Symbolic Regression Analysis on the EUROfusion JET-ILW Pedestal Database

F. Napoli[1], L. Frassinetti[2], S. Saarelma[3] and JET Contributors*

[1] *ENEA, Fusion and Nuclear Safety Department, C. R. Frascati, Via E. Fermi 45, 00044 Frascati (Roma), Italy*

[2] *Division of Fusion Plasma Physics, KTH Royal Institute of Technology, Stockholm SE*

[3] *CCFE, Culham Science Centre, Abingdon, Oxon OX14 3DB, UK*

**See the author list of E. Joffrin et al. accepted for publication in Nuclear Fusion Special issue 2019, https://doi.org/10.1088/1741-4326/ab2276*

Abstract

Our present understanding of the pedestal physics is still limited and its extrapolation to future JET-ILW experimental scenarios is a challenging task. To face this issue a new multi-objective genetic programming (GP) code has been implemented. This code is able to perform a multi-objective symbolic regression analysis on the EUROfusion JET-ILW pedestal database [1]. In order to improve our understanding of the obtained analytical expressions and to make easier the comparison with established results in power law form, we drive the GP search towards a class of scaling laws called *generalized power laws*. Possible new analytical nonlinear regression models for the pedestal thermal stored energy has been found. These results suggest that some of the new scaling laws might capture interesting experimental features that otherwise it would not be possible to obtain with ordinary power laws.

Introduction

The EPED model [2] can predict the H-mode JET-ILW pedestal within a relative error of about 20% when the pedestal is close to the Peeling-Ballooning (PB) boundary. However, when the pedestal is far from the PB boundary, our present understanding of the pedestal physics is still lacking [3,4]. Moreover, the extrapolation to future JET-ILW experimental scenarios and to different tokamaks is even more challenging.

In order to improve our modelling capabilities of the JET-ILW pedestal, a new multi-objective genetic programming (GP) [5] code has been implemented. GP is an evolutionary search algorithm inspired by natural evolution which encodes a solution to a problem as a computer program. When the instruction set used to build new solutions is made of mathematical functions only, GP becomes an instance of symbolic regression (SR) [5,6]. SR is a non-parametric nonlinear regression technique where, not only the model's parameters are fitted to the data, but also the regression model is selected in a data-driven way. Symbolic regression via GP (GPSR) is a powerful technique particularly suited to

discover nonlinear models with interaction, saturation and threshold effects, i.e. just the class of models we expect to be useful to describe the pedestal scaling.

In nuclear fusion it is very common the use of scaling laws in power law form. Power laws has been used successfully for the pedestal stored energy [1,7] even if there is no theoretical reason to assume *a priori* this kind of scaling. Reasons of their popularity are their extreme simplicity, their scale invariance and the availability of a simple fitting tool like log-regression. However, power laws present important limitations [6]. They are monotonic functions of their explanatory variables, thus they cannot describe common phenomena in plasma physics such as saturation and threshold effects. Moreover, the interaction among physical quantities is only multiplicative. Nevertheless, the use of power laws is well established and it is not reasonable to derive new scaling laws without a clear comparison with them. Thus, in order to improve our understanding of analytical expressions obtained by GPSR and to make easier the comparison with already established results in power law form, a novel SR methodology has been derived here. By this method we constrain the GP search in finding solutions in the form of *generalized power laws*, where exponents can be functions of explanatory variables.

Methods

In this work we look for meaningful and nontrivial data-driven analytical regression models for the thermal stored energy calculated using profiles evaluated at normalized poloidal flux coordinate $\psi_N = 0.9$, hereafter called W_{90}, as defined in the EUROfusion JET-ILW pedestal database [1]. Accordingly, we selected a subset of parameters of the pedestal database typically used in W_{90} scalings: plasma current (I_p), NBI power (P_{NBI}), toroidal magnetic field (B), electron line-averaged density (n_e), average triangularity (δ), effective mass (M_{eff}). We also included the gas flow rate of main species (Γ) as it is known that the gas fuelling affects the pedestal height. We log-transformat the dataset and then split it in two: training set (594 records) and validation set (394 records).

The multi-objective GP (MOGP) code here developed performs a multi-objective SR analysis on the dataset, finding regression models for W_{90}. Different analytical expressions and different sets of input decision variables are tested during the GPSR run. The multi-objective search in the model space is performed minimizing two objectives: the model complexity (the solution length) and the fraction of variance unexplained by the model ($1-R^2$). These are two conflicting objectives and a trade-off among them must be found. The Pareto dominance criterion [8] is used as a model selection criterion during the GPSR run: a

model is said to Pareto dominate another solution in the GP population if it is at least equal in both objectives to that solution and better in at least one objective. Thus, the MOGP code looks for a final set of non-dominated solutions, i.e. the Pareto set of the objective space. The minimization of the model complexity can be useful to avoid overfitting and can be considered as an implementation of the Occam's razor principle. Moreover, all evolved models are tested on a validation set to measure their generalization capabilities.

A combined search strategy is performed, implementing the concept known as Baldwin effect [9]. A *global search* is made by GP in the model space looking for the best analytical shape of the regression model and, when the selected model must be evaluated, a *local search* is performed in the model's parameter space: the model is fully parametrized and the numerical values of its parameters are determined by standard nonlinear least square regression methods.

Among all possible models that can be derived by GPSR, we limited the GP search space to a general class of power laws, called *generalized power laws,* where exponents can be functions of explanatory variables. In a first stage of the SR analysis, we limit the function set to linear functions, allowing the derivation of simpler expressions. In a second stage, interesting solutions from the previous stage are used to seed a new GPSR run and the function set is extended to nonlinear functions. In this way, even if we introduce nonlinear functions and allow more complex interactions among physical quantities, the model complexity is taken under control avoiding overly complex models since the very beginning of the GP search and keeping our solutions in power law form.

Results and Conclusions

In the first stage of SR analysis, the function set is limited to simple arithmetic operators {+,-,*}. In Figure 1 it is reported the final Pareto set (red points) computed aggregating solutions from 4 GPSR runs. We also consider as a reference solution that one obtained by a GPSR run with the same GP settings but with a simpler function set {+,-}. This solution encodes the standard power law that can be obtained by an ordinary least square method when all the explanatory variable are considered. To select a final pool of solutions from the Pareto set, we consider only those solutions that Pareto dominate the reference solution (green points) and one of them is selected (Figure 1). This solution is interesting because the exponent of triangularity is a decreasing function with plasma current while W_{90} is still an increasing function with plasma current due to the strong correlation among n_e and I_p. This scaling seems to be consistent with the experimental JET-ILW results where, so far, the high

triangularity has led to a pedestal pressure improvement mainly at low I_p [10]. However, if we look at the residual plot of this regression model, we discover that this fit is less reliable when the plasma current is greater than 2.6 MA because there are few records in the database with high plasma currents. Nonetheless, the accuracy of the fit is good ($R^2 \approx 0.85$) and it is comparable with that one from a standard power law fit [1]. In the second stage of this SR analysis, we use the previous selected solution to seed a new SRGP run with a richer function set {+,-,*,Exp,Sigmoid} and a weighted sum of squared residuals as fitness function, with higher weights for higher plasma currents. We obtained a final Pareto set made of solutions that are variations of the initial solution and more accurate for $I_p \geq 2.6$ MA. As an example, we examine here an interesting solution ($R^2 \approx 0.86$) where the isotopic mass has an effect on W_{90} (Figure 2). If we remove the sigmoid function present in this solution, the fit is completely lost, demonstrating it is relevance. Moreover, if we consider W_{90} as a function of M_{eff} only (using the correlation matrix), we get a figure qualitatively similar to measurements [11]. This type of dependence cannot be obtained with a standard power law (Figure 2).

In conclusion, this SR analysis suggests that in the future it might be possible to find accurate and interpretative models for the pedestal stored energy using a small function set and simple model selection criteria and, at the same time, to take under control the model complexity.

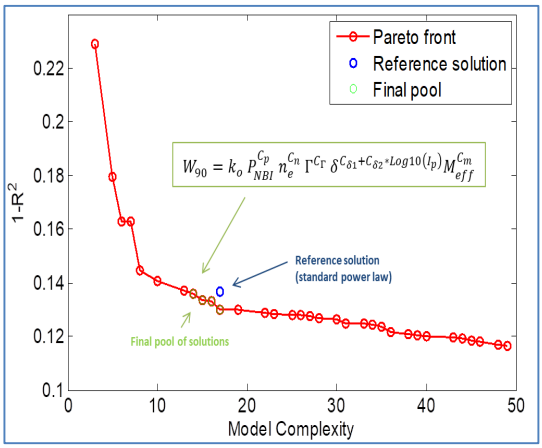

Figure 1 – Pareto front computed over 4 GPSR runs

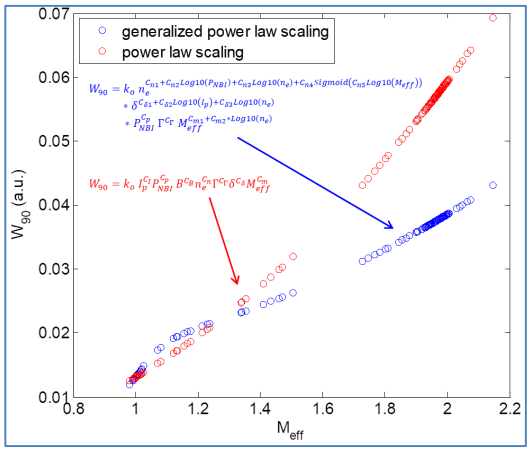

Figure 2 – Isotopic mass effect on W_{90} scaling

Acknowledgement. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

[1] L. Frassinetti, et al., 45th EPS Conf. on Plasma Physics, Prague (2018)
[2] P.B. Snyder, et al., Nuclear Fusion 51, 103016 (2011)
[3] L. Frassinetti et al., Nucl. Fusion 59, 076038 (2019)
[4] S. Saarelma, et al., 60th Annual Meeting of the APS Division of Plasma Physics, Portland, Oregon, USA
[5] E. Zitzler et al., EMO 2001, March 7-9 (2001), Zurich; J. R. Koza, GP, MIT Press, Cambridge (1992)
[6] A. Murari, et al., Nucl. Fusion 53, 043001 (2013); A. Murari, et al., Nucl. Fusion 57, 126017 (2017)
[7] J. C. Cordey, et al., Nucl. Fusion 43, 670 (2003)
[8] K. Deb, Multi-Objective Oprtimization Using Evoltionary Algorithms, Wiley (2002)
[9] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press (1998)
[10] Beurskens, et al., Nucl. Fusion (2013); Giroud, et al., Nucl. Fusion (2013)
[11] D. B. King, et al., 44th EPS Conf. on Plasma Physics, Belfast (2017)