LUND UNIVERSITY

Multi-step-ahead Multivariate Predictors: A Comparative Analysis

Cescon, Marzia; Johansson, Rolf

Published in:
Proc. 49th IEEE Conf. Decision and Control (CDC2010)

2010

Link to publication

Citation for published version (APA):
Cescon, M., & Johansson, R. (2010). Multi-step-ahead Multivariate Predictors: A Comparative Analysis. In *Proc. 49th IEEE Conf. Decision and Control (CDC2010)* (pp. 2837-2842)

Total number of authors:
2

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

Multi-step-ahead Multivariate Predictors: a Comparative Analysis

Marzia Cescon and Rolf Johansson

***Abstract*— The focus of this article is to undertake a comparative analysis of multi-step-ahead linear multivariate predictors. The approach considered for the estimation will be based on geometrically reliable linear algebra tools, resorting to subspace identification methods. A crucial issue is quantification of both bias error and variance affecting the estimate of the prediction for increasing values of the look ahead when only a small number of samples is available. No complete theory is available so far, nor sufficient numerical experience. Therefore, the analysis of this paper aims at shading some lights on the topic providing some insights and help to develop some intuitions.**

I. INTRODUCTION

Predictor estimation is a significant problem in many practical situations, from economics to social sciences, to medicine. Traditionally, the design of an optimal linear predictor for discrete-time systems is based on a model, either from first principles or identified from data relying on Prediction Error Methods (PEM) [1], [2]. In real application such a model may not be attainable or the underlying dynamics too complicated to be modeled by differential equations. Furthermore, there may be the need of estimating the output on a longer prediction horizon than just one sample look ahead.

The identification of multi-step-ahead optimal predictors of certain observed, i.e., available to measurements, variables has been analyzed in the past by several authors (e.g. [3], [4], [5], [6], [7]). However, results are still partial and not conclusive.

Motivated by the above, in this paper we will consider the problem of estimating data-driven multi-step-ahead linear multivariate predictors, without any prior knowledge of the underlying mechanism generating the data sequences and accounting for multiple-input multiple-output (MIMO) systems. The core of our approach resorts to the so-called state space identification methods (e.g. [8], [9], [10]) and it is based on projection operations of certain structured data matrices onto suitable subspaces spanned by the data. The key interpretation lies on the first step of such methods, namely, the construction of the state space. It is, indeed, the procedure utilized to estimate a basis for the state space that contains all the basic objects we need to our purposes ([1, Appendix 4.A]).

When evaluating the capability of predicting new data a crucial issue is quantification of both bias error and variance affecting the estimates for increasing values of the look ahead when the number of available samples is small, i.e., for a finite observation interval. In particular, it is of interest to see what happens in the presence of feedback and when data are noisy. Thus far, no complete theory is available nor sufficient numerical experience. Therefore we believe this state of the art is worth investigating.

Two illustrative examples are provided targeting a performance-based evaluation of such predictors: the first one is a simulation of a second-oder artificial model, whereas the second is the case of a real Type 1 Diabetes Mellitus patient.

The paper is organized as follows. Section II introduces the notation and provides the background for the sequel. In Sec. III the derivation of the multi-step-ahead multivariate predictors is carried out. Some results of computer simulations are presented in Sec. IV. The discussion on the achievements is presented in Sec. V. Finally, Sec. VI concludes the paper.

II. PRELIMINARIES

Consider a discrete-time time-invariant system $\mathcal{S}_n(A, B, C, D, K)$ in innovation form

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + Ke_k \\ y_k = Cx_k + Du_k + e_k \end{cases} \tag{1}$$

with input $u_k \in \mathbb{R}^m$, output $y_k \in \mathbb{R}^l$, state vector $x_k \in \mathbb{R}^n$ and zero-mean white noise innovation process, i.e., one-step ahead prediction error, $e_k \in \mathbb{R}^l$, k representing the current time instant in the identification problem. Let us define $\bar{A} = A - KC$ for future reference. Without loss of generality we assume that the system is minimal in the sense that it cannot be described by a state-space model of order less than n. We let the input signal u_k influence the output y_{k+1} but not earlier outputs, i.e., $D = 0$. Furthermore, we assume that the joint input-output process denoted by $z_k = [u_k^T y_k^T]$ is purely non-deterministic and has spectral density matrix $S_{zz}(e^{j\omega})$ bounded and bounded away from zero on the unit circle [11], [12], [13] in order to guarantee persistently exciting (PE) condition of sufficiently high order and let the projection operations in the sequel be well defined. The available data sequences u_k, y_k and the innovation process e_k will be organized in Hankel matrices and will be denoted by capital letters. Subscript indices $[i, j]$ of a matrix will be used to indicate the argument of the upper-left and the lower-left element, respectively, with the meaning of the interval of time considered, e.g., $U_{[t_1, t_1+t_2]}$ will contain in the first column the inputs from time t_1 to time $t_1 + t_2$. Accordingly, process tails of finite length N will be represented by the block rows of the block Hankel data matrices, e.g.,

$$U_k := [u_k \quad u_{k+1} \quad \cdots \quad u_{k+N-1}]$$

$$Y_k := [y_k \quad y_{k+1} \quad \cdots \quad y_{k+N-1}]$$

Dept. Automatic Control, Lund University, SE-221 00 Lund, Sweden; Email Marzia.Cescon@control.lth.se, Rolf.Johansson@control.lth.se

In the sequel, the orthogonal projection of the rows of a given matrix A onto the row space of a given matrix B will be denoted by $\hat{E}\{A \mid B\}$, whereas the symbol $\hat{E}_{\|C}\{A \mid B\}$ will denote the oblique projection of the row space of A onto the row space of B along the row space of C, the projection operator being $\hat{E}\{\cdot\}$. Past and future horizon in the identification problem which represent the number of block rows in the processed data Hankel matrices will be p and f, respectively, with $p \geq f$. Finally, the number of steps in the look ahead will be denoted by τ.

III. MULTI-STEP-AHEAD MULTIVARIATE PREDICTOR ESTIMATION

A. Construction of the predictors

Let a finite input sequence $\{u_k\}_{k=t-p}^{t+1+f}$ and a corresponding output sequence $\{y_k\}_{k=t-p}^{t+1+f}$ be generated by the system (1). Consider the following block Hankel matrices:

$$\begin{aligned} U^p : &= U_{[t-p,t]} \in \mathbb{R}^{p\cdot m\times N} \qquad (2)\\ &= \begin{bmatrix} u_{t-p} & u_{t-p+1} & \cdots & u_{t-p+N} \\ u_{t-p+1} & u_{t-p+2} & \cdots & u_{t-p+1+N} \\ \vdots & \vdots & \cdots & \vdots \\ u_t & u_{t+1} & \cdots & u_{t+N} \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} U^f : &= U_{[t+1,t+f+1]} \in \mathbb{R}^{f\cdot m\times N} \qquad (3)\\ &= \begin{bmatrix} u_{t+1} & u_{t+2} & \cdots & u_{t+1+N} \\ u_{t+2} & u_{t+3} & \cdots & u_{t+2+N} \\ \vdots & \vdots & \cdots & \vdots \\ u_{t+1+f} & u_{t+2+f} & \cdots & u_{t+1+f+N} \end{bmatrix} \end{aligned}$$

called the past and future input data matrices, respectively. Similarly, consider

$$Y^p := Y_{[t-p,t]} \in \mathbb{R}^{p\cdot l\times N} \qquad (4)$$

and

$$Y^f := Y_{[t+1,t+f+1]} \in \mathbb{R}^{f\cdot l\times N} \qquad (5)$$

past and future output data matrices, respectively. The integer p is chosen so to satisfy the following condition:

$$p \geq \max\{n, \tau\} \qquad (6)$$

in order for the system (1) to be observable and to guarantee predictions up to the largest future horizon we wish to investigate. Introduce the unknown, i.e., not yet observed, future output tails:

$$Y_j, \qquad t+f+2 < j \leq M \qquad (7)$$

Our aim is to predict such future tails using linear combination of the rows of (2), (3), (4), (5). We remind that, although the names may lead to misunderstandings, future input-output matrices contain data already measured hence belonging to the actual past.

As first step, we will look for an estimator of the known future tails:

$$Y_i, \qquad t+1 < i \leq t+1+f \qquad (8)$$

Mathematically, this can be formulated as the following least-squares problem:

$$\hat{\Gamma}, \hat{\Lambda} = \underset{\substack{\Gamma \in \mathbb{R}^{lp\times(l+m)p} \\ \Lambda \in \mathbb{R}^{lp\times mf}}}{\operatorname{argmin}} \; || \, Y_i - \begin{bmatrix}\Gamma & \Lambda\end{bmatrix} \begin{bmatrix} Z^p & U^f \end{bmatrix}^T \, ||_F^2 \qquad (9)$$

where $||\cdot||_F$ stands for the Frobenius norm of a matrix.

Geometrically, it can be interpreted as the orthogonal projection of Y_i onto $\begin{bmatrix} Z^p & U^f \end{bmatrix}^T$, i.e.

$$\hat{Y}_i = \hat{E}[Y_i \mid \begin{bmatrix} Z^p & U^f \end{bmatrix}^T] \qquad (10)$$

As a matter of fact, the orthogonal projection (10) corresponds to the sum of two oblique projections ([12], Lemma 1):

$$\hat{Y}_i = \hat{E}_{\|U^f}[Y_i \mid Z^p] + \hat{E}_{\|Z^p}[Y_i \mid U^f] \qquad (11)$$

Now, taking conditional expectation up to time t to eliminate the effect of yet unknown inputs, the output predictor for each vector Y_i as linear combination of past input and output amounts to computing an oblique projection, i.e.,

$$\begin{aligned} \hat{Y}_{i|t} &= \hat{E}_{\|U^f}[Y_i \mid Z^p] \qquad (12)\\ &\simeq \mathcal{O}_i X_i \qquad (13) \end{aligned}$$

where $\mathcal{O}_i$ is the extended observability matrix and X_i is the state sequence. Stacking all the predictors on top of each other, the sought matrix of multi-step ahead predictors is obtained:

$$\hat{Y}^f = \begin{bmatrix} \hat{Y}_{t+1|t} \\ \hat{Y}_{t+2|t} \\ \vdots \\ \hat{Y}_{t+f+1|t} \end{bmatrix} \qquad (14)$$

Subsection III-B will deal with the computational details, suffices it here to mention that $\hat{Y}^f$ will be calculated as follows:

$$\hat{Y}^f = \hat{\Gamma}_1 Z^p \qquad (15)$$

In the sequel the approach illustrated above applying (10) and (11) will be referred to as projection-based.

Let us now consider the innovation model (1) in predictor form:

$$\begin{cases} x_{k+1} = \bar{A}x_k + Bu_k + Ky_k \\ \hat{y}_{k|k-1} = Cx_k \end{cases} \qquad (16)$$

By iteration of (16) it is possible to express the output tails at time i, $t+1 \leq i \leq t+f-1$ making use of the *true* Markov parameters:

$$Y_i = \bar{\mathcal{O}}_i \bar{A} X_p + \Xi_i Z^p + \bar{\Psi}_i Z^f + E_i \qquad (17)$$

where the first term depends on the initial conditions of the state, the second term depends upon past input-output data and the third on future input-output data. Matrices $\bar{\mathcal{O}}_i$, $\bar{\Psi}_i$ and Ξ_i are given in (20) and (21). Solving a least-squares problem equivalent to that of the estimation of the long *VARX* model:

$$Y_i = \Xi_0 Z^p + E_i \qquad (18)$$

$$\bar{\mathcal{O}}_i = \begin{bmatrix} C \\ C\bar{A} \\ C\bar{A}^2 \\ \vdots \\ C\bar{A}^{i-1} \end{bmatrix}, \Psi_i = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ C[B \quad K] & 0 & \cdots & \cdots & \cdots & 0 \\ C\bar{A}[B \quad K] & C[B \quad K] & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ C\bar{A}^{i-2}[B \quad K] & \cdots & \cdots & \cdots & C[B \quad K] & 0 \end{bmatrix}, \tag{20}$$

$$\Xi_i = \begin{bmatrix} C\bar{A}^{p-1}[B \quad K] & C\bar{A}^{p-2}[B \quad K] & \cdots & \cdots & \cdots & C[B \quad K] \\ 0 & C\bar{A}^{p-1}[B \quad K] & \cdots & \cdots & \cdots & C\bar{A}[B \quad K] \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C\bar{A}^{p-1}[B \quad K] & \cdots & C\bar{A}^{i-1}[B \quad K] \end{bmatrix} \tag{21}$$

$$\hat{\Xi}_0 = \underset{\Xi_0}{\operatorname{argmin}} \; || \, Y_i - \Xi_0 Z^p \, ||_F^2 \tag{19}$$

where $|| \cdot ||_F$ stands for the Frobenius norm of a matrix, the Markov parameters of the system (1) are obtained.

Assuming that the effects of the unknown initial states in (19) vanishes for sufficiently large p and, as before, taking conditional expectation up to time t, the output predictors are expressed by:

$$\begin{bmatrix} \hat{Y}_{t+1|t} \\ \hat{Y}_{t+2|t} \\ \vdots \\ \hat{Y}_{t+f+1|t} \end{bmatrix} = \begin{bmatrix} \Gamma_{t+1} \\ \Gamma_{t+2} \\ \vdots \\ \Gamma_{t+f+1} \end{bmatrix} Z^p \tag{22}$$

with

$$\Gamma_i = \hat{\Xi}_i + \sum_{j=0}^{i-1} \hat{C}\hat{\bar{A}}^{i-j-1}\hat{K}\Gamma_j^P \qquad t+1 \leq i \leq t+f+1 \tag{23}$$

obtained from the estimated $\hat{\Xi}_0$. In a more compact notation:

$$\hat{Y}^f = \hat{\Gamma}_P Z^p \tag{24}$$

This approach will be referred to as PBSID-based.

Once the operator $\hat{\Gamma}_1$ and $\hat{\Gamma}_P$ have been estimated, they can be applied to new, still generated by the same underlying mechanisms, data to forecast the actual future.

B. Numerical Implementation: computing projections

As we have seen in Subsec. III-A, as far as the projection-based approach is concerned, the calculation of multi-step-ahead predictors from given input-output data corresponds to the computation the oblique projection of future outputs Y^f onto the joint input-output past Z^p along the future inputs U^f (12). The starting point is the linear regression type of problem (9) which may be numerically implemented by means of an LQ decomposition [10], [8]. In this section we will present two such implementation methods.

Consider the LQ decomposition:

$$\begin{bmatrix} Z^p \\ U^f \\ Y^f \end{bmatrix} = \begin{bmatrix} \mathcal{L}_{11} & 0 & 0 \\ \mathcal{L}_{21} & \mathcal{L}_{22} & 0 \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix} \begin{bmatrix} \mathcal{Q}_1{}^T \\ \mathcal{Q}_2{}^T \\ \mathcal{Q}_3{}^T \end{bmatrix} \tag{25}$$

Define

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{31} & \mathcal{L}_{32} \end{bmatrix} \begin{bmatrix} \mathcal{L}_{11} & 0 \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix}^\dagger = \begin{bmatrix} \hat{\Gamma}_1 & \hat{\Lambda}_1 \end{bmatrix} \tag{26}$$

where $\hat{\Gamma}_1 = \mathcal{L}_{31}\mathcal{L}_{11}{}^\dagger$ and $\hat{\Lambda}_1 = (\mathcal{L}_{32} - \mathcal{L}_{21}\mathcal{L}_{11}{}^\dagger\mathcal{L}_{31})\mathcal{L}_{22}{}^{-1}$. According to subsection III-A, the future output can, thus, be expressed in terms of the joint input-output past Z^p in the following way:

$$\hat{Y}^f = \hat{\Gamma}_1 Z^p \tag{27}$$

Alternatively, organizing the data differently:

$$\begin{bmatrix} U^f \\ Z^p \\ Y^f \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1{}^T \\ Q_2{}^T \\ Q_3{}^T \end{bmatrix} \tag{28}$$

Define

$$L = \begin{bmatrix} L_{32} & L_{31} \end{bmatrix} \begin{bmatrix} L_{22} & 0 \\ L_{21} & L_{11} \end{bmatrix}^\dagger = \begin{bmatrix} \hat{\Gamma}_2 & \hat{\Lambda}_2 \end{bmatrix} \tag{29}$$

where $\hat{\Gamma}_2 = L_{32}L_{21}{}^\dagger$ and $\hat{\Lambda}_2 = (L_{31} - L_{32}L_{22}{}^\dagger L_{21})L_{11}{}^{-1}$. Then

$$\hat{Y}^f = \hat{\Gamma}_2 Z^p \tag{30}$$

Last implementation issue is concerned with the least-squares problem (19). Similarly to (9) we perform an LQ-decomposition

$$\begin{bmatrix} Z^p \\ Y_i \end{bmatrix} = \begin{bmatrix} \mathscr{L}_{11} & 0 \\ \mathscr{L}_{21} & \mathscr{L}_{22} \end{bmatrix} \begin{bmatrix} \mathscr{Q}_1{}^T \\ \mathscr{Q}_2{}^T \end{bmatrix} \tag{31}$$

from which the estimate $\hat{\Xi}_0$ can be computed as

$$\hat{\Xi}_0 = \mathscr{L}_{21}\mathscr{L}_{11}^{-1} \tag{32}$$

IV. EXAMPLES

Two algorithms were used for simulation:

- *proj-1*: compute the predictions according to (27)
- *pbsid*: compute the predictions according to (24)

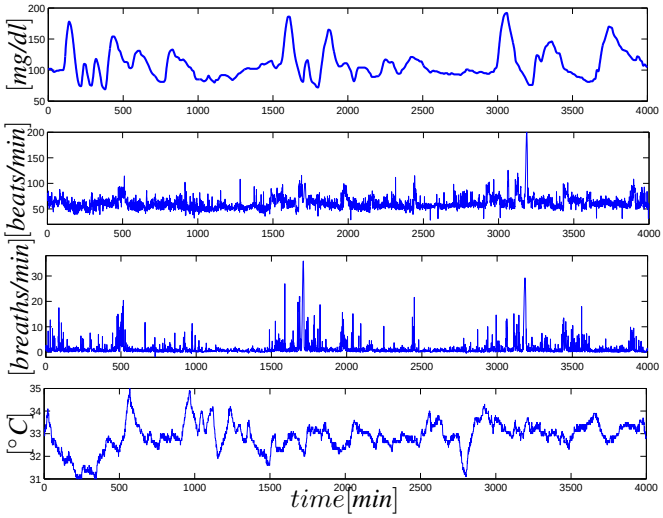

Fig. 2. Example 2. Patient data records. *Top* Continuous Glucose Monitoring Sensor data [*mg/dl*], *Upper Center* Heart Rate [beats/min], *Lower Center* Respiration [breaths/min], *Bottom* Skin Temperature [°C] vs. Time [*min*]

A. Example 1: Artificial Data

We tested the algorithms on two different setups: first, when the system is operating in open loop and second, when the system is operating under negative unity feedback. Inputs to the system are either zero-mean unit-variance white noise, or zero-mean unit-variance white noise passed through the filter

$$H(z) = \frac{1 + 0.8z^{-1} + 0.55z^{-2}}{1 - 0.5z^{-1} + 0.9z^{-2}}$$

for the first case and $u_k = r_k - y_k$ for the second case, where the reference signal r_k is white noise with unit variance. The system considered is described by

$$y_k = \frac{0.21z^{-1} + 0.07z^{-2}}{1 - 0.6z^{-1} + 0.8z^{-2}} u_k + \frac{1}{1 - 0.5z^{-1}} w_k$$

where w_k is Gaussian white noise with mean 0.5 and variance 2. We report results obtained over 100 Monte Carlo simulations, the number of input-output data points generated at each run being $N = 3000$. At any run, the estimates $\hat{\Gamma}_1$, $\hat{\Gamma}_P$ were obtained exploiting the first 1500 samples. Then, the predictions were computed by applying the above mentioned operators to the remaining data. Past and future horizons in the identification were set to 50 and 30, respectively. A 5-th order model was estimated with *N4SID* [16] and then used to derive predictors by means of the Matlab® System Identification Toolbox *compare.m* routine. Performances were assessed with respect to the average empirical prediction error variance. Figure 1 shows such variances on $\tau = 1, 10, 20, 30$ steps ahead for all of the three simulation setups.

B. Example 2: Type 1 Diabetes Mellitus Patient

A major challenge for a person with diabetes is to adapt insulin dosage regimens, food intake and exercise to keep blood glucose within tolerable limits during daily life activities. As a matter of fact the diabetic subject has usually inadequate understanding and overview of the actual physiological state at any time. The development of a reliable and accurate blood glucose predictor will provide the patients with invaluable informations for appropriate on-the-spot decision making concerning the management of the disease. The example shows how the predictors developed in Sec. III can be use to the purpose. We consider data records of a representative patient (Fig. 2) collected complying with a signed protocol over a 72-hours in-hospital trial which belongs to DIAdvisor, a major European research project [14]. For the duration of the whole study, the subject was equipped with an Abbott FreeStyle Navigator^{TM} Continuous Glucose Measurement Sensor (CGMS) [15] which provides estimation of glycemia levels from interstitial glucose measurements every 10 minutes. In addition, the Clinical LifeShirt® from VivoMetrics was used to gather several vital signs measured by sensors woven into the shirt around chest and abdomen. Standard meals were served for breakfast, lunch and dinner. Our aim was to estimate multi-step-ahead predictors of blood glucose levels from measurements of heart rate, respiration and skin temperature. We compared the different estimation strategies with respect to prediction capability on new data, i.e., not used for the construction of the prediction matrices. In particular we are interested in evaluating the performances on different look ahead τ, $10[min] \leq \tau \leq 30[min]$, with respect to the magnitude of the prediction errors.

The sample size was $N = 4000$ data points, of which the first 2000 were used for identification and the second 2000 for validation. Lacking prior knowledge of the underlying physiological system, the choice of the horizons in the identification step was made by a trial-and-error procedure. First, they were set to $p = f = 40$, second, to $p = 60, f = 30$, the lower bound on p being determined by the maximum value of τ. Furthermore, a state space model of order $n = 4$ was estimated from input-output data with *N4SID* [16] and then used to derive predictors by means of the Matlab® System Identification Toolbox *compare.m* routine, with the purpose of comparison.

Figure 3 reports the prediction error using validation data on a 10-min-ahead prediction horizon (left) and 40-min-ahead prediction horizon (right), with the horizons set to $p = f = 40$. Figure 4 illustrates the predictions on a 10-min-look ahead (top) and on a 30-min-look ahead (bottom), with the horizons set to $p = 60, f = 30$.

V. DISCUSSION

This paper dealt with a comparative analysis of subspace-based multi-step-ahead multivariate predictors. Predictor coefficients were directly obtained from input-output data, with no need of the determination of the model structure. However, parameter p representing the size of the past Hankel data matrices has strong connections with model order, and is given by the user from knowledge or intuitions on the system dynamics. Moreover, the lower bound on p is represented by the maximum prediction horizon τ one wish to investigate. Experience from simulation suggests that for

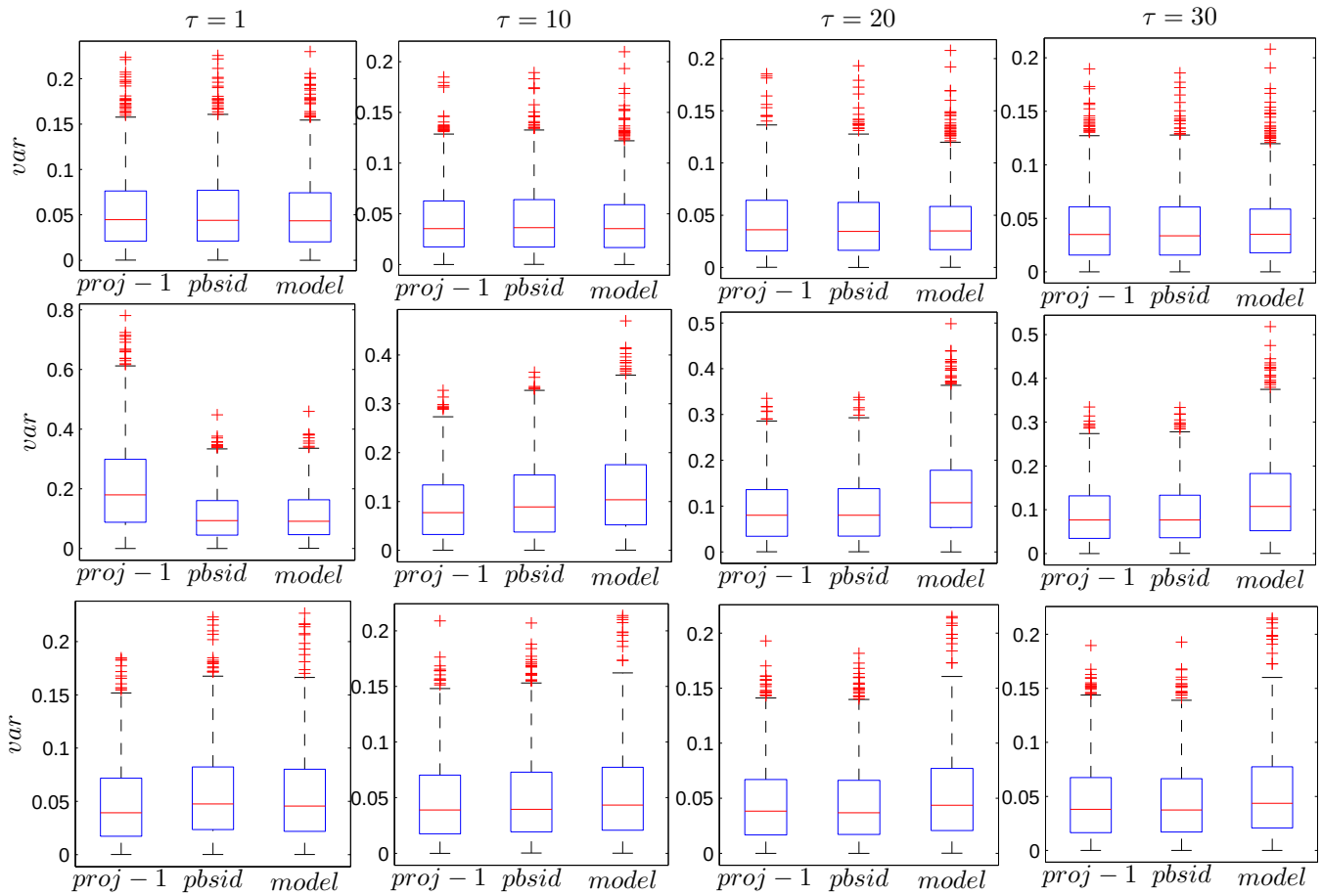

Fig. 1. Example 1. Prediction errors for performances on validation data: *Left* 1-step-ahead prediction, *Center-Left* 10-step-ahead prediction, *Center-Right* 20-step-ahead prediction, *Right* 30-step-ahead prediction. Horizons $p = 50$, $f = 30$. *Top panel* First case: system operating in open loop. Input is zero-mean unit-variance white noise, *Middle panel* First case: system operating in open loop. Input is colored noise, *Bottom panel* Second case: system operating in closed loop

prediction a small number of samples ahead it suffices to choose p double of the expected model order. As far as a longer prediction horizon is concerned, much care needs to be taken. As may be expected, the longer the past horizon, the better the performances. However, the investigation that we carried out suggests instead the contrary and even if apparently less crucial also the choice of the future horizon matters. Example 1 gave a system operating in open loop driven by a white input and a colored input, respectively. Simulations results suggest that the proposed algorithms compare favorably in the first case and are in general less efficient in the second case. A set of experiments carried out with a sinusoidal input showed, albeit unsatisfactory, still better prediction capabilities of the proposed methods compared to the classical model-based predictor, in particular they suggested superiority of the PBSID approach. Moreover, the feedback seems not to affect the behavior of the predictions, when the reference signal is white. The second example considered a challenging system, namely, the human glucoregulatory system. On the short look-ahead the prediction errors associated with the multi-step data-based predictors are within the same order of magnitude of the ones associated with the model-based approach, namely $2-3$ [*mg/dL*]. On the contrary, for bigger τ, *proj-1* breaks down exhibiting very poor performance, whereas *pbsid* proves to be remarkably robust. Besides being of interest on its own, the quest for an optimal multi-step-ahead predictor is of importance in many control applications, being used in the so called receding horizon type of control strategies [17], [18] and in the estimation of the extended observability matrix in subspace identification methods [19], [20].

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

In this contribution we have considered multi-step-ahead linear multivariate output predictors. No fixed model structure was postulated but predictor coefficients were directly estimated from input-output data. From an implementation point of view, the approach is attractive, amounting only to LQ decomposition of appropriately organized input-output Hankel matrices. All the algorithms investigated shared the advantageous properties that they may be applied to MIMO systems where very little is known.

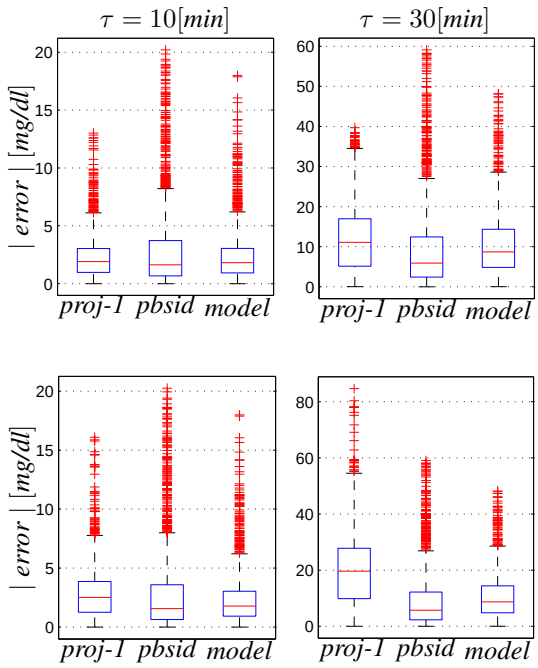

Fig. 3. Example 2. Absolute value of the prediction error. Performances on validation data. *Left* 10 min-ahead prediction, *Right* 30-min-ahead prediction, *Top* $p = f = 30$, *Bottom* $p = 60, f = 30$

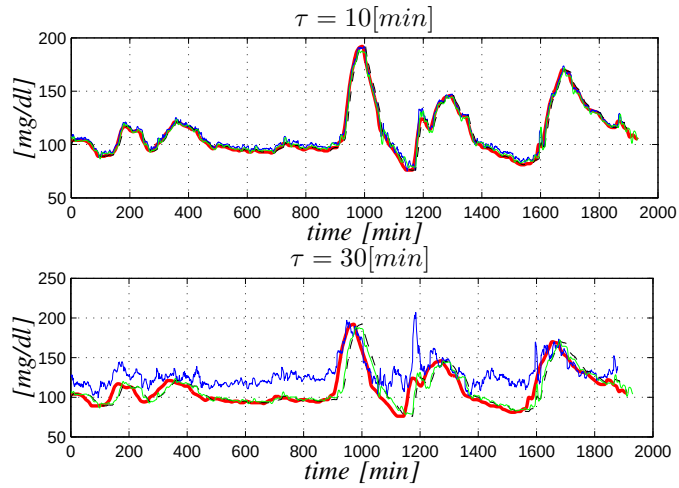

Fig. 4. Example 2. Continuous Glucose Monitoring Sensor data [*mg/dl*] (thick red), prediction with *proj-1* (blue) and prediction with *pbsid* (dashed black) vs. time [*min*]. *Top* 10-min-ahead prediction; *Bottom* 30-min-ahead prediction; $p = 60, f = 30$

B. Future Work

The results presented so far while providing insights and intuitions on how multi-step-ahead linear multivariate output predictors works in real applications do not lead to a rigorous mathematical analysis. In addition, they were obtained for particular applications thus being far from general. Whether or not these results depend on the particular example or can be generalized will be subject of future research.

VII. ACKNOWLEDGMENTS

This research was supported by the European project DIAdvisor ™, FP7 IST-216592 [14].

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.
[2] T. Söderström and P. Stoica, *System Identification*. Upper Saddle River, NJ: Prentice Hall, 1989.
[3] K. Åström, "Maximum likelihood and prediction error methods," *Automatica*, vol. 16, pp. 551–574, 1980.
[4] P. Stoica and A. Nehorai, "On multistep prediction error methods for time series models," *Journal of Forecasting*, vol. 8, pp. 357–368, 1989.
[5] B. D. Anderson and J. B. Moore, *Optimal Filtering*, T. Kailath, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[6] J. Holst, "Adaptive prediction and recursive estimation," Ph.D. dissertation, Department of Automatic Control, Lund Institute of Technology, Sweden, September 1977.
[7] G. Z. E Mosca and J. Lemos, "Robustness of multipredictor adaptive regulator: MUSMAR," *Automatica*, vol. 25, no. 4, pp. 521–529, 1989.
[8] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems : Theory-Implementation-Application*. Boston: Kluwer Academic Publishers, 1996.
[9] T. Katayama, *Subspace Methods for System Identification*. London: Springer-Verlag, 2005.
[10] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.
[11] A. Chiuso, "The role of vector autoregressive modeling in subspace identification," *Automatica*, vol. 43, no. 6, pp. 1034–1048, June 2007.
[12] T. Katayama and G. Picci, "Realization of stochastic systems with exogenous inputs and subspace identification methods," *Automatica*, vol. 35, pp. 1635–1652, 1999.
[13] E. Hannan and D. Poskitt, "Unit canonical correlations between future and past," *The Annals of Statistics*, vol. 16, no. 2, pp. 784–790, 1988.
[14] DIAdvisor www.diadvisor.eu.
[15] Abbott Diabetes Care www.abbottdiabetescare.com.
[16] P. Van Overschee and B. de Moore, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, pp. 75–93, 1994.
[17] J. Maciejovski, *Predictive Control with Constraints*, Pearson, Ed. Pearson Education Limited, 2002.
[18] J. Dong, M. Verhaegen, and E. Holweg, "Closed-loop subspace predictive control for fault tolerant MPC design," in *Proc. of the 17th IFAC World Congress*, Seoul, South Korea, July 2008, pp. 3216–3221.
[19] L. Ljung and T. McKelvey, "Subspace identification from closed loop data," *Signal Processing*, vol. 52, no. 2, pp. 209–215, 1996.
[20] M. Jansson, "Subspace identification and ARX modeling," in *Proc. IFAC Symp. on System Identification*, Rotterdam, August 2003.
[21] R. Johansson, *System Modeling and Identification*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[22] A. Chiuso and G. Picci, "Some algorithmic aspects of subspace identification with inputs," *Int. J. Applied Math. and Computer Science*, vol. 11, no. 1, pp. 55–75, 2001.
[23] A. Chiuso, "Choosing the future horizon in closed-loop cca-type subspace algorithms," *Automatica*, vol. 30, pp. 75–93, 2008.
[24] ——, "On the relation between CCA and predictor-based subspace identification," in *Proc. 44th IEEE Conference on Decision and Control and the European Control Conference*, Seville, December 2005, pp. 4976–4982.
[25] T. Kailath and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall, 2000.