A. Murari, G. Vagliasindi, S. DeFiore, E. Arena, P. Arena, L. Fortuna,Y. Andrew,
M. Johnson and JET EFDA contributors

Neural Computing Methods to Determine the Relevance of Memory Effects in Nuclear Fusion

Neural Computing Methods to Determine the Relevance of Memory Effects in Nuclear Fusion

A. Murari[1], G. Vagliasindi[2], S. DeFiore[2], E. Arena[2], P. Arena[2], L. Fortuna[2], Y. Andrew[3], M. Johnson[3] and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]*Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy.*
[2]*Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi-Università degli Studi di Catania, 95125 Catania, Italy*
[3]*EURATOM-UKAEA Fusion Association, Culham Science Centre, OX14 3DB, Abingdon, OXON, UK*
* See annex of F. Romanelli et al, "Overview of JET Results" ,
(Proc. 22 [nd] IAEA Fusion Energy Conference, Geneva, Switzerland (2008)).

ABSTRACT.

Dynamical systems are often considered immune from memory effects, the dependence of evolution from the previous history. This hypothesis has been tested for two phenomena in nuclear fusion, which are believed to sometimes show sensitivity to the previous evolution of the discharge: disruptions and the transition from the L to the H mode of confinement. To this end, two innovative neural network architectures, the Tapped Delay Lines and the Recurrent Networks of the Elman type, have been applied to JET database to extract these potential memory effects from the time series of the available signals. Both architectures can detect the dependence from the past history quite effectively. In the case of disruptions, only the ones triggered by locked modes seem to be influenced by the past 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 from the previous evolution has been detected.

1. INTRODUCTION

Very often dynamical systems are studied assuming that memory effects are completely negligible or, at last, of secondary importance. Conceptually this implicit assumption means that, to understand the physics involved or 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 which comply with this assumption are called without memory, in the sense that their future behaviour 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 non dissipative forces, which can be expressed as the derivative of a suitable potential function. The assumption that memory effects are not relevant to study the dynamics is almost always implicitly accepted also in magnetic confinement nuclear fusion, in which the history of the plasma is in general neglected. This hypothesis is maintained even if a lot of dissipative phenomena are present and also in cases when evidence to the contrary is sometimes found on present day machines. Two typical examples are disruption prediction [1] and the transition between the L 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 stabilising 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 which extent the entire evolution of the plasma, from the triggering event to the actual disruption, has to be taken into account in order 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 "recognises" 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 phenomena and the lack of established and fully general techniques to extract information about the history of a system from typical time series data. In this paper, the results of an investigation of memory effects in 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 non linear 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 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 which are going to disrupt or to discriminate between phases of L or H mode of confinement.

With regard to the organisation of the paper, in section 2 the main types of Neural Networks used in the following treatment are introduced. Both a simple modification of the traditional multilayer perceptron, called Tapped Delay Lines (TDL) networks [5], and a more substantial modification of the traditional network architecture, the recurrent networks of the so called Elman type (ERNN) [5], have been implemented. These specific network architectures have to be deployed because the original topology of the multilayer perceptron was explicitly devised to avoid memory effects, by eliminating internal loops. The aforementioned TDL 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 section 3), 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 section 4). Stock of the investigations performed so far is taken in the last section, together with some indications about the lines of further research.

2. TAPPED DELAY LINE AND ELMAN RECURRENT NETWORKS 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, called MultiLayer Perceptrons (MLP), 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 see how the performance of the network are modified when additional time slices are provided. 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 approach is sometimes called a "tapped delay line" since, from the hardware point of view, it can be implemented by storing intermediate time slices in a buffer.

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 propagated not only 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 at previous times are expected to have an effect on the convergence of the network. The Elman network is a recurrent network implementing this idea. It presents a hidden layer, with the topology shown in figure 1. 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 (BPTT) [6], which is a form of "unfolding". The recurrent weights are duplicated spatially for a suitable number of time steps indicated traditionally with the symbol τ. Therefore each node in a feedback loop is copied τ times, whose exact number 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 get 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 benchmark chosen has the form:

$$Z = aX_0 + bY_0 + cX_{-1} + dY_{-1} + eX_{-2} + fY_{-2} \tag{1}$$

The two inputs X_0 and Y_0 indicate the samples collected at the reference time, X_{-1} and Y_{-1} are the two inputs at the previous time, X_{-2} and Y_{-2} the values two time slices before the current one and so forth. The input variables can influence the output Z to the extent determined by the value of their multiplying coefficients (c,d,e,f etc).

The 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 improve when previous time slices are given as inputs. This is a regression problem, consisting of estimating the output Z of a system (or function) on the basis of the inputs X and Y. The results summarised in table I show this improvement in the regression capability when earlier time slices are given to the TDL networks.

The parameter used to quantify the increased in the success rate is the Root Mean Square Percentage (MSEP) error:

$$MSEP = \frac{\sum_{i=1}^{n} \left(\frac{Z_i - <Z_i>}{Z_i} \right)}{n} \tag{2}$$

where Z_i is the real value of Z, $<Z_i>$ is the estimated value of Z and n is the total number of samples. The MSEP is an absolute index and it is independent of the input range dimension. As an example, the values reported in table I indicate clearly that providing the TDL network with two additional time slices, corresponding to the memory effect generate by relation (1), has very beneficial effects. The improved performance testifies 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) but with a delay of three time steps has been used to generate synthetic signals. Time sequences up to four sequential time slices have been given to the "tapped delay line" 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 figure 2. The increase in performance, when the right number of time slices (three) is provided to the networks, is clearly seen as a minimum 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 (the results adding a forth time slice are shown in figure 2). This has been confirmed for all the various types of generating functions summarised in table I. It seems therefore that the TDL architecture is capable of identifying the right interval in which historical data are important.

Similar analysis has been performed to investigate the "memory effects capability" of the ERNN. Figure 3 shows the good capability of the ERNN to extract historical information from the input data obtained using relation (1) with a delay of three time steps with the generating function 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 train process. This behaviour has been confirmed for all the generating functions of table I. As for the TDL network, also the ERNN performance 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 the next two sections.

3. ASSESSMENT OF THE MEMORY EFFECTS BEFORE DISRUPTIONS

In this section, the two network architectures have been applied to the problem of identifying

4

disruptive discharges; this is typical classification problem that consists of determining which time slices in the database belong to discharges that are going to disrupt. The most relevant signals for disruption prediction, which have been retained for the study reported in this paper, are summarised in table II and were taken from literature as described in [7].

They were sampled at a sample rate of 20 ms and the entire database consists of 292 disruptive discharges and 220 non-disruptive cases.

For the results described in this paper, the signals reported in table II have been used as inputs to a set of 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 taking into account also the previous time slice, the third with data belonging to the two previous time slices and so on. The signals of the various time intervals have been multiplied by suitable weights, determined empirically to maximise performance and decreasing with increasing time to the disruption. The actual values of these weights are reported in the caption of figure 4; they are decreasing with the distance from the disruption, reflecting the fact that the information content of the time slices is decreasing the further 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, whose disruptions have been classified by the experts as due to a locked mode, has been used to train and then to test the TDL architecture. The reference time slice is between 300 and 320ms before the disruption. The performance of the network once earlier time slices are added as inputs is reported in figure 4. Including information of previous time slices (in the overall interval between 320 and 380 ms before the disruption) improves the performance of about 2%, which is quite significant given the high success rate of the network without historical data. In the figure the uncertainty intervals are due to the statistical fluctuations in the results obtained when changing randomly the training and test set. They do not have therefore to be considered error bars; when the training and test sets are kept constant, the improvements has always been consistently detected. The trend of the improvement in performance with time has been compared with the times before the disruptions 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 figure 5. The success rate of the TDL network increases significantly when the time slices corresponding to exactly 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.

To confirm these results, the same database has been analysed with ERNN networks. The indications about the memory effects are better than the ones derived from the tapped delay lines, as shown in figure 6. The improvements in the performance have again a maximum around 360 ms before the disruption. Moreover the improvement is even outside the confidence intervals due to the random choice of the training and test sets. The ERNN networks seem also to be capable of

detecting the second peak in the distribution of locked mode times, which is present around 420 ms before the disruption (see again figure 5). 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 due to 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 to cope and extract the details of the distribution function.

The same approach has been then applied to the entire database of JET disruptions, without any distinction about their causes. In this case, the 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 used database, there is not much information about an incoming disruption earlier than about 180 ms before its occurrence [8]. One example of the results is reported in figure 7 for the case of the TDL networks. 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 figure 7, 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. Therefore from the analysed 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, not clear indication of strong memory effects has been detected.

4. ASSESSMENT OF THE MEMORY EFFECTS AROUND THE TRANSITIONS FROM DIFFERENT CONFINEMENT REGIMES

Another important phenomenon, whose memory effects have been analysed with the neural networks described in section 2, is the transition between confinement regimes. A database of about 60 discharges has been prepared by the experts to provide clear and validated times for the transitions between the L and H mode. The most relevant signals to analyse this phenomenology have been identified with the nonlinear and unbiased method of the CART algorithm [9]. This is a supervised method which simply traverses the entire database to determine which variable and which value better divide the examples to be classified in two or more classes. After the most selective 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. The most

important quantities to study this type of plasma auto-organisation process are the Magneto-hydrodynamic energy, the Axial Toroidal Magnetic Field at 80% of the flux, the Electron Temperature, the Beta Normalized, the X-point Radial Position, the X-point Vertical Position [10]. The details of the database used can be found in [11]. 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. For both the training and the test sets, three couples of symmetric time windows around the transition have been defined (see figure 8 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. In this couple of intervals around the transition, the time slices have been chosen randomly for seven test sets, whereas an optimal training test 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 15ms, see figure 8) have been provided as inputs to the networks

The results indicate that historical information improves the performances of the networks only in the time interval [−100ms, 100ms] around the transition. Indeed, as can be seen in figure 8, only in this interval the improved performance is 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 it is likely to be the case for the intervals [−200ms, −100ms/100 ms, 200 ms] and [−300ms, -200ms/200ms, 300ms], historical information does not improve the performance of the networks and therefore memory effects seem to be not relevant any more. It then seems quite natural to conclude that some memory effects are present only very close to the transition. As in the previous section, the same database and the same training and test sets have been analyzed with ERNN network to confirm the results. The improving of the performance has been evaluated in the same time windows as the TDL case and the results are shown in figure 9 where again performance improves weakly and only in the time interval [-100ms,100ms] 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 seem to be not relevant any more. This result is coherent with previous experimental investigations [12], which have never found very strong evidence for hysteresis in JET plasmas.

5. PRELIMINARY CONCLUSIONS AND DIRECTIONS OF FUTURE INVESTIGATIONS

The potential of two neural networks architectures, the tapped delay lines and the recurrent networks of the Elman type, 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 for the disruptions preceded by a locked mode. For the general database, without discrimination about the causes, no statistically significant evidence of memory effects has been detected. With regard to the L to H transition, clear evidence of memory effects has been identified only for the time interval of +/-100ms around the time of the transition. Further 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.

With regard to the continuation of this line of research, other phenomena could be investigated. Among the most interesting, a part from the H-to-L transition, could be the formation of the various internal transport barrier, 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 theoretic techniques, based on signal entropies or conditional probabilities, could also be considered to investigate their potential to identify memory effects in time series signals.

REFERENCES

[1]. F.C. Schuller, "Disruptions in tokamaks," Plasma Phys. Control. Fusion, vol. **37**, no. 11A, pp. A135–A162, Nov. 1995.

[2]. Wagner, F. et al., Physical Review Letters **49**, 1408 (1982)

[3]. Fitzpatrick R. 1998 *Phys. Plasmas* **5** 3325

[4]. Thomas D. M. et al, ***Plasma Phys. Control. Fusion* 40 No 5 (May 1998) 707-712**

[5]. Mandic D. P. and Jonathon A. Chambers, *"Recurrent Neural Networks for Prediction"* John Wiley & Sons Ltd ISBNs: 0-471-49517-4

[6]. Rumelhart D.E., Hinton G.E. and Williams R.J. Nature 1986 323:533-536

[7]. Murari A. et al., IEEE Transactions on Plasma Science, Vol. **34**, No.3, June 2006

[8]. Murari A. et al, Nucl. Fusion **48** (2008) 0350

[9]. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1984 *Classification and Regression Trees* (Belmont, CA: Wadsworth Inc.) (1993, New York: Chapman and Hall)

[10]. G. Vagliasindi et al, Proceedings of the 11th European Symposium on Artificial Neural Networks (ESANN'08), Bruges (Belgium), 23-25 April 2008, pp. 517-522.

[11]. Meakins, A.J. (2008) "*A Study of the L-H Transition in Tokamak Fusion Experiments*". PhD Thesis. Imperial College London.

[12]. Y. Andrew et al., *Plasma Phys. Control. Fusion* **50** No 12 (December 2008) 124053 (8pp)

GF		MSEP (%)			MSEPmem (%)		
		All	Train	Test	All	Train	Test
1	`x=-5:0.1:5;` `y=-5:0.1:5;`	1.38	1.24	1.67	0.35	0.0000	1.03
2	`x=sin(-5:0.1:5);` `y=cos(-5:0.1:5);`	8.82	3.12	20.04	0.90	0.0000	2.67
3	`x=exp(-5:0.1:5);` `y=1./exp(-5:0.1:5);`	85.51	84.20	88.11	68.52	62.65	80.10
4	`x=tan(-5:0.1:5);` `y=sin(-5:0.1:5);`	68.24	55.03	94.65	0.24	0.20	0.67
5	`x=rand(1,100);` `y=rand(1,100);`	289.96	294.29	281.17	0.06	0.05	0.07

Table 1: Improvement of the predictions by TDL networks when historical information is provided. The historical evaluation has been performed for a memory effect of two time steps, i.e. two time slices before the reference time. In the first three columns the results obtained by the network without historical information are shown; the last three columns 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 X and Y

SIGNAL NAME	UNIT
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 Diamag. Energy Derivative dW_{dia}/dt	[W]
Safety factor at 95% of minor radius q_{95}	
Poloidal beta β_p	
Net power P_{net}	[W]

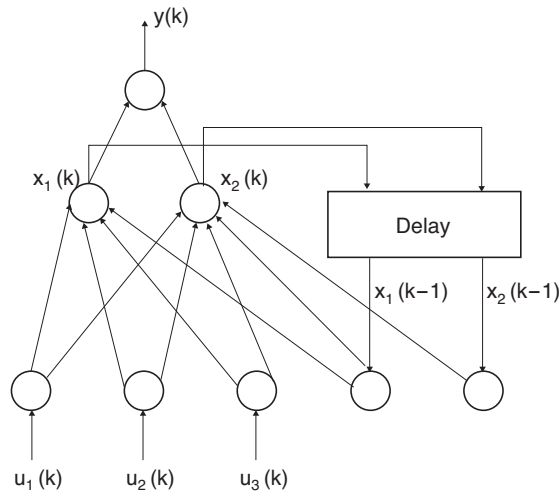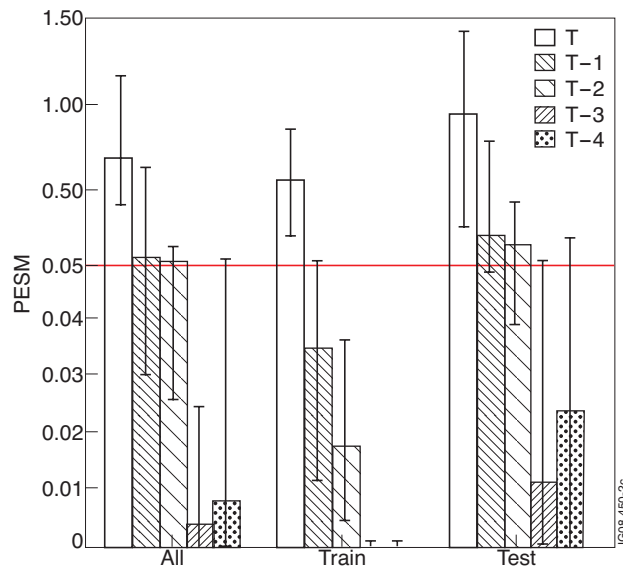Table 2: List of the signals used as predictors for the classification trees

Figure 1: Topology of the recurrent networks of the Elman type (ERNN) showing the internal feedback with delay. The symbol u identifies the inputs and the symbol x the internal status of the neurons in the intermediate layer.


	T	T−1	T−2	T−3	T−4
All	68.24	10.03	7.07	0.37	0.81
Train	55.03	3.48	1.75	0.01	0.01
Test	94.65	23.15	17.73	1.11	2.41

Figure 2: Evolution of the TDL classification errors for the system described by relation (1) with the generating function GF4 of the table I. The memory effect used to generate the synthetic data extends for three time slices. Two different scales were used for the plot in order to represent in a clearer way the data at low values of MSEP. The red line show the point at which there is the change in scale. The errors are expressed in percentage.

10
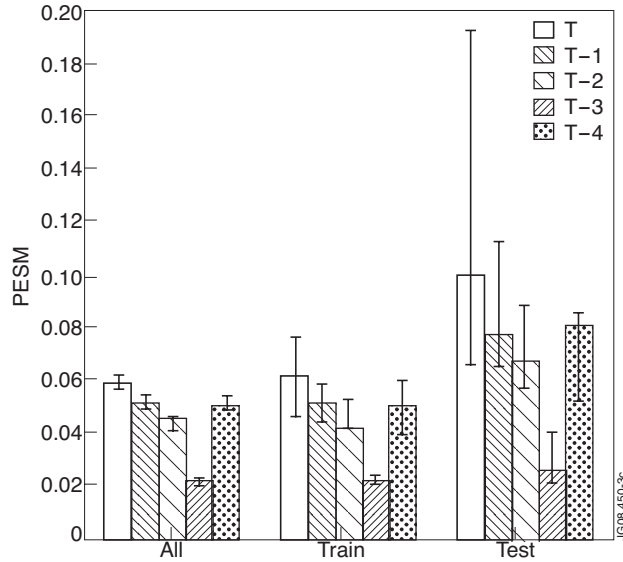
	T	T−1	T−2	T−3	T−4
All	5.84	5.06	4.51	2.14	5.05
Train	6.1	5.01	4.17	2.16	5.08
Test	9.95	7.69	6.85	2.58	8.03

Figure 3: 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 three time slices. The errors are expressed in percentage.
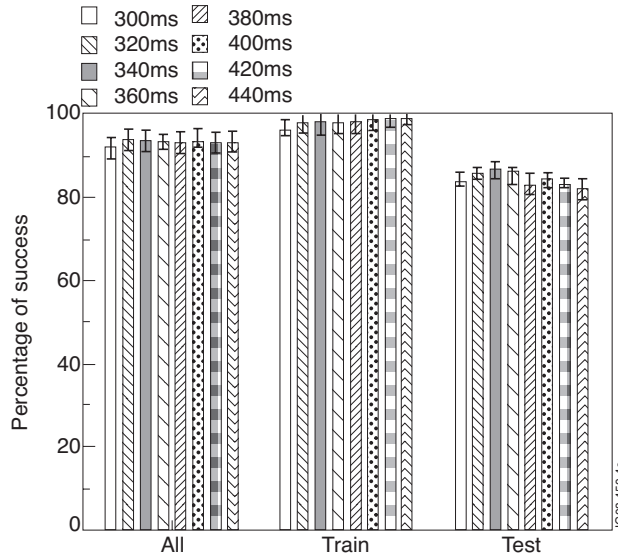
	300ms	320ms	340ms	360ms	380ms	400ms	420ms	440ms
All	92.0699	93.883	93.8121	93.6511	93.284	93.6702	93.3462	93.4043
Train	96.5258	98.1157	98.205	97.8854	98.2166	98.9119	98.8588	99.0064
Test	84.0232	85.8974	86.8298	86.2637	82.9837	84.6154	83.3333	81.9444

Figure 4: Improved performance of TDL networks with historical inputs for the case of disruptions triggered by a locked mode. The colour code indicates the times before the disruption 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 340 ms and 0.7 for the time slice 360 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 or not disruptive discharge.
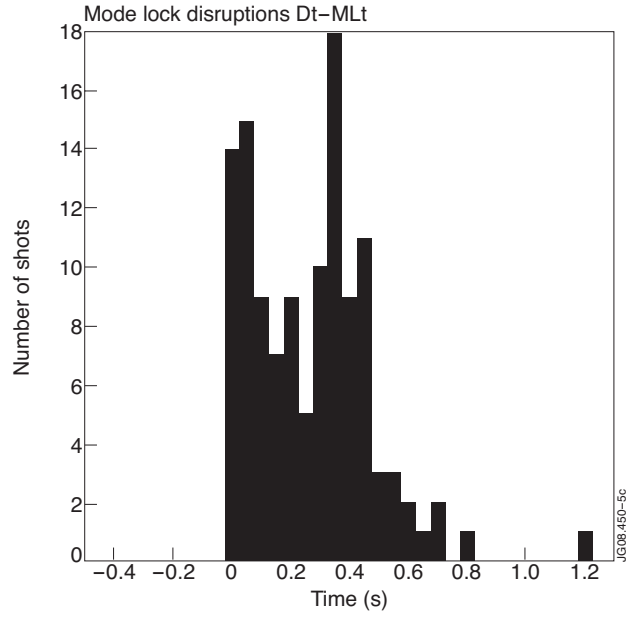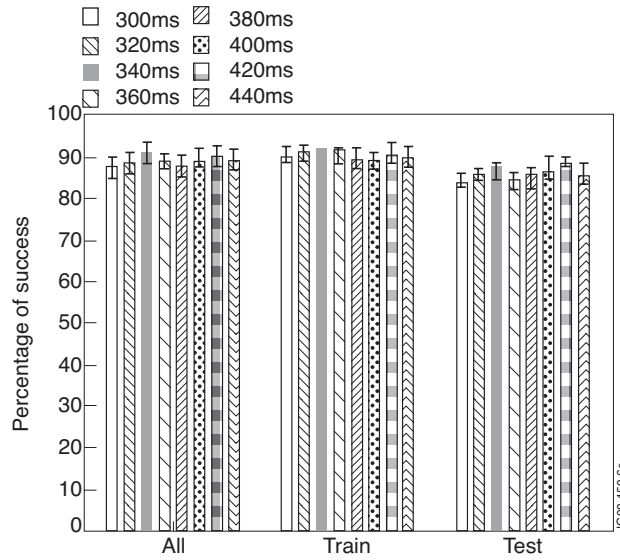
Mode lock disruptions Dt–MLt

Figure 5: Statistical distribution of the time which 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.
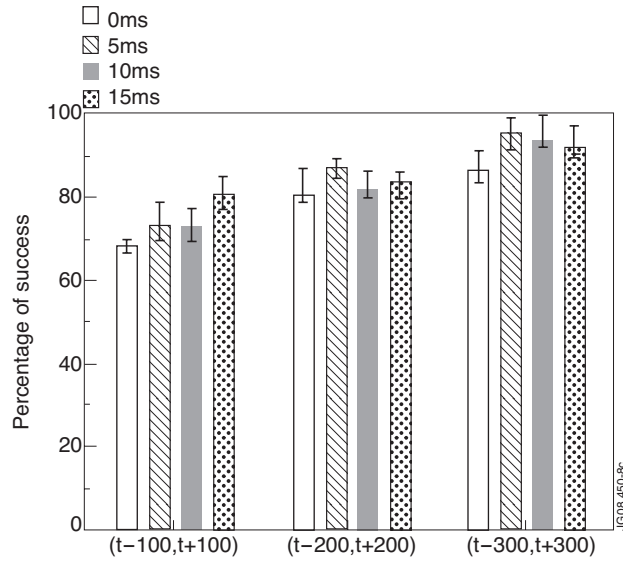

	300ms	320ms	340ms	360ms	380ms	400ms	420ms	440ms
All	87.0699	88.1277	90.0426	88.8936	87.6596	88.383	89.4468	88.5532
Train	89.5258	90.8705	91.8544	91.4013	89.2994	88.899	90.2548	89.5701
Test	84.2516	85.3896	87.6623	84.4156	85.7143	86.2709	87.7922	85.1577

Figure 6: Improved performances of ERNN networks with historical inputs. The same database and the same notation as in figure 5 have been used.
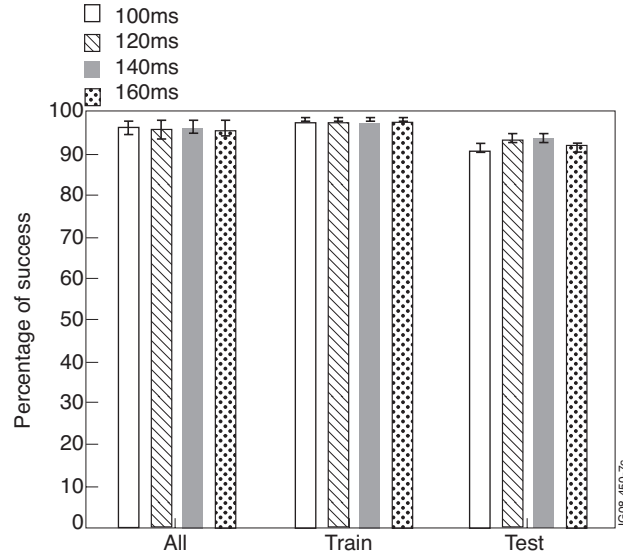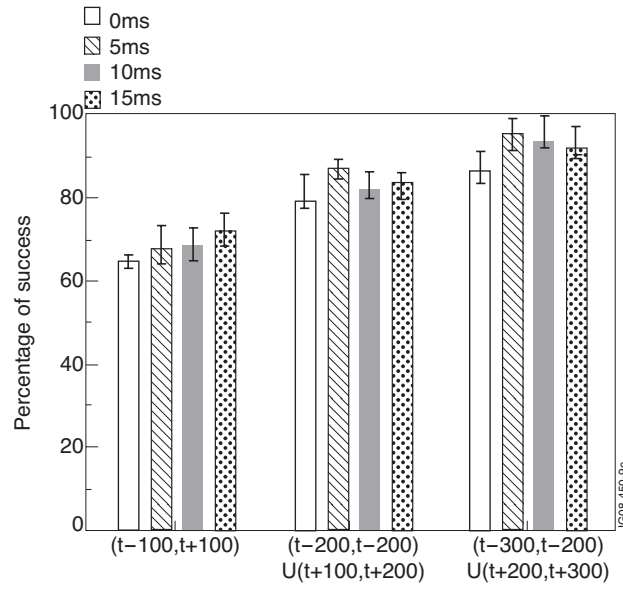
	100ms	120ms	140ms	160ms
All	96.7692	97.3516	97.75	97.0296
Train	99.5723	99.9648	100	100
Test	91.2573	92.4269	93.3723	92.1888

Figure 7: Performance of TDL architectures with historical inputs. No selection on the type of disruption has been performed. The nomenclature in the figure and the method to randomly select the various sets of discharges are the same as in figure 1.The results for the test set do not show any significant improvement

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

Figure 8: 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 which are properly classified as belonging to the L or H phase of the discharge. Intervals of various lengths around the transition and different integration times have been considered.

13

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

Figure 9: Success rate of the ERNN for the same database used in figure eight. The results are confirmed: the success rate improves only for the interval close to the transition.

14