# MODEL COMPARISON WITH ENERGY CONFINEMENT DATA FROM LARGE FUSION EXPERIMENTS

R. PREUSS, V. DOSE AND W. VON DER LINDEN
*Max-Planck-Institut für Plasmaphysik, EURATOM Association*
*D-85748 Garching b. München, Germany* †

**Abstract.** Bayesian probability theory is applied to the model selection problem in describing the energy confinement of a fusion plasma by scaling laws. Here several possible physical scenarios lead to different models with a certain number of free parameters which are used to assign dimensional information to the scaling law. Using a recently assembled database of the W7-AS stellarator magnetic confinement device a dimensionally constrained analysis is performed. Probabilities for all models and the associated exponents of a power law scaling function are presented.

**Key words:** Model Comparison, Plasma Energy Confinement, Scaling Law Exponents.

## 1. Introduction

Controlled thermonuclear fusion is considered to be one of the most promising future energy resources. The supply with deuteron fuel contained in the oceans would last for eons. However, a ready to use power station is far from being accomplished. The technical difficulties in getting a plasma dense and hot enough to ignite and in confining it long enough to produce an energy gain are huge. Thus, temperatures in the order of tens of million degrees Kelvin are necessary to get a deuteron plasma burning. Due to the high temperatures the plasma has to be kept away from any wall material. One way to achieve this is by employing large magnetic fields to confine the plasma.

Various experimental reactors have been build to examine the physical laws of plasma behavior under such extreme conditions. One line is the stellarator type of machines in shape of a doughnut, where the confinement of the plasma is achieved by external magnetic fields alone. The data used in this article is taken from such a machine, the stellarator W7-AS. Here the plasma is characterized by the particle density $n$, the magnetic field $B$, the heating power $P$, the minor radius $a$ and the rotational transform $\iota$ (which is the ratio of how often a magnetic field

†Email: preuss@ipp.mpg.de

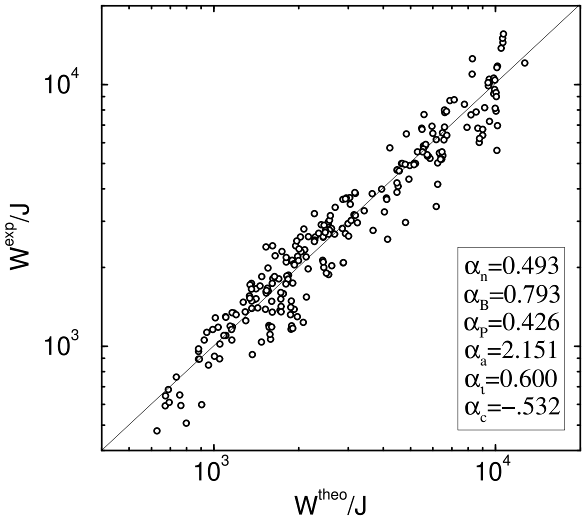


*Figure 1.* Unconstrained least squares fit of the $N = 250$ data. The plasma energy content of a machine predicted according to this fit would be found along the diagonal line.

line encircles the torus in poloidal and toroidal direction). Additional machine parameters like the major radius of the torus come only into play in comparisons with other stellarators of different size.

One decisive quantity in achieving a burning plasma is the energy content $W$ which depends in a certain way on the above parameters. Since an exact description of the plasma is a highly complicated many particle problem which is impossible to solve, global scaling laws are used to determine $W$ [1]. Introduced in the seventies scaling laws were of service for comparing different confinement regimes and inter-machine performances. Derived from small and mid sized machines, they were surprisingly successful in the performance prediction of much larger devices and are therefore of crucial importance in the planing of new machines (e.g. ITER).

Due to the fact that the exact plasma behavior is unknown one is restricted to assume a certain functional form where all the plasma determining parameters are fed into. The initial assumption of a power law dependence has survived up to now because of convenience and simplicity:

$$W_i = e^{\alpha_c} n_i^{\alpha_n} B_i^{\alpha_B} P_i^{\alpha_P} a_i^{\alpha_a} \iota_i^{\alpha_\iota} + \varepsilon_i \qquad , \tag{1}$$

where $W_i$ is the $i$-th measurement of the plasma energy content and $\varepsilon_i$ its error. The constant factor has been expressed as $e^{\alpha_c}$ for calculational convenience.

We are out for the scaling exponents $\vec{\alpha}^* = (\alpha_n, \alpha_B, \alpha_P, \alpha_a, \alpha_\iota, \alpha_c)^T$. The simplest way to do this is to perform a least squares fit where $\tilde{\chi}^2 = \sum_{i=1}^N (W_i^{exp} - W_i^{theo})^2$ is minimized over $\vec{\alpha}^*$ (see Fig. 1). The obtained exponents would then be used in the plasma energy content predictions of new machines.

| $M_j$ | plasma model | degrees of freedom $N_P$ | parameter vector $\vec{x}$ |
|---|---|---|---|
| **1** | collisionless low-$\beta$ | 1 | $x$ |
| **2** | collisional low-$\beta$ | 2 | $(x,y)^T$ |
| **3** | collisionless high-$\beta$ | 2 | $(x,z)^T$ |
| **4** | collisional high-$\beta$ | 3 | $(x,y,z)^T$ |
| **5** | ideal fluid | 1 | $x$ |
| **6** | resistive fluid | 2 | $(x,y)^T$ |

TABLE 1. Plasma models according to Connor and Taylor [2].

But this approach disregards conflicting dimensions of the quantities on the l.h.s. and the r.h.s. of Eq. (1). Moreover it ignores available information on the exponents from basic plasma theory. This was provided by Connor and Taylor [2] who 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. The examined six models are listed in table 1 together with the number of the remaining free parameters. The implementation of the models leads to constraints on $\alpha_n$, $\alpha_B$, $\alpha_P$ and $\alpha_a$, (further abbreviated as $\vec{\alpha}$). The number of degrees of freedom $N_P$ in choosing $\vec{\alpha}$ have reduced according to the complexity of the physical model from originally $N_M = 4$ exponents $\alpha_i$ to one, two or three. The corresponding parameter vectors are also shown in table 1. As an example the explicit dependences of $\vec{\alpha}$ are shown for the third model $M_3$:

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

with $\vec{\alpha} = \vec{c} + L\vec{x}$ and $\vec{x} = (x,z)^T$.

However, despite of the relative success in the past there is no reason to assume that a power scaling law represents the relevant plasma physics adequately [3]. Therefore we shall not require the power law exponents to satisfy the Connor/Taylor [CT] relations exactly. We therefore call our results dimensionally constrained as opposed to dimensionally exact.

In the following, Bayesian inference is employed to quantify the probability of each of the six CT models in the light of the data and to derive the respective exponent sets which approach the CT dimensional relation as close as possible at an as small as possible expense in data fit.

## 2. Model comparison

The probability of model $M_j$ in the light of the data $D$ and the prior information $I$ is given using Bayes' theorem.

$$p(M_j|D,I) = \frac{p(M_j|I)p(D|M_j,I)}{p(D|I)} = \frac{p(M_j|I)p(D|M_j,I)}{\sum_k^{\#M} p(M_k|I)p(D|M_k,I)} \quad . \tag{2}$$

Since a priori no model is preferred, a constant uniform prior is assigned to each model and $p(M_j|I)$ cancels out

$$p(M_j|D,I) = \frac{p(D|M_j,I)}{\sum_k^{\#M} p(D|M_k,I)} \quad . \tag{3}$$

What remains to be determined is the global likelihood of the data, $p(D|M_j,I)$. This is achieved in marginalizing over all quantities which contribute to the problem in the joint probability, as there are the unknown error level represented by the hyperparameter $\omega$, the exponent vector $\vec{\alpha}^*$ and its hyperparameter $\lambda$, and the parameter vector $\vec{x}$ and its hyperparameter $\mu$:

$$p(D|M_j,I) = \int p(D,\omega,\vec{\alpha}^*,\lambda,\vec{x},\mu|M_j,I)\ d\omega d\vec{\alpha}^* d\lambda d\vec{x} d\mu \tag{4}$$

$$= \int d\omega d\lambda d\mu\ p(\omega|I)p(\lambda|I)p(\mu|I) \int d\vec{\alpha}^* d\vec{x}\ p(D,\vec{\alpha}^*,\vec{x}|\omega,\lambda,\mu,M_j,I) \quad . \tag{5}$$

Let us postpone the integrals over the hyperparameters. In the following we are concerned with the evaluation of the last of the upper integrals, which is

$$p(D|\omega,\lambda,\mu,M_j,I) = \int d\vec{x}\ p(\vec{x}|\mu,I) \int d\vec{\alpha}^*\ p(\vec{\alpha}^*|\vec{x},\lambda,M_j,I)p(D|\vec{\alpha}^*,\omega,M_j,I) \ . \tag{6}$$

Now the remaining quantities have to be assigned. We start with the likelihood function. Since we assume a Gaussian error statistic with constant absolute error, we get

$$p(D|\vec{\alpha}^*,\omega,I) = \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N}(D_i - W_i(\vec{\alpha}^*))^2\right\} \quad . \tag{7}$$

The hyperparameter $\omega$ is related to the noise level by $\langle\varepsilon^2\rangle = 1/\omega$. Here we identify the data vector $D$ with the experimental measurement, and $W(\vec{\alpha}^*)$ with the theoretical power law.

$(\alpha_c,\ \alpha_\iota)$ are not determined by any model and therefore marginalized over. A direct integration is impossible and we therefore use the method of steepest descent to give

$$p(D|\vec{\alpha},\omega,I) \propto \left(\frac{\omega}{2\pi}\right)^{\frac{N-2}{2}} \exp\left\{-\frac{\omega}{2}\left(\tilde{\chi}_o^2 + \Delta\vec{\alpha}^T H \Delta\vec{\alpha}\right)\right\} \quad , \tag{8}$$

where a uniform distribution was assumed for the prior in $(\alpha_c,\ \alpha_\iota)$. Here and further on all the quantities which cancel out in the model comparison of Eq. 2 are dropped. $H$ is the reduced Hesse matrix over the remaining $\vec{\alpha}$, and $\Delta\vec{\alpha} = \vec{\alpha} - \vec{\alpha}^{ML}$ with $\vec{\alpha}^{ML}$ the position of the minimum of $\tilde{\chi}^2$.

Next we assign the prior functions. Using finite power assumption that $\lambda^{-1} = \langle||\vec{\alpha} - \vec{c} - L\vec{x}||^2\rangle$ the principle of maximum entropy gives [4]

$$p(\vec{\alpha}|\lambda,\vec{x},M_j,I) = \left(\frac{\lambda}{2\pi}\right)^{\frac{N_M}{2}} \exp\left\{-\frac{\lambda}{2}||\vec{\alpha} - \vec{c} - L\vec{x}||^2\right\} \quad . \tag{9}$$

The same assumption is made to assign the prior for $\vec{x}$:

$$p(\vec{x}|\mu,M_j,I) = \left(\frac{\mu}{2\pi}\right)^{\frac{N_{P_j}}{2}} \exp\left\{-\frac{\mu}{2}\vec{x}^T\vec{x}\right\} \quad . \tag{10}$$

We now assume that the likelihood function is sharply peaked as compared to the prior $p(\vec{\alpha}|H,I)$ and that the prior in $\vec{\alpha}$ is sharply peaked as compared to the prior $p(\vec{x}|H,I)$ [5]. This is the normal case, where the measurement contains information considerably more detailed than the prior, and in the same way that the $\vec{\alpha}$ are far better determined than the parameters $\vec{x}$. Thereby an hierarchy in the hyperparameters

$$\omega \gg \lambda \gg \mu \tag{11}$$

is established. The integral in Eq. (6) may consequently be approximated by

$$p(D|\omega,\lambda,\mu,M_j,I) = p(\hat{\vec{x}}|\mu,M_j,I)\int d\vec{x}\ p(\hat{\vec{\alpha}}|\vec{x},\lambda,M_j,I)\int d\vec{\alpha}\ p(D|\vec{\alpha},\omega,I)\ , \tag{12}$$

where $\hat{\vec{\alpha}}$ and $\hat{\vec{x}}$ denote the vectors for which the likelihood $p(D|\vec{\alpha},M_j,I)$ and the prior $p(\hat{\vec{\alpha}}|\vec{x},\lambda,M_j,I)$ attain their maximum, respectively.

First we want to integrate over the remaining nonlinear parameters in the likelihood.

$$\int d\vec{\alpha}\ p(D|\vec{\alpha},\omega,I) \propto \left(\frac{\omega}{2\pi}\right)^{\frac{N-N_M-2}{2}} \exp\left\{-\frac{\omega}{2}\tilde{\chi}_o^2\right\} \quad , \tag{13}$$

with the maximum of $p(D|\vec{\alpha},\omega,I)$ at $\hat{\vec{\alpha}} = \vec{\alpha}^{ML}$.

The integration of the prior in $\vec{\alpha}$ is straight forward:

$$\int d\vec{x}\ p(\hat{\vec{\alpha}}|\vec{x},\lambda,M_j,I) = \frac{1}{\sqrt{\det\left(L^T L\right)}}\left(\frac{\lambda}{2\pi}\right)^{\frac{N_M-N_{P_j}}{2}} \exp\left\{-\frac{\lambda}{2}\Delta\vec{c}^T S\Delta\vec{c}\right\} \quad , \tag{14}$$

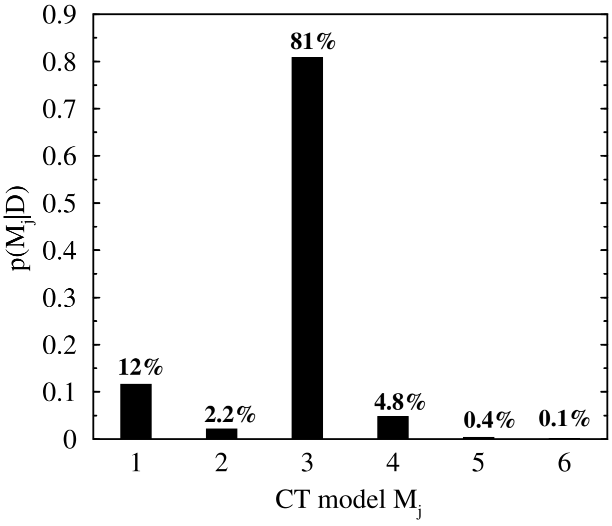


*Figure 2.* Probabilities for different CT models $M_j$.

with $S = \mathrm{I} - L\left(L^T L\right)^{-1} L^T$ and $\Delta\vec{c} = \vec{\alpha}^{ML} - \vec{c}$. Finally we need the maximum of $p(\hat{\vec{\alpha}}|\vec{x}, \lambda, M_j, I)$ at $\hat{\vec{x}} = \left(L^T L\right)^{-1} L^T \Delta\vec{c}$ to state the prior in the parameter vector:

$$p(\hat{\vec{x}}|\mu, M_j, I) = \left(\frac{\mu}{2\pi}\right)^{\frac{N_{P_j}}{2}} \exp\left\{-\frac{\mu}{2}\Delta\vec{c}^T K \Delta\vec{c}\right\} \quad , \tag{15}$$

with $K = L\left(L^T L\right)^{-2} L^T$.

What is left to be done is the marginalization over the hyperparameters using Jeffreys' Prior. This gives our final result for the model comparison:

$$p(D|M_j, I) \propto \frac{\Gamma\left(\frac{N-N_M-2}{2}\right)}{\{\tilde{\chi}_o^2\}^{\frac{N-N_M-2}{2}}} \frac{\Gamma\left(\frac{N_M-N_{P_j}}{2}\right)}{\{\Delta\vec{c}^T S \Delta\vec{c}\}^{\frac{N_M-N_{P_j}}{2}}} \frac{\Gamma\left(\frac{N_{P_j}}{2}\right)}{\{\Delta\vec{c}^T K \Delta\vec{c}\}^{\frac{N_{P_j}}{2}}} \quad , \tag{16}$$

where the first term already lost all dependences on the models and is given only for reasons of completeness. As a result (see Fig. 2) we have that the collisionless high-$\beta$ case is the most probable model to describe confinement in W7-AS. These results seem to indicate that collisions in the plasma 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 (5,6) are not supported by the data at all.

The results of Fig. 2 have been obtained with an approximation, which we call a mild one, namely a hierarchy in hyperparameters $\omega \gg \lambda \gg \mu$. Bayesian probability

| $M_j$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\langle\omega\rangle$ | ———— | | $9.4 \cdot 10^5$ | | ———— | |
| $\langle\lambda\rangle$ | 119 | 200 | 3065 | 1801 | 7.92 | 8.81 |
| $\langle\mu\rangle$ | 0.19 | 0.32 | 0.72 | 0.86 | 0.66 | 0.26 |

TABLE 2. Posterior expectation values of hyperparameters

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

$$p(\omega,\lambda,\mu|D,M_j,I) = \frac{p(\omega,\lambda,\mu|M_j,I)p(D|\omega,\lambda,\mu,M_j,I)}{p(D|M_j,I)} \quad , \tag{17}$$

which allows us to calculate expectation values of the hyperparameters. The second term in the numerator of (17) is known from (6). The denominator is identical to our result for model comparison and is given by (16). For the first term in the numerator of (17) we note that knowledge of a particular model $M_j$ does not imply any information on the values of the hyperparameters $\omega$, $\lambda$, $\mu$, that is, $p(\omega,\lambda,\mu|M_j,I)$ is logically independent of $M_j$. Considering further – from the prior information in I – that $\omega$, $\lambda$, $\mu$ are scale parameters we assign Jeffreys' prior to $p(\omega,\lambda,\mu|M_j,I)$. Then we are ready to calculate

$$\langle\xi\rangle = \int d\omega \; d\lambda \; d\mu \; \xi \; p(\omega,\lambda,\mu|D,M_j,I) \tag{18}$$

where $\xi$ stands for any of the three $\omega$, $\lambda$, $\mu$. The integration gives

$$\langle\omega\rangle = \frac{N - N_M - 2}{\tilde{\chi}_o^2} \quad , \tag{19}$$

$$\langle\lambda\rangle = \frac{N_M - N_{P_j}}{\Delta\vec{c}^T S \Delta\vec{c}} \quad , \tag{20}$$

$$\langle\mu\rangle = \frac{N_{P_j}}{\Delta\vec{c}^T K \Delta\vec{c}} \quad . \tag{21}$$

We can now turn to the results collected in table 2. $\langle\omega\rangle$ does of course not depend on the choice of model since it is entirely specified by the likelihood function (8). $\langle\lambda\rangle$ 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 (6) is very well justified for the present data set.

Finally the scaling exponents are calculated. We consider

$$\langle\Delta\vec{\alpha}\rangle = \int d\omega \; d\lambda \; d\mu \; p(\omega,\lambda,\mu|\vec{W},M_j,I) \cdot \langle\Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu} \quad . \tag{22}$$

| | $\alpha_n$ | $\alpha_B$ | $\alpha_P$ | $\alpha_a$ |
|---|---|---|---|---|
| MC | 0.499 | 0.781 | 0.412 | 2.183 |
| stdev | 0.025 | 0.047 | 0.024 | 0.081 |
| EB | 0.499 | 0.779 | 0.410 | 2.184 |
| stdev | 0.024 | 0.046 | 0.020 | 0.069 |

TABLE 3. Comparison of the results for the collisionless high-$\beta$ model. MC: numerical integration (Monte Carlo); EB: empirical Bayes result.

where we first have to calculate $\langle\Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu}$ given by

$$\langle\Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu} = \frac{\int d\vec{\alpha}\ \Delta\vec{\alpha}\ p(\vec{\alpha}|\omega,\lambda,\mu,D,M_j,I)}{\int d\vec{\alpha}\ p(\vec{\alpha}|\omega,\lambda,\mu,D,M_j,I)} \quad . \tag{23}$$

After some algebra we get

$$p(\vec{\alpha}|\omega,\lambda,\mu,D,M_j,I) \propto \exp\left\{-\frac{1}{2}\left(\Delta\vec{\alpha}-\vec{\alpha}_o\right)^T\left(\omega H+\lambda S\right)\left(\Delta\vec{\alpha}-\vec{\alpha}_o\right)\right\} = G(\vec{\alpha}), \tag{24}$$

with $\Delta\vec{\alpha} = \vec{\alpha} - \vec{\alpha}^{ML}$ and $\vec{\alpha}_o = -\lambda(\omega H + \lambda S)^{-1} S\Delta\vec{c}$. From the fact that $G$ is a Gaussian centered at $\Delta\vec{\alpha} = \vec{\alpha}_o$ we have

$$\int d\vec{\alpha}\ \left(\Delta\vec{\alpha}-\vec{\alpha}_o\right) G(\vec{\alpha}) = 0 = \int d\vec{\alpha}\ \Delta\vec{\alpha} G(\vec{\alpha}) - \vec{\alpha}_o \int d\vec{\alpha}\ G(\vec{\alpha}) \quad , \tag{25}$$

and therefore, since the factor of $\vec{\alpha}_o$ is just the normalization in (23),

$$\langle\Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu} = \vec{\alpha}_o = -\lambda(\omega H + \lambda S)^{-1} S(\vec{\alpha}^{ML} - \vec{c}) \quad . \tag{26}$$

Now that we know $\langle\Delta\vec{\alpha}\rangle_{\omega,\lambda,\mu}$ we have finally to marginalize over the hyperparameters in (22). 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 [4]. 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 3 and were obtained by substitution of (19), (20) and (21) into (26). Comparison of the parameter estimates to the exact Monte Carlo results shows that the empirical Bayes estimate performs extremely well in the present case.

## 3. 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, since we have no rigorous reason at all to assume that plasma behavior may be adequately represented by a power law. In our Bayesian analysis we found that the W7-AS plasma is best described by the collisionless high-$\beta$ Connor/Taylor model conditional on the assumption that the confinement function is of the power law type. 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.

## References

1. J. Hugill and J. Sheffield *Nucl. Fusion*, **18**, p. 15, 1978.
2. J. W. Connor and J. B. Taylor *Nucl. Fusion*, **17**, p. 1047, 1977.
3. R. J. Goldston *Plasma Physics and Controlled Nuclear Fusion*, **26**, p. 87, 1984.
4. G. L. Bretthorst *J. Magn. Reson.*, **88**, p. 552, 1990.
5. P. C. Gregory and T. J. Loredo *Astrophys. J.*, **398**, p. 146, 1992.