radioprotection	21	ASNR	XGBoost	D, L, P, G, X	CHF
Amentum	22	AMENTUM1	Neural network	D, L, P, G, X	CHF
	23	AMENTUM2	Neural network	D, L, P, G, X	CHF
	24	AMENTUM3	Neural network	D, L, P, G, X	CHF
Jozef Stefan Institute	25	JSI1	CatBoost	D, L, P, G, X	CHF
	26	JSI2	Gradient boosting	D, L, P, G, X	CHF
	27	JSI3	XGBoost	D, L, P, G, X	CHF
Korea Atomic Energy Research Institute / Korea Institute of Energy Technology	28	KAERI1	Dense neural network	D, L, P, G, X	CHF
	29	KAERI2	XGBoost	L/D, ρ_f/ρ_g, $We^{0.5}$, X	Bo
Karlsruhe Institute of Technology	30	KIT	Neural network	$\ln(\rho_f/\rho_g)$, ln(L/D), X, ln(Ah)	ln(Bo)
Royal Insitute of Technology	31	KTH1	Ennsemble neural network	D, L, P, G, X	CHF
	32	KTH2	Ensemble physics-informed NN	D, L, P, G, X	CHF
McMaster University	33	MCMASTER1	Neural network (standard)	D, P, G, X	CHF
	34	MCMASTER2	Neural network (normalized)	D, P, G, X	CHF
North Carolina State University	35	NCSU	Extremely randomized forest	D, L, P, G, X	CHF
NINE	36	NINE1	CatBoost	D, L/D, P, G, X	CHF
	37	NINE2	Neural network	D, L/D, P, G, X	CHF
US Nuclear Regulatory Commission	38	NRC	Neural network	D, L, P, G, X	CHF
Politecnico di Milano	39	POLIMI1	Convolutional neural network	D, L, P, G, X	CHF
	40	POLIMI2	Convolutional neural network	D, L, P, G, X	CHF
Politecnico di Torino	41	POLITO	Neural network ensemble	D, L, P, G, X	CHF
Paul Scherrer Institut	42	PSI	Neural network	D, L, L/D, P, G, X	CHF
Stern Laboratories	43	STERN	Neural network	D, L, P, G, X	CHF
Texas A&M University	44	TAMU	Neural network	D, L, P, G, X	CHF
Tractebel	45	TRACTEBEL1	Neural network	D, L, P, G, X	CHF
	46	TRACTEBEL2	Residual neural network	D, L, P, G, X	CHF
University of Massachusetts Lowell	47	UMASS	Bayesian neural network	D, L, P, G, X	CHF
Universidad Politecnica de Madrid	48	UPM	Random forest	D, L, P, G, X	CHF
The University of Tokyo	49	UTOKYO	Transformer	D, L, P, G, X	CHF
University of Wisconsin - Madison	50	UWM	Neural network	D, L, P, G, X	CHF

Blind Database – Interpolation vs Extrapolation Regions

- Convex hull approach
  - Smallest convex enveloppe enclosing database
  - Convex hull (5-D) ”volume” increase for each blind data
    - = 0 → Interpolation
    - > 0 → Extrapolation



- Applied to entire blind database
  - Identify ”true” interpolated points
  - Quantify level of extrapolation for remaining points ($L = V^{\frac{1}{5}}$)



Dimensionality Analysis (1)

- Optional task but highly recommended
- Not considered by most participants, except (to some extent):
  - CNEA
  - FRAMATOME
  - KIT
  - EDVANCE
  - KAERI
- Most models: $CHF = f(D, P, G, X, L)$
- Use of non-dimensional parameters (CNEA, KAERI, KIT) + *L*/*D* (NINE, PSI)
  - CNEA: trial & error ("limited success")
  - KAERI & KIT: fluid-to-fluid scaling considerations
- Not data driven, except
  - EDVANCE: Feature ranking & engineering
  - FRAMATOME: PCA

Dimensionality Analysis (2)

- Use of logarithms (KIT)
  - Provide more normal distribution of parameters
  - Improve model performance
- Heated length (*L*)
  - Clear correlation (but no causation!) in database
  - Used by all models, except CNEA & MCMASTER (but not replaced by other ”non-local” parameter)
  - Attempt to address this issue with PINN (KTH)

Model Optimization and Training

- Hyperparameter optimization
- Training strategies
- Documented in upcoming CHF bechmark Phase I results report

Conclusions for Phase 1

- ML/AI CHF benchmark has **48** submissions from **31** institutions (+ 2 reference models)
- Post-process of phase 1 on-going
- Preliminary conclusions
    - Up to factor 2 or 3 reduction in statistical metrics (as compared to LUT)
    - Limited success from dimensionality analysis
    - Tree-based algorithms seems to perform better
    - No obvious over-fitting
    - Good blind performance within interpolation range
    - Some models can reasonably extrapolate

Publications related to Phase 1

- List of benchmark related publications
  - Please inform benchmark coordinators about publications (wprs@oecd-nea.org)
  - https://git.oecd-nea.org/science/wprs/egmup-task-force-on-ai-and-machine-learning/phase1-chf-exercises/-/blob/main/publications/README.md
- NURETH-21 special sessions
  - 1 summary paper submitted by organisers
  - Many others from participants + related activivites based on US NRC database
- NEA working paper (2025)
  - Document generic results
  - All detailed results (per model) in appendices
- Journal publications
  - 1 summary paper in a selected journal
  - Other individual publications

Initial Plans for Phase 2 (2)

- Advanced uncertainty quantification
  - Field still under development
  - Example for deep NN: dropout-based, ensemble-based and Bayesian NN-based
  - Associate specific epistemic & aleatory uncertainties to each prediction
  - Demonstration using unseen CHF data
- Transfer learning to other geometries
  - Transfer of knowledge from one database to the other to improve performance
  - Example: Tube database → Annulus database → Rod bundle database
- Model interpretability vs explainability
  - Interpretability: Better understand how predictions are generated – *Model transparency*
  - Explainability: Discover meaning between inputs and outputs – *Understand behavior*

Initial Plans for Phase 2 (3)

- Fuel bundle benchmark
  - BFBT, PSBT and EPRI (11077 data points) CHF databases
  - Preliminary step: Generate local TH parameters (on-going in collaboration with UWM)
  - Combine all previously acquired knowledge from Phases 1 and 2
  - Assessment for reactor safety analysis (?)