Confinement scaling with machine size in the updated ITPA global H-mode energy confinement database

J. Hall, P. Zhang and G. Verdoolaege

Department of Applied Physics, Ghent University, 9000 Ghent, Belgium

Introduction

Based purely on experimental data, energy confinement scalings constitute an important instrument for assessing confinement performance in fusion devices. The International Tokamak Physics Activity (ITPA) global H-mode energy confinement database, which in 1998 resulted in the widely used IPB98(y,2) scaling [1], has since then been substantially expanded. In the most recent update DB5.2.3, new data was included from devices with fully metallic walls: JET with the ITER-like wall (JETILW) and ASDEX Upgrade with the full tungsten wall (AUGW). By means of regression analysis on the 'standard set' of the database, the ITPA20 scaling was obtained for ELMy H-mode plasmas, both in engineering and dimensionless variables [2]. The confinement time prediction of ITPA20 for a standard ITER operational point is 3.07 ± 0.46 s, about 15% lower than the estimate based on IPB98(y,2) (3.62 s). An 'ITER-like' scaling ITPA20-IL using a restricted subset was also formulated.

Among the most striking differences between the ITPA20 and IPB98 scalings is a significant reduction of the dependence of confinement on major radius R_{geo}, from an exponent $\alpha_R = 1.97$ in IPB98(y,2) to $\alpha_R = 1.71 \pm 0.32$ in ITPA20, and even $\alpha_R = 1.19 \pm 0.27$ in ITPA20-IL. In this paper, we show that this reduction is part of a trend that started well before the latest version of the database, with data from multiple devices contributing. In addition, we investigate the role of multicollinearity between predictor variables in reducing α_R and we identify a subset of the data that plays the most influential role in weakening the machine size dependence of confinement, compared to the IPB98 scaling.

Data sets

To demonstrate the evolution of the size dependence of the global energy confinement time, throughout this paper three versions of the ITPA global H-mode energy confinement database are compared: the 1998 DB2.8 [1], the 2004 DB4.5 [3] and finally the most recent iteration, the 2020 DB5.2.3 [2]. Only ELMy H-mode plasmas are considered occurring in the respective standard sets. To enhance the clarity of our findings, spherical machines were excluded from the present analysis, at the expense of reducing the range of inverse aspect ratio ε.

Effects of multicollinearity on the regression

Assuming a conventional power-law dependence of the confinement time on the engineering variables, we investigated the impact on the reduced size dependence of multicollinearity: the correlation (linear, on a logarithmic scale) between multiple predictor variables. In particular, multicollinearity can inflate the variance of the regression coefficients and render their estimates unstable against small changes in the data. Standard Pearson correlation only measures pairwise dependence, therefore we employed condition indices and variance decomposition proportions [4].

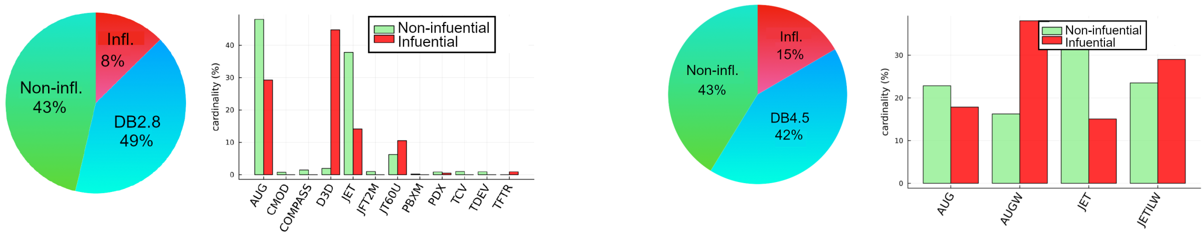
Judging from the condition indices, there are two directions of high multicollinearity across all datasets (DB2.8, DB4.5 and DB5.2.3). By calculating the variance decomposition proportions, it is seen that this is mainly due to correlations involving R_{geo}, plasma current I_{p} and inverse aspect ratio ε, although to some extent also the intercept, magnetic field B_{t} and density $\bar{n}_{\mathrm{e}}$. However, while multicollinearity is likely to be influential in variations of the scaling coefficients seen across the various versions of the database, it is probably not the only factor involved in reducing the size dependence, since it affects all versions to a similar extent.

Subset influencing size dependence

We next aimed at finding a subset of the data that has the most significant influence on the reduction of the size dependence. Starting from DB2.8, several optimization techniques were used to find a minimal subset in the new data added when establishing DB4.5 and later DB5.2.3, which would mainly cause the reduction. Thus, for the transition from DB2.8 to DB4.5, we focused on the points introduced in DB4.5 that were not present in DB2.8 (as DB2.8 is essentially a subset of DB4.5). Within the set of newly added points, the minimal subset with the largest influence on α_R was then identified using the optimisation techniques. This analysis was repeated for the transition from DB4.5 to DB5.2.3, eventually resulting in two distinct, non-overlapping sets of influential points from each database update considered. The subset consisting of the other, less influential points is termed 'non-influential', for simplicity. Fig. 1(a) shows the relative size of the subsets, as well

Table 1: *Main parameter estimates and ITER predictions using various database subsets (size n_{obs}).*

Data	n_{obs}	α_I	α_n	α_R	α_ε
DB2.8	1,244	0.77	0.44	2.2	0.58
+ infl.	219	0.99	0.17	1.4	0.28
+ non-infl.	1,096	0.68	0.54	2.3	0.79
DB4.5	2,540	0.84	0.38	1.9	0.60
+ infl.	896	1.2	0.062	1.1	-0.056
+ non-infl.	2,578	0.91	0.32	1.8	0.53



(a) (b)

Figure 1: Relative size of the original database (DB2.8 in (a), DB4.5 in (b)) and the influential and non-influential subsets among the newly added points (from DB4.5 in (a), DB5.2.3 in (b)), as well as device contributions.

as the contribution of each device to the influential and non-influential subsets in DB4.5, newly added to DB2.8. Fig. 1(b) is similar for the transition from DB4.5 to DB5.2.3. Note that multiple devices contribute significantly to the set of points that are influential in decreasing α_R.

Table 1 lists the parameter estimates obtained by least squares regression in the original DB2.8, as well as DB2.8 either supplemented with the influential points from DB4.5, or with the non-influential points. In accordance with our observations regarding multicollinearity, a number of other dependencies also change across the various subsets, the most salient ones included in the table as well.[1] One clearly sees a reduction of α_R from DB2.8 to DB4.5, as well as a strong reducing effect of the influential points.

Classification and interpretation

Having found a subset of points that are instrumental in reducing the size dependence, the next question is what these points have in common, either from the physical or statistical point of view. To that end, classification techniques were used to assign all points in the latest iteration of the database, DB5.2.3, either to a cluster of influential or non-influential points. This also allowed identifying a set of database variables providing the best separation of both clusters. All variables commonly included in both the engineering and dimensionless form of the confinement scaling were considered, as well as several categorical variables, such as divertor design, wall material and auxiliary heating mix.

Figure 2: *DB5.2.3 in dimensionless space with influential and non-influential subsets.*

[1]The estimates in DB2.8 are somewhat different to those in IPB98(y,2) due to different selection criteria and restrictions imposed on the scaling, but this does not affect the main message.