# Statistical analyses of confinement database

A. Kus[1], A. Dinklage[1], M. Yokoyama[2]
on behalf of all database contributors

[1]Max-Planck-Institut für Plasmaphysik, Euratom Assoc., Greifswald, Germany,
[2]National Institute for Fusion Science, Toki, Japan

CWGM 11, CIEMAT, Madrid, 11- 13 March 2013

# Goals of statistical analyses on ISHCDB

**General topic: global energy confinement scaling**
**Starting point: ISS04 scaling (2005)**

**Core idea of the ISS04 scaling procedure**

1. Split all data into 14 subgroups by devices & magnetic configurations (because of systematic offsets in the estimated confinement), and
2. Renormalize tauE (referring to ISS95, each subgroup with diff. factor)

**ISS04 approach implies some questions**

- ❑ Is this the only possible grouping?
- ❑ Do exist any “natural” subgroups (clusters) in the data?
- ❑ How the ISS04 scaling looks like, if the original 14 subgroups have been replaced with the identified clusters? (using dimensional/engineering variables)
- ❑ Could the acquired experience and know-how be applied for works towards a predictive scaling? (working with dimensionless variables)

**Used statistical tools**

- Cluster analysis (to identify possible clusters, w/o explanation their existence)
- Discriminant analysis (to identify variables causing cluster formation)

# Cluster analysis in engineering variables

Main results of cluster analysis of the ISS04 dataset (cf. CWGM 7, 8; EPS-2011) in the 7-dimensional space spanned by the ISS04 engineering variables {log_TAU, log_a, log_R, log_P, log_n, log_B, log_iota}

1. There exist cohesive subgroups (clusters) in the data
2. Clusters differ from ISS04 subgroups but resemble ISS04 grouping roughly
3. Geometry has largest impact on the cluster formation
4. Saturation in regression coefficients when increasing number of clusters

Idea

- ISS04 is a reference scaling for assessment of global stellarator confinement
- Results of cluster and discriminant analysis on the ISS04 data are encouraging

**?** Could we apply cluster and discriminant analysis for work towards a predictive scaling for future devices?

# About predictive confinement scalings

The base to assess the plasma confinement in differently sized devices is the principle of similarity: *if a set of dimensionless parameters fully decribes the plasma behaviour in one device, then it holds for all devices with the same dimensionless parameters, independently of the engineering quantities.*

→ *we need to work with dimensionless variables.*

How to obtain a proper set of variables to derive a predictive scaling formula?

1. Define a set of possible dimensional candidate-variables
2. Apply either
   a) *dimensional analysis* (Buckingham's $\Pi$ theorem ...), or
   b) use the *scale invariance principle* (Connor & Taylor)
3. Use similarity principle to define a relevant scaling model in dimensionless variables
4. Translate scaling model from dimensionless to dimensional+dimensionless variables

✓ Result is a dimensionally correct scaling law (without any dimensional dependencies)
✓ Scaling formula may be fully expressed in both dimensionless and engineering variables

Problems in the practice

- So far, no set of candidate–variables is defined/proposed
- Very laborious way when using dimensional analysis
- Not enough data (even to try some suggestions, e.g. Luce, et al., PPCF **50**, 043001 (2008))
- We can not follow this procedure

# Cluster analysis in different spaces

Attempt: cluster analysis in spaces determined by sets of variables related to dimensionally correct scalings

{ **rhostar**, **nustar**, **beta** } (ITER Phys., NF, 39, 2175 (1999)),
{ **bstar** = $B_t\, a^{5/4}$, **pstar** = $P_{heat}\, a^{3/4}$, **nstar** = n $a^{3/4}/B_t$ } (Lackner, FST, 54, 4, 989 (2008)),
{ $\mathbf{CT_1} = P_{heat}/n\, a^4 R\, B$, $\mathbf{CT_2} = a^3 B^4/n$, $\mathbf{CT_3} = n\, a^{3/4}/B_t$} (Dose, et al., Phys. Rev. Letters, 80, 16, (1998)).

Main results (all devices jointly)

- In all spaces clusters of data can be identified
- Separation of the data is not satisfactory clear (in single devices clusters are more distinct)
- As a consequence of unclean clustering, discriminant analysis supplies unconvincable results
- Regressions

1. Restricted (coll. high-beta constraint) on single clusters (no renormalization, no weighting)
   - o Individual clusters scale differently
   - o Coefficients for absorbed power, plasma density and magnetic field are partly in accordance with the ISS04 scaling
   - o By contrast, the rational transform coefficient shows a large variation, which may suggests that this parameter should be removed from the model
2. ISS04 procedure on clustered datasets (clusters used instead of ISS04 subgroups)
   - o A similar tendencies as for single clusters

- In all regresions: $a_a + a_R$ ~ (2.7 – 3.5) ISS04: $a_a + a_R = 2.28 + 0.64 = 2.92$

# Final comments

## Conclusions

- The ISHCDB dataset seems to be statistically inhomogeneous
- The ad-hoc separation in subsets as for the ISS04 scaling is not reflected by statistical cluster analysis in examined spaces
- Based on the existent database, it is not possible to make a point of a unified scaling law
- Predictive stellarator-heliotron scaling? A long-term task, if at all possible …

## Proposed next steps towards a predictive scaling

- ❑ Extension of the ISHCDB database by data from differently sized and shaped devices and for different operation/plasma regimes
- ❑ Specification of parameters allowing scalings with predictive qualities
  - Use Dimensional Analysis?
  - Use Connor-Taylor's invariance?
- ❑ (Improvement of statistical quality of analyzed data (errors in variables, …))

➢ Extended joint EPS-2012 paper now in preparation as a PPCF publication

# Regressions in three spaces: summary

| Case | Dataset | Nobs | RMSE | aa | ar | ap | an | ab | ai |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ISS04 scaling | 1721 | 0.0267 | 2.28 | 0.64 | -0.61 | 0.54 | 0.84 | 0.41 |
| 2 | ISHCDB_26xg_allData_DIML_c_5_0_CL_1 | 894 | 0.0782 | 2.27 | 1.16 | -0.73 | 0.70 | 1.06 | -0.20 |
| 3 | ISHCDB_26xg_allData_DIML_c_5_0_CL_2 | 669 | 0.1099 | 2.30 | 1.17 | -0.64 | 0.62 | 1.16 | -0.17 |
| 4 | ISHCDB_26xg_allData_DIML_c_5_0_CL_3 | 342 | 0.0669 | • | • | -0.54 | 0.34 | 0.66 | 1.35 |
| 5 | ISHCDB_26xg_allData_DIML_c_5_0_CL_4 | 533 | 0.0640 | • | • | -0.55 | 0.71 | 0.86 | -0.03 |
| 6 | ISHCDB_26xg_allData_DIML_c_5_0_CL_5 | 937 | 0.0409 | • | • | -0.83 | 0.28 | 1.23 | -0.26 |
| 7 | ISHCDB_26xg_allData_CT_c_3_0_CL_1 | 2819 | 0.1120 | 2.21 | 1.02 | -0.44 | 0.52 | 1.01 | -0.07 |
| 8 | ISHCDB_26xg_allData_CT_c_3_0_CL_2 | 501 | 0.0815 | 2.08 | 1.01 | -0.58 | 0.75 | 0.63 | -0.21 |
| 9 | ISHCDB_26xg_allData_CT_c_3_0_CL_3 | 126 | 0.1033 | 0.16 | -0.88 | -0.49 | 0.36 | -1.86 | 4.11 |
| 10 | ISHCDB_26xg_allData_KL_c_3_0_CL_1 | 2013 | 0.0978 | 1.78 | 1.20 | -0.61 | 0.58 | 0.83 | -0.21 |
| 11 | ISHCDB_26xg_allData_KL_c_3_0_CL_2 | 682 | 0.0894 | • | • | -0.61 | 0.41 | 0.90 | 2.21 |
| 12 | ISHCDB_26xg_allData_KL_c_3_0_CL_3 | 751 | 0.0476 | • | • | -0.79 | 0.40 | 1.02 | 0.13 |
| 13 | ISS04xg | 1721 | 0.2130 | 2.28 | 0.64 | -0.61 | 0.54 | 0.84 | 0.41 |
| 14 | ISHCDB_26xg_allData_DIML_c_5_0 | 3446 | 0.2124 | 2.05 | 0.87 | -0.48 | 0.43 | 0.94 | -0.21 |
| 15 | ISHCDB_26xg_allData_CT_c_3_0 | 3446 | 0.2066 | 1.59 | 1.15 | -0.43 | 0.39 | 0.82 | 0.14 |
| 16 | ISHCDB_26xg_allData_KL_c_3_0 | 3446 | 0.2066 | 1.90 | 1.22 | -0.52 | 0.56 | 0.91 | -0.03 |

DIML => {rhostar, nustar, beta}
CT => {$CT_1$, $CT_2$, $CT_3$}
KL => {bstar, pstar, nstar}

ISS04 regr. proc.: ISS04 dataset, clusters instead of ISS04 subgroups, increasing no. of clusters

| Case | Dataset | rmse | a0 | aa | ar | ap | an | ab | ai |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ISS04 | 0.2329 | 0.135 | 2.28 | 0.64 | -0.61 | 0.54 | 0.84 | 0.41 |
| 2 | ISS04dataset_c1 | 0.1083 | 0.026 | 2.02 | 1.13 | -0.57 | 0.55 | 0.98 | -0.25 |
| 3 | ISS04dataset_c2 | 0.2380 | 0.039 | 1.98 | 1.13 | -0.62 | 0.58 | 0.94 | -0.24 |
| 4 | ISS04dataset_c3 | 0.2647 | 0.125 | 2.29 | 0.50 | -0.60 | 0.49 | 0.80 | -0.05 |
| 5 | ISS04dataset_c4 | 0.2506 | 0.138 | 2.33 | 0.42 | -0.59 | 0.44 | 0.85 | -0.14 |
| 6 | ISS04dataset_c5 | 0.2354 | 0.156 | 2.47 | 0.59 | -0.66 | 0.60 | 0.89 | 0.25 |
| 7 | ISS04dataset_c6 | 0.2388 | 0.145 | 2.37 | 0.60 | -0.63 | 0.54 | 0.89 | 0.26 |
| 8 | ISS04dataset_c7 | 0.2372 | 0.140 | 2.33 | 0.59 | -0.61 | 0.52 | 0.87 | 0.25 |
| 9 | ISS04dataset_c8 | 0.2352 | 0.132 | 2.27 | 0.64 | -0.62 | 0.54 | 0.84 | 0.39 |
| 10 | ISS04dataset_c9 | 0.2352 | 0.132 | 2.26 | 0.64 | -0.61 | 0.53 | 0.84 | 0.38 |
| 11 | ISS04dataset_c10 | 0.2352 | 0.130 | 2.25 | 0.65 | -0.61 | 0.52 | 0.84 | 0.39 |
| 12 | ISS04dataset_c11 | 0.2342 | 0.134 | 2.27 | 0.63 | -0.61 | 0.54 | 0.83 | 0.40 |
| 13 | ISS04dataset_c12 | 0.2337 | 0.134 | 2.27 | 0.64 | -0.61 | 0.55 | 0.82 | 0.42 |
| 14 | ISS04dataset_c14 | 0.2340 | 0.130 | 2.26 | 0.65 | -0.61 | 0.54 | 0.82 | 0.42 |
| 15 | ISS04dataset_c16 | 0.2326 | 0.127 | 2.22 | 0.65 | -0.59 | 0.52 | 0.82 | 0.40 |
| 16 | ISS04dataset_c18 | 0.2308 | 0.126 | 2.22 | 0.66 | -0.60 | 0.52 | 0.82 | 0.41 |
| 17 | ISS04dataset_c20 | 0.2323 | 0.130 | 2.24 | 0.64 | -0.60 | 0.53 | 0.82 | 0.39 |
| 18 | ISS04dataset_c8_vc | 0.2372 | 0.122 | 2.25 | 0.64 | -0.65 | 0.56 | 0.81 | 0.37 |

ISS04 grouping
…
…
…
…

different cluster numbers
…
…
…

Saturation

partly manual clustering

Enigneering ISS04 variables

1) No clustering: differences to ISS04.
2) “Saturated” regression resembles ISS04 fairly.