PROCEEDINGS REPRINT

SPIE—The International Society for Optical Engineering

Reprinted from

Applications of Soft Computing

28–29 July 1997
San Diego, California

SPIE PROCEEDINGS SERIES

Volume 3165

Neuro-Fuzzy System for Chaotic Time Series Forecasting

Francesco Masulli & Léonard Studer

Istituto Nazionale di Fisica della Materia and
Dipartimento di Fisica - Università di Genova
Via Dodecaneso 33 - 16146 Genova - Italy
email : {masulli| studer}@ge.infm.it

ABSTRACT

We report on an on-going study to assess potential benefits using soft computing methods in forecasting problems. Our goal is to forecast natural phenomena represented by time series that show chaotic features. We use a Neuro-Fuzzy System for its ability to adapt to numerical data and for the possibility to input and extract expert knowledge expressed in words. We present results of experiments designed to study how to shape a Neuro-Fuzzy Systems to Forecast Chaotic Time Series. Our main conclusions are: 1) The Neuro-Fuzzy System is able to forecast a synthetic chaotic time series with high accuracy if the number of inputs and the time delay between them are chosen adequately. 2) The Takens-Mane theorem from chaos theory gives an useful lower bound on the minimal number of inputs. 3) The time delay between the inputs can not be set a priori. It has to be tuned for every different time series. 4) The number of fuzzy rules seems related to the size of the learning set and not to the structure of the chaotic dynamical system. We tentatively try to interpret the rules that the Neuro-Fuzzy System has learned. Finally we discuss the adequacy of the whole set of fuzzy rules to forecast locally the dynamical system.

Keywords: Non-Linear Time Series, Forecasting, Chaos, Neuro-Fuzzy Systems

1 INTRODUCTION

Time series measurements are widespread in science and engineering. The obvious advantage of this sort of measurement lies in the easy way to gain information on the system under consideration. For well-behaved systems, let's say linear systems, a complete analysis methodology is available. But it happens that not only linear systems are encountered in practice. In fact, certainly the majority of systems could show a wild behavior.

In the past, engineering practice and scientific research tried to avoid theses not well-behaved systems. But with the advent of computers, they began to be deeply studied, and a new land of science and technology opened. The field is young enough to be known under various name: chaos, non-linear systems, complex systems and so one.

Apart of this new land, computers allows also to study new methods in knowledge science and engineering. Knowledge acquisition and management becomes an experimental area where new paradigms can be tested rather easily. The knowledge related science is, in part, the ground subject of artificial intelligence (AI). Paradoxically, it appeared in AI that very complex and potentially very useful systems can constructed from the aggregation of simple elements as in the case of Artificial Neural Networks. It appeared also that sacrificing accuracy for uncertainty brings tractability in otherwise effectively unsolvable problems.

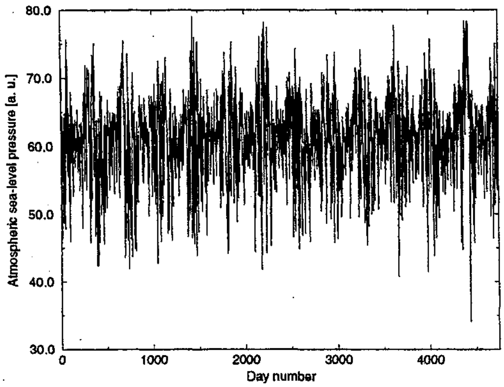


Figure 1: Time series of daily atmospheric sea-level pressure. The measurements were recorded at Chiavari (It) from 1977 to 1989. Unit of atmospheric pressure is arbitrary.

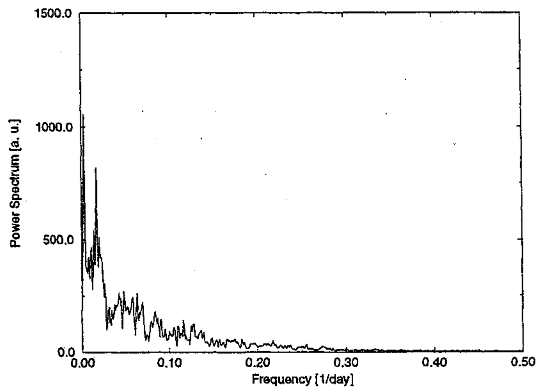


Figure 2: Power Spectrum of the pressure time series. A Blackman-Tuckey method with a Welch window has been used to compute the power spectrum. Notice the annual and the 50 days oscillations.

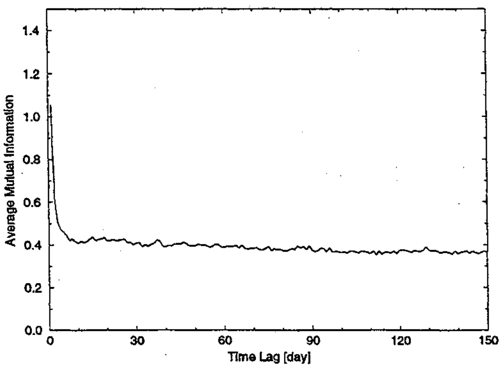


Figure 3: Average Mutual Information (in bits) of the pressure time series. The most interesting (non-linear) mutual information dependence is in the time lag range of 1-6 days. Quantitatively, the average mutual information is low (0.5 to 1 bit).

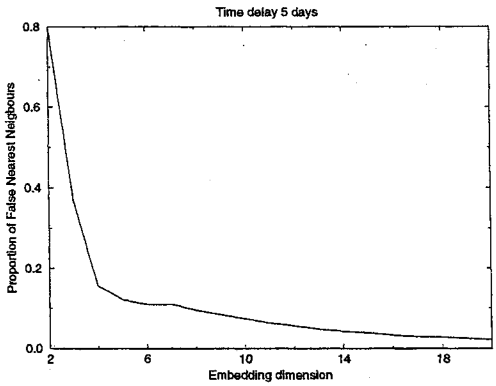


Figure 4: False nearest neighbours in function of the embedding dimension for the pressure time series. From this figure, it is clear that the dimension of the reconstructed space has to be greater than say 6 to unfold its attractor, if any deterministic signal underlies this time series. Keeping in mind previous figures, a important noise contribution has to be expected for this time series.

It can then be foreseen that at the core of chaos, computational experimentation and artificial intelligence lie some intellectual gems. As these lands are still rather new, it is somewhat uneasy to know how to move from one point to another. A kind of detailed road-map is still missing to help us.

In this paper, we show what kind of rocks we expect to be able to process to extract gems, what kind of cliffs can be met and we present some ways how to climb these barriers to better see the landscape of these new lands.

Let's present the typical rocks suspected to contain diamonds. In Fig 1, a time series of daily atmospheric sea-level pressure is shown. This time series has been recorded at Chiavari, a little city in the Italian Riviera. By visual inspection, some gross structure can be seen, such as a seasonal variation of the spread of the pressure. How are we lead to suspect some gems in this measurement ? First of all, some periodicities are known to exist in meteorological time series, enough to cite the famous El Niño phenomenon which is a multi year oscillation in the South Pacific. The power spectrum shows significant frequencies (Fig 2). Next, the Mutual Information Function presents some structure showing an information-theoretic mutual dependence on measurements separated by some time lag (Fig 3). Finally, a False Nearest Neighbors Analysis also presents some hint that the time series is more structured that its apparent randomness (Fig 4). All these methods will be briefly presented in our paper and/or, in these proceedings, in the papers by Abarbanel, Ghil, Gilmore or Morabito.[1–4]

To explore these new lands, we undertook a study to assess potential benefits using soft computing methods in forecasting problems.[5,6] Our goal is to forecast natural phenomena represented by time series that show chaotic features. We use a Neuro-Fuzzy System for its ability to adapt to numerical data and for the possibility to input and extract expert knowledge expressed in words.

In the second section of this paper, we present relevant elements picked from chaos theory. Next, we describe the Neuro-Fuzzy Systems. The fourth section is devoted to the description of our computer experiments. To make things as clear as possible, we use here a synthetic chaotic time series based on the Mackey-Glass equation. Results of our experiments are then presented. In the fifth section we tentatively present what kind of rules have been learned by the NFS and show how they fit in a AI perspective. Finally, we conclude and present some possible new directions.

2 ELEMENTS FROM CHAOS THEORY

A deterministic dynamical system is described by a set of differential equations. Its evolution is represented by the trajectory in phase space (of dimension N) of the vector $\mathbf{Q} = (x, \dot{x}, y, \dot{y}, z, \dot{z}, \ldots)^{\mathsf{T}}$ where $x, \dot{x}, y, \dot{y}, z, \dot{z}, \ldots$ are the variables of the system. The figure made in phase space by $\mathbf{Q}$ (after an initial transient period) is the attractor of the system. Strange attractors are typical of chaotic systems. These are figures whose geometry is fractal.

For non-linear systems such as chaotic ones, the dynamical variables $(x, y, z \ldots)$ are coupled. The evolution of one variable (let say x) is not independent of all the other ones $(y, z \ldots)$. Except for few simple phenomena, the set of differential equations is unknown. As variables are interdependent, the observation of only one of those brings information — maybe in an implicit way — on the other ones and consequently on the complete dynamical system. This coupling is the reason why time series of non-linear systems are so useful in the study of the latter.

Embedding

The question is: "How to reconstruct the complete dynamical system with only the one-variable time series $(x_1, x_2, x_3 \ldots)$?" Embedding is the method to solve this problem. In this paper, we use specifically the *time* *delay embedding* method. It consists in building m-dimensional vectors $\mathbf{x_i} = (x_i, x_{i-\tau}, x_{i-2\tau} \ldots x_{i-(m-1)\tau})^{\mathsf{T}}$.

In principles, it could suffice that $m \geq N$. But in general, the *effective* dynamical dimension is not directly related to the order of the dynamical system of dimension N. For example, this is the case for weak coupled variables.

There is, in chaos literature, a theorem based on topological considerations. It is due to Takens and Mañé and has been extended by Sauer *et al.*[7] It links the minimal embedding dimension d to the fractal dimension D of the strange attractor. If $d > 2D$, a fair reconstruction of the dynamical system is achieved in the sense that the reconstructed vectorial time delayed trajectory doesn't intersect itself and so doesn't give problems of ambiguity in its evolution. It is then judicious to set m — the dimension of the reconstructed vector — greater or equal to d — the value issued from the Takens-Mañé theorem. Of course, values in the range $[N, d[$ are also possible for m; the Taken-Mañé theorem is simply silent in this case.

The true dimension N of the dynamical system is lost by the process of measurement of the scalar time series. This is in analogy in looking to a planar circle (i.e. $D = 1$) from one axis; it will appear as a line. In other words, a point on the circle is given by a pair of coordinates. In this analogy, the scalar time series will be limited to a series of *one* of the both coordinates. The Takens-Mañé theorem states then that a list of 3-tuples built form the scalar time series will be a fair representation of the original figure. This fair representation could be, in our example, a closed loop in 3D space or even a (mathematical) knot, but without self-crossing !

In chaotic systems, two initially close trajectories will diverge exponentially in time. This makes the chaotic system very unstable. It also implies that the time delay τ used in the embedding is a parameter which has to be chosen carefully. If it is too long, the scalar variables $x_i, x_{i-\tau}, x_{i-2\tau} \ldots x_{i-(m-1)\tau}$ are not correlated and so the chaotic dynamical system can not be reconstructed. If it is too short, every scalar is essentially a copy of the previous one, bringing very little information on the dynamical system.

Chaos literature gives *prescriptions* on the optimal time delay τ to choose.[8] These prescriptions are not firmly grounded in theory and have to be seen as educated intuition. For example, τ can be the first zero of the autocorrelation function (FZA) or the first minimum of the average mutual information (MMI) of the time series.

As the system is deterministic, the evolution of state space point is well defined. So, by considering nearby points, the dynamical dimension N can be inferred by the following method. The points to be considered are built from the times series by an embedding. If the dimension of the embedding is too small, it will happen that some proportion of the pairs of point which are nearest neighbours will evolve to two very different future points, possibly loosing their neighbourood relations. By counting the proportion of False Nearest Neighbours in function of the embedding dimension a convenient embedding dimension m can be made evident (see[8,1]).

The characteristic feature of chaotic time series is that they show an exquisite dependence to initial conditions. Two initial points, even only slightly separated, will diverge from each other at a positive exponential rate in time. This divergence of trajectories makes infeasible a perfect forecaster, *i.e.* a forecaster that, given a finite set of initial conditions, will perfectly predicts the value of the system at any future time. Forecastings can only be made up to a limited time horizon. Nevertheless, building the best (but limited) forecaster for a given time series is not a definitely solved problem.

3 NEURO-FUZZY SYSTEM AS FORECASTER

Neuro-Fuzzy System (NFS)[9] have been introduced to easy the problem of tuning the parameters of fuzzy logic systems when big sets of numerical data are available. The structure of NFS can be made similar to Artificial Feed-Forward Neural Networks which have known capacities to learn complex relations in numerical data set.

On an other hand, the fuzzy set aspect of the NFS makes it very convenient to absorb (before the learning procedure) or give back (after the learning procedure) some linguistic knowledge. NFS can then been seen as an interface tool between numerical data set and refined knowledge.

The use of a NFS as a forecaster is appealing, because of what may be catch in the linguistic relations that it has learned. Nevertheless, they are aspects from chaos theory that can be used to shape *a priori* the NFS. In effect, it is no sense to forget what is already known from chaos theory when using a NFS-forecaster. In the previous section, we have already put forth the embedding dimension m as a very relevant element to be taken into account. We will see that m is linked to the number of input of the NFS.

Description of the Neuro-Fuzzy System

The Neuro-Fuzzy Systems is a function of the form $y = f(x_1, x_2 \ldots x_K)$. The NFS contains a set of fuzzy rules. The the j-th rule[1] is:

$$\begin{array}{rl} \textbf{IF} & x_1 \text{ is } \mathrm{A_j} \quad \textbf{AND} \quad x_2 \text{ is } \mathrm{B_j} \quad \textbf{AND} \quad \ldots \\ & \ldots \quad \textbf{AND} \quad x_K \text{ is } \mathrm{G_j} \\ \textbf{THEN} & y_j \text{ is } \mathrm{O_j}, \end{array}$$

where A_j, B_j, G_j and O_j are fuzzy sets. The membership $\mu_{jk}(x_k)$ of an element x_k to a fuzzy set A_j (x_k is $\mathrm{A_j}$) is given by a gaussian:

$$\mu_{jk}(x_k) = \exp\left(-\frac{(x_k - m_{jk})^2}{2\sigma_{jk}^2}\right), \tag{1}$$

where m_{jk} and σ_{jk}^2 are the means and the variances.

The **AND** connective is made by a product operation:

$$r_j = \prod_k \mu_{jk}(x_k), \tag{2}$$

where r_j is the activation results of the j-th rule, given the K inputs $\{x_k\}$.

The form of the fuzzy set O_j in the consequent part of each rules is arbitrary. We use fuzzy singletons s_j, *i.e.* $\mu_{O_j}(y_j) = 1$, if $y_j = s_j$.

By using singleton fuzzifier and height defuzzification,[9] the total output y is given by the weighted sum of the output fuzzy singletons. The weights are the activations r_j:

$$y = \frac{\sum_j r_j s_j}{\sum_j r_j} \tag{3}$$

It can be shown that the structure of this NFS can be mapped on the structure of a Radial Basis Function Network with K inputs, J hidden nodes (one node by rule) and 1 output node.[10]

With the above choices of membership function, product operator inference and sum composition rule, the NFS has been demonstrated to be an Universal Approximator.[11] In other word, it is a function that can approximate to arbitrary precision any continuous function (in R^K) on a compact subset of R^K. This property is a an important theoretical result, as it firmly grounds in mathematics the use of this tool as forecaster.

[1]Index k ranges from 1 to K. K is the total number of inputs. Index j ranges from 1 to J. J is the total number of rules.

Learning Numerical Input-Output Relations

The inputs-output relations computed by the NFS depend on the parameters of the gaussians (m_{jk} and σ_{jk}^2), the fuzzy singletons (s_j) and the number of inputs, rules and outputs ($K/J/1$). The exact values of these parameters are set during the learning phase.

In the learning phase, pairs of inputs-output are presented to the NFS. The actual output is computed and compared to the requested one. The difference is then back-propagated through the NFS after each presentations (on-line learning scheme). Formula of the gradients of m_{jk} and σ_{jk}^2 and s_j with respect to the difference are given elsewhere in the literature.[5,11] We used an adaptive learning-rate scheme, as proposed by Vogl *et al.* to speed up convergence in the learning phase.[12]

The data set consists of short sequences extracted from from the time series and grouped in patterns (or vectors). Of course, the input patterns are the m-dimensional vectors $(x_i, x_{i-\tau}, x_{i-2\tau} \dots x_{i-(m-1)\tau})$ built from the scalar time series by the embedding reconstruction method described above. The output pattern is simply the evolved point of the scalar time series i.e. x_{i+1}.

After that learning phase, the NFS is able to compute the input-output relation. Presenting new unseen inputs, the NFS then computes output that is the forecasted estimation from the sequence of inputs. The results achieved during the learning phase are checked by computing the root mean squared error (RMSE) on a test set of patterns independent of the learning set.

4 EXPERIMENTS

In this section we want to test the conjectures we mentioned above by some computer experiments. In particular, we want to show that the bound $d > 2D$ given by the Takens-Mañé theorem is observable with a NFS forecaster. Also, we want to show the dependence of the forecasting in function of the time delay τ. And finally, as the number of rules of the NFS remains a free parameter, we study its optimal value by a systematic scan.

For the experiments reported here we have used the Mackey-Glass equation[13]:

$$\dot{x}(t) = \frac{ax(t-\Delta)}{1+x(t-\Delta)^c} - bx(t), \tag{4}$$

with $a = 0.2$, $b = 0.1$ and $c = 10$.

Varying Δ from 17.0 to higher values allows to control the chaoticity of the generated time series. More exactly, the fractal dimension of the strange attractor is $D_{17} \approx 2.1$ for $\Delta = 17$ and raises to $D_{30} \approx 3.6$ for $\Delta = 30$.[14] As the actual value $x(t)$ is related to the past value $x(t-\Delta)$, the values x between $t-\Delta$ and t can be anything, leading to an infinite-dimensional system.

The Mackey-Glass equation was integrated by a Runge-Kutta method with a time step of 0.001, with $x(t < 0) = 0$ and $x(0) = 0.5$. The produced time series was then decimated by a factor 1000 and the the first 1000 values were discarded to let the transients die out (see figure 8 bottom).

6000 input patterns were presented in the learning phase and 1000 patterns made the test set. The learning procedure was left going for 500 epochs. At the end of the learning phase, the NFS was reaching a plateau. The initial decrease in convergence was fast, usually about 20 epochs were sufficient to achieve 90% of the total decrease of the RMSE observed in 500 epochs. Also, no sign of over-fitting were observed. Initial value for m_{jk}, σ_{jk}^2 and s_j were set at random.

With anticipation on the results presented in the 3 following paragraphs, we notice that the NFS can achieve an accurate modeling and forecasting of the Mackey-Glass time series. Anyhow, the quality of forecasting depends strongly on the actual values of the parameters (K, J, τ).

Number of inputs K

In a first experiment, we varied the number of inputs K. The NFS has here the structure K inputs, $2K$ rules (or hidden nodes), 1 output ($K/2K/1$). The time delay between the inputs was set to $\tau = 5$. As can be seen in the figure 5, a sharp reduction of RMSE is seen if the number K of inputs is greater than a threshold.

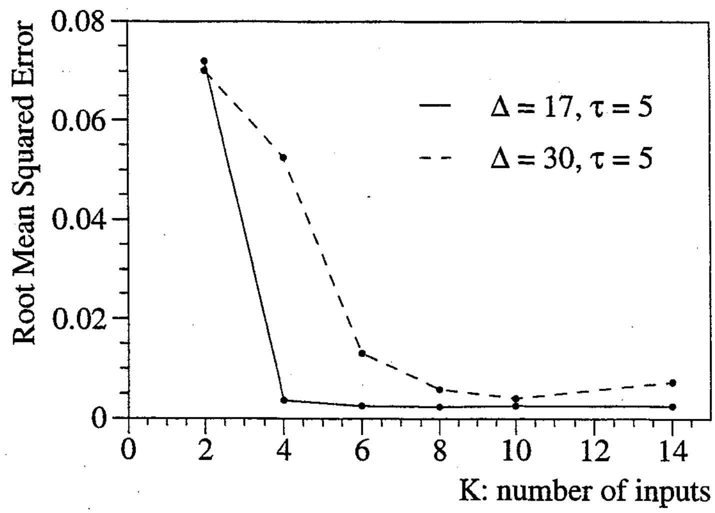


Figure 5: Decrease of the Root Mean Squared Error in function of the number K of inputs of the Neuro-Fuzzy System. Time delay τ is set to 5. The number of rules is set to 2 times the number of inputs. The Mackey-Glass time series are integrated with $\Delta = 17$ (solid line) or $\Delta = 30$ (dashed line).

The appearance of the sharp drop in figure 5 is clearly related to the Takens-Mañé bound which is $K_{17} \geq 5$ for the $\Delta = 17$ time series and $K_{30} \geq 8$ for the Mackey-Glass system with $\Delta = 30$.

Time delay τ

In a second experiment, the structure of the NFS was fixed to 6/12/1 (for a Mackey-Glass time series with $\Delta = 17$) or 10/20/1 (for a Mackey-Glass time series with $\Delta = 30$). In the figure 6, a mild rise of RMSE can be detected with increasing τ.

Unfortunately, concerning the optimal time delay τ, the prescriptions cited in the second section give no sensible results as the MMI point and the FZA point correspond both to $\tau = 11$. On the figure 6, it is clear that the optimal value for τ is 5 and not 11. Other prescriptions are given in the literature but clearly no definite answer exists and the optimal τ has to be tuned for every time series. It is also expected that noise should play a role.

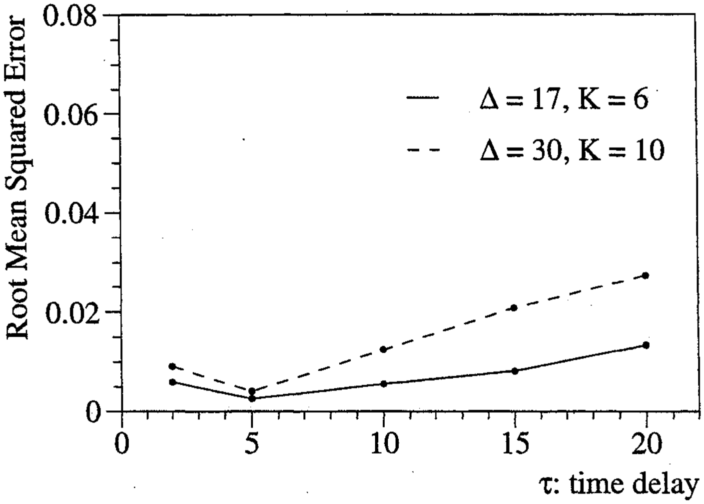


Figure 6: Mild increase of the Root Mean Squared Error in function of the time delay τ between the inputs. The Neuro-Fuzzy System is 6/12/1 for Mackey-Glass time series with $\Delta = 17$ (solid line) and 10/20/1 for a $\Delta = 30$ - Mackey-Glass time series (dashed line).

Number of rules J

In a third experiment, the number J of rules was varied from 2 to 18. We fixed the number of inputs to 6, the output was set to 1 and $\tau = 2$. A Mackey-Glass time series with $\Delta = 17$ was used. The RMSE in function of J is displayed in figure 7. RMSE decreases with increasing J and for $J = 14$ onwards reaches a plateau. In a complementary experiment (not displayed here) the number of presented patterns was reduced by a factor of 6. The RMSE values were then higher than what is displayed here but a plateau could also be detected for number of rules higher than 8.

The observed plateau beginning at $J = 14$ in figure 7 could be related to the structure of the strange attractor and the reason follows. Each rule is effectively active only in a local region of the input space. If the number of rules is insufficient, the NFS can not map the relevant details of the structure of the strange attractor; NFS captures only the gross features, leading to poor RMSE.

Our complementary experiment showed that the plateau began at a lower number of rules, $J = 8$. In this second case, the number of presented input patterns was smaller than in the first one, but the time series was extracted from the same $\Delta = 17$ - Mackey-Glass system. The strange attractor was then the same in both cases. This contradicts the hypothesis that the plateau phenomenon is related to the attractor.

The chaos literature is not very informative about the optimal number of rules. One prescription is to minimize a cost function which is the sum of the RMSE term and of a term proportional in the number of parameters. Again, it is only a prescription, this one based on the theory of information.

Let us recall that strange attractors show correlations at the local level. We suspect that the use of the whole fuzzy rule base is not well suited to capture this local dynamic. By construction, the gaussian fuzzy rules give a significant activation only for localized input domains. But the whole set of fuzzy rules contribute to the total output. In a sense, there is an interference between the global structure features (represented by the whole set of

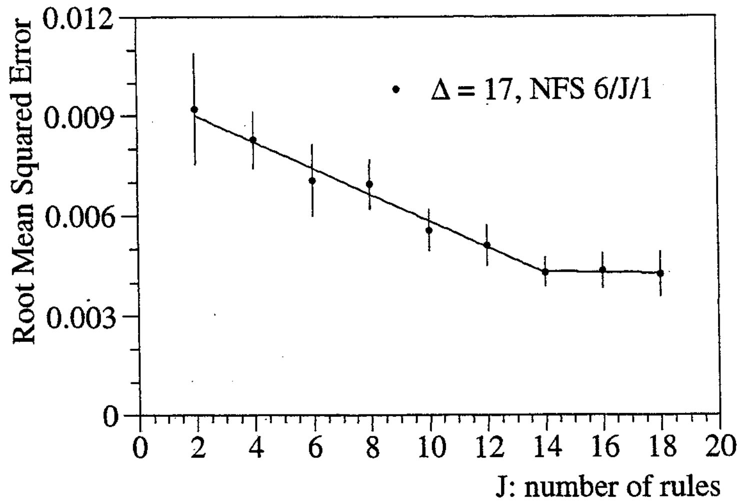


Figure 7: Decrease and plateau of the Root Mean Squared Error in function of the number of rules of a $6/J/1$ Neuro-Fuzzy System. Time delay τ between the inputs is set to 2. Mackey-Glass time series with $\Delta = 17$ are used. 6000 inputs patterns have been presented to the NFS during the learning phase.

rules) and the local structure (represented by the gaussians).

5 EXAMPLE OF LEARNED RULES

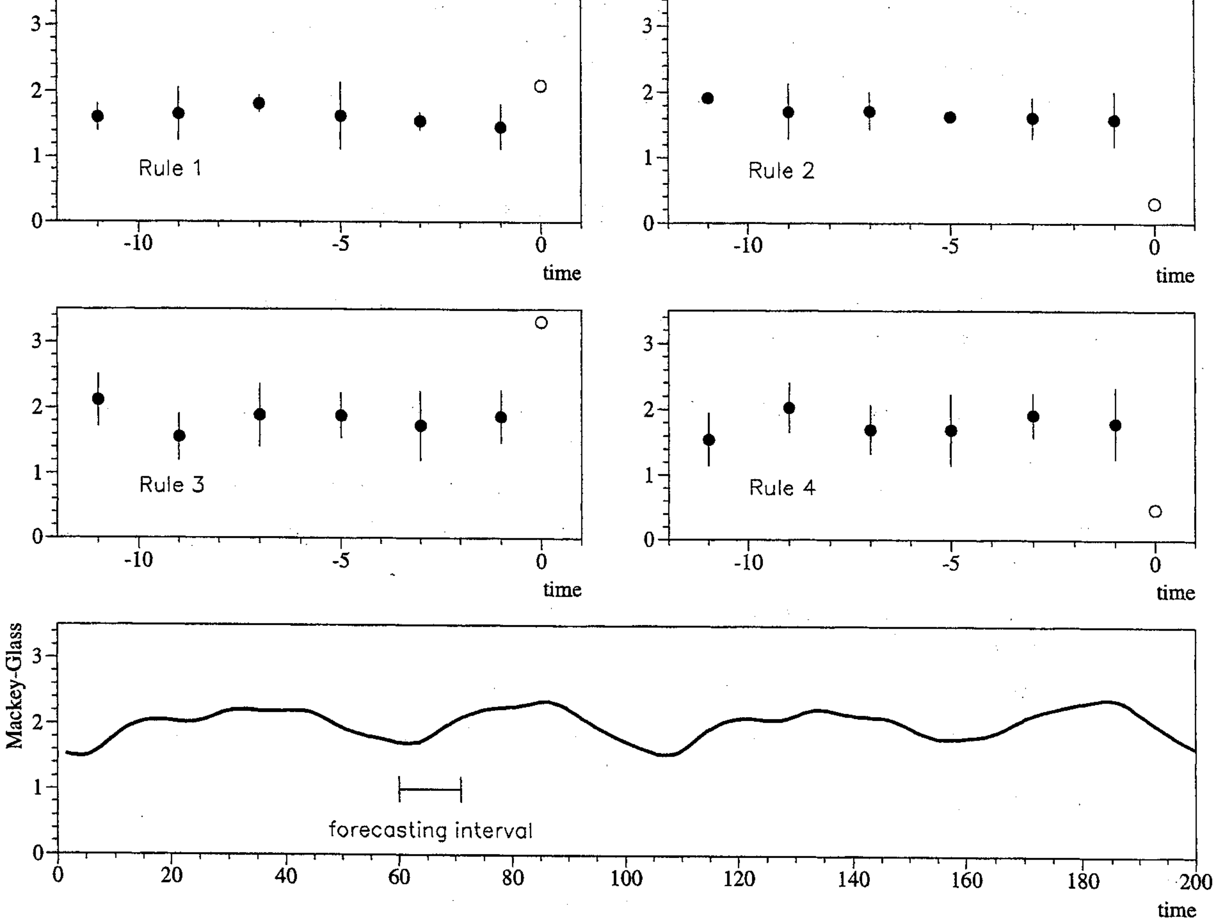
The fuzzy rules learned by the NFS can now be tentatively deciphered. In the figure 8 we present the 4 rules learned by a 6/4/1 NFS and a close view of the time series to be learned.

For rule 2, the first gaussian (at time $= -11$) and the fourth gaussian (at time $= -5$) have very small spreads (0.025 and 0.00002 resp.). This puts an effective constraint on the corresponding input points x_1 and x_4. They have to be very close to the centers of the respective gaussians to make the activation of the rule effectively non-zero. The other 4 gaussians have bigger spreads.

The linguistic version of rule 2 would then read: "IF $x_1 = 1.9$ AND $x_2 \approx 1.71$ AND $x_3 \approx 1.71$ AND $x_4 = 1.63$ AND $x_5 \approx 1.61$ AND $x_6 \approx 1.59$ THEN $y = 0.31$". Similar linguistic interpretations can easily be written for the other 3 rules.

We note that the fuzzy singletons of rules 2, 3 and 4 are out of the range of Mackey-Glass time series. This indicates that there are compensation effects between the rules to achieve an output in the range of the time series.

Of course, it remains to show how to use this extracted linguistic knowledge content in practical cases. We have found during our study that the meaningful use of learned rules is not as easy as we expected. One reason probably lies in the spread of the membership functions. Another reason is related to the particular choice of embedding we used. Time delay embedded vectors, although very convenient to compute, are not directly related to dynamical variables. Here, it is expected that a differential embedding (by example) would be more useful.



ure 8: Example of the rules learned by a 6/4/1 Neuro-Fuzzy System trained on inputs patterns from a $\Delta = 17$ - ckey-Glass time series. The 4 upper quadrants display the 4 rules. Each rule is characterized by 6 gaussians, h with its center value m_{jk} (solid dots) and spread σ_{jk} (vertical line). If the six inputs $x_1 \ldots x_6$ are equal to center values, the output activation for that rule is equal to the fuzzy singleton s_j displayed here by a open le at time=0. The wide quadrant at the bottom displays a partial view of the Mackey-Glass series. The time erval covered by the 6 inputs used in forecasting is also indicated.

Correlation between rules

We have noticed that some rules seem linearly correlated (or anti-correlated) to other ones. Rule 3 and rule 4 are manifestly anti-correlated in figure 8. As we don't know the distribution of m_{jk}, we have chosen to compute the Spearman's rank order correlation coefficient r_s which is a robust statistical estimator of the correlation coefficient and independent of the distribution of m_{jk}. For rule 3 and rule 4, we compute $r_s = -0.943$.

More generally, on a set of J rules, $H = (J^2 - J)/2$ distinct pairs of rule can be formed. Comparing the rank order of the K centers of the gaussians of both rule, we can compute the Spearman's r_s. With a confidence level of 95% (to not be wrong), about 10% of the H pairs are correlated whatever is the number of rules J. This indicates that there is some redundancy in the set of the learned fuzzy rules.

6 CONCLUSION

In this article we have presented results of three experiments designed to study how to shape a Neuro-Fuzzy Systems to Forecast Chaotic Time Series. We have shown:

- The Neuro-Fuzzy System is able to forecast a synthetic chaotic time series with high accuracy if the number of inputs K and the time delay τ are chosen adequately.
- The Takens-Mañé theorem gives an useful lower bound on the minimal number of inputs.
- The time delay τ between the inputs can not be set *a priori*. It has to be tuned for every different time series.
- The number of rules seems related to the size of the learning set and not to the strange attractor of the chaotic dynamical system.
- The meaningful and interpretation of the learned rules is not as easy as expected.
- Finally, it may be inadequate to compute output using the whole set of rules. Instead subsets of rules should be specialized to forecast only for localized region of the K-dimensional input space. We are currently studying this aspect.

ACKNOWLEDGMENTS

The work of Léonard Studer is supported by a grant from the Swiss National Science Foundation. Original source code of the Neuro-Fuzzy System used here was written by F. Casalino and R. Caviglia from the University of Genova. We thank L. Papa from the Istituto Idrografico della Marina (Genova, It) to have provided us the time-series of atmospheric sea-level pressure.

7 REFERENCES

[1] H. D. I. Abarbanel, "Predicting A from B: using it well", *These proceedings of SPIE* Vol. 3165, San Diego 1997.

[2] M. Ghil, "SSA-MMT toolkit: application to analysis and prediction of time series", *These proceedings of SPIE* Vol. 3165, San Diego 1997.

[3] R. Gilmore, "Topographical analysis of chaotic time series", *These proceedings of SPIE* Vol. 3165, San Diego 1997.

[4] F. Morabito, "Soft computing approach to plasma evolution tracking in Tokamak reactors", *These proceedings of SPIE* Vol. 3165, San Diego 1997.

[5] F. Masulli, F. Casalino, R. Caviglia, L. Papa, "Comparison of statistical methods and fuzzy systems in atmospheric pressure wave prediction (invited paper)" *Appl. and Science of Art. Neural Net. - SPIE Proc. Series*, (ed. S. K. Roger and D. W. Ruck, publ. SPIE - Bellingam, WA, USA) **2492**, pp 1050-1061 (1995)

[6] L. Studer & F. Masulli, "Building a Neuro-Fuzzy System to efficiently forecast chaotic time series " Nuclear Instr. & Methods A, in press (1997)

[7] T. Sauer, J. A. Yorke, M. Casdagli, "Embedology", *Journ. of Stat. Physics*, **65/3-4**, pp 579-616 (1991)

[8] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (publ. Springer Verlag, Berlin) (1996)

[9] L. X. Wang, *Adaptive Fuzzy Systems and Control*, (publ. Prentice Hall, 1994)

[10] J.-S. Roger Jang, C.-T. Sun, "Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems" *IEEE Transactions on Neural Networks*, **4**, pp 156-159 (1993)

[11] C. C. Jou, "On the Mapping Capabilities of Fuzzy Inference Systems", *Proc. of International Joint Conference on Neural Networks*, (ed. IEEE New-York, Baltimore, MD, USA) **2**, pp 708-713 (1992)

[12] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, D. L. Alkon, "Accelerating the Convergence of the Back-Propagation Method", *Biological Cybernetics*, **59**, pp 257-263 (1988)

[13] M. C. Mackey, L. Glass, "Oscillation and Chaos in Physiological Control Systems", *Science*, **197**, pp 287 (1977)

[14] J. D. Farmer, "Chaotic Attractors of an Infinite-Dimensional System", *Physica D*, **4**, pp 366-393 (1982).