NEURAL COMPUTING METHODS TO DETERMINE THE RELEVANCE OF MEMORY EFFECTS IN NUCLEAR FUSION

ANDREA MURARI,[a]* GUIDO VAGLIASINDI,[b] SEBASTIANO DE FIORE,[b] ELEONORA ARENA,[b] PAOLO ARENA,[b] LUIGI FORTUNA,[b] Y. ANDREW,[c] M. JOHNSON,[c] and JET-EFDA CONTRIBUTORS[d]†

[a]*Consorzio RFX-Associazione, EURATOM ENEA per la Fusione, I-35127 Padova, Italy*
[b]*Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi, Università degli Studi di Catania, 95125 Catania, Italy*
[c]*EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, United Kingdom*
[d]*JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, United Kingdom*

Received February 9, 2010
Accepted for Publication April 9, 2010

Dynamical systems are often considered immune from memory effects, i.e., the dependence of their time evolution on the previous history. This assumption has been tested for two phenomena in nuclear fusion that are believed to sometimes show sensitivity to the previous history of the discharge: disruptions and the transition from the L mode to the H mode of confinement. To this end, two neural network architectures, tapped delay lines and recurrent networks of the Elman type, have been applied to the Joint European Torus (JET) database to extract these potential memory effects from the time series of the available signals. Both architectures can detect the dependence on the previous evolution quite effectively. In the case of disruptions, only the ones triggered by locked modes seem to be influenced by the previous history of the discharge. With regard to the L-H transition, memory effects are present only in the time interval very close to the transition, whereas once the plasma has settled down in one of the two regimes, no evidence of dependence on the previous evolution has been detected.

KEYWORDS: *memory effects, recurrent neural networks, L-H transition*

Note: Some figures in this paper are in color only in the electronic version.

I. INTRODUCTION

Very often, dynamical systems are studied assuming that memory effects are completely negligible or, at least, of secondary importance. Conceptually, this implicit assumption means that to understand the physics involved or to predict the future evolution of an experiment, only the status of the system under study at a single moment in time is needed. The history leading to a certain state is considered irrelevant, and the physical phenomena that comply with this assumption are called without memory, in the sense that their future behavior can be predicted by simply knowing their state at any point in time of their evolution. This is of course the general case of all the systems acted upon by nondissipative forces, which can be expressed as the derivative of a suitable potential function. Developing techniques capable of detecting the presence of memory effects in experimental signals could therefore be useful not only to better understand the physics of these phenomena but also to define strategies for their control.

The assumption that memory effects are not relevant to study the dynamics is also almost always implicitly accepted in magnetic confinement nuclear fusion, in which the history of the plasma is generally neglected. This assumption is maintained even if many dissipative phenomena are present and in cases when evidence to the contrary is sometimes found in present-day machines. Two typical examples are disruptions[1] and the transition

*E-mail: murari@igi.pd.cnr.it
†See the Appendix of F. Romanelli et al., *Proceedings of the 22nd IAEA Fusion Energy Conference 2008*, Geneva, Switzerland.

between the L mode and H mode of confinement.[2] With regard to disruptions, no systematic analysis of memory effects on the occurrence of disruptions has ever been performed, even if some causes have a typical historical character; the most evident is the case of disruptions induced by previous locked modes. The locked mode consists of the deceleration of certain magnetic instabilities until they become stationary in the reference frame of the laboratory.[3] Once they are stationary, the stabilizing effect of the wall is far reduced, and these instabilities can grow to the point of affecting the entire discharge and even causing disruptions. It seems therefore appropriate to investigate to what extent the entire evolution of the plasma, from the triggering event to the actual disruption, has to be taken into account to understand the phenomenon. As far as the L-H transition is concerned, on some machines a significant hysteresis in the input power has been detected.[4] In these cases, the minimum power needed to reach the H mode is significantly higher than the power at which the opposite H to L transition takes place. Hysteresis is of course a paradigmatic case of memory effect, since it reveals that the system "remembers" its past history and somehow "recognizes" the direction from which it is approaching a certain transition point.

The neglect of memory effects is of course due in part to the difficulties inherent in the analysis of this type of phenomenon and the lack of established and fully general techniques to extract information about the history of a system from typical time series. In this paper, the results of an investigation of memory effects in the Joint European Torus (JET) using neural computing methods are reported. Various forms of neural networks have been tested because of their nonlinear and powerful character, leading to quite general and unbiased conclusions. In a certain sense, they are used as nonlinear identifiers to extract historical information from time series. They have been applied to the aforementioned problems of disruption prediction and the transition from the L mode to the H mode of confinement. In both cases, the networks have been designed and trained for classification purposes, i.e., either to identify discharges that are going to disrupt or to discriminate between phases of L or H mode of confinement.

In Sec. II the main types of neural networks used in the following treatment are introduced. Both a simple modification of the traditional multilayer perceptron (MLP), called tapped delay line (TDL) networks,[5] and a more substantial modification of the traditional network architecture, the recurrent networks of the so-called Elman type [Elman recurrent neural network[5] (ERNN)], have been implemented. These specific network architectures have to be deployed because the original topology of the MLP was explicitly devised to avoid memory effects by eliminating internal loops. The aforementioned TDL network and ERNN have been tested using synthetic data to show their potential to extract historical information from time series. Both types of networks have then been applied first to the evolution of the plasma before a disruption (see Sec. III) because in this case an independent method to test the quality of their predictions has been found. On the basis of the positive results obtained with the synthetic data and the real case of disruptions, the transition from the L to H mode of confinement has also been studied (see Sec. IV). Stock of the investigations performed so far is taken in Sec. V, together with some indications about the lines of further research.

II. TDL NETWORK AND ERNN FOR THE DETERMINATION OF MEMORY EFFECTS

The architecture of the traditional feedforward neural networks does not contain loops exactly for the purpose of avoiding internal feedback,[5] which is essential to introduce memory effects but which makes the training a much more difficult proposition. Indeed, in order to apply the original backpropagation algorithms, which were the first training methods devised, the network must not contain any internal loop. With these traditional neural networks, i.e., MLPs, the only way of assessing whether the history of the system plays a role in determining the output consists of providing the inputs at various times and seeing how the performance of the network is modified when additional time slices are added. With this approach the temporal information is in a certain sense converted into spatial information, and therefore, the traditional backpropagation algorithms can be used for the training. This network topology, shown in Fig. 1, is sometimes called a "TDL" since from the hardware point of view, it can be implemented by storing intermediate time slices in a buffer.

The activation functions chosen, for all the TDL applications described in this paper, are linear for the output layers and a bipolar sigmoid $sigmb(x) = 2/(1 + e^{-2x}) - 1$ for the hidden layer. The number of neurons in the hidden layers has been optimized on a case-by-case basis by finding the best trade-off between the success rate and overfitting.

In order to increase confidence in the results and test an alternative approach, a different type of architecture has also been considered. For the applications discussed in this paper, the main issue consists of being able to determine to what extent historical information is present in the time series of the acquired data. Recurrent networks[5] are modifications of the traditional MLP architecture, explicitly conceived to take into account short-term-memory effects. They operate not only on the input space but also on their previous internal state through suitable feedback loops. The inputs to a recurrent network are therefore not only propagated through a weight layer but also combined with the previous activation state, using one or more recurrent weight layers. If memory effects are present in the system, the values of the weights

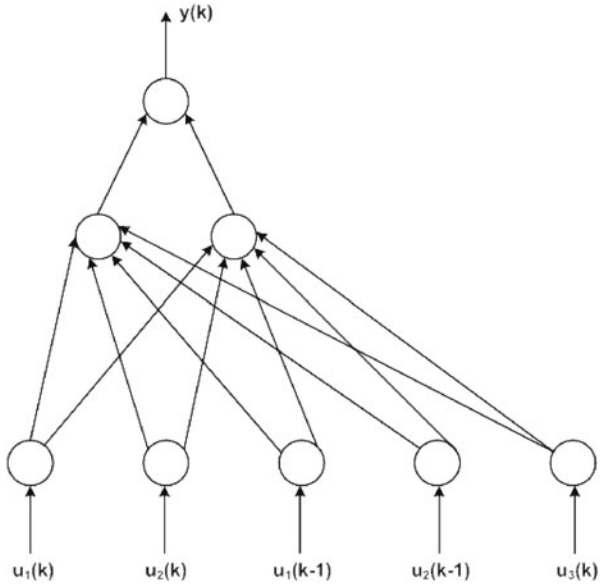


Fig. 1. Topology of the TDLs. The symbol u identifies the inputs, and the symbol y identifies the output. The neurons not labeled are the neurons of the hidden, intermediate layer. In the example shown in this figure, u_1 and u_2 present memory effects whereas u_3 does not.

at previous times are expected to have an effect on the convergence of the network. The ERNN is a recurrent network implementing this idea. It presents a hidden layer, with the topology shown in Fig. 2. In all the ERNNs used

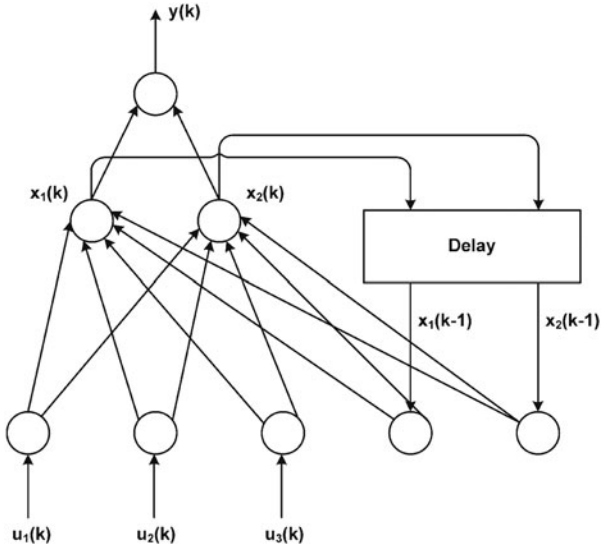


Fig. 2. Topology of the ERNNs showing the internal feedback with delay. The symbol u identifies the inputs, and the symbol x identifies the internal status of the neurons in the intermediate layer.

to obtain the results described in this paper, the activation function adopted is the unipolar sigmoid $sigmu(x) = 1/(1+e^{-x})$ for the neurons of both the hidden and output layers. Again, the number of neurons in the hidden layers has been optimized on a case-by-case basis by finding the best trade-off between the success rate and overfitting.

This type of architecture contains internal feedback loops that really embody short-term memory, contrary to the TDL solution, in which the historical information is taken into account by the past inputs presented to the network. This different approach, which is expected to be more powerful, on the other hand requires specific training procedures, basically more sophisticated versions of the traditional backpropagation. The training strategy adopted in this paper is called backpropagation through time,[6] which is a form of "unfolding." The recurrent weights are duplicated spatially for a suitable number of time steps indicated traditionally by the symbol τ. Therefore, each node in a feedback loop is copied τ times, the exact number of which depends on the memory requirements of the problem at hand. The backpropagation can then be applied to calculate the weights, taking into account the internal status of the network at previous τ time steps.

In order to become more familiar with the operation of these two architectures and to confirm the proper functioning of the software available, the two aforementioned architectures have been tested using synthetic data derived from a simple mathematical model. The formula used to benchmark the networks has the form

$$\begin{aligned} Y(K) = {} & au_1(k) + bu_2(k) + cu_1(k-1) + du_2(k-1) \\ & + eu_1(k-2) + fu_2(k-2) + g * u_1(k)u_2(k) \\ & + h * u_1(k-1)u_2(k-1) \ . \end{aligned} \tag{1}$$

The two inputs $u_1(k)$ and $u_2(k)$ indicate that the samples collected at the reference time $u_1(k-1)$ and $u_2(k-1)$ are the two inputs at the previous time $u_1(k-2)$ and $u_2(k-2)$, which are the values two time slices before the current one and so forth. The input variables can influence the output Y to the extent determined by the value of their multiplying coefficients (a, b, c, d, e, f, etc.), whose exact values are irrelevant to the results reported in the following but are reported in Table I.

Relation (1) has been used to generate a series of synthetic signals, which have then been given as input to the networks, to see to what extent their performance improves when previous time slices are given as inputs. This is a regression problem consisting of estimating the output Y of a system (or function) on the basis of the inputs u_1 and u_2. The results, summarized in Table I, refer to the application of the networks to test sets after appropriate training with completely independent examples. The reported results are meant to show the improvement in the regression capability when earlier time slices are given to the TDL networks. The parameter used to

TABLE I

Improvement of the Predictions by TDL Networks When Historical Information Is Provided*

Generating Function	All	MSEP Train	Test	MSEP with Memory		
				All	Train	Test
GF1 $u_1 = -5{:}0.1{:}5$ $u_2 = -5{:}0.1{:}5$	0.0119	0.0074	0.0208	0.0064	0.0001	0.0188
GF2 $u_1 = \sin(-5{:}0.1{:}5)$ $u_2 = \cos(-5{:}0.1{:}5)$	0.0767	0.0508	0.1278	0.0104	0.0001	0.0306
GF3 $u_1 = \exp(-5{:}0.1{:}5)$ $u_2 = [\exp(-5{:}0.1{:}5)]^{-1}$	0.0323	0.0029	0.0902	0.0141	0.0002	0.0417
GF4 $u_1 = \tan(-5{:}0.1{:}5)$ $su_2 = \sin(-5{:}0.1{:}5)$	0.5213	0.3978	0.7645	0.0084	0.0003	0.0242

*The historical evaluation has been performed for a memory effect of two time steps, i.e., two time slices before the reference time. The values of the constant in relation (1) are $a = b = 1$, $c = d = 0.9$, $e = f = 0.8$, and $g = h = 0.7$. In the second, third, and fourth columns, the results obtained by the network without historical information are shown; columns five, six, and seven report the improvement when the two previous time slices are provided. The results for both the training and the test sets have been reported for various generating functions of u_1 and u_2 (see third column).

quantify the increase in the success rate is the mean square error of predictions (MSEP):

$$\text{MSEP} = \frac{\sum_{i=1}^{n} \left(\frac{Y_i - \langle Y_i \rangle}{Y_i} \right)^2}{n} , \qquad (2)$$

where

Y_i = real value of Y

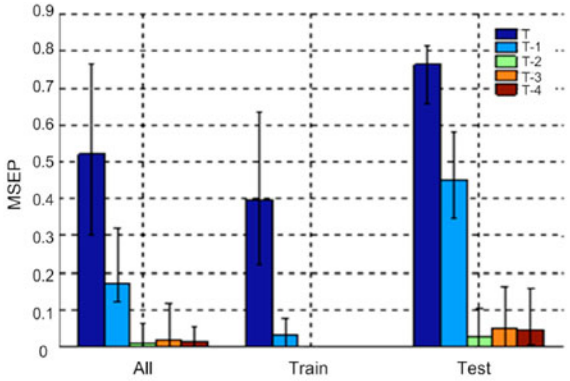
$\langle Y_i \rangle$ = estimated value of Y

n = total number of samples.

The number of neurons in the hidden layer is five for this application.

The MSEP is an absolute index, and it is independent of the input range dimensions. The values reported in Table I clearly indicate that providing the TDL network with two additional time slices, corresponding to the memory effect generated by relation (1), has very beneficial effects. The improved performance testifies to the ability of the TDL architecture to properly detect and accommodate historical information present in time series.

Additional analysis has been performed to investigate to what extent the TDL networks are able to identify the proper delay, which accounts for the memory effects in the data. To this end, again relation (1) has been used to generate synthetic signals. Time sequences up to four sequential time slices have been given to the TDL networks to see whether they can identify the right memory time in the system generating the data. The good capability of this architecture to extract historical information from the input data is shown in Fig. 3. The increase in performance, when the right number of time slices (three) is provided to the networks, is clearly seen as a minimum



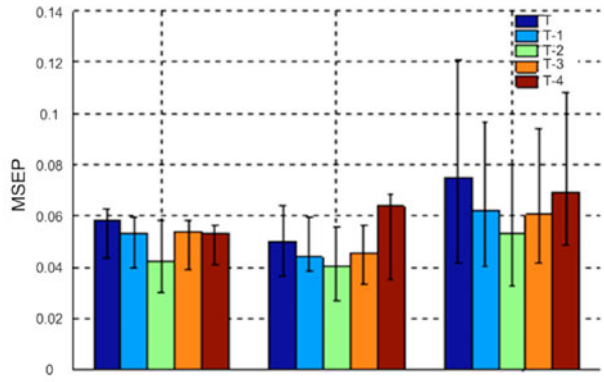
	T	T-1	T-2	T-3	T-4
All	0.5213	0.1725	0.0084	0.0165	0.0149
Train	0.3978	0.0307	0.0003	0.0003	0.0002
Test	0.7645	0.4519	0.0242	0.0486	0.0437

Fig. 3. Evolution of the TDL classification errors for the system described by relation (1) with the generating function GF4 of Table I. The memory effect used to generate the synthetic data extends for two time slices. The memory times in the legend are in the same order as the slots in the x-axis.

in the MSEP. On the other hand, the errors in the classification typically start increasing again if more than the right number of time slices is provided as input. This has been confirmed for all the various types of generating functions summarized in Table I. It seems therefore that the TDL architecture is capable of identifying the right interval in which historical data are important and that interval can be identified by the minimum of the indicator MSEP.

A similar analysis has been performed to investigate the "memory effect detection capability" of the ERNN. Figure 4 shows the good capability of the ERNN, with three neurons in the hidden layer, to extract historical information from the input data obtained using relation (1) with a memory effect of two time steps and the generating function (GF) GF4 of Table I. The MSEP in the classification decreases when the right number of time slices (again three) is considered in the training algorithm. Moreover, the errors start increasing again if more than the right number of time samples is provided during the training process. This behavior has been confirmed for all the generating functions of Table I. As for the TDL network, the ERNN performance also improves if inputs covering the right historical interval are provided.

After demonstrating the potential of the various network architectures to capture memory effects with synthetic data, the same tools have been applied to two important phenomena in tokamak plasmas—the disruptions and the transition between different modes of confinement—as described in detail in Secs. III and IV.



	T	T-1	T-2	T-3	T-4
All	0.058	0.0532	0.0425	0.0537	0.0535
Train	0.0498	0.044	0.0402	0.0452	0.0639
Test	0.075	0.0623	0.0529	0.0608	0.0689

Fig. 4. Evolution of the ERNN classification errors for the system described by relation (1) with the generating function GF4 in Table I. The memory effect used to generate the synthetic data extends for two time slices ($T-1$ and $T-2$). The memory times in the legend are in the same order as the slots in the x-axis.

III. ASSESSMENT OF THE MEMORY EFFECTS BEFORE DISRUPTIONS

This section describes how the two network architectures just described have been applied to the problem of identifying disruptive discharges; this is a typical classification problem that consists of determining which time slices in the database belong to discharges that are going to disrupt. Disruptions consist of unforeseen and sometimes very fast losses of plasma confinement, which abruptly terminate the discharge. The thermal quench, the phase in which the energy content of the plasma is deposited on the first wall, can occur in matters of a few milliseconds. The following current quench is slower but can typically occur in several tens of milliseconds. The typical temporal evolution of the main plasma quantities is shown in Fig. 5.

Disruptions are potentially very harmful events. First of all, they cause very high and localized thermal loads on the first wall. Second, the fast termination of the plasma current induces eddy currents on the surrounding metallic structures, which can give rise to high induced forces. The risk involved in disruptions is already quite significant in present-day large devices such as JET, and it is going to increase significantly in the next generation of machines, which will work at much higher plasma currents and thermal energy. Understanding their behavior to improve early prediction and appropriate intervention is therefore a very urgent issue.

The most relevant signals for disruption prediction, which have been retained for the study reported in this paper, are summarized in Table II and have been chosen on the basis of the nonlinear correlation method, called Classification and Regression Trees (CART), as described in Ref. 7. CART is a supervised method[8] that simply traverses the entire database to determine which variable and which value better divide the examples to be classified into two groups. After the most selective

TABLE II

List of the Signals Used as Inputs to the Disruption Predictors as Derived from CART

Signal Name
Plasma current, I_{pla} (A)
Mode lock amplitude, *Loca* (T)
Plasma density, *Dens* (m^{-3})
Total input power, P_{inp} (W)
Plasma internal inductance, L_i
Stored diamagnetic energy derivative, dW_{dia}/dt (W)
Safety factor at 95% of minor radius, q_{95}
Poloidal beta, β_p
Net power, $P_{net} = (P_{inp} - P_{rad})$ (W)

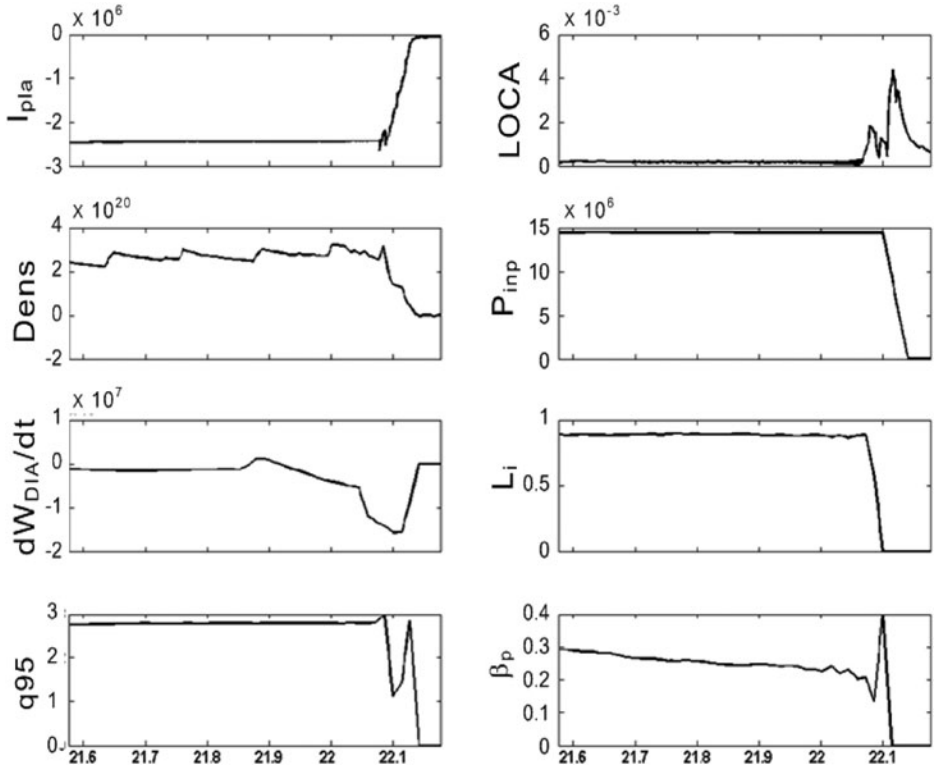
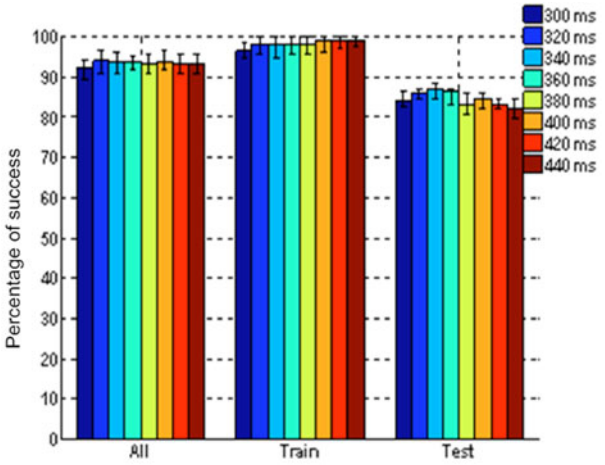


Fig. 5. Evolution of plasma quantities before and during a disruption for JET shot #52105. The disruption occurs around 22.08 s. Here, I_{pla} is the plasma current (in amperes), *Dens* is the plasma electron density (in particles per cubic meter), dW_{DIA}/dt is the time derivative of the diamagnetic energy (in joules per second), q_{95} is the safety factor at 95% of the plasma radius, *LOCA* is the magnetic signal proportional to the amplitude of the locked mode (in tesla), P_{inp} is the input power (in watts), L_i is the internal plasma inductance, and β_p is the poloidal plasma beta.

variable has been chosen, the procedure is repeated iteratively for the resulting subclasses until a perfect classification is obtained. The output of the method is represented as a tree whose nodes contain the variables in descending order of importance from the root down to the final leaves.

For the results described in this paper, the signals reported in Table II have been used as inputs to a set of TDL networks: The first network of the set has been trained with these signals taken only at one time, the second network has been trained with the same inputs but also taking into account the previous time slice, the third network has been trained with data belonging to the two previous time slices, and so on. The output of the networks is a Boolean value, indicating whether or not the plasma is going to disrupt (one Boolean value is used to indicate disruptive discharges, and the other is used to indicate nondisruptive discharges). The success rate is defined as the ratio in percentage of the number of properly classified time slices to the total number of time slices in the database (and this definition is the same for all the results quoted in the rest of the paper). The signals of the various time intervals have been multiplied by suitable weights, determined empirically to maximize performance and decreasing with increasing time to the disruption. The actual values of these weights are reported in the caption of Fig. 6; they are decreasing with the distance from the disruption, which reflects the fact that the information content of the time slices is decreasing the farther away from the time of the disruption.

To prove that the first architecture, the TDL, really extracts from the database information about the historical evolution of the discharge, this architecture has been applied first to the case of disruptions induced by a previous locked mode. A specific database of about 70 discharges, whose disruptions have been classified by the experts as all due to a locked mode, has been used to train and then to test the TDL architecture with ten neurons in the hidden layer. Approximately 70% of the discharges has been used for the training phase, and the remaining 30% has been used for the test phase. The stopping criterion is the threshold of 10 000 epochs. The reference time slice is between 300 and 320 ms before the disruption. The performance of the network—once earlier time slices, each one covering 20 ms, are added as inputs—is reported in Fig. 6. Including information of previous time slices (in the overall interval between 320 and 380 ms before the disruption) improves the performance almost



	All	Train	Test
300 ms	92.0699	96.5258	84.0232
320 ms	93.883	98.1157	85.8974
340 ms	93.8121	98.205	86.8298
360 ms	93.6511	97.8854	86.2637
380 ms	93.284	98.2166	82.9837
400 ms	93.6702	98.9119	84.6154
420 ms	93.3462	98.8588	83.3333
440 ms	93.4043	99.0064	81.9444

Fig. 6. Improved performance of TDL networks with historical inputs for the case of disruptions triggered by a locked mode. The different slots for each dataset (All, Train, and Test) indicate the times before the disruption when the various sets of inputs have been taken. The weights are 1 for the time slice at 300 ms, 0.9 for the time slice at 320 ms, 0.8 for the time slice at 340 ms, 0.7 for the time slice at 360 ms before the disruption, and so on. The success rate is the percentage of cases for which the networks properly manage to identify whether the time slice belongs to a disruptive discharge or a not disruptive discharge. The memory times in the legend are in the same order as the slots in the *x*-axis.

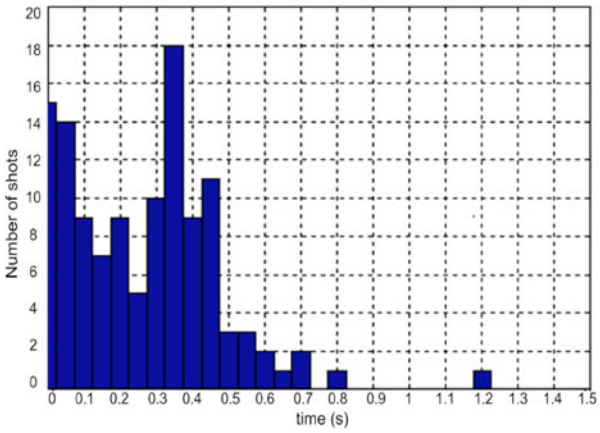
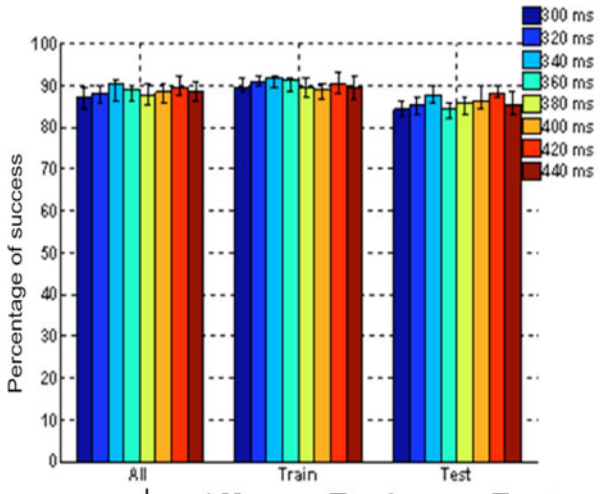


Fig. 7. Statistical distribution of the time that elapses between the locked mode and the disruption for our database. The *x*-axis is the time between the detection of the locked mode and the occurrence of the disruption. The time resolution in the determination of the time when the mode locks to the wall is better than 1 ms (the locked mode signals are sampled every 200 μs).

3%, which is quite significant given the high success rate of the network without historical data (already well above 80% as reported in Fig. 6). In Fig. 6 the uncertainty intervals are due to the statistical fluctuations in the results obtained when randomly changing the training and test sets. Therefore, uncertainty intervals do not have to be considered error bars; when the training and test sets are kept constant, the improvement has always been consistently detected. The trend of the improvement in performance with time has been compared with the times before the disruption when the locked modes occur. In this set of discharges, the frequency of locked modes has a significant peak around 360 ms before the disruption, as shown in Fig. 7. The success rate of the TDL network increases significantly when the time slices corresponding exactly to this interval are provided as inputs. This is a strong, experimental verification that the network, trained with the proposed method, is capable of extracting real historical information from the time series of the input signals. This potential of the network can contribute to determining how early in a discharge there is information about an incoming disruption.

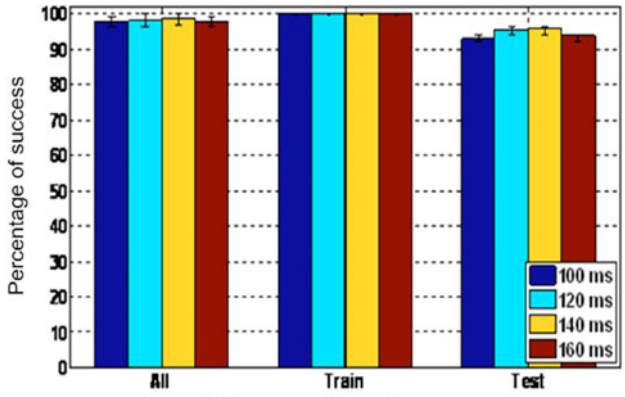
To confirm these results, the same database has been analyzed with ERNNs, also with ten neurons in the hidden layer. The indications about the memory effects are better than the ones derived from the TDLs, as shown in Fig. 8. The improvements in the performance again have a maximum around 360 ms before the disruption. Moreover, the improvement is even outside the uncertainty intervals due to the random choice of the training and test sets. The ERNNs also seem to be capable of detecting the second peak in the distribution of locked mode times, which is present around 420 ms before the disruption (again see Fig. 7). This feature of the input statistics has not been reproduced by the TDLs, which indeed show an inferior power compared to the ERNN architecture. The reason for the lower performance of the TDL approach is believed to be the excessive increase in the complexity of the network with the memory requirements of the problem. If the historical information to be considered extends too much into the past, the number of inputs becomes too high, and the TDL networks have problems in coping and extracting the details of the distribution function.

The same approach has then been applied to the entire database of JET disruptions, without any distinction about their causes. The used database consists of 292 disruptive discharges and 220 nondisruptive cases, whose signals are sampled at a rate of 20 ms. In this case, the



	All	Train	Test
300 ms	87.0699	89.5258	84.2516
320 ms	88.1277	90.8705	85.3896
340 ms	90.0426	91.8544	87.6623
360 ms	88.8936	91.4013	84.4156
380 ms	87.6596	89.2994	85.7143
400 ms	88.383	88.899	86.2709
420 ms	89.4468	90.2548	87.7922
440 ms	88.5532	89.5701	85.1577

Fig. 8. Improved performance of ERNNs with historical inputs. The same database and the same notation as in Fig. 6 have been used. The two peaks in the success rate (~340 to 360 ms and 420 ms before the disruption) correspond to the intervals of increased percentage of locked modes as shown in Fig. 7. The memory times in the legend are in the same order as the slots in the *x*-axis.



	All	Train	Test
100 ms	96.7692	99.5723	91.2573
120 ms	97.3516	99.9648	92.4269
140 ms	97.75	100	93.3723
160 ms	97.0296	100	92.1888

Fig. 9. Performance of TDL architecture with historical inputs. No selection on the type of disruption has been performed. The different slots for each dataset (All, Train, and Test) indicate the times before the disruption when the various sets of inputs have been taken. The definition of the success rate and the method to randomly select the various sets of discharges are the same as in Figs. 6 and 8. The results for the test set do not show any significant improvement. The memory times in the legend are in the same order as the slots in the *x*-axis.

interval between 100 and 180 ms before the disruption has been investigated. This choice is motivated by previous analyses with exploratory techniques, which have shown that in the database used, there is not much information about an incoming disruption earlier than ~180 ms before its occurrence.[9] One example of the results is reported in Fig. 9 for the case of the TDL networks with ten neurons in the hidden layer. Various time intervals have been chosen for the first time slice, but the sequence starting at 100 ms before the disruption—the one shown in Fig. 9—provides the most significant results. This analysis shows a consistent but very small trend of improved performance of the predictor when the earlier time slices are provided as additional inputs. Even if this trend has been consistently recovered in all the different cases performed with random training and test sets, the improvement in the performance is quite limited in absolute terms. These results indicate that some sort of memory effects cannot be completely excluded since the success rate of the TDLs is at least not worsened by including earlier time slices in the list of inputs, even if the information content of these time intervals is lower, being more distant from the disruption. On the other hand the trend is not very strong and difficult to address with the data available. Similar conclusions can be obtained with ERNNs. Therefore, from the analyzed database a picture emerges according to which the disruptions due to a locked mode present clear memory effects. On the other hand, in the general database without distinction about the disruption causes, no clear indication of strong memory effects has been detected.

IV. ASSESSMENT OF THE MEMORY EFFECTS AROUND THE TRANSITION TO THE HIGH CONFINEMENT REGIME

Another important phenomenon, whose memory effects have been analyzed with the neural networks described in Sec. II, is the transition between confinement regimes. In the ASDEX device it was discovered in 1982 that by increasing the input power above a certain threshold, the plasmas could be induced to transit to an enhanced confinement mode called the high confinement mode or H-mode.[2] The time evolution of the main plasma

quantities for a typical discharge with an L to H and an H to L transition is shown in Fig. 10. The H mode is characterized by the presence of a thin region of very low transport situated at the edge of the plasma. Steep gradients in the density and temperature profiles are observed across this region. This thin layer of increased gradients in the kinetic profiles is commonly referred to as the external transport barrier. Determining the scaling laws for the threshold to access the H mode is one of the most important research topics from the perspective of the next-generation international device ITER.

To study the relevance of memory effects on the plasma dynamics leading to the transition to the H mode, a database of about 60 discharges has been prepared by the experts. All these discharges present an L-mode and H-mode phase, and again, ~70% has been used for the training and the remaining 30% for the test. Also, for the networks described in this section, the stopping criterion is the threshold of 10 000 epochs. The details of this database can be found in Ref. 10. The signals most relevant to the analysis of this phenomenology have been identified again with the nonlinear and unbiased method of the CART algorithms. The most important quantities identified by CART are the magnetohydrodynamic (MHD) energy, the axial toroidal magnetic field at 80% of the flux, the electron temperature, the beta normalized, the *X*-point radial position, and the *X*-point vertical position.[11] For these signals, various time slices have been provided as input to TDL networks, and they have been trained to identify whether the plasma is in the L or H mode of confinement. The output of the networks is now a Boolean value, indicating whether or not a transition to the H mode has taken place (one Boolean value is used to indicate the L-mode phase of the discharges, and the other is used to indicate the H-mode phase). The number of neurons in the hidden layer is now eight. For both the training and the test sets, three couples of symmetric time windows around the transition have been defined (see Fig. 11 for the exact definition of these time intervals). Time slices on both sides of the transition from the L to the H mode are necessary for the networks to learn the difference between these two plasma states. The time of the transition is therefore considered the origin of the time axis in all the figures referring to the L-H transition.

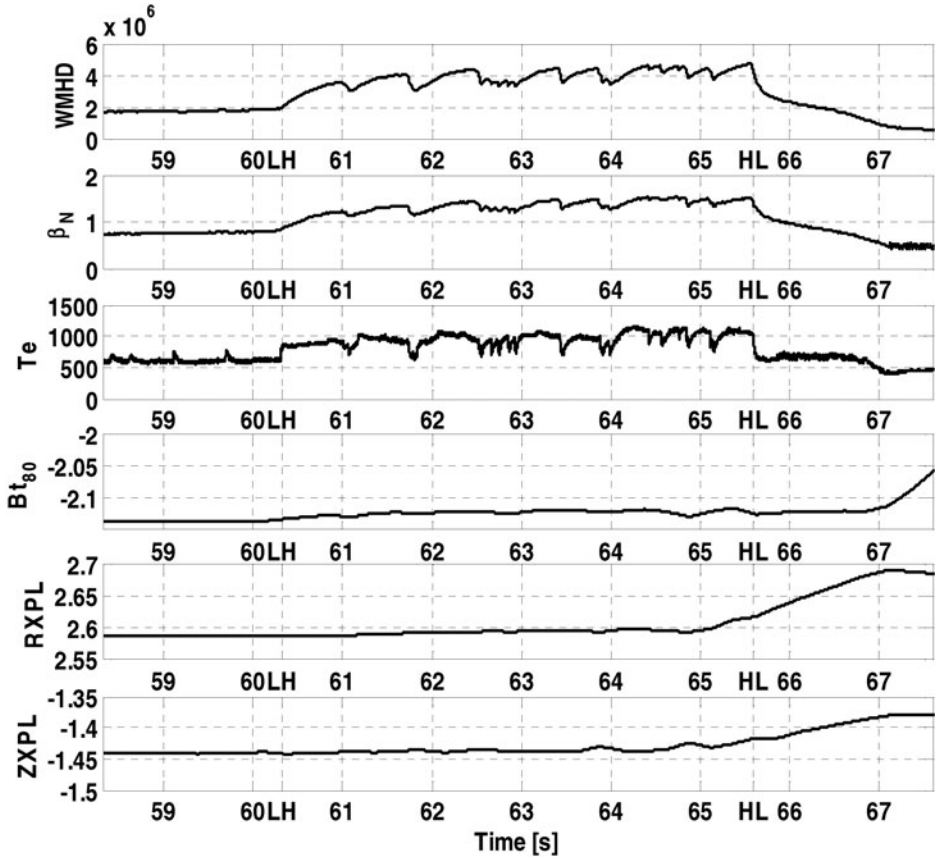
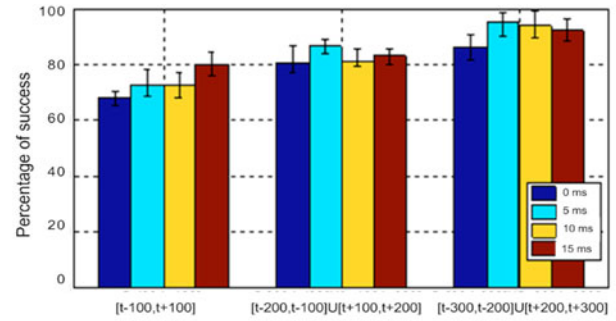
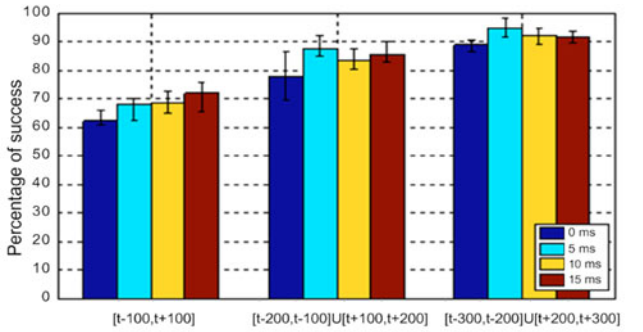


Fig. 10. Time evolution of the main plasma quantities for shot #58764. LH and HL indicate the times of the transition to and from the H mode, respectively; W_{MHD} is the internal energy in the MHD approximation (in joules), β_N is the normalized beta; *Te* is the electron temperature (in electron volts); B_{t80} is the toroidal field at 80% of the plasma radius (in teslas); *RXPL* is the horizontal coordinate of the *X* point (in meters); and *ZXPL* is the vertical coordinate of the *X* point (in meters).



	$[t_{LH}-100, t_{LH}+100]$		$[t_{LH}-200, t_{LH}-100]$ U $[t_{LH}+100, t_{LH}+200]$		$[t_{LH}-300, t_{LH}-200]$ U $[t_{LH}+200, t_{LH}+300]$	
	Mean	**Max**	**Mean**	**Max**	**Mean**	**Max**
0 ms	68.0413	70.0565	80.7088	86.7857	86.5627	90.7157
5 ms	72.7558	78.3366	86.6346	89.4231	95.1273	98.6193
10 ms	72.5198	76.9841	81.4046	86.0784	94.0589	99.4186
15 ms	80.3061	84.6939	83.2806	86.1167	92.2324	96.6203

Fig. 11. Performance of the TDL networks for the identification whether the plasma is in the L or H mode of confinement. The success rate indicates the percentage of time slices that are properly classified as belonging to the L or H phase of the discharge. Three intervals of various lengths around the transition and different integration times have been considered. The results shown refer to the test set. The memory times in the legend are in the same order as the slots in the *x*-axis.



	$[t_{LH}-100, t_{LH}+100]$		$[t_{LH}-200, t_{LH}-100]$ U $[t_{LH}+100, t_{LH}+200]$		$[t_{LH}-300, t_{LH}-200]$ U $[t_{LH}+200, t_{LH}+300]$	
	Mean	**Max**	**Mean**	**Max**	**Mean**	**Max**
0 ms	68.0413	70.0565	80.7088	86.7857	86.5627	90.7157
5 ms	72.7558	78.3366	86.6346	89.4231	95.1273	98.6193
10 ms	72.5198	76.9841	81.4046	86.0784	94.0589	99.4186
15 ms	80.3061	84.6939	83.2806	86.1167	92.2324	96.6203

Fig. 12. Success rate of the ERNN for the same database used in Fig. 11. The results are confirmed: The success rate improves only for the interval close to the transition. The memory times in the legend are in the same order as the slots in the *x*-axis.

In this couple of intervals around the transition, the time slices have been chosen randomly for seven test sets, whereas a single optimized training set has been prepared to properly cover the entire operational space. To assess the presence of memory effects in the data, time slices of increasingly longer periods (up to 15 ms; see Fig. 11) have been provided as inputs to the networks. The bin indicated with 0 ms contains the results obtained selecting single time slices symmetric around the transition. The bin called 5 ms has been calculated using two time slices around the transition, located 5 ms apart (and always symmetric with respect to the L-H transition time). The bin labeled 10 ms (yellow online) contains three values, symmetric in time around the L-H transition: one at a random time t, one the average between this random time and $t-5$ ms, and one the average between $t-5$ ms and $t-10$ ms (an analogous procedure has been adopted for the 15-ms case).

The results indicate that historical information improves the performance of the networks only in the time interval $[-100$ ms, 100 ms$]$ around the transition. Indeed, as can be seen in Fig. 11, only in this interval is the improved performance consistent and outside the statistical intervals due to the random choice of the training and test sets. The improvement also keeps increasing systematically as more time slices are provided to the network.

Once the plasma is stably in one of the two confinement regimes, as is likely to be the case for the intervals $[-200$ ms, -100 ms$] \cup [100$ ms, 200 ms$]$ and $[-300$ ms, -200 ms$] \cup [200$ ms, 300 ms$]$, historical information does not improve the performance of the networks, and therefore, memory effects seem not to be relevant any more. It then seems quite natural to conclude that some memory effects are present only very close to the transition. As in Sec. III, the same database and the same training and test sets have been analyzed with ERNNs to confirm the results. The optimal number of neurons in the hidden layer is now five. The improving of the performance has been evaluated in the same time windows as the TDL case, and the results are shown in Fig. 12, where again performance improves weakly and only in the time interval $[-100$ ms, 100 ms$]$ around the transition. Therefore, once the plasma is stably in one of the two confinement regimes, historical information does not improve the performance of the networks, and memory effects cannot be detected any more. These results are coherent with previous experimental investigations,[12] which have never found very strong evidence for hysteresis in JET plasmas.

V. PRELIMINARY CONCLUSIONS AND DIRECTIONS OF FUTURE INVESTIGATIONS

The potential of two neural network architectures—TDLs and ERNNs—to extract information about memory effects of time series has been investigated. The two

network topologies have been tested first using synthetic data to confirm their inherent sensitivity to the presence of historical information in their inputs. They have then been applied to the identification of memory effects in JET plasmas. Two main classes of phenomena have been studied: disruptions and the L to H transition. With regard to the first phenomenology, clear evidence for memory effects in the data has been found only for the disruptions preceded by a locked mode. For the general database, without discrimination about the causes of the disruptions, no statistically significant evidence of memory effects has been detected. Since in JET the mode locked is detected and taken into account in the prediction algorithms, the investigation presented in this paper supports the validity of the disruption avoidance strategy already implemented. With regard to the L to H transition, clear evidence of memory effects has been identified only for the time interval of ± 100 ms around the time of the transition. Farther away, when the plasma is more stably in one of the two confinement modes, there is no impact of the historical information on the output of the neural network classifiers. Therefore, even if the effect is not dramatic, theoretical models could be developed to accommodate some level of dependence from the history just before the transition.

With regard to the continuation of this line of research, other phenomena could be investigated. Among the most interesting, apart from disruptions and the L-H transition, could be the formation of the various internal transport barriers, which are routinely produced in JET. Instabilities, like sawteeth and neoclassical tearing modes, would also constitute an interesting subject of investigation. From a methodological point of view, some information theoretical techniques, based on signal entropies or conditional probabilities, could also be considered to investigate their potential to extract information about memory effects from time series signals.

REFERENCES

1. F. C. SCHULLER, "Disruptions in Tokamaks," *Plasma Phys. Control. Fusion*, **37**, *11A*, A135 (Nov. 1995).

2. F. WAGNER et al., *Phys. Rev. Lett.*, **49**, 1408 (1982).

3. R. FITZPATRICK, *Phys. Plasmas*, **5**, 3325 (1998).

4. D. M. THOMAS et al., *Plasma Phys. Control. Fusion*, **40**, *5*, 707 (May 1998).

5. D. P. MANDIC and J. A. CHAMBERS, *Recurrent Neural Networks for Prediction*, John Wiley and Sons.

6. D. E. RUMELHART, G. E. HINTON, and R. J. WILLIAMS, *Nature*, **323**, 533 (1986).

7. A. MURARI et al., *IEEE Trans. Plasma Sci.*, **34**, *3* (June 2006).

8. L. BREIMAN, J. H. FRIEDMAN, R. A. OLSHEN, and C. J. STONE, *Classification and Regression Trees*, Wadsworth, Inc., Belmont, California, Chapman & Hall, New York (1993).

9. A. MURARI et al., *Nucl. Fusion*, **48**, 0350 (2008).

10. A. J. MEAKINS, "A Study of the L-H Transition in Tokamak Fusion Experiments," PhD Thesis, Imperial College London (2008).

11. G. VAGLIASINDI et al., *Proc. 11th European Symp. Artificial Neural Networks (ESANN'08)*, Bruges, Belgium, April 23–25, 2008, p. 517 (2008).

12. Y. ANDREW et al., *Plasma Phys. Control. Fusion*, **50**, *12*, 124053 (Dec. 2008).