2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Stochastic Adaptive Learning Rate in an Identification Method: An Approach for On-line Drilling Processes Monitoring

A. BA, S. Hbaieb, N. Mechbal, and M. Vergé

Abstract—**On-line drilling processes monitoring is an essential task in enhancing their performances. In oilfield industry, dysfunctions that might occur have to be detected at the earliest possible stage in order to preserve drilling efficiency. This paper deals with a methodology for drilling processes monitoring by identifying time varying parameters. The basic idea behind the proposed algorithm is to improve the tracking ability of parameters change by means of an identification method using a new approach to adjust the forgetting factor. The effectiveness of the developed method is highlighted through experimental data obtained from tests campaign.**

I. Introduction

THE necessity to discover petroleum in hostile environments, justified by the decrease of world's oil reserve, attracts a great deal of interest in terms of scientific investigation and drilling processes (Fig. 1.) advancements, for economic success. In the context where the daily drilling rig cost can exceed 500.000 USD, drilling processes monitoring become important. Its benefits could improve operational performances as well as predictive maintenances. Nevertheless, one of the major requirements for efficient real-time monitoring is to receive and process data quickly in order to make a useful decision. This requirement is possible with the new progress made in telemetry technology that enables data transfer at frequencies exceeding 1 KHz. In addition, a better understanding of drilling processes behavior combined with the knowledge of a part of their theoretical modeling, identification and statistical signal processing [1], [2] are opening new ways for drilling processes monitoring [3]. However, drilling processes remain complicated systems and rocks being drilled are more often badly known. As a result, a model which takes into account accurately both process and rock is difficult to achieve and represents a research topic of grows interesting.
During drilling, fluid circulation plays a crucial role. It reduces the rock hardness by hydration process and allows cuttings transportation from downhole to surface. When cuttings process proceeds, a bit balling dysfunction (Fig. 2.) can occur. This scenario happens when cuttings and mud (drilling fluid) stuck on the teeth of the bit (head of the process) and cannot be washed by the fluid circulation. At a certain stage, the bit becomes completely covered; consequently the cutting process cannot be accomplished as expected [4]. For this reason, several studies concerning bit balling reduction were carried out and some results were found. Among them: drill bit design optimization, suitable drilling parameters and the best drilling mud choice are mostly used as solutions by drilling engineers. However, the main drawback of these strategies results from the lack of knowledge concerning the rock to drill before drilling operations. To overcome this difficulty, bit-rock interaction model has been developed [5]. The specificity of the achieved model lies in the possibility of not involving rock properties. Based on this model, a real-time approach for bit balling is developed in this paper. The methodology developed here falls into the categories of recursive least square algorithm (RLS) [1]. Note that, one of the noteworthy advantages of RLS arises out of the capability that it tracks time varying parameters by discarding old data and giving more weight to recent ones through a forgetting factor. In this case, the accuracy of the estimated parameters depends greatly on the choice of the forgetting factor. It is well-known that, the lowering the forgetting factor increase the rate of false alarm and the increasing the forgetting factor increase the lack of detection. To avoid such situations, a lot of efforts have been made to make the forgetting factor adjustable [6], [7].

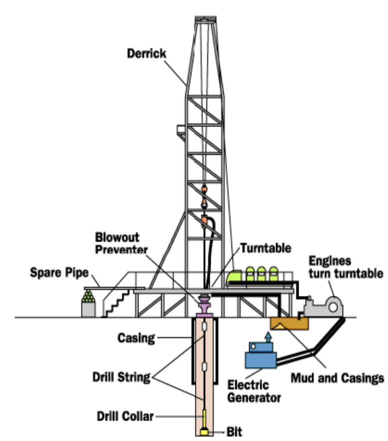

Fig. 1. Overall scheme of a drilling process

A. BA. is with both Laboratoire de Mécanique des Systèmes et des Procédés (UMR-CNRS), Ecole Nationale Supérieure d'Arts et Métiers, 151, Boulevard de l'hôpital, Paris, France and Schlumberger Riboud Product Centre, 1, rue Henri Becquerel, 92142 Clamart, Cedex, France (phone: 331-45-37-70-87; fax: 45-37-77-82; aba2@ slb.com).
S. Hbaieb. is with Schlumberger Riboud Product Centre, 1, rue Henri Becquerel, 92142 Clamart, Cedex, France (shbaieb@slb.com).
N. Mechbal. and M. Vergé are with Laboratoire de Mécanique des Systèmes et des Procédés (UMR-CNRS), Ecole Nationale Supérieure d'Arts et Métiers, 151, Boulevard de l'hôpital, Paris, France (Nazih.mechbal@paris.ensam.fr), (Michel.Verge@paris.ensam.fr).

978-1-4244-4524-0/09/$25.00 ©2009 AACC

It has been developed in [8] an algorithm which adjusts the forgetting factor by using stochastic gradient descent method. In this method, the learning rate plays an essential role. Nevertheless, it is widespread that the quality of the gradient descent method is improved if the learning rate is variable. In this context, in addition to develop a methodology for bit balling detection, it is the intent of this paper to extend the algorithm proposed in [8] by adjusting the forgetting factor through stochastic gradient descent method and making the learning rate adaptive [9]. The benefit of the proposed algorithm allows the forgetting factor to move along the direction which does not necessarily coincide with the gradient descent direction [10]. Thereby, the minimization process is accelerated and the fast convergence is achieved. Eventually, it is also shown by this method that the tracking ability is enhanced.

A brief introduction and summary of the bit-rock interaction model is pointed out in section II. The proposed identification method is presented in section III. In section IV some experimental data in case of bit balling are presented as well as the accurate estimate achieved by using the developed methodology. In section V concluding remarks are drawn.

Fig. 2. Drill bit – on the left without balling – on the right with balling

II. BIT-ROCK INTERACTION MODEL

Research activities regarding bit-rock interaction modeling have been investigated and significant contributions have been reported [5]. These models account for the cutting action of a single cutter, which is a Polycrystalline Diamond Compact (PDC). Cutting action of each cutter consists of two independent processes, the cutting process and the friction process. The torque T and the weight W are defined by:

$$T = T_c + T_f \ , \quad W = W_c + W_f$$

where:

$$T_c = \frac{\varepsilon \cdot a^2 \cdot d}{2} \ , \ W_c = \xi \cdot \varepsilon \cdot a \cdot d$$

$$T_f = \frac{\mu \cdot a \cdot \gamma \cdot W_f}{2} \ , \ d = \frac{2 \cdot \pi \cdot v}{\Omega}$$

The subscripts c and f denote cutting and friction respectively. For description of variables see TABLEI. The cutting components T_c and W_c correspond to the forces transmitted by the cutting face of each cutter and the friction components T_f and W_f correspond to others contacts generated from bit-rock interaction. The response of the bit is obtained by combining the cutting and frictional process [5]:

$$\frac{2 \cdot T}{a^2 \cdot d} = (1 - \beta) \cdot \varepsilon + \mu \cdot \gamma \cdot \frac{W}{a \cdot d} \quad (1)$$

with:

$$\beta = \mu \cdot \gamma \cdot \zeta$$

The specific energy which corresponds to the necessary energy to grind a given volume of rock is defined by [5]:

$$E = E_0 + \mu \cdot \gamma \cdot S \quad (2)$$

where:

$$E_0 = (1 - \beta) \cdot \varepsilon$$

Through (1) and (2) the following relations have been obtained:

$$S = \frac{W}{a \cdot d} \ , \ E = \frac{T}{a^2 \cdot d}$$

It is clear from (2) that the slope of E as a function of S is only dependent on the bit shape factor and the friction coefficient. Therefore, if the bit balling occurs these parameters will be adversely affected. As a consequence, the bit balling is considered as a change in this slope. Thereby, our objective amount to exploit the model (2) to develop a strategy for bit balling diagnosis. To achieve this aim, we propose an identification method based on a new adjustment of the forgetting factor. This adjustment is made according to the gradient descent method in the case where the learning rate is adaptive.

TABLE I
VARIABLES USED

Symbol	Quantity	Unity
a	Bit radius	m
d	Depth of cut	m
E	Specific energy	Mpa
S	Drilling strength	Mpa
T	Torque on bit	Nm
T_c	Torque on bit, cut	Nm
T_f	Torque on bit, friction	Nm
W	Weight on bit	N
W_c	Weight on bit, cut	N
W_f	Weight on bit, friction	N
ε	Intrinsic specific energy	Mpa
v	Rate of penetration (ROP)	m/h
Ω	Bit angular velocity	rad/s
μ	Friction coefficient	
ξ	Coefficient characterizing cutting force	
γ	Bit shape factor	

III. IDENTIFICATION METHOD

A. RLS Identification

The feature of the developed algorithm is to enhance the tracking ability in transient stage, by automatically shortening the value of the forgetting factor λ. It is performed according to the gradient descent method with an adaptive learning rate. We obtain an algorithm called stochastic adaptive learning rate gradient variable forgetting factor in recursive least square context (SALR-GVFF-RLS). The proposed algorithm is an extension of the gradient variable forgetting factor recursive least square (GVFF-RLS) developed in [9]. The SALR-GVFF-RLS enables the forgetting factor to achieve a fast descent owing to an adaptive learning rate which accelerates the convergence of the algorithm, when a transient stage happens. Consequently, the tracking ability is improved. The advantage of the proposed method allows being able to diagnose the bit balling when it is incipient.

Let us take again the equation (2) and rewrite it in a state space representation:

$$\theta(k+1) = \theta(k)$$
$$y(k) = \varphi^T(k)\cdot\theta(k) + \upsilon(k)$$

Where $\theta \in R^n$ is the parameter vector to be estimated:

$$\theta(k) = \begin{bmatrix}\theta_1(k)\\ \theta_2(k)\end{bmatrix} = \begin{bmatrix}\mu\cdot\gamma\\ E_0\end{bmatrix}$$

$y \in R^m$, is the output vector, here it is the specific energy *E*, $\upsilon(k)$ is the observation noise vector, $\varphi^T(k)$ stands for the regressor vector, and is constructed from drilling strength *S*. The classic recursive least square estimation with forgetting factor (FF-RLS) is given by [1]:

$$K(k+1) = \frac{P(k)\cdot\varphi(k+1)}{\lambda + \varphi^T(k+1)\cdot P(k)\cdot\varphi(k+1)} \tag{3}$$
$$P(k+1) = \lambda^{-1}\cdot[P(k) - K(k+1)\cdot\varphi^T(k+1)\cdot P(k)]$$
$$\varepsilon(k+1) = y(k+1) - \varphi^T(k+1)\cdot\hat{\theta}(k)$$
$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1)\cdot\varepsilon(k+1)$$

Our objective is to improve the adaptation quality by using a variable forgetting factor $\lambda(k)$ on the evidence of the gradient descent method as given by:

$$\lambda(k+1) = \lambda(k) - \eta\cdot\nabla_\lambda(k+1) \tag{4}$$

In (4), η denotes the learning rate. However, it is well-known that the performances of the gradient descent method depend on the choice of the learning rate. Specially, its value needs to be in a certain range so that to perform a good convergence. Moreover, it is not an easy task to choose the best learning rate value. In this context, we propose a new algorithm that makes the learning rate adaptive consequently to improve the effectiveness of the identification method in terms of tracking ability. At each stage, the determined learning rate depends on the current and the previous gradient value as shown in [9].

$$\eta(k+1) = \eta(k) + \alpha\cdot\langle\nabla_\lambda(k), \nabla_\lambda(k+1)\rangle \tag{5}$$

In this relation $\langle\cdot,\cdot\rangle$ represents the usual inner product in IR^n, α is called the meta-learning rate. The gradient $\nabla_\lambda(k+1)$ is obtained by adopting the following methodology [8]: first, the following cost function is minimized:

$$J(k+1) = \frac{1}{2}\cdot E\left[\varepsilon^2(k+1)\right] \tag{6}$$

In (6), $E[\cdot]$ represents an expectation operator. Then,

$$\begin{aligned}\nabla_\lambda(k+1) &= \frac{\partial J(k+1)}{\partial\lambda}\\ &= E\left[\frac{\partial\varepsilon(k+1)}{\partial\lambda}\cdot\varepsilon(k+1)\right]\end{aligned} \tag{7}$$

The relationship (7) is obtained by differentiating (6) with respect to λ.
by defining:

$$\psi(k+1) = \frac{\partial\hat{\theta}(k+1)}{\partial\lambda}$$

we have :

$$\begin{aligned}\frac{\partial\varepsilon(k+1)}{\partial\lambda} &= \frac{\partial}{\partial\lambda}[y(k+1) - \varphi^T(k+1)\cdot\hat{\theta}(k)]\\ &= -\varphi^T(k+1)\cdot\psi(k)\end{aligned} \tag{8}$$

using (7) and (8) we obtain:

$$\nabla_\lambda(k+1) = -E[\psi^T(k)\cdot\varphi(k+1)\cdot\varepsilon(k+1)]$$

In accordance with [8] and after some calculations, $K(k+1)$ can be rewritten as:

$$K(k+1) = P(k+1)\cdot\varphi(k+1)$$

therefore, the parameter vector in the equation (3) becomes:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + P(k+1)\cdot\varphi(k+1)\cdot\varepsilon(k+1)$$

by defining:

$$S(k+1) = \frac{\partial P(k+1)}{\partial\lambda}$$

and, after some calculations, the following relation is obtained:

$$\psi(k+1) = (I - K(k+1)\cdot\varphi^T(k+1))\cdot\psi(k) + S(k+1)\cdot\varphi(k+1)\cdot\varepsilon(k+1)$$

We have also:

$$S(k+1) = \lambda^{-1}\cdot(I - K(k+1)\cdot\varphi^T(k+1))\cdot S(k)\cdot (I - \varphi(k+1)\cdot K^T(k+1)) - \lambda^{-1}\cdot P(k+1) + \lambda^{-1}\cdot K(k+1)\cdot K^T(k+1)$$

The SALR-GVFF-RLS algorithm can be summarized as follows:

$$K(k+1) = \frac{P(k)\cdot\varphi(k+1)}{\lambda(k) + \varphi^T(k+1)\cdot P(k)\cdot\varphi(k+1)}$$

$$P(k+1) = \lambda^{-1}(k)\cdot[P(k) - K(k+1)\cdot\varphi^T(k+1)\cdot P(k)]$$

$$\varepsilon(k+1) = y(k+1) - \varphi^T(k+1)\cdot\hat{\theta}(k)$$

$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1)\cdot\varepsilon(k+1)$$

$$\lambda(k+1) = [\lambda(k) + \eta(k)\cdot\psi^T(k)\cdot\varphi(k+1)\cdot\varepsilon(k+1)]_{\lambda-}^{\lambda+}$$

$$\eta(k+1) = \eta(k) + \alpha\cdot\langle\nabla_\lambda(k), \nabla_\lambda(k+1)\rangle$$

$$S(k+1) = \lambda^{-1}(k+1)\cdot(I - K(k+1)\cdot\varphi^T(k+1))\cdot S(k)\cdot (I - \varphi(k+1)\cdot K^T(k+1)) - \lambda^{-1}(k+1)\cdot P(k+1) + \lambda^{-1}(k+1)\cdot K(k+1)\cdot K^T(k+1)$$

$$\psi(k+1) = (I - K(k+1)\cdot\varphi^T(k+1))\cdot\psi(k) + S(k+1)\cdot\varphi(k+1)\cdot\varepsilon(k+1)$$

In this algorithm, $\lambda(k)$ needs to be in the range of $0 < \lambda(k) \le 1$. The bracket followed by the floor λ_- and the ceiling λ_+ in the forgetting factor relation describes the truncation that restricts the forgetting factor to the interval $[\lambda_-, \lambda_+]$.

IV. Experimental Data

It is recognized in oilfield industry that the main cause of bit balling dysfunction is due to inappropriate drilling parameters (TABLEII) compared with rock properties and drill bit design. It has been intentionally chosen drilling parameters that cause bit balling dysfunction, in order to understand its effect on real data. These measurements stand for the first step for interpretation of bit balling behavior.

TABLE II
PARAMETERS DURING DRILLING TEST

Symbol	Quantity	Value and unity
Q	Flow rate	5 gpm
v	Rate of penetration	5 m/h
Ω	Bit angular velocity	300 rpm
UCS	Rock property	35 Mpa

To understand how bit balling dysfunction can interfere with the process operation, measurements of E, S and rate of penetration (ROP) over time are given in Fig. 3. Before sample number 1500, ROP was adjusted continuously to find a set of points that allows constructing a slope. It can be noticed that during this time interval S is almost constant. It is due to the fact that dysfunction has not yet occurred. In this case, the mean value of the strength comes from rock hardness. In spite of rock homogeneity and a constant ROP in the sample interval between [2450 5000] S and E increase. Their increase comes from T and W and is due to the change in the friction coefficient μ or/and the bit shape factor γ. These changes reveal a dysfunction on the bit and correspond to the bit balling. However, in real drilling conditions it is not a straightforward task to diagnose this kind of faults by looking at the progress in measurements. It is mainly caused by the similar behavior viewed on S and E either in case of changeover from slight rock to hard rock or the bit balling dysfunction. For this reason, we have developed an on-line identification method to determine the slope which characterizes E as a function of S and consequently detecting bit balling.

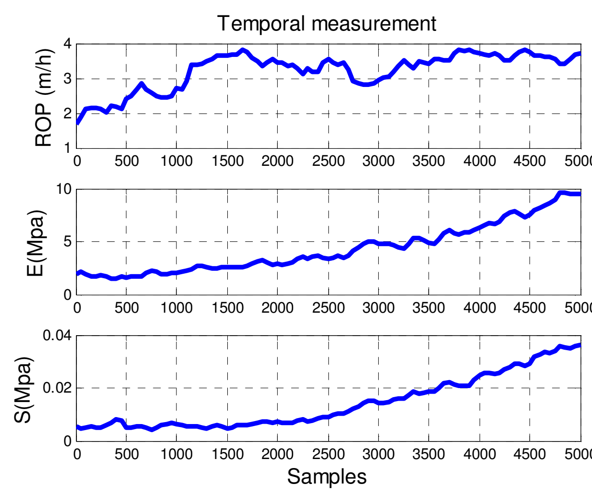

Fig. 3. Behavior of measurements in case of bit balling

A. *Methodology adopted*

An early treatment of data was accomplished before to start identification method. This treatment includes elimination of missing values. This operation is made by fixing a threshold and comparing it to the data received at each time step. After this, the data are filtered by a low pass filter to reduce the noise level. Then, the identification method is applied to data to characterize the slope of E as a function of S. At last, in order to generate automatically a fault indicator statistic Wald's tests were applied to the slope. The basic Wald's tests [2] were chosen owing to its effectiveness. Moreover, Wald's tests are computationally cheap, very easy to implement and robust to modeling errors. The previous steps are illustrated in Fig. 4. To start identification method many initializations are needed.

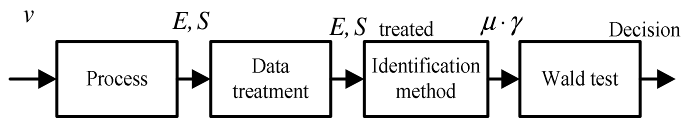

Fig. 4. Monitoring method for bit balling detection

This include initial estimate of states, initial covariance matrix, initial forgetting factor, initial learning rate and initial meta-learning rate. In this application, initialization of the forgetting factor is chosen according to the bit balling progress. Correlating the initial value of the forgetting factor with the progress of bit balling behavior is useful in order to enable bit balling detection in the sample interval from which its reversibility is conceivable by tacking appropriate corrective action. After several experimentations, it is interesting to point out, that the tuning parameters for the identification method developed here, depend directly on drilling parameters. The best choice of parameters at the initialization is the following:

TABLE III
PARAMETERS AT THE INITIALIZATION

Symbol	Value	Signification
$\hat{\theta}_i(0)$	0	Parameter vector
$\psi(0)$	0	Derivative parameter vector with respect to λ
P	$200 \cdot I_2$	Covariance matrix
$\lambda(1)$	0.97	Forgetting factor
$\eta(1)$	10^{-5}	Learning rate
α	10^{-9}	Meta learning rate

B. *Identified slopes*

To emphasize the performances of the proposed algorithm a comparison of its tracking capability with both the GVFF-RLS and the FF-RLS is achieved. Slopes obtained from the FF-RLS, the GVFF-RLS and the SALR-GVFF-RLS algorithms are shown in Fig. 5. In this figure and at the initialization, the parameter vector, the covariance matrix and the forgetting factor have the same values in all algorithms. Note that, in this figure, superior tracking performances are obtained by the SALR-GVFF-RLS algorithm. However, at the starting up to 1000 samples, due to insufficient data enabling slopes to have a regular behavior, a high level of fluctuations is present. In other words, this samples interval is required before to obtain a good slope behavior. Here, whatever the developed method these fluctuations are presents. Then, a transient regime happens and is the first step toward a good identification method. At the sample 2450 the slope starts to decrease, this decrease corresponds to the bit balling occurrence. It is shown in this figure that the proposed algorithm yields good

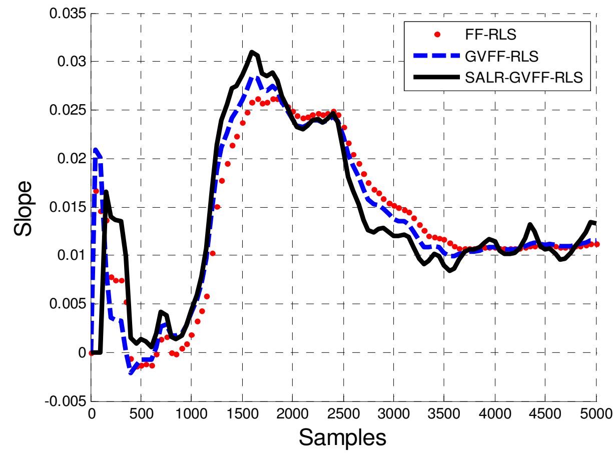

Fig. 5. Slopes progresses

results during transient stage owing to its fast convergence. For instance, during the samples interval between [2450 3500] the change from the safety operation to the bit balling dysfunction is achieved more quickly by the SALR-GVFF-RLS. Note that booth algorithms converge toward a same value. In the Fig. 6 the forgetting factor obtained by the GVFF-RLS and the SALR-GVFF-RLS are plotted. From this figure, the SALR-GVFF-RLS exhibits a faster descent. In addition, it is useful to point out that during the transient stage the adjustment of the forgetting factor is made automatically. However, at a certain stage, it could lead to several false alarms because of its low value. Then, it has been necessary to detect the end of the change in the forgetting factor descent and to restore it to its initial value. This detection is made by fixing a threshold and comparing it to the slope generated during the forgetting factor descent. One of the possible consequences of the forgetting factor reestablishment is the decrease of fluctuations generated in the slope because of the low value of the forgetting factor outside the transient domain. During this study several data base have been tested. In all cases, the better tracking is achieved by the SALR-GVFF-RLS.

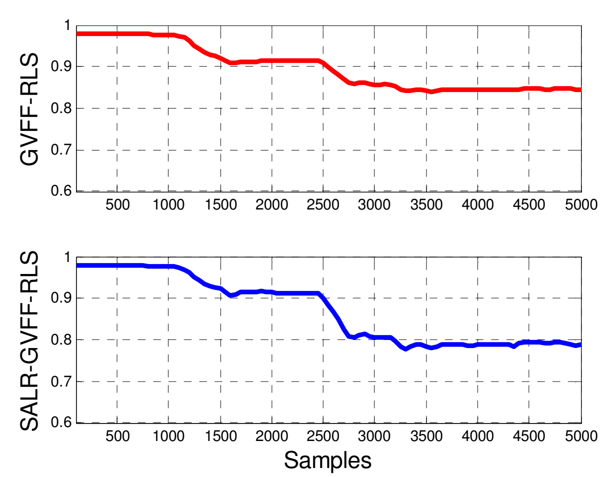

Fig. 6. Forgetting factor progress

With the proposed method, the most sensitive tuning parameters are both learning rate and meta-learning rate. To perform a good tracking ability by applying our algorithm to experimental data, their values at the initialization need to be in the range of $10^{-10} \le \alpha \le 10^{-8}$ and $10^{-6} \le \eta(1) \le 10^{-4}$. The same values are used in all cases. In this study the sampling frequency was equal to 50 Hz.

C. Decision to be taken

To improve the reliability on decision to be taken a classic Wald's tests [2] were applied to data. Its parameters are determined while the system was in safety operation. It is started once the data are sufficient and the slope having a regular behavior. In this case, it proceeds from sample number 1000. In the Fig. 7. we present results obtained from Wald's tests.

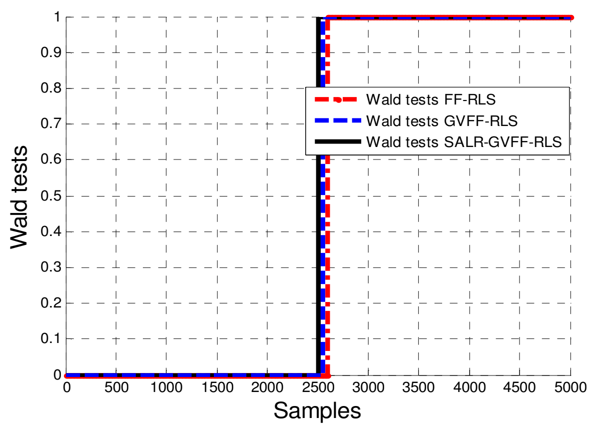

Fig. 7. Wald's tests results

It have been achieved the first, the second and the last detection respectively from slopes obtained by the SALR-GVFF-RLS, the GVFF-RLS and the FF-RLS. The first detection is achieved at the sample number 2500. Then, the second detection at the sample number 2555 and the last detection at the sample number 2600. The TABLEIV shows the results obtained by testing four data base and the number of samples necessary from which the detection is made by Wald's tests. Here, there is no false alarm, because the Wald's test is fairly robust to modeling errors and uncertainties. In this study, the late in detection is around 50 samples. It corresponds to the necessary time before to take a decision. Through the results reported here and other tests investigated, we have shown that with our methodology there is a gain higher than 50 samples, which is a no negligible result in oilfield area.

TABLE IV
WALD'S TESTS RESULT

	SALR-GVFF-RLS	GVFF-RLS	FF-RLS
Data base1	2500	2555	2600
Data base2	1830	1882	1937
Data base3	1790	1842	1891
Data base4	2785	2838	2900

V. CONCLUSION

In oilfield industry, drilling processes monitoring is an important issue. In this paper, an identification method with its capabilities to track parameters change by using a new strategy to adjust the forgetting factor according to the SALR-GVFF-RLS has been developed. This method shows good results regarding bit balling detection. Compared with both the GVFF-RLS and the FF-RLS it exhibits better performances in terms of tracking ability during transient stage. The demonstration of the feasibility of on-line bit balling detection has been carried out by using several real data obtained from tests campaign. With the developed method a gain of at least 50 samples is achieved. Moreover, this methodology remains valuable for diagnosis of others kinds of dysfunctions that might occur on the bit. This study stands for the first step in the further development and improvement of drilling processes optimization. The requirement to diagnose bit balling when it is incipient is useful concerning an appropriate corrective action to take and that brings the system to an optimal drilling condition. For instance, it would enable to avoid lost of times, by cleaning the drill bit more quickly. The next step of this study is to develop a fault tolerant control algorithm in order to automatically adjust the input of the drilling process in case of bit balling detection.

ACKNOWLEDGMENT

The authors would like to thank the Schlumberger Riboud Product centre for supporting this work. Special thanks are extended to the two anonymous reviewers for their comments helping us to improve the quality of the paper.

REFERENCES

[1] L. Ljung, *System Identification*: Theory for the User (2nd). Prentice-Hall PTR:Upper Sadle River, NJ. 1999.

[2] M. Basseville and I. V, Nikiforov, *Detection of abrupt changes*: Theory and application. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[3] G. Nygaard and G. Naevdal, "Nonlinear model predictive control scheme for stabilizing annulus pressure during oil well," *Journal o f Process Control*, vol. 16, pp. 719–732, Jannuary 2006.

[4] C. Cheatham and J. Nahm, "Bit balling in water-reactive shale during full-scang rate test," *Society of Petrolium and Engineers*, Texas, February, 1990.

[5] T. Richard and E. Detournay, "Influence of bit rock interaction on stick slip vibrations of PDC bits," *Society of Petroleum and Engineers*, Texas, October, 2002.

[6] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification-A survey," *Automatica*, 9, pp. 151-162, 1990.

[7] F. Gustafson, *Adaptive filtering and Change Detection*. Willey: WestSussex, England, 2000.

[8] M. B. Malik, "State space RLS with adaptive memory," *Signal Processing*, vol. 86, pp. 1365–1374, March 2006.

[9] L. B. Almeida, T. Langlois, J. D. Amaral, and A. Plankhov, "Parameter adaptation in Stochastic Optimization," *In On-line Learning in Neural Networks*, Cambridge University Press, pp. 111-134, 1998.

[10] J. Nocedal and S.J. Wright, *Numerical optimization*: Springer 1999.