Dimensionally constrained energy confinement analysis of W7-AS data

V. Dose, R. Preuss, and W. von der Linden

Max-Planck-Institut für Plasmaphysik, EURATOM Association

D-85748 Garching b. München, Germany

(February 12, 1998)

Abstract

A recently assembled W7-AS stellarator database has been subject to dimensionally constrained confinement analysis. The analysis employs Bayesian inference. Dimensional information is taken from the Connor-Taylor similarity transformation theory which provides six possible physical scenarios with associated dimensional conditions. Bayesian theory allows to calculate the probability for each model and we find that the present W7-AS data are most likely described by the collisionless high-beta case. Probabilities for all models and the associated exponents of a power law scaling function are presented.

I. INTRODUCTION

The description of plasma energy confinement via global scaling expressions dates back to the seventies [1,2]. Numerous scaling relations have been published since derived from ever increasing data sets. From the experimental point of view scaling laws are useful for comparing different confinement regimes and for inter-machine performance comparison. The functional form of the scaling law is unknown. Initial assumption of a power law dependence of the plasma energy content on variables as density, magnetic field, geometry scale and others has survived up to now because of convenience and simplicity. In fact in a log linear form the tempting beauties of linear algebra can be recovered. The remarkable persistence of these simple expressions is also due to the good job that several of them, derived from small and mid sized machines, have done in predicting the performance of much larger devices. In fact W7-X and ITER performance predictions rest on extrapolation of currently available power law scaling expressions. Uncertainties in the exponents of the variables have recently raised a discussion [3] on ITER performance prediction. These should not have come as a surprise to the community since already Riedel [4] has pointed out, that the equations, from which the exponents are derived, are severely ill conditioned. This means that one or more eigenvalues of the matrix are very small compared to the trace. In such a case a large fraction of the associated eigenvectors may be added to the exponent solution vector $\vec{\alpha}$ with only minute sacrifices in the goodness of fit. We note in passing that the incorporation of ever more variables in the global scaling expression drives this situation worse and needs suitable attention. Despite of their relative success in the past, there is no reason to assume that power scaling laws represent the relevant plasma physics [5] and other relations like offset linear scaling have been shown to perform equally well [6].

In addition to providing engineering guide lines for next generation machine construction, confinement scaling studies "have the potential of giving critical information for understanding the underlying nature of radial transport" [5]. In fact, attempts to relate observed scaling laws to basic and strict physical constraints dates back to the mid seventies [7]. In particular Connor and Taylor [8] explored the invariances of the basic equations of plasma behavior under similarity transformations and derived constraints on exponents of power law scaling expressions characteristic for different basic physics assumptions. Comparison of their relations with the then known empirical results of Hugill and Sheffield [1] showed "that the values (of exponents) are incompatible with any of the plasma models (treated in this paper) and that the discrepancies exceed quoted uncertainties". This situation has largely improved with the currently available much more precise data. Consequently dimensionally exact scaling studies have been performed in the last years. We mention in particular the work of Christiansen *et al.* [6]. These authors imposed the dimensional constraint, resulting from the Connor/Taylor invariance argument as a stiff auxiliary condition in their regression analysis and tried to infer the physical model represented by the data adopting the F-test of classical statistics. The essential quantity in this test is the difference between the mean square error of a free and a dimensionally exact fit. As a result they found that in their "subset of data" exponents of a power law could always be arranged to satisfy the 3 degrees of freedom collisional high-β model in the Connor/Taylor terminology. Results on hierarchically lower models with 2 and 1 degrees of freedom could not be obtained. This is the purpose of the present paper. Our aim is to provide and demonstrate a procedure which quantifies the extent to which (the presently used) W7-AS data follow a particular model out of the six choices offered by Connor and Taylor. The data set consists of the W7-AS part of the recently published international stellarator data base [9]. We shall not require the power law exponents to satisfy the Connor/Taylor [CT] relations exactly. Instead, after construction of an appropriate measure, we shall investigate how closely the data satisfy a particular model. We consider this to be a better suited strategy in view of the fact that we have no rigorous reason at all to assume that plasma behavior may be adequately represented by a power law. We therefore call our results dimensionally constrained as opposed to dimensionally exact. Let us mention in this context that Connor and Taylor never claimed a single power law term whose exponents satisfy dimensional constraints as a function representing plasma confinement. In fact they express a general dimensionless function in a series of dimensionally exact power law terms. We are currently exploring this route.

In the subsequent analysis we shall assume that energy confinement in W7-AS may be described in form of a simple power law term. We shall then decide to what extent this single power law term satisfies the CT relations. We shall finally employ Bayesian inference [10] in order to derive for each CT model exponent sets which approach the CT dimensional relation as close as possible at an as small as possible expense in data fit. Bayesian theory allows further and most importantly to quantify the probability of each of the six CT models in the light of the data and therefore answers for the first time in the by now long history of confinement analysis the question which physical model explains the given data best. Since we assume that this probability theory is not common knowledge among the readers of nuclear fusion we shall devote an entire chapter to derive and explain the ideas and procedures. On the other hand, readers who are familiar with Bayesian inference may of course skip this section and proceed directly to chapter III.

II. BAYESIAN INFERENCE

Statistical inference in the Bayesian formulation is a calculus based on two simple axioms: the sum rule

$$p(H_i|I) + p(\overline{H_i}|I) = 1 \tag{1}$$

and the product rule

$$p(H_i, D|I) = p(H_i|I)p(D|H_i, I) \qquad . \tag{2}$$

We consider hypotheses, which we might have reason to formulate in the light of some background information I. The sum rule simply states that the probability that a particular hypothesis H_i is true plus the probability that the negation $\bar{H}_i$ of H_i is true add up to one. Similarly, the product rule states that the probability for H_i and D being true given the background information I may be expressed as the probability for H_i being true conditional on I times the probability for D given that H_i is true. Sum and product rules may be written in a slightly different way. In the course of this work we shall be dealing with mutually exclusive hypotheses so that if one particular hypothesis is true then all the others are false. For such a case the sum rule generalizes to

$$p(\sum_i H_i|I) = \sum_i p(H_i|I) = 1 \quad . \tag{3}$$

We call this the normalization rule. Reformulation of the product rule follows from an alternative expansion of (2). Due to the symmetry in the arguments (H_i, D) we may also write

$$p(H_i, D|I) = p(D|I)p(H_i|D, I) \quad . \tag{4}$$

Combining (2) and (4) we arrive at Bayes' theorem

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)} \quad . \tag{5}$$

Bayes' theorem constitutes a recipe of learning. The probability $p(H_i|I)$ is called the prior probability for H_i, that is the probability which we assign to H_i being true before an experiment designed to test H_i has been performed and new data D become available. $p(H_i|D, I)$ is correspondingly the posterior probability for H_i being true in the light of the newly acquired data D. $p(D|H_i, I)$ is the probability for the data, given that H_i is true and is normally called the likelihood. Finally, $p(D|I)$ is the global likelihood for the entire class of hypotheses and is usually called the evidence. $p(D|I)$ is not independent of the other probabilities but may be obtained by application of the sum and product rule:

$$p(D|I) = \sum_i p(H_i, D|I) = \sum_i p(H_i|I)p(D|H_i, I) \quad . \tag{6}$$

In this form $p(D|I)$ is the sum over all possible numerators in (5) and thus has the meaning of a normalization constant.

Bayes' theorem leads us directly to the solution of the important problem of model comparison. If we associate with H_i the six models resulting from the Connor/Taylor invariance relations and with D the confinement data set of a particular experiment then our prominent goal will be to determine

$$\frac{p(H_i|D,I)}{p(H_k|D,I)} = \frac{p(H_i|I)}{p(H_k|I)} \frac{p(D|H_i,I)}{p(D|H_k,I)} = \frac{p(H_i|I)}{p(H_k|I)} B_{ik} = O_{ik} \qquad . \tag{7}$$

The ratio of posterior probabilities $p(H_i|D,I)$ and $p(H_k|D,I)$ is also called the odds ratio O_{ik}. Note that $p(D|I)$ has dropped out. The odds ratio factors into the prior odds ratio and the Bayes factor B_{ik}. The prior odds ratio is in many cases equal to one. This is because our experiment is usually performed in order to reveal whether we should prefer $p(H_i|D,I)$ or $p(H_k|D,I)$ and consequently a natural prior state of knowledge is that we have no reason to favor one model against another. We shall adopt this standpoint in treating the Connor/Taylor (CT) models. Having calculated the odds ratios O_{k1} of all models k with respect to model 1 we can then find the posterior probability for each model from

$$p(H_k|D,I) = \frac{O_{k1}}{\sum_j O_{j1}} \qquad , \tag{8}$$

where the sum is over all models and of course $O_{11} = 1$. We postpone a further discussion of the model comparison problem to the end of this paragraph and turn meanwhile to the parameter estimation problem. The parameters in our particular problem are of course the exponents of the power law scaling expression. So within each model we have a further class of hypotheses which is not given by discrete numbers but rather by the continuous values of a set of parameters $\vec{\xi}$ which characterize model H. The previous probabilities change accordingly to probability densities such that

$$p(\vec{\xi}|H,I)d\vec{\xi} \tag{9}$$

is the probability that the true values of the parameter set lies in the interval $[\vec{\xi}, \vec{\xi} + d\vec{\xi}]$. The above product and sum rules hold for densities with all sums replaced by integrals. The previous normalization (6) becomes accordingly

$$p(D|H,I) = \int d\vec{\xi}\ p(\vec{\xi}|H,I)\ p(D|\vec{\xi},H,I) \qquad . \tag{10}$$

We shall use (10) also in the form

$$p(D,\vec{\xi}_1|H,I) = \int d\vec{\xi}_2 \; p(\vec{\xi}_2|H,I) \; p(D,\vec{\xi}_1|\vec{\xi}_2,H,I) \qquad , \tag{11}$$

which we call marginalization. Marginalization is an exceedingly important ingredient of Bayesian inference. It allows us to eliminate parameters from our calculation which are essential but whose particular values are uninteresting. Such parameters are also called nuisance parameters. We shall make extensive use of the marginalization technique in the rest of this paper. We note in passing that marginalization constitutes the probably most important advantage of Bayesian inference over standard frequentists' statistics.

We shall now continue and finish the discussion of model comparison which is vital for the topic of this paper. Consider in particular two models H_1 with a single parameter and H_2 with two parameters. You might associate H_1 with the collisionless low-β and H_2 with either the collisionless high-β or the collisional low-β CT models. In order to evaluate (7) we have need to calculate (10). We follow Gregory and Loredo [11] in a qualitative evaluation of (10). Let us denote by $\hat{\vec{\xi}}$ the parameter vector for which the likelihood $p(D|\vec{\xi},H,I)$ attains its maximum. We now assume that the likelihood function is sharply peaked as compared to the prior $p(\vec{\xi}|H,I)$. This is the normal case, where the measurement contains information considerably more detailed than the prior. The integral in (10) may consequently be approximated by

$$p(D|H_k,I) \approx p(\hat{\vec{\xi}}|H_k,I) \int d\vec{\xi} \; p(D|\vec{\xi},H_k,I) \qquad . \tag{12}$$

If the parameter space $\vec{\xi}$ is k-dimensional then (12) may be further approximated by

$$p(D|H_k,I) \approx p(\hat{\vec{\xi}}|H_k,I) p(D|\hat{\vec{\xi}},H_k,I) \, (\Delta\xi)^k \qquad . \tag{13}$$

where $\Delta\xi$ specifies the "width" of the likelihood-function. From the normalization requirement for the prior,

$$1 = \int d\vec{\xi} \; p(\vec{\xi}|H_k,I) = p(\hat{\vec{\xi}}|H_k,I) \, (\delta\xi)^k \qquad , \tag{14}$$

we finally obtain

$$p(D|H_k, I) \approx p(D|\hat{\vec{\xi}}, H_k, I) \left(\frac{\Delta\xi}{\delta\xi}\right)^k \quad . \tag{15}$$

Now we are ready to discuss the Bayes factor B_{kj} in (7) for two hierarchical models H_k and H_j with k and j parameters respectively and $j < k$,

$$B_{kj} = \frac{p(D|\hat{\vec{\xi}}, H_k, I)}{p(D|\hat{\vec{\xi}}, H_j, I)} \left(\frac{\Delta\xi}{\delta\xi}\right)^{k-j} \quad . \tag{16}$$

Since $j < k$, it follows strictly that $p(D|\hat{\vec{\xi}}, H_j, I) \leq p(D|\hat{\vec{\xi}}, H_k, I)$ since H_k contains more parameters than H_j and all parameters of H_j are contained in the set of H_k and more free parameters allow a better fit to the data. The first ratio in (16) is therefore always ≥ 1. We then turn to the ratio of volume elements. We consider the normal case that our measurement sharpens appreciably our prior knowledge. The amount of uncertainty of our prior knowledge is expressed as the range $\delta\xi$ over which the prior is appreciably different from zero. For reasonably accurate data we then have always $\Delta\xi \ll \delta\xi$. In the subsequent specific calculations the ratio will turn out to be of the order 10^{-3}. The second term in (16) is therefore always < 1 and becomes even smaller as the difference in dimension of model k and model j increases. This parameter space factor penalizes consequently the inclusion of more parameters and is termed Occam's razor. In order that the Bayes factor becomes larger than one, e.g. the probability for the complexer model k exceeds that of the simpler model j, it is necessary that an increase in the likelihood ratio overcompensates the penalty in parameter space volume.

III. THE LIKELIHOOD FUNCTION

Let W_i be the i-th measurement of the plasma energy content of a toroidal magnetic confinement device. Similarly let n_i, B_i, P_i and a_i be density, magnetic field, heating power and minor radius. For the data we are treating in this paper there is a further dimensionless variable ι_i, the rotational transform. The data set consists of N=250 measurements and has been described in detail in [9]. Our summary presented in Table I is drawn from this work. We then wish to model the plasma energy content W_i as

$$W_i = e^{\alpha_c} n_i^{\alpha_n} B_i^{\alpha_B} P_i^{\alpha_P} a_i^{\alpha_a} \iota_i^{\alpha_\iota} + \varepsilon_i \qquad . \tag{17}$$

The constant factor has been expressed as e^{α_c} for calculational convenience. Let us further abbreviate the power law expression as $f(\vec{\alpha}^*, t_i)$ where $\vec{\alpha}^*$ is the exponent vector and t_i summarizes the variables $n_i, B_i, P_i, a_i, \iota_i$. We assume a Gaussian error statistics with constant absolute error leading to [12]

$$p(\vec{W}|\vec{\alpha}^*, \omega, I) = \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N} (W_i - f(\vec{\alpha}^*, t_i))^2\right\} \qquad . \tag{18}$$

The hyperparameter ω is related to the noise level by $\langle\varepsilon^2\rangle = 1/\omega$. As compared to (10) we identify the data vector $\vec{W}$ with D. A comment on the choice of variables is in order at this place. Of course from the plasma energy content W the confinement time is derived as $\tau = W/P$, and (17) is usually written in terms of confinement times τ_i. We prefer to use (17) in terms of the directly measured quantity W. The two versions are exactly equivalent in the case that we deal with the logarithms of equation (17) as is usually done. In this paper we shall deal with the basic nonlinear form of (17) since it's solutions are bias free [13]. In this case it does matter whether τ or W is chosen as the dependent variable since division of (17) by P introduces error correlations in the measurements of left and right hand side quantities.

Returning to (17) we see that the vector $\vec{\alpha}^*$ may be partitioned into $\vec{\alpha} = (\alpha_n, \alpha_B, \alpha_P, \alpha_a)$ and (α_c, α_ι). While the components of $\vec{\alpha}$ will enter our later dimensional constraints, the particular values of α_c and α_ι are uninteresting in the frame of the present analysis. The Bayesian way to get rid of them is to marginalize over. From the product rule we have

$$p(\vec{W}, \alpha_c, \alpha_\iota|\vec{\alpha}, \omega, I) = p(\alpha_c, \alpha_\iota|I) p(\vec{W}|\alpha_c, \alpha_\iota, \vec{\alpha}, \omega, I) \qquad , \tag{19}$$

and from the sum rule it follows

$$p(\vec{W}|\vec{\alpha}, \omega, I) = \int d\alpha_\iota d\alpha_c \; p(\alpha_c, \alpha_\iota|I) p(\vec{W}|\alpha_c, \alpha_\iota, \vec{\alpha}, \omega, I) \qquad . \tag{20}$$

Assuming complete ignorance about the values of α_c and α_ι the appropriate prior probability distribution is a uniform distribution over large but finite intervals V_c and V_ι, respectively [14]. With this assignment we can proceed to evaluate the integral in (20). To perform this integration directly is impossible and we therefore use the method of steepest descent. Define

$$\tilde{\chi}^2 = \sum_{i=1}^{N} \left(W_i - f(\vec{\alpha}^*, t_i)\right)^2 \qquad . \tag{21}$$

Then up to second order terms

$$\tilde{\chi}^2 = \tilde{\chi}_o^2 + \sum_k \frac{\partial \tilde{\chi}^2}{\partial \alpha_k^*} \Delta\alpha_k^* + \frac{1}{2} \sum_{k,l} \frac{\partial^2 \tilde{\chi}^2}{\partial \alpha_k^* \partial \alpha_l^*} \Delta\alpha_k^* \Delta\alpha_l^* \qquad , \tag{22}$$

with

$$\Delta\alpha_k^* = \alpha_k^* - \alpha_k^{*m} \qquad . \tag{23}$$

Now let α_k^{*m} denote the position of the minimum of $\tilde{\chi}^2$. Numerical values for α_k^{*m} together with the associated errors are collected in table II. At this point the gradient term in (22) vanishes and

$$\tilde{\chi}^2 = \tilde{\chi}_o^2 + \frac{1}{2} \sum_{k,l} \tilde{H}_{kl} \Delta\alpha_k^* \Delta\alpha_l^* \qquad . \tag{24}$$

Comparison of (22) and (24) defines the elements of the Hesse matrix $\tilde{H}$. We shall now separate the quadratic form (24) into one part containing the variables of $\vec{\alpha}$ and another containing α_c and α_ι only. To this end we partition $\tilde{H}$ into

$$\tilde{H} = \begin{pmatrix} U_o & V \\ V^T & W_o \end{pmatrix} \qquad , \tag{25}$$

and correspondingly $\Delta\vec{\alpha^*}^T = (\Delta\vec{\alpha}, \vec{y})$, with $\vec{y} = (\Delta\alpha_c, \Delta\alpha_\iota)$. This yields

$$\Delta\vec{\alpha^*}^T \tilde{H} \Delta\vec{\alpha}^* = \Delta\vec{\alpha}^T U_o \Delta\vec{\alpha} + 2\vec{y}^T V^T \Delta\vec{\alpha} + \vec{y}^T W_o \vec{y} \qquad . \tag{26}$$

The integral (20) now becomes

$$p(\vec{W}|\vec{\alpha},\omega,I) = \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}} \frac{1}{V_c}\frac{1}{V_\iota} \exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \exp\left\{-\frac{\omega}{2}\Delta\vec{\alpha}^T U_o \Delta\vec{\alpha}\right\} \cdot \int d\vec{y}\ \exp\left\{-\frac{\omega}{2}\left(\vec{y}^T W_o \vec{y} + 2\vec{y}^T V^T \Delta\vec{\alpha}\right)\right\} \quad . \tag{27}$$

Next we complete the square in the integrand of (27). With V_c and V_ι sufficiently large, such that the finite limits may be replaced by infinite limits of integration, we obtain the marginal distribution

$$p(\vec{W}|\vec{\alpha},\omega,I) = \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}-1} \frac{1}{V_c V_\iota} \frac{1}{\sqrt{\det(W_o)}} \exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \exp\left\{-\frac{\omega}{2}\Delta\vec{\alpha}^T H \Delta\vec{\alpha}\right\} \quad . \tag{28}$$

with

$$H = U_o - V W_o^{-1} V^T \quad , \tag{29}$$

the reduced Hesse matrix, which is of central importance for the rest of this paper.

IV. DIMENSIONAL CONSTRAINTS

The invariance principles invoked by Connor and Taylor [8] dictate that if the equations describing the plasma physics in a toroidal magnetic confinement device remain invariant under similarity transformations, then any transport quantity derived from them no matter whether the calculation is tractable or not must exhibit the same scale invariance. In terms of the notation of the present paper the CT results can be summarized as

$$W \propto n a^4 R B^2 \left(\frac{P}{n a^4 R B^3}\right)^p \left(\frac{n}{a^3 B^4}\right)^q \left(n a^2\right)^r \quad , \tag{30}$$

with exponents p, q, r as given in table III. In addition to the previously defined quantities the major radius of the device R shows up in (30). Since we consider data of a single machine in this paper, the R dependence of (30) can be absorbed in the proportionality. If we alternatively write

$$W \propto n^{\alpha_n} B^{\alpha_B} P^{\alpha_P} a^{\alpha_a} \quad , \tag{31}$$

comparison of exponents between (30) and (31) leads to linear relationships

$$\vec{\alpha} = \vec{c} + L\vec{x} \qquad . \tag{32}$$

For the example of a collisionless high-β plasma we have explicitly

$$\begin{aligned} \alpha_n &= 1 - x - z \\ \alpha_B &= 2 - 3x \\ \alpha_P &= x \\ \alpha_a &= 4 - 4x - 2z \end{aligned} \qquad \vec{c} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 4 \end{pmatrix} \qquad L = \begin{pmatrix} -1 & -1 \\ -3 & 0 \\ 1 & 0 \\ -4 & 2 \end{pmatrix} \qquad . \tag{33}$$

Depending upon the CT model the number of degrees of freedom $N_{\rm P}$ reduces from originally $M = 4$ exponents α_i to one, two or three. The corresponding parameter vectors are $\vec{x} = x$, $\vec{x} = (x, y)^T$ or $\vec{x} = (x, y, z)^T$. In (32) the new parameters in $\vec{x}$ can be eliminated in favor of $N_{lr} = M - N_{\rm P}$ (see table III) linear relationships for the exponents $\vec{\alpha}$.

$$d_\nu = \hat{b}_\nu^T \vec{\alpha} \qquad , \qquad \nu = 1 \ldots N_{lr} \qquad . \tag{34}$$

The equations represent normal forms of hyperplanes. The normal vectors $\hat{b}_\nu$ form an orthonormal set of vectors. Equation (34) may be used to construct a measure of how close a particular exponent vector $\vec{\alpha}$ falls to the dimensionally correct form of a particular plasma model:

$$Q = \sum_{\nu=1}^{N_{lr}} \left(\hat{b}_\nu^T \vec{\alpha} - d_\nu \right)^2 \qquad . \tag{35}$$

The second column in table IV gives Q/N_{lr} for the components of $\vec{\alpha}^m$ at the maximum of the likelihood function. Our Q-measure (35) shows that the maximum likelihood vector $\vec{\alpha}^m$ falls closest to the collisionless high-β model. Models 1, 2 and 4 exhibit similarly small Q values while the two fluid models lie far from the maximum likelihood solution $\vec{\alpha}^m$. We may also ask the more detailed question of how far the individual exponents (α_i^m) are away from the dimensionally exact case. To this end we seek the smallest deviations $\Delta\alpha_i = \alpha_i - \alpha_i^m$ from the maximum likelihood solution which fulfill the dimensional constraints (34). This is achieved upon minimizing

$$Q^* = \sum_{i=1}^{M} \Delta\alpha_i^2 + \sum_{\nu=1}^{N_{lr}} \lambda_\nu \left(\hat{b}_\nu^T \vec{\alpha} - d_\nu \right)^2 \qquad , \tag{36}$$

where λ_ν are Lagrange parameters which are determined such that (34) is fulfilled.

Since minimization of (36) corresponds to the solution of the underdetermined system of equation (34) in singular value approximation we call $\Delta\alpha_i$ the singular value deviates. The $\Delta\alpha_i$ need not be small here and will in fact turn out to be of comparable magnitude as $\vec{\alpha}^m$ for the two fluid models. They are given in columns 3 through 6 in table IV. As an alternative to this procedure the dimensional constraints have been rigidly imposed in the determination of a constrained maximum likelihood solution and the corresponding χ_o^2 [6] has been used as an indicator of how well a particular model is represented by the data. This procedure does only rule out the fluid models but fails to discriminate convincingly between the four kinetic models for the present data. Both procedures are however ad hoc. There is no reason why we should expect the Q-measure (35) to be zero for exponent vectors $\vec{\alpha}$ derived from noisy measurements even if the physics was correctly modeled by a single dimensionally exact power law term. Likewise, if we choose to rigidly impose the dimensional constraints already in minimizing χ^2 the noise on the measurement will always lead to the smallest χ^2 for the most flexible model, even if the physics was correctly described by a simpler one.

V. DIMENSIONALLY CONSTRAINED SCALING

In the previously derived likelihood function the data $\vec{W}$ were conditional on $\vec{\alpha}$. We are now going to distinguish between several choices for $\vec{\alpha}$ according to the different models specified in the preceding paragraph. Accordingly we are dealing in the following with

$$p(\vec{W}|\vec{\alpha}, H_j, \omega, I) \qquad . \tag{37}$$

Our first and prominent goal is of course the calculation of $p(\vec{W}|H_j, I)$ in order to perform the model comparison (7). In order to proceed we invoke the sum rule to obtain

$$p(\vec{W}|H_j, \omega, I) = \int d\vec{\alpha}\; p(\vec{W}, \vec{\alpha}|H_j, \omega, I) \qquad , \tag{38}$$

which we transform using the product rule to

$$p(\vec{W}|H_j,\omega,I) = \int d\vec{\alpha}\; p(\vec{\alpha}|H_j,I)p(\vec{W}|\vec{\alpha},H_j,\omega,I) \qquad . \tag{39}$$

We have again dropped the ω condition in $p(\vec{\alpha}|H_j,\omega,I)$ because of logical independence: knowledge of the noise level of the likelihood function does not add to our prior knowledge on $\vec{\alpha}$ which is purely dimensional. We assume once more that equation (32) constitutes testable information, in particular that

$$\langle||\vec{\alpha}-\vec{c}-L\vec{x}||^2\rangle = \lambda^{-1} \qquad . \tag{40}$$

Here $\vec{c}$ and matrix L depend upon the model H_j under consideration. Then, by the principle of maximum entropy [12],

$$p(\vec{\alpha}|H_j,\lambda,\vec{x},I) = \left(\frac{\lambda}{2\pi}\right)^{\frac{M}{2}} \exp\left\{-\frac{\lambda}{2}||\vec{\alpha}-\vec{c}-L\vec{x}||^2\right\} \qquad . \tag{41}$$

We insert this prior and the likelihood function (28) in (39) and perform the (M=4)-dimensional Gaussian integral. After some elementary though tedious algebra we obtain

$$\begin{aligned} p(\vec{W}|H_j,\omega,\lambda,\vec{x},I) = &\left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}-1}\left(\frac{\lambda}{2\pi}\right)^{\frac{M}{2}} \frac{1}{V_c V_\iota}\frac{1}{\sqrt{\det(W_o)}} \exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \\ &\cdot\frac{(2\pi)^{\frac{M}{2}}}{\sqrt{\det(\lambda \mathrm{I}+\omega H)}}\cdot\exp\left\{-\frac{\lambda}{2}\Delta\vec{l}^T R\Delta\vec{l}\right\} \qquad , \end{aligned} \tag{42}$$

with

$$\Delta\vec{l} = \vec{\alpha}^m - \vec{c} - L\vec{x} = \Delta\vec{c} - L\vec{x} \tag{43}$$

and

$$R = \mathrm{I} - \lambda\left(\lambda\mathrm{I}+\omega H\right)^{-1} = \mathrm{I} - \frac{\lambda}{\omega}\left(H+\frac{\lambda}{\omega}\mathrm{I}\right)^{-1} \qquad . \tag{44}$$

$\vec{\alpha}^m$ specifies as before the position of the maximum likelihood (see also Eq. 23). From (42) we wish to remove $\vec{x}$, ω, λ by marginalization. We consider $\vec{x}$-marginalization first. Using the finite power assumption for $\vec{x}^T\vec{x}$ we obtain for the prior (see e.g. Bretthorst [15])

$$p(\vec{x}|\mu,H_j,I) = \left(\frac{\mu}{2\pi}\right)^{\frac{P_j}{2}} \exp\left\{-\frac{\mu}{2}\vec{x}^T\vec{x}\right\} \qquad . \tag{45}$$

Then

$$p(\vec{W}|H_j,\omega,\lambda,\mu,I) = \int d\vec{x}\; p(\vec{x}|\mu,H_j,I)p(\vec{W}|\vec{x},H_j,\omega,\lambda,I) \qquad , \tag{46}$$

and P_j is the number of degrees of freedom in model H_j, e.g. $P_j = 1$ for the collisionless low-β and $P_j = 3$ for the collisional high-β cases. The integral (46) can be performed by the same techniques used to obtain (42) and yields

$$\begin{aligned} p(\vec{W}|H_j,\omega,\lambda,\mu,I) = & \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}-1}\left(\frac{\lambda}{2\pi}\right)^{\frac{M}{2}}\left(\frac{\mu}{2\pi}\right)^{\frac{P_j}{2}}\frac{1}{V_cV_\iota}\frac{1}{\sqrt{\det(W_o)}}\exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \\ & \cdot\ \frac{(2\pi)^{\frac{M}{2}}}{\sqrt{\det(\lambda\mathrm{I}+\omega H)}}\frac{(2\pi)^{\frac{P_j}{2}}}{\sqrt{\det\left(\mu\mathrm{I}+\lambda L^TRL\right)}} \\ & \cdot\ \exp\left\{-\frac{\lambda}{2}\Delta\vec{c}^T\left[R-\lambda RL\left(\mu\mathrm{I}+\lambda L^TRL\right)^{-1}L^TR\right]\Delta\vec{c}\right\} \qquad . \end{aligned} \tag{47}$$

We have nearly arrived at the wanted quantity except that we need to know the values for ω, λ and μ, the so-called hyperparameters. These values are of course not known and, moreover, they are entirely uninteresting. All three are scale parameters and the least informative prior for a scale parameter is Jeffreys' prior [12]

$$p(\mu)d\mu = \frac{d\mu}{\mu} \qquad . \tag{48}$$

$p(\mu)d\mu$ is obviously scale invariant since a scale transformation $\mu' = \alpha\mu$ leads to $p(\mu')d\mu' = d\mu/\mu = p(\mu)d\mu$. It therefore represents a state of complete ignorance about the particular values that μ may take on in the problem. Assuming Jeffreys' prior for ω, λ and μ, we could proceed to marginalize over them in (47). This is only possible numerically and even so very complicated. We shall therefore introduce at this stage a second very mild approximation. Since a large value for the hyperparameter means a sharply peaked probability distribution function we require that [15]

$$\omega \gg \lambda \gg \mu \qquad , \tag{49}$$

meaning that the information in the likelihood function is much more precise than that in the prior on $\vec{\alpha}^j$ and the latter in turn much more specific than in the prior for $\vec{x}$. We shall now use the hierarchical assumption to expand composite terms in (47):

$$\det\left(\lambda \mathrm{I} + \omega H\right) = \omega^M \det\left(H + \frac{\lambda}{\omega}\mathrm{I}\right) \simeq \omega^M \det\left(H\right) \tag{50}$$

$$\det\left(\mu \mathrm{I} + \lambda L^T R L\right) = \lambda^{P_j} \det\left(L^T R L + \frac{\mu}{\lambda}\mathrm{I}\right) \simeq \lambda^{P_j} \det\left(L^T L\right) \quad , \tag{51}$$

where we have used $R \approx \mathrm{I}$ after (44) for the last equality. We finally simplify the exponent in (47):

$$R - \lambda R L \left(\mu \mathrm{I} + \lambda L^T R L\right)^{-1} L^T R \simeq \mathrm{I} - L\left(L^T L\right)^{-1} L^T + \frac{\mu}{\lambda} L \left(L^T L\right)^{-2} L^T \quad . \tag{52}$$

With these approximations the marginal likelihood $p(\vec{W}|H_j, \omega, \lambda, \mu, I)$ factorizes

$$\begin{aligned} p(\vec{W}|H_j, \omega, \lambda, \mu, I) = & \frac{1}{V_c V_\iota \sqrt{\det(W_o)\det(H)\det\left(L^T L\right)}} \left(\frac{\omega}{2\pi}\right)^{\frac{N-M}{2}-1} \exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \\ & \cdot \left(\frac{\lambda}{2\pi}\right)^{\frac{M-P_j}{2}} \exp\left\{-\frac{\lambda}{2}\Delta\vec{c}^T S \Delta\vec{c}\right\} \\ & \cdot \left(\frac{\mu}{2\pi}\right)^{\frac{P_j}{2}} \exp\left\{-\frac{\mu}{2}\Delta\vec{c}^T \left[L\left(L^T L\right)^{-2} L^T\right] \Delta\vec{c}\right\} \quad , \end{aligned} \tag{53}$$

with

$$S = \mathrm{I} - L\left(L^T L\right)^{-1} L^T \quad . \tag{54}$$

It is now straight forward to marginalize over hyperparameters ω, λ, μ using Jeffreys' prior (48). The integrations are elementary and yield our final result for the model comparison:

$$\begin{aligned} p(\vec{W}|H_j, I) = & \frac{2\pi}{V_c V_\iota} \left(\frac{1}{\pi}\right)^{\frac{N}{2}} \frac{1}{\sqrt{\det(W_o)\det(H)\det\left(L^T L\right)}} \\ & \cdot \frac{\Gamma\left(\frac{N-M-2}{2}\right)}{\left\{\tilde{\chi}_o^2\right\}^{\frac{N-M-2}{2}}} \cdot \frac{\Gamma\left(\frac{M-P_j}{2}\right)}{\left\{\Delta\vec{c}^T S \Delta\vec{c}\right\}^{\frac{M-P_j}{2}}} \cdot \frac{\Gamma\left(\frac{P_j}{2}\right)}{\left\{\Delta\vec{c}^T \left[L\left(L^T L\right)^{-2} L^T\right] \Delta\vec{c}\right\}^{\frac{P_j}{2}}} \quad . \end{aligned} \tag{55}$$

We now substitute the numerical values of (55) into equations (7) and (8) and obtain the probability for model j represented by hypothesis H_j which we display as a bar diagram in figure 1. Note that (55) was obtained from marginalizing (37) over $\vec{\alpha}$ and hyperparameters ω, λ, μ. (55) is the global likelihood of the data in view of model (hypothesis) H_j, a quantity which is generic to Bayesian probability theory.

As a result we have that the collisionless high-β case is the most probable model to describe confinement in W7-AS. These results seem to indicate that collisions are only of minor importance. Inclusion of collisions makes up the difference between models (1,2) and models (3,4). In both cases the inclusion is accompanied by an additional degree of freedom in parameter space. This additional flexibility is penalized automatically since it obviously does not lead to a better explanation of the data. The fluid models are not supported by the data at all, as already expected from their rather large Q-measures (Table IV, second column).

The results of fig. 1 have been obtained with an approximation, which we called a mild one, namely a hierarchy in hyperparameters $\omega \gg \lambda \gg \mu$. Bayesian probability theory allows us to test a posteriori whether this assumption is justified. Let us consider the probability for the hyperparameters in the light of the data. Using Bayes' theorem we have

$$p(\omega, \lambda, \mu|\vec{W}, H_j, I) = \frac{p(\omega, \lambda, \mu|H_j, I)p(\vec{W}|H_j, \omega, \lambda, \mu, I)}{p(\vec{W}|H_j, I)} \quad , \tag{56}$$

which allows us to calculate expectation values of the hyperparameters. The second term in the numerator of (56) is known from (53). The denominator is identical to our result for model comparison and is given by (55). For the first term in the numerator of (56) we note that knowledge of a particular model H_j does not imply any information on the values of the hyperparameters ω, λ, μ, that is, $p(\omega, \lambda, \mu|H_j, I)$ is logically independent of H_j. Considering further – from I – that ω, λ, μ are scale parameters we again (see Eq. (48)) assign Jeffreys' prior to $p(\omega, \lambda, \mu|H_j, I)$:

$$p(\omega, \lambda, \mu|H_j, I) = \frac{1}{\omega\lambda\mu} \quad . \tag{57}$$

With this assignment we are ready to calculate

$$\langle x \rangle = \int d\omega \; d\lambda \; d\mu \; x \; p(\omega, \lambda, \mu|\vec{W}, H_j, I) \quad , \tag{58}$$

where x stands for any of the three ω, λ, μ. The integrations (58) are elementary and need no further comment. The results are:

$$\langle\omega\rangle = \frac{N-M-2}{\tilde{\chi_o}^2} \quad ,$$
$$\langle\lambda\rangle = \frac{M-P}{\Delta\vec{c}^T S \Delta\vec{c}} \quad , \tag{59}$$
$$\langle\mu\rangle = \frac{P}{\Delta\vec{c}^T \left[L\left(L^T L\right)^{-2} L^T\right]\Delta\vec{c}} \quad .$$

For $\langle\omega\rangle$ this resembles the unbiased estimator used in orthodox statistics for the noise level $\sigma^2 = \tilde{\chi_o}^2/(N-M)$. Before turning to the numerical results we go back to (51), where we approximated $\det\left(\mu\mathrm{I} + \lambda L^T L\right)$ for $\mu \ll \lambda$. In order that numerical values for $\langle\mu\rangle$ for different models be meaningful, it is necessary that the matrix norm of I, $\|I\|$, be comparable to $\|L^T L\|$. Choosing the Schur norm we have that $\|I\| = P_j$ and normalize accordingly $\|L^T L\|$ to P_j. We can now turn to the results collected in table V. $\langle\omega\rangle$ does of course not depend on the choice of model since it is entirely specified by the likelihood function (28). $\langle\lambda\rangle$ and $\langle\mu\rangle$ on the other hand do depend on the chosen model. We find that in all six cases the assumption $\omega \gg \lambda \gg \mu$ is excellent. This confirms that the approximation $\omega \gg \lambda \gg \mu$ which we made in arriving at (53) is very well justified for the present data set.

What remains to be done in order to finish the problem, is the calculation of the exponent vector $\vec{\alpha}$ and its error given the data and a particular model. To this end we consider

$$\langle f(\vec{\alpha})\rangle = \int d\vec{\alpha} f(\vec{\alpha}) p(\vec{\alpha}|\vec{W}, H_j, I) \quad . \tag{60}$$

Employing the sum rule

$$\langle f(\vec{\alpha})\rangle = \int d\vec{\alpha} f(\vec{\alpha}) \int d\omega\ d\lambda\ d\mu\ p(\vec{\alpha}, \omega, \lambda, \mu|\vec{W}, H_j, I) \quad , \tag{61}$$

and the product rule we get

$$\begin{aligned}\langle f(\vec{\alpha})\rangle &= \int d\omega\ d\lambda\ d\mu\ p(\omega, \lambda, \mu|\vec{W}, H_j, I) \int d\vec{\alpha} f(\vec{\alpha}) p(\vec{\alpha}|\omega, \lambda, \mu, \vec{W}, H_j, I) \\ &= \int d\omega\ d\lambda\ d\mu\ p(\omega, \lambda, \mu|\vec{W}, H_j, I) \cdot \langle f(\vec{\alpha})\rangle_{\omega,\lambda,\mu} \quad .\end{aligned} \tag{62}$$

We first have to calculate $\langle f(\vec{\alpha})\rangle_{\omega,\lambda,\mu}$. From Bayes theorem we have

$$p(\vec{\alpha}|\vec{W}, \omega, \lambda, \mu, H_j, I) = \frac{p(\vec{\alpha}|\omega, \lambda, \mu, H_j, I) p(\vec{W}|\vec{\alpha}, \omega, \lambda, \mu, H_j, I)}{p(\vec{W}|\omega, \lambda, \mu, H_j, I)} \quad . \tag{63}$$

The denominator in (63) is already known from (53) and the second term in the numerator which is logically independent of λ and μ from (28). So we need to calculate $p(\vec{\alpha}|\omega, \lambda, \mu, H_j, I)$ where we can drop the ω-dependence due to logical independence, too. Using the sum and product rules again we have

$$p(\vec{\alpha}|\lambda, \mu, H_j, I) = \int d\vec{x}\ p(\vec{\alpha}, \vec{x}|\lambda, \mu, H_j, I) = \int d\vec{x}\ p(\vec{x}|\mu, H_j, I) p(\vec{\alpha}|\vec{x}, \lambda, \mu, H_j, I) \quad . \quad (64)$$

In $p(\vec{x}|\mu, H_j, I)$ we have used the fact that it is logically independent of λ. The first term in the integrand of (64) is given by (45) and the second factor by (41). The integral (64) is again of the multivariate Gaussian type and yields after some elementary but cumbersome algebra

$$p(\vec{\alpha}|\lambda, \mu, H_j, I) \propto \exp\left\{-\frac{\lambda}{2} (\Delta\vec{\alpha} + \Delta\vec{c})^T S (\Delta\vec{\alpha} + \Delta\vec{c})\right\} \quad , \quad (65)$$

where we have again used $\Delta\vec{c} = \vec{\alpha}^m - \vec{c}$ and $\Delta\vec{\alpha} = \vec{\alpha} - \vec{\alpha}^m$ as before (23). The probability for $\vec{\alpha}$ using (63) and (65) is

$$\begin{aligned} p(\vec{\alpha}|\omega, \lambda, \mu, \vec{W}, H_j, I) &\propto \exp\left\{-\frac{1}{2}\left[(\Delta\vec{\alpha})^T \omega H (\Delta\vec{\alpha}) + (\Delta\vec{\alpha} + \Delta\vec{c})^T \lambda S (\Delta\vec{\alpha} + \Delta\vec{c})\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2} (\Delta\vec{\alpha} - \vec{\alpha}_o)^T (\omega H + \lambda S) (\Delta\vec{\alpha} - \vec{\alpha}_o)\right\} = G(\vec{\alpha}) \quad , \end{aligned} \quad (66)$$

from which we obtain $\langle f(\vec{\alpha})\rangle_{\omega,\lambda,\mu}$, the inner integral in (62) as

$$\langle f(\vec{\alpha})\rangle_{\omega,\lambda,\mu} = \frac{\int d\vec{\alpha} f(\vec{\alpha}) G(\vec{\alpha})}{\int d\vec{\alpha} G(\vec{\alpha})} \quad . \quad (67)$$

By comparison of coefficients in (66) we arrive at

$$\vec{\alpha}_o = -\lambda(\omega H + \lambda S)^{-1} S \Delta\vec{c} \quad . \quad (68)$$

Now let $f(\vec{\alpha}) = \Delta\vec{\alpha}$. From the fact that G is a Gaussian centered at $\Delta\vec{\alpha} = \vec{\alpha}_o$ we have

$$\int d\vec{\alpha}\ (\Delta\vec{\alpha} - \vec{\alpha}_o) G(\vec{\alpha}) = 0 = \int d\vec{\alpha}\ \Delta\vec{\alpha} G(\vec{\alpha}) - \vec{\alpha}_o \int d\vec{\alpha}\ G(\vec{\alpha}) \quad , \quad (69)$$

and therefore, since the factor of $\vec{\alpha}_o$ is just the normalization in (67),

$$\langle \Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu} = \vec{\alpha}_o = -\lambda(\omega H + \lambda S)^{-1} S \Delta\vec{c} \quad . \quad (70)$$

Now that we know $\langle \Delta\vec{\alpha} \rangle_{\omega,\lambda,\mu}$ we have finally to marginalize over the hyperparameters in (62). This is a very complicated integral which can only be done numerically. We employ Monte Carlo integration with importance sampling. This allows us also to check a frequently employed conceptual approximation to the strict straight forward Bayesian theory. In this approximation it is assumed that integrating over a hyperparameter is equivalent to estimating that hyperparameter from the data and then constraining it in the posterior distribution to that value [15]. This procedure, also called the empirical Bayesian estimate, is believed to perform best when many well conditioned data are available which in turn allow robust estimates of the hyperparameters. Empirical Bayes (EB) parameter estimates are included in table VI and were obtained by substitution of (59) into (70). Comparison of the parameter estimates to the exact Monte Carlo results shows that the empirical Bayes estimate performs extremely well in the present case.

Let us return once more to (67) with the choice $f(\vec{\alpha}) = (\Delta\vec{\alpha} - \vec{\alpha}_o)^2$ which allows us to put error limits on the values of the exponents in (70):

$$\langle (\Delta\vec{\alpha} - \vec{\alpha}_o)^2 \rangle_{\omega,\lambda,\mu} = \frac{1}{Z} \int d\vec{\alpha}\ (\Delta\vec{\alpha} - \vec{\alpha}_o)^2\, G(\vec{\alpha}) \qquad , \tag{71}$$

where Z is just the normalization integral. Due to the Gaussian functional form of $G(\vec{\alpha})$ this may be recast into

$$\langle (\Delta\alpha_k - \alpha_{ok})^2 \rangle_{\omega,\lambda,\mu} = -\frac{2}{Z} \frac{\partial}{\partial C_{kk}} Z \qquad , \tag{72}$$

with $C = \omega H + \lambda S$. In this notation we arrive at

$$Z = \frac{(2\pi)^{\frac{M}{2}}}{\sqrt{\det(C)}} \qquad , \tag{73}$$

and obtain after carrying out the differentiation

$$\left(\langle \Delta\alpha_k^2 \rangle - \alpha_{ok}^{\,2} \right)_{\omega,\lambda,\mu} = \left[(\lambda S + \omega H)^{-1} \right]_{kk} \qquad . \tag{74}$$

This expression for the variance given the hyperparameters ω, λ, μ has again to be averaged according to (62). As above the integral is done employing Monte Carlo techniques. Substitution of (59) into (74) provides empirical Bayes estimates of the variances, which turn out by inspection of table VI to differ only slightly from the exact answer. The agreement with the exact variance is definitely inferior to the agreement between parameters obtained by either way. Just for definiteness we mention that all Monte Carlo integrations have been carried out to a precision better than the least significant quoted decimal place of the numbers presented.

Numerical values for exponents $\vec{\alpha}$ and associated errors obtained from the Monte Carlo route are collected for all six CT models in table VII. We have repeated the maximum likelihood data and their root mean square errors for comparison. For each model the following lines present the exponent values according to (70,62), their rms-error after (74,62) and the singular value deviates of the exponents. Remember, that the singular value deviates were the smallest possible correction to $\vec{\alpha}$, which would turn it into a set of dimensionally exact exponents. We note that for models 2, 3 and 4 all singular value deviates remain smaller than one standard deviation. Model 1 fails marginally in this respect while for the two fluid models 5 and 6 all singular value deviates are larger than the quoted error. It is also interesting to calculate again the "dimensional misfit" Q/N_{lr} of (35). This number, given in parentheses after the model identifier H_j, is in every case smaller than the initial value given in table IV. This is in accord with expectation, since the analysis has incorporated dimensional information. The variation of exponents across the six models remains small and from the practical point of view unimportant. More important and interesting is the variation of the standard deviations. All of them are smaller or equal to the unrestricted least squares fit case. Going back to Bayes' theorem (5) it is easy to understand why this should be so. The posterior probability density $p(H|D,I)$ is obtained from the likelihood $p(D|H,I)$ by multiplication with the prior $p(H|I)$. For a uniform (constant) prior of arbitrary large range $< \infty$ the posterior is – apart from a constant factor – identical to the likelihood and any inferences drawn from it are identical to inferences from the likelihood. In particular the exponent vector $\vec{\alpha}$ for which the posterior distribution peaks is the unrestricted least squares solution. The prior probability used in this work is however much more informative (41). Correspondingly, the posterior $p(H|D,I)$ is narrower in $\vec{\alpha}$-space than the likelihood.

The reduction of the standard deviation is of course most impressive for the most probable model H_3. But note that the error in the exponent of the density dependence is reduced by at most 10%, while the others diminish by 25% (α_B), 41% (α_P) and 59% (α_a). This means that the density dependence of the prior probabilities, regardless which model we consider, fails to improve the description of the density dependence of the data. If, as argued by others [9], the density dependence of experimental data may saturate at higher densities, only an analysis (prior plus model function) allowing for this behavior will be able to describe the data correctly. Since we have no reason to question the CT-prior, we are lead to question the single power law representation of the energy content. This concludes our dimensionally constrained confinement scaling analysis.

VI. SUMMARY AND CONCLUSION

In this paper we have addressed the question of dimensionally constrained energy confinement analysis for a set of W7-AS data. We distinguish between dimensionally exact and dimensionally constrained. This is an important conceptual difference. In this terminology all previous work was devoted to dimensionally exact scaling functions of the power law type. Information on plasma physics was then drawn from the data misfit related to different dimensional conditions. This procedure gives only very coarse information and rules out in the case of the present data the two fluid models which lead to an increase in misfit by 60% compared to the unrestricted least squares case. A further discrimination between the four kinetic models is not possible.

An alternative route consists of using the unrestricted least squares fit solution and measuring the “distance” of it to the dimensionally exact power law expressions for the six CT models. Such a distance measure is defined in (35). Also according to this criterion the two fluid models separate distinctly from the kinetic models as being unreasonable.

Amalgamation of the misfit and the distance criterion is provided by Bayesian probability theory which incorporates the misfit as the likelihood function and the dimensional relations as prior information. Without any additional assumptions the theory adjusts for the relative weights of these two ingredients. Unlike regularization theory there is no need for choosing a coupling parameter. Bayesian theory can only answer the questions we ask and nothing else. The question that was resolved here is: if we assume a power law functional form for the W7-AS stellarator energy confinement, what is the most probable set of equations describing global transport. The answer of our analysis is that the W7-AS plasma is best described by the collisionless high-β Connor/Taylor model conditional on the assumption that the confinement function is of the power law type.

The quantitative difference between probabilities for CT model H_3 and the next smaller H_1 is however only one order of magnitude. Readers with experience in Bayesian analysis will find this not very pronounced in view of N=250 data. Usually in tests like this the most likely model will stick out in probability by several orders of magnitude against the others. Failure to do so shades doubt on the basic power law functional form. Also the very rigid density exponent and its variance, as discussed in detail at the end of section V, are presently taken as a hint that something goes wrong. In fact very recent single parameter density scans in W7-AS [16] show very clearly that the density dependence cannot be described by a simple n^{α_n} form. Similar conclusions of density dependence of energy confinement are drawn in [9]. A more complex function is therefore most likely needed for a correct description of energy confinement. This problem will be treated on the basis of Bayesian inference in a forthcoming paper.

VII. ACKNOWLEDGMENT

We are indebted to Dr. U. Stroth for providing the W7-AS data set and to Dr. G. Venus for a very careful reading of the manuscript.

REFERENCES

[1] HUGILL, J. and SHEFFIELD, J., Nucl. Fusion **18** (1978) 15.

[2] PFEIFFER, W. and WALTZ, R. E., Nucl. Fusion **19** (1979) 51.

[3] KOTSCHENREUTHER, M. and DORLAND, W., preprint (1995). See also Science **274** (1996) 1602.

[4] RIEDEL, K. S., Nucl. Fusion **30** (1990) 755.

[5] GOLDSTON, R. J., Plasma Physics and Controlled Nuclear Fusion **26** (1984) 87.

[6] CHRISTIANSEN, J. P. et al., Nucl. Fusion **31** (1991) 2117.

[7] KADOMTSEV, B. B., Sov. Phys.-J. Plasma Phys. **1** (1975) 295.

[8] CONNOR, J. W. and TAYLOR, J. B., Nucl. Fusion **17** (1977) 1047.

[9] STROTH, U. et al., Nucl. Fusion **36** (1996) 1063.

[10] LOREDO, T. J., in *Maximum Entropy and Bayesian Methods*, edited by FOUGÈRE, P. F., page 81, Kluwer Academic, Dordrecht, 1990.

[11] GREGORY, P. C. and LOREDO, T. J., Astrophys. J. **398** (1992) 146.

[12] JAYNES, E. T., Prior probabilities, in *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by ROSENKRANTZ, R. D., page 114, Reidel, Dordrecht, 1983.

[13] DOSE, V. et al., Nucl. Fusion **36** (1996) 735.

[14] ZELLNER, A., *An Introduction to Bayesian Inference in Econometrics*, pages 41–53, John Wiley & Sons, INC., New York, London, Sidney, Toronto, 1971.

[15] BRETTHORST, G. L., J. Magn. Reson. **88** (1990) 552.

[16] STROTH, U. et al., Physica Scripta **51** (1995) 655.

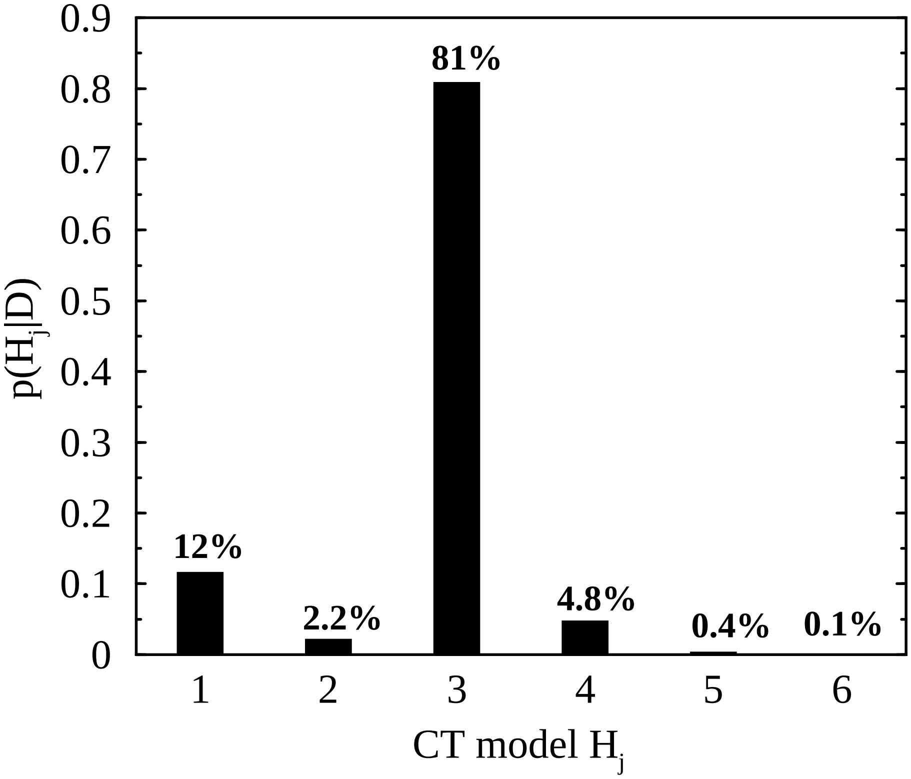

FIG. 1. Probabilities for the different CT models H_j.

W $[kJ]$	n $[10^{19}m^{-3}]$	B $[T]$	P $[MW]$	a $[m]$	ι
0.475 – 15.6	0.831 – 18.5	1.24 – 2.55	0.119 – 1.18	0.112 – 0.176	0.331 – 0.527

TABLE I. Data range

	α_c	α_n	α_B	α_P	α_a	α_ι
LSF	-0.532	0.493	0.793	0.426	2.151	0.600
stdev		0.028	0.063	0.041	0.200	

TABLE II. Unrestricted nonlinear least squares fit exponents

CT model	p	q	r	N_{lr}
1. collisionless low-β	x	0	0	3
2. collisional low-β	x	y	0	2
3. collisionless high-β	x	0	z	2
4. collisional high-β	x	y	z	1
5. fluid ideal	x	0	$\frac{3}{2}x-1$	3
6. fluid resistive	x	y	$\frac{3}{2}x-1$	2

TABLE III. Parameters of CT models

Model	Q/N_{lr}	$\Delta\alpha_n$	$\Delta\alpha_B$	$\Delta\alpha_P$	$\Delta\alpha_a$
1	0.0084	0.064	−0.121	0.016	0.079
2	0.0050	−0.026	−0.051	0.058	0.059
3	0.0003	−0.010	−0.010	−0.021	0.005
4	0.0006	−0.015	−0.010	−0.013	0.008
5	0.1263	−0.309	0.102	−0.058	−0.519
6	0.1135	−0.365	0.078	0.191	−0.226

TABLE IV. Singular value deviates

Model	1	2	3	4	5	6
$\langle\omega\rangle$			$9.406 \cdot 10^5$			
$\langle\lambda\rangle$	118.8	199.6	3065	1801	7.917	8.812
$\langle\mu\rangle$	0.1891	0.3188	0.7070	0.8556	0.6559	0.2595

TABLE V. Posterior expectation values of hyperparameters

	α_n	α_B	α_P	α_a
MC	0.499	0.781	0.412	2.183
stdev	0.025	0.047	0.024	0.081
svdev	0.000	−0.002	−0.005	0.000
EB	0.499	0.779	0.410	2.184
stdev	0.024	0.046	0.020	0.069
svdev	0.000	−0.001	0.003	0.000

TABLE VI. Comparison of the results for the collisionless high-β model. MC: numerical integration (Monte Carlo); EB: empirical Bayes result.

	α_n	α_B	α_P	α_a
LSF	0.493	0.793	0.426	2.151
stdev	0.028	0.063	0.041	0.200
H_1 ($Q/N_{lr} = 0.00257$)	0.501	0.762	0.422	2.265
stdev	0.026	0.052	0.038	0.123
svdev	0.070	−0.048	0.006	0.020
H_2 ($Q/N_{lr} = 0.00054$)	0.494	0.777	0.434	2.255
stdev	0.028	0.054	0.038	0.116
svdev	0.004	−0.008	0.029	0.012
H_3 ($Q/N_{lr} = 0.00001$)	0.499	0.781	0.412	2.183
stdev	0.025	0.047	0.024	0.081
svdev	0.000	−0.002	−0.005	0.000
H_4 ($Q/N_{lr} = 0.00003$)	0.497	0.785	0.417	2.189
stdev	0.028	0.051	0.040	0.117
svdev	−0.003	−0.002	−0.003	0.002
H_5 ($Q/N_{lr} = 0.08575$)	0.490	0.815	0.426	2.025
stdev	0.028	0.061	0.041	0.175
svdev	−0.303	0.065	−0.052	−0.398
H_6 ($Q/N_{lr} = 0.08885$)	0.489	0.811	0.428	2.045
stdev	0.028	0.062	0.041	0.185
svdev	−0.347	0.051	0.143	−0.184

TABLE VII. Dimensionally constrained scaling exponents