*31st EPS Conference on Plasma Phys. London, 28 June - 2 July 2004 ECA Vol.**28G**, P-5.145 (2004)*

Analysis of the bias in H-mode confinement scaling expressions related to measurement errors in variables

Knud Thomsen[1], J.G. Cordey[2], O.J.W.F. Kardaun[3] and ITPA H-mode Database WG[4].

1. EFDA CSU - Garching, Boltzmannstrasse 2, 85748 Garching, Germany.
2. EURATOM-UKAEA Fusion Association, Culham Science Centre, United Kingdom.
3. Association EURATOM-IPP, MPI für Plasmaphysik, 85748 Garching, Germany.
4. Alcator C-Mod: J.A. Snipes, M. Greenwald; ASDEX/ ASDEX Upgrade: O.J.W.F. Kardaun, F. Ryter, J. Stober; DIII-D: J. C. DeBoo, C. C. Petty; JET/ EFDA: J. G. Cordey, K. Thomsen, D. C. McDonald; JFT-2M/ JT-60U: Y. Miura, K. Shinohara, K Tsuzuki, Y. Kamada, T. Takizuka, H. Urano; MAST/ COMPASS-D/ START: M. Valovic, R. Akers, C. Brickley, A. Sykes, M. J. Walsh; PBX-M/ PDX/ TFTR/ NSTX: S. M. Kaye, C. Bush; TCV: Y. Martin; TdeV: A. Cote, G. Pacher; TEXTOR: J. Ongena; TUMAN-3M: S. Lebedev; T-10: A. Chudnovskiy.

1. Introduction: In the first Nuclear Fusion article on the International Global H-mode Confinement Database [1], the bias in standard Ordinary Least Squares (OLS) regressions related to measurement errors in the regression variables was estimated to be negligible. However, it is now found that, using the available measurement errors as provided by each Tokamak represented in the H-mode database, this bias is actually not quite negligible. Therefore, more sophisticated regression techniques that take the measurement errors into account are used, as in [2]. The first results of using such techniques [2,3,4] on the ELMy H-mode data subsets of version DB3v13 of the ITPA Global H-mode Confinement Database are discussed in this paper. It is shown that the β and ν^* dependences in the energy confinement time scaling expressions are rather sensitive to the errors in the stored energy, δW_{th} and the loss power, δP.

2. Estimate of the bias related to measurement errors: The variables used in the analyses are as follows: W_{th}, P, R, a (minor radius), A (cross section area), n, I, B, M. The reason for choosing A and a, rather than κ (elongation) and ε (inverse aspect ratio), is that with this set the number of cross correlations between the errors are smaller. Table I lists the measurement errors that the tokamaks with ELMy H-mode data in DB3v13 have provided. The average error for each variable is given in the last row. The bias in standard OLS regression related to these errors is estimated as $\sim (\lambda_e / \lambda_{pc})^2$, where λ_{pc} is the standard deviation (STD) of a Principal Component (PC) and λ_e is the STD of the measurement error along that PC. If $\lambda_{pc} > 4\ \lambda_e$ the bias is estimated to be less than 6%, and considered negligible. From PC analysis the estimate of the STD of a PC is an estimate of $\lambda_{pc} + \lambda_e$, if errors are present. The ratio ERR = $\lambda_e / (\lambda_{pc} + \lambda_e)$ can be estimated, if the errors are known or assumed. Hence, if ERR < 0.2 the bias is less than 6%, and is negligible.

	δR	δa	δA	δB	δI	δn	δM	δP	δW_{th}
ASDEX	1.0%	1.5%	2.0%	1.0%	1.0%	3.0%	10.0%	19.1%	14.2%
AUG	0.2%	1.1%	3.0%	1.0%	1.0%	3.0%	10.0%	12.0%	11.5%
CMOD	0.6%	2.0%	3.0%	1.0%	2.0%	5.0%	3.0%	12.7%	17.4%
COMPASS	2.0%	6.0%		2.0%	1.0%	5.0%	10.0%	13.3%	15.0%
D3D	0.6%	0.5%	3.0%	1.0%	1.0%	4.0%	1.0%	11.5%	11.6%
JET	1.0%	3.0%	6.0%	1.0%	1.0%	8.0%	20.0%	15.0%	7.2%
JFT2M	0.8%	3.0%	5.0%	1.0%	1.0%	2.0%	20.0%	12.0%	18.1%
JT60U	0.5%	1.0%	5.0%	1.0%	0.5%	10.0%	5.0%	12.8%	15.0%
PBXM	0.7%	3.0%	10.0%	1.0%	1.0%	5.0%	5.0%	16.3%	13.8%
PDX	0.8%	3.0%	5.0%	1.0%	1.0%	5.0%	5.0%	25.5%	16.2%
TCV	1.0%	2.0%	1.0%	1.0%	1.0%	5.0%	1.0%	10.2%	10.0%
TFTR	0.4%	1.3%	5.0%	2.0%	2.0%	5.0%	10.0%	21.9%	23.7%
TDEV	0.6%	5.0%	5.0%	1.0%	2.0%	2.0%	10.0%	7.0%	10.0%
START	6.9%	9.1%	10.0%	6.0%	2.0%	5.0%	15.0%	17.9%	16.5%
MAST	1.4%	2.6%	2.8%	1.5%	1.0%	7.0%	5.0%	10.0%	15.0%
NSTX	2.0%	3.0%	5.0%	1.0%	2.0%	6.0%	5.0%	10.0%	10.0%
Average	1.3%	2.9%	4.7%	1.5%	1.3%	5.0%	8.4%	14.2%	14.1%

Table I. *Estimates of the measurements errors on each Tokamak with ELMy data in the ITPA Global H-mode Confinement Database version DB3v13. The estimates δP and δW_{th} are based on the errors on the variables used to compute PLTH and WTH in the database.*

The PC's for an extended ELMy H-mode dataset of DB3v13 (SELDB3X III.15) are listed in Table II together with the estimates of $\lambda_e + \lambda_{pc}$ and ERR. It is clear that the bias related to the errors is not negligible for 2 PC's, PC6 and PC8. To avoid the problem with PC6 ~ M, a Deuterium only ELMy dataset (SELDB3X III.17) has been identified (if δM < 3% would

	ln(I)	ln(n)	ln(B)	ln(P)	ln(R)	ln(A)	ln(a)	ln(M)	$\lambda_e + \lambda_{pc}$	ERR
PC1	0.463	-0.026	0.079	0.532	0.217	0.609	0.274	0.049	1.632	0.05
PC2	0.246	0.724	0.358	0.313	-0.218	-0.317	-0.190	0.071	0.664	0.09
PC3	-0.277	-0.386	0.645	0.340	0.368	-0.307	-0.004	-0.112	0.380	0.15
PC4	0.499	-0.084	0.531	-0.665	0.045	0.065	0.092	0.066	0.330	**0.29**
PC5	-0.314	0.548	0.002	-0.243	0.672	0.205	0.102	-0.193	0.176	**0.28**
PC6	-0.207	0.022	0.042	-0.018	0.144	0.049	-0.021	0.965	0.157	**0.52**
PC7	-0.487	0.124	0.358	-0.040	-0.537	0.360	0.445	-0.052	0.105	**0.25**
PC8	0.143	0.050	-0.193	0.023	0.085	-0.506	0.819	0.069	0.060	**0.57**

Table II. *Principal components of ELMy H-mode subset of DB3v13 with estimates of $\lambda_e + \lambda_{pc}$ and ERR as explained in the text. Weights $\propto 1/(2+\sqrt{Nj/4})$ have been applied in the analysis. Nj is the number of points from device j. Values of ERR > 0.2 are shown in bold face.*

yield ERR6 < 0.2). PC7 and PC8 depend mainly but not only on qcyl. In fact, multiplying a scaling with a combination of PC7 and PC8 can change a β dependence into a dependence on ν* [5]. The bias associated with the errors on PC4 (due to δP), PC5 (due to δP and δn) and PC7 (due to δa and δA) are all only marginally significant.

3. Regression analysis with error in variable methods: The general method is to normalise the variables with respect to their errors and use a principal component analysis (PCA) on the normalised variables to determine the regression plane [3].

3.1 A SAS program, made by S. Pantula [4] that also takes error cross correlations into account has been used on the extended ELMy H-mode dataset. As reference standard OLS regression gives $\tau_E = 3.33\mathrm{s} \times I^{0.91}\ B^{0.18}\ P^{-0.68}\ n^{0.37}\ R^{1.30}\ a^{-1.06}\ A^{0.85}\ M^{0.07}$ with the variables in ITER units: I {15MA}, B {5.3T}, R {6.2m}, a {2m}, A {22m^2}, n {10^{20} m^{-3}}, P{87MW}. Using three correlations, -1 × δB δR, –0.5 × δa δn and +0.5 × δa δA, in the error model for the set of variables used together with the average errors from Table I, the Pantula program gives $\tau_E = 3.29\mathrm{s} \times I^{0.86}\ B^{0.22}\ P^{-0.64}\ n^{0.34}\ R^{1.31}\ a^{-1.43}\ A^{1.02}\ M^{0.07}$. Without the correlation between δa and δA the relative change in the I and B exponents would have been even larger and the ITER prediction smaller. Increasing the strength to +0.75 × δa δA gives $\tau_E = 3.34\mathrm{s} \times I^{0.89}\ B^{0.17}\ P^{-0.63}\ n^{0.34}\ R^{1.30}\ a^{-1.22}\ A^{0.90}\ M^{0.09}$. The effects of the other two correlations are much weaker. Finally, increasing δP also reduces the resulting power degradation. This effect is pursued in the next analysis.

3.2 A simpler version [2] of the PCA error in variable method that ignores cross correlation errors has been used to analyse a Deuterium only ELMy H-mode dataset of DB3v13. The regression result is $\tau_E = 3.36\mathrm{s} \times I^{0.77}\ B^{0.28}\ P^{-0.57}\ n^{0.32}\ R^{1.15}\ a^{-1.64}\ A^{1.19}$ for the average errors given in Table I . This expression is $B\tau_E \sim \beta^{-0.42}\ \nu^{*\,-0.19}\ \rho^{*\,-2.7}$ in non-dimensional form. The result is rather sensitive to the errors in the stored energy δW_{th} and the loss power δP. For fixed δW_{th} = 13% with δP ranging from 8% to 25% the scaling changes from $B\tau_E \sim \beta^{-0.77}\ \nu^{*\,-0.09}\ \rho^{*\,-2.6}$ for δP = 8% to $B\tau_E \sim \beta^{0.16}\ \nu^{*\,-0.36}\ \rho^{*\,-2.7}$ for δP = 25%. Fig. 1. shows the variation of the resulting β and ν* exponents as δP is varied. It is remarkable that the variation of δP from 8 to 25% only results in a variation of the prediction for ITER from 3.33 to 3.50s.

4. Conclusion: The strong degradation with β in the ELMy H-mode confinement scaling expressions, eg. the IPB(y,2) scaling expression [6], obtained with standard OLS regression is not observed experimentally [7,8]. This discrepancy may be due to measurement error related bias in the OLS regression. However, a reduction in the β degradation has so far only been achieved by assuming an error on the loss power that is as large or larger than the error on the energy. The range in point predictions for ITER resulting from the regression analysis using the PCA errors in variable methods remains in the interval [-30%, +10%] times 3.6s, predicted by IPB(y,2) [6]. The D only subset without Ohmic data leads to the lowest values.

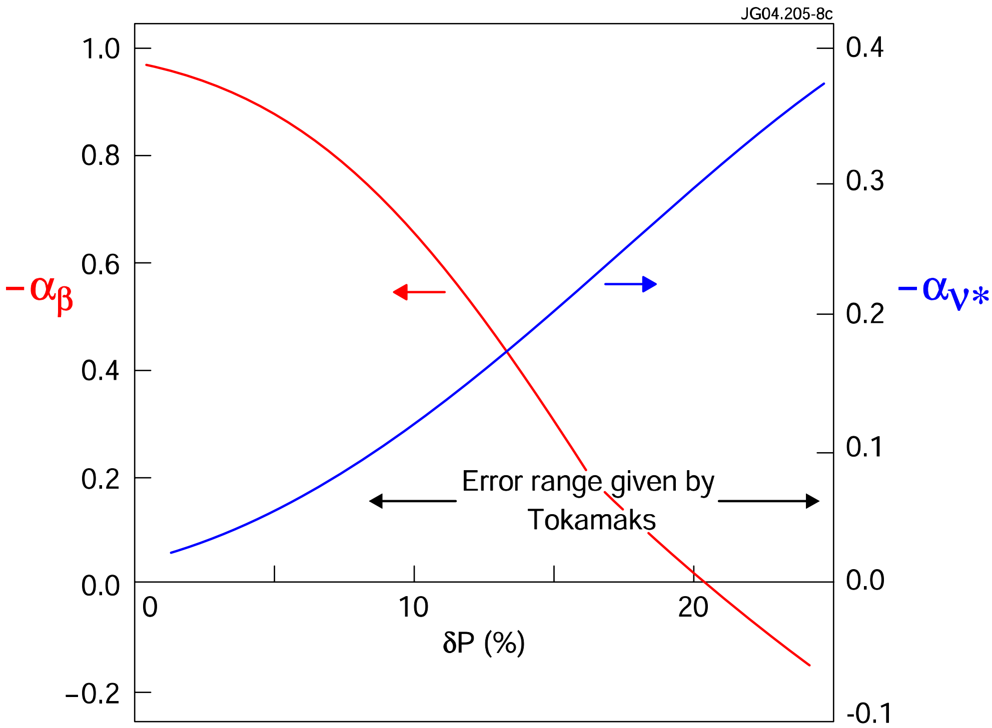

Fig.1. *The resulting α_β and $\alpha_{\nu*}$ of the confinement scaling expression $\propto \beta^{\alpha_\beta}\, \nu*^{\alpha_{\nu*}}$ as the assumed error δP on the loss power is varied in the PCA errors in variable method.*

5. References:

[1] Christiansen, J. P., et. al., Nuclear Fusion 32 (1992) 291.
[2] Kardaun, O.J.W.F., et.al., EPS XVI, Venice (1989), Part I, 253.
[3] Anderson, T.W., Annals of Statistics, Vol. 12, Issue 1 (1984) 1.
[4] Fuller, W.A., Measurement Error Models, Wiley, 1987.
[5] Kardaun, O.J.W.F., et.al., 9th IAEA TM on H-mode Physics and Transport Barriers, San Diego, California, 2003.
[6] ITER Physics Expert Groups, Nuclear Fusion 39 (1999) 2175.
[7] Petty C C, et.al., Physics of Plasmas 11 (2004) 2514.
[8] McDonald D C, et.al., Plasma Phys. Controlled Fusion 46 (2004) A215.