

# A New Evolutionary Method for Time Series Forecasting

Tiago A.E. Ferreira  
Center for Informatics  
Federal Univ. of Pernambuco  
Av.Prof. Luiz Freire, s/n, CDU  
50732-970, Recife - PE - Brazil  
taef@cin.ufpe.br

Germano C. Vasconcelos  
Center for Informatics  
Federal Univ. of Pernambuco  
Av.Prof. Luiz Freire, s/n, CDU  
50732-970, Recife - PE - Brazil  
gcv@cin.ufpe.br

Paulo J.L. Adeodato  
Center for Informatics  
Federal Univ. of Pernambuco  
Av.Prof. Luiz Freire, s/n, CDU  
50732-970, Recife - PE - Brazil  
pjla@cin.ufpe.br

## ABSTRACT

This paper presents a new method — the Time-delay Added Evolutionary Forecasting (TAEF) method — for time series prediction which performs an evolutionary search of the minimum necessary number of dimensions embedded in the problem for determining the characteristic phase space of the time series. The method proposed is inspired in F. Takens theorem and consists of an intelligent hybrid model composed of an artificial neural network (ANN) combined with a modified genetic algorithm (GA). Initially, the TAEF method finds the most fitted predictor model for representing the series and then performs a behavioral statistical test in order to adjust time phase distortions.

**Categories and Subject Descriptors:** I.6.5 [Model Development]: Modeling methodologies; I.2.8 [Problem Solving, Control Methods, and Search]: Heuristic methods

**General Terms:** Experimentation.

**Keywords:** Genetic Algorithms, Neural Network, Time Series, Forecasting.

## 1. INTRODUCTION

In this work, a systematic procedure based on an hybrid intelligent system approach is proposed for the automatic search of the important parameters that solve time series prediction problems. The method consists of a combination of a standard ANN with a modified genetic algorithm (GA) [2] which efficiently searches and defines 1. the minimum number of (and the specific - F. Takens [6]) temporal lags, 2. the best neural network structure in terms of the number of processing units, 3. the most fitted training algorithm that boosts the prediction performance, and 4. a behavioral statistical test carried out at the prediction model output to fix relative phase distortions in the series representation.

## 2. THE TAEF METHOD

The method proposed in this work — Time-delay Added Evolutionary Forecasting (TAEF) method — tries to reconstruct the phase space of a given time series by carrying out a search for the minimum dimensionality necessary to reproduce, to a certain accuracy, the phenomenon generator of the times series and its subsequent values. The proposed procedure is a intelligent hybrid system based on an artificial neural network (ANN) architecture (multi-layer perceptron) trained with a modified genetic algorithm (GA) [2] which not only searches for a number of the ANN parameters but also for the adequate embedded dimension represented in the

lags. The accuracy of the prediction generated is initially set by the user but is automatically changed by the training algorithm.

The scheme describing the proposed algorithm, Figure 1, is based on the iterative definition of the three main elements: 1. the underlying information necessary to predict the series; 2. the structure of the model capable of representing such underlying information for the purpose of prediction (the number of units in the ANN structure); and 3. the appropriate algorithm for training the model (the most appropriate algorithm among several candidates).

Following this principle, the important parameters defined by the algorithm are: 1. **The number of time lags to represent the series:** initially, a maximum number of lags (*MaxLags*) is defined by the user and a GA can choose any number of lags in the interval [1, *MaxLags*] for each individual of the population; 2. **The number of units in the ANN hidden layer:** the maximum number of hidden layer units (*NHiddenmax*) is determined by the user and the GA chooses, for each candidate individual, the number of units in the hidden layer (in the interval [1, *NHiddenmax*]); 3. **The training algorithm for the ANN:** RPROP [5], Levenberg-Marquardt [4], Scaled Conjugate Gradient [3], One Step Secant Conjugate Gradient [1] are candidates for the best algorithm for training the ANN and the GA defines these algorithms as individuals in the population.

In order to conclude the definition of the method a last aspect had to be considered. While the representations of some series were developed by the model with a very close approximation between the actual series and the predicted series (“in-phase” matching), the predictions of some other series were always presented with a one step shift (delayed) with respect to the original data (“out-of-phase” matching).

In order to make the TAEF method robust for representation of any time series, after the best model is chosen when training is finished, an statistical test is employed to check if the network representation has reached an in-phase or out-of-phase matching. If this test (for example the t-test) accepts the in-phase matching hypothesis, the elected model is ready for practical use. Otherwise, the method carries out a new procedure to adjust the relative phase between the prediction and the actual time series. The validation patterns are presented to the ANN and the output of these patterns are re-arranged to create new inputs that are both presented to the ANN and set as the output (prediction) target. The approximation results for both the in-phase and out-of-phase models are measured and the best model (smaller MSE error) is elected as the final model.

## 3. EXPERIMENTAL RESULTS

The series investigated <sup>1</sup> were normalized to lie within the interval [0,1] and divided in three sets, training set (50% of the points),

<sup>1</sup>A total of 6 series were investigated but only 2 are presented here

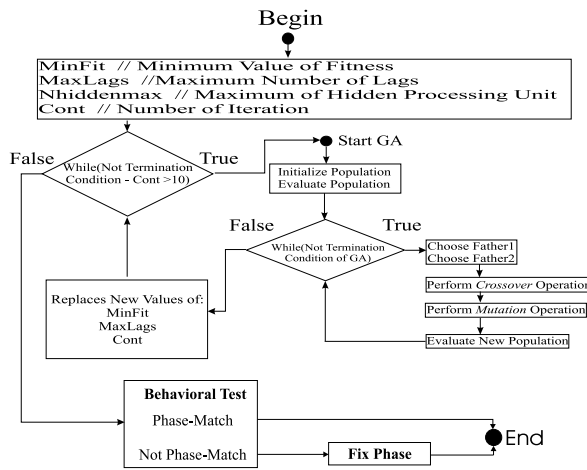


Figure 1: Algorithm for the TAEF method

validation set (25% of the points) and test set (25% of the points). The AG parameters are the same for all the series, with a mutation probability of 10%.

### 3.1 Sunspot Series

The sunspot series used consisted of the total annual measures of the sun spots from the years of 1700 to 1988, generating a database of 289 examples.

For the prediction of the Sunspot series (with 1 step ahead of prediction horizon), the TAEF method identified the lags 1 to 4 as the relevant to the problem, defined 4 processing units in the hidden layer of the network, elected the Levenberg-Marquardt algorithm as the most fitted for the ANN training and classified the model as “in-phase” matching. Figure 2(a) and Table 1 shows the results with all the performance measures for both cases: “in-phase” matching and if the prediction model had been chosen as “out-of-phase” matching.

### 3.2 Down Jones Series

The Dow Jones Industrial Average Index (DJIA) series corresponds to daily observations from 1st January 1998 to 26th of August 2003 of the DJIA index, constituting a database of 1420 points.

The hybrid model proposed automatically chose the lags 2, 4, 8, 6, 9 and 10 as the relevant lags for the series representation, defined 10 processing units for the hidden layer of the ANN, once again selected the algorithm Levenberg-Marquardt as the most fitted for the ANN training and classified the model as “out-of-phase” matching. Figure 2(b) and Table 1 shows the results with all the performance measures for both cases: out-of-phase matching and if the prediction model had been chosen as “in-phase” matching.

Table 1: Experimental Results.

	Sunspot		DJIA	
	In Phase	Out of Phase	In Phase	Out of Phase
MSE	$8.600 \cdot 10^{-3}$	0.3070	$8.4183 \cdot 10^{-4}$	$2.6841 \cdot 10^{-5}$
MAPE	34.03 %	82.55 %	1.15 %	0.20%
U of Theil	0.3218	1.2225	1.0006	0.0318
POCID	84.29%	65.22%	47.58%	97.14%
AIC	-289.6	-195.9	-2206.1	-3408.5
BIC	-211.3	-117.9	-1510.6	-2713.4

for lack of space

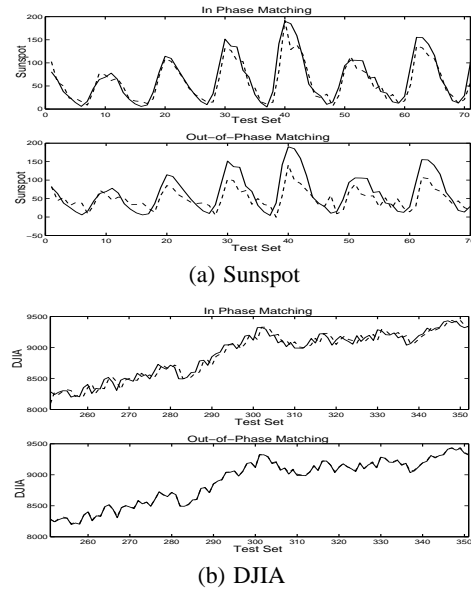


Figure 2: Prediction results (test set): actual values (solid lines) and predicted values (dashed lines).

## 4. CONCLUSIONS

This paper has presented an intelligent hybrid system approach, the Time-delay Added Evolutionary Forecasting (TAEF) method, which consists of an ANN combined with a modified GA and a behavior test of phase match hypotheses for the solution of time series forecasting problems.

The experimental results were measured according to a set of 6 consistent performance metrics (MSE, MAPE, U of Theil Statistics, POCID, AIC and BIC).

With the introduction of the behavior test for identifying whether the prediction model is “in-phase” or “out-of-phase” with the series to be forecasted, the TAEF method was able to classify if a given time series tends or not to a Random Walk like model. However, a systematic study is yet necessary to determine any possible limitations of the method when dealing with other types of components found in other different real world time series such as trends, seasonality, impulses, steps, and other non-linearities. Taking that into account, other time series with those components are being collected to carry out a broader investigation.

## 5. REFERENCES

- [1] R. Battiti. One step secant conjugate gradient. *Neural Computation*, 4:141–166, 1992.
- [2] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K. S. Tam. Tuning of the structure and parametrs of the neural network using an improved genetic algorithm. *IEEE Transaction on Neural Networks*, 14(1):79–88, January 2003.
- [3] M. F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533, 1993.
- [4] J. J. Mor. The levenberg-marquardt algorithm: Implementation and theory. In *Proceedings of Springer-Verlag in Numerical Analysis (Lecture Notes in Mathematics)*, pages 105–116, 1977.
- [5] M. Reidmiller and H. Braun. A direct adaptive method for faster backpropagation learnng: The rprop algorithm. In *Proceedings of the IEEE Int. Conf. on Neural Networks (ICNN)*, pages 586–591, San Francisco, 1993.
- [6] F. Takens. Detecting strange attractor in turbulence. In A. Dold and B. Eckmann, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lect. Notes in Maths.*, pages 366–381, New York, 1980. Springer-Verlag.