Plasma Phys. Control. Fusion **43** (2001) 405–416 www.iop.org/Journals/pm PII: S0741-3335(01)16536-X

A framework for the investigation of multiparametric dependences applied to total radiated power of JET plasmas

J Svensson[1,2], R Mohanti[2], K Lawson[2], M von Hellermann[3] and A Meigs[2]

[1] Department of Atomic and Molecular Physics, Royal Institute of Technology (KTH), Sweden
[2] UKAEA/Euratom Fusion Association, Culham Science Centre, Abingdon, OX14 3EA, UK
[3] FOM Institute for Plasma Physics, Rijnhuizen, 3430BE Nieuwegein, The Netherlands

Received 22 August 2000

Abstract

A framework is developed for investigating complex multivariate relationships in a dataset. This is based on using the universal approximation abilities of a multi-layer perceptron (MLP) neural network to predict a quantity of interest from a large set of parameters. A measure of redundancy is derived, and used in such a way that the average influence on the predicted quantity from any parameter can be estimated. Input parameters can be ordered in terms of increasing redundancy and therefore assist in finding the most important parameters a phenomenon of interest depends upon. In spite of the problem being multi-dimensional, the functional form of the one-to-one relationship between a parameter and a quantity of interest can be visualized. This framework is then used together with sensitivity analysis to investigate the dependence of the total radiated power of JET plasmas on a large number of parameters, leading to the identification of a much smaller set of parameters to be used in an effective MLP predictor of total radiated power.

1. Introduction

Phenomena that depend on a large number of parameters can be very difficult to analyse. The many inter-dependences between parameters can cause low correlation between a specific parameter and a quantity of interest even though the parameter is necessary for prediction of the quantity. We develop a method for statistically determining the relative importance of a large number of parameters influencing a plasma quantity of interest. This is based on creating a predictive model between a set of parameters and a quantity. This is described in section 2. In section 3 we discuss multivariate dependences and derive the measure to be used later for calculating redundancies. Section 4 brings the redundancy measure together with the predictive multi-layer perceptron (MLP) neural network model of section 2, exemplifying the method on two simple problems. In section 6 the framework is applied to the total radiated power of a set of JET plasmas and an effective predictor based on a small subset of the original parameters is modelled.

0741-3335/01/040405+12$30.00

Table 1. List of parameters used as inputs, together with short descriptions.

	No	Name	Description
Shape	1	gp1	See figure 4
	2	gp2	See figure 4
	3	gp3	See figure 4
	4	gp4	See figure 4
	5	gp5	See figure 4
	6	gp6	See figure 4
	7	gp7	See figure 4
	8	rig	See figure 4
	9	rog	See figure 4
	10	zup	See figure 4
	11	area	Area
	12	volm	Volume
	13	elon	Elongation
	14	tril	Lower triangularity
	15	triu	Upper triangularity
In/Out flux	16	dav	D_α vertical line of sight
	17	dah	D_α horizontal line of sight
	18	dao	D_α outer divertor view
	19	dai	D_α inner divertor view
	20	c3v	CIII vertical line of sight, visible
	21	c3h	CIII horizontal line of sight, visible
	22	c3o	CIII outer divertor view
	23	c3i	CIII inner divertor view
	24	c3	CIII horizontal line of sight, VUV
	25	c4	CIIV horizontal line of sight, VUV
	26	eler	Total electron count rate gas puffing
	27	majr	Deuterium puff rate
Divertor	28	rsol	Major radius outer strike point
	29	rxpl	Major radius X-point
	30	rsil	Major radius inner strike point
	31	zsol	Height outer strike point
	32	zsil	Height inner strike point
	33	zxpl	Height X-point
Density	34	nel3	Line integrated electron density
	35	nela	Line averaged electron density
	36	avl	Volume averaged density
	37	ax	Axial electron density
	38	yto	Total input power
Random	39	rand1	Random number 1
	40	rand2	Random number 2
	41	rand3	Random number 3

2. Predictive model

We look for a relationship between a quantity of interest, y (in the example in section 5, this is the total radiated power) and a number of other parameters, denoted by the vector $\boldsymbol{x}$ (when y is the total radiated power, $\boldsymbol{x}$ are the parameters in table 1). Such a relationship we define generally by

$$f(\boldsymbol{x}) = E[y|\boldsymbol{x}]\left(= \int yp(y|\boldsymbol{x})\,\mathrm{d}y\right) \tag{1}$$

where E denotes the conditional expectation value of y given a set of parameters $\boldsymbol{x}$, written out as an integral for clarity. To approximate the function f we will use a MLP neural network (Bishop 1995) of the following form:

$$f(\boldsymbol{x};\boldsymbol{w}) = \sum_{j=1}^{N_{\text{hid}}} w_j a_j(\boldsymbol{x}) + w_0 \tag{2a}$$

where

$$a_j(\boldsymbol{x}) = \left\{1 + \exp\left\{-\sum_{i=1}^{N_{\text{in}}} w_{ji}x_i + w_{j0}\right\}\right\}^{-1}. \tag{2b}$$

All w are free parameters to be determined during a training procedure described next. N_{in} is the number of inputs (the dimension of $\boldsymbol{x}$) and N_{hid} the number of basis functions $a_j(\boldsymbol{x})$ in (2a), denoted hidden neurons.

By minimizing the cost function

$$E = \sum_{p=1}^{N}(y^p - f(\boldsymbol{x}^p;\boldsymbol{w}))^2 \tag{3}$$

in the space of the free parameters $\boldsymbol{w}$ of (2), the mapping between $\boldsymbol{x}$ and y is determined. The sum in (3) is over a large set of examples of the mapping from $\boldsymbol{x}$ to y; p is the number of the example. Some precautions with respect to overfitting the function f has to be taken, but are not described here (see, for example, Bishop (1995)). The MLP function (2) is chosen for its ability to model high-dimensional dependences using few parameters $\boldsymbol{w}$, necessary to prevent overfitting when a limited dataset is used. It can be shown that a function of the form (2) can approximate any continuous function by just varying the number of terms in the inner summation (Hornik *et al* 1989, Ripley 1996). To test the final accuracy of model (1), a separate test set is put aside.

If model (1) can successfully predict the parameter of interest from the set $\boldsymbol{x}$ of input parameters, we can use this 'black-box' model to investigate aspects of the mapping from $\boldsymbol{x}$ to y, such as the relative importance of different inputs.

3. Multivariate dependences

If a parameter $y = f(x_1, x_2, x_3)$ is dependent on the three variables x_1, x_2 and x_3, a plot of observations of y against only one of the variables, say x_1, will give a graph with a spread in the y-direction that is caused by the variation of x_2 and x_3 at that x_1, such as in figure 1. We can describe the spread of y at a particular x_1 by a probability density

$$p(y|x_1). \tag{4}$$

The expectation value

$$E[y|x_1] \tag{5}$$

will give the average value of y at x_1. The uncertainty in y caused by trying to predict its value from only one variable x_1 can be measured by the variance of y given x_1:

$$\sigma^2[y|x_1]. \tag{6}$$

Turning the argument around, we could instead define a similar measure for the situation in which we know all variables *except* x_1. This would then give a measure of the extra variation of y caused by leaving that parameter (x_1) out:

$$\sigma^2[y|x_2, x_3]. \tag{7}$$

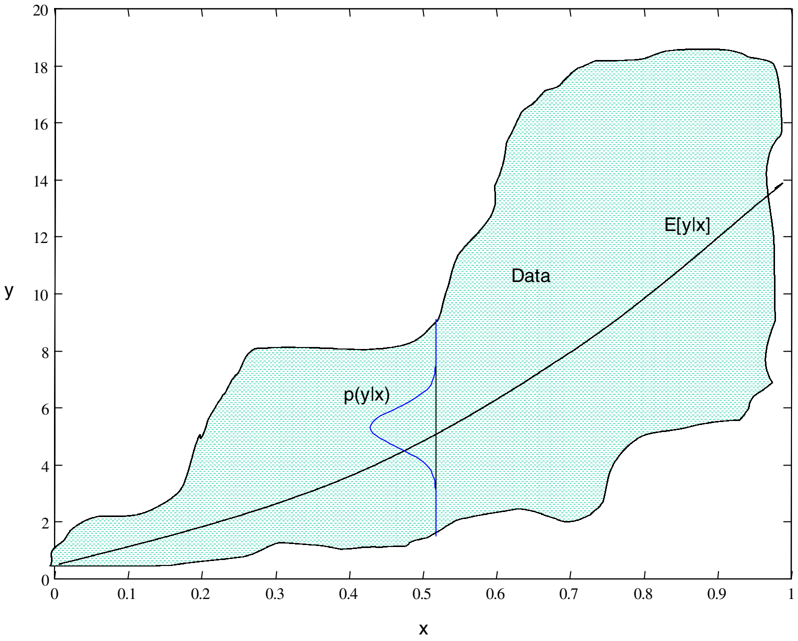

Figure 1. Illustration of form of scatter plot between x_1 and y with two variables (x_2 and x_3) missing.

This will be a local measure of the uncertainty of y at a specific x_2 and x_3, caused by lack of information about x_1. We can define a global measure by taking the expectation value of (7) over x_2 and x_3:

$$E_{x_2,x_3}[\sigma^2[y|x_2,x_3]]. \tag{8}$$

A more intuitive and dimensionless measure (even though not defined for $E[y|x_2,x_3] = 0$) can be defined by using instead of the variance, the standard deviation divided by the expected value of y:

$$R = E_{x_2,x_3}\left[\frac{\sigma[y|x_2,x_3]}{E[y|x_2,x_3]}\right] \tag{9}$$

We denote this measure by R. R gives the average relative variation of y around its average value when it is predicted from all parameters except x_1. To avoid the denominator being zero a sufficient bias in y can be added. This does not change the relative order of R for the different parameters (x_1, x_2, x_3, etc). We suggest that this is a good measure of how redundant x_1 is in predicting y, and we therefore refer to this value as the redundancy (of x_1). A low value of R will imply high redundancy, since small variation of y from leaving x_1 out implies that x_1 does not participate much in the prediction of y.

Another important measure (Bishop 1995) that will be used later is the influence small changes in one parameter, for example x_1, will have on y. This is measured by the partial derivatives

$$S = \frac{\partial f}{\partial x_1} \tag{10}$$

of the function with respect to one of the parameters.

The schemes for calculating redundancies and sensitivities are based on knowing the function $y = f(x_1, x_2, x_3)$ that maps to y. In the next section we approximate this function by an MLP neural network, after which (9) and (10) can be calculated following the procedure described in section 4.

4. Calculating redundancies and sensitivities without knowledge of the underlying model

4.1. Method

This section describes a procedure based on the previous two sections, to deal with the following problem:

> If all we have is a set of measurements, can we know whether one of the parameters, y, can be predicted from a set of other measurements (x_1, x_2, x_3), and if so, what the relative importance of the different parameters (x_1, x_2, x_3) is in predicting y.

As explained in section 2, the relationship between (x_1, x_2, x_3) and y can be modelled by an MLP neural network. Section 3 gives a recipe for calculating the redundancies of and sensitivities to the different inputs of the model. The sensitivities (10) are easily calculated by differentiating the MLP function (2) with respect to the inputs (Svensson *et al* 1999), yielding

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{M} w_j a_j (1 - a_j) w_{ji}. \tag{11}$$

The redundancies (9) demand a little more work. The integrals in (9) correspond to very complex analytical expressions, and so (9) is more easily approximated by a Monte Carlo integration over the range of the input parameters. This is done by sampling vectors of (x_2, x_3), and for each such sample, varying the parameter x_1 over its allowed range, calculating the output y for each step of x_1. It would be more accurate to sample x_1 from the distribution $p(x_1|x_2, x_3)$, but we have chosen to instead vary x_1 stepwise over its range, since we are often interested in the influence of a uniform variation of one parameter, rather than the variation dictated by the particular range of experiments performed when the training dataset was assembled. To assure that unphysical parameter regions of (x_1, x_2, x_3) are not visited during the Monte Carlo calculations, the probability density in the space of the inputs is first calculated using a mixture of Gaussians (Svensson *et al* 1999):

$$p(\boldsymbol{x}) = \sum_{j=1}^{M} \frac{P_j}{(2\pi)^{d/2} |\boldsymbol{S}_j|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}_j)^T \boldsymbol{S}_j^{-1} (\boldsymbol{x} - \boldsymbol{x}_j) \right). \tag{12}$$

That is, the probability density of $\boldsymbol{x}$ is described as a sum of M normal distributions with a dimensionality d equal to the dimension of $\boldsymbol{x}$. Each such Gaussian has covariance matrix $\boldsymbol{S}_j$, centre $\boldsymbol{x}_j$ and is weighted by a prior P_j. These parameters are determined from the data.

For each sample, a check is then made that the probability density of the sample is higher than a lower limit for the density in the training set. This is similar to the novelty detection mechanism in Bishop (1994). If the probability density of the sample is too low, that point is rejected. From the non-rejected samples, the quantities $E[y|x_2, x_3]$, $\sigma[y|x_2, x_3]$ and $E_{x_2,x_3}[\sigma[y|x_2, x_3]/E[y|x_2, x_3]]$ can be calculated.

We now demonstrate this scheme on two simple problems before proceeding to the analysis of the plasma radiated power.

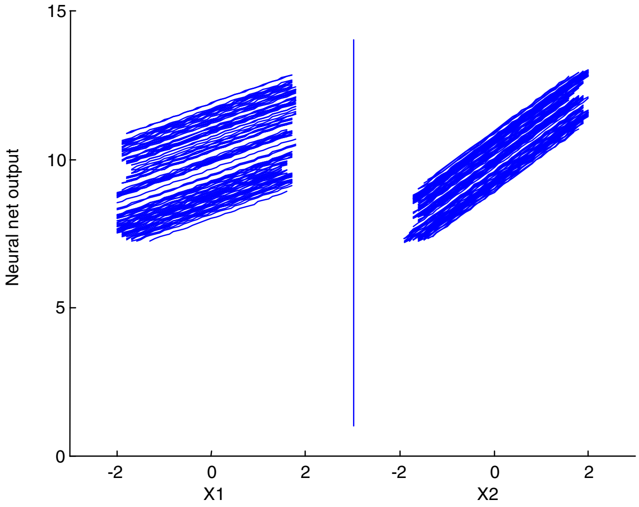

Figure 2. Trajectories on MLP surface when one variable is varied and the other held at constant values. No correlations between x_1 and x_2.

4.2. Examples

An MLP network was trained from a set of examples of the following mapping:

$$y = x_1 + 2x_2 + 10. \tag{13}$$

and were sampled independently from two identical uniform distributions. The calculation of S and R by the methods in section 4.1 for x_1 and x_2 gave the following values.

	S	R
x_1	1.0	6%
x_2	2.0	12%

The partial derivatives S were approximated accurately by the neural network, and the redundancies R are of double importance to x_2. The percentages should be interpreted as the variation of the output that is caused by the respective input. Because of linearity and independence, the sensitivities and redundancy measures give similar results. A visualization of the dependences can now be made by following trajectories of the neural network function where only one input is varied over its range (figure 2) for different values of the other parameter, thus yielding curves that show clear relationships between two variables rather than the spread of points as in figure 1. Recall that for this example, all we have is a set of cases of the relation (13), not the actual function.

In the second example, a correlation is introduced between x_1 and x_2, and a third independent variable x_3 is added:

$$y = x_1 + 2x_2 + x_3 + 20. \tag{14}$$

All variables are sampled from normal distributions with mean zero and unit standard deviation. A 90% correlation is introduced between x_1 and x_2. Also, a fourth totally random variable x_4 is added to the inputs of the neural network. In this case the sensitivities and redundancies are as follows.

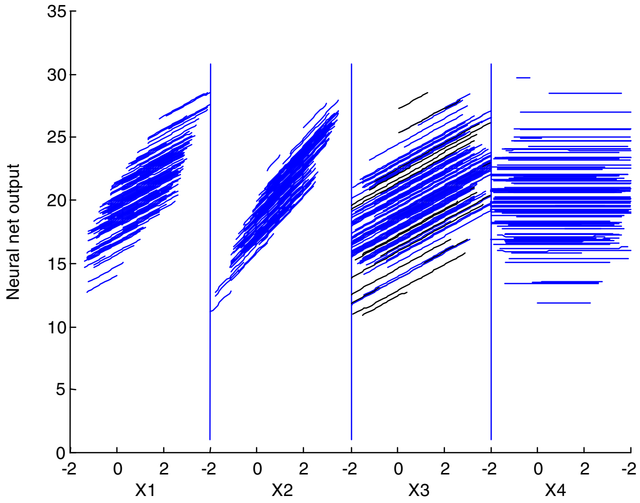

Figure 3. Trajectories on MLP surface when one variable is varied and the other held at constant values. Correlations between x_1 and x_2.

	S	R
x_1	1.0	3.8%
x_2	2.0	7.5%
x_3	1.0	7.2%
x_4	0.002	0.01%

The S column again just gives the partial derivatives of the MLP approximation of (14). From the R values it can be seen that x_2 still changes y twice as much as x_1, while x_3, even though it has the same partial derivative as x_1, has a lower redundancy value. This is because information in x_1 can also be found in x_2 since they are correlated. In this case, therefore, x_1 is the most redundant of the parameters, after the random input x_4. So, even though x_1 and x_3 have the same partial derivatives, it can be seen from the R values that if we had to leave one parameter out, the best choice would be x_1 (after the random parameter x_4). The trajectories for the four variables on the MLP surface are shown in figure 3. As can be seen, the unrelated parameter x_4 will give horizontal trajectories, x_2 will give have twice the slope of x_1 and x_3, and because of the relationship between x_1 and x_2, the length of the trajectories are shorter (the rest of the parameters are kept at constant values when one parameter is varied over its allowed range: relationships between variables will therefore give shorter trajectories). Similar graphs will be used for the total radiated power in section 5.

5. Application to total radiated power

We start by modelling the total radiated power from the 38 parameters in table 1, after which the sensitivities and redundancies are calculated as described in section 4. This will then form the basis for reducing the number of parameters to six, giving a new and much simpler model. As a validation of the whole method we have used three random numbers as extra inputs, increasing the number of total input parameters to 41. Short explanations of the parameters can be found

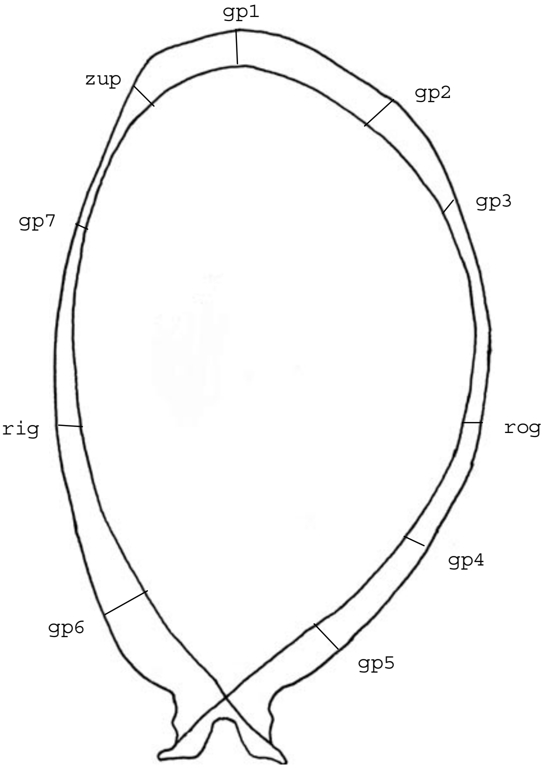

Figure 4. Plasma shape parameters.

in table 1 and in accompanying figure 4. The parameters are divided into different groups for plasma shape, in/out flux, divertor and density. The in/out flux group contain parameters only *related* to the fluxes, since intensities are used rather than calibrated densities.

An MLP network with 41 inputs and 20 hidden units was trained on a dataset consisting of 2192 examples. The examples were collected from mostly ELMy H-mode pulses from the MarkIIAP divertor configuration (Horton *et al* 1999). The final performance of the predictor was tested on a separate test set not used for the training. The test set contained data from 20 JET pulses collected at a later time than the data used in the training set. The overall error on this test set was 17%, giving a quite accurate predictor of the total radiated power. A redundancy and sensitivity analysis as described above was then done, the results of which are shown in figure 5. As can be seen, the redundancy analysis and the sensitivity analysis do not give the same results, as is expected. Sensitivity measures how sensitive the total radiated power is on average to small changes in a parameter, and redundancy is a measure of how much that parameter contributes to the prediction. We can thus see, for example, that most of the parameters describing the shape of the plasma are very important (since they determine the proximity to the wall, and therefore the influx of impurities), but are also highly redundant, since they will all be interrelated. If we were to pick the most important parameters we would not necessarily choose the ones with highest sensitivities, since those could be redundant because of interrelationships (so we might not need them all). Figure 6 shows another view of the same data, where the sensitivity and redundancy are plotted against each other. The interesting parameters should be sought for mainly in the upper right corner of the distribution of input

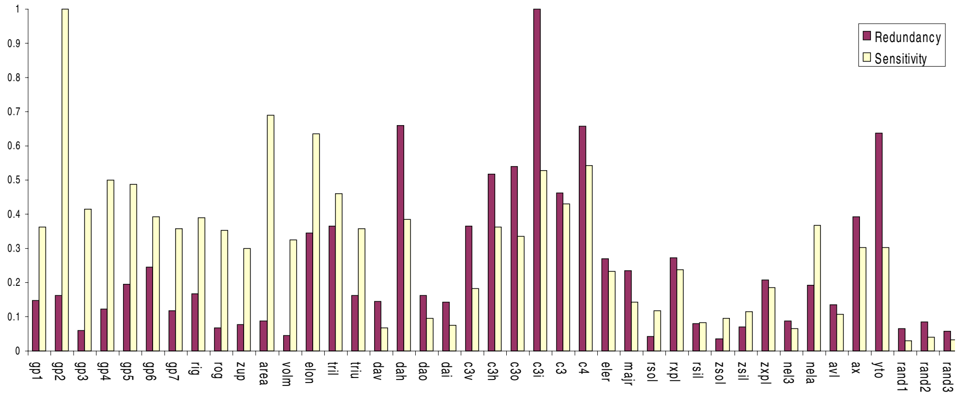

Figure 5. Redundancies and sensitivities for all parameters in table 1. Both sets are scaled so that the maximum value is 1.

Table 2. Parameters for the second MLP predictor.

No	Name	Description
1	Elon	Elongation
2	tril	Lower triangularity
3	triu	Upper triangularity
4	c4	CIV horizontal line of sight
5	nela	Line averaged electron density
6	yto	Total input power

parameters. Those will have both a high sensitivity and low redundancy. The parameters in the lower left corner will have very little relevance for the prediction of total radiated power. In this region we find the three random values, thus verifying the applicability of the model, and most divertor parameters. If we produce similar plots as in figures 2 and 3 for a subset of the parameters, we will be able to visualize one-to-one relationships between total radiated power and different plasma parameters (for different sets of the remaining parameters). Figure 7 shows this for some of the parameters in table 1.

When we picked a subset of the 41 parameters to create a new predictor, we looked at both the trajectories for different parameters, and the S/R plot in figure 6. Parameters in the lower left corner of the S/R plot for example were not considered. On the other hand, parameters in the upper right corner and parameters that had trajectories that showed a clear and simple relationship to the radiated power, such as c4, were chosen over parameters that had more complex dependences, such as dah. This gave us the parameters listed in table 2. A new MLP network trained with only those inputs was then trained and gave 18% error on the same 20 pulse test set as used above. The loss in precision due to exclusion of 32 parameters was thus only marginal. It should be pointed out that the choice of a combination of six parameters from a set of 38 parameters can be done in principle in $38 \times 37 \times 36 \times 35 \times 34 \times 33$ ways, so methods such as those developed here are necessary to hint at what variables are relevant. Figure 8 shows a plot of the total radiated power and the total radiated power predicted from those six parameters for the 20 pulses in the test set. As can be seen, the correspondence is remarkably good, except for the absolute value of the large spikes, where the error is somewhat larger. Figure 9 shows the correlation plots of each of the six parameters and the radiated power. As can be seen, the method can very well choose parameters that have not too high correlation

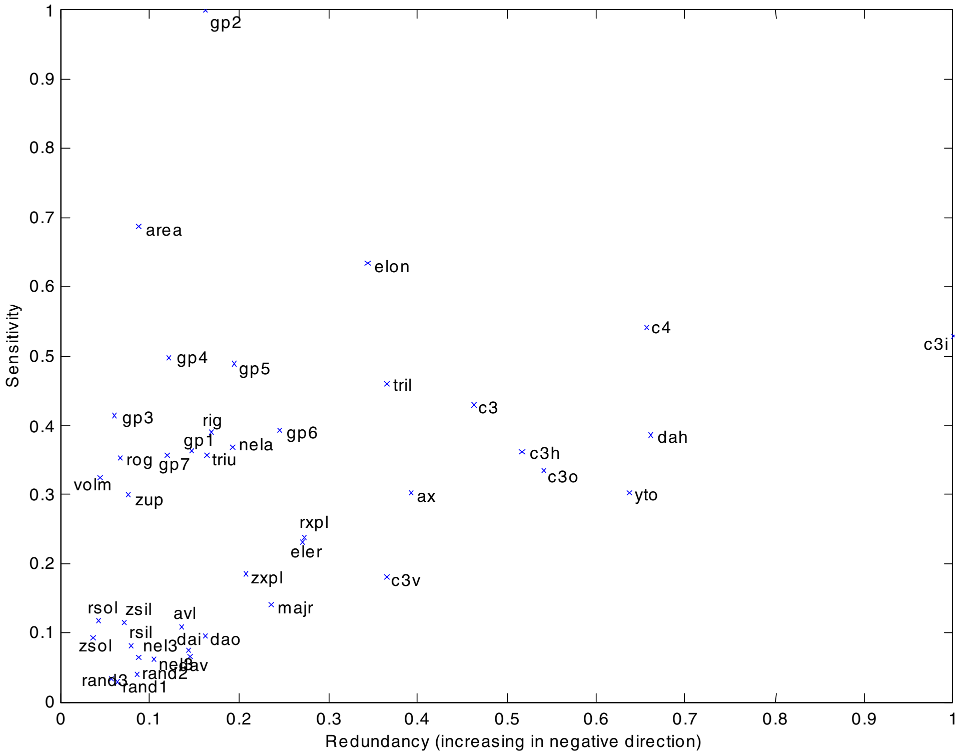

Figure 6. Scatter plot of sensitivity versus redundancy for the parameters in table 1. Both sets are scaled so that the maximum value is 1.

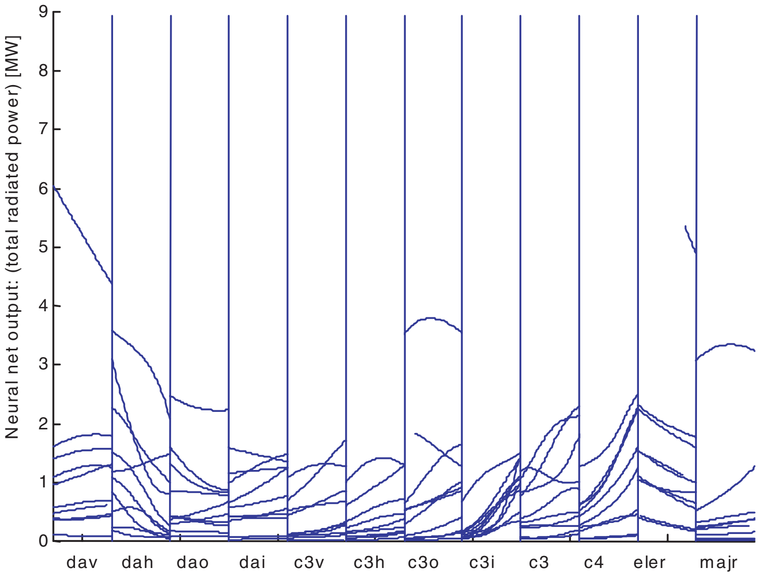

Figure 7. Trajectories on the MLP for the variation of some of the parameters in table 1.

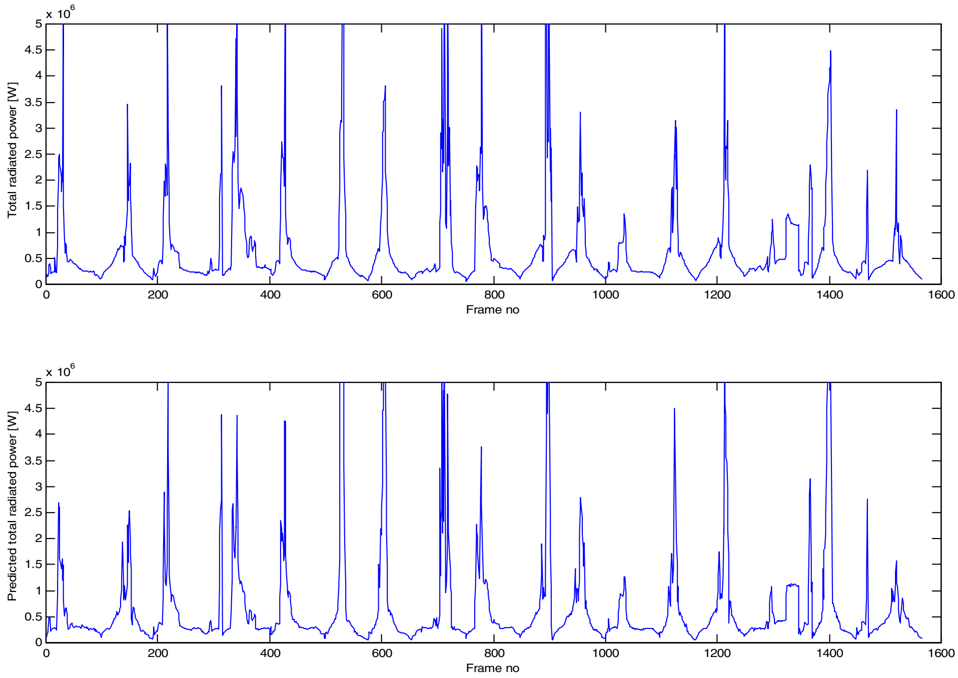

Figure 8. Above: time evolution of the total radiated power for all 20 pulses in the test set, plotted one after the other. Below: the 6-input MLP network predicting the same quantity from the parameters in table 2.

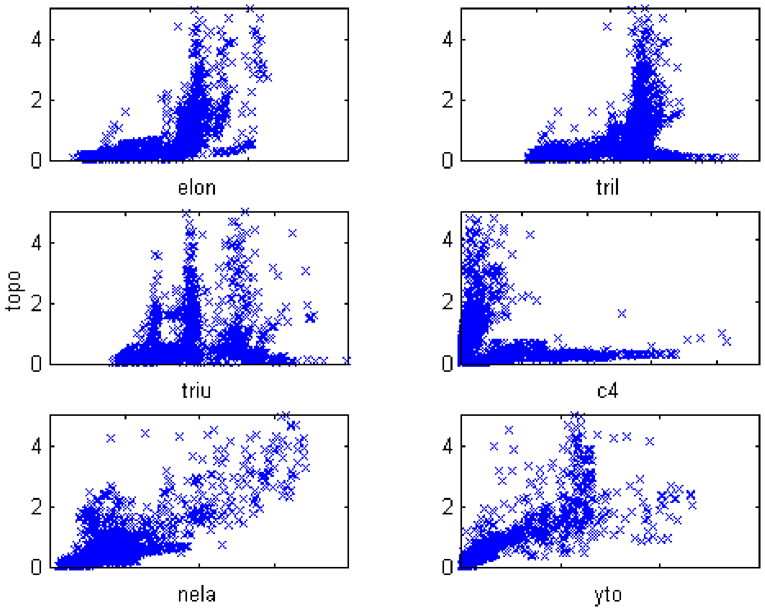

Figure 9. Correlation plots for the six parameters chosen as inputs for the second MLP network.

with the total radiated power, since dependence does not necessarily imply correlation. In fact, training a network on the six highest correlated variables gave 23% error, and on the six highest sensitivities 27% error. The method we have presented thus outperforms these approaches.

6. Conclusions

We have introduced a framework for analysing multivariate dependences based on MLP neural networks. A measure has been introduced that can be used together with a sensitivity analysis over the MLP network to deduce the importance of different parameters in determining a quantity of interest. This is done without any explicit physical model of the quantity, and can be used to look for relevant parameters for different phenomena in a much stronger way than could be achieved with a standard correlation analysis. The framework has been tested on the task of prediction of total radiated power from 38 plasma parameters. A subset of significant variables could be identified, resulting in an effective predictor based on only six of those parameters. Apart from a working predictor of total radiated power, the method will list the relative redundancies, and therefore importance, of the parameters used for 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

Bishop C M 1994 Novelty detection and neural network validation *IEE Proc. Vision, Image and Signal Process* **141** 217–22

——1995 *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press)

Hornik K, Stinchcombe M and White H 1989 Multilayer feed-forward networks are universal approximators *Neural Networks* **2** 359–66

Horton L D *et al* 1999 Studies in JET divertors of varied geometry I: non-seeded plasma operation *Nucl. Fusion* **39** 1

Ripley B D 1996 *Pattern Recognition and Neural Networks* (Cambridge: Cambridge University Press)

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