Joint EPS paper

Singularization of data subgroups in the International Stellarator-Heliotron Confinement Database

A. Kus[1], A. Dinklage[1,] E. Ascasibar[2], C.D. Beidler[1], T. Estrada[2], H. Funaba[3], J. Geiger[1], J.H. Harris[4,5], C. Hidalgo[2], D. Lopez-Bruna[2], A. Lopez-Fraguas[2], H. Maaßberg[1], T. Minami[3], T. Mizuuchi[6], S. Murakami[6], N. Nakajima[3], S. Okamura[3], D. Pretty[4], F. Sano[6], Y. Suzuki[3], Y. Takeiri[3], J. Talmadge[7], K. Thomsen[8], Yu. A. Turkin[1], K.Y. Watanabe[3], R. Wolf[1], Y. Xu[9], H. Yamada[3], M. Yokoyama[3]

[1]Max-Planck-Institut für Plasmaphysik, Euratom Assoc., Greifswald, Germany, [2]Laboratorio Nacional de Fusión, Asociación Euratom/Ciemat, Madrid, Spain, [3]National Institute for Fusion Science, Toki, Japan, [4]Australian National University, Canberra, Australia, [5]Oak Ridge National Laboratory, Oak Ridge, USA, [6]Kyoto University, Kyoto, Japan, [7]University of Wisconsin, Madison, USA, [8]European Commission, DG Research & Innovation, Unit K6, Brussels, Belgium, [9]Southwestern Institute of Physics, ChengDu City, China

A. Kus, A. Dinklage, Y. Xu

CWGM 10, IPP-Greifswald, 6-8 June 2012

The goal of the work

- With the ISS04 we have a reference scaling for the existing stellarator-heliotron devices

- The next step would be to ask whether it is possible to assess the energy confinement for the future devices;

  Following the principles of <u>similarity</u> and <u>scale invariance</u> of confinement, we have to work with dimensionless variables
  (after transformation coefficients of a *dimensionally correct* model, calculated in engineering variables, into dimensionless quantities)

Further issues:

- Is a common scaling possible at all?
- What is the correct set of scaling parameters?
- Is the ISHCDB dataset sufficient for the required analyses?
- Which parameters distinguish different experimental scenarios (high beta, CERC, high performance, …) – new data necessary

Proposed approach

1. Define the set of relevent dimensionless variables
2. Apply statistical tools to check if there exist natural subgroups in the data (cluster analysis)
3. Identify variables causing the cluster formation (discriminant analysis)
4. Check whether the clusters scale differently

   (Because of small, or missing, variation in the individual variables, different models have to be used for different clusters – model comparison is then possible only via R_{adj} and *RMSE* parameters)
5. How to derive a common scaling?

   Are renormalization factors still justified?

Used variables

- There exist many different sets of possible dimensionless variables (C.C.Petty, Phys. Plasmas 15, 080501 (2008), T.C. Luce, et al., PPCF 50 043001 (2008), …). But not enough data are available in ISHCDB to apply all of them

- Variables used in this paper
  - RHOSTAR, NUSTAR, BETA
  - Lackner's *dimless engineering* variables (FST, 54, 4, 989 (2008))

$$KL_1 = B_t\, a^{5/4}, \quad KL_2 = P_{heat}\, a^{3/4}$$

  - Derived from the Connor-Taylor constraints (NF 17, 1047 (1977)) the following variables may be incorporated in the power law ansatz:

$$CT_1 = \frac{P}{n\,a^4 B^3}, \quad CT_2 = \frac{a^3 B^4}{n}, \quad CT_3 = \frac{1}{n\,a^2}, \qquad W \sim CT_1^{\,x1}\, CT_2^{\,x2}\, CT_3^{\,x3}$$

Data in {TAU, RHOSTAR, NUSTAR, BETA}

Dataset: ISHCDB_26xg_allData,
Variables: TAU, RHOSTAR, NUSTAR, BETA

01.06.2012 14:04

Hierarchical Clustering

Method = Ward

Dendrogram

Clustering History

Data in clusters

No.	Data	Nobs	C1	C2	C3	C4	C5
1	ATF	229	67	162	0	0	0
2	CHS	196	172	24	0	0	0
3	HELE	120	39	81	0	0	0
4	HELJ	54	35	19	0	0	0
5	LHD	1362	0	84	341	0	937
6	TJ-II	318	284	34	0	0	0
7	W7-A	13	12	0	0	1	0
8	W7-AS	1083	285	265	1	532	0
9	W7-X	0	0	0	0	0	0
		.	.	.	.	.	.
		3375	894	669	342	533	937

Discriminator: TAU, BETA

Clusters 3-4-5 created by LHD and W7AS

(W7-X: predictive data)

Data in {TAU, KL1, KL2}

Dataset: ISHCDB_26xg_allData,
Variables: TAU, KL1, KL2

01.06.2012 14:13

Hierarchical Clustering

Method = Ward

Dendrogram

Clustering History

Data in clusters

No.	Data	Nobs	C1	C2	C3	C4
1	ATF	229	229	0	0	0
2	CHS	196	196	0	0	0
3	HELE	120	120	0	0	0
4	HELJ	54	54	0	0	0
5	LHD	1362	0	962	400	0
6	TJ-II	318	318	0	0	0
7	W7-A	13	13	0	0	0
8	W7-AS	1083	1082	0	1	0
9	W7-X	71	0	0	7	64
		.	.	.	.	.
		3446	2012	962	408	64

Discriminator: TAU, KL2

- LHD in two clusters
- W7-AS in one cluster
- W7-X mostly in one cluster

Data in {TAU, CT_1, CT_2, CT_3}

Dataset: ISHCDB_26xg_allData,
Variables: TAU, CT1, CT2, CT3

04.06.2012 11:10

Hierarchical Clustering

Data in clusters

No.	Data	Nobs	C1	C2	C3
1	ATF	229	214	15	0
2	CHS	196	151	45	0
3	HELE	120	113	7	0
4	HELJ	54	25	29	0
5	LHD	1362	1297	0	65
6	TJ-II	318	1	317	0
7	W7-A	13	0	13	0
8	W7-AS	1083	1008	75	0
9	W7-X	71	10	0	61
		.	.	.	.
		3446	2819	501	126

Discriminator: TAU, CT1, CT3

Small part of LHD and the main part of W7-X in one cluster

Summary: clusters in different spaces

Dataset	Nobs	Space	Correlations	nCL	Discriminator /device	Discriminator /gbeta
all Data	3461	{TAU, RHOSTAR, NUSTAR, BETA}	---	5	BETA, TAU (65)	---
LHD	1362	{TAU, RHOSTAR, NUSTAR, BETA}	RHOSTAR, BETA	2	---	RHOSTAR, TAU (85)
TJ-II	318	{TAU, RHOSTAR, NUSTAR, BETA}	---	4	---	---
W7AS	1098	{TAU, RHOSTAR, NUSTAR, BETA}	---	4	---	BETA (99)
	•			•		
all Data	3461	{TAU, CT1, CT2, CT3}	---	3	TAU, CT1, CT3 (63)	
LHD	1362	{TAU, CT1, CT2, CT3}	---	5	---	CT2 (93)
TJ-II	318	{TAU, CT1, CT2, CT3}	CT2, CT3	3	---	---
W7AS	1098	{TAU, CT1, CT2, CT3}		7	---	CT3 (96)
	•			•		
all Data	3461	{TAU, KL1, KL2}	---	4	TAU, KL2 (75)	---
LHD	1362	{TAU, KL1, KL2}	---	5	---	KL1 (91)
TJ-II	318	{TAU, KL1, KL2}	---	4	---	---
W7AS	1098	{TAU, KL1, KL2}	---	4	---	KL2 (99)

(Numbers in columns *Discriminator* stay for the percentage of correctly predicted assignments => discrim. quality.)

- In all spaces: no clear clustering in „allData“ (hence difficulties in discrim. anal.)
- Different discriminators for high/low beta in LHD and W7AS
- Not to be seen here: W7AS high-beta data tends to form one cluster; it is not the case in LHD

Planned further steps

Next todos (in collaboration with Y. Xu)

- Regressions on the clustered data
- Choosing the possible model and the proper scaling procedure for a common scaling for all clusters/devices;
  Identification of needed ISHCDB extensions
- Extend ISHCDB with new data
- When we have enough dimensionless variables (suitable to stellarators/heliotrons) – how to transform scaling formulas from dimensional to dimensionless exponents? (For tokamaks, such transformations are listed in T.C. Luce, et al., PPCF 50 043001 (2008))
- Regressions using errors-in-variable techniques (premised estimated errors are available – partly already evaluated)