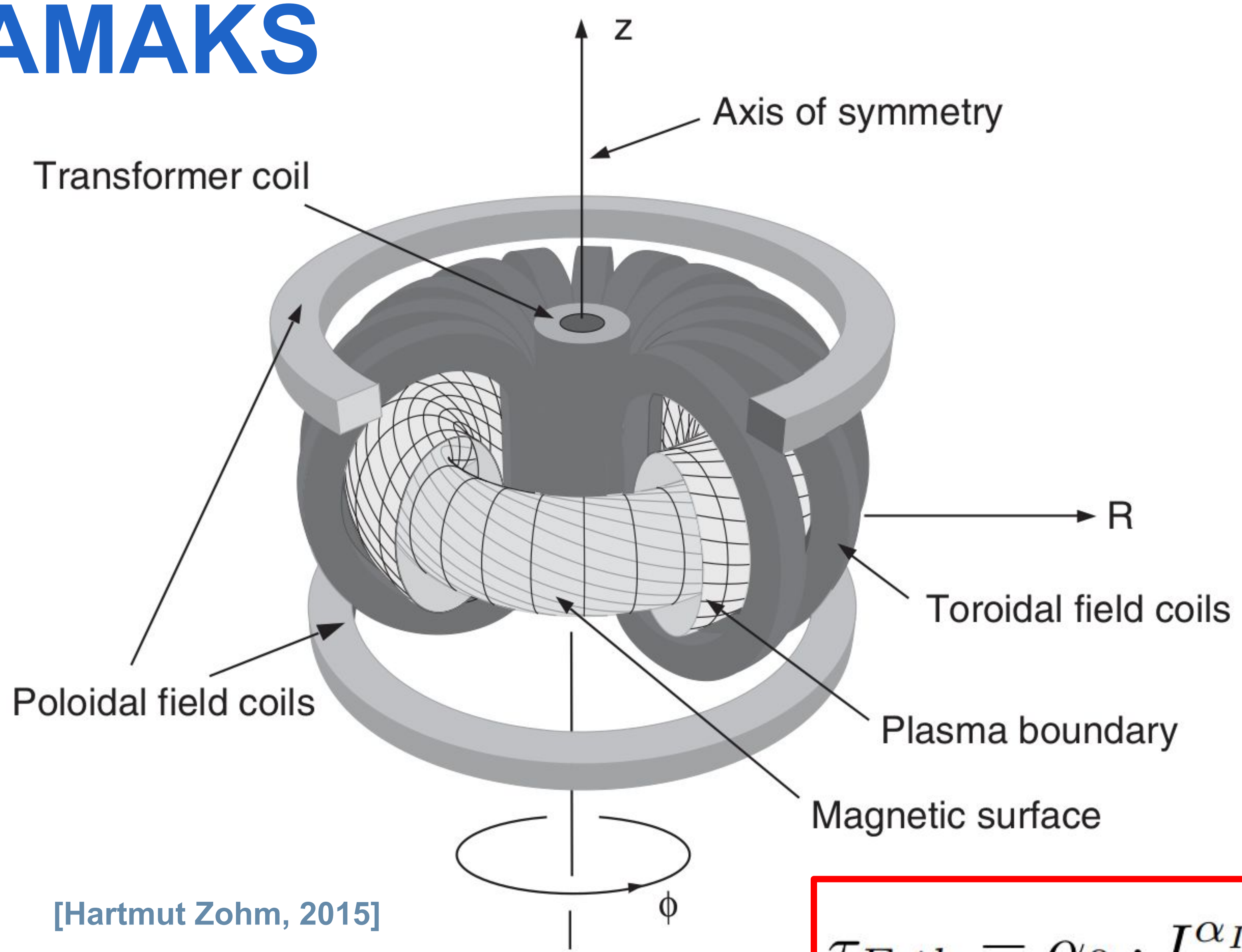
INVESTIGATING THE DEPENDENCE ON MACHINE SIZE OF THE ENERGY CONFINEMENT IN TOKAMAKS USING DATA-DRIVEN METHODS

M.Sc. Student Karina Chiñas Fuentes

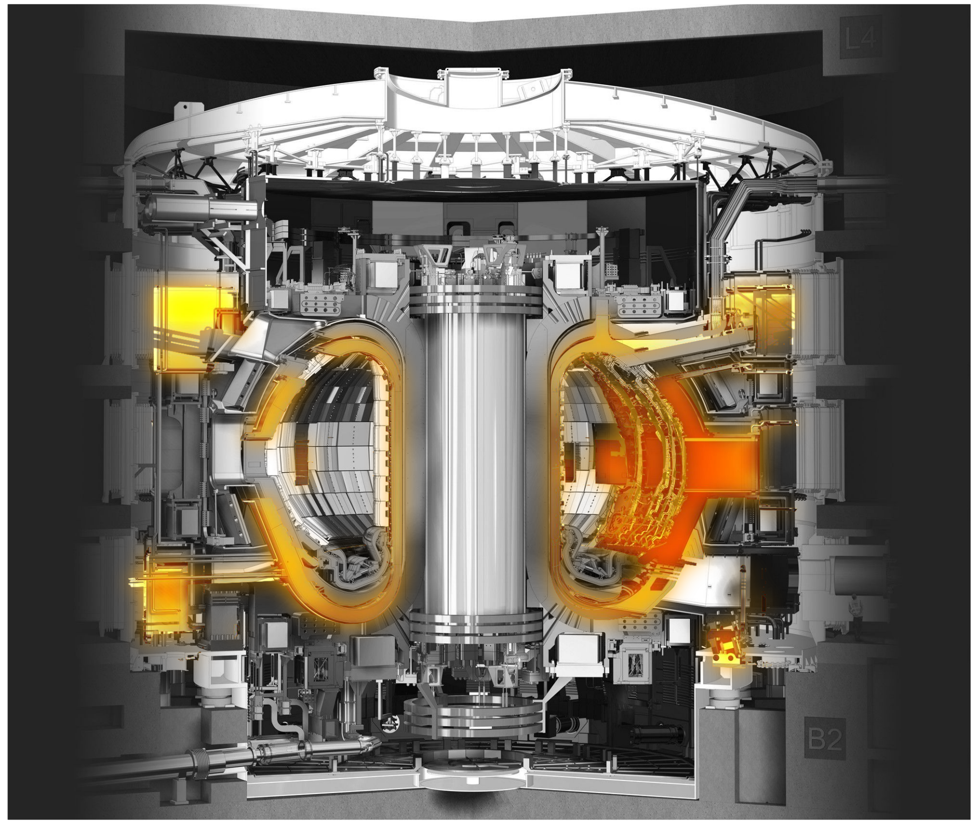
Prof. Dr. Geert Verdoolaege
PhD Student Joseph Hall

11/07/2023

TOKAMAKS



[Hartmut Zohm, 2015]

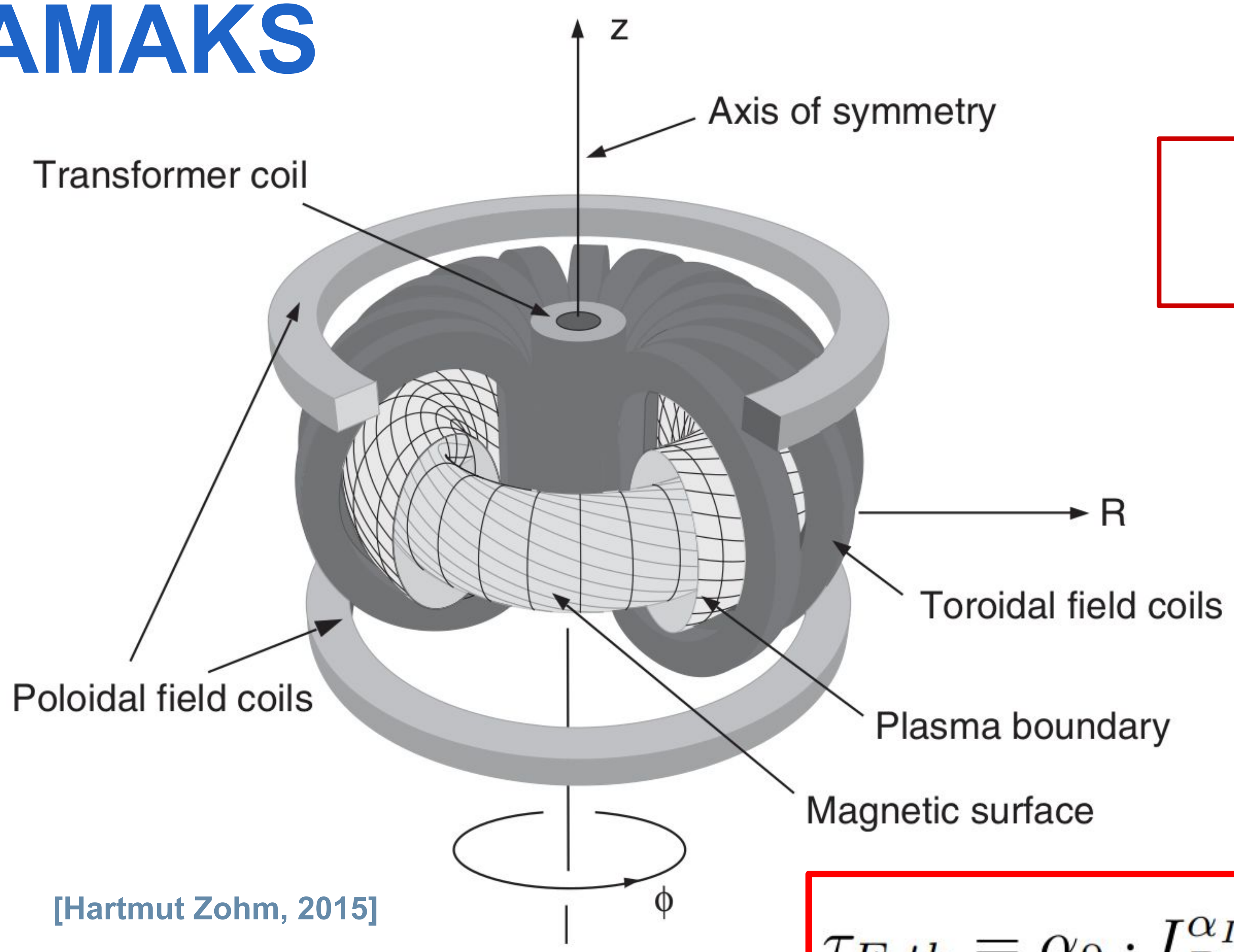


[ITER Organization, 2023]

log($\tau_{E,th}$) → OLS

$$\tau_{E,th} = \alpha_0 \cdot I_P^{\alpha_I} \cdot B_t^{\alpha_B} \cdot \bar{n}_e^{\alpha_n} \cdot P_{l,th}^{\alpha_P} \cdot R_{geo}^{\alpha_R} \cdot \kappa_a^{\alpha_\kappa} \cdot \epsilon^{\alpha_\epsilon} \cdot M_{eff}^{\alpha_M}$$

TOKAMAKS



[Hartmut Zohm, 2015]

H-mode (EMLy subset) confinement database

STDB2
1998

$\alpha_R \sim 2.22$

WHY?

STDB5
2021

$\alpha_R \sim 1.45$

DATA?

PHYSICS?

$$\tau_{E,th} = \alpha_0 \cdot I_P^{\alpha_I} \cdot B_t^{\alpha_B} \cdot \bar{n}_e^{\alpha_n} \cdot P_{l,th}^{\alpha_P} \cdot R_{geo}^{\alpha_R} \cdot \kappa_a^{\alpha_\kappa} \cdot \epsilon^{\alpha_\epsilon} \cdot M_{eff}^{\alpha_M}$$

WORKFLOW

WHICH REGISTERS IN STDB5 ARE **DECREASING** α_R?

→

IS THIS DECREASE DUE TO **DATA ISSUES**?

→

CAN WE **PREDICT** WHETHER A NEW REGISTER WILL **DECREASE** α_R?

→

RELATE THE FINDINGS **TOKAMAK CHARs & PHYSICS**

WORKFLOW

WHICH REGISTERS IN STDB5 ARE **DECREASING** α_R?

→

IS THIS DECREASE DUE TO **DATA ISSUES**?

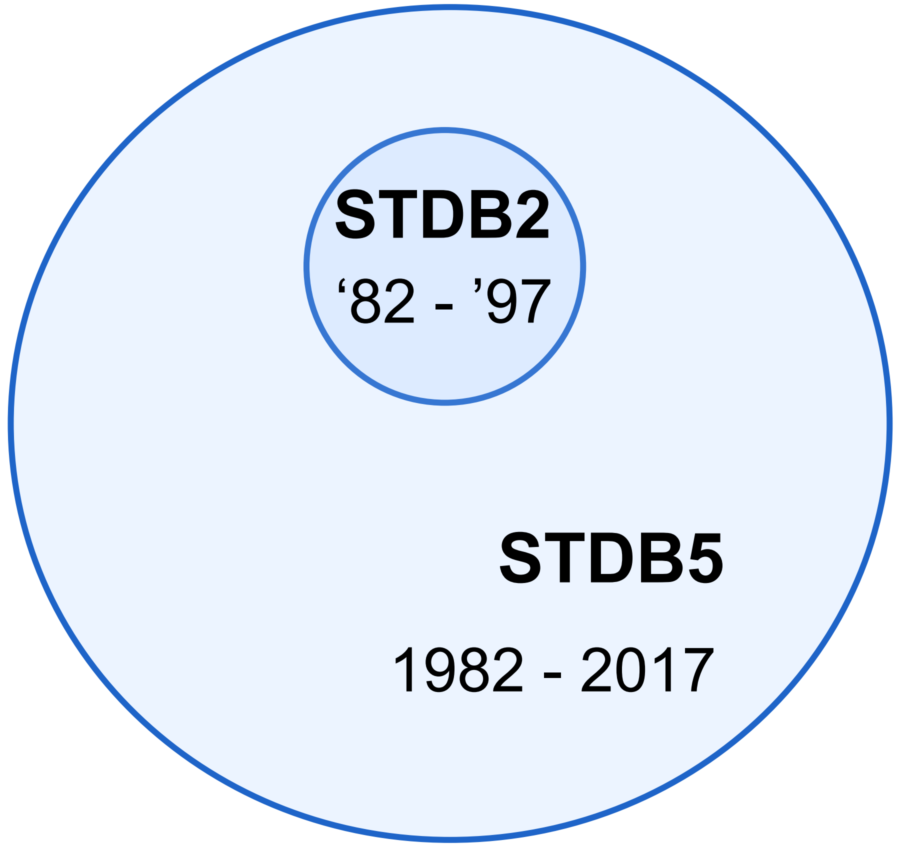
→

CAN WE **PREDICT** WHETHER A NEW REGISTER WILL **DECREASE** α_R?

→

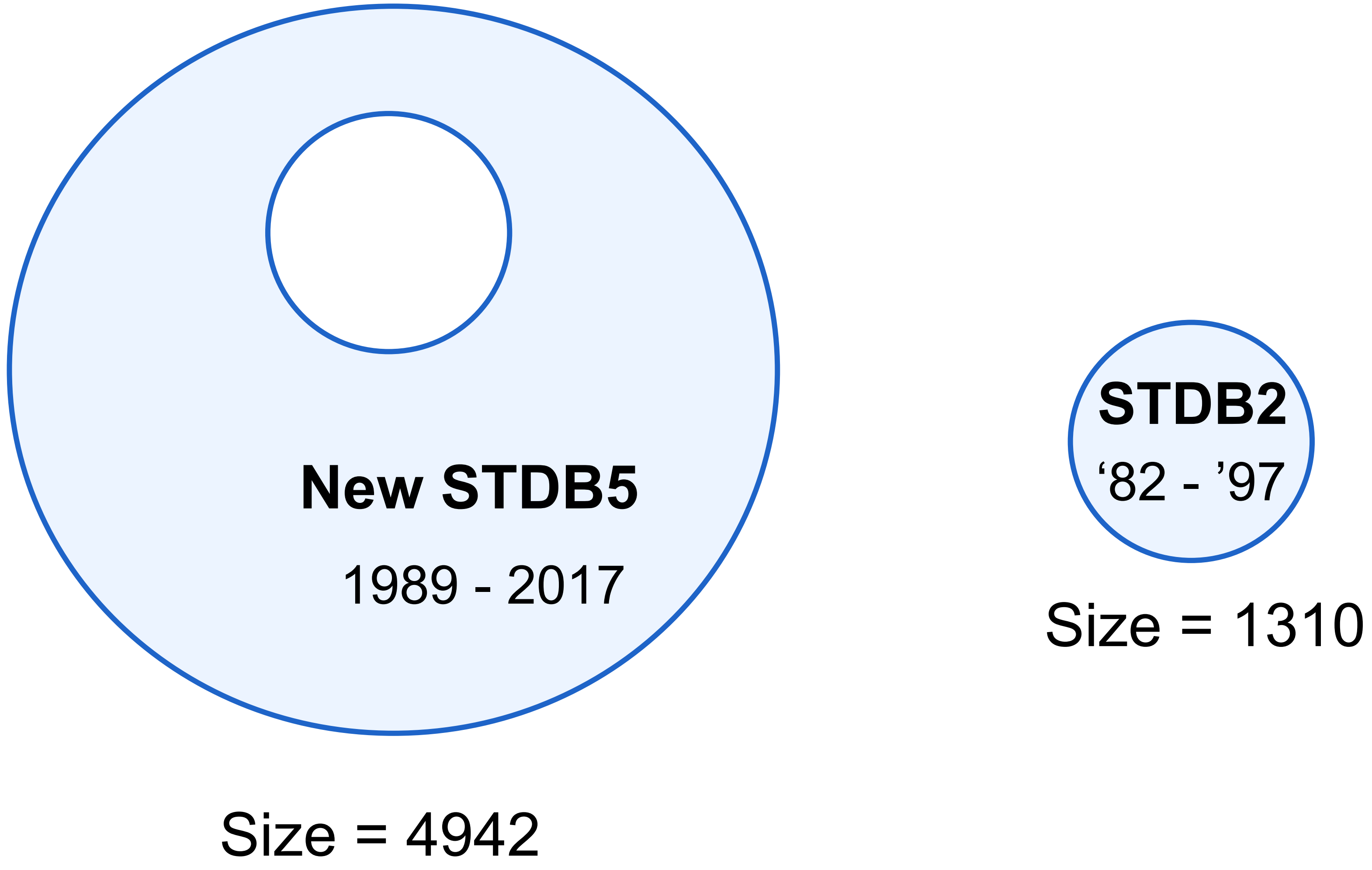
RELATE THE FINDINGS **TOKAMAK CHARs & PHYSICS**

FINDING THE DECREASING REGISTERS

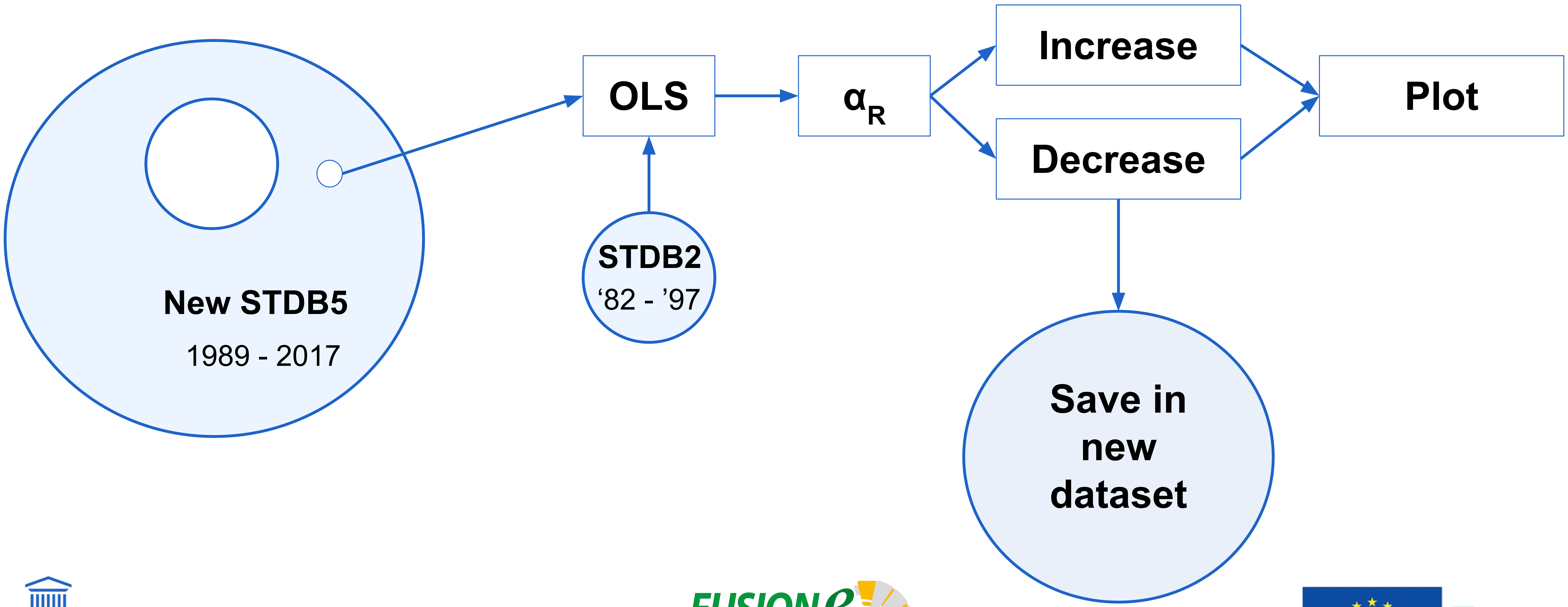


Dataset size = 6252

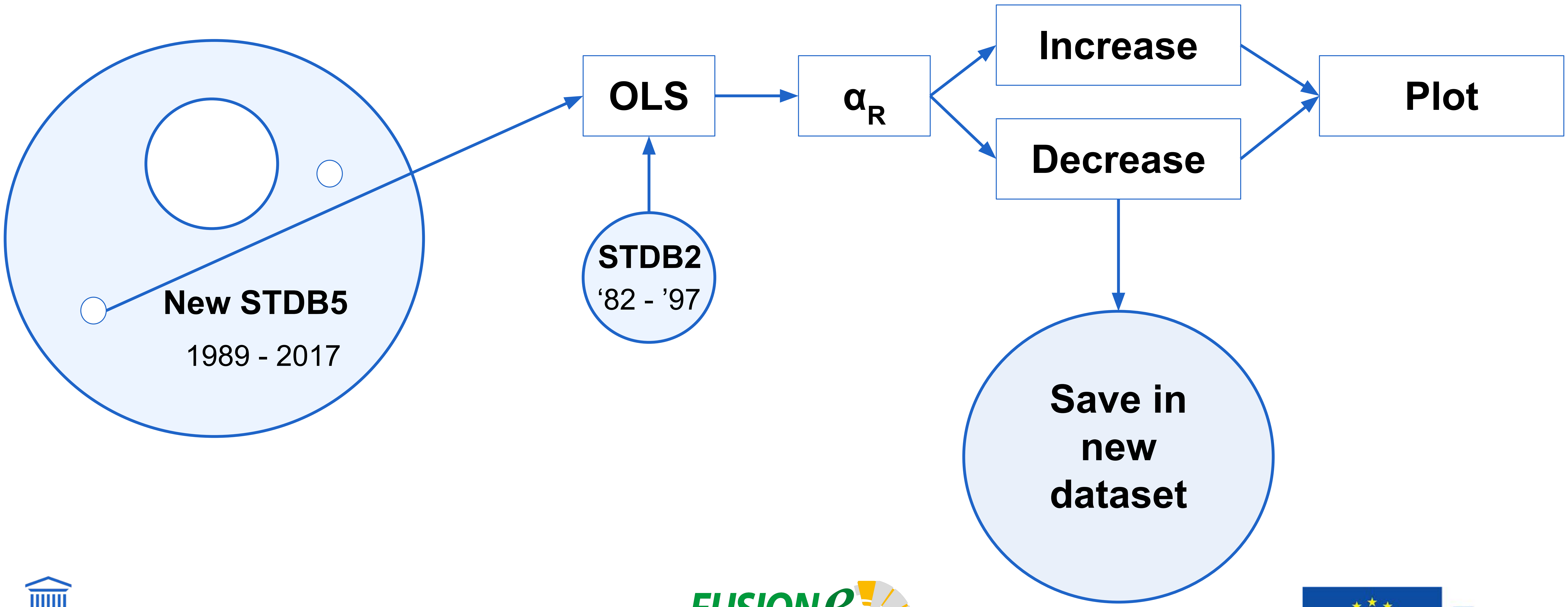
FINDING THE DECREASING REGISTERS



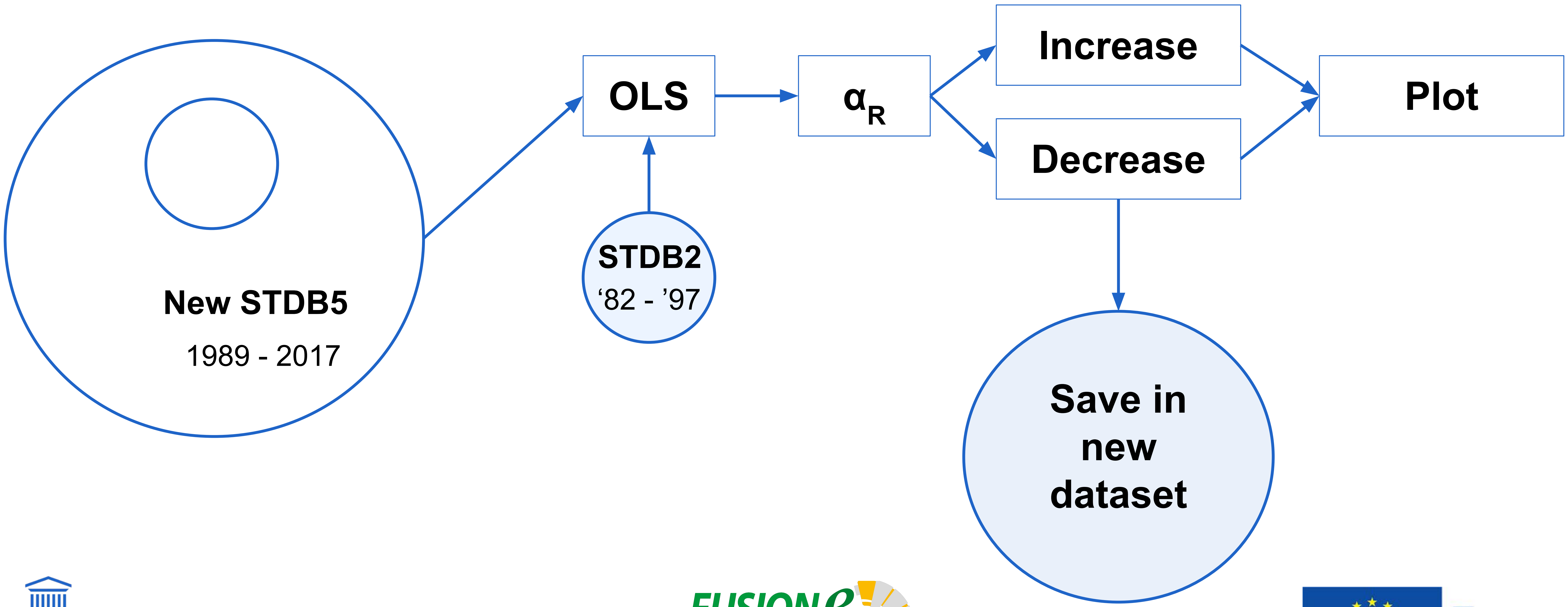
FINDING THE DECREASING REGISTERS



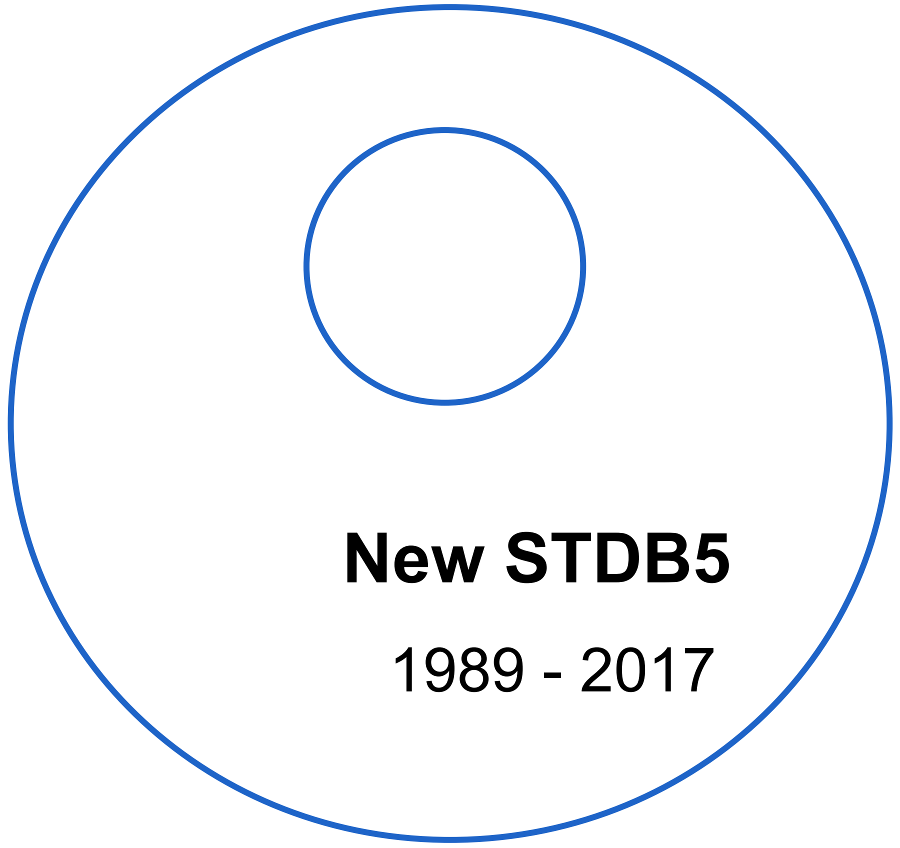
FINDING THE DECREASING REGISTERS

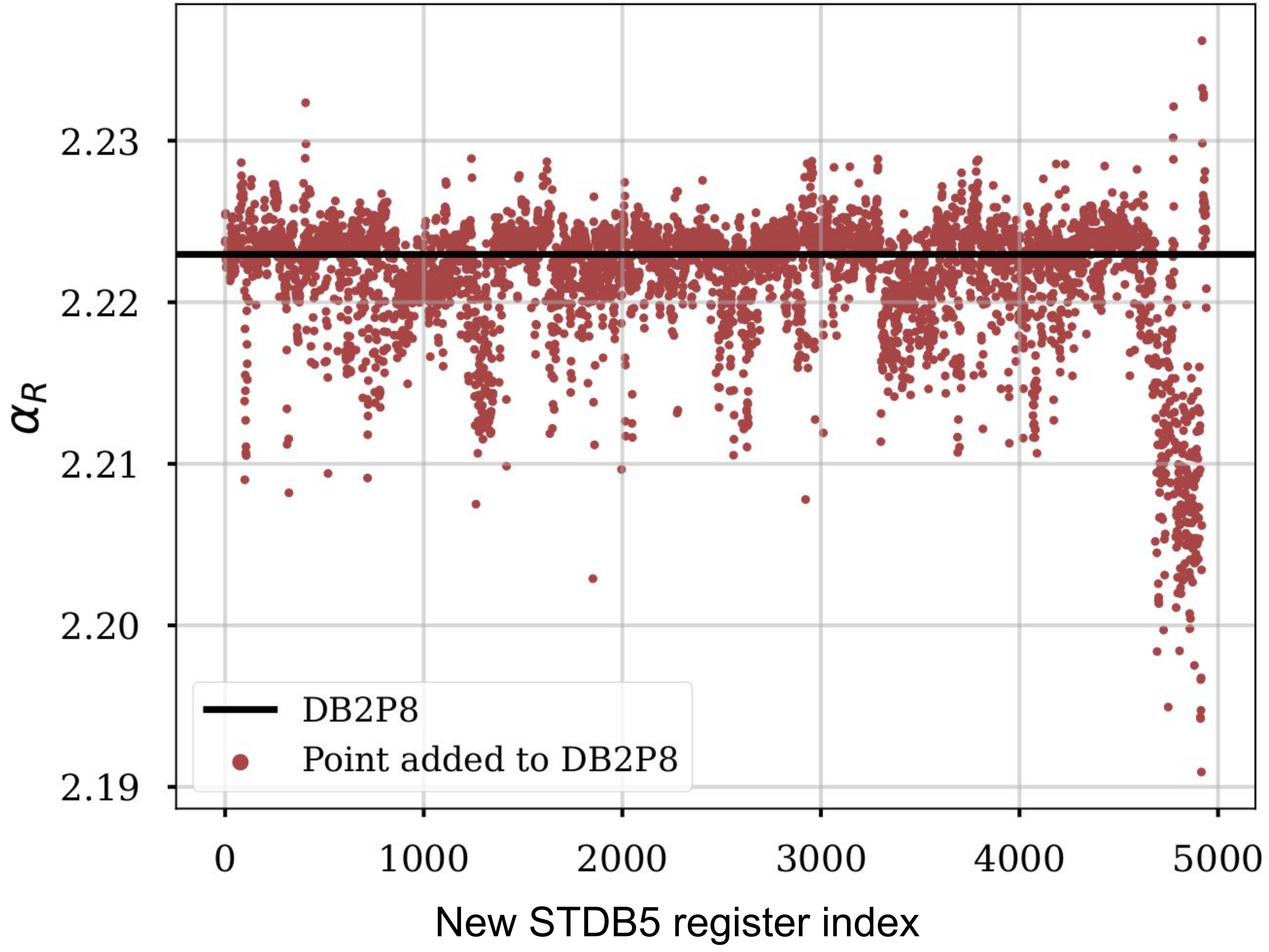


FINDING THE DECREASING REGISTERS



DECREASING REGISTERS

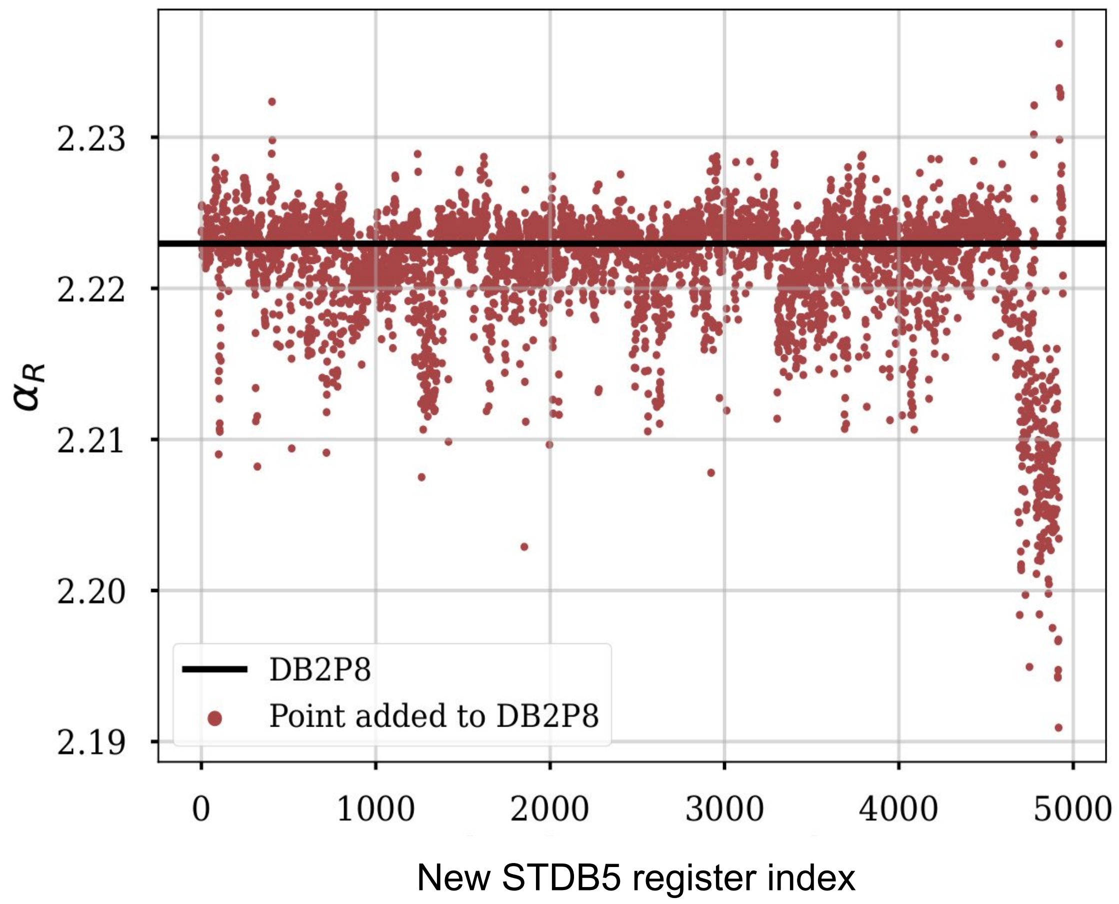




Plot

Random Sampling

DECREASING REGISTERS



Random Sampling

"Small dataset"	"Big dataset"
Smallest subset for $\alpha_R < 1$	**Smallest α_R**
$\alpha_R \sim 0.9998$	$\alpha_R \sim 0.6379$
Subset size = 618	Subset size = 1459
9.88% decreased α_R 90.12% did not	23.34% decreased α_R 76.66% did not

STDB5

STDB5

DECREASING AND UNAFFECTING REGISTERS

"Small dataset"	"Small dataset"	"Big dataset"	"Big dataset"
Smallest subset for $\alpha_R < 1$	**Complementary** for $\alpha_R < 1$	**Complementary** for smallest α_R	**Smallest** $\boldsymbol{\alpha_R}$
$\alpha_R \sim 0.99$	$\alpha_R \sim 1.71$	$\alpha_R \sim 2.16$	$\alpha_R \sim 0.64$
dataset size = 618	dataset size = 5634	dataset size = 4793	dataset size = 1459

DECREASING AND UNAFFECTING REGISTERS

“Small dataset”	“Small dataset”	“Big dataset”	“Big dataset”
Smallest subset for $\alpha_R < 1$	**Complementary** for $\alpha_R < 1$	**Complementary** for smallest α_R	**Smallest α_R**
$\alpha_R \sim 0.99$	$\alpha_R \sim 1.71$	$\alpha_R \sim 2.16$	$\alpha_R \sim 0.64$
dataset size = 618	dataset size = 5634	dataset size = 4793	dataset size = 1459

+ → Together they make **STDB5** ← +

dataset size = 6252

WORKFLOW

WHICH REGISTERS IN STDB5 ARE **DECREASING** α_R?

Depends on how you classify / label them.

IS THIS DECREASE DUE TO **DATA ISSUES**?

CAN WE **PREDICT** WHETHER A NEW REGISTER WILL **DECREASE** α_R?

RELATE THE FINDINGS **TOKAMAK CHARs & PHYSICS**

MULTICOLLINEARITY?

CONSEQUENCES

- Increase in the standard errors of each feature
- Numerical instability
- Unreliable models

VIF | **Condition Index** → **VD**

Variance Inflation Factor

Variance Decomposition

Used as predictors for the **target**

Features / Variables

I_p [MA]	B_t [T]	….	ϵ	$\tau_{E,th}$ [s]
0.2959	2.205	….	0.242	0.051
….	….	….	….	….
0.9794	4.795	….	0.327	0.1731
0.3740	1.744	….	0.286	0.0340

[W.-M Lee, 2021]
[D.Besley, E. Kuh, and R. Welsch, 2004]
[R. M. O'Brien, 2007]

VARIANCE INFLATION FACTOR

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$VIF_k = \frac{1}{1 - R_k^2}$$

$$Var[\beta_k] = \frac{\left(1 - R^2\right)\cdot\sum(y_i - \bar{y})^2}{n - M - 1}\cdot\frac{1}{\sum x_k^2}\cdot VIF_k$$

I_p [MA]	**B_t** [T]	….	**ϵ**	~~**$T_{E,th}$** [s]~~
0.2959	2.205	….	0.242	~~0.051~~
….	….	….	….	~~….~~
0.9794	4.795	….	0.327	~~0.1731~~
0.3740	1.744	….	0.286	~~0.0340~~

[D.Besley, E. Kuh, and R. Welsch, 2004]
[R. M. O'Brien, 2007]

VARIANCE INFLATION FACTOR

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$VIF_k = \frac{1}{1 - R_k^2}$$

$$Var[\beta_k] = \frac{\left(1 - R^2\right) \cdot \sum (y_i - \bar{y})^2}{n - M - 1} \cdot \frac{1}{\sum x_k^2} \cdot VIF_k$$

I_p [MA]	B_t [T]	….	ϵ	~~$\tau_{E,th}$ [s]~~
0.2959	2.205	….	0.242	~~0.051~~
….	….	….	….	~~….~~
0.9794	4.795	….	0.327	~~0.1731~~
0.3740	1.744	….	0.286	~~0.0340~~

[D.Besley, E. Kuh, and R. Welsch, 2004]
[R. M. O'Brien, 2007]

VARIANCE INFLATION FACTOR

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$VIF_k = \frac{1}{1 - R_k^2}$$

$$Var[\beta_k] = \frac{\left(1 - R^2\right)\cdot\sum(y_i - \bar{y})^2}{n - M - 1}\cdot\frac{1}{\sum x_k^2}\cdot VIF_k$$

I_p [MA]	B_t [T]	….	ϵ	~~$\tau_{E,th}$ [s]~~
0.2959	2.205	….	0.242	~~0.051~~
….	….	….	….	~~….~~
0.9794	4.795	….	0.327	~~0.1731~~
0.3740	1.744	….	0.286	~~0.0340~~

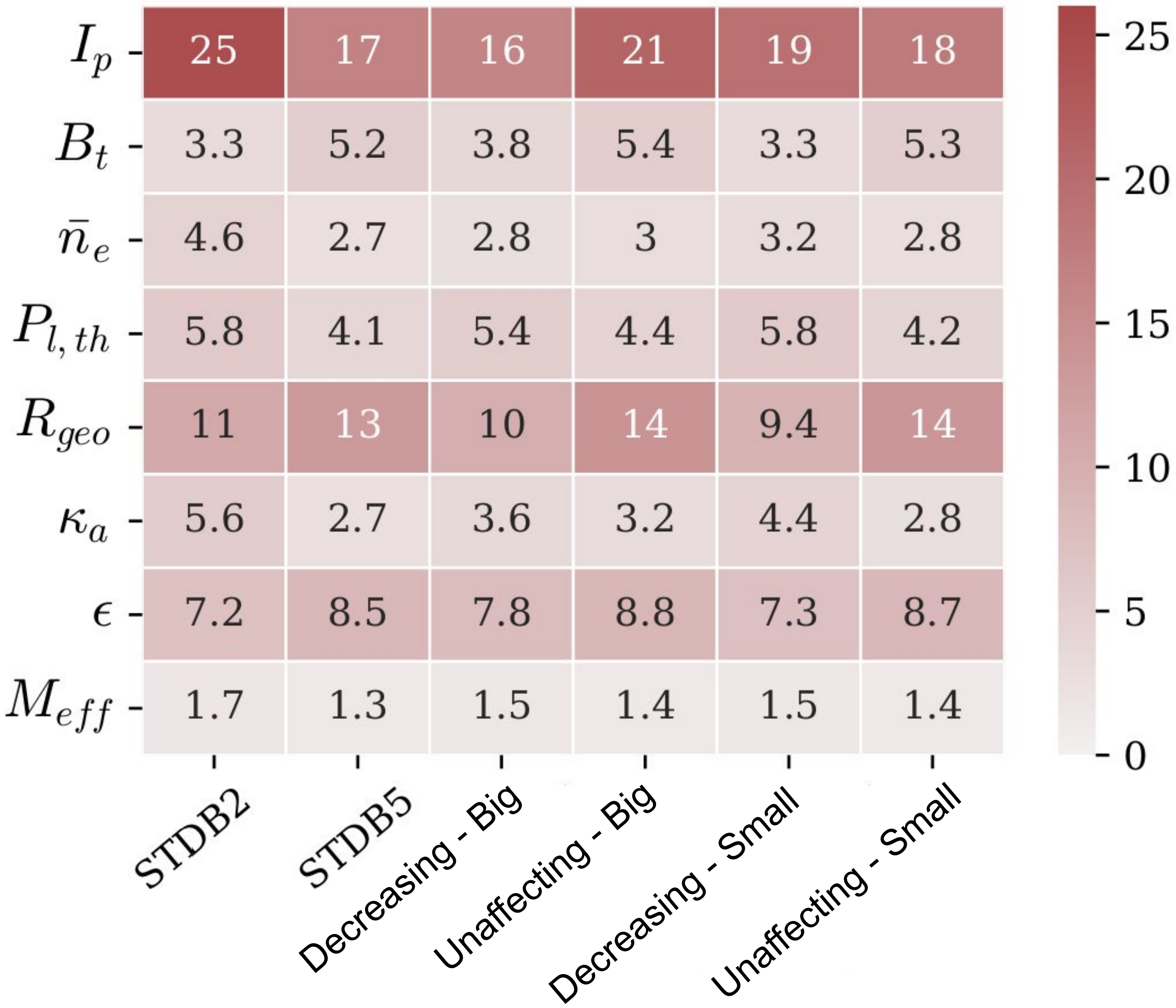
[D.Besley, E. Kuh, and R. Welsch, 2004]
[R. M. O'Brien, 2007]

VIF: RESULTS

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$VIF_k = \frac{1}{1 - R_k^2}$$

$$Var[\beta_k] = \frac{\left(1 - R^2\right) \cdot \sum (y_i - \bar{y})^2}{n - M - 1} \cdot \frac{1}{\sum x_k^2} \cdot VIF_k$$

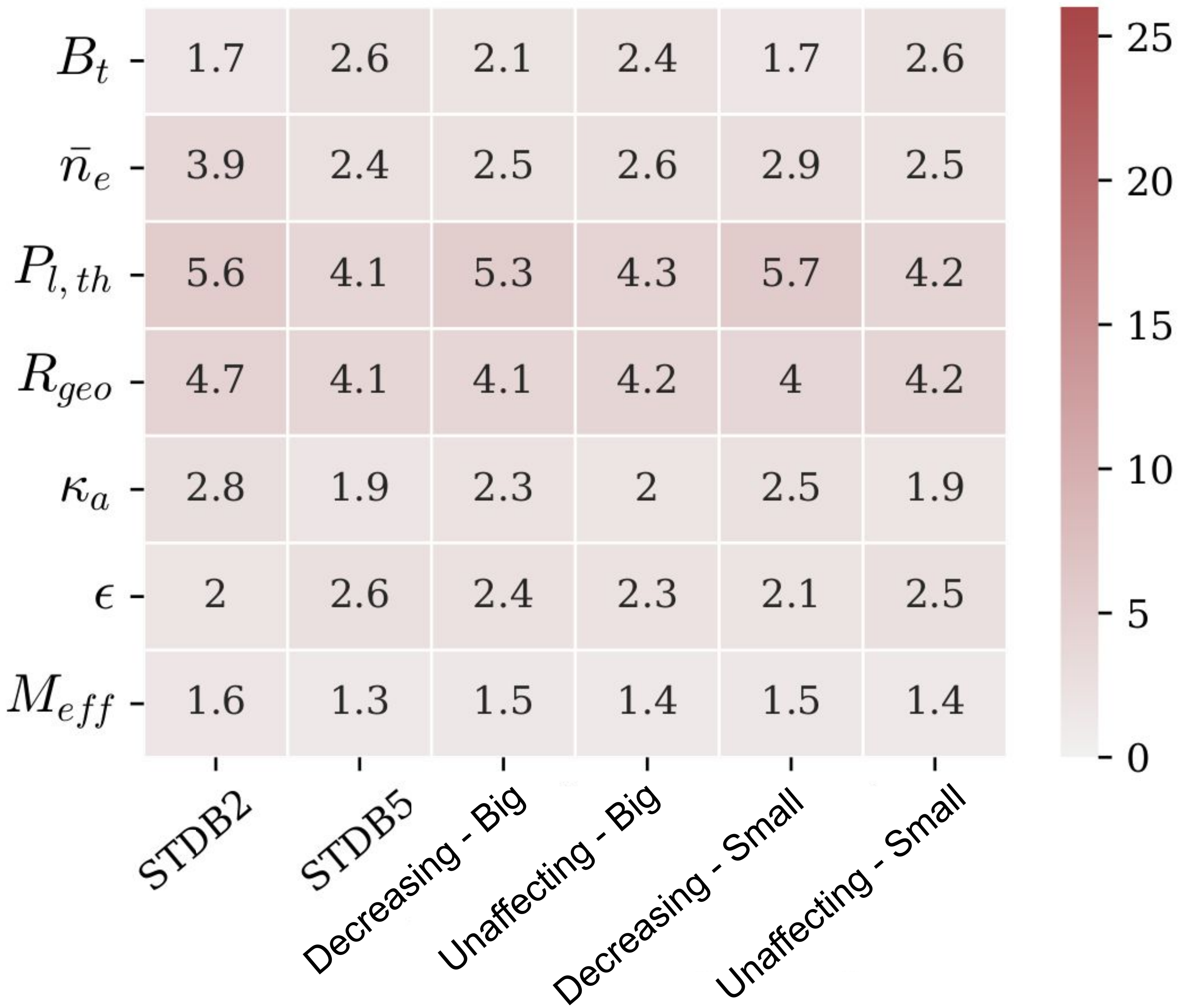


VIF: RESULTS

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$VIF_k = \frac{1}{1 - R_k^2}$$

$$Var[\beta_k] = \frac{\left(1 - R^2\right) \cdot \sum (y_i - \bar{y})^2}{n - M - 1} \cdot \frac{1}{\sum x_k^2} \cdot VIF_k$$



	STDB2	STDB5	Decreasing - Big	Unaffecting - Big	Decreasing - Small	Unaffecting - Small
B_t	1.7	2.6	2.1	2.4	1.7	2.6
$\bar{n}_e$	3.9	2.4	2.5	2.6	2.9	2.5
$P_{l,th}$	5.6	4.1	5.3	4.3	5.7	4.2
R_{geo}	4.7	4.1	4.1	4.2	4	4.2
κ_a	2.8	1.9	2.3	2	2.5	1.9
ϵ	2	2.6	2.4	2.3	2.1	2.5
M_{eff}	1.6	1.3	1.5	1.4	1.5	1.4

WORKFLOW

WHICH REGISTERS IN STDB5 ARE **DECREASING** α_R?

Depends on how you classify / label them.

→ IS THIS DECREASE DUE TO **DATA ISSUES**?

Yes, but not entirely.

→ CAN WE **PREDICT** WHETHER A NEW REGISTER WILL **DECREASE** α_R?

→ **RELATE** THE FINDINGS **TOKAMAK CHARs & PHYSICS**

FEATURES OF INTEREST: Entropy

47 columns

35 | 12

$$E = -\sum_{\substack{i,j=1 \\ i \neq j}}^{N} \left[S_{ij} \log(S_{ij}) + \left(1 - S_{ij}\right) \log(1 - S_{ij}) \right]$$

$$S_{ij} = \exp\left(-\gamma \cdot D_{ij}\right), \quad D_{ij} = \left[\sum_{k=1}^{M} \left(\frac{x_{ik} - x_{jk}}{max(F_k) - min(F_k)} \right)^2 \right]^{1/2}$$

$$S_{ij} = \frac{1}{M} \sum_{k=1}^{M} \delta_{ij}(x^k); \quad \text{with } \delta_{ij}(x^k) = \begin{cases} 1, \text{ if } x_i^k = x_j^k \\ 0, \text{ if } x_i^k \neq x_j^k \end{cases}$$

IDEA

Keep features that increased the entropy **when removed**.

[M. Dash and H. Liu, 2020]

FEATURES OF INTEREST: Entropy + Low MCL

47 columns

35 | 12

$$E = -\sum_{\substack{i,j=1 \\ i\neq j}}^{N} \left[S_{ij}\log(S_{ij}) + \left(1 - S_{ij}\right)\log(1 - S_{ij}) \right]$$

$$S_{ij} = \exp\left(-\gamma \cdot D_{ij}\right), \quad D_{ij} = \left[\sum_{k=1}^{M} \left(\frac{x_{ik} - x_{jk}}{max(F_k) - min(F_k)} \right)^2 \right]^{1/2}$$

$$S_{ij} = \frac{1}{M} \sum_{k=1}^{M} \delta_{ij}(x^k); \quad \text{with } \delta_{ij}(x^k) = \begin{cases} 1, \text{ if } x_i^k = x_j^k \\ 0, \text{ if } x_i^k \neq x_j^k \end{cases}$$

MULTICOLLINEARITY

Removing features with high VIF and causing high condition indices.

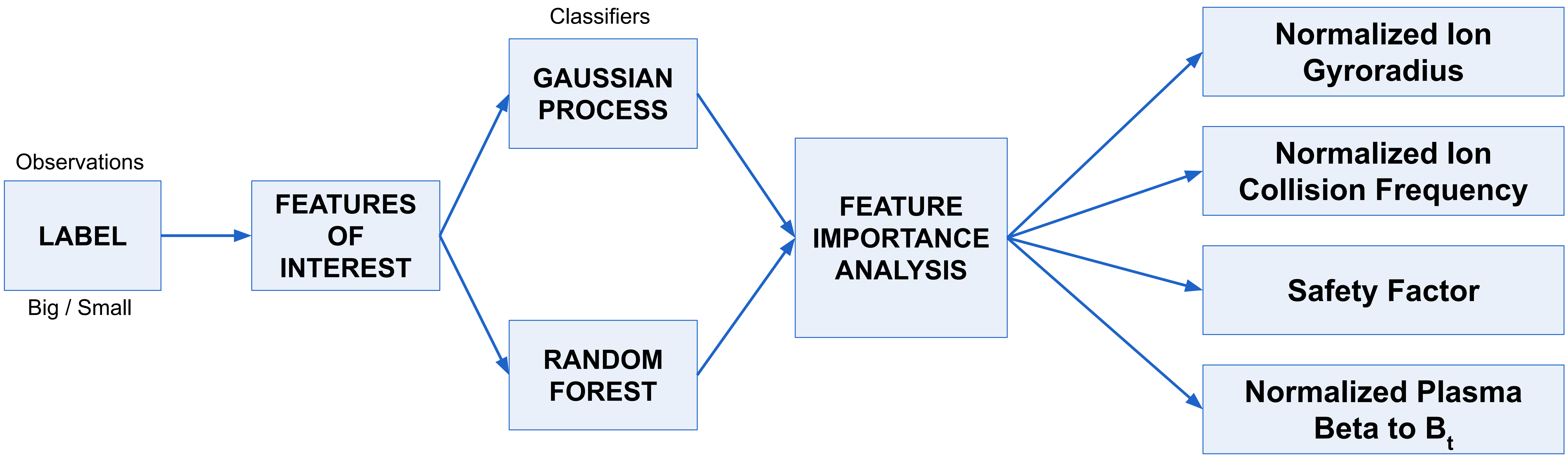
FEATURES OF INTEREST: Entropy + Low MCL

FEATURE	DESCRIPTION	FEATURE	DESCRIPTION
BETA	Plasma pressure normalised to $\mathbf{B}_t$	WFICFORM	Total fast ion energy due to ICRH estimated from approximate formula
RHOSTAR	Normalised Ion Gyroradius	WFFORM	Total fast ion energy due to NBI
NUSTAR	Normalised Ion collision frequency	ZEFFNEO	Line average plasma effective charge, from Bremsstrahlung
Q95	Plasma safety factor at the 95% poloidal flux surface	DWDIA	Time rate of change of the total plasma stored energy
PFLOSS	NBI power that is lost from the plasma through charge exchange and unconfined orbits		

Description of all columns in STDB5 found in [Princeton, 2021, http://arks.princeton.edu/ark:/88435/dsp01m900nx49h]

FEATURES FOR PREDICTION

Best Performance with “**big dataset**”



PREDICTIONS IN NEW DEVICES

	I_P	B_t	$\bar{n}_e$	$P_{l,th}$	R_{geo}	κ_a	ϵ	M_{eff}	ρ_*	β_t	ν_*	q_{95}	$\tau_{E,th}$
ITER	15	5.3	1.03	87	6.2	1.8	0.32	2.5	0.002	2.24	0.014	3	3.5
SPARC	8.7	12.2	3.1	25	1.85	1.97	0.31	2.5	0.003	1.20	0.03	3.2	0.77

UNAFFECTED

$$\tau_{E,2.16} = 0.06 \cdot I_p^{0.78} \cdot B_t^{0.24} \cdot \bar{n}_e^{0.41} \cdot P_{l,th}^{-0.75} \cdot R_{geo}^{2.16} \cdot \kappa_a^{0.5} \cdot \epsilon^{0.79} \cdot M_{eff}^{0.22}$$

WORKFLOW

WHICH REGISTERS IN STDB5 ARE **DECREASING** α_R?

Depends on how you classify / label them.

IS THIS DECREASE DUE TO **DATA ISSUES**?

Yes, but not entirely.

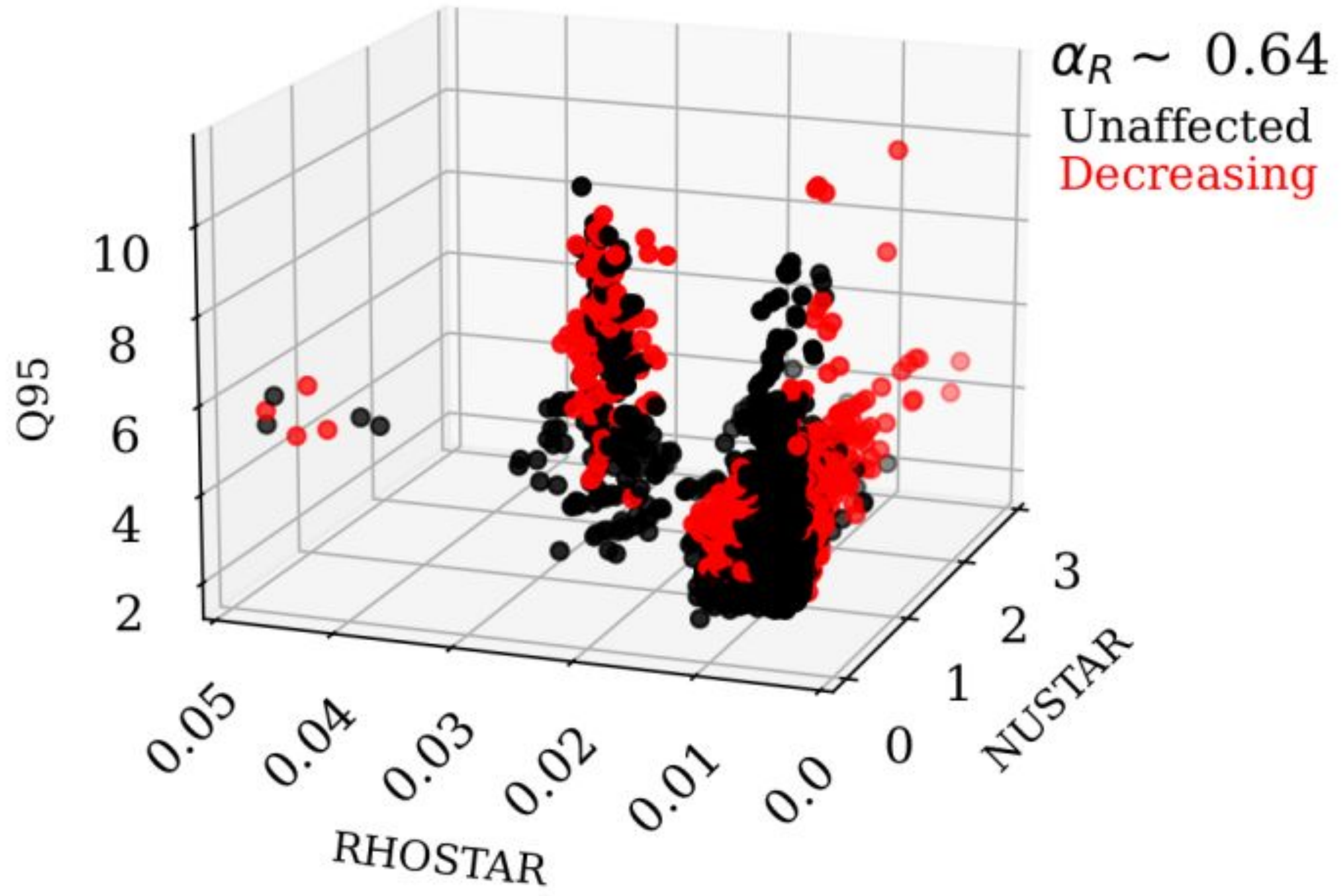
CAN WE **PREDICT** WHETHER A NEW REGISTER WILL **DECREASE** α_R?

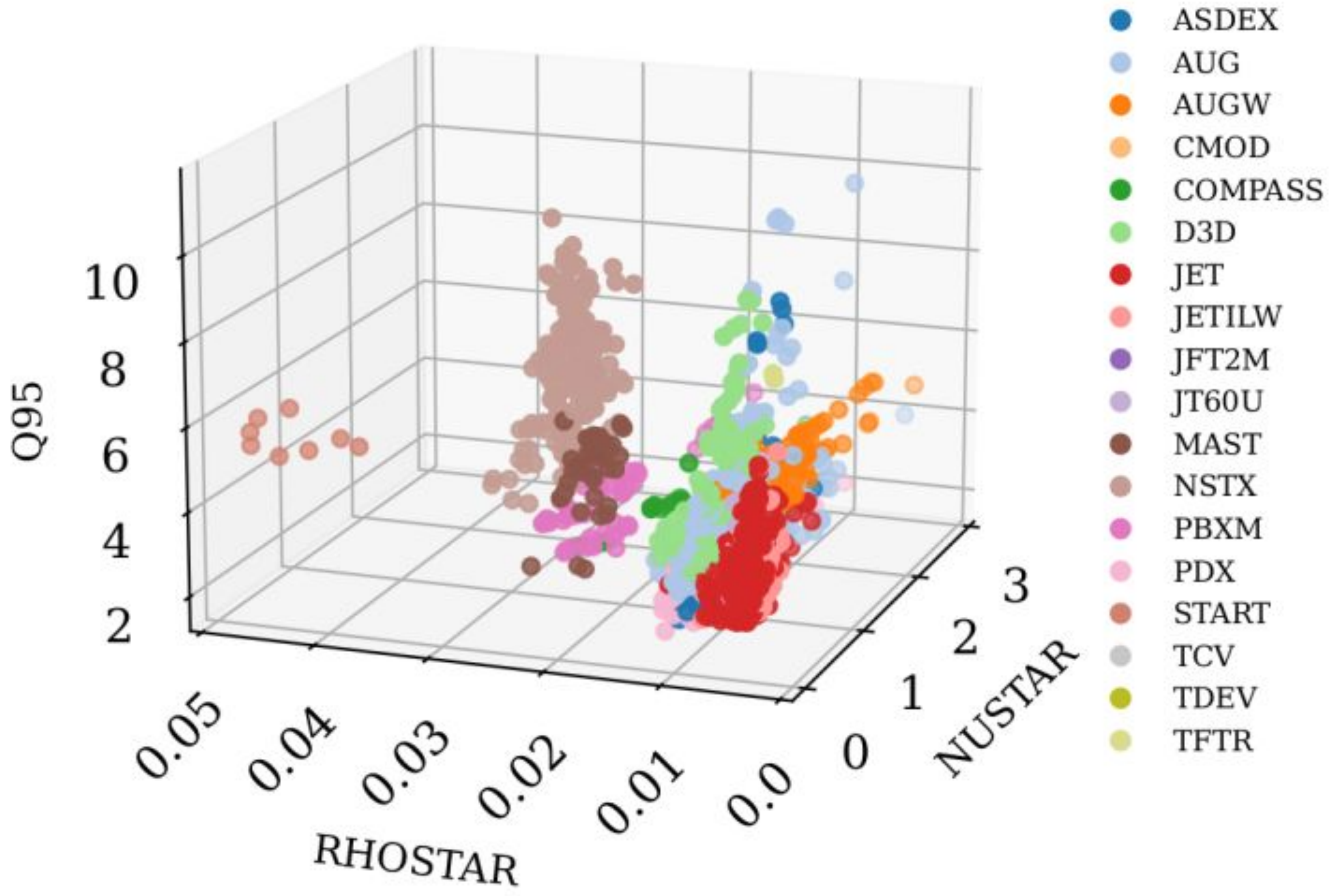
Yes, with only four variables.

RELATE THE FINDINGS **TOKAMAK CHARs & PHYSICS**

Characteristic Clusters: main features

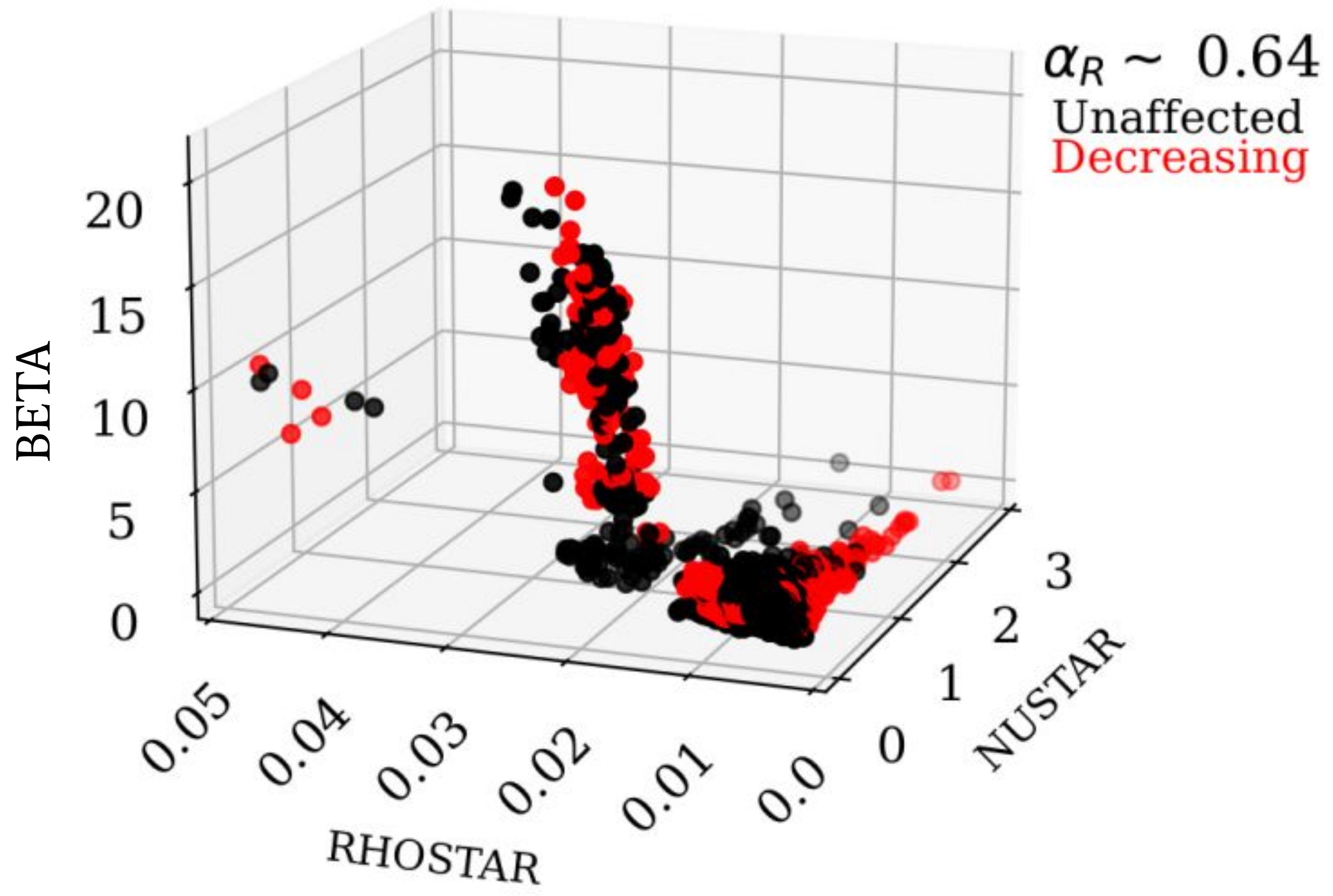
New STDB5

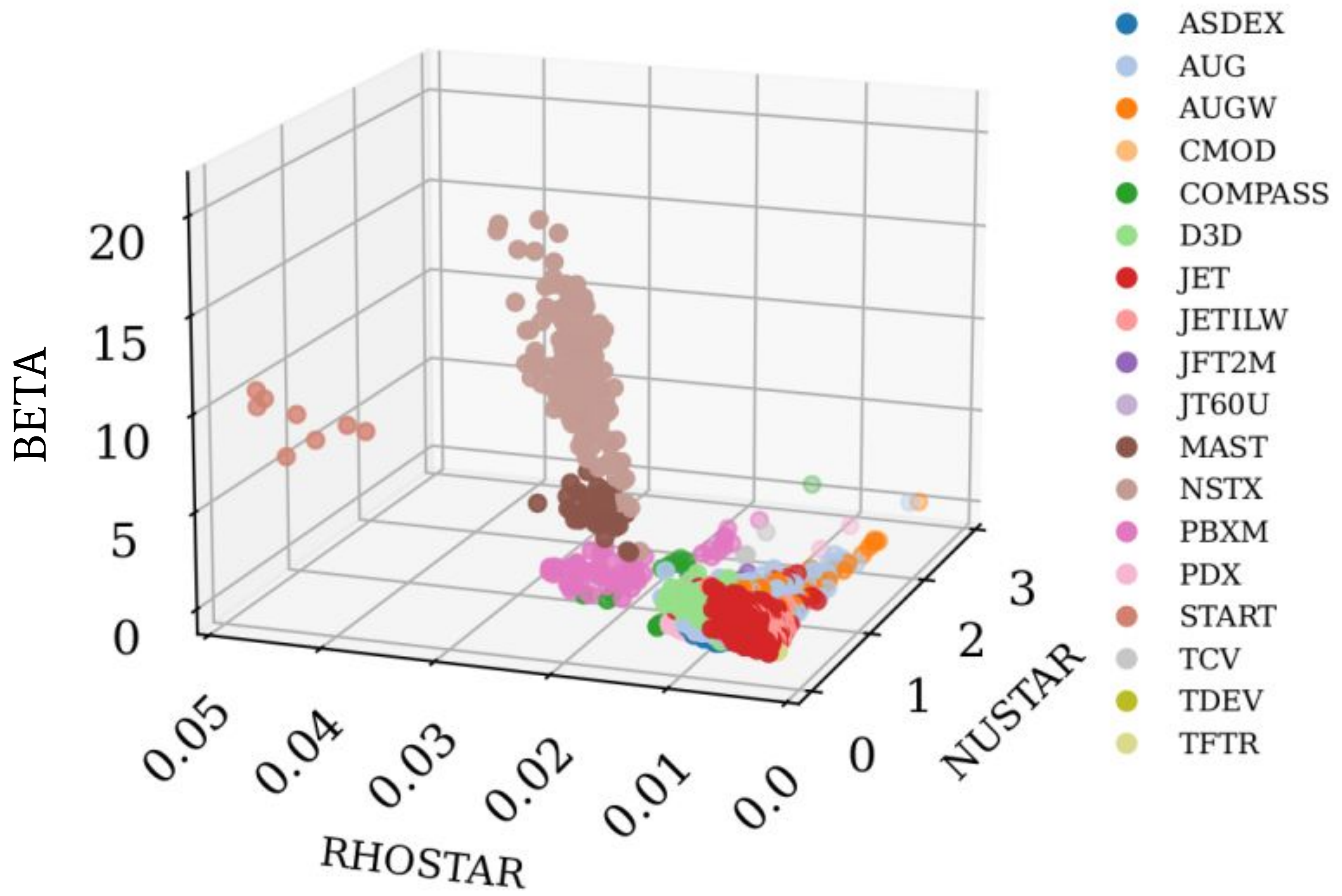




Characteristic Clusters: main features

New STDB5





Dimensionless Scaling and the Normalized Ion Gyroradius

$$\omega_i \cdot \tau_{E,the} \propto \rho_*^{-(2+\chi'_\rho)} \cdot F(\nu_*, \beta_t, \{p_i\})$$ with $0 \leq \chi'_\rho \leq 1$

$\boxed{\chi'_\rho = 0.9 \pm 0.3}$

H-mode

characteristic turbulence scale length $\ell \approx \rho_*^{\chi'_\rho} \cdot a^{1-\chi'_\rho}$

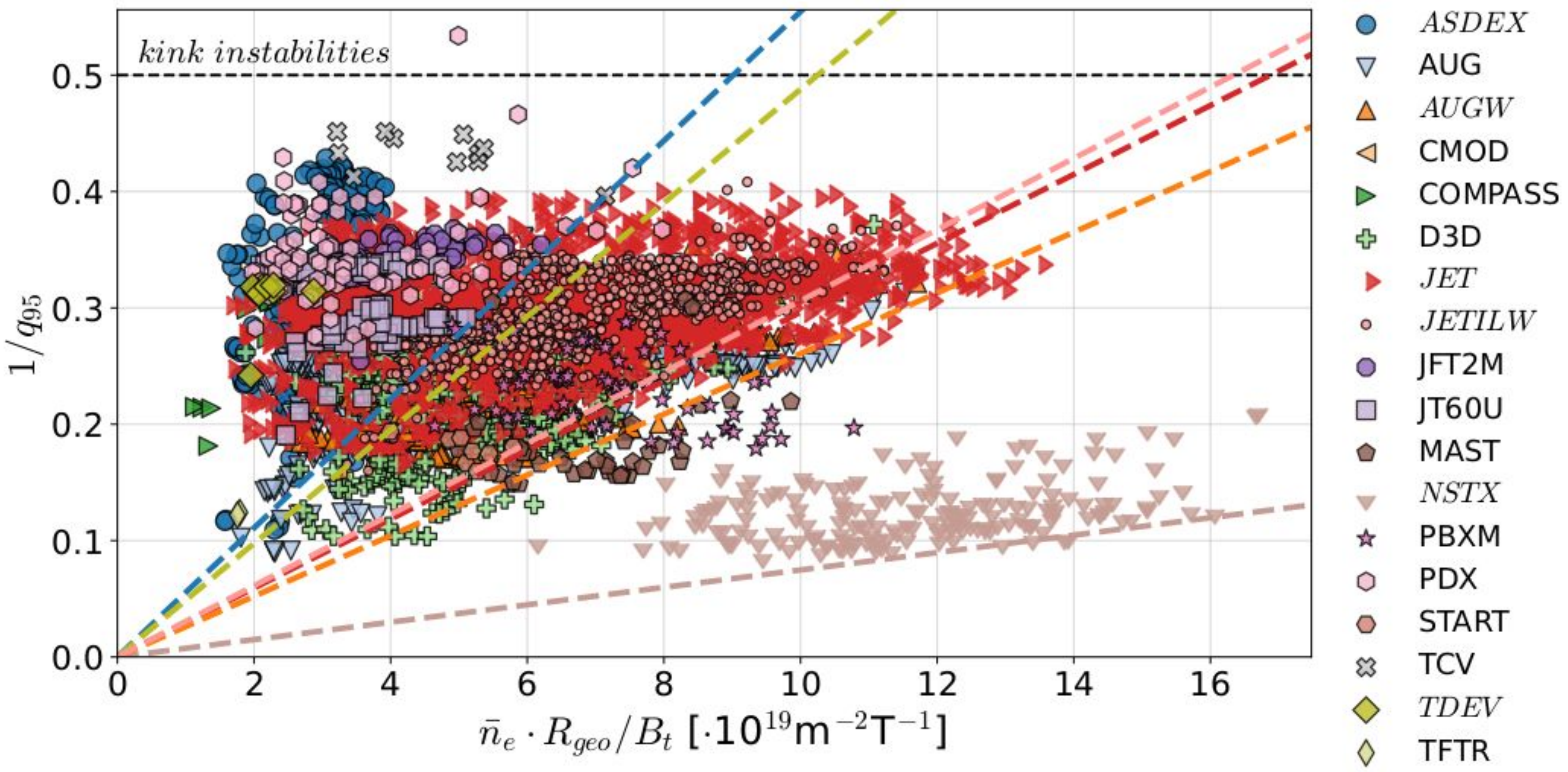
				big_ds		small_ds	
		DB2	STDB5	$\tau_{E,0.64}$	$\tau_{E,2.16}$	$\tau_{E,0.99}$	$\tau_{E,1.71}$
	χ_ρ	-3.09	-1.80	-1.31	-2.63	-1.5	-2.08
	χ'_ρ	1.09	0.2	0.69	0.63	-0.5	0.08
ITER	ℓ [m]	0.0012	7.99	227.43	0.026	64.99	1.15
SPARC	ℓ [m]	0.0017	1.67	22.33	0.019	8.46	0.37

[G. Verdoolaege et al., 2021]
[I. P. E. G. et al., 1999]

Density Limits

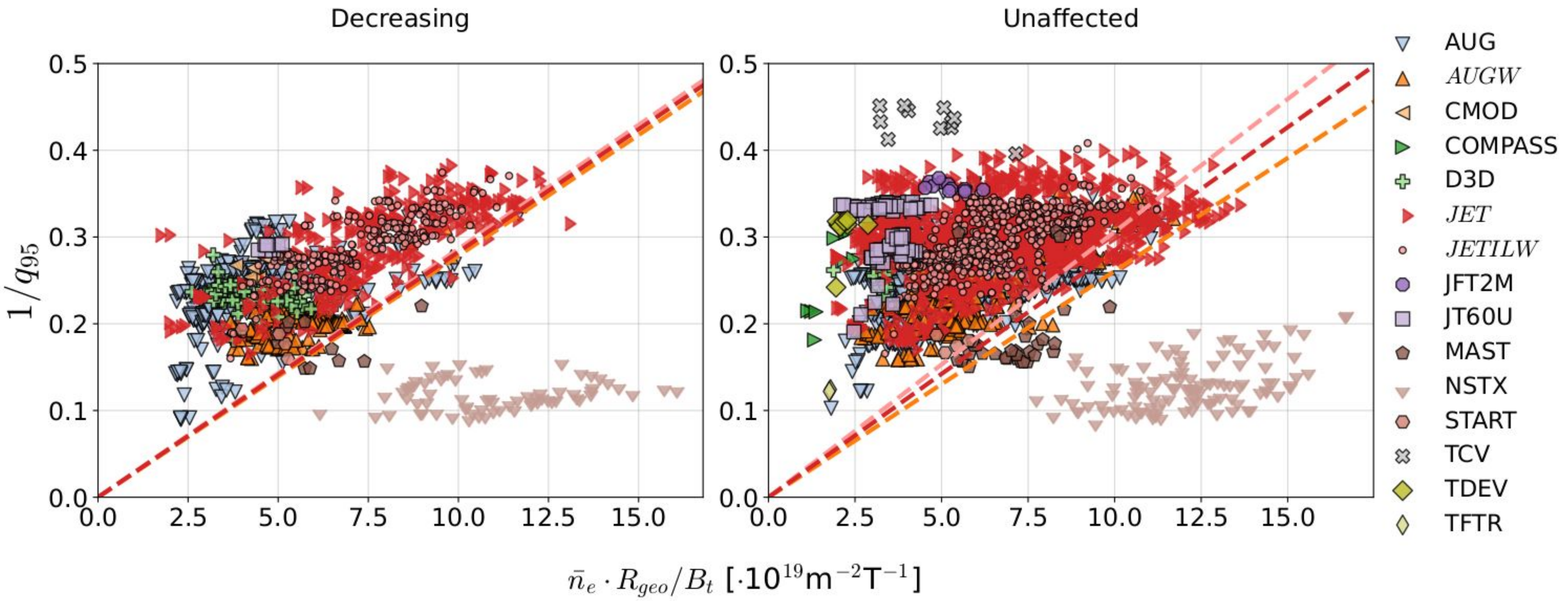
STDB5

$$n_{GW} = \frac{10 \cdot I_p}{\pi a^2}$$



Density Limits

New STDB5



Conclusions and Further Work

CONCLUSIONS

- Multicollinearity is **not** the sole factor influencing α_R.
- It is possible to predict influencing observations with
- Chara $\rho_* \quad \beta_t \quad \nu_* \quad q_{95}$ r spherical tokamaks.
- It is possible to tell if a dataset lacks of representative physics.
- Most of the observations that surpassed the GW limit are classified as unaffecting.
- ITER and SPARC are expected to follow a scaling law similar to the 1998 scaling (great news!).

FURTHER WORK

- Obtain more data on spherical tokamaks.
- Discern between spherical and non spherical.
- Take time-series data of relevant shots and subject to other ML algorithms; e.g. **surrogate modelling**.

Underlying Physics

Plasma Facing Components

Karina Chiñas Fuentes
Master of Science Student

FUSION-EP Programme

E karina.chinasfuentes@ugent.be
E nuclear.fusion@ugent.be
L Sint Pietersnieuwstraat 41, 9000 Gent

www.ugent.be
https://nuclearfusion.ugent.be/

infusion

Universiteit Gent
@ugent
@ugent
Ghent University

Thank you! Questions?

Complete References

- Zohm, Hartmut. “Magnetohydrodynamic stability of tokamaks”. John Wiley & Sons, 2015.
- “Machine,” ITER, 2023. Available: https://www.iter.org/mach. [Accessed: Jul. 05, 2023]
- W.-M Lee, “Statistics in Python — Collinearity and Multicollinearity”. Medium, 2021. [Online]. Available: https://towardsdatascience.com/statistics-in-python-collinearity-and-multicollinearity-4cc4dcd82b3f
- D. Belsley, E. Kuh, and R. Welsch, “Regression Diagnostics”. Canada: Wiley-Interscience, 2004.
- R. M. O’Brien, “A caution regarding rules of thumb for variance inflation factors,” Qualitative & Quantitative, vol. 41, pp. 673–690, Oct. 2007, Published on March 13, 2007. doi: 10.1007/s11135-006-9018-6.
- G. Verdoolaege et al., “The updated ITPA global h-mode confinement database: Description and analysis,” Princeton Plasma Physics Laboratory, Princeton University, 2021. [Online]. Available: http://arks.princeton.edu/ark:/88435/dsp01m900nx49h.
- M. Dash and H. Liu, “Feature selection for clustering,” Arizona State University, Jun. 2020. [On- line]. Available: https://www.public.asu.edu/~huanliu/papers/pakdd00clu.pdf.
- G. Verdoolaege et al., “The updated ITPA global H-mode confinement database: Description and analysis,” Nuclear Fusion, vol. 61, May. 2021. doi: 10.1088/1741-4326/abdb91.
- I. P. E. G. on Confinement, Transport, I. P. E. G. on Confinement Modelling, Database, and I. P. B. Editors, “Chapter 2: Plasma confinement and transport,” Nuclear Fusion, vol. 39, no. 12, p. 2175, Dec. 1999. doi: 10.1088/0029-5515/39/12/302.