W7-X Physics Meeting

Prediction of Neutral Gas Pressure in Wendelstein 7-X: Statistical Analysis and Machine Learning

D. Angelis, F. Sofos, S. Misdanitis, C. Dritselis, T. E. Karakasidis, D. Valougeorgis, V. Haak, D. Naujoks, G. Schlisio, S. A. Bozhenkov, V. Perseo, W7-X Team

Dimitrios Angelis

University of Thessaly

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Contents

Data preprocessing

Juice database - OP1.2b experimental campaign

- *Standard* (EJM) ⟶ *AEI port*
- *High iota (FTM)* ⟶ *AEP port*
- *High mirror (KKM)* ⟶ *AEI port*

- n_e $[10^{19}\ m^{-2}]$
- P_{heat} $[MW]$
- P_{rad} $[MW]$
- I_{cc} $[kA]$
- I_{tor} $[kA]$
- T_c $[keV]$
- T_e $[keV]$

} $p_n\ [10^{-6}\ mbar]$

$f_{rad} \in (0, 0.8)$

- **E**lectron **C**yclotron **R**esonance **H**eating
- Without:
  - neutral beam injection
  - pellet and impurity injection

Table 1: Range, mean and standard deviation of measured neutral gas pressure and plasma parameters for the standard (AEI port), high iota (AEH) and high mirror (AEI) configurations.

	p_n $[10^{-6}$ mbar$]$	n_e $[10^{19}$ m$^{-2}]$	P_{heat} [MW]	P_{rad} [MW]	I_{tor} [kA]	I_{cc} [kA]	T_c [keV]	T_e [keV]
Standard - AEI (No. of datapoints: 8856)								
Min	7.60	0.82	0.34	0.05	-2.73	-0.50	0.33	0.10
Max	1138.04	13.52	7.06	4.97	10.00	2.50	10.98	3.17
Mean	225.41	5.85	3.29	1.00	2.38	0.29	2.98	0.78
Std.	148.63	2.30	1.12	0.67	1.99	0.67	1.63	0.42
High iota - AEP (No. of datapoints: 3797)								
Min	28.45	0.91	0.46	0.14	-3.33	0.00	0.28	0.18
Max	1568.78	12.02	6.24	3.16	1.70	0.00	11.57	2.72
Mean	251.29	4.56	2.31	0.74	-0.67	0.00	3.14	0.70
Std.	221.87	1.82	1.16	0.43	0.73	0.00	1.38	0.34
High mirror - AEI (No. of datapoints: 1587)								
Min	14.16	1.38	0.75	0.20	-0.08	0.00	0.03	0.00
Max	245.16	8.21	5.26	3.35	7.93	0.00	9.04	1.34
Mean	111.17	4.34	2.91	0.85	1.25	0.00	3.81	0.62
Std.	36.26	1.19	0.90	0.44	0.93	0.00	1.46	0.18

Data preprocessing

Scatterplots (lower triangle)

Histograms (diagonal)

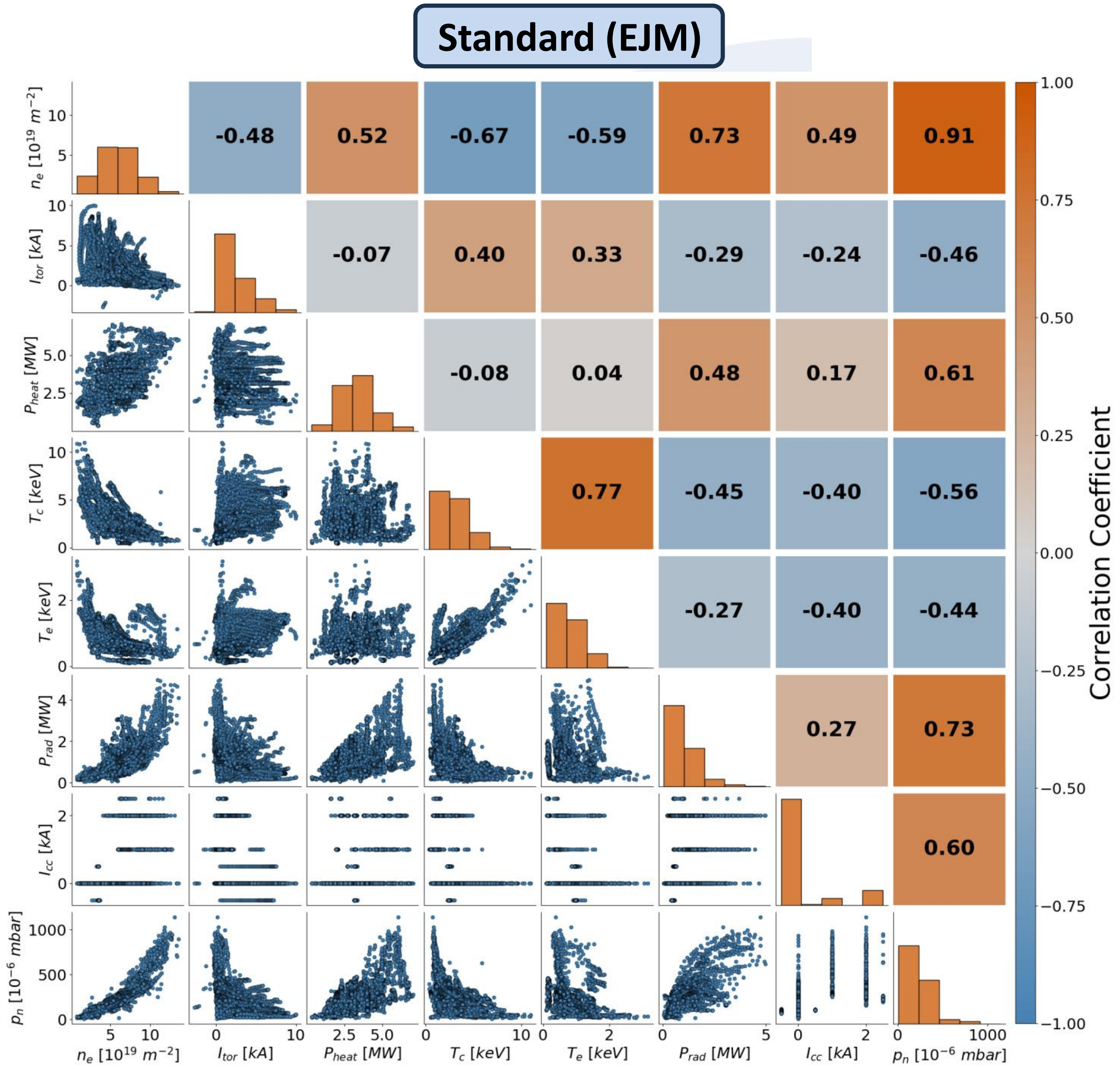
- Qualitative assessment
- Guide to the eye

Pearson Correlation ($r_{x,y}$) (upper triangle)

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

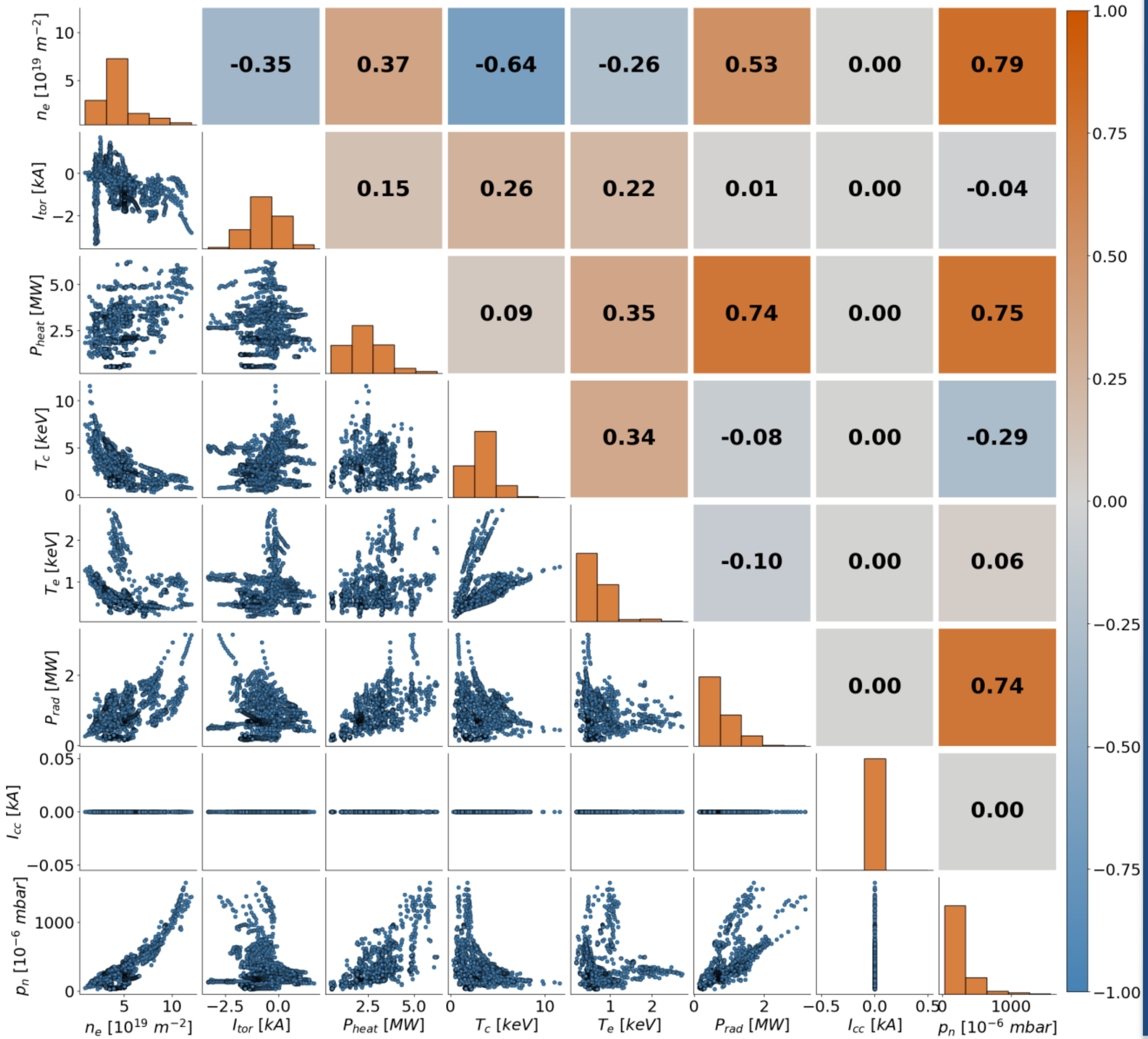
- Quantitative assessment
- Measure of linear dependence

$r_{x,y}$ { **+1** perfect positive linear correlation; **0** no linear correlation; **-1** perfect negative linear correlation

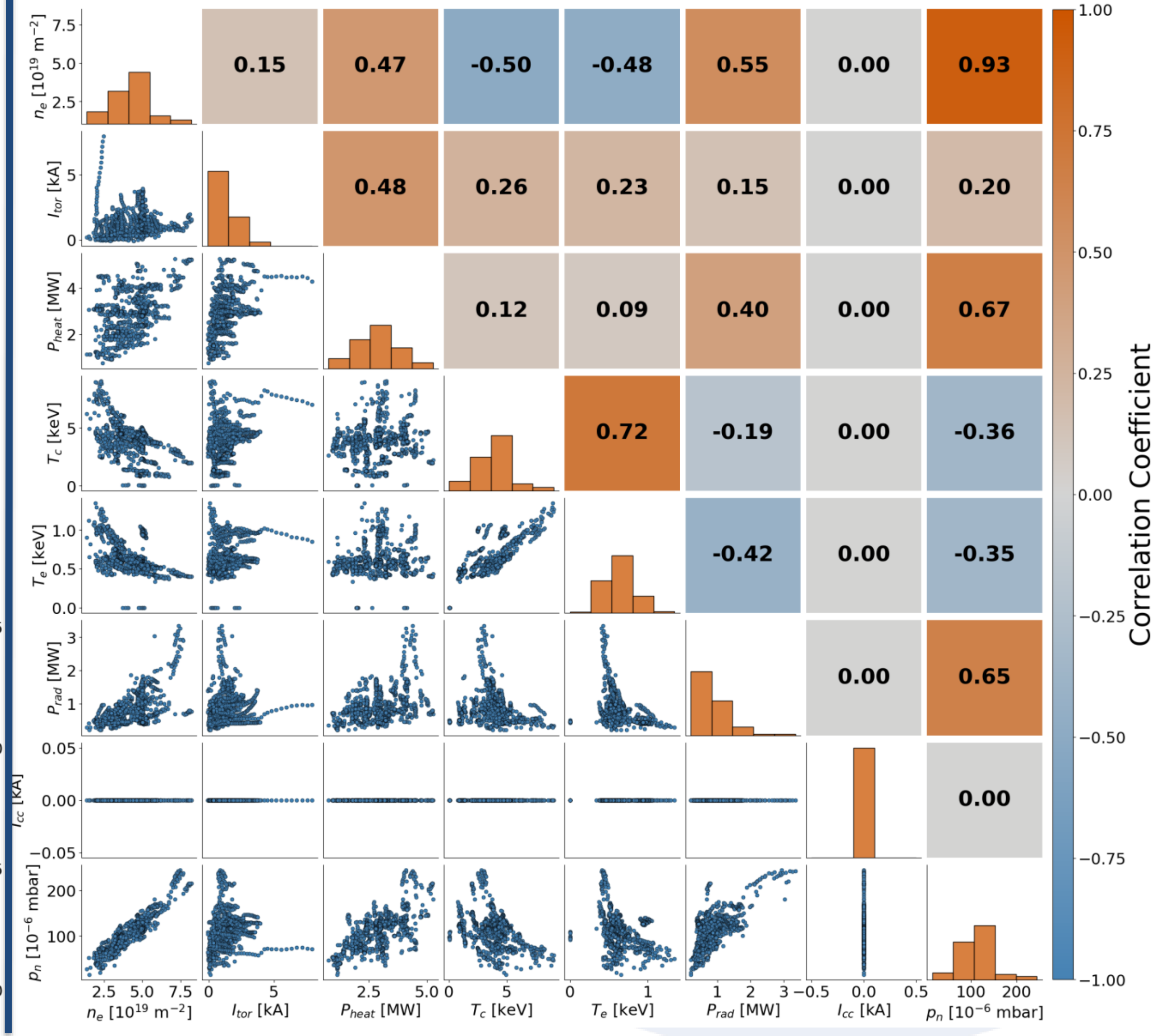


Data preprocessing

High iota (FTM)



High mirror (KKM)



Data preprocessing

Standard *(EJM)* **High iota** *(FTM)* ***High mirror*** *(KKM)*

Proportional	Inversely proportional
n_e	I_{tor}
P_{heat}	T_c
P_{rad}	T_e

- I_{cc} increases with p_n (not easily observed)
- I_{cc} = 0
- I_{cc} = 0

$$\boldsymbol{r_{n_e,p_n}} = 0.91$$

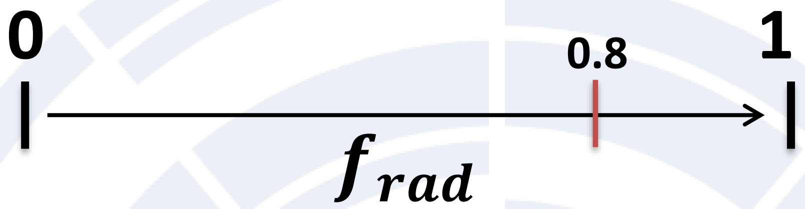
$$\boldsymbol{r_{n_e,p_n}} = 0.79$$

$$\boldsymbol{r_{n_e,p_n}} = 0.93$$

Almost linear scaling between the sub-divertor neutral number density and the line integrated electron density

[U. Wenzel *et al* 2024 *Nucl. Fusion* 64 034002]

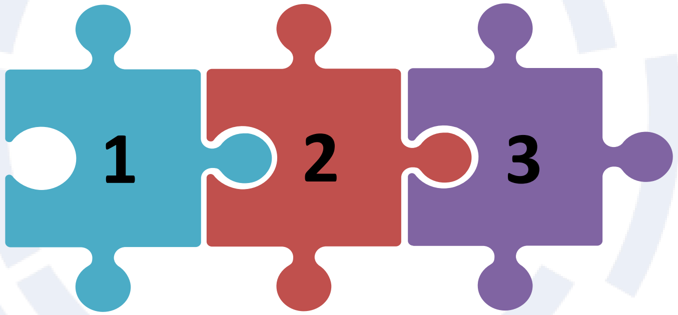
Clustering



Change of the effect of the plasma parameters on p_n

[V. Haak et al 2023 Plasma Phys. Control. Fusion 65 055024]

f_{rad} clusters



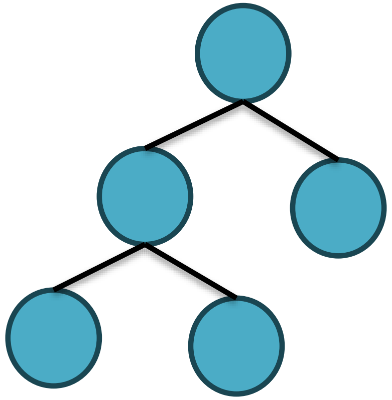
[D. Angelis et al 2025 Plasma Phys. Control. Fusion 67 075004]

Standard	**(0, 0.25)**	**(0.25, 0.50)**	**(0.50, 0.80)**
High iota	**(0, 0.30)**	**(0.30, 0.50)**	**(0.50, 0.80)**
High mirror	**(0, 0.30)**	**(0.30, 0.50)**	**(0.50, 0.80)**

Machine learning models

Extra trees

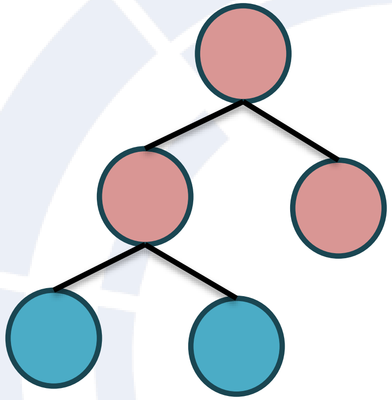
- n_e
- P_{heat}
- P_{rad}
- I_{cc}
- ⋮



- Directly uses the data to make predictions
- Black box
- Captures complex interconnections

Symbolic regression

- n_e
- P_{heat}
- P_{rad}
- I_{cc}
- ⋮



- $+$
- $\sqrt{}$
- $\exp$
- $\div$
- ⋮

- Uses mathematical operators to form functions
- Transparent
- Can reveal underlying governing laws

Feature importance analysis (Based on Extra Trees)

SHapley Additive exPlanations

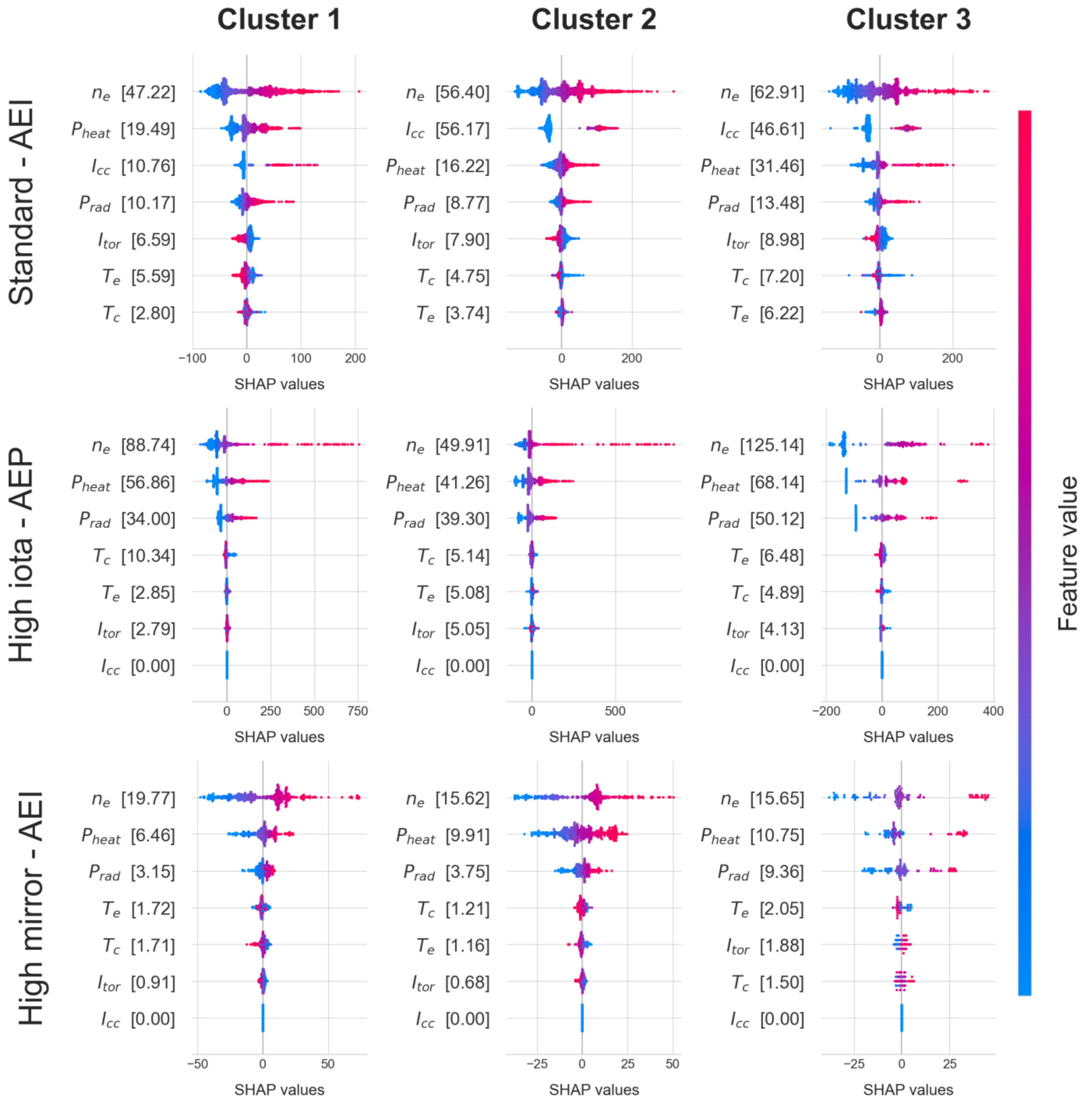
1. Train our model
2. Test how predictions change when a feature (e.g., n_e) is included or excluded
3. Repeat for many combinations of features
4. Average the results to get the feature importance

Table 2: Most important parameters in each magnetic configuration and cluster with their mean absolute SHAP values in brackets.

Standard configuration - AEI port				
Cluster	n_e	P_{heat}	P_{rad}	I_{cc}
1	[47.22]	[19.49]	[10.17]	[10.76]
2	[56.40]	[16.22]	[8.77]	[56.17]
3	[62.91]	[31.46]	[13.48]	[46.61]

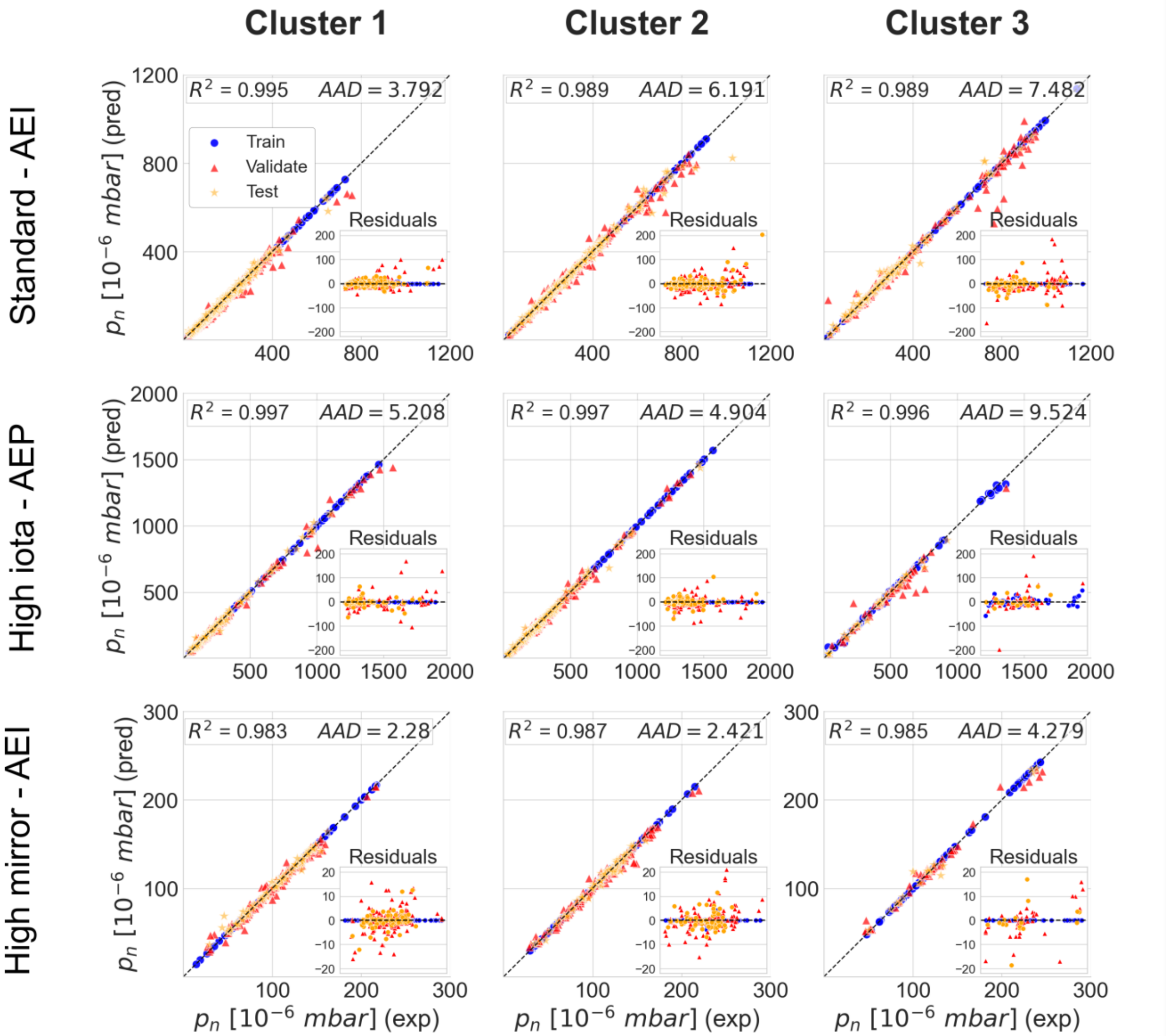
High iota configuration - AEP port			
Cluster	n_e	P_{heat}	P_{rad}
1	[88.74]	[56.86]	[34.00]
2	[49.91]	[41.26]	[39.30]
3	[125.14]	[68.14]	[50.12]

High mirror configuration - AEI port			
Cluster	n_e	P_{heat}	P_{rad}
1	[19.77]	[6.46]	[3.15]
2	[15.62]	[9.91]	[3.75]
3	[15.65]	[10.75]	[9.36]

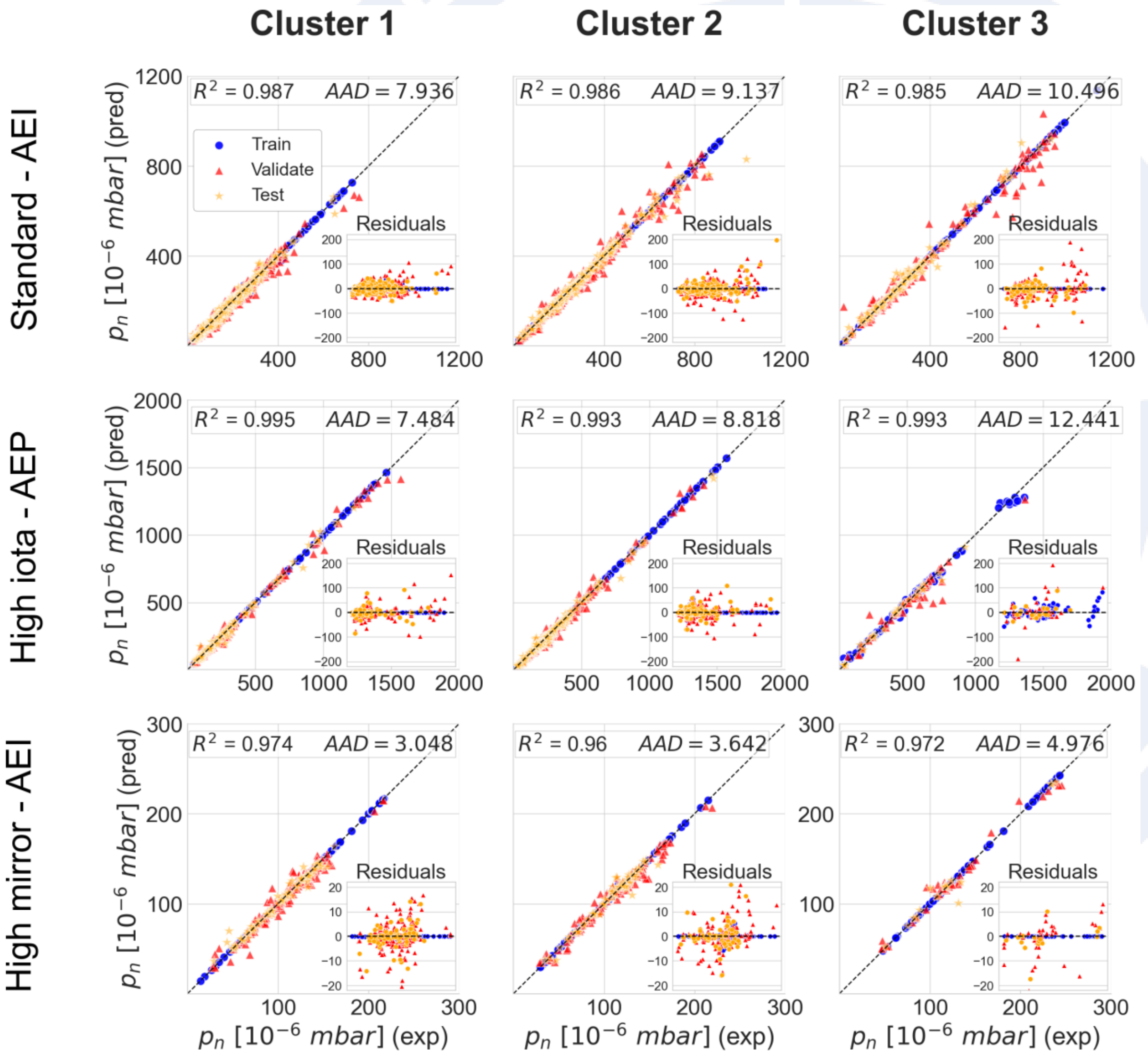


Extra Trees Results

All Parameters



Important Parameters



Symbolic Expressions

$$p_n = \alpha_1 \, n_e^{\alpha_2} \, P_{heat}^{\alpha_3}$$

[*D. Angelis et al 2025 Plasma Phys. Control. Fusion 67 075004*]

Standard - AEI

$$p_n = \alpha_1 \, n_e^{\alpha_2} \, (P_{heat} + \alpha_3)^{\alpha_4} \, (I_{cc} + \alpha_5)^{\alpha_6}$$

Table 3: Standard configuration (AEI port) - Constants α_i, $i = 1, 2, \ldots, 6$ of the SR expressions (6), with a 95% confidence interval, and the associated R^2 and AAD values.

Cluster	a_1	a_2	a_3	a_4	a_5	a_6	R^2	AAD	R^2	AAD
1	$4.34_{2.9}^{6.23}$	$1.15_{1.11}^{1.18}$	$0.25_{-0.53}^{3.13}$	$0.57_{0.38}^{0.82}$	$4.07_{3.28}^{4.81}$	$0.66_{0.53}^{0.77}$	0.933	18.850	0.911	21.612
2	$5.15_{2.85}^{10.0}$	$1.27_{1.2}^{1.32}$	$3.23_{1.84}^{4.64}$	$0.69_{0.53}^{0.8}$	$1.49_{0.34}^{3.34}$	$0.33_{0.15}^{0.61}$	0.918	32.511	0.872	40.545
3	$2.92_{2.4}^{3.83}$	$1.51_{1.38}^{1.65}$	$3.66_{3.08}^{4.75}$	$0.7_{0.57}^{0.85}$	$1.85_{0.37}^{3.06}$	$0.27_{0.12}^{0.4}$	0.917	36.927	0.892	42.020

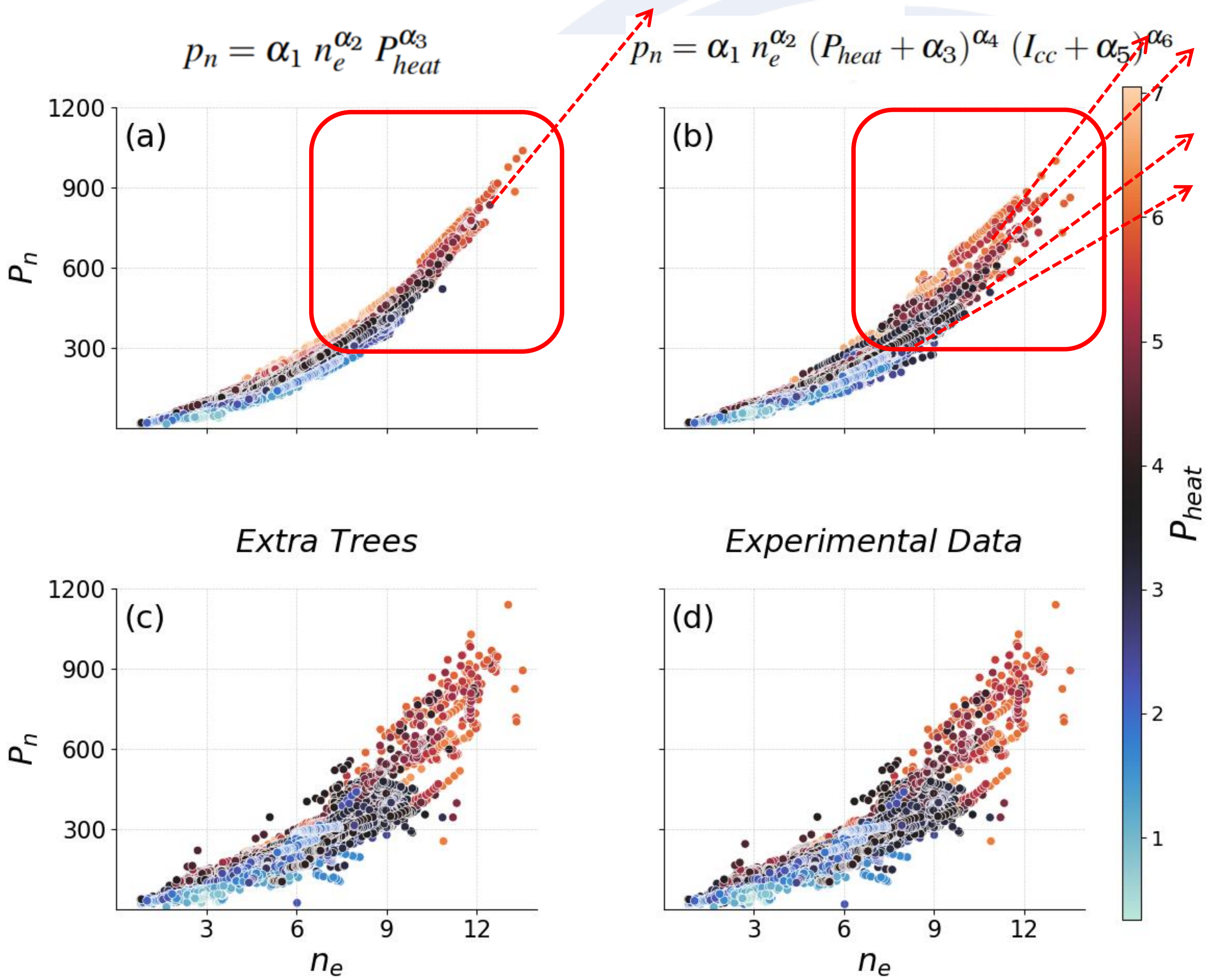
High iota AEP
High mirror AEI

$$p_n = \alpha_1 \, n_e^{\alpha_2} \, P_{heat}^{\alpha_3}$$

[*D. Angelis et al 2025 Plasma Phys. Control. Fusion 67 075004*]

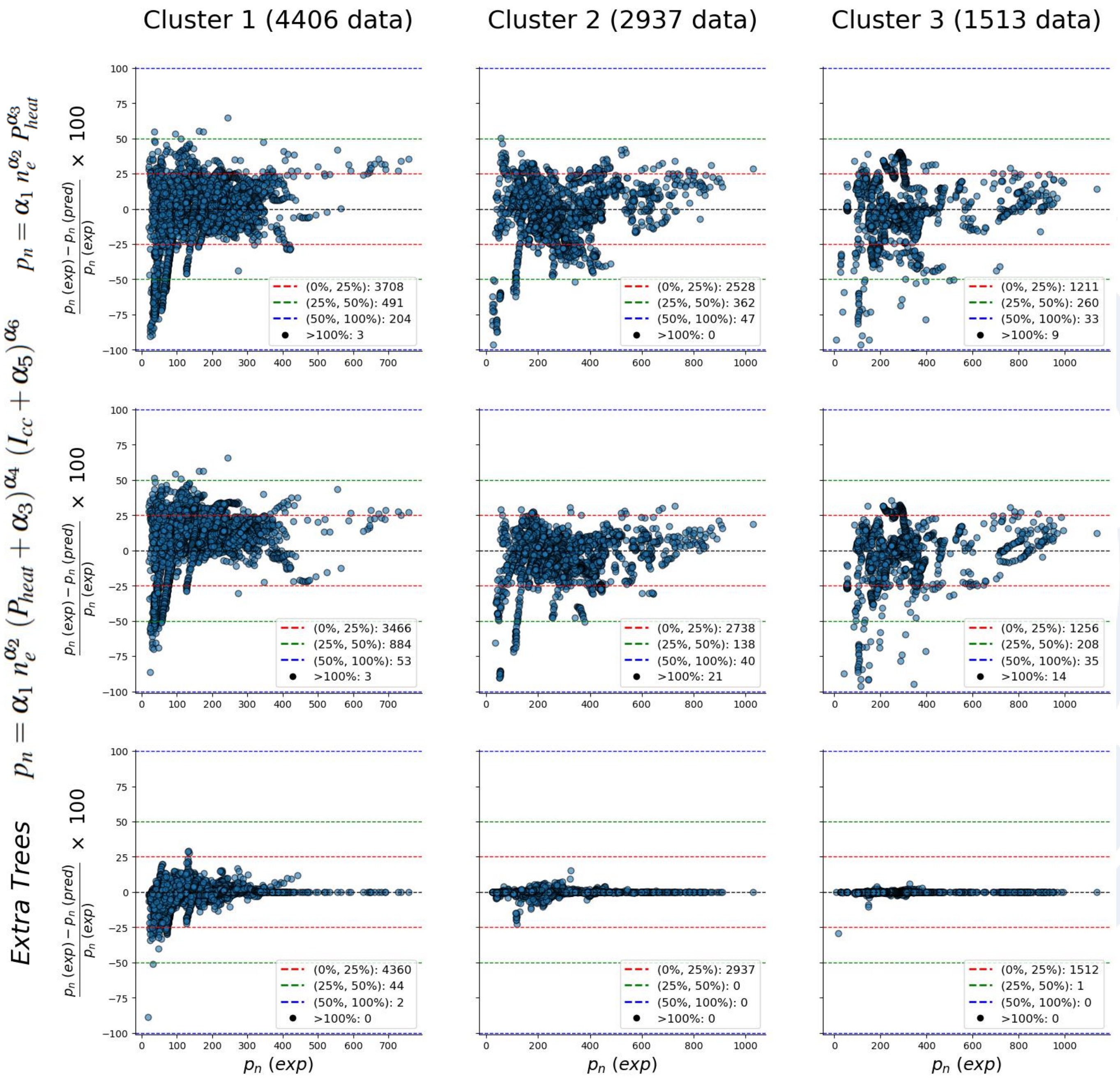
Standard Configuration – AEI Port - p_n vs n_e and P_{heat}

- Extra trees predictions are very accurate
- Both symbolic expressions capture the average trend of data
- The two-parameter expression proposes a sharper increase of p_n at higher n_e and P_{heat} values
- The three-parameter expressions shows "behavior streamlines" at that region
- Statistical error remains acceptable
- May pose problems when extrapolating as the two expressions may exhibit different behavior



Standard Configuration – AEI Port – Relative errors

- Majority of predictions fall within a 25% error range
- Extra trees
  - Cluster 1: largest number of errors above 25%
  - Clusters 2 and 3: predictions are very good
- Symbolic Expressions
  - Clusters 2 and 3: three-parameter better than two-parameter (as expected)
  - Cluster 1: two-parameter more accurate predictions within the 25% threshold, but this is reversed for errors >25%
  - I_{cc} is needed more in clusters 2 and 3 ($f_{rad} > 0.25$)
  - Three-parameter better addresses the error region > 50%
- Most errors are at low pressures

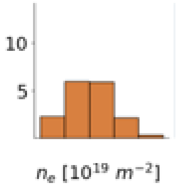


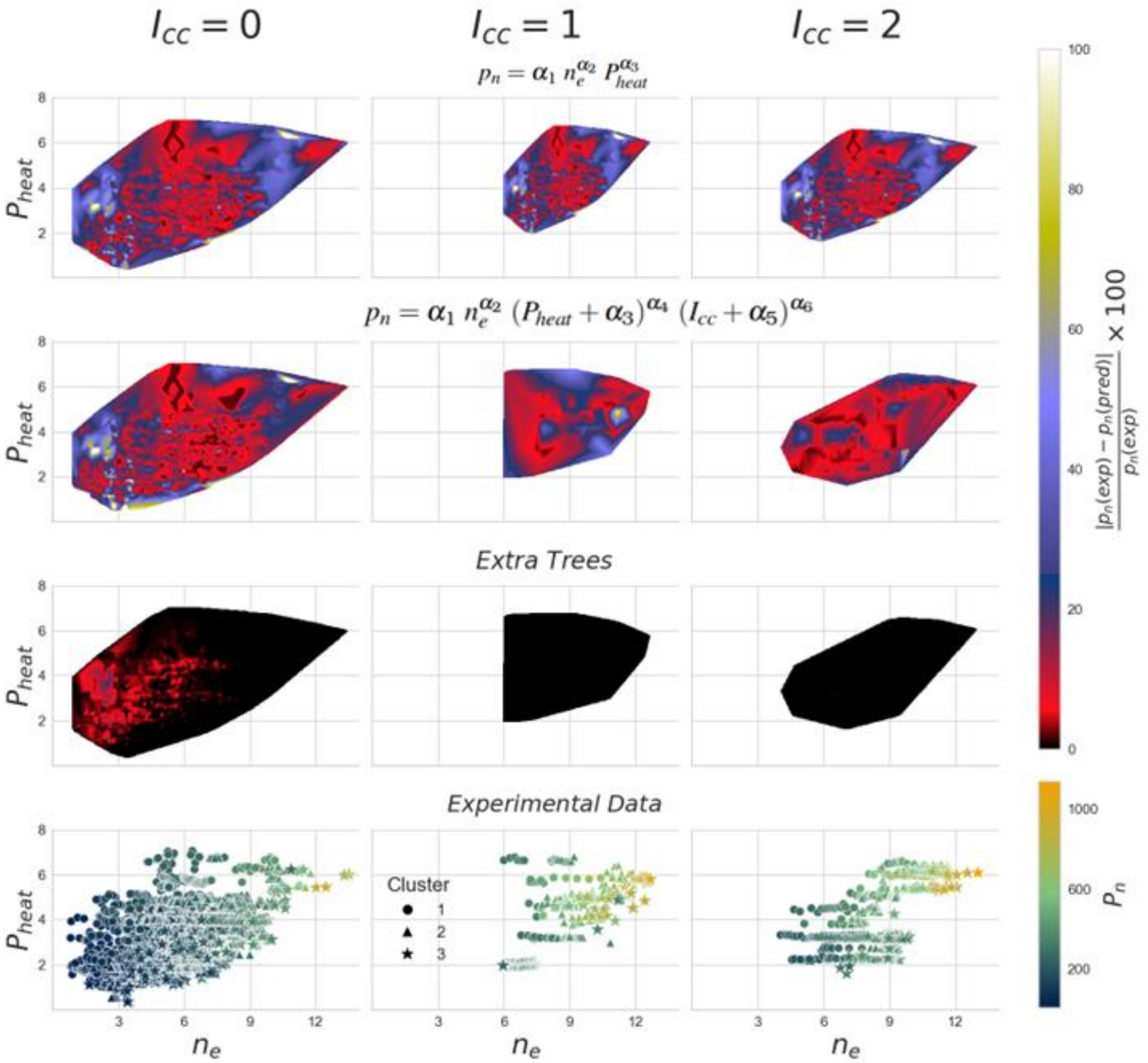
Standard Configuration – AEI Port – Contour relative error plots

- Extra trees predictions are very accurate

- $I_{cc} = [1,2]$: three-parameter better than two-parameter (as expected)

- $I_{cc} = [0]$: more comparable but still, three-parameter is better
  - Attributed to α_3 (P_{heat} correction)
  - Both expressions show deviations at small and large n_e values
    - Attributed to data distribution





Concluding remarks

- Data-driven approaches to model neutral pressure for partially attached conditions ($0 < f_{rad} < 0.8$)
  - Cluster data to better approach the changing f_{rad} behavior (from attached to detached)
  - Preprocessing analysis to acquire data information about $[p_n, n_e, P_{heat}, P_{rad}, I_{cc}, I_{tor}, T_c, T_e]$
  - Use Extra Trees to directly predict p_n
  - Find symbolic expressions to correlate p_n with other parameters

- Employ a feature importance analysis to identify key extra trees parameters
  - Standard: $[n_e, P_{heat}, P_{rad}, I_{cc}]$
  - High iota and high mirror: $[n_e, P_{heat}, P_{rad}]$
  - Negligible accuracy loss by omitting the rest parameters

- Symbolic expressions
  - Standard: $p_n = a_1 n_e^{a_2} \times (P_{heat} + a_3)^{a_4} \times (I_{cc} + a_5)^{a_6}$
    - $a_i, i = 1, \dots, 6$ optimized to the specific cluster
  - High iota and high mirror: $p_n = a_1 n_e^{a_2} \times P_{heat}^{a_3}$

Concluding remarks

- Majority of predictions fall within a 25% relative error threshold
  - Most deviations at low pressures
  - Expression parameters are influenced by where the data is most concentrated.

- Interpolation vs extrapolation
  - Two parameter expression exhibit a more constant behavior pattern
    - No significant statistical error
    - May pose problems on extrapolation
  - Three parameter expression shows behavior streamlines
    - Enhanced extrapolation potential, provided the underlying physical laws are sufficiently captured
  - Extra trees predictions are almost perfect
    - Black box model – may struggle on unseen values
    - More suitable for interpolation

Future work

- ❑ Employment of corresponding data of the OP2.2 and OP2.3 experimental campaigns
  - ➢ Require further post-processing before being used
  - ➢ Better tune the ML models under different experimental conditions (e.g., calibrate expression parameters)

Acknowledgements

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grand Agreement No 101052200 - EUROfusion), as well as the National Program of Controlled Thermonuclear Fusion. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. This work was supported by computational time granted from the EUROfusion HPC Marconi Fusion.

Thank you for your attention