EUROPEAN FUSION DEVELOPMENT AGREEMENT

JET

EFDA-JET-CP(05)02-10

A.J.Meakins, D.C.McDonald, Y.Andrew, J.G.Cordey, M.Coppins,
ITPA H-mode Threshold Database WG and JET EFDA contributors

# Applying Advanced Statistical Techniques to Tokamak L-H Threshold Data

# Applying Advanced Statistical Techniques to Tokamak L-H Threshold Data

A.J. Meakins[1], D.C. McDonald[2], Y. Andrew[2], J.G. Cordey[2], M. Coppins[1],
ITPA H-mode Threshold Database WG and JET EFDA contributors*

*[1]Imperial College, Prince Consort Road, London, SW7 2BZ, UK*
*[2]EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, OX14 3DB, UK*
** See annex of J. Pamela et al, "Overview of JET Results ",*
*(Proc.20 th IAEA Fusion Energy Conference, Vilamoura, Portugal (2004).*

Preprint of Paper to be submitted for publication in Proceedings of the
EPS Conference,
(Tarragona, Spain 27th June - 1st July 2005)

## INTRODUCTION

The L-H transition power threshold is often predicted using log-linear power laws fitted to experimental data. Here we present two methods aimed at improving such predictions for both future machines such as ITER and current machines such as JET. Extrapolation to future machines is addressed using Errors-in-Variables (EIV) techniques. Interpolative single machine prediction is addressed using Neural Networks (NNs).

## EIV REGRESSION APPLIED TO MULTI-MACHINE THRESHOLD SCALINGS

Log-linear power law scalings produced from the H-mode Threshold database are often fitted using Ordinary Least Squares (OLS) regression. The OLS statistical model contains the assumption that the error on the dependent variable is much larger than the error on any of the independent variables. Applying OLS to a dataset which violates this assumption results in an undesirable biasing of the fit as shown in fig.1.

In this project the measurement errors on the data used to produce the most recent threshold scalings [3] were evaluated to determine if the use of OLS had introduced bias. Through applying an alternate fitting procedure which does not suffer the assumptions of OLS, Errorsin- Variables Orthogonal Regression (EVOR) [1, 2], a set of unbiased scalings were produced. The results of the analysis of the IAEA04R [3] data set are presented here. In fitting the H-mode threshold scalings, the threshold power ($P_{L\text{-}H}$) is the dependent variable. The toroidal magnetic field strength (B), line integrated density (n) and a geometry variable are the independent variables. Plasma surface area (S) or the combination of plasma minor radius (a) and major radius (R) are the commonly used geometry variables. The relative (percentage) errors on each of these variables for every tokamak in the database were assembled (see table 1). These errors were combined in weighted quadrature with the tokamaks weighted by the number of points they contributed to the IAEA04R data set. This produced a single error estimate for each variable in the data set (see row marked IAEA04R in table 1).

These errors reveal that the OLS assumption is only weakly satisfied by the IAEA04R data for both common forms of scaling.With the errors on n and S at 41.5% and 26.0% of the error on $P_{L\text{-}H}$, respectively, a moderate skew in the threshold scalings is to be expected. Similar results were obtained for a number of different data sets derived from the threshold database, suggesting that an alternate fitting model would be more suitable.

The alternate fitting method employed in this analysis, EVOR, is an Errors-in-Variables model which can handle significant errors on all the variables without introducing bias to the fit (fig.1). Its use has been demonstrated with the confinement database [4]. Like OLS, EVOR attempts to minimise the least squared error to fit the data. Unlike OLS, however, the least squared error is calculated along a vector orthogonal to the fitting function. The mathematical procedure for EVOR is only slightly more complicated compared to OLS. Fitting power law scalings with EVOR was performed with the following procedure: 1) Take the Log of the data and normalise each variable with associated relative error, 2) Subtract data means from each variable for the entire dataset (mean centre), 3) Calculate Principal Components (PCs) using Principal Component Analysis, 4) Select smallest PC as it represents the normal to fitting plane through data, 5) Transform back out of log space to produce scaling.

Fitting the IAEA04R data with both OLS and EVOR produced the following scalings:

$$P_{L-H} = 0.077 \quad n^{0.44} \quad B^{0.65} \quad S^{0.80} \quad \text{(OLS)} \tag{1}$$

$$P_{L-H} = 0.075 \quad n^{0.56} \quad B^{0.58} \quad S^{0.85} \quad \text{(EVOR)} \tag{2}$$

The exponents on n and B change between the OLS fit and the EVOR fit with the exponent on n increasing by 27% and the exponent on B decreasing by 11%. The effect, for this scaling, is an increase in the predicted $P_{L\text{-}H}$ for ITER by 13% from 33.7MW to 38.1MW. The new prediction still falls within the ITER design parameters [5]. A similar trend was observed for other data selections from the threshold database.

The average errors used in the EVOR calculation (table 1) contain a degree of uncertainty as they are based on a combination of estimated errors from each tokamak in the database. To determine the sensitivity of the scalings to this uncertainty plots describing how each exponent changes when varying a single error about its estimate were produced. Figs. 2(a)-(d) show the results for the scaling listed above. While varying the errors on a, R, S and B had little effect, the plots demonstrated the importance of producing accurate errors for both n and $P_{L\text{-}H}$ as the exponents were observed to be very sensitive to these errors. The effect the uncertainty in the $P_{L\text{-}H}$ error has on the ITER prediction for the above scaling is shown in fig.2(e).

In conclusion it was found that the use of OLS introduces bias into the threshold scalings which can be corrected for by using an alternate fitting procedure. Refining the errors on $P_{L\text{-}H}$ and n will improve the reliability of the predictions. The effects of correlated errors, which neither EVOR nor OLS treats, still needs to be investigated for the threshold database.

## NEURAL NETWORKS APPLIED TO JET L-H DATA

When planning experiments on JET, log-linear power laws are often used to predict the threshold power required to achieve a L-H transition. The JET L-H threshold data, however, is known to contain regions of operating space that can not adequately be described by power law fits, such as the minima in the threshold at low density or the behaviour of the threshold with X-point height. In this project NNs were applied to the JET L-H data as NNs have a more flexible functional form allowing them to describe more complex trends in the data. NNs consist of a set of interconnected processing nodes where the structure of the NN determines the relationship between the network inputs and outputs. It has been shown that NNs are able to fit any well behaved single-mapping function [6].

Data from dedicated ramp shots performed with the the MkII-GB (Septum) and MkII-GBSRP divertors were assembled with the data sampled in L-mode 10-30ms prior to the transition. The variables collected included $P_{L\text{-}H}$, B, n and X-point height. In total 164 shots were collected for the septum divertor and 106 shots for the SRP divertor. To remove known systematic differences between divertors the data from each divertor was fit separately. Due to the low number of shots the number of variables that could be fitted with was severely limited. The threshold power was fitted as a function of just n and B which required filtering the range of the remaining variables to reduce residual scatter. After filtering, 73 points remained for the septum and 42 points for the SRP.

The fit to the septum data is shown in figure 3. A 5-10% reduction of RMSE compared to power law fits to the septum and SRP data was observed. The small amount of data limited the analysis of the low density behaviour with only a slight turn up in threshold seen in the septum data, a feature the power law did not fit. Similar results were observed for the SRP data.

The small quantity of data and limited range of parameters limited improvement of the NNs over the power law scalings for the JET L-H data. Applying this method to other databases such as the ELM frequency, for which there is a large set of data, should produce better results. With sufficient data the improved predictions provided by NNs should allow for potential savings in experimental time on future tokamaks such as ITER.

## ACKNOWLEDGEMENTS

This work was funded jointly by the UK EPSRC and by the European Communities under the contract of Association between EURATOM and UKAEA. The views and opinions expressed herein do not necessarily reflect those of the European Commission. This work was carried out within the EFDA framework.

## REFERENCES

[1]. Anderson T.W. Annals of Statistics **12** (1984) 1
[2]. Kardaun O.J.W.F. et al. EPS Conf. (1989) 253
[3]. Martin Y.R. et al. IAEA Int. Conf. (2004) IT/P3-35
[4]. Cordey J.G. et al. “The scaling of confinement in ITER with b and collisionality” submitted to Nucl. Fusion
[5]. Aymar R. et al. Plasma Phys. Control. Fusion **44** (2002) 519
[6]. Bishop C. M “Neural Networks for Pattern Recognition” (Oxford University Press, 1995) p128

| Tokamak | $S_R$ | $S_a$ [1] | $S_S$ | $S_B$ | $S_n$ | $S_{PLH}$ [1] |
|---|---|---|---|---|---|---|
| ASDEX | 1.0% | 1.5% | 2.1% | 1.0% | 3.0% | 19.2% |
| AUG | 0.2% | 1.1% | 1.3% | 1.0% | 3.0% | 23.3% |
| CMOD | 0.6% | 2.0% | 2.2% | 1.0% | 5.0% | 15.3% |
| COMPASS | 2.0% | 6.0% | 9.6% | 2.0% | 5.0% | 11.3% |
| D3D | 0.6% | 0.5% | 1.1% | 1.0% | 4.0% | 9.9% |
| JET | 1.0% | 3.0% | 4.9% | 1.0% | 8.0% | 12.2% |
| JFT2M | 0.8% | 3.0% | 7.3% | 1.0% | 2.0% | 10.1% |
| JT60U | 0.5% | 1.0% | 1.8% | 1.0% | 10.0% | 19.3% |
| PBXM | 0.7% | 3.0% | 8.3% | 1.0% | 5.0% | 21.2% |
| TCV | 1.0% | 2.0% | 2.7% | 1.0% | 5.0% | 11.1% |
| MAST | 1.4% | 2.6% | 3.0% [3] | 1.5% | 7.0% | 10.0% [2] |
| NSTX | 2.0% | 3.0% | 6.5% | 1.0% | 6.0% | 6.9% |
| TUMAN-3N | 2.0% | 5.0% | 5.9% | 3.0% | 10.0% | 6.3% |
| IAEA04R | 0.87% | 2.50% | 4.03% | 1.07% | 6.45% | 15.55% |

JG05.319-1c

*Table 1: Errors quoted or calculated for variables used in scalings for each tokamak. [†]Those errors which were calculated. [+]Quoted by MAST team. [*]Low estimate.*

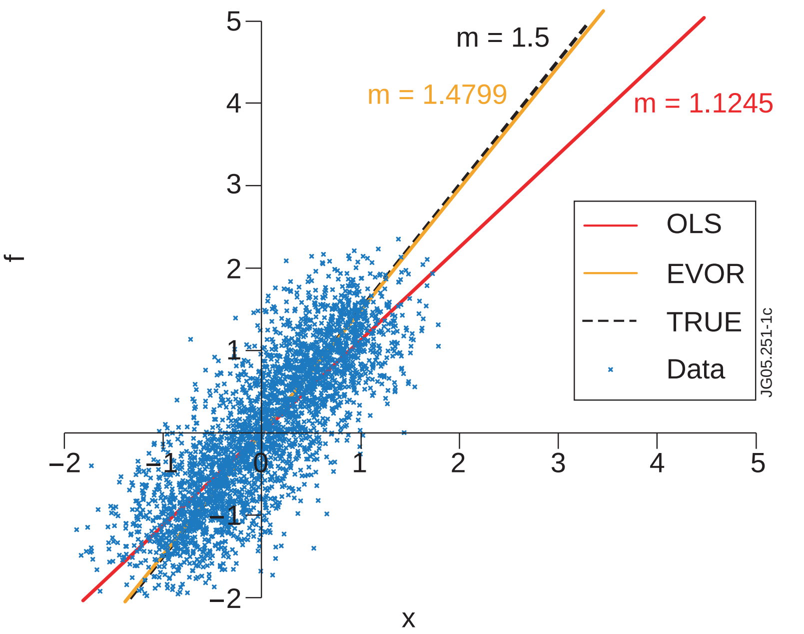



*Figure 1: A Comparison of EVOR and OLS fits to a generated data set. Data set contains 3000 points sampled uniformly over [-1, 1] from the line marked TRUE, given by f = 1.5x. Gaussian errors of s = 1 were added to the data on both the dependent variable ( f ) and the independent variable (x). m is the gradient of each line.*

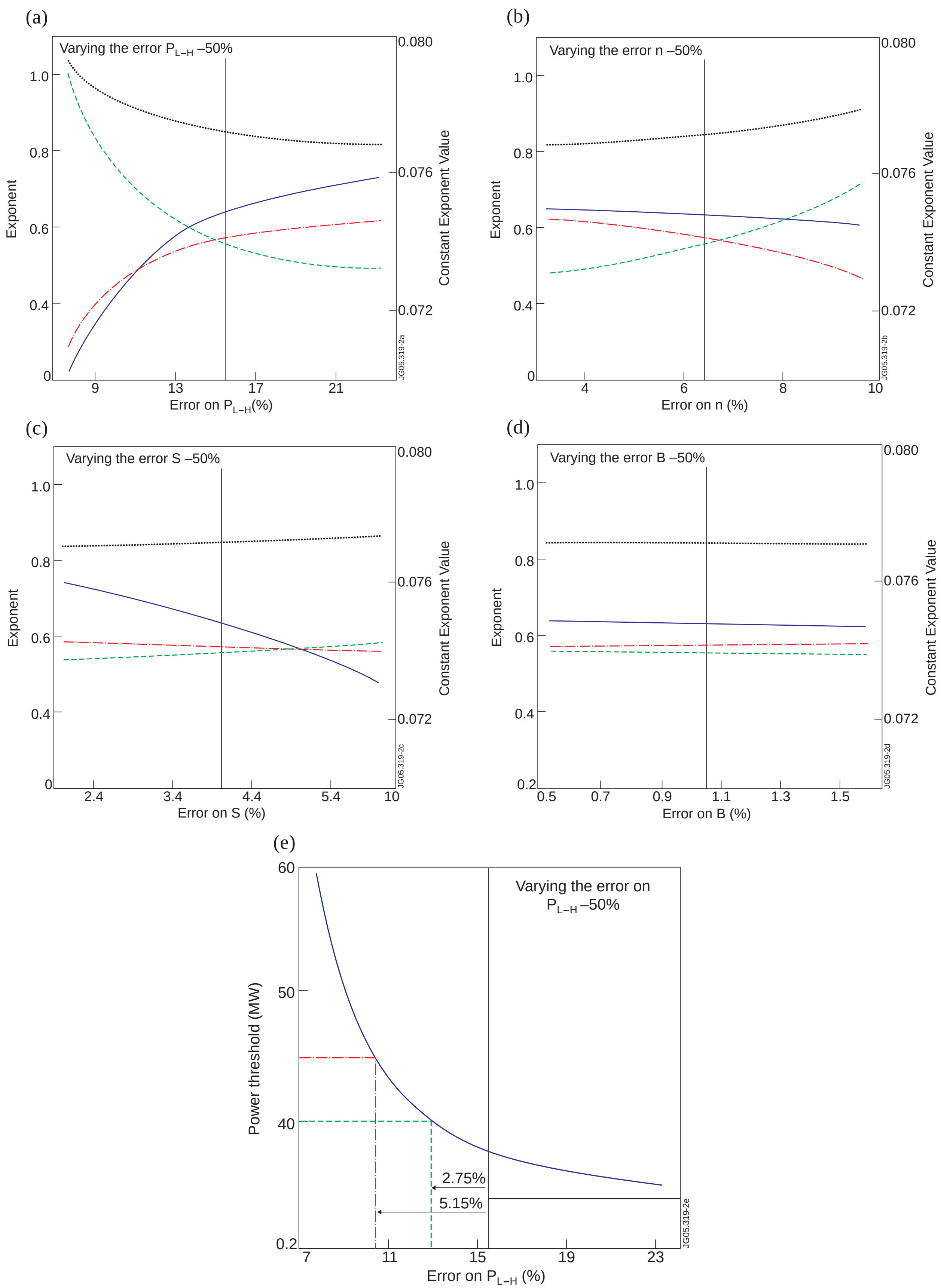
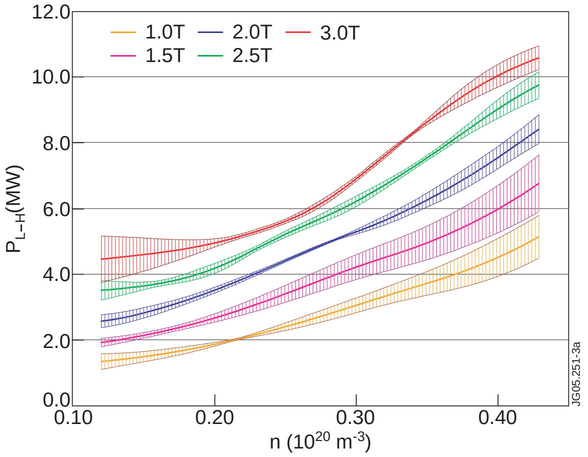



*Figure 2: (a)-(d) Stability of scaling to errors on each variable. Exponents: B (red), n (green), S (orange). Scaling constant is blue (reads off right axis). (e) Effect on ITER estimate of varying error on $P_{L\text{-}H}$.*



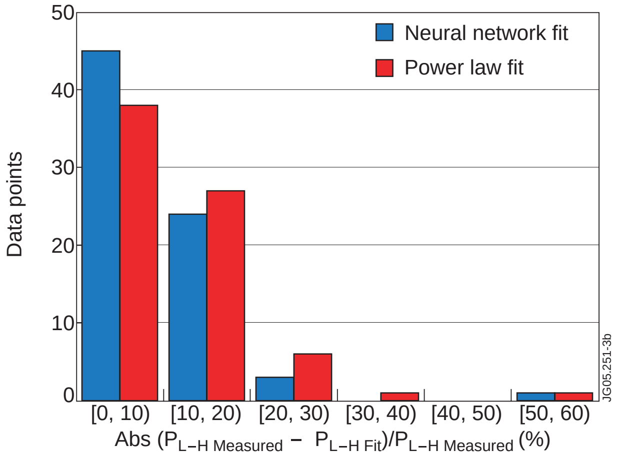




*Figure 3: (a) Results of fit to septum data with $P_{L\text{-}H}$ a function of B and n. Error bands show level of agreement between 5 NNs fitted to the same data - narrow bands denote greater agreement. (b) Histogram of relative error between measured $P_{L\text{-}H}$ and the values estimated from the power law and neural network fits.*