Available online at www.sciencedirect.com
ScienceDirect
Physics Procedia 62 (2015) 113 – 117
Physics Procedia

3rd International Conference Frontiers in Diagnostic Technologies, ICFDT3 2013

A Statistical Analysis of the Scaling Laws for the Confinement Time distinguishing between Core and Edge

E.Peluso[1], M.Gelfusa[1], A.Murari[2], I.Lupelli[3,1] and P.Gaudio[1]

[1] Associazione EURATOM-ENEA - University of Rome "Tor Vergata" , Roma, Italy
[2] Associazione EURATOM-ENEA per la Fusione, Consorzio RFX, 4-35127 Padova, Italy
[3] EURATOM/CCFE Fusion Association, Culham Science Centre, Abingdon, Oxon OX14 3DB, United Kingdom

Abstract

The H mode of confinement in Tokamaks is characterized by a thin region of high gradients, located at the edge of the plasma and called the Edge Transport Barrier. Even if various theoretical models have been proposed for the interpretation of the edge physics, the main empirical scaling laws of the plasma confinement time are expressed in terms of global plasma parameters and they do not discriminate between the edge and core regions. Moreover all the scaling laws are assumed to be power law monomials. In the present paper, a new methodology is proposed to investigate the validity of both assumptions. The approach is based on Symbolic Regression via Genetic Programming and allows first the extraction of the most statistically reliable models from the available experimental data in the ITPA database. Non linear fitting is then applied to the mathematical expressions found by Symbolic regression. The obtained scaling laws are compared with the traditional scalings in power law form.

Peer-review under responsibility of the ENEA Fusion Technical Unit

Keywords: H mode scaling; symbolic regression; edge and core confinement

1. Energy confinement time in H mode plasmas

The extrapolation of the energy confinement time to the next generation of devices remains a research subject of high interest, both theoretical and experimental, in the Tokamak community. Various scaling laws have been proposed using dimensional and dimensionless quantities. They typically use global quantities, without discriminating between core and edge, and are all based on the assumption that the scalings are in power law form. In the present paper, a new methodology is described which has the potential to assess the validity of both assumptions. The proposed approach allows extracting the scaling laws for the energy confinement in Tokamaks directly from experimental databases, without any previous assumption about the mathematical form of the scalings. The technique is based on genetic programming and nonlinear regression. It has been applied to the ITPA database

Peer-review under responsibility of the ENEA Fusion Technical Unit
doi:10.1016/j.phpro.2015.02.020

of H-mode discharges of both edge and core data and the results have been validated with a series of established statistical tools.

The typical scaling laws of the confinement time in Tokamaks are expressed in terms of global plasma parameters. This approach is based on the implicit assumption that the confinement in the core and in the edge scale in the same way and that the two regions do not have to be discriminated and analyzed separately. This hypothesis should be confirmed since it could have a significant impact on the reliability of the extrapolations to ITER. In its turn, the power law assumption on the form of these scaling equations is not justified neither theoretically nor experimentally. Indeed theoretically there is no justification to limit the analysis to scalings obeying power law expressions. Experimentally, the vast majority of the used databases do not present statistical distributions compatible with power laws [1,2]. Moreover, it must be considered that, in general, the formulation of the scaling laws as power laws can be unsatisfactory for several reasons. In power law scaling laws, there is no saturation of the effects, even when the independent variables grow to infinity or decay to zero. These scalings are also monotonic and tend to overestimate the relevance of the variables with the longest tails. The interaction of all the variables is also assumed to be multiplicative, which tends to result in non-integer exponents of the independent variables not always easy to reconcile with theory. In the case of Tokamak physics, the inadequacies of the power law scaling have clearly been shown in [2].

With regard to the organization of the paper, in the next section 2, the main statistical elements of the proposed methodology are introduced. The following section 3 is devoted to presentation of the results derived for the latest version of the ITPA database. Conclusions and lines of future investigations are the subject of the last section 4 of the paper.

2. The statistical analysis method based on symbolic regression

As mentioned in the introduction, this paper describes the application of advanced techniques of symbolic regression (SR) via genetic programming (GP) to the problem of deriving scaling laws for the confinement time from large databases. The main advantage of the proposed approach consists of practically eliminating any assumption on the form of the scaling laws. The methods developed indeed allow identifying the most appropriate mathematical form for the regression equations. They also permit to demonstrate that the obtained expressions have the potential to better interpret the present experimental data for the confinement time in comparison with power laws (PLs).

As mentioned before, the main objective of SR is to identify for a given finite dataset, a class of appropriate model structures to describe the system under study without "a priori" hypotheses. Solutions of varying levels of complexity can be generated and evaluated to obtain the best trade-off between accuracy and computational complexity. In this study, SR analysis has been conducted by using a GP approach. GP is a systematic, domain-independent method that merely creates and searches for the best individual computer program (CP), i.e. a mathematical expression for the regression of the database in this particular application, among several possible randomly generated ones. The first step in a GP is the generation of the initial population of CPs (formulas in our case) and then the algorithm finds out how well an element of the population works on the basis of some appropriate metrics. This comparison is quantified by a numeric value called fitness function (FF). Fitness can be measured in many ways. To derive the results presented in this paper, the AIC criterion has been adopted [3] for the FF.

In the second phase of GP, as with most evolutionary algorithms, genetic operators (Reproduction, Crossover and Mutation) are applied to individuals that are probabilistically selected on the basis of the FF, in order to generate the new population. That is, better individuals are more likely to have more descendants than inferior individuals. When a stable and acceptable solution, in terms of complexity, is found or some other stopping condition is reached, the algorithm provides the solution with best performance in terms of the FF. In this work, CPs are composed of functions and terminal nodes and can be represented as a combination of syntax trees. The function nodes can be standard arithmetic operations and/or any mathematical functions, squashing terms as well as user-defined operators [2]. The function nodes, included in the symbolic regression used to derive the results presented in this paper, are reported in Table 1; x_i and x_j are the generic independent variables. Furthermore the logistic function is defined as $F = a \cdot (1 + m \cdot e^{-b \cdot x_i})/(1 + n \cdot e^{-b \cdot x_i})$.

Table 1: Types of function nodes included in the analysis performed in this paper.

Function class	List
Arithmetic	c (real and integer constants),+,-,*,/
Operations	power(x_i, x_j), power(x_i, c), power($f(x_i)$,x_j)
Squashing	logistic(x_i)$\vert_{a=1;m=0;n=1;b\neq 0}$ logistic($f(x_i)$)$\vert_{a=1;m=0;n=1;b\neq 0}$

The initial population of CP is formed by stochastic combining functions and terminals nodes and then it is evolved stochastically, generation by generation, to be converted into new, better populations. As mentioned before, evolution is achieved by using genetic operations such as reproduction, crossover and mutation. Reproduction involves selecting the program from the current population and allowing it to survive by copying it into the new population. The crossover operation involves choosing nodes in two parent trees and swapping the respective branches thus creating two new offsprings. A mutation operation consists of selecting a random node from the parent tree and substituting it with a newly generated random tree within the terminals and functions available.

However, the best result provided by the algorithm is not necessarily the best solution to the given problem. A more solid output can be found computing the Pareto Frontier (PF) using the number of nodes and the FF of the pool of individuals formed by the first, second and third best ranked models of each iteration during the run. In more detail, the PF is a reduced collection of models where each element represents the best model (the best FF) for the subgroup of individuals having the same number of nodes. In order to select more accurately and provide further information, these models are then classified using a Bayesian criterion. Indeed, while the AIC criterion is used for the FF, the Bayesian Information Criterion (BIC) [3] is used for the PF to increase the generalization capability of the model selection.

3. Results for the ITPA Database

The ITPA database contains only 199 usable entries for a comparison between the confinement in the core and at the edge. Using this data, the best models identified with the symbolic regression are:

$$W_{core}^{[SR]}[\text{Mj}] = 0.0106_{0.0098}^{0.0113}\frac{I^{1.126_{1.027}^{1.225}}R^{1.574_{1.198}^{1.950}}}{\epsilon^{2.588_{2.280}^{2.895}}} + 0.045_{0.030}^{0.061}P + 0.141_{0.065}^{0.218}M - 0.263_{0.161}^{0.364}R \quad (1)$$

$$W_{pedestal}^{[SR]}[\text{Mj}] = 0.022_{0.020}^{0.024}I\cdot P\cdot M + 0.289_{0.220}^{0.357}I + 0.067_{0.054}^{0.079}P\cdot R - 0.347_{0.282}^{0.413}\frac{P}{F_q^{1.273_{1.231}^{1.315}}} \quad (2)$$

These scaling laws are to be compared with the reference ones reported in the literature [4]:

$$W_{core}^{[4]}[\text{Mj}] = 0.103\cdot I^{0.88}B^{0.11}P^{0.25}n^{0.49}M^{0.23}\cdot R^{2.02}\epsilon^{1.22}\kappa_a^{0.24} \quad (3)$$

$$W_{pedestal}^{[4]}[\text{Mj}] = 6.43\cdot 10^{-4}I^{1.58}B^{0.06}P^{0.42}n^{-0.08}\cdot M^{0.2}R^{1.08}\epsilon^{-2.13}\kappa_a^{1.81}F_q^{2.09} \quad (4)$$

The statistical quality of these scaling laws has been investigated using the following statistical indicators: MSE, AIC, BIC and Kullback-Leibler divergence (KL) [3]. The values of these indicators, for the four different scaling laws previously reported, are shown in Table 2.

Table 2. Comparison of the statistical indicators for the power law and non power scaling law reported above.

	MSE	k	AIC	BIC	KLD
(1)	0.089	11	-457.41	-421.68	0.0217
(2)	0.058	13	-539.80	-496.05	0.0240
(3)	0.145	10	-363.67	-330.54	0.0276
(4)	0.111	11	-414.58	-423.59	0.1575

Table 3. Extrapolations to ITER for the power law and non power scaling law.

(1)	$76.177_{38.951}^{146.762}$ Mj
(2)	$80.31_{71.784}^{88.751}$ Mj
(3)	171.53 Mj
(4)	167.05 Mj

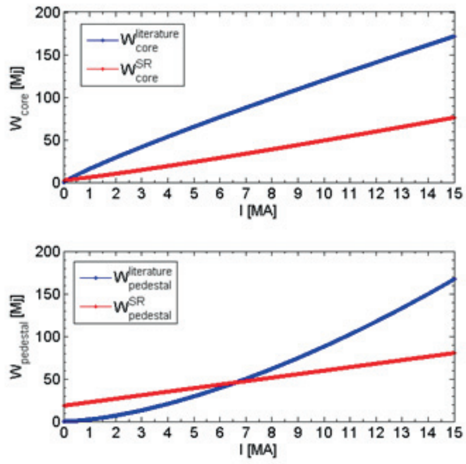

Fig. 1. Trends of the energy content versus plasma current for an ITER class plasma $P = 87[\mathrm{MW}], n_{core} = 10.3 \cdot [10^{19} m^{-3}]$; $n_{pedestal} = 8.24 \cdot [10^{19} m^{-3}]$; $B = 5.3[\mathrm{T}]; R = 6.2[\mathrm{m}]$; $\epsilon \simeq 0.32$; $M = 2$; $F_q = 1.54; \kappa_a = 1.7$. The blue lines are the trends of the relation reported in literature [4].

This shows that, at least in a statistical sense, the non power law scalings manage to interpret much better the available database than the power laws. The difference between the two types of scalings are also quite relevant when extrapolated to ITER. Indeed the values of the internal energy for a typical ITER discharge are quite different; the obtained numerical values for W are reported in Table 3. The distribution of the energy between core and edge is very similar for the two types of scalings (basically 50% in each region). On the other hand, the non power law scalings predict a much lower energy content in ITER plasma. The overall energy content predicted by the non power laws is indeed about 50% of that of the power laws. The main reason for this difference can be derived by the comparison of the overall trend of the two types of scaling laws with the various parameters. The trends versus the plasma current are reported in Figure 1 for the core and the edge regions. As can be seen by the figures, the power laws tend to show a much more steep increase of the plasma energy content with the plasma current (and also with the other main parameters). The non power law scalings, on the contrary, predict a much slower increase in the plasma energy content with the various engineering parameters so, in a certain sense, they represent a more conservative and cautious prediction.

4. Conclusions

The proposed technique of symbolic regression via genetic algorithms provides a versatile exploration methodology to extract mathematical equations directly from large databases. The application of this approach to the scaling laws of the energy content of Tokamaks has allowed investigating two main assumptions: the similar scaling of the energy and the fact that the scalings are power laws. The first assumption is confirmed by the analysis of an ITPA database. The energy seems to be more or less equally distributed between the core and the edge in both present day devices and in ITER. It is therefore reasonable to adopt global scaling laws which do not distinguish between these two plasma regions. On the other hand, the assumption of power law scaling is clearly falsified by the approach. The best scaling law, in a statistical sense, are certainly non power laws and predict a much smaller energy content in ITER. A few words of caution are in any case in place. It must be remembered that the available database is very limited and therefore of very poor statistical relevance. Moreover, the quality of its content is also far from ideal. The collinearities are very high and the region represented very limited. The obtained results do not therefore have to be interpreted as definitive scaling laws for the extrapolations to ITER. The main message of this study is that, whereas the distinction between core and edge does not seems a very important aspect, on the contrary the assumption of power law scaling is not justified neither theoretically nor experimentally. Therefore relaxing this hypothesis and investigating the laws of different mathematical form are to be strongly recommended.

References

[1] A. Murari et al 2013 Nucl. Fusion 53 043001

[2] A. Murari et al 2012 Nucl. Fusion 52 063016

[3] Kenneth P. Burnham, David R. Anderson (2002), Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer. (2nd ed)

[4] McDonald D.C. at al, Nuclear Fusion 47(2007).