Scalings and Plasma Profile Parameterisation of ASDEX High Density Ohmic Discharges

P. J. Mc Carthy*, K. S. Riedel†, O. J. W. F. Kardaun,
H. Murmann, K. Lackner, the ASDEX Team

IPP 5/34 May 1990

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

8046 GARCHING BEI MÜNCHEN

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

GARCHING BEI MÜNCHEN

Scalings and Plasma Profile Parameterisation of ASDEX High Density Ohmic Discharges

P. J. Mc Carthy*, K. S. Riedel†, O. J. W. F. Kardaun,
H. Murmann, K. Lackner, the ASDEX Team

IPP 5/34 May 1990

*On attachment from University College, Cork, Ireland
† New York University, 251 Mercer Str., New York, USA

Die nachstehende Arbeit wurde im Rahmen des Vertrages zwischen dem Max-Planck-Institut für Plasmaphysik und der Europäischen Atomgemeinschaft über die Zusammenarbeit auf dem Gebiete der Plasmaphysik durchgeführt.

SCALINGS AND PLASMA PROFILE PARAMETERISATION OF ASDEX HIGH DENSITY OHMIC DISCHARGES

P.J. Mc Carthy[1], K.S. Riedel [2], O.J.W.F. Kardaun,
H. Murmann, K. Lackner, the ASDEX Team

Max-Planck-Institut für Plasmaphysik, EURATOM Association,
D-8046 Garching bei München, Fed. Rep. Germany

[1]On attachment from University College, Cork, Ireland
[2] New York University, 251 Mercer Str., New York, USA

Abstract

A database of high density ($.3 < \overline{n_e}/10^{20}\mathrm{m}^{-3} < .8$), low q_a ($1.9 < q_a < 3.4$), Ohmic discharges from the ASDEX experiment is analysed statistically [1] . Bulk parameter scalings and parameterised temperature and density profile shapes are presented. The total plasma kinetic energy, assuming $T_i = T_e$, scales as $\overline{n_e}^{.54\pm.01} I_p^{.90\pm.04}$ and is independent of B_t. The electron temperature profile peaking factor scales as $\frac{T_0^{3/2}}{<T^{3/2}>} = .94(\pm.04) q_a^{1.07\pm.04}$ in close agreement with the assumption of classical resistive equilibrium. In the inner half of the plasma, the inverse fall-off length for both temperature and density has a strong dependence on q_a, with the temperature dependence being more pronounced. Outside the half radius the q_a dependence disappears but the density profile broadens near the edge with increasing plasma current. A second database of moderate density, moderate q_a discharges is presented for comparison.

1. INTRODUCTION

In most high magnetic field tokamaks, the experimental Ohmic energy confinement time increases roughly linearly with plasma density over a large range of densities. In general, moderate field devices observe a similar but weaker increase in confinement time at small to moderate densities. Unfortunately, the Ohmic energy confinement time saturates at higher values of the Murakami parameter $\overline{n_e}R/B_t$ [2] . This phenomenon is termed density roll-over and, under usual operating conditions in ASDEX, occurs at line-averaged densities around $.25 \times 10^{20}\mathrm{m}^{-3}$ and magnetic fields on the order of two Tesla. (Pellet fuelled discharges and the recently discovered Improved Ohmic Confinement regime on ASDEX [3] are not considered here.) We present a statistical analysis of the bulk parameter scalings and profile shapes in this saturated high density regime. For a detailed discussion of the theoretical aspects of profile shape determination, the reader is referred to [1], which we abbreviate as KRML.

The high density database consists of 105 pairs of temporally compressed experimental electron temperature and density profiles from 50 ASDEX deuterium discharges with a corresponding set of bulk plasma parameters and equilibrium flux surface information. Each compressed profile typically consists of the time average of 12 consecutive 'raw' profiles sampled at 17 msec intervals using the ASDEX Nd:YAG ('YAG') Thomson scattering system [4].

All profiles are measured during the current flat-top Ohmic phase of the discharge. Apart from making the data analysis more manageable, the preliminary compression had the effect of strongly reducing profile fluctuations due to sawteeth. Such averaging is free from bias as long as the sawtooth period ($\simeq$ 10msec) is incommensurate with the sampling period. Some typical parameter values for this database are the line-averaged electron density $\overline{n_e} = .4 \times 10^{20}\mathrm{m}^{-3}$, the plasma current $I_p = 400\mathrm{kA}$, and the cylindrical edge $q_a = 2.5$. ASDEX discharges have a fixed circular geometry with $R_{plasma} = 1.65\mathrm{m}$ and $a_{minor} = .40\mathrm{m}$. Table I(a) presents a more detailed summary of the database. All datapoints consist of double-null divertor discharges in deuterium from three consecutive shot days (404 to 406) spanning a one week period in February 1986 during which titanium gettering was extensively used (about every third discharge). One of the experimental goals for these shot days was a density limit investigation, resulting in the presence of a density ramp of .2 to $.4 \times 10^{20}\mathrm{m}^{-3}\mathrm{s}^{-1}$ for 10 out the 50 discharges, which accounted for $\simeq$ 50% of the datapoints. In selecting the discharges, care was taken to cover as wide a range in ASDEX control parameter operating space as possible. The main toroidal field is relatively seldom varied, however, and two thirds of the datapoints are clustered at toroidal magnetic field values

of 1.8T or 2.2T . Nevertheless, the distinction between q_a and plasma current scaling will be quite apparent. Bolometric measurements showed that radiation losses for these discharges ($10\% < P_{rad}/P_{Ohmic} < 40\%$, increasing with density) were strongly localised at the plasma edge.

For comparison, we present a complementary analysis of a second database (38 datapoints from 38 deuterium discharges) of moderate density ($\overline{n_e} \simeq .3 \times 10^{20}\mathrm{m}^{-3}$), moderate q_a ($q_a \simeq 3.1$), non-gettered, Ohmic discharges which is summarised in Table I(b) and includes discharges with densities centred around the transition to the roll-over regime. These discharges were made over a six month period spanning 64 shot days from August 1985 to February 1986 and were selected from days where the vacuum vessel was in a nominally normal state (no gettering, stainless steel walls). At other times within this period, however, the experiment was run under a variety of conditions including carbonised walls and three days of operating with He as working gas. In addition, the vessel was open for several weeks in December 1985. The YAG system was calibrated using Raman scattering from hydrogen gas on two occasions during this period. Our database contains no data quantifying the resulting changes or residual effects of these operating periods on the wall condition and, in particular, the condition of the YAG diagnostic window itself. Hence we expect the residual error to the fits to these experimental data to be larger. For these reasons, the scalings of the second database are included largely for purposes of comparison. The distribution of datapoints for each dataset is displayed in the combined Hugill plot (Fig. 1).

The two databases differ in three parameters, $\overline{n_e}$, q_a, and wall condition (gettered vs. non-gettered). Thus the combined database is largely clustered in two cells of the eight possible combinations. Furthermore, some scaling differences between the two sets of data were visually apparent in preliminary efforts at fitting the combined dataset. By analysing each database separately, we are able to estimate secondary, weaker effects that are not immediately apparent in a joint analysis. In our case, these effects are weak current and density dependencies of the outer section of the normalised density profiles. If the two data sets were combined, these weaker effects would be obscured by artificial secondary scalings arising from the clustering mentioned above.

As is typical of single machine databases, geometric parameters such as the plasma position and cross-sectional area vary very little (about 1%) and, once the channel positions have been mapped onto the normalised flux surface coordinate ($0 \leq r \leq 1$), these variables are ignored in the analysis. We assume that the macroscopic plasma state is essentially determined by the line-averaged density $\overline{n_e}$, the total plasma current I_p, and either q_a or the toroidal magnetic field B_t. These parameters represent the

major control parameters which the experimentalist utilises to vary the plasma state in an Ohmic plasma.

The condition of the plasma wall may significantly influence plasma performance. In an effort to include these effects, an additional plasma variable such as the effective ion charge Z_{eff}, or the total Ohmic power P_Ω is sometimes used as an extra independent variable. Both Z_{eff} and the loop voltage U_{loop} depend on the control parameters $\overline{n_e}$, I_p and q_a , and therefore are not purely measures of the condition of the plasma wall.

To examine the extent to which the Ohmic power and Z_{eff} vary independently of the control parameters $\overline{n_e}$, I_p and q_a , we carry out a principal component analysis (PCA) [5]: The correlation matrix of the logarithms of $\overline{n_e}$, I_p, q_a, Z_{eff}, and P_Ω is calculated and diagonalised (see Table II). Each eigenvalue is interpreted as the sample variance of its corresponding principal component over the database, and a small eigenvalue indicates a near collinearity between the original variables. Since the principal components, by construction, are statistically uncorrelated, the sum of a subset of eigenvalues gives the cumulative variance explained by the corresponding subset of principal components. We find that, for both databases, over 97% of the data variance can be explained by the first three components. The residual 3% of the total variance is ascribed to noise in the temperature measurements which affects Z_{eff} and noise in V_{loop} which affects both Z_{eff} and P_Ω . Hence, we discard the principal components having the two smallest eigenvalues.

The remaining eigenvectors span a three-dimensional subspace. This subspace can be spanned by any set of three well-conditioned linear functions of the five variables. That $\overline{n_e}$, I_p and q_a in fact constitute such a well-conditioned set is checked by re-doing the PCA for these three alone. The eigenvalues of the 3 x 3 correlation matrix, also listed in Table II, are all of the same order of magnitude, indicating that, unlike the larger set, no near collinearity exists between them.

For scalings of the bulk parameters, we preferred to use the logarithms of $\overline{n_e}$, B_t, and I_p because the correlation between I_p and B_t was less than that of either variable with q_a. The dominance of q_a in determining plasma profile shapes motivated us to replace B_t with q_a for the profile shape analysis.

2. BULK SCALINGS

In this analysis, we use the Spitzer value for the effective ion charge Z_{eff} , which is calculated from the electron temperature profile assuming resistive equilibrium (well

satisfied for the current flat-top profiles selected) and Spitzer conductivity. We note that it cannot be used as an independent parameter in the temperature profile regressions below. Recent Bremstrahlung measurements on ASDEX [6] show that for the saturated Ohmic confinement regime (to which our high density database belongs), Z_{eff} is very flat over most of the plasma radius, though tending to rise strongly near the boundary. We adopt the conventional assumption of zero radial dependence here. The ion density is calculated with the assumption that the sole impurity is oxygen. The plasma kinetic energy, $\mathrm{Wp}_{kin} = \frac{3}{2}\int (n_e T_e + n_i T_i) dv$, is calculated by assuming the ion temperature is equal to the electron temperature, $T_i = T_e$. This assumption is justified when the electron-ion energy exchange time is much shorter than the energy confinement time. In the high density database, the typical values are $\tau_{ei} = 5$ msec and $\tau_E = 75$ msec . In some of the moderate density discharges, this condition is at best marginally satisfied.

As an independent measure of the plasma energy content we also make use of the diamagnetic flux measurement on ASDEX whose interpretation is not affected by the uncertainty in T_i and is further simplified, in the case of Ohmic plasmas, by the absence of pressure anisotropy. The extreme sensitivity of the measurement ($\frac{\psi_{dia}}{\psi_{tor}} \simeq 10^{-4}$) to such factors as slight mechanical displacements of the diamagnetic loop means, however, that the typical error associated with the derived value for beta poloidal is $\delta(\beta_{p(dia)}) = \pm.05$. For $\beta_{p(dia)} \simeq 0.3$ (the lower limit in each database) this implies an error of some $\pm 15\%$ in the diamagnetic energy and confinement time. For these databases, the diagmagnetic energy is systematically greater than the kinetic energy with an energy-independent offset $\simeq 6 \pm 2$ kJ.

The models considered are of the form $y = \sum_i \alpha_i x_i + \epsilon$ where y and x denote the logarithms of the dependent and independent bulk plasma variables respectively. The root mean square error (RMSE) of the fit is $\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2/(N-p)}$ where p is the number of independent variables (including the intercept) and $\hat{y}_i$ is the fitted value of y_i. As our response variables are natural logarithms, the RMSE corresponds to a relative error in the physical variable. We also quote the squared multiple correlation coefficient R^2, which represents the fraction of total variance about the mean accounted for by the fit: $\mathrm{R}^2 = \sum_i (\hat{y}_i - \overline{y})^2 / \sum_i (y_i - \overline{y})^2$. In the bulk scaling results that follow, all plasma variables are expressed in the units of Table I.

Total Plasma Energy Scaling

The plasma energy content from kinetic data for the high density database satisfies

$$\mathrm{Wp}_{kin} = 133(\pm 6)\overline{n_e}^{\,.54\pm.01} I_p^{\,.90\pm.04} B_t^{\,.06\pm.04} \quad \mathrm{RMSE} = .04 \quad \mathrm{R}^2 = .965$$

Similarly, the scaling law based on the moderate density database is

$$\mathrm{Wp}_{kin} = 128(\pm 6)\overline{n_e}^{\,.53\pm.03} I_p^{\,.80\pm.05} B_t^{\,.13\pm.05} \quad \mathrm{RMSE} = .04 \quad \mathrm{R}^2 = .959$$

The plasma energy content inferred from the diagmagnetic signal for the high density database is described by

$$\mathrm{Wp}_{dia} = 130(\pm 6)\overline{n_e}^{\,.40\pm.01} I_p^{\,.88\pm.04} B_t^{\,.13\pm.04} \quad \mathrm{RMSE} = .04 \quad \mathrm{R}^2 = .943$$

whereas the moderate density database gave a poorer fit:

$$\mathrm{Wp}_{dia} = 83(\pm 10)\overline{n_e}^{\,.21\pm.05} I_p^{\,.77\pm.07} B_t^{\,.33\pm.08} \quad \mathrm{RMSE} = .06 \quad \mathrm{R}^2 = .878$$

The significance of each regressor in the least squares model can be interpreted in terms of $\hat{\alpha}/\hat{\sigma}(\hat{\alpha})$, the ratio of the fitted coefficient to its standard error estimate. Under standard least squares assumptions, including (a) the correctness of the regression model and (b) normally distributed errors in the dependent variable, the ratio $\hat{\alpha}_j/\hat{\sigma}(\hat{\alpha}_j)$ has a Student's t distribution under the null hypothesis that $\alpha_j = 0$. When the probability of t exceeding this ratio is less than some small specified significance level:

$$P\left\{ t \notin \left(-\left|\frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)}\right|, \left|\frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)}\right| \right) \right\} < \epsilon \,, \tag{1}$$

then the null hypothesis is rejected and the regressor is considered significant. For many degrees of freedom, as in the present case, Student's t can be well approximated by the normal distribution. Using $\epsilon = .05$ i.e. a 5% rejection level, we then have $t_{\epsilon=.05} = 2.0$ and hence a significance criterion of $|\hat{\alpha}/\hat{\sigma}(\hat{\alpha})| > 2.0$. This result is no longer exact, though still approximate, for mild violations of the normality assumption. A seriously deficient model can, however, invalidate this interpretation of the coefficient standard errors.

It is often useful to have an estimate of the contribution to the overall R^2 from each independent variable. If the regressors are (approximately) uncorrelated, we have the following useful relationship (see Appendix A):

$$\Delta(\mathrm{R}^2)_j = \frac{t_j^2}{N-p}(1-\mathrm{R}^2) \,, \tag{2}$$

where the LHS denotes the decrease in R^2 if the j^{th} regressor is removed from the model.

Using these criteria we see, for instance, that the B_t scaling for Wp is insignificant for the high density database, while in the moderate density case it is marginally significant. In both cases, its omission scarcely affects the goodness of fit (eq. (2) gives $\Delta(\mathrm{R}^2) \simeq .001, .007$ respectively).

Volume Averaged Electron Temperature

Both databases are in rough agreement with the Pfeiffer-Waltz [7] and JET Ohmic scalings [8] . The high density database satisfies

$$< T_e >= 0.639(\pm 0.03)\overline{n_e}^{\,-.52\pm.01} I_p^{\,.97\pm.04} B_t^{\,-.04\pm.04} \quad \text{RMSE} = .04 \quad \text{R}^2 = .947$$

Similarly, the moderate density database satisfies

$$< T_e >= 0.605(\pm 0.06)\overline{n_e}^{\,-.61\pm.04} I_p^{\,.93\pm.06} B_t^{\,.04\pm.07} \quad \text{RMSE} = .05 \quad \text{R}^2 = .913$$

Here, B_t is insignificant for both databases and its omission has a very small ($\simeq .001$) effect on R^2 . We do not present regressions of the temperature versus the Spitzer Z_{eff}, since the latter quantity is derived from the temperature profile:

$$Z_{eff} \propto \frac{\langle T^{\frac{3}{2}} \rangle_{area} V_{loop}}{I_p R_{plas}} \tag{3}$$

(neglecting variations in the Coulomb logarithm).

Temperature Profile Peaking Factor

The lack of electron temperature profile variation with respect to changes in plasma parameters is an active research topic. In its purest form [9], profile consistency postulates a Gaussian shape (for consistency with [9] we use a dimensioned radial coordinate here):

$$T_e(r) \simeq T_e(0) \exp(-\frac{2}{3} b \frac{r^2}{a^2}) \tag{4}$$

Making the classical resistive equilibrium ansatz $\eta_{\|} \propto T^{-3/2}$, except for weak radial variations (due, in our case for example, to a possible Z_{eff} dependence on the radius), the current density profile then obeys

$$J_{\|}(r) \simeq J(0) \exp(-b \frac{r^2}{a^2}) \tag{5}$$

The constraint $\int J(r) 2\pi r dr = I$, the plasma current, together with the relations $q_a = \frac{2\pi a^2 B_{tor}}{\mu_0 I R}$ and $q_0 = \frac{2 B_{tor}}{\mu_0 R J_0}$ then leads to the approximation $b \simeq q_a/q_0$ (where $\exp(-q_a/q_0) \ll 1$ is assumed). The local shape parameter $L_{T_e}^{-1} = \frac{d}{dr} ln T_e(r) = 4 q_a r / 3 q_0 a^2$ for this model is accordingly predicted to be proportional to both r and, at any fixed r, to q_a (assuming fixed q_0 for sawtoothing profiles). As will be clear later from the profile parameterisation results, both these predictions are in disagreement with ASDEX Ohmic electron temperature profiles, the second one grossly so. Instead

of assuming the particular profile shape (4), we consider the following more general relationship between the temperature and the q profile,

$$q(r) = \frac{rB_{tor}}{RB_{pol}(r)} = \frac{2\pi r^2 B_{tor}}{\mu_0 R I(r)} = \frac{2B_{tor}}{\mu_0 R < J >_r} \tag{6}$$

where $< J >_r = \frac{1}{\pi r^2} \int_{r'=0}^{r} J(r') 2\pi r' dr'$. Assuming classical resistive equilibrium, we immediately have

$$\frac{< T^{3/2} >_{r_1}}{< T^{3/2} >_{r_2}} \simeq \frac{q_{r_2}}{q_{r_1}} \tag{7}$$

where the inexactitude reflects the radial variations already mentioned. Note that no assumptions are made here about the functional form of the temperature profile.

Apart from the edge safety factor q_a, the only other independent information on the q profile is the assumption that, since all profiles are sawtoothing, $q(r < r_{q=1}) \simeq q_0$, a constant for all profiles with a value close to unity. Thus, with the sole assumption of resistive equilibrium, we would expect the peaking ratio (7) evaluated at $r \simeq 0$ and $r = a$ to scale as

$$\frac{< T^{3/2} >_{r\simeq 0}}{< T^{3/2} >_{r=a}} \simeq \frac{q_{r=a}}{q_{r\simeq 0}} = q_a/q_0 \tag{8}$$

and to be independent of other bulk parameters. Since the YAG channel closest to the magnetic axis lies typically on the 14% flux surface, T_0 is essentially an extrapolated quantity. We present, therefore, scaling results for the LHS of expression (8) at both $r = 0$ and $r = .2a$ where the latter radius satisfies the double requirement of being safely inside the inversion radius for all profiles as well as lying in the data region. Due to the averaging procedure, however, it is also affected somewhat by the extrapolated section of the profile. For the high density database we obtained

$$\frac{T_0^{3/2}}{< T^{3/2} >_{r=a}} = 1.02(\pm.06)\overline{n_e}^{\,.03\pm.02} I_p^{\,.11\pm.09} q_a^{\,1.13\pm.06} \quad \text{RMSE} = .06 \quad R^2 = .874$$

$$\frac{< T^{3/2} >_{r=.2a}}{< T^{3/2} >_{r=a}} = .99(\pm.04)\overline{n_e}^{\,.03\pm.015} I_p^{\,.07\pm.06} q_a^{\,1.07\pm.04} \quad \text{RMSE} = .04 \quad R^2 = .931$$

while the moderate density database yielded

$$\frac{T_0^{3/2}}{< T^{3/2} >_{r=a}} = .96(\pm.17)\overline{n_e}^{\,.15\pm.11} I_p^{\,-.12\pm.26} q_a^{\,1.07\pm.14} \quad \text{RMSE} = .10 \quad R^2 = .795$$

$$\frac{< T^{3/2} >_{r=.2a}}{< T^{3/2} >_{r=a}} = 1.09(\pm.10)\overline{n_e}^{\,.13\pm.07} I_p^{\,-.07\pm.10} q_a^{\,.95\pm.09} \quad \text{RMSE} = .06 \quad R^2 = .883$$

Since for all four regressions, the $\overline{n_e}$ and I_p coefficients are at best marginally significant at the 95% level, the results indeed strongly indicate that the peaking factor (8) is determined solely by q_a . We now assume a sole q_a dependence and re-do the

regressions to get a closer confidence interval for the constant factor and the exponent. For the high density database we get

$$\frac{T_0^{3/2}}{<T^{3/2}>_{r=a}} = .94(\pm.04)q_a{}^{1.07\pm.04} \quad \text{RMSE} = .06 \quad R^2 = .866$$

$$\frac{<T^{3/2}>_{r=.2a}}{<T^{3/2}>_{r=a}} = .93(\pm.03)q_a{}^{1.03\pm.03} \quad \text{RMSE} = .04 \quad R^2 = .925$$

while the moderate density database yielded

$$\frac{T_0^{3/2}}{<T^{3/2}>_{r=a}} = .85(\pm.11)q_a{}^{1.13\pm.10} \quad \text{RMSE} = .10 \quad R^2 = .780$$

$$\frac{<T^{3/2}>_{r=.2a}}{<T^{3/2}>_{r=a}} = .97(\pm.07)q_a{}^{.99\pm.07} \quad \text{RMSE} = .07 \quad R^2 = .870$$

In none of the four cases can the null hypothesis that the q_a exponent is unity be rejected at the 95% confidence level. The second regression yields the narrowest 95% confidence interval: $.97 \leq q_a$ exponent ≤ 1.09 . The constant factors similarly are not inconsistent with a value of unity (with the marginal exception of the second regression: $.87 \leq$ const $\leq .99$). These results are consistent with, but more specific than the inequality

$$q_a^{2/3} \leq \frac{T(0)}{<T>} \leq q_a \tag{9}$$

derived by Waltz et al [10] also using classical resistivity plus sawtoothing. Since we are working with $< T^{3/2} >$ rather than $< T >^{3/2}$ as well as with sawtooth-averaged profiles, the upper and lower bounds of the inequality converge in our case. For completeness, we regressed the ratio of the two alternative normalising quantities on q_a to obtain, for the combined database,

$$\frac{< T^{3/2} >_{r=a}}{< T >^{3/2}_{r=a}} = 1.01(\pm.005)q_a{}^{.10\pm.005} \quad \text{RMSE} = .01 \quad R^2 = .699$$

The use of $< T >^{3/2}$ as normalising factor in the peaking factor regressions would accordingly have added $\simeq .10$ to each q_a exponent. It is interesting to note that, for the more accurate high density database regressions, the use of this, in our opinion less appropriate, normalising factor would have lead to the rejection of the null hypothesis of a q_a exponent equal to unity and hence to inconsistency with (8). We believe these results constitute striking evidence that the resistive equilibrium ansatz is sufficient to explain quantitatively the peaking factor, that is the global shape parameter, for ASDEX Ohmic temperature profiles. In later sections of this paper we investigate how the bulk parameter dependence of $L_{T_e}^{-1}$, the local shape parameter, varies radially.

Loop Voltage/Ohmic Power

We present the loop voltage scalings. The Ohmic power scalings differ from these by exactly one power of I_p. For the high density database, the loop voltage scales as

$$V_{loop} = 2.38(\pm 0.1)\overline{n_e}^{\,.36\pm.02} I_p^{\,.12\pm.05} B_t^{\,-.27\pm.04} \quad \text{RMSE} = .04 \quad R^2 = .864$$

Similarly, the moderate density database satisfies

$$V_{loop} = 2.40(\pm 0.1)\overline{n_e}^{\,.30\pm.02} I_p^{\,.19\pm.03} B_t^{\,-.31\pm.03} \quad \text{RMSE} = .03 \quad R^2 = .896$$

Due to the additional factor of I_p on both sides, we get necessarily higher R^2 values (.935 and .984) for the Ohmic power regressions.

Spitzer Z effective

For the high density database, the Spitzer Z_{eff} scales as

$$Z_{eff} = 2.90(\pm 0.18)\overline{n_e}^{\,-.48\pm.02} I_p^{\,.67\pm.05} B_t^{\,-.34\pm.05} \quad \text{RMSE} = 0.05 \quad R^2 = .893$$

Regressing the impurity contribution to Z_{eff} gives an entirely different scaling:

$$Z_{eff} - 1 = 2.21(\pm 0.33)\overline{n_e}^{\,-1.14\pm.04} I_p^{\,1.51\pm.12} B_t^{\,-.81\pm.12} \quad \text{RMSE} = .18 \quad R^2 = .844$$

For moderate densities, the Spitzer Z_{eff} scales as

$$Z_{eff} = 2.67(\pm 0.58)\overline{n_e}^{\,-.69\pm.08} I_p^{\,.64\pm.11} B_t^{\,-.22\pm.13} \quad \text{RMSE} = .09 \quad R^2 = .714$$

The impurity scaling yields

$$Z_{eff} - 1 = 1.83(\pm 0.71)\overline{n_e}^{\,-1.10\pm.14} I_p^{\,1.07\pm.19} B_t^{\,-.43\pm.21} \quad \text{RMSE} = .16 \quad R^2 = .709$$

Using the relation $n_{H;D} = n_e - \sum_i n_i Z_i$ (i-summation over impurity species only), $Z_{eff} \equiv \sum_j n_j Z_j^2 / n_e$ (j-summation over all species) can be re-expressed as

$$Z_{eff} = 1 + \frac{\sum_i n_i (Z_i^2 - Z_i)}{n_e} \tag{10}$$

With this representation, we see that the $\overline{n_e}$ exponents in the impurity scalings plausibly suggest an impurity density (almost) independent of the line density. The strong I_p and B_t scalings are not so readily interpretable.

Energy Confinement Time

The I_p and $\overline{n_e}$ dependencies of the total plasma energy and the Ohmic power approximately cancel to give a relatively weak dependence in τ_E. This makes the energy confinement scalings less pronounced. The high density database yielded the following scaling for the kinetic τ_E :

$$\tau_{Ekin} = 56(\pm 5)\overline{n_e}^{\,.18\pm.03} I_p^{\,-.21\pm.08} B_t^{\,.34\pm.08} \quad \text{RMSE} = .07 \quad R^2 = .392$$

The moderate density database fit yielded

$$\tau_{Ekin} = 53(\pm 6)\overline{n_e}^{\,.23\pm.05} I_p^{\,-.39\pm.06} B_t^{\,.44\pm.07} \quad \text{RMSE} = .05 \quad R^2 = .709$$

The equal and opposite tendency of the I_p and B_t coefficients, especially for the moderate density database, suggests a $\overline{n_e}\ I_p\ q_a$ or a $\overline{n_e}\ B_t\ q_a$ representation. We show the former:

$$\tau_{Ekin} = 72(\pm 5)\overline{n_e}^{\,.18\pm.03} I_p^{\,.12\pm.11} q_a^{\,.34\pm.07} \qquad \text{(High density database)}$$

$$\tau_{Ekin} = 74(\pm 6)\overline{n_e}^{\,.22\pm.05} I_p^{\,.05\pm.08} q_a^{\,.43\pm.07} \qquad \text{(Moderate density database)}$$

The confinement time derived from the diagmagnetic measurement of the energy content for the high density database scales as

$$\tau_{Edia} = 73(\pm 4)\overline{n_e}^{\,.04\pm.02} I_p^{\,.17\pm.09} q_a^{\,.41\pm.06} \quad \text{RMSE} = .06 \quad R^2 = .406$$

The moderate density database fit yielded

$$\tau_{Edia} = 55(\pm 5)\overline{n_e}^{\,.10\pm.05} I_p^{\,.23\pm.10} q_a^{\,.65\pm.08} \quad \text{RMSE} = .06 \quad R^2 = .704$$

3. PROFILE SHAPE ANALYSIS

The temperature and density measurements are obtained using the ASDEX sixteen channel YAG Thomson scattering diagnostic [4] with a sampling rate of 60 Hz. This system consists of sixteen spatial channels located in the vertical plane at R = 1.63 m . They are spaced at approximately 4 cm intervals from $Z = .200$ m to $Z = -.394$ m. We did not use the 16^{th} channel which lies very close to or on the separatrix, as the measurement failed for the majority of profiles in this database. The radius (averaged over all profiles) of the flux surface passing through each channel is presented in column 2 of Tables III and V for the high and moderate density databases respectively.

3.1. Radial Representations

In this section, we discuss continuous radial representations of the plasma profiles. We consider data consisting of n separate 'compressed profiles' of a spatially varying plasma variable such as temperature or density, at p distinct radial points. Each 'compressed profile' is the average of about $m = 12$ consecutive measurements taken at 17 msec intervals. We do, however, make use of the uncompressed profiles for the purpose of estimating the channel-by-channel raw measurement fluctuations within each discharge. Thus our temperature data can be described by $T_{i,j}(r'_l)$, where $i = 1, ..., m$ labels the uncompressed timepoint, $j = 1, ..., n$ is the compressed profile index, and $l = 1, ..., p$ denotes the radial channel number. A preliminary transformation of

the physical measurement locations r_l' to the corresponding flux surface radii r_l, is performed.

Continuous representations have the following characteristics: a) A large number of dependent variables, represented by point data, is replaced by a reasonably small number of coefficients which nevertheless will be sufficient to represent all the principal features of the profiles. b) Profiles measured at two different sets of radial locations may be compared. This is relevant, e.g., where we wish to compare YAG temperature measurements with electron cyclotron emission (ECE) data measured at different spatial locations. c) Smoothness is imposed in the belief that the profiles are in diffusive equilibrium.

Instead of fitting the profile itself, we choose to fit its natural logarithm Y . Minimising the residual of the logarithm of the plasma profile corresponds to minimising the relative rather than the absolute error. Preliminary comparisons with low order spline or polynomial fits to the actual profiles revealed that the logarithmic fit tended to have not only smaller residual errors on the logarithmic scale but also on the usual physical scale. This indicates that the 'exponentiated form' of the logarithmic fit is a better approximation to the actual plasma profiles than a comparable low order fit to the linear profile. We note that the difference between logarithmic and linear fits decreases as more fit parameters (either spline knots or higher order polynomial terms) are added. Logarithmic fits have several other advantages. Firstly, the predicted profile can never be negative. Secondly, well-known power law–type scalings reduce to linear models. Finally, if the noise level is proportional to the absolute value of the measurement (an admittedly idealised situation), then, on the logarithmic scale, unweighted least squares may be used.

Spline representations, which we employ here for profile parameterisation, give flexibility in choosing between local resolution and compact global representation. In addition, the parametric dependencies of the spline coefficients are less prone to influence by non-local features than polynomial representations. The profile parameterisations presented in this article are based on a twice continuously differentiable spline with selectable knot count ν , which is forced to be parabolic inside the first knot. The radius is decomposed into $\nu + 1$ regions with knots at r_1, r_2, . . . , r_ν . This profile may be parameterised explicitly by:

$$\mu(r) = \begin{cases} \mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 & \text{for } 0 \le r \le r_1 \text{ (Inner Region)} \\ \mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 + c_1(r - r_1)^3 & \text{for } r_1 < r \le r_2 \text{ (Region 1)} \\ \mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 + c_1(r - r_1)^3 & \\ \quad + c_2(r - r_2)^3) & \text{for } r_2 < r \le r_3 \text{ (Region 2)} \\ \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 + c_1(r - r_1)^3 & \\ \quad + c_2(r - r_2)^3) + \ldots + c_\nu(r - r_\nu)^3 & \text{for } r_\nu < r \le 1 \text{ (Outer Region)} \end{cases} \tag{11}$$

In more compact fashion, we express the spline $\mu(r)$ as

$$\mu(r) = \sum_{k=1}^{\nu+2} \alpha_k H(r_k) \varphi_k(r) \tag{12}$$

where ν is the number of knots used, α_k represent the spline parameters:

$\mu(r_1), \mu''(0), c_1, c_2, \ldots c_\nu$; $\varphi_k(r)$ are the polynomials: $\varphi_1(r) = 1$, $\varphi_2(r) = (r^2 - r_1^2)/2$, $\varphi_3(r) = (r - r_1)^3, \ldots \varphi_{\nu+2}(r) = (r - r_\nu)^3$, and

$$H(r_k) = \begin{cases} 0 & \text{for } r < r_k \\ 1 & \text{for } r \ge r_k \end{cases} \tag{13}$$

is the Heaviside unit stepfunction ($r_k = 0, 0, r_1, r_2, \ldots r_\nu$) . As well as the parabolic restriction near the axis, a 'natural' spline boundary condition, $\mu''(1) = 0$, can be imposed. This boundary condition is implemented in the SAS REG procedure using the RESTRICT statement [11] .

3.2. Parametric Dependencies of Plasma Profiles

Since the plasma profile shapes depend on the bulk plasma variables such as q_a, the spline coefficients will be functions of these bulk plasma variables. Since our database is not expected to contain sharp transitions in behaviour in parameter space, a low order parametric representation will be adequate. Therefore we approximate the smooth parametric dependencies of the profile shape by linear or possibly quadratic polynomials in the logarithms of the bulk variables.

Let $\vec{x}$ be the vector consisting of the bulk plasma variables $\overline{n_e}, q_a, I_p$, and let x_j be any component. We define the linear basis functions $g_0(\vec{x}) = 1$ (intercept), $g_j(\vec{x}) = ln(x_j/x_j^*)$, $j = 1,3$; where x_j^* is a representative value of the variable x_j in the database of interest. For ease of comparison, we choose for both databases the same normalising values of $q_a^* = 2.5$, $B_t^* = 2.0$ T, $I_p^* = .4$ MA and $\overline{n_e}^* = .4 \times 10^{20}\mathrm{m}^{-3}$, although these do not constitute a typical parameter set for the moderate density database. By normalising the bulk variables to x_j^*, the value of the response variable at the intercept in the regression becomes the predicted value at $\vec{x} = \vec{x}^*$.

To estimate the parametric dependencies of the profiles, we represent them as

$$Y(r,\vec{x}) = \sum_{k,j} \alpha_{k,j} g_j(\vec{x}) H(r_k) \varphi_k(r) \tag{14}$$

Higher order terms can be represented, if necessary, by extending the set of possible basis functions: $g_{j,l}(\vec{x}) = ln(x_j/x_j^*)ln(x_l/x_l^*)$.

3.3. Error Structures:

We distinguish between several categories of random profile variations. By time point to time point variations, we denote those processes which fluctuate on a time scale at least as fast as that of the diagnostic sampling rate. These include statistical noise from the measuring process and plasma fluctuations arising, in particular, from the $m = 1$ sawtooth instability. Discharge to discharge variations are changes in the plasma profiles not observed within a single discharge. These discharge variations include effects such as impurity accumulation on the diagnostic windows and the condition of the plasma wall. These impurity and plasma wall effects tend to vary to an even larger extent from one experimental operating period to the next. In addition, discharges separated by a recalibration of the YAG system can exhibit systematic differences in the measured profiles.

As discussed in appendix B, this hierarchy of temporal scales for plasma variation generates a compound error structure which can be treated statistically. For simplicity, we analyse the time averaged data and neglect the compound error structure. Since the profiles in our database are already averaged over twelve consecutive time samples (see introduction), they no longer possess the same variance as the original uncompressed observations. For time averaged datapoints, $\overline{Y}(r,\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} Y_i(r,\vec{x})$, the total unexplained variance can be decomposed into:

$$n\hat{\sigma}_{tot}^2 = n\frac{\hat{\sigma}_{int}^2}{m} + N_{dis}\hat{\sigma}_{dis}^2 \tag{15}$$

where $\hat{\sigma}_{int}$ is an estimate for the standard deviation of the within-discharge or 'internal' fluctuations, m is the number of time points in the compressed profile (in our case $m = 12$), n is the number of compressed datapoints, N_{dis} is the number of distinct discharges and $\hat{\sigma}^2_{dis}$ is an estimate for the variance due to discharge-to-discharge changes.

To estimate the within-discharge variance $\hat{\sigma}^2_{int}$, we analysed the original uncompressed data and, for each channel, calculated the empirical variance for each 12-point set separately. We estimated $\hat{\sigma}^2_{tot}$ by regressing the set of n (compressed) datapoints for each individual measurement channel against the bulk variables and noting the unexplained variance. The difference, as given by the third term in (13), is an estimate of the discharge-to-discharge variance ($\hat{\sigma}^2_{dis}$ is calculated for the high density database, for instance, using $n = 105$ and $N_{dis} = 50$) . Columns 3 and 4 of Tables III(b)-VI(b) display $\hat{\sigma}_{int}/\sqrt{12}$ and $\hat{\sigma}_{tot}$.

3.4. Coefficient Estimation

The weighted least squares formulation of the problem is the determination of that vector of coefficients $\underline{\alpha}$ which minimises the expression:

$$\sum_{j,l}\Big(Y_{j,obs}(r_l) - Y_{j,fit}(r_l,\underline{\alpha})\Big)^2 W(r_l) \tag{16}$$

where $W(r_l)$, $l = 1,...,15$ are appropriately chosen weights for each of the 15 YAG channels. We investigated two approaches for determining $W(r_l)$. In the first method, we rely on the total unexplained variance for each channel (as discussed in the previous subsection) as a measure of the channel weighting: $W(r_l) = \hat{\sigma}^{-2}_{tot}(r_l)$. A second approach to the selection of the regression weights, is the iterative estimation of the residual variance of the spline fit at each channel:

$$\hat{\sigma}^2_{(n+1)}(r_l) = \frac{1}{N}\sum_j\Big(Y_{j,obs}(r_l) - Y_{j,fit(n)}(r_l,\underline{\alpha}) - \delta\overline{Y}(r_l)\Big)^2 \tag{17}$$

where $\delta\overline{Y}(r_l)$ is a possible systematic bias in fitting the l^{th} measurement channel which can be estimated by including an indicator dummy variable for each of the 15 channels in the regression. The inclusion of this term prevents undue downweighting of channels where the parameterised profile may consistently fail to match the observed data. For the first iteration, the variances $\hat{\sigma}^2_{(1)}(r_l)$, are initialised to one (equal weights). The iteration is terminated after the third iteration. We then have $W(r_l) = \hat{\sigma}^{-2}_{(last\ iter.)}(r_l)$. If the regression model and the assumed error structure is correct, this is a likely to be

a consistent and efficient estimate (see e.g. [12]). Nevertheless, we regard $\hat{\sigma}_{tot}^{-2}(r_l)$ as a more robust estimate than $\hat{\sigma}_{(last\ iter.)}^{-2}(r_l)$ since it depends only on the bulk variables whereas the latter estimate also depends on the spline model and has the additional problem of the strong (anti)correlation of outer channel residuals. Therefore we prefer to use $W(r_l) = \hat{\sigma}_{tot}^{-2}(r_l)$. The dominant effect of this reweighting is to decrease the influence of the channels near the plasma boundary where the relative error is largest.

3.5. Criteria for Additional Free Parameters

Some relevant statistical tests for the significance of an additional variable are discussed in KRML. In the case of independent errors, these criteria are given by the F test [1] and Mallows C_p statistic [11] . Mallows C_p statistic is the sum of the total bias in the regression and the total variance of the predicted values. As more free parameters are added, the bias decreases but the variance increases. To determine whether to add another parameter, one can minimise the C_p statistic or apply the F test.

In practice, these statistical approaches often either under- or overestimate the significance of the additional variables due to the fact that correlations in the errors are neglected. In addition, these tests neglect systematic errors and assume that the ideal data, without measurement errors, is exactly describable by the regression equation under consideration. Thus many spurious dependencies may be included and real dependencies missed by unthoughtful or automated use of these methods.

4. PROFILE PARAMETERISATION TECHNIQUES

In this section, we present a number of techniques to improve the representation of profile shapes. Logarithmic representations are employed for the reasons outlined in section III.

4.1. Exponential Form for Low Order Representations

Initial efforts concentrated on fitting polynomial representations of the form:

$$T(r) = T_0 exp(ar^2 + br^4 + cr^6) \tag{18}$$

This model was successful in reproducing the general properties of the ASDEX profiles, but not detailed features. Sharp gradients and local flattenings, perhaps due to magnetic islands, were poorly modelled. A disadvantage of the above model, exacerbated by the addition of even higher order polynomial terms, is the insensitivity of the inner region to r^4 and higher powers as well as the high degree of stiffness of these polynomials. To enable us to fit each region of the profile in moderate powers of the radial

coordinate, albeit at the price of limited continuity at knot boundaries, we turned to spline representations.

4.2. Once and Twice Continuously Differentiable Cubic Splines

Following [13] , we first used a five parameter, two knot Hermitian spline, i.e. one with no continuity requirement on the second derivative. This turned out to be significantly better in parameterising steep gradients and abrupt spatial transitions in profile shape than the polynomial model. By experimentation, we found that a total of four knots, requiring seven regression parameters, gave a practical balance between fitting accuracy and significance of the spline coefficients.

A serious disadvantage to the Hermitian spline emerged, however. By allowing discontinuous second derivatives at the knots, continuous transitions in plasma behaviour were modelled as sharp transitions across the knot boundary. This effect was especially prominent in the slope of the inverse fall-off length as a function of q_a. After some investigation, we opted instead for the twice differentiable spline model (11) described in section 3.1 above.

4.3. Spline Knot Locations and Boundary Conditions

The knot positions are chosen such that the measuring channels are distributed roughly equally in the various regions between and outside of the knot positions. Too many knots result in spuriously oscillatory fits. The knot locations were varied manually to achieve a near 'optimal' fit as determined by the balance between goodness of fit and significance of the fit coefficients.

The innermost channel is typically located at $r = .14$ and the outermost channel at $r = .89$ (Table III column 2). When third degree polynomials were used in the innermost and outermost regions, the extrapolated curves (to $r = 0$ and $r = 1$ respectively) had unphysical oscillations. These oscillations were eliminated by reducing the number of free parameters for these regions. Near the origin, the profile was forced to be parabolic. The natural boundary condition, $\mu''(1) = 0$, was applied at $r = 1$, but only to the bulk parameter–dependent spline coefficients. Imposing it on the intercept coefficients causes a strong increase in the overall regression RMSE. We settled on the following 5-knot set: $r_{knot} = .2, .3, .4, .5, .65$. This yields a model with seven spline coefficients per bulk parameter basis function used in the fit, plus boundary conditions. The profile parameterisations presented later were carried out using the linear basis functions $g_j(\vec{x})$; $j = 0, 3$ only. Some quadratic and cross terms were very significant in preliminary regressions involving both databases simultaneously. For reasons given in the introduction, however, the results we present come from separate

profile shape analyses for each database. For these regressions, second order terms were rarely significant and the goodness of fit was scarcely affected by the restriction to linear terms. Using the three bulk parameters $\overline{n_e}$, I_p and q_a , we have a regression model with a total of (intercept + 3 bulk parameters) $\times 7 - 3$ restrictions, i.e. 25 degrees of freedom to fit $15 \times 105 = 1575$ individual temperature (or density) data. With the foregoing boundary conditions, this spline representation tended to be rather stable in extrapolating profile behaviour into regions where there were no measurement channels.

4.4. Normalisation

The goal of our profile analysis is to determine the radial parametric behaviour of the scale-independent local shape parameters $L_{T_e}^{-1} = \frac{1}{T_e}\frac{d}{dr}T_e$ and $L_{n_e}^{-1} = \frac{1}{n_e}\frac{d}{dr}n_e$. This is accomplished by first parameterising the logarithmic profiles themselves. This brings up the problem of fitting the profile magnitude. We found that unless we normalised the profiles before parameterisation, the residual sum of squares was dominated by the uncertainty in predicting this nuisance parameter.

Originally, we normalised each profile by its line-average, calculated from the spline fit. This normalisation greatly reduces but does not minimise the residual error in the profile parameterisation since the line-average is itself a function of the profile shape. Instead, we estimated the profile size parameters, using the SAS procedure GLM [11] , by treating the profile index as an indicator variable. This yielded as normalising factor the only radially independent parameter in our spline repesentation, i.e. $\mu(r_1)$, the profile value at the first knot which is sited at $r = .2$. Normalisation has the effect of reducing by one the number of degrees of freedom for a given spline model. This led to the number of degrees of freedom for the model consisting of a 5-knot spline with coefficients dependent on 3 bulk parameters described in the previous subsection being reduced from 25 to 21 for normalised profiles.

4.5. Operating Period Indicator Variable

In the course of determining $\hat{\sigma}^2_{tot}(r_l)$ for each YAG channel, plots of residuals versus shot number revealed that the moderate density database residuals, whose discharges spanned a period of over six months in contrast to the one week span of the high density database, fell into four distinct groupings which we ascribe to four distinct experimental operating periods (see introduction). This 4-cluster formation was observed for all 15 channels, although the pattern formed by the clusters differed for each regression. To remove this operating period contribution to the overall unexplained channel variances for this database, and hence to enable a comparison to be made between the two databases, separate indicator variables for each operating period were added in the

individual channel regressions used to determine $\hat{\sigma}^2_{tot}$. These indicator variables were not, however, included in the spline parameterising regressions. To do so would have required an additional 60 independent variables (4 for each channel) which, in our judgement, would have led to overfitting of the profiles.

4.6. Examination of Outliers

Apart from the operating period effects mentioned above, plots of raw versus fitted data for the same channel by channel regressions revealed that a small number of individual channel measurements from both databases produced strongly outlying residuals (the worst case was one of 8.5 standard deviations). To arrive at a quantitative criterion for identifying suspect data, we analysed the Studentised residuals. A Studentised residual is the error in the fitted value normalised to the RMSE. For normally distributed errors, they have approximately a standard normal distribution. If we consider a single Studentised residual, the probability that it lies outside $\pm c$ is ϵ , where $\epsilon \simeq 1 - (2\pi)^{-\frac{1}{2}} \int_{-c}^{+c} e^{-\frac{x^2}{2}} dx$. Considering now n uncorrelated residuals together, we have that the probability of all n lying *inside* $\pm c$ is $(1-\epsilon)^n$. Hence the probability of at least one among n Studentised residuals lying outside $\pm c$ is given by $1-(1-\epsilon)^n = \beta$, say. Provided the correct model is used to fit the data, we suspect any outlier whose Studentised residual exceeds $\pm c_\beta$ for a suitably small β (using $\epsilon = 1-(1-\beta)^{\frac{1}{n}} \simeq \beta/n$ for $\beta << 1$, we invert the probability integral to determine c_β). We chose $\beta = 1\%$ which, for $n = 105$ and $n = 38$, yields $c = 3.90$ and $c = 3.64$ for the high and moderate density databases respectively. Using this criterion, we identified 6 suspect outliers from the high density database and 18 from the moderate density, corresponding to 0.2% and 1.5% of the data respectively.

Profiles containing any suspect observations were now examined individually. In most cases it was visually obvious that the affected channel was inconsistent with the rest of the profile, and such observations were considered to be bad data. One discharge accounted for the majority of the suspect data in the moderate density database. On inspection, it was clear that the quality of the profile data for this discharge was so poor that it was excluded entirely from the subsequent analysis, thereby reducing the number of discharges from 38 to 37 for this database. On the other hand, several suspect observations from a single profile in the high density database were not visually inconsistent with the rest of that profile's data. On investigating further, it turned out that this discharge had the highest B_t value (2.73 T) in the database. This highlights the need to examine all suspect outliers individually, since the influential position of this data suggests an inadequacy in the model used to fit the data rather than in the data itself, and its rejection would be quite unjustified. The small number of observations

deemed to be faulty were deleted and replaced by their respective predicted values to maintain the block structure of the design matrix.

4.7. Measurement Asymmetries

Ten of the fifteen YAG channels in ASDEX are located in nearly symmetric positions with respect to the horizontal midplane. By examining the residual errors for each channel separately, an up-down asymmetry was found in the density profile (up-down difference $\simeq 7\%$) The asymmetry was nearly uniform on all five pairs of measurement channels and was independent of plasma parameters. No significant asymmetry was found in the temperature profiles.

A possible explanation for this asymmetry is a spatially nonuniform distribution of impurities on the diagnostic window. Assuming that for the range of scattered laser light detected by the system ($.8\mu m \leq \lambda_{scatt} \leq 1.06\mu m$), the impurity accumulation causes a spectrally uniform reduction in transmission, then the electron temperature, which is calculated from the ratio of the scattering signals of two spectral channels at the same spatial location, is relatively insensitive to such asymmetries and only the density is affected.

To estimate and correct for this asymmetry, we expanded the set of bulk parameter dependent radial basis functions to include a single asymmetry indicator variable which takes the value of one for the channels above the midplane and minus one below the midplane. The final regression model, the results of which are presented in the next section, accordingly increases by one degree of freedom to give a total of 22.

5. EXPERIMENTAL RESULTS

5.1. Temperature

Table III(a) presents the descriptive statistics of the temperature profile for the high density database on a channel by channel basis. Table V(a) displays the equivalent information for the moderate density database. The channels are numbered according to their vertical position with channel 1 at $Z = .200$ m, channel 6 at $Z = 0$ m and channel 15 at $Z = -.353$ m. For each channel, the mean normalised flux radius is presented, followed by the mean temperature and the spread in keV. The spread (this term is chosen to avoid any confusion with 'standard deviation' in the sense of 'regression error') is just the 'standard deviation from the mean': $spread = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})^2}$, where $\bar{y}$ is the sample mean. Recall that before the regression, the profiles were normalised by the size parameter from the GLM procedure, i.e. $T_e(r = .2)$ obtained from

fitting each profile individually. Columns 5 and 6 present the mean and spread of these normalised profiles. Channels 5 and 8 have very small spreads since they lie closest to the normalisation radius.

As we are fitting on the natural logarithmic scale, it is of particular interest to tabulate the logarithmic or relative spread of the normalised profiles as a measure of the total variation of the data for each channel. This quantity, together with the noise level estimates and the results of our parameterisation of ASDEX Ohmic temperature profiles for the high density database is presented in Table III(b) (or V(b) for the moderate density database). $\hat{\sigma}_{int}/\sqrt{12}$, the magnitude of the internal, within-discharge, relative fluctuations scaled for the time averaged database profiles, is displayed in column 3. Column 4 tabulates $\hat{\sigma}_{tot}$, the total fluctuation level on the relative scale.

The ratio of the two noise estimates lies in the range $.7 \leq \hat{\sigma}_{int}/\hat{\sigma}_{tot} \leq 1$ indicating that the inferred discharge to discharge variance will be smaller than $\hat{\sigma}_{int}$. Using the RMS values for all 15 channels we find from eq. (15) that

$$\hat{\sigma}_{dis}(rms) \simeq \sqrt{(\hat{\sigma}^2_{tot}(rms) - \hat{\sigma}^2_{int}(rms)) \cdot n/N_{dis}} = \sqrt{(.028^2 - .025^2) \cdot 105/50} = .018$$

For the moderate density database, $.5 \leq \hat{\sigma}_{int}/\hat{\sigma}_{tot} \leq .9$ and, using eq. (15), $\hat{\sigma}_{dis}(rms) \simeq .030$ so that, in contrast to the high density database, the total unexplained variance here is dominated by $\hat{\sigma}_{dis}$. This result holds, despite having used operating period indicator variables (in determining $\hat{\sigma}_{tot}$) to remove the variance contribution due to the four operating periods spanned by the moderate density database (see section 4.5). Table III(b) column 5 shows the (relative) RMSE calculated for each individual channel from the parameterised spline fit to the profiles. The generally close agreement with column 4 ($\hat{\sigma}_{fit}(rms) = .031$; $\hat{\sigma}_{tot}(rms) = .028$) indicates the adequacy of 5-knot spline model. For the moderate density database, in contrast, the fitting errors are, in general, much larger than the noise levels ($\hat{\sigma}_{fit}(rms) = .063$; $\hat{\sigma}_{tot}(rms) = .039$). This is explained by the fact that, unlike the $\hat{\sigma}_{tot}$ channel-by-channel regressions, we did not use operating period indicator variables in the spline parameterisation of the moderate density database profiles (section 4.5). The average channel regression errors on absolute scale for the normalised profiles may be calculated by multiplying column 5 by the appropriate elements in Table III(a) (or V(a)) column 6. The RMSE values for all 15 channels on the absolute scale are .024 and .042 for the high and moderate density databases respectively.

In Figs. 2 and 3 we present reference profiles and parametric dependencies of the minor radius–normalised (negative) inverse fall-off length (IFOL) as a function of radius. Dashed curves indicate local 95% confidence bands (see KRML for the calculation procedure as well as a discussion on local and global confidence bands). By reference

profile we mean the evaluation of the parameterised spline fit at the representative set of parameter values $\vec{x} = \vec{x}^*$ as discussed in section 3.2. Since the vector of basis functions $\vec{g}(\vec{x}) = (1, ln(x_1/x_1^*), ln(x_2/x_2^*), ln(x_3/x_3^*))$ reduces in this case to $\vec{g}(\vec{x}^*) = (1,0,0,0)$, the reference profile is just that described by the set of intercept spline coefficients. The parametric radial dependencies of the IFOL are obtained by differentiating the parameterised spline representation for the normalised profiles both spatially and with respect to the basis function of the desired parameter. Fig. 2(b), for instance, shows the radial behaviour of the q_a dependence which, using eq. (14) (with $g_1(\vec{x}) = ln(q_a) - ln(q_a^*)$), is given by:

$$\frac{\partial(-L_{T_e}^{-1})}{\partial ln(q_a)} = -\frac{\partial^2 Y(r,\vec{x})}{\partial g_1 \partial r} = -\frac{\partial}{\partial r}\sum_k \alpha_{k,1} H(r_k)\varphi_k(r) \qquad (19)$$

We see that at $r \simeq .35$, where the profile shape is most sensitive to q_a, a unit change in $ln(q_a)$ causes a change of $\simeq 5$ in the negative IFOL, corresponding to 12.5 m^{-1} for an ASDEX minor radius of .4 m. Parametric dependencies of 'experimental' point values are also displayed. These are calculated by differencing the measurement values of pairs of neighbouring channels: $L_i^{-1} = \frac{(T_{i+1}-T_i)}{.5(T_{i+1}+T_i)(r_{i+1}-r_i)}$ and regressing each such set of n observations on $\overline{n_e}$, I_p and q_a . Such a point value has the advantage that it is more local than the continuous function represented by eq. (19), but the disadvantage that the signal to noise ratio will be lower. The mean value will also be particularly affected by systematic errors in one or both of the adjacent channels.

The minimum, reference, and maximum q_a IFOL profiles displayed in Fig. 2(a) show that the temperature profile shape for the high density database is remarkably invariant outside $r = .6$. This behaviour is broadly similar for the profiles of the moderate density database. The larger error bands of the latter, reflecting the substantially higher regression error for this database, are due in large measure to the already discussed problem of multiple operating periods for this database. The parametric dependencies of the IFOL profiles are detailed in the remaining sub-figures. The q_a dependence of the inverse fall-off length of both databases increases rapidly and reaches a maximum near $r = .35$. The strength of this q_a dependence then decreases rapidly, and outside $r = .6$ the profile shape is independent of q_a. Between $.15 \leq r \leq .25$ the point estimates for the 'experimental' IFOL differences suggest a radially uniform q_a dependence. However, the large error bars allow for a slope between, say, -10 and 10 on the scale of the plot. Inside the first knot, our radial spline model consists of a parabola with $T_0' = 0$. This requires the IFOL as well as each individual parameter dependency to describe a straight line through the origin in this region. The I_p and $\overline{n_e}$ dependencies are much weaker, although over some portions of the radius they marginally avoid rejection at the 5% level.

Figs. 4 and 5 show reference profiles and parametric radial dependencies for the normalised profiles $ln(\hat{T}_e(r))$ (where $\hat{T}_e(r) = T_e(r)/T_e(r = .2)$). These are equivalent to the integrals of Figs. 2 and 3, with an integration constant chosen to give $T_e(r = .2) = 1$. The invariance of the profile shape outside $r = .6$ is reflected here by the fact that all profiles are parallel in this region and all parametric dependencies (almost) horizontal.

The coefficients describing the spline representations of T_e, n_e and p_e for the high density database are collected in Table VII. All profiles in Figs. 2 and 4 can be generated using the first block of spline coefficients in Table VII together with eq. (14) and recalling the knot layout: $r_{knot} = .2, .3, .4, .5, .65$ (Note that the $\mu(r_1)$ coefficient is not tabulated as it vanishes for our normalised profiles). Specifically, the profiles in Fig. 4 are reproduced by directly substituting the spline coefficients into eq. (14). To recover the profiles of Fig. 2, eq. (14) must first be differentiated spatially. Thus the reference normalised profile in Fig. 4(a) is given explicitly by (see eqs. (12) - (14))

$$\ln(\hat{T}_{e(ref)}(r)) = -1.248(r^2 - .2^2) - 3.611(r - .2)^3 H(.2) + \ldots + 31.383(r - .65)^3 H(.65) ,$$

while the reference IFOL profile (the negative of Fig. 2(a)) is described by

$$L^{-1}_{T_e(ref)}(r) = 2(-1.248)r + 3(-3.611)(r - .2)^2 H(.2) + \ldots + 3(31.383)(r - .65)^2 H(.65) .$$

5.2. Density

Since many of the results presented in the last section apply to the density profiles as well, we only mention the differences. Tables IV and VI contain the density statistics for the high and moderate density databases respectively. For both databases, as can be seen from the RMS values for $\hat{\sigma}_{int}$ and $\hat{\sigma}_{tot}$, the discharge to discharge variations in the measured density are comparable to, or sometimes even dominate the within-discharge variance. The overall regression RMS relative error for each database (.029 and .060) is very similar to the corresponding temperature value. As explained in the temperature results, the error on the absolute scale (for normalised profiles) can be easily computed, giving RMSE values for all 15 channels of .024 and .042, again very similar to the temperature results.

The density profiles are extremely flat inside $r = .4$ as shown in Figs. 6-9. The main feature of the IFOL profiles is a q_a dependence closely mirroring that of the temperature IFOL, though at a reduced magnitude $\left(-\frac{\partial(L^{-1})}{\partial ln(q_a)}(max) \simeq 5;\ -\frac{\partial(L_n^{-1})}{\partial ln(q_a)}(max) \simeq 2\right)$. The radial region over which it is significant ($.25 < r < .55$ for the high density database) is more localised than the equivalent region for the temperature. Whereas, outside the half radius, the temperature IFOL is monotonically growing as a function of radius, the density IFOL peaks around $r = .75$ and the profile shape tends to become

flatter towards the boundary. As was the case with the temperature profile, the shape dependencies on I_p and $\overline{n_e}$ are weaker than that on q_a. Near the edge, however, there is a statistically significant broadening of the density profile shape with increasing current, enhancing the broadening tendency already remarked upon (the 'experimental' IFOL datapoints suggest that this current dependence is due solely to the outermost channel). In addition, some flattening of the density profile occurs in the region $.5 < r < .7$ with increasing $\overline{n_e}$.

A number of density profiles in both databases are slightly hollow in the region $.3 < r < .4$. The use of the same set of reference profile parameters ($q_a^* = 2.5$, $I_p^* = .4$ MA and $\overline{n_e}^* = .4 \times 10^{20}\mathrm{m}^{-3}$) for both databases exaggerates this aspect for the moderate density database (Figs. 7 and 9).

5.3. Electron Pressure

The analysis of the (logarithmic) electron pressure profiles, defined as $ln(P_e) = ln(T_e) + ln(n_e)$, gives additional insight, as can be seen from the parametric dependencies shown in Figs. 10-13. For the high density database, the most striking feature is that the I_p dependence is significant over most of the radius and there is a general broadening of the outer half of the pressure profile with increasing current, the effect being strongest towards the plasma boundary. This can be viewed as a result of the weaker I_p effects on the temperature and density profile shapes reinforcing one another. The q_a dependence of the inside half of the profile is similarly enhanced, whereas the $\overline{n_e}$ radial dependence is little changed from that of the density profile.

6. ANALYSIS FOR RADIALLY CORRELATED ERRORS

When the random errors are correlated, the weighted least squares estimator is consistent but not efficient, i.e. as the number of datapoints approaches infinity, the estimates of the regression coefficients converge to their true value but the rate of convergence is not optimal.

To increase the precision of the estimate, one tries to model the actual covariance matrix for the errors. The closer the assumed or estimated $\underline{\underline{\Sigma}}$ is to the actual error structure, the more accurate the ensuing estimates for $\underline{\alpha}$ are. We continue to assume that the statistical fluctuations are temporally uncorrelated and neglect the parametric dependencies. However we now allow radial correlations in the fluctuations. Since the dataset consists of $15n$ datapoints, the entire covariance matrix is $15n \times 15n$. However we assume a block diagonal form for $\underline{\underline{\Sigma}}$ of the form:

$$\Sigma^{tot}_{i,k,i',l} = \Sigma_{k,l}\delta_{i,i'} \tag{20}$$

where i, i' index the time-averaged profile and k, l index the channel number.

This covariance matrix of the residual radial errors may be estimated by:

$$\hat{\underline{\underline{\Sigma}}}_{k,l} = \frac{1}{n}\sum_{i}(Y_i(r_k) - \underline{Y}_{i,fit}(r_k, \underline{\alpha}))(Y_i(r_l) - \underline{Y}_{i,fit}(r_l, \underline{\alpha})) \tag{21}$$

In the previous sections, we have assumed $\underline{\underline{\Sigma}}$ is diagonal. To examine whether our initial hypothesis of independent errors is reasonable, we perform a principal components analysis on $\hat{\underline{\underline{\Sigma}}}$.

If the condition number (the square root of the ratio of the largest to smallest eigenvalues) or the ratio of the arithmetic mean to the geometric mean of the eigenvalues is approximately one, then the previous analysis, based on radially uncorrelated fluctuations, is justified.

If the condition number is substantially larger than one, then the statistical estimating efficiency can usually be improved by prescribing a functional form, $\underline{\underline{\Sigma}}(\underline{\theta})$ to model the observed covariance. We can simultaneously estimate $\underline{\alpha}$ and $\underline{\theta}$ using maximum liklihood estimates [1].

7. DISCUSSION AND SUMMARY

By comparing the high density and moderate density databases we have identified which features of the plasma have approximately the same scaling and which features scale differently. The scalings for the total plasma energy and electron temperature hardly differ between roll-over and saturated Ohmic confinement regimes. The total plasma energy and average temperature depend on the plasma current I_p and average density $\overline{n_e}$, but, at constant I_p and $\overline{n_e}$, are practically unaffected by the toroidal magnetic field. The nearly linear current dependence for both the temperature and the total energy is reminiscent of L mode scaling.

The strong dependencies on the plasma current I_p of both the total plasma energy Wp and the Ohmic power approximately cancel to give a relatively weak current scaling for τ_E. Replacing B_t by q_a gives a τ_E scaling weakly dependent on $\overline{n_e}$, strongly dependent on q_a and with no statistically significant I_p dependence.

The Spitzer Z_{eff} depends on all three control variables. Regression of $Z_{eff} - 1$ indicates, for both databases, that the impurity density is (almost) $\overline{n_e}$ indepependent, but strongly dependent on both I_p and B_t .

The electron temperature profile peaking factor $\frac{T_0^{3/2}}{<T^{3/2}>}$ scales as $.94(\pm.04)q_a{}^{1.07\pm.04}$, a result which is in close agreement with the prediction of classical resistive equilibrium.

A careful statistical analysis is necessary to determine the radially varying parametric dependencies of the profile shapes on the bulk plasma variables. By simultaneously fitting all profiles with spline coefficients which depend on the plasma variables, we believe we have found a reasonably effective method for solving this problem.

An earlier study of ASDEX temperature profile shapes [14] (for both Ohmic and neutral beam heated discharges) revealed that the shape depends strongly on q_a inside the sawtooth mixing radius, but is almost independent of plasma parameters outside $r = .45$. The results of our profile parameterisation are roughly consistent with, and constitute a refinement of this analysis, for Ohmic profiles.

Except for a dependency of the outer region of the density profile shape on plasma current, the I_p and $\overline{n_e}$ dependencies on both the temperature and density profile shape are rather weak in general. In most cases the current and density dependencies can be set to zero within the error bars of the dataset.

In the interior, q_a is the most important bulk plasma parameter in determining the temperature shape. By $r = .5$, however, this dependence has weakened considerably and outside $r = .6$, as is clear from Fig. 2, the IFOL profile has an invariant shape. We note that the extent of the q_a sensitive region is reasonably consistent with the widest sawtooth inversion radius in the database ($r_{inv}(max) \simeq 1/q_a(min) = .54$ and $.42$ for the high and moderate density databases respectively).

Comparing Figs. 6 and 2, we see that the variation in the density profile shape, while significant, is much weaker than for the temperature. This follows from the result that the density IFOL q_a sensitivity over the inner half of the radius is only $\simeq 40\%$ that of the temperature. The q_a dependence is only significant for $.25 \leq r \leq .5$. In contrast to the temperature shape, which is unique outside $r = .6$, the density profile broadens significantly near the edge with increasing current.

The electron pressure IFOL exhibits a very strong q_a dependence in the inner half of the profile, while increasing I_p causes a general broadening of the outer half, a tendency which intensifies approaching the plasma boundary.

Our findings are in agreement with the hypothesis of electron temperature profile consistency [13,14] . However, we have not addressed the relative merits of profile

consistency versus local transport models [9,15] containing sawtooth effects. This issue should be addressed by a statistical comparison of experimental profile dependencies with the dependencies predicted by local transport models.

APPENDIX A
PARTITIONING R^2 USING THE t STATISTIC

Without loss of generality, we consider the multiple linear regression problem

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \ \ldots \ + \alpha_p x_p + \epsilon \tag{22}$$

with centered dependent and independent variables. If the independent variables are uncorrelated, i.e. if $< x_j, x_k > \equiv \sum_{i=1}^{N} x_{i,j} x_{i,k} = \|x_j\|^2 \delta_{j,k}$, it is easily shown that the least squares solution reduces to

$$\hat{\alpha}_j = \frac{< x_j, y >}{\|x_j\|^2}; \quad \hat{\sigma}^2(\hat{\alpha}_j) = \frac{\hat{\sigma}^2}{\|x_j\|^2} \tag{23}$$

Here $\hat{\sigma}(\hat{\alpha}_j)$ is the estimate of the standard error for the coefficient estimate $\hat{\alpha}_j$ and $\hat{\sigma}^2 = \|y - \hat{y}\|^2/(N-p)$ is the mean square regression error. From the definition of the t statistic, we have that

$$t_j \equiv \frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)} = \frac{< x_j, y >}{\|x_j\|\hat{\sigma}} = \frac{\sqrt{N-p}}{\|y - \hat{y}\|} \frac{< x_j, y >}{\|x_j\|} \tag{24}$$

In geometrical terms, $< x_j, y > /\|x_j\|$ is the projection of y onto x_j where y and x_j are vectors in $\Re^N$. Hence adding x_j to the regression model makes a fractional contribution to the total variance of

$$\Delta(R^2)_j = \frac{< x_j, y >^2}{\|x_j\|^2 \|y\|^2} \tag{25}$$

Combining this with

$$1 - R^2 = \frac{\|y - \hat{y}\|^2}{\|y\|^2} \tag{26}$$

and substituting into eq. (24) we obtain eq. (2). Furthermore, summing up all contributions, we have from Pythagoras' theorem that $\sum_{j=1}^{p} < x_j, y >^2 / \|x_j\|^2 = \|\hat{y}\|^2$ which gives

$$\sum_{j=1}^{p} t_j^2 = (N-p)\frac{\|\hat{y}\|^2}{\|y-\hat{y}\|^2} = (N-p)\frac{\mathrm{R}^2}{1-\mathrm{R}^2} \tag{27}$$

This formula is the analog of Weisberg's partition of C_p [16] . It provides a useful practical check on the applicability of eq. (2) when the regressors are correlated.

APPENDIX B

COMPOUND ERROR STRUCTURES: TEMPORAL HIERARCHY

To efficiently estimate the spline coefficients, $\underline{\alpha}$, we try to model the actual covariance matrix for the errors. The closer the assumed or estimated $\underline{\underline{\Sigma}}$ is to the actual covariance error structure, the more accurate the ensuing estmates for $\underline{\alpha}$ are.

The assumption of independent errors is not always justified. In general, tokamaks possess a compound error structure. The first level of errors are statistical fluctuations which vary from time point to time point within a given discharge. The next level consists of those errors which vary from discharge to discharge (we assume here that there is only one compressed datapoint per discharge) but remain constant within a given discharge. Finally, there are variations which only change between operating periods of a tokamak. We denote the covariance matrices of for the radial fluctuations of each of these three types of errors by $\underline{\underline{\Sigma}}^{int}, \underline{\underline{\Sigma}}^{disch}, \underline{\underline{\Sigma}}^{op}$ respectively.

We use a triple index, (p,i,t) to denote a given profile timepoint where p indexes the operating period, i the discharge number, and t the time. Within a single profile timepoint, the individual radial measurements are denoted by a fourth index, l. The cross-correlation of any two pairs of profile measurements, (p,i,t) and (p',i',t') is given by a 15×15 matrix, $\underline{\underline{\Sigma}}_{p,i,t,p',i',t'}$.

We assume that the errors do not depend on the plasma parameters and that the covariance structure does not vary between different blocks of data at each level. The most general error structure of this form is

$$\underline{\underline{\Sigma}}_{p,i,t,p',i',t'} = \underline{\underline{\Sigma}}^{int}\delta_{p,p'}\delta_{i,i'}\delta_{t,t'} + \underline{\underline{\Sigma}}^{disch}\delta_{p,p'}\delta_{i,i'} + \underline{\underline{\Sigma}}^{op}\delta_{p,p'} \tag{28}$$

We restrict our attention to datasets consisting of a single operating phase. In this case eqn. (28) reduces to $\underline{\underline{\Sigma}}^{int}\delta_{i,i'}\delta_{t,t'} + \underline{\underline{\Sigma}}^{disch}\delta_{i,i'}$. For simplicity, we assume that each discharge consists of n_t timepoints.

We estimate the within-discharge variance, $\underline{\underline{\Sigma}}^{int}$ empirically by calculating the time point average and the time point variance for each discharge separately:

$$\hat{\Sigma}^{int}_{k,l} = \frac{1}{n_d(n_t-1)} \sum_{i=1}^{n_d} \sum_{t=1}^{n_t} (Y_{i,t}(r_k) - \overline{Y}_i(r_k))(Y_{i,t}(r_l) - \overline{Y}_i(r_l)) \qquad (29a)$$

where

$$\overline{Y}_i(r_k) = \frac{1}{n_t} \sum_t Y_{i,t}(r_k) \qquad (29b)$$

The disadvantage of analysing only the time averaged profiles is that information about the statistical fluctuations is lost. In an optimal statistical analysis, all time points would be retained and analysed simultaneously. This analysis may become unwieldy when the number of timepoints is large. However, if the timepoint variations are comparable to n_t times the discharge variation, a statistical analysis based on structured covariance matrices is desirable.

Within a single operating period, the total variation between datapoints is estimated by

$$\hat{\Sigma}_{k,l} = \frac{1}{n_d n_t - f} \sum_{i=1}^{n_d} \sum_{t=1}^{n_t} (Y_{i,t}(r_k) - Y_{fit}(r_k;\underline{\alpha}))(Y_{i,t}(r_l) - Y_{fit}(r_l;\underline{\alpha})) \qquad (30)$$

where f denotes the number of fitted parameters. The fitted values, $Y_{fit}(r_l;\underline{\alpha})$ depend on the values of the plasma parameters and therefore implicitly on the indices, i and t. $Y_{fit}(r_l;\underline{\alpha})$ may be estimated either by regressing each measurement channel separately or by fitting all channels simultaneously using the spline representation. The latter method will inflate the variance if the profiles cannot be well approximated by the spline representation.

The discharge variance is computed by subtracting the within-discharge variance as defined in eqn. (29a) from the total datapoint variance defined in eqn. (30).

In describing nested error structures of this form, statisticians use the terms "within discharge variation" to refer to the time point to timepoint variation and "between discharge variation" for the discharge variation.

Several caveats must be placed on this procedure. First, using too many or too few terms in the regression analysis will artificially inflate the variance estimates. Second,

the errors in the estimates of the variances tend to be rather larger unless a substantial number of profiles are available.

ACKNOWLEDGEMENT

The work of PJM was performed under a EURATOM supported reciprocal research agreement between IPP and University College, Cork. The work of KSR was supported by the U.S. Department of Energy, Grant No DE-FG02-86ER53223.

REFERENCES

[1] Kardaun, O.J.W.F., Riedel, K.S., McCarthy, P.J., Lackner, K., (submitted to Nuclear Fusion).

[2] Murakami, M., Callen, J.D., Berry, B.A., Nucl. Fusion **16** (1976) 347.

[3] Müller, E.R., Söldner, F.X., Janeschitz, G., Murmann, H., Fußmann, G., Kornherr, M., Poschenrieder, W., Wagner, F., Würz and the ASDEX-Team in Controlled Fusion and Plasma Physics (Proc. 15th Europ. Conf. Dubrovnik, 1988 Europhys. Conf. Abst. **12B-I** (1988) 19.

[4] Röhr, H., Steuer, K.-H., Murmann H., Meisel D., Max-Planck-Institut für Plasma Physik, Report No. III/121 (1987) (also Nucl. Fusion **22** (1982) 1099).

[5] Mardia, K.V., Kent J.T., Bibby,J.M., Multivariate Analysis Academic Press Inc., London (1979).

[6] Steuer, K.-H., Röhr, H., Roberts, D.E., et al. in Controlled Fusion and Plasma Physics (Proc. 15th Europ. Conf. Dubrovnik, 1988 Europhys. Conf. Abst. **12B-I** (1988) 31.

[7] Pfeiffer W., Waltz R., Nucl. Fusion **19** (1979) 51.

[8] Barlett D.V., Bickerton R.J., Brusati M. et al., Nucl. Fusion **28** (1988) 73.

[9] Coppi, B., Comments on Plasma Phys. Controll. Fusion **5** (1980) 261.

[10] Waltz, R.E., Wong, S.K., Greene, J.M., Dominguez, R.R., Nucl. Fusion **26** (1986) 1729.

[11] SAS, User's guide: Statistics (5^{th} ed.), SAS Institute Inc., Cary, NC, (1985).

[12] Poortema, K., On the statistical analysis of growth, PhD thesis, Groningen (1989).

[13] Kardaun, O., Mc Carthy, P.J., Lackner, K., Riedel, K., Gruber, O., in Theory of Fusion Plasmas, Varenna (1987) 435.

[14] Murmann H., Wagner F. et al., Proceedings of the 13th European Conference on Plasma Physics Schliersee 1986. Europ. Conf. Abstr. **10c-I** (1986) 216.

[15] Christiansen, J.P., Callen, J.D., Cordey, J.G., Thomsen, K., Nucl. Fusion **28** (1988) 817.

[16] Weisberg, S., Technometrics **23** (1981) 27.

Table I(a)

Summary of High n Low q Database

Parameter	mean	minimum	maximum	
q_a	2.432	1.858	3.370	
I_p	0.388	0.321	0.483	MA
Ne_linav	.442	.326	.823	e+20m-3
Btor	2.011	1.754	2.730	T
Loop Voltage	1.325	1.124	1.760	V
Ohmic Power	.516	.362	.811	MW
Te_volav	.380	.271	.482	keV
Ne_volav	.391	.271	.712	e+20m-3
Zeff_Spitzer	1.800	1.391	2.155	
Rmag_axis	1.684	1.674	1.715	m
Zmag_axis	-0.000	-0.001	0.000	m
b_minor	0.400	0.398	0.406	m
Area	0.480	0.475	0.495	m**2
bpol+li/2	0.905	0.762	1.056	
bpol_diam	.386	.277	.528	
Wplas_diam	44.9	34.3	70.1	kJ
Wplas_kin	38.6	28.2	61.260	kJ
Taue_diam	87.6	71.5	113.0	msec
Taue_kinetic	74.8	55.9	96.7	msec

Table I(b)

Summary of Moderate n Moderate q Database

Parameter	mean	minimum	maximum	
q_a	3.088	2.382	4.195	
I_p	0.355	0.281	0.452	MA
Ne_linav	.306	.163	.399	e+20m-3
Btor	2.335	1.682	2.813	T
Loop Voltage	1.072	0.920	1.225	V
Ohmic Power	.383	.269	.495	MW
Te_volav	.490	.388	.678	keV
Ne_volav	.247	.128	.296	e+20m-3
Zeff_Spitzer	2.587	2.040	3.504	
Rmag_axis	1.690	1.678	1.712	m
Zmag_axis	0.001	-0.000	0.013	m
b_minor	0.398	0.391	0.404	m
Area	0.478	0.461	0.496	m**2
bpol+li/2	0.970	0.857	1.142	
bpol_diam	.398	.304	.542	
Wplas_diam	38.3	36.0	50.3	kJ
Wplas_kin	33.5	19.5	43.9	kJ
Taue_diam	101.1	80.0	124.0	msec
Taue_kinetic	88.0	70.5	107.1	msec

Table II(a)

Correlation Matrix of ln{ Ne_lav, I_p, q_a, Zeff_Spitzer, P_Ohmic}

High n Low q Database

	ln{Ne_lav}	ln{I_p}	ln{q_a}	ln{Zeff_Sp}	ln{P_Ohm}
ln{Ne_lav}	1.000	0.199	-0.135	-0.819	0.728
ln{I_p}	0.199	1.000	-0.694	0.208	0.750
ln{q_a}	-0.135	-0.694	1.000	-0.034	-0.636
ln{Zeff_Sp}	-0.819	0.208	-0.034	1.000	-0.359
ln{P_Ohm}	0.728	0.750	-0.636	-0.359	1.000

PCA for all five variables

Eigenvalues of 5 x 5 Matrix:	2.745	1.751	0.389	0.085	0.029
Fraction of Total Variance:	0.549	0.350	0.078	0.017	0.006
Cumulative Variance Fraction:	0.549	0.899	0.977	0.994	1.000

PCA for ln{Ne_lav} ln{I_p} ln{q_a}

Eigenvalues of 3 x 3 Matrix:	1.767	0.930	0.302
Fraction of Total Variance:	0.589	0.310	0.100
Cumulative Variance Fraction:	0.589	0.899	1.000

Table II(b)

Correlation Matrix of ln{ Ne_lav, I_p, q_a, Zeff_Spitzer, P_Ohmic}

Moderate n Moderate q Database

	ln{Ne_lav}	ln{I_p}	ln{q_a}	ln{Zeff_Sp}	ln{P_Ohm}
ln{Ne_lav}	1.000	0.293	-0.065	-0.724	0.528
ln{I_p}	0.293	1.000	-0.661	0.225	0.937
ln{q_a}	-0.065	-0.661	1.000	-0.371	-0.722
ln{Zeff_Sp}	-0.724	0.225	-0.371	1.000	0.040
ln{P_Ohm}	0.528	0.937	-0.722	0.040	1.000

PCA for all five variables

Eigenvalues of 5 x 5 Matrix:	2.728	1.808	0.333	0.121	0.009
Fraction of Total Variance:	0.546	0.361	0.067	0.024	0.002
Cumulative Variance Fraction:	0.546	0.907	0.974	0.998	1.000

PCA for ln{Ne_lav} ln{I_p} ln{q_a}

Eigenvalues of 3 x 3 Matrix:	1.749	0.952	0.299
Fraction of Total Variance:	0.583	0.317	0.100
Cumulative Variance Fraction:	0.583	0.900	1.000

Table III(a)

Temperature Descriptive Statistics for High n Low q Database
(Absolute Scale)

Channel	‹radius›	‹Te(kev)›	spread	‹Te/Te_r=.2›	spread
1	.515	0.555	.096	0.804	.104
2	.432	0.571	.089	0.825	.085
3	.345	0.636	.090	0.918	.041
4	.247	0.678	.100	0.977	.023
5	.180	0.694	.104	1.000	.019
6	.138	0.718	.111	1.034	.028
7	.164	0.705	.108	1.015	.020
8	.225	0.686	.102	0.988	.019
9	.309	0.646	.090	0.933	.046
10	.407	0.620	.100	0.897	.095
11	.489	0.573	.103	0.829	.114
12	.585	0.505	.085	0.732	.102
13	.682	0.407	.074	0.589	.089
14	.784	0.271	.046	0.393	.063
15	.885	0.166	.028	0.241	.038

Table III(b)

Logarithmic Spread, Noise estimates, and Spline Regression Errors
(Relative Scale)

Channel	Logarithmic Spread	Within-discharge noise	Total noise	Profile Regression ‹Relative Error›
1	.130	.026	.027	.029
2	.104	.018	.023	.037
3	.046	.018	.023	.031
4	.024	.020	.020	.021
5	.019	.018	.018	.017
6	.027	.027	.027	.029
7	.020	.020	.020	.019
8	.019	.016	.016	.017
9	.051	.015	.022	.024
10	.107	.018	.025	.034
11	.137	.024	.032	.033
12	.139	.035	.040	.042
13	.148	.029	.029	.028
14	.155	.031	.031	.031
15	.154	.041	.052	.050
RMS value for 15 channels:		.025	.028	.031

Table IV(a)

Density Descriptive Statistics for High n Low q Database
(Absolute Scale)

Channel	‹radius›	‹Ne(E+20 m-3)›	spread	‹Ne/Ne_r=.2›	spread
1	.515	0.523	.149	0.926	.055
2	.432	0.576	.168	1.018	.038
3	.345	0.581	.166	1.026	.023
4	.247	0.586	.166	1.037	.015
5	.180	0.579	.160	1.026	.015
6	.138	0.572	.160	1.013	.015
7	.164	0.556	.155	0.985	.017
8	.225	0.539	.153	0.954	.014
9	.309	0.542	.157	0.958	.018
10	.407	0.534	.152	0.944	.033
11	.489	0.501	.147	0.887	.053
12	.585	0.462	.146	0.813	.064
13	.682	0.419	.135	0.736	.058
14	.784	0.327	.107	0.573	.050
15	.885	0.272	.092	0.479	.057

Table IV(b)

Logarithmic Spread, Noise estimates, and Spline Regression Errors
(Relative Scale)

Channel	Logarithmic Spread	Within-discharge noise	Total noise	Profile Regression ‹Relative Error›
1	.060	.015	.021	.022
2	.038	.012	.018	.026
3	.023	.014	.015	.017
4	.014	.011	.014	.016
5	.015	.012	.014	.020
6	.015	.012	.014	.014
7	.018	.011	.018	.023
8	.015	.011	.014	.016
9	.019	.010	.017	.017
10	.035	.010	.015	.019
11	.060	.015	.025	.028
12	.079	.033	.047	.051
13	.079	.014	.027	.027
14	.087	.021	.031	.032
15	.116	.048	.060	.059
RMS value for 15 channels:		.019	.027	.029

Table V(a)

Temperature Descriptive Statistics for Moderate n Moderate q Database
(Absolute Scale)

Channel	‹radius›	‹Te(kev)›	spread	‹Te/Te_r=.2›	spread
1	.516	0.754	.135	0.711	.109
2	.434	0.788	.146	0.741	.094
3	.347	0.970	.151	0.914	.095
4	.252	1.074	.171	1.007	.051
5	.190	1.066	.153	1.001	.022
6	.154	1.108	.154	1.042	.044
7	.180	1.074	.168	1.007	.026
8	.238	1.034	.155	0.970	.027
9	.321	0.960	.145	0.905	.092
10	.417	0.819	.160	0.770	.103
11	.499	0.664	.139	0.624	.092
12	.594	0.555	.122	0.521	.083
13	.691	0.470	.098	0.444	.082
14	.793	0.285	.060	0.269	.049
15	.894	0.181	.033	0.171	.025

Table V(b)

Logarithmic Spread, Noise estimates, and Spline Regression Errors
(Relative Scale)

Channel	Logarithmic Spread	Within-discharge noise	Total noise	Profile Regression ‹Relative Error›
1	.152	.025	.027	.087
2	.130	.020	.023	.031
3	.105	.018	.032	.056
4	.050	.021	.027	.043
5	.021	.013	.013	.025
6	.042	.031	.032	.059
7	.026	.018	.018	.035
8	.027	.019	.024	.026
9	.103	.017	.029	.038
10	.137	.023	.042	.047
11	.151	.032	.036	.079
12	.164	.042	.059	.127
13	.186	.024	.045	.046
14	.187	.023	.050	.049
15	.151	.037	.076	.096
RMS value for 15 channels:		.025	.039	.063

Table VI(a)

Density Descriptive Statistics for Moderate n Moderate q Database
(Absolute Scale)

Channel	‹radius›	‹Ne(E+20 m-3)›	spread	‹Ne/Ne_r=.2›	spread
1	.516	0.346	.064	0.813	.082
2	.434	0.385	.069	0.904	.062
3	.347	0.406	.070	0.955	.048
4	.252	0.407	.065	0.959	.047
5	.190	0.434	.078	1.018	.024
6	.154	0.437	.083	1.021	.028
7	.180	0.437	.088	1.021	.036
8	.238	0.429	.088	1.001	.047
9	.321	0.376	.058	0.886	.047
10	.417	0.400	.087	0.933	.088
11	.499	0.343	.067	0.804	.075
12	.594	0.295	.062	0.690	.079
13	.691	0.235	.045	0.551	.065
14	.793	0.189	.040	0.444	.053
15	.894	0.132	.032	0.311	.063

Table VI(b)

Logarithmic Spread, Noise estimates, and Spline Regression Errors
(Relative Scale)

Channel	Logarithmic Spread	Within-discharge noise	Total noise	Profile Regression ‹Relative Error›
1	.100	.016	.036	.073
2	.069	.014	.025	.039
3	.051	.012	.027	.055
4	.049	.013	.013	.041
5	.023	.012	.012	.034
6	.028	.014	.015	.033
7	.035	.012	.019	.042
8	.048	.012	.019	.063
9	.054	.012	.015	.054
10	.095	.013	.019	.054
11	.093	.020	.028	.046
12	.117	.040	.041	.050
13	.119	.012	.030	.048
14	.119	.018	.045	.090
15	.213	.045	.108	.114
RMS value for 15 channels:		.020	.028	.060

Table VII

Spline Coefficients (with standard errors) for High n Low q Database

Normalised Temperature Profiles

	Intercept	$ln(q_a/q_a^*)$	$ln(I_p/I_p^*)$	$ln(\overline{n_e}/\overline{n_e}^*)$
$\mu''(0)/2$	− 1.248(.07)	− 2.214(.54)	− 1.782(.82)	− 0.004(0.20)
c_1	− 3.611(1.9)	− 85.812(14.)	2.326(21.)	− 9.151(5.1)
c_2	16.736(5.7)	259.152(41.)	39.709(62.)	28.949(15.)
c_3	− 23.810(7.8)	− 228.112(54)	− 88.397(82.)	− 30.122(20.)
c_4	− 7.549(6.2)	38.372(38.)	49.847(58.)	11.579(14.)
c_5	31.383(4.1)	16.181(13.)	− 2.712(20.)	− 1.881(4.9)

Normalised Density Profiles

	Intercept	$ln(q_a/q_a^*)$	$ln(I_p/I_p^*)$	$ln(\overline{n_e}/\overline{n_e}^*)$
$\mu''(0)/2$	− 0.466(.06)	0.607(.43)	0.191(.64)	0.383(0.15)
c_1	9.880(1.5)	− 50.252(11.)	− 27.318(16.)	− 4.605(3.9)
c_2	− 44.698(4.4)	113.355(32.)	99.895(48.)	0.618(11.)
c_3	75.616(5.9)	− 53.578(41.)	− 138.561(62)	31.504(15.)
c_4	− 69.438(4.9)	− 27.343(30.)	91.804(45.)	− 48.734(11.)
c_5	56.806(3.5)	18.483(11.)	− 31.147(17.)	24.538(4.1)

Normalised Pressure Profiles

	Intercept	$ln(q_a/q_a^*)$	$ln(I_p/I_p^*)$	$ln(\overline{n_e}/\overline{n_e}^*)$
$\mu''(0)/2$	− 1.633(.08)	− 1.100(.61)	− 0.827(.93)	0.358(0.22)
c_1	4.807(2.2)	− 150.187(16)	− 54.177(24.)	− 12.413(5.8)
c_2	− 24.171(6.6)	414.743(47.)	239.688(71.)	25.599(17.)
c_3	45.270(9.1)	− 335.527(62)	− 377.360(94)	8.229(23.)
c_4	− 67.789(7.3)	46.687(44.)	258.188(67.)	− 46.249(16.)
c_5	78.703(4.9)	23.340(16.)	− 76.694(24.)	28.796(5.8)

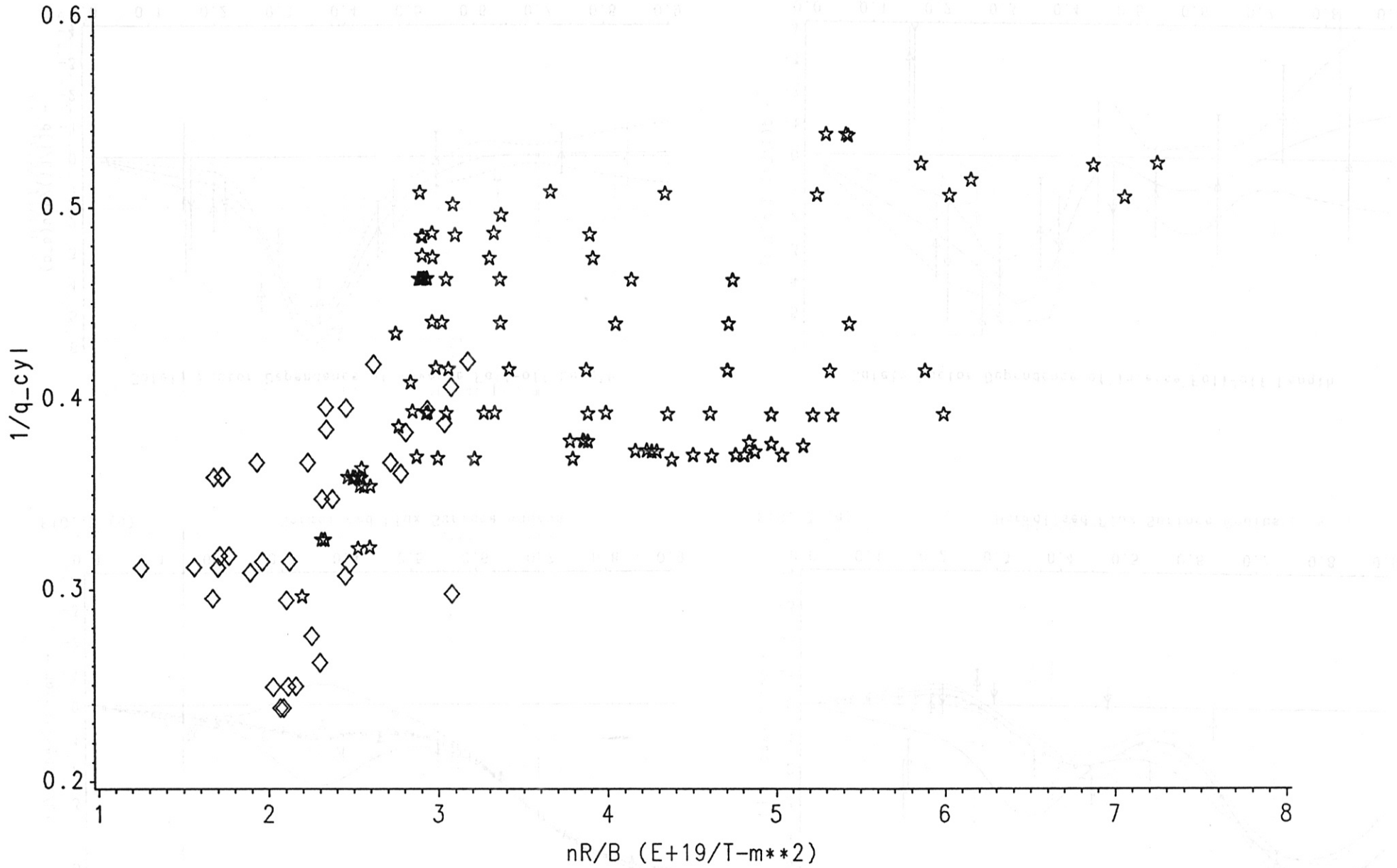

Fig 1. Hugill Diagram

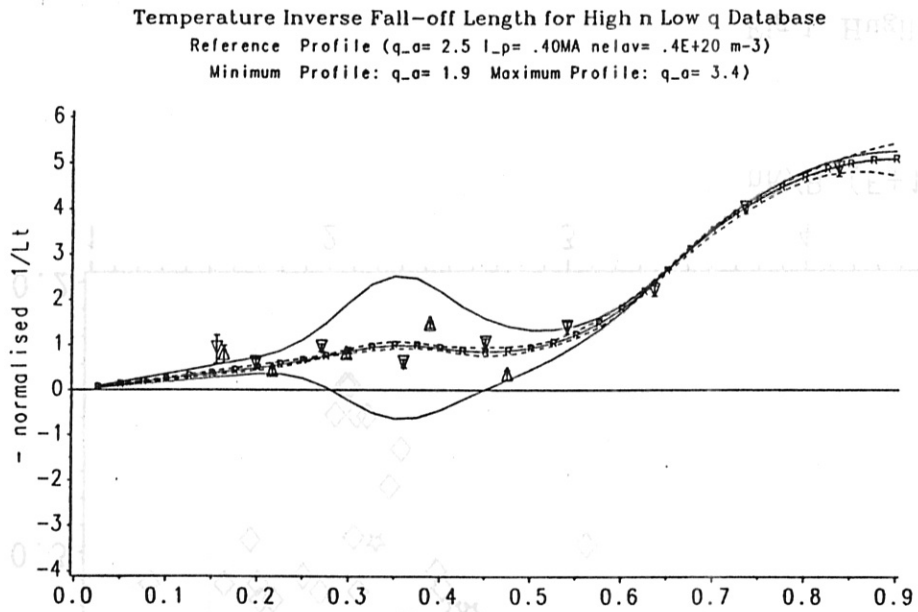

FIG. 2 (a)

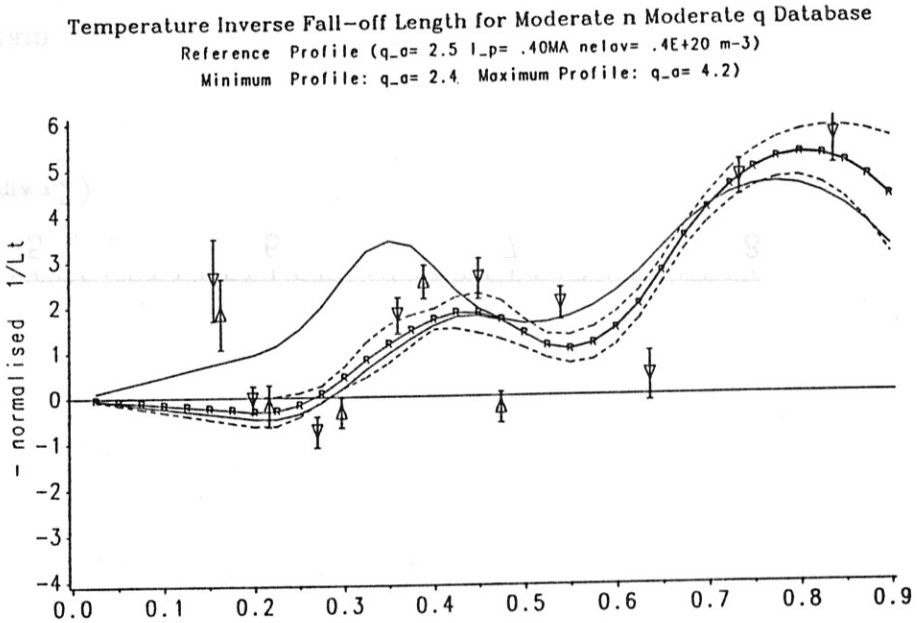

FIG. 3 (a)

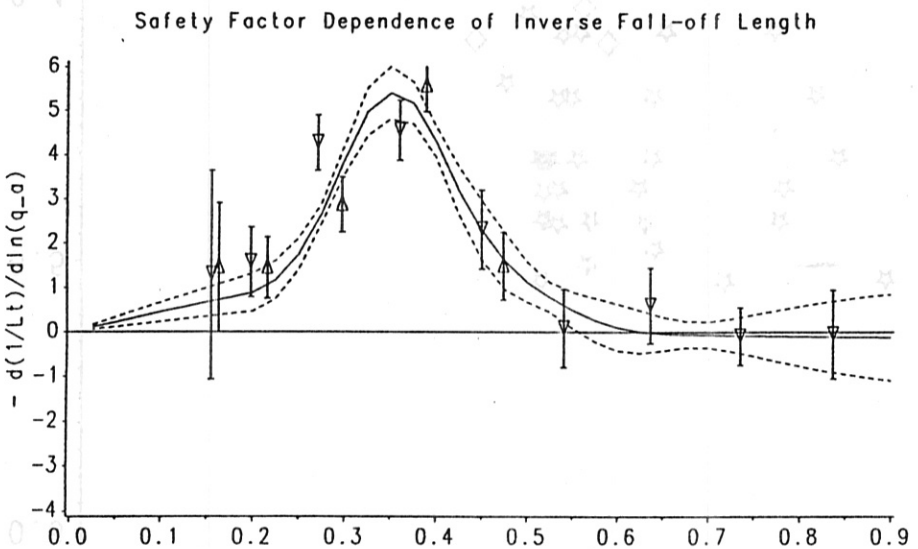

FIG. 2 (b)

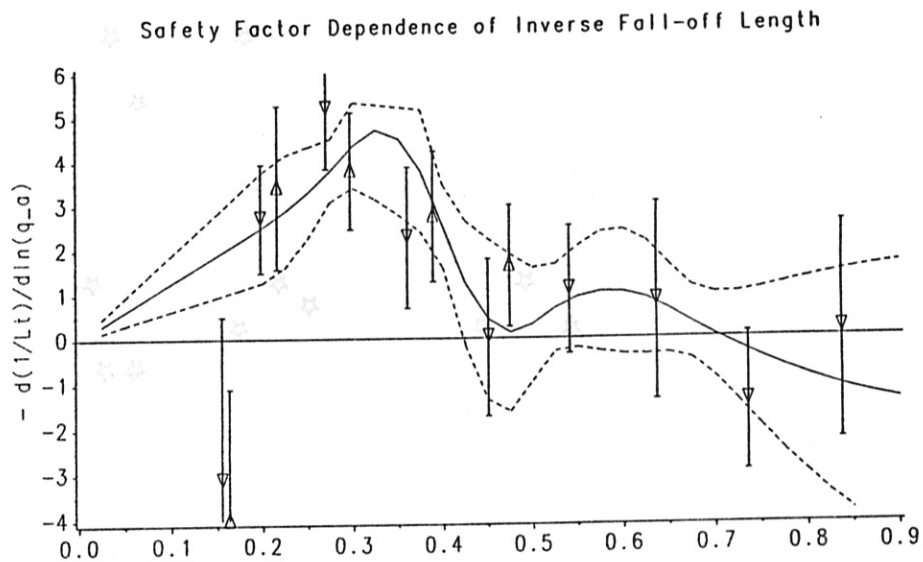

FIG. 3 (b)

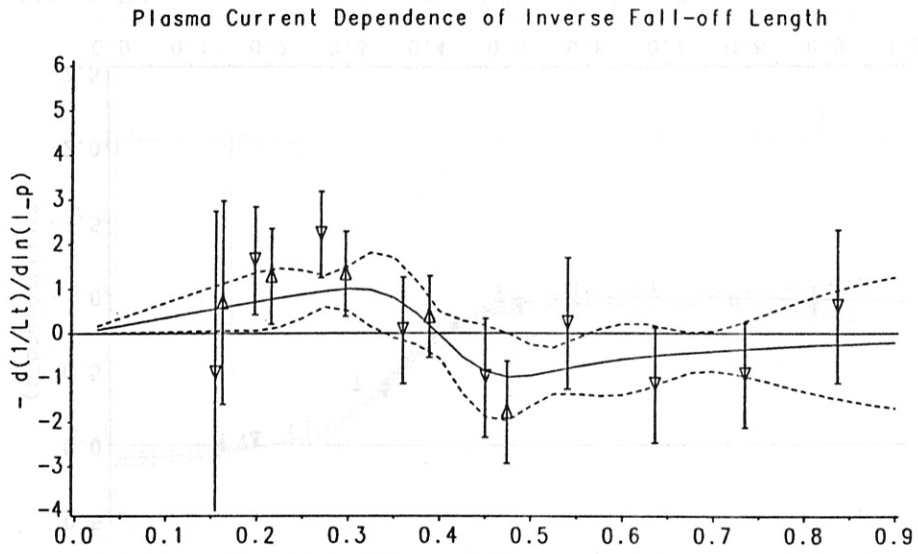

FIG. 2 (c)

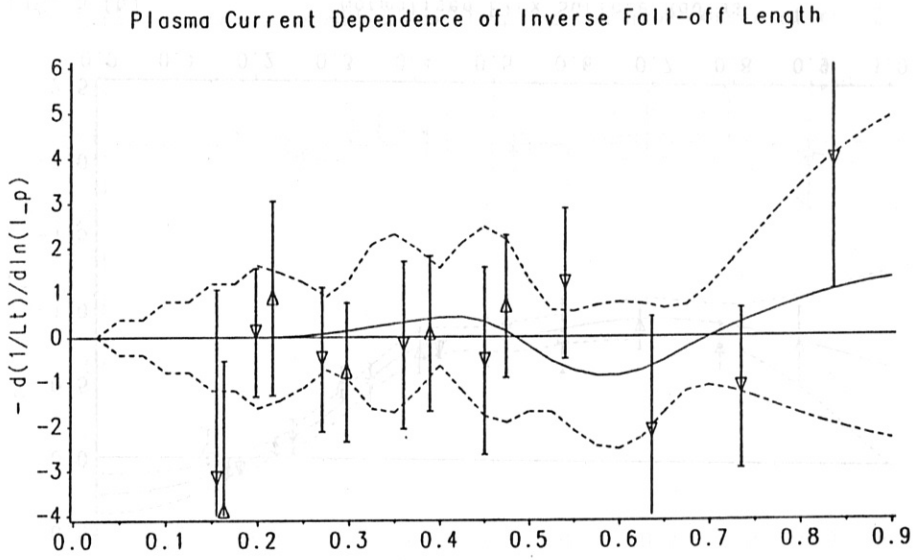

FIG. 3 (c)

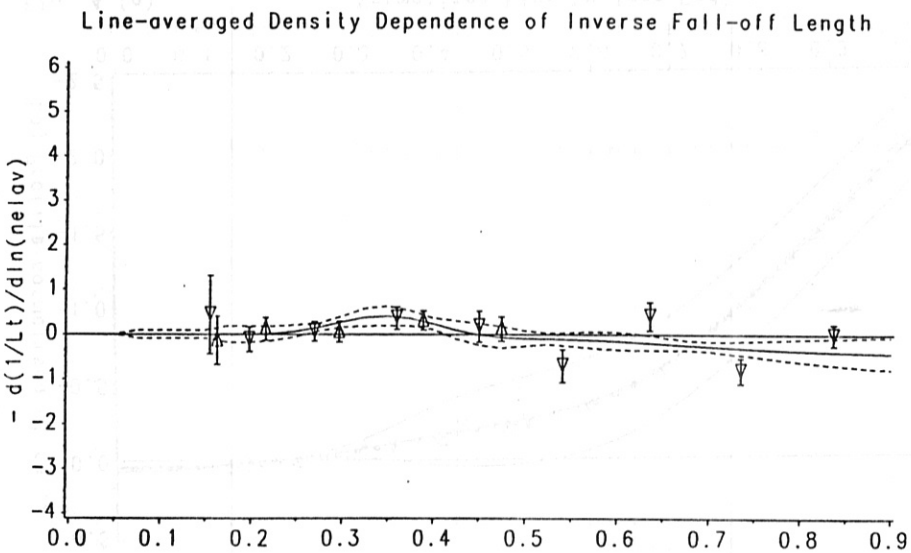

FIG. 2 (d)

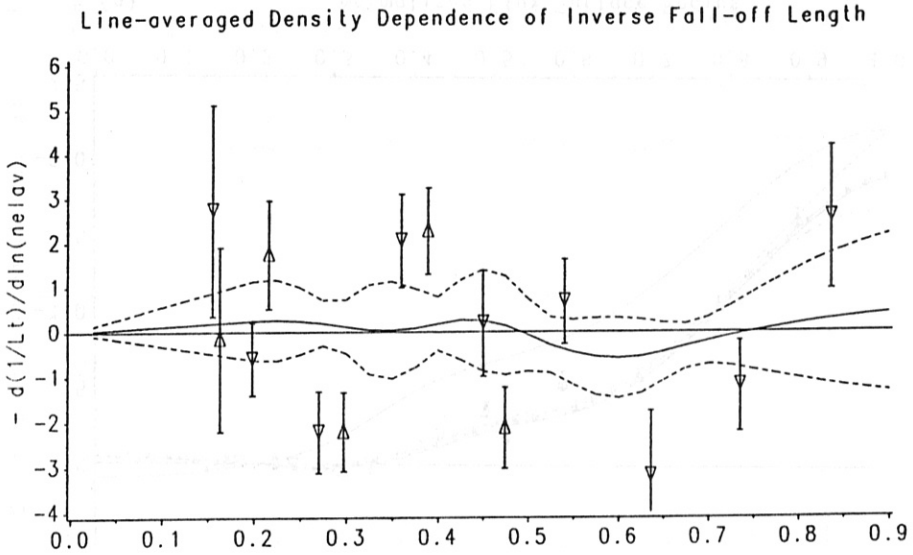

FIG. 3 (d)

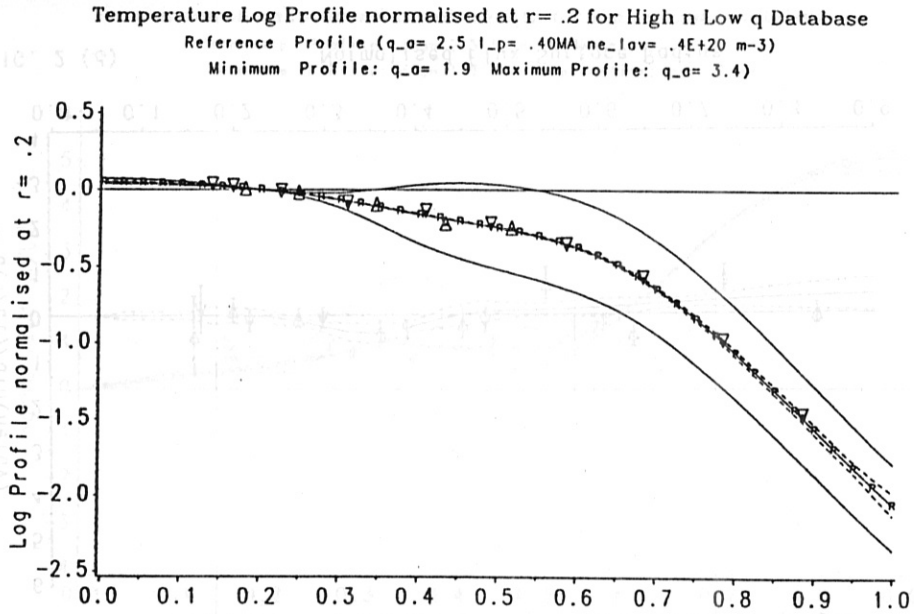

FIG. 4 (a)

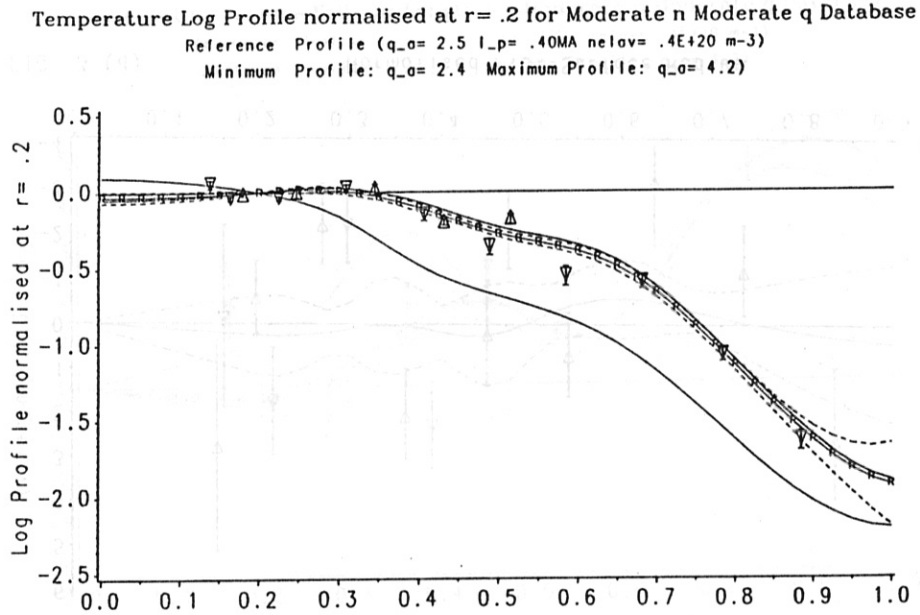

FIG. 5 (a)

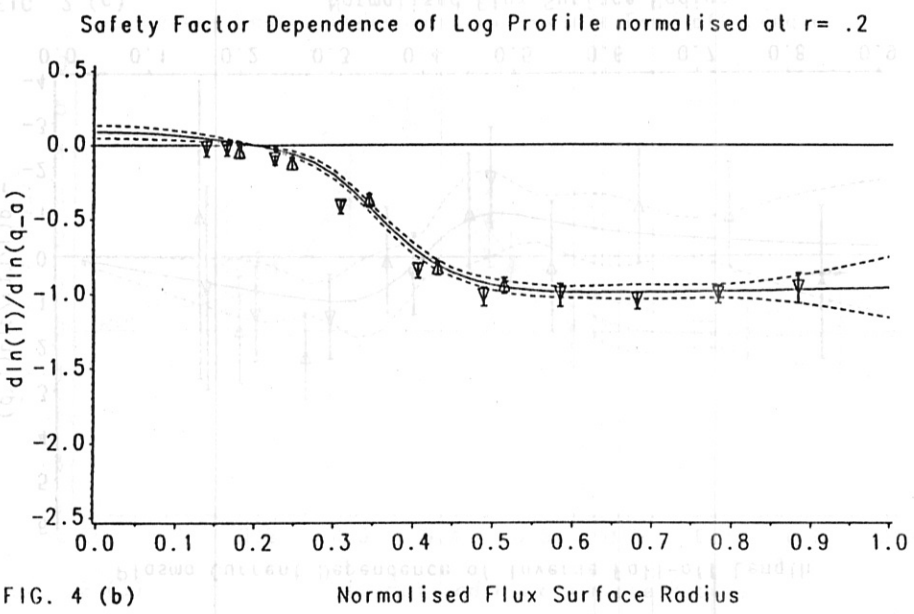

FIG. 4 (b)

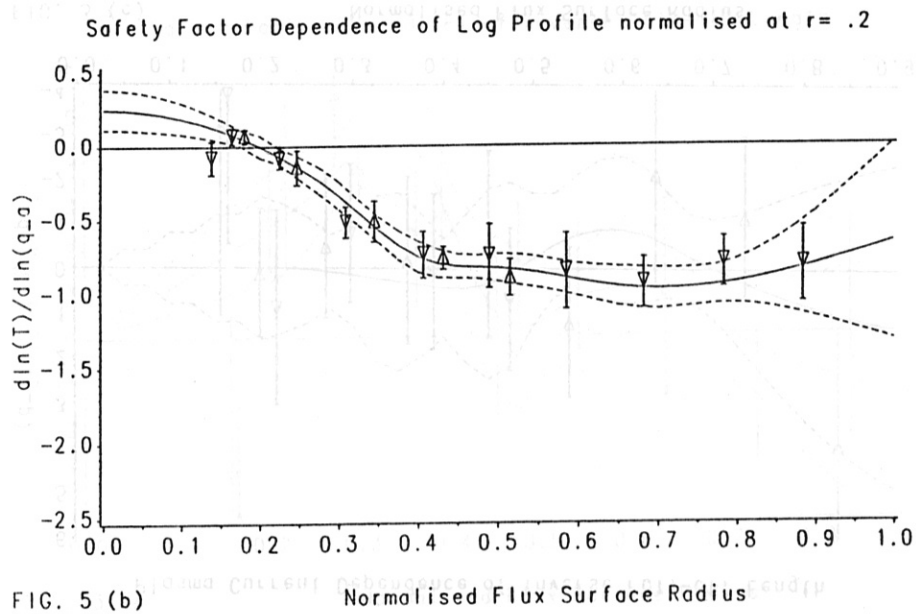

FIG. 5 (b)

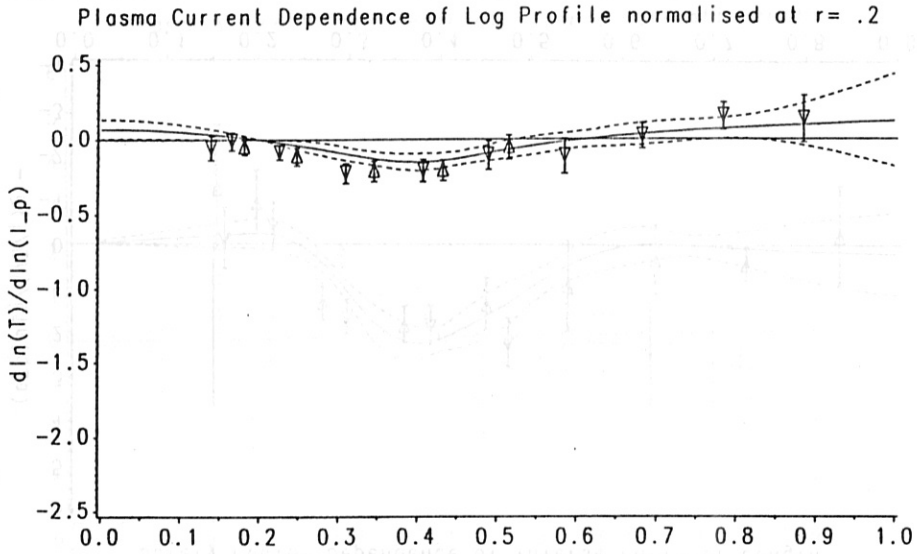

FIG. 4 (c)

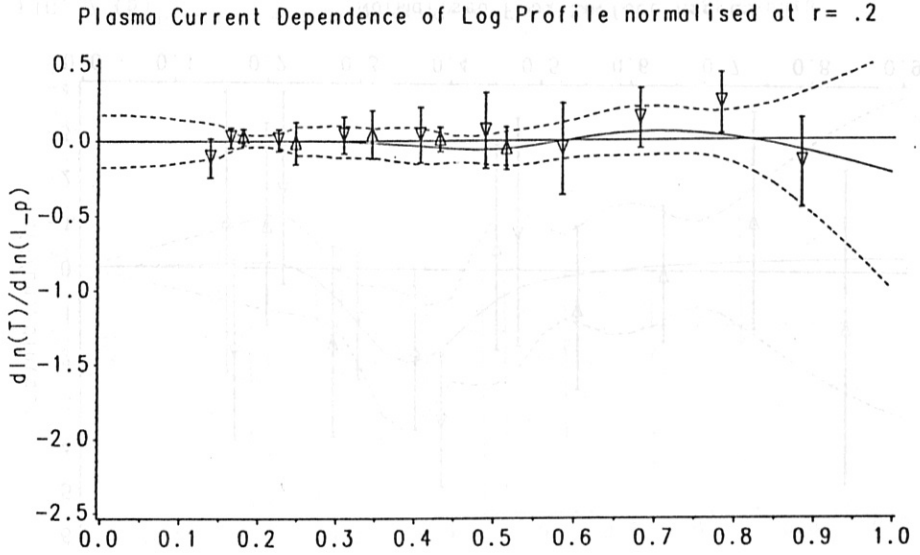

FIG. 5 (c)

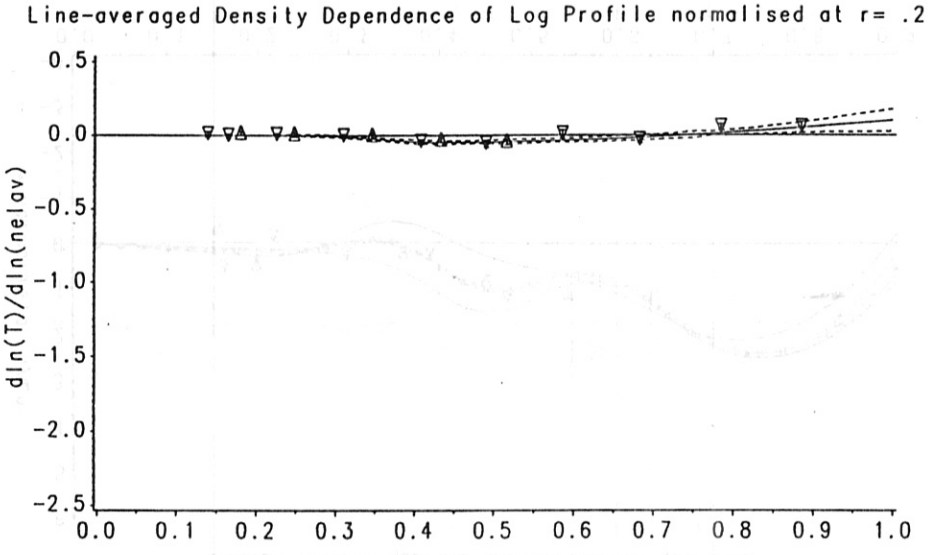

FIG. 4 (d)

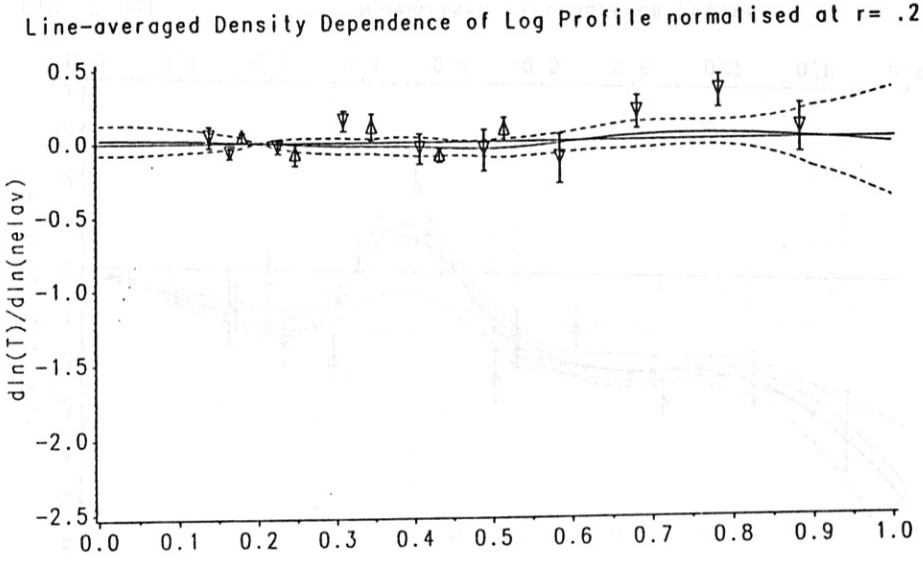

FIG. 5 (d)

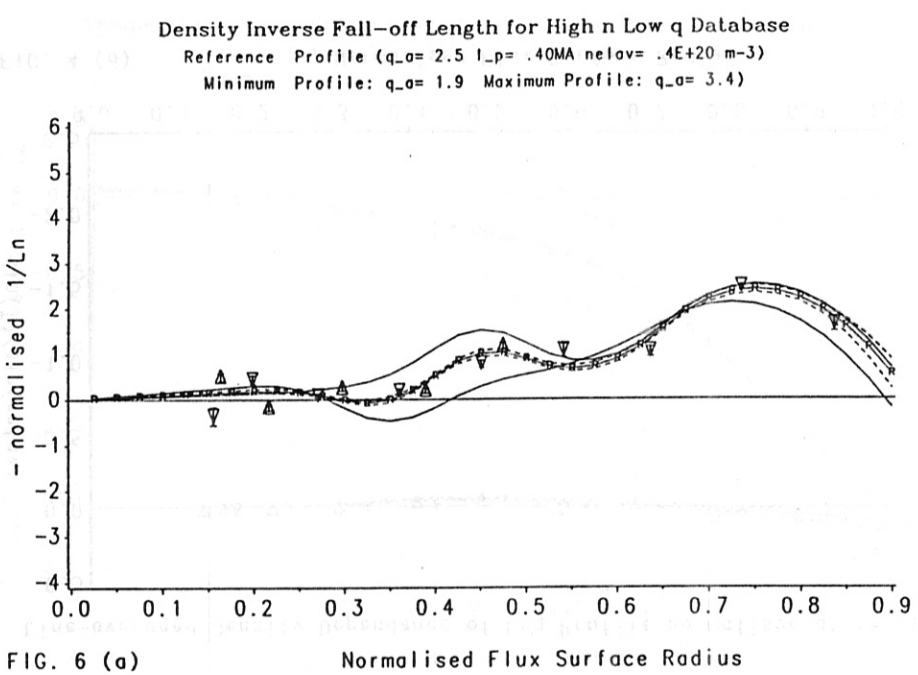

FIG. 6 (a)

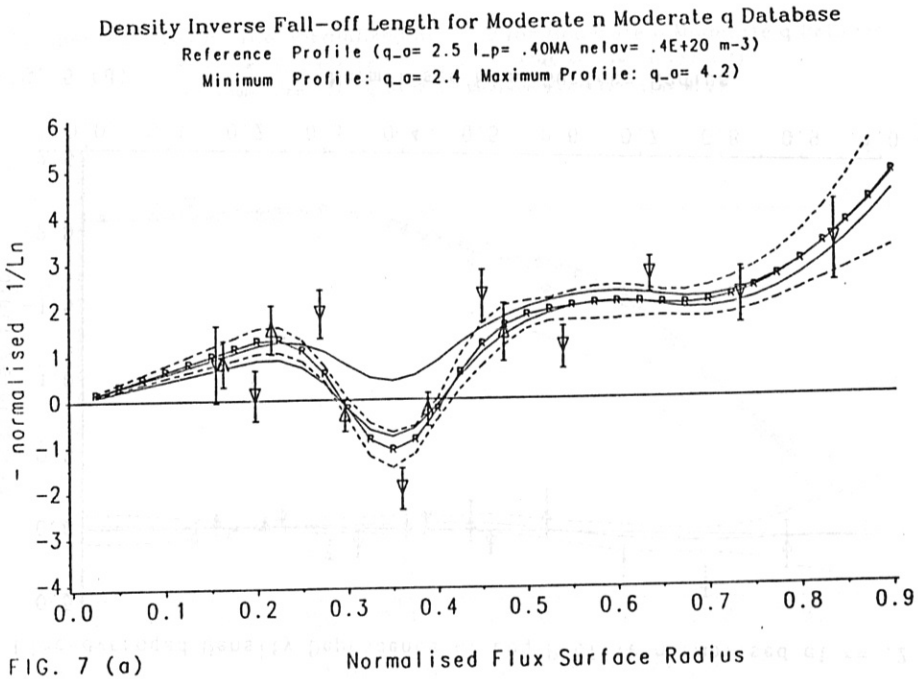

FIG. 7 (a)

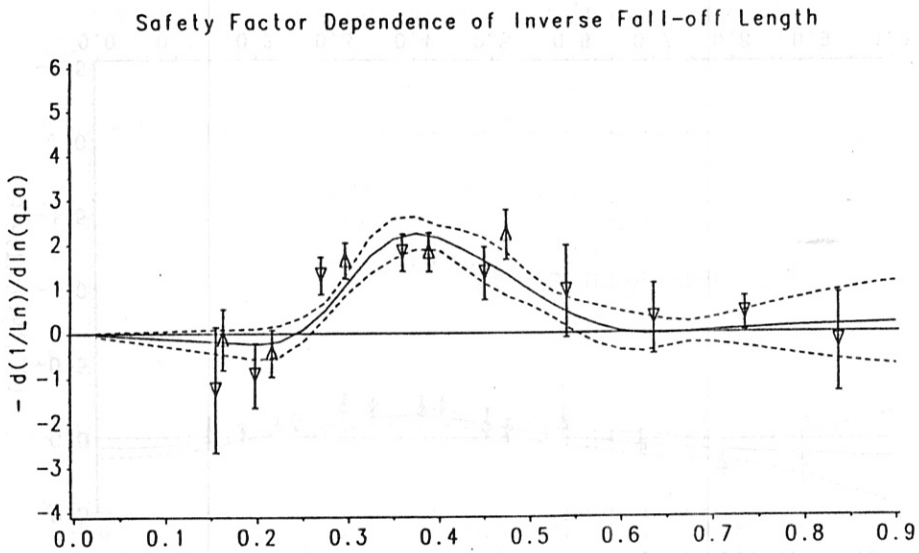

FIG. 6 (b)

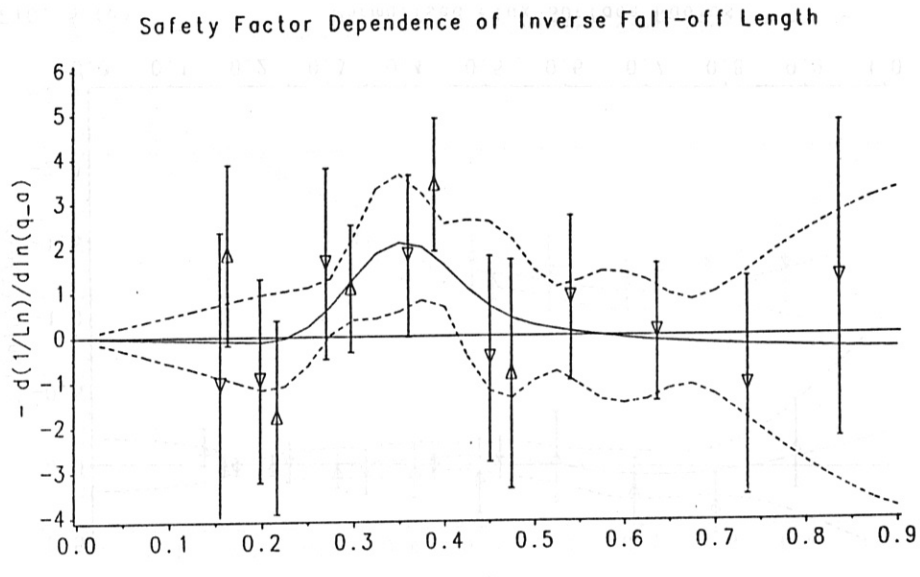

FIG. 7 (b)

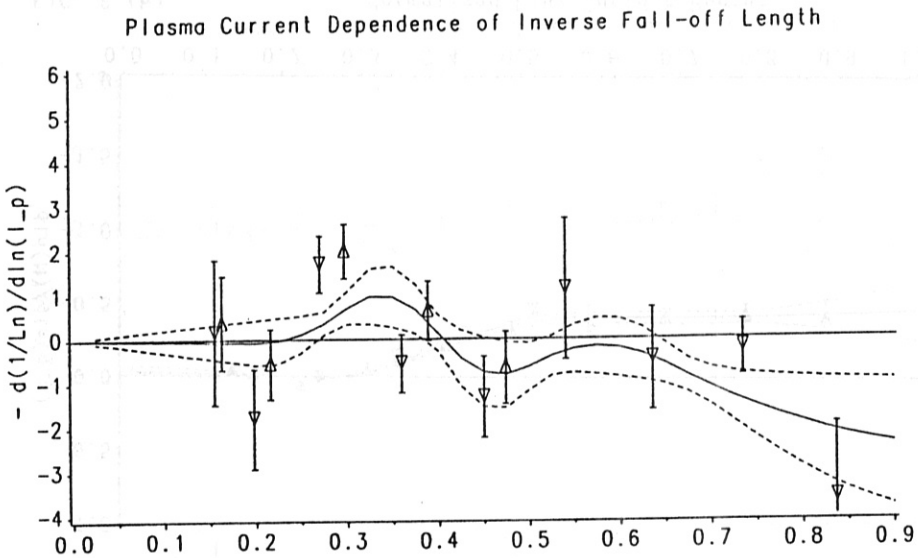

FIG. 6 (c)

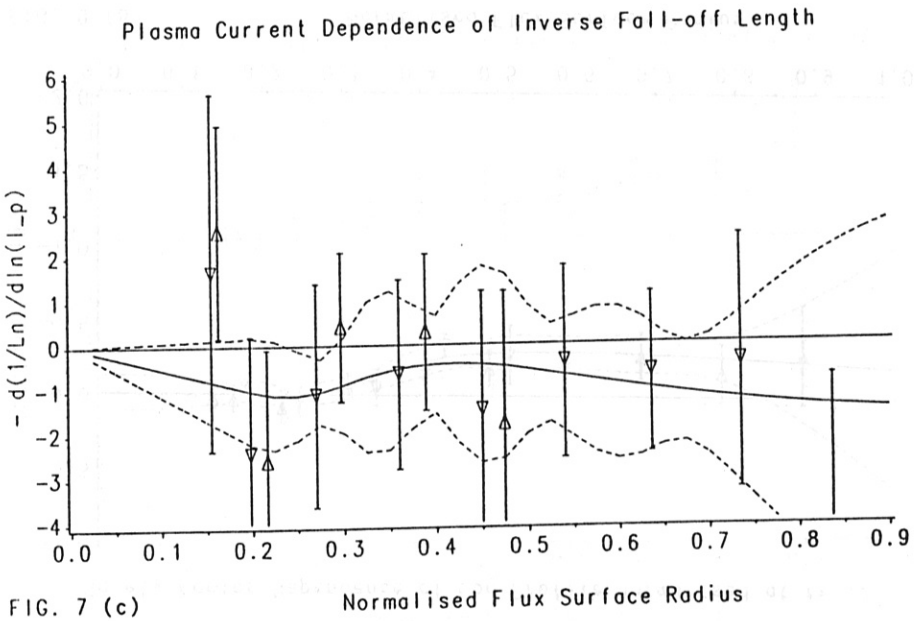

FIG. 7 (c)

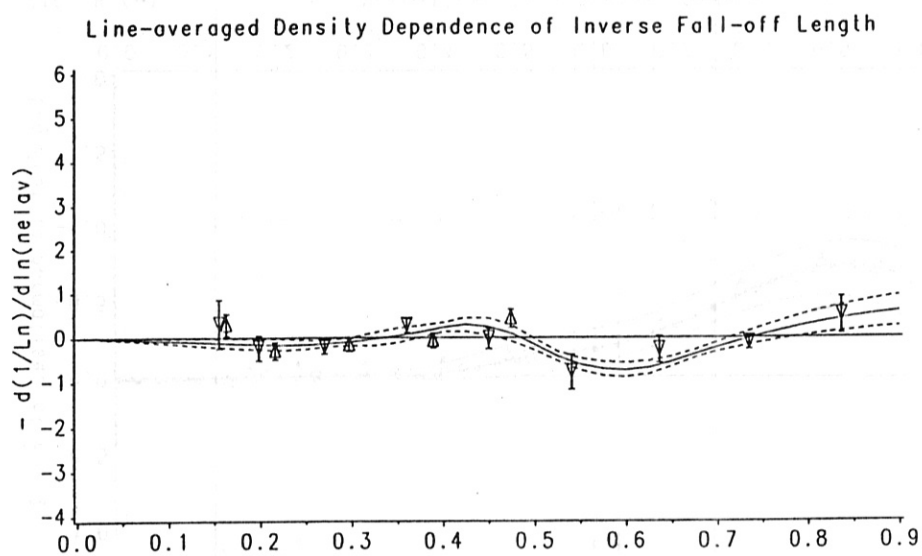

FIG. 6 (d)

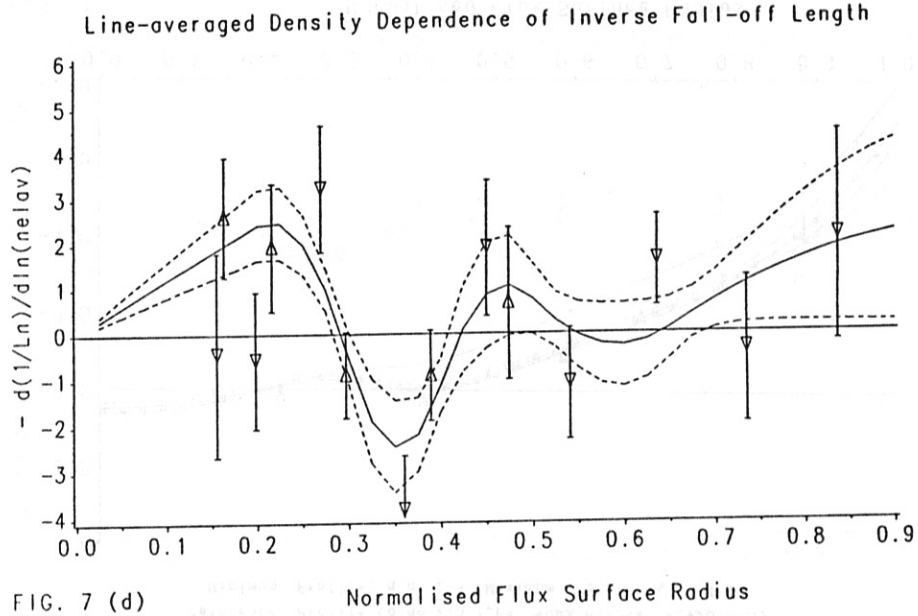

FIG. 7 (d)

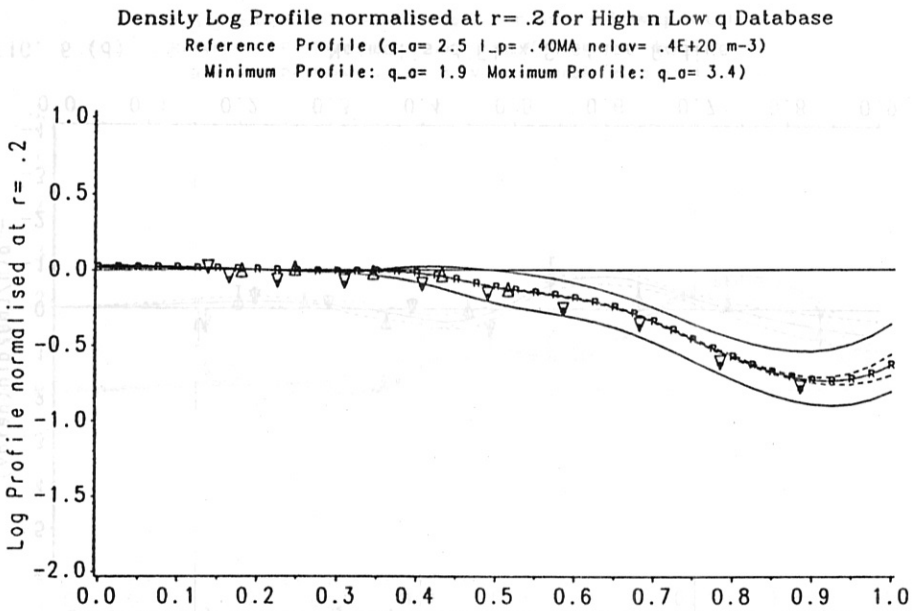

FIG. 8 (a)

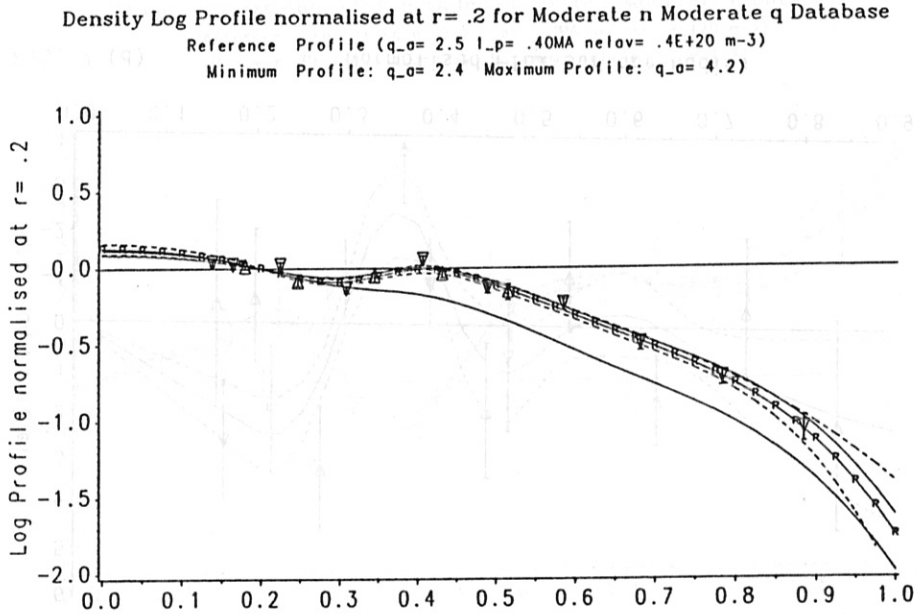

FIG. 9 (a)

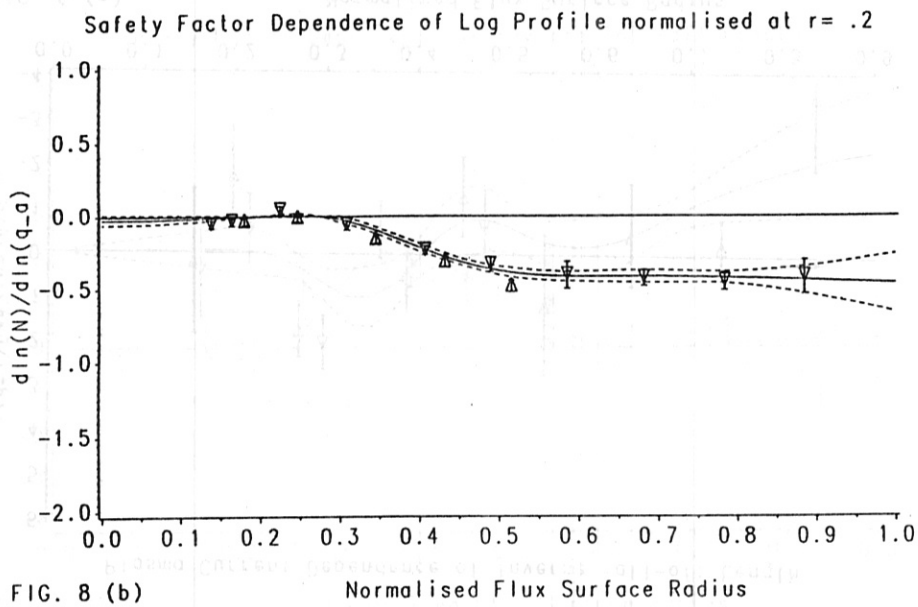

FIG. 8 (b)

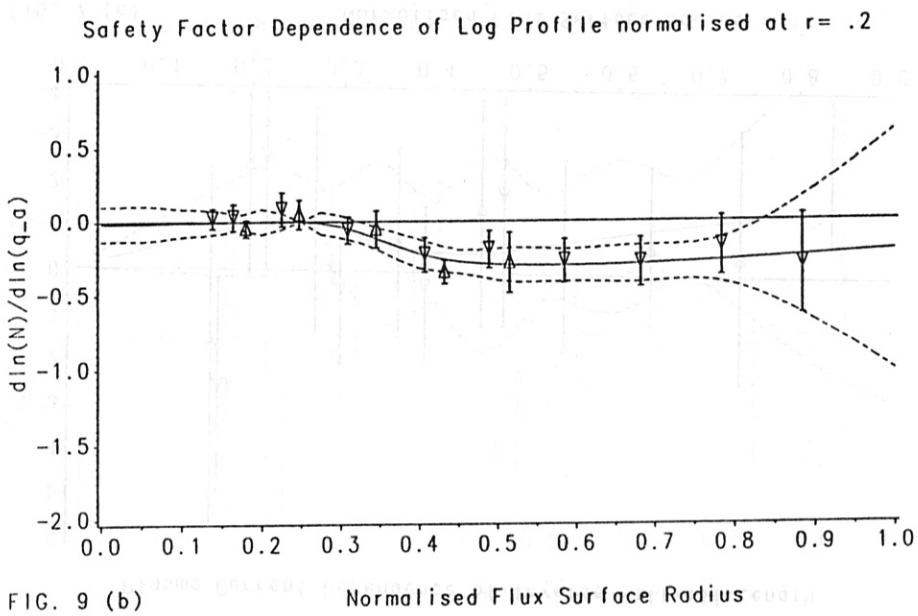

FIG. 9 (b)

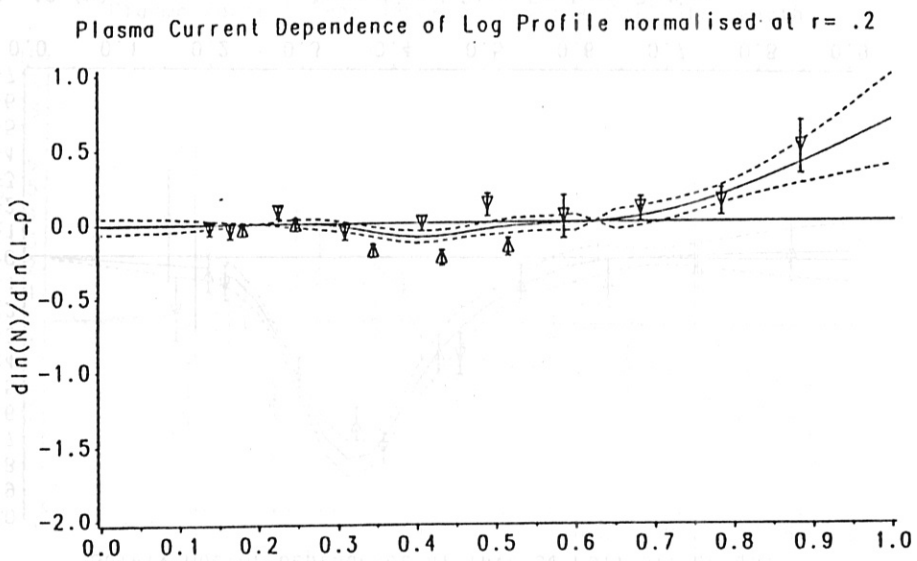

FIG. 8 (c)

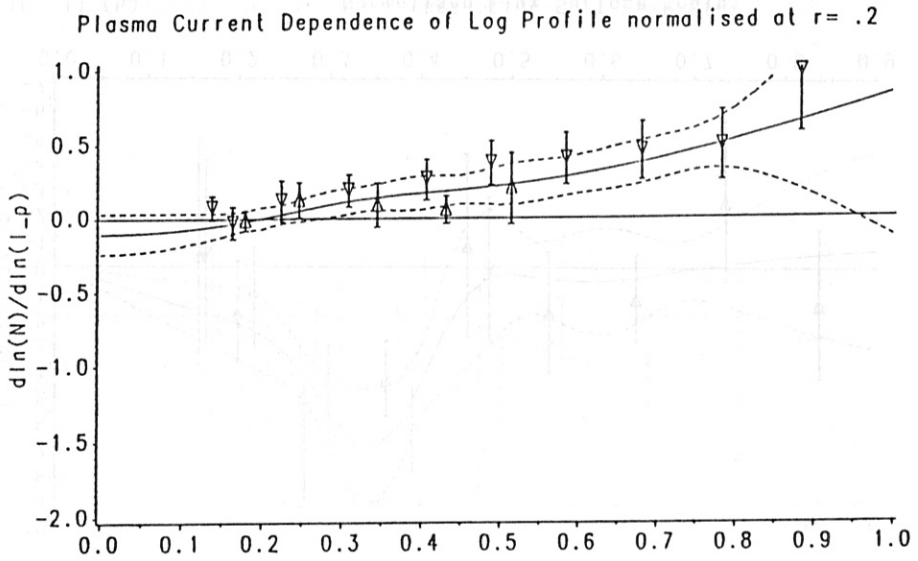

FIG. 9 (c)

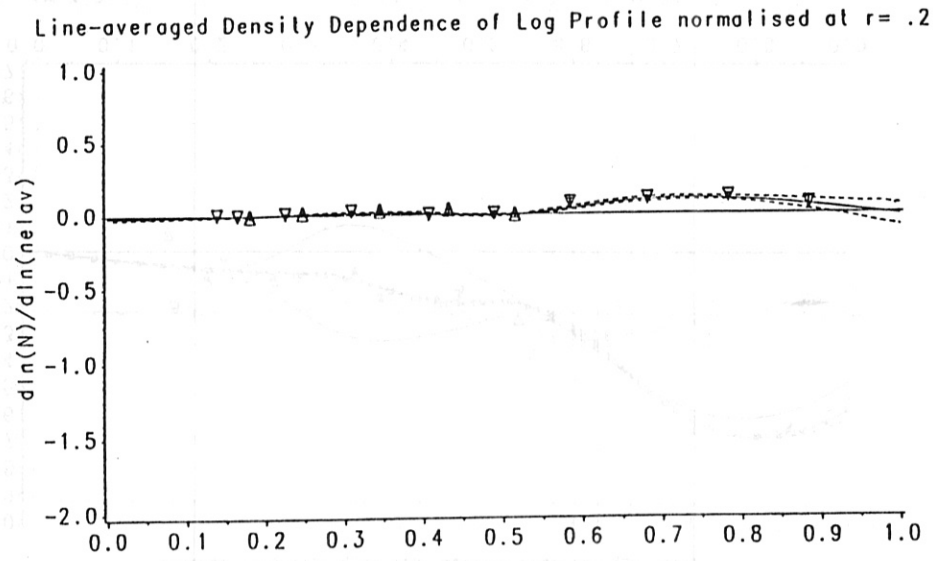

FIG. 8 (d)

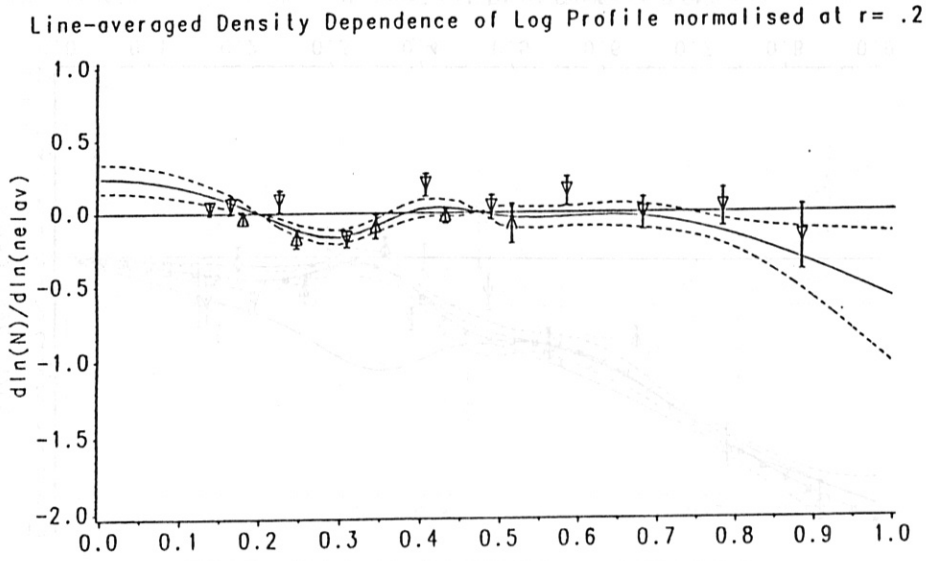

FIG. 9 (d)

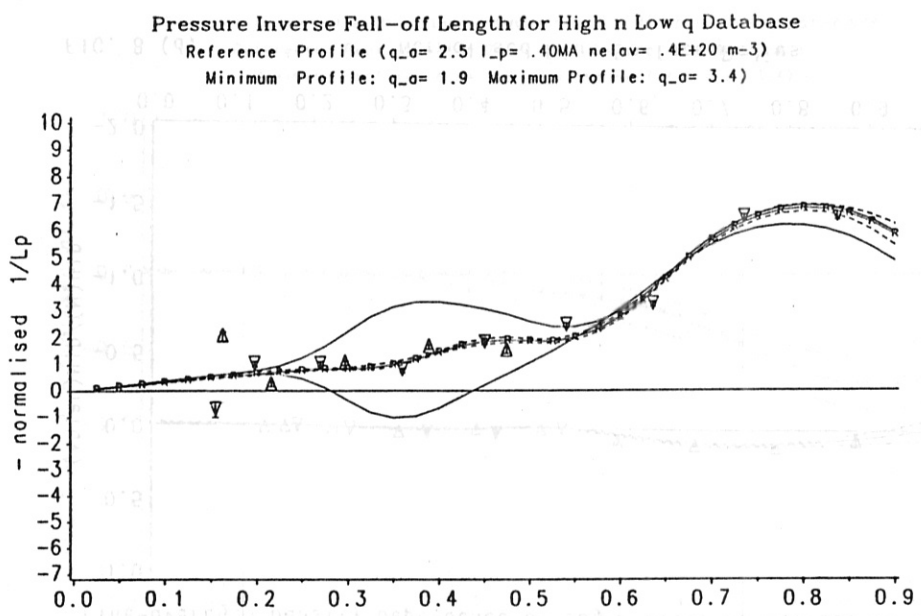

FIG. 10 (a)

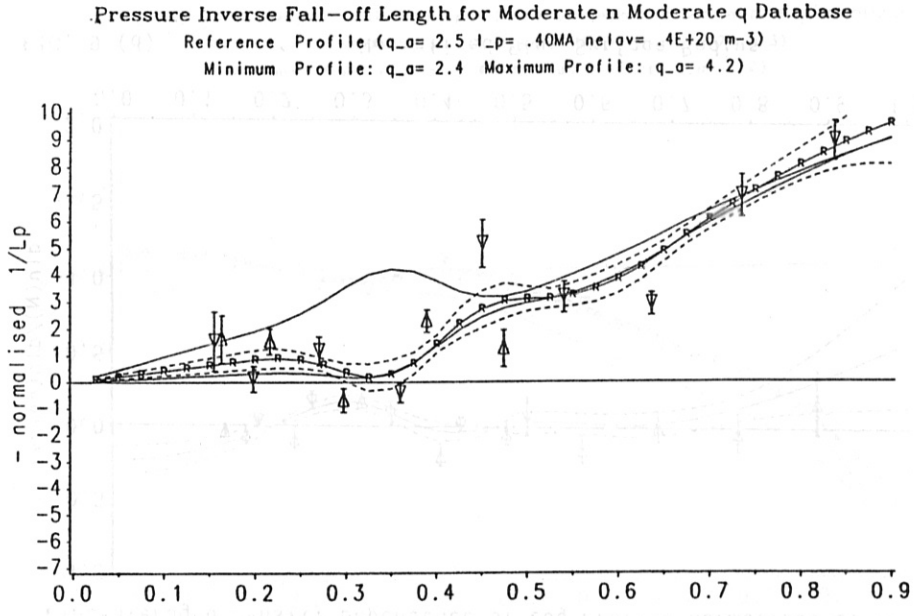

FIG. 11 (a)

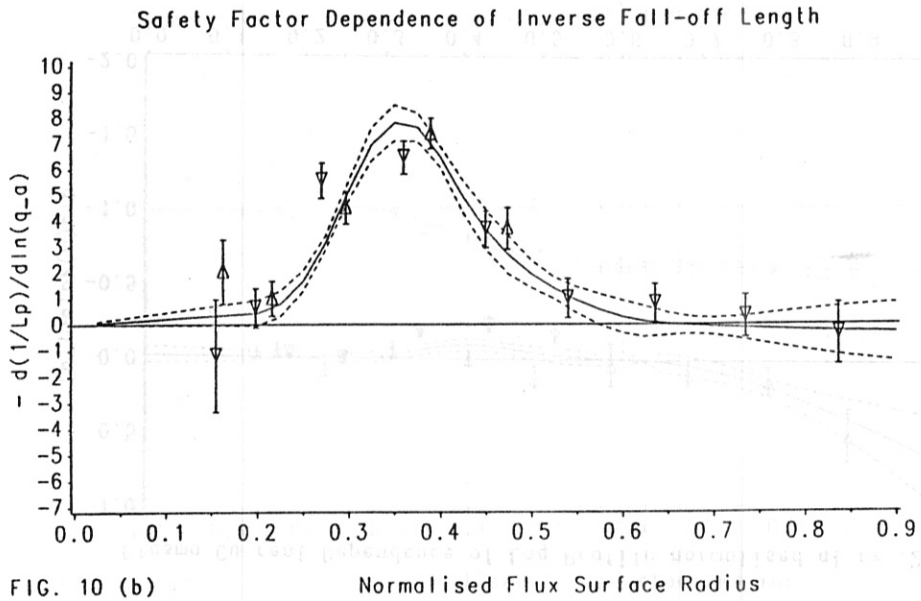

FIG. 10 (b)

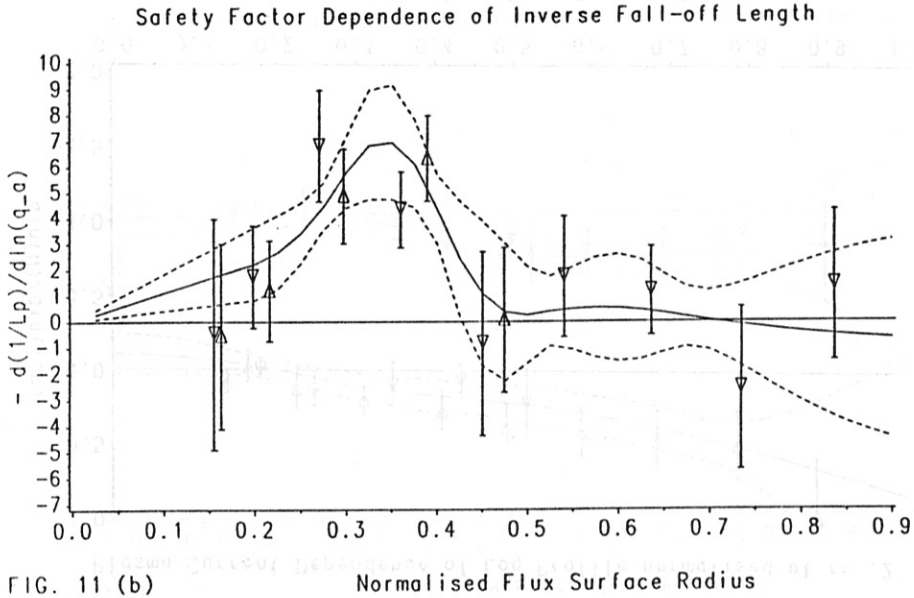

FIG. 11 (b)

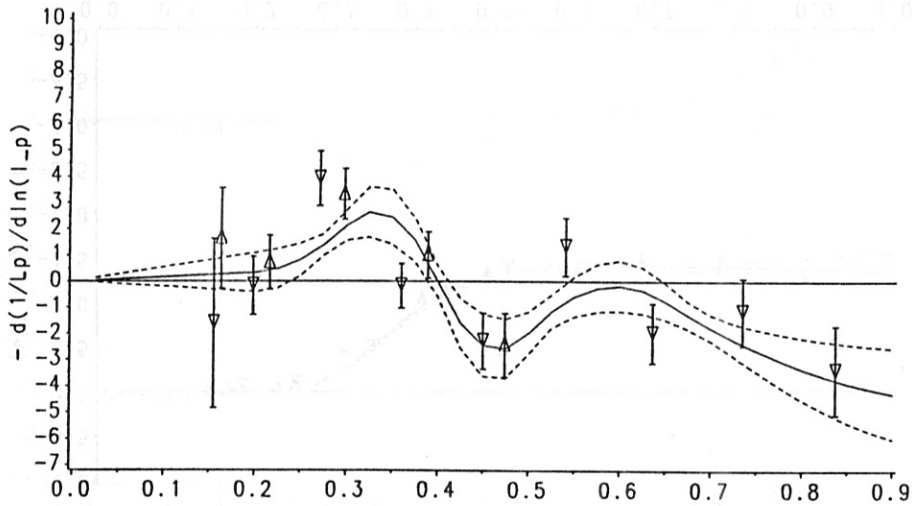

FIG. 10 (c)

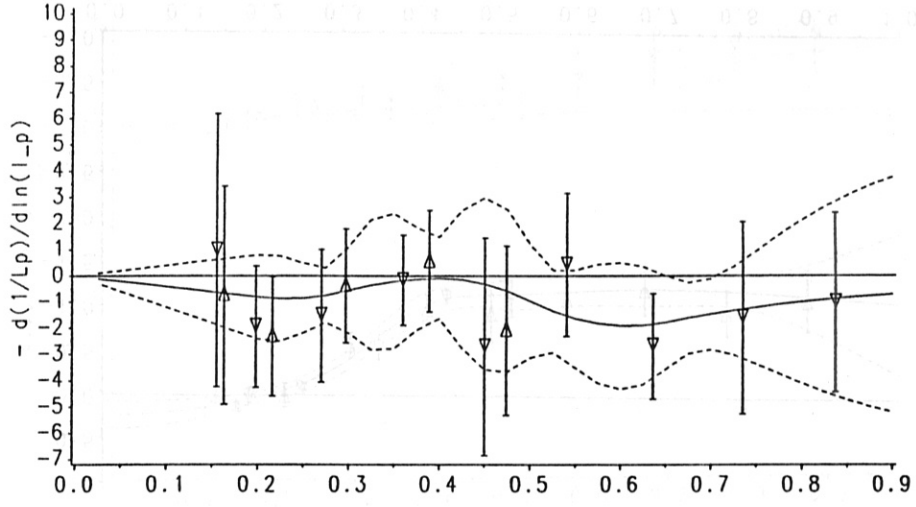

FIG. 11 (c)

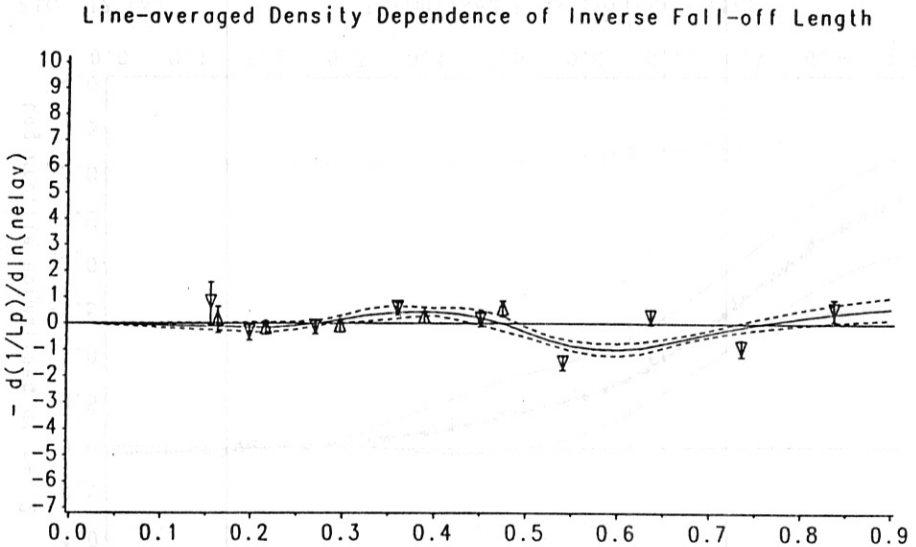

FIG. 10 (d)

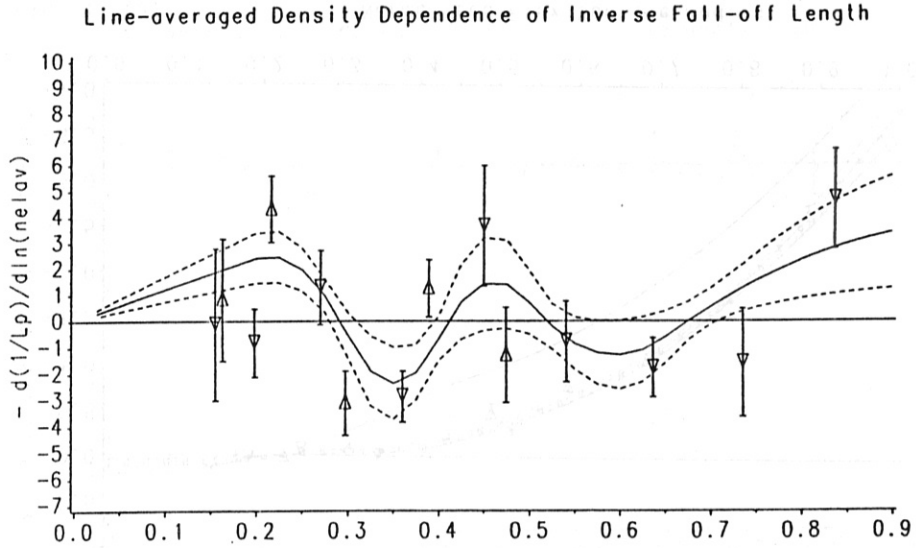

FIG. 11 (d)

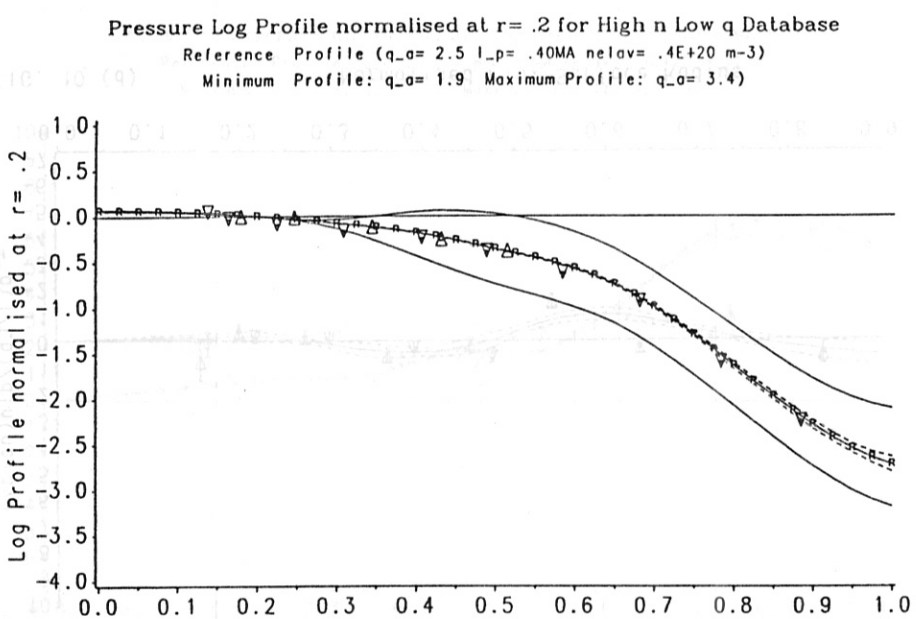

FIG. 12 (a)

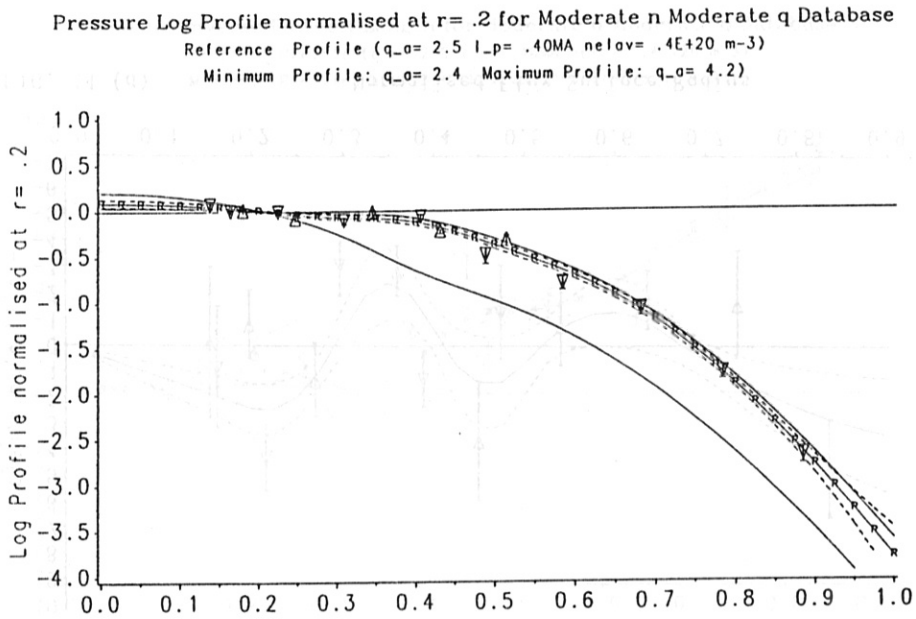

FIG. 13 (a)

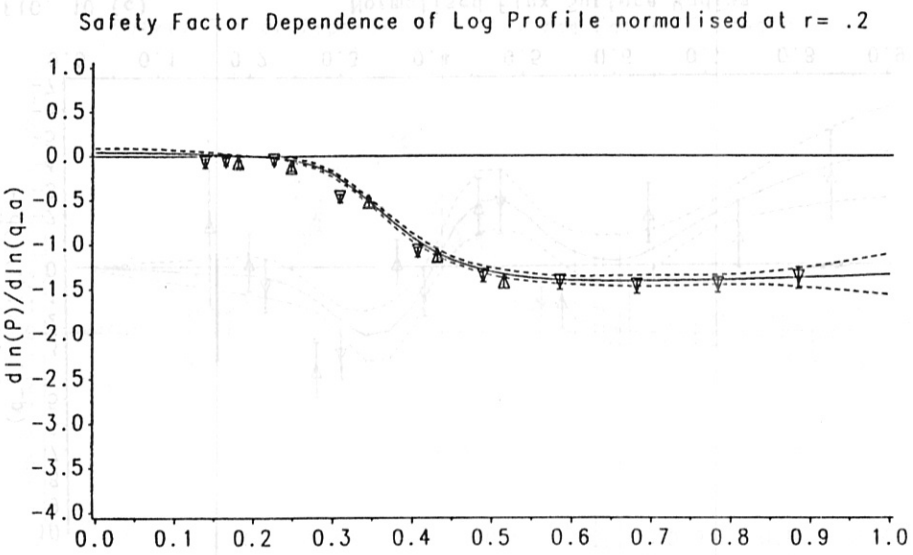

FIG. 12 (b)

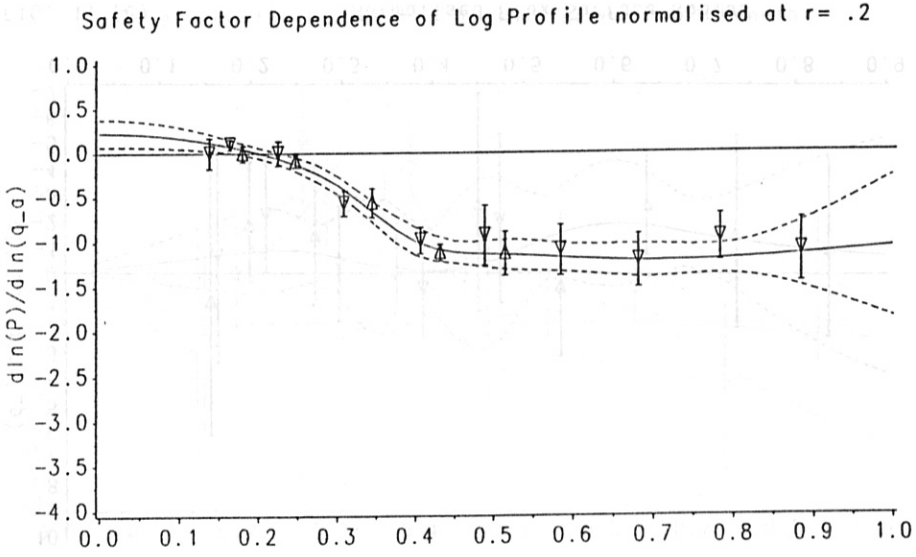

FIG. 13 (b)

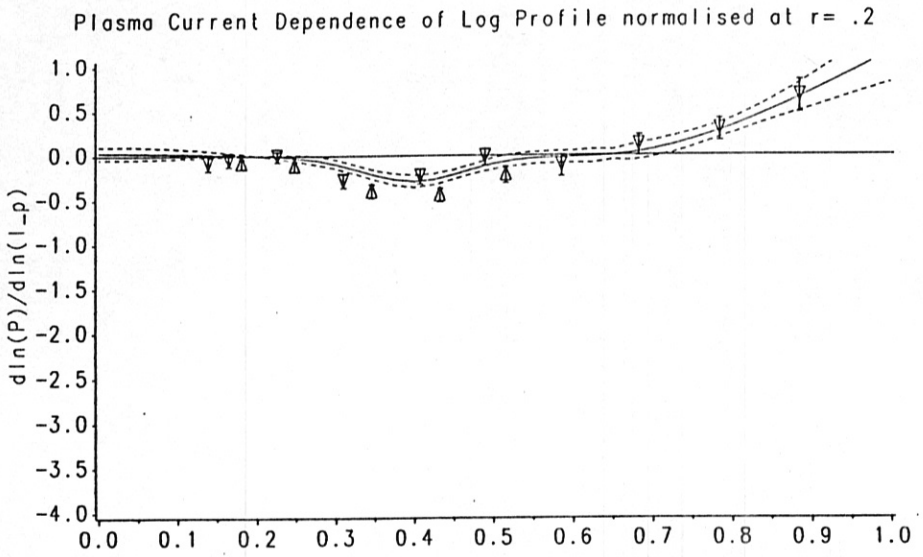

FIG. 12 (c)

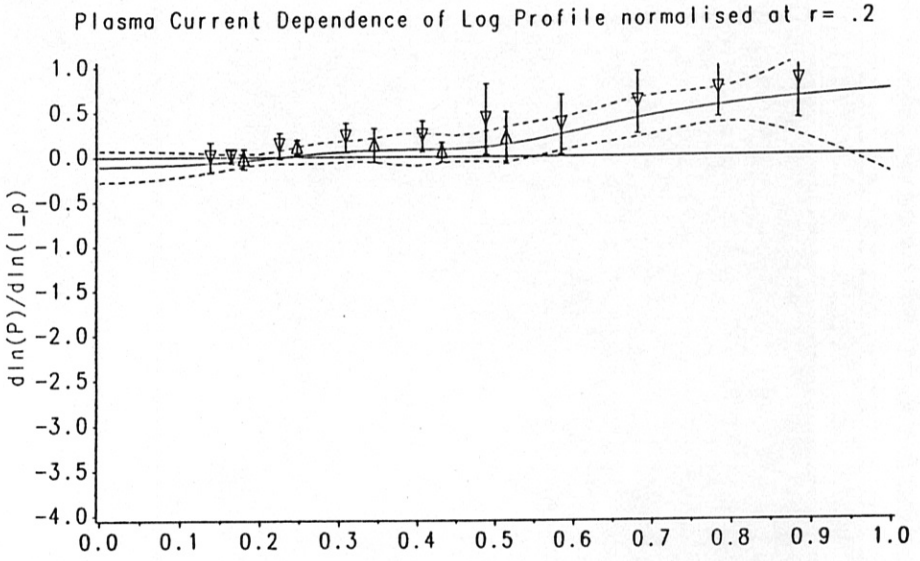

FIG. 13 (c)

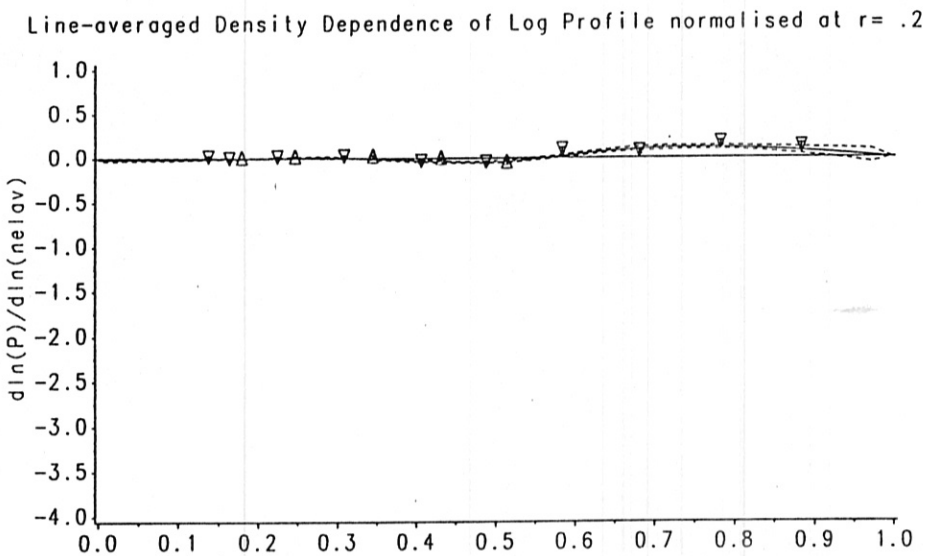

FIG. 12 (d)

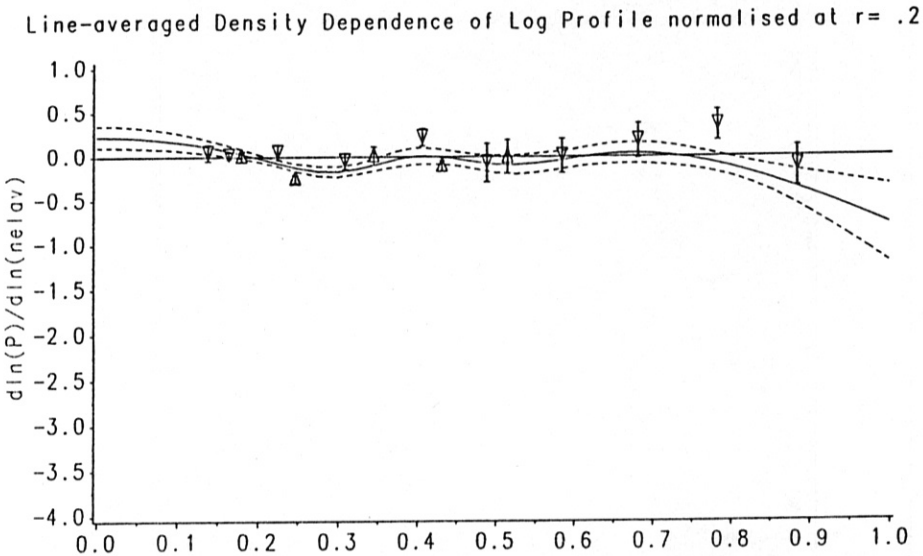

FIG. 13 (d)