EFDA–JET–PR(09)04

A. Murari, J. Vega, D. Mazon, N. Martin, G. Rattá, G.Vagliasindi
and JET EFDA contributors

Machine Learning Methods for Data Driven Theory in the Physical Sciences with Applications to Confinement Regime Identification in Nuclear Fusion

Machine Learning Methods for Data Driven Theory in the Physical Sciences with Applications to Confinement Regime Identification in Nuclear Fusion

A. Murari[1], J. Vega[2], D. Mazon[3], N. Martin[3], G. Rattá[2], G.Vagliasindi[4]
aand JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy.
[2]Asociación EURATOM-CIEMAT para Fusión, CIEMAT, Madrid, Spain
[3]Association EURATOM-CEA, CEA Cadarache DSM/IRFM, 13108 Saint-Paul-lez-Durance, France
[4]Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi-Università degli Studi di Catania, 95125 Catania, Italy
** See annex of F. Romanelli et al, "Overview of JET Results", (Proc. 22 nd IAEA Fusion Energy Conference, Geneva, Switzerland (2008)).*

Preprint of Paper to be submitted for publication in
Reports on Progress in Physics

ABSTRACT.
In various fields of science complex, non linear systems have to be investigated. Given the rapid development of information technology, in many cases a lot of data is available to study the phenomena of interest but, contrary to what happened in previous societies, now very often the bottleneck is the ability to extract the right knowledge from these large repositories of information. In this paper two machine learning tools, Neural Networks and Support Vector Machines, are used for data driven theory formulation, to derive mathematical expressions directly from the database with a minimum of a priori hypotheses. As an example, the methodology is applied to one of the major problems in present day magnetic confinement fusion: the understanding of the scaling laws for the access to the high confinement regime. First the two numerical tools are refined to extract numerical relations from an international database, which are compared with the most widely accepted theoretical models. Since the international database does not contains signals covering the entire evolution of the discharges, a most sophisticated process is developed, using a dedicated and validated database of JET Joint Undertaking discharges, to study in more detail the plasma evolution close to the transition. A specific approach is developed to determine the dimensionality of the problem and then to select the most relevant signals. Specific mathematical relations for the scaling of the temperature threshold are derived again using NNs and SVMs independently. Clear evidence is collected for the existence of two different types of transition to the H mode. Both NNs and SVMs converge on very similar equations which show a success rate in interpreting JET database of practically 100% once the error bars in the measurements are taken into account. The proposed methodology therefore seems to be able to extract, in an automatic, robust and user independent way, the right physics from the database. Additional refinements of the approach to provide also confidence intervals in the derived mathematical equations are also presented.

1. INTRODUCTION

In several fields of modern science, from high energy physics and astrophysics to medicine and sociology, the task of extracting useful information from massive databases has become increasingly common. This is a consequence of the recent, enormous developments in computing and information technologies, which allow generating and recording enormous amounts of information. These new capabilities have changed completely the quantity of data available from both modern societies and new experiments. In the last survey performed by Berkley University the world annual production of new data in 2002 has been estimated to be of the order of 4 exabytes [1]. The new Big Physics experiments show the same trend as modern societies; the Hubble Space Telescope has already sent Tbytes of data to earth and Large Hadron Collider at CERN is expected to produce Tbytes of data per year.

The plasmas produced in Magnetic Confinement Fusion are very complex, open systems. They are kept out of equilibrium by injection of material and energy in order to sustain configurations capable of maximising the fusion rate. To understand the behaviour of these plasmas, an increased

number of instruments called diagnostics take many measurements of all the most important physical parameters. In the largest devices these diagnostics can produce very large amounts of information. In the JET Joint Undertaking (JET), the largest fusion device in the world, more than 10 Gbytes of data, which is the equivalent of a 2 hours digital movie in terms of information content, can be generated per shot. JET entire database now exceeds 40 Tbytes and the extrapolations indicate that the volume of data will be orders of magnitude larger in the next generation of devices, like ITER. It is therefore increasingly difficult to analyse all this information manually or with traditional methods. Moreover nuclear fusion plasmas are not very accessible for measurement and therefore many parameters are derived indirectly or from spontaneous emissions. As a consequence, difficult inversion problems have often to be solved, like in the case of tomographic reconstructions, to determine the physical quantities of interest. It is also worth emphasizing that reactor relevant plasmas are integrated systems and therefore any serious attempt at understanding them should aim at a holistic view of these physical entities. The data analysis process is therefore qualitative different from high energy physics, where only a few significant events have to extract from a myriad of unwanted reactions, and is more similar to the task of the health sciences, where a complete view of the patient conditions is required to optimise treatments and cures.

All the aforementioned issues motivate the development of new exploratory and data analysis methods to be able to process automatically large amounts of information and derive the required knowledge from the information stored in the databases. On the other hand, the most widely used techniques to help in this respect have been originally developed for the private sector, to understand the market and the costumer behaviour. Therefore these methodologies are more suited to giving qualitative trends or, in any case, they are not explicitly conceived to provide a mathematical formulation of the phenomena. They need therefore to be explicitly refined and tuned to attack typical physical problems of the hard sciences like the determination of scaling laws.

Among the vast class of models which are studied in various fields of science, scaling laws are of particular relevance. In many fields of research they are expressed in the form of power law monomials:

$$A = \frac{B^{\beta} C^{\chi} \ldots D^{\delta}}{G^{\gamma} H^{\lambda} \ldots H^{\mu}} \tag{1}$$

where the capital letters represent physical quantities. The importance of these scaling laws is at least twofold. On the one hand, they allow extrapolating results from one range of dimensional parameters to another (typically from the model to the prototype in engineering applications [2]). On the other hand, they also reveal some important properties of the phenomena under consideration, in particular their self-similar character [3].

In this paper various machine learning methods, mainly Classification and Regression Trees (CART), Artificial Neural Networks (NNs) and Support Vector Machines (SVMs) are refined, originally developed and then used to perform a complete process of theory formulation, from the

assessment of the problem dimensionality and the variable selection to the final determination of the mathematical laws describing the scaling of the phenomenon with the physical quantities. This comprehensive approach is applied to the specific problem of identifying the main laws governing the access to the high confinement regime of the plasma, the so called H mode, which is introduced in section 2. First the developed method is applied to an international database covering all the major tokamak devices in the world. The mathematical formulations of the most credited theoretical models are systematically compared with the equations derived from the NNs and the SVM (see section 3). The relations of the independent data driven methods, NNs and SVMs, agree very well with each other and manage to discriminate the H form the L mode of confinement significantly better than the theoretical models, which cannot be considered fully satisfactory. Therefore, to improve the understanding of the physics behind the transition between the two confinement modes, a highly optimised JET database, carefully validated by experts, has been analysed. An original combination of CART, Principal Component Analysis (PCA) and Autoassociative Neural Networks (ANNs) has been devised to determine the dimensionality of the problem and to identify the variables with the highest information content (section 4). Using the variables determined with the procedure described, a systematic investigation has been performed to improve the performance of the ANNs and SVMs for the classification of the plasma confinement state. It is very interesting to note that again the mathematical relations derived from the NNs and the SVMs classifiers not only have very similar performance, significantly higher than the most credited theoretical models, but they also agree with each other within the experimental uncertainties, whereas they can significantly differ from the theoretical models. Moreover, since JET database contains more detailed, even if less general, information, it has been possible to further refine the analysis. The detailed evolution of the plasma around the time of the transition from the L to the H mode of confinement has been analysed. This has allowed investigating the effects of the measurement errors; once the uncertainties in the measurements are taken into account the data driven equations obtained with the NNs and the SVMs classify JET database with a success rate of practically 100%. Moreover a specific methodology has been developed to associate confidence intervals to the mathematical relations derived with the NNs and the SVMs, which is another significant advantage with respect to the formulas derived with traditional methods of theory formulation (see section 5). The detailed analysis of the plasma time evolution around the transition from the L to the H mode of confinement has also allowed identifying two different types of transitions, which is another original result of the proposed approach. This new evidence has been correlated to different collisionality regimes.

The fact that the equations derived with two completely independent machine learning techniques provide very coherent results, in terms of both performance and variable dependences, increases the confidence in the approach and supports the application of the proposed methodology to other physical problems. The very positive results obtained, for both the derivation of mathematical formulas and the extraction of new physical insight from the database, therefore motivate further research, whose main elements are mentioned in the last section together with a summary of the paper.

2. DIFFERENT CONFINEMENT REGIMES IN REACTOR RELEVANT PLASMAS: EMPIRICAL STUDIES AND THEORETICAL MODELS

2.1 THE H-MODE OF CONFINEMENT

In Magnetic Confinement Fusion (MCF), plasmas of hydrogen isotopes are heated to high temperatures in order for the nuclei to overcome the Coulomb barrier and reach distances small enough to provide a sufficient rate of fusion reactions. To increase the net energy output, the density and confinement time of the plasma have to be increased as much as possible at temperature of tens of KeVs. At these temperatures, the plasma must be kept away from the material walls and is indeed contained using suitably shaped magnetic fields. The today most studied configuration in the world remains the Tokamak, a Russian acronym for "Magnetic Bottle", originally introduced in the sixties [4]. Unfortunately, pushing the parameters in the direction of maximizing the economic viability increases also the vulnerability of Tokamak plasmas to various instabilities, which affect the confinement and can even cause the sudden and unplanned collapse of the configuration. Moreover, due to their inherent complexity and nonlinear nature, a sound theory of tokamak plasma behaviour is not available. Even basic trends are therefore derived by experimental investigations, taking place in various machines of different size located around the world. In the perspective of designing the next generation of devices, one of the main scientific and practical issues consists of understanding the confinement properties of the various plasma configurations and their scaling with the main physical and engineering parameters.

During the empirical study of tokamak plasma properties in many devices, a number of distinct regimes of confinement have been identified. Known as confinement modes, these regimes often offer unique confinement characteristics and different scaling of the performance with the input power. Of the modes identified, the Ohmic mode, the L-mode, the H-mode and the configurations involving internal transport barriers are the most robust and most widely investigated [4]. These regimes have been observed repeatedly and reproduced reliably in a great number of different machines.

Before introducing the H-mode, it is appropriate to discuss first the regimes which precede it both historically and in the natural evolution of a plasma discharge, the Ohmic and L-mode. In the first operational mode utilized in early tokamaks, the energy input to the plasma was in the form of pure ohmic heating. As such, it became known as the "Ohmic confinement mode". While at the beginning the confinement times achieved in this regime seemed very promising, it was later found that ohmic heating alone was too inefficient to achieve the high core parameters required by a thermonuclear reactor. Indeed, as the plasma temperature increases, its resistance drops, reducing the level of heating that can be achieved ohmically. The Ohmic mode was therefore quickly ruled out as a viable route to a fusion power plant. The use of additional heating schemes, ranging from the injection of neutral beams to radiofrequencies and microwaves, increased significantly the temperatures achievable but had a detrimental effect on the confinement time, since these forms of energy input tend to generate a series of instabilities, which can be severely penalising in terms of increased transport. The first achieved regime with additional heating is therefore now generically identified as the L-mode, which stands for low confinement mode.

In the ASDEX device it was then discovered in 1982 that, increasing further the input power, the plasmas tended to transit spontaneously to an enhanced confinement mode called the High confinement or H-mode[5]. The H-mode is characterized by the presence of a thin region of very low transport situated at the edge of the plasma. Steep gradients in the density and temperature profiles are observed across this region as illustrated in figure 1. This thin layer of increased gradients in the kinetic profiles is commonly referred to as the 'H-mode pedestal'. Inside of the pedestal region, the profiles remain broadly similar to the L-mode except for being shifted upwards by the pedestal. The thickness of the pedestal is typically 1-5% of the radius of the plasma, and therefore in JET the measured pedestal widths are of the order of a few centimetres. The low transport region at the edge of the H-mode plasma is known as an Edge Transport Barrier (ETB). Determining the scaling laws for the threshold to access the H-mode is one of the most important research topics in the perspective of the next generation international device ITER.

2.2 THE L-H TRANSITION AND THE THEORETICAL MODELS FOR ITS INTERPRETATION

L-H transitions are experimentally observed to occur only when the heating power applied to the plasma exceeds a critical value, which has become known as the L-H power threshold (Pth). The H-mode is typically reached transiting first through the L-mode of operation. Once the correct L-mode conditions are met, the transition from L to H mode occurs with the spontaneous formation of an ETB. When the ETB is starting to develop, the confinement at the edge of the plasma improves which consequentially sustains the further growth of the H-mode pedestal.

With regards to the time scales, the formation of an ETB occurs in matter of milliseconds, as can be seen by its effect on particle transport. Observation of the particle flux from the plasma edge, proportional to the emission of the D±. line of deuterium, shows a dramatic drop in particle loss, seen as a drop in D± intensity, corresponding to the ETB formation. To illustrate this aspect figure 2 shows the D± signal from the outer divertor during an H-mode shot in JET. Measurements from reflectometers and high speed visual imaging systems show sharp reductions in the amplitude of edge fluctuations coinciding with the onset of the ETB.

Due to the importance of achieving the H-mode on the next step device, ITER, a lot of experimental time has been devoted to identifying the scaling of the power threshold to access it. Without any established theoretical models of additional heating or confinement physics, determining in detail how a particular input power generates the plasma parameters required to access the H mode has not been possible. Currently, therefore the parametric scaling of the P_{th} is best understood through empirical analysis of large shot databases. On the other hand, a series of theoretical models exists, which try to interpret the onset of an ETB as the interplay between plasma instabilities and various stabilising factors. These models consider a wide range of different physical phenomena and provide some testable criteria for the onset of the H-mode in terms of measurable plasma parameters. These criteria are expressed in terms of a critical electron temperature Te, which is considered to be

somehow directly proportional to the input power, and are dependent on a wide range of edge-local plasma parameters. The most widely accepted theoretical models available, which express the temperature threshold to enter the H mode in terms of general plasma parameters measured in a wide range of machines, are summarise in table I [6].

The theoretical scaling laws for the access to the H mode of confinement are typically expressed in a monomial form of the type:

$$T_{ec} \propto k\, p_1^a p_2^b \ldots p_n^q \tag{2}$$

where T_{ec} is the temperature threshold and p_i are the various variables. These scaling laws are the results of the interplay between the main instabilities, believed to affect the plasma before the transition to the H mode, and some stabilising phenomena taking over when the right conditions are met.

The model developed in [7,8] assumes that the drift wave turbulence is the main transport mechanism in the L mode and the resistive skin effects are considered the main stabilising factor. In [9] a model based on drift wave instabilities has been particularised for two different assumptions about the transport (convective or conductive) in the Scrap Off Layer (SOL). Two different equations are therefore obtained for the temperature threshold, one for convective and one for conductive transport in the SOL. In the model described in [10] also effects of the neutral particles in the core are taken into account. Drift wave turbulence in toroidal geometry is the basis of the computational model developed in [11]. The mechanism of turbulence suppression assumed in [12] to trigger the onset of the H mode is considered to be the flow shear generated by the ion orbit losses. It is worth mentioning that all these models determine only a proportionality relation between Tec and the various physical quantities.

All these models, which have been derived from basic theoretical considerations on the main processes linking instabilities and transport in thermonuclear plasmas, will be compared with data driven equations obtained by specifically developed NNs and SVMs in the next section.

3. NEURAL NETWORKS AND SVM FOR THEORY FORMULATION AND MODEL SELECTION

3.1 NEURAL NETWORK ARCHITECTURE AND REGULARIZATION

Neural networks are a soft computing tool which was originally conceived for the simulation of the brain but their use has become widely spread in many domains of science. From a mathematical point of view, NNs with nonlinear activation functions, are universal approximators and can therefore be used to determine the boundary between the L and H mode operational space. Backpropagation is the method normally used to train the traditional Multilayer Perceptrons and it is the one adopted also in the vast majority of applications even today. But, despite the many successful applications of backpropagation, it still presents some delicate points that have to be addressed when an ambitious

task, like the one of deriving mathematical formulas from the network architecture and its coefficients, is to be tackled. In our case the most significant difficulties to be expected are the presence of local minima and overfitting. With regard to the first issue, it must be remembered that backpropagation employs a type of gradient descent; that is, it follows the slope of the error surface downward, constantly adjusting the weights toward a minimum. The error surface of a complex network is highly convoluted, full of hills, valleys, folds, and gullies in high-dimensional space. The network can get trapped in a local minimum (a shallow valley) even when there is a much deeper minimum nearby. From the limited viewpoint of the network, all directions are up, and it has no way to escape. Overfitting is another fault in the convergence of a network, whereby the error on the training set is driven to a very small value, but when new data is presented to the network the error rate becomes large. In this case, the network has memorized the training examples, but it has not learned to generalize to new situations.

One method for improving network generalization is to use architectures that are just large enough to provide an adequate fit. The larger the network, the more complex the functions the network can create. On the other hand, if a small enough network architecture is adopted, it will not have enough power to overfit the data. For our purposes, the derivation of manageable mathematical relations to understand the physics involved in the L-H transition and its scaling laws, a simple network architecture is particularly advantageous. As discussed later in more detail, the architecture shown in figure 3 has been proved to be adequate for the problem of the L-H transition. For the middle layer, a nonlinear function $f1 = \tanh(x) = 2/(1- e^{-2x}) -1)$ has been adopted. It has indeed been demonstrated that this configuration, one linear plus a nonlinear stage, constitutes a universal approximator [13].

Unfortunately, in general it is difficult to know beforehand how large a network should be for a specific application and indeed in our case converging on the architecture of figure 3 has been a relatively time consuming trial and error process. Moreover, simple network architecture is not always enough to protect against overfitting and local minima. That is the reason why other methods for improving generalization have been developed and the solution adopted in this paper is a form of Bayesian regularization. To understand the potential and the objective character of this approach, a short review of the basic framework for learning in networks is in place. The training set for the mapping to be learned is a set of input target pairs $D = \{p^m, t^m\}$, where m is a label running over the pairs, p are the inputs and t the targets. A neural network architecture A is adopted, consisting of the specification of the number of layers, the number of units in each layer, the type of activation function performed by each unit, and the available connections between the units. If a set of values w is assigned to the connections in the network, the network defines a mapping a(p;w,A) from the inputs p to the outputs a. The distance of this mapping to the training set is measured by some error function; for example the error for the entire data set is commonly taken to be

$$E_D\,(D, |w; A) = \sum_m \left[\mathrm{a}\,(p^m; w, A) - t^m \right]^2 \tag{3}$$

The task of 'learning' consists of finding a set of connections *w*, which gives a mapping that fits the training set well, i.e. has small error E_D; it is also hoped that the learned connections will 'generalize' well to new examples. Traditional backpropagation learns by performing gradient descent on E_D in the *w* space. In order to improve the generalization properties of the networks, by avoiding local minima and overfitting, it is possible to add extra regularizing terms $E_W(w)$ to E_D. A natural choice consists of introducing terms which penalize large weights in the hope of achieving a smoother or simpler mapping. These additional terms fall into the category of additive weight dependent energies.

A sample weight energy term is:

$$F = \alpha E_w (w \backslash A) + \beta E_D (D | w; A) \tag{4}$$

Gradient based optimization is then used to minimize the combined function:

$$F = \alpha E_w (w \backslash A) + \beta E_D (D | w; A) \tag{5}$$

where α and β are empirical parameters, typically optimised by trial and error for the problem at hand. Using this performance function will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to overfit. The relative size of the objective function parameters dictates the emphasis for training. If $\alpha \ll \beta$, then the training algorithm will result in smaller errors. If $\alpha \gg \beta$, training will emphasize weight size reduction at the expense of network errors, thus producing a smoother network response [14]. In our case the described regularisation has proved to be a very sound method to converge on stable configurations of the NNs. Several tests have been performed, repeating the training starting from different initial conditions and the regularisation has permitted to converge always on networks of very similar weights. From the point of view of local minima avoidance, this approach is therefore to be considered very satisfactory. It also allows overcoming the problem of overfitting since, as shown later, the regularization capabilities of the NNs trained in this way result in particularly good performance even in the test with new examples.

Using the network architecture and the activation function shown in figure 3, it is possible to derive from the network a mathematical formulation of the temperature threshold for accessing the H mode as a monomial expression of the type (2). The first step in this direction consists of providing to the network for training the signals as natural logarithms (normalised to assume values between 0 and 1 as it is normal praxis with NNs). The NN can be then trained to provide an output higher than a certain threshold c when the plasma is in the H mode of confinement. Typically again the output can be rescaled to remain in the interval range [0-1] and therefore the value of c can be chosen to be 0.5 as done for the analysis reported in the following. In any case, keeping the notation c for the threshold to access the H mode and remembering that the original inputs were provided as natural logarithms, the NN represented in figure 3 would indicate that the plasma is in the H mode when:

$$\frac{2k}{(1 + e^{-2b1}\, T_e^{-2w1}\ n_e^{-2w2}\ B_{T80}^{-2w3}\ q_{80}^{-2w4})} + (b_2 - k) > c$$

where ne is the electron density, B_{T80} and q_{80} are the toroidal field and the safety factor at 80 % of the plasma minor radius and the other symbols correspond to the notation of fig.3 and therefore the w_i are the network coefficients and the b_i the biases. This expression can be algebraically rearranged to extract the dependence in terms of T_e:

$$T_{ec}\, \gamma n_e^{\alpha} B_{T80}^{\beta}\, q_{80}^{\theta} \tag{6}$$

where

$$\gamma = \left[\frac{e^{-2bl}\,(C + b_2 + k)}{-C + b_2 + k}\right]^{1/Zwl} \qquad \alpha = -\,w_1/w_2 \quad \beta = -\,w_3/w_1 \quad \theta = -\,w_4/w_1 \tag{7}$$

This relation is exactly in the monomial form typical of the most widely accepted theoretical models. The term between the squares is a simple constant whereas the dependence from the physical variables is in the form of a quantity elevated to an exponent. All the numerical constants, the coefficients w_i and the biases b_i together with the threshold c, can be simply derived by the trained network, since they are the final results of the training process. To summarise, a NN with the architecture shown in figure three can be trained to discriminate time slices belonging to the H and L mode phases of the discharges. If the input signals are given in logarithmic form, the scaling law for the temperature threshold can then be derived using relations (7). Selecting the architecture of figure 3 for the NN implies assuming that the scaling law can be expressed as a monomial form of the type given by relation (2).

3.2 SUPPORT VECTOR MACHINES

The same functional expression derived using the NNs described in the previous section can be obtained also with a completely independent numerical method based on the SVM. SVMs are mathematical tools which can perform more general tasks, like regression, but which are used as classifier in the application described in this paper [15]. They basically consist of suitable kernels, which project the inputs into higher dimensional spaces, where the classification becomes a separable problem and can be solved with traditional quadratic programming methods. In more detail, given a training set of ℓ samples $(\mathbf{x}_1, y_1),\ldots(\mathbf{x}_\ell, y_\ell)$, $x_i \in \Re^n$, for a binary classification problem (*i.e.* $y_i \in \{+1,-1\}$, SVM estimates the following decision function:

$$D(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i\, y_i\, H(\mathbf{x}_i, \mathbf{x}) \tag{8}$$

whereH $(\boldsymbol{x}_i, \boldsymbol{x})$ is a kernel function [16] and the parameters α_i , $i = 1,\ldots,\ell$ are the solutions of the following quadratic optimization with linear constraints:

maximization of the functional

$$Q(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j H(\mathbf{x}_i, \mathbf{x}_j) \tag{9}$$

subject to the constraints

$$\sum_{i,j=1}^{\ell} y_i \alpha_j = 0, \quad 0 \quad \alpha_i \quad \frac{C}{\ell}, \; i = 1,...,\ell \tag{10}$$

where C is a regularization parameter [16].

The data points $\boldsymbol{x}_i$ associated with nonzero values of the coefficients α_i are called support vectors. Once the support vectors have been determined, the SVM decision can be expressed in the form

$$D(\mathbf{x}) = \sum_{\text{support vectors}} \alpha_i y_i H(\mathbf{x}_i, \mathbf{x}) \tag{11}$$

$D(\mathbf{x})$ is the distance from the input $\mathbf{x}$ to the hyper-plane that separates the two classes and, hence, the hyper-plane points satisfy $D(\mathbf{x}) = 0$.

The rule to clarify a feature vector $\mathbf{u}$ as L mode (class C_L) or H mode (class C_H) is given by:

if sgn $(D(\mathbf{u})) \geq 0$

$\mathbf{u} \in C_L$

otherwise

$\mathbf{u} \in C_H$

where sgn(t) is the sign function.

A polinomial kernel is of the form $H(\mathbf{x}, \mathbf{x}') = [\mathbf{x} \cdot \mathbf{x}')+1]^d$.For the application of discriminating between the L and H mode of confinement, it has been demonstrated that typically a linear kernel, obtained by setting $d = 1$ in the previous formula, is perfectly adequate. Indeed substituting the linear kernel with nonlinear ones, the improvement in the classification accuracy has turned out to be of the order of 1%, which is not significant for the applications discussed in this paper. If the linear kernel is used, the SVM provides the following expression for the separation hyperplane:

$$a\, x_1 + b\, x_2 + ...k\, x_n + C = 0 \tag{12}$$

where the coefficients a...k are the results of the quadratic minimization implemented by the SVM together with the constant C. At this point if the logarithms of the signals are provided as inputs $x_1...x_n$, the equation of the separation hyperplane identified by the SVM can be represented again as a monomial exactly of the form of relation (2).

3.3 RESULTS USING THE INTERNATIONAL DATABASE

To maximise the generality of the results obtained with the methodology described in the previous sections, an international database has been considered [17]. This database was explicitly conceived

to support advanced studies of the H mode and includes validated signals from the vast majority of Tokamak machines ever operated in the world. The entries of the database, for which all the variables used by the most widely accepted theoretical models are available, have been considered. Of the resulting 400 L mode and 400 H mode examples, about 70% have been used for the training and 30% for the test of the NNs and SVMs. The derived mathematical formulas and their success rate, in discriminating the H from the L mode time slices, are compared to the performance of the theoretical models in the following table II. For each theoretical model, both NNs and SVMs have been trained with exactly the same signals used in the scaling formula to allow a simple and direct comparison of the final results. It is important to mention that, due to lack of experimental information, the ratio τ of the ion and electron temperature has been considered equal to one in the analysis and, as the database includes only deuterium plasmas, the plasma atomic mass A is always two. This explains why these two parameters are not considered as scaling quantities in the final formulas obtained by the NNs and the SVMs.

To interpret the results reported in table II some comments are in place. First of all, it must be noted that, contrary to the case of the equation obtained with the NNs and the SVMs, all the theoretical models provide only proportionality between the electron temperature and the other scaling quantities. Therefore, for all the theoretical models analysed, the multiplicative constant necessary for a quantitative comparison has been obtained empirically by maximising their success rate on the given database. In certain sense therefore also the theoretical models have been tuned to optimise their performance on the international database used. In any case, even with this ad hoc factor, the performance of the theoretical models is significantly lower than the data driven models obtained with the NNs and the SVMs. In general, the relations provided with the data driven methods have success rates of classification in the range between 89 and 95%.

Moreover, the performance and the equations provided by the NNs and the SVMs are very similar to each other, whereas they can differ significantly from the theoretical models (both in terms of success rates and mathematical formulation). It should also be mentioned that, for each physical quantity, the NNs and SVMs converge always on a very similar functional dependence, irrespective of the theoretical model they have been meant to benchmark (for example the exponent of the magnetic field B is almost always around 1.2). This fact, together with the observation that the success rates remain always in a very limited range (the variability of the performance is in a 5% range), seems to indicate that the data driven methods identify consistently the same physics. The mathematical relations with the highest success rate, almost 95%, is provided by the SVM (equation 13a) and NNs (equation 13b) and read:

$$T_{ec} = 4026 * \frac{B^{1.12}\, a^{0.90}}{q^{0.4}\, R^{0.61}\, n^{0.02}} \tag{13 a}$$

$$T_{ec} = 3030 * \frac{B^{1.07}\, a^{0.8}}{q^{0.34}\, R^{0.36}\, n^{0.02}} \tag{13 b}$$

Given the high success rate and the consistency with the estimate of the SVMs and NNs, this functional dependence must be considered a strong candidate for further theoretical interpretation. On the other hand, even if the global dependences are similar to the ones of some theoretical models, the differences are still significant and therefore no single theoretical model is really validated by the proposed procedure and more work is needed to interpret relation (13) in terms of basic plasma physics. A first data driven approach to undertake this task is described in the next section 4. For the moment it is worth emphasising that the convergence of the two methods, NNs and SVMs, is particularly relevant. They represent not only two independent numerical approaches but they are also conceptually very different. The NNs exploits complex nonlinear functions in the original space whereas the SVMs are linear separators but in a higher dimensional space where the problem is separable.

3.4 CONFIDENCE INTERVALS IN THE MATHEMATICAL FORMULAS

One of the limitations of the typical approach of theory formulation, based on derivation from first principles, is the difficulty to associate confidence intervals to the obtained formulas. It has therefore been decided to develop a methodology to derive the uncertainties for the data driven techniques just described. To achieve this goal, a specific procedure has been refined to take into account all the major sources of uncertainties, from the error bars in the measurements to the bias and the irreproducibility in the training.

The calculating process consists of choosing randomly the discharges in the training and test sets and then performing the training again with randomly chosen initial weights. If the misclassification rates for the training and the test set differ by less than 2%, the resulting network coefficients are recorded. This process is repeated to achieve astable statistics (in our case 25 different NNs have been enough). The condition imposed on the difference between the success rates of the training and test sets is a comprehensive one, which is meant to take into account all the major sources of uncertainties. The just described procedure is illustrated graphically in figure 4.

Once the coefficients of the various NNs have been determined, they can be used to calculate the mathematical expression for the temperature threshold as described in section 3.3 (relation 6 and 7). For the international database it turns out that the exponents of the four most important variables are quite stable and they do not change significantly. Their probability distribution functions are reported in figure 5, which shows how they closely resemble a Gaussian, as expected since most of the errors are random.

The statistical distributions obtained allow determining a confidence interval for the exponents. Again for the case of the international database, the resulting expression for the temperature threshold can be written in the form

$$T_{ec} \propto \frac{B^{0.12\pm0.0035}\, a^{0.75\pm0.005}}{q^{0.30\pm0.0035}\, R^{0.15\pm0.097}\, n^{0.01\pm0.004}} \tag{14}$$

With the proposed approach, it is therefore possible to take into account all the major causes of

uncertainty affecting the output of the networks. The data driven theory formulation devised can therefore associate specific confidence intervals to the final mathematical relations, a significant added value compared to the more traditional modelling methods.

4. IDENTIFICATION OF INTRINSIC DIMENSIONALITY, VARIABLE SELECTION AND VARIOUS TYPES OF L TO H TRANSITION

The procedure of theory formulation with machine learning techniques described in the previous sections has focused more on the aspects of model selection, deriving mathematical relations to be compared with available theories. In this perspective, the signals used have been the same as the ones of the theoretical models against which the NNs and SVMs formulas have been benchmarked. In this section machine learning methods are applied to the preliminary stage of determining the dimensionality and the most relevant signals for the problem at hand. This additional analysis stage would permit to carry out a complete data driven process of theory derivation, from the identification of the important signals to the final formulas with confidence intervals, without any underlying assumptions.

With regard to the problem of determining the dimensionality of a system, engineers, physicists and in general scientists are every day more confronted with the task of extracting information about poorly-known processes from data. Discerning the significant patterns in experimental signals, as a first step in process understanding, can be greatly facilitated by reducing dimensionality. Dimensionality reduction is a method of obtaining information from a high dimensional feature space using fewer intrinsic dimensions and is very beneficial for efficient classification, regression, presentation and visualization. By representing data in the lower dimensional space, most of the time not much information that matters to the application is lost. Indeed the superficial dimensionality of the available signals, the number of individual observations constituting one measurement vector, is often much greater than the intrinsic dimensionality, the number of independent variables producing significant non-random changes in the physical phenomena under study. In the case of the confinement regime, it is worth mentioning that the H mode is routinely reached in most of the operational devices in the world and therefore, at least in principle, the amount of information available to study this phenomenon is enormous, since in some devices like JET even thousands of signals are acquired for each discharge. Therefore there is lot of scope and material to reduce the dimensionality of the signal input space.

The reduction of a data set from its superficial to intrinsic dimensions and the identification of the most relevant signals is therefore the initial task to be undertaken when starting from scratch the data driven study of the transition to the H mode. An important aspect to remember in our case is that the general scope of the present analysis consists of deriving mathematical formulas for theoretical analysis. Therefore the signal reduction process should converge on the lists of the most relevant physical signals and not on combinations of them, which can have high numerical power but loose the physical meaning.

With these considerations in mind, the first step has been performed with the CART approach.

CART is a supervised, non-parametric technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively [18]. To build the tree, the CART algorithms typically traverse the entire dabase to determine which variable better allows dividing it in two subclasses, in our case one with only L and one with only H mode time slices. Since no variable has such a exploratoy power to permit a full separation between the two confinement regimes, the algorithms select the most powerful and use it to split the database into two child branches. The process is then reperated iteratively for the two subsets until pure nodes are reached. In this context, a node is considered pure when it contain only examples, time slices of signals in our case, beloging to the same class.

One of CART most important features for our purposes is the evaluation of the importance of the different explanatory variables, i.e. the variable provided as input during the building of the tree, to describe the output through the so-called "variable ranking method". To estimate the importance of a variable, the sum across all nodes in the tree of the improvement scores, that the variable induces when it acts as a primary splitter, is performed [18]. The importance values so produced allow ranking the different input signals from high to low importance. In this way, CART can be used for feature selection, being able to identify the most important variables to describe the output.

The CART approach has been applied to a set of 42 JET discharges, analysed by the experts, who have agreed on the time points in which the transitions take place. The available discharges cover the shot range between 55211 (21/03/2002) and 62723 (28/01/2004) and they therefore refer mainly to JET divertor configuration with the Septum. Contrary to the information provided in the international database, in this case the entire evolution of the discharges is available with a time resolution of 10 ms (and does not contain only time points representative of well established confinement regimes). Therefore more detailed studies and careful analysis of the plasma evolution around the transition are possible. This validated JET database, whose main characteristics are described in detail in [19], has been divided in two sets, one for the training and one for the final test. The composition of these two sets has been varied randomly to achieve a sounder statistical basis for the derived conclusions. The results reported in the following are the average obtained from these random sets

For the discharges in this specific database, thousands of signals have been acquired and are stored in JET general database. In order to reduce the original dimensionality of the problem, the first step has been the selection of a subset of signals (about 40), which the experts consider as possibly linked to the confinement regime transition, given the basic understanding of the physics involved. This step has been undertaken manually to avoid the well known problem that, including an excessive number of variables in the automatic exploratory phase, can result in the algorithms detecting a high number of spurious correlations, which can be due to casual fluctuations in the data and are not relevant to the phenomenon to be analysed. After application of the CART algorithm, the eight variables with more than 90% of the explanatory power are summarised in table III.
It is worth mentioning that the CART technique, using a brute force approach, is a completely non

linear and unbiased method. The entire database of the original signals is completely traversed and no linearization or even normalization of the data is needed. On the other hand, the main limitations of this form of nonlinear correlation reside in the sequential character of the operations and in the lack of information about the mutual correlations between the resulting most important variables. Therefore, in order to overcome these drawbacks, a further selection process has been undertaken to further reduce the dimensionality of the problem. The first step consists of PCA analysis used as a linear clustering technique [16] i.e. as a method to group the signals in classes with the same characteristics so that only a subset of them can be chosen without major loss of information. PCA is originally a method for mapping multidimensional data into lower dimensions with minimal loss of information. It is mathematically defined as an orthogonal linear transformation, which projects the data to a new coordinate system such that the greatest variance of any projection of the data coincides with the first coordinate (called the first principal component (PC)), the second greatest variance with the second coordinate, and so on. PCA is theoretically the optimum linear transform for a given data in the least square sense. In the present application, the signals identified with CART have been divided into two main groups: the first one includes the quantities which change significantly during the same shot and are therefore called dynamic (≤n, Wmhd, Te, n_e), the second comprise the parameters which are almost constant during a discharge and are therefore identified as static (B_{T80}, q_{80}, RXPL, ZXPL). This distinction reduces significantly the computational resources required for the PCA analysis. The principal components of the two groups of variables are reported in figure 6 and 7. Bin charts represent the principal components and their cumulative energy is defined as the percentage of total variance in the data which is accounted for. It is calculated by summing the percentage of the total variance described by each component. Among the static signals, we have selected the three first variables B_{T80}, q_{80}, ZXPL, which have the higher projections onto the principle components (the so called principal components loadings [16]) and represent more than 90% of the variance in the original data. The principal components loadings are shown in the two dimensional plots on the right of figure 6 and 7. With regard to the dynamic variables, given the very high correlation between Wmhd, bn and n_e, only the first (because Wmhd presents the highest correlation loading) has been retained together with Te. On the basis of the results of the PCA analysis we are able to determine that the vast majority of the information in the signals is accounted for by the following 5 quantities: Wmhd, Te, B_{T80},q_{80}, ZXPL.

In our context the main limitation of PCA is its linear character, which limits its application to systems with significant nonlinear aspects, like thermonuclear plasmas. The five signals identified by PCA as the most relevant have been therefore analysed with a non linear principle component method (NLPCA). NLPC is used to identify correlations among problem variables as an aid to dimensionality reduction, visualization, and exploratory data analysis without any restriction on the degree of nonlinearity of the correlation in the data. In our application NLPCA has been implemented by training an Autoassociative Neural Network (ANN) to perform an identity mapping, the process whereby the network inputs are reproduced at the output [20]. The network contains an internal “bottleneck” layer

(containing fewer nodes than input or output layers), which forces the network to develop a compact representation of the input data, and two additional hidden layers, as shown in figure 8. The ability of these neural networks to detect non linear correlations depends on the presence of hidden layers with nonlinear nodes, typically sigmoidals of the type $\sigma(x) = 1/(1 + e^{-x})$.

Note that to achieve the universal fitting property, exactly two layers of sigmoidal nodes and three layers of weighted connections are required. In practice, sigmoid are often used also in the nodes of the output layer so that the network produces outputs in a fixed, finite range. Moreover, the sigmoid function can be scaled multiplicatively or translated without affecting the generality of the network. Since we frequently deal with mean-centred data sets, a sigmoid function rescaled into the range $(-1, 1)$ was used in this work ($\tan sig\,(x) = \frac{2}{1+e^{-2x}} - 1$).

In the approach adopted in this paper, the number of nodes in the bottleneck layer is reduced until further suppression of a neuron makes the ANN incapable of properly mapping the inputs on the outputs. At that point the number of neurons in the bottleneck layer is expected to represent the intrinsic dimensionality of the space. In more detail, the analysis has been started using a network with as many neurons in the bottleneck layer as input variables. Then, the number of neurons in the central layer has been reduced and the quality of the identity mapping of the network has been checked. Indeed, when removing an additional neuron in the central layer causes a significant increase in the errors, the degradation in performance indicates that all the neurons are essential to properly model the phenomenon. Figure 9 reports the results of this technique for the case of the five variables previously identified. The very drastic increase in the error rate after reducing the neurons in the bottleneck layer to two indicates clearly that the intrinsic dimensionality of the problem is three.

At this point, having identified the dimensionality of the problem, a strategy is required to finally select the subset of the most significant variables. In this respect, a new criterion, which is an original development working quite well for the present problem, consists of analysing the weights between the inputs and the neurons in the bottleneck layer. Indeed, during training, bias and weights are updated to project the data into the low-dimensional feature space and to minimize the error in the outputs. Therefore the whole information about the compression of data is contained in the weights, which are force to assume values suited to a compact representation of the inputs. The most relevant variables are therefore expected to be the ones with the highest sum of the weights between themselves and all the neurons in the bottleneck layer. To illustrate this aspect, table IV describes the coefficients obtained.

According to the criterion of the maximum sum of the weights, the variables Te, Wmhd and B_{T80} are the ones selected in this case. The entire methodology developed to determine the dimensionality of the problem and then to select the signals with the highest information content is summarised in figure 10.

The appropriate character of this choice has been confirmed with a brute force approach. A series of NNs have been trained with all the possible combination of three out of the five PCA variables to classify the time slices of JET databases in the two classes of H and L mode. The best results have

been obtained consistently by the network using exactly the three variables Te, Wmhd and B_{T80} identified with the method of the sum of the weights. Also the fact that the discarded variables have much lower information content has been confirmed again by a brute force approach. It has indeed been checked that NNs using also these signals do not present significantly better success rates than the ones using only the three most significant signals. The computation of the weights sum to identify the most relevant inputs is expected to be a procedure of general validity. In any case, the results for the JET database used in this paper are reported in table V. A series of NNs and SVMs have also been trained and tested on exactly the same JET database but using the variables of the various theoretical models summarised in table I. From table V it is clear that, even using less variables, the classification of the machine learning methods is much better when the signals identified by the data driven selection procedure are chosen. It is also important to emphasise one more time the very similar performance obtained by the NNs and the SVMs.

4.2 DETAILED ANALYSIS OF THE TRANSITION

JET database, providing signals with much higher temporal resolution, allows performing more detailed studies. In particular, the plasma evolution close to the transition can be analysed more specifically to try to get further insight into the dynamics of the H mode onset. To this end, the time slices analysed belong to the interval between plus and minus 300 ms around the time of the L to H transition (instead of the entire discharge as in the present section). At first æ of the discharges in the entire database have been used for training and the remaining º as test set. For easier interpretation of the physics, the signals T_e, n_e, B_{T80} and q_{80} have been used as inputs, in analogy with the theoretical models. In any case, it has been double-checked that, for this specific database and analysis, substituting Wmhd with n_e does not change significant the success rate of neither NNs nor SVMs and therefore the new set of variables is practically equivalent to the one of table V.

At the beginning, when trained with the entire database, both NNs and SVMs have presented a failure rate of around 15%. On the other hand a closer analysis of the results shows that the error rate changes significantly depending on the shot. Typically both NNs and SVMs have a very high success rate for about 75% of the discharges and a very poor one for the remaining 25%. Therefore it has been decided to develop a different classifiers one for the subset of discharges with a seemingly anomalous behaviour. With the first classifier, applied only to 75% of the database, an average failure rate of 4% has been obtained and the temperature threshold can be expressed by the relation:

$$T_{ec} \approx 575 \; \frac{B_{T80}^{0.8} \; a_{80}^{0.06}}{n_e^{0.01}} \tag{15a}$$

A very similar equation has been derived independently using the SVM numerical method:

$$T_{ec} \approx 3790 \; \frac{B_T^{1.026}}{n_e^{0.051} \; q_{80}^{0.052}} \tag{15b}$$

Using only the remaining discharges (50% of which used for the test), misclassified by the previous equation, a new specific classifier presents an average failure rate of 3.9% and provides the following equation:

$$T_{ec} \approx 1200 \frac{q_{80}^{3.95}}{n_e^{0.06} B_{T80}^{4.2}} \tag{16a}$$

Again a very similar result has been obtained independently using the SVM numerical method:

$$T_{ec} \approx \frac{q_{80}^{2.74}}{n_e^{1.106} . B_{T80}^{3.936}} . 7,81.10^{24} \tag{16b}$$

The previous analysis supports two conclusions. First of all the distinction between the two formulas seems to have a quite robust physical interpretation in terms of different plasma collisionality. Indeed, in figure 11 it is reported how the two groups of discharges, corresponding to the two different relations, can be perfectly discriminated by their values of collisionality. The data driven methods developed have allowed identifying two different types of transition form the L to the H mode in the discharges with the septum, a phenomenon which had escaped the analysis before. On the other hand, the results of the SVMs are not always very close to the ones of the NNs (see relation 16b). This is due to the limited number of discharges (only around 10) which are available at higher collisionality. In any case, the potential of the proposed methodology for data mining is to be considered a strong point in support of the proposed approach and will be further pursued in the near future.

The second important aspect is that the two classifiers can be considered to be almost 100 % accurate on their respective subsets of discharges once the confidence intervals in the measurements is taken into account, as discussed in more detail in the next section.

5. SUCCESS RATE OF THE CLASSIFICATION TAKING IT ACCOUNT THE CONFIDENCE INTERVALS IN THE MEASUREMENTS

In this section a detailed analysis of the confidence intervals to be associated to the formulas derived with the methods described in the previous sections is performed. The treatment is particularized for JET database, because it is the one which contains detailed information and high time resolution and therefore allows better verification of the results. The discussion is also mainly exemplified by the NNs results but similar conclusions can be derived using SVMs. Indeed the approaches introduced in this section can be considered of quite general validity.

Tokamak plasmas are very difficult physical objects to measure. In JET a systematic and continuous activity is in place to reduce the uncertainties in the diagnostics. Nonetheless even the most relevant measurements are affected by experimental errors, which have to be properly taken into account when attempting to draw conclusions about the physics. Therefore, to confirm the quality of the results in terms of classification accuracy, a specific effort has been devoted to the analysis of the discharges for which, even using the best variables, the NNs or the SVMs do not

perform a perfect classification, i.e. they do not coincide perfectly with the expert estimates. To this end, it has been decided to investigate whether, between the time slice of the transition as identified by the experts and the time slice identified by the data driven methods, there is a change in at least one of the signals outside the error bars. Only in this case there would be really evidence of a discrepancy between the two estimates. The idea is represented in figure 12.

For JET database, realistic uncertainties in the plasma parameters of interest, as obtained by careful analysis of the measuring instrumentation, are reported in table VI. A systematic analysis, using these values for the measurement uncertainties and performed on the two classes of NNs used to classify the discharges of JET database, allows drawing some very sound conclusions. First of all, the discrepancies between the NN estimates and the expert assessments occur typically very close to the time of the transition between the two modes of confinement. Moreover when the NNs make not perfectly correct classifications around the transition (about 4% of the cases overall), this is normally due to the fact that the signature in the data is not strong enough and the variation in the signals around the transition time is not outside the error bars. Typically the difference in the signals, between the time slice identified by the experts and the one estimated by the NNs, are typically well within the error bars. Therefore it can be said that the NNs transition times agree practically 100% with the expert assessment once the uncertainties in the measurements are properly taken into account. Since in general the SVMs provide results very similar to the NNs, this observation is valid for both data driven methods.

From completeness, it has also been decided to develop an objective criterion to associate a confidence interval to the transition times derived by the NNs. In order to assess how the error bars affect the estimates of the NNs, their output has been calculated not only for the nominal value of the measurements but also for the two extreme possible values, taking into account the experimental uncertainties. All the 12 combinations of the four inputs (T_e, n_e, B_{T80} and q_{80}) have been considered and the corresponding times of the output threshold crossing between the L and H mode have been calculated. The time interval between the earliest and latest instants, in which the NNs outputs cross the output threshold of 0.5, can be considered as the uncertainty to be associated to the NNs estimates of the transition time between the L and the H mode of confinement. An example is illustrated graphically in figure 13, where the blue line corresponds to the output using nominal values of the input variables. The two azure lines represent the earliest and latest onset of the H mode, using the extreme values of the input variables inside their error bars. The red lines show the uncertainty interval in the time of the onset of the H mode.

CONCLUSIONS AND FURTHER DEVELOPMENTS

A series of techniques has been developed to study complex, nonlinear phenomena, which require extracting information from massive databases. Two completely independent machine learning tools, NNs and SVMs, have been used for data driven formulation of equations describing the essence of the physics in the database. The analysis has been focused on the issue of determining

the scaling laws of the temperature threshold between the L and H mode of confinement in reactor relevant Tokamak plasmas but the methodology is absolutely general an can be deployed in any scientific domain. In the specific case of the confinement regime identification, the results of these two tools agree very well with each other and the resulting mathematical formulations interpret much better than the more widely accepted theoretical models both an international database and a specific set of JET discharges. To improve the theoretical understanding of the transition physics, a careful determination of the problem dimensionality and the identification of the most relevant quantities to take into consideration have been performed. The signals selected in this way can be used to derive theories more based on the experimental data. The developed techniques have also been applie to the detailed study of the plasma evolution around the time of the transition to the H-mode. They have allowed detecting two different types of plasma dynamics around the transition, whose detailed explanation requires further analysis. A general procedure has also been devised to take into account all the major sources of uncertainty and to attach confidence intervals to the final relations derived for the scaling law of the threshold in the electron temperature.

It is worth pointing out that the proposed methodologies are not only useful to analyse large amount of data but they also provide an objective, user independent procedure to achieve sound data driven results. The only phase which remains open to the intervention of the user is the selection of the database, an aspect which is outside the scope of the present paper. In any case techniques are under investigation to develop specific databases automatically, with minimum intervention from the user. Another area of future research is the application of the approach to more complex an more nonlinear physical phenomena, for which the simple network architecture of figure 3 and the linear kernel for the SVM are no more adequate. This will certainly require significant upgrades of the techniques and possibly the application of other machine learning methods. Another interesting aspect, which would require further attention, is an integrated treatment of the experimental errors at the level of the original NNs and SVMs used to derive the mathematical formulas. In the case of NNs, Bayesian networks seem to be an adequate tool, whereas for SVM recourse to hybrid methods, combining SVM with the Parzen window [21] are expected to provide satisfactory results.

ACKNOWLEDGEMENTS:

This work was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES

[1]. http://www2.sims.berkeley.edu/research/projects/how-much-info/

[2]. Crowe C.T, Elger F. and J.A. Roberson "*Engineering Fluid Dynamics*" John Wiley & Sons. 2001

[3]. Barenblatt G.I. "*Scaling*" Cambridge University Press 2003

[4]. Wesson J., *Tokamaks*. Oxford: Clarendon Press Oxford, 2004. Third edition.
[5]. Wagner, F. et al., Physical Review Letters **49**, 1408 (1982).
[6]. Connor J.W, Wilson H R 2000 Plas. Phys. Cont. Fus. **42** R1
[7]. Chankin A.V 1997 Plas. Phys. Cont. Fus. **39** 1059
[8]. Chankin A.V, Saibene G 1999 Plas. Phys. Cont. Fus. **41** 913
[9]. Kerner W. et al. 1998 Contrib. Plas. Phys. **38** 118
[10]. Rogister A.L 1994 Plas. Phys. Cont. Fus. **36** A219
[11]. Scott B. et al. 1996 Fusion Energy (Proc. 16th Int. Conf., Montreal, 1996) vol **2** (Vienna: IAEA) 649
[12]. Shaing K.C, Crume Jr. E C 1989 Phys. Rev. Lett. **63** 2369
[13]. Mandic D.P. and Chambers J.A. "*Recurrent Neural Networks for Prediction*" 2001 John Wiley and Sons Ltd
[14]. MacKay D.J.C., "*A Practical Bayesian Framework for Backprop Networks*" Computation and Neural Systems California Institute of Technology 139 **74**, Pasadena CA 91125 mackay@hope.caltech.edu
[15]. Vapnik V. "The Nature of Statistical Learning Theory". Second edition. Springer. (1999).
[16]. Cherkassky V., F. Mulier. "Learning from data". John Wiley & Sons, Inc. (1998)
[17]. http://efdasql.ipp.mpg.de/HmodePublic/
[18]. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1984 *Classification and Regression Trees* (Belmont, CA:Wadsworth Inc.) (1993, New York: Chapman and Hall)
[19]. Meakins, A.J. (2008) "*A Study of the L-H Transition in Tokamak Fusion Experiments*". PhD Thesis. Imperial College London.
[20]. M.A.Kramer, AIChE.- Massachusetts Institute of Technology, Cambridge: Laboratory for Intelligent Systems in Process Engineering, Dept. of Chemical Engineering, 1991.2: Vol. **37**. p233-243.
[21]. Vega J. , Murari A., VagliasindiG., Rattá G. A. "*Automated estimation of L/H transition times at JET by combining Bayesian statistics and Support Vector Machines*" to be submitted to Nuclear Fusion

Chankin	$T_{ec} \propto \left(\frac{\tau}{1+\tau}\right)^{\frac{2}{16}} \frac{(Rq)^{\frac{1}{2}} B^{\frac{1}{2}}}{A^{\frac{1}{16}}}$
Kernel et al. (collisinal) [9]	$T_{ec} \propto \frac{n^{-\frac{2}{15}} B^2}{(Rq)^{\frac{4}{5}} A^{\frac{2}{5}}}$
Kernel et al. (non collisinal) [9]	$T_{ec} \propto \frac{(Rq)^{\frac{2}{15}} n^{\frac{1}{15}} B^{\frac{2}{5}}}{A^{\frac{1}{15}}}$
Rogister et al. [10]	$T_{ec} \propto \frac{\tau A q R n^2}{a B^2}$
Scott et al. [11]	$T_{ec} \propto 0.135 \frac{\tau n^{-1} B^2}{(1+\tau) A}$
Shaing and Ceume [12]	$T_{ec} \propto \frac{\tau R^{\frac{5}{4}} (qn)^{\frac{1}{2}}}{A^{\frac{2}{4}}}$

Table I: The expression of the temperature threshold to access the H mode of confinement as provided by theoretical models using only widely available plasma quantities. T_e is the electron temperature, n_e the electron density, B the toroidal field, q the safety facto, R the major radius, a the minor radius, A the plasma atomic mass and τ the ratio of the ion and electron temperature.

	Theoretical models	Success rates	NN	Success rates	SVM	Success rates
Chankin [9,10]	$T_{ec} \propto (\frac{\tau}{1+\tau})^{3/16} \frac{(Rq)^{1/8} B^{1/2}}{A^{1/16}}$	87.9%	$T_{ec} = 298 * \frac{B^{1,2} R^{0,95}}{q^{0,35}}$	93.7%	$T_{ec} = 311 * \frac{B^{1,17} R^{0,86}}{q^{0,28}}$	93.1%
Kernel et al. (Collisionless) [11]	$T_{ec} \propto \frac{B^2}{(Rq)^{4/5} A^{3/5} n^{7/5}}$	91.3%	$T_{ec} = 315 * \frac{B^{1,19} R^{0,92}}{q^{0,36} n^{0,02}}$	93.5%	$T_{ec} = 821 * \frac{B^{1,14} R^{0,89}}{q^{0,36} n^{0,02}}$	93.4%
Kernel et al. (Collisional) [11]	$T_{ec} \propto \frac{(Rq)^{2/15} n^{1/15} B^{2/5}}{A^{1/15}}$	83.6%				
Rogister et al. [12]	$T_{ec} \propto \frac{AqRn^2}{aB^2}$	50%	$T_{ec} = 3030 * \frac{B^{1,07} a^{0,8}}{q^{0,34} R^{0,36} n^{0,02}}$	94.5%	$T_{ec} = 4026 * \frac{B^{1,12} a^{0,90}}{q^{0,4} R^{0,61} n^{0,02}}$	94.7%
Scott et al. [13]	$T_{ec} \propto \frac{\tau B^2}{nA(1+\tau)}$	90.5 %	$T_{ec} = 784 * \frac{B^{1,86}}{n^{0,02}}$	91.2%	$T_{ec} = 1090 * \frac{B^{1,67}}{n^{0,02}}$	91%
Shaing & Crume [14]	$T_{ec} \propto \frac{\tau R^{5/4} (qn)^{1/2}}{a^{3/4}}$	63.3%	$T_{ec} = 16300 * \frac{a^{1,58}}{q^{0,1} R^{1,57} n^{0,01}}$	89.2%	$T_{ec} = 11854 * \frac{a^{1,28}}{q^{0,19} R^{0,9} n^{0,17}}$	89.3%

Table II: Mathematical relations provided by the theoretical models, the NN and SVM classifiers using the same variables The success rates are also reported. T_e is the electron temperature, n_e the electron density, B the toroidal field, q the safety facto, R the major radius,a the minor radius, A the plasma atomic mass and τ the ratio of the ion and electron temperature.

Short name	Description
Bndiam	Beta normalised with respect to the diamagnetic energy
n_e	Outer interferometry channel
Wmhd	Magnetohydrodynamic energy
T_e	Electron temperature at the intersection between interferometry vertical view and magnetic surface ($\psi = 0.8$)
RXPL	X point horizontal coordinate
ZXPL	X point vertical coordinate
q_{80}	Safety factor on magnetic surface ($\psi = 0.8$)
B_{T80}	Axial toroidal Magnetic Field at a magnetic surface ($\psi = 0.8$)

Table III: The most relevant variables for the study of the L to H transition as provided by the CART method.

	Coefficients NLPCA				
Neuron	**W_{mhd}**	**T_e**	**Q80**	**ZXPL**	**B_{T80}**
A	1	0,85	0,4	0,15	0,45
B	1	0,85	0,35	0,15	0,45
C	1	1	0,4	0,2	0,55
Sum over all neurons	3	2.7	1.15	0.5	1.45

Table IV: Coefficients for the extraction of the most relevant variables: they are the sum of the weights from the inputs to all the three nodes in the bottleneck layer. Variables with the most important weights are the ones containing the most information after compression.

	Best theoretical model	NN using variables from the best theoretical model	NN using the three most relevant variables
Failure rate	18%	16.1%	5.2%
		SVM using variables from the best theoretical model	**SVM using the three most relevant variables**
Failure rate	18%	14%	5.9%

Table V: Error rates of the best theoretical model, NN and SVM using variables from theoretical model and NN using most relevant variables.

T_e	B_{T80}	q_{80}	n_e
±10%	±2%	±5%	±10%

Table VI: Error bars of the signals used as inputs to the NNs.

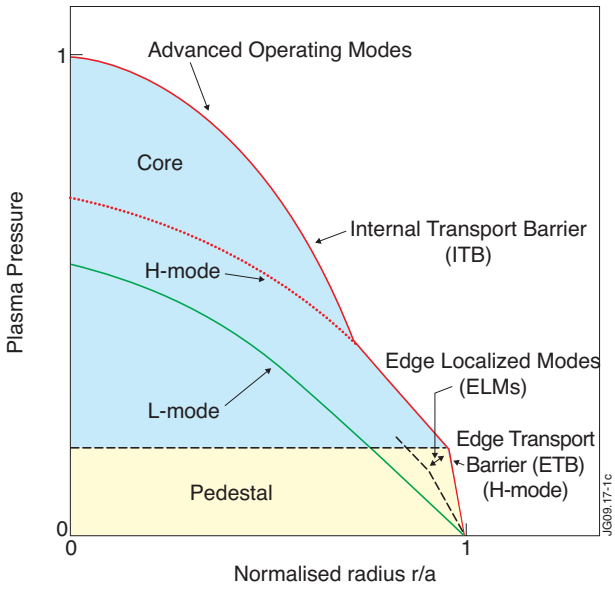

Figure 1. An illustration of the typical plasma pressures profiles obtained in the L-mode, H-mode and Internal Transport Barrier (ITB) confinement regimes.

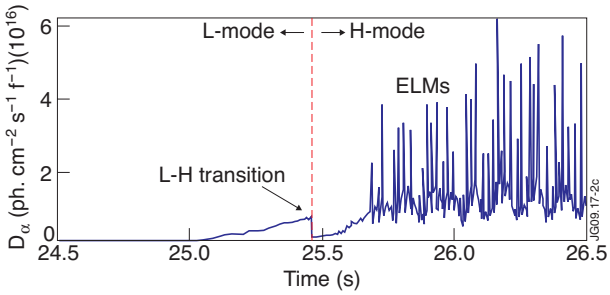

Figure 2: A deuterium line, Dα, signal from JET which demonstrates the sharp drop in the particle flux observed at the onset of the H-mode. The sharp increase in the Dα emission later on are the results of ELMs instabilities, a clear indication of a consolidated H-mode regime.

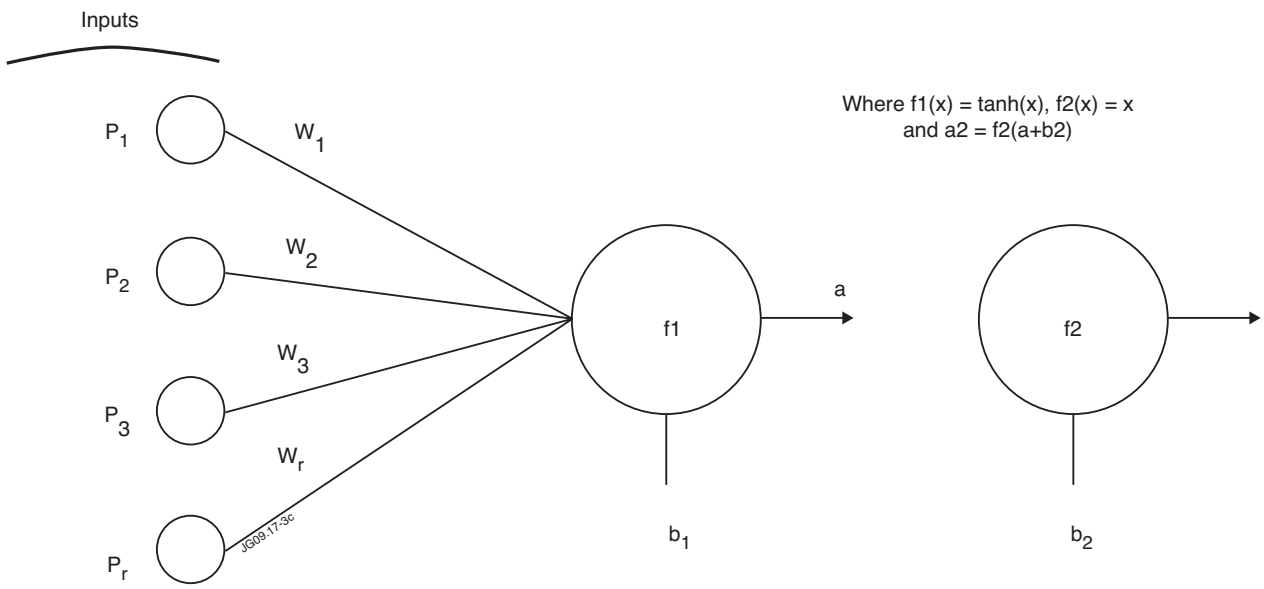

Figure 3: Architecture of the NNs used to obtain the results presented in this paper.

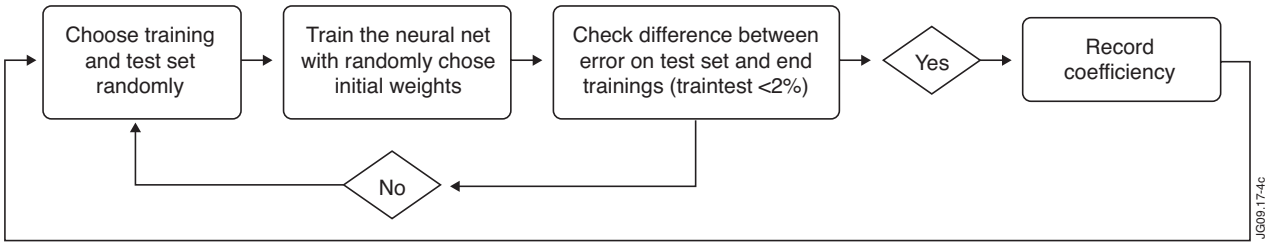

Figure 4: Flowchart of the procedure followed to determine the confidence intervals in the equations produced by the NNs.

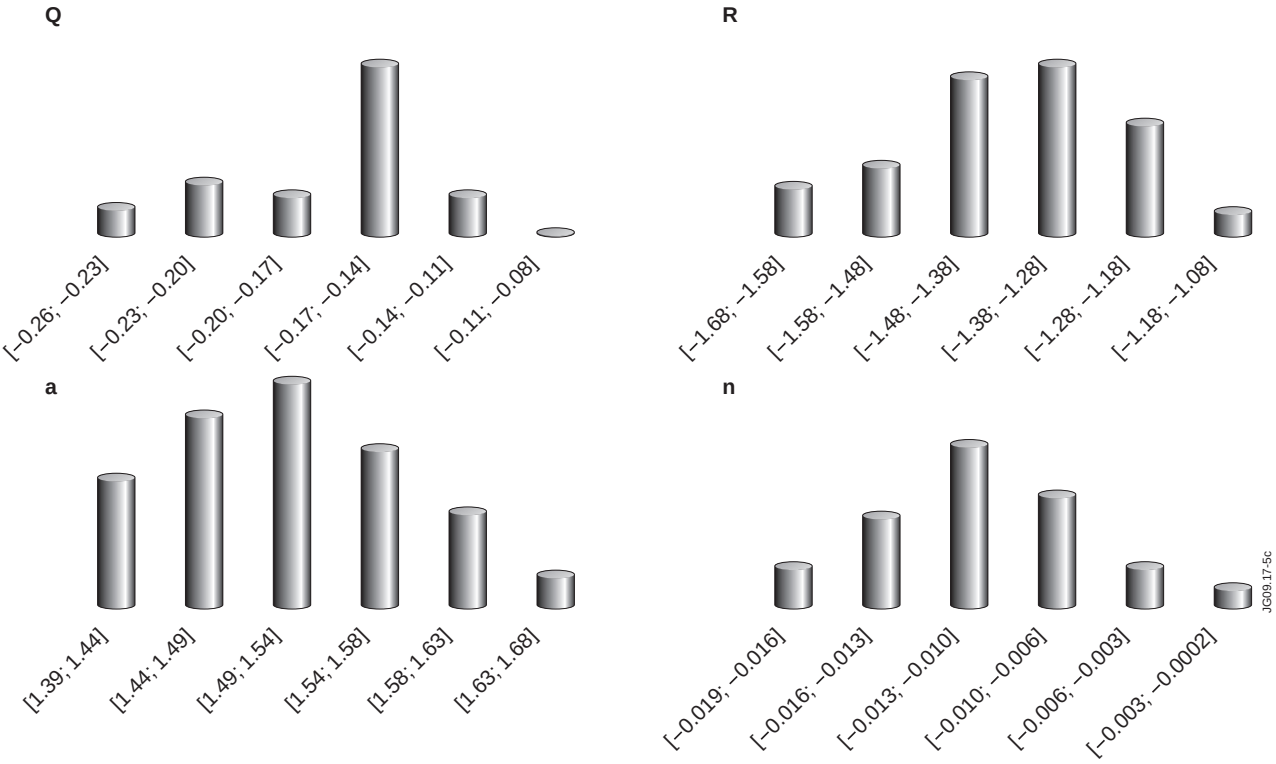

Figure 5: Bin charts representing the distribution of the exponents for the usual four variables used to discriminate the L from to H mode in the international database. All coefficients obtained from the 25 different training sets are plotted.

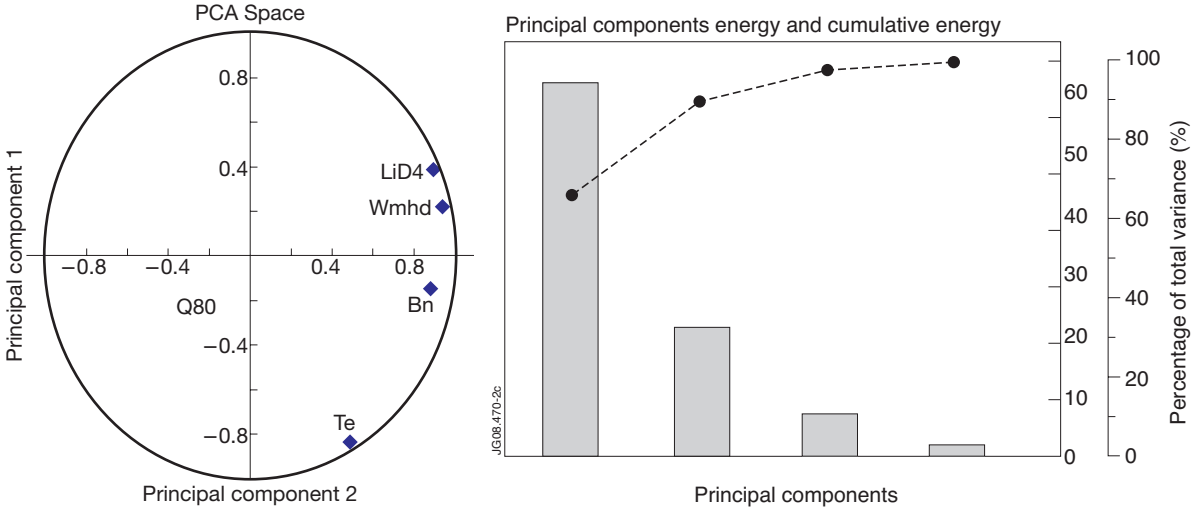

Figure 6: Correlation between the dynamic signals β_N, Wmhd, T_e and n_e. The bin chart gives the energy of the principal components and the red line shows the cumulative energy in terms of percentage of the total variance in the data. The plot on the right illustrates the degree of correlation of the variables with respect to the principal components (one and two in the figure), the so called principal components loadings

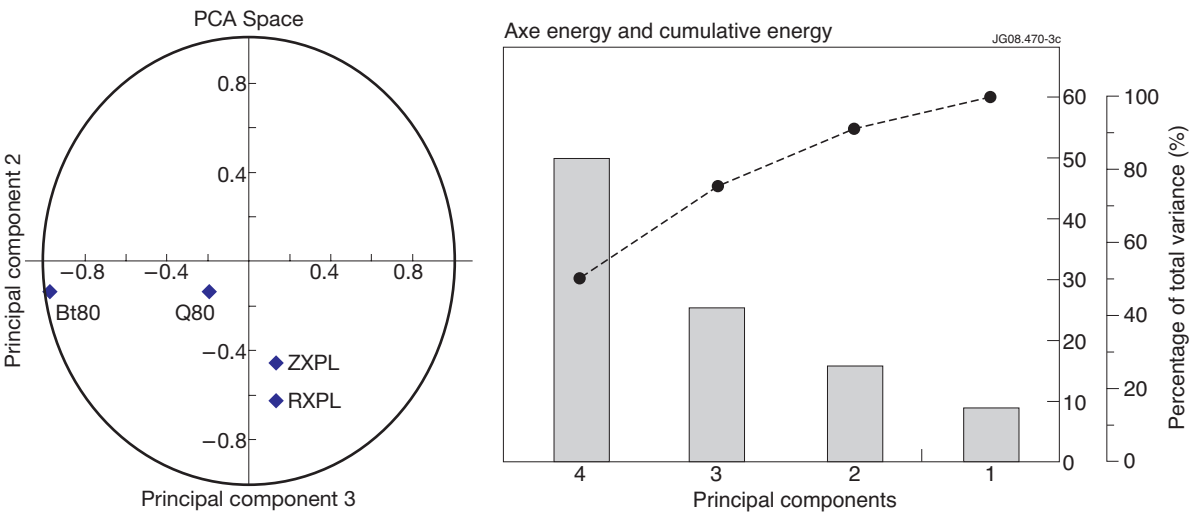

Figure 7: Correlation between the static signals B_{T80}, q_{80}, RXPL, and ZXPL. The top bin chart gives the energy of the principal components and the red line shows the cumulative energy in terms of percentage of the total variance in data. The plot on the right illustrates the degree of correlation of the variables with respect to the principal components (two and three in the figure), the so called principal components loadings.

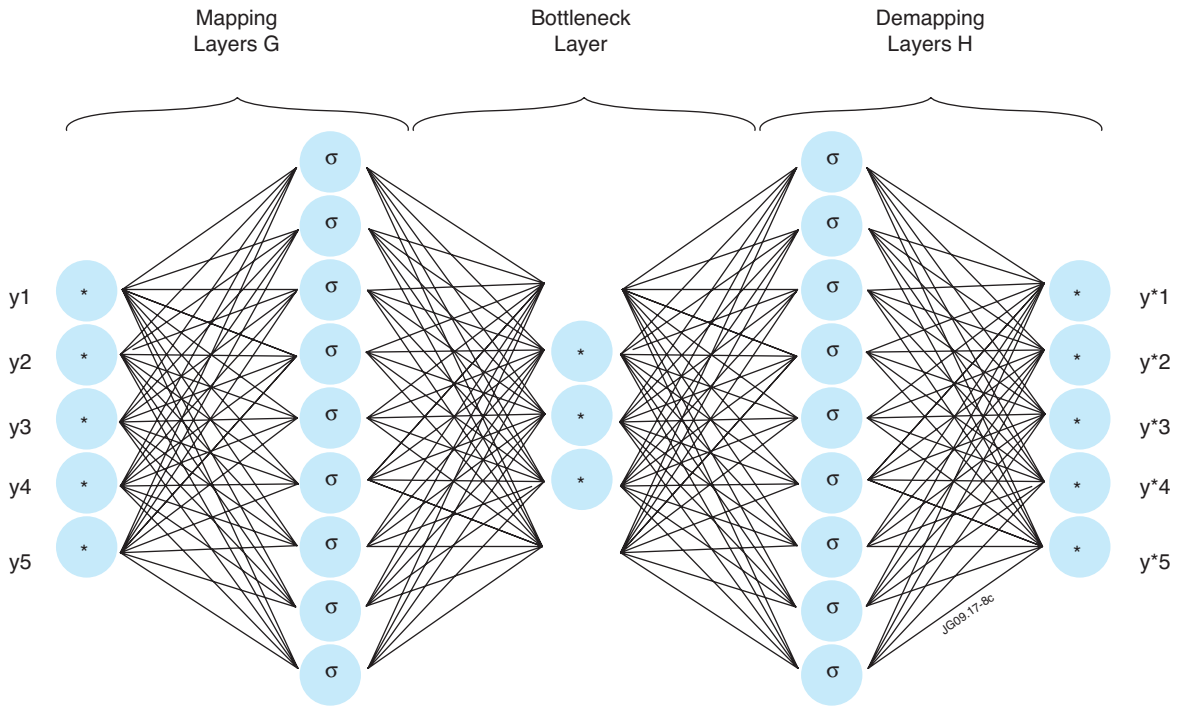

*Figure 8: The topology of Autoassociative Neural Networks The symbol σ indicates the non linear, sigmoid layers; the symbol * indicates the linear mapping layers.*

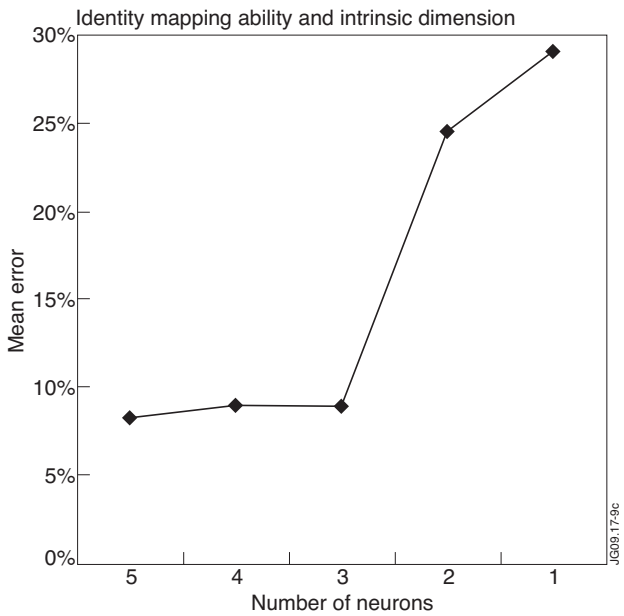

Figure 9: The increase in the error rate with the number of neurons in the bottleneck layer.

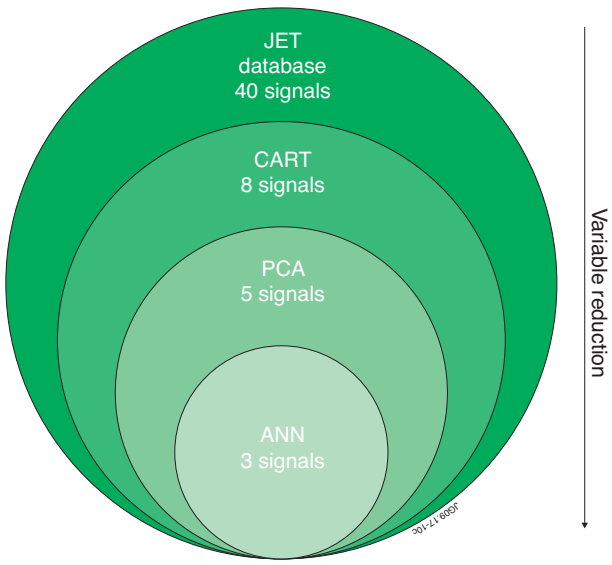

Figure 10: Summary of the methodology adopted to determine the dimensionality of the problem and to select the most relevant variables.

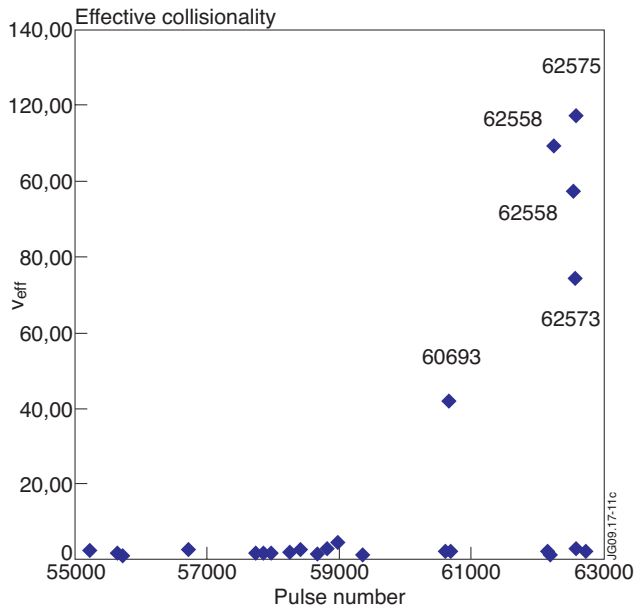

Figure 11: Effective collisionality for the different discharges in th test set. The five shots with collisionality higher than 20 are the ones well classified with the second formulation (relations 8).

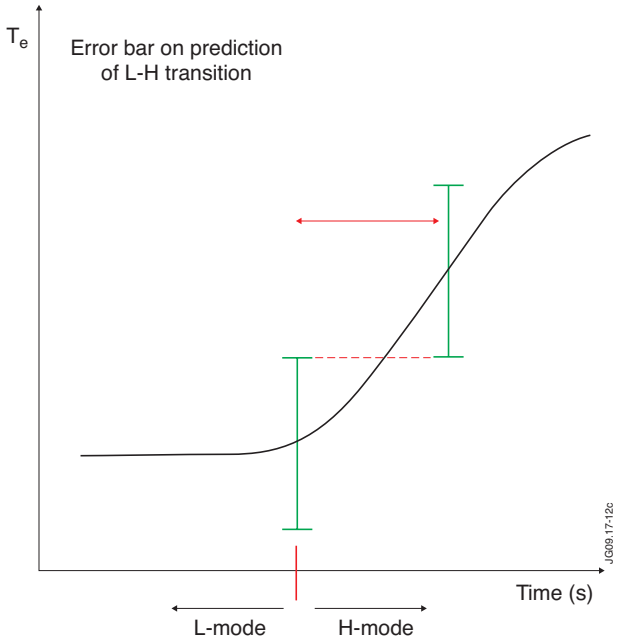

Figure 12: The plot illustrates in a simplified way how at least one of the four signals considered should vary outside the error bars between the expert and the NNs predictions to really have a discrepancy.

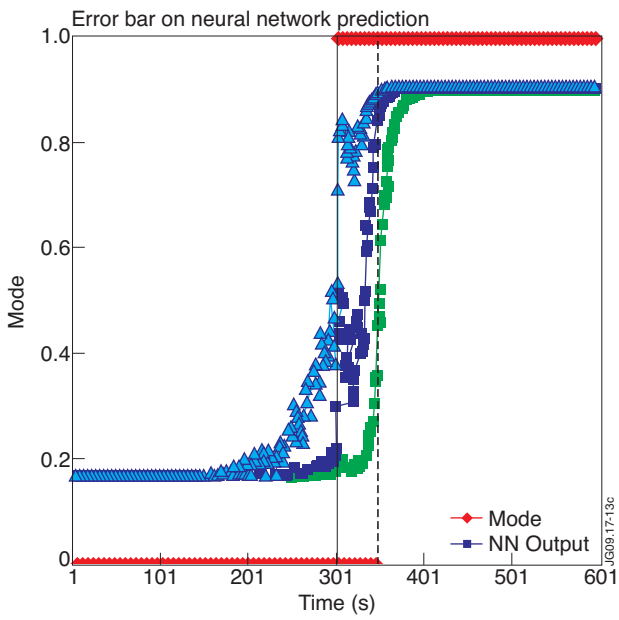

Figure 13: Propagation of error bars of input variable to prediction of the H mode on Pulse No: 56712. The red line corresponds to mode L or H (0 or 1) with the error bar associated (dotted line). The blue line corresponds to the nominal output of the network and the azure lines correpond to the earliest and latest onset of the H mode using variable values moved in their error bars. The y axis represents the confinement mode: H above the threshold of 0.5 and L below.