A Statistical Approach to Plasma Profile Analysis

O.J.W.F. Kardaun, K.S. Riedel*, P.J. Mc Carthy†,
K. Lackner

IPP 5/35 May 1990

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

8046 GARCHING BEI MÜNCHEN

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK
GARCHING BEI MÜNCHEN

A Statistical Approach to Plasma Profile Analysis

O.J.W.F. Kardaun, K.S. Riedel*, P.J. Mc Carthy†,
K. Lackner

IPP 5/35 May 1990

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

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

A STATISTICAL APPROACH TO PLASMA PROFILE ANALYSIS

O.J.W.F. Kardaun, K.S. Riedel*, P.J. Mc Carthy†, K. Lackner

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

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

Abstract

A general statistical approach to the parameterisation and analysis of tokamak profiles is presented. The modelling of the profile dependence on both the radius and the plasma parameters is discussed, and pertinent, classical as well as robust, methods of estimation are reviewed. Special attention is given to statistical tests for discriminating between the various models, and to the construction of confidence intervals for the parameterised profiles and the associated global quantities. The statistical approach is shown to provide a rigorous approach to the empirical testing of plasma profile invariance.

CONTENTS

1. SUMMARY AND INTRODUCTION

Since the advent of active beam methods, such as laser induced Thomson scattering to determine the electron density and temperature and neutral beam CX spectroscopy to determine the ion temperature, direct measurements of relevant plasma parameters at discrete radial positions, without the need of Abel inversion, have become available, and plasma physicists have been interested in 'smoothing' these profiles in order to use them in interpretation codes or to compare them with theoretical predictions. As the density and temperature profiles contain vital information on the confinement properties of the plasma, accurate empirical representation of the profile dependence on the basic plasma parameters is an important intermediate step in the understanding and controlling of confinement. As individual profile measurements exhibit often considerable scatter, and basic transport properties depend on the radial derivatives of these profiles, it is of importance to express clearly the statistical accuracy that is obtained from a series of profile measurements. Recently, some work has been performed on such a statistical analysis of plasma profiles [1], [2], and local transport coefficients [3], [4]. In this article, we extend this work by giving a thorough overview and discussion of the statistical techniques to analyse tokamak profiles and their associated global quantities.

In section 2, we describe the analysis of a single sample of n unsmoothed profiles, for a fixed value of the plasma variables I_p, B_t, etc., by standard multivariate statistical analysis. In this approach, each empirical profile, consisting of temperature or density measurements at p different radial locations, is considered as one observation in $\mathbb{R}^p$.

In section 3, we consider various continuous representations for the true underlying profile, and we discuss a number of relevant error structures.

In section 4, the influence of plasma variables, such as I_p, B_t, etc. on the profiles is modelled, and a mathematical definition of profile invariance is given.

In section 5, various methods of estimating the free parameters in the models of sections 3 and 4 are reviewed and discussed.

In section 6, several statistical tests for discriminating between different concrete profile representations are given, paying attention the the mean value structure as well as the error structure.

In section 7, confidence and prediction bands for the underlying profiles are constructed, as well as confidence intervals for the derived volume-averaged quantities.

To indicate the context of our investigation, we start to give an idealised view of tokamak operation as an input-output system. For each datapoint, a set of bulk plasma

parameters has been selected, on different time scales, by the Tokamak design group, by the committee setting the campaign period goals, and finally by the experiment leader.

All these parameters we call input variables, and denote them by $\underline{x} = (x_1, \ldots, x_w)$. For Ohmic discharges, the input variables are: the toroidal magnetic field B_t, the safety factor q_{cyl}, the total current I_p (for a fixed geometry, any two of these three variables determine the third one, in fact, using the units T, MA and m, $q_{cyl} = 5\frac{B_t}{I_p}\frac{ab}{R}$), and the line-averaged density $\overline{n}$; the minor horizontal radius a, the minor vertical radius b, the major radius R, and the vacuum field magnetic ripple; the plasma cross-section shape, typically characterised by discrete variables, such as the configuration type (limiter, single null divertor, double null divertor), well as by continuous variables such as ellipticity, triangularity, and the distance between the separatrix and the wall; the ion species mixture, and the wall conditioning (carbonisation, boronisation, gettering, etc.). For neutral beam heating, additional variables must be included such as the input power, angle of injection, energy per injected particle, species of injected particle. Radio frequency heating requires the specification of input power, the location and shielding of the antenna and the spectra of the injected waves. If pellet injection is used to refuel the plasma, additional variables are the size, frequency, angle of injection, species and velocity of the pellet.

The output variables that we will be principally interested in are the temperature and density (and hence pressure) profiles, as well as the derived local transport coefficients. Other output variables are the profiles of current, impurity density and radiation, the loop voltage, or, equivalently, bulk resistance, the toroidal rotation, the (poloidal or toroidal) plasma beta, the energy confinement time, the Shafranov shift as well as other moments of the magnetic field, and the characteristics of ELM activity (frequency, amplitude) and sawteeth (frequency, amplitude, and inversion radius).

2. DISCRETE PROFILE REPRESENTATIONS

In this and the next section, we consider data at single point in plasma parameter space, consisting of n separate observations of a plasma profile, such as temperature, at p distinct radial points, $T_j(r_l)$ with $1 \leq j \leq n$ and $1 \leq l \leq p$. Let the variable Y denote the logarithm of the profile variable (density or temperature). The n observed profiles can be represented by the basic data matrix:

$$\begin{pmatrix} Y_1(r_1) & \dots & Y_1(r_p) \\ \vdots & \ddots & \vdots \\ Y_n(r_1) & \dots & Y_n(r_p) \end{pmatrix} \tag{1}$$

All variables depend on time. We will restrict, however, our attention to steady state plasmas and assume that the final state is independent of the past history of the discharge. We further note that in a stationary phase all plasma profiles evolve more or less periodically, the periodicity being dictated by the sawtooth crashes. Our data will consist of n distinct time points.

Each profile measurement has essentially three sources of deviations from the 'ideal profile': 1) deterministic (systematic) errors, 2) random errors due to measurement noise, 3) fluctuations due to variations in the plasma. Systematic errors arise either from overlooking some physical effect, or from idealisation of the measuring process. For Thomson scattering experiments, these assumptions typically include perfectly collimated measurement devices, physical localization of the scattering, neglect of other nonresonant scattering mechanisms, and the assumption that the electron velocity distribution is Maxwellian. Often the size of such errors can be roughly estimated, but it is usually very difficult to determine them more precisely, so that the resulting bias can only rarely be corrected for. (In fact, the 'easier' types of systematic errors are already accounted for in the standard evaluation of the diagnostic.) The random errors due to measurement noise may often be estimated by an error-propagation analysis.

A special difficulty is our inability to characterise the condition of the wall. Aside from discrete categories such as gettered/ungettered and wall carbonisation, it is clear that the immediate past history of the device influences impurity recycling from the wall. Only to a rough approximation one may assume that the influence of the wall on the plasma can be parameterised by a single variable such as Z_{eff}.

In Ohmic discharges, the primary source of the intrinsic plasma variation is the m=1 sawtooth activity. The perturbations about the mean profile from sawteeth are highly correlated in space, since the profiles are flatter immediately after a sawtooth crash. This spatial correlation of the fluctuations should be taken into account in a careful profile analysis.

The profile fluctuations are generally temporally as well as radially correlated. The profiles of the ASDEX Thomson scattering experiments are sampled with a fixed 60 Hz frequency. We assume that this frequency is incommensurable with the sawtooth period, and hence that the sampling times are quasi-randomly distributed with respect

to the sawtooth phase $\frac{t-t_{crash}}{\tau_{saw}}$. This assumption of quasi-randomness with respect to the sawtooth phase allows us to neglect, for simplicity, the temporal correlation.

The spatial correlation is conveniently described by viewing the j^{th} profile measurement $\underline{Y}_j = (Y_j(r_1) \ldots Y_j(r_p))$ as one (random) vector of observations with mean μ and covariance matrix $\underline{\underline{\Sigma}}$. Assuming also normality, which can sometimes, if necessary, be relaxed somewhat, we write:

$$\underline{Y}_j \sim N_p(\underline{\mu}, \underline{\underline{\Sigma}}), \qquad j = 1, \ldots, m. \tag{2}$$

Note that the systematic errors are not described by the covariance matrix $\underline{\underline{\Sigma}}$. One can formally decompose

$$\underline{\underline{\Sigma}} = \underline{\underline{\Sigma}}_{noise} + \underline{\underline{\Sigma}}_{plasmavar.} \tag{3}$$

The number of free parameters is p for the μ and $p(p+1)/2$ for $\underline{\underline{\Sigma}}$. If the total number of observations, np is sufficiently large, (at least several times the number of free parameters) these parameters may be estimated by using the standard maximum likelihood estimators of $\underline{\mu}$ and $\underline{\underline{\Sigma}}$:

$$\underline{\hat{\mu}} = \frac{1}{m}\sum_{j=1}^{m} \underline{Y}_j, \qquad \underline{\underline{\hat{\Sigma}}} = \frac{1}{m}\underline{\underline{S}} = \frac{1}{n}\sum_{j=1}^{m} (\underline{Y}_j - \underline{\hat{\mu}})(\underline{Y}_j - \underline{\hat{\mu}})^t. \tag{4}$$

The matrix $\underline{\underline{S}}$ is sometimes called 'the matrix of residual sums of squares and sums of cross products', which we shall abbreviate by 'the residual SSCP matrix'. $\underline{\underline{\Sigma}}_{plasmavar.}$ can only be estimated if one has an independent estimate of $\underline{\underline{\Sigma}}_{noise}$ from the error-propagation analysis. A reasonable assumption for the ASDEX Thomson scattering experiment is that $\underline{\underline{\Sigma}}_{noise}$ is diagonal.

A special feature of the ASDEX YAG diagnostic is that 10 of the 16 channels are located symmetrically with respect to the horizontal midplane. This affords the following possibility to test for up-down asymmetry for circular plasmas with z=0. Radial measurement positions that have no symmetrical counterpart are dropped from the analysis. The remaining spatial locations are divided into two groups, corresponding to the upper and lower part of the plasma. The temperature measurement vector is partitioned accordingly: $(\underline{\mu}^t = (\underline{\mu}_1, \underline{\mu}_2)^t)$, and the null-hypothesis to be tested is $\underline{\mu}_1 = \underline{\mu}_2$. If the profiles are up-down symmetric, then $\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2$ has a multivariate normal distribution, whose covariance matrix is easily derived from the partitioned covariance matrix of $\underline{Y}$. Inserting the usual estimate for this covariance matrix, we find that

$$T = (\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)^t (\underline{\underline{S}}_{11} - \underline{\underline{S}}_{12} - \underline{\underline{S}}_{21} + \underline{\underline{S}}_{22})^{-1} (\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2) \tag{5}$$

is distributed as $p/(m-p)$ times the $F_{p,m-p}$ distribution. (Here, $\underline{\underline{S}}_{11}$ stands for the residual SSCP matrix corresponding to $\underline{Y}_1$, $\underline{\underline{S}}_{12}$ for the residual SSCP matrix 'between' $\underline{Y}_1$ and $\underline{Y}_2$, and so forth.)

Physically one expects that the electron temperature is constant on flux surfaces. Hence, for symmetric plasma discharges, the true, underlying radial profile should be up-down symmetric and an observed asymmetry should be due to an asymmetry in the measuring process. Supposing that the two types of asymmetry ('physical', and 'measurement error') are additive, the physical asymmetry of the density profile may be estimated under the assumption that the density and the temperature profile exhibit the same total amount of asymmetry from measurement error.

3. CONTINUOUS PROFILE REPRESENTATIONS

To obtain a physically meaningful continuous profile representation, a preliminary mapping of the physical measurement points r_l' to the corresponding flux radii, r_l must be performed. For low beta, large aspect-ratio devices, a simple calculation based on the Shafranov shift is sufficient. For smaller aspect ratios, a nonlinear correction based on function parameterisation can be used [5], [6]. We will not consider such corrections here, nor the possible errors associated with the flux-surface mapping.

For a complete specification of a statistical model for plasma profiles, we must specify both its mean-value structure (i.e., give a description of how the mean values of the profiles depend on the input variables) and its covariance structure (i.e., present a stochastic model for the errors in the observations).

3.1 Mean value structures

For every discrete set of profile measurements, $\{T(r_l),\ l = 1,\ldots,p\}$, we seek a continuous representation of the profile. This has several advantages. The profile may be described by a relatively small number of coefficients. Furthermore, it facilitates comparison between profiles that are measured at different sets of radial points. Finally, smoothness is imposed in the belief that the profiles are in diffusive equilibrium.

Various transformations of the plasma profile may be considered. A natural transformation is to take logarithms of the temperature, density and pressure profiles. Minimising the error in the fit on logarithmic scale corresponds to minimising the relative

error as opposed to the absolute error. If the relative error is more nearly constant over the database then the absolute error, then on logarithmic scale, in regressing the response variable (temperature, etc.) against the input variables, one can apply simple unweighted regression. (Often, however, the absolute error in the experimental measurements increases whereas the relative error decreases with increasing value of the response variable. In that case, a weighted regression is needed on either scale.) A logarithmic fit has the advantage that no restriction on the regression parameters is needed to ensure that the fitted temperature and density profiles are always positive. Furthermore, after logarithmic transformation, the concept of profile consistency can be defined as additivity of the radial dependence and the plasma-parameter dependence of the profiles. (The notion of 'additivity' or 'non-interaction' is a well-known and heavily investigated concept in statistics.) Some additional conveniences in using a logarithmic scale come into effect when a power-law type scaling can be postulated. Firstly, a power-law scaling is transformed into a linear regression model. This facilitates the estimation procedure. Furthermore, taking logarithms makes the scaling law dimensionless: The usual transition from dimensional variables to dimensionless variables (see e.g. [7]), by taking products and quotients, corresponds to a linear transformation on logarithmic scale, possibly with reduction of the dimensionality by linear constraints. Finally, since on logarithmic scale the pressure is just the sum of the density and the temperature, one can easily obtain the full regression information by multivariate regression of only two of these three variables against the plasma parameters.

We restrict our attention to models with parametric dependencies of the form:

$$\mu(r) = \sum_{h=1}^{p'} \alpha_h f_h(r), \tag{6}$$

where $f_1(r), \ldots, f_{p'}(r)$ are basis functions. For p radial measurement positions $r_1 \ldots r_p$, we can write

$$\underline{\mu} = \underline{\underline{X}}_{rad}\underline{\alpha}, \tag{7}$$

where $\underline{\mu} = (\mu(r_1), \ldots, \mu(r_p))^t$, $\underline{\alpha} = (\alpha_1, \ldots, \alpha_{p'})^t$ is a vector of regression coefficients, and

$$\underline{\underline{X}}_{rad} = \begin{pmatrix} f_1(r_1) & \ldots & f_{p'}(r_1) \\ \vdots & \ddots & \vdots \\ f_1(r_p) & \ldots & f_{p'}(r_p) \end{pmatrix}, \tag{8}$$

is the radial design matrix. We now consider three possible sets of basis functions for continuous representations of the plasma profiles.

3.1.1 Polynomials

Clearly polynomials constitute a simple representation. Assuming symmetry and analyticity of the plasma profiles, we expand in even polynomials in r. Expanding the logarithm of the temperature, the representation is

$$T(r) = T_0 \exp(\sum_{n=1}^{\infty} a_n r^{2n}) \tag{9}$$

The basis functions $(1, r^2, r^4, r^6, \cdots)$ have the property that the higher order polynomials are essentially localised near the outer edge of the plasma. Only a few terms are fitted in practice. It should be noted that these simple basis functions are highly non-orthogonal and hence their estimated coefficients are rather ill-determined and have relatively large standard deviations. This need not be a strong inconvenience if the emphasis is on predicting the profiles, and one is not especially interested in the regression coefficients themselves. The estimated coefficients of orthogonal polynomials are close to being statistically independent, at least if the radial positions are not too unwisely chosen, and the measurement errors are independent and radially uniform.

3.1.2 Perturbation Expansion

We write $T(r)$ as the product of a gaussian and a polynomial expansion. Allowing for asymmetry,

$$T(r) = T_o \exp(-d_o r^2) \left(1 + \sum_{n=1}^{\infty} b_n r^n\right). \tag{10}$$

Such a representation may work conveniently whenever the plasma profiles differ little from gaussians, in which case one hopes for a rapid convergence. Another possibility is to express the expansion by Hermite polynomials.

$$T(r) = T_o \exp(-c_o r^2) \left(1 + \sum_{n=1}^{\infty} a_n H_{n,c_0}(r)\right), \tag{11}$$

where $H_{n,c_0} = e^{c_0 r^2}(-d/dr)^n e^{-c_0 r^2}$.

For each c_0 we have a complete set of basis functions. The choice of c_0 affects how many terms are needed. A convenient procedure is to estimate c_0 by fitting a gaussian function to the temperature profile, and subsequently estimate the polynomial coefficients $a_1, a_2, \ldots$ for fixed c_0. Both steps can be carried out by applying linear regression. For any fixed number of coefficients $a_1, \ldots, a_m$, an optimal value of c_0 can be

determined by minimising the residual sum of squares with respect to c_0. The correlation between the estimated polynomial coefficients will be reduced if the polynomials are orthogonalised on the finite interval [0,1], i.e when generalised Hermite polynomials are used. (Whether this orthogonalisation will be worth the effort in practice, is another question.)

For any fixed value of c_0, the coefficients $a_1, a_2, \ldots$ are linearly related to the moments

$$m_n = \int r^n T(r) dr, \qquad n = 1, 2, \ldots \tag{12}$$

of the temperature distribution, see Appendix. The profiles are symmetric if the odd coefficients of the Hermite expansion, or equivalently, the odd temperature moments are zero. As the number of radial measurement positions is limited, only the lower moments can be tested. A convenient way is to test, e.g., $a_1 = a_3 = a_5 = 0$, directly from the asymptotic normal distribution of the fitted regression coefficients.

The kurtosis, K_u is the normalised fourth central radial moment of the plasma profile. For a symmetric profile (for which $m_1 = 0$), it is defined [8] as

$$K_u = \frac{m_4/m_0}{(m_2/m_0)^2}. \tag{13}$$

The kurtosis, which is 3 for a Gaussian profile, is a sensible measure of the the 'broadness' of a profile with respect to to a Gaussian. Expressed in the Hermitian regression coefficients, one has $K_u - 3 = \frac{6a_4 - 3a_2^2}{(a_2 + \frac{1}{4c_0})^2}$. The kurtosis can be used in scaling studies of density profile effects on energy confinement.

3.1.3 Splines

The physical domain is subdivided into a number of regions and low order polynomials are used to represent the the profile, i.e. the logarithm of the temperature, density or pressure, in each region. As a general spline representation, we consider

$$\mu(r) = \begin{cases} \phi_0(r) & \text{for } 0 \le r < r_1 \quad \text{(Inner Region)} \\ \phi_0(r) + \phi_1(r) & \text{for } r_1 \le r < r_2 \quad \text{('1-2' Region)} \\ \phi_0(r) + \phi_1(r) + \phi_2(r) & \text{for } r_2 \le r < r_3 \quad \text{('2-3' Region)} \\ \quad \ldots & \\ \phi_0(r) + \phi_1(r) + \phi_2(r) + \ldots \phi_l(r) & \text{for } r_l \le r \le 1 \quad \text{(Outer Region)}, \end{cases} \tag{14}$$

where $\phi_0(r) = \mu(0) + (-1)^s a_0 r + b_0 r^2 + c_0 r^3$ and $\phi_j(r) = b_j(r - r_j)^2 + c_j(r - r_j)^3$ for $j = 1, 2, \ldots, l$.

The representation is understood to be on the interval $[0,1]$, after the negative radial positions have been reflected. The linear term on $\phi_0(r)$ is intended to describe a possible, simple asymmetry. Hence $s = 0$ for, say, the positive radial positions and $s = +1$ for the negative radial positions. Clearly, $2b_j$ represents the jump of the second derivative, and $6c_j$ the jump of the third derivative at the j^{th} knot. If the spline is supposed to be twice continuously differentiable, then the coefficients b_j are constrained to be zero. We will call this a 'second order spline'. For a so-called Hermitian spline, the coefficients b_j are arbitrary. The order of continuity imposed at a knot position will be called the 'continuity index' of the knot, and the regions between the knots ('0-1', '1-2', etc.) will be called the 'knot regions'. Note that the spline coefficients $\mu(0), a_0, b_j, c_j$ for $j = 1,2,\ldots,l$ occur as linear parameters in this representation, which permits a simple direct estimation. The spline model is intended to represent only the global radial profile behaviour. Hence, the discrepancies between the observed and predicted smooth profile are to be attributed to model misspecification, experimental error and plasma profile fluctuations. The above model is quite general. Decisions have to be made about four (interrelated) aspects of a further model specification.

(1) Choosing the number of knots,

(2) choosing the continuity indices of the knots,

(3) choosing the polynomial degrees in the knot regions,

(4) choosing the knot positions.

We use the following notation (which reflects the first 3 aspects). A 2.3.2 spline model has two knots, continuity up to the first derivative at these knots, and a 2^{nd}, a 3^{rd} and a 2^{nd} degree polynomial in the 3 regions, respectively. A second order spline with the same number of knots and the same polynomial degrees, but second order continuity, is denoted by 2:3:2, etc. Note that requiring a continuity index of k for a knot between two polynomials of degree less or equal to k, corresponds to removal of the knot.

(1) The choice of the number of knots is influenced by the number of available radial measurement positions, the noise level, and the number of profiles available. Knots are needed where sudden changes in the second derivative of the profile are expected. Usually, one should have at least one measurement position between two knots to avoid collinearity problems. The lower the noise level, and, more or less equivalent, the larger the number of profiles that are simultaneously fitted, the larger the number of knots that can be used. The maximal number of knots is fixed by the requirement that the number of fitted parameters should not exceed the total number of distinct radial positions. For the 16 channel ASDEX YAG-laser measurements, 2 to 5 knot spline models have been extensively investigated [9].

(2) If one believes that diffusion must make the profiles twice differentiable, one can impose second order splines. We note that given the sharp discontinuity of the profiles

after a sawtooth crash, one should probably not assume a high degree of smoothness near the sawtooth inversion layer. Hence, requiring first order continuity near the sawtooth inversion radius, and second order continuity away from the inversion radius seems to be a plausible choice.

(3) Traditionally, one uses third degree polynomials. For regularisation, i.e. to avoid 'unphysical' wild behaviour, a boundary condition at the edge (for instance, $\mu''(1) = 0$) may be imposed. This reduces the effective degree in the outer region by 1. Second order splines with this boundary condition are historically called 'natural splines'. The following interrelationship between the various aspects of model specification is noted. If one wants to keep the same flexibility, then increasing a continuity constraint should be accompanied by increasing a near-by polynomial degree or the number of knots. If, for example, a 2.3.2 model is considered to more or less adequate, except for its discontinuity of the second derivative, then natural alternatives would be 'natural' 3:4:2, or 3:3:3:2, or 2:3:3:3:2 splines. Obviously, increasing the number of knots complicates the problem of finding the most suitable knot positions.

(4) A physical approach is to divide the tokamak into three regions, a sawtooth region, a "good confinement" region and an edge region. A natural choice would be to choose r_a near the sawtooth inversion radius. The exact sawtooth inversion radius, a regressed approximation, or the simple empirical relation $r_{inv} = 1/q_{cyl}$ may be used. The other knot positions may be chosen such that the measurement positions are roughly equally distributed in the various regions. This 'natural choice' is, however, open to the objection that the 'built-in' correlation with q_{cyl} may complicate the task of determining the full dependence of the profile shape on q_{cyl}. Another approach is to consider knot positions $r_1, r_2, \ldots$ as (nonlinear) free parameters which are estimated by numerically minimising the residual sum of squares. We have some preference for manually varying the knot positions, using the physical considerations mentioned above, and examining the sensitivity of the fit on the choice of the knots, by investigating the residuals (in particular the residual sum of squares) for various alternative fits.

An elaboration of the 2.3.2 Hermitian spline model, applied to ASDEX profile analysis, was given in [2]. In [9] more extensive investigations are made using up to 5 knot, second order splines. As we shall see in section 6, one can formally apply an F statistic to test, on the basis of n experimental profiles, whether the addition of an extra knot or, more generally, decreasing the continuity index at some radial position(s), leads to a statistically significant reduction of the residual sum of squares. The significance level of such an F test, as well as the precise fit of the final model, depend, however, on the assumed error structure.

3.2 Error structures

Combining the the mean value structure, described by (7) with the assumption of a multivariate normal distribution of the deviations, as described by (2), we write

$$\underline{Y}_j = \underline{\underline{X}}_{rad}\underline{\alpha} + \underline{E}_j, \qquad j = 1, \ldots, m, \tag{15}$$

where $\underline{E}_j \sim N_p(\underline{0}, \underline{\underline{\Sigma}})$ and $\underline{\underline{X}}_{rad}$ is given by (8). We will consider the following models with respect to the error structure:

Model I: $\underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{I}}$
Model II: $\underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{W}}_d + \underline{\underline{X}}_{rad} \underline{\underline{\Lambda}}\, \underline{\underline{X}}_{rad}^t$,
where $\underline{\underline{W}}_d$ is a known diagonal matrix, and $\underline{\underline{\Lambda}}$ is an arbitrary $p' \times p'$ covariance matrix,
Model III: $\underline{\underline{\Sigma}}$ arbitrary.

Model I is the simplest case, and corresponds to the usual assumption made in justifying the use of ordinary least-squares regression. This simple model for the error structure may not be a good approximation to reality, however.

Model III is the most general of the above three models. It has $\frac{1}{2}p(p+1) + p'$ free parameters, which may be a rather large number to estimate in practice. Hence, there is some need for a realistic model of the error structure, but with less degrees of freedom than model III.

Model II may be such a model with $\frac{1}{2}p'(p'+1) + 1 + p'$ free parameters. It has the following physical interpretation: $\sigma^2 \underline{\underline{W}}_d$ represents the independent measurement errors (arising from detector noise, etc.), whose relative magnitude can sometimes be estimated (independently from the profile data) by analysing the propagation of errors of the measurement process. The second term in Model II represents global variations of the plasma profile, which, by energy and particle conservation, cannot be independent. It is noted that II may looked upon as a random coefficient model, which can be written as

$$\underline{Y}_j = \underline{\underline{X}}_{rad}\underline{A} + \underline{E}_j, \qquad j = 1, \ldots, m, \tag{16}$$

where $\underline{A} \sim N_{p'}(\underline{\alpha}, \underline{\underline{\Lambda}})$, and $\underline{E}_j \sim N_p(\underline{0}, \sigma^2 \underline{\underline{W}}_d)$. So, correlated large-scale plasma variations are modelled by assuming that the underlying profiles vary globally, on a suitable timescale, according to an arbitrary multivariate normal distribution of the regression parameters.

Models I, II, and III are special cases of the covariance structure

$$\underline{\underline{\Sigma}} = \sum_{i=1}^{f} \theta_i \underline{\underline{G}}_i. \tag{17}$$

Estimation theory for model (14) with covariance structure (16) has been developed in [10].

The flux surface mappings, and hence the radial design matrices $\underline{\underline{X}}_{rad}$ are not precisely the same for the various profiles. Hence, a refinement of (15) is

$$\underline{Y}_j = \underline{\underline{X}}_{rad,j}\underline{\alpha} + \underline{E}_j, \qquad j = 1, \ldots, m. \tag{18}$$

As we will see in the next section, this model formulation is also suitable to describe the more general situation where the plasma profiles depend on the plasma variables.

The above described models of random errors are primarily adequate for describing statistical measurement noise. With some stretch of their original function, they may, however, also be used to describe quasi-random measurements, such as those associated with sawtooth activity. To see this, we present a simple model for the sawtooth activity inside the sawtooth inversion radius r_{inv}. We assume that the temperature is flat just after the sawtooth crash and grows linearly between the crashes. Thus $T(r,t) = T(0) + a_2(r_{inv}^2 - r^2)(\frac{t}{\tau})$ for $r \leq r_{inv}$. Assuming that the time sampling frequency is incommensurate with the sawtooth period, we have the random coefficient model

$$T(r,t) = T(0) + A_2(r_{inv}^2 - r^2), \tag{19}$$

where $A_2 \sim U(0, a_2)$, i.e. A_2 is uniformly distributed on the interval $(0, a_2)$. Clearly, $2A_2$, which can be interpreted as the random curvature, has mean a_2 and standard deviation $a_2/\sqrt{3}$. (The average curvature a_2 has the interpretation of being sawtooth amplitude in the center divided by r_{inv}^2.) Note that this sawtooth model assumes a (simple) spline for the temperature itself, instead of for the logarithm. Furthermore, the spline coefficient has a uniform instead of a normal distribution. The practical implications of these differences are likely to be small. It is noted that the sawtooth activity is confined to the inner region of the plasma. If this activity is the dominant source of plasma variation, then the matrix $\underline{\underline{\Sigma}}$ in the above discussed models will be nearly singular. In model II this is modelled by a small contribution of $\sigma^2\underline{\underline{W}}_d$ and zero elements for $\underline{\underline{\Lambda}}$, except for the row(s) and column(s) corresponding to the regression coefficient of $\mu'(r_1)$, where r_1 is near the sawtooth inversion radius.

4. PROFILE DEPENDENCE ON PLASMA PARAMETERS

4.1 Mean value and error structures

In principle, the profile shapes can be determined separately for each value of the plasma parameters $\underline{x}$. However, we are interested in parametric dependencies of the profile shapes in an entire domain of the plasma parameters. Initial statistical analyses should concentrate on relatively small domains of parameter space where one physical phenomenon is expected to be dominant. Such domains may be experimentally delineated. One speaks, for instance, of the runaway electron regime, the neo-Alcator regime, the density rollover regime, and the ELM-free H-mode regime. Theoretical parameter domains may be characterised by local instability boundaries for the η_i and ballooning modes. After these 'single phenomenon' regions have been understood, the transition regions may be more easily explored. One example of a general scaling is Goldston's expression $\tau_E^{-2} = \tau_{E,ohm}^{-2} + \tau_{E,aux}^{-2}$ for the energy confinement for both the Ohmic and the auxilary heating regime.

We now formulate approaches to determine the dependencies of the profiles on the plasma variables. Assuming that the coefficients α_j in (6) are functions of the plasma variables $\underline{x}$, we represent the profiles by $Y(r,\underline{x}) = \mu(r,\underline{x}) + E(r,\underline{x})$, where

$$\mu(r,\underline{x}) = \sum_{h=1}^{p'} \alpha_h(\underline{x}) f_h(r) = \sum_{h=1}^{p'} \sum_{k=0}^{w} \alpha_{h,k} g_k(\underline{x}) f_h(r), \qquad (20)$$

and $E(r,\underline{x})$ describes the measurement errors and plasma fluctuations. A simple, but natural choice for the plasma basis functions, at least in a single domain of parameter space where the sawtooth averaged profiles are expected to vary smoothly with the plasma parameters is $g_o(\underline{x}) = 1$ and $g_k(\underline{x}) = \ln(x_k/x_k^*)$, where x_k^* is a typical value of x_k in the database of interest ($k = 1,\ldots,w$). In a more extensive investigation one can, just as for the radial dependence, consider polynomial or spline models in $\ln(x_k/x_k^*)$.

Given m profile measurements at p radial positions, the discrete analogue of (20) can be written as

$$\underline{\underline{Y}} = \underline{\underline{X}}_{rad} \underline{\underline{\alpha}} \underline{\underline{X}}_{cov} + \underline{\underline{E}}, \qquad (21)$$

where $\underline{\underline{Y}}$ is a $p \times m$ data matrix, $\underline{\underline{E}}$ is a $p \times m$ error matrix, and $\underline{\underline{\alpha}}$ is the $p' \times w$ matrix of regression coefficients. Note that $f_h(r)$ is represented by the h^{th} column of the radial design matrix $\underline{\underline{X}}_{rad}$, and $g_k(\underline{x})$ by the k^{th} row of the 'covariate design matrix' $\underline{\underline{X}}_{cov}$.

Alternatively, the regression model can be written as

$$\underline{Y}_j = \underline{\underline{X}}_j \underline{\alpha} + \underline{E}_j, \qquad j = 1,\ldots,m \qquad (22)$$

Here, $\underline{Y}_j$ is the $p \times 1$ vector for the j^{th} profile, $\underline{\alpha}$ is the vector of regression coefficients (obtained by vertical concatination of the columns of $\underline{\underline{\alpha}}$), and $\underline{\underline{X}}_j = \underline{X}^t_{cov,j} \otimes \underline{\underline{X}}_{rad}$ is the tensor product of the transpose of the j^{th} column of $\underline{\underline{X}}_{cov}$ with $\underline{\underline{X}}_{rad}$. ($\underline{X}^t_{cov,j}$ contains typically the logarithms of plasma variables for the j^{th} profile. By definition, for any matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$, $(\underline{\underline{A}} \otimes \underline{\underline{B}})_{i,j} = A_{i,j}\underline{\underline{B}}$.)

With respect to the errors, we assume that $\underline{E}_1, \ldots, \underline{E}_m$ are independent and $\underline{E}_j \sim N_p(0, \underline{\underline{\Sigma}}_j)$. In the simplest case one has $\underline{\underline{\Sigma}}_1 = \ldots = \underline{\underline{\Sigma}}_m = \underline{\underline{\Sigma}}$ where $\underline{\underline{\Sigma}}$ satisfies model I, II, or III of the previous section. For models I and III, this assumption of 'homoskedasticity' can be tested using Bartlett's modification of the likelihood ratio test (see [10], Ch. 10).

The reader will have noticed that formula (22) has the same structure as formula (17), which described the situation with fixed plasma parameters, but different radial design matrices. Hence, the same techniques of estimation, testing and confidence intervals can be applied. The covariance structure may also depend on the plasma

4.2 Profile invariance

The concept of profile invariance can be formulated as follows. A family of plasma profiles is invariant with respect to some plasma variable if the profile shape, i.e. $L^{-1}(r, \underline{x}) = \partial / \partial r \mu(r, \underline{x})$, is independent of that plasma variable. From (20), with $g_0(\underline{x}) = f_1(r) = 1$, it follows that

$$L^{-1}(r, \underline{x}) = \sum_{h>1} \alpha_{h,0} f'_h(r) + \sum_{h>1, k>0} \alpha_{h,k} f'_h(r) g_k(\underline{x}). \tag{23}$$

Profile invariance holds for all plasma variables if the second term in this expression is 0, i.e. if all elements of $\underline{\underline{\alpha}}$ in (21), except those of the first row and the first column, vanish.

Similarly, one can consider the more general situation of profile invariance, with respect to some plasma variable (say, q_{cyl}), on an interval $(r_{min}, r_{max}) \subseteq [0,1]$. Now, the condition is $\sum_{h>1}^{p'} \alpha_{h,k} f'_h(r) = 0$ for all $r \in (r_{min}, r_{max})$ and for all k's that correspond to that plasma variable (e.g. $g_1(\underline{x}) = \log q_{cyl}$, $g_2(\underline{x}) = (\log q_{cyl})^2$, etc.). When a cubic spline model for $\mu(r, \underline{x})$ is used, this condition can be translated into the requirement that $i+2$ linear combinations of $\alpha_{2,k}, \ldots, \alpha_{p',k}$ are zero if i knot regions are needed to cover the interval (r_{min}, r_{max}). Such a hypothesis can statistically by tested as soon as the distribution of the estimate of $\underline{\underline{\alpha}}$ has been derived. This route will be undertaken in sections 5 and 6. Alternatively, the hypothesis can conveniently be tested from a global confidence band for the derivative of $L^{-1}(r, \underline{x})$ with respect to q_{cyl}, see section 7.

5. ESTIMATION OF REGRESSION COEFFICIENTS

5.1 Least squares and maximum likelihood

Initially, we consider only estimates for the model $\underline{Y}_j = \underline{\underline{X}}_{rad}\underline{\alpha} + \underline{E}_j, j = 1, \ldots, m$, where $\underline{\underline{X}}_{rad}$ is a *fixed* $p \times p'$ radial design matrix, and $\underline{E}_j \sim N_p(\underline{0}, \underline{\underline{\Sigma}})$ satisfies one of the error structures I, II, III, described in section 3. Then, we make some remarks on estimates for the more general situation, described in section 4, of different radial design matrices.

Let $\underline{\hat{\mu}} = m^{-1}\sum_j \underline{Y}_j$ denote the empirical average of the m profiles. It can easily be derived that for any $p \times p$ non-singular matrix, $\underline{\underline{G}}$ such that $\underline{\underline{X}}_{rad}^t \underline{\underline{G}}\, \underline{\underline{X}}_{rad}$ is also non-singular,

$$\underline{\hat{\alpha}} = \underline{\underline{Q}}\, \underline{\hat{\mu}}, \tag{24}$$

with

$$\underline{\underline{Q}} = \left(\underline{\underline{X}}_{rad}^t \underline{\underline{G}}^{-1} \underline{\underline{X}}_{rad}\right)^{-1} \underline{\underline{X}}_{rad}^t \underline{\underline{G}}^{-1} \tag{25}$$

is a normally distributed, linear, unbiased (i.e. $E(\underline{\hat{\alpha}}) = \underline{\alpha}$) estimator for $\underline{\alpha}$. The covariance matrix of $\underline{\hat{\alpha}}$ is

$$\underline{\underline{V}}(\underline{\hat{\alpha}}) = \frac{1}{m} \underline{\underline{Q}}\, \underline{\underline{\Sigma}}\, \underline{\underline{Q}}^t. \tag{26}$$

Such estimators for $\underline{\alpha}$ are called generalised least squares estimators. Notice that the choice $\underline{\underline{G}} = c\underline{\underline{I}}$ corresponds to ordinary least squares regression, and $\underline{\underline{G}}$ diagonal to weighted least squares regression, the weights being inversely proportional to the diagonal elements of $\underline{\underline{G}}$. Equations (24) and (25) can, for symmetric and positive definite G, be interpreted by the fact that the vector $\underline{\hat{\mu}} = \underline{\underline{X}}_{rad}\underline{\hat{\alpha}}$ is the projection of $\underline{\hat{\mu}}$ on the linear subspace generated by the columns of $\underline{\underline{X}}_{rad}$, using the inner product defined by $\underline{\underline{G}}$.

If $\underline{\underline{\Sigma}}$ is known, then among all possible matrices $\underline{\underline{G}}$, the optimal choice is $\underline{\underline{G}} = \underline{\underline{\Sigma}}$, since it gives the estimator that maximises the likelihood and has minimal covariance matrix among *all unbiased* estimators for $\underline{\alpha}$. The variance of this estimator reduces to

$$\underline{\underline{V}}(\underline{\hat{\alpha}}) = \frac{1}{m}\left(\underline{\underline{X}}_{rad}^t \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}}_{rad}\right)^{-1}. \tag{27}$$

In practice, $\underline{\underline{\Sigma}}$ is unknown and must be estimated as well. The 'best' estimator for $\underline{\underline{\Sigma}}$, and the simplest way to calculate this estimator depend on the assumed error structure.

Model I, corresponding to ordinary least squares regression, is the simplest case. Estimates for σ^2 are based on the residual sum of squares,

$$\hat{\sigma}^2 = \frac{1}{mp - p'} \sum_{j=1}^{m} (\underline{Y}_j - \underline{\underline{X}}_{rad}\underline{\hat{\alpha}}_I)^t (\underline{Y}_j - \underline{\underline{X}}_{rad}\underline{\hat{\alpha}}_I) \tag{28}$$

being the minimum variance unbiased estimator for σ^2, which is distributed as $\sigma^2/(mp-p')$ times $\chi^2_{mp-p'}$.

In model III an iterative procedure might seem to be needed, to solve simultaneously

$$\underline{\underline{\Sigma}} = \frac{1}{m-1} \sum_{j=1}^{m} (\underline{Y}_j - \underline{\underline{X}}_{rad}\underline{\alpha})(\underline{Y}_j - \underline{\underline{X}}_{rad}\underline{\alpha})^t \tag{29}$$

and (24–25), starting for example with the unweigthed least squares estimate for $\underline{\alpha}$. It can be proven, however, see e.g. [11] [12], that one gets directly the ML estimate $\underline{\hat{\alpha}}_{III}$ by inserting for $\underline{\underline{G}}$ in (25) the estimated variance in the unsmoothed model, i.e. $\underline{\underline{\hat{\Sigma}}} = m^{-1}\underline{\underline{S}}$ as given by (4). The adjusted ML estimate of $\underline{\underline{\Sigma}}$ in the smoothed model is then obtained by inserting $\underline{\hat{\alpha}}_{III}$ into (29). (Adjusted, because in the denominator $m-1$ has been used instead of m.)

An elegant generalisation exists for the situation with covariates, as expressed by (21). As shown in [11] [13], among others,

$$\underline{\hat{\alpha}} = (\underline{\underline{X}}_{rad}\underline{\underline{\tilde{\Sigma}}}^{-1}\underline{\underline{X}}^t_{rad})^{-1}\underline{\underline{X}}^t_{rad}\underline{\underline{\tilde{\Sigma}}}^{-1}\underline{\underline{Y}}\,\underline{\underline{X}}^t_{cov}(\underline{\underline{X}}_{cov}\underline{\underline{X}}^t_{cov})^{-1}, \tag{30}$$

with

$$\underline{\underline{\tilde{\Sigma}}} = f^{-1}\underline{\underline{Y}}\,(\underline{\underline{I}} - \underline{\underline{X}}^t_{cov}(\underline{\underline{X}}_{cov}\underline{\underline{X}}^t_{cov})^{-1}\underline{\underline{X}}_{cov})\underline{\underline{Y}}^t, \tag{31}$$

and $f = m - w - 1 - (p - p')$, is in that case the maximum likelihood estimator (the constant f was chosen to simplify the next formula). An unbiased estimate of its covariance matrix is given by [13]

$$\underline{\underline{\hat{V}}}\,(\underline{\hat{\alpha}}) = (\underline{\underline{X}}_{cov}\underline{\underline{X}}^t_{cov})^{-1} \otimes (\underline{\underline{X}}^t_{rad}\underline{\underline{\tilde{\Sigma}}}^{-1}\underline{\underline{X}}_{rad})^{-1}(m - w - 2)/(m - w - 2 - (p - p')), \tag{32}$$

where $\underline{\alpha}$ is the vertical concatination of the columns of $\underline{\alpha}$, and $\otimes$ denotes the tensor product.

The maximum likelihood estimates are asymptotically (i.e., as $m \to \infty$) normal and efficient. It is noted that $\underline{\underline{\hat{\Sigma}}}$ may deviate considerably from $\underline{\underline{\Sigma}}$, if the total number, mp, of observations is not much larger than the total number of free parameters ($\frac{1}{2}p(p+1)+p$ in the unsmoothed model). This leads, then, to an inefficient estimate for $\underline{\alpha}$ which is, in addition, not normally distributed. In such a case, it may be wise to use a model with fewer parameters.

We now consider model II in the simple case of a fixed design matrix with no covariates. If its special covariance structure is inserted for $\underline{\underline{G}}$, one can derive [14, 15, 16, 17] that (24–26) reduce to

$$\hat{\underline{\alpha}}_{II} = \left(\underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad}\right)^{-1} \underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \hat{\underline{\mu}} \tag{33}$$

and

$$\underline{\underline{V}}\,(\hat{\underline{\alpha}}_{II}) = \frac{1}{m}\left(\sigma^2 \left(\underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad}\right)^{-1} + \underline{\underline{\Lambda}}\right). \tag{34}$$

Note that the estimator $\hat{\underline{\alpha}}_{II}$ does not depend at all on the parameters $\underline{\underline{\Lambda}}$ and σ^2. Of course, the covariance matrix of $\hat{\underline{\alpha}}_{II}$ depends on these parameters, but it does so in a simple way. Equation (33) requires the inversion of a sometimes considerably smaller matrix than in the matrix $\underline{\underline{G}}$ in (24) and (25).

In Model II, (33) was constructed to be the minimum variance unbiased estimator for $\underline{\alpha}$. We now have to find an estimator for its variance $\underline{\underline{V}}\,(\hat{\underline{\alpha}}_{II})$. It can be derived (see e.g. [18, 19] that, if $\underline{\underline{W}}_d$ is known,

$$\hat{\underline{\underline{\Sigma}}}_{II} = \frac{1}{m-1} \sum_{j=1}^{m} (\underline{Y}_j - \underline{\underline{X}}_{rad} \hat{\underline{\alpha}}_{II})(\underline{Y}_j - \underline{\underline{X}}_{rad} \hat{\underline{\alpha}}_{II})^t \tag{35}$$

is the minimum variance unbiased estimator for $\underline{\underline{\Sigma}}$. Inserting $\hat{\underline{\underline{\Sigma}}}_{II}$ and

$$\underline{\underline{Q}}_{II} = \left(\underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad}\right)^{-1} \underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1}, \tag{36}$$

into (26), one gets the minimum variance unbiased estimator for $\underline{\underline{V}}\,(\hat{\underline{\alpha}}_{II})$, which will be denoted by $\hat{\underline{\underline{V}}}\,(\hat{\underline{\alpha}}_{II})$.

A particularly nice feature is that $\hat{\underline{\alpha}}_{II}$ can be looked upon as the sample mean of the estimated regression coefficients of the individual profile fits, and that $\hat{\underline{\underline{V}}}\,(\hat{\underline{\alpha}}_{II})$ can be rewritten as the empirical covariance matrix of this sample mean, i.e.,

$$\hat{\underline{\alpha}}_{II} = \frac{1}{m} \sum_{j=1}^{m} \hat{\underline{\alpha}}_{jII}, \tag{37}$$

and

$$\hat{\underline{\underline{V}}}\,(\hat{\underline{\alpha}}_{II}) = \frac{1}{m(m-1)} \sum_{j=1}^{m} (\hat{\underline{\alpha}}_{jII} - \hat{\underline{\alpha}}_{II})(\hat{\underline{\alpha}}_{jII} - \hat{\underline{\alpha}}_{II})^t, \tag{38}$$

where $\hat{\underline{\alpha}}_{jII} = \underline{\underline{Q}}_{II} \underline{Y}_j$.

Notice that one can construct separate estimators for σ^2 and $\underline{\underline{\Lambda}}$. The usual estimator for σ^2 equals the weighted sample average of the squared residuals from the fitted individual profiles, with a correction for the fact that mp' parameters are estimated, i.e.

$$\hat{\sigma}^2 = \frac{1}{m(p-p')} \sum_{j=1}^{m} (\underline{Y}_j - \underline{\underline{X}}_{rad} \hat{\underline{\alpha}}_{jII})^t \underline{\underline{W}}_d^{-1} (\underline{Y}_j - \underline{\underline{X}}_{rad} \hat{\underline{\alpha}}_{jII}). \tag{39}$$

Subsequently, $\underline{\underline{\Lambda}}$ may be estimated by the relation

$$\hat{\sigma}^2 \left(\underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad}\right)^{-1} + \hat{\underline{\underline{\Lambda}}} = m\hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}_{II}), \tag{40}$$

which follows directly from (34). This estimator for $\underline{\underline{\Lambda}}$ has the disadvantage that, because of the subtraction, it may not be positive definite. If this occurs, one can consider (1) to reformulate the model, assuming that at least some elements of $\underline{\underline{A}}$ are non-random, and (2) to estimate $\underline{\underline{\Lambda}}$ and σ^2 numerically by maximum likelihood. It is stressed that in order for (33) and (38) to be sensible estimators, the weights associated with $\underline{\underline{W}}_d$ must be known, or at least be independently estimable.

Being the UMVU estimate in a linear model, $\hat{\underline{\alpha}}_{II}$ has the property that for every $\underline{f}(r)$, $\hat{\underline{\alpha}}_{II}^t \underline{f}(r)$ is the minimum variance unbiased estimate for $\underline{\alpha}^t \underline{f}(r)$. This property is useful for predicting future profiles, see section 7. If $\underline{\underline{Q}}_w = \underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad}$ is rather close to a singular matrix, the restriction to unbiasedness leads to large variances, however, and one may be inclined to allow for some bias in order to get a substantial reduction in variance. Procedures to do this are ridge regression, which replaces $\underline{\underline{Q}}_w$ by $\underline{\underline{Q}}_w + \underline{\underline{R}}$ for some positive definite $\underline{\underline{R}}$ (frequently, $\underline{\underline{R}} = \lambda \underline{\underline{I}}$), and latent root regression [20] which discards small principal components of $\underline{\underline{Q}}_w$ that are, in addition, uninformative for the regression. In the context of model II, empirical Bayes estimates have been developed [21], see also [22], that minimise the mean squared error of prediction (i.e. the variance plus the square of the bias), summed over the profiles in the data base. Here, we will restrict attention to ML and UMVU estimates, which is reasonable as long as the matrix $\underline{\underline{Q}}_w$ is rather well conditioned.

Now we will discuss estimation procedures for the general case of unequal design matrices, where

$$\underline{Y}_j = \underline{\underline{X}}_j \underline{\alpha} + \underline{E}_j, j = 1, \ldots, m, \tag{41}$$

and $\underline{E}_{j=1}^m \sim N_p(\underline{0}, \underline{\underline{\Sigma}}_j)$. For any set of non-singular matrices $\underline{\underline{G}}_1, \ldots, \underline{\underline{G}}_m$, such that $\sum_j \underline{\underline{X}}_j^t \underline{\underline{G}}_j^{-1} \underline{\underline{X}}_j$ is invertible, an unbiased estimator for $\underline{\alpha}$ is given by

$$\hat{\underline{\alpha}} = \left(\sum_{j=1}^m \underline{\underline{X}}_j^t \underline{\underline{G}}_j^{-1} \underline{\underline{X}}_j\right)^{-1} \sum_j \underline{\underline{X}}_j^t \underline{\underline{G}}_j^{-1} \underline{Y}_j \tag{42}$$

This property holds independent of the error structure of $\underline{E}_j$, the only condition being that $\underline{E}_j$ has expectation zero for $j = 1, \ldots, m$. Note that, although each $\underline{\underline{X}}_j^t \underline{\underline{G}}_j^{-1} \underline{\underline{X}}_j$ may be singular, $\hat{\underline{\alpha}}$ can formally be regarded as a weighted average of the individual least squares estimates $\hat{\underline{\alpha}}_1, \ldots, \hat{\underline{\alpha}}_m$, where each $\hat{\underline{\alpha}}_j$ is weighted according to $\underline{\underline{X}}_j^t \underline{\underline{G}}_j^{-1} \underline{\underline{X}}_j$.

For known $\underline{\underline{\Sigma}}_1, \ldots, \underline{\underline{\Sigma}}_m$, the choice $\underline{\underline{G}}_j = \underline{\underline{\Sigma}}_j$ leads to the minimum variance unbiased estimator, which in this case coincides with the maximum likelihood (ML) estimator.

Furthermore, in that case, $\underline{\underline{X}}_j^t \underline{\underline{\Sigma}}_j^{-1} \underline{\underline{X}}_j$ equals the inverse of the covariance matrix of $\hat{\underline{\alpha}}_j$, and

$$\underline{\underline{V}}(\hat{\underline{\alpha}}) = (\sum_j \underline{\underline{X}}_j^t \underline{\underline{\Sigma}}_j^{-1} \underline{\underline{X}}_j)^{-1}. \tag{43}$$

In the random coefficient model, i.e. if $\underline{\underline{\Sigma}}_j = \sigma^2 + \underline{\underline{X}}_j^t \underline{\underline{\Lambda}} \underline{\underline{X}}_j$, insertion of $\underline{\underline{G}}_j = \underline{\underline{\Sigma}}_j$ in (42) leads to

$$\hat{\underline{\alpha}}_{II} = (\sum_j \underline{\underline{H}}_j)^{-1} \sum_j \underline{\underline{H}}_j \hat{\underline{\alpha}}_{(j)}, \tag{44}$$

where $\hat{\underline{\alpha}}_j = \left(\underline{\underline{X}}_j^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_j\right)^{-1} \underline{\underline{X}}_j^t \underline{\underline{W}}_d^{-1} \underline{Y}_j$ and $\underline{\underline{H}}_j^{-1} = \underline{\underline{\Lambda}} + \left(\underline{\underline{X}}_j^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_j\right)^{-1} \sigma^2$. Obviously, now, $\underline{\underline{V}}(\hat{\underline{\alpha}}_{II}) = (\sum_j \underline{\underline{H}}_j)^{-1}$. Note that for $\underline{\underline{\Lambda}} = 0$, the individual estimates $\hat{\underline{\alpha}}_{(j)}$ are averaged with weights $\underline{\underline{X}}_j^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_j$, whereas for $\underline{\underline{\Lambda}} \to \infty$ they are averaged with equal weights. For equal design matrices (44) reduces to (37) for any $\underline{\underline{\Lambda}}$. In general, $\hat{\underline{\alpha}}_{II}$ depends on $\underline{\underline{\Lambda}}$ and another equation is needed to estimate $\underline{\underline{\Lambda}}$. One possibility is to consider

$$\hat{\sigma}^2 \sum_j \left(\underline{\underline{X}}_j^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_j\right)^{-1} + m\hat{\underline{\underline{\Lambda}}} = \sum_j (\hat{\underline{\alpha}}_{jII} - \hat{\underline{\alpha}}_{II})(\hat{\underline{\alpha}}_{jII} - \hat{\underline{\alpha}}_{II})^t, \tag{45}$$

which is the analogue of (40). Obviously, σ^2 is estimated from the residual sum of squares from the individual regressions. Iterative solution of (44) and (45), starting e.g. with $\underline{\underline{\Lambda}} = 0$, gives consistent estimates of both $\underline{\alpha}$ and $\underline{\underline{\Lambda}}$. These estimates can relatively easily be calculated, but they may not be 'efficient', i.e. they may not have asymptotically minimal variance, as $\underline{\underline{\Lambda}}$ is not estimated by maximum likelihood.

We will discuss the method of maximum likelihood in a more general formulation, which contains the random coefficient model as special case. We assume that the covariance matrices are parameterised by a finite dimensional parameter $\underline{\theta}$, and for brevity we write $\underline{\underline{\Sigma}}(\underline{\theta})$ for $(\underline{\underline{\Sigma}}_1(\underline{\theta}), \ldots, \underline{\underline{\Sigma}}_m(\underline{\theta}))$. The log likelihood of the observations $\underline{Y}_1, \ldots, \underline{Y}_m$ is

$$l(\underline{\alpha}, \underline{\underline{\Sigma}}(\underline{\theta})) = C(m,p) - \frac{1}{2}(\sum_{j=1}^m \log \left|\underline{\underline{\Sigma}}_j\right| + \sum_{j=1}^m (\underline{Y}_j - \underline{\underline{X}}_j \underline{\alpha})^t \underline{\underline{\Sigma}}_j^{-1} (\underline{Y}_j - \underline{\underline{X}}_j \underline{\alpha}) \tag{46}$$

The likelihood equations are

$$(\partial/\partial \underline{\alpha}, \partial/\partial \underline{\theta})\, l(\underline{\alpha}, \underline{\underline{\Sigma}}(\underline{\theta})) = (\underline{0}, \underline{0}). \tag{47}$$

The maximum likelihood estimates are those values of $\underline{\alpha}$ and $\underline{\theta}$ that maximise (46). If the maximum does not occur at the boundary of the parameter region, then (47) constitutes a necessary condition. If the log-likelihood is a (strictly) concave function of $(\underline{\alpha}, \underline{\theta})$, then (47) is a sufficient condition for the (unique) maximum. Concavity is not

always easy to demonstrate in practice, however. For a large number of parameters, it pays to look for an efficient computational algorithm [23, 24, 25].

As an example, we present the computational equations for the still rather general special case that each $\underline{\underline{\Sigma}}_j = \sum_{i=1}^{f} \theta_i \underline{\underline{G}}_{ij}$ for $j = 1, \ldots, m$. The derivative of $l(\alpha, \underline{\underline{\Sigma}}(\underline{\theta}))$ with respect to $\underline{\alpha}$ gives

$$\sum_j \underline{\underline{X}}_j^t \underline{\underline{\Sigma}}_j^{-1} (\underline{Y}_j - \underline{\underline{X}}_j \underline{\alpha}) = \underline{0}, \tag{48}$$

whose solution corresponds precisely to (29) with $\underline{\underline{G}}_j = \underline{\underline{\Sigma}}_j$. Derivation with respect to $\underline{\theta}_i$ gives [25, 19]

$$\sum_{j=1}^{m} tr \underline{\underline{\Sigma}}_j^{-1} \underline{\underline{G}}_{ij} (I - \underline{\underline{\Sigma}}_j^{-1} \underline{e}_j \underline{e}_j^t) = 0, \qquad i = 1, \ldots, f, \tag{49}$$

where $\underline{\underline{\Sigma}}_j = \sum_{i=1}^{f} \theta_i \underline{\underline{G}}_{ij}$ and $\underline{e}_j = \underline{Y}_j - \underline{\underline{X}}_j \underline{\alpha}$. Equations (48) and (49) are to be solved iteratively. The covariance matrix of the ML estimator $(\underline{\hat{\alpha}}, \underline{\hat{\theta}})$ is estimated by the negative of the inverse of the matrix of second derivatives of the log likelihood at the maximum likelihood solution $(\underline{\hat{\alpha}}, \underline{\hat{\theta}})$.

The maximum likelihood estimator tends to be rather sensitive to the validity of the distributional assumption (multivariate normality) for the errors. Sometimes, it is more robust to iterate the likelihood equations only a few times using the ordinary least squares estimates as starting values. In [25] three algorithms are discussed to solve (48) and (49) numerically. They are implemented in the program BMDP5V [26]. Under certain conditions, asymptotic normality and consistency has been proven [19], even for estimators obtained by the first few iteration steps of (48) and (49).

5.2 Robust estimation

The idea of robust statistics is to use estimators and test procedures that are insensitive to (moderately large) deviations from the probabilistic assumptions. They tend to give better fits to the *bulk* of the data than classical procedures do, and hence robust methods are also suitable for outlier detection. Here, we restrict attention to robust estimation of the mean value structure (i.e., the parameter $\underline{\alpha}$), under the assumption $\underline{\underline{\Sigma}}_1 = \cdots = \underline{\underline{\Sigma}}_m = \underline{\underline{\Sigma}}(\underline{\theta})$. where $\underline{\underline{\Sigma}}(\underline{\theta})$ is assumed either to be known, or independently estimable. We will discuss the multivariate situation of outlying profiles from the regression on the plasma parameters, rather than the somewhat simpler univariate case of single outlying measurement points (channels) from individual profile fits.

Generalised least squares estimation minimises $-l(\underline{\alpha}, \underline{\underline{G}})$ from (46) as a function of $\underline{\alpha}$ for some (symmetric, positive definite) matrix $\underline{\underline{G}}$. In practice, one has to insert some sensible estimate of $\underline{\underline{\Sigma}}(\underline{\theta})$ for G. A generalisation of this procedure (see e.g. [27], [28]) is to estimate $\underline{\alpha}$ by minimising

$$\sum_{j=1}^{m} \rho\left((\underline{Y}_j - \underline{\underline{X}}_j\underline{\alpha})^t \underline{\underline{G}}_j^{-1}(\underline{Y}_j - \underline{\underline{X}}_j\underline{\alpha})\right) \tag{50}$$

for some suitable, non-decreasing function ρ, and symmetric, positive definite matrices $\underline{\underline{G}}_1, \ldots, \underline{\underline{G}}_m$. Such estimates are called M-estimates (from 'Minimisation'/'Maximisation').

Remarks: (1) Note that (50) permits arbitrarily weighted regression. Choosing $\underline{\underline{G}}_j = \underline{\underline{W}}_j\underline{\underline{\Sigma}}\underline{\underline{W}}_j^t$, with $\underline{\underline{W}}_j$ diagonal and positive, corresponds to assigning the diagonal elements of $\underline{\underline{W}}_j$ as weights for the j^{th} profile, *in addition to* the weighting imposed by $\underline{\underline{\Sigma}}$, $j = 1, \ldots, m$. These weights can be used to used to robustify the estimation procedure, at the cost of some loss of efficiency in case the multivariate normal error distribution happens to be a good approximation. (2) Equation (50) can be viewed as a log likelihood equation for a very specific probability distribution. More importantly, it has been derived (see [27, 29]), that for symmetric error distributions, in a number of situations, solution of (50) yields consistent and asymptotically normal parameter estimates.

The quantity $D_j = (\underline{Y}_j - \underline{\underline{X}}_j\underline{\hat{\alpha}})^t \underline{\underline{G}}_j^{-1}(\underline{Y}_j - \underline{\underline{X}}_j\underline{\hat{\alpha}})$ can be interpreted as a squared residual, i.e. the squared distance between the observed and the fitted j^{th} profile, in the metric defined by $\underline{\underline{G}}_j$. A large value of D_j indicates an outlying profile. (If m profiles are tested simultaneously, the cutt-off point for outlier detection should be based on the distribution function of $\max_{j=1}^{n} D_j$, rather than on that of D_j.) An outlying profile is a candidate for a bad measurement, and should be checked on physical grounds. A weak point of least squares is that large *actual* residuals (i.e. deviations from the actual postulated model) that are, in addition, influential (i.e. have a large influence on the regression fit, because they are in plasma parameter space far away from the bulk of the datset), can distort the regression fit in such a way, that the corresponding *fitted* residuals are quite moderate. Hence, such points go undetected as outliers. One way out is to calculate the j^{th} residual from a dataset from which just the j^{th} datapoint has been excluded. (This can efficiently be done, see e.g. [30].) Outlier detection is an important part in practical regression. Justified deletion of outliers is an important step in robustifying a least squares regression fit. Another expedient is to use robust (e.g. M-type) estimates that automatically downweight influential data points [28].

Least squares corresponds to $\rho(u) = u$, and is rather sensitive to outliers, as the residuals from the fit are squared. In robust regression, sensitivity to outliers is diminished

by using functions for ρ that grow less strongly than linearly for large values of the residual. We present some common examples. The mean absolute deviation (MAD) estimator corresponds to $\rho(u) = \sqrt{u}$. To avoid uniqueness problems, the absolute deviation norm may be replaced by $\rho(u) = \|u\|^r$ for some $0.5 < r < 0.7$, say. The 'Huber estimator' is defined by $\psi(u) = \rho'(u) = \sqrt{(u)}$ for $u < cst$ and $\psi(u) = \sqrt{(cst)}$ for $u > cst$. For cst between 0 and ∞, it ranges between the mean absolute deviation and the least squares estimator. It turns out that this estimator is 'optimally robust' from two points of view, see [28], Ch. 2.6 and 2.7, under the assumption that only the response variable is contaminated with outliers. A problem is the choice of cst. The optimal choice depends on the (unkown) amount of contamination. For practical purposes, it seems sensible to choose cst near $\chi^2_{(p-p');0.05/m}$, which is the classical cut-off from normal least squares. Finally, one can apply iteratively reweighted least squares, downweighting the profiles with large values of D_j (by choosing e.g. the weights $\underline{\underline{W}}_j = w_j \underline{\underline{I}}$, with w_j inversely proportional to D_j). It is easily shown that these are, in fact, M-estimates in a slightly disguised form [28].

The need of robust statistics depends on the quality of the dataset, and on the amount of effort one is willing to spend on residual analysis. Broadly speaking, for 'high quality' datasets, standard least squares is all right. For 'medium quality' datasets, standard least squares is only justified if one spends considerable effort to detect and repair (wherever justified) any outlying data. Robust estimation may then be useful for better pinpointing these outliers. For 'low quality' datasets, robust estimation is indispensible for fitting the bulk of the data and for outlier identification. In general, solution of the estimating equations and determination of the covariance matrix of the regression estimates is more difficult for robust estimation than it is for least squares. The reader is referred to [28] for a comprehensive survey of the available methods, and for reference to some computer programs.

6. MODEL TESTING

We present some statistical tests for testing the null-hypothesis that a given model for the observations is correct. The general procedure may heuristically be described as follows. First, the parameters for the model to be tested, say M_2, are estimated, as are the parameters for a more comprehensive, competing, model M_1, which contains model M_2 as a special case. Then the distance between these two sets of parameters, suitably normed, is calculated. Under the null-hypothesis that model M_2 holds, this distance has some probability distribution. The characteristics of this distribution,

or of an asymptotic approximation thereof, are determined. If the observed value of this distance is in the tail of the distribution, then the null-hypothesis is considered to be implausible. If one rejects at the 5% level, then the probability to reject the null-hypothesis incorrectly (when in fact it is true) is at most 5%. The probability to make an error of the second kind, i.e. to accept the null-hypothesis when in fact it is false, depends of course on how much the true underlying model differs from the hypothesised model M_2. Let us denote this probability by $\beta(\underline{\alpha}_{12})$, where $\underline{\alpha}_{12} \in V_{12}$ denote the parameters in model M_1 *in excess* of those of model M_1, i.e. the parameterisation is such that the origin of the space V_{12} corresponds to model M_2. Broadly speaking, the tests to be discussed have the property that in comparison with other tests $\beta(\underline{\alpha}_{12})$ is *relatively low* in *every direction* of the parameter space V_{12}. Statistically, the corresponding optimality criterion, which constitutes a practical compromise, is called 'asymptotically most stringent, against unrestricted alternatives' [31, 32, 33]. In the following, M_1 and M_2 are identifiers that indicate in each situation the more comprehensive model and the more restricted model, respectively. Which specific models are chosen for them, changes with the context.

6.1 Discrete versus continuous profile representations

As in section 3, we assume that, for fixed values of the plasma parameters, the available data consist of a single sample of m profile measurements $\underline{Y}_1, \ldots, \underline{Y}_m$, each at p radial positions. For simplicity we assume that the radial positions correspond to constant flux-surface radii. As null-hypothesis we consider a continuous (e.g. spline) model with p' free, linear regression parameters and with a general $p \times p$ covariance matrix $\underline{\underline{\Sigma}}$ (i.e., (14) with covariance model III in section 3). The more comprehensive model M_1 is the unsmoothed model $\underline{Y}_j = N_p(\underline{\mu}, \underline{\underline{\Sigma}})$, where $\underline{\mu}$ denotes the unknown underlying profile vector. Let $\hat{\underline{\underline{\Sigma}}}$ be the ML estimator of the covariance matrix under model M_1, see (4), and $\hat{\underline{\underline{\Sigma}}}_{III}$ the ML estimator under model M_2, see (29). The statistical test is based on the following result, see e.g. [12] and [34], Ch.5.3, A.2.3: If the null-hypothesis holds, and $m > p - p'$, then

$$T_1 = \frac{\left|\hat{\underline{\underline{\Sigma}}}_{III}\right| - \left|\hat{\underline{\underline{\Sigma}}}\right|}{\left|\hat{\underline{\underline{\Sigma}}}\right|} = \left(\hat{\underline{\mu}} - \underline{\underline{X}}_{rad}\hat{\underline{\alpha}}_{III}\right)^t \hat{\underline{\underline{\Sigma}}}^{-1} \left(\hat{\underline{\mu}} - \underline{\underline{X}}_{rad}\hat{\underline{\alpha}}_{III}\right) \tag{51}$$

is distributed as

$$\frac{p - p'}{m - p + p'} F(p - p', m - p + p'). \tag{52}$$

For large values of T_1 the continuous model will be rejected. The test statistic T_1 can be interpreted in two ways:

(a) The right-hand side gives the distance, in $p-$dimensional space, between the estimated underlying profile under model M_1 and under model M_2, in the metric defined by the (estimated) residual covariance matrix under the more comprehensive model M_1.
(b) The determinant of a covariance matrix is a scalar measure of the total residual variation associated with a given model. (This determinant is sometimes called the generalised variance.) Hence, the statistic T estimates the fractional increase in generalised variance in going from M_1 to model M_2.

6.2 Different covariance structures

If a given smoothed profile model with general covariance matrix $\underline{\underline{\Sigma}}$ is not rejected using the previous test, then we can proceed to test the more parsimonious covariance structure II within the smoothed model, i.e. $M_1 : \underline{\underline{\Sigma}} = \underline{\underline{\Sigma}}_{II}$ is tested against $M_2 : \underline{\underline{\Sigma}} = \underline{\underline{\Sigma}}_{III}$. One can do this by applying a large-sample likelihood ratio test. The likelihood ratio statistic is

$$T_2 = -2 \ln \frac{\max_{M1} L(\underline{\alpha}, \underline{\underline{\Sigma}})}{\max_{M2} L(\underline{\alpha}, \underline{\underline{\Sigma}})}, \tag{53}$$

where L(.,.) stands for the likelihood function, which has to be maximised under model M_1 and M_2, respectively. By writing down the multivariate normal densities, one derives that

$$-2 \ln L(\hat{\underline{\alpha}}_{III}, \hat{\underline{\underline{\Sigma}}}_{III}) = m(\ln \left| 2\pi \hat{\underline{\underline{\Sigma}}}_{III} \right| + p), \tag{54}$$

where $\hat{\underline{\alpha}}_{III}$ and $\hat{\underline{\underline{\Sigma}}}_{III}$ are the ML estimates under Model 2, see (29). For the smoothed model M_1, the likelihood has to be maximised numerically, using e.g. the program BMDP5V. Assuming that some regularity conditions are satisfied, it follows from general theory developed by Wilks [35] and Wald [31] (see e.g. [36], Ch. 4 and [37], Ch. 6e) that, asymptotically,

$$T_2 \sim \chi^2_{\frac{1}{2}p(p+1) - \frac{1}{2}p'(p'+1) - 1}. \tag{55}$$

In practice, it is convenient to apply an asymptotically equivalent version of T_2, which is obtained by inserting the UMVU instead of the ML estimates of $\underline{\underline{\Sigma}}$ into the likelihood ratio (53). In that case, we get, after some simplification,

$$\tilde{T}_2 = m \ln \frac{\left| \underline{\underline{S}}_{II} \right|}{\left| \underline{\underline{S}}_{III} \right|}, \tag{56}$$

where $\underline{\underline{S}}_{III} = m \hat{\underline{\underline{\Sigma}}}_{III}$ and $\underline{\underline{S}}_{II} = (m-1) \hat{\underline{\underline{\Sigma}}}_{II}$, see (35). For large m, $\tilde{T}_2$ has the same distribution as T_2. As in (51), a computationally more convenient form of the ratio of

the generalised variances is

$$\frac{\left|\underline{\underline{S}}_{II}\right|}{\left|\underline{\underline{S}}_{III}\right|} = 1 + m(\hat{\underline{\alpha}}_{II} - \hat{\underline{\alpha}}_{III})^t \underline{\underline{X}}_{rad}^t \underline{\underline{S}}_{III}^{-1} \underline{\underline{X}}_{rad} (\hat{\underline{\alpha}}_{II} - \hat{\underline{\alpha}}_{III}). \tag{57}$$

It should be mentioned that test statistics based on other combinations of the eigenvalues of $\underline{\underline{S}}_{II}$ and $\underline{\underline{S}}_{III}$ than those implied by (56) are in use, see e.g. [38], Ch.5. For reasonable sample sizes, the practical differences between these multivariate test statistics tend to be small. Special care is required if the (unconstrained) ML estimates of the parameters under Model 2 lead to an estimate of $\underline{\underline{\Sigma}}$ which is not positive definite [17, 39, 40].

6.3 Different continuous profile representations

Although the theory presented holds for general linear profile representations, we will, for concreteness, orient the discussion towards spline representations. Suppose, we have a Hermitian spline model with p' free parameters for the mean value structure and with one of the errors structures I, II, III. We want to test the null-hypothesis that a spline sub-model holds. A spline sub-model is defined as a spline model where, with respect to the original model, the continuity index (see section 3.1.3) of some of the knots has been increased. (Recall that imposing third order continuity for a third degree spline amounts to removing the knot.) To test such a null-hypothesis, which can be stated as linear restrictions on the parameters, there exist standard F-tests, especially in the case of the error structures I and III, see e.g. [41, 34, 37]. Here, we will directly discuss the more relevant and interesting case of error structure II. We consider the general model $M_1 : \underline{Y}_j = \underline{\underline{X}}_{rad}\underline{A} + \underline{E}_j$, where $\underline{E}_j \sim N_p(0, \sigma^2 \underline{\underline{W}}_d)$ and $\underline{A} \sim N_{p'}(\underline{\alpha}, \underline{\underline{\Lambda}}_1)$ have a multivariate normal distribution. Within this model, we test the sub-model $M_2 : A \sim N_{p''}(\underline{\underline{D}}\,\underline{\beta}, \underline{\underline{D}}\,\underline{\underline{\Lambda}}_2\underline{\underline{D}}^t)$ for some $p' \times p''$ matrix D.

We again use the likelihoood ratio as an asymptotic test statistic. Consider the random variable

$$T_3 = m \ln \frac{\left|\underline{\underline{S}}_{II,2}\right|}{\left|\underline{\underline{S}}_{II,1}\right|}, \tag{58}$$

where $\underline{\underline{S}}_{II,2}$ and $\underline{\underline{S}}_{II,1}$ are the the residual SSCP matrices for the restricted model and for the general model, respectively. If the null-hypothesis holds, then asymptotically (i.e. as $m \to \infty$),

$$T_3 \sim \chi^2_{\frac{1}{2}p'(p'+1)+p'-\frac{1}{2}p''(p''+1)-p''}. \tag{59}$$

If the observed value of T_3 is the tail of this χ^2 distribution, then the null-hypothesis has to be rejected. Similar to formula (51), the ratio of the generalised variances can be rewritten as

$$\frac{|\underline{\underline{S}}_{II,2}|}{|\underline{\underline{S}}_{II,1}|} = 1 + m(\underline{\hat{\alpha}}_{II} - \underline{\underline{D}}\,\underline{\hat{\beta}}_{II})^t \underline{\underline{X}}_{rad}^t \underline{\underline{S}}_{II,1}^{-1} \underline{\underline{X}}_{rad} (\underline{\hat{\alpha}}_{II} - \underline{\underline{D}}\,\underline{\hat{\beta}}_{II}). \tag{60}$$

As a concrete example, we consider testing the null-hypothesis that the underlying profiles can be described by gaussian functions (with random coefficients) against the alternative that a general Hermitian spline model (also with random coefficients) holds, see (13). For the gaussian model we have $\underline{\beta} = (\mu(0), b_0)^t$, the random gaussian coefficients $\underline{B}$ satisfying $\underline{B} \sim N(\underline{\beta}, \underline{\Lambda}_2)$, with $\underline{\Lambda}_2$ a general 2×2 covariance matrix. For the general spline model we have $\underline{\alpha} = (\mu(0), a_0, b_0, c_0, b_1, c_1, b_2, c_2, \ldots)^t$, and $\underline{A} \sim N(\underline{\alpha}, \underline{\Lambda}_1)$. Obviously,

$$\underline{\underline{D}}^t = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}. \tag{61}$$

The gaussian model has the simple design matrix

$$\underline{\underline{X}}_{rad,g} = \underline{\underline{X}}_{rad}\underline{\underline{D}} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ r_1^2 & r_2^2 & r_3^2 & \ldots & r_p^2 \end{pmatrix}. \tag{62}$$

Using the methods described section V, one can estimate, within each of the random coefficient models, $\underline{\alpha}$ and $\underline{\beta}$ (the estimates are denoted by $\underline{\hat{\alpha}}_{II}$ and $\underline{\hat{\beta}}_{II}$, respectively) as well as the residual SSCP matrices from both fits (denoted by $\underline{\underline{S}}_{II,1}$ and $\underline{\underline{S}}_{II,2}$). Hence, one can calculate the test statistic T_3 from either side of (60). (The left-hand side may computationally be somewhat more cumbersome for large matrices.) Note that for the gaussian model, $\underline{\underline{D}}^t \underline{\underline{X}}_{rad}^t \underline{\underline{S}}_{II,1}^{-1} \underline{\underline{X}}_{rad} \underline{\underline{D}}$ in (60) can be rewritten as $\underline{\underline{X}}_{rad,g}^t \underline{\underline{S}}_{II,1}^{-1} \underline{\underline{X}}_{rad,g}$, which is readily calculated and can be interpreted as a constant times the estimated covariance matrix for the regression coefficients $\underline{\hat{\beta}}_{II}$, *provided* the covariance matrix of the p profile measurements is estimated from the residual SSCP matrix of the *full* spline model M_1.

6.4 Profile invariance

We consider the model with covariates and error structure III, i.e.,

$$\underline{\underline{Y}} = \underline{\underline{X}}_{rad}\underline{\underline{\alpha}}\,\underline{\underline{X}}_{cov} + \underline{\underline{E}}, \tag{63}$$

with the columns of $\underline{\underline{E}}$ independently $N_p(\underline{0}, \underline{\underline{\Sigma}})$ distributed, see (21). As argued in section 4.2, the hypothesis of (full) profile invariance means that all elements of $\underline{\underline{\alpha}}$ except for the first row and the first column are zero. This can be expressed as

$$H_0 : \underline{\underline{C}}\,\underline{\underline{\alpha}}\,\underline{\underline{M}} = 0, \tag{64}$$

where $\underline{\underline{C}}$ and $\underline{\underline{M}}$ are diagonal matrices with 0 on the (1,1) position and unity on all other diagonal elements. The hypothesis is tested by a similar criterion as in the previous cases, namely

$$T_4 = \widetilde{m} \ln \frac{|\underline{\underline{S}}_{II,2}|}{|\underline{\underline{S}}_{II,1}|}, \tag{65}$$

where $\underline{\underline{S}}_{II,2}$ denotes the SSCP matrix under H_0 and the $\underline{\underline{S}}_{II,1}$ the SSCP matrix under the less restricted model. Now, we have

$$\underline{\underline{S}}_{II,2} = \underline{\underline{C}}(\underline{\underline{X}}_{rad}^t (f\tilde{\underline{\underline{\Sigma}}})^{-1} \underline{\underline{X}}_{rad})^{-1} \underline{\underline{C}}^t, \tag{66}$$

with $f = m - w - 1 - (p - p')$. The expression for $\underline{\underline{S}}_{II,1}$ is slightly more complicated. It can be written as [13] $\underline{\underline{S}}_{II,2} + \underline{\underline{S}}_{II,H}$, with

$$\underline{\underline{S}}_{II,H} = (\underline{\underline{C}} \,\underline{\alpha}\, \underline{\underline{M}})(\underline{\underline{M}}^t \underline{\underline{R}}\, \underline{\underline{M}})^{-1} (\underline{\underline{C}} \,\underline{\alpha}\, \underline{\underline{M}})^t, \tag{67}$$

where

$$\underline{\underline{R}} = \underline{\underline{Q}} + f^{-1} \underline{\underline{Q}}\, \underline{\underline{X}}_{cov} \underline{\underline{Y}}^t (\tilde{\underline{\underline{\Sigma}}}^{-1} (\underline{\underline{I}} - \underline{\underline{P}}_{\tilde{\underline{\underline{\Sigma}}}})) \underline{\underline{Y}}\, \underline{\underline{X}}_{cov} \underline{\underline{Q}} \tag{68}$$

with $\underline{\underline{Q}} = (\underline{\underline{X}}_{cov} \underline{\underline{X}}_{cov}^t)^{-1}$ and

$$\underline{\underline{P}}_{\tilde{\underline{\underline{\Sigma}}}} = \underline{\underline{X}}_{rad} (\underline{\underline{X}}_{rad}^t \tilde{\underline{\underline{\Sigma}}}\, \underline{\underline{X}}_{rad})^{-1} \underline{\underline{X}}_{rad}^t \tilde{\underline{\underline{\Sigma}}}^{-1}. \tag{69}$$

Note that $\underline{\underline{P}}_{\tilde{\underline{\underline{\Sigma}}}}$ is the projection operator with respect to the inner product defined by $\tilde{\underline{\underline{\Sigma}}}$. For $\widetilde{m} = m - (w+1) - (p-p') - \frac{1}{2}(r_c - r_m + 1)$, where r_c and r_m denote the rank of $\underline{\underline{C}}$ and $\underline{\underline{M}}$, respectively, T_4 is approximately distributed as a χ^2 variate with $r_c r_m$ degrees of freedom.

The hypothesis of partial profile invariance (for instance, with respect to q_{cyl}, but only for $r > 0.6$), can also be expressed as $H_0 : \underline{\underline{C}} \,\underline{\alpha}\, \underline{\underline{M}} = 0$, for suitable matrices $\underline{\underline{C}}$ and $\underline{\underline{M}}$, and hence be tested by the just described procedure.

7. CONFIDENCE BANDS AND PREDICTION

The presentation of just a single, smoothed profile estimate gives no information on which *other* profiles would *also* be plausible for the true underlying profile or for a new experimentally determined profile. Such information is given by confidence, and prediction bands, respectively. Although confidence bands can be considered as a limiting

case of prediction bands, we shall discuss them consecutively, as the construction of prediction bands is more complicated.

It will be useful to make a distinction between 'local' and 'global' confidence bands. A local confidence band for an underlying unknown profile $\mu(r)$ is defined as the region between two random boundaries $\mu_{loc}^{(l)}(r)$, and $\mu_{loc}^{(h)}(r)$ such that, for any $r \in [0,1]$,

$$P\{\mu(r) \in \left(\mu_{loc}^{(l)}(r), \mu_{loc}^{(h)}(r)\right)\} = 1 - \alpha, \tag{70}$$

i.e., for any $r \in [0,1]$, the probability that the random interval $\left(\mu_{loc}^{(l)}(r), \mu_{loc}^{(h)}(r)\right)$ includes $\mu(r)$ equals $1-\alpha$. A global confidence band for $\mu(r)$ is defined by

$$P\{\mu(r) \in \left(\mu_{loc}^{(l)}(r), \mu_{loc}^{(h)}(r)\right) \quad \text{for all} \quad r \in [0,1]\} = 1 - \alpha. \tag{71}$$

The quantity $1-\alpha$ is called the confidence coefficient of the band, which is frequently chosen to be 67% or 95%. The null-hypothesis that the 'true' plasma profile equals some predescribed profile $\mu_o(r)$ is not rejected at the $100\alpha\%$ level if and only if $\mu_o(r)$ is everywhere contained in a *global* confidence band with confidence coefficient $1-\alpha$.
Remark: For each confidence band with confidence coefficient $1-\alpha$, there exists a statistical test rejecting at level α. Usually, one tries to select tests, and confidence bands with some 'optimality criterion', for instance, to confidence bands with minimal area. As, under our model assumptions, the estimates for $\mu(r)$ are symmetrically distributed and (usually) unbiased, we will consider only symmetrical confidence bands that can be written as $\hat{\mu}(r) \pm \delta_{loc}(r)$, and $\hat{\mu}(r) \pm \delta_{gl}(r)$, and that correspond to well-known and rather efficient (two-sided) statistical tests.

We will discuss now the construction of confidence bands for fixed plasma parameters and error structures I and II. The methods are easily adapted to error structure III, however. As in section III, we have the following model for the m stochastically independent profiles $\underline{Y}_1, \ldots, \underline{Y}_m$ at p radial positions: $\underline{Y}_j = \underline{\underline{X}}_{rad}\underline{A} + \underline{E}_j$, with $\underline{A}$ either deterministically equal to $\underline{\alpha}$ (model I) or $\underline{A} \sim N_{p'}(\underline{\alpha}, \underline{\underline{\Lambda}})$ (model II) and $\underline{E}_j \sim N_p(\underline{0}, \sigma^2\underline{\underline{W}}_d)$, for $j = 1, \ldots, m$.

7.1 Local confidence bands

A local confidence band for $\mu(r) = \sum_{h=1}^{p'} \alpha_h f_h(r)$ is constructed by calculating for each $r \in [0,1]$ a confidence interval for the linear combination $\underline{\alpha}^t \underline{f}(r)$ of the regression parameters, where $\underline{\alpha} = (\alpha_1, \ldots, \alpha_{p'})^t$ and $\underline{f}(r) = (f_1(r), f_2(r), \ldots, f_{p'}(r))^t$ From the fact that $\hat{\underline{\alpha}}^t$ has a multivariate normal distribution it follows that $\hat{\underline{\alpha}}^t \underline{f}(r)$ is normally distributed

with expectation $\underline{\alpha}^t \underline{f}(r)$ and variance $\underline{f}^t(r) \underline{\underline{V}}(\hat{\underline{\alpha}}) \underline{f}(r)$. By studentising (i.e. by taking into account the effect that the estimated instead of the true variance of $\hat{\underline{\alpha}}$ is inserted) we get

$$\hat{\underline{\alpha}}_I^t \underline{f}(r) \pm (\underline{f}^t(r) \hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}_I) \underline{f}(r))^{\frac{1}{2}} t_{mp-p';\alpha/2}, \tag{72}$$

where $\hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}_I) = m^{-1} \left(\underline{\underline{X}}_{rad}^t \underline{\underline{W}}_d^{-1} \underline{\underline{X}}_{rad} \right)^{-1} \hat{\sigma}$ with $\hat{\sigma}$ given by (28), as the required confidence interval under model I, and

$$\hat{\underline{\alpha}}_{II}^t \underline{f}(r) \pm (\underline{f}^t(r) \hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}_{II}) \underline{f}(r))^{\frac{1}{2}} t_{m-1;\alpha/2}, \tag{73}$$

where $\hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}_{II})$ is given by (38), as the corresponding confidence interval under model II. As usual, t_f stands for the Student distribution with f degrees of freedom. Note that the unbiased estimation of the covariance structure in model II costs a considerable number of degrees of freedom. This does not always matter in practice, as for $f > 20$ and usual confidence levels, the Student distribution is very close to the standard normal.

7.2 Global confidence bands

A global confidence band for the unknown profile is derived from a k-dimensional confidence ellipse for $\underline{\alpha}$, which consists of all values $\underline{\alpha}_o$ such that

$$(\underline{\alpha}_o - \hat{\underline{\alpha}})^t \hat{\underline{\underline{V}}}(\hat{\underline{\alpha}})^{-1} (\underline{\alpha}_o - \hat{\underline{\alpha}}) \leq c(m, p, p'; \alpha), \tag{74}$$

where $c(m, p, p'; \alpha)$ equals $F_{p', mp-p';\alpha}$ for model I and $\frac{(m-1)p'}{m-p'} F_{p', m-p';\alpha}$ for model II. (Asymptotically, both expressions tend to $\chi^2_{p';\alpha}$.) The extreme values of $\underline{\alpha}_o^t \underline{f}(r)$ under the restriction (74) are found to be (using e.g. the method of Lagrange multipliers)

$$\hat{\underline{\alpha}}^t \underline{f}(r) \pm (\underline{f}^t(r) \hat{\underline{\underline{V}}}(\hat{\underline{\alpha}}) \underline{f}(r))^{\frac{1}{2}} (c(m, p, p'; \alpha))^{\frac{1}{2}}. \tag{75}$$

Evidently, a global band is wider than a local band with the same confidence coefficient. Notice from (72–73) and (75) that in this case the bands are 'proportional', i.e. a global band with coefficient $1 - \alpha$ corresponds to a local band with coefficient $1 - \alpha' > 1 - \alpha$.

Under model I, similar formulae can be derived for local and global bands for unequal design matrices, and in particular for the regression model (20–22) that includes plasma variables as covariates. (In that case one may be particularly interested in a band that is global with respect to r, but local with respect to the plasma variables.) Under model II one has then to be satisfied with asymptotic confidence bands, based on some maximum likelihood estimate of $\underline{\underline{V}}(\hat{\underline{\alpha}}_{II})$.

A practical application of confidence band construction is the following. As indicated in section 4.2, profile invariance is expressed by the requirement that the profile shape $L^{-1}(r,\underline{x}) = \partial/\partial r \mu(r,\underline{x})$ does not depend on the plasma parameters $\underline{x}$. For a graphical evaluation, it is useful to plot $\partial g_k(\underline{x})\hat{L}^{-1}(r,\underline{x})$, which is the mixed derivative of $\hat{\mu}(r,\underline{x})$, as a function of r, for $g_1(\underline{x}) = \ln I_p$, $g_2(\underline{x}) = \ln q_{cyl}$, etc. The hypothesis of profile invariance is not rejected for those plasma variables, and for those radii, for which the plotted mixed derivative of $\hat{\mu}(r,\underline{x})$ does not 'significantly' differ from zero. For each given plasma variable, this can directly by tested once a global confidence band for this mixed derivative is constructed. Under the assumption that it is sufficient to the expand $\mu(r,\underline{x})$ only up to the first powers in the logarithms of the plasma parameters, $\partial/\partial g_k(\underline{x})\partial/\partial r\mu(r,\underline{x}) = \sum_{h=1}^{p'} \alpha_{h,k} f_h'(r)$ for $k = 1,\ldots w$ (see (20)), and hence precisely the type of gobal confidence band as discussed above can be used. A practical elaboration of this for ohmic ASDEX shots is found in [9].

7.3 Prediction bands

We now turn to the somewhat more complicated case of prediction bands. The discussion will be tuned to error model II, where we have random measurement noise and intrinsic plasma variations. We start by defining a fairly general construct, which contains several useful special cases. Let $\langle Y\rangle_{m_0}(r,t)$ be the average of m_0 hypothetical, new measurements (on logarithmic scale) at time t and at some normalised flux-surface radius r. Similarly, let $\langle Y\rangle_{m_0,m_1}(r,t) = m_1^{-1}\sum_{j=1}^{m_1}\langle Y\rangle_{m_0}(r,t_j)$, where $t_1,\ldots,t_{m_1}$ are equidistant timepoints over the time interval $(t-\frac{1}{2}\Delta t, t+\frac{1}{2}\Delta t)$. In other words, $\langle Y\rangle_{m_0,m_1}(r)$ 'estimates' a time-averaged profile over the just mentioned time interval on the basis of m_0 hypothetical, repeated measurements at each of m_1 equidistant timepoints. We assume Δt to be so large that, for $m_1 \to \infty$, $\langle Y\rangle_{m_0,m_1}(r,t) \to \mu(r)$. (The assumed model implies that, for each fixed r, $\langle Y\rangle_{m_0}(r,t)$ is a stationary stochastic process. Hence, the right-hand side is independent of time.)

A prediction band for $\langle Y\rangle_{m_0,m_1}(r,t)$ is based on the the variance of the difference between $\langle Y\rangle_{m_0,m_1}(r,t)$ and its estimate $\underline{\hat{\alpha}}_{II}^t\underline{f}(r)$:

$$\begin{aligned} V(\langle Y\rangle_{m_0,m_1}(r,t) - \underline{\hat{\alpha}}_{II}^t\underline{f}(r)) &= V(\langle Y\rangle_{m_0,m_1}(r,t) - \underline{\alpha}^t\underline{f}(r)) + V((\underline{\hat{\alpha}}_{II}-\underline{\alpha})^t\underline{f}(r)) \\ &= (\frac{1}{m_1}+\frac{1}{m})(\underline{f}^t(r)\underline{\underline{\Lambda}}\,\underline{f}(r)) + \\ &\quad + (\frac{1}{m_1 m_0}+\frac{1}{m})\underline{f}^t(r)\left(\underline{\underline{X}}_{rad}^t\underline{\underline{W}}_d^{-1}\underline{\underline{X}}_{rad}\right)^{-1}\underline{f}(r)\sigma^2, \end{aligned} \tag{76}$$

see (34). (To understand this formula, it may be useful to consider first the various special cases: $m_1 = 1, m_0 \to \infty$; $m_1 = 1$, m_0 arbitrary, etc.) Note that, because of the stationarity assumption, the right-hand side of (76) is independent of time.

For brevity, we denote the variance in (76) by V. In practice, the quantities $\underline{\underline{V}}(\underline{\hat{\alpha}}_{II})$, σ^2, and $\underline{\underline{\Lambda}}$ are unknown, and have to be estimated, for instance from (38)-(40), and inserted into (76). From the resulting estimate $\hat{V}$, asymptotic prediction bands for $\langle Y \rangle_{m_0,m_1}(r,t)$ are constructed (with some radial interpolation assumption for the first term if the variance is not assumed to be constant for the p measurement channels). Asymptotically (i.e. if m and p sufficiently large), a local m_0, m_1 prediction band with confidence coefficient $1-\alpha$ is given by $\underline{\hat{\alpha}}_{II}^t \underline{f}(r) \pm \hat{V}^{1/2} u_{\alpha/2}$. For $m_0 \to \infty$, we have a confidence band for the underlying profile at a certain time t (if $m_1 = 1$), or of the time-averaged profile (if $m_1 \to \infty$). An asymptotic global *confidence* band is obtained by replacing $u_{\alpha/2}$ by $(\chi^2_{p';\alpha})^{1/2}$, and letting $m_0 \to \infty$. Studentised confidence bands, which take into account the error in estimating V, are constructed by replacing $u_{\alpha/2}$ by $t_{p-p';\alpha/2}$, and $\chi^2_{p';\alpha}$ by $p' F_{p',p-p';\alpha}$.

Although somewhat more complicated, performing predictions from the covariate model goes along similar lines. Such an analysis is well worth the effort, as it enables one to make predictions for any value (within a reasonable range) of the plasma parameters $\underline{x}$, and because the accuracy in the prediction will be increased due to the extra information contained in the additional density and temperature profiles.

7.4 Confidence intervals for global plasma variables

There may be special interest in estimating volume-averaged profiles, instead of the profiles themselves. In particular, we consider

$$< n_e >= \int_0^1 n_e(r) h(r) dr, \quad < p_e >= \int_0^1 n_e(r) T_e(r) h(r) dr, \quad < T_e >=< p_e > / < n_e >, \tag{77}$$

where r denotes the flux-surface radius and $h(r)$ is some weight function arising from the transformation from geometrical radius to flux surface radius. For concreteness, we give the discussion for $< n_e >$. For $< p_e >$ the calculations are completely analogous, and as will be explained in the end, an approximate confidence interval for $< T_e >$ can be constructed from those for $< p_e >$ and $< n_e >$. For simplicity of notation, we suppose, as in section III, that $\ln n_e(r) = \mu(r) = \sum_{h=1}^{p'} \alpha_h f_h(r) = \underline{\alpha}^t \underline{f}(r)$. The formulae are, however, analogous in the physically more general case $\mu(r) = \sum_h \sum_k f_h(r) g_k(\underline{x})$. We assume that we have unbiased and approximately normally distributed estimators for $\underline{\alpha}$ and $\mu(r)$, which are denoted by $\underline{\hat{\alpha}}$ and $\hat{\mu}(r)$, respectively. We introduce

$$\hat{n}(r) = e^{\hat{\mu}(r)}, \quad \text{and} \quad < \hat{n}_e >= \int_0^1 e^{\hat{\mu}(r)} h(r) dr. \tag{78}$$

It is assumed that $< n_e >$ is accurately approximated (say, within 1 %) by a numerical integration routine, so that the difference between $< \hat{n}_e >$ and its numerical estimate is negligible with respect to both the variance and the bias in $< \hat{n}_e >$. It can be derived, see Appendix, that

$$var < \hat{n}_e >= \int_0^1 \int_0^1 e^{\underline{\alpha}^t(\underline{f}(r)+\underline{f}(r'))} \underline{f}^t(r) \underline{\underline{V}}(\underline{\hat{\alpha}}) \underline{f}(r') h(r) h(r') dr dr' \qquad (79)$$

To estimate this variance, estimators for $\underline{\alpha}$ and $\underline{\underline{V}}(\underline{\hat{\alpha}})$ have to be inserted in this expression (see section V), and the integration has to be carried out either analytically (which may be rather complicated) or numerically.

As (78) is a non-linear transformation, an unbiased estimate $\hat{\mu}(r)$ will *not* lead to an unbiased estimate for $< n_e >$. In fact, as derived in the Appendix, the bias is approximately

$$bias < \hat{n}_e >= \frac{1}{2} \int_0^1 e^{\underline{\alpha}^t \underline{f}(r)} \underline{f}^t(r) \underline{\underline{V}}(\underline{\hat{\alpha}}) \underline{f}(r) h(r) dr. \qquad (80)$$

To determine $< \hat{n}_e >$, bias $< \hat{n}_e >$, and var $< \hat{n}_e >$ numerically, the same type of integration routine is required. If $\mu(r)$ is approximated by splines, a low-order integration routine should be used to cope with the third derivative discontinuities. As, otherwise, the smoothed profiles behave very neatly, and only a relatively low accuracy is needed, the numerical integration should present no problem.

In practice, it may turn out that $bias^2 < \hat{n}_e >$ is negligible with respect to $var < \hat{n}_e >$. In that case, an approximately 95% confidence interval for $< n_e >$ is given by $< \hat{n}_e > \pm 2\, var^{1/2} < \hat{n}_e >$. If this is not the case, one can correct for the bias (which should be small anyhow), and assume that the variance is only negligibly influenced by this operation.

Finally, we discuss how to obtain a confidence interval for the density weigthed volume-averaged temperature. We assume that we fit the density and the temperature profiles on logarithmic scale, so that we have

$$\hat{n}_e(r) = e^{\underline{\hat{\alpha}}_n^t \underline{f}_n(r)} \qquad \hat{T}_e(r) = e^{\underline{\hat{\alpha}}_T^t \underline{f}_T(r)},$$

where $\underline{\hat{\alpha}}_n$ and $\underline{\hat{\alpha}}_T$ both have a multivariate normal distribution. For some reason, one might prefer in practice another set of basis functions for the density fit than for the temperature fit, so for generality we do not suppose that $\underline{f}_n(r) = \underline{f}_T(r)$.

Under the above model assumption,

$$< \hat{p}_e >= \int_0^1 e^{\underline{\hat{\alpha}}_n^t \underline{f}_n(r) + \underline{\hat{\alpha}}_T^t \underline{f}_T(r)} h(r) dr, \quad < \hat{n}_e >= \int_0^1 e^{\underline{\hat{\alpha}}_n^t \underline{f}_n(r)} h(r) dr,$$

and we consider the estimate $< \hat{T}_e >=< \hat{p}_e > / < \hat{n}_e >$, which, for $var(\underline{\hat{\alpha}}_n^t \underline{f}_n(r)) \ll 1$ and $var(\underline{\hat{\alpha}}_T^t \underline{f}_T(r)) \ll 1$, has approximately zero bias and variance

$$\frac{1}{< n_e >^2} var < \hat{p}_e > + \frac{< p_e >^2}{< n_e >^4} var < \hat{n}_e > - 2\frac{< p_e >}{< n_e >^2} cov\,(< \hat{n}_e >, < \hat{p}_e >).$$

(If necessary, the bias can be estimated. We will concentrate on the variance.) Each term in this expression, except for $cov(< \hat{n}_e >, < \hat{p}_e >)$, can be estimated according to the formulas derived above. For the cross term we have

$$cov(< \hat{n}_e >, < \hat{p}_e >) = \int_0^1 \int_0^1 cov(\hat{n}_e(r), \hat{p}_e(r)) h(r) h(r') dr dr',$$

with

$$cov(\hat{n}_e(r), \hat{p}_e(r')) = n_e(r) p_e(r') (\underline{f}_n^t(r) \underline{\underline{V}}\,(\underline{\hat{\alpha}}_n) \underline{f}_n^t(r') + \underline{f}_n^t(r) \underline{\underline{V}}\,(\underline{\hat{\alpha}}_n, \underline{\hat{\alpha}}_T) \underline{f}_T^t(r')).$$

Here, $\underline{\underline{V}}\,(\underline{\hat{\alpha}}_n, \underline{\hat{\alpha}}_T)$ denotes the matrix of covariances between $\underline{\hat{\alpha}}_n$ and $\underline{\hat{\alpha}}_T$. Obviously, this is zero if the temperature and the density measurements are independent, which is for instance not the case for the Nd:YAG laser experiments, where $\underline{\underline{V}}\,(\underline{\hat{\alpha}}_n, \underline{\hat{\alpha}}_T)$ has to be estimated from a simultaneous regression of the temperature and density profiles. For volume-averaged profiles up to flux-surface radius r_0, all of the above formulae hold if only the upper integration limit is changed from 1 to r_0.

The important point to see is that although the expressions look somewhat complicated, their derivation and structure is quite straightforward. Given some tools for numerical integration, practical implementation is not a tremendous task.

It should be noted that the best (i.e. minimum variance) unbiased predictor for the local profile is not the best unbiased predictor for the volume averaged profile. In the latter case, because of the typical behaviour of $rn(r)$, the profile has to be known with a higher precision in the middle region (say $.25 < r < 0.75$) than at both ends of the interval $[0, 1]$. On the other hand, an optimised profile fit with respect to MSE ($< \hat{n}_e >$) may locally be bad, and for instance over- or underestimate the central density considerably. In a larger profile study, where the profiles are used for various purposes and where it is useful to have the estimates of the volume averages consistent with the fitted profiles, we prefer therefore local profile fits with good overall properties, such as those discussed in section 5, and allow for the fact that the estimates for the volume averages do not have minimal variance.

8. DISCUSSION AND CONCLUSIONS

In this article, we have presented a systematic approach to the parameterisation of temperature and density profile shapes. Our approach is to treat profile shape parameterisation as a problem in statistical regression. The profiles are expanded in a double series of radial and plasma-parameter basis functions. Each product basis function is a covariate and the corresponding regression coefficients have to be estimated. The 'best' method of estimation, depends on the postulated error structure. Partly motivated by the ASDEX YAG measurements, the simplifying assumption was made that the measurement variations at a certain channel are uncorrelated in time. On the other hand, the deviations from an ideal profile fit may be radially correlated.

In Secs. 2, 3, and 4, we have presented a variety of different representations for the plasma profiles and the error covariance, $\underline{\underline{\Sigma}}$. These statistical models often form a hierarchy with the simple models embedded in the complex models as special cases. The more realistic models have more free parameters and therefore require more data. In a sense, the discrete point model of Sec. 2 constitutes the most accurate radial representation. However, a spline model with roughly equal numbers of measurement channels between radial knots yields a more effective representation for many datasets. The random coefficient model of $\underline{\underline{\Sigma}}$, where the spline coefficients vary randomly in addition to statistical noise, is expected to be a realistic model of the spatial correlations of the fluctuations.

In Sec. 5, three methods to estimate the spline coefficients are presented. Firstly, by generalised least squares. Direct and elegant solutions are possible for relatively simple covariance structures, and even for the random coefficient model in the simple situation that one has a dataset of profiles for fixed values of the plasma parameters.

Secondly, by maximum likelihood, which, in simple situations and for normally distributed errors, coincides with least squares. In more complex situations, differentiation of the log likelihood with respect to the mean value and the covariance parameters, gives two coupled sets of equations, which have to be solved iteratively. In such a case, least squares can be considered as one such iteration step which maximises the likelihood with respect to the mean value parameters for a fixed (known or sensibly estimated) covariance matrix.

Finally, a discussion has been given of robust regression. The so-called M-estimators can be viewed as a generalisation of maximum likelihood estimators, in as far as they minimise some general function of the residuals, which may have nothing to do with the actual likelihood function. For relatively simple covariance structures, the asymptotic

distribution of general M-estimators has been rigorously derived [28]. The purpose is to diminish the influence of erroneous, outlying datapoints on the regression estimates, by downweighting large residuals. As a robust regression fits closely the bulk of the data, it also suitable for clearly detecting outlying (i.e. suspicious) datapoints.

Sec. 6 presents a number of statistical tests to check if a simplification of the model, i.e. either of the profile representation or of the model for $\underline{\underline{\Sigma}}$, produces significantly more than the expected increase in the residual error. As a special application, the empirical testing of profile invariance is considered.

Sec. 7 describes the construction of local and global confidence bands for the true profile shapes, and of prediction bands for observing new profiles. Furthermore, estimates and confidence intervals are given for corresponding volume-averaged quantities.
A parameterised representation provides a compact summary of an experimental investigation: a table of spline coefficients is much more usable than a large database of all discharges. Knowledge of the dependence of profiles on the plasma parameters may lead to new physical insight. In view of the considerable experimental variation in individual profile measurements, it is, however, essential to express clearly at least the statistical uncertainties associated with fitting a set of carefully measured profiles. Such parameterised profiles are then in a form suitable to be input into transport and stability simulations.

9. APPENDIX

9.1 Profile representations by perturbation expansion

We consider the Hermite polynomial representation given by (11). For any fixed value of c_0, the coefficients $a_1, a_2, \ldots$ are linearly related to the moments of the temperature distribution. To see this, we write (11) as

$$T(r) = T_0 \sum_{n=0}^{\infty} (-1)^n a_n \phi_{c_0}^{(n)}(r) \qquad \text{with} \quad \phi_{c_0}^{(n)}(r) = (d/dr)^n e^{-c_0 r^2}, \tag{81}$$

with $a_0 = 1$. From this it follows that the two-sided Laplace transform equals

$$\int T(r) e^{-sr} dr = T_0 (\frac{\pi}{c_0})^{1/2} e^{\frac{s^2}{4c_0}} \sum_{n=0}^{\infty} (-1)^n a_n s^n \tag{82}$$

Since the Laplace transform generates the moments of the temperature profile, we get, by expanding (13) as a power series in s,

$$T_0(\frac{\pi}{c_0})^{1/2}(a_{2k} + \frac{a_{2k-2}}{4c_0} + \cdots + \frac{a_0}{k!(4c_0)^k}) = \frac{m_{2k}}{(2k)!}. \tag{83}$$

A similar expression holds for the odd moments. Hence, for each fixed value of c_0 one can, from the fitted regression coefficients $a_1, a_2, \ldots$ (and their estimated covariance matrix), easily calculate the moments $m_1, m_2, \ldots$ (and their covariance matrix). Reversely, by estimating the moments by some method, one could derive the corresponding estimates for the polynomial coefficients.

Remark. Note that for symmetric profiles, the Fourier transform of (11) is a gaussian function times a power series with coefficients $a_1, a_2, \ldots$, i.e. assumes the form (10). In this sense, the representations given by (10) and (11) are dual.

9.2 Variance and bias for volume-averaged global quantities

It is recalled that, by definition,

$$MSE(< \hat{n}_e >) = E(< \hat{n}_e > - < n_e >)^2 = var\ < \hat{n}_e > + bias^2\ < \hat{n}_e >, \tag{84}$$

where $var\ < \hat{n}_e > = E(< \hat{n}_e > - E(< \hat{n}_e >))^2$, and $bias\,(< \hat{n}_e >) = E(< \hat{n}_e >) - < n_e >$, and $E(X)$ denotes the mathematical expectation of a random variable X. The variance and the bias of $< \hat{n}_e >$ depend in a simple manner on the covariance function

$$cov(\hat{n}(r), \hat{n}(r')) = E\,(\hat{n}(r) - E(\hat{n}(r))(\hat{n}(r') - E(\hat{n}(r')).$$

By Taylor expansion, it is derived that, for $var\,\hat{\mu}(r) \ll 1$ and $var\,\hat{\mu}(r') \ll 1$,

$$cov(\hat{n}(r), \hat{n}(r')) = n(r)n(r')cov(\hat{\mu}(r), \hat{\mu}(r')). \tag{85}$$

Because of the linear structure of our linear model for $\mu(r)$, we have

$$cov(\hat{\mu}(r), \hat{\mu}(r')) = \underline{f}^t(r)\underline{\underline{V}}\,(\underline{\hat{\alpha}})\underline{f}(r'). \tag{86}$$

As $< \hat{n}_e > = \int \hat{n}(r)h(r)dr$, it is easily derived that the variance of $< \hat{n}_e >$ is obtained by integrating $cov(\hat{n}(r), \hat{n}(r'))$ over $h(r)dr$ and $h(r')dr'$. Hence, the full expression is given by (79).

The bias of $< \hat{n}_e >$ can clearly be written as $\int_0^1 bias\,\hat{n}(r)h(r)dr$. Again by Taylor expansion, it is derived that, for $var\,\hat{\mu}(r) \ll 1$,

$$bias\,\hat{n}(r) \simeq \frac{1}{2} var\,\hat{n}(r),$$

where $var\,\hat{n}(r) = cov(\hat{n}(r), \hat{n}(r))$ can be estimated from (85). Hence, the full formula is given by (80).

ACKNOWLEDGEMENTS

OJK likes to acknowledge the comments of Prof. Dr. W. Schaafsma and Dr. K. Poortema on the manuscript in an embryonic stage. KSR wishes to thank Prof. Dr. D. Pfirsch for helpful discussions. The work of KSR was supported by the U.S. Department of Energy, Grant No. DE-FG02-86ER53223. The work of PJM was performed under a EURATOM supported reciprocal research agreement between University College Cork and IPP.

REFERENCES

[1] MINER, W., ROSS, D., WILEY, J.C., Poster No. 2C5, Sherwood Theory Conference, 1988.

[2] KARDAUN, O., McCARTHY, P.J., LACKNER, K., RIEDEL K.S., GRUBER, O., in: Theory of Fusion Plasmas (Proc. Workshop Varenna 1987), Società Italiana di Fisica, Bologna (1988) 435.

[3] McCARTHY, P.J., RIEDEL, K.S., KARDAUN O., LACKNER K., GRUBER O., Poster No. 1D16, Sherwood Theory Conference, 1988.

[4] CHRISTIANSEN, J.P., CALLEN, J.D., CORDEY J.G., THOMSEN K., Nucl. Fusion **28** (1988) 817.

[5] BRAAMS, B.J., *Computational Studies in Tokamak Equilibrium and Transport*, thesis, Utrecht (1986).

[6] McCARTHY, P.J., SEXTON, M.C., *Plasma Profile Recovery by Function Parameterisation*, Max-Planck-Institut für Plasmaphysik, Report No. 5/12.

[7] CONNOR, J.W., Plasma Phys. Contr. Fusion **30** 6 (1988) 619.

[8] VON MISES, R., *Mathematical Theory of Probability and Statistics*, p. 135, Academic Press, New York (1964).

[9] McCARTHY, P.J., RIEDEL, K.S., KARDAUN O., MURMANN, H., LACKNER K., the ASDEX Team, *Scalings and Plasma Profile Parameterisation of ASDEX High Density Ohmic Discharges*, Max-Planck-Institut für Plasmaphysik, Report No. 5/34.

[10] ANDERSON, T.W., Ann. of Stat. **1** (1973) 135.

[11] KHATRI, C.G., Ann. of the Inst. of Stat. Math. **18** (1966) 75.

[12] JOHANSEN, S., Lecture Notes in Statistics Vol. 22, Springer, Heidelberg (1983).

[13] GRIZZLE, J.E., ALLEN, D.M, Biometrics **25** (1969) 357.

[14] RAO, C.R., Proc. Fifth Berkeley Symp. on Math. Stat. and Prob. **1** (1967) 355.

[15] SZATROWSKI, T.H., MILLER, J.J. Ann. of Stat. **8** (1980) 802.

[16] SCHAAFSMA, W., *Multivariate Analyse*, Lecture Notes, Groningen (1981).

[17] AZZALINI, A., *Growth Curve Analysis for Patterned Covariance Matrices*, technical report. Dept. of Statistical Sciences, University of Padua (1985).

[18] LEHMANN, E., *Theory of Point Estimation*, Wiley, New York (1983).

[19] POORTEMA, K., *On the Statistical Analysis of Growth*, thesis, Groningen (1989).

[20] WEBSTER, J.T., GUNST, R.F., MASON, R.L., Technometrics, **16** (1974) 513.

[21] RAO, C.R., Biometrics **31** (1975) 545.

[22] STRENIO, J.F., WEISBERG, H.I., BRYK, A.S., Biometrics **39** (1983) 71.

[23] GILL, P.E., MURRAY, W., WRIGHT, M.H., *Practical Optimization*, Academic Press, London (1981).

[24] PŠENIČNYI, B.N., DANILIN, Ju.M., *Numerische Methoden für Extremalaufgaben*, VEB, Berlin (1982).

[25] JENNRICH, R.I., SCHLUCHTER, M.D., Biometrics **42** (1986) 805.

[26] BMDP Statistical Software Manual, W.J. Dixon ed., University of California Press, Berkeley (1988).

[27] HUBER, P.J., *Robust Statistics*, Wiley, New York (1981).

[28] HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., STAHEL, W.A., *Robust Statistics, The Approach based on Influence Functions*, Wiley, New York (1986).

[29] RONNER, A.E., *P-norm Estimators in a Linear Regression Model*, thesis, Groningen, (1977).

[30] SAS, User's guide: Statistics, SAS Institute Inc., Cary NC (1985).

[31] WALD, A., Trans. Amer. Math. Soc. **14** (1943) 426.

[32] LEHMANN, E., *Testing Statistical Hypotheses*, Wiley, New York (1986).

[33] SNIJDERS, T.A.B., *Asymptotic Optimality Theory for Testing Problems with Restricted Alternatives*, Mathematical Centre Tracts 113, Amsterdam (1979).

[34] MARDIA, K.V., Appl. Statist. **24** (1975) 163.

[35] WILKS, S.S., Ann. Math. Stat. **9** (1943) 60.

[36] SERFLING, R.J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York (1980).

[37] RAO, C.R., *Linear Statistical Inference and its Applications*, Wiley, New York (1973).

[38] MORRISON, D.F., *Multivariate Statistical Methods*, Mcgraw-Hill, London (1976).

[39] SCHAAFSMA, W., *Hypothesis Testing Problems with the Alternative Restricted by a Number of Inequalities*, Noordhoff, Groningen (1966).

[40] HUMAK, K.M.S., *Statistical Inference in Linear Models*, O. Bunke and H. Bunke eds., Wiley, New York (1986).

[41] ANDERSON, T.W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1984).