Tomography using neural networks

G. Demeter
Department of Plasma Physics, Research Institute for Particle & Nuclear Physics of the Hungarian Academy of Sciences, XII. Konkoly Thege Miklòs ùt 29-33, H-1525 Budapest, P.O. Box 49, Hungary

(Received 16 July 1996; accepted for publication 15 November 1996)

We have utilized neural networks for fast evaluation of tomographic data on the MT-1M tokamak. The networks have proven useful in providing the parameters of a nonlinear fit to experimental data, producing results in a fraction of the time required for performing the nonlinear fit. Time required for training the networks makes the method worth applying only if a substantial amount of data are to be evaluated. © *1997 American Institute of Physics.* [S0034-6748(97)00603-5]

I. INTRODUCTION

Impurity injection using laser accelerated pellets[1–3] and the study of the transport of these injected impurities has been the major field of investigation in recent years on the MT-1M tokamak.[4,5] In some experiments, two 16-channel microchannel plate (MCP) cameras[6] were placed at various cross sections of the torus, one horizontally, the other vertically. Micro pellets of impurities were injected using a laser blow-off device vertically from below and these impurities emitted radiation in the soft x-ray domain as the pellets ablated in the plasma. (Figure 1 shows the experimental setup.) With various filters added, the cameras provided information on the distribution of the injected impurity ions in a cross section of the torus. More precisely, each channel of these cameras measured an integral of the radiation of the injected impurity ions along a linear domain of the cross section. These signals were digitized every 10 μs and thus a time evolution of the cloud of impurity ions could be investigated. The signal on the channels of the two cameras in a typical experiment can be seen in Fig. 2. It can be seen corresponding with a localized distribution around the center of the tokamak cross section in the horizontal direction, and moving from the edge of the plasma toward the center in a vertical direction. The problem is to restore the original distribution of impurity ions in the cross section of the torus from these measurements as accurately as possible.

II. THE PROBLEM OF TOMOGRAPHIC RECONSTRUCTION

The problem of tomographic reconstruction is to reconstruct a two dimensional source function from a set of integrals of this source function.[7] Obviously, a lot of information is lost because each channel integrates the radiation along a line, and it is impossible to restore the original distribution exactly. Therefore we attempted to approximate the source function $\Phi(x,y)$ from the measurements on the channels of the cameras, which are integrals of this source function:

$$M_i=\int \Phi(x,y)\times\omega_i(x,y)dx\ dy.$$

The functions $\omega_i(x,y)$ contain information on the measurement setup (geometry, etc.) and are assumed to be known. For the approximation, we used a given class of test functions $F(x,y,p_k)$ containing parameters p_k. The aim is to find the set of parameters that minimize the error function

$$E(p_k)=\sum\left(\int F(x,y;p_k)\times\omega_i(x,y)dx\ dy-M_i\right)^2.$$

We tried to minimize the measurable difference between the original source function and the approximating function $F(x,y;p_k)$ belonging to a prescribed class of test functions. The integrals containing the functions $\omega_i(x,y)$ must be evaluated numerically. If $F(x,y;p_k)$ functions are linear sums of a set of base functions $\varphi_k(x,y)$, i.e.,

$$F(x,y;p_k)=\sum_k p_k\times\varphi_k(x,y)$$

then

$$E(p_k)=\sum\left(\sum_k p_k\times\int \varphi_k(x,y)\times\omega_i(x,y)dx\ dy-M_i\right)^2,$$

i.e., the parameters can be taken out of the numerical integrals containing the characteristics of the measurement setup. The minimization will therefore be that of a quadratic function of the parameters, which is simple and computationally efficient. If an iterative minimization of the error function is implemented, the integrals have to be performed only once for each measurement setup and choice of base functions. Alternatively, the error function, being a quadratic one, may be minimized using a matrix inversion method.[7] The choice of a linear superposition of some base functions as a test function, however, works best if there are a larger number of measurements (integrals of the source function) than there are base functions needed to adequately describe the source function. If the number of base functions (and hence the number of parameters) is larger than the number of measurements, the conditions for minimal error will only define a subspace of the entire parameter space on which the minimization takes place. This means that additional criteria have to be added to select a point in this subspace of parameters, which may complicate calculations considerably. Such criteria are used, for example, by maximum entropy tomography and minimum undulation tomography.[7] In our case the problem is that there is no simple set of base functions which corresponds well to a localized distribution moving in the cross section of the tokamak. We may try to use a linear superposition of localized functions as a base (for example we may use two-dimensional step functions), however, for

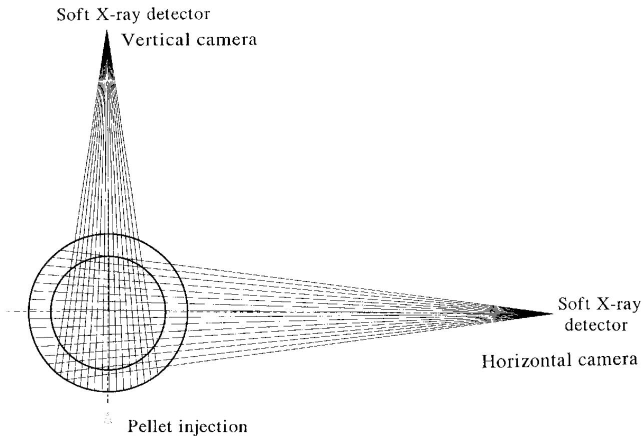

FIG. 1. Experimental setup. Both the horizontal and the vertical camera had 16 viewing chords each.

an adequate description of the source function we need much more than 32 base functions (the number of measurements available). Another natural choice would be to fit a two dimensional Gaussian distribution to the measurements

$$F(x,y;p_k)=\frac{A}{2\pi\sigma_x\sigma_y}\times\exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2}-\frac{(y-y_0)^2}{2\sigma_y^2}\right),\quad p_k\in\{x_0,y_0,A,\sigma_x,\sigma_y\}$$

with parameters for the position of the center of the distribution in the horizontal and vertical directions, the widths of the distribution in the two directions, and an amplitude parameter. The problem is that this function contains its parameters nonlinearly, meaning that the numerical integrals have to be evaluated in every step of an iterative minimization of the error function. This makes fitting nonlinear functions to tomographic data extremely inefficient and cumbersome.

III. NEURAL NETWORKS

Neural networks have recently evolved into a powerful method of problem solving applied in a variety of fields. The basic idea behind neurocomputing, derived from analogies of the human brain, is to use a large number of primitive processors to evaluate data in parallel.[8–10] The single processor (or neuron) would perform a simple task of multiplying each of a number of input values by an internal weight value corresponding to that input, creating a linear sum of the weighted inputs and producing some simple function (called the transfer function) of the linear sum as its output. With a large number of neurons organized into a network, the network as a whole may be able to perform complicated tasks. One frequently used architecture of these networks is a multilayered, feedforward type network called the Backpropagation Network. This network consists of layers of neurons, with the neurons of the first layer receiving the input (which is an n-dimensional vector), and each successive layer receiving the output of the previous layer as its input. At the end, the output layer of this network produces an m-dimensional output vector. The neurons within the individual layers are not interconnected, i.e., the input of a neuron of a layer consists only of the outputs of neurons of the previous layers. Thus this network realizes a mapping from an n-dimensional input space to an m-dimensional output space. There are mathematical theorems to prove this network to be a universal function approximator under certain conditions.[8]

One of the most important virtues of such systems is, that so called learning strategies may be utilized to change the weights of individual neurons to adjust the performance of the network. These strategies can be used to teach a network to solve a problem using examples of desired output to specific inputs. The Backpropagation Network is simple enough for a straightforward learning strategy for training from examples to be formulated. Another important virtue of this network (and many others as well) is its resistance to noise—its ability to perform well in a noisy environment. Neural networks have been used for fast measurement evaluation in plasma physics previously, including nonlinear curve fitting to experimental data.[11–16]

IV. APPLICATION OF THE NETWORKS

The question now arises as to whether such neural networks could be trained to ''guess'' the parameters that a conventional nonlinear curve fitting would produce on tomographic data. In other words, we may try to find a com-

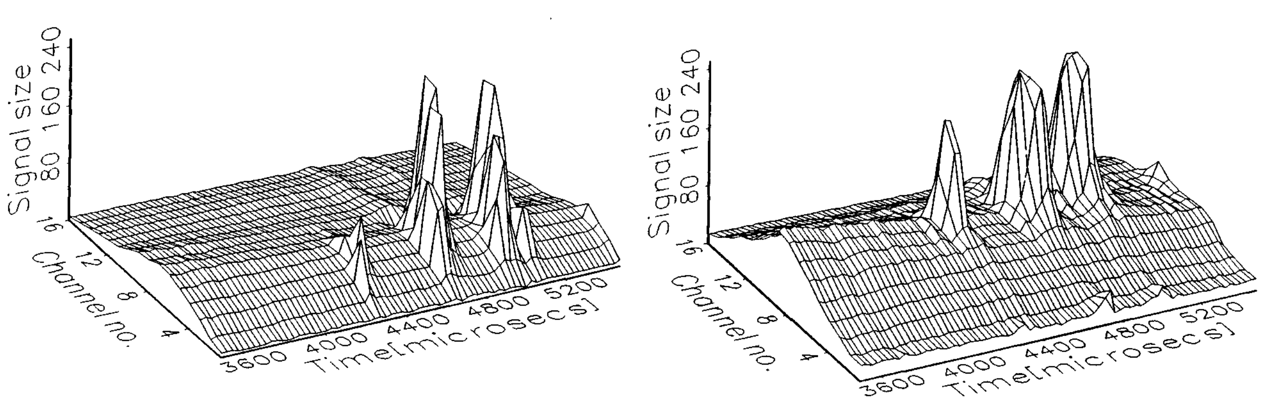

FIG. 2. The signal on the channels of the cameras in a typical experiment, horizontal camera on the left, vertical camera on the right.

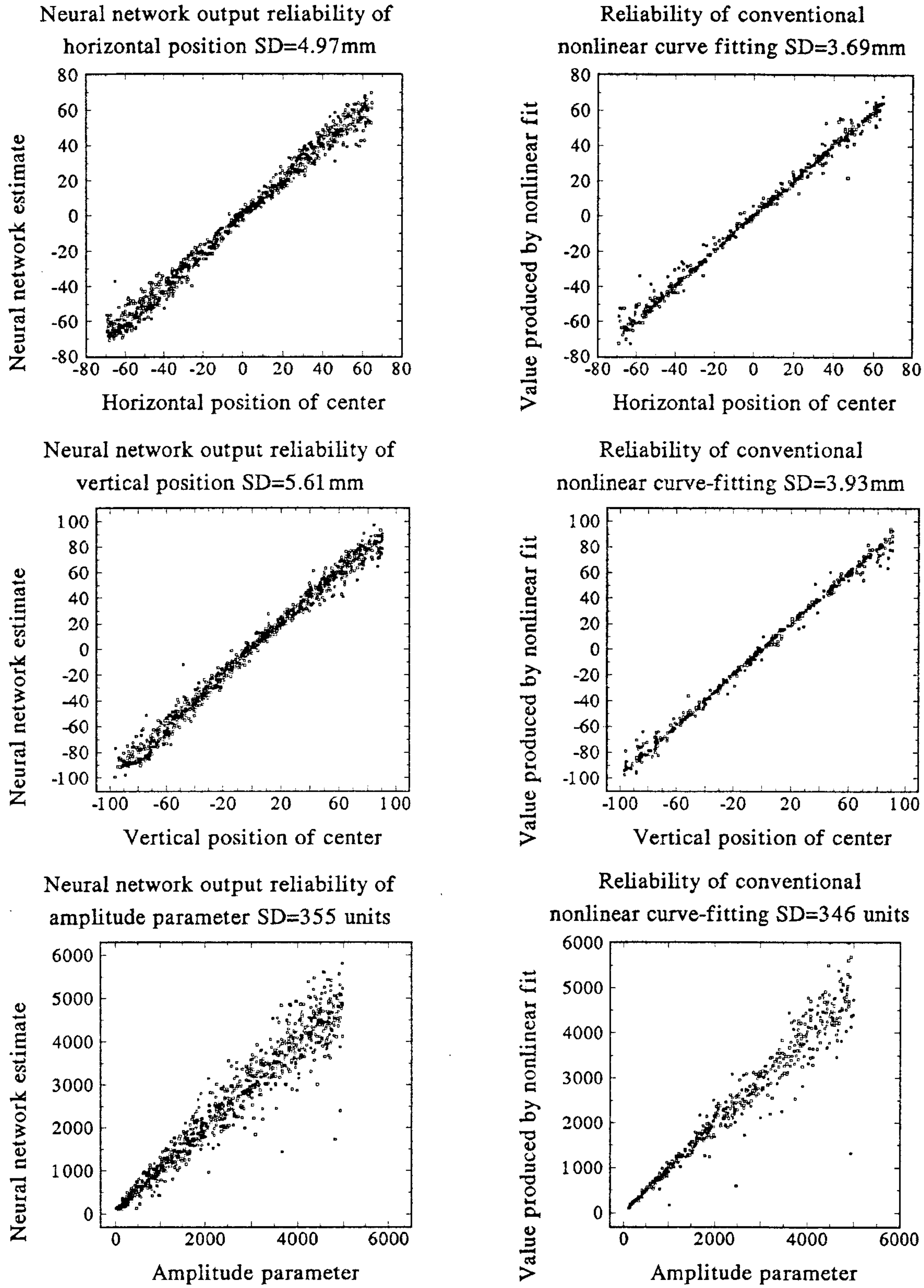

FIG. 3. Performance of the neural networks and conventional curve fitting on position and amplitude parameters. On each of the figures the network output or the results of the curve fitting are plotted on the vertical axis against the generating parameters of the samples (desired output). Standard deviation from the desired output is written on top. Note that the plots of the conventional fit contain only half the number of points as the neural network plots.

plicated mapping that returns the parameters that a nonlinear curve fitting would produce given the tomographic data. A database of samples of the input space with corresponding desired outputs to train the networks may be obtained by taking two dimensional Gaussian distributions with various parameters, making these generating parameters the desired output, and calculating the signals that the detectors would measure from the knowledge of the experimental setup. Thus

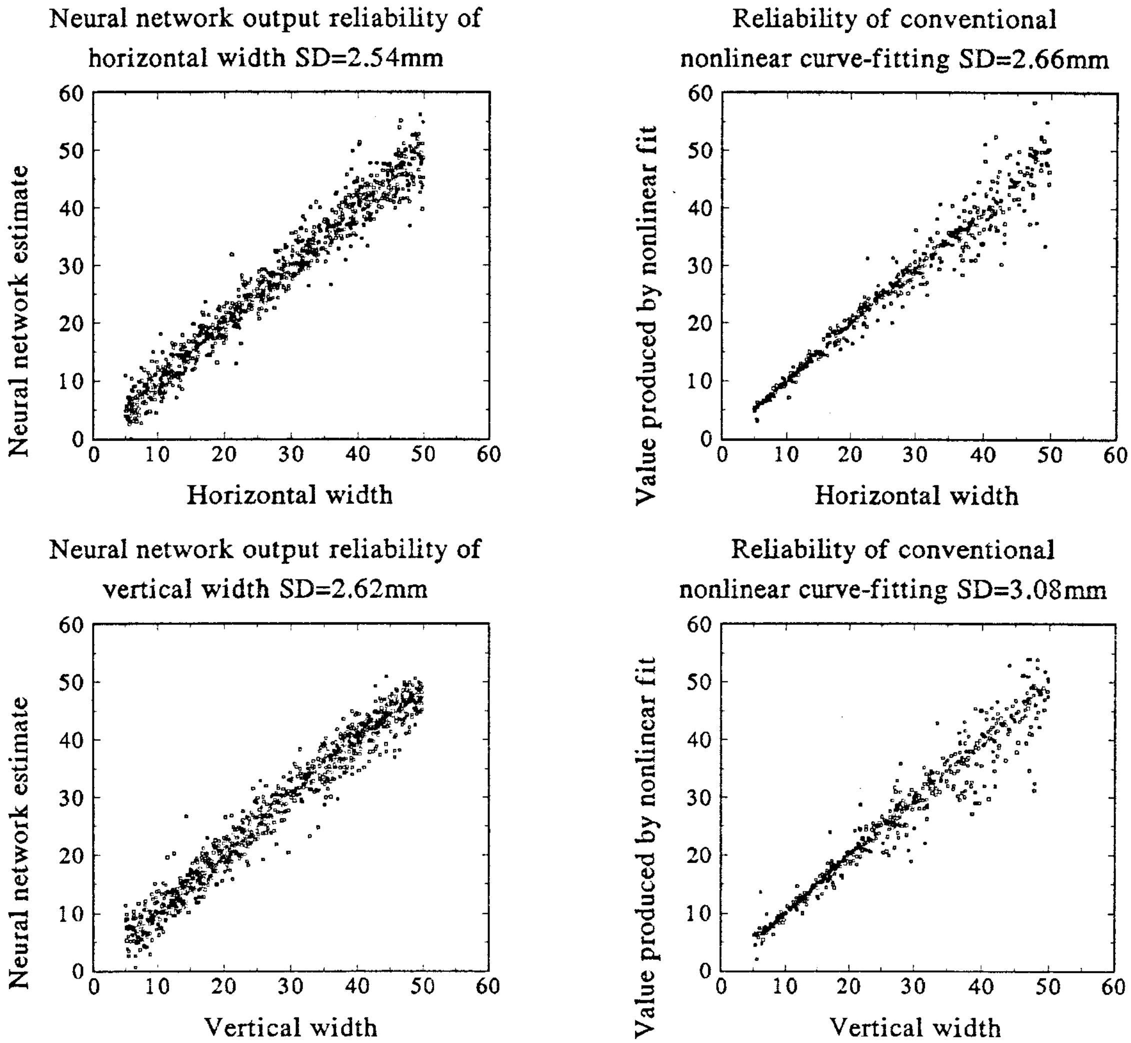

FIG. 4. Performance of the neural networks and conventional curve fitting on width parameters. On each of the figures the network output or the results of the curve fitting are plotted on the vertical axis against the generating parameters of the samples (desired output). Standard deviation from the desired output is written on top.

data for training is readily available, and this fact makes the application of the Backpropagation Network straightforward. This method of data evaluation was tried on the MT-1M tokamak.

The Backpropagation Networks utilized had two layers—one hidden and one output layer. The hidden layer neurons had a sigmoid transfer function, $f(x)=1/(1+e^{-x})$, which is a convenient continuous approximation of the step function and the output neurons simply produced their linear sum as their output. The output of the five parameters (corresponding to the two position, two width, and the amplitude parameter of the Gaussian) were all scaled to be in the interval [0, 10]. It was found to produce better results, when training five networks with one output neuron each, to estimate one parameter of the distribution rather than to train one with five output neurons to estimate all five parameters simultaneously. It is also possible to exploit the fact that there is a linear dependence between the channel values and the amplitude parameter at given position and width parameters. Since the networks should return to the same position and width parameters from channel values that span a large dynamic range, it is sensible to get rid of this problem when trying to estimate position and width parameters by normalizing the channel values. Therefore when training networks to estimate the position and width parameters, the channel values of the samples (i.e., the inputs to the networks) were normalized so that the maximum channel value was unity on each sample. This was found to increase performance considerably. The networks were found to function best with around 20 hidden layer neurons. Each hidden layer neuron had 32 inputs, corresponding to the 32 measuring channels of the cameras.

A database on which the networks were trained was set up, and a separate database was used for testing the performance of the networks and to compare it to the performance of a conventional nonlinear curve fitting. This conventional algorithm consisted of calculating the error function described in Sec. II and minimizing this error as a function of the parameters using a conjugate gradient method for each sample separately. A substantial amount of noise was also added to the channel values of the samples to simulate realistic experimental conditions. This noise consisted of two

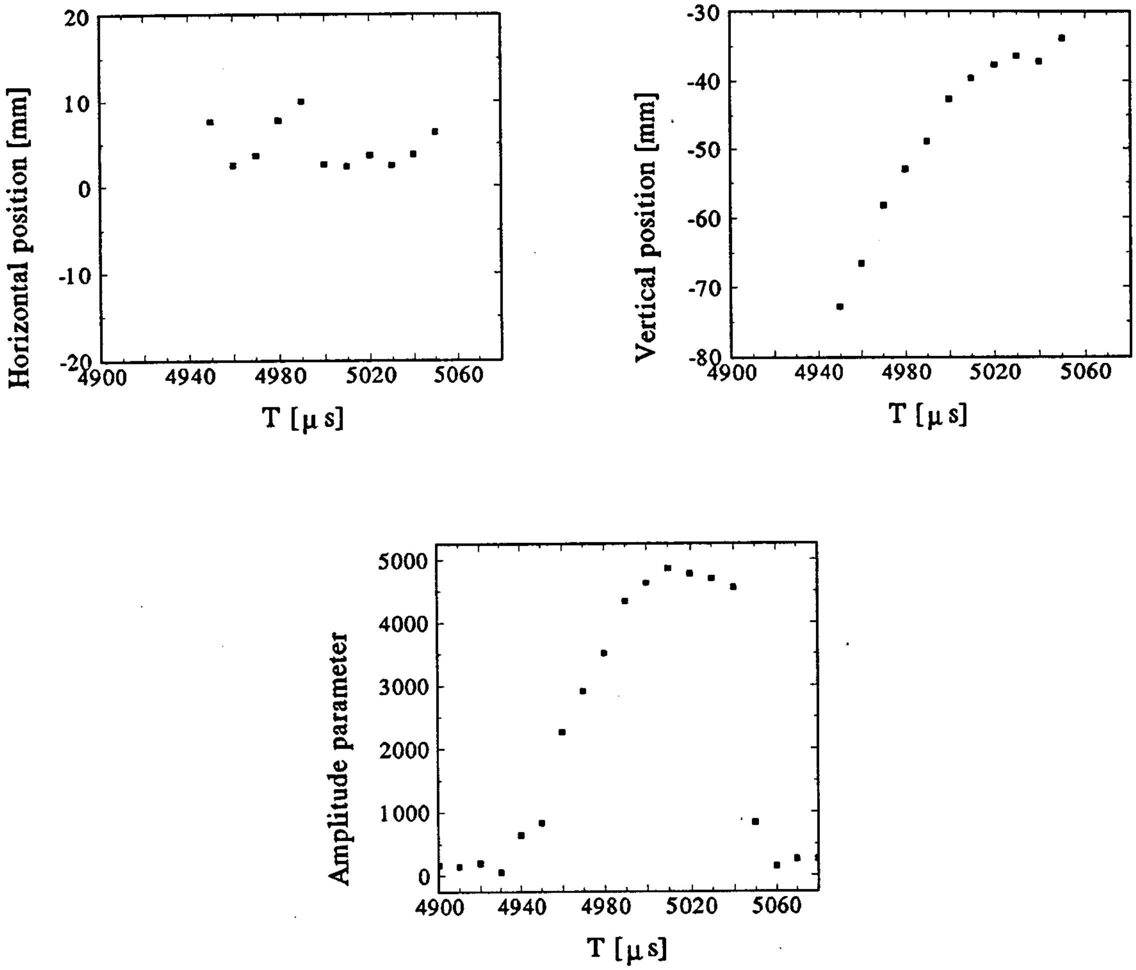

FIG. 5. Results of tomographic data processed by neural networks. The position parameters can be seen to correspond to a localized distribution moving into the plasma from below in the vertical direction, while the amplitude parameters correspond to a sudden increase in radiation from impurity ions.

components. One was a Gaussian distribution noise added to each channel value, whose standard deviation was 10% of that channel value. The other was a white noise, whose amplitude was 2% of the largest channel value in the sample. Results of the comparison can be seen in Figs. 3 and 4. The left-hand side figures show how well the neural networks estimated the generating parameters of the samples (the desired output), while the right-hand side figures show the same for a conventional nonlinear curve fitting. Standard deviations from the generating parameters can be seen at the top of each of the figures. It can be seen in the figures that the results for position parameters are slightly worse for the neural network estimate, while the results for amplitude parameters are practically the same for the conventional nonlinear fit and the neural networks. While the standard deviations for the width parameters are smaller for the neural networks, the structure of the error is different, as the relative error for narrow distributions is much larger for the networks. By changing the circumstances of learning, this can be changed, and it is possible to train the network, so that the relative error of the parameters is constant. The overall performance of the networks thus makes them suitable for fast measurement evaluation, and the values returned by the networks may be used later as a starting point for a conventional nonlinear fit if greater accuracy is desired, reducing time needed for convergence. It must be stressed, however, that while a conventional nonlinear fit may provide information on the applicability of our chosen test functions (the error at the end of the minimization may still be large indicating that the chosen test function does not describe the distribution adequetly), the neural networks have no such capability. The networks will provide an output whatever the input is, and if the measurements correspond to a different distribution, (e.g., one corresponding to several pellets at the same time) the output will provide false information.

The price to be paid for utilizing neural networks lies in training the networks. Training the Backpropagation Network involves a nasty nonlinear minimization involving a large number of parameters. This is, of course, extremely time consuming and training networks only pay off if there are large numbers of tomographic data to be evaluated. The computational load needed for training varies with the number of hidden layer neurons and training samples used. For the numerical minimization of the error function we tried the conjugate gradient, and the Broyden–Fletcher–Goldfarb–Shanno algorithm,[17] both of which were found to work well. A set of 2000–3000 samples to train the networks was found to give good results. The minimization was found to converge in a few hundred iterations. For 30 hidden layer neurons and 2000 training samples, the training required 4–6 days on an IBM PC compatible computer equipped with a 33 MHz 386 DX processor and with an arithmetic coprocessor. After training, processing of the samples by the neural networks is almost instantaneous. The nonlinear fitting of a single sample took around 1 h on the same machine. Therefore, in our setup the computational load of training the networks was equal to that of completing a few hundreds of conventional nonlinear fits. Since using the networks also involves some experimenting to which the architecture is most suitable for solving a given problem, it is wise to invest in neural network training only if the amount of tomographic data exceeds a few times that amount. It must also be mentioned, that a hardware implementation of these networks would be suitable for real-time data evaluation if desired.

Real experimental data was also processed by the networks. The results on a series of tomographic data involving a pellet injection into the plasma can be seen in Fig. 5. In these experiments, aluminium pellets were injected into the plasma and there were no filters in front of the cameras. From previous experiments we know that the signal comes mainly from the Al I,II,III ions, which are present in the plasma only in the vicinity of the pellet. It can be seen that the position parameters returned by the networks do indeed correspond to a pellet moving into the plasma from bellow, while the amplitude rises sharply as the pellet enters the plasma.

In conclusion, we can say that neural networks are suitable for fast processing of tomographic data, but it is worth investing in training such networks only if either there are large numbers of tomographic data to be processed (which is, however most often the case) or if real time evaluation is needed for some reason.

ACKNOWLEDGMENTS

The author would like to express his gratitude to his collegue, Sándor Zoletnik, who directed his attention to this interesting topic and has offered valuable advice throughout the work. This work was sponsored by the Hungarian Science Foundation (OTKA) under contract number 17489 and 14068.

[1] S. Zoletnik, S. Kálvin, and G. Bürger, Rev. Sci. Instrum. **65**, 426 (1994).
[2] S. Zoletnik, G. Kocsis, G. Bürger, P. N. Ignácz, B. Kardon, S. Kálvin, and J. S. Bakos, Rev. Sci. Instrum. **66**, 2904 (1995).
[3] J. S. Bakos, B. Kardon, S. Kálvin, G. Kocsis, G. Petravich, and S. Zoletnik, IEEE Trans. Plasma Sci. **24**, 29 (1996).
[4] S. Zoletnik, J. S. Bakos, B. Kardon, S. Kálvin, G. Kocsis, and G. Veres, 20th EPS Conference on Controlled Fusion and Plasma Physics, Lisbon, 1993, p. III-1171.
[5] S. Zoletnik, J. S. Bakos, P. N. Ignácz, B. Kardon, S Kálvin, G. Kocsis, J. Szigeti, and G. Veres, International Atomic Energy Agency, Technical Committee Meeting on Research Using Small Tokamaks, Würzburg, 1992, p. 86.
[6] S. Kálvin, J. S. Bakos, G. Bürger, B. Kardon, G. Petravich, and S. Zoletnik, Rev. Sci. Instrum. **60**, 2857 (1989).
[7] S. Zoletnik and S. Kálvin, Rev. Sci. Instrum. **64**, 1208 (1993).
[8] Robert Hecht-Nielsen, *Neurocomputing* (Addison-Wesley, New York, 1991).
[9] J. A. Freeman and D. M. Skapura, *Neural Networks* (Addison-Wesley, New York, 1991).
[10] Albert Bos, doctoral thesis, 1993.
[11] C. M. Bishop and C. M. Roach, AEA Fusion Report (1991).
[12] C. Bishop, I. Strachan, J. O'Rourke, G. Maddison, and P. Thomas, Neural Computing and Applications, v. 1. n. 1 (1992) Reconstruction of Tokamak Density Profiles using Feedforward Networks.
[13] J. T. Larsen, W. L. Morgan, and W. H. Goldstein, Rev. Sci. Instrum. **63**, 4775 (1992).
[14] C. Bishop, Neural Networks for Real-time Plasma Diagnostics, IAEA TCM on the Avoidance and Control of Disruption in Tokamaks, Culham Laboratory, (1991).
[15] C. M. Bishop, C. M. Roach, and M. G. von Hellermann, Plasma Phys. Control. Fusion **35**, 765 (1993).
[16] L. Allen and C. M. Bishop, Plasma Phys. Control. Fusion **34**, 1291 (1991).
[17] D. F. Shanno, Math. Op. Res. **3**, 244 (1978).