ICA AND NONLINEAR TIME SERIES PREDICTION FOR RECOVERING MISSING DATA SEGMENTS IN MULTIVARIATE SIGNALS

Torbjørn Eltoft

Institute of Physics
University of Tromsø
9037 Tromsø, Norway

Ørjan Kristiansen

Institute of Physics
University of Tromsø
9037 Tromsø, Norway

ABSTRACT

In this paper we introduce a new method for filling in gaps in a time series belonging to a set of simultaneously recorded, statistically dependent signals. By combining the properties of the independent component analysis (ICA) transform with those of the dynamical-functional artificial neural network (D-FANN), we have developed a predictor that effectively exploits the mutual dependency between the component signals. This is done by performing the predictions in the ICA-domain, whereas the prediction errors, which are used to update the model parameters, are calculated in the observation domain.
We have shown that the ICA D-FANN predictor is capable of accurately filling in gaps in both synthetic and real time series. Our tests show that the new approach outperforms a predictor based on a standard multilayer perceptron (MLP) network, and a predictor based on the finite impulse response (FIR) network.

1. INTRODUCTION

In time series prediction we are given a set of measurements of some observables, and the goal is to be able to learn the evolution in time. A popular approach to this problem is to reconstruct the dynamics of the underlying attractor of the phase space in an embedding space. This method is mainly based on Takens theorem [1], which assures that future dynamics of a system can be captured by embedding vectors composed of a number of past values. Hence, given a dynamic system $\boldsymbol{y}_{t+1} = \boldsymbol{g}(\boldsymbol{y}_t)$, we define the measurement $x_t = f(\boldsymbol{y}_t)$. Takens theorem then states that for an embedding dimension q of the reconstruction vectors $\boldsymbol{v}_t = (x_t, x_{t-1}, \cdots, x_{t-q})$, there exists a mapping $\boldsymbol{v}_{t+1} = \boldsymbol{F}(\boldsymbol{v}_t)$ which has the same dynamical characteristics as the original system $\boldsymbol{y}(t)$. The dimension q should be at least twice the dimension of the strange attractor.
The problem in time series prediction then is to learn the mapping $\boldsymbol{F}$ from the observations. In recent years this has been proposed done using various neural network models [2, 3, 4].
In this paper we consider a specific problem that may occur in some applications of multivariate time series analysis. Given a set of correlated signals, e.g. time series of geophysical observations or economical indexes, we find that in one series (or a few) some parts are missing. We assume that the missing sections extend over some tens of samples. The problem of recovering the missing parts may be defined as a prediction and interpolation task, where we want to use past samples to predict future values, but we also want to take advantage of having observations of the other signals in the set during the time of the missing section. We address this problem by combining the properties of independent component analysis (ICA) [5] with those of the dynamical-functional artificial neural network (D-FANN) [4, 6]. Our approach assumes that the observed signals are generated from independent sources through a linear mixing. Let $\boldsymbol{x}(t)$ and $\boldsymbol{s}(t)$ be the observed signals and the original source signals, respectively, and let $\mathcal{A}$ denote the mixing matrix. Then the model assumes that $\boldsymbol{x}$ is related to $\boldsymbol{s}$ as:

$$\boldsymbol{x}(t) = \mathcal{A}\boldsymbol{s}(t). \tag{1}$$

Let $\boldsymbol{u}(t)$ be an estimate of the source signals $\boldsymbol{s}(k)$, obtained by first estimating a demixing matrix $\mathcal{W}$ using some ICA technique [7, 8], and then calculating $\boldsymbol{u}(t)$ as:

$$\boldsymbol{u}(t) = \mathcal{W}\boldsymbol{x}(t) = \mathcal{W}\mathcal{A}\boldsymbol{x}(t). \tag{2}$$

The idea is that the signals $\boldsymbol{u}(k)$ generally have a much simpler structure than the observed signals $\boldsymbol{x}(k)$ that makes them easier to predict [9].

We use the DCT-based D-FANN predictor developed by Eltoft and deFigueiredo [4, 6] to perform nonlinear one-step predictions on each of the individual component signals. By doing the predictions in the ICA domain, i.e. on $\boldsymbol{u}(t)$, while calculating prediction errors for updating the model in the observation domain, i.e. using $\boldsymbol{x}(t)$, we are effectively utilizing the dependencies between the observed signals in the prediction process.
The method is tested in experiments using both synthetic and real time series data, and the preliminary results are very promising. We compare the performance of our method with the performances of a predictor using a standard multilayer perceptron (MLP) network [10] and a predictor based on the finite impulse response (FIR) network, which uses FIR filters in the synapses [2].

2. MULTIVARIATE SIGNAL MODEL AND THE ICA TRANSFORM

We assume in our model that the observed signal $\boldsymbol{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ at time t is given by,

$$\boldsymbol{x}(t) = \mathcal{A}\boldsymbol{s}(t). \tag{3}$$

This model is generative, meaning that it describes how the observed data are generated by a process of mixing a set of independent source signals $\boldsymbol{s}(t) = [s_1(t), \ldots, s_M(t)]^T$, where the mixing is assumed to be instantaneous. $\boldsymbol{s}(t)$ is applied to a linear system whose input-output characterization is defined by a nonsingular

$N \times M$ *mixing matrix* $\mathcal{A}$, with elements a_{ij}. With both $\boldsymbol{s}(t)$ and $\mathcal{A}$ being unknown, we formulate the objective of the ICA transform: Based on the mixture vector $\boldsymbol{x}(t)$, the goal of ICA is finding a linear transformation $\mathcal{W}$ of the dependent signals $\boldsymbol{x}(t)$, that makes the output as independent as possible,

$$\boldsymbol{u}(t) = \mathcal{W}\boldsymbol{x}(t) = \mathcal{W}\mathcal{A}\boldsymbol{s}(t). \tag{4}$$

The matrix $\mathcal{W}$ is often referred to as the *separation matrix*, with elements w_{ij}, while $\boldsymbol{u}(t) = [u_1(t), \ldots, u_M(t)]^T$ is an estimate of the sources. We note that the sources are exactly recovered when $\mathcal{W}$ is the inverse of $\mathcal{A}$ except for a possible permutation and scaling difference, i.e.

$$\mathcal{Q} = \mathcal{R}\mathcal{S} = \mathcal{W}\mathcal{A}, \tag{5}$$

where in (5) $\mathcal{R}$ is a permutation matrix and $\mathcal{S}$ is a scaling matrix. These two matrices define the performance matrix $\mathcal{Q}$. If $\mathcal{Q}$ is normalized and reordered, a perfect separation leads to the identity matrix.

For the linear mixing and unmixing model, the following four assumptions are adopted [5, 11]:

1. The sources $\boldsymbol{s}(t)$ are, at each time instant, mutually independent. This is expressed by the condition:

$$p(\boldsymbol{s}(t)) = p(s_1(t), \ldots, s_M(t)) = \prod_{i=1}^{M} p_i(s_i(t)). \tag{6}$$

2. The number of mixtures is greater than, or equal to the number of sources, $N \geq M$. This is a necessity to make $\mathcal{A}$ a full rank matrix. But even in the case when $M > N$, the mixing matrix $\mathcal{A}$ may be identified; but the realizations of the independent components can not, because the $\mathcal{A}$-matrix is not invertible. However, most of the existing theory for ICA is not valid in this latter case.
3. At most one of the sources is allowed to be normally distributed. This is because the unmixing of two Gaussian sources is ill posed when the sources are white random processes. Nevertheless, non-white Gaussian processes may be recovered with time -decorrelation methods if they have different spectra.
4. No sensor noise, or only low additive noise signals are permitted. This is necessary to satisfy the infomax condition, in which the mutual information between outputs is minimized. However, noise can be considered as an independent source itself, and segregated from the mixtures.

In the approach to solve the ICA problem problem, the temporal structure of the time series is omitted, and $\boldsymbol{x}(t)$ and $\boldsymbol{s}(t)$ of (3) are regarded as realizations of random vectors $\boldsymbol{x}$ and $\boldsymbol{s}$, i.e.

$$\boldsymbol{u} = \mathcal{W}\boldsymbol{x} \tag{7}$$

In [12] a review of various information theoretic contrast functions for solving $\mathcal{W}$ is given, including *mutual information*, *negentropy*, *maximum entropy*, and *infomax*. We have chosen to use the FastICA algorithm to solve for $\mathcal{W}$. This algorithm is based on maximizing the negentropy, and obtains convergence in a small number of iterations as a result of the its cubic convergence rate [7].

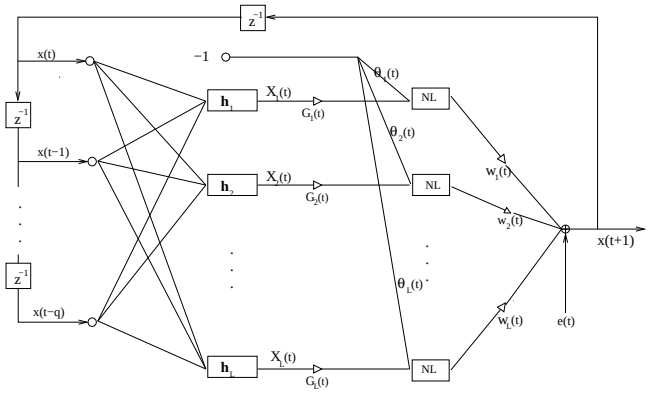


Fig. 1. D-FANN model for time series prediction.

3. TIME SERIES PREDICTION USING A DCT BASED D-FANN

Time series prediction is defined as: Given a finite sequence of a discrete time series $x(t)$, i.e. $x(t), x(t-1), \cdots, x(t-q)$, find the continuation $x(t+1), x(t+2), \cdots$. This involves finding a scalar q and a function F such that $x(t+1)$ can be estimated by

$$\hat{x}(t+1) = -F(x(t), x(t-1), \cdots, x(t-q)) \,. \tag{8}$$

This is equivalent to modeling the time series as:

$$\begin{aligned} x(t+1) + F(x(t), x(t-1), \cdots, x(t-q)) \\ = n(t+1) \,, \end{aligned} \tag{9}$$

with $n(t+1)$ being a white noise process (see Fig.1). If the statistics of the time series $x(t)$ are non-Gaussian or the time series is the result of some nonlinear operation, the function F is nonlinear. The model in (8) then defines *a generic nonlinear AR model*. Let the vector $\boldsymbol{x}^t = \{x(t), x(t-1) \cdots, x(t-q)\} \in E^{q+1}$ (where E^{q+1} is a $q+1$-dimensional real Euclidean Space) represent our time segment of length $q+1$. In our formulation of this modeling problem, we consider the function $F: E^{q+1} \rightarrow R$ as a bounded analytic functional belonging to a generalized Fock Space, F_ρ, where the subindex ρ refers to a specific sequence of positive numbers $\rho = \{\rho_0, \rho_1, \rho_2, \cdots\}$ defining the functional space [13]. The solution for F is achieved in two stages. First, a best structural approximation $\hat{F}$ of F is found using some proper representation $\boldsymbol{h}^i, i = 1, 2, \cdots, L$ of the data vector $\boldsymbol{x}^t$. From Hilbert Space theory it follows that the best approximation $\hat{F}$ is the projection of F in the closed subspace of F_ρ spanned by the reproducing kernels $K(\boldsymbol{h}_1, \cdot), K(\boldsymbol{h}_2, \cdot), \cdots, K(\boldsymbol{h}_L, \cdot)$ [13]. The corresponding approximate nonlinear model for the process $x(t)$ is then:

$$x_L(t+1) + \sum_{i=1}^{L} w_i(t) K(\boldsymbol{h}_i, \boldsymbol{x}^t) = n(t+1) \,. \tag{10}$$

In the second stage we use a nonlinear predictive coding approach to calculate the values of $w_i(t), i = 1, 2, \cdots, L$, which minimize the mean square error

$$\begin{aligned} \mathcal{E}[e^2(t+1)] &= \mathcal{E}[(x(t+1) - \hat{x_L}(t+1))^2] \\ &= \mathcal{E}[(x(t+1) + \sum_{i=1}^{L} w_i(t) K(\boldsymbol{h}_i, \boldsymbol{x}^t))^2] \,, \end{aligned} \tag{11}$$

where in (11) $\mathcal{E}[\cdot]$ refers to ensemble averaging.

$K(\boldsymbol{h}_i, \boldsymbol{x}^t)$ can be represented by a scalar-valued analytic function φ of the argument $< \boldsymbol{h}_i, \boldsymbol{x}^k >$, thus $K(\boldsymbol{h}_i, \boldsymbol{x}^t) = \varphi(< \boldsymbol{h}_i, \boldsymbol{x}^t >)$, where $< \cdot, \cdot >$ denotes the inner product in E^{q+1}. Also, the function φ can be chosen to be the hyperbolic tangent function $tanh(\cdot)$, which appears as the sigmoid function in many frequently used neural networks. The vectors $\boldsymbol{h}_i \stackrel{\text{def}}{=} \{h_{i1}, h_{i2}, \cdots, h_{iq}\},\ i = 1, 2, \cdots, L$ may be thought of as the impulse responses of a bank of FIR filters. Fig.1 shows this filter bank interpretation of the time series model. Note that the model in Fig.1 has also been given adjustable biases θ_i and gains G_i. With these modifications the prediction $\hat{x}(t+1)$ is given by:

$$\hat{x}(t+1) = -\sum_{i=1}^{L} w_i(t)\varphi(G_i(t) < \boldsymbol{h}_i, \boldsymbol{x}^t > -\theta_i(t)) \ . \quad (12)$$

where the impulse responses $\boldsymbol{h}_i,\ i = 1, 2, \cdots, L$ are fixed. We call this model a *dynamical-functional artificial neural network (D-FANN) predictor* in accordance with the name of the underlying neural network model, as proposed by deFigueiredo in [14].

In the modeling (training) phase the problem of finding the weights of the first layer reduces to that of finding a best set of basis vectors by which the signal $\boldsymbol{x}^t$ can be represented. The optimal weights of the second layer are then adjusted in a supervised learning process by minimizing the error function $\mathcal{E}[e^2(t+1)]$ of (11).
From communication theory it is known that the optimal way (in terms of minimum distortion) of transform coding a signal is by using the Karhunen-Loéve Transform (KLT). The Discrete Cosine Transform (DCT) is the approximation to the KLT that empirically has been found to have the best variance distribution. For our signal segment $\boldsymbol{x}^t$ the DCT is defined as:

$$\boldsymbol{X}^t = \mathcal{C}\, \boldsymbol{x}^t, \quad (13)$$

where the matrix denoted by $\mathcal{C}$ has elements defined as:

$$C_{in} = (\frac{2}{N})^{1/2}[k_i \cos(\frac{(i-1)(n-\frac{1}{2})\pi}{q+1})], \quad i, n = 1, 2, \cdots, q+1, \quad (14)$$

where $k_i = \frac{1}{\sqrt{2}}$ if $i = 1$, else $k_i = 1$. Each component of $\boldsymbol{X}^t$ is given by

$$X_i(t) = k_i(\frac{2}{N})^{1/2} \sum_{l=1}^{q+1} cos(\frac{(i-1)(l-\frac{1}{2})\pi}{q+1})\, x(t-l), \quad i = 1, 2, \cdots, q+1. \quad (15)$$

The DCT is equivalent to a bank of narrow-band all-zero filters operating in parallel, that decomposes the input signal into a set of frequency components. It has the appealing properties that the basis vectors are orthogonal, and since the basis set are not signal dependent it can be predefined. There also exist fast algorithms for its implementation on a computer. In the DCT-based D-FANN model the impulse response of filter i is hence defined as $\boldsymbol{h}_i = \{C_{i1}, C_{i2}, \cdots C_{i\,q+1}\},\ i = 1, 2, \cdots, L$.
The parameters of the second layer of the D-FANN, i.e. the w's, the θ's, and the G's, may be obtained by minimizing the instantaneous squared prediction error using the *back-propagation algorithm*, or minimizing the exponentially weighted sum of squared prediction error using *the nonlinear recursive least squares algorithm*. In this paper we have used the latter.

4. THE ICA D-FANN PREDICTOR

In this section we describe how we combine the D-FANN predictor and the ICA transform to recover a missing section in one of a set of multivariate time series. We assume that $N = M$, i.e. the number of mixtures is equal to the number of sources, that up to time $t \leq K$ we are given the complete set $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_N(t)]$, and for $t > K$ time series $x_1(t)$ is missing samples. We make use of the dependencies between the different time series in the following way:
Using the FastICA algorithm, we estimate $\mathcal{A}$ and $\boldsymbol{s}(t)$ from the mixtures $\boldsymbol{x}(t)$. The estimated source signals are now given by

$$\boldsymbol{u}(t) = \hat{\boldsymbol{s}}(t) = \hat{\mathcal{W}}\boldsymbol{x}(t) = \hat{\mathcal{A}}^{-1}\boldsymbol{x}(t), \quad (16)$$

where $\hat{\mathcal{W}} = \hat{\mathcal{A}}^{-1}$ is the estimated separating matrix, and $\hat{\mathcal{A}}$ is the estimated mixing matrix. We will in the following assume that $\mathcal{A}$, and consequently $\mathcal{W}$, is stationary, and for the sake of convenience we will refer to the observation space as the x-space, and the ICA space (space of separated sources), as the u-space.
The procedure is to use $\hat{\mathcal{W}}$ to transform $\boldsymbol{x}(t)$ to u-space, and perform predictions in u-space. Then the prediction results are transformed back to the x-space. The prediction error is computed in x-space, and re-transformed to the u-space. Based on these errors in u-space, we update the predictor parameters. Specifically, let

$$\hat{\boldsymbol{u}}(t+1) = [\hat{u}_1(t+1), \hat{u}_2(t+1), \cdots, \hat{u}_N(t+1)] \quad (17)$$

denote the one-step predictions of $\boldsymbol{u}(t+1)$. Then we obtain the predictions of $\boldsymbol{x}(t+1)$ as

$$\hat{\boldsymbol{x}}(t+1) = \hat{\mathcal{A}}\hat{\boldsymbol{u}}(t+1). \quad (18)$$

Let $\boldsymbol{e}_x(t+1)$ denote the prediction error in x-space. The components of $\boldsymbol{e}_x(t+1)$ are then defined as

$$e_{xj}(k+1) = \begin{cases} 0 & j = 1,\ k \geq K \\ x_j(k+1) - \hat{x}_j(t+1) & \text{otherwise,} \end{cases} \quad (19)$$

i.e. for the component where data is missing we set the prediction error to zero. Next, we calculate:

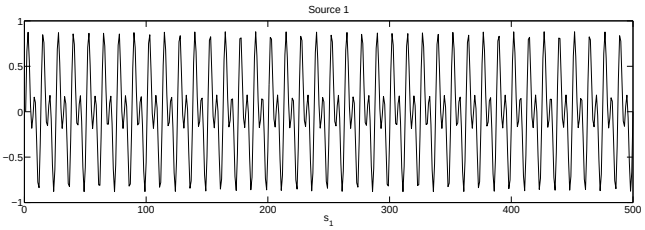
$$\boldsymbol{e}_u(t+1) = \hat{\mathcal{W}}\boldsymbol{e}_x(t+1), \quad (20)$$

which is used to adjust the weights of the second layer of the D-FANN.
Note that because of the assumption of perfect prediction of $x_1(t+1)$ for $t \geq K$, we will modify the prediction error when transforming back to u-space. Intuitively, the more mixtures we have available, the less importance the approximation of $e_{x1}(t+1) = 0$ is expected to have on the predictions.

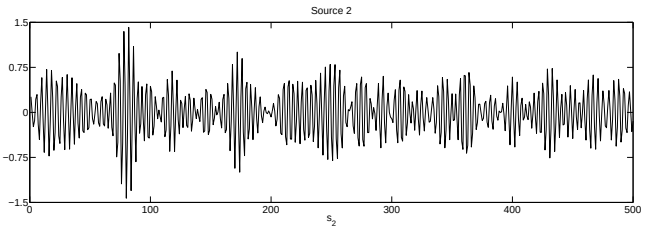
5. RESULTS

5.1. Synthetic data

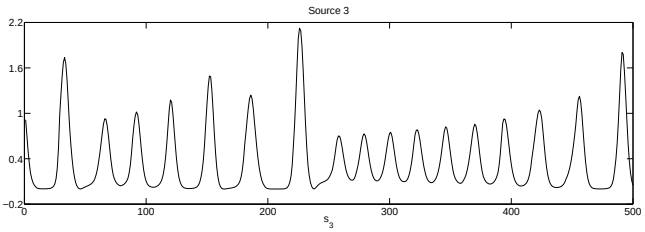
In Fig.2 we have plotted the four synthetic time series, which we will refer to as source 1, source 2, source 3, and source 4. Source 1 is a multiplication of two random sinusoids. Source 2 results from filtering standardized white Gaussian noise through a Butterworth bandpass filter. Source 3 is a solution of the Lorenz equations,



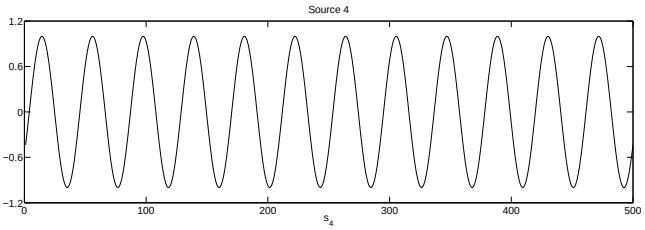
(a) Source 1



(b) Source 2

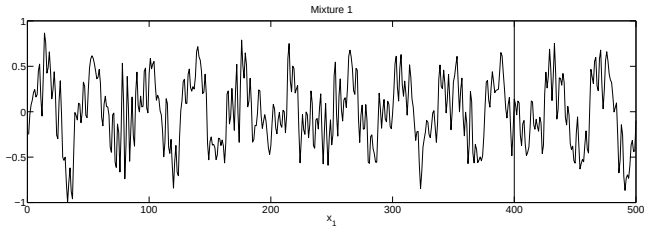


(c) Source 3

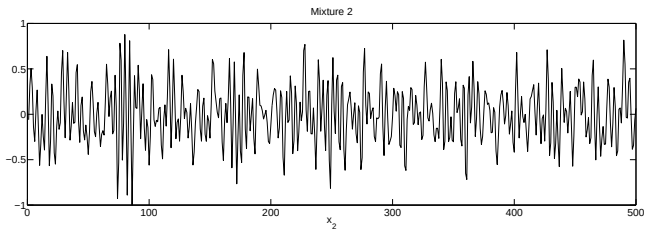


(d) Source 4

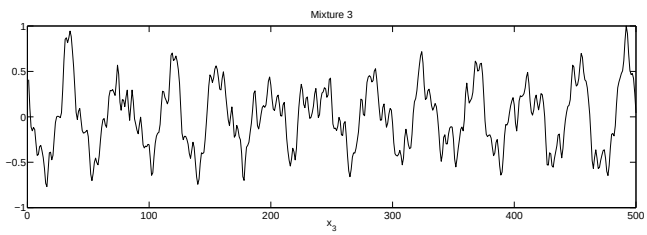
Fig. 2. Original source signals.



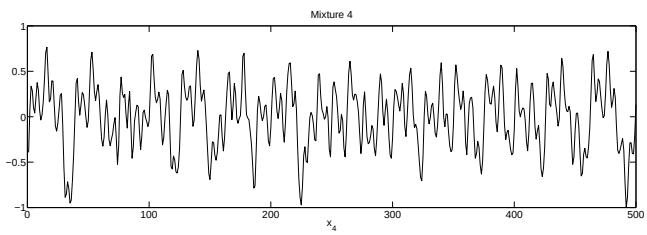
(a) Mixed signal 1



(b) Mixed signal 2



(c) Mixed signal 3



(d) Mixed signal 4

Fig. 3. The four resulting mixed signals.

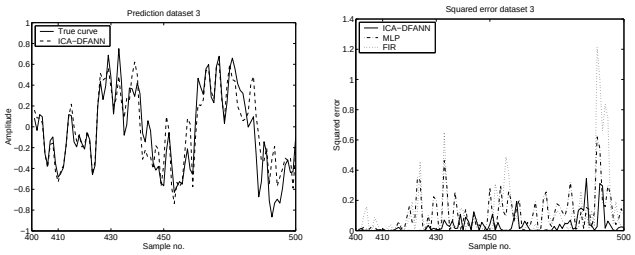
while source 4 is a randomly chosen pure sinusoid.
These signals were mixed using the mixing matrix in (21),

$$\mathcal{A} = \begin{bmatrix} 1.2 & 1.5 & -1.4 & 1.4 \\ 0.8 & -1.3 & 0.3 & -0.1 \\ -1.1 & 0.4 & 1.8 & -1.9 \\ 2.1 & -0.6 & -2.5 & 0.6 \end{bmatrix}, \tag{21}$$

and the resulting signals are shown in Fig.3. The mixtures in Fig.3 have been scaled between -1 and 1. This scaling is important to weigh the signal components equally heavy in the prediction process.
We now assume these mixtures to be the available data. We have no information about $\boldsymbol{s}(t)$ and $\mathcal{A}$, and we assume that the sample range 401-500 of the first mixture is missing. This is the segment succeeding the vertical line on mixture 1 in Fig.3. Our objective is to recover this signal segment using the strategy outlined above.
As a quantitative performance measure of the three predictors we have used

$$R_p(T) = 10 \log_{10} \left(\frac{\delta_s^2(T)}{\delta_p^2(T)} \right),$$

where $\delta_s^2(T)$ is the mean-square value of the true curve over the T samples to be predicted, and $\delta_p^2(T)$ is the corresponding value of the prediction error for these samples. The result is presented in Fig.7, a and b. Fig.7a shows the true and predicted signals over the actual period. In this case the predictor has an embedding dimension of $q = 50$. We notice that the predicted signal quite nicely has captured the overall dynamics of the true signal over the whole segment of 100 samples. Fig.7b shows the instantaneous mean squared errors of our predictor compared to the results using an ordinary multilayer perceptron (MLP) network, and a FIR MLP network. The R_p-values in this case are 6.03, .081, -.77 for the ICA D-FANN, the MLP and the FIR MLP, respectively. Our method hence outperforms a traditional strategy using single a pre-



(a) Predicted and true signals (b) Squared error performance.

Fig. 4. Prediction result for synthetic data.

dictive neural network.

5.2. Real data

The real data we have used are ground measurements of meteorological data from the state Minnesota of the United States. These data are available at National Climatic Data Center (NCDC) [15] and include sequential "time biased corrected" monthly average temperatures, precipitation and Palmer indices. The indices are the Palmer drought severity index (PDSI), Palmer hydrological drought index (PHDI), modified Palmer drought index (PMDI) and the Palmer "Z" index (ZNDX). The period of recording is from 1895 to 1998. The data have been plotted in Fig.5. (For more information about the data, we refer to [16, 15].)

From Fig.5 we see that the temperature series, and perhaps also the precipitation series, has a kind of periodicity, while the Palmer drought indices seem more or less stochastic. It may look like the PDSI, PHDI and the PMDI are equivalent, but this is not the case [15].

For prediction, we will assume that we do not have available the last ten years of the PHDI, i.e. the years from 1988 to 1998, which corresponds to 120 samples. The beginning of the prediction period is marked by a vertical line in the PHDI series in Fig.5.

Note that we have scaled the data between -1 and 1. This is to obtain equal weighing when calculating the errors in x-space. In this experiment we used an embedding dimension of 25.

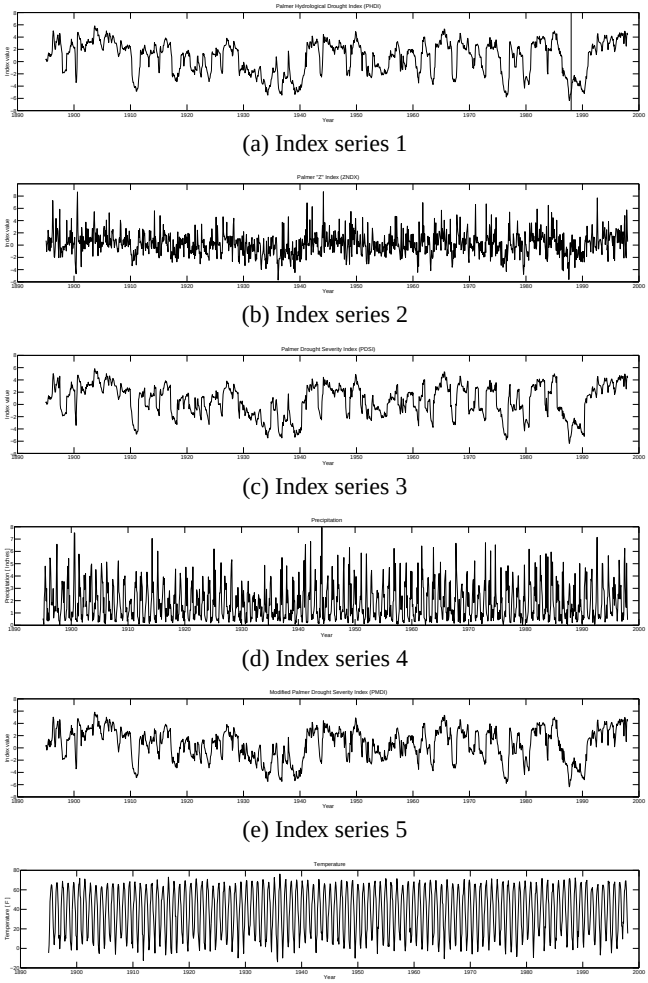
Fig.6 shows the estimated source signals. We note that sources 1 and 2 look like impulse noise series, whereas source 5 has periodic fluctuations that may be related to seasonal variations. The physical significance of these time series are however hard to interpret.

Fig.7a shows the results of predicting the missing segment using the ICA D-FANN, the MLP, and the FIR MLP predictors, respectively. Also plotted is the true signal. Also in this case we observe that our proposed method is superior to the other two. This is confirmed by the instantaneous mean squared error plotted in Fig.7b. The numerical R_p-values are 5.82, -.092, -.75 for the ICA D-FANN, the MLP and the FIR MLP, respectively.

6. CONCLUSION

In this paper we have proposed a new prediction tool for filling in gaps in a time series belonging to a set of mutually statistically dependent signals. This predictor is a combination of the independent component analysis (ICA) transform and the dynamical-functional artificial neural network (D-FANN). The idea is to use the DCT based D-FANN predictor to do nonlinear predictions of the signals in the ICA domain, where they are assumed to be easier to predict, and exploiting the dependencies between the signal components by calculating the available prediction errors in the observation domain.

This strategy has been tested on a computer generated time series, and on a real time series (meteorological ground measurements from the state Minnesota of USA). We found that the ICA D-FANN predictor was capable of recovering the missing gaps quite accurately, and it could follow the dynamics of the true curve quite well. Compared to the MLP and FIR networks, it was occasionally superior, especially for the real time series. An appealing property of the ICA D-FANN predictor is that it does not seem to be very sensitive of the embedding dimension. For the MLP and FIR MLP networks the choice of embedding dimension was crucial for their prediction performance.



(a) Index series 1

(b) Index series 2

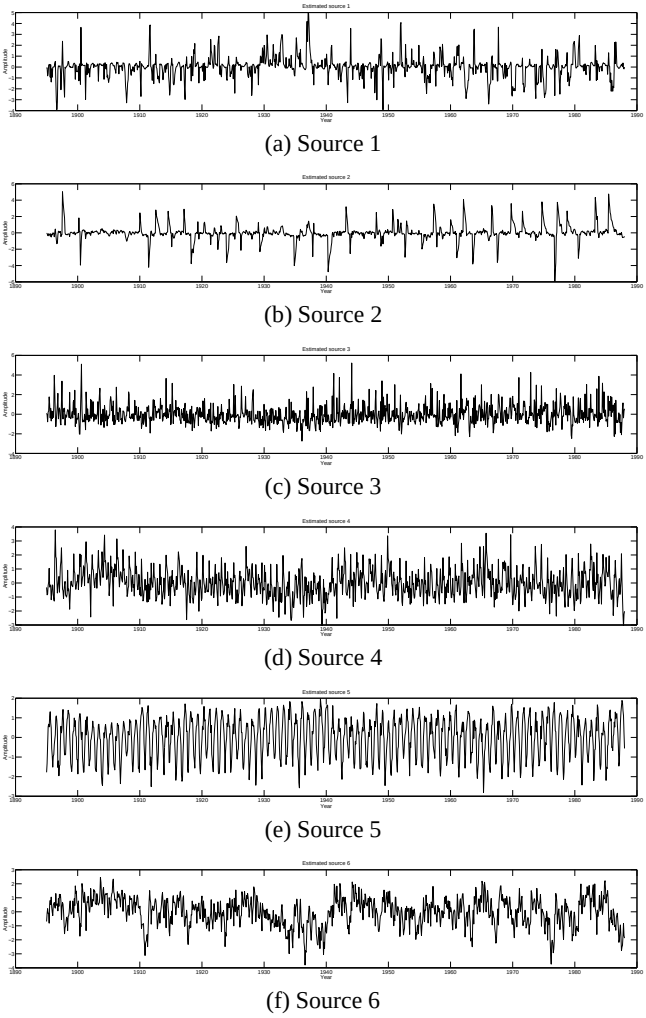
(c) Index series 3

(d) Index series 4

(e) Index series 5

(f) Index series 6

Fig. 5. From a) to f) we have: Palmer hydrological drought index, Palmer "Z" index, Palmer drought severity index, precipitation, modified Palmer severity index, and temperature.



(a) Source 1

(b) Source 2

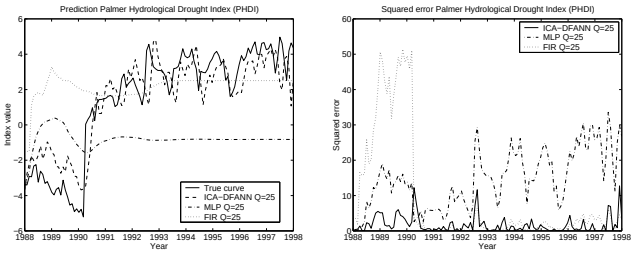
(c) Source 3

(d) Source 4

(e) Source 5

(f) Source 6

Fig. 6. Estimated source signals.



(a) Predicted and true signals (b) Squared error performance.

Fig. 7. Prediction result for real data.

7. REFERENCES

[1] F. Takens, *Detecting strange attractors in turbulence*, vol. 898, Springer, Berlin, 1980, (*Warwick,1980*).

[2] E. A. Wan, "Time series prediction by using a connectionist network with internal delay lines," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, Eds., pp. 195–217. Addison-Wesley Publishing Company, Inc, Reading, MA, 1994.

[3] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 526–535, 1995.

[4] T. Eltoft and R. J. P. deFigueiredo, "A dct-based d-fann for nonlinear adaptive time series prediction," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 1131–1134, 2000.

[5] P. Comon, "Independent component analysis - a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[6] T. Eltoft and R. J. P. deFigueiredo, "Nonlinear adaptive time series prediction with a dynamical-functional artificial neural network," *Submitted to IEEE Trans. Neural networks*, 2000.

[7] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.

[8] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 6, pp. 1129–1159, 1995.

[9] S. Malaroiu K. Kiviluoto and E. Oja, "Ica preprocessing for time series prediction," in *Proceedings of the second international workshop on "Independent Component Analysis and Blind Signal Separation"*, P. Pajunen and J. Karhunen, Eds., Helsinki, 2000, pp. 453–457.

[10] P. Werbos, "Generalization of backpropagation with application to a recurrent gas market model," *Neural Networks*, vol. 1, pp. 339–356, 1988.

[11] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, 1996.

[12] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, 1998.

[13] R. J. P. deFigueiredo, "A generalized fock space framework for nonlinear system and signal analysis," *IEEE Trans. Circuits Syst.*, vol. 30, pp. 637–647, 1983.

[14] R. J. P. deFigueiredo, "Optimal interpolating and smoothing functional artificial neural networks (fann's) based on a generalized fock space framework," *Circuits Syst. Signal Processing*, vol. 17, pp. 271–287, 1998.

[15] NCDC, "Temperature, precipitation, drought, data," *Http:/www.ncdc.noaa.gov/onlineprod/drought/ftppage.html*, 2000.

[16] W. K. Johnson and R. W. Kohne, "Susceptibility of reservoirs to drought using palmer index," *Journal of Water Resources Planning and Management*, vol. 119, pp. 367–387, 1993.