ISSN 0915-633X

NATIONAL INSTITUTE FOR FUSION SCIENCE

Discriminant Analysis to Predict the Occurrence of ELMS in H-Mode Discharges

O.J.W.F. Kardaun, S.-I. Itoh, K. Itoh and J.W.P.F. Kardaun

(Received – Aug. 18, 1993)

NIFS-242 Aug. 1993

RESEARCH REPORT
NIFS Series

NAGOYA, JAPAN

DISCRIMINANT ANALYSIS TO PREDICT THE OCCURRENCE OF ELMS IN H-MODE DISCHARGES

O.J.W.F. Kardaun, S.-I. Itoh*, K. Itoh** and J.W.P.F. Kardaun†

Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2,
D-85748 Garching bei München

* Kyushu University, Kasuga, 816 Japan

** National Institute for Fusion Science, Nagoya 464-01, Japan

† Central Bureau of Statistics, Voorburg, The Netherlands‡

Keywords: Discriminant Analysis, H-mode, Small ELMs, Giant ELMs, Statistical Analysis

---

‡The statements in this paper are the sole responsibility of the authors. In particular, they are not to be construed as official or reflecting the views of the Netherlands Central Bureau of Statistics.

CONTENTS

Abstract

After an exposition of its theoretical background, discriminant analysis is applied to the H-mode confinement database to find the region in plasma parameter space in which H-mode with small ELMs (Edge Localized Modes) is likely to occur.
The boundary of this region is determined by the condition that the probability of appearance of such a type of H-mode, as a function of the plasma parameters, should be (1) larger than some threshold value and (2) larger than the corresponding probability for other types of H-mode (i.e., H-mode without ELMs or with giant ELMs).
In practice, the discrimination has been performed for the ASDEX, JET and JFT-2M tokamaks (a) using four instantaneous plasma parameters (injected power P_{inj}, magnetic field B_t, plasma current I_p and line averaged electron density $\bar{n}_e$) and (b) taking also memory effects of the plasma and the distance between the plasma and the wall into account, while using variables that are normalised with respect to machine size.
Generally speaking, it is found that there is a substantial overlap between the region of H-mode with small ELMs and the region of the two other types of H-mode. However, the ELM-free and the giant ELM H-modes relatively rarely appear in the region, that, according to the analysis, is allocated to small ELMs. A reliable production of H-mode with only small ELMs seems well possible by choosing this regime in parameter space.
In the present study, it was not attempted to arrive at a unified discrimination across the machines. So, projection from one machine to another remains difficult, and a reliable determination of the region where small ELMs occur still requires a training sample from the device under consideration.

1 INTRODUCTION

Progress in tokamak physics led to the conceptual design of the international thermonuclear experimental reactor (ITER) [1]. Due to the improvement in energy confinement time τ_E of the plasma in H-mode, see e.g. [2,3], with respect to what one expects from L-mode scalings, see e.g. [4,5], the machine size needed for achieving the mission of ignition could substantially be reduced. In [6] it was shown that this improvement in confinement time over L-mode also alleviates the necessary engineering R&D. The parameter dependence of τ_E for H-mode plasmas from various machines has been enthusiastically investigated [7–15] and several scalings, under varying assumptions, have been presented.

The phenomenon ELM, which stands for Edge Localized Mode, has been observed from the beginning of H-mode research [16] and has aroused active interest ever since [17,18]. In a very rough empirical classification, based on solely the H_α (or D_α) signal, one can distinguish between three types of H-mode: ELM-free H-mode, H-mode with giant ELMs, and H-mode with small ELMs. The discrimination of these three H-modes is an important concern for future machines, because they exhibit a quite different behaviour. ELM-free H-mode discharges exhibit the largest increase in confinement time, but are most vulnerable to the accumulation of impurities and the resultant radiation collapse [2]. Long sustainment of ELM-free H-mode seems at present impossible [19]. The presence of giant ELMs during the H-mode moderates this problem, but the repetition of the large heat pulses will be a major obstacle in designing the divertor plate and the first wall. Large heat pulses cause a transient peak in the temperature of the divertor plate, so as to enhance the erosion rate and impurity generation. The H-mode with small ELMs is presently the best candidate for future plasmas aiming at sustained thermonuclear burning. Hence, we need a guiding principle to avoid the appearance of ELM-free H-mode as well as H-mode with giant ELMs, in order to enhance the probability of (quasi) steady-state operation in future experimental fusion reactors. Several experimental efforts have been made to look for the most efficient control parameter to realise the H-mode with small frequent ELMs [20–22].

The ITER H-mode database (ITERH.DB1) [7,8] contains global confinement data from the ASDEX, DIII-D, JET, JFT-2M, PBX-M, and PDX tokamaks. The three types of H-mode that were discussed above have been distinguished in this database. The scaling of the energy confinement time with respect to the plasma parameters is not precisely the same for the various types of H-mode [8,11,15]. Hence, also from this point of view, it is useful to identify the plasma parameter regimes where the various types of H-mode occur.

To analyse this question, we apply the method of discriminant analysis [23–27] to this H-mode database. We want to study the regions in plasma parameter space where two classes of H-mode, i.e., class-1 (ELM-free or with Giant ELMs) and class-2 (with Small ELMs) occur. The analysis is done for each device separately. In this report, attention is restricted to the analysis of the ASDEX, JET, and JFT-2M data of the database. These are the tokamaks

the authors are most familiar with. The restriction to ITERH.DB1 was made because of the general availability of the dataset, and the fact that precisely this dataset has been used for the global confinement time analysis published in [8].

The flow of this article is as follows. In Section 2, we present a guided tour to those theoretical aspects of discriminant analysis which are needed to understand the background of the practical sections. In particular, the relationship between discriminant analysis and regression analysis is payed attention to. In Section 3, we describe the datasets used, present a preliminary graphical analysis, and apply discriminant analysis in various settings to predict the occurrence of the various types of ELMs. In particular, we compare 'parametric discrimination' (using linear as well as quadratic discriminant functions) with 'non-parametric discrimination' (using kernel density estimates and the multinomial independence model, respectively). We discuss the explicit forms of the linear and the quadratic boundaries, and compare the performance of the various methods. In Section 3.2, this is done using four or five instantaneous plasma parameters that are usually also applied in global confinement time analyses, i.e, plasma current, magnetic field, injected power, electron density, and isotope composition, all considered at the same time-point as that of the occurrence of ELMs. In Section 3.3, the elapsed time since the L-H transition is taken into account, and the ohmic target density replaces the instantaneous density as a discriminating variable. In other words, we examine the effect of the plasma's 'memory'. Furthermore, the plasma-wall distance is used as an additional variable, and all variables are normalised with respect to machine size. Condensed results of the analyses are presented in tables and physical interpretations are discussed in the main text whenever they are deemed to be appropriate. A Summary and Discussion is presented in Section 4.

Wir machen uns Bilder der Tatsachen
Wittgenstein, Tractatus

2.1 Informal Introduction to Discriminant Analysis

In order to assess quantitatively in which regions of plasma parameter space the various types of ELM are likely to occur, we must formulate relevant parts of the mathematical theory of discrimination. Here we present some heuristic considerations. A more precise description of the theory is given in the next sub-section. An important aspect is to determine the mathematical form of the boundary of the regions of interest.

For simplicity, let us consider two classes of shots with ELMs. We want to determine the boundary between the regions in plasma parameter space where these two classes are expected to occur. One way to find this boundary is by estimating the probability distributions of the two classes of ELMy shots over plasma parameter space. We consider boundaries such that, locally, class-1 (e.g. non-HSELM) discharges are more probable than class-2 (e.g. HSELM) discharges on one side of the boundary, and vice versa on the other side. The probability density to find class-j discharges is denoted by

$$f_j(\underline{x}), \quad j = 1, 2, \tag{1}$$

where $\underline{x}$ is an n-dimensional vector of variables (i.e., plasma parameters), such as the injected power P_{inj}, plasma current I_p, magnetic field B_t, electron density $\bar{n}_e$, etc. These densities have to be estimated from the available data. The estimates depend on the assumed class of probability distributions, see Section 2.2. The boundary $\mathbf{B}$ is defined by $\mathbf{B} = \{\underline{x} \,|\, f_1(\underline{x}) = f_2(\underline{x})\}$. The region $\mathbf{R}_1$ satisfying the condition $\mathbf{R}_1 = \{\underline{x} \,|\, f_1(\underline{x}) > f_2(\underline{x})\}$ is the region where class-1 shots are expected, whereas class-2 shots are expected in the region $\mathbf{R}_2 = \{\underline{x} \,|\, f_1(\underline{x}) < f_2(\underline{x})\}$. The boundary $\mathbf{B}$ and the regions $\mathbf{R}_{1,2}$ can be calculated for each set of plasma variables $\underline{X} = (X_1, \ldots, X_p)$. The dimension p can be as large as the number of parameters which fully characterise the discharge. Among this large set of parameters, we look for key parameters which determine the boundary $\mathbf{B}$. In statistics, this is called the problem of variable selection. Obviously, this requires in practice a good interplay between physical considerations and statistical methods. Given a boundary in a higher dimensional space, one can identify which linear combinations of plasma parameters are locally important to discriminate between the classes: obviously, they consist of those linear combinations of which the hyperplanes of constant values are 'close' to the tangent plane of the boundary. (This means that the gradient of the linear combination should have a large component in the direction of the gradient of the ratio $f_1(\underline{x})/f_2(\underline{x})$.)

The allocation regions $\mathbf{R}_1$ and $\mathbf{R}_2$ are influenced by considerations that take the risks associated with the various types of misallocation into account. Usually, misclassifying a

(new) observation as class-1 has different consequences than misclassifying a (i.e. another!) new observation as class-2. For instance, mispredicting ELM-free H-mode as ELMy H-mode is more harmful, from the viewpoint of impurity accumulation, than mispredicting ELMy H-mode as ELM-free. In that case, the boundary $\mathbf{B}$ between the regions $\mathbf{R}_1$ and $\mathbf{R}_2$ is no longer optimal, and improvement is possible by constructing boundaries so as to reduce the probability of dangerous errors, which amounts to minimise, in some sense, the expected losses.

Another extension is to consider an additional area in parameter space. For instance, one region is assigned to class-1, another to class-2, and a third region is the 'grey' area, in which the probabilities associated with the two classes are either not distinctively different or are both negligibly low. (In fact, distinguishing between these two reasons for non-allocation, one has two additional regions.) By using such 'non-allocation regions', the prediction of 'class-1' (or 'class-2') can be made with less expected error. This kind of analysis is sometimes called 'discrimination in non-forced decision situations'.

Each of these cases require more precise formulations, which are discussed in the next Section.

2.2 Statistical Aspects of Discriminant Analysis

The theory of discriminant analysis is a well developed branch of statistics and at the same time still a field of active research, see e.g. [28–35]. Part of the algorithms are implemented in special or general statistical packages [36–39]. Here, we will discuss only those parts of the theory that are needed to understand the background of the practical analyses in Section 3.

One can approach discriminant analysis from a purely data-descriptive point of view and from a probabilistic point of view. (Both approaches, but most easily the latter one, can be incorporated into a decision theoretical framework.) In the latter approach, a probabilistic model is used to describe the situation. The applicability of such a model in non-random situations may be questioned from a fundamental point of view. However, such a probabilistic framework is almost indispensible if one wants to estimate the performance of procedures in future situations, and to express uncertainties in various estimates. Moreover, it often leads to procedures that are also sensible from a data-descriptive point of view. Or reversely: A specific procedure can often be viewed upon as a data-descriptive one, with little further interpretation, and as a probabilistic one, with considerably more interpretation, the validity of which is of course dependent on the adequacy of the framework. Sometimes a procedure developed in one probabilistic framework can also be interpreted in another probabilistic framework, which may be more relevant for the data at hand. We will return to this when discussing the relation between discriminant analysis and regression analysis. To our opinion, the use of such probabilistic models is justified as long as one does not take them 'too seriously',

i.e., one should realise their approximative model character. Here we shall discuss basically probabilistic methods, but we take care to introduce the various probabilistic structures only step by step.

With the term (multivariate) observation one generally denotes a basic item (consisting of various components) that is liable to repetitive measurement. In this context we use it for a vector of plasma parameters, say (P_{inj}, I_p, B_t, $\bar{n}_e$), measured at a certain time-point (or even at several time-points) during a discharge. The principal object of the excercise is to predict the occurrence of the type of ELMs as a function of the plasma parameters on the basis of historical 'training samples' of such observations for each class of ELMs.

In the case of two kinds of observations that are distributed according to two elliptical multivariate distributions, there is a geometrically evident family of boundaries that can be used for prediction: Assign a new observation to group 1 if $D_1 < cD_2$, where D_1 and D_2 are distances of the particular observation to the centers of gravity of group 1 and 2, respectively. Obviously, these distances have to be measured in the metric defined by the ellipses. Intuitively, the choice $c = 1$ seems best. (In a decision theoretic formulation, the constant c depends on the losses one wants to minimise. In a symmetric loss situation, the choice $c = 1$ is appropriate.) In this case the boundary becomes simple if two distributions differ only by their first moment: it is the line conjugate to the line joining the expectation values of the two distributions. ('Conjugate' is the same as 'perpendicular in the metric given by the ellipses'. Such a line, obviously, passes through the two (real or imaginary) intersection points of any pair of ellipses with the same 'radius'.)

A specialisation of the above elliptical situation is the case of two multivariate normal distributions, and a generalisation is the situation that we have two shifted multivariate distributions, with convex contours of equal probability density that define a seminorm on $\mathbf{R}^p$. For analogy with the normal case, it is convenient to 'scale' the seminorm according to the relation $f(\underline{x}) \propto exp[-\frac{1}{2}D^2(\underline{x})]$, where $f(\underline{x})$ denotes the probability density of the multivariate distribution. In the language of mechanics, $-2\log f(\underline{x}) = D^2(\underline{x})$ is called the potential $V(\underline{x})$.

We now describe discriminant analysis while introducing some more probabilistic structure. Again for simplicity we do this for two groups. Generalisation to $k > 2$ groups is in principle straightforward, though in practice sometimes unwieldy because of combinatorial complexity.

The basic method is to estimate the probability density for the two groups, and to allocate a new observation according to a rule based on (an estimate of) the above described family of boundaries. Actual allocation requires the determination of a single member of this family. This choice, other than the obvious one, $f_1(x) = f_2(x)$, may be influenced by two considerations: (a) The assignment of losses for the various types of error that can occur and

(b) the assignment of prior probabilities with respect to the occurrence of the two types of observations. Let $\rho_h, h = 1, 2$, denote these prior probabilities (obviously, $\rho_1 + \rho_2 = 1$), and let $f_h(\underline{x})$ denote the probability density for group h. Then, according to Bayes' theorem, the conditional or 'posterior' probability $\rho_{h|\underline{x}}$ to belong to group h, given the vector of observations $\underline{x}$, can be written as

$$\rho_{h|\underline{x}} = \frac{\rho_h f_h(\underline{x})}{\sum_{h'=1}^{2} \rho_{h'} f_{h'}(\underline{x})}. \tag{2}$$

The above class of allocation rules considers regions of which the boundaries can, equivalently, be characterised by (i) the difference in metric or potential, (ii) the ratio of probability densities, and (iii) the posterior probabilities. The relationship between the characterisations is easily derived to be

$$V_2(\underline{x}) - V_1(\underline{x}) = D_2^2(\underline{x}) - D_1^2(\underline{x}) = 2d \tag{3}$$

$$f_1(\underline{x})/f_2(\underline{x}) = e^d \tag{4}$$

$$\log \frac{\rho_{1|\underline{x}}}{1 - \rho_{1|\underline{x}}} = \log \frac{\rho_1}{1 - \rho_1} + d. \tag{5}$$

The latter expression means that d equals the shift from the prior to the posterior probabilities on the logit scale. (The S-shaped logit function $\text{logit}(p) = \log(p/(1-p))$ transforms the interval $(0, 1)$ into $(-\infty, +\infty)$.) Note that in going from (i) to (iii) gradually more probabilistic structure is introduced. A basic question is the choice of d. One choice is $d = 0$, another choice is d such that $\rho_{1|\underline{x}} = 50\%$ (in that case, d depends on ρ_1, and is obviously 0 for the 'Laplacian' choice of initial ignorance: $\rho_1 = 1/2$). A third choice is based on a loss-function formulation. Losses are associated with actions. In the two group situation, there can be two or more relevant actions. We consider: a_1: assign to group 1, a_2: assign to group 2, and a_0: no assignment. The actions are to be based on the value of the experimental variable $\underline{x} \in \mathbf{R}^p$. The mapping $d : \mathbf{R}^p \to \{a_0, a_1, a_2\}$ is called a decision function. A decision procedure fixes a decision function, which in the above case can be characterised by three regions, which we denote by $\mathbf{R}_0$, $\mathbf{R}_1$, and $\mathbf{R}_2$, respectively, where

$$\mathbf{R}_0 = d^{-1}(a_0) = \{x \epsilon \mathbf{R}^p \mid d(x) = a_0\}, \tag{6}$$

and similarly for $\mathbf{R}_1$ and $\mathbf{R}_2$. (If $\mathbf{R}_0$ is empty, we are in the two decision situation described above.) The losses depend on the actions as well as on the true state of nature, and are conveniently summarised by a loss matrix. For instance:

	a_0	a_1	a_2
1	ℓ_{01}	0	ℓ_{21}
2	ℓ_{02}	ℓ_{12}	0

(The first index of ℓ denotes the action, and the second index of ℓ indicates the true state of nature.) Some losses have to be specified, be it $\ell_{01} = \ell_{02} = \infty$, and $\ell_{21} = \ell_{12} = 1$. Then the risk ('expected loss') can be expressed as a function of the true state of nature and the decision procedure:

$$\mathrm{R}(d,1) = \ell_{01} \int_{R_0} f_1(\underline{x})d\underline{x} + \ell_{21} \int_{R_2} f_1(\underline{x})d\underline{x} \tag{7}$$

$$\mathrm{R}(d,2) = \ell_{02} \int_{R_0} f_2(\underline{x})d\underline{x} + \ell_{12} \int_{R_1} f_2(\underline{x})d\underline{x} \tag{8}$$

One principle is to choose a decision rule d_{mm} such that the maximum expected loss is minimised, i.e.,

$$\max_h \mathrm{R}(d_{mm}, h) = \min_d \max_h \mathrm{R}(d,h) \qquad h = 1,2. \tag{9}$$

This is called the minimax rule. Another principle is to minimise some linear combination of the two risks. A natural candidate for the weights of this linear combination is the set of prior probabilities, i.e.,

$$\mathrm{R}(d,\rho) = \rho_1 R(d,1) + \rho_2 R(d,2) \tag{10}$$

is minimised. This is called a Bayes decision rule with respect to the prior distribution $\rho = (\rho_1, \rho_2)$. Its definition can be reformulated as: a Bayes rule with respect to ρ minimises, as a function of $\underline{x}$ the sum of the probability densities, weighted by both the losses and the prior probabilities, i.e., it minimises the 'risk densities'

$$\sum_{h=1}^{2} \ell(d(\underline{x}), h)\rho_h f_h(\underline{x}). \tag{11}$$

One can easily see that, in the special case of 0-1 losses, and no doubt region, the Bayes rule assigns according to the largest value of $\rho_h f_h(\underline{x})$, i.e. to the largest posterior probability. Unfortunately, prior probabilities are sometimes hard to assess, or even make little physical sense. Therefore it is useful to look at the Bayes rule with respect to a prior distribution ρ_{lf} which is 'least favourable' in the sense that

$$\mathrm{R}(d,\rho_{lf}) = \max_\rho \mathrm{R}(d,\rho). \tag{12}$$

It can be shown that under some regularity conditions, the Bayes rule with respect to the least favourable prior distribution coincides with the minimax rule [40].

Let us now consider some special cases. We consider the trivial (0-1) loss-function formulation. For multivariate normal densities with equal covariance matrices, the Bayes rule with respect to equal prior probabilities allocates an observation $\underline{x}$ to that group h for which the squared Mahalanobis distance (or 'potential')

$$D_h^2(\underline{x}) = (\underline{x} - \underline{\mu}_h)^t \underline{\underline{\Sigma}}_{(h)}^{-1} (\underline{x} - \underline{\mu}_h) \tag{13}$$

is minimal. For arbitrary prior probabilities, the 'effective' distance is changed into:

$$D_h'^2(\underline{x}) = (\underline{x} - \underline{\mu}_h)^t \underline{\underline{\Sigma}}_{(h)}^{-1}(\underline{x} - \underline{\mu}_h) - 2\log \rho_h, \tag{14}$$

and for unequal covariance matrices into:

$$D_h''^2(\underline{x}) = (\underline{x} - \underline{\mu}_h)^t \underline{\underline{\Sigma}}_{(h)}^{-1}(\underline{x} - \underline{\mu}_h) - 2\log \rho_h + \log(\det(\underline{\underline{\Sigma}}_{(h)})). \tag{15}$$

As usual, superscript t is used to denote vector transposition. In general, it requires rather elaborated numerical procedures to determine the minimax rule, and for simplicity one calculates a Bayes rule with respect to equal prior probabilities, or according to prior probabilities that are proportional to the observed relative frequencies of occurrence of the two groups in the sample.

The discrimination surfaces between group 1 and group 2 are given by

$$D_2^2(\underline{x}) - D_1^2(\underline{x}) = c, \tag{16}$$

where c depends on the ratio of the prior probabilities, the ratio of the determinants of the covariance matrices $\underline{\underline{\Sigma}}_{(1)}$ and $\underline{\underline{\Sigma}}_{(2)}$, and, in general, also on the ratio of the losses ℓ_{21} and ℓ_{12} (at least if $\ell_{01} = \ell_{02} = \infty$). In general, these surfaces are quadratic surfaces, in practice ellipses or hyperbolae. If and only if $\underline{\underline{\Sigma}}_{(1)} = \underline{\underline{\Sigma}}_{(2)} = \underline{\underline{\Sigma}}$, they constitute parallel hyperplanes $\underline{w}^t \underline{x} = c'$, characterised by a vector of weights

$$\underline{w} = \underline{\underline{\Sigma}}^{-1}(\underline{\mu}_2 - \underline{\mu}_1). \tag{17}$$

In discriminant analysis, these are often called the weights of Fisher's linear discriminant function. They share a number optimality properties and interpretations. Weight vectors belong mathematically to the dual of the data-space $\mathbf{R}^p$. Often they are only important up to a multiplicative factor, so that, mathematically speaking, they constitute a $p-1$ dimensional projective space. Plotting the weight vectors in the original data-space $\mathbf{R}^p$, they mark directions on which the data can be perpendicularly projected. For any random vector $\underline{X}_i$ in $\mathbf{R}^p$, $\underline{w}^t \underline{X}_i$ divided by the length of the vector $\underline{w}$ is such a projection. We give three interpretations, which hold if the origin of the coordinate system in which the vector $\underline{w}$, given by Eq. (17) with $\underline{\underline{\Sigma}}$, $\underline{\mu}_1$, and $\underline{\mu}_2$ estimated by their sample analogues, is chosen somewhere on the line connecting the centers of gravity of the two groups.

(i) The above weights correspond to directions for which the ratio of the between-class variance and the sum of the within-class variances is maximal.

(ii) Considering data vectors $\underline{x}_1, \ldots, \underline{x}_p$ in the data space $R^{(n_1+n_2)}$, where n_1 and n_2 denote the sample sizes, $w_1\underline{x}_1 + \cdots + w_p\underline{x}_p$ can be viewed as that linear combination (up to a proportionality factor) which maximises the (sample) correlation coefficient with the vector $\underline{y}$ denoting the class membership. As the correlation coefficient is independent of location and

scale transformations of $\underline{y}$ and (uniformly) of $\underline{x}_1, \ldots, \underline{x}_p$, the coding of the class membership and any proportionality factor in $\underline{w}$ are immaterial.
(*iii*) The weights in Eq. (17) can be obtained (up to a linear transformation) by linear regression of the vector $\underline{y}$ on $\underline{x}_1, \ldots, \underline{x}_p$. If the vector $\underline{y}$ is coded as $c_1 = -\frac{n_2}{n_1+n_2}$ for group 1 and $c_2 = c_1 + 1$ for group 2, then precisely the weights in Eq. (17) are formally recovered by the linear regression. (Note that the usual probabilistic assumptions used in linear regression models are different.) This equivalence, which is useful in computational practice, goes back to R.A. Fisher [41], see also [27] for a vivid practical account.

Discriminant analysis of $k > 2$ groups can be based on separate discriminant analyses between each of the $k(k-1)/2$ pairs of groups. However, the practical complexity of this approach increases rapidly with k. A more parsimoneous description is given by canonical correlation analysis between $k-1$ binary group membership codes and the explanatory data vectors $\underline{x}_1, \ldots, \underline{x}_p$. This is implemented in the procedure DISCRIM of the package SAS [38]. (For two groups, it reduces to multiple linear regression analysis of a single group-membership vector.) In view of the general situation, the regression coeffcients are often called canonical coefficients. They depend on how the vectors $\underline{x}_1, \ldots, \underline{x}_p$ are (separately) normalised. Often, one divides by the square root of the diagonal elements of some suitable covariance matrix. For instance, the 'total sample' covariance matrix or the 'pooled within-class' covariance matrix. The latter is estimated by $(n_1 + n_2 - 2)^{-1}(\underline{\underline{S}}_{(1)} + \underline{\underline{S}}_{(2)})$, where $\underline{\underline{S}}_{(1)}$ and $\underline{\underline{S}}_{(2)}$ are the sum of squares and cross-product (SSCP) matrices, corrected for the mean.

For simplicity, we shall consider in the practical sections only the 'parametric case', with multivariate normal densities, the 'non-parametric case', with arbitrary multivariate densities, and the 'multinomial independence' model, with a discretised data structure (and a global association factor).

In practice, the densities $f_h(\underline{x})$ have to be estimated. In the parametric case, this is done by substituting the standard estimators for $\underline{\mu}$ and $\underline{\underline{\Sigma}}$. In the non-parametric case, this can be done by so-called non-parametric density (e.g. kernel) estimators. For a discussion of such type of estimators, we refer to [42] (which describes mainly the univariate situation) and to the description of PROC DISCRIM in the SAS manual [38] and the references therein. The third approach starts with discretising the distribution by dividing each axis into k, say 5 or 10, intervals. In general, one then obtains a multinomial distribution. However, as the number free parameters, $(k^p - 1)$ with p the number of discriminating variables, quickly exceeds the number of available data points, one has to make further simplifying assumptions. One possibility is to impose the multinomial independence model, which approximates the distribution by the product of the marginal multinomial distributions. In that case the number of free parameters is $p \times (k-1)$. It obviously only works well if the original continuous distribution can be thought to be generated by independent random variables, which will usually not be the case in practice. In the case of two groups, one could discretise the 'joint

principal components' [27], which are at least uncorrelated, and then transform back to the original scale. The multinomial independence model is implemented in the program INDEP [36, 43].

3 PRACTICE

Bilder bergen die Gefahr,
uns gefangenzuhalten
Wittgenstein, Philosophische Untersuchungen

3.1 Description of the Dataset and Preliminary Analysis by Visual Inspection

We define two classes of H-mode plasmas, class-1 (without ELMs and with giant ELMs) and class-2 (with small ELMs). We look for the boundary in plasma parameter space which separates the regime of class-1 discharges from that of class-2 discharges.

The ITERH.DB1 dataset contains data from ASDEX, DIII-D, JET, JFT-2M, PBX-M, and PDX. It consists of about 2000 H-mode timeslices ('observations') of some 1000 discharges. For each timeslice, 76 global plasma parameters ('variables') have been included. Details of the dataset and its variables can be found in [8]. One of those variables is a label of the type of H-mode: ELM-free, with small ELMs, and with giant ELMs (denoted as H, HGELM, and HSELM, respectively). The characterisation of a timeslice as H, HGELM or HSELM has been made by each experimental group providing the data. The distinction between HSELM and HGELM was based on the height of the spikes in the H_α signal. If it was lower than or equal to the H_α level during the L-mode, the timeslice was labeled 'HSELM'. From the H-mode database, we dropped the pellet shots, and kept only the semi-stationary H-mode timeslices, for which $\dot{W}_{mhd}$ and, if available, $\dot{W}_{dia}$ are between -0.05 and $+0.35$ times the injected power. Otherwise we applied none of the restrictions used in the 'standard dataset' used in [8].

We do not intend to present here an exhaustive practical analysis. We selected subsets from the original dataset, each consisting of shots in a single magnetic configuration (either DN or SN) from one of the machines ASDEX, JET, or JFT-2M. The reason for this is that DN and SN seems indeed to make a clear difference. In some of the six possible groups, the plasma parameters have been varied rather little or only a few datapoints were available. Hence, for convenience, we restrict attention to the ASDEX (DN), JET(SN), JFT2M (SN) data in the database. The total numbers of observations for each tokamak are given in the following summary table.

In addition, in Figs. 1 to 3, we made a graphical representation showing the distributions of class-1 and class-2 discharges in plasma parameter space for the three tokamaks. Attention is restricted to discharges on (a) the $P_{inj} - I_p$ plane, (b) the $P_{inj} - \bar{n}_e$ plane, (c) the $\bar{n}_e - I_p$ plane, and (d) the Seplim/a$-P_{inj}$ plane. (P_{inj} is the injected power, I_p is the plasma current,

$\bar{n}_e$ is the line averaged plasma density, a is the minor radius, and Seplim is the distance between the separatrix and either the vessel or the limiter.) The symbols $\cdot$, $\star$, and $\odot$ denote ELM-free H-mode (i.e. H), H-mode with giant ELMs (i.e. HGELM) and H-mode with small ELMs (i.e. HSELM), respectively. The symbols $\cdot$ and $\star$ correspond to class-1 shots, whereas the symbol $\odot$ corresponds to class-2 shots.

Device	*Number of Observations*	
ASDEX	351	(only DN: 206)
JET	522	(only SN: 466)
JFT-2M	384	(only SN: 373)
total	1257	(main subsets: 1045)

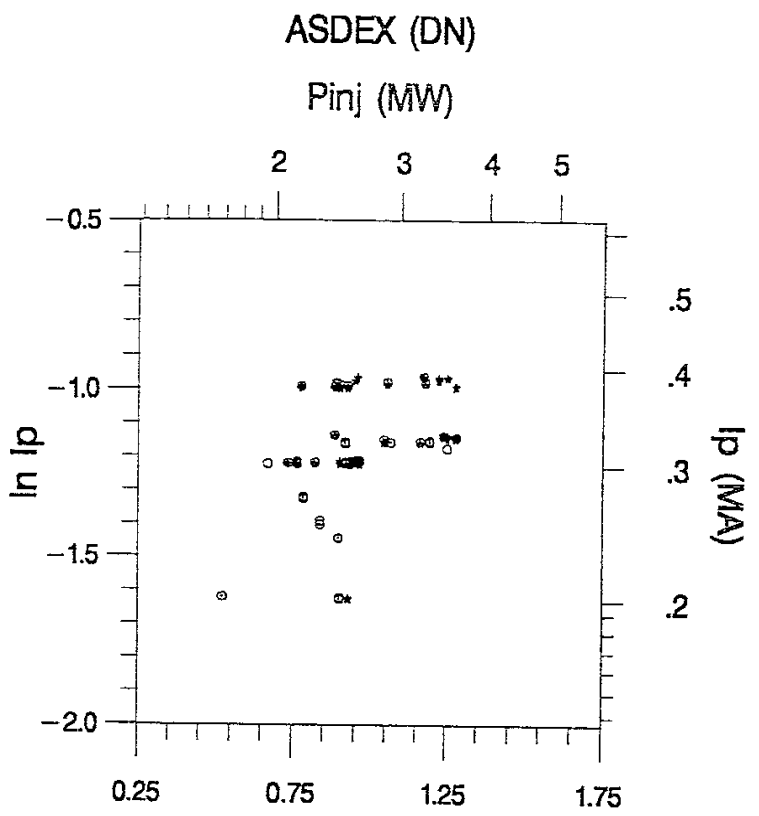
Figure 1 gives the plots for ASDEX. On first sight, it appears from this figure that the two classes considerably overlap. Some distinction, however, seems possible if we look more carefully. From Fig. 1 (a), we see that H-mode with small ELMs (i.e., class-2) are predominant in the region of low P_{inj} and low I_p. For higher P_{inj} and I_p both classes occur, but class-1 occurs more frequently. From Fig. 1 (b) one can see that in the $P_{inj} - \bar{n}_e$ plane, the domain of class-1 discharges surrounds the region HSELM, except at low P_{inj}. The plot of the distribution in the $\bar{n}_e - I_p$ plane (c), indicates that the density is of discriminatory value in addition to the current. (The figure suggests that no class 1 discharges occur at low current, at least if the density is not very low.)

Figure 2 gives the plots for JET. From Fig. 2 (a), one can see that HSELM discharges tend to occur at high current (at any medium to high input power) and at medium current (3 MA) only for low input power. From Fig. 2 (b), one can see that HSELM and HGELM, though they occur less often than ELM-free H-mode discharges, are well scattered over the density and power range. At low density and power, there tends to be a region where ELM-free discharges are less likely to occur. HGELM does not seem to occur for $\bar{n}_e < 1.5 P_{inj}^{0.4}$. From Fig. 2 (c), it appears that also in the case of JET, the combination of current and density gives a better discrimination than each of the variables alone. The HSELM discharges do not occur at low current (2 MA).

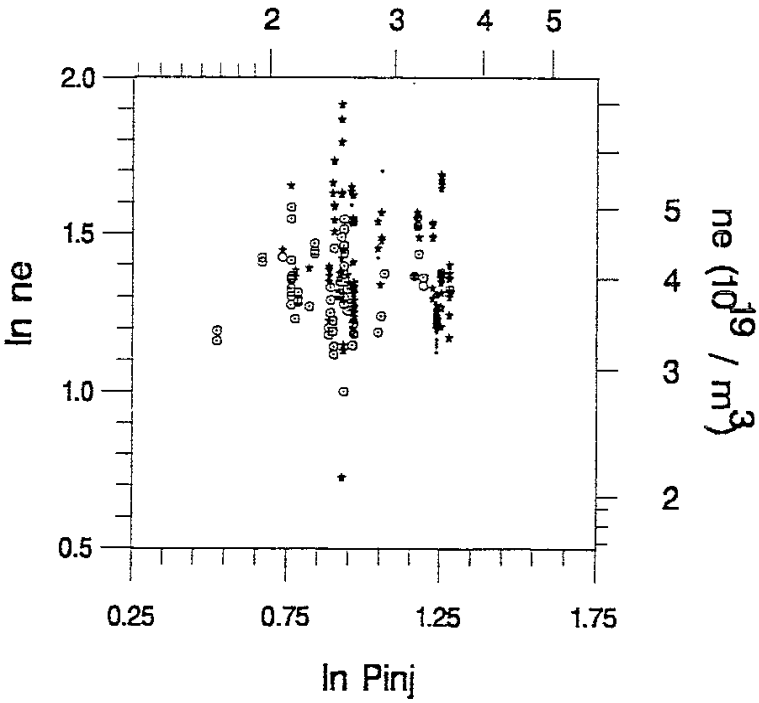
Figure 3 gives the plots for JFT-2M. In the present dataset, no discharge from JFT-2M showed large ELM's. From Fig. 3 (a) it appears that HSELM discharges do not tend to occur at low current and at low power. Otherwise, they are fairly mixed with the ELM-free H-mode discharges. In Figure 3 (b), the HSELM discharges are also rather dispersed, but tend to avoid the low density region. One might ask whether this may be due to a (high) correlation between density and current. From Fig. 3 (c) one can see that this is not the case, so that in summary, HSELM at JFT-2M seems to avoid the region of low values for I_p, P_{inj} as well as $\bar{n}_e$. In Fig. 3, the deuterium injected ($D \rightarrow H$) discharges have been marked by symbols that are about a factor 1.5 larger than those for the hydrogen injected ($H \rightarrow H$) discharges.

Fig. 1. ASDEX Double Null (DN), $H \rightarrow D$ discharges. The symbols $\cdot$, $\star$, and $\odot$ indicate ELM-free H-mode, H-mode with Giant ELMs, and H-mode with small ELMs, respectively. The first two types of H-mode belong to class-1 and the last type of H-mode to class-2. The data are projected onto the following planes: (a) $P_{inj} - I_p$, (b) $P_{inj} - \bar{n}_e$, (c) $\bar{n}_e - I_p$, and (d) Seplim/a-P_{inj}. Seplim/a denotes the distance between the separatrix and the vessel, normalised by the minor radius.

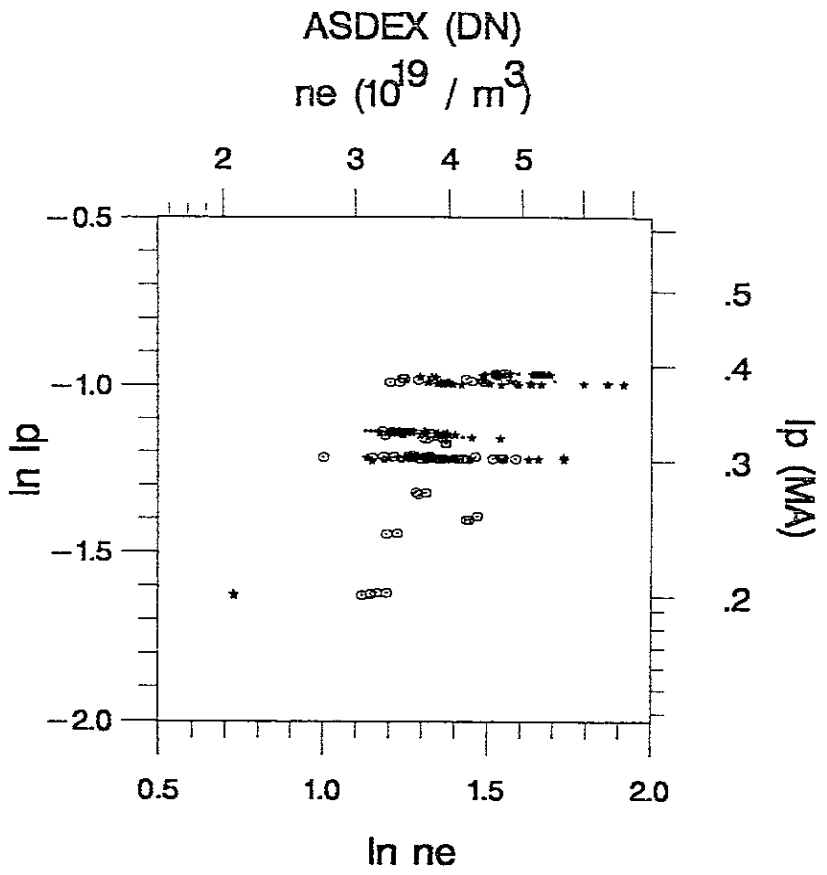
(a)



(b)



(c)


ASDEX (DN)
ne (10^{19} / m^3)
2
3
4
5
−0.5
−1.0
−1.5
−2.0
ln Ip
.5
.4
.3
.2
Ip (MA)
0.5
1.0
1.5
2.0
ln ne

(d)

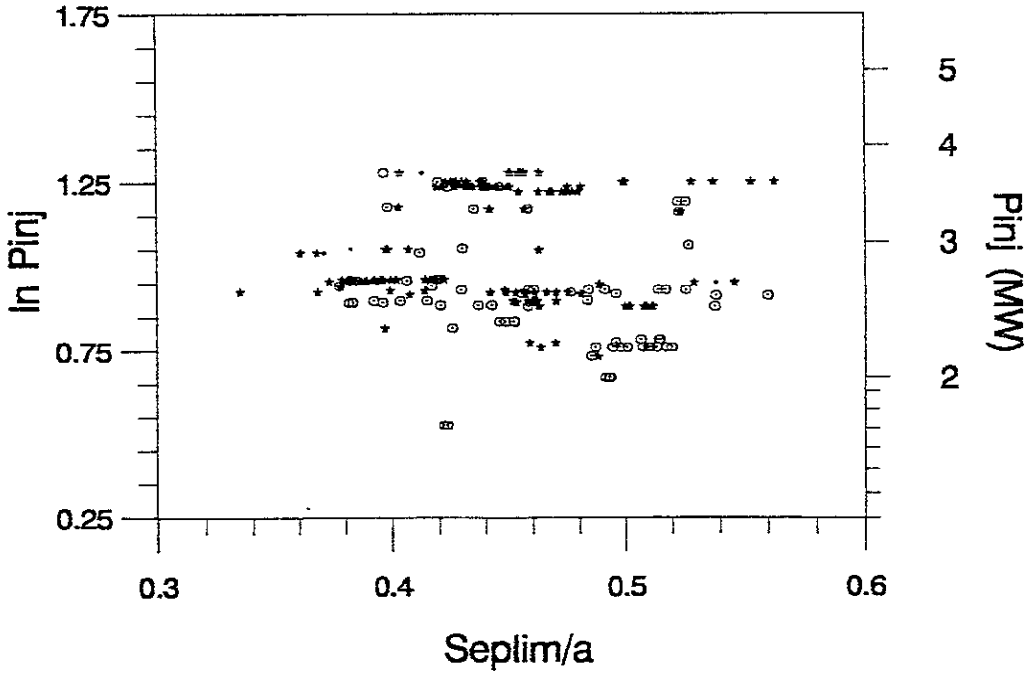
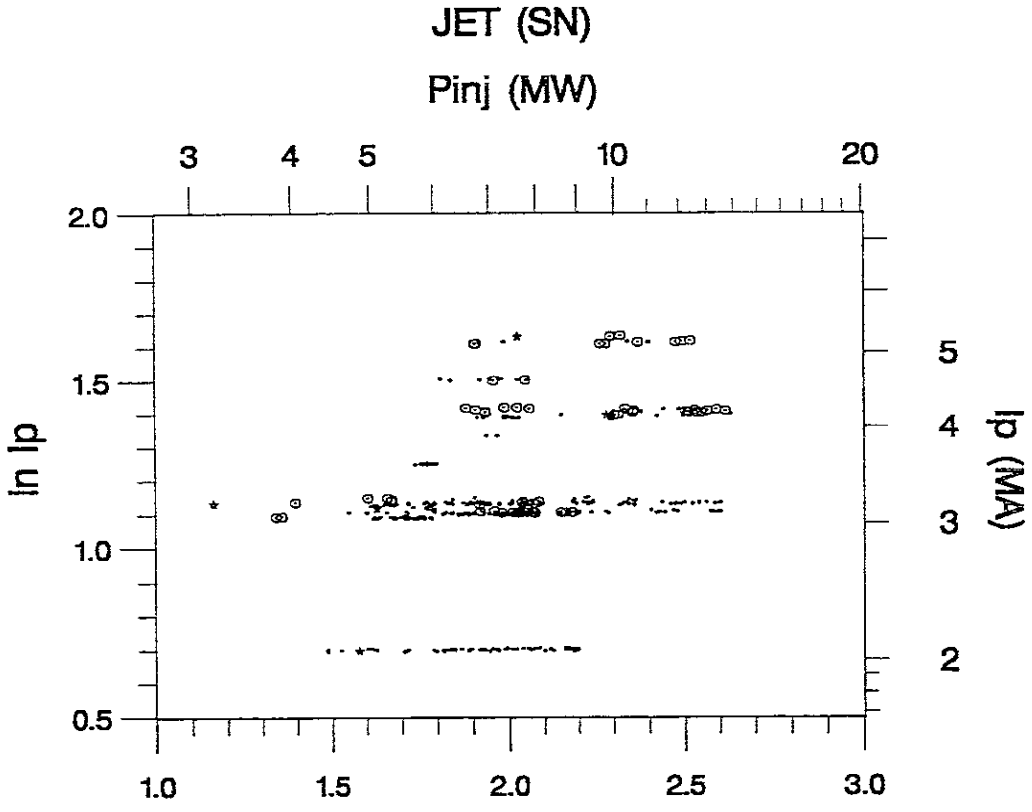
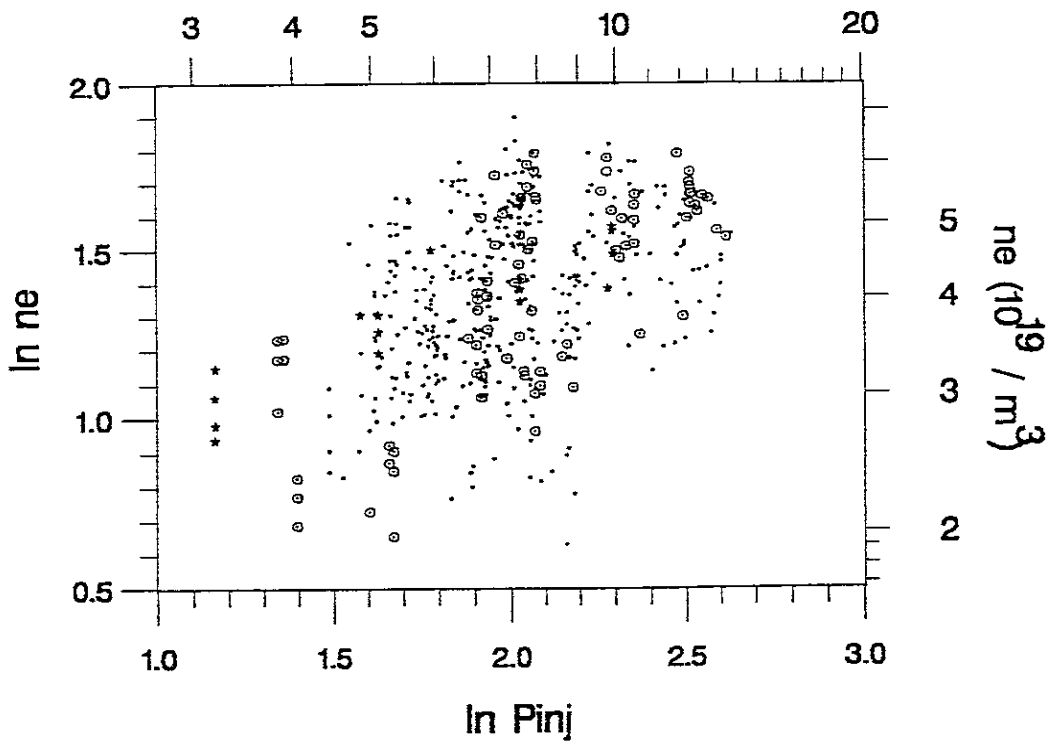

1.75
1.25
0.75
0.25
ln Pinj
5
4
3
2
Pinj (MW)
0.3
0.4
0.5
0.6
Seplim/a

Fig. 2. JET Single Null (SN), $D \rightarrow D$ discharges. Seplim/a denotes the distance between the separatrix and the limiter, normalised by the minor radius. For further explanation, see Fig. 1.

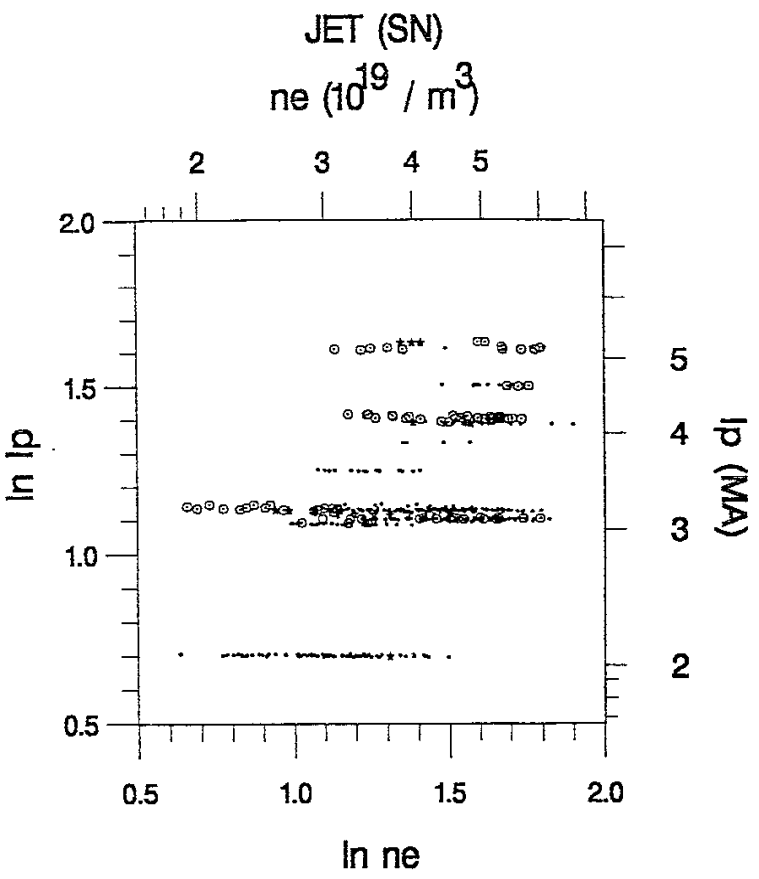
(a)



(b)



(c)


JET (SN)
ne (10^{19} / m^3)
2
3
4
5
2.0
1.5
1.0
0.5
ln Ip
5
4
3
2
Ip (MA)
0.5
1.0
1.5
2.0
ln ne

(d)

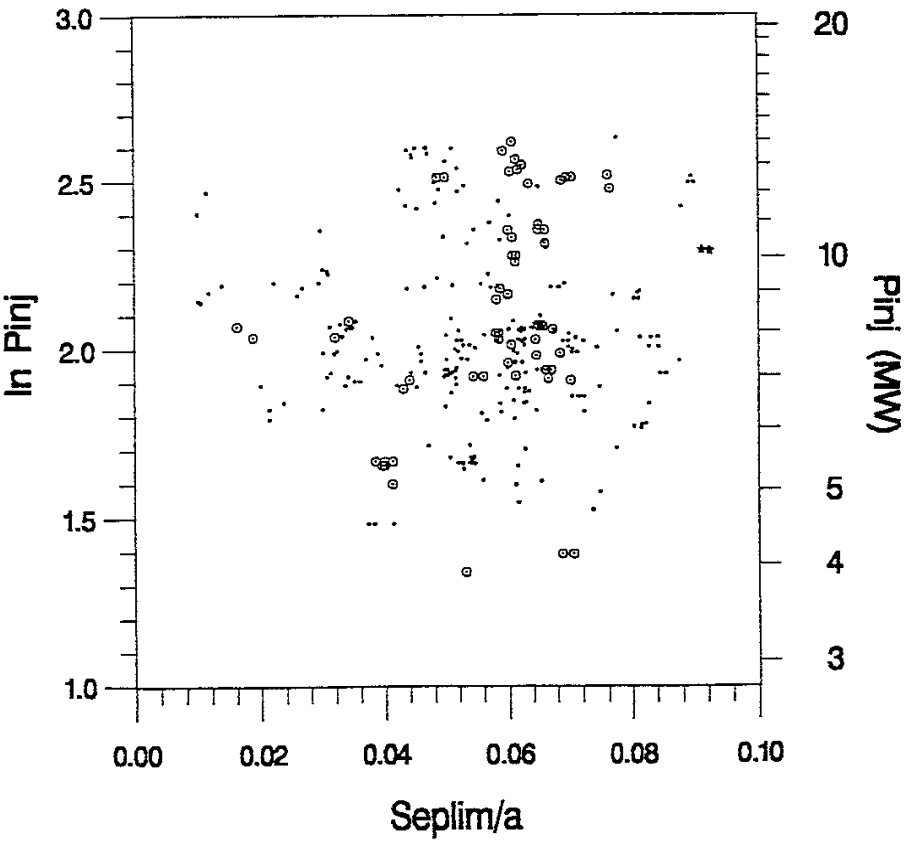
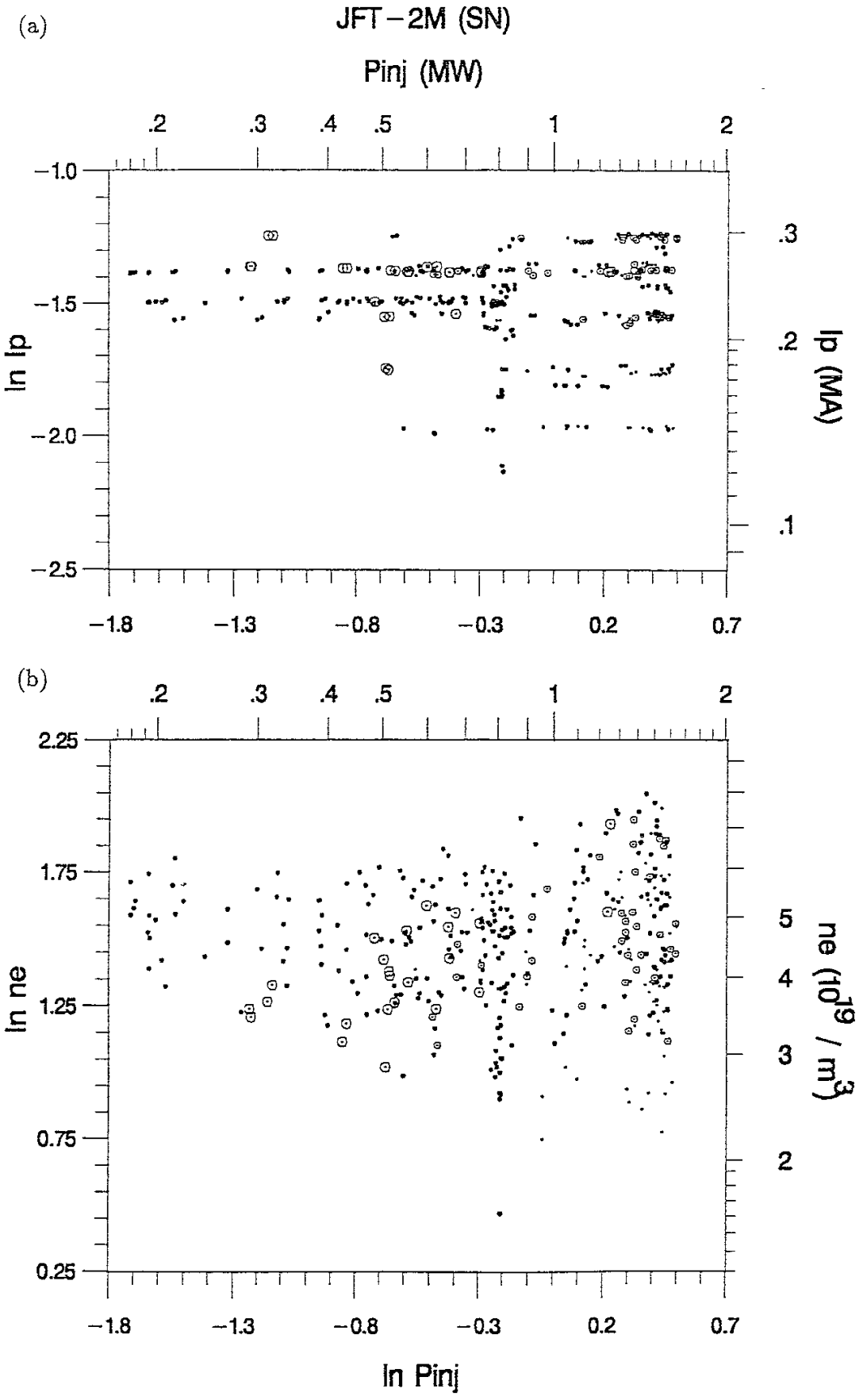
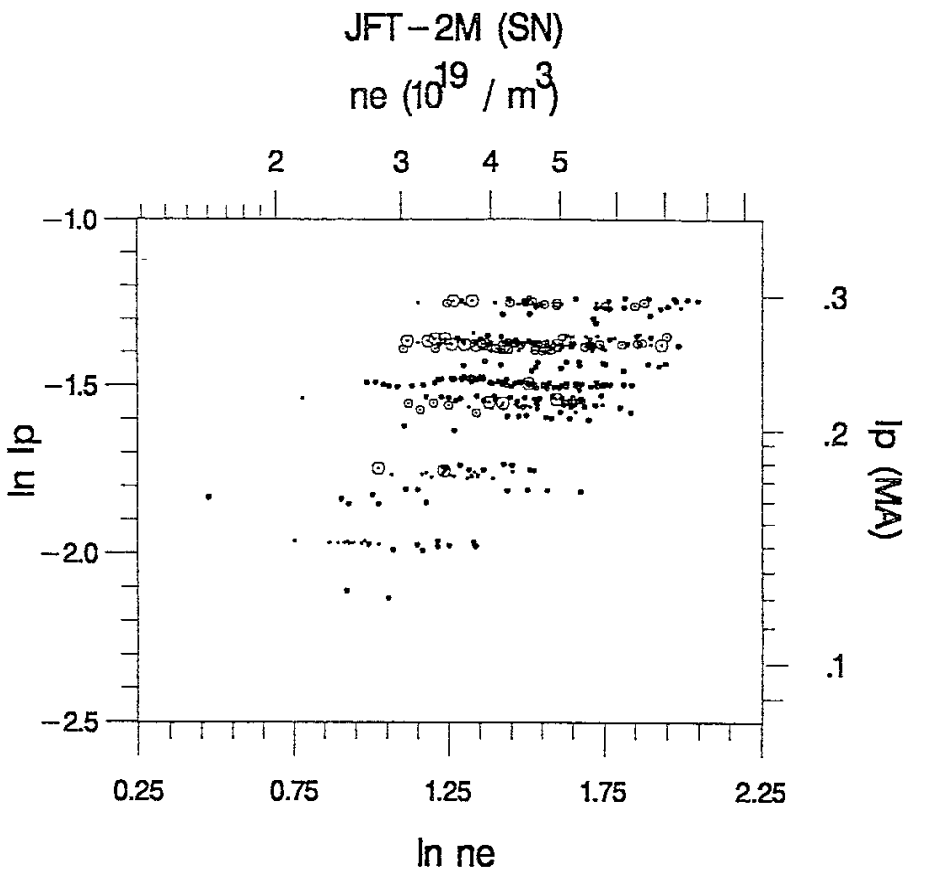

3.0
2.5
2.0
1.5
1.0
ln Pinj
20
10
5
4
3
Pinj (MW)
0.00
0.02
0.04
0.06
0.08
0.10
Seplim/a

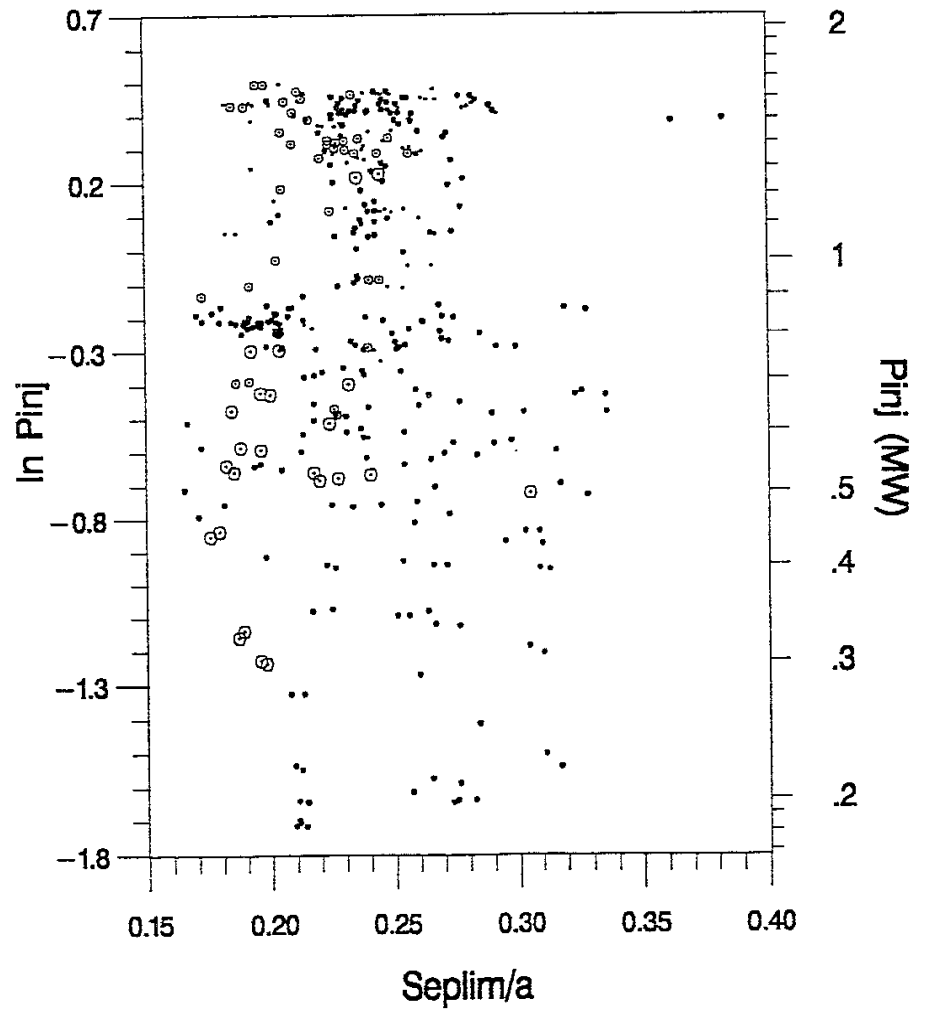
Fig. 3. JFT-2M Single Null (SN) discharges. The smaller variant of each symbol denotes $H \rightarrow H$, and the larger variant $D \rightarrow H$. Seplim/a denotes the distance between the separatrix and the limiter, normalised by the minor radius. For further explanation, see Fig. 1.



(c)


JFT – 2M (SN)
ne (10^{19} / m^3)
2
3
4
5
−1.0
−1.5
−2.0
−2.5
ln Ip
.3
.2
.1
Ip (MA)
0.25
0.75
1.25
1.75
2.25
ln ne

(d)


0.7
0.2
−0.3
−0.8
−1.3
−1.8
ln Pinj
2
1
.5
.4
.3
.2
Pinj (MW)
0.15
0.20
0.25
0.30
0.35
0.40
Seplim/a

The latter discharges turn out to have been performed with relatively high power. For both types of discharge it appears that HSELM does not occur at low I_p, nor at low $\bar{n}_e$.

The above interpretation of the graphs should be viewed as tentative, because of an incompleteness and an essential limitation.

The incompleteness is that other plasma variables, such as the magnetic field, and, sofar, the distance of the plasma to the wall have not been considered. From other investigations, the last variable is expected to influence the occurrence of ELM's at least at ASDEX and JFT-2M. Here, as an example, it is plotted against $\ln P_{inj}$ in the Figs. 1 (d) - 3 (d). It can be seen that for JET and JFT-2M, there is an upper limit of this distance, above which no HSELM is found (at Seplim / a $\approx$ 0.075 and 0.25, respectively). For ASDEX, the distance (which is measured to the wall and not to the limiter) seems, according to this 2-D projection, not to have a clear influence on the occurence of HSELM. The reader is referred to Section 3.3.1, where the influence of this distance is discussed when q_{cyl}, the plasma currrent and the time since the L-H transition are also (simultaneously) taken into account.

From this it is clear that interpreting for each tokamak all 10 graphs from all of the above 5 plasma variables in a similar way still has the limitation that only two-dimensional projections are considered. (And from these, only those on the coordinate planes of the variables considered.) The above graphical approach, how illustrative and supportive it may be, is limited by the fact that we cannot easily represent the data in more than 2 or 3 dimensions. Of course, an importance objective is to try and find lower dimensional subspaces and surfaces in $\mathbf{R}^5$ that separate the data. In the following we will do this systematically by applying discriminant analysis.

3.2 Discriminant Analysis using Four Instantaneous Plasma Parameters

We first perform discriminant analysis taking instantaneous plasma parameters (on logarithmic scale) as discriminating variables. Choosing the engineering parameters P_{inj} (injected power of the neutral beams), I_p (plasma current), B_t (toroidal magnetic field), and $\bar{n}_e$ (line averaged electron density), we define

$$X_P = \ln P_{inj} \quad X_I = \ln I_p \quad X_B = \ln B_t \quad X_n = \ln \bar{n}_e \tag{18}$$

The variables P_{inj}, I_p, B_t, and $\bar{n}_e$ are assumed to be measured in units of MW, MA, Tesla, and $10^{19}/m^3$, respectively. We use the symbol $S_I = \{P, I, B, n\}$ to denote the set of indices corresponding to these instantaneous plasma parameters. The above choice of variables implies that we assume, for the time being, that the probability of appearance of each class of H-mode is not influenced by the history of the discharge. (The influence of the discharge history, or 'plasma memory', will be discussed in Section 3.3.)

Two discrete variables in the dataset that may influence the occurrence of ELMs require special attention. One is the equilibrium configuration: there exist single-null (SN) and double-null (DN) discharges. In view of the possible qualitative differences associated with these configurations, we analyze SN and DN configurations separately. The other variable is the gas mixture. We label the type of gas by the parameter A. A distinction is made between proton (H, A=1), deuteron (D, A=2), and mixed proton-deuteron (A is assumed to be 1.5) plasmas. Table 1 gives a break-down of the total number of discharges with respect to device, configuration and ion species. One can see that in this respect each Tokamak tends to concentrate somewhat on its own 'favourite' type of discharges.

Table 1 — H-mode Database (from ITERH.DB1)
Numbers of timeslices per tokamak, gas composition, and divertor configuration.

	Gas=1		Gas=1.5		Gas=2	All
	SN	DN	SN	DN	SN	
ASDEX	•	206	145	•	•	351
JET	•	•	•	56	466	522
JFT2M	104	11	269	•	•	384
Total	104	217	414	56	453	1257

Gas: 1=H, 2=D, 1.5 = mixture H and D

In our analysis we consider only DN discharges from ASDEX and SN discharges from JET. The reason is that in the case of ASDEX SN, the class-2 (i.e. HSELM) discharges in the database occur only at a particular value of the magnetic field, whereas for JET all of the DN discharges in the database belong to class-1. An analysis with the same variables is difficult for ASDEX (SN) and JET (DN). We therefore choose the ASDEX (DN) and JET (SN) subsets.

We first apply parametric discriminant analysis using quadratic discriminant functions. Subsequently, non-parametric discriminant analysis is performed: (1) by SAS using uniform kernel estimates (with a certain radius that determines the amount of smoothing) for the probability densities, and (2) by the program INDEP which uses a product multinomial model for the discretised plasma variables (for each variable 10 classes have been taken, roughly according to the deciles of the distribution).

In each case, a (mis-) classification summary is given, which counts the numbers of correctly and of the various types of incorrectly classified observations in the present dataset. This gives information about the performance of the procedure. If the summary would be based on the discriminant function fitted to all data, it would yield an optimistic view (bias) of the performance of the procedure since the same data are used twice. Therefore, the leaving-one-out method (also called 'jackknife') has been used, which classifies the j^{th} observation

according to a discriminant function based on all observations except for observation j. This largely eliminates the bias (but increases the variance of the discriminant function estimates). If the performance of the parametric procedure is close to those of the non-parametric procedures, then, at least for the set of plasma variables considered, the assumption of a quadratic form of the boundary seems to be appropriate, and we have good explicit formulas describing the region in which we can expect small ELMs.

3.2.1 ASDEX (DN)

The basic frequency table describing the occurrence of ELMs in the dataset is

	non-HSELM	HSELM	Total
Number of observations	134	72	206
Ratio (%)	65	35	100

The distributions of the two groups of data with respect to the variables $X_i (i \in S_I)$ have to be estimated. Therefore, we present the centers of mass, the standard deviations, and the correlations for both groups, see Table 2. (Together, they determine the first two moments of the distributions.)

Table 2 — ASDEX (DN) data

(a) Mean values and standard deviations

	non-HSELM, N=134		HSELM, N=72			
log: ↓	Mean	SD	Mean	SD	T	F
P_{inj}	1.07	0.16	0.91	0.17	6.7	•
I_p	–1.10	0.12	–1.22	0.16	5.4	**
B_t	0.77	0.07	0.77	0.07	–0.8	•
$\overline{n}_e$	1.39	0.19	1.33	0.12	3.0	***

***: P < 0.001, **: 0.001 < P < 0.01, •: 0.4 < P < 0.6

(b) Correlation coefficients

		non-HSELM, (STD_0= 0.09)		HSELM, (STD_0= 0.12)	
log: ↓	log: →	P_{inj}	I_p	B_t	$\overline{n}_e$
P_{inj}		1	0.42	0.18	0.02
I_p		0.14	1	0.31	0.32
B_t		–0.22	0.16	1	0.13
$\overline{n}_e$		–0.23	0.60	0.19	1

The sample correlation coefficients of the HSELM class are displayed in the right upper corner and those of the non-HSELM class in the left lower corner.
STD_0 is one standard deviation of the sample correlation coefficient under the hypothesis of no actual correlation.

From the table, one can see that HSELM occurs at lower current and lower injected power than non-HSELM. One can also see that the correlations are not very large, the largest being

$r = 0.6$ between $\log I_p$ and $\log \bar{n}_e$ in the non-HSELM group. (However many of them are more than two standard deviations different from zero, and not the same for the HSELM as for the non-HSELM group.) The other columns of Table 2 (a) quantify how significant the differences of the mean values and the standard deviations are between the two groups. For each plasma variable X_i, $i = 1, \ldots, 4$, the difference in mean value between the two classes is to be compared with the standard deviation of this difference, which is estimated by $\hat{\sigma}_i = \sqrt{SD_{i,1}^2/N_1 + SD_{i,2}^2/N_2}$, where N_1 and N_2 denote the sample sizes. If this difference is significant, i.e., if $T_i = |\bar{X}_i^1 - \bar{X}_i^2|/\hat{\sigma}_i > 2$, then this variable is effective for (univariate) discrimination between the two classes. (The over-bar is used to denote the mean value and the second subscript j to denote the class: 1 = non-HSELM, 2 = HSELM). The t-values T_i are displayed in the fifth column of Table 2 (a). In fact, when the null-hypothesis that the two variances are equal was rejected, in column F denoted by one or more *'s, the t-value was calculated using separate estimates for σ_1 and σ_2 [38]. (The corresponding loss in degrees of freedom for the approximate t distribution is negligible in this case.)

So, on the basis of this statistical analysis, we can infer from the table that X_P and X_I are significantly larger in the non-HSELM class than in the HSELM class. This is not the case for X_B and X_n. However, the variation of the density is significantly larger for the non-HSELM class than for the HSELM class. These observations are in accordance with the visual inspection of Fig. 1.

Of course, the above considerations are only univariate, taking the 'influence' of only one variable at a time into account. (Due to correlations, such 'influences' may be confounded with those of other variables.)

The simultaneous influence of P_{inj}, I_p, B_t, $\bar{n}_e$ is investigated by performing discriminant analysis for these variables (on logarithmic scale). From the output of the program DISCRIM of SAS [38], we will discuss (*i*) the distance matrix between the (two) groups, (*ii*) the standardised canonical coefficients, and (*iii*) the (mis-) classification tables for quadratic and non-parametric discriminant analysis.

The (squared) distances between the two groups are calculated using as metrics (yardsticks) the covariance matrices of both groups of data. These distances are summarised in a matrix $\underline{\underline{\hat{D}}}$, which allows an effective generalisation to k groups. Specifically,

$$\hat{D}_{ij} = (\underline{\bar{X}}_i - \underline{\bar{X}}_j)^t \underline{\underline{\hat{\Sigma}}}_{(j)}^{-1} (\underline{\bar{X}}_i - \underline{\bar{X}}_j) \tag{19}$$

is used as an estimator of

$$D_{ij} = (\underline{\mu}_i - \underline{\mu}_j)^t \underline{\underline{\Sigma}}_{(j)}^{-1} (\underline{\mu}_i - \underline{\mu}_j). \tag{20}$$

For ASDEX (DN), the matrix $\underline{\underline{\hat{D}}}$ equals

$$\underline{\underline{\hat{D}}} = \begin{pmatrix} 0 & 1.38 \\ 1.72 & 0 \end{pmatrix}. \tag{21}$$

The fact that $\hat{D}_{1,2}$ is smaller than $\hat{D}_{2,1}$ means that the estimated variance along the line connecting the means, which is (for group j) proportional to

$$(\underline{\bar{X}}_1 - \underline{\bar{X}}_2)^t \underline{\underline{\hat{\Sigma}}}_{(j)} (\underline{\bar{X}}_1 - \underline{\bar{X}}_2),$$

has a smaller value for group 1 than for group 2. This implies that $\underline{\underline{\hat{\Sigma}}}_1$ and $\underline{\underline{\hat{\Sigma}}}_2$ are unequal. Of course, the difference between $\hat{D}_{1,2}$ and $\hat{D}_{2,1}$ is not tremendously large. The question is whether it is statistically significant, i.e. whether $D_{1,2} = D_{2,1}$. Assuming multivariate normal distributions for the two samples and neglecting the sampling variation of $\underline{\bar{X}}_1$ and $\underline{\bar{X}}_2$, the ratio between $D_{2,1}$ and $D_{1,2}$ is asymptotically (for $N \gg p$) distributed as an F-distribution with N_2 and N_1 degrees of freedom [25]. In our case, $N_1 = 134$, $N_2 = 72$, and $p = 4$. The critical value of the corresponding F-distribution at 5% is about 1.4. The ratio between $D_{1,2}$ and $D_{2,1}$ being $1.72/1.38 \simeq 1.25$, the effect is not statistically significant at the 5% level. Hence, from that point of view, one cannot conclude that $\underline{\underline{\Sigma}}_1$ and $\underline{\underline{\Sigma}}_2$ are unequal. (Note that taking the sampling distribution of $\underline{X}_1$ and $\underline{X}_2$ into account will lead to an even larger critical value and hence also not to statistical significance of the value 1.25.) Of course, this univariate consideration works only in one direction: the fact that $D_{1,2} = D_{2,1}$ does not imply at all that $\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2$. A multivariate test is required to determine whether or not the hypothesis $\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2$ can be accepted. In fact, the modified likelihood ratio statistic (see e.g, [26], Ch. 7.4), which compares the pooled covariance matrix with the individual ones, and programmed in SAS [38], gives a highly significant result: $45 > \chi^2_{10,.05} = 18$. Hence, the null-hypothesis $\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2$ is indeed rejected in favour of $\det(\underline{\underline{\Sigma}}_2) < \det(\underline{\underline{\Sigma}}_1)$, though along the line connecting the two centers of mass, the two distributions do not have a significantly different spread.

The pooled within-class standardised canonical coefficients are given by

$$\underline{\hat{C}} = \begin{pmatrix} 0.72 \\ 0.45 \\ -0.19 \\ 0.25 \end{pmatrix}. \tag{22}$$

The vector $\underline{\hat{C}} = (\hat{C}_P, \hat{C}_I, \hat{C}_B, \hat{C}_n)^t = (0.72, 0.45, -0.19, 0.25)^t$ is proportional to

$$D_{ws}^{1/2} \underline{\underline{\hat{\Sigma}}}_{ws}^{-1} (\underline{\bar{X}}_1 - \underline{\bar{X}}_2), \tag{23}$$

where $\underline{\underline{\hat{\Sigma}}}_{ws} = (n_1 + n_2 - 2)^{-1}((n_1 - 1)\underline{\underline{\hat{\Sigma}}}_1 + (n_2 - 1)\underline{\underline{\hat{\Sigma}}}_2)$ is the pooled within-class covariance matrix and D_{ws} its diagonal part. The subscript denotes the class of discharge (1=non-HSELM, 2=HSELM). The word standardised means in this case that the coefficients pertain to the plasma variables $\ln P_{inj}$, $\ln I_p$, $\ln B_t$, $\ln \bar{n}_e$ normalised by their (pooled within sample) standard deviations. For that reason, the unstandardised (raw) coefficients are multiplied

by $D_{ws}^{1/2}$, i.e. by the standard deviations of the corresponding plasma variables. As we have only two groups, the canonical coefficients are (up to a multiplicative constant) the same as the weights of Fisher's discriminant functions, and also the same as the coefficients obtained from linear regression of the group membership vector (arbitrarily coded). The t-values (the ratio between the estimated coefficients and their estimated standard deviations), calculated from standard linear regression, are $(4.9, 2.6, -1.3, 1.5)$. This indicates that C_P and C_I are significantly (at least 2 standard deviations) different from zero, whereas C_B and C_n are not.

Linear discriminating boundaries are in terms of the standardised plasma variables formed by the hyperplanes perpendicular to the vector $\underline{\hat{C}}$, and in terms of the unstandardised plasma variables by the hyperplanes perpendicular to $\underline{\bar{X}}_1 - \underline{\bar{X}}_2$ in the metric given by $\underline{\underline{\hat{\Sigma}}}_{ws}$. The pooled within-class covariance matrix is used as a metric to measure distances and angles. This should be viewed as a approximate data-descriptive procedure, since we know that the hypothesis $\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2$ is rejected by the data, in which case no single metric is satisfying from all points of view, and some improvement is expected by considering quadratic boundaries. At least, from the vector $\underline{\hat{C}}$ one can see that the (standardised) variables X_P and X_I are more important for discrimination than the (standardised) variables X_B and X_n, and that small ELMs ($j = 2$) are to be sought in the region of low current and injected power. (More precisely, for low values of $\underline{\hat{C}}^t \underline{X}$, where $\underline{X}$ is the (column-) vector of standardised variables X_p, X_I, X_B, X_n. However, the coefficients for X_B and X_n are not significantly different from zero.)

Table 3 (a) shows the jackknifed ('cross-validated') resubstitution summary of the data in the data set. Now, a quadratic discriminant function has been used. In the 2×2 performance table, the (1,2)-component is small (14.9%), but the (2,1)-component is large (47.2%). This result indicates that most of the non-HSELM discharges are correctly classified as 'non-HSELM', whereas a relatively large fraction of the HSELM discharges is misclassified as 'non-HSELM'. This is because, as is seen in Fig. 1 (a), except for two almost identical outlying points at low current, the non-HSELM discharges are observed in a fairly narrow region, $\mathbf{R}_1$, in parameter space, where also SELM discharges are possible. Below this region is a region, $\mathbf{R}_2$, where only SELM discharges are possible. This result allows us to avoid the region where the non-HSELM discharges occur, which would be the more dangerous ones from the viewpoint of sustaining good confinement under stationary conditions.

The few outlying HGELM points (which are, according to the agreed definition, indeed HGELM, but somewhat borderline with HSELM), make the dataset from a statistical point of view an interesting example to try and compare various robust versions of discriminant analysis (based on concepts in [44, 45]), which downweight the points that do not conform to the majority. We shall refrain here from doing so. To ease our conscience, we 'spied' at the additional ASDEX data from the ITERH.DB2 dataset. There, an ELM-free point occurs very close to the outlying HGELM points, indicating that that spot does not belong to the

Table 3 — Classification performance for various models
ASDEX (DN)

Prior probability: HSELM = 0.35

True class	Allocated class: non-HSELM	HSELM	OTHER	TOTAL
a) Quadratic boundaries				
non-HSELM	114	20		134
Row %	85.1	14.9		
HSELM	34	38		72
Row %	47.2	52.8		
Total	148	58		206
Row %	71.8	28.2		
b1) Kernel density estimation (r=1)				
non-HSELM	130	4	0	134
Row %	97.0	3.0	0.0	
HSELM	43	29	0	72
Row %	59.7	40.3	0.0	
Total	173	33	0	206
Row %	84.0	16.0	0.0	
b2) Kernel density estimation (r=0.5)				
non-HSELM	116	9	9	134
Row %	86.6	6.7	6.7	
HSELM	34	32	6	72
Row %	47.2	44.4	8.3	
Total	150	41	15	206
Row %	72.8	19.9	7.3	
c) Multinomial independence model				
non-HSELM	111	23		134
Row %	82.8	17.2		
HSELM	27	45		72
Row %	37.5	62.5		
Total	138	68		206
Row %	67.0	33.0		

region of (exclusively) HSELM. Hence, from a physical point of view it seems sensible not to downweight the outlying points in this case.

Table 3 ($b1$) shows the jackknifed resubstitution summary of the discrimination based on kernel estimates of the probability densities (with a uniform kernel with radius 1 on natural logarithmic scale). In comparison with Table 3 (a), the (2,1) component of the matrix (i.e., the probability to misclassify HSELM discharges as non-HSELM) changes from 47.2 to 59.7%. The (1,2) component of the matrix (i.e., the probability to misclassify non-HSELM discharges

as HSELM) reduces from 14.9%, which is the value for the parametric discrimination, to 3.0%. This indicates that even (multi-dimensional) elliptical contours do not demarcate very well the region where only HSELM discharges occur, and that some further improvement of prediction is possible, at the cost of a more complicated boundary, if one wants to avoid entirely the non-HSELM discharges. This more complicated boundary does not classify all HSELM shots correctly, possibly because the HSELM shots are rather scattered throughout 'non-HSELM region'. By adjusting the kernel radius, one can get some trade-off between the numbers of (1,2)-type and (2,1)-type misclassifications. This is illustrated in Table 3 ($b2$), where the same type of discrimination is performed, but now with kernel radius 0.5. One sees a reduction of the number of (2,1)-type and an increase of the number of (1,2)-type of misclassification. Note also that 9 non-HSELM and 6 HSELM observations are not allocated to either group, and hence fall into the category 'other'. This is because those observations are outside a radius 0.5 of all other observations in the dataset, whence, by this method, both estimated probability densities are zero. Such a situation does not occur if the radius of the kernel is increased to 1.0. Which kernel radius will be 'optimal' depends on the relative losses associated with the two types of misclassification.

In Table 3 (c), the results from applying the multinomial independence model, obtained by using the program INDEP [36], are shown. For simplicity, the discretisation was done on the original variables, not on the (joint) principal components. (As one can see from Table 2 (b), the four discriminating variables are not very highly correlated.) Ten groups per variable were used, as far as sensible roughly according to the deciles. Zero cells were avoided by using a 'flattening constant' [46], which has some Bayesian justification. One can see that the performance of the multinomial independence model is comparable with that of the quadratic boundary model.

A very simple way to compare the three approaches is to look at the Crude Error Rates (C.E.R.), i.e the total fraction of misclassifications. From this point of view, the estimated performance is similar (C.E.R. = 26%, 23%, and 24%, respectively). (In Table 3 ($b2$), the C.E.R. would be 24.5% if not classifying an observation is considered half as serious as making a misclassification.) The two types of misclassification occur in a different ratio for the kernel density estimate approach than for the other two approaches, however. Clearly, the C.E.R is a sensible criterion only if the two types of misclassification are considered to be about equally serious, which is not the case here. To cope with this situation, one has either to associate losses to the misclassifications, or to analyse both types of misclassifications jointly. Some theory and analysis using the last approach, albeit in another physical context, can be found in [47].

3.2.2 JET (SN)

The basic frequency table is

	non-HSELM	HSELM	Total
Number of observations	383	83	466
Ratio (%)	82	18	100

The centers of mass and the standard deviations of the distributions are shown in Table 4. The column T of this table gives, for each of the four variables, the difference in mean value divided by its estimated standard deviation, and the column F roughly indicates the significance of the difference between the standard deviations.

Table 4 — JET (SN) data

(a) Mean values and standard deviations

	non-HSELM, N=383		HSELM, N=83			
log: ↓	Mean	SD	Mean	SD	T	F
P_{inj}	1.97	0.26	2.07	0.34	–2.5	**
I_p	1.08	0.22	1.30	0.19	–8.4	–
B_t	0.90	0.15	1.03	0.11	–8.7	**
$\overline{n}_e$	1.36	0.24	1.38	0.30	–0.4	**

**: $0.001 < P < 0.01$, –: $0.05 < P < 0.2$

(b) Correlation coefficients

		non-HSELM, (STD_0= 0.05)		HSELM, (STD_0= 0.11)	
log: ↓	log: →	P_{inj}	I_p	B_t	$\overline{n}_e$
P_{inj}		1	0.55	0.60	0.70
I_p		0.18	1	0.30	0.45
B_t		0.45	0.51	1	0.52
$\overline{n}_e$		0.35	0.58	0.45	1

The sample correlation coefficients of the HSELM class are displayed in the right upper corner and those of the non-HSELM class in the left lower corner.
STD_0 is one standard deviation of the sample correlation coefficient under the hypothesis of no actual correlation.

In contrast to the case of ASDEX, X_P and X_I are larger for HSELM than for non-HSELM discharges. Also, the average value of X_B is larger for HSELM than for non-HSELM (for ASDEX there was no significant difference). The question whether the density distributions of the HSELM and non-HSELM shots can be considered to be the same is addressed in the same way as for ASDEX. The distance between the two groups, in the two corresponding metrics, is given by

$$\underline{\underline{\hat{D}}} = \begin{pmatrix} 0 & 3.33 \\ 1.75 & 0 \end{pmatrix}. \tag{24}$$

Now $D_{1,2}$ is larger than $D_{2,1}$, and significantly so, since the ratio is about 1.9 which is larger than $F_{379,79;.05} \simeq 1.4$. This indicates that, for JET, group 1 (non-HSELM) does not have the same covariance matrix as group 2 (HSELM). This is confirmed by the modified likelihood ratio test [26, 38], which gives a highly significant result: $72 \gg \chi^2_{10,0.05} = 18$, $\det(\underline{\underline{\Sigma}}_2)$ being smaller than $\det(\underline{\underline{\Sigma}}_1)$. Hence, the discrimination boundary cannot accurately be expressed by linear combinations of X_j.

Table 5 — Classification performance for various models JET (SN)

Prior probability: HSELM = 0.18

True class	Allocated class non-HSELM	HSELM	OTHER	TOTAL
a) Quadratic boundaries				
non-HSELM	370	13		383
Row %	96.6	3.4		
HSELM	44	39		83
Row %	53.0	47.0		
Total	414	52		466
Row %	82.2	11.2		
b1) Kernel density estimation (r=1.0)				
non-HSELM	362	19	0	383
Row %	94.5	5.0	0.0	
HSELM	24	59	0	83
Row %	28.9	71.1	0.0	
Total	386	78	0	466
Row %	82.2	17.8	0.0	
b2) Kernel density estimation (r=0.7)				
non-HSELM	358	17	8	383
Row %	93.5	4.4	2.1	
HSELM	14	69	0	83
Row %	16.9	83.1	0.0	
Total	327	86	8	466
Row %	79.8	18.5	1.7	
c) Multinomial independence model				
non-HSELM	325	58		383
Row %	84.9	15.1		
HSELM	35	48		83
Row %	42.2	57.8		
Total	360	106		466
Row %	77.3	22.7		

The pooled within-class standardised canonical coefficients are

$$\underline{\hat{C}} = \begin{pmatrix} 0.11 \\ 0.87 \\ 0.51 \\ -0.72 \end{pmatrix}, \tag{25}$$

where $\underline{\hat{C}} = (C_P, C_I, C_B, C_n)^t$. Using this simple yardstick, we see that the injected power is, in comparison with ASDEX (DN), more important and the current is less important for discrimination between HSELM and non-HSELM. Also, the canonical coefficients for the magnetic field and the density are somewhat larger in absolute value than in the case of ASDEX (DN). The t-values of these coefficients, calculated from linear regression analysis, are $(1.0, 7.1, 4.1, -5.6)$. Hence, except for C_P, all coefficients are clearly significant. Table 5 shows the jackknifed (mis-) classification summary for (a) the quadratic discriminant function and (b) non-parametric discriminant functions based on density estimates with uniform kernels and radii $r = 1$ and $r = 0.7$, respectively. The estimated probability of (1,2) misclassification is 3.4% in case (a), 5% in case $(b1)$, and 5.5% in case $(b2)$. (The latter applies if making no classification is considered half as serious as making a wrong classsification.) This means that the quadratic fit is better suited to exclude the non-HSELM shots than in the ASDEX (DN) case. Therefore, we will describe this boundary more explicitly. The squared Mahalanobis distance, or potential, of an observation at point $\underline{x}$ to the center of gravity, $\underline{\mu}_j$, of group j can be written as

$$D_j^2(\underline{x}) = (\underline{x} - \underline{\mu}_j)^t \underline{\underline{A}}_j (\underline{x} - \underline{\mu}_j) + \underline{b}_j^t (\underline{x} - \underline{\mu}_j) + c_j, \tag{26}$$

where the index $j = 1, 2$ indicates the class, and $\underline{\underline{A}}_j$ and $\underline{b}_j$ are a matrix and a vector of coefficients, respectively. The posterior probability density for an observation at point $\underline{x}$ to belong to group j is given by

$$f_j(\underline{x}) = e^{-\frac{1}{2}D_j^2(\underline{x})} / \Sigma_{k=1}^2 e^{-\frac{1}{2}D_k^2(\underline{x})}. \tag{27}$$

An observation is allocated to group j if the posterior probability density to belong to group j is larger than 50% $(j = 1, 2)$.

The boundary **B** separating the two allocation regions is given by

$$B = \{\underline{x} \,|\, f_1(\underline{x}) = f_2(\underline{x})\}. \tag{28}$$

The coefficients $\{\underline{\underline{A}}_j, \underline{b}_j^t, c_j\}$, multiplied by $-\frac{1}{2}$, where $\underline{\underline{A}}_j = (A_{jik})_{i,k \in S_I}$ and $\underline{b}_j^t = (b_{jk})_{k \in S_I}$, for $j = 1, 2$, are tabulated in Table 6. In both cases, among the quadratic coefficients $A_{i,k}$, the diagonal ones, $A_{k,k}$, are dominant $(i, k \in S_I)$.

Table 6 — Quadratic discrimination

JET (SN)

Coefficients of the quadratic form representing –0.5 times the squared Mahalanobis distance

To: non-HSELM

log: ↓	log: →	P_{inj}	I_p	B_t	$\overline{n}_e$
P_{inj}		–9.7	–2.3	7.6	2.8
I_p		–2.3	–18.4	10.2	7.9
B_t		7.6	10.2	–38.8	2.5
$\overline{n}_e$		2.8	7.9	2.5	–14.9
Linear term		21.8	9.2	11.0	7.9
Constant		–30.2			

To: HSELM

log: ↓	log: →	P_{inj}	I_p	B_t	$\overline{n}_e$
P_{inj}		–11.5	5.7	10.2	5.4
I_p		5.7	–20.1	–2.2	1.7
B_t		10.2	–2.2	–62.4	5.0
$\overline{n}_e$		5.5	1.7	5.0	–11.3
Linear term		–3.1	28.8	78.0	–5.9
Constant		–46.4			

3.2.3 JFT-2M (SN)

The original ITERH.DB1 database contains only a few JFT-2M observations with small ELMs. After having made some analyses with them, we turned back to the original time traces plotted on the review sheets [8]. Looking at the H_α signal, it was noticed that in a number of cases, small ELMs did occur just after the last time-point that had been selected for inclusion in the database. Since for the present investigation, in contrast to the confinement analysis in [8], it is important to know whether or not small ELMs were produced during a substantial period of time during the shot, which may be outside the two time-points selected for the confinement analysis, we reclassified some of the shots as being 'HSELM'. A description of the criteria used and a list of shots and time intervals that were classsified as 'HSELM' are given in the Appendix.

The reclassification was based solely on the H_α signal. It should be noted that sometimes the H_α spikes are provoked by sawteeth. Hence, the label HSELM is, more precisely speaking, used to denote 'shots with a spiky H_α signal indicating either (small) ELMs or sawteeth'. A further study in which the H_α signals are compared with those from interferometry in order to separate the ELMy from the sawtooth cases, would be very valuable.

For the ASDEX and JET data, the reclassification problem is less pressing. In those cases usually three, sometimes four, time-points per shot have been taken, covering the in-

terval between the beginning of the H-mode and the time that W_{dia} is maximal somewhat better than in the case of JFT-2M. Nevertheless, in a future investigation, one could consider to investigate as response variable the total duration of the periods (standardised in some way) with and without small ELMs during the discharge. At present, no such information is available in the database. The basic frequency table for JFT-2M (SN) is:

	non-HSELM	HSELM	Total
Number of observations	312	61	373
Ratio (%)	84	16	100

The centers of mass and the standard deviations of the distributions are shown in Table 7. Again, the column F indicates which SD's, and the column T indicates which mean values are to be considered different between HSELM and non-HSELM.

Table 7 — JFT-2M (SN) data

(a) Mean values and standard deviations

	non-HSELM, N=312		HSELM, N=61			
log: ↓	Mean	SD	Mean	SD	T	F
P_{inj}	–0.18	0.57	–0.13	0.51	–0.7	•
I_p	–1.52	0.19	–1.39	0.11	–7.0	***
B_t	0.22	0.07	0.24	0.01	–4.5	***
$\overline{n}_e$	1.48	0.26	1.46	0.22	0.7	-
Gas	0.32	0.17	0.16	0.20	6.6	-

***: P < 0.001, -: 0.05 < P < 0.2, •: 0.2 < P < 0.5

(b) Correlation coefficients

		non-HSELM, (STD_0= 0.06)			HSELM, (STD_0= 0.13)	
log: ↓	log: →	P_{inj}	I_p	B_t	$\overline{n}_e$	Gas
P_{inj}		1	0.16	0.15	0.51	–0.77
I_p		–0.13	1	0.02	0.40	–0.22
B_t		–0.09	0.18	1	0.11	–0.25
$\overline{n}_e$		–0.02	0.62	0.17	1	–0.26
Gas		–0.41	0.15	–0.11	0.21	1

The sample correlation coefficients of the HSELM class are displayed in the right upper corner and those of the non-HSELM class in the left lower corner.
STD_0 is one standard deviation of the sample correlation coefficient under the hypothesis of no actual correlation.

From the table one can see that the HSELM discharges generally occur at higher values of I_p, B_t, and at a lower value of GAS than the non-HSELM discharges do. The last fact can also be seen from Table 8: For the H into H (GAS=1) discharges, a larger fraction HSELM observations occurs than for the D into H discharges (GAS=1.5). Of course, such considerations are only univariate, taking the 'influence' of only one variable at a time into account. Due to correlations, such 'influences' may be confounded with those of other variables.

Table 8 — JFT-2M (SN) data
Numbers of timeslices per ELM type and gas composition

Gas	H→H	D→H	Total
non-HSELM	67	245	312
HSELM	37	24	61
Total	104	269	373

The simultaneous influence of P_{inj}, I_p, B_t, $\bar{n}_e$, and GAS is investigated by performing discriminant analysis for these variables on logarithmic scale. We summarise the main results.

The estimated pairwise distances between two groups are given by the matrix

$$\underline{\underline{\hat{D}}} = \begin{pmatrix} 0 & 4.77 \\ 1.82 & 0 \end{pmatrix}. \tag{29}$$

The fact that $\underline{\underline{D}}_{1,2}$ is significantly larger than $\underline{\underline{D}}_{2,1}$ indicates that, like at JET and unlike at ASDEX, see Eqs. (21) and (24), the HSELM shots occur in a smaller region of plasma parameter space than the non-HSELM shots. The imbalance between $\underline{\underline{D}}_{1,2}$ and $\underline{\underline{D}}_{2,1}$ implies that the discrimination boundary is expected not to be efficiently representable by a linear combination of the above five, logarithmically transformed, variables.

The (pooled within-class) standardised canonical coefficients are estimated by

$$\underline{\hat{C}} = \begin{pmatrix} 0.19 \\ -0.87 \\ -0.06 \\ 0.47 \\ 0.83 \end{pmatrix}. \tag{30}$$

The t-values of the estimates $\underline{\hat{C}}^t = (C_P, C_I, C_B, C_n, C_G)$ are $(1.4, -6.0, -0.5, 3.2, 6.1)$. As in the univariate analysis, we have a significant dependence on plasma current and on gas composition. However, in contrast to the univariate analysis, the effect of the plasma density turns out to be significant. This (significant) density dependence is to be interpreted when the four other plasma parameters are kept constant. In the univariate analysis, this fact was masked by the correlation between density and current, which is about 0.6 for non-HSELM, and 0.4 for HSELM, see Table 7 (*b*). Compared with the univariate analysis, the injected power remains insignificant, and the magnetic field has become insignificant in the presence of the other variables. To understand the latter, we noted that two magnetic field scans exist in the present dataset, but no HSELM shots occurred in these scans. This is in accordance with B_t being significant from univariate considerations. However, the two magnetic field scans were made only with D into H, i.e. GAS=1.5, and are somewhat negatively correlated

($r = -0.3$) with I_p. The coefficients of the simultaneous analysis express the fact that the non-occurrence of HSELM can be ascribed to the high value of GAS and the relatively low value of I_p for the 16 observations from the 2 scans. In addition to these two effects, which are estimated from all data, the magnetic field does not exhibit additional discriminatory value. Obviously this holds within the simple power-law type model used so far.

Table 9 — Classification performance for various models
JFT-2M (SN)

Priors HSELM = 0.16

True class	Allocated class: non-HSELM	HSELM	OTHER	TOTAL
a) Quadratic discriminant analysis				
non-HSELM	239	73		312
Row %	76.6	23.4		
HSELM	11	50		61
Row %	18.0	82.0		
Total	250	123		373
Row %	67.0	33.0		
b1) Kernel density estimation (r=1.0)				
non-HSELM	263	42	7	312
Row %	84.3	13.5	2.2	
HSELM	17	42	2	61
Row %	27.9	68.9	3.3	
Total	280	84	9	373
Row %	75.1	22.5	2.4	
b2) Kernel density estimation (r=0.5)				
non-HSELM	238	12	62	312
Row %	76.3	3.9	19.9	
HSELM	22	21	18	61
Row %	36.1	34.4	29.5	
Total	260	33	80	373
Row %	69.7	8.9	21.5	
c) Multinomial independence model				
non-HSELM	296	16		312
Row %	94.9	5.1		
HSELM	45	16		61
Row %	73.8	26.2		
Total	341	32		373
Row %	91.4	8.6		

Table 9 shows the jackknifed (mis-) classification summary of the discriminant analysis using: (a) a quadratic discriminant function, (b) non-parametric density estimates with

uniform kernels and radii $r = 1$ and $r = 0.5$, respectively, and (c) the multinomial independence model with a global association factor. It is noted that the discrimination with the uniform kernel density estimates improves the number of (dangerous) misclassifications of non-HSELM shots. Nevertheless, a fair amount of misclassifications remains for $r = 1$, whereas for $r = 0.5$ there is a large number of unclassified observations. The multinomial independence model reduces the number of non-HSELM misclassifications, albeit at the cost of a larger number of the (less dangerous) HSELM misclassifications.

3.3 Discriminant Analysis taking Plasma Memory and Plasma-Wall Distance into Account

We now want to investigate the effect of 'plasma memory' on the class of plasma discharges that will occur. First, we only replace the instantaneous plasma density by the target plasma density, i.e., we use as fourth discrimination variable the logarithm of the line-average density in the Ohmic phase,

$$X_{\bar{n}_{e,ohm}} = \ln(\bar{n}_{e,ohm}), \tag{31}$$

and keep the other three components the same as in Section 3.2. The effect of this replacement will be investigated for ASDEX (DN) plasmas only.

The estimated canonical coefficients are

$$\underline{\hat{C}} = \begin{pmatrix} 0.53 \\ 0.99 \\ -0.12 \\ 0.64 \end{pmatrix}. \tag{32}$$

Comparing these values with the canonical coefficients in Section 3.2, we see the importance of P_{inj} and I_p and the unimportance of B_t for linear discrimination confirmed. However, the density dependence stands out more clearly now. This means that the appearance of HSELM discharges depends on the target plasma density rather than on the instantaneous density.

The difference between the target density and the instantaneous density in their effect on the discrimination suggests the importance of the elapsed time after the transition to the H-mode. Furthermore, the distance between the separatrix and the wall has been known by the experimentalists to influence the occurrence of ELMs [16, 8].

To see these effects quantitatively, we choose as new set of variables

$$X_t = \ln(t - t_{LH}), \quad X_{PV} = \ln(P_{inj}/V_p), \quad X_{\bar{n}_{e,ohm}} = \ln(\bar{n}_{e,ohm}) \tag{33}$$

$$X_d = d_{SW} \text{ or } d_{SL}, \quad X_q = \ln(q_{cyl}), \quad X_{BIV} = \ln(B_t I_p/V_p), \tag{34}$$

where $t - t_{LH}$ is the elapsed time after the L-H transition (expressed in units of 50 ms for ASDEX and JFT-2M, and of 500 ms for JET), d_{SL} is the distance between the separatrix and 'limiter' and d_{SW} the distance between the separatrix and the outer wall, both normalised by the minor radius and defined more precisely below, $q_{cyl} = 10\frac{B_t(T)}{I_p(MA)}\frac{ab}{R}$ is the cylindrical q-value, and $V_p = 2\pi^2 Rab$ is the plasma volume (in m^3). The impact of the parameter X_d has been experimentally extensively studied, see e.g. [20]. As some values of X_d were close to zero, it was decided not to use the logarithm of this variable. Note that for constant V_p, the transformation from (X_B, X_I) to (X_q, X_{BIV}) can be described by a simple rotation of the coordinate frame. With these variables, we perform discriminant analysis for ASDEX (DN) and JET (SN).

Table 10 — ASDEX (DN) data

(a) Mean values and standard deviations

	non-HSELM, N=134		HSELM, N=72			
log: ↓	Mean	S.D.	Mean	S.D.	T	F
$t - t_{LH}$	0.10	0.54	–0.15	0.53	3.1	•
P_{inj} / V_p	–0.61	0.16	–0.76	0.17	6.5	•
$\overline{n}_{e,\,ohm}$	1.25	0.17	1.23	0.24	–0.6	**
sepwall / a	0.44	0.04	0.46	0.05	–3.0	•
q_{cyl}	1.15	0.13	1.28	0.17	–5.6	*
$B_t I_p / V_p$	–2.02	0.16	–2.12	0.21	3.7	**

***: P < 0.001, **: 0.001 < P < 0.01, *: 0.01 < P < 0.05, –: 0.05 < P < 0.3, •: 0.3 < P

(b) Correlation coefficients

		non-HSELM, (STD_0= 0.09)				HSELM, (STD_0= 0.12)	
log: ↓	log: →	$t - t_{LH}$	P_{inj} / V_p	$\overline{n}_{e,\,ohm}$	sepwall / a	q_{cyl}	$B_t I_p / V_p$
$t - t_{LH}$		1	0.00	0.20	0.19	–0.02	0.10
P_{inj} / V_p		–0.44	1	0.17	–0.15	–0.39	0.45
$\overline{n}_{e,\,ohm}$		0.25	0.14	1	0.31	–0.49	0.69
sepwall / a		0.03	0.16	0.55	1	–0.13	0.33
q_{cyl}		0.00	–0.27	–0.69	–0.58	1	–0.73
$B_t I_p / V_p$		0.26	0.08	0.69	0.60	–0.59	1

The sample correlation coefficients of the HSELM class are displayed in the right upper corner and those of the non-HSELM class in the left lower corner. STD_0 is one standard deviation of the sample correlation coefficient under the hypothesis of no actual correlation.

3.3.1 ASDEX (DN)

Table 10 describes and compares the distributions of the non-HSELM and HSELM observations in plasma parameter space. We can see, for instance, that the difference in X_t is 0.25, i.e., on average, the non-HSELM mode appears $e^{0.25} \times 50 \simeq 65$ ms later during the shot than the HSELM mode. The difference in d_{SW} means that the distance between the separatrix and the outer wall for the HSELM observations is on average $0.02 \times 40 = 0.8$ cm larger

than for the non-HSELM observations. Discriminant analysis applied to the $N = 206$ DN observations from ASDEX gives the following estimated (pooled within-class standardised) canonical coefficients:

$$\underline{\hat{C}}^t = (\,0.53, 0.54, -0.52, -0.49, -0.66, 0.29\,)\,. \tag{35}$$

The t-values of $\underline{\hat{C}}^t = (\hat{C}_t, \hat{C}_{PV}, \hat{C}_{\bar{n}_{e,ohm}}, \hat{C}_d, \hat{C}_q, \hat{C}_{BIV})$, obtained from standard linear regression, are $(-4.5, -4.5, 3.2, 3.9, 4.2, -1.6)$. These results indicate that the variables $t-t_{LH}$, $\bar{n}_{e,ohm}$, and d_{SW}) are about as important for discrimination as the heating power.

The interpretation of the last two coefficients in Eq. (35) is: for fixed values of the four other plasma parameters, the direction of highest discrimination is given by the logarithm of $q^{-0.66/0.14}(B_t I_p/V_p)^{0.29/0.18}$, where 0.14 and 0.18 are the pooled within-class standard deviations of X_q and X_{BIV}, estimated from Table 10 (a). By comparing Eq. (35) with Eqs. (22) and (32) one can see that the q-value is more important for discrimination than the absolute value of the toroidal magnetic field.

Table 11 shows the result of the jackknifed (mis-) classification summary. The performance is better than the performance of the discriminant analysis based on the four instantaneous variables. The (2,1) component of the (mis-) classification table is 9.0% for the quadratic discriminant analysis, and 7.8% for the analysis using uniform kernel density estimation with $r = 1.5$. The assumption of the quadratic fitting is better than in Section 3.2.1.

It is noted that for X_d the variable SEPLIM as available in ITERH.DB1 was used, which denotes for ASDEX an estimate of the distance between the plasma boundary and the outside torus wall (in the horizontal plane), not taking the position of the ICRH antenna into account.

We also did the discriminant analysis while taking the position of the ICRH antenna into account, as well as with an estimate of the closest distance between the plasma and the wall (not necessarily in the horizontal plane). These data are presently not available in the ITER H-mode Database. In both cases, the performance of the discrimination turned out to be less than in the above case. This suggests that for instance the magnetic field ripple [48] may be of more importance for the occurrence of ELM's than the closest distance between the plasma and the wall. Such a suggestion would not be in contradiction with [49], where ballooning-type instabilities with toroidal modes numbers $n = 8$ to 15 were considered to be likely candidates for ELM precursors. This topic can be investigated more precisely as soon as more accurate estimates of the minor plasma radius than presently available in ITERH.DB1 will be at our disposition.

The coefficients for the Mahalanobis distances to the two centers of gravity are given in Table 12. The discrimination surfaces are surfaces of constant difference between the two Mahalanobis distances. Note that they can be multidimensional ellipses and that they can also

Table 11 — Classification performance for various models ASDEX (DN)

Priors HSELM = 0.35

True class	Allocated class non-HSELM	HSELM	OTHER	TOTAL
a) Quadratic discriminant analysis				
non-HSELM	122	12		134
Row %	91.0	9.0		
HSELM	23	49		72
Row %	31.9	68.1		
Total	145	61		206
Row %	70.4	29.6		
b1) Kernel density estimation (r=1, threshold=0.5)				
non-HSELM	108	10	16	134
Row %	80.6	7.5	11.9	
HSELM	16	36	20	72
Row %	22.2	50.0	27.8	
Total	124	46	36	206
Row %	60.2	22.3	17.5	
b2) Kernel density estimation (r=1.5, threshold=0.5)				
non-HSELM	122	9	3	134
Row %	91.0	6.7	2.2	
HSELM	27	43	2	72
Row %	37.5	59.7	2.8	
Total	149	52	5	206
Row %	72.3	25.2	2.4	

have an hyperbolic character. The reader is referred to [50] for a graphical two-dimensional section of the discriminant surface for these data.

3.3.2 JET (SN)

For JET, we used for d_{SL} the variable SEPLIM as available in ITERH.DB1, which is an estimate of the minimum distance between the separatrix and the 'limiter', i.e. any part of the wall. In Table 13, the univariate summary statistics and the correlation matrices are given. One can see that the ohmic density and the variable $B_t I_p / V_p$ are the most important ones for discrimination. Also, the HSELM shots seem to occur at somewhat higher values of P_{inj}/V_p and at lower values of q_{cyl} than the non-HSELM discharges do. The time since the onset of the H-mode does not exhibit a significant effect.

However, these univariate considerations do not necessarily give the correct estimates for the simultaneous influence of these variables. In fact, the estimated standardised canon-

Table 12 — Quadratic discrimination

ASDEX (DN)

Coefficients of the quadratic form representing –0.5 times the squared Mahalanobis distance

To: non-HSELM

Quadratic term	log: →	$t-t_{LH}$	P_{inj} / V_p	$\overline{n}_{e,\,ohm}$	sepwall / a	q_{cyl}	$B_t I_p$ / V_p
log: ↓							
$t-t_{LH}$		–2.61	–3.80	2.65	–4.01	1.86	2.16
P_{inj} / V_p		–3.80	–26.84	2.91	–1.33	–6.21	0.40
$\overline{n}_{e,\,ohm}$		2.65	2.91	–44.18	14.28	–24.50	16.60
sepwall / a		–4.01	–1.33	14.28	–477.34	–42.96	50.68
q_{cyl}		1.86	–6.21	–24.50	–42.96	–65.69	–8.05
$B_t I_p$ / V_p		2.16	0.40	16.60	50.68	–8.05	–47.38
Linear term		–2.75	–22.05	224.10	689.55	209.63	–259.17
Constant		–669.29					

To: HSELM

Quadratic term	log: →	$t-t_{LH}$	P_{inj} / V_p	$\overline{n}_{e,\,ohm}$	sepwall / a	q_{cyl}	$B_t I_p$ / V_p
log: ↓							
$t-t_{LH}$		–1.91	0.33	1.06	3.51	0.46	–0.49
P_{inj} / V_p		0.33	–26.86	–4.40	–30.16	–1.05	15.08
$\overline{n}_{e,\,ohm}$		1.06	–4.40	–18.96	1.23	–0.00	16.36
sepwall / a		3.51	–30.16	1.23	–290.90	15.80	41.59
q_{cyl}		0.46	–1.05	–0.00	15.80	–39.62	–24.61
$B_t I_p$ / V_p		–0.49	15.08	16.36	41.59	–24.61	–48.28
Linear term		–9.16	64.56	108.64	357.55	–19.43	–198.11
Constant		–313.37					

ical coefficients from discriminant analysis, which are, as we have seen before, the multiple regression coefficients after standardising all variables by the pooled within-class variances, are ($N = 277$)

$$\hat{\underline{C}}^{t} = (-0.004, 0.038, 0.217, 0.003, -0.037, 0.785)\,. \tag{36}$$

The corresponding t-values are $(-0.04, 0.26, 0.89, 0.02, -0.25, 3.2)$. This means that, except for $B_t I_p/V_p$, none of the coefficients are statistically significant, nor large in absolute value! This illustrates the effect of confounding, due to correlations between the discrimination variables. Indeed, HSELMs are associated with a larger ohmic density than non-HSELMs, but a higher ohmic density is also correlated with a higher value $B_t I_p/V_p$ ($r = 0.85$). At a constant value of $B_t I_p/V_p$, the ohmic density has no statistically significant predictive value for the occurrence of small ELMs. The correlations between $B_t I_p/V_p$ and P_{inj}/V_p and q_{cyl} are smaller ($r = 0.4$ and $r = -0.4$, respectively), but apparently sufficient, as one can see by comparing Table 13 (a) with Eq. (36), to provoke a smaller amount of confounding.

The question may arise whether the correlations between the variables are sufficiently high to make the dataset ill-conditioned for simultaneous regression. This is investigated by

Table 13 — JET (SN) data

(a) Mean values and standard deviations

log: ↓	non-HSELM, N=218 Mean	non-HSELM, N=218 S.D.	HSELM, N=59 Mean	HSELM, N=59 S.D.	T	F
$t - t_{LH}$	0.54	0.58	0.65	0.62	–1.2	•
P_{inj} / V_p	–2.68	0.26	–2.55	0.30	–3.2	–
$\overline{n}_{e,\,ohm}$	0.39	0.32	0.72	0.21	–9.5	***
seplim / a	0.055	0.017	0.059	0.012	–1.9	**
q_{cyl}	1.10	0.17	1.01	0.19	3.6	–
$B_t I_p / V_p$	–2.72	0.34	–2.33	0.23	–10.4	***

***: P < 0.001, **: 0.001 < P < 0.01, *: 0.01 < P < 0.05, –: 0.05 < P < 0.3, •: 0.3 < P

(b) Correlation coefficients

log: ↓	log: →	non-HSELM, (STD_0= 0.05) $t - t_{LH}$	P_{inj} / V_p	$\overline{n}_{e,\,ohm}$	seplim / a	HSELM, (STD_0= 0.11) q_{cyl}	$B_t I_p / V_p$
$t - t_{LH}$		1	0.29	0.22	–0.04	0.05	–0.01
P_{inj} / V_p		–0.07	1	0.59	0.24	–0.29	0.62
$\overline{n}_{e,\,ohm}$		0.18	0.16	1	0.28	–0.24	0.68
seplim / a		0.16	–0.03	0.21	1	–0.13	0.52
q_{cyl}		–0.25	0.08	–0.48	0.01	1	–0.58
$B_t I_p / V_p$		0.16	0.33	0.83	0.13	–0.34	1

The sample correlation coefficients of the HSELM class are displayed in the right upper corner and those of the non-HSELM class in the left lower corner. STD_0 is one standard deviation of the sample correlation coefficient under the hypothesis of no actual correlation.

principal component analysis. The square root of the smallest eigenvalue of the correlation matrix equals 0.36, and the associated eigenvector is mainly associated with $B_t I_p/V_p$ and $\bar{n}_{e,ohm}$. As the measurement accuracy of these 2 quantities is better than, say, 0.2 times 0.36 (i.e., than about 7%), the estimated bias in the estimates of the canonical coefficients induced by neglecting such (random) measurement errors is not more than a few percent of those estimates. The square root of second smallest eigenvalue is considerably larger (0.74). From these considerations, the condition is sufficiently good for a well-behaved simultaneous linear discriminant analysis.

The canonical coefficients suggest that the minimum distance between the separatrix and the wall is not at all important for (linear) discrimination. This should, however, not be misinterpreted. About 150 observations with missing values for d_{SL} have not been used in the analysis. They are, in overwhelming majority, shots for which at least one of the X-points is outside or nearly outside the vessel, so that the missing values correspond to $d_{SL} \leq 0$. Using an indicator variable for those missing values (0=missing, 1=non-missing), instead of the variable d_{SL}, gives a significant discrimination coefficient ($t = 2.6$), while the other coefficients remain roughly the same! Also, coding the missing values as $d_{SL} = 0$, and retaining the positive values of d_{SL} as they are in the database, gives a significant coefficient ($t = 2.45$). So the negative values of d_{SL} seem to be far more interesting for discrimination

than the positive ones. HSELMs are associated (in the simultaneous regression) with positive d_{SL}, i.e. with X points within the vessel, but the precise value of d_{SL} is rather unimportant. A fortunate coincidence is that restricting attention to the observations with both X-points inside the vessel leads a region where the desirable HSELM discharges are expected, and at the same time simplifies the discriminant analysis, by fading out the influence of all plasma variables, except for $B_t I_p/V_p$.

Table 14 — Quadratic discrimination
JET (SN)
Coefficients of the quadratic form representing –0.5 times the squared Mahalanobis distance

To: non-HSELM

Quadratic term	log: →	$\bar{n}_{e,\,ohm}$	$B_t\, I_p\, /\, V_p$
log: ↓			
$\bar{n}_{e,\,ohm}$		–16.0	12.6
$B_t\, I_p\, /\, V_p$		12.6	–14.3
Linear term		81.0	–87.6
Constant		–132.3	

To: HSELM

Quadratic term	log: →	$\bar{n}_{e,\,ohm}$	$B_t\, I_p\, /\, V_p$
log: ↓			
$\bar{n}_{e,\,ohm}$		–21.0	12.8
$B_t\, I_p\, /\, V_p$		12.8	–17.1
Linear term		89.8	–98.2
Constant		–144.7	

Hence, as a compact representation of the case that both X-points are inside the vessel, we present in Table 14 the quadratic discriminant function using the variables $B_t I_p/V_p$ and $\bar{n}_{e,ohm}$ only. The table is based on the $N = 277$ observations with $d_{SL} > 0$. In two dimensions one can easily make plots of the discriminant curves such as in [50]. In our situation, one can see that the difference between the two Mahalanobis distances is positive definite. Hence, the discriminant curves are ellipses.

Finally, we check that, using only $\bar{n}_{e,ohm}$ and $B_t I_p/V_p$, the canonical coefficients are $\underline{C}^t = (0.22, 0.81)$, with t-values $(0.97, 3.6)$. This is not very different from the corresponding coefficients in the linear analysis with the 6 variables discussed above.

Table 15 (a) shows the result of the jackknifed (mis-) classification summary of the quadratic analysis with the six and with the two variables described above. One can see

Table 15 — Classification performance for various models JET (SN)

Priors HSELM = 0.21

True class	Allocated class: non-HSELM	HSELM	OTHER	TOTAL
a) Quadratic discriminant analysis				
a1) with the 6 variables from table 13				
non-HSELM	199	19		218
Row %	91.3	8.7		
HSELM	24	35		59
Row %	40.7	59.3		
Total	223	54		277
Row %	80.5	19.5		
a2) with $\bar{n}_{e,\,ohm}$ and $B_t\, I_p\, /\, V_p$ only				
non-HSELM	200	18		218
Row %	91.7	8.3		
HSELM	26	33		59
Row %	44.1	55.9		
Total	226	51		277
Row %	81.6	18.4		
b1) Kernel density estimation (r=1, threshold=0.5)				
non-HSELM	191	4	23	218
Row %	87.6	1.8	10.6	
HSELM	10	36	13	59
Row %	17.0	61.0	22.0	
Total	201	40	36	277
Row %	78.7	14.4	13.0	
b2) Kernel density estimation (r=1.5, threshold=0.5)				
non-HSELM	203	10	5	218
Row %	93.1	4.6	2.3	
HSELM	12	43	4	59
Row %	20.3	72.9	6.8	
Total	215	53	9	277
Row %	77.6	19.1	3.3	

that the quadratic discrimination with the six variables performs relatively well compared to the non-parametric discrimination. With the two variables $\bar{n}_{e,ohm}$ and $B_t I_p / V_p$ only, the performance is somewhat worse, but still reasonable for practical purposes. The overall misclassification rate is 15.9% compared to 15.5% for the six variables, and 12.2% for the analysis based on the four instantaneous variables (see Table 5). The latter analysis included the points for which $d_{SL} < 0$. We now classify the HSELM shots better than in the analysis with the

four instantaneous variables, but at the cost of including more non-HSELM shots in the region attributed to HSELM. We suspect that these costs could be reduced to a considerable extent if the negative values of d_{SL} would become available in the database.

4 SUMMARY AND DISCUSSION

In this paper, we looked at various methods to determine the plasma parameter regime where H-mode discharges with small ELMs can be expected.

After a description of its theoretical aspects, discriminant analysis was applied to the ASDEX, JET, and JFT-2M data of the ITERH.DB1 database. We divided the H-mode discharges into two classes: class-1 (ELM-free or with giant ELMs) and class-2 (with small ELMs). This distinction was motivated by the fact that H-mode with small ELMs is favourable from the viewpoint of a long sustenance of improved confinement against the contamination by impurities. The distributions of the two classes of discharges in plasma parameter space overlap each other. The general statistical methodology to describe these distributions, and to discriminate between the two classes of H-mode has been discussed in Section 2.

In Section 3, linear and quadratic discriminant analysis on logarithmic scale was used to find explicit expressions for combinations of variables that are efficient to predict the region where small ELMs will occur. A general aspect based on the analysis of ASDEX (DN), JET (SN) and JFT-2M (SN) data is that linear discrimination (on logarithmic scale) is not very accurate since the covariance matrices for the two classes are significantly different. Hence, the boundary $\mathbf{B}$ cannot effectively be expressed by a simple power law in terms of the engineering variables, and hence also not in terms of the dimensionless plasma parameters.

Instead, the discriminant surfaces, i.e. the surfaces on which the difference in Mahalonobis distance to the centers of gravity of the two groups of H-mode discharges is constant, are on logarithmic scale explicitly described by quadratic equations.

The performance of the quadratic discriminant analysis was estimated by using the jackknife method on the available datasets of ASDEX, JET and JFT-2M, and expressed in (mis-) classification tables. A comparison was made with the performance of discrimination based on non-parametric density estimates and of discrimination using a multinomial independence model. These more flexible non-parametric methods showed a better performance than the quadratic discriminant analysis, however considerably less so when the 'plasma memory' and the plasma-wall distance were taken into account. The non-parametric methods do not permit a simple representation of the discrimination surfaces.

It was found that for quadratic discriminant functions a larger fraction of non-SELM discharges is (correctly) classified as non-SELM, than SELM discharges are classified as SELM. The last feature is seen by inspecting the (mis-) classification tables and can be explained by

assuming, as is partly seen from the scatter plots, that there is a mixed region where both classes of discharges (HSELM and non-HSELM) occur, and a region where predominantly HSELM discarges occur. This feature of the data is of importance for operating future machines since non-HSELM discharges have undesirable properties for long burning plasmas. (The ELM-free H-mode is vulnerable to impurity accumulation, and giant ELMs may produce damage from strong repetitive heat loads on the divertor plates.) By using discriminant analysis we have presented a method to avoid these unfavourable modes to a considerable extent.

In the analysis of the ASDEX (DN) H into D discharges, the injected power P_{inj} and the plasma current (or q-value) were shown to be important for discriminating between HSELM and non-HSELM. For JFT-2M (SN) discharges, plasma current and gas composition were the most important discriminating variables (with some additional effect of density), and for JET (SN) D into D discharges, it was plasma current, magnetic field and density. Quantitative estimates of the linear discriminant functions have been given in the main text. They can be used as simple approximations to the discriminant surfaces, though it was shown that quadratic surfaces give a more accurate description. Explicit quadratic coefficients have been presented for ASDEX (DN) H into D discharges in Table 12. A comparison shows that the discriminant surfaces are not the same for the three devices. For instance, low P_{inj} is favourable to get small ELMs in ASDEX, but in JET and JFT-2M, the variable P_{inj} does not play an important role. This sets a limit to the ability to predict, from the present analysis, the presence of small ELM's in future devices. In ensuing analyses, the change of the discriminant surfaces with machine size must be investigated more closely.

An important role of the 'plasma memory' during the discharge was found. The target plasma density is more important for the prediction of the class of H-mode than the instantaneous density. The elapsed time is in ASDEX as important for discrimination as is the injected power. In JET, the elapsed time does not seem important. These contrasting results may be due to the fact that the timeslices at JET have been chosen differently, usually more at the end of the H-mode phase, than the timeslices at ASDEX. They also suggests an improvement of the database and the ensuing analysis. Ideally, all the time-points of transition between H-mode with SELM and the other types of H-mode should be recorded. From that one can estimate, as a function of the plasma parameters and the elapsed time since the L-H transition, the 'hazard rate' of transition of HSELM to another type of H-mode, or the fraction of time the discharge dwells in H-mode with small ELM's.

Two results from the analysis for JET are the following. (1) The distance between the separatrix and the limiter, d_{SL}, is not important for predicting the type of H-mode, unless it is negative, which corresponds to the X-point being (nearly) outside the vessel. The negative values of d_{SL} are missing in the present database. (2) For discharges with the X-point inside the vessel, the discrimination is much simplified by the fact that only the target density and $B_t I_p / V_p$ are by far the most important variables.

This result from JET may be in agreement with the fact that for ASDEX we found a somewhat better discrimination using d_{SW} instead of d_{SL}, and it will be interesting to investigate empirically, with improved data from both machines, made possible by the work of McCarthy [51] and O'Brian et al. [52], whether the magnetic field ripple, rather than the closest distance between the plasma the wall, is the factor to influence the ocurrence of small ELM's.

It has not been the purpose of the present article to present an exhaustive study, but to provide the background and to illustrate the practical use of discriminant analysis to identify the regions in plasma parameter space where various types of H-mode can be produced. The approach can also be useful for identiying e.g. the plasma parameter regions where H-mode and L-mode occur, which is one of the objectives of the Threshold Database [15].

The necessity to establish a guideline to get rid of the impurity accumulation in the H-mode is now widely recognised [53]. The control of the type of H-mode remains therefore an urgent and interesting task.

5 APPENDIX

From the plotted graphs of the H_α signal, the following shots and time intervals were classified as 'leading to small ELMs and/or sawteeth'. The criterion for this classification was that approximately within one confinement time after the beginning of the time interval stated, small 'ELMy' spikes did occur on the H_α signal for some period of time during which W was 'non-decreasing', i.e. $dW/dt > -0.05W/\tau_E$, or during which a 'semi-stationary' W, i.e. $|dW/dt| < 0.2W/\tau_E$ and $W > 0.8W_{max}$, was obtained. In the following list, dt denotes a constant which was set to 0.005 s.

(shot=37740) and (0.75-dt < time < 0.77+dt)
(shot=44973) and (0.77-dt < time < 0.83+dt)
(shot=44958) and (0.64-dt < time < 0.75+dt)
(shot=44959) and (0.64-dt < time < 0.82+dt)
(shot=44960) and (0.78-dt < time < 0.82+dt)
(shot=44961) and (0.66-dt < time < 0.71+dt)
(shot=44963) and (0.66-dt < time < 0.71+dt)
(shot=44974) and (0.65-dt < time < 0.71+dt)
(shot=45057) and (0.65-dt < time < 0.67+dt)
(shot=45057) and (0.74-dt < time < 0.90+dt)
(shot=45061) and (0.65-dt < time < 0.73+dt)
(shot=45080) and (0.75-dt < time < 0.95+dt)
(shot=45082) and (0.92-dt < time < 0.95+dt)
(shot=45098) and (0.77-dt < time < 0.82+dt)
(shot=45142) and (0.64-dt < time < 0.66+dt)
(shot=45143) and (0.65-dt < time < 0.68+dt)
(shot=45144) and (0.65-dt < time < 0.70+dt)
(shot=45430) and (0.69-dt < time < 0.75+dt)
(shot=45431) and (0.69-dt < time < 0.75+dt)
(shot=45438) and (0.89-dt < time < 0.94+dt)
(shot=45691) and (0.65-dt < time < 0.71+dt)
(shot=45714) and (0.73-dt < time < 0.82+dt)
(shot=45716) and (0.65-dt < time < 0.70+dt)
(shot=45734) and (0.71-dt < time < 0.80+dt)
(shot=45737) and (0.71-dt < time < 0.80+dt)
(shot=45784) and (0.72-dt < time < 0.80+dt)
(shot=48427) and (0.69-dt < time < 0.79+dt)
(shot=48428) and (0.64-dt < time < 0.80+dt)
(shot=48461) and (0.66-dt < time < 0.72+dt)
(shot=48465) and (0.65-dt < time < 0.71+dt)
(shot=48480) and (0.69-dt < time < 0.80+dt)

(shot=48481) and (0.69-dt < time < 0.80+dt)
(shot=48482) and (0.69-dt < time < 0.80+dt)
(shot=48483) and (0.70-dt < time < 0.80+dt)
(shot=48486) and (0.70-dt < time < 0.80+dt)
(shot=49716) and (0.78-dt < time < 0.90+dt)
(shot=49719) and (0.81-dt < time < 0.90+dt)
(shot=56850) and (0.80-dt < time < 0.82+dt)
(shot=56851) and (0.78-dt < time < 0.90+dt)
(shot=56854) and (0.78-dt < time < 0.82+dt)
(shot=56855) and (0.79-dt < time < 0.90+dt)
(shot=56861) and (0.79-dt < time < 0.88+dt)
(shot=56867) and (0.79-dt < time < 0.92+dt)
(shot=56868) and (0.80-dt < time < 0.92+dt)
(shot=56870) and (0.80-dt < time < 0.83+dt)
(shot=56871) and (0.85-dt < time < 0.90+dt)
(shot=56872) and (0.86-dt < time < 0.94+dt)
(shot=57597) and (0.80-dt < time < 0.83+dt)

ACKNOWLEDGEMENTS

The authors acknowledge the work of the ASDEX, DIII-D, JET, JFT-2M, PBX-M, and PDX teams, which made the compilation of the ITERH.DB1 database possible, and they are grateful to the other members of the H-mode Database working group, in particular Dr. Y. Miura and Dr. K. Thomsen, for various discussions. This work was initiated when two of the authors (SII and KI) stayed at the Max-Planck-Institut für Plasmaphysik for an ITER Specialist Meeting. Part of the work was done when one (OK) or more of the authors were visiting JAERI and JET. Thanks are due to the ASDEX, JFT-2M, and JET groups for their hospitality, and to Dr. K. Lackner, Dr. H. Maeda, and Dr. G. Cordey for encouragement and practical realisation of the visits. This work has been partially supported by the collaboration program between JAERI and Universities on Fusion, and by the Grant-in-Aid for Scientific Research of the Ministry of Education in Japan.

REFERENCES

[1] International Atomic Energy Agency, ITER Conceptual Design Report, ITER Documentation Series No. 18, IAEA, Vienna (1991).

[2] The ASDEX Team, The H-mode of ASDEX, Nuclear Fusion **29** (1989) 1959-2040.

[3] Stambaugh, R.D., Wolfe, S.M., Hawryluk, R.K., et al., Enhanced Confinement in Tokamaks, Physics of Fluids B **2** (1990) 2941-2960.

[4] Yushmanov, P.N., Takizuka, T., Riedel, K.S., et al., Scalings for Tokamak Energy Confinement, Nuclear Fusion **30** (1990) 1999-2006.

[5] Kaye, S.M., Barnes, C.W., Bell, M.G., et al., Status of Global Energy Confinement Studies, Physics of Fluids B **2** (1990) 2926-2940.

[6] Itoh K., Itoh, S.-I., and Fukuyama, K., The Impact of Improved Confinement on Fusion Research, Fusion Engeneering and Design **15** (1992) 297-308.

[7] Cordey, J.G., DeBoo, J.C., Kardaun, O., et al., ITER: Preliminary Analysis of Energy Confinement H-mode Database, in Plasma Physics and Controlled Nuclear Fusion Research 1990 (Proc. 13th Int. Conf. Washington, DC, 1990), Vol 3, IAEA, Vienna (1991) 443-452.

[8] Christiansen, J.P., Cordey, J.G., Thomsen, K., et al., Global Energy Confinement Database for ITER, Nuclear Fusion **32** (1992) 291-338, Corrigendum 1281.

[9] Kardaun, O., Thomsen, K., Christiansen, J.P., et al., On Global H-mode Scaling Laws for JET, in: Controlled Fusion and Plasma Heating (Proc. 16th Eur. Conf. Venice 1989), Vol. 13B, Part I, European Physical Society (1989) 253-256.

[10] Kardaun O., Thomsen, K., Cordey, J.G., et al., Global H-mode Scalings based on JET and ASDEX Data, in: Controlled Fusion and Plasma Heating (Proc. 17th Eur. Conf. Amsterdam 1990), Vol. 14B, Part I, European Physical Society (1990) 110-113.

[11] Kardaun O., H-mode Scalings for Energy Confinement (based on the ITERH.DB1 Database), in: IPP Annual Report, Max-Planck-Institut für Plasmaphysik, Garching bei München (1990) 78, and (1991) 38.

[12] Schissel, D., DeBoo, J.C., Burell, K.H., et al., H-mode Energy Confinement Scaling from the DIII-D and JET Tokamaks, Nuclear Fusion **31** (1991) 73-82.

[13] Riedel, K., Random Coefficient H-mode Confinement Scalings, Nuclear Fusion **32** (1992) 1270-1280.

[14] Miura, Y., Takizuka, T., Tamai, H., et al., Geometric Dependence of the Energy Confinement Time Scaling for H-mode Discharges, Nuclear Fusion **32** (1992) 1473-1479.

[15] The H-mode Database Working Group (presented by O.J.W.F. Kardaun), ITER: Analysis of the H-mode Confinement and Threshold Databases, preprint IAEA/F-1-3 (Proc. 14th Int. Conf. on Plasma Physics and Contr. Nuclear Fusion, Würzburg, 1992).

[16] Keilhacker, M., et al., Confinement Studies in L and H-type ASDEX Discharges, in: Plasma Physics and Controlled Fusion **26** (1984) 49-63.

[17] Zohm, H., Wagner, F., Endler, M., et al., Studies of Edge Localized Modes on ASDEX, Nuclear Fusion **32** (1992) 489-494.

[18] Zohm, H., Osborne, T.H., Burrell, K.H., et al., ELM Studies on DIII-D and a Comparison to ASDEX Results, GA report A20895, 1992.

[19] Itoh, S.-I., Maeda, H., and Miura Y., Improved Operating Mode and the Evaluation of Confinement Improvement. Fusion Engeneering and Design **15** (1992) 343-352.

[20] Wagner, F., Ryter, F., Field, A.R., et al., Recent Results of H-mode Studies at ASDEX, in: Plasma Physics and Controlled Nuclear Fusion Research 1990 (Proc. 13th Int. Conf. Washington, DC, 1990), Vol 1, IAEA, Vienna (1991) 277-290.

[21] Miura, Y., Aikawa, H., Hoshino, K., et al., Studies on Improved Confinement on JFT-2M, in: Plasma Physics and Controlled Nuclear Fusion Research 1990 (Proc. 13th Int. Conf. Washington, DC, 1990), Vol 1, IAEA, Vienna (1991) 325-333.

[22] The DIII-D Team (presented by R.D. Stambaugh), DIII-D Research Program Progress, in: Plasma Physics and Controlled Nuclear Fusion Research 1990 (Proc. 13th Int. Conf. Washington, DC, 1990), Vol 1, IAEA, Vienna (1991) 69-91.

[23] Kshirsagar, A.M. (1972). *Multivariate Analysis*, Marcell Dekker, New York.

[24] Krishnaiah, P.R., and Kanal, L.N., eds. (1982). *The Handbook of Statistics, Vol. 2*, North Holland, Amsterdam.

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

[26] Morrison, D.F. (1990). *Multivariate Statistical Methods*, McGraw-Hill International Editions, New York.

[27] Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach*, Chapmann and Hall, London.

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

[29] Kendall, M.G., Stuart A., and Ord, J.K. (1983). *The Advanced Theory of Statistics, Vol. III, 4^{th} ed.* Charles Griffin and Co., London.

[30] Lackenbruch, P.A. (1975). *Discriminant Analysis*, Hafner, New York.

[31] Schaafsma, W. and Van Vark, G.N., Classification and Discrimination Problems with Applications, part II., Statistica Neerlandica **31** (1979) 91-126.

[32] Ambergen, A.W. (1989). *Statistical Uncertainties in Posterior Probabilities*, Thesis, University of Groningen.

[33] Steerneman A.W. (1987). *On the Choice of Variables in Discriminant Analysis and Regression Analysis*, Thesis, University of Groningen.

[34] Schmitz, P.I.M. (1986). *Logistic Regression in Medical Decision Making and Epidemiology*, Thesis, Erasmus University Rotterdam.

[35] McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Interscience.

[36] Gelpke, G.J, and Habbema, J.D.F. (1981). *User's Manual for the INDEP-SELECT Discriminant Analysis Program.* Department of Medical Statistics, Leiden University, Leiden.

[37] Van der Sluis, D.M., Schaafsma, W., and Ambergen, A.W. (1989). *POSCON user manual, A Decision Support System in Diagnosis and Prognosis.* University of Groningen.

[38] SAS Institute Inc., PROC DISCRIM in: *SAS/STAT User's Guide,* Version 6, 1989, Cary, NC.

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

[40] Ferguson, T.S. (1967). *Mathematical Statistics, A Decision Theoretic Approach,* Academic Press, New York.

[41] Fisher, R.A., The use of multiple measurements in taxonomic problems, Ann. of Eugenics **7** (1936) 179-188.

[42] Härle, W. (1990). *Smoothing Techniques with Implementation in S,* Springer Series in Statistics, Springer-Verlag, Heidelberg.

[43] Habbema, J.D.F. and Gelpke, G.J., A computer Program for Selection of Variables in Diagnostic and Prognostic Problems, Computer Programs in Biomedicine **13** (1981) 251-270.

[44] Huber, P.J. (1981). *Robust Statistics,* Wiley, New York.

[45] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics, the Approach based on Influence Functions,* Wiley, New York.

[46] Fienberg, S.E. and Holland, P.W., On the Choice of Flattening Constants for Estimating Multinomial Probabilities, Journal of Multivariate Analysis, **2** (1972) 127-134.

[47] Kardaun, J.W.P.F. and Kardaun, O.J.W.F., Comparative Diagnostic Performance of Three Radiological Procedures for the Detection of Lumbar Disk Herniation, Methods of Information in Medicine **29** (1990) 12-22.

[48] Sadler, G. Barabaschi, P., Bertolini,E., et al., Effects of Enhanced Toroidal Field Ripple on JET Plasmas, in: Plasma Physics and Controlled Fusion **34** (1992) 1971-1976.

[49] Huysmans, G.T.A., de Blank, H.J., Kerner, W., Goedbloed, J.P., and Nave, M.F.F. (1992), MHD Stability Models of Edge Localized Modes in JET Discharges, in: Controlled Fusion and Plasma Heating (Proc. 19th Eur. Conf. Innsbruck 1992), Vol. 16B, Part I, European Physical Society (1992) 247-250.

[50] Kardaun, O.J.W.F, Kardaun, J.W.P.F., Itoh, S.-I., and Itoh, K., Discriminant Analysis of Plasma Fusion Data, in: Computational Statistics (Proc. 10th Symposium on Computational Statistics 1992), ed. by Y. Dodge and J. Whittaker, Physica-Verlag Heidelberg, Vol. 1, 163-170. Also NIFS research report 156.

[51] McCarthy, P.J. (1992). *An Integrated Data Interpretation System for Tokamak Discharges*, Thesis, University of Cork.

[52] O'Brian, D.P., Ellis, J.J., and Lingertat, J., Local Expansion Method for Fast Plasma Boundary Identification in JET, Nuclear Fusion **33** (1993) 467-474.

[53] Engelmann, F., Fujisawa, N., Luxon, J., et al., ITER: Physics R and D Programme, in: Plasma Physics and Controlled Nuclear Fusion Research 1990 (Proc. 13th Int. Conf. Washington, DC, 1990), Vol 3, IAEA, Vienna (1991) 435-442.

Recent Issues of NIFS Series

NIFS-189 K. Itoh, S. -I. Itoh and A. Fukuyama, *Cross Field Ion Motion at Sawtooth Crash* ; Oct. 1992

NIFS-190 N. Noda, Y. Kubota, A. Sagara, N. Ohyabu, K. Akaishi, H. Ji, O. Motojima, M. Hashiba, I. Fujita, T. Hino, T. Yamashina, T. Matsuda, T. Sogabe, T. Matsumoto, K. Kuroda, S. Yamazaki, H. Ise, J. Adachi and T. Suzuki, *Design Study on Divertor Plates of Large Helical Device (LHD)* ; Oct. 1992

NIFS-191 Y. Kondoh, Y. Hosaka and K. Ishii, *Kernel Optimum Nearly-Analytical Discretization (KOND) Algorithm Applied to Parabolic and Hyperbolic Equations* : Oct. 1992

NIFS-192 K. Itoh, M. Yagi, S.-I. Itoh, A. Fukuyama and M. Azumi, *L-Mode Confinement Model Based on Transport-MHD Theory in Tokamaks* ; Oct. 1992

NIFS-193 T. Watari, *Review of Japanese Results on Heating and Current Drive* ; Oct. 1992

NIFS-194 Y. Kondoh, *Eigenfunction for Dissipative Dynamics Operator and Attractor of Dissipative Structure* ; Oct. 1992

NIFS-195 T. Watanabe, H. Oya, K. Watanabe and T. Sato, *Comprehensive Simulation Study on Local and Global Development of Auroral Arcs and Field-Aligned Potentials* ; Oct. 1992

NIFS-196 T. Mori, K. Akaishi, Y. Kubota, O. Motojima, M. Mushiaki, Y. Funato and Y. Hanaoka, *Pumping Experiment of Water on B and LaB_6 Films with Electron Beam Evaporator* ; Oct., 1992

NIFS-197 T. Kato and K. Masai, *X-ray Spectra from Hinotori Satellite and Suprathermal Electrons* ; Oct. 1992

NIFS-198 K. Toi, S. Okamura, H. Iguchi, H. Yamada, S. Morita, S. Sakakibara, K. Ida, K. Nishimura, K. Matsuoka, R. Akiyama, H. Arimoto, M. Fujiwara, M. Hosokawa, H. Idei, O. Kaneko, S. Kubo, A. Sagara, C. Takahashi, Y. Takeiri, Y. Takita, K. Tsumori, I. Yamada and H. Zushi, *Formation of H-mode Like Transport Barrier in the CHS Heliotron / Torsatron* ; Oct. 1992

NIFS-199 M. Tanaka, *A Kinetic Simulation of Low-Frequency Electromagnetic Phenomena in Inhomogeneous Plasmas of Three-Dimensions* ; Nov. 1992

NIFS-200 K. Itoh, S.-I. Itoh, H. Sanuki and A. Fukuyama, *Roles of Electric Field on Toroidal Magnetic Confinement*, Nov. 1992

NIFS-201 G. Gnudi and T. Hatori, *Hamiltonian for the Toroidal Helical Magnetic Field Lines in the Vacuum*; Nov. 1992

NIFS-202 K. Itoh, S.-I. Itoh and A. Fukuyama, *Physics of Transport Phenomena in Magnetic Confinement Plasmas*; Dec. 1992

NIFS-203 Y. Hamada, Y. Kawasumi, H. Iguchi, A. Fujisawa, Y. Abe and M. Takahashi, *Mesh Effect in a Parallel Plate Analyzer*; Dec. 1992

NIFS-204 T. Okada and H. Tazawa, *Two-Stream Instability for a Light Ion Beam -Plasma System with External Magnetic Field;* Dec. 1992

NIFS-205 M. Osakabe, S. Itoh, Y. Gotoh, M. Sasao and J. Fujita, *A Compact Neutron Counter Telescope with Thick Radiator (Cotetra) for Fusion Experiment*; Jan. 1993

NIFS-206 T. Yabe and F. Xiao, *Tracking Sharp Interface of Two Fluids by the CIP (Cubic-Interpolated Propagation) Scheme*, Jan. 1993

NIFS-207 A. Kageyama, K. Watanabe and T. Sato, *Simulation Study of MHD Dynamo : Convection in a Rotating Spherical Shell*; Feb. 1993

NIFS-208 M. Okamoto and S. Murakami, *Plasma Heating in Toroidal Systems*; Feb. 1993

NIFS-209 K. Masai, *Density Dependence of Line Intensities and Application to Plasma Diagnostics*; Feb. 1993

NIFS-210 K. Ohkubo, M. Hosokawa, S. Kubo, M. Sato, Y. Takita and T. Kuroda, *R&D of Transmission Lines for ECH System* ; Feb. 1993

NIFS-211 A. A. Shishkin, K. Y. Watanabe, K. Yamazaki, O. Motojima, D. L. Grekov, M. S. Smirnova and A. V. Zolotukhin, *Some Features of Particle Orbit Behavior in LHD Configurations*; Mar. 1993

NIFS-212 Y. Kondoh, Y. Hosaka and J.-L. Liang, *Demonstration for Novel Self-organization Theory by Three-Dimensional Magnetohydrodynamic Simulation*; Mar. 1993

NIFS-213 K. Itoh, H. Sanuki and S.-I. Itoh, *Thermal and Electric Oscillation Driven by Orbit Loss in Helical Systems*; Mar. 1993

NIFS-214 T. Yamagishi, *Effect of Continuous Eigenvalue Spectrum on Plasma Transport in Toroidal Systems*; Mar. 1993

NIFS-215 K. Ida, K. Itoh, S.-I.Itoh, Y. Miura, JFT-2M Group and A. Fukuyama, *Thickness of the Layer of Strong Radial Electric Field in JFT-2M H-mode Plasmas*; Apr. 1993

NIFS-216 M. Yagi, K. Itoh, S.-I. Itoh, A. Fukuyama and M. Azumi, *Analysis of Current Diffusive Ballooning Mode*; Apr. 1993

NIFS-217 J. Guasp, K. Yamazaki and O. Motojima, *Particle Orbit Analysis for LHD Helical Axis Configurations* ; Apr. 1993

NIFS-218 T. Yabe, T. Ito and M. Okazaki, *Holography Machine HORN-1 for Computer-aided Retrieve of Virtual Three-dimensional Image* ; Apr. 1993

NIFS-219 K. Itoh, S.-I. Itoh, A. Fukuyama, M. Yagi and M. Azumi, *Self-sustained Turbulence and L-Mode Confinement in Toroidal Plasmas* ; Apr. 1993

NIFS-220 T. Watari, R. Kumazawa, T. Mutoh, T. Seki, K. Nishimura and F. Shimpo, *Applications of Non-resonant RF Forces to Improvement of Tokamak Reactor Performances Part I: Application of Ponderomotive Force* ; May 1993

NIFS-221 S.-I. Itoh, K. Itoh, and A. Fukuyama, *ELMy-H mode as Limit Cycle and Transient Responses of H-modes in Tokamaks* ; May 1993

NIFS-222 H. Hojo, M. Inutake, M. Ichimura, R. Katsumata and T. Watanabe, *Interchange Stability Criteria for Anisotropic Central-Cell Plasmas in the Tandem Mirror GAMMA 10 ;* May 1993

NIFS-223 K. Itoh, S.-I. Itoh, M. Yagi, A. Fukuyama and M. Azumi, *Theory of Pseudo-Classical Confinement and Transmutation to L-Mode;* May 1993

NIFS-224 M. Tanaka, *HIDENEK: An Implicit Particle Simulation of Kinetic-MHD Phenomena in Three-Dimensional Plasmas;* May 1993

NIFS-225 H. Hojo and T. Hatori, *Bounce Resonance Heating and Transport in a Magnetic Mirror;* May 1993

NIFS-226 S.-I. Iton, K. Itoh, A. Fukuyama, M. Yagi, *Theory of Anomalous Transport in H-Mode Plasmas;* May 1993

NIFS-227 T. Yamagishi, *Anomalous Cross Field Flux in CHS ;* May 1993

NIFS-228 Y. Ohkouchi, S. Sasaki, S. Takamura, T. Kato, *Effective Emission and Ionization Rate Coefficients of Atomic Carbons in Plasmas;* June 1993

NIFS-229 K. Itoh, M. Yagi, A. Fukuyama, S.-I. Itoh and M. Azumi, *Comment on 'A Mean Field Ohm's Law for Collisionless Plasmas;* June 1993

NIFS-230 H. Idei, K. Ida, H. Sanuki, H. Yamada, H. Iguchi, S. Kubo, R. Akiyama, H. Arimoto, M. Fujiwara, M. Hosokawa, K. Matsuoka, S. Morita, K. Nishimura, K. Ohkubo, S. Okamura, S. Sakakibara, C. Takahashi, Y. Takita, K. Tsumori and I. Yamada, *Transition of Radial Electric Field by Electron Cyclotron Heating in Stellarator Plasmas;* June 1993

NIFS-231 H.J. Gardner and K. Ichiguchi, *Free-Boundary Equilibrium Studies for the Large Helical Device,* June 1993

NIFS-232 K. Itoh, S.-I. Itoh, A. Fukuyama, H. Sanuki and M. Yagi, *Confinement Improvement in H-Mode-Like Plasmas in Helical Systems,* June 1993

NIFS-233 R. Horiuchi and T. Sato, *Collisionless Driven Magnetic Reconnection,* June 1993

NIFS-234 K. Itoh, S.-I. Itoh, A. Fukuyama, M. Yagi and M. *Azumi, Prandtl Number of Toroidal Plasmas;* June 1993

NIFS-235 S. Kawata, S. Kato and S. Kiyokawa , *Screening Constants for Plasma;* June 1993

NIFS-236 A. Fujisawa and Y. Hamada, *Theoretical Study of Cylindrical Energy Analyzers for MeV Range Heavy Ion Beam Probes;* July 1993

NIFS-237 N. Ohyabu, A. Sagara, T. Ono, T. Kawamura and O. Motojima, *Carbon Sheet Pumping;* July 1993

NIFS-238 K. Watanabe, T. Sato and Y. Nakayama, *Q-profile Flattening due to Nonlinear Development of Resistive Kink Mode and Ensuing Fast Crash in Sawtooth Oscillations;* July 1993

NIFS-239 N. Ohyabu, T. Watanabe, Hantao Ji, H. Akao, T. Ono, T. Kawamura, K. Yamazaki, K. Akaishi, N. Inoue, A. Komori, Y. Kubota, N. Noda, A. Sagara, H. Suzuki, O. Motojima, M. Fujiwara, A. Iiyoshi, *LHD Helical Divertor;* July 1993

NIFS-240 Y. Miura, F. Okano, N. Suzuki, M. Mori, K. Hoshino, H. Maeda, T. Takizuka, JFT-2M Group, K. Itoh and S.-I. Itoh, *Ion Heat Pulse after Sawtooth Crash in the JFT-2M Tokama;* Aug. 1993

NIFS-241 K. Ida, Y.Miura, T. Matsuda, K. Itoh and JFT-2M Group, *Observation of non Diffusive Term of Toroidal Momentum Transport in the JFT-2M Tokamak;* Aug. 1993