A Bayesian Approach to the Global Confinement Time Scaling in W7AS

V. Dose, W. von der Linden, and A. Garrett
Max-Planck-Institut für Plasmaphysik, EURATOM Association
(Postal address: PO BOX 1533, D-85740 Garching bei München, Germany)

Abstract. A recently assembled W7AS stellarator database, consisting of 143 electron cyclotron resonance (ECR) heated discharges, 51 discharges heated by neutral beam injection (NI), and 5 discharges using combined ECR and NI heating, is used to test three scaling models for the energy confinement time in terms of other physical parameters. The three models are: Lackner-Gottardi; ion plateau neoclassical transport; and ITER 89-P. The analysis employs Bayesian reasoning. Based on these data, Lackner-Gottardi scaling is 37 dB more likely to be correct than ITER 89-P in W7AS, and is over 21 dB more probable than the ion plateau neoclassical transport model. Values for the scaling exponents in the light of the data are given.

1 Introduction

In the absence of *ab initio* theoretical expressions for the energy confinement parameters in fusion devices, empirical scaling rules of power law type remain vital, especially in the construction of the next generation of machines. Confinement properties of the proposed W7X machine have accordingly been estimated on the basis of confinement time data from the existing W7AS stellarator at Garching [1]. We adopt the expression

$$\tau = \exp(\alpha_c)\; n^{\alpha_n} B^{\alpha_B} P^{\alpha_P} a^{\alpha_a} \iota^{\alpha_\iota} \equiv f(\boldsymbol{\alpha}, \mathbf{X}) \tag{1}$$

where τ, n, B, P, a, ι respectively denote the confinement time, the line-averaged density, the toroidal magnetic field, the injected power, the minor radius and the rotational transform. The usual procedure is to convert (1) into a linear problem by taking the logarithm. This procedure is, however, improper for two reasons. Firstly, it assumes implicitly a constant relative error of the data. This is hard to conceive for an electrical measurement for e.g. the plasma diamagnetic energy content W_{diag}. An electrical measurement is in error by the square root of the signal size if carried out in the shot noise limit. If on the other hand the measurement is deteriorated by an external source not associated with W_{dia} itself, the noise level should be independent of W_{dia}, i.e. of constant absolute value. Secondly, even in the inconceivable case of constant relative error, the nonlinear problem and the problem linear in the logarithms are only equivalent if the errors are negligible, as will be shown in detail in Appendix I. The situation of negligible errors does of course not apply to confinement time scaling in nuclear fusion. Consequently, we employ nonlinear least squares fitting assuming constant absolute errors on the experimental data throughout this paper. In addition we employ Bayesian probability theory which allows to incorporate prior information. In particular, it enables to test the W7AS performance against various slightly differing current scaling expressions. Again this is of importance for the prediction of confinement properties of W7X. Model testing is a particularly interesting capability of the Bayesian approach since the fit problem (1) will always result in a rather flat hypersurface in parameter space whose absolute minimum has correspondingly little significance. The reason is quite simple. A particular data set may conform more or less with basic plasma physical relations in which case the exponents in (1) satisfy to a certain degree one or more linear relationships. These in effect reduce the set of independent variables (n, B, P, a)

to a number smaller than 4 and make the fit problem (1) progressively singular. Accidentally, a second source of singularity may come into play. If the sampling of the input data space (n, B, P, a, ι) suffers from collinearity between two or more variables it becomes progressively difficult to determine the dependence of the confinement time τ on these variables. This point will be discussed in the next section. In summary, the capability of Bayesian analysis to perform model comparison and to deal with ill conditioned problems makes it an ideal tool to analyze confinement properties of nuclear fusion devices.

2 The data set

The data set investigated in this paper is the same as in [2] and consists of three types of discharges. 143 discharges employ ECR heating only, 51 discharges employ NI heating only while 5 discharges combine NI and ECR. The range of the variables is displayed in table I. In order to investigate collinearities between the "input" data n, B, P, a, ι we first set up their covariance matrix and perform a principle component analysis [3]. Since the input variables have different dimensions and vastly different mean values this makes sense only after transformation to the dimensionless variables $v \rightarrow v/\overline{v}$, where v stands for one of the interesting variables and the $\overline{v}$ denotes its mean value.

Table II summarizes the dominant correlation coefficients. Let us first discuss the ECR-only and NI-only data subsets. The largest correlation in the NI subset shows up between density and power. This is expected since at a fixed beam energy the deposited power is directly given by the number of injected particles and so to the density n. The largest correlation in the ECR data set shows up between density and magnetic field. This does not come as a surprise either. ECR heating is performed at $70GHz$ and at a magnetic field of roughly $1.5T$ and $2.5T$. In the low field case heating proceeds via the second harmonic and is restricted to roughly half the density range of the high field case. Concomitantly we shall see that fits for these two subsets suffer from instabilities in the exponents (α_n, α_P) and (α_n, α_B), respectively. The third column in table II displays the correlations in the global data set. Caution has to be given to the calculation of these correlations. Simple straight forward pooling of the ECR and NI data leads to the paradoxical result that the correlation coefficient for n and p of the global data set is $r(n,p) = 0.728$, which is a greater value than that derived for either the ECR or the NI subsets. The reason is that in particular the density n covers a significantly greater range in the NI discharges than in the others. The concept of correlation coefficients is, however, only invariant under scaling of all data by a common factor. If on the other hand a data set consists of several groups, which cover different volumes of the data space, we must account for this grouping. Details are given in Appendix II. The properly performed correlation analysis of the pooled data shows a significantly reduced $r(n, B)$ and a negligible $r(n, P)$. This means that we do need a data set combining NI and ECR discharges in order to compute the dependence of the confinement time τ on the individual variables n, B, P. The remaining correlation coefficients are quite small. Unlike the above discussed correlations they are not dictated by physics and could in principle be reduced by the choice of an appropriate shot program.

The good natured behavior of the data set is also reflected in the range of eigenvalues of the properly pooled covariance matrix. The ratio of the largest eigenvalue to the smallest is only 23. Accordingly the structure of the associated normalized eigenvectors is such that four of them have one component larger than 0.9 on n, P, a, ι while the largest component of the fifth is 0.79 on B. Altogether we conclude that the global data set is robust and well suited to investigate confinement performance of W7AS.

3 Bayesian Analysis

The joint probability distribution $p(\boldsymbol{\alpha}, \tau|M)$ for the parameters $\boldsymbol{\alpha}$ characterizing our scaling function and the confinement time data τ, given a model M, may be expanded according to the product rule [4]:

$$p(\boldsymbol{\alpha}, \tau|M) = p(\tau|\boldsymbol{\alpha}, M)p(\boldsymbol{\alpha}|M) = p(\boldsymbol{\alpha}|\tau, M)p(\tau|M) \qquad . \tag{2}$$

Here $p(\tau|\boldsymbol{\alpha}, M)$ is the probability of the data given parameters $\boldsymbol{\alpha}$ and model M; $p(\boldsymbol{\alpha}|M)$ is the prior probability for the parameters $\boldsymbol{\alpha}$ given model M; and $p(\tau|M)$, the normalizing factor, is the probability of the data given model M. Rearranging the right hand side of (2) we solve for $p(\boldsymbol{\alpha}|\tau, M)$:

$$p(\boldsymbol{\alpha}|\tau, M) = p(\boldsymbol{\alpha}|M)\frac{p(\tau|\boldsymbol{\alpha}, M)}{p(\tau|M)} \tag{3}$$

where the denominator $p(\tau|M)$ can be written in the form of a normalizing factor:

$$p(\tau|M) = \int d\boldsymbol{\alpha}\ p(\tau, \boldsymbol{\alpha}|M) = \int d\boldsymbol{\alpha}\ p(\boldsymbol{\alpha}|M)p(\tau|\boldsymbol{\alpha}, M) \tag{4}$$

Equation (3) is Bayes' theorem for updating our prior probability distribution for $\boldsymbol{\alpha}$ in the light of the data τ. Equation (4) tells us how to remove parameters (here, $\boldsymbol{\alpha}$) from a Bayesian calculation by averaging, or *marginalizing* over them; it follows from the product rule for probabilities combined with the sum rule.
If the measurement errors ϵ_k of the data τ_k, $k = 1 \dots K$ are logically independent, and the only noise parameters of significance are the mean $\langle\epsilon_k\rangle = 0$ and the variance $\langle\epsilon_k^2\rangle = \sigma^2 = z$, then by the maximum entropy principle the likelihood $p(\tau|\boldsymbol{\alpha}, M)$ takes the familiar Gaussian form [5]

$$p(\tau|\boldsymbol{\alpha}, z, M) = (2\pi z)^{-K/2}\exp(-L/2z)\ ; \qquad L = \sum_{k=1}^{K}\epsilon_k^2 = \sum_{k=1}^{K}(\tau_k - f(\boldsymbol{\alpha}, \mathbf{X}_k))^2 \qquad . \tag{5}$$

Here we have displayed the dependence on the noise level z, which is hardly ever known in advance in real experiments. The familiar least squares procedure follows by maximizing the likelihood function (5) with respect to the parameters $\boldsymbol{\alpha}$, and does not require any knowledge of z. However the correct procedure is to take the maximum of the posterior distribution for $\boldsymbol{\alpha}$, which according to Bayes' theorem is proportional to the product of the likelihood and the prior distribution. Unless the noise level z is very small it follows (reasonably!) that our prior information about $\boldsymbol{\alpha}$ is relevant. We are aware of global confinement scaling laws for stellarators that have been derived from theory [1], semi-empirically [6], and empirically [2]. We encode this prior knowledge as the distribution

$$p(\boldsymbol{\alpha}|\lambda, M) = (\frac{\lambda}{2\pi})^{N/2}\exp(-\tfrac{1}{2}\lambda T)\ ; \qquad T = \sum_{i=1}^{N}(\alpha_i - m_i)^2 \tag{6}$$

where again a new parameter λ has been introduced which specifies our prior confidence, according to model M, that the N components of $\boldsymbol{\alpha}$ are close to value $\mathbf{m}$. Like z, λ is usually unknown *a priori*. Large values of λ shift the maximum of the posterior (3) towards the *a priori* value $\boldsymbol{\alpha} = \mathbf{m}$; small values for z enforce a closer fit to the data. Both λ and z emerge as by-products of the following analysis, in which the data are used to estimate $\boldsymbol{\alpha}$ simultaneously with the values of these parameters. Accordingly we seek the joint posterior probability distribution $p(\boldsymbol{\alpha}, \lambda, z|\tau, M)$. This is to be synthesised from $p(\tau|\boldsymbol{\alpha}, z, M)$ in (5), $p(\boldsymbol{\alpha}|\lambda, M)$ in (6), and the prior probability distribution for λ and z, $p(\lambda, z|M)$. We are free to include λ as further conditioning information in $p(\tau|\boldsymbol{\alpha}, z, M)$, since it makes no difference, and likewise z as conditioning information in $p(\boldsymbol{\alpha}|\lambda, M)$. From Bayes' theorem and the product rule,

$$p(\boldsymbol{\alpha}, \lambda, z|\tau, M) = \frac{p(\tau|\boldsymbol{\alpha}, \lambda, z, M)p(\boldsymbol{\alpha}|\lambda, z, M)p(\lambda, z|M)}{p(\tau|M)} \tag{7}$$

Using the freedom just mentioned, this gives

$$p(\boldsymbol{\alpha}, \lambda, z|\tau, M) = \frac{p(\tau|\boldsymbol{\alpha}, z, M)p(\boldsymbol{\alpha}|\lambda, M)p(\lambda, z|M)}{p(\tau|M)} \tag{8}$$

which is the desired result, since the denominator is expressible as the numerator integrated over $\boldsymbol{\alpha}, \lambda, z$. We now turn to the prior $p(\lambda, z|M)$. Clearly λ and z are logically independent so that z is redundant conditioning information in $p(\lambda|z, M)$, and by the product rule $p(\lambda, z|M) = p(\lambda|z, M)p(z|M) = p(\lambda|M)p(z|M)$. These distributions are independent also of the physical model M that is used. We will write them as $p(\lambda)$ and $p(z)$, though strictly all probabilities are conditional; there is more prior information which we are not bothering to make explicit.
Let us suppose that we have an estimate of the expectation values $\lambda_0 = \langle\lambda\rangle$ and $z_0 = \langle z\rangle$ [2]. Application of the maximum entropy principle provides, given z_0, the prior distribution function [5]

$$p(z) = \frac{1}{z_0}\exp(-z/z_0) \tag{9}$$

Similarly

$$p(\lambda) = \frac{1}{\lambda_0}\exp(-\lambda/\lambda_0) \tag{10}$$

where we defer any estimate of λ_0. We arrive finally at

$$p(\boldsymbol{\alpha}, \lambda, z|\tau, M) = \frac{p(\lambda)p(z)}{p(\tau|M)}p(\boldsymbol{\alpha}|\lambda, M)p(\tau|\boldsymbol{\alpha}, z, M) \tag{11}$$

where we now have analytical expressions for all the probabilities involved. (Recall that the denominator is just the integral of the numerator.) From this joint posterior we can marginalize out any of $\boldsymbol{\alpha}$, λ or z. Ultimately we are interested in estimates

$$p(\lambda, z|\tau, M) = \frac{p(\lambda)p(z)}{p(\tau|M)}\int d\boldsymbol{\alpha}\ p(\boldsymbol{\alpha}|\lambda, M)p(\tau|\boldsymbol{\alpha}, z, M) \tag{12}$$

then the maximum of the integrand gives the required best estimate for $\boldsymbol{\alpha}$, supposing λ and z are known. We shall find this maximum and then carry out the integration in Gaussian approximation. The resulting posterior for λ and z is very sharply peaked, so that we can accurately substitute the peak values back in to find the best estimate for $\boldsymbol{\alpha}$. Formally, this is an approximation to the procedure of marginalizing over λ and z and then finding the peak of the resulting posterior for $\boldsymbol{\alpha}$.
From (5) and (6), maximizing the integrand with respect to $\boldsymbol{\alpha}$ is equivalent to minimizing

$$\phi \equiv \frac{\lambda}{2}T + \frac{1}{2z}L \tag{13}$$

The minimum falls where $\partial\phi/\partial\alpha_i = 0$; we denote its location $\boldsymbol{\alpha}_0$ and find it in practice by Newtonian iteration. By Taylor expanding $\partial\phi/\partial\alpha_i$ about $\boldsymbol{\alpha}_0$ we have, given an estimate $\boldsymbol{\alpha}^e$ of $\boldsymbol{\alpha}_0$,

$$0 = \frac{\partial\phi}{\partial\alpha_i}|_{\alpha_0} = \frac{\partial\phi}{\partial\alpha_i}|_{\alpha^e} + \sum_j \frac{\partial^2\phi}{\partial\alpha_i\partial\alpha_j}|_{\alpha^e}\Delta\alpha_j + O(\Delta\alpha^2) \tag{14}$$

where $\Delta\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^e$. To leading order in $\Delta\boldsymbol{\alpha}$, substituting (13) for ϕ and (6) for T,

$$(H_{\alpha^e} + \mu I)\Delta\boldsymbol{\alpha} = \mu(\mathbf{m} - \boldsymbol{\alpha}^e) - \frac{1}{2}\frac{\partial L}{\partial \boldsymbol{\alpha}}|_{\alpha^e} \tag{15}$$

where $H_{\alpha_0} \equiv \frac{1}{2}\partial^2 L/\partial\alpha_i\partial\alpha_j|_{\alpha_0}$ is the Hessian matrix of L, I denotes the unit matrix, and $\mu = \lambda z$. Iterative solution of (15) for successive corrections $\Delta\boldsymbol{\alpha}$ provides $\boldsymbol{\alpha}_0$, the maximum of the integrand if μ were known, and the Hessian matrix. A suitable starting estimate $\boldsymbol{\alpha}^e$ is the *a priori* value $\boldsymbol{\alpha}^e = \mathbf{m}$. For $\mu = \lambda z \to \infty$ we obtain $\boldsymbol{\alpha}_0 = \mathbf{m}$, so that the data do not modify our prior knowledge. For $\mu \to 0$, on the other hand, the solution is that of the maximum likelihood problem. The parameter μ is inverse to the distance in $\boldsymbol{\alpha}$-space that the solution travels in going from the prior to the posterior.
Let β_i be the eigenvalues of the Hessian matrix, which we now denote H without the subscript. The condition of H is

$$\frac{\mathrm{cond}(H) \equiv \beta_{max}}{\beta_{min}} \tag{16}$$

while the condition of $H + \mu I$ is

$$\frac{\mathrm{cond}(H + \mu I) = \beta_{max} + \mu}{\beta_{min} + \mu} \tag{17}$$

Therefore $\mathrm{cond}(H + \mu I)$ is always smaller than $\mathrm{cond}(H)$, implying that the solution is less sensitive to the noise introduced with the data in the right hand side of (15). This brings us back to an estimate of $\langle\lambda\rangle$. The larger is $\mathrm{cond}(H)$, the more necessary is the regularization in (15) in order to reduce the risk of overfit. When $\beta_{max}/\beta_{min} \gg 1$ we would expect $\beta_{min} \ll \mu \ll \beta_{max}$. A possible choice is the geometric mean of the eigenvalues:

$$\langle\lambda\rangle = \frac{1}{z_0}\langle\mu\rangle = \frac{1}{z_0}(\prod_i \beta_i)^{1/N} \tag{18}$$

The choice is not critical: changing $\langle\lambda\rangle$ by an order of magnitude changes the results presented in the next section only in the third decimal place.
The assignment of all probabilities is now complete.
Having found the maximum of the integrand in (12), we now perform the integration. In Gaussian approximation only the second order terms in deviation from $\boldsymbol{\alpha}_0$ are retained in the exponent. Their coefficient is $\frac{1}{2}\partial^2\phi/\partial\alpha_i\partial\alpha_j|_{\alpha_0}$, a matrix which was evaluated above; integration of the Gaussian form gives the inverse square root of the corresponding determinant:

$$\int d\boldsymbol{\alpha}\; p(\boldsymbol{\alpha}|\lambda, M)p(\tau|\boldsymbol{\alpha}, z, M) = (\frac{\lambda}{2\pi})^{N/2}(\frac{1}{2\pi z})^{K/2}e^{-\phi(\boldsymbol{\alpha}_0)}\frac{\pi^{N/2}}{\{\det(\lambda I + \frac{1}{z}H)\}^{1/2}} \tag{19}$$

On substituting this result and expressions (9) and (10) for the priors into (12), we determine fully the form of the posterior probability distribution $p(\lambda, z|\tau, M)$ for z, λ. (Only the normalization factor $p(\tau|M)$, which is easy to calculate *a posteriori*, is yet to be found.) The most important feature of this distribution is the location of its maximum, which we now find. Define for convenience

$$\psi(\lambda, z) \equiv \ln p(\lambda, z|\tau, M)p(\tau|M) \tag{20}$$

$$= \mathrm{const} - \frac{\lambda}{\lambda_0} - \frac{z}{z_0} + \tfrac{1}{2}N\ln\lambda - \tfrac{1}{2}K\ln z - \phi(\boldsymbol{\alpha}_0) - \tfrac{1}{2}\ln(\det(\lambda I + \frac{1}{z}H)) \tag{21}$$

or, in terms of μ and z,

$$\psi = \mathrm{const} - \frac{\mu}{\lambda_0 z} - \frac{z}{z_0} + \tfrac{1}{2}N\ln\mu - \tfrac{1}{2}K\ln z - \phi(\boldsymbol{\alpha}_0) - \tfrac{1}{2}\sum_{i=1}^{N}\ln(\mu + \beta_i) \tag{22}$$

where β_i are the eigenvalues of H as before. The maximum of the posterior probability distribution $p(\lambda, z|\tau, M)$ is given by $\partial\psi/\partial\mu = 0$ and $\partial\psi/\partial z = 0$. These equations are [7]

$$\lambda(T + \frac{2}{\lambda_0}) = \sum_{i=1}^{N} \frac{\beta_i}{\mu + \beta_i} \tag{23}$$

and

$$\mu T + L = Kz + 2(\frac{\mu}{\lambda_0} - \frac{z^2}{z_0^2}) \tag{24}$$

These equations determine the location of the maximum, λ_{max}, z_{max} of ψ; in practice we iterate them together with (15). Equation (23) is a slightly modified form for the conventional estimate of the noise level [7]. If there are many more data K than parameters N, and given a well conditioned Hessian matrix H_α for which the right hand side of (22) approaches N, then we recover the ordinary least squares rule that the noise level is equal to $L/(K-N)$. The right hand side of (22) is equal to the number of 'large' eigenvalues of H_{α_0}; that is, to the number of parameters determined by the data essentially regardless of the prior probability. This expression becomes smaller, the less new information that is contained in the data.

Having obtained the maximum of the posterior distribution $p(\lambda, z|\tau, M)$, we now determine its normalization constant, $p(\tau|M)$ from (12); our purpose will emerge shortly. From the definition of ψ in (20) it follows by marginalizing over λ, z that

$$p(\tau|M) = \int d\lambda \int dz \; \exp\psi(\lambda, z) \tag{25}$$

By expanding ψ about its maximum, which is defined to be at $\hat{\mu}, \hat{z}$, we obtain

$$\psi(\mu, z) = \psi(\hat{\mu}, \hat{z}) + \Delta\mathbf{V}^T H_{\hat{\mu}\hat{z}} \Delta\mathbf{V} \tag{26}$$

where $\Delta\mathbf{V} = (\mu - \hat{\mu}, z - \hat{z})$ and $H_{\hat{\mu}\hat{z}}$ is the Hessian of ψ. Since $\psi(\mu, z)$ is sharply peaked about its maximum we may accurately integrate (25) over $(z, \lambda = \mu/z)$ in Gaussian approximation. The result is

$$p(\tau|M) = \exp\psi(\hat{\mu}, \hat{z}) \frac{\pi}{\{\det(-H_{\hat{\mu}\hat{z}})\}^{1/2}} \tag{27}$$

Since

$$p(\tau|M)p(M) = p(\tau, M) = p(\tau)p(M|\tau) \tag{28}$$

we have the result

$$p(M|\tau) = p(M)\frac{p(\tau|M)}{p(\tau)} \tag{29}$$

which can be used to compare models M [8] by taking the ratio

$$\frac{p(M_1|\tau)}{p(M_2|\tau)} = \frac{p(M_1)}{p(M_2)} \frac{p(\tau|M_1)}{p(\tau|M_2)} \tag{30}$$

(Of course the priors $p(M_i)$ are conditional on information we have not written explicitly.) The models compared in the next section are: M_1, the semi-empirical Lackner-Gottardi scaling rewritten for stellarators; M_2, the scaling derived from neoclassical transport theory with the ion plateau; and a third model M_3, the ITER 89-P scaling rewritten for stellarators. Models M_1 and M_2 have the same prior values for the exponents and differ only by the constant factor $\exp(\alpha_c)$. Model M_3 also exhibits different exponents. Since we have no prior preference for either of the three expressions we assign them equal prior probabilities $p(M_1) = p(M_2) = p(M_3) = 1/3$, so that (30) simplifies to the ratio of expressions (25) for the models. This is our reason for evaluating $p(\tau|M)$.

We complete our theoretical analysis by examining the uncertainties in the α_i. Define the integral in (12) as

$$Z \equiv \int d\boldsymbol{\alpha}\ p(\boldsymbol{\alpha}|\lambda, M)p(\tau|\boldsymbol{\alpha}, z, M) \tag{31}$$

which was evaluated in Gaussian approximation at (19):

$$Z \propto [\det(\frac{1}{2}\lambda I + \frac{1}{2\sigma^2}H)]^{-1/2} \tag{32}$$

where we have now replaced z by σ^2 for clarity. The determinant is that of the covariance matrix in the exponent of the integrand, so that

$$\langle \Delta\alpha_i \Delta\alpha_j \rangle = \frac{2\sigma^2}{Z} \partial Z / \partial(\mu I + H)_{ij} \tag{33}$$

On carrying out the differentiation we obtain

$$\langle \Delta\alpha_i \Delta\alpha_j \rangle = \sigma^2\ (\mu I + H)^{-1}_{ij} \tag{34}$$

The inverse matrix in this formula is readily computed along with the solution of (15).

4 Results and Discussion

Our discussion will focus first on the analysis of the global data set comprising ECR, NI and ECR+NI discharges. In Table III we compare parameters obtained by the customary logarithmic fit (logf), and the unrestricted nonlinear least squares fit (nlf), with Bayesian analysis based respectively on the Lackner-Gottardi model (BLG), the ion plateau neoclassical theoretical model (BNC), and the ITER 89-P scaling reformulated for the stellarator (BIT). (The 'B' stands for Bayesian.) Numbers in brackets denote the values for the model parameters. We stress that we estimate the scaling parameters from the data and do not regard them as cast in stone; the state of the art is still semi-empirical! The last column indicates the sum of the squared deviations between the best fit and the data.

From evaluation of (29) the data favor the Lackner-Gottardi model over ITER 89-P by a factor of 37dB, and over the neoclassical model by as much as 21dB. We have adopted the usual decibel measure in which the ratio of two probabilities P_1/P_2 corresponds to $10\log_{10}(P_1/P_2)dB$. For easy use of this log scale remember that 3 db represent a factor of two. The difference between ion plateau neoclassical transport and Lackner-Gottardi lies only in the numerical factor, which arises from theory in the first and rests on the analysis of tokamak confinement in the latter model. The 21dB superiority of the empirically determined LG scaling constant for the present stellarator data set suggests that confinement physics does not differ greatly between tokamaks and stellarators. Major differences arise if we compare the three sets with the logfit results. Even the overall measured L is about 10% worse. The data confirm that the present fitting problem is intrinsically nonlinear. Substantially closer to our 'best' BLG data is the unrestricted nonlinear least squares fit. However we notice major discrepancies in the constant and in the B, a and ι exponents. BLG agrees within error with the model parameters. Next we turn to the analysis of the individual classes of discharges on the basis of the successful Lackner-Gottardi model. The corresponding data are assembled in Table IV. Results from the analysis of the global set are repeated for easier comparison. Interesting features show up in the density dependence. While both NI and ECR data alone suggest dependence as $n^{0.7}$, the global set

predicts $n^{0.58}$. Similarly, the power exponent is larger for the global data set than for either the pure NI or the pure ECR data set.
The reason for this seeming paradox lies in the collinearities of (n, B) in the ECR data and of (n, p) in the NI data which are mutually broken in the global data set.
We close this section with a warning. The subtleties we have encountered with different data sets from a single fusion device arise also in a comparison of the performance of different machines. Our results warn that such comparisons should be performed with Bayesian analysis and with care.

Acknowledgements

We are indebted to Dr. U. Stroth for providing the W7AS data set and for raising our interest in global confinement scaling. We are grateful to Prof. Karl Lackner for providing the ITER 89-P model rewritten for stellarators.

References

[1] Wendelstein Project Group, IPP EURATOM Association, Wendelstein VII-X Application for Preferential Support, 1990

[2] U. Stroth, G. Kühner, H. Maassberg, M. Murakami and R.A. Dory, Report IPPIII/195 (1994)

[3] W.J. Krzanowski, "Principles of Multivariate Analysis", Clarendon Press, Oxford 1988

[4] A.J.M. Garrett, in "Maximum Entropy in Action", eds. B. Buck and V.A. Macaulay, Oxford Science Publications 1991

[5] M. Tribus, "Rational Descriptions, Decisions and Designs", Pergamon Press 1969

[6] K. Lackner and N.A.O. Gottardi, Nuclear Fusion **30** (1990) 767

[7] S.F. Gull, in "Maximum Entropy and Bayesian Methods", ed. J. Skilling, Kluwer Academic Publishers 1989

[8] P.C. Gregory and T.J. Loredo, Astrophysical Journal **398** (1992) 146

Table Captions

Table I: Range of input variables for the three different groups of data.

Table II: Dominant correlation coefficients of ECR, NI data and pooled data.

Table III: Exponents of the scaling function obtained from the customary logarithmic fit (logf), from an unrestricted nonlinear least squares fit (nlf) and from Bayesian analysis based on the Lackner-Gottardi model (BLG), the ion plateau neoclassical model (BNC), and the ITER 89-P model (BIT).

Table IV: Scaling exponents from Bayesian analysis based on the L-G model for the global data set (ECR+NI) and for the individual ECR and NI data groups.

Appendix I

It is tempting to estimate the parameters of power-law problems, as given in (1), by taking the logarithm and invoking linear regression theory. Here we show that this approach leads to a biased estimator for the parameters due to the nonlinear error transform. Consider for the sake of simplicity that we want to fit a function $f(x,\alpha) = x^\alpha$ of a single parameter α to measured data f_i obtained at settings x_i of the independent variable. We assume further that the data are distorted by a constant relative error (this is physically unrealistic but convenient for the following proof). Let $\hat{f}_i = f(x_i, \hat{\alpha})$ be the "true" values obtained from evaluating f with the "true" value $\hat{\alpha}$ for the parameter α, then

$$f_i = \hat{f}_i + \varepsilon_i \hat{f}_i \qquad \text{with } \varepsilon_i = N(0,\sigma) \qquad . \tag{35}$$

Taking the logarithm yields

$$\ln f_i = \ln(\hat{f}_i(1+\varepsilon_i)) = \ln \hat{f}_i + \ln(1+\varepsilon_i) \qquad . \tag{36}$$

Let $\overline{(\ln(1+\varepsilon)^2)} = \eta^2$. The least squares solution is obtained upon minimizing

$$\phi = \frac{1}{\eta^2} \sum_i (\ln f_i - \alpha \ln x_i)^2 \tag{37}$$

which yields

$$\Delta\alpha = \overline{(\ln(1+\varepsilon)\ln x)} / \overline{(\ln x)^2} \qquad . \tag{38}$$

The expectation value in the numerator can be broken apart since ε_i and x_i are uncorrelated and we have

$$\Delta\alpha = \overline{\ln(1+\varepsilon)}\, \frac{\overline{\ln x}}{\overline{(\ln x)^2}} \qquad . \tag{39}$$

But now the random variable $\ln(1+\varepsilon_i)$ has no longer zero mean

$$\ln(1+\varepsilon_i) = \overline{(\varepsilon - \frac{1}{2}\varepsilon^2 + \ldots)} \leq -\frac{1}{2}\sigma^2 \qquad . \tag{40}$$

We find that $\Delta\alpha$ does not converge to zero in the limit of large samples but is biased towards smaller values by

$$\Delta\alpha \leq -\frac{1}{2}\sigma^2 |\overline{\ln x} / \overline{(\ln x)^2}| \tag{41}$$

An unbiased estimator is only obtained upon minimizing the non-linear problem

$$\phi(\alpha) = \frac{1}{2\sigma^2} \sum_i \frac{(f_i - f(x_i,\alpha))^2}{\hat{f}_i^2} \qquad . \tag{42}$$

We will prove that for a sufficiently large data sets the estimator for α approaches the true value $\hat{\alpha}$. It is therefore sufficient to expand $f(x_i,\alpha)$ up to second order in $\Delta\alpha = \alpha - \hat{\alpha}$. Upon differentiating with respect to $\Delta\alpha$ and keeping only terms in leading order we obtain

$$\Delta\alpha = \overline{(\varepsilon \hat{f}'/\hat{f})} / \left(\overline{(\hat{f}'/\hat{f})^2} + \overline{(\varepsilon \hat{f}''/\hat{f}i)} \right) \qquad . \tag{43}$$

Since the data $\hat{f}_i$ and the errors ε_i are uncorrelated

$$\overline{(\varepsilon \hat{f}'/\hat{f})} \to \overline{\varepsilon}\, \overline{\hat{f}'/\hat{f}} \to 0 \qquad , \tag{44}$$

for sufficiently large data sets and viz $\alpha \to \hat{\alpha}$.

Appendix II

In this paper we dealt with a data set comprising three groups characterized by different ways of plasma heating. In this appendix we present the proper covariance matrix for pooled data pairs originating from different experimental sources. Consider the covariance of any two variables (u, v) over the whole set of N data pairs

$$(N-1)\mathrm{cov}(u,v) = \sum_{\alpha=1}^{g} \sum_{i=1}^{N_\alpha} (u_{\alpha,i} - \overline{u})(v_{\alpha,i} - \overline{v}) \qquad , \tag{45}$$

where the first sum runs over the different groups of data pairs stemming from g types of experiments. The second index i enumerates the data pairs within each group. $\overline{u}$ ($\overline{v}$) represents the average value of u (v) over all groups

$$\begin{aligned} \overline{u} &= \frac{1}{N} \sum_{\alpha,i} u_{\alpha,i} = \frac{1}{N} \sum_{\alpha=1}^{g} N_\alpha \overline{u_\alpha} \\ \overline{u_\alpha} &= \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} u_{\alpha,i} \qquad , \end{aligned} \tag{46}$$

while $\overline{u_\alpha}$ ($\overline{v_\alpha}$) is the "group average" of u (v) in the α^{th} group. With the help of the group averages (45) may be rewritten as

$$(N-1)\mathrm{cov}(u,v) = \sum_{\alpha=1}^{g} \sum_{i=1}^{N_\alpha} (u_{\alpha,i} - \overline{u_\alpha})(v_{\alpha,i} - \overline{v_\alpha}) + \sum_{\alpha=1}^{g} N_\alpha (\overline{u_\alpha} - \overline{u})(\overline{v_\alpha} - \overline{v}) \qquad . \tag{47}$$

The covariance of the pooled data (45) consists therefore of two contributions: the first term in (47) describes the average of covariances within the individual groups while the second term describes the covariances of the group averages. In fact, in the data set at hand the average density in the ECR group is lower by more than a factor of 4 than that of the NI group. Consequently this term really matters in the present case. It is, however, entirely uninteresting since it depends on translations. The covariance transforms under a rigid translation of the groups $u_{\alpha,i} \to u_{\alpha,i} + a_\alpha$ into

$$(N-1)\mathrm{cov}(u,v) \to \sum_{\alpha=1}^{g} \sum_{i=1}^{N_\alpha} (u_{\alpha,i} - \overline{u_\alpha})(v_{\alpha,i} - \overline{v_\alpha}) + \sum_{\alpha=1}^{g} N_\alpha \Big(\overline{u_\alpha} - \overline{u} + a_\alpha - \frac{1}{N} \sum_{\alpha'} N_{\alpha'} a_{\alpha'}\Big)(\overline{v_\alpha} - \overline{v}) \qquad . \tag{48}$$

Obviously, the intra-group covariances are invariant under rigid translations of the groups while, in general, the inter-group covariances are affected if the relative distances between the groups are modified, i.e. $a_\alpha \neq$ const. Correlation and principal component analyses have therefore to be based on the average intra-group covariance matrix in order to be meaningful, since the inter-group covariance will be dominated by external experimental effects which have nothing to do with the intrinsic u-v-dependence.

	Heating	P/MW	$\overline{n}_e/10^{19}m^{-3}$	B/T	ι_a	a/m
5	ECR+NI	0.42 – 1.11	2.25 – 6	2.53 – 2.56	0.35 – 0.5	0.16 – 0.18
143	ECR only	0.51 – 2	0.42 – 8.4	1.31 – 2.57	0.31 – 0.59	0.12 – 0.19
51	NI only	0.35 – 3	3.17 – 23.3	1.10 – 2.59	0.32 – 0.63	0.12 – 0.18

Table I

	ECR	NI	global
$r(n,B)$	-0.599	-0.203	-0.449
$r(n,P)$	-0.070	-0.649	-0.114
$r(B,a)$	-0.156	-0.352	-0.198
$r(a,\iota)$	-0.294	-0.430	-0.335

Table II

.

	α_c	α_n	α_B	α_P	α_a	α_ι	L
logf.	-1.509	0.527	0.718	-0.563	1.846	0.294	$1.238\,10^{-3}$
	±0.277	±0.029	±0.056	±0.034	±0.123	±0.085	
nlf.	-1.417	0.604	0.671	-0.652	2.073	0.665	$1.105\,10^{-3}$
	±0.421	±0.035	±0.086	±0.033	±0.187	±0.107	
BLG	-1.755	0.590	0.728	-0.637	1.982	0.529	$1.118\,10^{-3}$
	±0.082	±0.030	±0.059	±0.029	±0.049	±0.058	
	(-1.75)	(0.6)	(0.8)	(-0.6)	(2.0)	(0.4)	
BNC	-2.210	0.581	0.759	-0.635	1.770	0.485	$1.130\,10^{-3}$
	±0.082	±0.030	±0.059	±0.029	±0.049	±0.058	
	(-2.408)	(0.6)	(0.8)	(-0.6)	(2.0)	(0.4)	
BIT	-1.367	0.584	0.738	-0.637	1.901	0.608	$1.112\,10^{-3}$
	±0.233	±0.034	±0.080	±0.032	±0.108	±0.084	
	(-1.668)	(0.1)	(1.05)	(-0.5)	(2.0)	(0.85)	

Table III

.

	α_c	α_n	α_B	α_P	α_a	α_ι
ECR+NI	-1.755 $\pm$0.082	0.590 $\pm$0.030	0.728 $\pm$0.059	-0.637 $\pm$0.029	1.982 $\pm$0.049	0.529 $\pm$0.058
ECR	-1.64 $\pm$0.121	0.722 $\pm$0.042	0.688 $\pm$0.074	-0.614 $\pm$0.032	1.870 $\pm$0.067	0.525 $\pm$0.075
NI	-1.779 $\pm$0.117	0.700 $\pm$0.068	0.757 $\pm$0.092	-0.543 $\pm$0.059	2.057 $\pm$0.074	0.396 $\pm$0.092

Table IV