

923

On the use of variograms for the prediction of time series

Michel GEVERS

Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, Australia

Received 17 December 1984
Revised 6 March 1985

Abstract: We consider the prediction of stationary stochastic processes with non-zero mean. When the covariance of the process is known, but the mean is not, the classical approach is to first estimate the mean from the past data, and then apply an optimal predictor to the zero-mean residuals. Bastin and Henriët [1] showed that an alternative was to use a predictor based on 'variograms' rather than covariance information, thus avoiding the estimation of the mean. We show here that the two predictors are identical when the unknown mean is replaced by its minimum variance estimate. We also examine, through simulation, how the two predictors compare when the statistics are unknown.

Keywords: Time series analysis, Prediction, Variogram, Minimum variance prediction.

1. Introduction¹

We consider a discrete scalar stochastic process $\{y_0, y_1, \dots, y_t, \dots\}$ with constant but not necessarily zero mean:

$$E\{y_t\} = m. \quad (1)$$

We define the 'variogram' of the process as the semi-variance of the increments

$$\gamma(t, t - \tau) = \frac{1}{2}E\{(y_t - y_{t-\tau})^2\}. \quad (2)$$

We shall assume throughout that the variogram is stationary, i.e.

$$\gamma(t, t - \tau) = \gamma(\tau) = \frac{1}{2}E\{(y_t - y_{t-\tau})^2\}. \quad (3)$$

A process with the properties (1) and (3) is called

¹ Because this paper is an extension and correction of [1], we shall for the most part use the same notations to facilitate comparison.

intrinsic (see e.g. [2]). If the process is wide-sense stationary, then the (auto)covariance can be defined as

$$R(\tau) = E\{(y_t - m)(y_{t-\tau} - m)\}. \quad (4)$$

For a wide-sense stationary process, the covariance function and the variogram are related by

$$\gamma(\tau) = \sigma^2 - R(\tau), \quad \text{with } \sigma^2 = R(0). \quad (5)$$

Note that an intrinsic random process is always wide-sense stationary, but the converse is not true. The class of intrinsic processes is larger and includes Wiener processes as a special case. Except when specifically stated, we shall from now on assume that the process $\{y_t\}$ is wide-sense stationary.

In this paper we shall derive different minimum variance unbiased (MVU) expressions for $(d + 1)$ -step ahead predictions of the $\{y_t\}$ process under a variety of assumptions. In Sections 2 to 4 we shall assume known second-order statistics and consider Levinson predictors with growing memory: the predicted value at time $t = N$ uses all available past data. In Section 2 we shall briefly recall the expressions of the classical Levinson predictor (CLP) for y_{N+d} , given $\{y_0, y_1, \dots, y_{N-1}\}$, under the assumptions that the constant mean m and the covariance (or the variogram) are known. In Section 3 we consider the case where the mean is unknown. Two different MVU predictors can be used in this case:

(1) one can compute an unbiased estimate \hat{m} , and then replace the mean by its estimate \hat{m} in the expressions of the CLP of Section 2. This predictor will be called Approximate Levinson Predictor (ALP);

(2) alternatively, one can use a MVU predictor of the form $\sum_{i=1}^N b_i Y_{N-i}$, in which the b_i are computed from the variogram function. This predictor, derived by Bastin and Henriët [1], who called it the modified Levinson predictor, does not involve the mean or its estimate in any way.

In Section 4 we show that the two predictors are identical if the MVU estimate is used for \hat{m} in the ALP. This may appear as a surprising result, considering that the first involves an estimate of the mean, while the other does not.

In Sections 5 and 6, we shall consider the more realistic situation where the statistics of the process are unknown. We now consider predictors of fixed order N , whose coefficients are determined from sample estimates of the covariance function or the variogram. These are computed at time t using all t available data ($t \gg N$). Since a theoretical analysis appears extremely difficult, the two predictors (MLP and ALP) are compared on the basis of heuristic arguments (in Section 5) and simulations (in Section 6).

2. Processes with known mean and covariance function

When the mean is known, the classical $(d+1)$ -step ahead Levinson predictor (CLP) for Y_{N+d} given y_0, \dots, y_{N-1} has the form (see e.g. [3])

$$\hat{y}_{N+d} = m + \sum_{i=1}^N a_i (y_{N-i} - m) = m + \mathbf{a}^T (Y - mU) \tag{6}$$

where

$$\mathbf{a}^T = (a_1, \dots, a_N), \quad U^T = (1, 1, \dots, 1),$$

$$Y^T = (y_{N-1}, \dots, y_0).$$

The prediction error $\tilde{y}_{N+d} \triangleq \hat{y}_{N+d} - y_{N+d}$ is unbiased. The minimization of the prediction error variance $E\{(\tilde{y}_{N+d})^2\}$ with respect to the a_i leads to the following system of N linear equations:

$$R\mathbf{a} = R_0 \tag{7}$$

where

$$R = E\{(Y - mU)(Y - mU)^T\}, \tag{8a}$$

$$R_0 = E\{(Y - mU)(y_{N+d} - m)\}. \tag{8b}$$

Using (5), we can also write

$$R = \sigma^2 U U^T - \Gamma, \quad R_0 = \sigma^2 U - \Gamma_0, \tag{9}$$

where

$$\Gamma = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & & \gamma(N-2) \\ \vdots & & \ddots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \dots & \gamma(0) \end{bmatrix}, \tag{10a}$$

$$\Gamma_0 = \begin{bmatrix} \gamma(d+1) \\ \gamma(d+2) \\ \vdots \\ \gamma(d+n) \end{bmatrix}. \tag{10b}$$

Note that $\gamma(0) = 0$. Substituting (9) in (7) yields an alternative system of equations for \mathbf{a} in terms of the variogram:

$$\begin{bmatrix} \Gamma & U \\ U^T & \sigma^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \alpha \end{bmatrix} = \begin{bmatrix} \Gamma_0 \\ 1 \end{bmatrix}, \tag{11}$$

where

$$\alpha \triangleq 1 - \mathbf{a}^T U = 1 - \sum_{i=1}^N a_i. \tag{12}$$

Using matrix and vector notations, (11) can be rewritten as $A_1 \lambda_1 = c$, with obvious definitions for A_1 , λ_1 and c . The optimal prediction error variance, $V_c \triangleq E\{(\tilde{y}_{N+d})^2\}$, can be written in a number of ways:

$$\begin{aligned} V_c &= \sigma^2 - \mathbf{a}^T R_0 = \alpha \sigma^2 + \mathbf{a}^T \Gamma_0 \\ &= \sigma^2 - R_0^T R^{-1} R_0 = c^T A_1^{-1} c. \end{aligned} \tag{13}$$

3. Processes with unknown mean and known second-order statistics

We now consider the case where the constant mean m is unknown, but where the covariance function $R(\tau)$ or the variogram $\gamma(\tau)$ are assumed known. Recall that they are related by (5).

3.1. The approximate Levinson predictor

The most obvious strategy is to replace the mean in (6) by a linear unbiased estimate based on the past data:

$$\hat{m} = \sum_{i=1}^N \zeta_i y_{N-i} = \zeta^T Y \tag{14}$$

with the unbiasedness condition

$$\xi^T U = 1. \quad (15)$$

The CLP of (6) is then replaced by the approximate Levinson predictor (ALP):

$$y_{N+d}^* = \hat{m} + a^T(Y - \hat{m}U) \quad (16)$$

where a is the solution of (7), as before. This estimator was called approximate classical Levinson predictor (ACLIP) in [1]. The prediction error can be written

$$\begin{aligned} \tilde{y}_{N+d} &\triangleq y_{N+d}^* - y_{N+d} \\ &= m + a^T(Y - mU) - y_{N+d} + \alpha(\hat{m} - m). \end{aligned} \quad (17)$$

The sum of the first two terms in (17) is the optimal CLP, so that the sum of the first three terms is the CLP prediction error. Therefore, the prediction error variance

$$V_a \triangleq E\{(y_{N+d}^* - y_{N+d})^2\}$$

is

$$\begin{aligned} V_a &= V_c + \alpha^2 \sigma_{\hat{m}}^2 \\ &\quad + 2\alpha E\{[m + a^T(Y - mU) - y_{N+d}][\hat{m} - m]\} \\ &= V_c + \alpha^2 \sigma_{\hat{m}}^2 \\ &\quad + 2\alpha E\{[a^T(Y - mU) - y_{N+d}][(Y - mU)^T \xi]\} \end{aligned}$$

since \hat{m} is unbiased, $\sigma_{\hat{m}}^2$ denotes the variance of \hat{m} . The third term in the above expression is zero by (7). Therefore

$$V_a = V_c + \alpha^2 \sigma_{\hat{m}}^2. \quad (18)$$

It follows from (18) that V_a will be minimum if $\sigma_{\hat{m}}^2$ is minimum, i.e. if \hat{m} is chosen as the MVU estimate of m . The corresponding estimator will be called MVALP, for Minimum Variance Approximate Levinson Predictor. The MVU estimate for \hat{m} is obtained as follows:

$$\begin{aligned} \sigma_{\hat{m}}^2 &= E\{(\hat{m} - m)^2\} = E\{[\xi^T(Y - mU)]^2\} \\ &= \xi^T R \xi = \sigma^2 - \xi^T \Gamma \xi. \end{aligned} \quad (19)$$

Minimizing $\xi^T R \xi$ w.r.t. ξ subject to (15) yields

$$\xi = (U^T R^{-1} U)^{-1} R^{-1} U. \quad (20)$$

After lengthy manipulations, using (5), ξ can also

be expressed in terms of the variogram function $\gamma(\tau)$:

$$\xi = (U^T \Gamma^{-1} U)^{-1} \Gamma^{-1} U. \quad (21)$$

The corresponding minimum error variance for the mean estimate will be denoted $\bar{\sigma}_m^2$:

$$\bar{\sigma}_m^2 = (U^T R^{-1} U)^{-1} = \sigma^2 - (U^T \Gamma^{-1} U)^{-1}. \quad (22)$$

Expression (18) shows that the ALP yields a larger error variance than the CLP, and that the two error variances coincide when $\sigma_{\hat{m}}^2 \rightarrow 0$, as should be expected. We shall denote by \bar{V}_a the prediction error variance of the MVALP. Using (12), (7), (22) and (13), we can write

$$\bar{V}_a = V_c + \alpha^2 (U^T R^{-1} U)^{-1} \quad (23)$$

$$\begin{aligned} &= V_c + (1 - R_0^T R^{-1} U)^2 (U^T R^{-1} U)^{-1} \\ &= \sigma^2 - R_0^T R^{-1} R_0 \\ &\quad + (1 - R_0^T R^{-1} U)^2 (U^T R^{-1} U)^{-1}. \end{aligned} \quad (24)$$

Note that this last expression is entirely in terms of the covariance function of $\{y_i\}$.

3.2. The modified Levinson predictor (MLP)

When the mean m is unknown, an alternative strategy is to seek the MVU predictor of the following form:

$$\hat{y}_{N+d} = \sum_{i=1}^N b_i y_{N-i} = \mathbf{b}^T \mathbf{Y} \quad (25)$$

together with the constraint, imposed by unbiasedness, that

$$\sum_{i=1}^N b_i = 1, \quad \text{i.e. } \mathbf{b}^T \mathbf{U} = 1. \quad (26)$$

Minimization of $E\{(\tilde{y}_{N+d})^2\}$ w.r.t. \mathbf{b} subject to (26) yields the following linear system of $(N+1)$ equations:

$$\begin{bmatrix} R & -U \\ U^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mu \end{bmatrix} = \begin{bmatrix} R_0 \\ 1 \end{bmatrix}, \quad (27)$$

where μ is a Lagrange coefficient. Alternatively, using (5), we get

$$\begin{bmatrix} \Gamma & U \\ U^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mu \end{bmatrix} = \begin{bmatrix} \Gamma_0 \\ 1 \end{bmatrix}, \quad (28a)$$

or, with obvious matrix and vector notations,

$$A_2 \lambda_2 = c. \quad (28b)$$

The prediction error variance can be written in a number of ways:

$$V_m = \sigma^2 + \mathbf{b}^T R \mathbf{b} - 2 \mathbf{b}^T R_0 = 2 \mathbf{b}^T \Gamma_0 - \mathbf{b}^T \Gamma \mathbf{b} \quad (29a)$$

$$= \mu + \sigma^2 - \mathbf{b}^T R_0 = \mu + \mathbf{b}^T \Gamma_0 = \mathbf{c}^T A_2^{-1} \mathbf{c}. \quad (29b)$$

The predictor MLP was proposed by Bastin and Henriot [1], who showed that

$$V_m = V_c + \alpha^2 (U^T R^{-1} U)^{-1} \quad (30)$$

with α defined by (12). This, together with (23), shows that

$$V_m = \bar{V}_a \leq V_a. \quad (31)$$

The inequality in (31) follows from the fact that $\bar{V}_a = \min_{\xi} V_a$. We show in the next section that the MVALP and the MLP are identical, which of course explains the equality in (31).

4. The MVALP and the MLP are identical

Proof. It follows from (16) and (14) that the MVALP can be written

$$\begin{aligned} y_{N+d}^* &= (1 - \mathbf{a}^T U) \hat{m} + \mathbf{a}^T Y \\ &= (\alpha \xi^T + \mathbf{a}^T) Y. \end{aligned} \quad (32)$$

Therefore, we need to prove that

$$\mathbf{a} + \alpha \xi = \mathbf{b} \quad (33)$$

where \mathbf{a} , ξ , \mathbf{b} are solutions of (7), (20) and (27) respectively, and with α given by (12).

The solution of (27) is unique, because it is the vector \mathbf{b} that minimizes

$$E \left\{ [y_{N+d}]^2 \right\} = \mathbf{b}^T R \mathbf{b} + \sigma^2 - 2 \mathbf{b}^T R_0$$

subject to (26). It is also trivial to see that $\mathbf{a} + \alpha \xi$ satisfies the last equation of (27). Therefore, to prove (33), it remains to be shown that there exists a μ such that

$$R(\mathbf{a} + \alpha \xi) - \mu U = R_0. \quad (34)$$

It follows immediately from (7), (20) and (22) that this is so for

$$\mu = \alpha \bar{\sigma}_m^2. \quad (35)$$

This concludes the proof.

In the process of proving our main result, we have shown that the solution μ of (27) and (28) is $\alpha \bar{\sigma}_m^2$. This yields some closed form expressions for \mathbf{b} and some new expressions for μ in terms of either $R(\tau)$ or $\gamma(t)$. From (35), (27) and (28) we get

$$\mathbf{b} = R^{-1} (R_0 - \alpha \bar{\sigma}_m^2 U) = \Gamma^{-1} (\Gamma_0 - \alpha \bar{\sigma}_m^2 U). \quad (36)$$

Using (35), (12) and (22), we have

$$\mu = \frac{1 - U^T R^{-1} R_0}{U^T R^{-1} U} = \frac{U^T \Gamma^{-1} \Gamma_0 - 1}{U^T \Gamma^{-1} U}. \quad (37)$$

The second equality in (37) is obtained by multiplying (36) to the left by U^T , and using (26), (12) and (22). This then allows us to write two expressions for \mathbf{b} in terms of $R(\tau)$ or $\gamma(\tau)$ only:

$$\mathbf{b} = R^{-1} R_0 + \frac{1 - U^T R^{-1} R_0}{U^T R^{-1} U} R^{-1} U \quad (38a)$$

$$= \Gamma^{-1} \Gamma_0 + \frac{1 - U^T \Gamma^{-1} \Gamma_0}{U^T \Gamma^{-1} U} \Gamma^{-1} U. \quad (38b)$$

The formulas (38a) and (38b) are remarkable because the expressions in terms of (R, R_0) and (Γ, Γ_0) are identical. Note that the first term of (38a) is \mathbf{a} , while the second term is $\alpha \xi$. (Recall (33).) This is not so, however, for (38b), because $R^{-1} R_0 \neq \Gamma^{-1} \Gamma_0$. Finally, note that the expressions (37) and (38) can also be obtained from (27) using a matrix inversion lemma.

In Bastin and Henriot [1], it was claimed that $\bar{V}_a - V_m = \alpha^2 \sigma^2 \geq 0$, and that therefore the MLP was better than the MVALP. This was based on the fact that V_a had been erroneously computed as $V_a = V_c + \alpha^2 (\sigma_m^2 + \sigma^2)$ rather than the correct expression (18). Note, however, that the equality $V_m = \bar{V}_a$ holds only if the unbiased estimator \hat{m} is chosen to be the minimum variance estimator. If the sample mean estimator

$$\hat{m} = \frac{1}{N} \sum_{i=1}^N y_{N-i} \quad (39)$$

is used in (16), then $V_m \leq V_a$.

5. The case of unknown statistics: Heuristic comparison

In this section we want to compare the ALP and the MLP in the case where m , $R(\tau)$ and $\gamma(\tau)$

are unknown and must be estimated from the data. In such cases it is impossible to use the growing memory Levinson filters described above, because it does not make sense to estimate N predictor coefficients from N data. We shall therefore compare finite-order predictors (of order N) whose coefficients a_i , b_i are computed as before, but with m , $R(\tau)$ and $\gamma(\tau)$ replaced by their sample estimates \hat{m} , $\hat{R}(\tau)$ and $\hat{\gamma}(\tau)$. The ALP (16) and the MLP (25) are now replaced, respectively, by

$$y_{t+d}^* = \hat{m} + \sum_{i=1}^N a_i (y_{t-i} - \hat{m}) \quad (40)$$

and

$$\hat{y}_{t+d} = \sum_{i=1}^N b_i y_{t-i} \quad (41)$$

with the a_i and b_i computed as indicated. The sample estimates are obtained using all t data, assuming $t \gg N$:

$$\hat{m} = \frac{1}{t} \sum_{k=1}^t y_{t-k}, \quad (42a)$$

$$\hat{R}(\tau) = \frac{1}{t-\tau} \sum_{k=1}^{t-\tau} (y_{t-k} - \hat{m})(y_{t-k-\tau} - \hat{m}), \quad (42b)$$

$$\hat{\gamma}(\tau) = \frac{1}{2(t-\tau)} \sum_{k=1}^{t-\tau} (y_{t-k} - y_{t-k-\tau})^2. \quad (42c)$$

Note that in the growing memory predictors of Section 3 and 4, the predictor length coincides with the number of data upon which the mean estimate is based. This is not so in the fixed length predictor (40). Therefore the result (33) of Section 4 no longer holds.

However, if we redefine Y as

$$Y^T = (y_{t-1}, \dots, y_{t-N})$$

and R and R_0 accordingly, and if we denote by V_c , V_a and V_m the prediction error variances, respectively, of the finite length predictors CLP, ALP and MLP using exact covariances for the computation of the a_i and b_i and (42a) for the estimation of \hat{m} , then the expression (30) for V_m is still valid, and therefore

$$\text{for } t = N: \quad V_a \geq V_m, \quad (43a)$$

$$\text{for } t \rightarrow \infty: \quad \sigma_{\hat{m}}^2 \rightarrow 0 \text{ and hence } V_a \rightarrow V_c \leq V_m. \quad (43b)$$

On the basis of this observation, we would like to compare the finite length ALP (40) and the MLP (41) in the practical case where the unknown m , $R(\tau)$ and $\gamma(\tau)$ are estimated from increasing data using (42)². We would also like to know whether the predictor coefficients should be estimated using $\hat{R}(\tau)$ or $\hat{\gamma}(\tau)$. (Recall that all predictors can be computed using either covariance or variogram formulas.)

An exact comparison of predictor error variances is extremely complicated, but the following heuristic arguments can be made.

(1) Clearly, for $t \rightarrow \infty$, the estimators \hat{m} , $\hat{R}(\tau)$ and $\hat{\gamma}(\tau)$ converge to their true values, and therefore the ALP will converge to the CLP. Hence, asymptotically the ALP is better than the MLP: see (43b).

(2) However, for small sample sizes (t small), there is a good reason to think that the MLP, with its coefficients computed from the variogram formulas (38b), should be preferred over the ALP with \hat{m} replaced by its sample mean. One argument is that the MLP does not require a preliminary estimate of the mean, and that the $\hat{\gamma}(\tau)$ are unbiased estimates that are independent of the mean. The other reason is that, for $t = N$, $V_a \geq V_m$ when second-order statistics are known. One would think that this result would still hold for small sample sizes when exact covariances or variograms are replaced by estimated ones, although this would be extremely hard to prove.

(3) The computation of the MLP using the variogram formulas (38b) requires $N + d$ statistics, $\{\hat{\gamma}(\tau), \tau = 1, \dots, N + d\}$, while the computation of the MLP using the covariance formulas (38a), as well as the computation of the ALP, require $N + d + 2$ statistics.

(4) For intrinsic (but not necessarily wide-sense stationary) processes with known variograms, the MLP estimator (38b) is the optimal linear predictor: it can be derived directly without going through (5). In such case $R(\tau)$ and σ^2 may not exist. Therefore the MLP (38b) covers a wider class of random processes, and should be preferred in case of doubt on the wide-sense stationarity of the process. As it turns out, stationarity is very difficult to validate in practice. If covariance

² Of course, in such case other predictors might be used such as maximum likelihood or least squares predictors derived directly from the data, but our aim in this discussion is to compare the ALP and MLP predictors.

estimates are used under the false belief that the process is stationary, this may lead to completely erroneous results, as the following example, due to Matheron [5], shows. Consider a Wiener process with variogram $\gamma(\tau) = |\tau|$, and suppose we have observed $\{y_0, y_1, \dots, y_t\}$. If the user believes the process is stationary and computes \hat{m} and $\hat{R}(\tau)$ by (42), then it can be shown that, for $\tau \geq 0$,

$$E\{\hat{R}(\tau)\} = \frac{t^2 - 1}{3t} - \frac{4}{3}\tau + \frac{2}{3}\frac{\tau^2}{t}. \quad (44)$$

It is a parabola; an apparent variance of $\hat{R}(0) = (t^2 - 1)/3t$ will be found, whereas the true variance is infinite. The sample variogram (42c), on the other hand, is unbiased.

6. Simulations

The ALP (with the sample mean used for \hat{m} and the coefficients a_i computed from the sample covariances) and the MLP (with the coefficients computed from the sample variograms) have been compared on simulated data generated from a large number of ARMA(p, q) models, using short time series. For each run, the ND available data have been split up into ND = NE + NP data. The first NE data were used to estimate the predictor coefficients; the remaining NP data were then used for validation: the mean square prediction error obtained on these NP data was computed. For each model, different combinations were chosen for the order of the predictor N , the prediction horizon d , the number of data used for estimation NE, and the number of data used for prediction NP. For each one of those, the mean square prediction errors V_a (for the ALP) and V_m (for the MLP) were computed as the average over NR independent runs. The conclusions drawn from these simulations are as follows.

(1) The simulations do not confirm the heuristic arguments (2) of Section 5: even though the MLP was almost always better than the ALP, particularly for short time series (NE small), there were cases the opposite was true.

(2) For processes with exponentially decaying covariances, the MLP turned out to be consistently better. For processes with damped oscillatory covariances, the ALP was sometimes better.

(3) As the process gets closer to white noise (i.e.

Table 1

Model used to generate data	Test	V_a	V_m
$(y_k - 5) - 0.95(y_{k-1} - 5) = e_k$	1	11.08	5.50
	2	5.68	3.12
$(y_k - 3) - 0.7(y_{k-1} - 3) = e_k$	1	4.33	3.14
	2	3.86	2.70
$(y_k - 5) - 0.5(y_{k-1} - 5) = e_k$	1	2.50	2.44
	2	2.98	2.45
$(y_k - 5) + 0.95(y_{k-1} - 5) = e_k$	1	3.58	9.76
	2	3.48	3.00
$(y_k - 3) + 0.7(y_{k-1} - 3) = e_k$	1	1.63	1.69
	2	2.63	2.35
$(y_k - 5) + 0.5(y_{k-1} - 5) = e_k$	1	1.45	1.64
	2	2.16	1.93

Test 1: $N = 5, d = 3, NE = 30, NP = 20, NR = 10$.

Test 2: $N = 3, d = 1, NE = 15, NP = 10, NR = 20$.

e_k = zero mean white Gaussian noise with unit variance.

V_a = average over NR runs of the mean square prediction error using the ALP and covariance estimates.

V_m = average over NR runs of the mean square prediction error using the MLP and variogram estimates.

with a covariance decaying rapidly to zero) the superiority of one predictor over the other (see point 2) decreases. Typical results are shown in Table 1: they are a good illustration of conclusions 2 and 3 above. Simulations with higher-order models have also been performed: they confirm these conclusions.

7. Concluding remarks

We have shown that, with known second-order moments but unknown mean, the MLP, introduced in [1], is identical to the classical Levinson predictor in which the unknown mean is replaced by its minimum variance unbiased estimate. It is therefore better than any approximate Levinson predictor using any other mean estimate. One might have expected this property to carry over to the case of fixed-length predictors used on short time series with unknown (and hence estimated) second-order statistics. Simulations show that this is true in most cases, but not all. This result indicates that, even though the variogram based predictor is not consistently better than the covariance based one, in the case of estimated statistics it deserves to be considered and perhaps further

studied by the control community. It should also be recalled that, when the process is not wide-sense stationary but has independent increments, the variogram based predictor is optimal, while the covariance based predictor is not.

Acknowledgements

The author wishes to thank Georges Bastin and Albert Benveniste for interesting comments on an earlier draft of this paper, and Marie-Antoinette Poubelle for a clever suggestion.

References

- [1] G. Bastin and E. Henriot, A minimum variance property of Levinson predictors using variograms, *Systems Control Lett.* **2** (1982) 25-29.
- [2] G. Matheron, *Les Variables Régionalisées et leur Estimation* (Masson, Paris, 1965).
- [3] B.D.O. Anderson and J.B. Moore, *Optimal Filtering* (Prentice Hall, Englewood Cliffs, NJ, 1979).
- [4] T.W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1958).
- [5] G. Matheron, *The Theory of Regionalized Variables and its Applications*, Cahiers du Centre de Morphologie Mathématique de Fontainebleau (1971).

