INSTITUTE OF PHYSICS PUBLISHING PLASMA PHYSICS AND CONTROLLED FUSION

Plasma Phys. Control. Fusion **43** (2001) 389–403 www.iop.org/Journals/pm PII: S0741-3335(01)16537-1

Direct measurement of JET local deuteron densities by neural network modelling of Balmer alpha beam emission spectra

J Svensson[1,2], M von Hellermann[3] and R W T König[4]

[1] Department of Atomic and Molecular Physics, Royal Institute of Technology (KTH), 100 44 Stockholm, Sweden
[2] UKAEA/Euratom Fusion Association, Culham Science Centre, Abingdon, OX14 3EA, UK
[3] FOM Institute for Plasma Physics, Rijnhuizen, 3430BE Nieuwegein, The Netherlands
[4] Max Planck Institute for Plasma Physics, Boltzmannstraße 2, D-85748 Garching, Germany

Received 22 August 2000

Abstract

The analysis of Balmer alpha beam emission spectra represents a challenging task, both in terms of the wealth of information hidden in them, and in terms of complex spectral features. The shape of a spectrum depends on many local plasma parameters, such as magnetic field strength and direction, beam density, effective charge and deuteron density. This paper is concerned with the deduction of local plasma deuteron densities from Balmer alpha emission from plasma atoms following charge exchange with the beam atoms. The model we will use is statistical, and is based on multi-layer perceptron neural networks (Hertz *et al* 1991, Bishop 1995). The use of neural networks makes the deconvolution task fully automatic and fast enough for real-time calculation of complete deuteron density profiles. It is shown that the spectra themselves and local electron densities are the only data necessary for accurate inference of local deuteron densities. This result is partly inferred from a sensitivity analysis of dependences on different plasma parameters. Proper error bars for the model predictions will be derived using Bayesian probability theory.

1. Introduction

Local deuteron densities at JET are currently inferred indirectly from the analysis of impurity line emission induced by charge exchange along the heating beam (von Hellermann 1995). Using local charge neutrality gives the deuteron density

$$n_{\mathrm{D}} = n_{\mathrm{e}} - \sum_{Z>1} Z_{n_Z} \tag{1}$$

where Z is the charge of impurity species with density n_Z. This paper deals with the problem of deducing local deuteron densities directly from a D_α charge exchange line on a background motional Stark beam emission pattern (figure 1). This is accomplished here by modelling the mapping between spectrum and local deuteron density by a multi-layer perceptron (MLP)

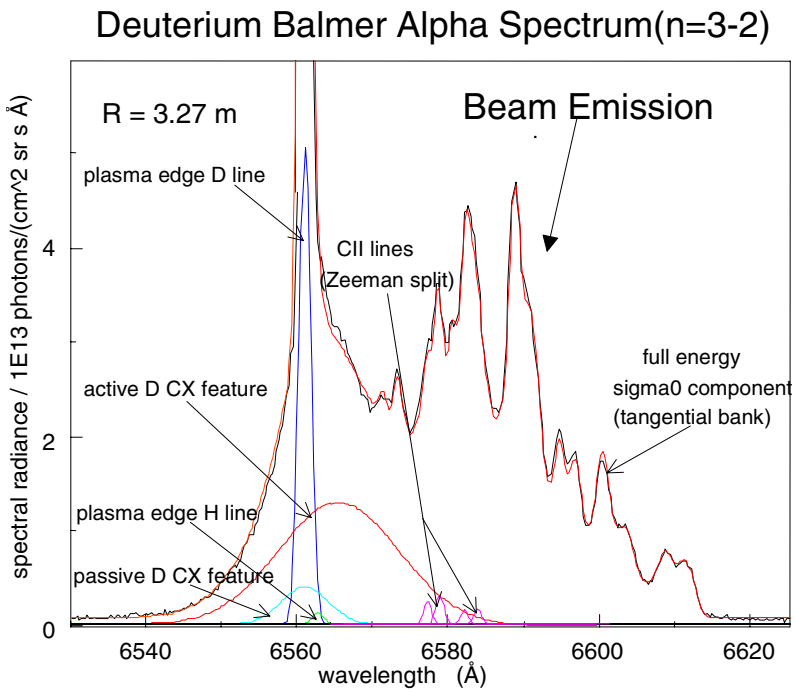

Figure 1. Example of Balmer alpha spectrum. JET pulse 40346.

neural network. The model thus created will be fully automatic and could provide deuteron density profiles in real-time during a pulse. Section 2 gives the physics background for this analysis. In section 3 the validity of the statistical approach taken in this paper is discussed. Section 4 explains the particular neural network model deployed, and error bars on the prediction of that model are then discussed in section 5. The model will be tested on pulses not included in the tuning of the parameters of the statistical model.

2. Calculation of local deuteron densities

The current indirect scheme for inferring deuteron densities from (1) involves analysis of impurity lines (normally C, He, and Be) from several spectra (figure 2). For each spectrum, a charge exchange line is chosen from which the total flux

$$\Phi_Z^{\mathrm{CX}} = \frac{1}{4\pi} \sum_{i=1}^{N_{\mathrm{PINI}}} \sum_{j=1}^{N_{\mathrm{E}}} n_{\mathrm{b}}^{i,j} n_Z q_{\mathrm{CX}}^j L^i \tag{2}$$

gives the density n_Z for one impurity. q_{CX}^j is the effective emission rate (Summers and von Hellermann 1993) (discussed further in section 3) for charge exchange induced line emission and L^i the effective intersection length of the viewing line through beam component i, calculated from an integration along the line of sight, assuming a Gaussian distribution of the neutral beam density over a beam cross section. The sums are over beam components and energy species. The beam densities, $n_{\mathrm{b}}^{i,j}$ for each beam (i) and energy component (j) are

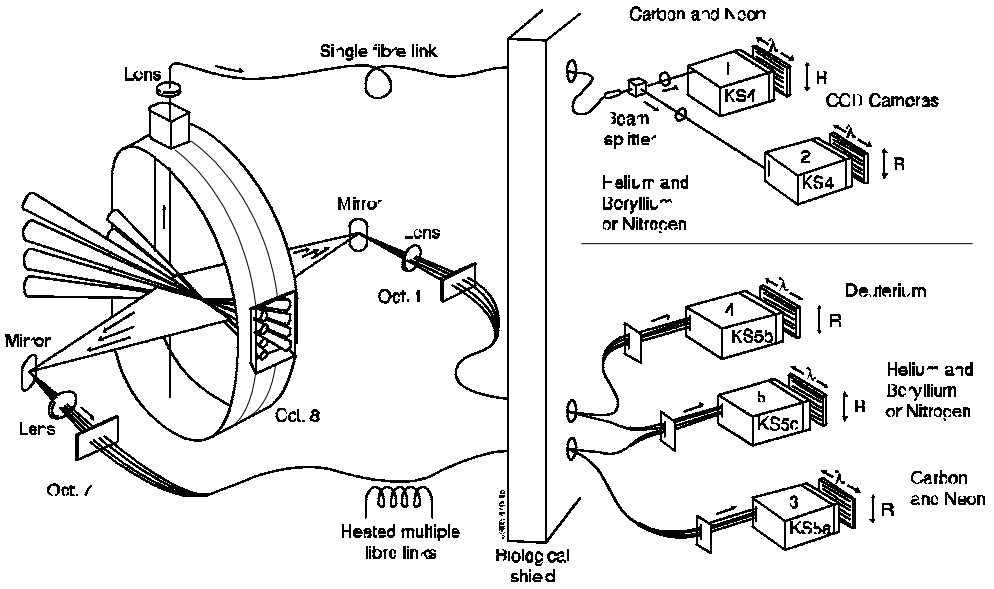

Figure 2. Diagnostic system for active Balmer alpha spectroscopy. Spectrometer 4 is used for beam emission spectra. Spectrometers 1,2,3 and 5 provide data for the indirect analysis.

deduced by calculating the attenuation of the beam along its path

$$n_{\mathrm{b}}^{i,j}(l) = n_{\mathrm{b}}^{i,j}(0)\exp\left\{-\int_0^l n_{\mathrm{e}}\frac{\langle\sigma_{\mathrm{e}}v\rangle}{v_{\mathrm{b}}}\,\mathrm{d}l - \sum_{Z=1}^{N_{\mathrm{ion}}}\int_0^l n_Z\left(\frac{\langle\sigma_Z^{\mathrm{CX}}v\rangle}{v_{\mathrm{b}}} + \frac{\langle\sigma_Z^{\mathrm{ion}}v\rangle}{v_{\mathrm{b}}}\right)\mathrm{d}l\right\} \tag{3}$$

where l is the distance along the beam, v_{b} the beam particle velocity, and $\langle\sigma_{\mathrm{e}}v\rangle/v_{\mathrm{b}}$, $\langle\sigma_Z^{\mathrm{CX}}v\rangle/v_{\mathrm{b}}$, $\langle\sigma_Z^{\mathrm{ion}}v\rangle/v_{\mathrm{b}}$ the velocity-averaged cross sections for electron impact ionization, charge exchange and ion impact ionization of the beam neutrals.

In order to calculate the impurity densities from (2), $n_{\mathrm{b}}^{i,j}$ must first be calculated using (3), but (3) is dependent on all impurity density profiles in the plasma, and so (1)–(3) must be solved together in an iterative self-consistent way (von Hellermann 1995, Mandl 1992). It should also be noted that the total line emissions in (2) are the outcomes of in themselves fairly complex spectral fits with many overlapping lines (von Hellermann and Summers 1993). Thus, while being able to infer the local deuteron densities, the above procedure relies on indirect inference and is a complex non-automatic procedure.

A way to deduce local deuteron densities directly is to measure the local D_α emission from plasma deuterium ions following charge exchange with the beam neutrals (Mandl *et al* 1993). Knowledge of the beam density will then give the local plasma deuteron density directly from

$$\Phi_{\mathrm{CX}} = \frac{1}{4\pi}\sum_{i=1}^{N_{\mathrm{PINI}}}\sum_{j=1}^{N_{\mathrm{E}}} n_{\mathrm{b}}^{i,j} n_{\mathrm{D}} q_{\mathrm{CX}}^{j} L^{i}. \tag{4}$$

Still we would have to calculate n_{D}, $n_{\mathrm{b}}^{i,j}$ and all impurity densities together from (2), (3) and (4), but the deuteron density calculations would in this case be based on direct measurements of deuterium emissions.

A third possibility for the calculation of local deuteron densities comes from using the background beam emission spectrum (figure 1) (described in more detail in section 2), to deduce the local beam densities directly (Mandl *et al* 1993), thus taking away the need for the

self-consistent impurity densities/beam attenuation calculation described above. Combined with (4) this would give the deuteron density directly without involving (2) and (3). This scheme forms the basis of the neural network analysis and is described in more detail in the next sections.

3. Balmer alpha beam emission spectra

3.1. Deduction of deuteron densities

Figure 1 shows an example D_α/beam emission spectrum. The main features are:

- D_α and H_α emission from the edge,
- Prompt D_α emission from thermal deuterium following charge exchange with beam atoms,
- Halo D_α emission from neutralized deuterium around the beam,
- D_α emission from charge exchange with the fast deuteron distribution,
- Motional Stark splitted D_α from beam atoms,
- Passive D_α charge exchange from recycled neutrals.

The active D_α emission will be Doppler shifted relatively to the edge emission. Due to the additional beam halo emission, the geometrically defined precise localization of the active spectrum will be diminished (Mandl *et al* 1993). The beam emission spectrum is characterized by the motional Stark pattern for each of the three energy components of the beam. Nine Stark transitions are visible for each of the three energy components, giving a total of 27 Stark lines. Both the halo emission, edge emission and passive edge charge exchange emission will normally overlap the active D_α charge exchange emission. The total flux for each energy component j gives the neutral density $n_{\mathrm{b}}^{i,j}$ for beam i:

$$\Phi_{\mathrm{BE}}^{i,j} = \frac{1}{4\pi} n_{\mathrm{e}} n_{\mathrm{b}}^{i,j} q_{\mathrm{BE}}^{j} L^{i} \tag{5}$$

if the electron density n_{e} and the effective emission rate q_{BE}^{j} are known. Again, is an effective penetration length of the viewing line through beam component i. q_{BE}^{j} is an effective emission rate in the sense that it takes into account the population of different energy levels for a given set of plasma parameters.

The intersection points between a single horizontal line of sight and different beams are assumed to fall approximately on identical flux surfaces. It is then reasonable to assume that at such intersection points, different beams will have gone through the same n_{e} density profiles, implying that the beam emissions at such positions will differ only by a geometrical factor. This factor is determined by the angle between line of sight and beam and also the nearest distance between viewing line and beam. We can thus write the beam emissions as

$$\Phi_{\mathrm{BE}}^{i,j} = g^{i} \Phi_{\mathrm{BE}}^{i,j} \tag{6}$$

where g^i is the geometrical factor for beam i, giving the ratio of $\Phi_{\mathrm{BE}}^{i,j}$ to $\Phi_{\mathrm{BE}}^{1,j}$ with $g^1 = 1$. Note that, since the plasma rotation velocities at the intersection points are assumed to be identical, the Doppler shifts will be equal and therefore the only observable quantity will be the sum of the beam emissions, not the individual components:

$$\Phi_{\mathrm{BE}}^{j} = \sum_{i=1}^{N_{\mathrm{PINI}}} g^{i} \Phi_{\mathrm{BE}}^{1,j} = \Phi_{\mathrm{BE}}^{1,j} \sum_{i=1}^{N_{\mathrm{PINI}}} g^{i}. \tag{7}$$

From (4)–(7) we can then find an explicit formula for calculating n_{D} from total line fluxes and n_{e}:

$$n_{\mathrm{D}} = n_{\mathrm{e}} \left\{ \frac{\Phi_{\mathrm{CX}}}{\sum_{j=1}^{N_{\mathrm{E}}} (q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}})\Phi^j_{\mathrm{BE}}} \right\}. \tag{8}$$

We see that we will need (apart from the spectrum itself), the electron density n_{e} as input in order to determine n_{D}.

3.2. Dependence of effective emission rate on local plasma parameters

Since the effective rate takes into account all excitation processes and, for the case of $q^{\mathrm{j}}_{\mathrm{CX}}$, subsequent redistribution following charge capture, the ratios $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ will be dependent on several local plasma parameters. Thus, it is a function of ion temperature (T_{i}), effective charge (Z_{eff}), magnetic field (B) and collision energy (E_{coll}) as well as the electron density (n_{e}). In principle, we would therefore have to supply the local values of all these to calculate n_{D} from the spectrum (8). We will now show that either there is information in the spectrum to account for these dependences, *or* the actual sensitivity of $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ to them is low. Even though $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ is in principle dependent on all five parameters, it is not as sensitive to variations of some of them as to others. Figure 3 shows a comparison of the relative change of $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ when different parameters are varied. This graph has been produced by looking upon the $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ dependence on four out of five parameters as probability densities of the form

$$p\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| B, Z_{\mathrm{eff}}, T_{\mathrm{i}}, E_{\mathrm{coll}}\right) \tag{9a}$$

$$p\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| n_{\mathrm{e}}, B, Z_{\mathrm{eff}}, T_{\mathrm{i}}\right) \tag{9b}$$

$$p\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| E_{\mathrm{coll}}, n_{\mathrm{e}}, B, Z_{\mathrm{eff}}\right) \tag{9c}$$

$$p\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| T_{\mathrm{i}}, E_{\mathrm{coll}}, n_{\mathrm{e}}, B\right) \tag{9d}$$

$$p\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| Z_{\mathrm{eff}}, T_{\mathrm{i}}, E_{\mathrm{coll}}, n_{\mathrm{e}}\right). \tag{9e}$$

This follows from the observation that when one parameter is taken away, $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ can only be approximately predicted from the rest and the range of predictions can therefore be described by a probability density. Randomly sampling from the parameter space of the four dependent variables in (9*a*)–(9*e*) and for each such sample varying the fifth parameter over its range, will give us a Monte Carlo estimate of the quantity

$$E\left[\sigma^2\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| x_2, x_3, x_4, x_5\right)\right] \tag{10}$$

where $E[\,]$ denotes expectation value and x_2, x_3, x_4, x_5 represents one group of parameters as in (9*a*)–(9*e*). Equation (10) will give us the average variation of $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ with the excluded parameter. In figure 3 the more intuitive measure

$$E\left[\sigma\left(\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| x_2, x_3, x_4, x_5\right) \Big/ E\left[\frac{q^j_{\mathrm{CX}}}{q^j_{\mathrm{BE}}}\middle| x_2, x_3, x_4, x_5\right]\right] \tag{11}$$

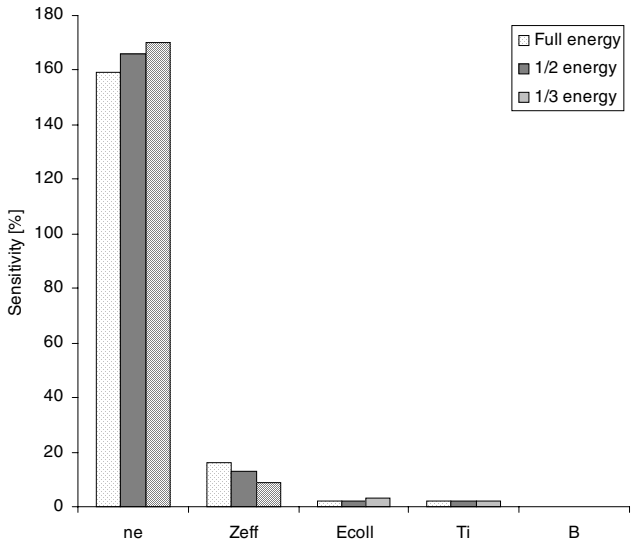

Figure 3. Influence of different plasma parameters on the ratio $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ between the effective rates for different beam energies. The dependency on E_{coll} is due to q^j_{CX} only.

is used, which will roughly give the average relative variation of $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ when the corresponding excluded parameter is varied over its range. Observe that the inner expectation will be over $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ and the outer over x_2, x_3, x_4, x_5.

As can be seen, $q^j_{\mathrm{CX}}/q^j_{\mathrm{BE}}$ does not vary much with E_{coll}, T_{i} and B. The main variation comes from n_{e} and Z_{eff}. We must therefore make sure that information about these is available on the inputs to the networks. Since we already know (from (8)) that we need n_{e}, we only have to deal with Z_{eff}. The weak dependences on T_{i} and E_{coll} can be accounted for in the spectrum from the width and Doppler shift of the D_α charge exchange line corresponding to local ion temperature and plasma rotation velocity respectively. The Z_{eff} dependence demands somewhat more analysis since there is no direct Z_{eff} dependence in the spectrum. We will therefore use a statistical argument. The bremsstrahlung level in the spectrum,

$$\Phi_{\mathrm{brems}} \propto \int \frac{n_{\mathrm{e}}^2}{\sqrt{T_{\mathrm{e}}}} Z_{\mathrm{eff}}\, \mathrm{d}l \tag{12}$$

is dependent on the Z_{eff} and n_{e} profiles (Hutchinson 1987). Even though local Z_{eff} cannot be in principle deduced directly from Φ_{brems} and local n_{e} alone, it is valid to ask how much *information* is provided about local Z_{eff} from just knowledge about Φ_{brems} and *local* n_{e}. To answer this we fitted the surface

$$y = E[Z_{\mathrm{eff}}|\Phi_{\mathrm{brems}}, n_{\mathrm{e}}] \tag{13}$$

using a large dataset of triples $(Z_{\mathrm{eff}}|\Phi_{\mathrm{brems}}, n_{\mathrm{e}})$. $E[\,]$ again denotes expectation value and was modelled by a MLP network. By investigating the variation around this surface it can be deduced how accurately local Z_{eff} can be predicted from Φ_{brems} and local n_{e} alone. We calculated this prediction accuracy by using (13) on a separate set of $(\Phi_{\mathrm{brems}}, n_{\mathrm{e}})$ followed by comparison of the result with the true Z_{eff}. It turned out that it was possible to predict those local Z_{eff} with about 12% average error. The datasets used was from a whole year of experiments.

To find how 12% average error of Z_{eff} will influence the ratio $q^j_{\text{CX}}/q^j_{\text{BE}}$, we again apply the model (9)–(11) but this time introducing a 12% variation on Z_{eff} and from that calculate (11). This gave about a 2% average change of $q^j_{\text{CX}}/q^j_{\text{BE}}$ under the assumption that Z_{eff} was known with 12% accuracy. This analysis was done for a central line of sight, and the results might therefore differ somewhat for other radial positions. The low 2% extra variation of $q^j_{\text{CX}}/q^j_{\text{BE}}$ for the line of sight analysed, points in the direction that up to a high degree of accuracy, the only extra parameter needed to account for the $q^j_{\text{CX}}/q^j_{\text{BE}}$ dependence on different plasma parameters is n_{e}, which will be used as auxiliary input. The quantities needed to determine n_{D} should therefore, apart from n_{e}, be found from information in the spectrum alone.

4. MLP neural network model

MLP neural networks are a family of parametrized functions capable of effectively handling function approximations in high-dimensional spaces (Bishop 1995). It has been previously used for spectral fitting in, for example, Bishop *et al* (1993), Baker *et al* (1994), Svensson *et al* (1999) and Larsen (1992). The advantages of extracting physical parameters from measurements with MLPs are very high processing speeds, full automation and sometimes (like in the present case) a possibility to extract parameters without a fully specified physical model of the data. The disadvantages have traditionally been the difficulty of providing error bars for the MLP predictions that take into account both inherent noise on the data and the uncertainty imposed by the MLP being only an approximation of the true mapping. Particularly, if the MLP tries to make predictions on data outside the range of data used for creating the specific MLP model, the error bars must reflect this. There are many ways of providing such error bars (Jacobs *et al* 1991, Nix and Weigend 1994, MacKay 1992, Neal 1996), and here we will use an approach based on Bayesian inference (MacKay 1992, Neal 1996) (section 5 deals with this further). Our approach will be slightly different from the most commonly used Bayesian approach to error bars though, since that approach will overestimate the errors on predictions for applications like the current one (Svensson 2000).

4.1. Statistical model

We will here refer to the neural network merely as a parametrized function mapping inputs (beam emission spectrum, electron density) to outputs (e.g. deuteron density):

$$y = f(\boldsymbol{x}; \boldsymbol{w}) \tag{14}$$

where y is the network output, $\boldsymbol{x}$ the vector of inputs, and $\boldsymbol{w}$ the parameters in the network to be determined from a set of examples of the mapping $\boldsymbol{x} \rightarrow y$, referred to as the *training set (D)*. It turns out to be advantageous to train the network on

$$\sum_{Z>1} Z n_Z \tag{15}$$

rather than n_{D} (see section 4.2) and then calculate n_{D} in a separate step from

$$n_{\text{D}} = n_{\text{e}} - \sum_{Z>1} Z n_Z. \tag{16}$$

The likelihood of the data D under model (14) is given by the product of the likelihoods of the individual targets, t^p, where p is the number of the target:

$$p(t^p|\boldsymbol{x}^p, \boldsymbol{w}) = 1/\sqrt{2\pi}\sigma^p(\boldsymbol{x}^p) \exp\left(- \frac{(y^p - f(\boldsymbol{x}^p; \boldsymbol{w}))^2}{2\sigma^p(\boldsymbol{x}^p)^2} \right). \tag{17}$$

Since there will be an error associated with both the input values (spectrum and n_e), and the target value ($\sum_{Z>1} Zn_Z$), the model should optimally take into account contributions from both inputs and outputs. In Svensson (2000) a scheme for doing this is introduced, and we will use that model here. The basis for the model is to let the standard deviations $\sigma^p(\boldsymbol{x}^p)$ reflect output uncertainty both from the target values and from the inputs, achieved by assuming a local linear approximation of the network:

$$\sigma^p(\boldsymbol{x}^p) = \sqrt{(\sigma_t^p)^2 + \sum_{i=1}^{N_{IN}} (\sigma_{x_i}^p)^2 \left(\frac{\partial f}{\partial x_i^p}\right)^2}. \tag{18}$$

Here $\sigma_{x_i}^p$ refers to the standard deviation of input number i: for example, number p. Taking the negative logarithm of the total likelihood gives the cost function

$$E = \frac{1}{2} \sum_{p=1}^{N_P} \log(\sigma^p(\boldsymbol{x}^p, \boldsymbol{w})^2) + \frac{(t^p - f(\boldsymbol{x}^p, \boldsymbol{w}))^2}{\sigma^p(\boldsymbol{x}^p, \boldsymbol{w})^2} \tag{19}$$

to be minimised in $\boldsymbol{w}$. Using (20) in place of the standard cost function for neural network training

$$E = \frac{1}{2} \sum_{p=1}^{N_P} (t^p - f(\boldsymbol{x}^p, \boldsymbol{w}))^2 \tag{20}$$

improved the accuracy by up to about 20%, since it represents a more accurate model of the data. The solution will also suffer much less from the problem of overfitting when this model is used (Svensson 2000).

4.2. Architecture

We use one network for each viewing line of the spectrometer, giving nine networks altogether, ranging in radial positions from $R = 3.14$ m to $R = 3.69$ m (figure 4). Each network takes as input the corresponding beam emission spectrum and the electron density for that position. The output of each position is $\sum_{Z>1} Zn_Z$. The deuteron density is thus calculated by subtracting the network output from the electron density as in (16). This will give a higher accuracy than if we were to train the network on the deuteron densities directly, since in expression (1) the electron density will be the dominating term, the risk being therefore that the network fits the noise on n_e.

In fact, the relative error of n_D will be

$$\varepsilon^{n_D} = \varepsilon^y \left(\frac{n_e}{n_D} - 1\right) \tag{21}$$

where ε^y is the relative error of the network output. In the limit when $n_e \to n_D$, $\varepsilon^y \to \infty$, since the target value $\sum_{Z>1} Zn_Z = n_e - n_D$ is the denominator in the relative error ε^y. Performing the limit calculations gives $\varepsilon^{n_D} = y/n_D$ for $n_e \to n_D$ (y is the network output). For a majority of scenarios $n_e < 2n_D$, and (21) will reduce the error a factor $(n_e/n_D - 1)$ compared to the error of the network output.

The total number of inputs to each network is 137, and the number of hidden neurons (Bishop 1995) is 20, giving 2781 free parameters for each network.

4.3. Preprocessing

It is very important that the large training set is consistently analysed and that outliers are detected. For this dataset we have, for example, excluded spectra recorded during large edge-localized modes (ELMs), where extra lines are likely to add to the complexity of the charge

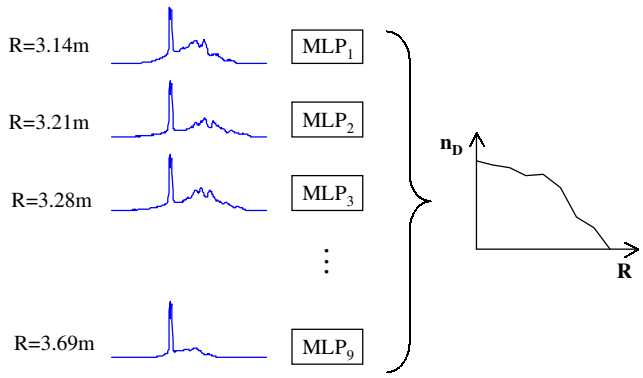

Figure 4. Overview of neural net n_D-analysis system. Each of the nine networks analyse spectra from one radial position, after which the full profile is reconstructed.

exchange spectra from which the densities n_Z are deduced, making their corresponding error possibly very large. Because of the existence of n_e on the inputs, it was necessary to include a wide range of plasma impurity levels in the training set, so that $\sum_{Z>1} Zn_Z$ would not be deducible from n_e alone. We have also cross-checked the densities n_Z inferred from charge exchange spectra by comparison with measurements of line-averaged Z_{eff} from bremsstrahlung radiation and also compared the plasma energy calculated from those densities to the total kinetic energy as deduced from measuring the plasma diamagnetism (Hutchinson 1987). Datapoints where the predictions and measurements disagree considerably were excluded from the training set. The criterion used was to keep values where

$$|Z_{\text{eff}}^{\text{CX}} - Z_{\text{eff}}^{\text{brems}}| < 1 \tag{22a}$$

and

$$|W_{\text{dia}}^{\text{CX}} - W_{\text{dia}}^{\text{magn}}| < 10^9 \text{ J}. \tag{22b}$$

These and a few other simpler preprocessing rules left a training set of about 6000 points per viewing line. A separate test set was put aside, containing pulses not in the training set, to be used to assess the final performance of the networks (see section 6).

5. Error bars on model predictions

5.1. Bayesian formalism

Bayesian probability theory represents a stand-alone approach to probability theory. Central to the Bayesian approach is the view that probabilities are to be looked upon as *uncertainties* representing the amount of information available at a certain point in an inference process (about, for example, an unknown parameter). This means that for a standard parameter estimation problem, there will be associated a probability distribution to the parameter about which inferences are to be made, in contrast to standard probability theory, where only the measurements would have an associated probability distribution (and the most likely value (for example) of the parameter would be calculated using an estimator). With standard statistical methods for parameter estimation one thus usually chooses an *estimator* for the parameter of interest, for example the average, or the maximum of the likelihood. In the Bayesian view one does not primarily use estimators for making inferences about parameters. The final outcome

of Bayesian inference is always the *posterior* probability distribution of the parameter. For a given model, new data D will update the posterior distribution in accordance with Bayes' rule:

$$p(\boldsymbol{w}|D) = \frac{p(D|\boldsymbol{w})p(\boldsymbol{w})}{p(D)} \tag{23}$$

where $\boldsymbol{w}$ is the parameter whose values are to be inferred from the data D. It would be far outside the scope of this paper to thoroughly motivate the validity of the Bayesian approach described above, but there are a number of textbooks (Sivia 1996, Gelman *et al* 1995, Jaynes 1998, Jeffreys 1939), general papers (Cox 1946, Jaynes 1986) and applications (Sivia and Carlile 1992, Thodberg 1993) available. For our purposes, if $\boldsymbol{w}$ in (23) is the vector of free parameters in the network and D the training set, then (23) will give the uncertainty of the parameters $\boldsymbol{w}$ in the light of the training data and the network model. Once the network is trained, these uncertainties will influence the accuracy to which the function (14) makes predictions of $\sum_{Z>1} Zn_Z$. This uncertainty will thus be inherent to the network, and must be added to the error caused by the noise on the input for a particular prediction (see section 5.2).

The distribution $p(\boldsymbol{w})$ in (23) is the *prior* distribution for $\boldsymbol{w}$. It represents the information of the distribution of $\boldsymbol{w}$ that is known prior to the observation of D and will reduce overfitting effectively by working as a regularizer (Bishop 1995). Even though there are schemes for deducing the form of a parameterized prior (MacKay 1992), we will here assume a flat, non-informative, prior because of the shorter training times associated with such models. Also, the specific model we use (Svensson 2000) will in itself regularize the solution well.

5.2. Error bars

If an uncertainty is associated with the free parameters $\boldsymbol{w}$ of the network, subsequent predictions $f(\boldsymbol{x}; \boldsymbol{w})$ will have to take these uncertainties into account, as well as taking into account noise on the inputs $\boldsymbol{x}$ from which the current predictions are made. The network error bars will thus have two contributions. The uncertainty in the prediction caused by the noise on the inputs $\boldsymbol{x}$ can be calculated by making a linear approximation of the network over the variation of $\boldsymbol{x}$ (Svensson *et al* 1999)

$$\sigma_{\boldsymbol{x}}^2(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{IN}}} (\sigma_{x_i})^2 \left(\frac{\partial f}{\partial x_1} \right)^2. \tag{24}$$

For calculating the contribution to the error from the uncertainties on $\boldsymbol{w}$, we follow the approach in Svensson (2000), where the variation of the network over the posterior

$$\sigma_{\boldsymbol{w}}^2 = \int (f(\boldsymbol{x}; \boldsymbol{w}) - m(\boldsymbol{x}))^2 p(\boldsymbol{w}|\boldsymbol{D})\, \mathrm{d}\boldsymbol{w} \tag{25}$$

is used. m is here the specific network function $f(\boldsymbol{x}; \boldsymbol{w}_{\mathrm{opt}})$ found by the minimization of (19). After some manipulations, including Laplace's approximation (Gelman *et al* 1995), (25) can be shown (Svensson 2000) to reduce to

$$\sigma_{\boldsymbol{w}}^2(\boldsymbol{x}) = (\nabla_{\boldsymbol{w}} f)^T \boldsymbol{A}^{-1} (\nabla_{\boldsymbol{w}} f) \tag{26}$$

where $\boldsymbol{A}$ is the Hessian matrix of the negative logarithm of the posterior (23) at $\boldsymbol{w}_{\mathrm{opt}}$. Since the dimensionality of $\boldsymbol{w}$ is very large (see section 4.2), we will use a diagonal approximation of $\boldsymbol{A}$, keeping only the terms on the diagonal (Bishop 1995). This corresponds to assuming a spherical Gaussian distribution at the maximum of the posterior. The total error for a network prediction will thus be

$$\sigma_{\mathrm{f}}(\boldsymbol{x}) = \sqrt{\sigma_{\boldsymbol{w}}^2(\boldsymbol{x}) + \sigma_{\boldsymbol{x}}^2(\boldsymbol{x})} \tag{27}$$

where $\sigma_{\boldsymbol{w}}^2(\boldsymbol{x})$ is given by (26) and $\sigma_{\boldsymbol{x}}^2(\boldsymbol{x})$ by (24).

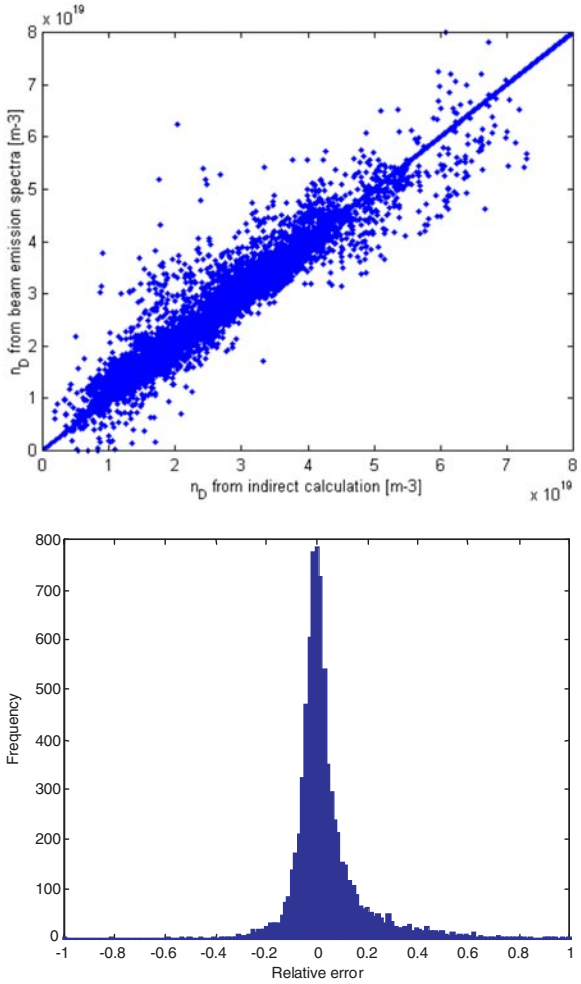

Figure 5. Comparison of about 8000 analysed beam emission spectra with corresponding n_{D} from indirect analysis. The average relative error is 4.3%.

In the final step, n_{D} is calculated from

$$n_{\mathrm{D}} = n_{\mathrm{e}} - f(n_{\mathrm{e}}, \boldsymbol{s}) \tag{28}$$

where $f(n_{\mathrm{e}}, \boldsymbol{s})$ is the network function, showed here as a function of n_{e} and spectrum $\boldsymbol{s}$. The error on n_{D} will thus have contributions from n_{e}, $f(n_{\mathrm{e}}, \boldsymbol{s})$ and the covariance between n_{e} and the network function $f(n_{\mathrm{e}}, \boldsymbol{s})$. Since $f(n_{\mathrm{e}}, \boldsymbol{s})$ is a differentiable function of n_{e}, the final error on n_{D} can be calculated to be

$$\sigma_{n_{\mathrm{D}}} = \sqrt{\sigma^2(n_{\mathrm{e}}) + \sigma_{\mathrm{f}}^2 - 2\frac{\partial f}{\partial n_{\mathrm{e}}}\sigma^2(n_{\mathrm{e}})}. \tag{29}$$

where $-2(\partial f/\partial n_{\mathrm{e}})\sigma^2(n_{\mathrm{e}})$ is the covariance term and σ_{f} is given by (27). Equation (29) is the form of the error bar that will be used in section 6.

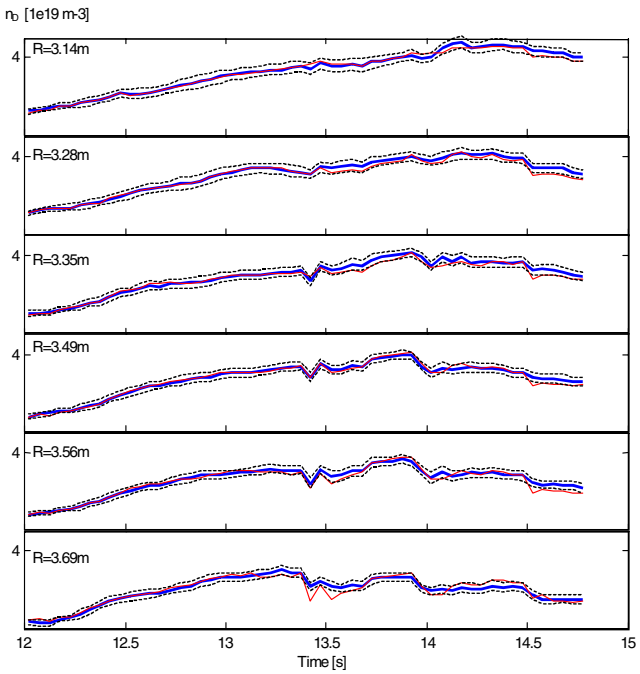

Figure 6. Time traces of neural net n_{D} (thick curve), indirect n_{D} (thin curve) and error bars (30) (dashed curves). JET pulse 40352.

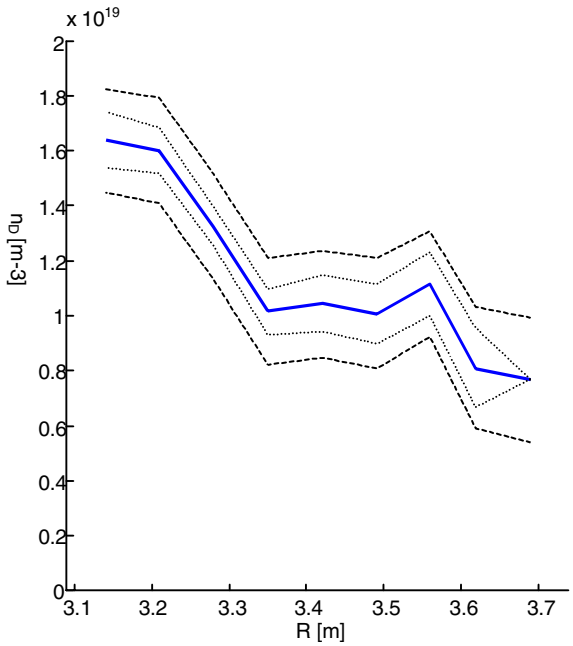

Figure 7. n_{D} profile with error bars. Dashed curves are total error (29), and dotted curves the posterior error contribution (26). For clarity, 2σ are shown for both error bars. JET Pulse 42243, $t = 7.83$ s.

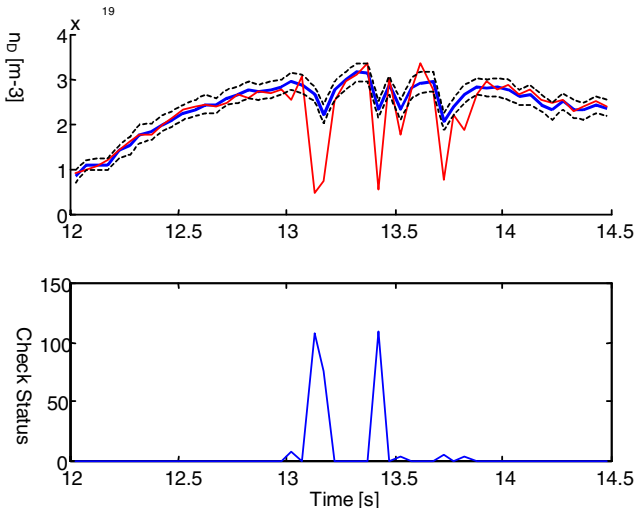

Figure 8. Above: neural net time trace (thick curve), indirectly calculated n_D (thin curve) and the error bars (29) (dashed curves). Below: a measure of the number of consistency checks that failed. JET pulse 40490. $R = 3.69$ m.

6. Results

The results in this section are all based on a test set with pulses not used for the training of the networks. Altogether 42 JET pulses were in the test set, providing about 8000 spectra. The average relative error over all radial positions on this set was 4.3%. A correlation plot and histogram are shown in figure 5. It should be noted that datapoints that did not pass the consistency tests described in section 4.3 were not included in the comparison. This means that neural network results were only compared with the indirect calculations that were not in doubt. In actuality, the 4.3% figure should be taken with some care, since there is always the possibility that some of the datapoints in the test pulses could be very similar to datapoints in the training set since the test set was collected from the same time period as the training set.

Figure 6 shows timetraces for different radii together with error bars as derived in section 5. In figure 7 is shown a n_D profile with the different contributions to the error bars resolved. Please observe that the error bars for the neural net prediction will generally be much larger than the 4.3% error mentioned above. This is because the latter measures the proximity of the neural net to the density calculated with the indirect method. Both methods use the same n_e and so influences of n_e will be correlated between the results of the two models. The error bars (30) take into account both the uncertainty due to n_e and to the uncertainty due to the limited training set used for training—it is not a comparison with the indirect model. For a discussion of errors on the indirect model from which the neural network is trained, see von Hellermann and Summers (1993).

Finally, figure 8 gives an example of the importance of the preprocessing described in section 4.3. The largest discrepancies between the neural net model and the indirect calculations occur for datapoints that were checked to be inconsistent by applying rules like those in section 4.3. The low average error gives credence to the model developed in this paper, but a final assessment of the predictive error should optimally involve a larger test set, based on data collected and analysed during a period of many months or years after the period from which training data was collected.

7. Conclusions

It has been shown that a statistical model for the analysis of beam emission spectra based on MLP neural networks can calculate local deuteron densities with below 5% error. Even though the statistical model will inherit average systematic errors from the underlying indirect model, the basis for predictions is a direct deuterium emission line, making it currently the only direct measure of local deuteron densities at JET. Also, the model works fully automatically and the speed with which predictions are made is so high that a diagnostic for real-time fuel density profiles would be possible. The validity of this approach has been outlined and error bars based on Bayesian probability theory were implemented as a measure of the uncertainty of the model predictions.

Acknowledgments

This work was supported by the European Community under an association contract between EURATOM and the Swedish Natural Science Research Council NFR.

References

Baker D R, Groebner R J and Burrell K H 1994 A neural network for the analysis of DIII-D charge exchange recombination data *Plasma Phys. Control. Fusion* **36** 109

Bishop C M 1995 *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press)

Bishop C M, Roach C M and von Hellermann M 1993 Automatic analysis of JET charge exchange spectra using neural networks *Plasma Phys. Control. Fusion* **35** 765

Cox R T 1946 Probability, frequency and reasonable expectation *Am. J. Phys.* **14** 1–13

Gelman A, Carlin J B, Stern H S and Rubin D B 1995 *Bayesian Data Analysis* (London: Chapman and Hall)

Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation (Santa Fe Institute Studies in the Sciences of Complexity. Lecture Notes)* vol 1 (Reading, MA: Addison-Wesley)

Hutchinson I H 1987 *Principles of Plasma Diagnostics* (Cambridge: Cambridge University Press)

Jacobs R A, Jordans M I, Nowlan S J and Hinton G E 1991 Adaptive mixtures of local experts *Neural Comput.* **3** 79

Jaynes E T 1986 Bayesian methods: an introductory tutorial *Maximum Entropy and Bayesian Methods in Applied Statistics* ed J H Justice (Cambridge: Cambridge University Press)

——1998 *Probability Theory: The Logic of Science* not yet published, but can be found on ftp://bayes.wustl.edu/Jaynes.book

Jeffreys H 1939 *Theory of Probability* (Oxford: Clarendon)

Larsen J T, Morgan W L and Goldstein W H 1992 Artificial neural networks for plasma X-ray spectroscopic analysis *Rev. Sci. Instrum.* **63** 4775

MacKay D J C 1992 Bayesian interpolation *Neural Comput.* **4** 415

Mandl W 1992 Development of active Balmer-alpha spectroscopy at JET *PhD Thesis* JET-IR(92)05, JET Joint Undertaking, Abingdon, UK

Mandl W, Wolf R, von Hellermann M and Summers H P 1993 Beam emission spectroscopy as a comprehensive plasma diagnostic tool *Plasma Phys. Control. Fusion* **35** 1373

Neal R M 1996 *Bayesian Learning for Neural Networks* (Berlin: Springer)

Nix D A and Weigend A S 1994 Estimating the mean and variance of the target probability distribution *Proc. IEEE Int. Conf. on Neural Networks (Orlando, FL)* IEEE-ICNN94 (Piscataway, NJ: IEEE) p 55

Sivia D S 1996 *Data Analysis A Bayesian Tutorial* (Oxford: Oxford University Press)

Sivia D S and Carlile C J 1992 Molecular spectroscopy and Bayesian spectral analysis—how many lines are there? *J. Chem. Phys.* **96** 170

Summers H P and von Hellermann M 1993 Atomic and molecular data needs for spectroscopic diagnostics of fusion plasmas *Atomic and Plasma Material Interaction Processes in Controlled Thermonuclear Fusion* ed Janev (Amsterdam: Elsevier) p 87

Svensson J 2000 An MLP training algorithm taking into account known errors on inputs and outputs *JET Report*

Svensson J, von Hellermann M and König R W T 1999 Analysis of charge exchange spectra using neural networks *Plasma Phys. Control. Fusion* **41** 315

Thodberg H H 1993 *Ace of Bayes: Application of Neural Networks with Pruning* (The Danish Meat Research Institute, Roskilde) Manuscript 1132E

von Hellermann M 1995 Feasibility of quantitative spectroscopy on ITER *Proc. 1st Int. Workshop on ITER Diagnostics (Varenna)* (New York: Plenum)

von Hellermann M and Summers H P 1993 Active beam spectroscopy at JET *Atomic and Plasma Material Interaction Processes in Controlled Thermonuclear Fusion* ed Janev (Amsterdam: Elsevier) p 135