Cross-Validation Estimates of Statistical Uncertainties in Empirical Tokamak Scalings

Greg Hammett, PPPL

http://w3.pppl.gov/~hammett

W. Dorland, Univ. Maryland, M. Kotschenreuther, IFS

Ackn: C. Kessel, D. Meade

TTF April, 2003 Madison

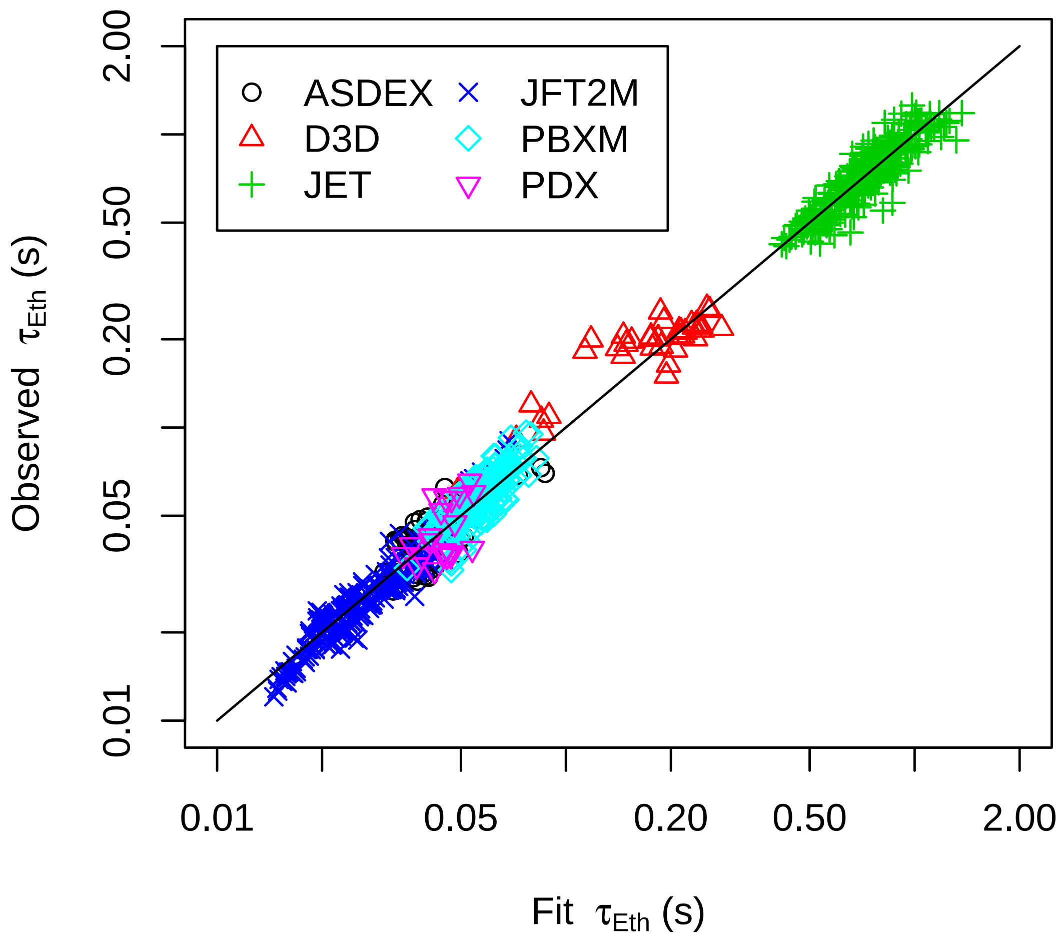
In process of writing up previous results (summarized in 1997 FESAC report) using cross-validation etc. to estimate non-ideal statistical uncertainties in ITER-93H empirical scaling. Extending analysis to more recent databases.

Simple description of effect of non-ideal correlated errors, cross-validation method to estimate them.

Example application to ITER-93H database, extension to H98(y,2) database.

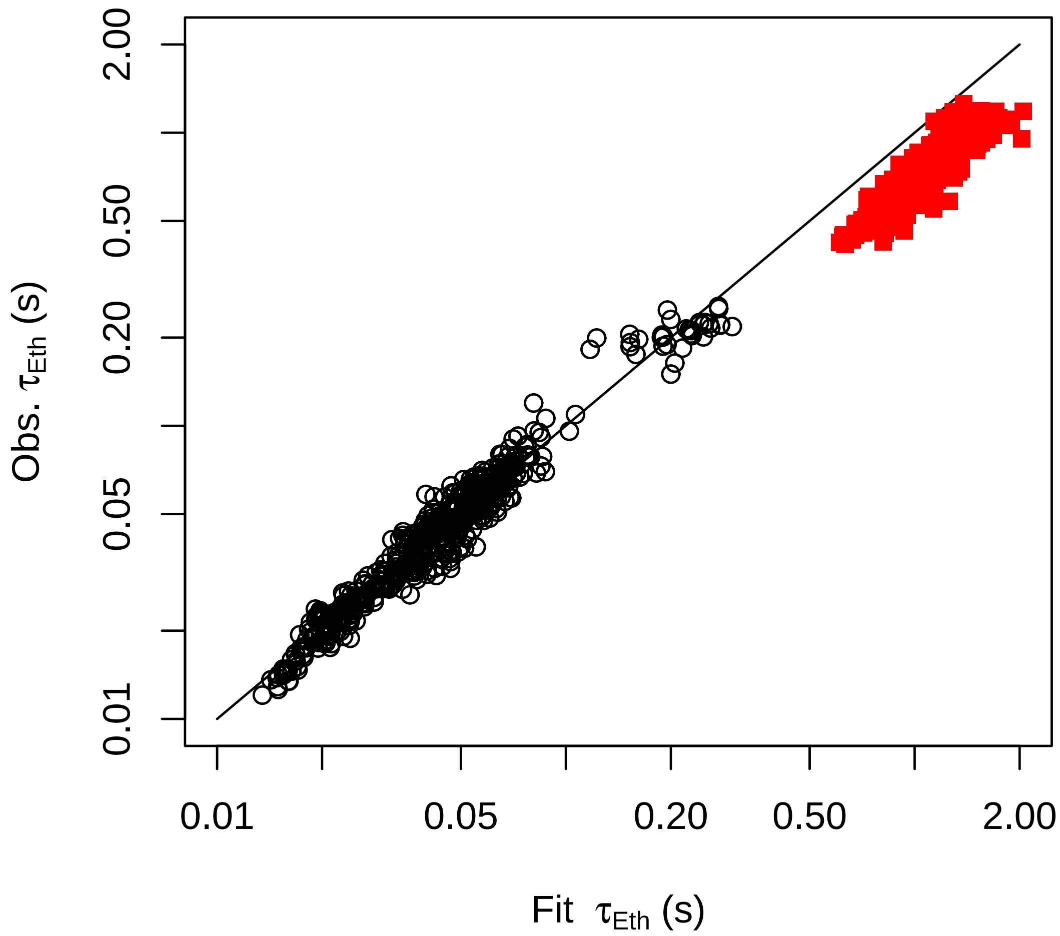
Major problems with original ITER-96 design essentially boiled down to the issue that it had to operate at 1.5 $\times$ Greenwald density limit: significant uncertainties in extrapolation.

New, improved fusion reactor designs: higher elongation & triangularity: below Greenwald density, reduces uncertainties to more manageable level.



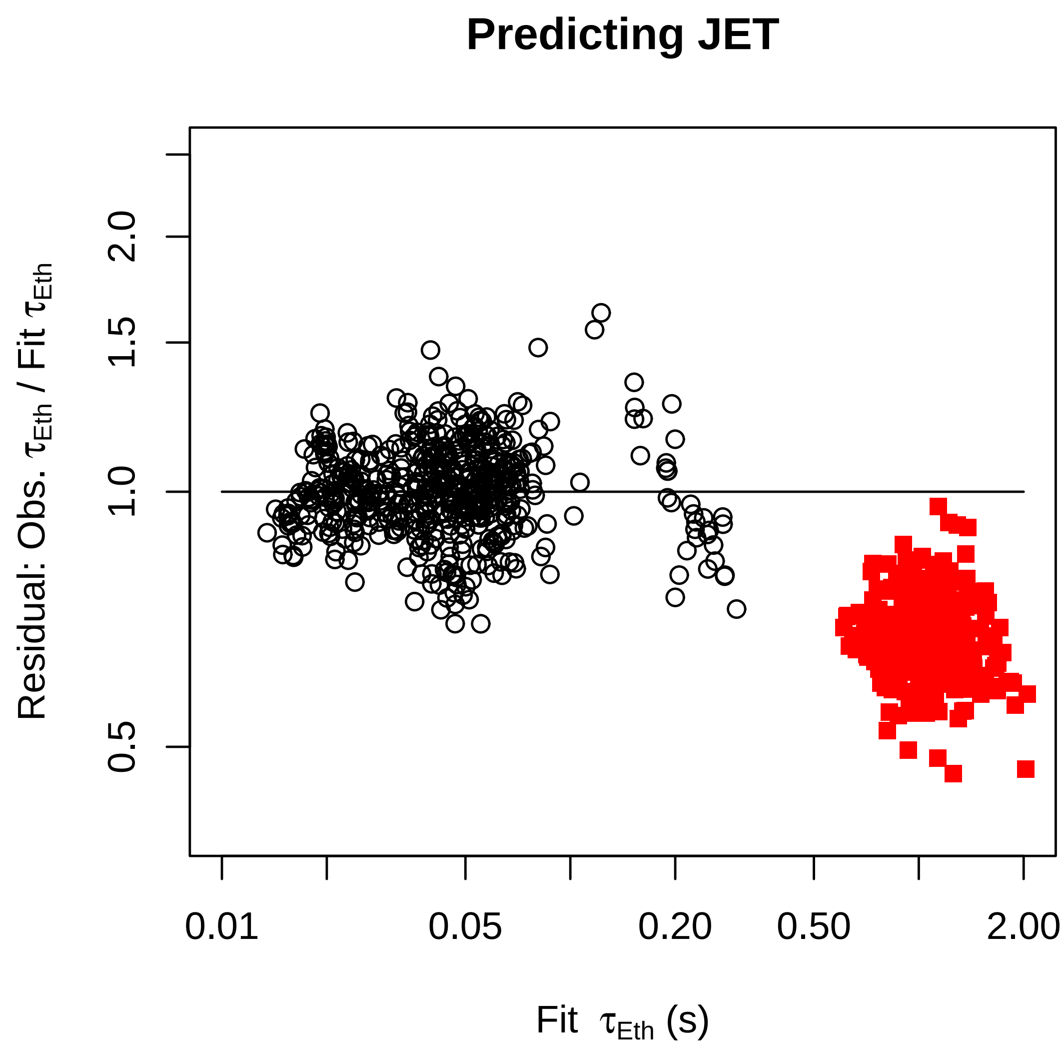
ITER-93H scaling for the tokamak energy confinement time τ_E compared with the data from six tokamaks in DB2.5?? to which it was fit. The RMSE of the fit in log(τ_E) is 0.123.

Predicting JET



Prediction of JET data (red), using a fit to the other five tokamaks excluding JET data. The RMSE of the fit to the data excluding JET is 0.125. The RMSE of predicting JET data is $\Delta_{JET} = 0.408$, which is significantly larger than ideal expected $\hat{\Delta}_{JET} = 0.138$ if all errors were statistically independent.

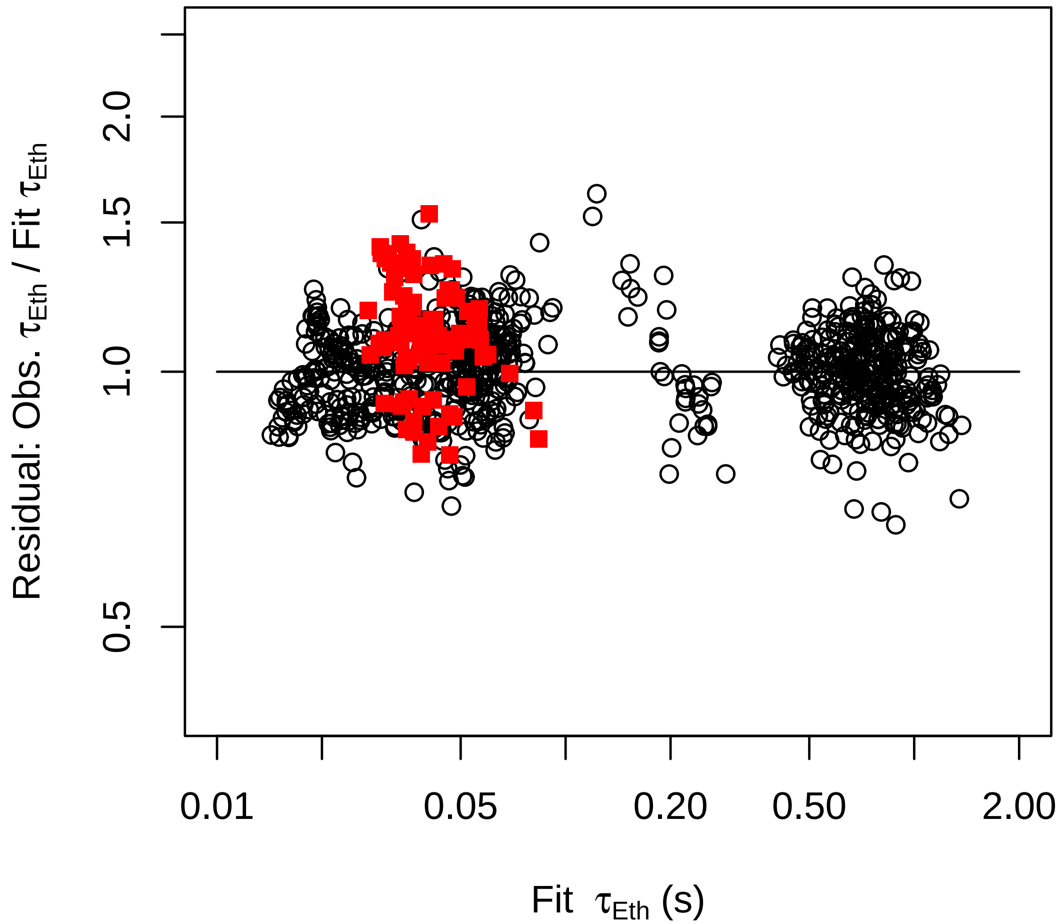
JET data is systematically low, mean prediction error $\bar{\Delta}_{JET} = -0.393$, significantly larger than expected ideal error in the mean $\sigma_{\bar{\Delta}_{JET}} = \pm 0.060$.



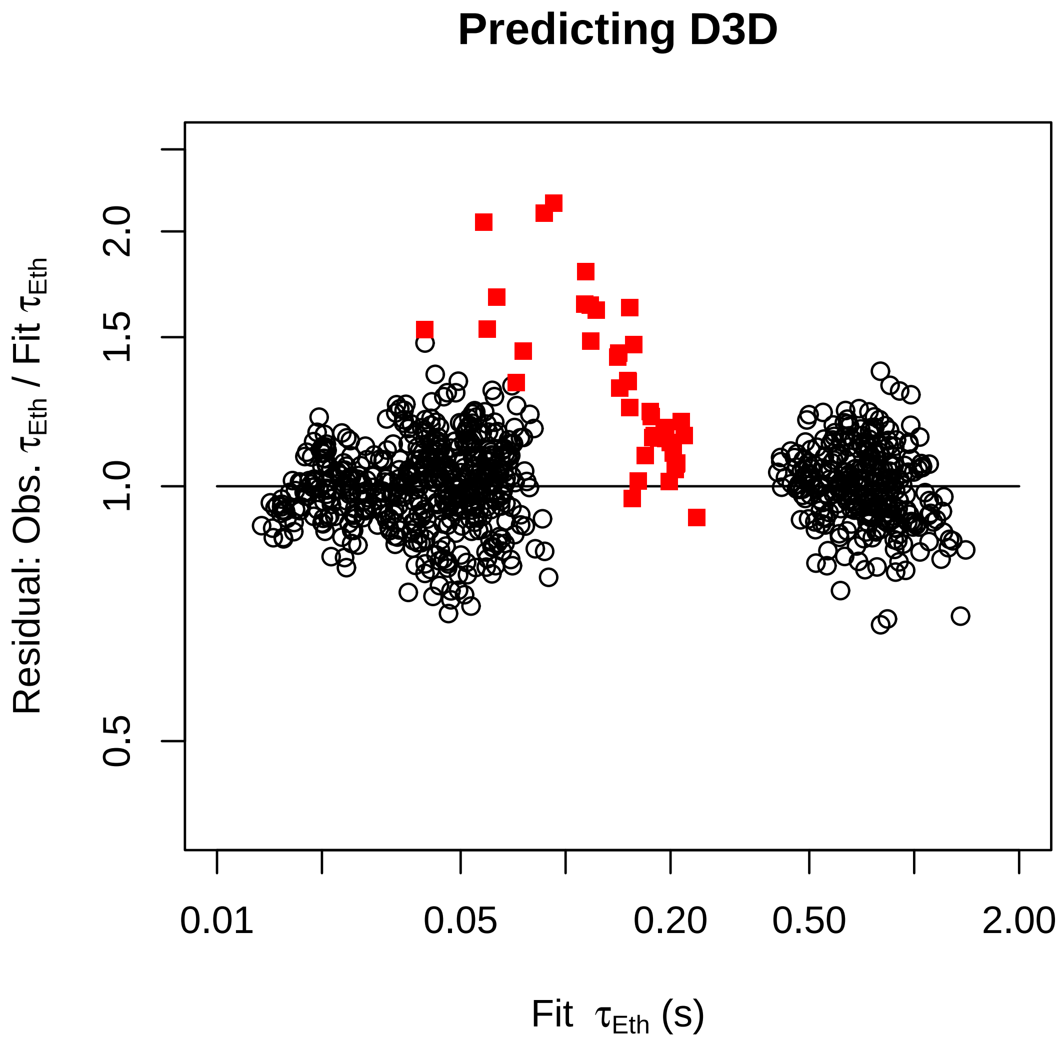
(Like previous plot, but easier to see residuals.) Prediction of JET data (red) using fit to other five tokamaks excluding JET. RMSE of fit to data excluding JET is 0.125. RMSE of predicting JET data is $\Delta_{JET} = 0.408$, significantly larger than ideal expected $\hat{\Delta}_{JET} = 0.138$ if all errors were statistically independent.

JET data systematically low, mean prediction error $\bar{\Delta}_{JET} = -0.393$, significantly larger than expected ideal error in mean $\sigma_{\bar{\Delta}_{JET}} = \pm 0.060$.

Predicting ASDEX

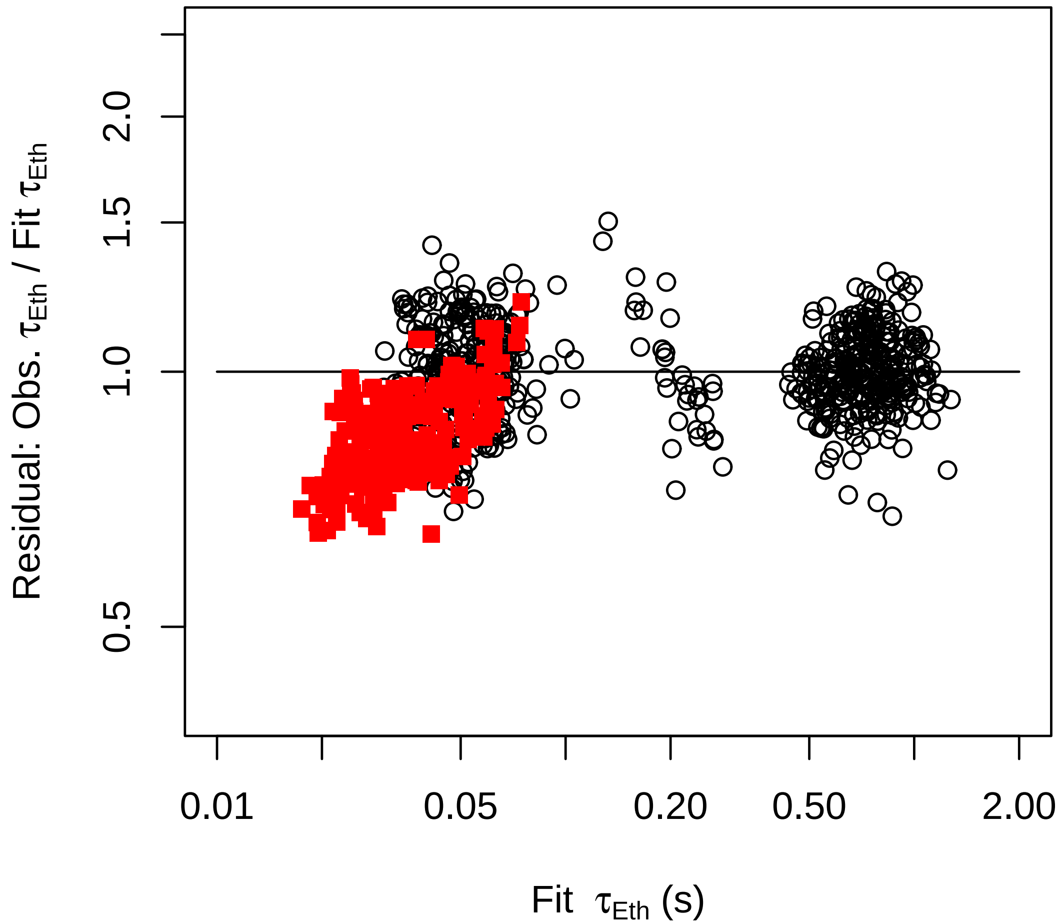


Residual errors of the fit to data excluding ASDEX (black, RMSE = 0.119). The resulting fit is then used to predict ASDEX (red, RMSE=0.175, mean error=-0.097).

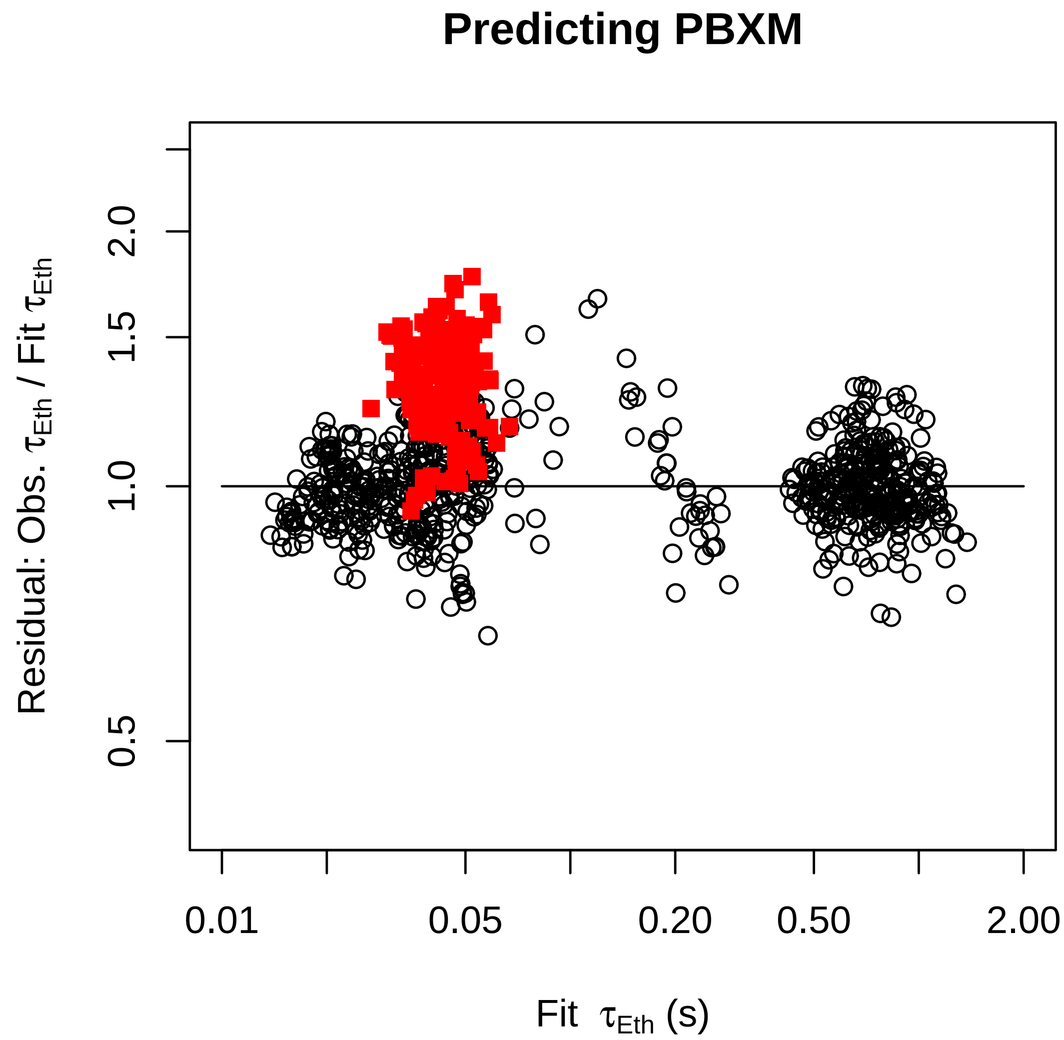


Residual errors of the fit to data excluding DIII-D (black, RMSE = 0.113). The resulting fit is then used to predict DIII-D (red, RMSE=0.350, mean error=-0.277).

Predicting JFT2M

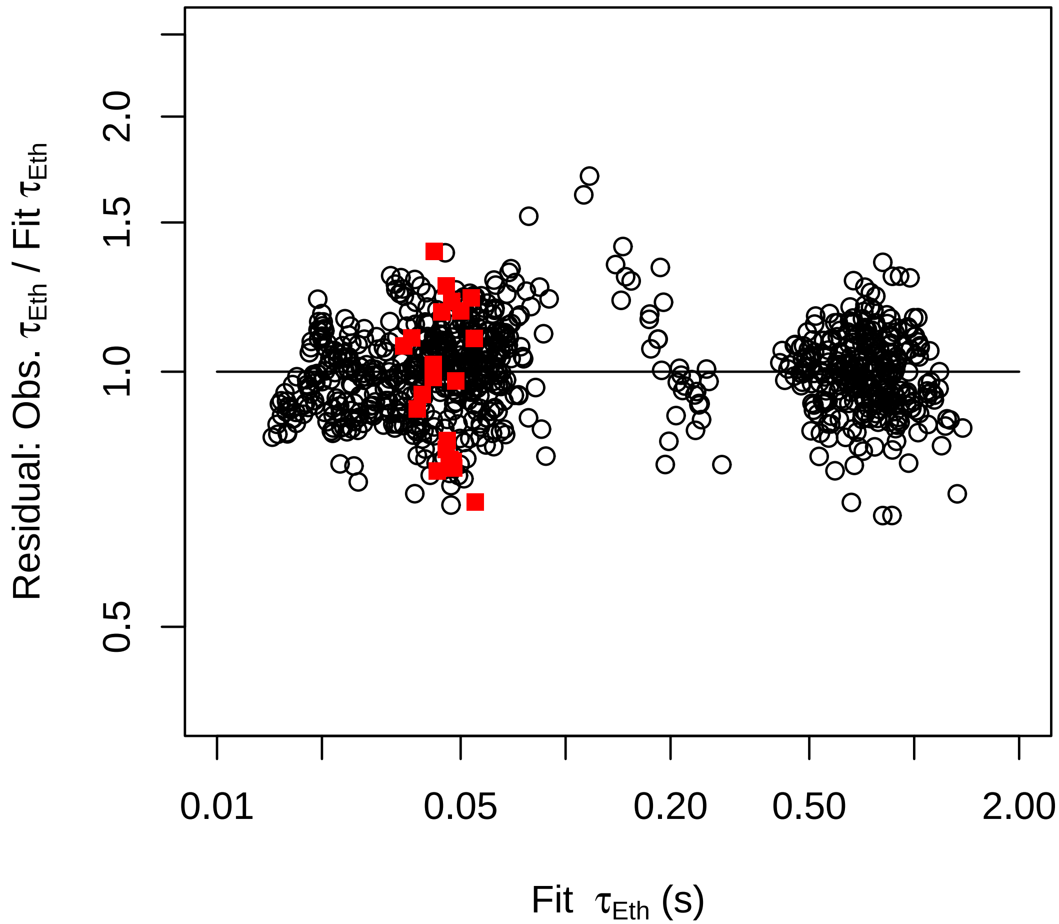


Residual errors of the fit to data excluding PBX-M (black, RMSE = 0.118). The resulting fit is then used to predict PBX-M (red, RMSE=0.307, mean error=-0.277).



Residual errors of the fit to data excluding PBX-M (black, RMSE = 0.118). The resulting fit is then used to predict PBX-M (red, RMSE=0.307, mean error=-0.277).

Predicting PDX



Residual errors of the fit to data excluding PDX (black, RMSE = 0.120). The resulting fit is then used to predict PDX (red, RMSE=0.196, mean error=0.030).

Cross-validation motivation

Most elementary statistics textbooks focus on standard techniques for ideal case of uncorrelated errors with normal Gaussian distribution, etc.

Many ways in which ideal assumptions can be violated. Many specialized techniques developed to handle various special cases. Cross-validation, jackknife, described in classic textbook by Mosteller and Tukey, advanced texts and research papers.

Cross-validation is straightforward technique to directly test how well a formula predicts new data the formula wasn't fit to. Avoids "overfitting".

Cross-validation extensively used for neural networks: train on one set of data, judge how well it performs on an independent set of data.

Ideal statistics w/ uncorrelated errors

Simple model $y_i = \mu + \epsilon_i$

ideal uncorrelated errors $\langle \epsilon_i \epsilon_j \rangle = \sigma^2 \delta_{ij}$

Estimated mean $\bar{y} = \frac{1}{N} \sum_i y_i$

Estimated variance
$$s^2 = \frac{1}{N-1} \sum_i (y_i - \bar{y})^2$$
$$\langle s^2 \rangle = \sigma^2$$

Uncertainty in mean $\sigma_{\bar{y}}^2 = \langle (\bar{y} - \langle \bar{y} \rangle)^2 \rangle = \frac{\sigma^2}{N}$

Non-Ideal statistics w/ correlated errors

For simplicity assume observations contain C perfectly correlated copies. **(can be generalized to partial correlations)**

$$\langle \epsilon_i \sum_j \epsilon_j \rangle = C\sigma^2$$

Effective number of indep. observations $N_{eff} = N/C$

Unbiased est. variance $s_C^2 = \frac{1}{N_{eff}-1} \sum_{i=1}^{N_{eff}} (y_{iC} - \bar{y})^2$

Usual variance estimate is now biased:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2 = \frac{C}{N-1} \sum_{i=1}^{N/C} (y_{iC} - \bar{y})^2 = \frac{N-C}{N-1} s_C^2$$

Uncertainty in mean $\sigma_{\bar{y}}^2 = \frac{s_C^2}{N_{eff}} = C\frac{s^2}{N}\frac{N-1}{N-C}$

Error estimates for Multiple Regression

Uncertainty in regression formula's predicted mean $y(\vec{x})$

$$\sigma_{\hat{y}}^2 = \frac{s_C^2}{N_{eff}}\left(1 + \Lambda^2\right)$$

where $\Lambda^2 = \Sigma_j (x_j - \bar{x}_j)^2/\sigma_{x_j}^2$ (in orthogonal basis) is distance of extrapolation from center of database, normalized to std. dev.

Total expected variance Δ_g^2 of a new set of data g predicted by a regression formula:

$$\begin{aligned}\Delta_g^2 &= s_C^2 + \frac{s_C^2}{N_{eff}}\left(1 + \Lambda^2\right) \\ &= \frac{N - N_p}{N - CN_p} s^2 + C\frac{N - N_p}{N - CN_p}\,\frac{s^2}{N}\left(1 + \Lambda^2\right)\end{aligned}$$

Compare with actual error of prediction $\rightarrow$ measures C.

Cross-Validation for estimating non-ideal errors in Multiple Regression

Drop one group of data (for example, from tokamak g) from the database.

Fit a regression formula $y_{(g)}(\vec{x})$ to the data excluding group g.

Use this regression formula to predict data in group g, measure the actual error of prediction Δ_g^2.

Repeat for each group g. Average results to determine the effective number of correlated observations C

$$C = \frac{\overline{\Delta_g^2 N_{(g)}} - \overline{s_{(g)}^2 (N_{(g)} - N_p)}}{\overline{\Delta_g^2} N_p + \overline{\langle\langle s_{\hat{y}}^2 \rangle\rangle_g (N_{(g)} - N_p)}}$$

and effective number of independent obs. $N_{eff} = N/C$.

Monte Carlo tests with synthetic data used to verify these formulas.

Cross-validation summary table for ITER-93H scaling

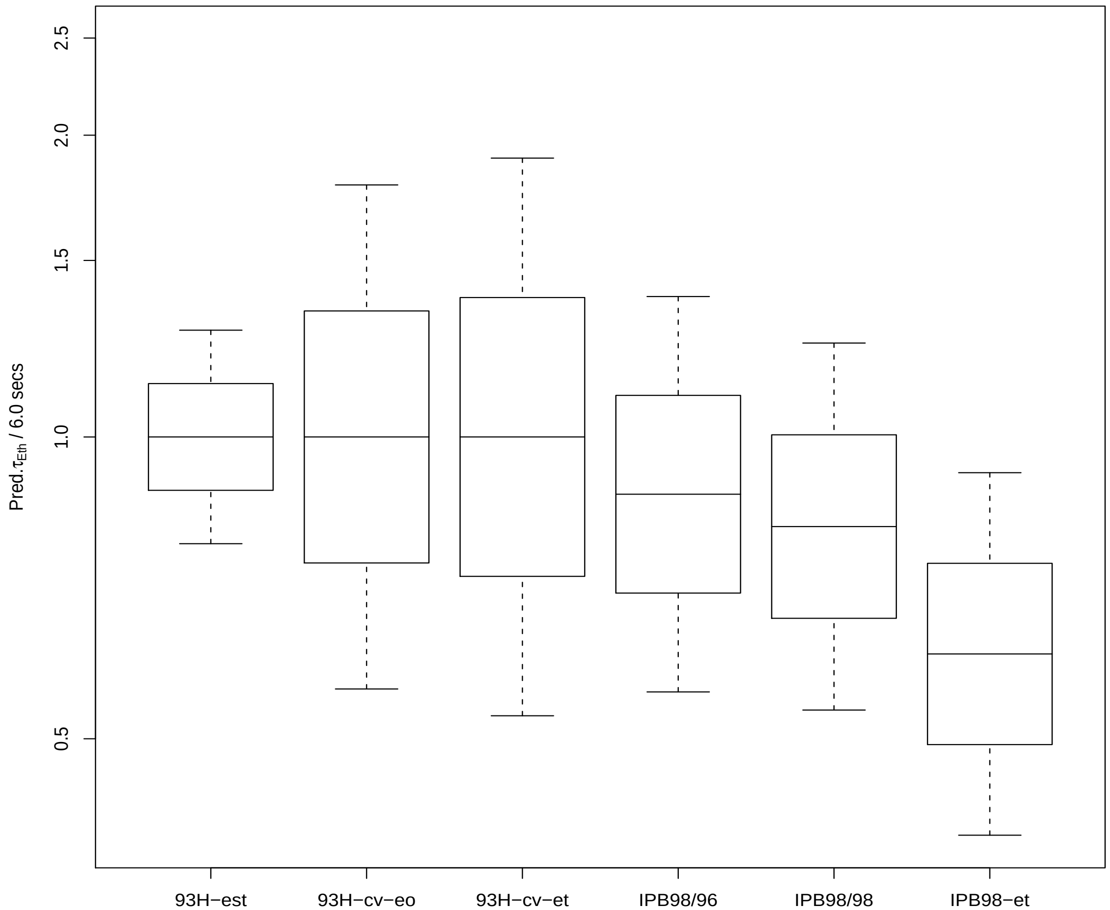
Tokamak (Group) predicted	RMSE prediction actual	ME pred	RMSE fit	ideal sd(pred mean)	ideal RMSE pred	ideal sd(ME)	RMS extrap. dist.	# of obs.	# of eff. corr. obs.
g	Δ_g	$\overline{\Delta}_g$	$s_{(g)}$	$\langle\langle s^2_{\hat{y}}\rangle\rangle^{1/2}_g$	$\hat{\Delta}_g$	$\sigma_{\overline{\Delta},g}$	Λ	N_g	C_g
ASDEX	0.175	0.097	0.119	0.027	0.122	0.030	6.3	91	15.0
D3D	0.350	0.277	0.113	0.031	0.117	0.036	7.8	40	47.7
JET	0.408	-0.393	0.125	0.059	0.138	0.060	11.1	303	24.6
JFT2M	0.215	-0.177	0.123	0.037	0.128	0.037	7.3	247	15.6
PBXM	0.307	0.277	0.118	0.038	0.124	0.039	8.5	155	30.7
PDX	0.196	-0.030	0.120	0.021	0.122	0.033	4.8	22	29.2
Avg.	0.288	0.241	0.120	0.037	0.126	0.040	7.9	143	27.1

More careful averaging yields $C = 28.5$ for equal tokamak weighting and $C = 24.7$ for equal observation weighting. The resulting uncertainty in any predicted mean response is enhanced by a factor of 6.3 (equal tokamak weighting) or 5.7 (equal observation weighting) over the standard result with ideal statistics.

Cross-validation summary table for IPB98(y,2) scaling

Tokamak (Group) predicted	RMSE prediction actual	ME pred	RMSE fit	ideal sd(pred mean)	ideal RMSE pred	ideal sd(ME)	RMS extrap. dist.	# of obs.	# of eff. corr. obs.
g	Δ_g	$\overline{\Delta}_g$	$s_{(g)}$	$\langle\langle s^2_{\hat{y}}\rangle\rangle^{1/2}_g$	$\hat{\Delta}_g$	$\sigma_{\overline{\Delta},g}$	Λ	N_g	C_g
ASDEX	0.148	-0.049	0.145	0.021	0.147	0.022	4.1	431	1.5
AUG	0.135	-0.054	0.143	0.009	0.144	0.017	2.0	102	-10.2
CMOD	0.313	0.299	0.142	0.041	0.147	0.047	10.3	37	33.0
D3D	0.191	-0.102	0.136	0.018	0.137	0.020	4.3	270	27.9
JET	0.296	0.267	0.139	0.026	0.142	0.028	6.0	246	48.3
JFT2M	0.122	0.090	0.144	0.023	0.146	0.030	5.6	59	-9.1
JT60U	0.246	-0.243	0.142	0.014	0.142	0.049	3.3	9	67.3
PBXM	0.245	0.176	0.140	0.039	0.146	0.043	9.9	59	20.6
PDX	0.196	0.069	0.139	0.017	0.140	0.022	4.1	97	33.7
Avg.	0.220	0.176	0.141	0.025	0.143	0.033	6.1	145.6	23.7

More careful averaging yields $C = 28.7$ for equal tokamak weighting and $C = 26.7$ for equal observation weighting. Uncertainty in predicted mean enhanced by factor of 6.0 (equal tokamaks) or 5.7 (equal observations) over standard ideal result. Errors also drop because $N = 858$ for ITER-93H, $N = 1310$ observations for $IPB-98(y,2)$, and dataset covers wider parameter range.



Predicted τ_E for ITER-96 design (or slightly lower density ITER-98) with estimated 1 and 2 σ error bars:

"93H-est" $\sigma = 0.12$ original ITER assumption $N_{eff} = N/4$

"93H-cv-eo" $\sigma = 0.29$ cross-validation est. equal obs. weights

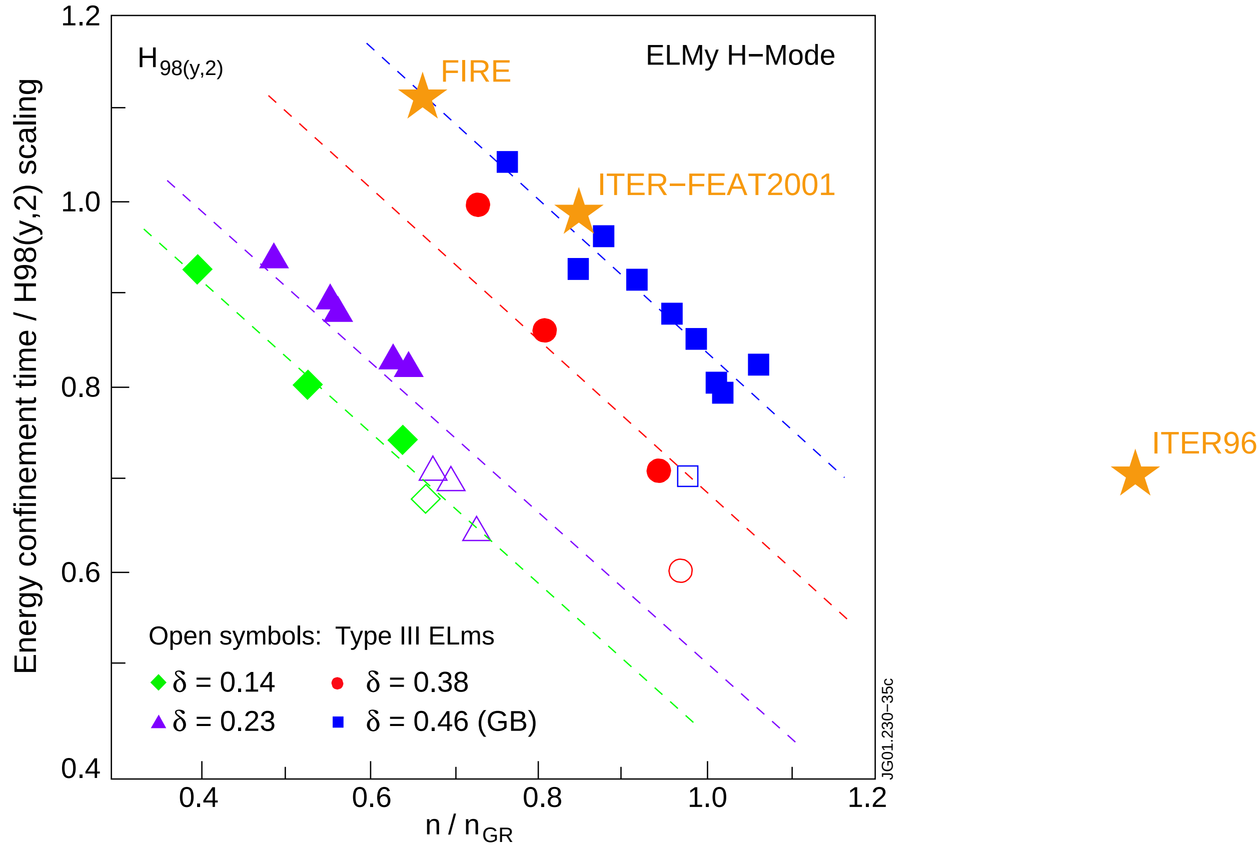
"93H-cv-et" $\sigma = 0.32$ cross-validation est. equal tok. weights

"IPB98/96" $\sigma = 0.23$ cross-validation est. orig. ITER96 design

"IPB98/98" $\sigma = 0.21$ cross-validation est. ITER98

"IPB98-et" $\sigma = 0.21$ equal tok. weight in fit and CV

Stronger plasma shaping improves performance



JET data from G. Saibene, EPS 2001, J. Ongena, PPCF 2001. Seen in other tokamaks also.

Confinement degrades if density too large relative to empirical Greenwald density limit $n_{Gr} = I_p/\pi a^2$, improves with higher triangularity.

Relative to original 1996 ITER design, new ITER-FEAT 2001 and FIRE designs have significantly higher triangularity and elongation, & designed to operate at significantly lower density relative to Greenwald density limit.

Improvements in new fusion designs $\downarrow$ uncertainties

Density and pressure limits improve with stronger shaping elongation κ & triangularity δ:

Empirical Greenwald density limit $$n_{Gr} = \frac{I_p}{\pi a^2} \propto \frac{B_T}{R q_{95}} \left[1 + \kappa^2 (1 + 2\delta^2)\right]$$

Pressure limit $$\beta_{Troyon} = \frac{p}{B^2/8\pi} = \frac{I_p}{a B_T} \propto \frac{a}{R q_{95}} \left[1 + \kappa^2 (1 + 2\delta^2)\right]$$

New ITER-FEAT design uses segmented central solenoid to increase shaping.

FIRE pushes to even stronger shaping.

	R m	a m	B T	I_p MA	n_{Gr} $10^{20}/m^3$	$\frac{\langle n_e \rangle}{n_{Gr}}$	κ_x	δ_x	P_{fusion} keV	$P_\alpha/(2\pi R)$
ITER-96	8.14	2.80	5.68	21.0	0.85	1.50	1.75	0.35	1500	5.9
ITER-FEAT	6.20	2.00	5.30	15.1	1.19	0.85	1.85	0.48	400	2.0
FIRE	2.14	0.60	10.0	7.7	6.92	0.66	2.00	0.70	150	2.2
Aries-AT (a goal)	5.20	1.30	5.86	12.8	2.41	1.00	2.18	0.84	1760	9.0

Caveats: There are still some remaining uncertainties regarding confinement, edge pedestal scaling, ELMs, disruptions, & heat loads, tritium retention, neoclassical beta limits, but also reasonable possibilities for dealing with potential problems or further improving performance.