Local Learning for Iterated Time Series Prediction

Gianluca Bontempi
Iridia
Université Libre de Bruxelles
1050 Bruxelles - Belgium
http://iridia.ulb.ac.be
e-mail: gbonte@ulb.ac.be

Mauro Birattari
Iridia
Université Libre de Bruxelles
1050 Bruxelles - Belgium
http://iridia.ulb.ac.be
e-mail: mbiro@ulb.ac.be

Hugues Bersini
Iridia
Université Libre de Bruxelles
1050 Bruxelles - Belgium
http://iridia.ulb.ac.be
e-mail: bersini@ulb.ac.be

Abstract

We introduce and discuss a local method to learn one-step-ahead predictors for iterated time series forecasting. For each single one-step-ahead prediction, our method selects among different alternatives a local model representation on the basis of a local cross-validation procedure. In the literature, local learning is generally used for function estimation tasks which do not take temporal behaviors into account. Our technique extends this approach to the problem of long-horizon prediction by proposing a local model selection based on an iterated version of the PRESS leave-one-out statistic. In order to show the effectiveness of our method, we present the results obtained on two time series from the Santa Fe competition and on a time series proposed in a recent international contest.

1 INTRODUCTION

The use of local memory-based approximators for time series analysis has been the focus of numerous studies in the literature (Farmer & Sidorowich, 1987; Yakowitz, 1987). Memory-based approaches do not estimate a global model of the dynamic process underlying the time series but defer the processing of data until a prediction is explicitly requested. A database of observed values is stored in memory and the prediction is derived from an interpolation based on a neighborhood of the current state (*locally weighted regression*). These methods have been recently grouped under the denomination of *lazy learning* methods (Aha, 1997). A key issue in local learning is the selection of the local model structure which will have the best generalization performance given a set of sparse and noisy data (*bias/variance* dilemma). Under the assumption that data is described locally by a linear model, enhanced linear statistical procedures can be used to validate the local approximator. One example is the PRESS (Predicted Sum of Squares) statistic (Myers, 1994): this statistic returns the leave-one-out cross-validation error of a linear model as a by-product of the least-squares regression. As a consequence, the performance of a local linear model can be assessed with no additional computational burden, making easier the task of selecting the best local approximator. Once the most promising structure in cross-validation is selected, the local model returns the prediction.

The goal of the paper is to present a method for long term prediction based on the iteration of a one-step-ahead local learning predictor. Forecasting the value of a time series multiple steps ahead requires either a *direct* prediction of the desired value or the *iteration* of a one-step-ahead estimator. In order to emulate the system, a direct predictor often requires higher functional complexity than a one-step-ahead predictor. On the other hand it is more difficult for a one-step-ahead model to deal with the problem of accumulation of errors due to the iteration (Farmer & Sidorowich, 1987). The novelty of our approach consists in defining an iterated formulation of the PRESS leave-one-out statistic as a criterion for local model selection in multiple-step-ahead forecasting. We will show how the iterated PRESS outperforms a non iterated criterion, by assessing the generalization performance of a one-step predictor on a horizon longer than a single step, yet preserving nice properties of computational efficiency. To our knowledge there are no methods in local modeling literature which adopt cross-validation methods for iterated time series prediction. This technique makes local learning an effective alternative to state-of-theart iterated techniques like recurrent neural networks approaches which are typically based on time consuming tuning procedures (e.g. back propagation through time (Rumelhart *et al.*, 1986) or real-time recurrent learning (Williams & Zipser, 1989)).

2 ITERATED AND DIRECT METHODS FOR PREDICTION

A time series is a sequence of measurements φ^t of an observable φ at equal time intervals. The Takens theorem (Packard *et al.*, 1980) implies that for a wide class of deterministic systems, there exists a *diffeomorphism* (one-to-one differential mapping) between a finite window of the time series $\{\varphi^{t-1}, \varphi^{t-2}, \ldots, \varphi^{t-m}\}$ (*lag vector*) and the state of the dynamic system underlying the series. This means that in theory it exists a multi-input single-output mapping (*delay coordinate embedding*) $f : R^m \to R$ so that:

$$\varphi^{t+1} = f(\varphi^t, \varphi^{t-1}, \ldots, \varphi^{t-m+1}) \tag{1}$$

where m (*dimension*) is the number of past values taken into consideration. This formulation returns a state space description, where in the m dimensional space the time series evolution is a trajectory, and each point represents a temporal pattern of length m.

A model of the mapping (1) can be used for two objectives: *one-step* prediction and *iterated* prediction. In the first case, the m previous values of the series are assumed to be available and the problem is equivalent to a problem of function estimation. In the case of iterated prediction, the predicted output is fed back as an input to the following prediction. Hence, the inputs consist of predicted values as opposed to actual observations of the original time series. A prediction iterated for k times returns a *k-step-ahead* forecasting. However, iterated prediction is not the only way to perform *k-step-ahead* forecasting. Weigend (Weigend, 1996) classifies the methods for k-step-ahead prediction, according to two features: the horizon of the predictor and the horizon of the training criterion. He enumerates three cases:

1. the model predicts k steps ahead by iterating a one-step-ahead predictor (Eq. 1) whose parameters are optimized to minimize the error on one-step-ahead forecast (one-step-ahead training criterion)

2. the model predicts k steps ahead by iterating a one-step-ahead predictor whose parameters are optimized to minimize the error on the iterated k-step-ahead forecast (k-step ahead training criterion)

3. the model performs a direct forecast at time $t+k$: $\varphi^{t+k} = f^k(\varphi^t, \varphi^{t-1}, \ldots, \varphi^{t-m+1})$

Methods of class 1 have low performance in long horizon task. This is due to the fact that they are essentially models tuned with a one-step-ahead criterion and therefore they are not able to take temporal behavior into account. Methods like recurrent neural networks (Williams & Zipser, 1989) are an example of class 2. Their recurrent architecture and the associated training algorithm (temporal backpropagation) are more able to handle the time-dependent nature of the data. Direct methods of class 3 often require high functional complexity in order to emulate the system. An example of combination of local techniques of type 1 and 3 is provided by Sauer (Sauer, 1994). In the next section we will present our local technique as a member of the second class of predictors.

3 A LOCAL METHOD FOR ITERATED PREDICTION

We propose a locally weighted regression method to estimate a one-step-ahead predictor trained and selected according to a k-step-ahead criterion. Typically, the data analyst who adopts a local regression approach, has to take a set of decisions related to the model (e.g. the number of neighbors, the kernel function, the parametric family, the distance metric). However, in local learning literature different methods exist which automatically select the adequate configuration (Atkeson *et al.*, 1997; Birattari *et al.*, 1999) by adopting tools and techniques from the field of linear statistical analysis. The most important of these tools is the PRESS statistic which is a simple, well-founded and economical way to perform *leave-one-out* cross-validation and to assess the performance in generalization of local linear models. By assessing the performance of each linear model, alternative configurations can be tested and compared in order to select the best one in terms of expected prediction.

Our method adopts the local learning procedure by replacing the one-step-ahead criterion for model selection, represented by the *conventional PRESS*, with a k-step-ahead criterion, defined as *iterated PRESS*. This criterion is used to select the best structure of a one-step-ahead estimator with the aim of capturing the long term dynamics underlying the available set

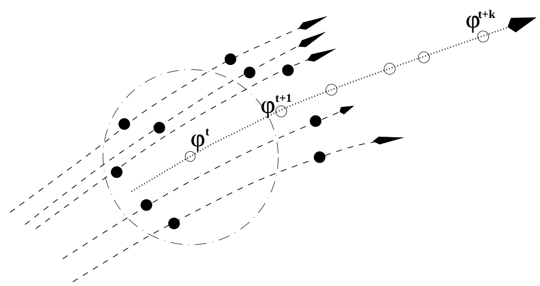

Figure 1: Local modeling in the embedding space

of observations. In this paper we will present experiments where the method is used to select query-by-query (i.e. at each time step) the best number of neighbors by keeping fixed the shape of the regression kernel (tricube) and the distance metric (euclidean). However, this approach can be extended to the selection of any desired model features, such as the set of regressors or the degree of the local approximator (Bersini *et al.*, 1998). Our local learning procedure can be summarized as follows:

1. The available realization of the time series is described in terms of a dataset D_N of N ordered pairs (x_i, y_i), where x_i is a temporal pattern of length m, and the concatenation of x_i and the scalar y_i is a temporal pattern of length $m+1$. Following this formulation, the current state is itself described by a pattern $\bar{x}$ of length m.
2. The one-step-ahead predictor is a local estimate of the mapping $f : \Re^m \rightarrow \Re$ from the dataset D_N. The estimate is performed by a locally weighted regression which selects the neighbors of $\bar{x}$ in the input space $\Re^m$.
3. The k-step-ahead prediction is performed by iterating a one-step-ahead estimator.
4. The local model is selected in a space of alternative model configurations, each characterized by a different number of neighbors.
5. The prediction ability of each alternative model is assessed by an iterated formulation of the cross-validation PRESS statistic (k-step-ahead criterion).

The basic idea of cross-validation for local modeling is represented in Fig. 1 which plots a time series in the embedding space. The open circles denote the current state φ^t and its unknown future values φ^{t+1} and φ^{t+k} while the dashed lines are a set of n observed trajectories in the neighborhood of the current state. Conventional cross-validation consists in assessing the accuracy of the local predictor by repeating n times the following procedure: the i^{th} trajectory is removed, a one-step-ahead prediction of the removed trajectory is computed with the observations of the i^{th} trajectory set aside and finally the leave-one-out error is obtained by comparing the prediction with the real value. The iterated PRESS generalizes this procedure by returning the leave-one-out error along the i^{th} trajectory in the case of a prediction iterated for $k > 1$ steps. In the next section we will show how the analytic expression of the iterated statistic can be derived from the one-step-ahead leave-one-out PRESS formula.

3.1 THE CONVENTIONAL AND ITERATED PRESS STATISTIC

For a more concise formulation and a simpler graphical illustration, some simplifications are required. First, we present the case of a local model where the neighbors are weighted according to a rectangular kernel. Second, we discuss only the one-dimensional case (i.e. model in Eq. (1) with $m = 1$: $\varphi^{t+1} = f(\varphi^t)$), leaving the general case in the appendix. Furthermore, given a generic regressor r, the notation in bold characters $\boldsymbol{r}$ stands for a vector where a constant value 1 is appended to the value r in order to consider the offset term in the regression. Denote the values of the series at time t and $t+1$ with the variables $x = \varphi^t$ and $y = \varphi^{t+1}$, respectively. Suppose that the value of the series at time t is $\varphi^t = \bar{x}$, that a set of n neighbors $\{x_i\}_{i=1}^n$ of the current state is selected and that a linear regression is used to model locally the data. Let $e_{xy}(i)$ $i = 1 \ldots n$ be the residual of the linear model and $e_{xy}^{cv}(i)$ the leave-one-out residual

$$e_{xy}^{cv}(i) = y_i - \hat{y}_{-i} = y_i - \boldsymbol{x}_i^T \beta_{xy}^{-i} \tag{2}$$

that is the difference between the real value of the mapping and the prediction $\hat{y}_{-i}$ in x_i of the linear model estimated with the i^{th} point aside (see Fig. 2).

Results in linear statistical analysis say that a simple relation (PRESS statistic) exists between $e_{xy}(i)$ and $e_{xy}^{cv}(i)$ that is

$$e_{xy}^{cv}(i) = \frac{e_{xy}(i)}{1 - h_{ii}} \tag{3}$$

where h_{ii} is the i^{th} diagonal element of the *Hat* matrix (see Appendix). Once the vector $e_{xy}^{cv}(i)$ is available,

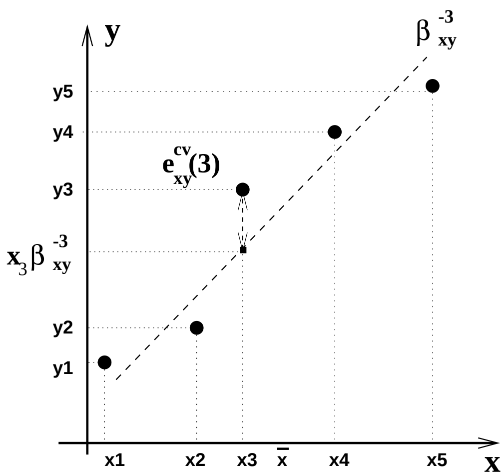

Figure 2: Conventional Press in the input-output space $X \rightarrow Y$;

it is possible to compute a leave-one-out measure (e.g. the mean squared error) of the performance of the local model for a single step prediction. Unfortunately, this index gives no hints about the expected performance of this model when used for iterated prediction. In fact, it is based on the representation of the time series as an input-output static mapping and it cannot take into account the behavior of the model as a dynamic process. To this aim it is necessary to compute a k-step-ahead criterion in order to measure the quality of the prediction on a longer horizon.

Here, we present an iterated formulation of the PRESS statistic for the horizon $k = 2$. Notice that an iterated prediction from time t to time $t + 2$ is the composition of two predictions: $\hat{\varphi}^{t+1} = \hat{f}_t(\varphi^t)$ and $\hat{\varphi}^{t+2} = \hat{f}_{t+1}(\hat{\varphi}^{t+1})$ with $\hat{f}_t$ local model of the embedding (1) at time t.

Denote the values of the series at time t, $t+1$ and $t+2$ with the variables $x = \varphi^t$, $y = \varphi^{t+1}$ and $z = \varphi^{t+2}$, respectively. Hence, the prediction problem is analogous to the estimation of a composed mapping $z = g(h(x))$ whose components are $z = g(y)$ and $y = h(x)$ (see Fig. 3). Let β_{yz}^{-i} be the vector of the regression parameters estimated on the set $\{(y_i, z_i)\}_{i=1}^{n}$ with the i-th sample aside and let $e_{yz}^{cv}(i)$ be the leave-one-out residual of the mapping g. We will define the *iterated cross-validation residual* for $k = 2$ as

$$e_{xz}^{it}(i) = z_i - (\hat{\boldsymbol{y}}_{-i})^T \beta_{yz}^{-i} \tag{4}$$

that is, the difference between the $k = 2$ steps ahead real value and the iteration of two one-step leave-one-out predictions.

We illustrate the idea in Fig. 3. Here we have $n = 5$ triples $\{(x_i, y_i, z_i)\}_{i=1}^{5}$ and we want to estimate the iterated cross-validation residual in $x_i = x_3$.

According to Eq. (2), $e_{xy}^{cv}(3)$ is the difference between y_3 and $\boldsymbol{x}_3^T \beta_{xy}^{-3}$, and $e_{yz}^{cv}(y_3)$ is the difference between z_3 and $(\hat{\boldsymbol{y}}_{-3})^T \beta_{yz}^{-3}$. From (2) and (4), the iterated cross-validation residual $e_{xz}^{it}(3)$ is the difference between z_3 and the regression value $(\hat{\boldsymbol{y}}_{-3})^T \beta_{yz}^{-3}$ where $\hat{y}_{-3} = y_3 - e_{xy}^{cv}(x_3)$ obtained by shifting y_3 by the error $e_{xz}^{cv}(3)$ incurred at the previous step. The formula of the iterated PRESS for a generic dimension m and a generic horizon k is derived in the Appendix.

The iterated cross-validation error returns an estimate of the performance of an iterated one-step estimator from time t to time $t + k$. As far as long horizon predictions are required, this statistic returns richer information than the simple one-step PRESS statistic and allows more reliable local model selection for iterated prediction.

4 EXPERIMENTS AND FINAL CONSIDERATIONS

The iterated PRESS approach has been applied both to the prediction of a *real-world* data set (A) and to a computer generated time series (D) from the *Santa Fe Time Series Prediction and Analysis Competition*. The A time series has a training set of 1000 values and a test set of 10000 samples: the task is to predict the continuation for 100 steps, starting from different points. The D time series has a training set of 100000 values and a test set of 500 samples: the task is to predict the continuation for 25 steps, starting from different points.

We adopt a local learning iterated prediction method where the selection of neighbors is made according to the iterated PRESS. The horizon of the iterated criterion is $h = 5$ for the series A and $h = 25$ for the series D. The number of neighbors is limited to range from 4 to 12 for both series. We adopt for the series A an embedding model having the same dimension $m = 16$ proposed in (Sauer, 1994) and for the series D an embedding model with $m = 20$ as reported in (Zhang & Hutchinson, 1994). Each prediction of the local model, inclusive of the modeling phase, takes about one second of computation on a Pentium machine.

Table 1 compares the NMS (Normalized Mean Squared) prediction errors on the A test set of the local

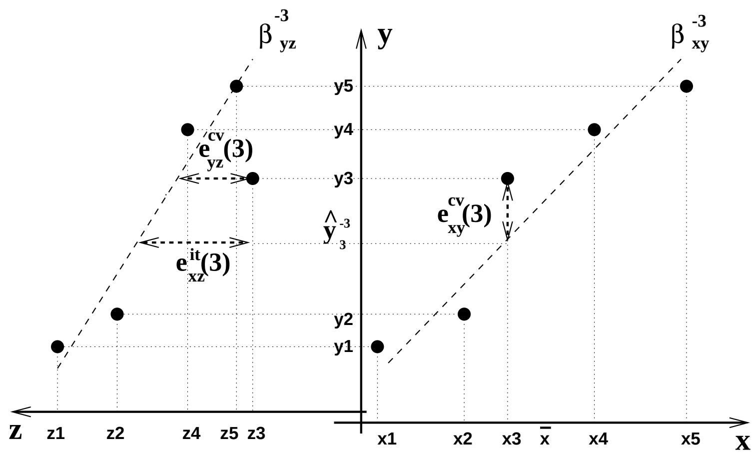

Figure 3: Iterated Press in the composed space $X \to Y \to Z$

Table 1: NMSE of the predictions for time series A

Test data	Non iter. PRESS	Iter. PRESS	Sauer	Wan
1-100	0.350	0.029	0.077	0.055
1180-1280	0.379	0.131	0.174	0.065
2870-2970	0.793	0.055	0.183	0.487
3000-3100	0.003	0.003	0.006	0.023
4180-4280	1.134	0.051	0.111	0.160

Table 2: RMSE of the predictions for time series D

Test data	Non iter. PRESS	Iter. PRESS	Zhang Hutchinson
0-24	0.1255	0.0492	0.0665
100-124	0.0460	0.0363	0.0616
200-224	0.2635	0.1692	0.1475
300-324	0.0461	0.0405	0.0541
400-424	0.1610	0.0644	0.0720

predictor (iterated and non iterated) with the performance statistics reported by Sauer (Sauer, 1994) and Wan (Wan, 1994). Sauer used a combination of iterated and direct local linear models with a fixed number of four neighbors on a data set obtained by interpolating the original one. Wan used a recurrent network (FIR-network) with one input unit, two layers of 12 hidden units each, and one output unit. The architecture was trained with temporal backpropagation, and according to Wan, a training run typically took over night on a Sun Sparc2.

Table 2 compares the local RMS (root mean squared) errors on the series D with the results obtained by (Zhang & Hutchinson, 1994). Zhang and Hutchinson adopted a combination of multilayer perceptrons for iterated and direct predictions which were trained on a Connection Machine for a total period of about 100 hours.

The iterated approach has been applied also to the prediction of a time series proposed in the competition of the *International Workshop on Advanced Black-box techniques for nonlinear modeling* (Leuven, Belgium; 1998). The data set[1] consists of 2000 values and the task is to predict the continuation for the next 200 steps (see Fig. 4). Two predictions submitted by the authors with different embedding orders (m = 20 and m = 24) ranked second and fourth, respectively (Suykens & Vandewalle, 1998).

[1]see http://www.esat.kuleuven.ac.be/sista/workshop/

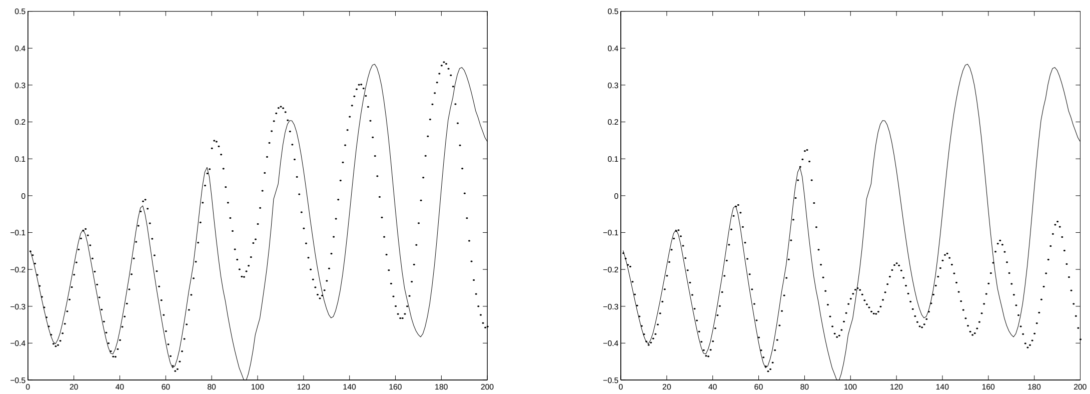

Figure 4: Predictions performed by the iterated PRESS predictor for the Leuven time series competition (full line: to be predicted, dotted: prediction): a) embedding order m=20 ($NMSE = 0.0475$) b) m =24 ($NMSE = 0.0667$)

The experiments show that for long horizons the iterated PRESS can improve the performance of the local approach based on the conventional leave-one-out cross-validation. At the same time it emerges that in iterated tasks the local approach is able to outperform complex nonlinear architectures at a reduced computational cost.

Appendix

The PRESS leave-one-out formula for a linear regression model with X input matrix $[n, m+1]$ and y output vector $[n, 1]$ is derived in (Myers, 1994) as

$$\begin{aligned}
e^{cv}_{xy}(i) &= y_i - \mathbf{x}_i^T \beta^{-i}_{xy} = \\
&= y_i - \mathbf{x}_i^T \big[P + \frac{P\mathbf{x_i}\mathbf{x}_i^T P}{1-h_{ii}}\big] X^T_{-i} y_{-i} = \\
&= y_i - \mathbf{x}_i^T P X^T_{-i} y_{-i} - \frac{h_{ii}\mathbf{x}_i^T P X^T_{-i} y_{-i}}{1-h_{ii}} = \\
&= \frac{(1-h_{ii})y_i - \mathbf{x}_i^T P X^T_{-i} y_{-i}}{1-h_{ii}} = \qquad (5)\\
&= \frac{(1-h_{ii})y_i - \mathbf{x}_i^T P (X^T y - \mathbf{x}_i y_i)}{1-h_{ii}} = \\
&= \frac{(1-h_{ii})y_i - \hat{y}_i + h_{ii} y_i}{1-h_{ii}} = \frac{e_{xy}(i)}{1-h_{ii}}
\end{aligned}$$

where $P = (X^T X)^{-1}$, h_{ii} is the diagonal element of the Hat matrix $H = XPX^T$, $\mathbf{x}_i^T$ is the i-t row of X, $\hat{y}_i = \mathbf{x}_i^T P X^T y$ is the prediction of the linear regression, $e_{xy}(i) = \hat{y}_i - y_i$ is the residual and β^{-i}_{xy} denotes the vector $[m+1, 1]$ of least-squares coefficients computed with the i^{th} data point aside. Note that the derivation of the PRESS formula in (5) makes use of the following equivalence $X^T_{-i} y_{-i} + \mathbf{x}_i y_i = X^T y$.

We derive now the iterated PRESS formula for a problem of prediction from time t to time $t+k$ having m as embedding order . Let $\bar{\mathbf{x}}$ be the embedding vector at time t and $\mathbf{x}_i$ a neighbor of $\bar{\mathbf{x}}$. Consider the trajectory in the m dimensional state space passing through $\mathbf{x}_i$: let $\mathbf{y}_i$ the embedding vector $k-1$ steps ahead with respect to x_i and z_i the value of the series k steps later than $\mathbf{x}_i$. The k-iterated leave-one-out residual at time t in the i^{th} point is defined as the value of the leave-one-out regression β^{-i}_{yz} computed in the point $\mathbf{y}_i$ previously shifted of the vector $\boldsymbol{\delta}_i$ of errors cumulated in the $k-1$ previous steps. Then the iterated PRESS formula is derived as follows:

$$\begin{aligned}
&e^{it}_{xz}(i) = z_i - (\mathbf{y}_i - \boldsymbol{\delta}_i)^T \beta^{-i}_{yz} = \\
&= z_i - \mathbf{y}_i^T \big[P + \frac{P\mathbf{y}_i\mathbf{y}_i^T P}{1-h_{ii}}\big] Y^T_{-i} z_{-i} + \\
&+ \boldsymbol{\delta}_i^T \big[P + \frac{P\mathbf{y}_i\mathbf{y}_i^T P}{1-h_{ii}}\big] Y^T_{-i} z_{-i} = \\
&= \frac{e_{yz}(i) + \boldsymbol{\delta}_i^T \big[(1-h_{ii}) + P\mathbf{y}_i\mathbf{y}_i^T\big] P Y^T_{-i} z_{-i}}{1-h_{ii}} = \\
&= \frac{e_{yz}(i) + \boldsymbol{\delta}_i^T \big[(1-h_{ii}) + P\mathbf{y}_i\mathbf{y}_i^T\big] P (Y^T z - \mathbf{y}_i z_i)}{1-h_{ii}} = \\
&= \frac{e_{yz}(i) + \boldsymbol{\delta}_i^T \big[(1-h_{ii}) P Y^T z\big]}{1-h_{ii}} + \\
&+ \frac{\boldsymbol{\delta}_i^T \big[-(1-h_{ii}) P\mathbf{y}_i z_i + P\mathbf{y}_i\mathbf{y}_i^T P Y^T z - P\mathbf{y}_i\mathbf{y}_i^T P\mathbf{y}_i z_i\big]}{1-h_{ii}} = \\
&= \frac{e_{yz}(i) + \boldsymbol{\delta}_i^T \big[(1-h_{ii}) P Y^T z - P\mathbf{y}_i z_i + P\mathbf{y}_i \hat{z}_i\big]}{1-h_{ii}} = \\
&= \frac{e_{yz}(i) + \boldsymbol{\delta}_i^T \big[(1-h_{ii})\beta_{yz} - P\mathbf{y}_i e_{yz}(i)\big]}{1-h_{ii}}
\end{aligned}$$

where we used the following equivalences: $P = (Y^T Y)^{-1}$, $Y^T_{-i} z_{-i} + \mathbf{y}_i z_i = Y^T z$, $\mathbf{y}_i^T P Y^T z = \hat{z}_i$, $e_{yz}(i) = z_i - \hat{z}_i$, $\mathbf{y}_i^T P \mathbf{y}_i z_i = h_{ii}$ and $\beta_{yz} = PY^T z$.

Acknowledgements

The work of Gianluca Bontempi was supported by the European Union TMR Grant FMBICT960692. The work of Mauro Birattari was supported by the FIRST program of the Région Wallonne, Belgium.

References

Aha D. W. 1997. Editorial. *Artificial Intelligence Review*, **11**(1–5), 1–6.

Atkeson C. G. , Moore A. W. & Schaal S. 1997. Locally weighted learning. *Artificial Intelligence Review*, **11**(1–5), 11–73.

Bersini H. , Birattari M. & Bontempi G. . 1998. Adaptive memory-based regression methods. *Pages 2102–2106 of: Proceedings of the 1998 IEEE International Joint Conference on Neural Networks*.

Birattari M. , Bontempi G. & Bersini H. 1999. Lazy learning meets the recursive least-squares algorithm. *In:* Kearns M. S. , Solla S. A. & Cohn D. A. (eds), *Advances in Neural Information Processing Systems 11*. Cambridge: MIT Press.

Farmer J. D. & Sidorowich J. J. 1987. Predicting chaotic time series. *Physical Review Letters*, **8**(59), 845–848.

Myers R. H. 1994. *Classical and Modern Regression with Applications*. second edition edn. Boston, MA: PWS-KENT Publishing Company.

Packard N. H. , Crutchfeld J. P. , Farmer J. D. & Shaw R. S. 1980. Geometry from a time series. *Physical Review Letters*, **45**(9), 712–716.

Rumelhart D. E. , Hinton G. E. & Williams R. K. 1986. Learning representations by backpropagating errors. *Nature*, **323**(9), 533–536.

Sauer T. 1994. Time series prediction by using delay coordinate embedding. *Pages 175–193 of:* Weigend A. S. & Gershenfeld N. A. (eds), *Time Series Prediction: forecasting the future and understanding the past*. Harlow, UK: Addison Wesley.

Suykens J. A. K. & Vandewalle J. (eds) 1998. *Nonlinear Modeling: Advanced Black-Box Techniques*. Kluwer Academic Publishers. Chap. The K. U. Leuven Time Series Prediction Competition, pages 241–251.

Wan E. A. 1994. Time series prediction using a connectionist network with internal delay lines. *Pages 195–217 of:* Weigend A. S. & Gershenfeld N. A. (eds), *Time Series Prediction: forecasting the future and understanding the past*. Harlow, UK: Addison Wesley.

Weigend A. S. 1996. Time Series Analysis and Prediction. *Chap. 12, pages 395–449 of:* P. Smolensky M. C. Mozer & Rumelhart D. E. (eds), *Mathematical Perspectives on Neural Networks*. Lawrence Erlbaum Associates.

Williams R. & Zipser D. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, **1**, 270–280.

Yakowitz S. 1987. Nearest-neighbour methods for time series analysis. *Journal of Time Series Analysis*, **8**(2).

Zhang X. & Hutchinson J. 1994. Simple architectures on fast machines: practical issues in nonlinear time series prediction. *Pages 219–241 of:* Weigend A. S. & Gershenfeld N. A. (eds), *Time Series Prediction: forecasting the future and understanding the past*. Addison Wesley.