SCALINGS AND PLASMA PROFILE PARAMETERIZATION
OF ASDEX HIGH DENSITY OHMIC DISCHARGES

P.J. McCarthy*, K.S. Riedel**, O.J.W.F. Kardaun,
H.D. Murmann, K. Lackner, and the ASDEX Team
Max-Planck-Institut für Plasmaphysik,
IPP-Euratom Association,
Garching bei München,
Germany

ABSTRACT. A database of high density (0.3 < n_e [10^{20} m^{-3}] < 0.8), low q_a (1.9 < q_a < 3.4), Ohmic discharges from the ASDEX experiment is analysed statistically. Bulk parameter scalings and parameterized temperature and density profile shapes are presented. The total plasma kinetic energy, assuming T_f = T_e, scales as \( n_e^{-5.34 \pm 0.01} I_p^{0.90 \pm 0.06} \) and is almost independent of the toroidal magnetic field. The electron temperature profile peaking factor scales as \( T_{eq}^{0.94 \pm 0.04} q_a^{0.7 \pm 0.04} \), 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 (0.2 < n_e/[10^{20} m^{-3}] < 0.4, 2.4 < q_a < 4.2), 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. Generally, in moderate field devices a similar but weaker increase in confinement time at small to moderate densities is observed [1]. Unfortunately, the Ohmic energy confinement time saturates at higher values of the Murakami parameter \( \bar{n}_e R/B_t \) [2]. This phenomenon is termed density rollover and, under usual operating conditions in ASDEX, occurs at line averaged densities of about 0.25 \times 10^{20} m^{-3} and magnetic fields of the order of 2 T. (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. A second database of moderate density and moderate edge safety factor q_a is presented for comparison. For a detailed discussion of the theoretical aspects of profile shape determination, the reader is referred to Ref. [4].

In our databases, neither of which extends down to the linear regime, no significant change was observed in the scaling of the total kinetic plasma energy as the rollover density was crossed and exceeded. Plasma energy and the Ohmic power scalings nearly cancel, leaving a moderate q_a dependence (\( T_e \sim q_a^{0.3} \)) in the confinement time.

To analyse the profile shapes, we fit all the measured profiles simultaneously by means of a radial spline function, each of whose coefficients depends on the plasma parameters \( q_a, I_p \) and \( \bar{n}_e \). This powerful technique enables us to quantify the various parametric dependences as a function of radius. For the electron temperature profile, we find that the profile shape variation consists almost exclusively of a \( q_a \) dependence confined to \( r/a \leq 0.60 \). The density profile also exhibits a \( q_a \) dependence, though it is weaker than that of the temperature. Unlike the temperature, however, we find a broadening of the density profile near the plasma edge with increasing plasma current.

Our research systematizes and validates earlier graphical and qualitative studies of 'profile consistency' (see Ref. [5] for a review of this topic). The statistical approach enables us to make quantitative statements about the relative strength of the interior \( q_a \) dependence and the weaker exterior parametric dependences of the profile shape.
Other authors have previously reported a variety of different results on the peaking factor and the interior domain scalings. The multiplicity of results arises from a combination of tokamak to tokamak differences, different physics regimes, systematic measurement errors, and sometimes also from the use of relatively unsophisticated analysis methods. We find, for instance, that normalizing by \((T_3/2)^{2/3}\) instead of the conventional \(T\) in the calculation of the temperature profile peaking factor permits a more critical test of the assumption that the electron temperature and current density profiles are in resistive equilibrium. Using the result that the (inverse) fall-off length is solely dependent on \(q_a\), we show that the temperature profile shape cannot be Gaussian.

In Section 2, we describe our low \(q_a\), high density and moderate \(q_a\), moderate density databases. In Section 3, bulk plasma parameter regressions are presented including new results for the scalings of the plasma energy \(W_p\) and \(Z_{\text{eff}}\). In Section 4, peaking factors and central temperature scalings are examined. In Section 5, the spline model and the statistical methods used in the profile analysis are summarized. In Section 6, a number of practical techniques to improve the analysis are discussed. One of these is to separate the parameterization of the profile shape and magnitude. In Section 7, our experimental shape analysis results are presented in detail. Section 8 gives a discussion and summary.

In the Appendix, we discuss the statistical significance of each regression variable in a least squares regression model. Sections 5 and 6 and the Appendix may be omitted by those readers who are interested only in the results and not in the statistical methods.

### 2. DATABASE DESCRIPTION

The main (high density) database consists of 105 pairs of time 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 ms intervals using the ASDEX Nd:YAG Thomson scattering system [6]. 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 has the effect of strongly reducing profile fluctuations due to sawteeth. Such averaging is free from bias as long as the sawtooth period (\(\approx 10\) ms) is incommensurate with the sampling period.

### TABLE I(a). SUMMARY OF THE HIGH \(n\), LOW \(q\) DATABASE

| Parameter               | Mean   | Minimum | Maximum |
|-------------------------|--------|---------|---------|
| \(q_a\)                 | 2.432  | 1.858   | 3.370   |
| \(I_p\) (MA)            | 0.388  | 0.321   | 0.483   |
| \(n_e\) (10^{20} m^{-3})| 0.422  | 0.326   | 0.823   |
| \(T_e\) (T)             | 2.011  | 1.754   | 2.730   |
| Loop voltage (V)        | 1.325  | 1.124   | 1.760   |
| Ohmic power (MW)        | 0.516  | 0.362   | 0.811   |
| \(T_{\text{eff}}\) (keV)| 0.380  | 0.271   | 0.482   |
| \(\tilde{n}_e\) (10^{20} m^{-3})| 0.391 | 0.271 | 0.712 |
| \(Z_{\text{eff}}\) Spitzer| 1.800 | 1.391  | 2.155   |
| \(R_{\text{magn}}\) (m) | 1.684  | 1.674   | 1.715   |
| \(Z_{\text{magn}}\) (m) | -0.000 | -0.001  | 0.000   |
| \(b_{\text{minor}}\) (m) | 0.400  | 0.398   | 0.406   |
| Area (m^2)              | 0.480  | 0.475   | 0.495   |
| \(\beta_p + \ell/2\)    | 0.905  | 0.762   | 1.056   |
| \(\beta_{\text{p,dis}}\) | 0.386  | 0.277   | 0.528   |
| \(W_{\text{kin}}\) (kJ) | 38.6   | 27.8    | 60.9    |
| \(W_{\text{dis}}\) (kJ) | 44.9   | 34.3    | 70.1    |
| \(W_{\text{pmhd}}\) (kJ) | 46.4   | 35.6    | 69.6    |
| \(\tau_{\text{kin}}\) (ms) | 74.8  | 55.1    | 97.9    |
| \(\tau_{\text{dis}}\) (ms) | 87.6   | 71.5    | 113.0   |
| \(\tau_{\text{pmhd}}\) (ms) | 90.3   | 71.4    | 107.0   |
TABLE I(b). SUMMARY OF THE MODERATE n, MODERATE q DATABASE

| Parameter                  | Mean    | Minimum | Maximum |
|----------------------------|---------|---------|---------|
| $q_a$                      | 3.088   | 2.382   | 4.195   |
| $I_p$ (MA)                 | 0.355   | 0.281   | 0.452   |
| $n_e^{\text{av.}}$ (10^{20} m^{-3}) | 0.306   | 0.163   | 0.399   |
| $B_t$ (T)                  | 2.335   | 1.682   | 2.813   |
| Loop voltage (V)           | 1.072   | 0.920   | 1.225   |
| Ohmic power (MW)           | 0.383   | 0.269   | 0.495   |
| $T_e^{\text{av.}}$ (keV)   | 0.490   | 0.388   | 0.678   |
| $n_e^{\text{vol.}}$ (10^{20} m^{-3}) | 0.247   | 0.128   | 0.296   |
| $Z_{\text{eff Spitzer}}$   | 2.587   | 2.040   | 3.504   |
| $R_{\text{mag axis}}$ (m)  | 1.690   | 1.678   | 1.712   |
| $Z_{\text{mag axis}}$ (m)  | 0.001   | -0.000  | 0.013   |
| $b_{\text{minor}}$ (m)     | 0.398   | 0.391   | 0.404   |
| Area (m^2)                 | 0.478   | 0.461   | 0.496   |
| $\beta_p + \ell/2$         | 0.970   | 0.857   | 1.142   |
| $\beta_\perp$              | 0.398   | 0.304   | 0.542   |
| $W_{\text{pin}}$ (kJ)      | 33.5    | 18.8    | 42.9    |
| $W_{\text{dia}}$ (kJ)      | 38.3    | 26.0    | 50.3    |
| $W_{\text{mhd}}$ (kJ)      | 40.8    | 26.9    | 51.7    |
| $\tau_{\text{kin}}$ (ms)   | 88.0    | 68.0    | 104.8   |
| $\tau_{\text{dia}}$ (ms)   | 101.1   | 80.0    | 124.0   |
| $\tau_{\text{mhd}}$ (ms)   | 106.2   | 95.2    | 124.7   |

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 data points are clustered at toroidal magnetic field values of 1.8 T or 2.2 T. Nevertheless, the distinction between $q_a$, $B_t$, and $I_p$ scaling will be quite apparent. Bolometric measurements showed that radiation losses for these discharges ($10\% < P_{\text{rad}}/P_{\text{Ohmic}} < 40\%$, increasing with density) were strongly localized at the plasma edge.

For comparison, we present a complementary analysis of a second database (38 compressed data points from 38 deuterium discharges) of moderate density ($n_e = 0.3 \times 10^{20} \text{ m}^{-3}$), moderate $q_a$ ($q_a = 3.1$), non-gettered, Ohmic discharges, which is summarized in Table I(b). On ASDEX, these densities are centred around the transition to the rollover 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 carbonized walls and three days of operating with helium 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 in the fits to these experimental data to be larger. For these reasons, we include the scalings of the second database largely for purposes of comparison. The distribution of data points for each dataset is displayed in the combined Hugill plot (Fig. 1).

The two databases differ in the three parameters: \( \bar{n}_e \), \( q_a \) and wall condition (gettered/ non-gettered). Thus, we can view the combined database as being clustered in two cells of eight possible combinations. Indeed, 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 dependences of the outer section of the normalized density profiles. If the two databases 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 normalized flux surface co-ordinate (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 \( \bar{n}_e \) and by two of the three parameters \( I_p \), \( q_a \) and \( B_t \). These parameters represent the major control parameters which the experimentalist utilizes 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 \( Z_{\text{eff}} \) or the total Ohmic power \( P_0 \) is sometimes used as an extra independent variable. Both \( Z_{\text{eff}} \) and \( P_0 \) depend on the control parameters \( \bar{n}_e \), \( I_p \) and \( q_a \), and therefore are not purely measures of the condition of the plasma wall.

### TABLE II(a). CORRELATION MATRIX FOR THE HIGH \( n \), LOW \( q \) DATABASE

|          | \( \ln \bar{n}_e \) | \( \ln I_p \) | \( \ln q_a \) | \( \ln Z_{\text{eff}} \) | \( \ln P_0 \) |
|----------|----------------|-------------|-------------|----------------|-------------|
| \( \ln \bar{n}_e \) | 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 |
| \( \ln Z_{\text{eff}} \) | -0.819 | 0.208 | -0.034 | 1.000 | -0.359 |
| \( \ln P_0 \) | 0.728 | 0.750 | -0.636 | -0.359 | 1.000 |

PCA for \( \ln \bar{n}_e \), \( \ln I_p \), \( \ln q_a \), \( \ln Z_{\text{eff}} \), \( \ln P_0 \)

| Eigenvalues of 5 \( \times \) 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 \( \ln \bar{n}_e \), \( \ln I_p \), \( \ln q_a \)

| Eigenvalues of 3 \( \times \) 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 FOR THE MODERATE \( n \), MODERATE \( q \) DATABASE

|          | \( \ln \bar{n}_e \) | \( \ln I_p \) | \( \ln q_a \) | \( \ln Z_{\text{eff}} \) | \( \ln P_0 \) |
|----------|----------------|-------------|-------------|----------------|-------------|
| \( \ln \bar{n}_e \) | 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 |
| \( \ln Z_{\text{eff}} \) | -0.724 | 0.225 | -0.371 | 1.000 | -0.040 |
| \( \ln P_0 \) | 0.528 | 0.937 | -0.722 | 0.040 | 1.000 |

PCA for \( \ln \bar{n}_e \), \( \ln I_p \), \( \ln q_a \), \( \ln Z_{\text{eff}} \), \( \ln P_0 \)

| Eigenvalues of 5 \( \times \) 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 \( \ln \bar{n}_e \), \( \ln I_p \), \( \ln q_a \)

| Eigenvalues of 3 \( \times \) 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 II(c). SOME CORRELATIONS INVOLVING \( I, B, q, \) and \( n \)

| Database | \( \ln(I, B) \) | \( \ln(I, q) \) | \( \ln(B, q) \) | \( \ln(IB, q) \) | \( \ln(B, n) \) | \( \ln(BI, n) \) |
|----------|----------------|-------------|-------------|---------------|-------------|-------------|
| High \( n \) | -0.07 | -0.70 | 0.76 | 0.08 | -0.04 | 0.32 |
| Moderate \( n \) | 0.29 | -0.67 | 0.53 | -0.14 | 0.21 | 0.11 |
To examine the extent to which the Ohmic power and $Z_{\text{eff}}$ vary independently of the control parameters $n_e$, $I_p$, and $q_a$, we carry out a principal component analysis (PCA) [7]: The correlation matrix of the logarithms of $n_e$, $I_p$, $q_a$, $Z_{\text{eff}}$, and $P_n$ is calculated and diagonalized (see Table II). Each eigenvalue is interpreted as the sample variance of the 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 (standardized) data variation can be explained by the first three components. The residual 3% of the total variance is attributed to noise in the temperature measurements, which affects $Z_{\text{eff}}$, and noise in the loop voltage, which affects both $Z_{\text{eff}}$ and $P_n$. Hence, we discard the principal components having the two smallest eigenvalues.

The remaining eigenvectors span a three-dimensional subspace. This subspace can be efficiently represented by linear combinations of any well conditioned set of three of the five variables. That $n_e$, $I_p$, and $q_a$ constitute such a well conditioned set is checked by re-doing the PCA for these three parameters alone. The eigenvalues of the $3 \times 3$ 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 generally preferred to use the logarithms of $n_e$, $B$, and $I_p$ because the correlation between $I_p$ and $B$ was much less than that of either variable with $q_a$ (Table II(c)). However, some scalings are given in terms of $n_e$, $q_a$, and $B_i$, where the latter parameter was chosen to maintain a set of nearly uncorrelated bulk variables. The dominance of $q_a$ and the weak $I_p$ influence in determining plasma profile shapes motivated us to use $n_e$, $I_p$, and $q_a$ for the profile shape analysis.

3. BULK SCALINGS

In this section, we examine the scalings of the bulk plasma variables. Let us briefly summarize our results: The scaling of the total plasma energy as determined by kinetic measurements is virtually the same in both databases, i.e. no transition was observed going from the rollover regime to the saturated Ohmic confinement (SOC) regime. However, the diamagnetic and equilibrium MHD estimates of this parameter show strong differences between the two databases. The plasma energy and the Ohmic heating power have roughly similar parametric dependences, which results in a confinement time scaling consisting mainly of a rather moderate $q_a$ dependence. Power law scalings for $Z_{\text{eff}}$ and for $Z_{\text{eff}} - 1$ are also presented. The latter parameter scales like $1/n_e$, indicating that the impurity density is independent of $n_e$.

We use the Spitzer value for $Z_{\text{eff}}$, 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 bremsstrahlung measurements on ASDEX [8] show that for the SOC regime (to which our high density database belongs), $Z_{\text{eff}}$ is very flat over most of the plasma radius, though tending to rise strongly near the boundary. We adopt here the conventional assumption of zero radial dependence. The ion density is calculated with the assumption that the sole impurity is oxygen. The plasma kinetic energy, $W_{\text{kin}} = \frac{1}{2} \int (n_e T_e + n_i T_i) \, dv$, is calculated by assuming that the ion temperature is equal to the electron temperature, $T_i = T_e$. This assumption is justified when then electron–ion energy exchange time is much shorter than the energy confinement time. In the main database, the typical values are $\tau_{ei} = 5 \text{ ms}$ and $\tau_E = 75 \text{ ms}$. In some of the moderate density discharges, $\tau_{ei} \ll \tau_E$ is not satisfied (for five of these discharges, $\tau_{ei}/\tau_E$ at the plasma centre lies in the range 0.25 to 0.38).

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 ($\psi_{\text{dia}}/\psi_{\text{tor}} = 10^4$) to such factors as slight mechanical displacements of the diamagnetic loop means, however, that the typical error associated with the derived value for poloidal beta is $\delta(\beta_{\text{pol}}) = \pm 0.05$. For $\beta_{\text{pol}} = 0.3$ (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 diamagnetic energy is systematically greater than the kinetic energy with an energy independent offset of $\approx 6 \pm 2 \text{ kJ}$ (see Fig. 2).

A third measure of the plasma energy is derived from equilibrium magnetic measurements, from which the parameter $\beta_p + \ell_i/2$ can be recovered. To isolate $\beta_p$, we need an estimate for $\ell_i/2$, which, as is well known,
cannot be determined from equilibrium data in the case of circular plasmas. From an earlier investigation, we use the following empirical relation from a parameterization of current density profiles derived from experimental temperature profiles and the assumption of Spitzer resistive equilibrium: $\ell/2 = 0.332 + 0.199 \ln q_a$.

The regression models considered are of the form

$$y = \sum \alpha_i x_i + \epsilon$$

where $y$ and $x$ denote the logarithms of the dependent and the independent bulk plasma variables, respectively. The root mean square error (RMSE) of the fit is

$$\sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N - p}}$$

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

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

3.1. Total plasma energy scaling

The plasma energy content from kinetic data for the main database satisfies

$$W_{p_{\text{kin}}} = 133(\pm 6) n_e^{0.54 \pm 0.01} I_p^{0.92 \pm 0.04} B^0.06 \pm 0.04$$

RMSE = 0.04, $R^2 = 0.965$

Similarly, the scaling law based on the secondary database is

$$W_{p_{\text{kin}}} = 128(\pm 6) n_e^{0.33 \pm 0.03} I_p^{0.80 \pm 0.05} B^0.13 \pm 0.05$$

RMSE = 0.04, $R^2 = 0.959$

The plasma energy content inferred from the diamagnetic flux for the main database is described by

$$W_{p_{\text{dia}}} = 130(\pm 6) n_e^{0.40 \pm 0.01} I_p^{0.88 \pm 0.04} B^0.13 \pm 0.04$$

RMSE = 0.04, $R^2 = 0.943$

whereas the secondary database gives a poorer fit:

$$W_{p_{\text{dia}}} = 83(\pm 10) n_e^{0.21 \pm 0.05} I_p^{0.77 \pm 0.07} B^0.33 \pm 0.08$$

RMSE = 0.06, $R^2 = 0.878$

The plasma energy content derived from equilibrium magnetic measurements for the main database is described by

$$W_{p_{\text{mhd}}} = 156(\pm 5) n_e^{0.41 \pm 0.01} I_p^{0.94 \pm 0.03} B^0.01 \pm 0.03$$

RMSE = 0.03, $R^2 = 0.968$

whereas the secondary database gives

$$W_{p_{\text{mhd}}} = 255(\pm 28) n_e^{0.26 \pm 0.04} I_p^{1.19 \pm 0.06} B^0.36 \pm 0.07$$

RMSE = 0.05, $R^2 = 0.940$

The criterion for ‘significance’ of a regressor is whether or not the regression coefficient is at least twice its standard deviation. In the Appendix, the statistical background of this simple rule and its relation with $R^2$ is discussed.

From the indicated uncertainties (one standard deviation) of the regression coefficients, we see, for instance,
that the B, scaling for \( \text{W}_{\text{pin}} \) is insignificant for the main database, while in the secondary database it is (just about) statistically significant. In both cases, omission of B, affects the goodness of fit only slightly in absolute terms (Eq. (A.7) gives \( \Delta(R^2) = 0.001 \) and 0.007, respectively).

Because of the design of the database, the correlations between \( n_e \), \( I_p \) and B, are small (see Table II(c)). This results in small correlations between the regression estimates of the exponents of \( n_e \), \( I_p \) and B,. In our regressions, all these correlations are less than 0.15 for the main database and less than 0.35 for the secondary database. For assessing the significance of the difference between a postulated scaling law and our empirical scalings, such low correlations between the estimates can be neglected in practice.

The toroidal field shows the most extreme variation in the six \( \text{W}_p \) scalings presented above. There is also, however, a strong tendency for the \( I_p \) coefficient to move in the direction opposite to B, when going from one scaling to another. This suggests a representation with \( q_a \). To preserve the near independence of the regression estimates, we choose BI as the conjugate variable [9]. In the new representation, we have the following regression coefficients for the six cases (same order as above):

\[
\begin{align*}
\text{BI}_{Ip}: & \quad (0.49, 0.46, 0.52, 0.55, 0.48, 0.42) \\
q_a: & \quad (-0.42, -0.32, -0.37, -0.20, -0.47, -0.75)
\end{align*}
\]

The errors in the coefficients are roughly the same as in the \((I_p, B_s)\) scaling. Obviously, to a good approximation, the simple transformation \( q_{\text{BI}} = \frac{1}{2}(\alpha_B + \alpha_I) \) and \( q_a = \frac{1}{2}(\alpha_B - \alpha_I) \) holds. We roughly summarize:

\[
\begin{align*}
\text{W}_{\text{pmhd}}: & \quad (\text{BI})^{-0.5 \pm 0.05} q_{\text{BI}}^{0.5 \pm 0.25}, \text{ where here the approximate ranges are indicated for the regression coefficients in the six cases above. Hence, at constant } n_e, \\
\text{BI scaling is nearly constant, but the } q_a \text{ scaling varies considerably with the type of measurement and with the database.}
\end{align*}
\]

The observed \( q_a \) scaling differences may be partly due to physics and partly to systematic errors. \( \text{W}_{\text{pmhd}} \) is particularly vulnerable to errors in the \( \pm 50\% \) correction to \( \beta_p + \ell/2 \). To get an idea of the influence of this error, we have performed a sensitivity analysis. Using \( \ell/2 = 0.332 + \delta_1 + (0.199 + \delta_2) \) in \( q_a \), we made a number of \( \text{W}_{\text{pmhd}} \) regressions for different choices of \( \delta_1 \) and \( \delta_2 \) in the range \(-0.1 \) to \(+0.1 \). (Note that \( \delta_1 \) can describe an error in \( \beta_p + \ell/2 \) as well as in \( \ell/2 \).) We find that the exponent for \( q_a \) in the \( \text{W}_{\text{pmhd}} \) regression varies as \(-0.47 + 1.3\delta_1 - 1.6\delta_2 \) for the main database and \(-0.75 + 0.43\delta_1 - 2.0\delta_2 \) for the secondary database. Thus, an error of \(-0.1 \) in the \( \ln q_a \) coefficient for \( \ell/2 \) (we would expect this to be an extreme case) gives \( q_a \) exponents of \(-0.31 \) and \(-0.55 \) for the main and secondary databases, respectively.

We conclude that this error source cannot reconcile the scaling of \( \text{W}_{\text{pmhd}} \) with \( q_a \) for the secondary database with the remaining five \( \text{W}_p \) scalings.

Now we turn to \( \text{W}_{\text{pdia}} \). Suppose we have an offset \( \delta_B \) in \( \beta_{\text{pdia}} \). A sensitivity analysis showed that this roughly gives an offset of \(-2\delta_B \) in the \( q_a \) coefficient (for both databases). The density dependence is offset by \(-\delta_3 \) for the main database and by \(-0.5\delta_3 \) for the secondary database. It is noted that the observed systematic difference of 6 kJ between the \( \text{W}_{\text{pin}} \) and \( \text{W}_{\text{pdia}} \) measurements corresponds (at \( I_p = 0.4 \text{ MA} \)) to a difference in \( \beta_p \) of 0.05, which is consistent with the observed difference in the \( n_e \) dependence. It does not, however, offer a satisfactory explanation for the \( \text{W}_{\text{pdia}} \) or \( \text{W}_{\text{pmhd}} \) scaling differences between the databases, for which we must refer to the above mentioned differences in operating regimes (gettered/non-gettered walls, high densities/rollover densities, etc.).

### 3.2. Volume averaged electron temperature

Both databases are in rough agreement with the Pfeiffer-Waltz [10] and JET [11] Ohmic scalings. The main database satisfies

\[
\langle T_e \rangle = 0.639(\pm 0.03) \quad n_e^{-0.52 \pm 0.01} \quad I_p^{-0.97 \pm 0.04} \quad B_{\text{tr}}^{-0.2 \pm 0.04} \quad \text{RMSE} = 0.04, \quad R^2 = 0.947
\]

Similarly, the secondary database satisfies

\[
\langle T_e \rangle = 0.605(\pm 0.06) \quad n_e^{-0.61 \pm 0.04} \quad I_p^{0.93 \pm 0.06} \quad B_{\text{tr}}^{0.04 \pm 0.07} \quad \text{RMSE} = 0.05, \quad R^2 = 0.913
\]

Here, B, is insignificant for both databases and its omission has a very small (\( \approx 0.001 \)) effect on \( R^2 \). Though a natural candidate for determining the temperature, we have not included \( Z_{\text{eff}} \) in the list of regressors, since, in our case, it is derived, assuming Spitzer resistivity, from the temperature profile itself:

\[
Z_{\text{eff}} N(Z_{\text{eff}}) \propto \frac{(T^{3/2})_{\text{area}} V_{\text{loop}}}{I_p R_{\text{plas}}}
\]
where $N(Z_{\text{eff}})$ is the Spitzer–Harm factor (ranging from $N(1) = 1$ to $N(\infty) = 0.58$) and the radial variation of the Coulomb logarithm is neglected.

**3.3. 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 main database, the loop voltage scales as

$$V_{\text{loop}} = 2.38(\pm 0.1) \, n_e^{-0.36 \pm 0.02} \, I_p^{0.12 \pm 0.05} \, B_t^{-0.27 \pm 0.04}$$

RMSE = 0.04, $R^2 = 0.864$

Similarly, the secondary database satisfies

$$V_{\text{loop}} = 2.40(\pm 0.1) \, n_e^{-0.30 \pm 0.02} \, I_p^{0.19 \pm 0.03} \, B_t^{-0.31 \pm 0.03}$$

RMSE = 0.03, $R^2 = 0.896$

Because of the additional factor of $I_p$ on both sides, we obtain necessarily higher $R^2$ values (0.935 and 0.984) for the Ohmic power regressions (the RMSE values are unchanged).

**3.4. Spitzer $Z_{\text{eff}}$**

For the main database, the Spitzer $Z_{\text{eff}}$ scales as

$$Z_{\text{eff}} = 2.90(\pm 0.18) \, n_e^{-0.48 \pm 0.02} \, I_p^{0.67 \pm 0.05} \, B_t^{-0.34 \pm 0.05}$$

RMSE = 0.05, $R^2 = 0.893$

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

$$Z_{\text{eff}} - 1 = 2.21(\pm 0.33) \, n_e^{-1.14 \pm 0.04} \, I_p^{1.51 \pm 0.12} \, B_t^{-0.81 \pm 0.12}$$

RMSE = 0.18, $R^2 = 0.844$

For the secondary database, the Spitzer $Z_{\text{eff}}$ scales as

$$Z_{\text{eff}} = 2.67(\pm 0.58) \, n_e^{-0.69 \pm 0.08} \, I_p^{0.64 \pm 0.11} \, B_t^{-0.22 \pm 0.13}$$

RMSE = 0.09, $R^2 = 0.714$

The impurity scaling yields

$$Z_{\text{eff}} - 1 = 1.83(\pm 0.71) \, n_e^{-1.10 \pm 0.14} \, I_p^{0.07 \pm 0.19} \, B_t^{-0.43 \pm 0.21}$$

RMSE = 0.16, $R^2 = 0.709$

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

$$Z_{\text{eff}} = 1 + \frac{\sum n_i (Z_i^2 - Z_i)}{n_e}$$

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

**3.5. Energy confinement time**

The $I_p$ and $n_e$ dependences of the total plasma energy and the Ohmic power approximately cancel, leaving a relatively weak $q_a$ dependence of $\tau_E$. This makes the energy confinement scalings less pronounced. The main database yields the following scaling for the kinetic $\tau_E$:

$$\tau_{E, \text{kin}} = 56(\pm 5) \, n_e^{0.18 \pm 0.03} \, I_p^{-0.21 \pm 0.08} \, B_t^{0.34 \pm 0.08}$$

RMSE = 0.07, $R^2 = 0.392$

The secondary database fit yields

$$\tau_{E, \text{kin}} = 53(\pm 6) \, n_e^{0.23 \pm 0.05} \, I_p^{-0.39 \pm 0.06} \, B_t^{0.44 \pm 0.07}$$

RMSE = 0.05, $R^2 = 0.709$

As explained in the discussion of the $W_p$ scaling, we prefer the representation in terms of $B_t I_p$ and $q_a$:

$$\tau_{E, \text{kin}} = 68(\pm 4) \, n_e^{0.18 \pm 0.03} \, (B_t I_p)^{0.06 \pm 0.06} \, q_a^{0.28 \pm 0.05}$$

(high density database)

$$\tau_{E, \text{kin}} = 72(\pm 6) \, n_e^{0.22 \pm 0.05} \, (B_t I_p)^{0.02 \pm 0.04} \, q_a^{0.41 \pm 0.05}$$

(moderate density database)

The confinement time derived from the diamagnetic measurement of the energy content for the main database scales as

$$\tau_{E, \text{dia}} = 69(\pm 3) \, n_e^{0.04 \pm 0.02} \, (B_t I_p)^{0.08 \pm 0.05} \, q_a^{0.33 \pm 0.04}$$

RMSE = 0.06, $R^2 = 0.403$
The secondary database fit yields

\[
\tau_{\text{Edia}} = 50(\pm 5) \, n_a^{-0.10 \pm 0.05} \, (B)^{0.11 \pm 0.05} \, q_a^{0.53 \pm 0.06}
\]

RMSE = 0.06, \( R^2 = 0.699 \)

The confinement time derived from equilibrium magnetic measurements for the main database scales as

\[
\tau_{\text{Emhd}} = 77(\pm 3) \, n_a^{0.04 \pm 0.02} \, (B)^{0.04 \pm 0.04} \, q_a^{0.23 \pm 0.04}
\]

RMSE = 0.06, \( R^2 = 0.260 \)
The secondary database fit yields

$$T_{\text{Emhd}} = 103(\pm 9) \, n_e^{-0.04 \pm 0.05} \, (B_l)^{-0.03 \pm 0.05} \, q_a^{-0.02 \pm 0.06}$$

$$\text{RMSE} = 0.06, \quad R^2 = 0.036$$

The difference in $q_a$ scaling between the two databases suggests the possibility of a quadratic (i.e. $\ln q_a \ln q_a$) dependence [9]. No curvature in the $\ln q_a$ dependence was apparent, however, in a plot of $\ln T_{\text{Edia}}$ for both databases with their respective $n_e$ and $B_l$ dependences removed. Figure 3 shows $\ln T_{\text{Edia}}/(B_l/0.8)^{0.08}(n_e/0.4)^{0.04}$ for the main database and $\ln T_{\text{Edia}}/(B_l/0.8)^{0.11}(n_e/0.4)^{-0.10}$ for the secondary database. A similar picture holds for $\ln T_{\text{Emhd}}$, whereas for $\ln T_{\text{Emhd}}$ there appeared to be a deterioration at high $q_a$ (figures not shown). Three points at low $q_a$ seem to be somewhat outlying in Fig. 3. However, their removal does not change the $\ln q_a$ regression coefficient by more than one standard deviation. Unlike some previous reported results (see, e.g. Ref. [12], Section 4), we found no significant confinement time deterioration with density in the SOC regime.

Figure 4 shows $T_{\text{Ekin}}$ (with the $I_p$ and $B_l$ dependences removed) versus $n_e$ for both databases, i.e. $\ln T_{\text{Ekin}}/(I_p/0.4)^{0.21}(B_l/2)^{0.34}$ for the main database and $\ln T_{\text{Ekin}}/(I_p/0.4)^{-0.39}(B_l/2)^{0.44}$ for the secondary database. A similar behaviour is observed for both $T_{\text{Edia}}$ and $T_{\text{Emhd}}$. For the difference in the $q_a$ scaling, the reader is referred to the sensitivity discussion in Section 3.1.

4. MAGNITUDE OF PROFILE AND TEMPERATURE PEAKING FACTOR

4.1. Electron temperature, density and pressure at $r = 0.2$

To enable reconstruction of the absolute profiles from the profile shape scalings presented later, we have regressed $T_e$, $n_e$, and $p_e = n_e T_e$ at the 20% flux radius. This normalization radius lies inside the inversion radius for our $q_a$ range and has data points on either side of it. (The YAG channel closest to the magnetic axis lies typically on the 14% flux radius.) The high density, low $q_a$ database satisfies

$$T_e^{20\%} = 0.369(\pm 0.01) \, n_e^{-0.49 \pm 0.01} \, I_p^{-0.29 \pm 0.04} \, B_l^{-0.71 \pm 0.04}$$

$$\text{RMSE} = 0.04, \quad R^2 = 0.937$$

Most noteworthy here is the $\sqrt{n_e}B_l$-like dependence of $p_e^{20\%}$ for both databases. On regressing $p_e$ at each of the five most central YAG channels, which typically lie between the 14% and 25% flux radii, we found a similar absence of an $I_p$ dependence (coefficients ranged from 0.0 to 0.05, with a typical standard deviation of 0.05 for the main database, and from -0.13 to 0.18, with a typical standard deviation of 0.075 for the secondary database). The remaining ten channels all exhibited strong $I_p$ dependences (coefficients up to 2.0). Hence, we note that the central electron pressure is independent of the total plasma current. Since the onset of $I_p$ dependence occurs for those channels whose radii roughly correspond to minimum values of the sawtooth inversion radius ($r_{\text{inv}}(\min) = 1/q_a(\max)$ = 0.30 and 0.24 for the main and secondary databases, respectively), we speculate that this $I_p$ independence is related to sawtooth stability and sawtooth induced transport.

4.2. Temperature profile peaking factor

Though we later analyse, as a function of radius, the bulk parameter dependences of the local shape parameter $L^1$, we present here results for the usual single-parameter measure of the temperature profile shape, namely the temperature profile peaking factor. We show that the observed peaking factor is very close to that expected assuming a Spitzer resistive equilibrium.
We first express the local cylindrical safety factor \( q(r) \) as

\[
q(r) = \frac{rB_r}{B_p(r)} = \frac{2\pi r^2 B_r}{\mu_0 R J(r)} = \frac{2B_r}{\mu_0 R} \langle J \rangle_r
\]

where

\[
\langle J \rangle_r = \frac{1}{\pi r^2} \int_0^r J(r')2\pi r'dr'
\]

is the current density averaged up to radius \( r \).

Assuming a classical resistive equilibrium, we immediately obtain (recall that \( r \) is normalized to 1 at the edge)

\[
\frac{\langle T^{3/2} \rangle_{r=0.2}}{\langle T^{3/2} \rangle_{r=1}} = \frac{q_{r=1}}{q_r}
\]

where the uncertainty arises from the fact that we neglect radial variations (assumed to be weak) in \( Z_{\text{eff}} \) or in the Coulomb logarithm. Since all profiles are sawtoothing, we have \( q(r < r_{\text{q}} = 1) = q_0 \), a constant with a value close to unity. Hence we expect the 3/2 moments peaking factor, \( \langle T^{3/2} \rangle_{r=0.2}/\langle T^{3/2} \rangle_{r=1} \) to scale as \( q_a \). Since the YAG channel closest to the magnetic axis lies typically on the 14\% flux radius, \( T_0 \) is essentially an extrapolated quantity. We present, therefore, scalings for the left-hand side of Eq. (2) at both \( r = 0 \) and \( r = 0.2 \), with the latter radius satisfying the double requirement of being safely inside the inversion radius for all profiles as well as lying in the data region. For the main database we obtain

\[
\frac{T^{3/2}_{0.2}}{\langle T^{3/2} \rangle_{r=1}} = 1.02(\pm 0.06) n_e^{-0.03 \pm 0.02} I_p^{-0.11 \pm 0.09} q_a^{0.13 \pm 0.06}
\]

RMSE = 0.06, \( R^2 = 0.874 \)

while the secondary database yields

\[
\frac{T^{3/2}_{0.2}}{\langle T^{3/2} \rangle_{r=1}} = 0.99(\pm 0.04) n_e^{-0.03 \pm 0.15} I_p^{-0.07 \pm 0.06} q_a^{0.13 \pm 0.04}
\]

RMSE = 0.04, \( R^2 = 0.931 \)

In all four cases, the \( q_a \) exponent is unity to within two standard deviations. Similarly, the constant factors are within two standard deviations of unity (with the marginal exception of the second regression: 0.87 \( \leq \) const \( \leq \) 0.99). These results are consistent with, but more specific than, the inequality

\[
q_a^{2/3} \leq \frac{T(0)}{T} \leq q_a
\]

developed by Waltz et al. [13] using classical resistivity plus sawtoothing. Since we use \( \langle T^{3/2} \rangle \) rather than \( T^{3/2} \), the upper and lower bounds of the inequality coincide in our case. To enable a comparison with previous
work, we have regressed the more traditional peaking factor $T(0)/T$ to obtain, for the combined database, the scaling: $T(0)/T = 1.00(\pm0.028) q_a^{0.73\pm0.030}$, whereas our normalization gives (also for the combined database) $T(0)/T^{(2/3)^2} = 0.99(\pm0.028) q_a^{0.67\pm0.029}$. We believe that these results constitute strong evidence that the resistive equilibrium ansatz is sufficient to explain quantitatively the scaling of the peaking factor for ASDEX Ohmic temperature profiles.

5. PROFILE SHAPE ANALYSIS

The next two sections are devoted to a detailed description of profile parameterization techniques and the ensuing methods of data analysis, tailored to our case. For a more theoretical background, the reader is referred to Ref. [4]. The experimental results are presented in Section 7.

We assume that the logarithm of the temperature satisfies $\ln T(r, q_a, I_p, n_e) = \mu(r, X) + \epsilon$, where $\epsilon$ is a random error. The deterministic part, $\mu(r, X)$, is represented as a spline with reasonably high resolution in the radial direction, and a simple (polynomial-type) dependence on the plasma parameters. The coefficients of the representation are determined by fitting all profiles simultaneously in a weighted least squares regression.

The temperature and density measurements are obtained using the ASDEX 16-channel YAG Thomson scattering diagnostic [6] with a sampling rate of 60 Hz. This system consists of 16 spatial channels located in the vertical plane at $R = 1.63$ m. They are spaced at approximately 4 cm intervals from $Z = 0.200$ m to $Z = -0.394$ m. We did not use the 16th channel, which lies very close to or on the separatrix, since the measurement has very poor signal to noise ratio 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(a) and V(a) for the main and secondary databases respectively.

5.1. Radial representations

We discuss here continuous radial representations of 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 $m = 12$ consecutive measurements taken at 17 ms 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 $T_{ij}(r_f)$, where $i = 1, \ldots, m$ labels the uncompressed time point, $j = 1, \ldots, n$ is the compressed profile index and $f = 1, \ldots, p$ denotes the radial channel number. We make a preliminary transformation of the physical measurement locations $r_f$ to the corresponding flux surface radii $r_f$.

Continuous representations have the following characteristics: (a) A large number of dependent variables, represented by point data, is replaced by a small number of coefficients which nevertheless will be sufficient to represent all significant features of the profiles. (b) Profiles measured at two different sets of radial locations may be compared. This is relevant, for example, 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 diffuse equilibrium.

Instead of fitting the profile itself, we choose to fit its natural logarithm $Y$. Minimizing the residual of the logarithm of the plasma profile corresponds to minimizing the relative error rather than the absolute error. Preliminary comparisons with low order spline or polynomial fits to the actual profiles have revealed that the logarithmic fit tends to have smaller residual errors not only 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 regression parameters (either spline knots or higher order polynomial terms) are added. Logarithmic fits have several other advantages: (1) The predicted profile can never be negative. (2) Well known power law type scalings reduce to linear models. (3) If the noise level is proportional to the absolute value of the measurement (an admittedly idealized situation), then, on the logarithmic scale, unweighted least squares may be used.

Spline representations, which we employ here for profile parameterization, give flexibility in choosing between local resolution and compact global representation. The profile parameterizations presented here are based on twice continuously differentiable splines with a selectable number of knots, $\nu$. The profile is forced to be parabolic inside the first knot, i.e. the region enclosing the magnetic axis. The radius is decomposed into $\nu + 1$
regions with knots at \( r_1, r_2, \ldots, r_v \). Such a profile can be parameterized explicitly by:

\[
\mu(r) = \begin{cases} 
\mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 & \text{for } 0 \leq r \leq r_1 \text{ (inner region)} \\
\mu(r_1) + \mu''(0)(r_2^2 - r_1^2)/2 + c_1(r - r_1)^3 & \text{for } r_1 < r \leq r_2 \text{ (region 1)} \\
\mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 + c_1(r - r_1)^3 + c_2(r - r_2)^3 & \text{for } r_2 < r \leq r_3 \text{ (region 2)} \\
\mu(r_1) + \mu''(0)(r^2 - r_1^2)/2 + c_1(r - r_1)^3 + \cdots + c_v(r - r_v)^3 & \text{for } r_v < r \leq 1 \text{ (outer region)} 
\end{cases}
\tag{4}
\]

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

\[
\mu(r) = \sum_{k=0}^v \alpha_k H(r - r_k) \varphi_\nu(r) \tag{5}
\]

where \( \nu \) is the number of knots used, \( \alpha_k \) represent the spline parameters: \( \mu(r_1), \mu''(0), c_1, c_2, \ldots, c_\nu \); \( \varphi_\nu(r) \) are the polynomials: \( \varphi_0(r) = 1 \), \( \varphi_0(r) = (r^2 - r_1^2)/2 \), \( \varphi_1(r) = (r - r_1)^3, \ldots, \varphi_\nu(r) = (r - r_\nu)^3 \), and

\[
H(r - r_k) = \begin{cases} 
0 & \text{for } r < r_k \\
1 & \text{for } r \geq r_k 
\end{cases}
\tag{6}
\]

is the Heaviside unit step function \( (r_k = 0, 0, r_1, r_2, \ldots, r_v) \).

Besides the parabolic restriction near the axis, a 'natural' spline boundary condition, \( \mu''(1) = 0 \), was imposed in practice. The spline fits were carried out using the SAS REG procedure [14], which contains convenient possibilities for restricted regression. A similar spline model for transport analysis of individual profiles has been used in Ref. [15].

5.2. Parametric dependences 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 parameters. Since our database is not expected to contain sharp transitions in behaviour in parameter space, a low order parametric representation is expected to be adequate. Therefore, we approximate the smooth parametric dependences of the profile shape by linear or possibly quadratic polynomials in the logarithms of the bulk variables.

Let \( \vec{x} = (x_1, x_2, x_3) \) be the vector consisting of the bulk plasma variables \( q_a, I_p, n_e \). We define the linear basis functions \( g_0(\vec{x}) = 1 \) (intercept), \( g_j(\vec{x}) = \ln x_j/x_j^* \), \( j = 1, 2, 3 \); 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 normalizing values \( q_a^* = 2.5, I_p^* = 0.4 \text{ MA} \) and \( n_e^* = 0.4 \times 10^{20} \text{ m}^{-3} \), although these do not constitute a typical parameter set for the secondary database. By normalizing the bulk variables to \( x_j^* \), the value of the response variable at the intercept in the regression becomes the predicted value at \( \vec{x} = \vec{x}^* \).

The full profile representation can be written as

\[
Y(r, \vec{x}) = \sum_{k,j} \alpha_{k,j} g_j(\vec{x}) H(r - r_k) \varphi_\nu(r) + \epsilon = \mu(r, \vec{x}) + \epsilon \tag{7}
\]

where \( \epsilon \) is an error term whose structure is discussed in Section 5.3. The basis functions are as defined before. Note that higher order terms can be included, if necessary, by extending the set of possible basis functions: \( g_{j,l}(\vec{x}) = \ln x_j/x_j^* \ln x_l/x_l^* \), etc.

5.3. Error structures

Here, we present and motivate the splitting up of the total regression error into contributions that are attributable to different physical sources. An efficient analysis should take into account the particular features of such an error structure. In Section 5.4, we discuss the methods of estimation we applied in our case.

We distinguish between several categories of random profile variations. We use the term 'internal' variations to denote fluctuations 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, for example, 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 Ref. [16], Appendix B, this hierarchy of temporal scales for plasma variation generates a compound error structure which can be treated statistically. We give a simplified discussion here. For convenience, we assume normally distributed errors, although this assumption can be relaxed in most of the discussion. Since the profiles in our database are already averaged over twelve consecutive time samples (see Section 1), they no longer possess the same variance as the original uncompressed observations. For time averaged data points,

$$\bar{Y}(r, x) = \frac{1}{m} \sum_{i=1}^{m} Y_i(r, x)$$

the total unexplained variance can be decomposed into:

$$\sigma_{\text{tot}}^2 = \frac{\sigma_{\text{int}}^2}{m} + \sigma_{\text{dis}}^2$$

where $\sigma_{\text{int}}^2$ is the variance of the internal or 'within-discharge' fluctuations of the uncompressed profiles, $m$ is the number of time points in the compressed profile (in our case $m = 12$) and $\sigma_{\text{dis}}^2$ denotes the variance due to discharge-to-discharge variations.

TABLE III(a). TEMPERATURE DESCRIPTIVE STATISTICS FOR THE HIGH $n$, LOW $q$ DATABASE (absolute scale)

| Channel | $<T_e>$ (keV) | Spread | $<T_e/T_e(r = 0.2)>$ | Spread |
|---------|---------------|--------|----------------------|--------|
| 1       | 0.515         | 0.096  | 0.804                | 0.104  |
| 2       | 0.432         | 0.089  | 0.825                | 0.085  |
| 3       | 0.345         | 0.090  | 0.918                | 0.041  |
| 4       | 0.247         | 0.100  | 0.977                | 0.023  |
| 5       | 0.180         | 0.104  | 1.000                | 0.019  |
| 6       | 0.138         | 0.111  | 1.034                | 0.028  |
| 7       | 0.164         | 0.108  | 1.015                | 0.020  |
| 8       | 0.225         | 0.102  | 0.988                | 0.019  |
| 9       | 0.309         | 0.090  | 0.933                | 0.046  |
| 10      | 0.407         | 0.100  | 0.897                | 0.095  |
| 11      | 0.489         | 0.103  | 0.829                | 0.114  |
| 12      | 0.585         | 0.085  | 0.732                | 0.102  |
| 13      | 0.682         | 0.074  | 0.589                | 0.089  |
| 14      | 0.784         | 0.046  | 0.393                | 0.063  |
| 15      | 0.885         | 0.028  | 0.241                | 0.038  |

TABLE III(b). LOGARITHMIC SPREAD, NOISE ESTIMATES AND SPLINE REGRESSION ERRORS (relative scale)

| Channel | Logarithmic spread $\sigma$ | Within-discharge $\sigma$ | Unnormalized channel regression $\sigma$ | Normalized channel regression $\sigma$ | Normalized profile spline regression $\sigma$ |
|---------|-----------------------------|---------------------------|----------------------------------------|---------------------------------------|---------------------------------------------|
| 1       | 0.130                       | 0.026                     | 0.044                                  | 0.027                                 | 0.027                                       |
| 2       | 0.104                       | 0.018                     | 0.039                                  | 0.023                                 | 0.041                                       |
| 3       | 0.046                       | 0.018                     | 0.041                                  | 0.023                                 | 0.029                                       |
| 4       | 0.024                       | 0.021                     | 0.040                                  | 0.020                                 | 0.019                                       |
| 5       | 0.019                       | 0.022                     | 0.040                                  | 0.018                                 | 0.019                                       |
| 6       | 0.027                       | 0.034                     | 0.048                                  | 0.027                                 | 0.030                                       |
| 7       | 0.020                       | 0.022                     | 0.041                                  | 0.020                                 | 0.019                                       |
| 8       | 0.019                       | 0.017                     | 0.038                                  | 0.016                                 | 0.016                                       |
| 9       | 0.051                       | 0.015                     | 0.036                                  | 0.022                                 | 0.023                                       |
| 10      | 0.107                       | 0.018                     | 0.044                                  | 0.025                                 | 0.039                                       |
| 11      | 0.137                       | 0.024                     | 0.049                                  | 0.032                                 | 0.035                                       |
| 12      | 0.139                       | 0.035                     | 0.054                                  | 0.040                                 | 0.041                                       |
| 13      | 0.148                       | 0.035                     | 0.048                                  | 0.029                                 | 0.029                                       |
| 14      | 0.155                       | 0.031                     | 0.044                                  | 0.031                                 | 0.032                                       |
| 15      | 0.154                       | 0.041                     | 0.057                                  | 0.052                                 | 0.051                                       |
| RMS for 15 channels: | 0.027 | 0.045 | 0.028 | 0.030 |
To estimate the within-discharge variance $\sigma_{\text{int}}^2$, we have analysed the original uncompressed data and, for each channel, calculated the empirical variance for each 12-point set separately. We have estimated $\sigma_{\text{int}}$ by regressing the set of $n$ (compressed) data points for each individual measurement channel against the bulk variables and noting the unexplained variance. The difference, as given by the third term in Eq. (8), is an estimate of the discharge-to-discharge variance.

In Tables III(b) to VI(b), columns 3 and 4, one can see the estimates of $\sigma_{\text{int}}$ (scaled for compressed profiles) and $\sigma_{\text{tot}}$ for each of the 15 channels. It is clear from the large channel-to-channel variation displayed in these tables that it would not be justified to make model assumptions that $\sigma_{\text{int}}$ and/or $\sigma_{\text{tot}}$ are the same for all channels.

5.4. Coefficient estimation

We determine the spline coefficients, including parametric dependences, by fitting all profiles simultaneously in a weighted least squares regression where the weights, $W(r_f)$, depend on the channel location.

We wish to determine that vector of the coefficients $\alpha$ which minimizes

$$\sum_{j,t} [Y_{j,\text{obs}}(r_f) - Y_{j,\text{fit}}(r_f, \alpha)]^2 W(r_f)$$

where $W(r_f), f = 1, \ldots, 15$ are appropriately chosen weights for each of the 15 YAG channels. We have investigated two approaches for determining $W(r_f)$.

### TABLE IV(a). DENSITY DESCRIPTIVE STATISTICS FOR THE HIGH $n$, LOW $q$ DATABASE (absolute scale)

| Channel | $\langle n_e \rangle$ ($10^{20}$ m$^{-3}$) | Spread $\langle n_e/n_e(r = 0.2) \rangle$ | Spread $\langle n_e/n_e(r = 0.2) \rangle$ |
|---------|--------------------------------------|----------------------------------------|----------------------------------------|
| 1       | 0.515 0.523                        | 0.149 0.926                           | 0.055 0.038                           |
| 2       | 0.432 0.576                        | 0.168 1.018                           | 0.023 0.038                           |
| 3       | 0.345 0.581                        | 0.166 1.026                           | 0.021 0.023                           |
| 4       | 0.247 0.586                        | 0.166 1.037                           | 0.015 0.015                           |
| 5       | 0.180 0.579                        | 0.160 1.026                           | 0.015 0.015                           |
| 6       | 0.138 0.372                        | 0.160 1.013                           | 0.015 0.015                           |
| 7       | 0.164 0.536                        | 0.155 0.985                           | 0.017 0.007                           |
| 8       | 0.225 0.539                        | 0.153 0.954                           | 0.014 0.014                           |
| 9       | 0.309 0.542                        | 0.157 0.958                           | 0.018 0.018                           |
| 10      | 0.407 0.534                        | 0.152 0.944                           | 0.033 0.033                           |
| 11      | 0.489 0.501                        | 0.147 0.887                           | 0.053 0.053                           |
| 12      | 0.585 0.462                        | 0.146 0.813                           | 0.064 0.064                           |
| 13      | 0.682 0.419                        | 0.135 0.736                           | 0.058 0.058                           |
| 14      | 0.784 0.327                        | 0.107 0.573                           | 0.050 0.050                           |
| 15      | 0.885 0.272                        | 0.092 0.479                           | 0.057 0.057                           |

| Channel | Logarithmic spread $\sigma$ | Within-discharge $\sigma$ | Unnormalized channel regression $\sigma$ | Normalized channel regression $\sigma$ | Normalized profile spline regression $\sigma$ |
|---------|-----------------------------|---------------------------|----------------------------------|---------------------------------|----------------------------------|
| 1       | 0.060 0.015                 | 0.020 0.021               | 0.023                            | 0.026                            |
| 2       | 0.038 0.012                 | 0.021 0.018               | 0.026                            | 0.026                            |
| 3       | 0.023 0.014                 | 0.019 0.015               | 0.017                            | 0.017                            |
| 4       | 0.014 0.011                 | 0.018 0.014               | 0.016                            | 0.016                            |
| 5       | 0.015 0.012                 | 0.020 0.014               | 0.020                            | 0.020                            |
| 6       | 0.015 0.012                 | 0.018 0.014               | 0.014                            | 0.014                            |
| 7       | 0.018 0.011                 | 0.022 0.018               | 0.023                            | 0.023                            |
| 8       | 0.015 0.011                 | 0.018 0.014               | 0.017                            | 0.017                            |
| 9       | 0.019 0.010                 | 0.022 0.017               | 0.017                            | 0.017                            |
| 10      | 0.035 0.010                 | 0.021 0.015               | 0.019                            | 0.019                            |
| 11      | 0.060 0.015                 | 0.028 0.025               | 0.028                            | 0.028                            |
| 12      | 0.079 0.033                 | 0.045 0.047               | 0.047                            | 0.047                            |
| 13      | 0.079 0.014                 | 0.025 0.027               | 0.034                            | 0.034                            |
| 14      | 0.087 0.021                 | 0.027 0.031               | 0.041                            | 0.041                            |
| 15      | 0.116 0.048                 | 0.071 0.060               | 0.073                            | 0.073                            |
| RMS for 15 channels: | 0.020 | 0.030 | 0.027 | 0.032 |
In the first method, we rely on the total unexplained variance for each channel (as discussed in Section 5.3) as a measure of the channel weighting: \( W(r_i) = \sigma^2_{i_{total}}(r_i) \).

A second approach to the selection of the regression weights is the iterative estimation of the residual variance at each channel. At the k-th iteration we have:

\[
\sigma^2_{(k+1)}(r_i) = \frac{1}{n} \sum_{j=1}^{n} (Y_{j\text{obs}}(r_i) - Y_{j\text{fit}(k)}(r_i, \delta_k) - \delta Y_k(r_i))^2
\]

where \( \delta Y(r_i) \) is a possible systematic bias in fitting the \( f \)-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 parameterized profile may consistently fail to match the observed data. For the first iteration, the variances \( \sigma^2_{i_{total}}(r_i) \) are initialized to unity (equal weights). The iteration is terminated after the third iteration. We then have \( W(r_i) = \sigma^2_{i_{total}}(r_i) \). If the regression model and the assumed error structure is correct, this is likely to be a consistent and efficient estimate (see, e.g. Ref. [17]). Nevertheless, we regard \( \sigma^2_{i_{total}}(r_i) \) as a more robust estimate than \( \sigma^2_{i_{last iter}}(r_i) \) since it depends only on the fit by 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. Accordingly, we preferred to use \( W(r_i) = \sigma^2_{i_{total}}(r_i) \). The dominant effect of this reweighting is to decrease the influence of the channels near the plasma boundary where the relative error is largest.

### TABLE V(a). TEMPERATURE DESCRIPTIVE STATISTICS FOR THE MODERATE \( n \), MODERATE \( q \) DATABASES

| Channel (Radius) | \( (T_e) \) (keV) | Spread | \( (T_e/T_e(r = 0.2)) \) | Spread |
|------------------|------------------|--------|------------------------|--------|
| 1                | 0.516            | 0.754  | 0.135                  | 0.711  | 0.109 |
| 2                | 0.434            | 0.788  | 0.146                  | 0.741  | 0.094 |
| 3                | 0.347            | 0.970  | 0.151                  | 0.914  | 0.095 |
| 4                | 0.252            | 1.074  | 0.171                  | 1.007  | 0.051 |
| 5                | 0.190            | 1.066  | 0.153                  | 1.001  | 0.022 |
| 6                | 0.154            | 1.108  | 0.154                  | 1.042  | 0.044 |
| 7                | 0.180            | 1.074  | 0.168                  | 1.007  | 0.026 |
| 8                | 0.238            | 1.034  | 0.155                  | 0.970  | 0.027 |
| 9                | 0.321            | 0.960  | 0.145                  | 0.905  | 0.092 |
| 10               | 0.417            | 0.819  | 0.160                  | 0.770  | 0.103 |
| 11               | 0.499            | 0.664  | 0.139                  | 0.624  | 0.092 |
| 12               | 0.594            | 0.555  | 0.122                  | 0.521  | 0.083 |
| 13               | 0.691            | 0.470  | 0.098                  | 0.444  | 0.082 |
| 14               | 0.793            | 0.285  | 0.060                  | 0.269  | 0.049 |
| 15               | 0.894            | 0.181  | 0.033                  | 0.171  | 0.025 |

### TABLE V(b). LOGARITHMIC SPREAD, NOISE ESTIMATES AND SPLINE REGRESSION ERRORS

| Channel | Logarithmic spread | Within-discharge | Unnormalized channel regression | Normalized channel regression | Normalized profile spline regression |
|---------|--------------------|------------------|---------------------------------|------------------------------|-----------------------------------|
|         | \( \sigma \)       | \( \sigma \)     | \( \sigma \)                     | \( \sigma \)                   | \( \sigma \)                        |
| 1       | 0.152              | 0.025            | 0.054                           | 0.027                        | 0.100                             |
| 2       | 0.130              | 0.020            | 0.044                           | 0.023                        | 0.026                             |
| 3       | 0.105              | 0.018            | 0.054                           | 0.032                        | 0.065                             |
| 4       | 0.050              | 0.021            | 0.048                           | 0.027                        | 0.048                             |
| 5       | 0.021              | 0.021            | 0.047                           | 0.013                        | 0.019                             |
| 6       | 0.042              | 0.031            | 0.053                           | 0.032                        | 0.058                             |
| 7       | 0.026              | 0.021            | 0.045                           | 0.018                        | 0.026                             |
| 8       | 0.027              | 0.019            | 0.049                           | 0.024                        | 0.033                             |
| 9       | 0.103              | 0.017            | 0.046                           | 0.029                        | 0.043                             |
| 10      | 0.137              | 0.023            | 0.054                           | 0.042                        | 0.044                             |
| 11      | 0.151              | 0.032            | 0.043                           | 0.036                        | 0.094                             |
| 12      | 0.164              | 0.042            | 0.064                           | 0.059                        | 0.123                             |
| 13      | 0.186              | 0.024            | 0.044                           | 0.045                        | 0.052                             |
| 14      | 0.187              | 0.023            | 0.047                           | 0.050                        | 0.061                             |
| 15      | 0.151              | 0.037            | 0.079                           | 0.076                        | 0.117                             |
| RMS for 15 channels: | 0.026              | 0.052            | 0.039                           | 0.039                        | 0.069                             |

NUCLEAR FUSION, Vol.31, No.9 (1991)
5.5. Criteria for additional free parameters

In the context of profile parameterization, some relevant statistical tests for the significance of including additional variables are discussed in Ref. [4]. In the case of independent errors, these criteria are given by the F test [7] and the Mallows $C_p$ statistic [18]. The 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 look at the change in the $C_p$ statistic or apply the F test.

It should be noted that, in practice, these statistical approaches may either underestimate or overestimate the significance of the additional variables since the correlations in the errors are neglected. In addition, these tests neglect systematic errors and assume that the ideal data, without measurement errors, can be exactly described by the regression equation under consideration. Thus, many spurious dependences may be included and real dependences missed by unthoughtful or automatic use of these methods.

6. PROFILE PARAMETERIZATION TECHNIQUES

In this section, we discuss a number of practical aspects encountered during our investigation, which are expected to be useful in any profile analysis. Logarithmic representations are employed throughout, for the reasons outlined in Section 5.1.

TABLE VI(a). DENSITY DESCRIPTIVE STATISTICS FOR MODERATE $n$, MODERATE $q$ DATABASE (absolute scale)

| Channel | (Radius) $\langle n \rangle$ ($10^{20}$ m$^{-3}$) | Spread $\langle n / n_0 \rangle (r = 0.2)$ |
|---------|--------------------------------------------------|------------------------------------------|
| 1       | 0.516                                            | 0.813                                    |
| 2       | 0.434                                            | 0.904                                    |
| 3       | 0.347                                            | 0.955                                    |
| 4       | 0.252                                            | 0.959                                    |
| 5       | 0.190                                            | 1.018                                    |
| 6       | 0.154                                            | 1.021                                    |
| 7       | 0.180                                            | 1.021                                    |
| 8       | 0.238                                            | 1.001                                    |
| 9       | 0.321                                            | 0.886                                    |
| 10      | 0.417                                            | 0.933                                    |
| 11      | 0.499                                            | 0.804                                    |
| 12      | 0.594                                            | 0.690                                    |
| 13      | 0.691                                            | 0.551                                    |
| 14      | 0.793                                            | 0.444                                    |
| 15      | 0.894                                            | 0.311                                    |

TABLE VI(b). LOGARITHMIC SPREAD, NOISE ESTIMATES AND SPLINE REGRESSION ERRORS (relative scale)

| Channel | Logarithmic spread $\sigma$ | Within-discharge $\sigma$ | Unnormalized channel regression $\sigma$ | Normalized channel regression $\sigma$ | Normalized profile spline regression $\sigma$ |
|---------|-----------------------------|---------------------------|------------------------------------------|----------------------------------------|---------------------------------------------|
| 1       | 0.100                       | 0.016                     | 0.040                                    | 0.036                                  | 0.073                                       |
| 2       | 0.069                       | 0.014                     | 0.028                                    | 0.025                                  | 0.039                                       |
| 3       | 0.051                       | 0.012                     | 0.032                                    | 0.027                                  | 0.055                                       |
| 4       | 0.049                       | 0.013                     | 0.021                                    | 0.013                                  | 0.041                                       |
| 5       | 0.023                       | 0.012                     | 0.016                                    | 0.012                                  | 0.034                                       |
| 6       | 0.028                       | 0.014                     | 0.021                                    | 0.015                                  | 0.033                                       |
| 7       | 0.035                       | 0.012                     | 0.019                                    | 0.019                                  | 0.042                                       |
| 8       | 0.048                       | 0.012                     | 0.021                                    | 0.019                                  | 0.063                                       |
| 9       | 0.054                       | 0.012                     | 0.017                                    | 0.015                                  | 0.054                                       |
| 10      | 0.095                       | 0.013                     | 0.019                                    | 0.019                                  | 0.054                                       |
| 11      | 0.093                       | 0.020                     | 0.029                                    | 0.028                                  | 0.044                                       |
| 12      | 0.117                       | 0.040                     | 0.045                                    | 0.041                                  | 0.048                                       |
| 13      | 0.119                       | 0.012                     | 0.025                                    | 0.030                                  | 0.048                                       |
| 14      | 0.119                       | 0.018                     | 0.041                                    | 0.045                                  | 0.091                                       |
| 15      | 0.213                       | 0.045                     | 0.101                                    | 0.108                                  | 0.151                                       |

RMS for 15 channels: 0.020 0.038 0.028 0.065
6.1. Polynomial models

Initial efforts concentrated on fitting polynomial representations of the form

\[ T(r) = T_0 \exp(a r^2 + b r^4 + c r^6) \]

This model was successful in reproducing the general properties of the ASDEX profiles, but not detailed features. Sharp gradients and local flattenings (due, perhaps, to magnetic islands) were poorly modelled. A disadvantage of the above model, exacerbated by the addition of 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 co-ordinate, we turned to spline representations.

6.2. Once and twice continuously differentiable cubic splines

Following Ref. [19], 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 clearly better in parameterizing 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 of the Hermitian spline emerged, however. By allowing discontinuous second derivatives at the knots, continuous transitions in plasma behaviour were modelled as sharp jumps across knot boundaries. 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 continuously differentiable spline model, Eq. (4).

6.3. 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. We decided on a set of five knots at the following locations: \( r_{\text{knot}} = 0.2, 0.3, 0.4, 0.5, 0.65 \).

6.4. Normalization

The innermost channel is typically located at \( r = 0.14 \) and the outermost channel at \( r = 0.89 \) (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 (this is already enforced in Eq. (4)). The so-called natural boundary conditions, \( \mu'(1) = 0 \), was applied at \( r = 1 \). We investigated the regression fits with the natural boundary conditions applied to (a) all spline coefficients and (b) bulk parameter dependent spline coefficients only. Case (a) resulted in considerably higher fitting errors than case (b) for, in particular, the outer channels of the main database density profile fit. However, we concluded that the improvement in case (b) was achieved at the expense of overfitting of the outer channels and we present here only the results of the fully applied boundary conditions. In a preliminary version of this work [16], the results of case (b) are presented. Thus our five-knot set yields a model of seven spline coefficients, with one boundary constraint per bulk parameter basis function used in the fit. The profile parameterizations presented later were carried out using only the linear basis functions \( g_j(\bar{x}) \), \( j = 0, 1, 2, 3 \). Some quadratic and cross terms were very significant in preliminary regressions involving both databases simultaneously. For reasons given in Section 1, 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 \( q_n, I_p \) and \( n_e \), we have a regression model with a total of (intercept + three bulk parameters) \( \times \) (seven radial coefficients — one boundary condition) = 24 degrees of freedom to fit (e.g. for the main database) \( 15 \times 105 = 1575 \) individual temperature (or density) data. With the foregoing boundary conditions, this spline representation tended to be stable in extrapolating the profile behaviour into regions where there were no measurement channels.

The goal of our profile analysis is to determine the dependence of the profile shapes, i.e. the functions

\[ L_{\gamma_e}(r) = \frac{1}{T_e(r)} \frac{d}{dr} T_e(r) \]
on the plasma parameters. This brings up the problem of fitting the profile magnitude. We found that when we made a simultaneous fit of the magnitude and the shape of the profile, the residual sum of squares was dominated by the uncertainty in fitting the profile magnitude. Therefore, a model was used which provided a free parameter to fit each individual profile magnitude. This normalization procedure, justified because the profile magnitude scales out of the profile shape definition above, causes very significant reductions in the residual sum of squares of the shape regressions.

Originally, we normalized each profile by its line average, calculated from the spline fit. This greatly reduced, but did not minimize, the residual error in the profile parameterization since the line average is itself a function of the profile shape. Instead, we estimated the profile magnitude parameters, using the SAS procedure GLM [14], by treating the profile index as an indicator variable. This yielded as normalizing factor the radially independent term in our spline representation, i.e. \( \mu(\tau_1) \), the profile value at the first knot which is sited at \( \tau = 0.2 \). Normalization has the effect of reducing by one the number of degrees of freedom for each individual spline. Hence, the total number of degrees for the five-knot spline, with coefficients dependent on three bulk parameters (as described in Section 6.3), is reduced from 24 to 20.

6.5. Operating period indicator variable

In the course of determining \( \delta^2_{\text{ext}}(\tau_1) \) for each YAG channel, plots of residuals versus shot number revealed that the secondary database residuals, whose discharges spanned a period of over six months in contrast to the one week span of the discharges of the main database, fell into four distinct groupings which we ascribe to four distinct experimental operating periods (see Section 1). This four-cluster formation was observed for all 15 channels, although the pattern formed by the clusters differed for each channel. 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 \( \delta^2_{\text{ext}} \). These indicator variables were not, however, included in the profile shape regressions. To do so would have required an additional 60 independent variables (four for each channel) which, in our judgement, would have led to overfitting of the profiles.

6.6. Examination of outliers

Apart from the operating period effects mentioned above, plots of raw data 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 Studentized residuals. A Studentized residual is the difference between the observed and fitted value, normalized to the RMSE. For normally distributed errors, these residuals have approximately a standard normal distribution. If we consider a single Studentized residual, the probability that it lies outside \( \pm c \) is \( \epsilon \), where \( \epsilon = 1 - (2\pi)^{-n/2} \int e^{-x^2/2} \, dx \). Considering now \( n \) uncorrelated residuals together, we find that the probability of all \( n \) residuals lying inside \( \pm c \) is \((1 - \epsilon)^n\). Hence the probability of at least one among \( n \) Studentized 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 Studentized residual exceeds \( \pm c \) for a suitably small \( \beta \) (using \( \epsilon = 1 - (1 - \beta)^{1/n} = \beta/n \) for \( \beta \ll 1 \), we invert the probability integral to determine \( c_\beta \)). We chose \( \beta = 1\% \) which, for \( n = 105 \) and \( n = 38 \), yields \( c = 3.90 \) and \( c = 3.64 \) for the main and secondary databases, respectively. Using this criterion, we identified six suspect outliers from the main database and 18 from the secondary database, amounting 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 marked as bad data. One discharge accounted for the majority of the suspect data in the secondary 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 main 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, (2.73 T) in the database. This highlights the need to examine all suspect outliers individually, since the influential position of these 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 finally deemed to be faulty (a total of four points affecting three profiles in the main database; twelve points affecting nine profiles in the secondary database) were deleted from the regression.

6.7. 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 $\approx 7\%$). The asymmetry was nearly uniform on all five pairs of measurement channels and was independent of the plasma parameters. No significant asymmetry was found in the temperature profiles.

A likely explanation for this asymmetry was found after a routine inspection in 1990. A slight misalignment of the laser beam direction against the entrance slits of the detectors may have decreased the sensitivity of the radial channels below the equatorial plane. Another possible cause is a spatially non-uniform distribution of impurities on the diagnostic window. Assuming that for the range of scattered laser light detected by the system ($0.8 \mu m \leq \lambda_{scant} \leq 1.06 \mu m$) the impurity accumulation causes a spectrally uniform reduction in transmission, it is found that the temperature, which is calculated from the ratio of the scattering signals of two spectral channels at the same spatial location, is insensitive to such asymmetries and only the density is affected.

To estimate and correct for this density profile asymmetry, we expanded the set of bulk parameter dependent, radial basis functions to include a single asymmetry indicator variable which takes the value of $+1$ for the channels above the midplane and the value of $-1$ for the channels below the midplane. In the final model, the results of which are presented in Section 7, the degrees of freedom are accordingly increased by one, to give a total of 21 for the density profile regressions (the temperature model remains at 20 degrees of freedom).

7. EXPERIMENTAL RESULTS

7.1. Temperature profile

Here we present a detailed graphical representation of our results. The dominant shape dependence is a peaking of the profile inside $r = 0.6$ with increasing $q_a$. The profile shape shows very little dependence on any bulk parameters outside $r = 0.6$. Thus, our result is similar to that of Fredrickson et al. [20] (profile shape invariant for $r \geq 0.4$) and that of Murmann et al. [21] (profile shape invariant for the region outside the influence of the $q = 1$ surface), though our more elaborate model enables us to examine more sensitively the extent of the $q_a$ dependent region.

Figure 5 and 6 display, as a function of radius, the reference profiles and the parametric dependence of the negative inverse fall-off length (IFOL), normalized to the inverse minor radius. The dashed curves indicate local 95% confidence bands. The corresponding global 95% confidence bands are, in this case, approximately 1.5 times wider (see Ref. [4] for a discussion of local and global confidence bands). By the term reference profile we mean the evaluation of the parameterized spline fit at the representative set of parameter values $\bar{x} = \bar{x}$ as discussed in Section 5.2. Since the vector of basis functions $g(\bar{x}) = (1, \ln x/y, \ln x/y^2, \ln x/y^3)$ reduces in this case to $g(\bar{x}) = (1, 0, 0, 0)$, the reference profile is just that described by the set of intercept, i.e. bulk parameter independent, spline coefficients.

The parametric radial dependences of the IFOL are obtained by differentiating the parameterized spline representation for the normalized profiles both radially and with respect to the basis function of the desired parameter. Figure 5(b), for example, shows the radial behaviour of the $q_a$ dependence which, using Eq. (7) (with $g_i(\bar{x}) = \ln q_a - \ln q_a^*$), is given by

$$\frac{\partial (-L_i^{-1})}{\partial \ln q_a} = -\frac{\partial^2 \mu(r, \bar{x})}{\partial g_i \partial r}$$

$$= -\frac{\partial}{\partial r} \sum_k \alpha_k H(r - r_k) \phi_k(r) \quad (11)$$

We see that at $r = 0.35$, where the profile shape is most sensitive to $q_a$, a unit change in $\ln q_a$ causes a change of $-5$ in the negative IFOL, corresponding to 12.5 m$^{-1}$ for an ASDEX minor radius of 0.4 m. Reference values and parametric dependences of discrete IFOLs are also displayed. These are calculated by differencing the measurement values of pairs of neighbouring channels:

$$L_i^{-1} = \frac{T_{i+1} - T_i}{0.5(T_{i+1} + T_i)(r_{i+1} - r_i)}$$
FIG. 5. Reference negative IFOL profiles for the main (high density) database $T_e$, and parametric dependence profiles for $q_a$, $I_p$, and $n_e$ with local 95% confidence bands. Reference values for discrete (two-channel) IFOLs and the corresponding parametric dependences are also plotted as individual data points. The central solid curve in (a) is the predicted negative IFOL at the reference parameter values $q_a = 2.5$, $I_p = 0.4$ MA, $n_e = 0.4 \times 10^{20}$ m$^{-3}$. The local 95% confidence bands are plotted as dashed curves. The flattest and steepest predicted negative IFOL profiles in the database (i.e. those for minimum ($q_a = 1.9$) and maximum ($q_a = 3.4$) values of $q_a$) are plotted as solid curves without confidence bands. The individual data points are predictions for discretely calculated IFOLs (Section 7.1), again evaluated at the reference parameter values and accompanied by 95% confidence bands. (b) Parametric dependence of $L_f^T$ on $q_a$. The solid curve is the radial profile of the change (in units of 1/minor radius) in the negative IFOL per unit change in $\ln q_a$ (see Eq. (11)). The individual data points give similar information on the discrete IFOLs. (c) Parametric dependence of $L_f^T$ on $I_p$ and $n_e$. (d) Parametric dependence of $L_f^T$ on $n_e$. 
FIG. 6. Reference negative IPOL profiles for the secondary (moderate density) database and parametric dependence profiles for \( n_e \), and \( \nu_e \).
and regressing these approximate IFOLs on \( q_a \), \( I_p \) and \( \bar{n}_e \). Such a point value has the advantage that it is more local than the continuous function represented by Eq. (11), 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 negative IFOL profiles for minimum, reference and maximum \( q_a \), displayed in Fig. 5(a), show that the temperature profile shape for the main database is remarkably invariant outside \( r = 0.6 \). This behaviour is broadly similar for the profiles of the secondary 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 dependences of the IFOL profiles are detailed in the remaining figures.

The \( q_a \) dependence of the inverse fall-off length of both databases increases rapidly and reaches a maximum near \( r = 0.35 \). The strength of this \( q_a \) dependence then decreases sharply, and outside \( r = 0.6 \) the profile shape is independent of \( q_a \). (Between \( 0.15 \leq r \leq 0.25 \), the point estimates of the discrete IFOLs suggests a radially uniform \( q_a \) dependence. However, the large error bars allow for a slope between roughly -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 parametric dependence to describe a straight line through the origin in this region.

The \( I_p \) and \( \bar{n}_e \) dependences are much weaker. Although, over some portions of the radius, rejection at the 5% level is marginally avoided, temperature profile invariance with respect to \( I_p \) and \( \bar{n}_e \) is generally seen to hold, within the experimental error limits, for all radii. Figures 7 and 8 show reference profiles and parametric radial dependences for the normalized profiles \( \ln T_e(r) \) (where \( T_e(r) = T_{e,0}(\bar{n}_e, I_p, q_a) \exp(-f(q_a)r^2) \))

\[
-L^{-1} = -\frac{d}{dr} \ln T_e(r) = 2rf(q_a)
\]

which describes a family of straight lines fanning out from the origin. Examining Fig. 5(a) (and Fig. 6(a)), we see that far from fanning out with increasing radius, our \( L^{-1} \) profiles converge to a unique curve for \( r > 0.6 \). This invariance occurs where one would expect most \( q_a \) variation (see Eq. (13)). It is clear from this evidence that the Gaussian hypothesis is false for our data. It is also visually obvious that even the \( q_a \) independent part of the profile \( (r > 0.6) \) does not lie on a Gaussian. To quantify this, we made a regression of channels 12 to 15 (i.e. those roughly satisfying \( r \geq 0.6 \)), normalized to their line averaged value, versus a Gaussian function, which gave \( T_e(r \geq 0.6) = 3.76 T_{e,av}^{line} \exp(-2.53r^2) \), with an RMS relative error of 5.4%. Adding a cubic term already strongly reduces the error: \( T_e(r \geq 0.6) = 1.88 T_{e,av}^{line} \exp(1.48r^2 - 3.62r^3) \), with an RMS relative error of 3.8% (F value for the cubic term = 400). Thus, even for the outlying portion of the profile, we can categorically rule out a Gaussian shape.

Figures 9 and 10 show a sample experimental \( T_e \) profile from each database, representing extremes of the \( q_a \) range covered by our data. In each case, the predicted profile with 95% global confidence bands is also shown. Since we are concerned here with both profile magnitude and shape recovery, these predictions come from a parameterized spline regression of unnormalized experimental profiles. Hence the confidence bands are wider (by \( \approx 50\% \)) than those of the corresponding normalized profile predictions.

### 7.2. Temperature profile variance

We now examine how well the fitted profiles describe the data, both in terms of root mean squared error and the fraction of data variance explained by the model. We also give a breakdown of the profile fluctuations for within-discharge and discharge-to-discharge contributions.

Table III(a) presents some descriptive statistics for the temperature profiles in the main database on a channel-by-channel basis. Table V(a) displays the equivalent information for the secondary database. The channels are numbered according to their vertical position, with channel 1 at \( Z = 0.200 \) m, channel 6 at...
FIG. 7. Reference $T_e$ profiles (normalized at $r = 0.2$) for the main database, and parametric dependence profiles for $q_a$, $I_p$ and $n_r$.

(a) (b) (c) (d)
FIG. 8. Reference $T_e$ profiles (normalized at $r = 0.2$) for the secondary database, and parametric dependence profiles for $q_0$, $b$, and $n_r$. 

(a) 

(b) 

(c) 

(d)
MCCARTHY et al.

FIG. 9. Sample experimental and predicted temperature profile with global 95% confidence bands from the main database. Parameters: $q_a = 1.94$, $I_p = 0.452$ MA, $n_e = 0.704 \times 10^{20}$ m$^{-3}$.

FIG. 10. Sample experimental and predicted temperature profile with global 95% confidence bands from the secondary database. Parameters: $q_a = 4.01$, $I_p = 0.281$ MA, $n_e = 0.299 \times 10^{20}$ m$^{-3}$.

Z = 0 m, and channel 15 at Z = -0.353 m. For each channel, the mean normalized flux radius is presented, followed by the mean temperature and the spread in keV. The spread (this term is chosen to avoid possible confusion with standard deviation in the sense of ‘regression error’) is just the ‘standard deviation from the mean’, i.e.

$$\text{spread} = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})^2}$$

where $\bar{y}$ is the sample mean. Recall that, before the regression, the profiles were normalized by the magnitude parameter from the SAS GLM procedure, i.e. $T_e(r = 0.2)$, obtained from fitting each profile individually. Columns 5 and 6 present the mean and the spread of these normalized profiles. Channels 5 and 8 have very small spreads since they lie closest to the normalization radius.

As we are fitting on the natural logarithmic scale, it is of particular interest to tabulate the logarithmic, or relative, spread of the normalized profiles as a measure.
of the total variation of the data for each channel. This quantity, together with the channel-by-channel noise level estimates and the RMSEs resulting from our spline parameterization of the temperature profiles, is presented for the main database in Table III(b) and for the secondary database in Table V(b). In column 3, we display $\frac{\delta_{\text{int,comp}}}{\sqrt{12}} = \delta_{\text{int,comp}}$, the estimated standard deviation of the within-discharge noise scaled for time compressed profiles. Column 4 tabulates $\delta_{\text{tot}}$, the total noise level (of the compressed profiles), estimated by regressing $\ln T_e$ for each channel on the bulk parameters. The differences of the squares of the entries in columns 3 and 4 are an estimate of the discharge-to-discharge variance.

The ratio of the two noise estimates lies in the range $0.2 \leq \frac{\delta_{\text{int,comp}}^2}{\delta_{\text{tot}}^2} \leq 0.5$, indicating that the discharge-to-discharge variance is the dominant contribution to $\delta_{\text{tot}}$. Using Eq. (8) and the RMS values (over the 15 channels) for $\delta_{\text{int,comp}}$ and $\delta_{\text{tot}}$, we find that, for the main database, $\delta_{\text{disrms}} = 3.6\%$. For the secondary database, $0.1 \leq \frac{\delta_{\text{int,comp}}^2}{\delta_{\text{tot}}^2} \leq 0.6$, and we have $\delta_{\text{disrms}} = 4.5\%$.

To indicate how much bias is introduced by our spline model, we also carried out channel-by-channel regressions of the normalized profiles. This is equivalent to using an interpolating radial spine. The RMSEs for these regressions appear in Table III(b), column 5. The RMSEs calculated for each channel from the parameterized spline regressions appear in column 6. The generally close agreement between columns 6 and 5 ($\delta_{\text{spline rms}} = 0.030$; $\delta_{\text{channel rms}} = 0.028$) confirms the adequacy of the five-knot spline model. The largest discrepancies occur for channels 2 and 10, which form one of the five up-down symmetric channel pairs. On inspecting the bias (systematic deviation from the regression line) for each channel, it was found that channels 2 and 10 had by far the largest bias (−3.3% and +2.9%, respectively; the next largest channel bias was 1.2%), indicating that at least one of these channels suffered from a systematic error of up to 6%.

We also checked for possible bias arising from the use of multiple profiles from the same discharge (see Section 2). The $T_e$ profile analysis was repeated for a subset of the main database where only one profile per discharge was admitted. We chose the profiles to cover the same density range as the full database. The results were very consistent with those presented here, the main difference being that the confidence bands were wider, as expected from the smaller number of profiles used.

In contrast to the main database, the spline regression errors in the secondary database (see Table V(b), column 6) are, in general, much larger than the channel-by-channel errors in column 5 ($\delta_{\text{spline rms}} = 0.069$; $\delta_{\text{channel rms}} = 0.039$). This is explained by the fact that, unlike the channel-by-channel regressions, we did not use operating period indicator variables in the spline parameterization of the secondary database profiles (Section 6.5).

For both databases, the regression errors generally increase for outlying channels, reflecting a progressively deteriorating signal-to-noise ratio. This is due to the decrease in scattered laser light signal intensity with decreasing electron density. Note that channel 12 is an exception, with errors similar to those for channel 15. This is consistent with the fact that, whereas all other channels had three distinct spectral filters (normally offering the choice of the less noisy of two independent determinations of the temperature), channel 12 had only two at the time the discharges for our databases were made.

7.3. Density

Since many of the results presented in Section 7.2 apply to the density profiles as well, we only mention the differences. Tables IV and VI contain the density statistics for the main and secondary databases, respectively.

Figures 11 and 12 portray graphically our parameterization of the local shape parameter of the density profile, $-\frac{L_{q_a}}{n_e} = -d \ln n_e/dr$, for each database. Figures 13 and 14 show the integrals of Figs 11 and 12, i.e. the normalized density profile parameterizations. Figures 15 and 16 show sample experimental $n_e$ profiles (with prediction profiles and confidence bands) for the same discharges which provided the sample $T_e$ profiles shown in Figs 9 and 10. Note the 'jump' in the predicted profiles at $r = 0$, which arises from the presence in the regression of the density up-down asymmetry variable (Section 6.7).

The dominant feature of the density IFOL profiles (Figs 11 and 12) is a $q_a$ dependence closely mirroring that of the temperature IFOL, though at a reduced magnitude:

$$-\frac{\partial L_{q_a}^{-1}}{\partial \ln q_a} (\text{max}) = 5 \quad \text{and} \quad -\frac{\partial L_{q_a}^{-1}}{\partial \ln q_a} (\text{max}) = 2$$

The radial region over which it is significant ($0.25 < r < 0.55$ for the main database) is smaller than the equivalent region for the temperature. Thus, the variation of the density profile shape with $q_a$, while not being as dramatic as that of the temperature, is nonetheless considerable, as is evident in the contrast
FIG. 11. Reference negative IPOL profiles for the main database $n_e$, and parametric dependence profiles for $q_p$, $I_p$, and $T_e$. 

(b) 

(c)
FIG. 12. Reference negative FOL profiles for the secondary database $n_e$, and parametric dependence profiles for $\psi$, $I$, and $n_t$. 
FIG. 13. Reference $n_e$ profiles (normalized at $r = 0.2$) for the main database, and parametric dependence profiles for $q_p$, $I$, and $n_e$. 

McCARTHY et al.
FIG. 14. Reference $n_e$ profiles (normalized at $r = 0.2$) for the secondary database, and parametric dependence profiles for $\varepsilon_r$, $I_p$, and $n_e$. 

ASDEX SCALING AND PLASMA PROFILE PARAMETERIZATION
between Figs 15 and 16. The magnitude of the density IFOL remains smaller than that of the temperature over the entire profile. At \( r = 0.9 \) it takes values between 2 and 3 (fall-off lengths between 20 cm and 13.3 cm for a minor radius of 0.4 m), whereas the temperature fall-off length at \( r = 0.9 \) is nearly fixed at \( L^{-1} = 5.5 \) (see Fig. 5(a)), i.e. a fall-off length of about 7 cm for \( a = 0.4 \) m. As is the case with the temperature profile, the shape dependences of \( I_p \) and \( \bar{n}_e \) are weaker than that of \( q_a \). Near the edge, however, there is a statistically significant broadening of the density profile shape with increasing current. (We note with caution, however, that the discrete IFOL data points suggest that this current dependence is due solely to the outermost channel.) In addition, some flattening of the density profile with increasing \( \bar{n}_e \) occurs in the region \( 0.5 < r < 0.7 \).
FIG. 17. Reference negative FOL profiles for the main database $P_x$ and parametric dependence profiles for $Q_x$ and $R_x$.
FIG. 18. Reference negative IFOL profiles for the secondary database $p_s$, and parametric dependence profiles for $q_o$, $I_p$, and $n_e$. 
FIG. 19. Reference $P_x$ profiles (normalized at $r = 0.2$) for the main database, and parametric dependence profiles for $q_a$, $I$ and $n_e$. 

(a) 

(b) 

(c) 

(d)
FIG. 20. Reference in p profiles (normalized at $r = 0.2$) for the secondary database, and parametric dependence profiles for $q_a$, $L$, and $n_e$. 

(a) 

(b) 

(c) 

(d)
For both databases (as can be seen from the RMS values for $\delta_{\text{int,comp}}$ and $\delta_{\text{ini}}$), the discharge-to-discharge variance ($\delta_{\text{dis, rms}} = 2.2\%$ for the main database and $\approx 3.2\%$ for the secondary database) represents the largest contribution to the total variance, as is the case for the temperature profiles. The overall spline regression RMS relative error for each database (0.032 and 0.065) is very similar to the corresponding temperature value.

A number of density profiles in both databases are slightly hollow in the region $0.3 < r < 0.4$. Since the set of reference profile parameters ($q_a^* = 2.5$, $I_p^* = 0.4$ MA and $\bar{n}_e^* = 0.4 \times 10^{20}$ m$^{-3}$) was not very typical for the secondary database, this feature looks somewhat exaggerated for that database (see Figs 12(a) and 14(a)).

**7.4. Electron pressure**

The analysis of the (logarithmic) electron pressure profiles, defined as $\ln P_e = \ln T_e + \ln n_e$, offers additional insight, as can be seen from the parametric dependences shown in Figs 17-20. For the main database, the most striking feature is that the $I_p$ dependence is more significant over most of the radius for the pressure profile than either the temperature profile or the density profile alone. In the outer region of the plasma, there is a clear broadening of the pressure profile with the current, while peaking occurs in the region $0.2 < r < 0.4$. The $q_a$ dependence of the inner half of the profile is very strong (temperature and density profile dependences reinforce each other), whereas the $\bar{n}_e$ dependence is little changed from that of the density profile.

**8. DISCUSSION AND SUMMARY**

**8.1. Bulk scalings**

We have first presented and compared the scalings of various global plasma parameters for two complementary Ohmic datasets. For the main, high density database, the volume averaged temperature and three independently measured values of the total plasma energy depend on the plasma current $I_p$ and the line averaged density $\bar{n}_e$, but, at constant $I_p$ and $\bar{n}_e$, they are practically unaffected by the toroidal magnetic field. For the secondary database, $W_{\text{pdia}}$ and $W_{\text{pmhd}}$ show strong $q_a$ scaling differences, while $W_{\text{pkin}}$ and $\langle T_e \rangle$ scalings are almost the same as in the main database. The nearly linear current dependence for both the temperature and the total energy is reminiscent of L-mode scaling.

The Spitzer $Z_{\text{eff}}$ depends on all three control variables. Regression of $Z_{\text{eff}} - 1$ indicates, for both databases, that the impurity density is almost independent of $\bar{n}_e$ but strongly dependent on both $I_p$ and $B_t$. The electron temperature profile peaking factor $T_e^{3/2}/\langle T_e \rangle$ scales as $0.94(\pm 0.04)q_a^{0.07} \pm 0.06$, in close agreement with the prediction of classical resistive equilibrium.

The strong dependences on the plasma current $I_p$ of both the total plasma energy $W_p$ and the Ohmic power approximately cancel to give a relatively weak current scaling for $\tau_E$. Replacing $(B_t, I_p)$ by $(q_a, B_t)$ gives a $\tau_E$ scaling with a weak dependence on $\bar{n}_e$ and a moderate dependence on $q_a$, and with no statistically significant dependence on $B_t I_p$. In the range $0.3 < \bar{n}_e/10^{20}$ m$^{-3} < 0.8$, we have detected little dependence of the global confinement time on the density. In particular, a decrease in the rollover regime with density, as reported in Ref. [12], was not found. It should be remarked that the SOC data for $\bar{n}_e > 0.5 \times 10^{20}$ m$^{-3}$ in Ref. [12], first reported in Ref. [22], are based on a single scan with $I_p = 0.42$ MA and $B_t = 2.2$ T. Obviously, some scans show confinement deterioration, while others show a flat or even weakly improving $\tau_E$ versus $\bar{n}_e$ dependence. The same phenomenon can be observed in plots from the Doublet III analysis [23]. The reason for this variability in SOC density scaling is at present unclear and requires further investigation.

**8.2. Profile analysis**

A careful statistical analysis is necessary to determine the radially varying parametric dependences of the profile shapes on the bulk plasma variables. By simultaneously fitting all profiles with spline coefficients which depend on the plasma variables, we have been able to examine the profile dependences on a detailed, quantitative level. On the basis of this spline model, a convenient graphical representation has been used to inspect visually the influence of the various plasma parameters on the profile shapes.

An earlier study of ASDEX temperature profile shapes [21] (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 'the influence of the $q = 1$ surface' . The results of our profile parameterization are roughly consistent with and constitute a refinement of this analysis for Ohmic profiles.

Except for a dependence (due solely to the outermost channel) of the density profile shape on plasma current (and to a lesser extent on $\bar{n}_e$), the $I_p$ and $\bar{n}_e$ dependences of both the temperature profile and the density shapes
are rather weak in general. Over most of the radius, the current and density dependences are not significant, given the error bars of the databases.

In the interior, $q_a$ is the dominant bulk plasma parameter in determining the temperature shape. By $r = 0.5$, however, this dependence has weakened considerably, and outside $r = 0.6$ the IFOL profile has an invariant shape (see Fig. 5). We note that the extent of the $q_a$ sensitive region is reasonably consistent with the widest sawtooth inversion radius in each database ($r_{\text{inv}}(\text{max}) = 1/q_a(\text{min}) = 0.54$ and 0.42 for the main and secondary databases, respectively).

Comparing Figs 11 and 5, we see that the variation of the density profile shape, while significant, is much weaker than that of the temperature. This follows from the result that, over the inner half of the radius, the sensitivity of the density IFOL to $q_a$ is only $\approx 40\%$ of that of the temperature. The $q_a$ dependence is only significant for $0.25 \leq r \leq 0.5$. In contrast to the temperature shape, which is unique outside $r = 0.6$, 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 broadening of the outer half — a tendency which intensifies approaching the plasma boundary.

Our findings are well described in terms of profile invariance [1, 19] and are in quantitative agreement with important criteria for profile consistency as described in Ref. [5] and developed by many authors. However, we have not addressed the relative merits of profile consistency versus local transport models [11, 24] containing sawtooth effects. This issue could be addressed by a statistical comparison of experimental profile dependences with the dependences predicted by local transport models.

Appendix

STATISTICAL SIGNIFICANCE
OF REGRESSION VARIABLES

The significance of a regressor $x_j$ in the least-squares model can be interpreted in terms of $\hat{\alpha}_j/\hat{\sigma}(\hat{\alpha}_j)$, 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 $\alpha_j = 0$. For any statistic $T$ that has a $t$ distribution with $f$ degrees of freedom, the following relation between the critical value $t_{f, \epsilon}$ and the 'exceedence probability' or significance level $\epsilon$ holds:

$$ P\{ |T| > t_{f, \epsilon} \} = \epsilon \quad (A.1) $$

The null hypothesis $\alpha_j = 0$ is rejected and the regressor is considered significant if $|\hat{\alpha}_j/\hat{\sigma}(\hat{\alpha}_j)| > t_{f, \epsilon}$ for some small value of $\epsilon$, say 5%. For many degrees of freedom ($f > 30$ usually suffices), as in the present case, the Student's $t$ can be well approximated by the normal distribution. Thus, we have $t_{0.05} = 2.0$, and the significance criterion is $|\hat{\alpha}/\hat{\sigma}(\hat{\alpha})| > 2.0$. 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. Without loss of generality, we consider the multiple linear regression problem

$$ y = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_p x_p + \epsilon \quad (A.2) $$

with centred dependent and independent variables. If the independent variables are uncorrelated, i.e. if $\langle x_j x_k \rangle = \sum_{i<j} x_{ij} x_{ik} = ||x_j||^2 \delta_{j,k}$, it is easily shown that the least-squares solution reduces to

$$ \hat{\alpha}_j = \frac{\langle x_j, y \rangle}{||x_j||^2} \quad \text{and} \quad \hat{\sigma}^2(\hat{\alpha}_j) = \frac{\hat{\sigma}^2}{||x_j||^2} \quad (A.3) $$

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

$$ t_j = \frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)} = \frac{\langle x_j, y \rangle}{||x_j|| \hat{\sigma}} = \frac{\sqrt{N - p} \langle x_j, y \rangle}{||y - \hat{y}|| ||x_j||} \quad (A.4) $$

In geometrical terms, $(x_j, y)/||x_j||$ is the projection of $y$ onto $x_j$, where $y$ and $x_j$ are vectors in $\mathbb{R}^N$. Hence, adding $x_j$ to the regression model makes a fractional contribution to the total variance of

$$ \Delta(R^2)_j = \frac{\langle x_j, y \rangle^2}{||x_j||^2 ||y||^2} \quad (A.5) $$

1632
Noting that

\[ 1 - R^2 = \frac{\|y - \hat{y}\|^2}{\|y\|^2} \]  
(A.6)

we can use Eqs (A.4) and (A.6) to eliminate all terms involving \( x_j \) or \( y \) in Eq. (A.5), and we finally obtain

\[ A(R^2)_j = -\frac{t^2}{N - p} (1 - R^2) \]  

(A.7)

where the left-hand side denotes the decrease in \( R^2 \) if the \( j \)-th regressor is removed from the model. Note that this relationship strictly holds only for uncorrelated regressors. If we now sum up all contributions, we obtain, using Pythagoras' theorem, \( \sum_j (x_j, y)^2 / \|x_j\|^2 \), which leads to the equality

\[ \sum_{j=1}^p t_j^2 = (N - p) \frac{\|y\|^2}{\|y - \hat{y}\|^2} \]

\[ = (N - p) \frac{R^2}{1 - R^2} \]  
(A.8)

This formula is the analogue of Weisberg's partition of \( C_p \) [25]. It provides a useful practical check on the applicability of Eq. (A.7) when the regressors are correlated, which is usually the case.

**ACKNOWLEDGEMENTS**

The work of PJM was performed under a Euratom supported reciprocal research agreement between the Max-Planck-Institut für Plasmaphysik, Garching, and University College, Cork. The work of KSR was supported by the United States Department of Energy, under Grant No. DE-FG02-86ER53223.

The authors are indebted to the referees, who made several critical suggestions from which the manuscript has profited considerably.

**REFERENCES**

[1] BESSENRODT-WEBERPALS, M., MCCORMICK, K., SÖLDNER, F.X., Nucl. Fusion 31 (1991) 155.
[2] MURAKAMI, M., CALLEN, J.D., BERRY, L.A., Nucl. Fusion 16 (1976) 347.
[3] MÜLLER, E.R., SÖLDNER, F.X., JANESCHITZ, G., et al., in Controlled Fusion and Plasma Heating (Proc. 15th Eur. Conf. Dubrovnik, 1988), Vol. 12B, Part I, European Physical Society (1988) 19.
[4] KARDAUN, O.J.W.F., RIEDEL, K.S., McCARTHY, P.J., et al., A statistical approach to plasma profile analysis, submitted to Nucl. Fusion.
[5] ARUNASALAM, V.D., BRETTZ, N.L., EFFTHIMION, P.C., et al., Nucl. Fusion 30 (1990) 2111.
[6] RÖHR, H., STEUER, K.H., MURMANN, H., MEISEL, D., Nucl. Fusion 22 (1982) 1099.
[7] MARDIA, K.V., KENT, J.T., BIBBY, J.M., Multivariate Analysis, Academy Press, London (1979).
[8] STEUER, K.H., RÖHR, H., ROBERTS, D.E., et al., in Controlled Fusion and Plasma Heating (Proc. 15th Eur. Conf. Dubrovnik, 1988), Vol. 12B, Part I, European Physical Society (1988) 31.
[9] KARDAUN, O., THOMSEN, K., CORDEY, J., et al., in Controlled Fusion and Plasma Heating (Proc. 17th Eur. Conf. Amsterdam, 1990), Vol. 14B, Part I, European Physical Society (1990) 110.
[10] PFIEFFER, W., WALTZ, R., Nucl. Fusion 19 (1979) 51.
[11] BARTLETT, D.V., BICKERTON, R.J., BRUSATI, M., et al., Nucl. Fusion 28 (1988) 73.
[12] MERTENS, V., BESSENRODT-WEBERPALS, M., DODEL, G., et al., in ASDEX Contributions to the 17th European Conference on Controlled Fusion and Plasma Physics, Rep. 1/6A, Max-Planck-Institut für Plasmaphysik, Garching (1990) 1.
[13] WALTZ, R.E., WONG, S.K., GREENE, J.M., DOMINGUEZ, R.R., Nucl. Fusion 26 (1986) 1729.
[14] SAS/STAT, User's Guide: Statistics Release 6.03 Edition, Cary, NC: SAS Institute Inc. (1998) 1028 pp.
[15] ST. JOHN, H., JAHNS, G.I., BURRELL, K.R., DeBOO, J.C., Bull. Am. Phys. Soc. (1987) 9.
[16] McCARTHY, P.J., RIEDEL, K.S., KARDAUN, O., et al., Scales and Plasma Profile Parameterisation of ASDEX High Density Ohmic Discharges, Rep. 5/3A, Max-Planck-Institut für Plasmaphysik, Garching (1990).
[17] POORTEMA, K., On the Statistical Analysis of Growth, PhD Thesis, University of Groningen (1989) 98 pp.
[18] DRAVER, H., SMITH, N., Applied Regression Analysis, 2nd edn, Wiley, New York (1981).
[19] KARDAUN, O.J.W., McCARTHY, P.J., LACKNER, K., RIEDEL, K., GRUBER, O., in Theory of Fusion Plasma (Proc. Joint. Varenna-Lausanne Int. Workshop Chexbres, 1988), Editec Compositori, Bologna (1989) 435.
[20] FREDRICKSON, E.D., McGUIRE, K.M., GOLDSTON, R.J., et al., Nucl. Fusion 27 (1987) 1897.
[21] MURMANN, H., WAGNER, F., BECKER, G., et al., in Controlled Fusion and Plasma Heating (Proc. 13th Eur. Conf. Schliersee, 1986), Vol. 10C, Part I, European Physical Society (1986) 216.
[22] WAGNER, F., GRUBER, O., BARTIROMO, R., et al., in: ASDEX Contributions to the 12th European Conference on Controlled Fusion and Plasma Physics, Rep. 3/107, Max-Planck-Instit für Plasmaphysik, Garching (1985) 16.
[23] EJIMA, S., PETRIE, T.W., RIVIERE, A.C., et al., Nucl. Fusion 22 (1982) 1627.
[24] CHRISTIANSEN, J.P., CALLEN, J.D., CORDEY, J.G., THOMSEN, K., Nucl. Fusion 28 (1988) 817.
[25] WEISBERG, S., Technometrics 23 (1981) 27.

(Manuscript received 6 April 1990
Final manuscript received 25 April 1991)