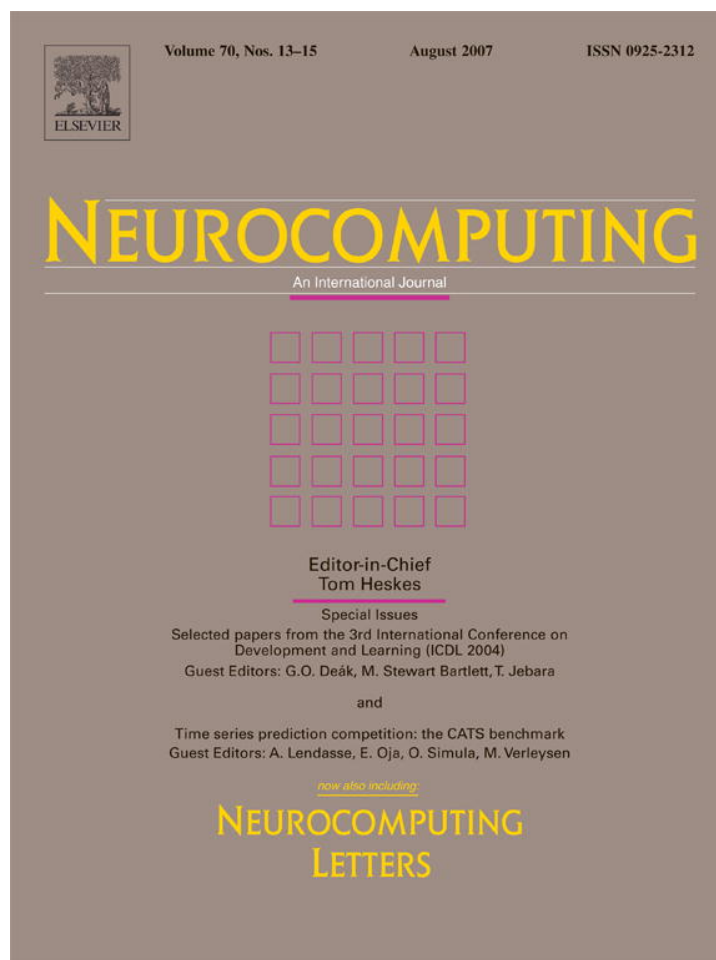


Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

Symmetric-embedding prediction of the CATS benchmark

P.F. Verdes^{a,*}, P.M. Granitto^b, M.I. Széliga^c, A. Rébola^c, H.A. Ceccatto^c

^a*Institut für Umweltphysik, Universität Heidelberg, Im Neuenheimer Feld 229, D-69120 Heidelberg, Germany*

^b*Istituto Agrario di San Michelle a/A, Via E. Mach 2, I-38010 San Michelle a/A, Italy*

^c*Instituto de Física Rosario, CONICET and Universidad Nacional de Rosario, Blvd. 27 de Febrero 210 Bis, S2000EZF Rosario, Argentina*

Available online 22 February 2007

Abstract

We present a general strategy for filling the missing data of the CATS benchmark time series prediction competition. Our approach builds upon a time-symmetric embedding of this time series and the use of a one-shot forecasting for each missing value inside the gaps from distant-enough delayed and forwarded predictors. In the extrapolation region we perform standard, non-iterated forward predictions. For modeling purposes we consider bagging of multi-layer perceptrons (MLPs). We discuss two different implementations of this strategy: The first one is based on a simultaneous modeling of both large- and short-scale dynamics information, using (suitably delayed and forwarded) original CATS values and their first differences as inputs to MLPs. The second one follows a two-stage strategy, in which behaviors at different scales are modeled separately. First, the overall behavior at large scales is fitted with a smooth curve obtained by repeated application of a Savitzky-Golay filter. Then, the remaining short-scale variability is approximated using bagged MLPs. Expected error levels for these two implementations are provided according to performance on test data.

© 2007 Elsevier B.V. All rights reserved.

Keywords: CATS benchmark; Time series analysis; Multilayer perceptrons; Ensembles

1. Introduction

In this paper we explore possible forecasting strategies for the CATS benchmark time series prediction competition. The usual practice assumes that the underlying process can be modeled in a d -dimensional pseudo-phase space according to

$$x_{t+1} = f(\mathbf{x}_t) + \eta_t, \quad (1)$$

where $\mathbf{x}_t = (x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau})$ is the standard homogeneous time-delayed embedding defined by lag τ ($= 1$ in the following) and dimension d [1], f is the — possibly non-linear — dynamics to be estimated from the data and η_t is some residual noise. Within this general framework, after some exploratory data analysis we propose a prediction method based on two distinctive characteristics:

- the use of past and future information on the series behavior in the modeling process, by inclusion of time-delayed and time-forwarded predictors; and

- the consideration of different non-homogeneous embeddings, each one specifically tailored for a particular predictee position inside the missing-data intervals.

Guided by a study of the time-delayed mutual information of the CATS series, we select relevant predictors resulting in a non-homogeneous embedding space. This general embedding is later mirrored in time to utilize past and future dynamics information, which is naturally suggested by the gap nature of most of the missing data (only 20% of the test points corresponds to an extrapolation in time). Finally, to avoid uneasy iterated predictions in a non-homogeneous space, the general time-symmetric embedding is custom-tailored to each predictee position, thus allowing a one-shot prediction strategy. For modeling purposes we use both single and bagged multi-layer perceptrons (MLPs).

This work is organized as follows: In Section 2 we first discuss the need for a non-linear modeling of the unknown dynamical function f and the selection of relevant predictors through a mutual information study. Then, we test different embedding frameworks suggested by this exploratory data analysis and consider the inclusion of

*Corresponding author. Tel.: +49 6221 546334; fax: +49 6221 546405.
E-mail address: Pablo.Verdes@iup.uni-heidelberg.de (P.F. Verdes).

information from the series' future behavior. Finally, we discuss the concrete one-shot prediction method used in the CATS benchmark competition. In Section 3 we present the results obtained for the original values of the benchmark and in Section 4 we comment on the advantages and disadvantages of our prediction method. Then, in Section 5 we propose a slightly different approach suggested by the exploratory data analysis of Section 2.1 and in Section 6 we apply it to the new values of the CATS benchmark. Finally, in Section 7 we draw a few conclusions.

2. Method for prediction

2.1. Exploratory data analysis

In this subsection we will answer the following questions: Should non-linear models be employed to approximate the dynamical function f ? If so, which predictors (time-delayed coordinates) are potentially useful for modeling the data?

2.1.1. Linearity vs. non-linearity

Following Casdagli and Weigend [3,4], we use local linear models to test for non-linearity. They can be considered as the local Taylor expansion of the unknown f , and are easily determined by minimizing

$$\sigma^2 = \sum_{x_j \in N(x_t, \varepsilon)} (x_{t+1} - \mathbf{a}_t \mathbf{x}_j - b_t)^2 \quad (2)$$

with respect to \mathbf{a}_t and b_t , where $N(x_t, \varepsilon)$ is the ε -neighborhood of x_t , excluding x_t . This minimization can be solved through a set of coupled linear equations, which is a standard linear algebra problem. Then, the prediction is $\hat{x}_{t+1} = \mathbf{a}_t \mathbf{x}_t + b_t$. We compute the normalized mean squared error (NMSE) between the original and predicted values, i.e., the MSE divided by the data variance, as a function of the size ε of the neighborhood on which the local linear model is fitted [8]. If the minimum NMSE occurs at large neighborhood sizes, then the data will be (at least in the embedding space considered) best described by a global linear model. In contrast, if the optimum occurs at

small neighborhood sizes, then some non-linear model may be a better choice. Notice that we are disregarding a possible non-stationary behavior of the system generating the data.

In the right panel of Fig. 1 we plot the results for the original time series using homogeneous time-delayed embedding vectors of dimensions $d = 2, 3, \dots, 20$. As we can see in this figure, for all dimensions considered we find the best modeling performances for maximum neighborhood sizes. This implies, according to the discussion above, that a global linear model best describes the raw data. However, this could be simply due to the high autocorrelation of the large-scale behavior of the signal. To study the character of the small-scale fluctuations of x_t , we conducted the same analysis on the series of first differences $\Delta x_t = x_t - x_{t-1}$ (shown in the left panel of Fig. 1). We highlight some interesting features: first, notice that for $d = 2$ (upper curve) the data look like random noise — recall that unpredictability is characterized by a value of $\text{NMSE} \simeq 1$. However, in higher dimensions there is some gain in NMSE for small neighborhood sizes (notice that this series is overall much less predictable than x_t). In conclusion, in this case the results indicate the presence of a weak non-linear determinism. This suggests that the best strategy might well consist of building first a linear autoregressive model to account for the large-scale behavior, and then modeling the remaining unexplained variability with a non-linear predictor (the NMSE levels in the left panel of Fig. 1 indicate that we might expect to explain up to 20–25% of the remaining variance if we use local linear models). However, in a first attempt we will prefer the alternative path of simply gathering together large and small-scale information (x_t and Δx_t , respectively) in a unique state-space description of the system to build a non-linear model of the dynamics.

2.1.2. Selection of relevant predictors

Since the large-scale behavior seems to be only linearly autocorrelated, for the selection of relevant predictors we focus on the more interesting non-linear series of first

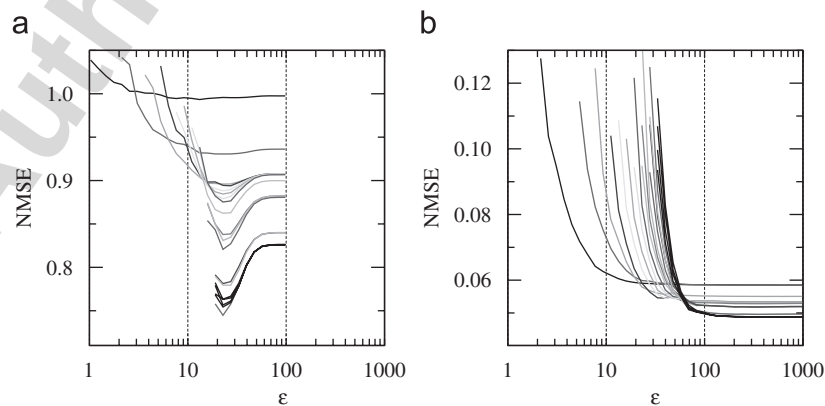


Fig. 1. Normalized mean squared error as a function of neighborhood size ε for $\tau = 1$ and embedding dimensions $d = 2$ to 20: (a) series of first differences; (b) original time-series.

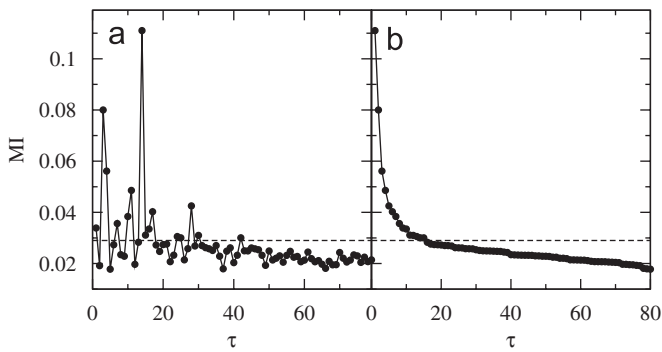


Fig. 2. (a) Time-delayed mutual information (MI) as a function of delay τ for the series of first differences. (b) Rank-ordered values of MI. The dashed horizontal line indicates the threshold 0.029 employed for selecting predictors.

differences. In the left panel of Fig. 2 we show the decay of the time-delayed mutual information (MI) of this series for increasing time lags τ . MI is a natural generalization of linear autocorrelation to investigate possible non-linear dependencies between lagged variables [5].

We will consider Δx_{t-n} , $n = 1, 2, \dots$, to be a relevant predictor of Δx_t if the corresponding MI value exceeds a certain threshold. To fix a concrete value, we rank-ordered the obtained results for MI (shown in the right panel of Fig. 2). From the overall behavior of this curve, it seems reasonable to set a threshold near 0.03. We have chosen the particular value 0.029, indicated with a dashed horizontal line in both graphs, since at this value there is a small step down in relevance as measured by MI. Then, this procedure pins out $n \in D_n = \{1, 3, 4, 7, 10, 11, 14, 15, 16, 17, 24, 25, 28, 30, 42\}$. Notice that we could have chosen a threshold near 0.33, where there is also a noticeable step down in MI. This would have eliminated the lags $n = 15, 24, 25, 30$ and 42. However, we preferred to include the corresponding long-delayed predictors as inputs to the model.

Although these time lags were chosen by analysing $\{\Delta x_t\}$, hereafter we will assume that they also provide a sufficiently well sampled representation for the more slow, autocorrelated $\{x_t\}$. We thus propose to employ $\{x_{t-n}, \Delta x_{t-n}\}$ as predictors for all delays $n \in D_n$.

2.2. Testing the embedding framework

In the first part of this subsection we assess the above proposed embedding by t -testing the

- null hypothesis: “The performance of a non-linear model obtained from the set of inputs $\{x_{t-n}, \Delta x_{t-n}, n \in D_n\}$ is not better than the one obtained from $\{x_{t-n}, n \in D_n\}$ alone”.

Then, given the gap nature of most of the missing data, in the last part of this subsection we address a key question: Could a better modeling be obtained by including information from the future? In all cases we have used single MLPs as non-linear modeling technique.

2.2.1. Inclusion of Δx_{t-n}

To see how much it can be gained by including $\{\Delta x_{t-n}, n \in D_n\}$ as predictors, we have trained MLPs with and without these first differences as model inputs. In particular, we considered architectures $15:h:1$ and $30:h:1$, varying the number of hidden units h between 5 and 20. For this study, we randomly selected 33% of the points in the whole record for testing purposes (set \mathcal{T}) and the remaining 67% data points for learning purposes (set \mathcal{L}). The MLPs were trained on bootstrap re-samples of \mathcal{L} , using the out-of-sample data (approximately 37% of the points in \mathcal{L}) to validate the models. All the networks were trained until the minimum of the corresponding validation-set error was achieved (“early-stopping” criterion). We found that no performance improvement was obtained for h larger than 10, so we selected this number of hidden units for model comparison.

In order to produce a paired t -test, both embedding settings were run on parallel on 100 different random splits of the data in \mathcal{T} and \mathcal{L} sets. We computed the relative performance differences

$$\frac{\Delta E}{E} = \frac{(\text{NMSE}_{\{x\}} - \text{NMSE}_{\{x, \Delta x\}})}{\text{NMSE}_{\{x\}}} \quad (3)$$

between the minimum validation errors obtained using the $\{x_{t-n}, n \in D_n\}$ set of inputs ($\text{NMSE}_{\{x\}}$) and using the full set of inputs $\{x_n, \Delta x_n, n \in D_n\}$ ($\text{NMSE}_{\{x, \Delta x\}}$). The results of this t -test were 100% conclusive: in all of the 100 experiments the MLPs trained with $\{x_n, \Delta x_n, n \in D_n\}$ outperformed the $\{x_n, n \in D_n\}$ fitting. In particular, the predictions were, on average, 2.4% better.

2.2.2. Inclusion of data from the future

The fact that only 20% of the test points correspond to an extrapolation in time naturally suggests including information from the future in the modeling strategy. For this, which “forward” predictors should be employed? To be consistent, we apply the same selection criterion described above. Since MI is by definition symmetrical, i.e., $\text{MI}(x, y) = \text{MI}(y, x)$, the time-lagged mutual information of x_t is invariant under time reversal: $\text{MI}(x_t, x_{t-\tau}) = \text{MI}(x_{t-\tau}, x_t)$. Thus, the picture for threshold selection of forward predictors is again given by Fig. 2, and the same set of 15 relevant delays must be mirrored into the future.

Now, suppose we were to predict a particular missing value within a gap, say, e.g., the point in the 10th position counting from the gap’s left edge. Avoiding the common exponential error growth associated to iterated predictions, we choose a one-shot strategy: we restrict ourselves to considering only predictors $\{x_{t-n}, \Delta x_{t-n}\}$ for $n = 10, 11, 14, \dots, 42$ to forecast x_t . In this context, we ask ourselves: Can we improve the modeling performance by enlarging the “past” embedding framework with the addition of $\{x_{t+n}, \Delta x_{t+n}\}$, $n = 11, 14, 15, \dots, 42$? To answer this question we have run 100 experiments with different random splits of the data as above, training MLPs with architectures $22:10:1$ and $42:10:1$. The results of the 100 runs are plotted

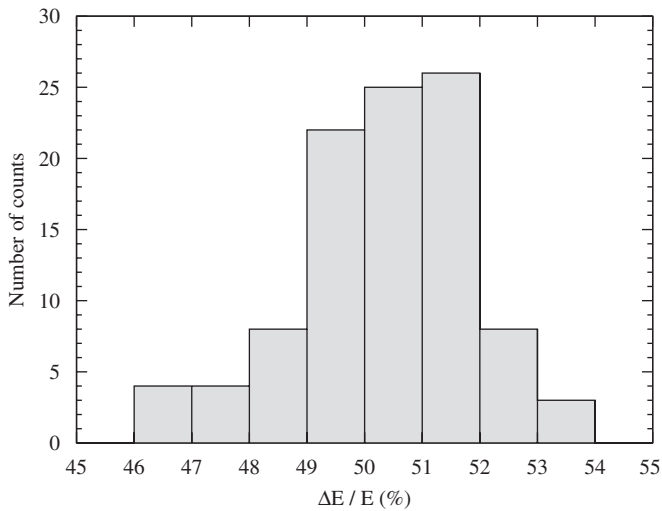


Fig. 3. Relative performance difference $\Delta E/E$ between results prior to and after inclusion of data from the future. The results of 100 independent runs are plotted as a histogram.

as a histogram in Fig. 3. It can be seen from this figure that the inclusion of information from the future leads to a dramatic error reduction. In particular, the relative performance improvement is, on average, 50.4%.

2.3. The proposed prediction method

From the studies in the two previous subsections, as a concrete prediction strategy for the original CATS benchmark competition we propose the following method:

- use a non-linear technique (MLPs) to model the unknown dynamical function f ;
- consider a general non-homogeneous time-symmetric embedding framework $\{x_{t\pm n}, \Delta x_{t\pm n}, \text{ with } n \in D_n\}$, selected by the study of the CATS series time-delayed MI behavior; and finally
- from this set of predictors, custom-tailor the particular embedding setting for each predictee position in a gap to avoid iterated predictions.

Numbering from 1 to 20 the points within the gaps, in Table 1 we summarize the temporal locations of the predictors involved in each case. Needless to say, for the last 20 points of the whole record (extrapolation region) we are forced to use only the available past delays quoted in the second column of this table.

3. Results obtained on the initial values of the benchmark

To produce the final predictions we considered ensembles of MLPs, since they are known to perform better than single networks [10]. Using again t -testing, we compared the advantages of different aggregation strategies [6,7]. We concluded that bagging [2] was best suited for the problem at hand, showing an improvement over single networks that

Table 1

Relative positions n of the predictors x_{t+n} involved in the “casting” of each missing value within a gap at temporal positions $t+1, \dots, t+20$

Position	Past predictors	Future predictors
1	-42, -30, ..., -1	24, 25, ..., 42
2, 3	-42, -30, ..., -3	24, 25, ..., 42
4	-42, -30, ..., -4	17, 24, ..., 42
5	-42, -30, ..., -7	16, 17, ..., 42
6	-42, -30, ..., -7	15, 16, ..., 42
7	-42, -30, ..., -7	14, 15, ..., 42
8, 9	-42, -30, ..., -10	14, 15, ..., 42
10	-42, -30, ..., -10	11, 14, ..., 42
11	-42, -30, ..., -11	10, 11, ..., 42
12, 13	-42, -30, ..., -14	10, 11, ..., 42
14	-42, -30, ..., -14	7, 10, ..., 42
15	-42, -30, ..., -15	7, 10, ..., 42
16	-42, -30, ..., -16	7, 10, ..., 42
17	-42, -30, ..., -17	4, 7, ..., 42
18, 19	-42, -30, ..., -24	3, 4, ..., 42
20	-42, -30, ..., -24	1, 3, ..., 42

ranged between 1% and 5%. We found that the error decrease becomes negligible after aggregation of a few networks, in such a way that 10 MLPs were judged to be enough for building the composite regressors.

The modeling errors obtained in this way over 500 randomly selected test samples are plotted in Fig. 4 as a function of the predictee position within the missing intervals. In the left panel we depict the expected uncertainties over the first four gaps, using predictors from both the past and future. In the right panel we compare this curve against the expected performance over the extrapolation region, which can only be modeled from past data.

Finally, in Fig. 5 we illustrate the performance of our method on five equally-spaced block-intervals of length 20 that we had excluded from the modeling data for this purpose from the beginning. We can see in this figure that, as expected, predictions using time-symmetric embedding (indicated in black) show a smaller mismatch with targets than one-sided forecasts (grey) near the gap edges. Black and grey curves exemplify possible performances over the four missing intervals and the last extrapolation region, respectively.

4. Analysis of the results: advantages and disadvantages of the proposed method

The estimated generalization errors in our approach, plotted in Fig. 4, deserve a few comments. First, in the left panel of this figure, besides the obvious overall symmetry we notice a non-monotonous increase in NMSE as we incursion into a gap from its edges towards the center. For example, the error in the 7th position is smaller than in the 6th one. To understand this behavior we refer to Table 1, where we find that in passing from position 6 to 7 we do not lose any past predictor but gain two forward ones instead, namely x_{t+14} and Δx_{t+14} . In moving one step further on, to

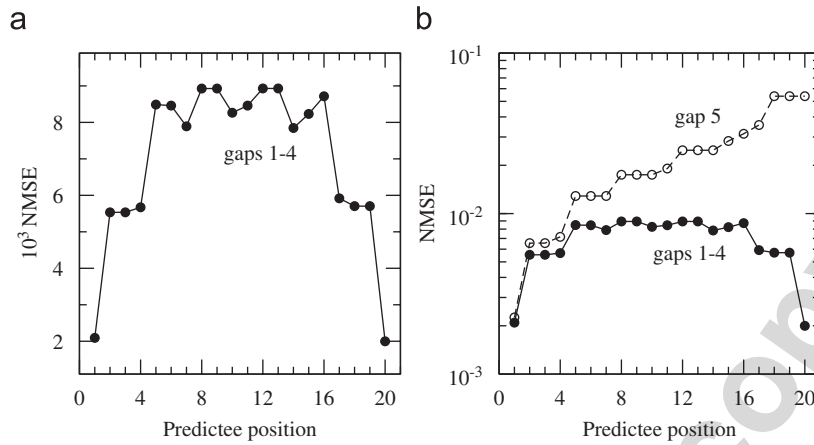


Fig. 4. Estimated distribution profile of modeling errors: (a) first four gaps, symmetric embedding; (b) last interval, one-sided embedding (open circles). The curve shown in (a) is also included here for comparison (full circles).

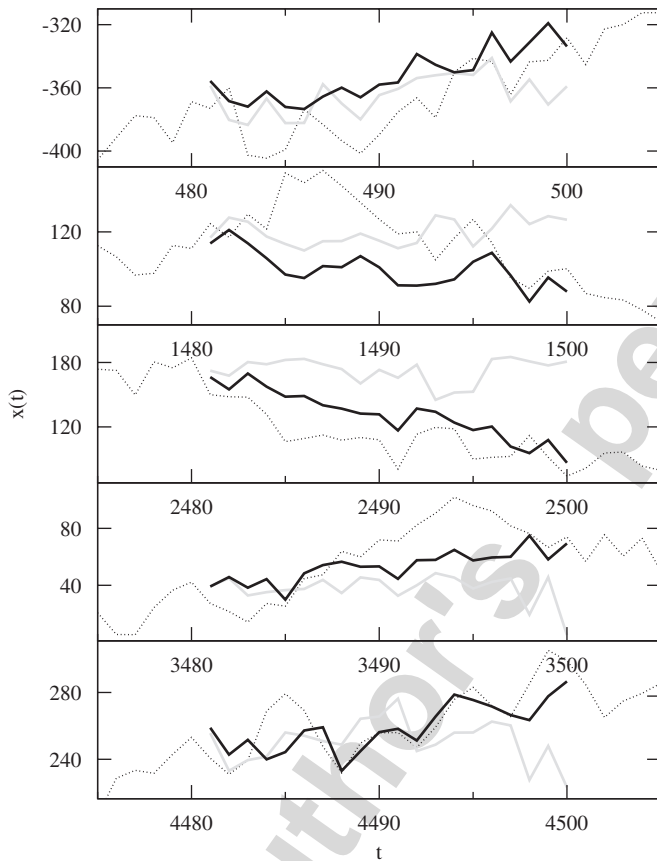


Fig. 5. Examples of validation gaps that were completed according to our procedure. True data are indicated by a dotted thin black line, time-symmetric predictions with a full thick black line, and past-based predictions with a full thick grey line.

position 8, the useful information located at $t - 7$ is lost and the error increases again. In the right panel we observe a monotonous growth in the prediction error on the extrapolation region as the predictee distance to the last-known datum increases. The stepwise nature of this behavior is related to the discrete loss of available past

predictors (see the second column of Table 1). As expected, the results over the gaps are consistently better than over this last interval.

Our exploratory data analysis in Section 2.1 suggested a two-stage modeling strategy. Due to time constraints, its implementation and comparison against the prediction method proposed in Section 2.3 was not possible before the competition deadline. This question will be discussed in the next Section, and the resulting method later applied to the new values of the CATS benchmark on Section 6.

5. Improved method for prediction

At the end of Section 2.1 we concluded that the performance of local linear models of the CATS series as a function of the neighborhood size did not show an intrinsic non-linear structure, but the same study on the first differences Δx_t indicated the presence of a weak non-linear determinism. This suggested the use of a two-stage modeling: first, a smooth global fit to account for the large-scale behavior and, secondly, a non-linear prediction of the remaining unexplained short-scale variability. In this section we explore this possibility.

For the first modeling stage (smooth global fit), the simplest possibility is to use any available low-pass filtering scheme. In particular, in this work we have considered the Savitzky-Golay filter [9]. According to this method, for each value x_t of the series one considers a symmetric window $[x_{t-W}, x_{t+W}]$ of semi-width W centered at this point, and fits a polynomial of order k . Then, the original central value x_t is replaced by the polynomial value at this point. This process is repeated as many times as required to obtain a smooth enough curve. In our case, we have proceeded as follows:

- In addition to the four missing intervals of the competition, we considered four extra gaps of the same size to be used as a validation of this procedure.

- The original and new validation gaps were filled with a straight line connecting the first known points at the gap edges.
- The resulting series was smoothed with a Savitzky-Golay filter, using polynomials of order $k = 1, 2, 4$ and windows of semi-widths $W = 5, 6, \dots, 100$.
- We computed the NMSEs between the curves obtained in this way and the original data at the new validation gaps.
- The combination of polynomial order and window size corresponding to the smallest validation NMSE was chosen.
- This process was repeated several times, using at each step the previously predicted values to fill the gaps, until no further sensible reduction on the validation NMSE was obtained.

In the first step, this process selected second-order polynomials ($k = 2$) and a semi-window size $W = 26$, with

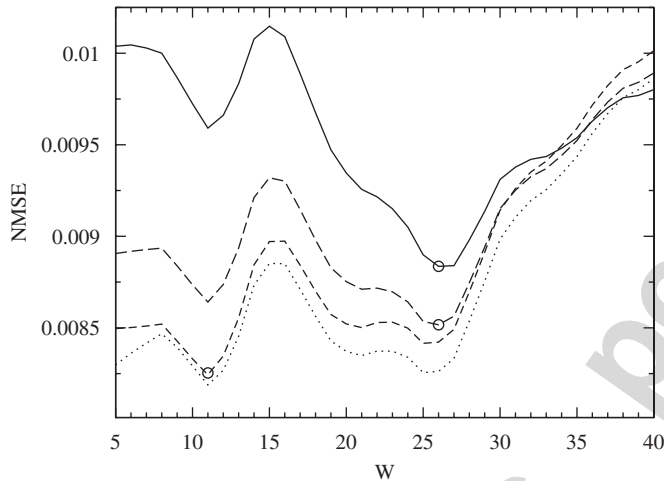


Fig. 6. NMSE obtained by filling the new validation gaps using the Savitzky-Golay filter with parabolas and varying window semi-width W . From top to bottom, lines correspond to the first, second, third and fourth applications of the filter. Circles indicate the optimal W in the first three iterations.

a validation $\text{NMSE} = 8.84 \times 10^{-3}$. Keeping always $k = 2$ for simplicity, in the second step the optimal semi-width size was again $W = 26$ ($\text{NMSE} = 8.52 \times 10^{-3}$) and, in the third step, $W = 11$ ($\text{NMSE} = 8.24 \times 10^{-3}$). In Fig. 6 we plot the validation NMSE at every step as a function of W for fixed $k = 2$. From this figure we see that a further application of the filter does not lead to a sensible validation error reduction, so we finally kept the values obtained after application of the 26-26-11 filter with $k = 2$ as the large-scale fit of the CATS series. For the extrapolation at the final interval of the missing data we have simply fitted a parabola to the last 26 points of this smooth curve.

Once this slowly-varying approximation to the large-scale behavior was obtained, we computed the remaining variability shown by the original CATS series. The predictors for the resulting short-scale dynamics were subsequently selected along the same lines of Section 2.1. Fig. 7 shows the results of this selection. Then, very much like what was done before, we trained bagged MLPs for each predictee position. The performance obtained with this new method and its comparison with the method proposed in Section 2 are presented in the next section.

6. New results obtained on the CATS benchmark

The performance of the new modeling approach was estimated on 500 random samples as before. Fig. 8 shows the estimated generalization errors for the different predictee positions inside the gaps and in the extrapolation region for the *residuals* prediction task. The error level $\text{NMSE} = 1$ should be here interpreted as achieved by the prediction of the smooth Savitzky-Golay curve. If we want to compare these results against those in Fig. 4, which gives similar information on the generalization capability of the first method, we must bear in mind that the error levels of Fig. 8 are an improvement relative to the ground performance of $\text{NMSE} = 8.24 \times 10^{-3}$ given by the Savitzky-Golay slow-dynamics approximation. This

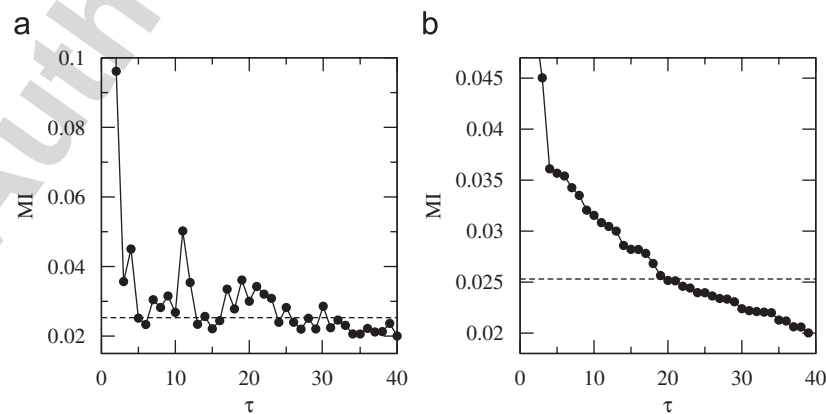


Fig. 7. (a) Time-delayed MI of predictors used in the second stage of the new modeling approach. (b) Rank-ordered values of MI, used to define a threshold for inclusion (dashed line).

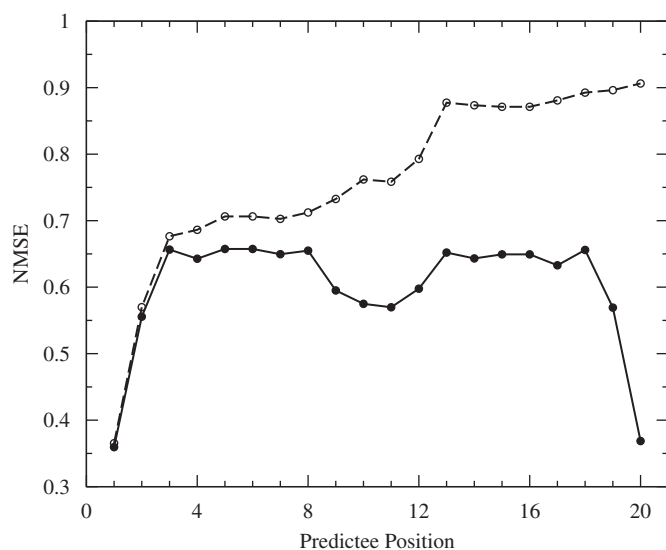


Fig. 8. Estimated error distribution profile in predicting the residuals in the second stage of the new prediction method. Open circles: extrapolation region, one-sided embedding. Full circles: gaps, two-sided embedding.

Table 2

Mean square error between the actual values of the missing data in the competition and their predictions using the previously proposed method of Section 2, the Savitzky-Golay smoothed curve, and the new method of Section 5

Error	Previous method	Savitzky-Golay curve	New method
E_2	442	359	343
E_1	660	634	668

E_2 corresponds to the error on the 4 gaps and E_1 to the total error including the extrapolation region.

comparison shows that the new method is sensibly more accurate than the previous one.

How does the prediction error of the new method compare with that of the previous one on the original missing data from the competition? Using the true data values which are now available, we have computed the corresponding errors and give the results in Table 2. It can be seen from this table that the new method is more accurate for the prediction of the four gaps, but the strategy followed for extrapolation into the future leads actually to a larger total error value than the one corresponding to the method of Section 2. However, according to Figs. 4 and 8, this should only be a misfortune due to peculiarities of the missing data. We stress the fairly good prediction performance of the simple Savitzky-Golay smoothed curve.

Finally, in Fig. 9 we present the predictions of the new method on the (still unknown) missing values of the new CATS benchmark.

7. Conclusions

We have presented a possible forecasting procedure for the CATS benchmark time series prediction competition.

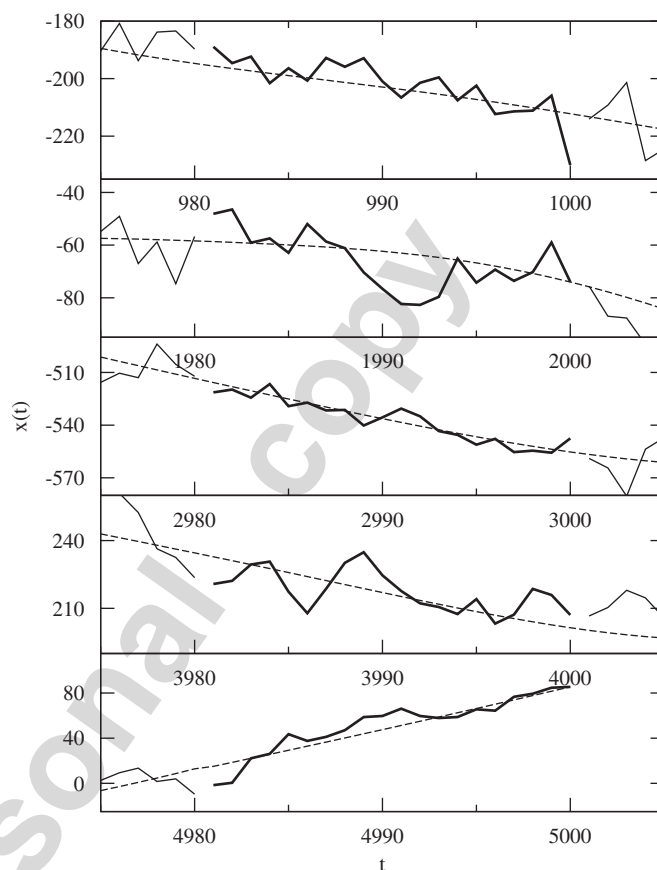


Fig. 9. Predicted values for the missing gaps of the new data, completed according to our improved two-stage procedure. Known data are indicated by a thin full line. Our first-stage model, i.e., the Savitzky-Golay approximation, is plotted with a thin dashed line. Final predictions are depicted with a thick full line.

Our approach is simple and can be summarized as a time-symmetric embedding of the given time series (except for the extrapolation region where future information is naturally unavailable) plus the use of specific information for each missing value to produce one-shot forecastings. As modeling technique we have employed bagged MLPs.

We have discussed two different implementations of this procedure: The first one was based on a simultaneous modeling of both large- and short-scale dynamics information, using (suitably delayed and forwarded) original CATS values and their first differences as inputs to MLPs. The second one followed a two-stage strategy, in which behaviors at different scales were modeled separately. First, the overall behavior at large scales was fitted with a smooth curve obtained by repeated application of Savitzky-Golay filters. Then, the remaining short-scale variability was approximated using bagged MLPs.

Future research should clarify the suitability of the several choices that had to be made throughout the modeling process. Pending investigations include a more careful selection of possible predictors (perhaps with different delays for the series of CATS values and first differences), a comparison against iterated forecasting, etc.

Acknowledgment

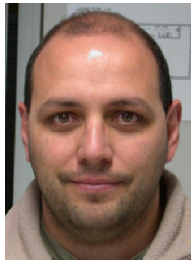
PFV is supported by the Alexander von Humboldt Foundation. PMG acknowledges financial support from PAT project SAMPPA.

References

- [1] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.Sh. Tsimring, The analysis of observed chaotic data in physical systems, *Rev. Mod. Phys.* 65 (1993) 1331–1392.
- [2] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (1996) 123–140.
- [3] M. Casdagli, Chaos and deterministic versus stochastic non-linear modeling, *J. R. Stat. Soc. B* 54 (1991) 303–328.
- [4] M. Casdagli, A.S. Weigend (Eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past*, SFI Studies in the Sciences of Complexity, vol. XV, Addison-Wesley, Reading, MA, 1993.
- [5] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A* 33 (1986) 1134–1140.
- [6] P.M. Granitto, P.F. Verdes, H.D. Navone, H.A. Ceccatto, A late-stopping method for optimal aggregation of neural networks, *Int. J. Neural Syst.* 11 (2001) 305–310.
- [7] P.M. Granitto, P.F. Verdes, H.A. Ceccatto, Neural Networks Ensembles: Evaluation of Aggregation algorithms, *Artif. Intell.* 163 (2005) 139–162.
- [8] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of non-linear time series methods: the TISEAN package, *Chaos* 9 (1999) 413–435.
- [9] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, 1992.
- [10] A.J.C. Sharkey (Ed.), *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*, Springer, London, 1999.



Pablo Verdes obtained a M.Sc. and Ph.D. in Physics in 1997 and 2003, respectively, both from Universidad Nacional de Rosario, Argentina. He is currently a Post-Doc at the Institute of Environmental Physics, University of Heidelberg, Germany. His main research interest is non-stationary time series analysis and prediction.



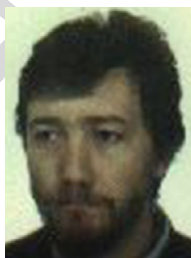
Pablo Granitto obtained a M.Sc. and Ph.D. in Physics in 1997 and 2003, respectively, both from Universidad Nacional de Rosario, Argentina. He is currently a Post-Doc at the Istituto Agrario San Michele all'Addige, Trento, Italy. His research interests include the application of modern machine learning techniques to agroindustrial problems and the development of feature selection and ensemble methods.



Maria Ines Szeliga received a Degree in Physics from Universidad Nacional de Rosario, Argentina, in 2002. She is currently a Ph.D. candidate at Universidad Nacional de Rosario, where she also serves as a teaching assistant within the Electrical Engineering Department. Her research interests include applications of neural networks to modeling and forecasting, statistics, time series analysis, and statistical physics.



Alejandro Rebola received a Degree in Physics in 2002 from Universidad Nacional de Rosario, Argentina, where he is currently a Ph.D. candidate under the supervision of Prof. H. A. Ceccatto.



H. Alejandro Ceccatto obtained a M.Sc. in Physics in 1979 from Universidad Nacional de Rosario, Argentina, and a Ph.D. in Physics in 1985 from Universidad Nacional de La Plata, Argentina. He completed postdoctoral studies at the Department of Applied Physics of Stanford University, USA, 1988, and at the Institut fuer Theoretische Physik, Universitaet zu Koeln, Germany, 1989. He has been the Director of the Intelligent Systems Group of Instituto de Fisica Rosario since 1995.

He has supervised 20 M.Sc. and 11 Ph.D. Theses and is currently a full professor at Universidad Nacional de Rosario.