RECEIVED
NOV 15 1994
T I

INSTITUTE FOR FUSION STUDIES

DE-FG05-80ET-53088-676

IFSR #676

Neural Net Prediction of Tokamak Plasma Disruptions

J.V. HERNANDEZ, A. VANNUCCI,^{a)} T. TAJIMA,^{b)} Z. LIN,
W. HORTON, and S.M. MCCOOL^{c)}

Institute for Fusion Studies
The University of Texas at Austin
Austin, Texas 78712

^{a)}Instituto de Fisica Universidade de Sao Paulo, Sao Paulo, Brazil

^{b)} also at Advanced Science Research Center, JAERI, Ibaraki, Japan

^{c)} Fusion Research Center, The University of Texas at Austin, Austin, Texas 78712

October 1994

THE UNIVERSITY OF TEXAS

AUSTIN

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Neural net prediction of tokamak plasma disruptions

J.V. Hernandez, A. Vannucci,* T. Tajima,[†] Z. Lin,
W. Horton, and S.C. McCool[‡]

*Institute for Fusion Studies, The University of Texas at Austin
Austin, Texas 78712*

Abstract

The computation based on neural net algorithms in predicting minor and major disruptions in TEXT tokamak discharges has been performed. Future values of the fluctuating magnetic signal are predicted based on L past values of the magnetic fluctuation signal, measured by a single Mirnov coil. The time step used ($= 0.04$ ms) corresponds to the experimental data sampling rate. Two kinds of approaches are adopted for the task, the contiguous future prediction and the multi-timescale prediction. Results are shown for comparison. Both networks are trained through the back-propagation algorithm with inertial terms. The degree of this success indicates that the magnetic fluctuations associated with tokamak disruptions may be characterized by a relatively low-dimensional dynamical system.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

*Instituto de Fisica Universidade de Sao Paulo, Sao Paulo, C.P. 20516-S.P.-01498, Brazil

[†]also at Advanced Science Research Center, Japan Atomic Energy Research Institute, Ibaraki, Japan

[‡]Fusion Research Center, The University of Texas at Austin, Austin, Texas 78712 U.S.A.

I. INTRODUCTION

Disruptions in tokamak plasmas have been, for decades, a concern among the fusion research scientists. Commonly, these harmful instabilities have been classified as major and minor disruptions, according to their potential of completely or partially destroying the plasma current. Due to the fast decay of the plasma current following a major disruption, strong electromagnetic forces are produced that stress the whole tokamak structure. In large devices such as JET (Joint European Torus), displacements of the whole mechanical structure by several millimeters have been already reported.¹ For the future fusion reactors such as ITER (International Thermonuclear Experimental Reactor), requiring very high plasma currents be dealt with, mechanisms that can possibly avoid or even attenuate the occurrence of major disruptions will be important.

Many investigations, both theoretical and experimental, have already been carried out in an attempt to better understand the physical mechanisms that are responsible for triggering plasma disruptions.²⁻⁹ Magnetic islands contacting the limiter,^{2,3} MHD mode coupling⁴⁻⁷ and ergodization of field lines within magnetic islands,⁸⁻⁹ for example, are some of the explanations that were already explored. Experiments based on the application of resonant¹⁰ and nonresonant¹¹ magnetic fields, in an attempt to control the occurrence of major disruptions, have been also performed but only with some partial success.

In this paper, a different approach in how to possibly avoid the occurrence of fast disruptions, by means of using an adaptable algorithmic strategy called a neural network¹² to predict in advance the instant of time a disruption will occur, is being proposed. Since the $m = 2$ MHD mode, as it is widely accepted, plays an important role in triggering the disruptive instabilities, the fluctuating magnetic field signal from one of the Mirnov coils installed in TEXT (Texas Experimental Tokamak) tokamak was considered the appropriate experi-

mental signal to be used as input to the program. Although neural nets have already been used for fitting plasma parameters and general data analysis,¹²⁻¹⁶ we believe that this is the first time that a neural network is used for predicting in advance the moment a disruption will take place.

II. NEURAL NETWORKS

In the presence of dissipation, the evolution of a high-dimensional dynamical system may become trapped into a low-dimensional region of the phase space, an attractor. If that is the case, then the underlying state of the system can be reconstructed from the observation of a time series $x(t)$ using time-delayed vectors,^{12,17,18,19}

$$X_t = (x_t, x_{t-1}, \dots, x_{t-(n-1)}) , \quad (1)$$

as long as the dimension $n > d$, where d is the dimension of the attractor, and $x_{t-j} \equiv x(t-j\tau)$ where τ is the time delay. This implies that there exists a smooth nonlinear function Ψ which relates n past measurements of $x(t)$ to a future value of $x(t)$ ^{12,17,18,19}:

$$x(t+1) = \Psi [x_t, x_{t-1}, \dots, x_{t-(n-1)}] . \quad (2)$$

In order to predict the future evolution of the system, it is necessary to find some way to approximate Ψ .

For this work the TEXT magnetic fluctuation data are used in feed-forward perceptron-type neural networks with nonlinear activation functions²⁰ to approximate the smooth function Ψ and to perform the prediction task. Neural nets were originally used in a prediction context by Lapedes and Farber.²¹

The reason for using neural nets with at least one hidden layer and nonlinear activation functions is that these networks can uniformly approximate any continuous function, provided enough hidden units are used.²¹ The generic architecture of a feed-forward neural

net contains L inputs, one or several hidden layers with one or several units each, and one output layer. A unit in one layer is fully connected with the units in the previous layer, and the output $h_i^{(k)}$ (unit i in the k -th layer) is a function $g_i^{(k)}$ (the activation function) of the weighted sum of the outputs of all the units in the previous layer:

$$h_i^{(k)} = g_i^{(k)} \left(\sum_{j=1}^N W_{ij}^{(k-1)} h_j^{(k-1)} \right) \quad (3)$$

where $W_{ij}^{(k-1)}$ is the strength of the connection between unit j in the $(k-1)$ -th layer and unit i in the k -th layer.

If the activation function $g(x)$ is a nonlinear saturating function, such as $\tanh(x)$, the network should be able to approximate a nonlinear function Ψ , provided the appropriate values of the weights $\{W_{ij}\}$ can be found, that is, provided the network is properly trained.

Supervised feed-forward networks learn by example. The weights between the connections are determined by presenting a set of actual input-output values (the training set) and comparing, through some error or cost function (W), the output of the net $\hat{O}$ with the actual output O from the system. The supervised learning approach has to be contrasted with the traditional expert system approach where the performance of the system depends on the previous while often incomplete knowledge of the program by the programmer. For this work the cost function was chosen to be the mean square error:

$$\mathcal{E}(W) \equiv \frac{1}{M} \sum_{j=1}^M (O_j - \hat{O}_j)^2, \quad (4)$$

where M is the number of input-output pairs of training patterns and the error is minimized by adjusting the weights according to the back propagation algorithm,²³ which is a gradient descent recipe, and including an inertial term to accelerate convergence. In particular

$$W_{ij}^{\text{new}} = W_{ij}^{\text{old}} + \Delta W_{ij}, \quad \Delta W_{ij} = -\eta \frac{\partial \mathcal{E}}{\partial W_{ij}} + \alpha \Delta W_{ij}, \quad (5)$$

with η the learning rate, being a small positive constant ($0 < \eta \ll 1$) and with α the inertial term, a number between 0 and 1.

In order to characterize the goodness of the model, the ARV (average relative prediction error variation) given by

$$ARV = \frac{\sum_{j=1}^M (O_j - \hat{O}_j)^2}{\sum_{j=1}^M (O_j - \langle O_j \rangle)^2}, \quad (6)$$

and the correlation coefficient ρ given by

$$\rho = \frac{\sum_{j=1}^M (O_j - \langle O \rangle)(\hat{O}_j - \langle \hat{O} \rangle)}{\sigma_O \sigma_{\hat{O}}}, \quad (7)$$

are introduced. In these equations, $\langle O \rangle$ is the mean value of the set of M actual outputs, σ_O and $\sigma_{\hat{O}}$ are the standard deviations of the actual and the predicted outputs, respectively.

III. PREDICTING DISRUPTIONS IN THE TEXT TOKAMAK

For the exploration of a method to predict disruptions in the TEXT tokamak, two approaches are tried here. One is a supervised feed-forward neural net which consists of 15 units in the input layer corresponding to the L past values of the fluctuating field $\tilde{B}_{t-j}$ ($j = 1, \dots, L$), and has two hidden layers (the first with nine units and the second with three) and one output unit for the predicted $\tilde{B}_{t+L}$. The activation function for the hidden units are nonlinear, $g(x) = \tanh(x)$, and the activation function for the output unit is linear, $g(x) = x$, in order to have a wide dynamical range. The number of hidden units was chosen by monitoring the out-of-sample performance of the network. The learning rate and the inertial term are $\eta = 0.001$ and $\alpha = 0.9$, respectively.

The training of the network was carried out by using two TEXT plasma discharges, one disruptive and the other nondisruptive. In order to stop the training of the net, we monitored its performance on a third (nondisruptive) shot (the validation set) so that the prediction error reached a minimum when evaluated over the validation set. Finally, the trained network was used on an independent (disruptive) shot: the prediction set.

In this work the neural net was used to perform both single and multistep predictions. In single step predictions, actual past values of the time derivative of the fluctuating poloidal magnetic field were used to predict its next value with a time step $\tau = 0.04\text{ms}$, which corresponded to the sampling rate of the experimental data. In multistep predictions, the predicted output at time 1 is fed-back into the input and it is used to predict a new output, at time 2, and so on.

The other approach is a supervised feed-forward neural network consisting of one hidden layer with 8 hidden units, and one output unit. The activation function is $g(x) = \tanh(x)$ for all connections. The number of the input units will be decided by the requirements of detailed prediction problems. Through monitoring the performance of the network, the learning rate and the inertia term are set to be 0.02 and 0.2 respectively in this approach. The training of the network is done by using one TEXT tokamak data shot including both the general oscillating part and the disruption part. The number of training epochs is of the order of 10^6 . For every 5000 times, a resulting set of weights are recorded and the value of the ARV and the correlation coefficient are calculated as a measurement of the performance of the network. The final resulting weights are chosen to be ones with a global minimum of ARV and maximum of the correlation function. The real prediction work is done when this optimum weights set is applied onto other data shots. Another scheme for multistep prediction is developed here. For example, for a L -step prediction, we will use the magnetic field data before time t as input source and take the data $\tilde{B}_{t+L}$ as the target. Both the playing back onto the training data shot itself and the result of the weights applied to other data shots are shown as follows.

IV. RESULTS OBTAINED

A. First approach — Contiguous future prediction

After the best architecture has been found and all the weights have been properly set, the neural network is used in forecasting disruptions in a different plasma pulse other than the ones used for training. This is done for the discharge shown in Fig. 1. It corresponds to a $I_p \sim 180$ kA shot which disrupted at $t \sim 290$ ms (Fig. 1a) and was preceded by a minor disruption, approximately 8 ms earlier. After the major disruption has taken place, the plasma position control system was no longer able to maintain the plasma column centered (Fig. 1c) and several minor disruptions occurred while the plasma current decayed. In Fig. 1d the derivative of the fluctuating poloidal magnetic field from one of the Mirnov coils is shown, which was used for feeding the neural network.

The predicted $\tilde{B}_\theta(t+n\tau)$ values by the neural network for the last 200 ms of the discharge are shown in Fig. 2. In Fig. 2a, the experimental data is plotted and can be compared to results obtained for the single ($n = 1$) and multistep ($n = 15, 21, 28$) predictions, presented in Figs. 2b and 2c, 2d, 2e. These values of n correspond to predictions made 0.04 ms and 0.6, 0.84 and 1.12 ms ahead in time, respectively. Expansions of these signals in regions before the disruption time, as shown in Figs. 3, indicate that the predictions done by the neural net are consistent, even for higher values of n . In Figs. 4a and 4b the corresponding prediction error and the correlation coefficient, as calculated from equations (5) and (6), for a different number of iterations are shown. As it is noted, the performance of the neural network, in determining the future values of the perturbed poloidal magnetic field, tends to degrade exponentially, as the number of iterations increases.

In Fig. 5, expansions of the signals shown in Fig. 2 are plotted within the disruption time region. Although there are some perceptible discrepancies between the wave forms of the experimental data (Fig. 5a) and the results of the prediction (Figs. 5b–5e), the occurrence

of both the minor disruption (at $t \sim 183$ ms) and the major disruption (at $t \sim 191$ ms) are clearly identified by the neural network, even for the cases corresponding to higher numbers of iterations. It is interesting to note, on the oscillatory portions of the output results from the neural net, that the signal frequency tends to differ considerably from the actual data, as the number of iterations increases. This behavior probably can be understood as due to the typical frequency decrease of the Mirnov oscillations during the period of time that preceded a disruption. In TEXT-U, for example, close to a disruption, this frequency change varies from shot to shot, usually decreasing from about 12 kHz to 1-2 kHz; but in many cases reaching values of only some tens of Hz.²³

B. Second approach — Multi-timescale prediction

In the second approach, for a simple short-term prediction, the number of input units is set to be 5, organized as follows:

$$\ddot{B}(t), \dot{B}(t - \Delta t), \dots, \dot{B}(t - 4\Delta t) .$$

The target output is $\ddot{B}(t + 5\Delta t)$ for the goal of 5-step prediction. We processed four data shots (ss1-ss4) in this way and four weights sets (weights1-weights4) were properly obtained accordingly. We apply these weights sets to data shots (SS1-SS4) alternatively and the obtained ARV and correlation coefficient is so given in the Table 1.

As the time resolution of the data shots is $\Delta t = 0.04$ ms, we achieved the 0.2 ms prediction by the 5-step model. From the table, we can see that the performance of Weights1 on ss3 and vice versa are better. Figure 6 shows that the playing back of weights2 on ss4 reproduces the shape of ss4 to its best extent. By inspection of the original data ss1-ss4, we find that ss1 and ss3 are similar to each other, while ss2 and ss4 are similar to each other. As expected, when the training data is similar to the forecasted one, the neural network is afforded to master the essential characteristics more accurately and will do a better job.

Furthermore, we attempt to make a long-term prediction. Following the same procedure as described above, we substitute the target output to be $\dot{\tilde{B}}(t + 25\Delta t)$ to obtain a 25-step prediction. But this leads to a failure with the appearance of a certain phase shift in oscillatory portion between the original data and the predictive one. The ARV now is up to 1 even for the playing back on the training data itself. In this situation, we surmise that the data shot, even long before the disruption time has more than one characteristic frequency. When we do the 5-step prediction, the prediction time is not long enough to cover one time period of the data shot, so that no out-of-phase phenomenon occurs, while it does in the 25-step prediction. This shows that our single frequency input organization can no longer accommodate the multifrequency (wide-band) data.

Solving this problem, we use Fourier analysis to figure out if there are any characteristic frequencies in the data shot and what they are. We first deal with ss4, which is shown to be the most difficult one to be modeled. The Fourier transform of ss4 is shown in Fig. 7. The index n (value of x -axis) is related to the characteristic frequency we are searching for, $f = \frac{n}{2N\Delta t}$, where N is the total number of data points processed. In the case for ss4 ($N = 6000$) the indices n of three peaks are 1103, 1316, and 1590, respectively; these correspond to the frequencies of 2298Hz, 2741Hz, and 3311Hz, accordingly. For convenience, we count the time period in step units, and these correspond to 10.88, 9.12, and 7.55 steps, in respectively. Considering this feature of the frequency spectra, we introduce another sampling period with another five inputs in addition to the former five as:

$$\dot{\tilde{B}}(t - 6\Delta t), \dot{\tilde{B}}(t - 2 \times 6\Delta t), \dots, \dot{\tilde{B}}(t - 5 \times 6\Delta t) .$$

We choose 6 because the maximum time period is 10.88 steps and 6 will cover the half period. With this introduction of multiple sampling periods all the major characteristic frequencies of this system's behaviors should be mastered by our model.

Training the network with this new input organization, we derive a new set of weights

with SS4. Playing back on ss4 with multiperiods shows much improvement over the former performance with a single sampling period. The ARV decreases to 0.39, comparing to the unacceptable level of order of unity 1 before. With detailed comparison of the general oscillatory portion, the phase shift appears to have vanished (Fig. 8). In the disruptive part, the phase shift is also reduced in comparison with the case with a single sampling period (Fig. 9). Fourier analysis of the ss1-ss3 is shown in Table 2 and the same procedure is used for ss1 and ss3. The result shows the multifrequency model actually works for the complicated disruptive data shot.

After several attempts, 5 is chosen to be the most convenient number of steps in our two-frequencies model. Training the network with ss1, the improvement is obvious. With playing back onto ss1 itself, the ARV is now 0.13 and correlation coefficient 0.92, even better than the former 5-step prediction result. The result of applying weights1 onto ss3 is also promising (Fig. 10). The phase shift is absent and the ARV and correlation coefficient accordingly is 0.27 and 0.82.

Given these outcomes, we can conclude that at least the phase shift problem in disruption prediction can be partly solved with our new input organization. Now more data shots are processed to implement this algorithm.

An attempt to feed the actual neural net program with data recently taken of the perturbed magnetic fields measured by an internal Mirnov coil in TEXT-U has been already undertaken. However, the magnetic coils recently installed in TEXT-U were observed to pick up rather intense electromagnetic noise from the ambient environment which made the prediction analysis by the neural net not very satisfactory.

The results presented here demonstrate the possibility of predicting the occurrence of disruptions in tokamak discharges more than 1 ms before they actually happen. This is encouraging. Certainly this prediction time period is not yet enough to allow, in a context of

real time tokamak operation, any action to be taken to modify the plasma current profile in order to avoid the upcoming disruption or, at least, minimize its harmful consequences. The predictive time span needs to be extended. For this purpose, plans are already on course to modify the program in order to include, together with the derivative of the fluctuating poloidal magnetic fields, other experimental signals such as soft X-ray emission, locked mode, H_α emission, etc., so that more information about the dynamic evolution of the plasma discharge can be incorporated into the neural net input channels. This improvement hopefully will increase the input dimensionality and the neural network ability to immediately perceive any change on the discharge conditions that could trigger a major disruption. Extending the prediction time to 3 ms, for example, would already be enough to forecast some of the major disruptions in TEXT-U (fast disruptions) before the characteristic amplitude growth of the $m = 2$ MHD precursor mode initiates.²³

V. CONCLUSIONS

The utilization of a neural net for predicting plasma disruptions in TEXT discharges more than 1 ms before they actually take place has been successfully demonstrated. The neural net code constructed was the feed-forward type. Both one-hidden layer and two-hidden layer topologies are used with different multistep prediction strategies. The program was trained through the back-propagation algorithm with inertial terms by using plasma discharges (including both disruptive and nondisruptive parts) other than the one used for forecasting. The code will now be reconstructed to accept the input of other experimental signals, like the soft X-ray, locked mode, H_α , etc., in an attempt to extend the prediction time even further.

The utility and power of the present approach are that the physical mechanism of disruptions is not essential, and thus, the approach is general and flexible. The success of neural net prediction of the rather stiff temporal phenomenon of disruptions is encouraging, as they

are believed to be among the most difficult to predict.

The work was supported in part by the U.S. Department of Energy and in part by JAERI and FAPESP.

REFERENCES

- ¹Noll, P., Sonnerup, L., Froger *et al.*, in 8th Topical Meet. on Tech. of Fusion (Proc. Fus. Tech., Salt Lake, 1989) p. 259.
- ²Karger, F., Lackner, K., Fussman, G. *et al.*, in *6th International Conference on Plasma Physics and Controlled Nuclear Fusion Research* (Berchtesgaden, Germany, 1976) International Atomic Energy Agency-CN-35/A7.
- ³Roberts, D.E., DeVilliers, J.A.M., Fletcher, J.D. *et al.*, Nucl. Fus. **26**, 785 (1986).
- ⁴Tsuji, S., Nagayama, Y., Miyamoto, K. *et al.*, Nucl. Fus. **25**, 305 (1985).
- ⁵Vannucci, A., Nascimento, I.C., and Caldas, I.L., Plasma Phys. Contr. Fusion **31**, 147 (1989).
- ⁶Vannucci, A., and Gill, R.D., Nucl. Fus. **31**, 1127 (1991).
- ⁷Waidman, G., and Kuang, G., Nucl. Fus. **32**, 645 (1992).
- ⁸Finn, J.M., Nucl. Fus. **15**, 845 (1975).
- ⁹Itoh, K., Itoh, S.I., Fukuyama, A., and Lichtenberg, A.J., Nucl. Fus. **32**, 1851 (1992).
- ¹⁰Hender, T.C., Fitzpatrick, R., Morris, A.W. *et al.*, Nucl. Fus. **32**, 2091 (1992).
- ¹¹Vallet, J.C., Pouchy, L. *et al.*, Cadarache Report EUR-CEA-FC 1427 (1991).
- ¹²Ruelle D., Ann. N.Y. Acad. Sci. **1**, 357 (1980).
- ¹³Allen, L., and Bishop, C.M., Plasma Phys. Contr. Fusion **34**, 1291 (1992).
- ¹⁴Lister, J.B. and Schnurrenberger, H., Nucl. Fus. **31**, 1291 (1991).

- ¹⁵Bishop, C.M., Roach, C.M., and Von Hellerman, M.G., JET Report JET-P(92)90, U.K. (1992).
- ¹⁶Hernandez, J.V., Tajima, T., Horton, W., Geophys. Res. Lett. **20**, 2707 (1993).
- ¹⁷Takens, F., *Dynamical Systems and Turbulence* (Springer, Berlin, 1981).
- ¹⁸Sauer, T., Yorke, J.A., and Casdagli, M., J. Stat. Phys. **65**, 579 (1991).
- ¹⁹Ekman, J.P., and Ruelle, D., Rev. Mod. Phys. **57**, 617 (1985).
- ²⁰Hertz, J., Krogh, A., and Palmer, R.G., *Introduction to Theory of Neural Computation* (Addison-Wesley, Reading, 1991) Ch. 6.
- ²¹Lapedes, A.S., and Farber, R.M., Tech. Report LAUR-87-2662, (Los Alamos National Laboratory, 1987).
- ²²Hornik, K., Stinchcombe M., and White. H., Neural Networks **215**, 359 (1989).
- ²³Rumelhart, D.E., Hinton, G.E., and Williams, R.J., Nature **323**, 533 (1986).
- ²⁴Vannucci, A., and McCool, S.C., FRC Report, (The University of Texas at Austin, 1994).

Table I — ARV and Correlation Coefficient

	Shot #	Weights 1	Weights 2	Weights 3	Weights 4
SS1	133651	0.20, 0.90	1.76, -0.21	0.28, 0.84	1.03, 0.35
SS2	133652	5.30, -0.13	0.47, 0.73	0.45, -0.06	2.06, 0.08
SS3	133654	0.25, 0.87	1.84, -0.29	0.21, 0.86	0.56, 0.65
SS4	133655	1.61, 0.29,	2.30, -0.41	2.41, 0.11	0.46, 0.71

In each box, the left value correspond to the ARV, while the right one corresponds to the correlation coefficient.

Table II — Fourier Transform Result for SS1 and SS3 ($N = 7000$)

	SS1 (133651)		SS3 (133654)	
	f_1	f_2	f_1	f_2
Index (n)	844	2160	939	1953
Frequency (kHz)	1.58	3.86	1.68	3.49
Time period (steps)	15.84	6.48	14.91	7.17

f_1 and f_2 correspond to the two major characteristic frequencies of the data shot, respectively.

FIGURE CAPTIONS

Fig. 1. The discharge shown above corresponds to a $I_p \sim 180$ kA shot which disrupted at $t \sim 290$ ms and was preceded by a minor disruption approximately 8 ms earlier. After the major disruption has taken place, the plasma position control system was unable to maintain the plasma column centered (horizontal position) and several minor disruptions occurred. The derivative of the fluctuating poloidal magnetic field from one of the Mirnov coils is shown, which was used for feeding the neural network.

Fig. 2. The predicted $\dot{B}_\theta(t+n\tau)$ values by the contiguous future prediction last 200 ms of the discharge are shown. In Fig. 2a, the experimental data is plotted for comparison with results of single timestep ($n = 1$) and multiple time step ($n = 15, 21, 28$) predictions, presented in 2b, 2c, 2d, and 2e. These values of n correspond to predictions made 0.04 ms and 0.6, 0.84, 1.12 ms ahead in time, respectively.

Fig. 3. Expansion of the experimental data (3a) and contiguous future prediction results (3b-3e) in regions before the disruption time are shown to indicate the consistency of the neural net prediction.

Fig. 4. a) and b) For contiguous future prediction, prediction error ARV and correlation coefficient, as calculated from Eqs. (6) and (7), are shown. As noted, the performance of the network degrades exponentially as the number of iterations increases.

Fig. 5. Expansion of the signals shown in Fig. 2, are plotted within the disruption range. It is noted that both the minor disruption ($t \sim 183$ ms) and the major disruption ($t \sim 191$ ms) are clearly identified by the neural network.

Fig. 6. With the single-frequency input organization, the multi-timescale prediction of shot 133655 (Fig. 6b) is compared with the experimental data (Fig. 6a). The predictive

result is scaled here to range $[-1, 1]$ for the sake of numerical stability. It is shown that both major disruptions and minor disruptions are identified, though there is a small time shift between the experimental data and the predictive one. ($\Delta t \sim 0.3$ ms)

Fig. 7. Three peaks can be got from the figure: 1103, 1316 and 1590, which correspond to frequencies: 2298Hz, 2714Hz and 3311Hz.

Fig. 8. Expansions of the original data 133655 and the multi-timescale prediction with multi-frequency input organization are shown with the nondisruptive range. The prediction is 1.0 ms ahead in time. It shows that the adoption of the multifrequency input organization removes the time shift between the experimental data and the neural net prediction.

Fig. 9. Expansions of the signals in Fig. 8 are shown within the disruption range. It shows that both the major disruptions and the minor disruptions are identified by the neural net and the time-shift is reduced.

Fig. 10. The same neural net organization as used in Figs. 8 and 9 is used for prediction of another shot (133654). Expansion of experimental and predictive data are shown at the boundary of disruption and nondisruption portion. All major and minor disruptions are identified and time-shift is removed except the small range ($\Delta t \sim 2.2$ ms) very near the boundary point.

Fig. 1 Plasma Disruption (133655)

Fig. 2 Neural prediction of shot 133655

Pooidal magnetic field fluctuation

Fig. 3 Neural prediction (Nondisruptive)

Figure 4.a Error varies with iteration times

Figure 4.b Correlation Coefficient varies with iteration times

Poloidal magnetic field fluctuation

Fig. 5 Neural prediction (Disruptive)

Figure 6. TEXT shot ss4 (133655)

Figure 7. Fourier analysis of ss4 (133655)

Figure 8. Nondisruptive portion for ss4 (133655)

Figure 9. Disruptive portion of ss4 (133655)

Figure 10. Neural net prediction for ss3 (133654)