Forecasting Using Non-Linear Techniques In Time Series Analysis: An Overview Of Techniques and Main Issues

Michel Camilleri

Department of Computer Science and AI,
University of Malta

Abstract. The development of techniques in non linear time series analysis has emerged from its time series background and developed over the last few decades into a range of techniques which aim to fill a gap in the ability to model and forecast certain types of data sets such a chaotic determinate systems. These systems are found in many diverse areas of natural and human spheres. This study outlines the background within which these techniques developed, the fundamental elements on which they are based and details some of the predictive techniques. This study aims to provide some insight into their mechanisms and their potential.

1 Introduction: The Need for Forecasting

No one knows the future, but there is a lot of benefit to be derived from attempting to form a picture of what the future could be - even if it is imprecise. It is one of the task most decision makers are expected asked to do - from the hunter-farmers of prehistoric age to the stock analysts of today.

[1] Almost all managerial decisions are based on forecasts. Every decision becomes operational at some point in the future, so it should be based on forecasts of future conditions.

A significant amount of research has been carried out using time series techniques in such diverse fields as [13] modelling and predicting solar sun spot activity, [5] analysis of Ethernet traffic, Business cycles, [3] Intelligent analysis of clinical time series in Diabetes Mellitus. Non linear techniques are emerging as powerful tool in the hands of the analyst to leverage patterns of behaviour underlying hiding dynamical systems - which may elude analysis using other traditional techniques.

[1] A model is an external and explicit representation of a part of reality as it is seen by individuals who wish to use this model to understand, change, manage and control that part of reality. Analysts try to model the system or process which is being studied in a variety of ways.

Regression Is the study of relationships among variables, a principle purpose of which is to predict, estimate the value of one variable from known or assumed values of other variables related to it. Using one predictor it is called *simple linear regression*. Using two or more predictors, *multiple regression* analysis is employed

A Scatter diagram is a graphical representation of the pairs of data to gain an overall view of the problem — is there an apparent relationship? Direct, inverse? If the points lie within a band described by parallel lines we can say there is a *linear relationship* between the pair of x and y values. If the rate of change is generally not constant then the relationship is *curvilinear*.

Linear Models: One would determine if a linear relationship exists between t and y and then attempt to produce a linear equation stating y as a function of x in the form of $y = a + bx + e$

where a is the intercept, b is the slope and e is the error term accounting for variables that affect y but are not included as predictors, and/or otherwise unpredictable and uncontrollable factors.

Least squares method: To predict the mean y-value for a given x-value, we need a line which passes through the mean value of both x and y and which minimizes the sum of the distance between each of the points and the predictive line — the best fit to the sample data. The least squares method achieves this by calculating the minimum averaged squared deviations between sample y points and the estimated line.

A procedure is used for finding values of a and b which reduces to the solution of simultaneous linear equations. Shortcut formula have been developed as an alternative.

Solution methods: Techniques of matrix algebra can be manually employed to solve simultaneous linear equations especially useful when there are more than two equations and two unknowns. Several computer packages using matrix algebra to solve simultaneous equations are widely available and can be utilised to relieve the computational problems - can solve both linear and polynomial equations

Variables of interest: To make predictions or estimates we must identify the effective predictors of the variable of interest which variables are important indicators? And can be measured at the least cost? Which carry only a little information? Which are redundant?

2 Time Series Analysis

[10] A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally (not always) spaced intervals with a consistency in the activity and the method of measurement [1].

Time Series Analysis comprises methods for the quantitative characterization, the manipulation, and the understanding of time series data and the underlying systems.

Time Series Forecasting Approaches There are two basic approaches to forecasting time series: the self-projecting time series and the cause-and-effect approach. It is possible that both approaches still lead to the creation of accurate and useful forecasts. The first is often the more difficult to implement and validate than the second.

Self-projecting time series: This approach uses only the time series data of the activity to be forecast to generate forecasts. The basic idea behind self-projecting time series forecasting model is to find a mathematical formula that will approximately generate the historical patterns in a time series.

Linear correlations in time are the most common and best studied sources of predictability. Linear statistical methods assume a random nature of the data since the underlying auto-regressive and moving average models are stochastic. Auto-regressive models and moving average models are stochastic due to the assumed random nature of the data [10]. The randomness prevents one from making any precise predictions, even with an absolute knowledge of the present state, but the strong correlation assures one that the next observation will be given by approximately by a linear combination of the preceding observations, except for additive noise ie:

$$s_{n+1} = a_0 + \sum_{j=1}^{m} a_j s_{n-m+j}$$

Autoregressive models is one of a group of linear prediction formulae that attempt to predict an output of a system based on the previous outputs and inputs. Autoregressive processes as their name implies, regress on themselves. The current value of the series is a linear combination of the

p most recent past values of itself plus an error term which incorporates everything new in the series at time t that is not explained by the past values.

An autoregressive model AR is a model which depends only on the previous outputs of the system is called. A moving average model MA. Is a model which depends only on the inputs to the system. AN autoregressive-moving-average model ARMA is a model based on both inputs and output.

Deterministic dynamical systems: Pure dynamical systems are described either by discrete time maps where:

$$x_{n+1} = f(x_n) \text{ (n is a discrete unit of time)}$$

Or by first order ordinary differential equations:

$$dx(t0)/dt = f(t, x(t)) \text{ (t is a real value of time)}$$

Both are descriptions of the fact that all future states of such nonlinear deterministic dynamical systems are unambiguously determined by specifying its present state at some time n (or t). Thus there also exists a deterministic forecasting function.

Nonlinear deterministic dynamical systems: Another type of correlation in time is present in non-linear deterministic dynamical systems. The underlying systems would have a deterministic nature but display irregularities. Chaos hence is an alternative source for unpredictability and irregularity in observed data, an alternative with respect to stochastic inputs.

These irregularities in the measurements arise from processes which co-exist with the system under study or are due to inaccuracies (noise) in the measurement process itself.

Any irregularities in the knowledge of the present state will evolve over time, in in the case of a chaotic system, will grow exponentially. However the uncertainty is amplified only at a finite rate and one can still hope to make reasonable short term forecasts.

Nonlinear prediction techniques exploit correlations over time visible only by using non-linear statistics.

3 Non-Linear Time Series Analysis

[11] The pioneering work by Packard et al. and the seminal paper by F. Takens stimulated, about 20 years ago, a new approach towards complex dynamical phenomena, nowadays called nonlinear time series analysis. Extensions towards nonlinear stochastic models and towards nonlinear statistical analysis are manifold and have been considered within various frameworks.

Stationarity: [10] A signal is called stationary if all transition probabilities from one state of the system to another are independent of time within the observation period. All relevant parameters have to be kept constant and that phenomena belonging to the dynamics are contained in the time series sufficiently frequently so that the probabilities or other rule scan be inferred properly.

Unpredictability: [10] of the future is the most striking feature of chaos despite a deterministic evolution. This unpredicatbiity is a consequence of the inherent instability of the solutions reflected by what is called sensitive dependence on initial conditions. The tiny deviations between the intial conditions of all the trajectories are blown up in a few time steps and every unobserved detail of the state at time zero is important for the precise path of the trajectory in state space.

This lead to two concepts (i) loss of information related to unpredictability; and (ii) nearby trajectories separate fast — exponentially fast over time.

If separation over time is very slow this is not in itself dramatic — typical of predominantly periodic systems. If this separation of trajectories over time is fast, exponentially fast — exponential divergence — we can speak of chaos. This increase of separation over time can be a characteristic of the underlying system.

Measurement technique and Sampling Rates: [10] A time series (as any other measurement) has to provide enough information to determine the quantity of interest unambiguously. Algorithms have some minimal requirements as to how long and how precise the time series must be and how frequently measurements have to be taken in order to obtain meaningful results. If meaningful results are not obtained then a smaller sub-phenomenon is to be studied. Sufficient sampling is required in general not just to particular methods of time series analysis

Example of sufficient sampling: Measurement of temperature changes on roof of house. How many single measurements are needed? Answer: it depends:

For one day: check one per hour for one week to get a fair average of the temperature profile
For seasonal fluctuations: fixed times, once a day for a year
For year to year: couple of measurements every year
For a complete description of phenomenon: once an hour, every day, for a couple of years

Modelling and Prediction depend on the measurement process, sampling technique and also the form of representation of the data.

4 Data Representation

One can create different perspectives of the raw data in order to be in a better position to examine and discover relationships over time. One may also apply techniques, such as — temporal abstraction, Phase space embedding, Poincare Map to render the data more suitable for prediction prior to prediction itself.

Scalar measurement/time chart: A dynamical system can be represented by a measurement which changes value over time e.g. in the shape a periodic function $\sin(t)$. The observations in an experiment are usually a time series - a scalar sequence of measurements over time. [10]

Describing dynamical systems in phase space: [10] A dynamical system can be described either by an m-dimensional map or by an explicit system of m first-order differential equations. Dynamical systems can be described by equations of motion which are defined in phase space.

It is important to establish a vector space (state space or phase space) for the system. Specifying a point in phase space specifies the state of the system and vice-versa. In a purely deterministic system once the present state is fixed, the states at all future points are determined as well.

By studying the dynamics of the corresponding phase space points one can thus study the dynamics of the system. There can be different choices for what phase space of a system can be.

Non linear prediction techniques, determination of Lyaponuv Exponents, noise filtering and other related applications use approximations of the dynamical equations to represent the underlying dynamical system. Such equations are defined in phase space so it is most natural to use a phase space description for these approximations too.

A trajectory in phase space of the dynamical system is a sequence of points x_n or $x(t_0)$ solving the above equations. An interesting exercise is to plot trajectories which start with similar phase

states (at least the first three phase states) on top of each other and observe the variation as time progresses. In the case of non-linear prediction, one is interested in those trajectories which stay in a bounded area forever. Other trajectories may run away to infinity as time proceeds. One is also interested in the rate at which they separate away from each other as time progresses

When irregularity is present in the signal we can talk of chaos in a deterministic system.

4.1 Phase Space Reconstruction

One plots the phase state of the function at different time points. For a single coordinate pair, the value of y will be the value of the function at time point n is plotted and the value of x will be the value of the function at the previous time point $n-1$ the x coordinate. All phase states i.e. the coordinate pairs $s(n)$, $s(n-1)$ are evaluated for all n.

The resulting chart will be a plot of the state of change of the system at different time points ie. the phase state map. If stationarity is present in the process, such a map of phase states will demonstrate this aspect more evidently than a normal time series plot of $F(t)$ vs t.

Phase Space Reconstruction Is at the core of nonlinear time series analysis. In such reconstructed spaces, the deterministic features of systems can be identified and studied, such as empirical invariant measures, attractor dimensions, entropies, Lyapunov exponents, equations of motion, and short-term forecasts can be made.

4.2 Elements of Phase Space Embedding

[10] The time series is initially a sequence of scalar measurements of some quantity which depends on the current state of the system taken at multiples of a fixed sampling time:

$$S_n = s(x(ndt)) + N_n$$

Some measurement function s and make observations only up to some random fluctuations N_n, the measurement noise. For a moment we will disregard the effect of noise.

A delay reconstruction in m dimensions is then formed by the vectors s_n, given as

$$S_n = (s_n - (m-1)v, s_n - (m-2)v, \ldots, s_{n-v}, s_n)$$

The lag or delay time is the time difference in number of samples v (or in time units vdt) between adjacent components of the delay vectors.

A number of embedding theorems are concerned with the question under which circumstances and to what extent the geometrical object formed by the vectors s_n, is equivalent to the original trajectory x_n. Under quite general circumstances if the dimension m of the delay coordinate space is sufficiently large, the attractor formed by s_n is equivalent to the attractor in the unknown space in which the original system is living.

The most important embedding parameter is the product (m x t) — embedding dimensions x delay time. It is the time span represented by the embedding vector. This can be derived after finding the two parameters separately.

Finding a good embedding in phase space: [10] In an ideal noise-free data there exists a dimension m such that the vectors s_n are equivalent to phase space vectors Choose too large m — will add redundancy and degrade performance of many algorithms — predictions, Lyaponuv Exponent. Choose too small an m and the vector will not represent the system and prediction errors will

be high. One must carry out analysis increasing the values of m until the typical behavour of deterministic data appears.

Finding value of smallest m with no False Neighbours: Consider an attractor with an embedding dimension m_0 i.e. it represents the system. If one were to generate a number of predicted values for the attractor itself using the coordinates of the attractor one on each other - the predicted values should all be close neighbours of the actual attractor points. If one were to reduce the dimension of the attractor by one — it could be that the attractor no longer represents the system and the predicted values will not be close or neighbours of the actual attractor coordinates. This number of false neighbours are generated by this $(m_0 - 1)$ dimension attractor.

Turning this concept into an algorithm one would start with small values of m and calculating the number of false neighbours generated. Gradually increase the number of dimensions by one till the number of false neighbours is zero. This means that this number of dimensions is the best embedding for the system - the one with the least dimensions required to correctly represent the system.

Finding a good estimate of the lag or time delay: A good estimate for the lag time is more difficult. Embeddings with different r but same m are the same mathematically and for noise-free data. However a good choice of t facilitates analysis.

If t is small compared to the internal time scales of the system, successive elements of the delay vector are strongly correlated. All vectors s_n are then clustered around the diagonal in Rm unless m is very large. If t is very large, successive elements are already almost independent and the points fill a large cloud in the Rm, where the deterministic structures are confine to the very small scales.

5 Examination of Data

Prior to applying any modelling or predictive techniques, the analyst is to examine the data using a variety of techniques amongst which is visual inspection of various plots which can be obtained prior to processing.

Visual examination is still a powerful tool which must not be dismissed. Human visual interpretation of data and results still lies at the top of the methods to understand systems.

A number of mathematical/computational techniques also exist which assist the analyst in obtaining a deeper understanding of the dynamics of the system being studied. Here are two of the principal ones.

Lyapunov exponent [10] can be a characteristic of the underlying system. It will indicate the change in distance of trajectories with initially neighbouring points. can be described by the properly averaged exponent of the increase of separation of trajectories over time. A number of these exponents can be calculated for various dimensions m.

Autocorrelation sum is a measure of how many neighbours each point has in phase space. One may superficially deduce that the more neighbours the better the predictability. [10] However autocorrelation of signals from deterministic chaotic systems decay exponentially with increasing lag. One has to be very cautious about interpreting these values.

6 A Simple Non-Linear Prediction Technique using Delay Vectors

[10] Here a simple prediction technique exploits the deterministic structures in the signal. The data is assumed to originate from measurement of an underlying dynamical system.

One needs to predict a future state at a desired time point $n+1$ using the last value at time point n. One scans the list of all states in phase space and tries to find the one closest (according to some norm) to the phase state value at time point n.

If a time point n_0 Is found where the phase state is similar to that at n (this means x_{n0} it is close to x_n). Then the continuity of the underlying dynamical system and the representation guarantees that x_{n0+1} will also be close to x_{n+1}.

However usually it is not the actual states that are measured but one or more quantities which depend on these states. More often than not the measurement function is as unknown as the underlying equation of motion.

Thus we have scalar measurements $S_n = s(x_n)$, where x_n is the actual state at time point n and n can take the values from 1 to N. One can then make use of a set of such measurements taken at a particular time point n and $(m-1)$ time points prior to n to represent the original sequence of phase states. This is called a delay reconstruction and the sequence of phase state values representing is called the delay vector:

$$S_n = (s_{n-(m-1)}v, s_{n-(m-2)}v, \ldots, s_n - v, s_n)$$

This introduces two adjustable parameters into the prediction method (which in principle is parameter free). The number of dimensions m and the time delay which is the time delay v between successive measurements. For all measurements $s_1, \ldots, s_n$ available so far, one can construct the corresponding delay vectors $s_{(m-1)}v + 1, \ldots, s_n$.

In order to predict a future measurement $s_n + d_n$, one can find the embedding vector closest to s_n and use s_{no+dn} as a predictor. This scheme has been used in tests for determinism by Kennel and Isabelle (1992). The idea may go back to Pikovsky 1986. The method used by Sugihara and May (1990) is closely related.

This means that one scans all measurements and for each measurement the whole delay vector of s_n is compared with the delay vectors of each measurement in the phase state data i.e. for points 1 to $n-1$. To compare delay vectors of two points one could check the distance between corresponding elements of the two delay vectors. If each of the distances between the pairs are all within a desired distance then the two delay vectors are close.

Finding just the closest delay vector close to the s_n is not enough. Any measurement of a continuous quantity makes sense within a finite boundary or resolution. The size of this limit or resolution depends on the differences arising from the measurement equipment and the fact that the values are eventually discretised. This implies that it is probable there would be more than one delay vector which would come reasonably close the delay vector of s_n. Each of these have to be considered as valid predictors of the future point. They have to be considered as neighbours of s_n Every neighbour of s_n will be used to predict a future value. The average of these predicted future values will be the finally accepted prediction

Thus the neighbourhood distance eps is another is another important parameter which has an effect on the success of prediction. This determines how many neighbours are found, which in turn effects the average predicted value.

This method is very inefficient if more than a few predictions are needed, e.g. because prediction error needs to be determined accurately to find the optimal set of parameters of time delay and dimension. For each prediction all points in the reconstruction are considered. All points are tested and most are rejected because they are too far away.

7 Prediction Technique with Box Assisted Neighbour Search

[10] Here will describe a method that will reduce this problem significantly by proposing a modification to the process of the repeated search for neighbours.

Creation of Phase State Map Array : An imaginary map with two axes y representing s_n and x representing s_{n-1} is drawn and subdivided into a number of boxes of equal sizes forming a grid. Each box represents a defined area of the phase state and each box is delimited by a unique set of lower and upper limits for both y and x coordinates. The size of the box will be defined by these limits may be varied. Every measurement taken is converted to the phase state form. Each phase state pair is then placed in the box according to the coordinate values.

Eventually all points are mapped into this grid. This grid can be represented in a number of ways. One of which is a simple array representing a list of pointers to actual data points. With a simple algorithm it is possible to sequence the array to reflect the boxes in the grid and create another pointer list to point at the start pointer and end pointer of each box. Another alternative, possible faster to run, is using a tree structure.

Reduced computation in finding neighbours: According to the prediction technique proposed earlier one has to scan the whole set of points for neighbours of the selected point. Using this phase state map of the points one can home in to the box in the grid related to the phase state of the selected point and determine which of the points in that box are actually neighbours. This would involve only a small subset of the total data set and would result in a huge reduction in computation cost.

Searching in neighbouring boxes: Some points may be neighbours of a selected point but will fall in a neighbouring box because they fall just outside the limit values. So one has to include the neighbouring boxes (8 in all) in the search for neighbours. Although this adds to the computation cost, overall there is still a significant gain over the previous method.

Defining an appropriate box size and number of boxes: A further condition to this approach is that the number of boxes has to be defined. If the boxes are too sparsely populated that there is little gain. If there are too few boxes but densely populated then there is little gain too. One may attempt to define the box size after examining the distribution of the phase state map. Once the box size and number of boxes are defined one will find that a number of points may refer to boxes outside the boundaries of the whole grid. In this case one has to mirror the coordinates of the 'outlying' point back into the grid using a suitable mod function. This is why one still has to check for closeness of each point in the selected boxes because there may be points which actually belong to a box outlying the grid.

Creating a list of neighbours: Once the array or tree representing the data set in phase state form is created it is a relatively smaller effort to create a list of neighbours for a selected point. In the case of a delay vector — each member of that delay vector is compared to the corresponding members of the potential neighbour point before deciding if the point being compared is a neighbour to the selected point.

Prediction using the list of neighbours: Once the list of neighbours is found the average predicted value for the future point or points can be determined. 8 Measuring success of prediction

[10] An important part of the prediction process is the measurement of the difference between the predicted value and the actual value. One can separate the data set into two parts — the training set and a hidden set to be used to test the success or not of the predictive technique being used. Usually a technique is used to predict values for a number of time points in the future. So it is necessary to first compute the error for each prediction and then carry out some form of summation to get a single number.

[21] These are some methods to calculate the degree of success of prediction, each of which contributes in a different manner to the evaluation of the prediction: mean error, mean absolute error, sum of squared error, root mean square error, percentage error, mean percentage error (MPE), mean absolute percentage error (MAPE).

Measurement of prediction error is important to improve on the prediction process by helping to identify the sets of input parameters which gave the best prediction results.

8 Number of Variables Used in Analysis

[10] Univariate time series (measurement of one observable) has been considered so far. Phase space has been reconstructed using delay vectors based on combinations of these measurements. However it is possible to form an embedding using a set of different variables obtained at the same time. One could think of other ways of representing the underlying system.

Multichannel measurements is analysis using multiple sources of simultaneous data — several measurements of the same type — at the same time — but at different segments of the process — which may give a more complete representation of the underlying system

Equivalent variables at different positions — In other situations one may find it useful to measure similar physical quantities simultaneously at different positions in a spatially extended system.

Variables with different physical meanings - One may also want to measure several observables with different physical meanings simultaneously. E.g. a 9-dimensional space with three different observables each at three different times.

Distributed systems multivariate time series are very frequent in physiology, economics or climatology, but are generally too complicated for a systematic investigation of the interrelation between the observables. There are research problems which are still unsolved. Take the example of a bivariate time series. There is a problem in deciding which observable to use in order to define the delay embedding based on one coordinate. Which observable yields more additional information?

9 Methodology

Forecasting methods: A number of methods exist which help the analyst to develop a forecasting model with a degree of objectivity, in a way which is measurable and reproducible.

These methods usually involve:

- the preparation of the data for analysis — smoothing, identification of cyclic (seasonal) components, trends, outliers, noise reduction;
- the application of a selected technique or set of techniques;
- measurement of forecasting error;
- refinement of parameters of model;
- updating of time series data as the current data becomes out of date;
- repeated evaluation of accuracy of model and refinement of model.

Standards and Practices for forecasting: [17] A set of standards is proposed by Scott Armstrong. From a detailed study in the use of standards a model with 139 points falling under these 16 categories was formulated, covering formulating problems, obtaining information, implementing methods, evaluating methods and using forecasts.

References

1. Hossein Arsham, *Time-Critical Decision Making for Economics and Finance*. Available from `http://home.ubalt.edu/ntsbarsh/stat-dta/Forecast.htm`.
2. R. Bellazzi, C. Larizza, P. Magni, S. Montani and G. De Nicolao, *Intelligent Analysis of Clinical Time Series by combining Structural Filtering and Temporal Abstractions*, Artificial Intelligence in Medicine '99. LNAI 1620, 1999.
3. R. Bellazzi, C. Larizza, P. Magni, S. Montani and M. Stefanelli, *Intelligent Analysis of Clinical Time Series: an Application in the Diabetes Mellitus Domain*, AI in Medicine, vol 20, no. 1, 2000.
4. Michael Breakspear, *Perception of Odors by a Nonlinear Model of the Olfactory Bulb*, International Journal of Neural Systems, Vol 11, No.2:101-124, 2001.
5. Kavitha Chandra, Chun You, Gbenga Olowoyeye and Charles Thompson, Non-Linear Time-Series Models of Ethernet Traffic, CACT, Tech. Rep., June 1998.
6. Ana Beatriz Galvao, *Univariate Non-Linear Time Series Models and the Business Cycle Stylised Facts*, 2000.
7. Ben Goertzel, *Delay Vectors as Perceptual Chunks: Understanding Nonlinear Time Series Analysis as Self Organising Cognition* Dynamical Psychology An International, Interdisciplinary Journal of Complex Mental Processes, 1999.
8. R. Hegger, H. Kantz, and T. Schreiber *Practical Implementation of nonlinear time series methods: The TISEAN package*, CHAOS 9, 413, Winter School on Nonlinear Time Series Analysis, 1998.
9. B.P.T. Hoekstra, C.G.H. Diks, M.A. Allessie and J. DeGoede, *Non Linear Time Series Analysis: Methods and Applications to Atrial Fibrillation, Analysis and Processing of Cardiac Electrograms* in Atrial Fibrillation, Annali dell'Istituto Superiore di Sanita', Roma, Ref Ann. 1st Super. Sanita' vol 37, no.3, 2001.
10. H. Kantz and T. Scheiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.
11. Holger Kantz, *Time Series Analysis* in Reconstructed State Spaces, Stochastics and Dynamics, Vol 1, No. 1, 2001.
12. Kulkarni D. R., A. S. Pandya, and J. C. Parikh, *Dynamic Predictions from Time Series Data — An Artificial Neural Network Approach*, International Journal of Modern Physics C, in press, 1997.
13. Kulkarni D. R., A. S. Pandya, and J. C. Parikh, *Modeling and Predicting Sunspot Activity — State Space Reconstruction + Artificial Neural Network Methods*, GeoPhys Lett. 25, 457, 1998
14. P. Magni, R. Bellazi, and F. Locatelli, *Using uncertainty management techniques in medical therapy planning: a decision theoretic approach.* In Applications of Uncertainty Formalisms, Springer-Verlag, Berlin, 1998.
15. Pakalapati Rajani, *Network Traffic Analysis: Identifying Invariant Features in Ethernet Traffic*, M.Sc. Thesis, Dept of Elec. Eng. Univ of Massachusetts Lowell, 2000.
16. C. P. Price, D. Prichard, E. A. Hogenson, *Do the Sunspot Numbers Form a "Chaotic" Set?* Available from `citeseer.ist.psu.edu/price92do.html`.
17. J. Scott Armstrong, *Standards and Practices for Forecasting taken from Principles of Forecasting: A handbook for Researchers and Practitioners,* Kluwer Academic Publishers, 2001.
18. Sello Stefano, *Time Series Forecasting: A Multivariate Stochastic Approach*, Topic Note Nr. NSG/260199, Los Alamos National Laboratories Preprint Archive, Physics/9901050, 1999.
19. Sello Stefano, *Time Series Forecasting: A nonlinear Dynamics Approach*, Los Alamos National Laboratories, preprint archive Physics, 9906035, or Thermo Fluid Dynamics Research Centre Enel Research USG/180699.
20. Andreas Seyfang, Silvia Miksch, Werner Horn, Michael S. Urschitz, Christian Popow, Christian F. *Poets Using Time-Oriented Data Abstraction Methods to Optimize Oxygen Supply for Neonates*, Lecture Notes in Computer Science, 2001.
21. Statsoft Inc, *Time Series Analysis — a review of techniques Electronic Textbook*, Statsoft Inc. Available from `http://www.statsoftinc.com/textbook/stathome.html`.
22. Statsoft Inc, *Statistica, Product Information*. Available from `http://www.statsoftinc.com/downloads`.
23. David M. Walker, Nicholas B. Tufillaro and Paul Gross, *Raidal-Basis Models for Feedback Systems with Fading Memory*, IEE Transactions on Circuits and Systems, Vol.48, No. 9, 2001.
24. Chun You and Kavitha Chandra, *Time Series Models for Internet Data Traffic*, Proc. 24th Conference on Local Computer Networks, IEEE Press, p164-171, 1999.
25. Xiru Zhang and Kurt Thearling, *Non-Linear Time-Series Prediction by Systematic Data Exploration on a Massively Parallel Computer*, Available from `http://www./thearling,com/Text/sfitr.html`, Santa Fe Technical Report 94-07-045.