Confinement Scaling with Machine Size in the Updated ITPA Global H-Mode Energy Confinement Database

J. Hall[1] · Y. Zhang[1] · P. Zhang[1] · C. Angioni[2] · G. Verdoolaege[1] · JET Contributors*. · EUROfusion Tokamak Exploitation Team**

[1]*Dept. of Applied Physics, Ghent University, 9000 Ghent, Belgium*
[2]*Max–Planck–Institut für Plasmaphysik, Boltzmannstrasse 2, D–85748 Garching, Germany*

** See the authorlist of C. Maggi et al., Nucl. Fusion 64, 112012 (2024)*
*** See the authorlist of E. Joffrin et al., Nuclear Fusion 64, 112019 (2024)*

1 Background & motivation

The IPB98(y, 2) scaling law was derived in 1999 [1].

$$\tau_E^{\mathrm{IPB98(y,2)}} = 0.056 I_p^{0.93} B_t^{0.15} \overline{n}_e^{0.41} P_{\mathrm{l,th}}^{-0.69} R_{\mathrm{geo}}^{1.97} \kappa^{0.78} \varepsilon^{0.58} M_{\mathrm{eff}}^{0.19}$$

Some 23 years later, an updated regression on the expanded version of the same database showed [2]:

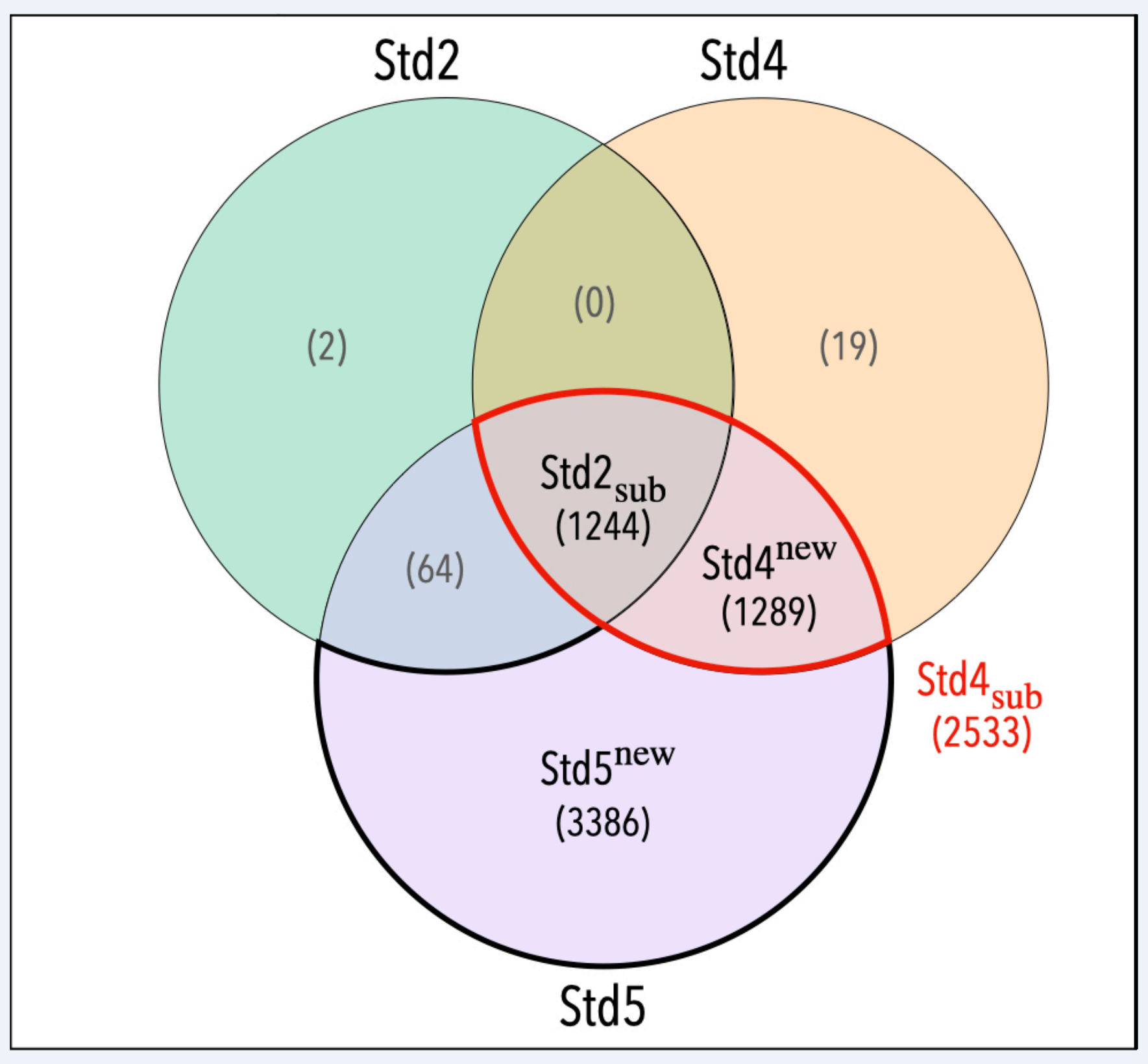
$$\tau_E^{\mathrm{ITPA20}} \propto R_{\mathrm{geo}}^{1.71\pm0.32} \qquad \tau_E^{\mathrm{ITPA20-IL}} \propto R_{\mathrm{geo}}^{1.19\pm0.27}$$

The size exponent α_R ***weakened***.
Two candidate explanations were investigated:
(i) Statistical artefacts, in particular multicollinearity
(ii) Physical properties of the additional discharges

2 ITPA H-mode database

Three database iterations are used:

$\mathrm{Std2_{sub}}$: 1244 pts (8 machines)	$\alpha_R = 1.7$	(2.1)	[1]
$\mathrm{Std4_{sub}}$: 2533 pts (13 machines)	$\alpha_R = 1.4$	(2.0)	[3]
Std5: 5919 pts (14 machines)	$\alpha_R = 1.3$	(1.6)	[2]



Spherical devices excluded (distinct transport physics).

Leading to a reduced range in ε

Hence, ε was removed from the regression model.

3 Multicollinearity diagnostics

Each regression coefficient has an associated standard deviation which is governed by:

- The model
- Predictors range
- Multicollinearity
- Measurement uncertainty

$$\sigma_i \equiv \sqrt{\mathrm{Var}\left(\hat{\beta}_i\right)} = \sqrt{\frac{\sigma_\epsilon^2}{\mathrm{SS}_i} \cdot \mathrm{VIF}_i}, i = 1, \ldots, p$$

Variation inflation factors (VIF)

Variable	$\mathrm{Std2_{sub}}$	$\mathrm{Std4_{sub}}$	Std5
I_p	7.0	6.9	7.8
R_{geo}	5.1	5.0	6.6

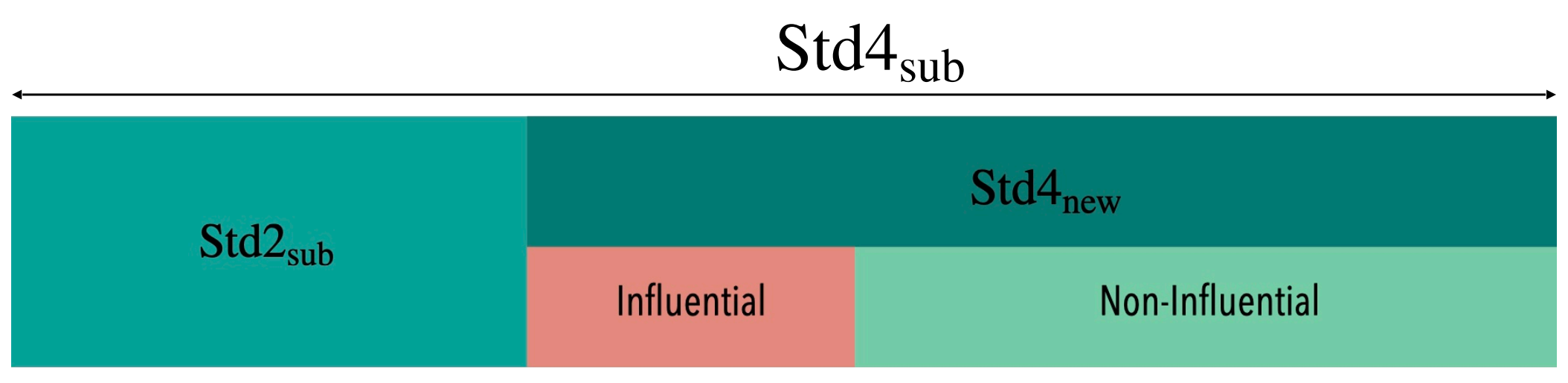
VIF > 5 are indicative of strong multicollinearity

- While multicollinearity is pronounced, the effect on coefficient variance remains limited
- ***We deduce that multicollinearity is not the sole cause of reduced size scaling.***

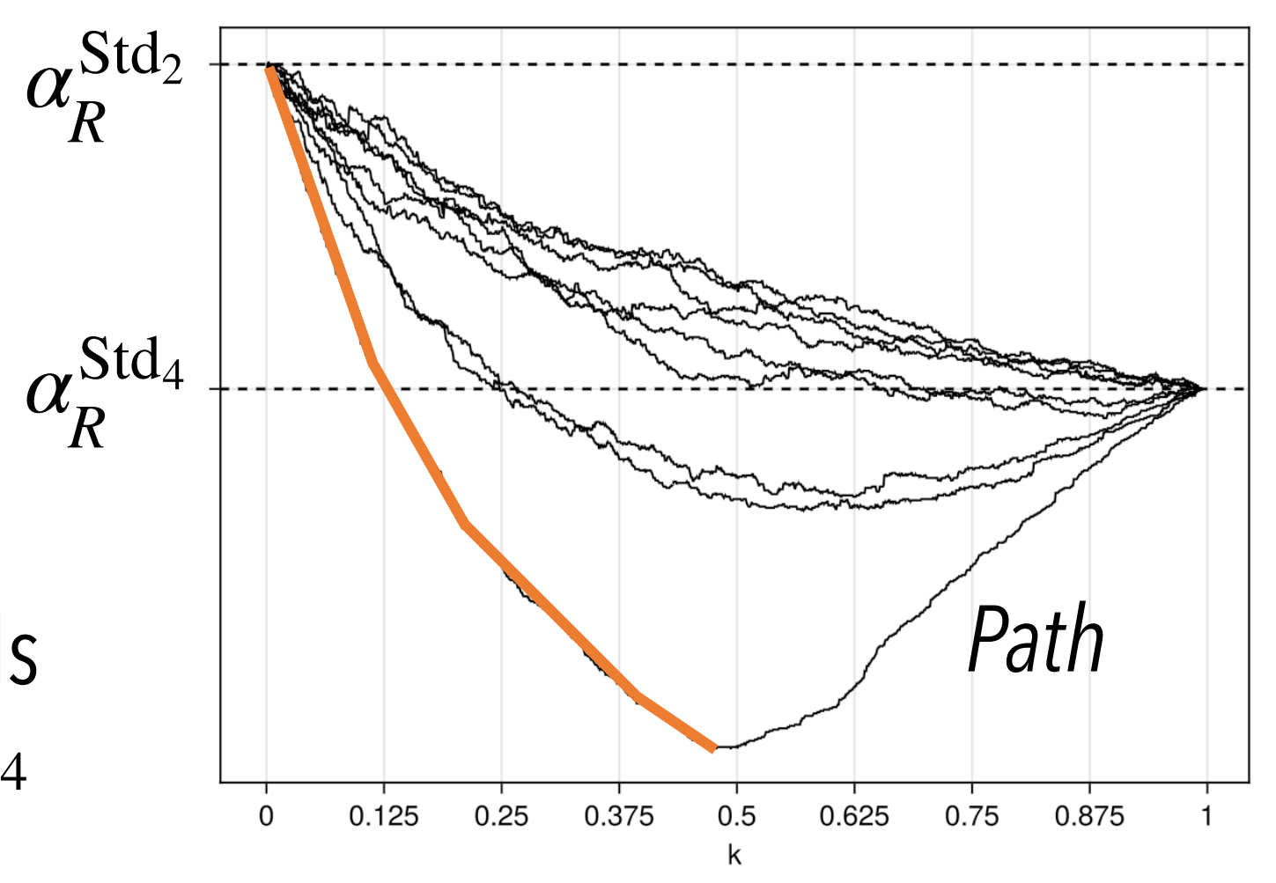
4 Identification of the Influential Set

To avoid confusion of abstraction: *we explain the optimisation procedure by focusing on the example of the transition from $\mathrm{Std2_{sub}}$ to $\mathrm{Std4_{sub}}$. The transition of $\mathrm{Std4_{sub}}$ to Std5 follows analogously.*

1 ORDER

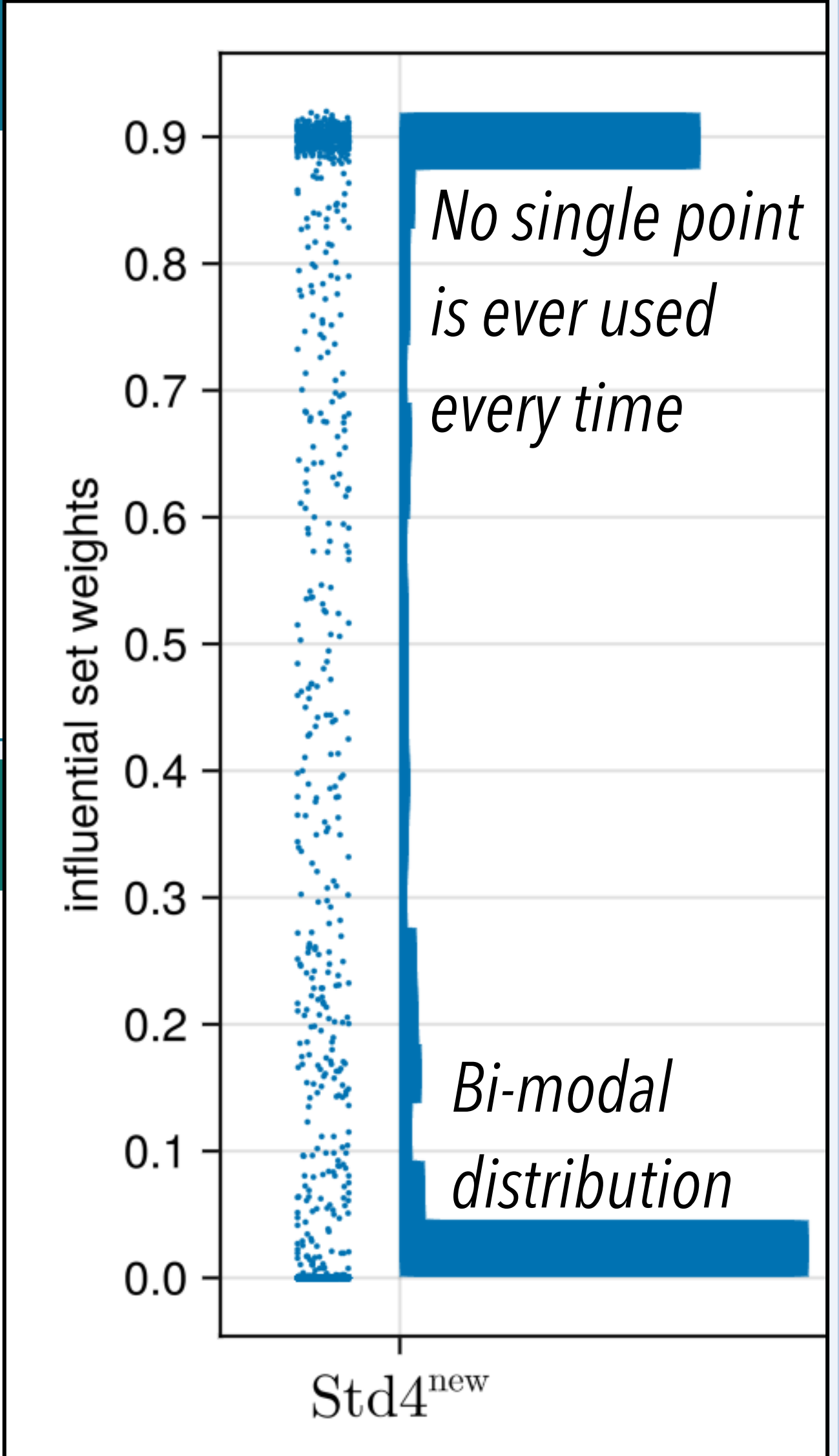


- Aim: determine the smallest subset of $\mathrm{Std4_{new}}$ that results in the largest reduction in α_R (influential set)
- Iteratively add points from $\mathrm{Std4_{new}}$ into $\mathrm{Std2_{sub}}$
- Perform a regression in each instance, recording α_R
- Optimal ordering of points leads to an intermediate $\alpha_R < \alpha_R^{\mathrm{Std_4}}$

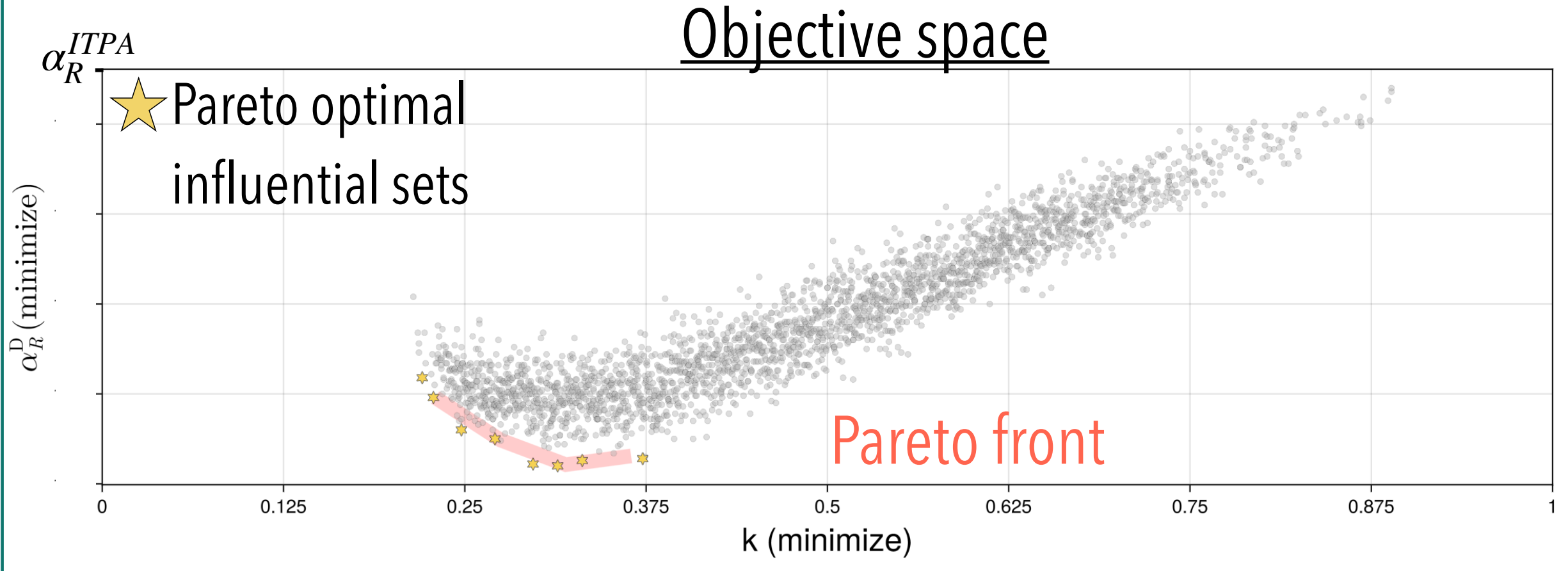


2 Quantify

- For each path generated
  - Assign 1 to each point leading to the minimum of the path (highlighted in orange)
  - Assign 0 otherwise
- Average over the number of paths generated
- Consider the mean value of each data-point a probability of inclusion in the influential set

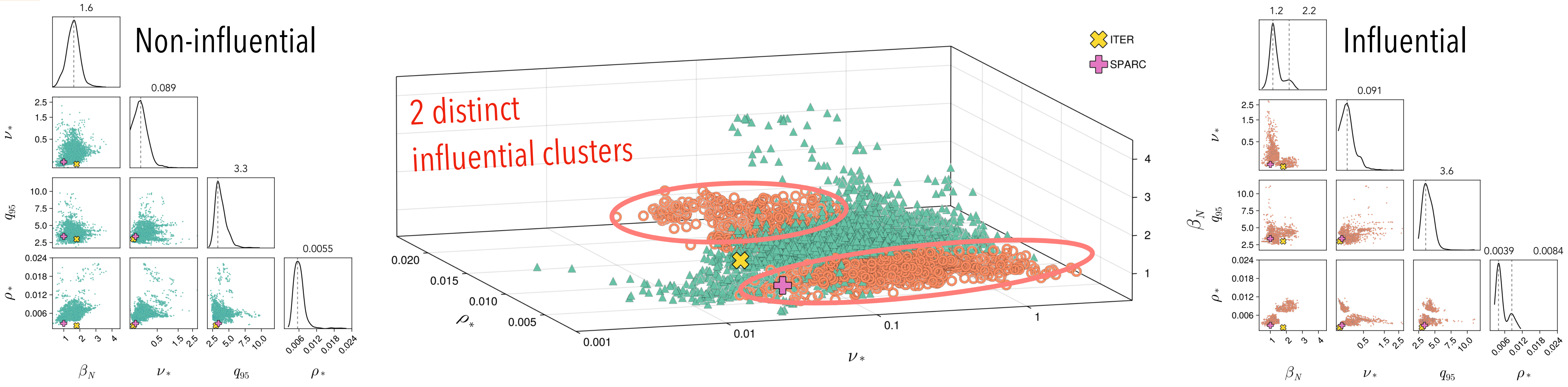


3 Optimise



- Extract subsets of size k through weighted sampling.
- Evaluate based on α_R^D and subset size, k.

4 Classify



5 Two distinct scaling laws

Non-Influential Scaling (α_R = 1.7)

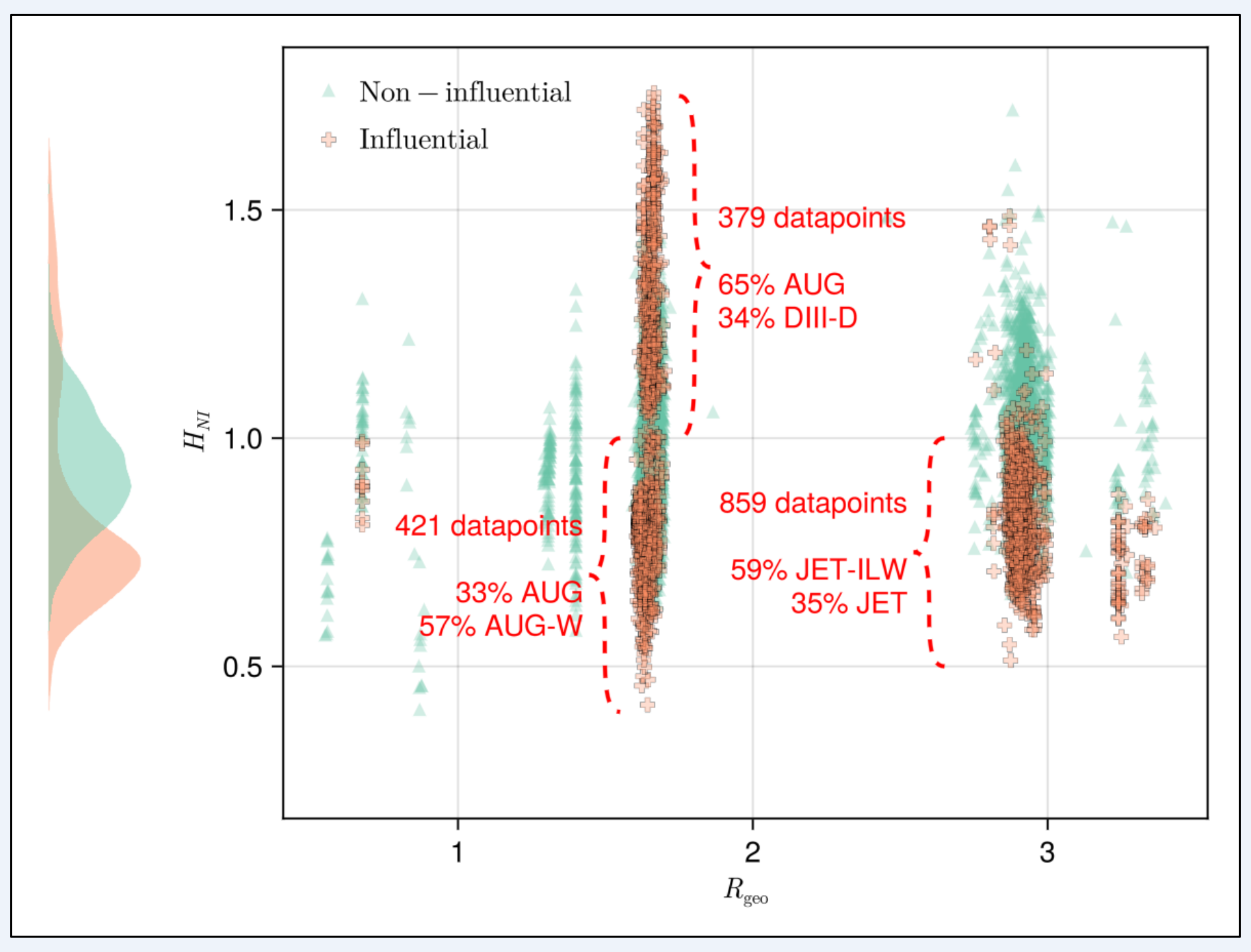
$$\tau_{\mathrm{E,th}}^{\mathrm{NI}} = 0.04 I_{\mathrm{p}}^{1.1} B_{\mathrm{t}}^{0.065} \overline{n}_{\mathrm{e}}^{0.35} P_{l,th}^{-0.71} R_{\mathrm{geo}}^{1.7} \kappa^{0.77} M_{\mathrm{eff}}^{0.14}$$

RMSE = 0.14 · Regained size scaling · Negligible field scaling (α_B = 0.065) · Strong power degradation $\alpha_P = -0.71$

Influential Scaling (α_R= 0.56)

$$\tau_{\mathrm{E,th}}^{\mathrm{I}} = 0.177 I_{\mathrm{p}}^{1.5} B_{\mathrm{t}}^{-0.25} \overline{n}_{\mathrm{e}}^{-0.19} P_{l,th}^{-0.62} R_{\mathrm{geo}}^{0.56} \kappa^{0.73} M_{\mathrm{eff}}^{0.54}$$

RMSE = 0.15 · Weak size scaling · Negative density exponent ($\alpha_n = -0.19$) · **Descriptive role only**



6 Conclusions

1. Multicollinearity is present but cannot solely account for α_R reduction
2. Maximally influential set has been determined with an optimisation procedure
3. Influential set well localised in dimensionless ($\rho^* \cdot \nu^* \cdot \beta_N \cdot q_{95}$) space, forming two clusters
4. Virtually all observations excluded from the database to form the ITER-like subset used for deriving ITPA20-IL scaling are assigned to the non-influential subset
5. Non-influential scaling recovers IPB98-like size dependence α_R = 1.7.

ITER & SPARC: 84% non-influential probability
→ NI scaling is the relevant prediction

Tokamak	$\tau_{\mathrm{E,th}}^{\mathrm{I}}(s)$	$\tau_{\mathrm{E,th}}^{\mathrm{NI}}(s)$	$\tau_{\mathrm{E,th}}^{\mathrm{reference}}(s)$
ITER	1.9	3.8	3.6
SPARC	0.46	0.88	0.77

References
[1] ITER Physics Basis Editors, Nucl. Fusion 39, 2175 (1999)
[2] G. Verdoolaege et al., Nuclear Fusion 61, 076006 (2021)
[3] D.C McDonald et al., Nuclear Fusion, 47, 147 (2007)
[4] J. Hall et al., Nuclear Fusion (2026) (under review)
[5] C. Angioni et al., Nuclear Fusion 66 026006 (2026)

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 (Grant Agreement No 10105220 –EUROfusion). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission