An example of failure detection:design and comparative study of some algorithms

Michèle Basseville, Albert Benveniste

IRIA

CENTRE DE RENNES

IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: 954 90 20

Rapports de Recherche

N° 73

AN EXAMPLE OF FAILURE DETECTION: DESIGN AND COMPARATIVE STUDY OF SOME ALGORITHMS

Michèle BASSEVILLE
Alain BENVENISTE

Mai 1981

IRIA

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél. 954 90 20

Rapports de Recherche

N° 73

AN EXAMPLE OF FAILURE DETECTION: DESIGN AND COMPARATIVE STUDY OF SOME ALGORITHMS

Michèle BASSEVILLE
Alain BENVENISTE

Mai 1981

AN EXAMPLE OF FAILURE DETECTION:

DESIGN AND COMPARATIVE STUDY OF SOME ALGORITHMS

Michèle BASSEVILLE
IRISA/CNRS

Alain BENVENISTE
IRISA/INRIA

Résumé :

Dans cet article, on analyse le comportement de plusieurs algorithmes de détection de ruptures lorsqu'on les applique aux mêmes données réelles (signaux géophysiques), et on compare ces algorithmes selon les points de vue suivants : complexité, efficacité, robustesse, aptitude à caractériser les ruptures détectées.

Trois types d'algorithmes sont étudiés : détecteurs de type "dérivée filtrée", tests basés sur des sommes cumulées ("cusum"), test du rapport de vraisemblance généralisé (TRVG) de Willsky. Une version modifiée de ce dernier test est élaborée, et un nouveau détecteur, mélangeant les tests TRVG et "cusum", est présenté.

Abstract:

The purpose of this paper is to analyze the behaviour of several failure detection algorithms when applied to the same real data (geophysical signals) and to compare these algorithms from different points of view: complexity, efficiency, robustness, ability to characterize the detected failures. Three types of algorithms are investigated: "filtered derivatives" detectors, cumulative sum ("cusum") tests, Willsky's generalized likelihood ratio (GLR) algorithm. A modified version of this last test is elaborated, and a new detector, mixing GLR and cusum tests, is presented.

AN EXAMPLE OF FAILURE DETECTION:

DESIGN AND COMPARATIVE STUDY OF SOME ALGORITHMS

M. BASSEVILLE, A. BENVENISTE
IRISA

I. INTRODUCTION

The purpose of this paper is to analyze the behaviour of several failure detection algorithms when applied to the same real data (signals of resistivity of the earth) and to compare these algorithms from different points of view: processing time, complexity, efficiency, robustness, ability to characterize the detected failures.

Let us first emphasize that the word "failure" should be interpreted as abrupt changes in one or several parameters of a model representing the observed signals. This model is usually a linear stochastic one, and is designed with the specific purpose of failure detection, and not necessarily in the framework of descriptive modelling. This point will be further developped. As far as the formulation of the failure detection problem is concerned, the interested reader is referred to the survey papers [3] and [14].

The problem to be solved is the detection of abrupt changes in some geophysical signals. The processing time and computing load are not the most crucial issues of the study; therefore, somewhat complicated algorithms have been considered or elaborated in order to obtain robust and accurate detection of the jumps or failures and identification of their main characteristics. However, some more simple algorithms have been studied too, in order to get a new insight on the tradeoff in detector complexity versus performance.

The model, which has been chosen for failure detection in these signals $\{y_k\}_{k>0}$ is simply a noisy slope on the mean, i.e., when no failure is present:

$$(1)\begin{cases} x_{k+1} = \tau\, x_k + \mu_k + w_k^1 \\ \mu_{k+1} = \mu_k + w_k^2 \\ y_k = x_k + e_k \end{cases}$$

or, with $\phi = \begin{pmatrix} \tau & 1 \\ 0 & 1 \end{pmatrix}$, $H = (1 \quad 0)$, $w_k = \begin{pmatrix} w_k^1 \\ w_k^2 \end{pmatrix}$

and the state $X_k = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix}$:

$$(2) \quad \begin{cases} X_{k+1} = \phi X_k + W_k \\ y_k = H X_k + e_k \end{cases},$$

where $\{W_k\}_{k>0}$ and $\{e_k\}_{k>0}$ are two gaussian white noise sequences, stochastically independent of each other, and such that:

$$\begin{cases} E(W_k) = 0 \\ E(W_k W'_k) = Q = \begin{pmatrix} q^1 & 0 \\ 0 & q^2 \end{pmatrix} \\ E(e_k) = 0 \\ E(e_k^2) = \sigma^2 \end{cases}$$

The "stabilization" coefficient τ as well as the variances of the two state noises q^1 and q^2 are supposed to be known (or arbitrarily chosen), while the measurement noise variance σ^2 is unknown and estimated at each step.

This model only takes into account the slope actually present in the observed signals (see *figure n°15*); the state noises allow some adaptation of the model to slow fluctuations. Obviously this simple model is not sufficient to describe the behaviour of the signals, which present, in many parts, a rather complicated structure; but it will be shown later that this model, when coupled to adequate detectors, allows accurate detection of the interesting failures or jumps. These jumps are supposed to be jumps in the mean, that is to say the actual model is:

$$(3) \quad \begin{cases} x_{k+1} = \tau x_k + \mu_k + w_k^1 + \nu \, \delta_{\theta,k+1} \\ \mu_{k+1} = \quad k + w_k^2 \\ y_k = x_k + e_k \end{cases}$$

where θ is the instant of jump and ν the "amplitude" of the jump.

The problem is then the following:

i) give the alarm as soon as possible after an actual θ, but don't give raise to too many false alarms;

ii) estimate the amplitude ν of the jump;

iii) reactualize the observation procedure according to the informations so collected by the detector.

In section II, the generalized likelihood ratio (GLR) algorithm of A.S. WILLSKY and H.L. JONES ([15], [16]) will be derived in detail. It will be shown that some drawbacks of this method, such as high sensitivity to the designed threshold and window length, can be overcome, and a new version of this algorithm will be presented. This detector is more robust and easier to handle for the practical user, and allow some hierarchical classification of the detected failures.

In section III, some simpler algorithms are considered. First, the classical "filtered derivatives" algorithm are shown to be not very efficient; second the Hinkley's or Page's algorithm ([11], [13]), which has been shown to be more robust and efficient ([2]), is presented with several modifications concerning the design of the threshold. Finally, a new algorithm, mixing Willsky's and Hinkley's detectors, is elaborated in order to reduce computing time while keeping the efficiency of the GLR algorithm.

Detailed comparisons between the different detectors and general conclusions are given in section IV.

II. GENERALIZED LIKELIHOOD RATIO (GLR) ALGORITHMS.

II.1. The original test.

This test is a sequential probability ratio test of the hypothesis H_0: *"no failure is present, i.e. all the observations* y_i *until time* k *are governed by the same probability law* P_o*"* against the hypothesis H_1: *" a failure occured before time* k*, i.e. the observations are governed by the law* P_o *until time* $\theta-1$*, and by the law* P_1 *from time* θ *to time* k*"*. But the failure time θ and the "amplitude" of the jump, being unknown, are replaced by their maximum likelihood estimates.

This test, which is usual in statistics, was first introduced for failure detection in space applications, and was derived in its most general form by A.S. WILLSKY and H.L. JONES ([15], [16]), whose presentation will be followed in this section. The GLR algorithm has already been described in the survey paper [3], but another description is given here in order to derive the new version of the algorithm which we propose. Some applications of the GLR detector should be mentioned in various areas: digital aircraft control ([5], [6], [7]), electrocardiography ([9], [10]), freeways supervision [16].

Consider a system whose possibly disturbed behaviour can be modeled by the following equations:

$$(4)\quad \begin{cases} x_{k+1} = \phi(k+1,k)\, x_k + \Gamma_k\, w_k + \delta_{\theta,k+1}\,\nu \\ y_{k+1} = H_{k+1}\, x_{k+1} + e_{k+1} \end{cases}$$

where the state x_k belongs to $\mathbb{R}^n$, the observation y_k belongs to $\mathbb{R}^p$, and $\{w_k\}_k$ and $\{v_k\}_k$ are two white gaussian sequences, stochastically independent of each other, with respective covariance matrices Q_k and R_k definite positive. θ is the instant of jump and ν is its "amplitude". The GLR algorithm is essentially intended for detecting failures which have a "linear" effect on the system and can be modeled as jumps or ramp in the state or in the observed variables. It seems to be poorly suited to detect failures in the dynamics of the system (changes in the parameters of ϕ, Γ, H), because, by inserting these failures in an augmented state space for example, one looses

the linearity of the model, which is the key point in the derivation of the algorithm (additive decomposition of the innovation - see further on). The main interest of this method lies first in the recursive manner in which the estimates of the failure time θ, the amplitude ν and the generalized likelihood ratio can be computed, as it will be shown, and, second, in its ability to give a compensation scheme for updating the filter estimates after a system change has been detected.

Under the null hypothesis H_0 ($\theta=\infty$), the Kalman filter, corresponding to the model (4), is described by the following equations:

$$(5) \quad \begin{cases} \hat{x}_{k+1|k} = \phi(k+1,k)\ \hat{x}_{k|k} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \gamma_k \end{cases}$$

where γ_k is the innovation:

$$(6) \quad \gamma_k = y_k - H_k\ \hat{x}_{k|k-1},$$

and where the gain K_k, the error covariance $P_{k|k}$ and the covariance V_k of the innovation are given by:

$$(7) \quad \begin{cases} K_k = P_{k|k-1}\ H'_k\ V_k^{-1} \\ P_{k+1|k} = \phi(k+1,k)\ P_{k|k}\ \phi'(k+1,k) + \Gamma_k\ Q_k\ \Gamma'_k \\ P_{k|k} = (I - K_k\ H_k)\ P_{k|k-1} \\ V_k = H_k\ P_{k|k-1}\ H'_k + R_k. \end{cases}$$

Now, x_k, $\hat{x}_{k|k}$, y_k and γ_k can be expressed in a form that explicitly involves θ and ν. Indeed because of the linear effect of the failure, we can write:

$$(8) \quad \begin{aligned} x_k &= x_k^1 + x_k^2 & \hat{x}_{k|k} &= \hat{x}^1_{k|k} + \hat{x}^2_{k|k} \\ y_k &= y_k^1 + y_k^2 & \gamma_k &= \gamma_k^1 + \gamma_k^2 \end{aligned}$$

where the uppperscript 1 indicates all the effects but θ and ν, and the upperscript 2 indicates <u>only</u> the θ and ν effects. From the filter equations, we yield:

$$(9)\quad \begin{cases} x_k^2 = \phi(k,\theta)\ \nu \\ y_k^2 = H_k\ \phi(k,\theta)\ \nu \\ \hat{x}_{k|k}^2 = F(k,\theta)\ \nu \\ \gamma_k^2 = G(k,\theta)\ \nu \end{cases}$$

where:

$$(10)\quad \begin{cases} F(k,\theta) = K_k\ G(k,\theta) + \phi(k,k-1)\ F(k-1,\theta) \\ G(k,\theta) = H_k[\ \phi(k,\theta) - \phi(k,k-1)\ F(k-1,\theta)] \\ G(k,k) = H_k \end{cases}$$

(For $k<\theta$, ϕ, F, and G are obviously zero).
G is called the signature of the failure.

Therefore, the two hypothesis to be tested are:

$$(11)\quad \begin{aligned} H_0&:\ \gamma_k = \gamma_k^1 \\ H_1&:\ \gamma_k = \gamma_k^1 + G(k,\theta)\ \nu \end{aligned}$$

where γ_k^1 is a white gaussian sequence, with covariance matrix V_k. The generalized likelihood ratio test computes the likelihood ratio:

$$(12)\quad \Lambda_k = \frac{L(\gamma_1,\ldots,\gamma_k \mid H_1,\ \theta=\hat{\theta}_k,\ \nu=\hat{\nu}_k)}{L(\gamma_1,\ldots,\gamma_k \mid H_0)}$$

where $\hat{\theta}_k$ and $\hat{\nu}_k$ are the maximum likelihood estimates of θ and ν under the hypothesis H_1:

$$(\hat{\theta}_k,\hat{\nu}_k) = \arg\max_{(\tilde{\theta},\tilde{\nu})} L(\gamma_1,\ldots,\gamma_k \mid H_1,\ \theta=\tilde{\theta},\ \nu=\tilde{\nu}).$$

Under gaussian asssumptions, the log-likelihood ratio is:

$$(13)\quad \begin{aligned} \ell(k,\hat{\theta}_k) &= 2\ \mathrm{Log}\ \Lambda_k \\ &= \sum_{j=1}^{k} \gamma'_j\ V_j^{-1}\ \gamma_j - \sum_{j=1}^{k} (\gamma_j - G(j,\hat{\theta}_k)\hat{\nu}_k)'\ V_j^{-1} (\gamma_j - G(j,\hat{\theta}_k)\hat{\nu}_k) \end{aligned}$$

and $\hat{\nu}_k$ minimizes:

$$f(y) = \sum_{j=1}^{k} (\gamma_j - G(j,\hat{\theta}_k)y)'\ V_j^{-1} (\gamma_j - G(j,\hat{\theta}_k)y).$$

so:

$$(14) \quad \hat{\nu}_k = C^{-1}(k,\hat{\theta}_k)\, d(k,\hat{\theta}_k)$$

where:

$$C(k,\theta) = \sum_{j=\theta}^{k} G'(j,\theta)\, V_j^{-1}\, G(j,\theta)$$

and: (15)

$$d(k,\theta) = \sum_{j=\theta}^{k} G'(j,\theta)\, V_j^{-1}\, \gamma_j .$$

Besides, $\hat{\theta}_k$ is the time between 1 and k which maximizes the log-likelihood ratio:

$$(16) \quad \ell(k,\tilde{\theta}) = d'(k,\tilde{\theta})\, C^{-1}(k,\tilde{\theta})\, d(k,\tilde{\theta}).$$

The decision rule is as follows:

$$(17) \quad \ell(k,\hat{\theta}_k) \underset{H_0}{\overset{H_1}{\gtrless}} \varepsilon$$

The first drawback of this method is the increasing length of the filters bank it requires (searching for $\hat{\theta}_k$ between 1 and k). In practice, the search is constrained to a finite window of length M: $k-M < \tilde{\theta} \leq k$. (In fact, $k-M < \tilde{\theta} \leq k-n+1$, where n is the state space dimension).

For the particular model that we consider here:

$$(18) \quad \begin{cases} X_{k+1} = \phi\, X_k + W_k + \begin{pmatrix} \nu \\ 0 \end{pmatrix} \delta_{k+1,\theta} \\ y_k = H\, X_k + e_k \end{cases}$$

where: $X_k = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix}$, $\phi = \begin{pmatrix} \tau & 1 \\ 0 & 1 \end{pmatrix}$, and $H = (1 \quad 0)$,

$$E(e_k^2) = \sigma^2, \qquad E(W_k\, W'_k) = \begin{pmatrix} q^1 & 0 \\ 0 & q^2 \end{pmatrix},$$

the gain $K_k = \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}$ and the error covariance $P_{k|k} = \begin{pmatrix} a_k & b_k \\ b_k & c_k \end{pmatrix}$ are given by:

$$
(19)\quad \begin{cases}
\alpha_{k+1} = \dfrac{v_k}{\hat{\sigma}^2_{k+1} + v_k} \\
\beta_{k+1} = \dfrac{\tau b_k + c_k}{\hat{\sigma}^2_{k+1} + v_k} \\
a_{k+1} = (1 - \alpha_{k+1})\, v_k \\
b_{k+1} = -\beta_{k+1}\, v_k + \tau b_k + c_k \\
\qquad = (1 - \alpha_{k+1})\,(\tau b_k + c_k) \\
c_{k+1} = -\beta_{k+1}\,(\tau b_k + c_k) + c_k + q_k^2
\end{cases}
$$

where : $v_k = \tau^2 a_k + 2\tau b_k + c_k + q_k^1$.

The estimates have the following form:

$$
(20)\quad \begin{cases}
\hat{x}_{k+1} = \hat{x}_{k+1|k+1} = \tau \hat{x}_k + \hat{\mu}_k + \alpha_{k+1}\, \gamma_{k+1} \\
\hat{\mu}_{k+1} = \hat{\mu}_{k+1|k+1} = \hat{\mu}_k + \beta_{k+1}\, \gamma_{k+1}
\end{cases}
$$

where:

$$
(21)\quad \gamma_{k+1} = y_{k+1} - \tau\, \hat{x}_k - \hat{\mu}_k,
$$

and:

$$
(22)\quad \hat{\sigma}^2_{k+1} = \frac{k-2}{k-1}\, \hat{\sigma}^2_k\, 1_{\{k>2\}} + \frac{k}{(k+1)(k+2)}\, (y_{k+1} - \tau \hat{x}_k)^2.
$$

The weighting coefficients in (22) are chosen in order to obtain an unbiased estimate with exponential forgetting of past observations. But the key point in this estimate is the presence of the "delayed approximation error" $y_{k+1} - \tau\hat{x}_k$. The choice of $y_{k+1} - \tau\hat{x}_k$, instead of $y_k - \tau\hat{x}_k$ or even $y_{k+1} - \tau\hat{x}_k - \hat{\mu}_k$, is intended to contend with the undesirable under estimation of the variance σ^2 of the measurement noise. (See also [12] for another approach of this problem. But R.K. MEHRA's approach seems to involve too many computations). An extra reason for using the estimate (22) is the need for a large variance σ^2 in the highly noise areas (see *fig. n° 1* or *10* for example) where the hypothesis "white measurement noise" is clearly unrealistic

(notice that 2000 sample points are present). The estimate (22) leads to increase σ^2 in the serrate areas, by including the effect of the local slope in the measurement noise.

The signature and test functions are as follows:

let be: $F(k,\theta) = \begin{pmatrix} f^1(k,\theta) \\ f^2(k,\theta) \end{pmatrix}$. Then, for $k-M+1 < \theta < k+1$:

$$(23)\begin{cases} f^1(k+1,\theta) = (1-\alpha_{k+1})\,[\tau\, f^1(k,\theta) + f^2(k,\theta)] + \tau^{k-\theta+1}\,\alpha_{k+1} \\ f^2(k+1,\theta) = -\beta_{k+1}\,[\tau\, f^1(k,\theta) + f^2(k,\theta)] + f^2(k,\theta) + \tau^{k+1-\theta}\,\beta_{k+1} \\ g(k+1,\theta) = \tau^{k+1-\theta} - [\tau\, f^1(k,\theta) + f^2(k,\theta)] \end{cases}$$

and:

$$(23b)\begin{cases} f^1(k+1,k+1) = \alpha_{k+1} \\ f^2(k+1,k+1) = \beta_{k+1} \end{cases}$$

For the test function: $\ell(k,\theta) = d'(k,\theta)\, C^{-1}(k,\theta)\, d(k,\theta)$, we have in this case:

$$(24)\quad \ell(k,\hat{\theta}_k) = \frac{[d(k,\hat{\theta}_k)]^2}{C(k,\hat{\theta}_k)},$$

where the functions C and d are given by:

$$(25)\begin{cases} C(k+1,\theta) = C(k,\theta) + \dfrac{[g(k+1,\theta)]^2}{\hat{\sigma}^2_{k+1} + v_k} \\ d(k+1,\theta) = d(k,\theta) + \dfrac{g(k+1,\theta)}{\hat{\sigma}^2_{k+1} + v_k}\,\gamma_{k+1} \end{cases}$$

$(k-M+1 < \theta < k+1)$ and:

$$(25b)\begin{cases} C(k+1,k+1) = \dfrac{1}{\hat{\sigma}^2_{k+1} + v_k} \\ d(k+1,k+1) = \dfrac{1}{\hat{\sigma}^2_{k+1} + v_k}\,\gamma_{k+1} \end{cases}$$

Finally:

$$(26)\quad \hat{v}_k = \frac{d(k,\hat{\theta}_k)}{C(k,\hat{\theta}_k)}.$$

The main drawback of this algorithm lies in the difficulty to choose the window length M and the threshold ε. By a simulation study as well as on real data, we observed that, in this special case, the log-likelihood ratio $\ell(k,k-i)$ is a linearly increasing function of i, and therefore ε should linearly increase with M. This fact has been corroborated by some theoretical computations (see the Appendix). Furthermore, for noisy observations, the number of false alarms and missed detections is highly depending upon the threshold ε. (see *figures n° 1, 2*).

However, the systematic investigation of the behaviour of the log-likelihood ratio ℓ and the estimate $\hat{\nu}$ shows that the failure time θ is correctly estimated (even if the jump occurs in more than one step), and that the estimate $\hat{\nu}_k = \hat{\nu}(\hat{\theta}_k)$ is a good estimate of the amplitude of the failure. As it shall be seen later, this properties of the failure estimates have been explicitly used for deriving a new version of the algorithm.

Before doing this, let us just mention some other studies of the algorithm and describe the compensation scheme for the filter, as it was introduced by A.S. WILLSKY and H.L. JONES ([15], [16]).

A simplified version of the algorithm, where the maximization on θ is cancelled, has been studied by E.Y. CHOW [7]: in this case, the estimate $\hat{\theta}_k$ is constrained to the value k-M+1.

Besides, the problem of the existence of $C^{-1}(k,\theta)$, which is the covariance matrix of the estimate $\hat{\nu}_k$, can be rather difficult for non observable systems and has been studied by C.B. CHANG and K.P. DUNN [6] who proposed a new algorithm in which the inversion of C is avoided and replaced by the inversion of a matrix, the dimension of which is generally less than that of C.

The sensitivity of the algorithm to modelling errors has been studied in [4]. It seems that an oscillatory behaviour of ℓ and a jump in the estimate $\hat{\theta}$ are characteristics of model errors.

The compensation of the filter after detection of a failure works as follows (see [15]): let the subscript n indicate "new" estimates and the subscript o indicate "old" estimates, then one natural choice of the new estimates is:

$$
\begin{aligned}
(\hat{x}_{k|k})_n &= \hat{x}^1_{k|k} + x^2_k \\
&= (\hat{x}_{k|k})_o + x^2_k - \hat{x}^2_{k|k} \qquad (28) \\
&= (\hat{x}_{k|k})_o + [\phi(k,\hat{\theta}_k) - F(k,\hat{\theta}_k)]\ \hat{\nu}_k
\end{aligned}
$$

and for the error covariance matrix:

$$
(P_{k|k})_n = (P_{k|k})_o + [\phi(k,\hat{\theta}_k)-F(k,\hat{\theta}_k)]\ C^{-1}(k,\hat{\theta}_k)\ [\phi(k,\hat{\theta}_k)-F(k,\hat{\theta}_k)]' . \qquad (29)
$$

In order to avoid successive false alarms after a failure detection, the test is reinitialized, i.e. the innovations are reset to zero, and the error covariance is increased essentially because the estimate $\hat{\nu}_k$ is not always accurate and can lead to instabilities in the algorithm.

For our particular model, we shall see later that this reactualization procedure has to be modified in order to take into account the fact that the slope of the observations is scarcely the same before and after a jump in mean.

II.2. A new version of the GLR algorithm.

In fact, this new algorithm is not a GLR test but uses all the informations given by this test as it was introduced by A.S. WILLSKY. It has been derived in order to obtain a more robust detector, and to allow an easier adjustment by the practical user.

We use the estimate $\hat{\nu}_k = \hat{\nu}(\hat{\theta}_k)$, which has been seen to be a rather good one, and decide that a jump has occured when this estimated amplitude

exceeds (in square norm for example) an a priori fixed minimum amplitude ν_m by a quantity which is of course a function of the variance of $\hat{\nu}_k$. In other words, we test "$\nu \geq \nu_m$" against "$\nu < \nu_m$" by a likelihood ratio technique. The matrix $C^{-1}(k,\hat{\theta}_k)$ is the covariance matrix of $\hat{\nu}_k$; but, because this fact is true only under the hypothesis H_1 and under the hypothesis that $\hat{\theta}_k$ is an unbiased estimate of θ, and because of the frequent multistep jumps that occur in practice, we prefer to smooth the estimate $\hat{\nu}_k$ in a finite window and estimate the covariance matrix S_k of all the estimates $\hat{\nu}_k$ present in this window. The length p of this window is obviously taken to be less than the length M of the window introduced in Willsky's algorithm.

Our test is then the following:

$$(30) \quad N_k(\bar{\nu}_k-\nu_m)' \; S_k^{-1} \; (\bar{\nu}_k-\nu_m) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda,$$

where $N_k = p \cdot 1_{\{k>p+1\}} + (k-1) \cdot 1_{\{k\leq p+1\}}$, and where the threshold λ and the minimum amplitude of jump to be detected ν_m have to be chosen. Up to now, we obviously assumed that ν_m and $\bar{\nu}_k$ are positive. In case where $\bar{\nu}_k$ is negative, we replace it by $-\bar{\nu}_k$ in the preceeding test statistic.

For the special slope model we consider, this test becomes:

$$(31) \quad \frac{N_k(\pm\, \bar{\nu}_k-\nu_m)^2}{S_k} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

where:

$$(32)\begin{cases} \bar{\nu}_k = \frac{1}{p} \sum_{i=0}^{p-1} \hat{\nu}_{k-i} = \bar{\nu}_{k-1} + \frac{\hat{\nu}_k-\hat{\nu}_{k-p}}{p} \\ S_k = \frac{1}{p-1} \sum_{i=0}^{p-1} (\hat{\nu}_{k-i}-\bar{\nu}_k)^2 \\ \quad = S_{k-1} + \frac{\hat{\nu}_k-\hat{\nu}_{k-p}}{p} \left[\frac{p+1}{p-1}(\hat{\nu}_k-\bar{\nu}_k) + \hat{\nu}_{k-p} + \bar{\nu}_k\right] \end{cases}$$

if $k > p+1$, and:

$$(33)\quad \begin{cases} \bar{\nu}_k = \dfrac{1}{k-1} \displaystyle\sum_{i=0}^{k-2} \hat{\nu}_{k-i} = \dfrac{k-1}{k}\,\bar{\nu}_{k-1} + \dfrac{1}{k-1}\,\hat{\nu}_k \\ S_k = \dfrac{1}{k-2} \displaystyle\sum_{i=0}^{k-2} (\hat{\nu}_{k-i} - \bar{\nu}_k)^2 \\ \quad = \dfrac{k-3}{k-2}\, S_{k-1} + \dfrac{k-1}{(k-2)^2} (\hat{\nu}_k - \bar{\nu}_k)^2 \end{cases}$$

if $2 \leq k \leq p+1$.

The minimum amplitude ν_m and the threshold λ, which measures a kind of deviation between the estimated and the expected jump amplitudes, are clearly more meaningful for the user than the threshold ε on the log-likelihood ratio ℓ. Furthermore, it quickly appeared, during application of the two algorithms to real data, that the modified one is less sensitive to the choice of the window size M and of the threshold (ε or λ) and that these two choices are less dependent upon each other. For example, for M=20, $\tau=q^1=1$, $q^2=0$, changing ε from 30 to 40 may heavily affect the number of alarms obtained with Willsky's previous GLR algorithm (see *figures n° 1 and 2*), whereas for the same parameters and p=15 and $\nu_m=40$, changing λ from 20 to 100 is less discernible for our modified version. See *figures n° 3, 4.*

In order to understand the figures, it should be noticed that the original signal is always the upper one, the lower being the filtered and segmented one. The vertical lines indicate the detected "failures". Except for the two last figures (*n° 20 and 21*) where the processing is reversed, for all the figures the signal is processed from right to left. Each figure corresponds to 2000 sample points.

Let us conclude this paragraph by some comments concerning the adjustment of the different parameters in our slope model: τ, q^1, q^2, and in the GLR detectors: (M,ε) or (M,p,ν_m,λ). Parameter of "stabilization" τ appeared to be of poor interest within the scope of failure detection and so was cancelled ($\tau=1$). On the contrary, a noise on the mean state x_k is necessary because it allows a quicker convergence of the Kalman filter and increases the Kalman gain and thus the forgetting ability. For our 2^8 levels quantized real data, $q^1=1$ seems to be a good value (0.1 is obviously insufficient, and 10 too large).

Taking into account a noise on the slope μ_k should allow a better following of the filter, especially in very noisy areas where our simple model is inadequate, and should be chosen according to the sampling frequency of the data. However, as far as the only aim of the study is the detection and estimation of the failures, it seems that the introduction of such a slope-noise does not improve the accuracy of the GLR detectors: the variations it involves on the estimates $\hat{\theta}_k$ and $\hat{\nu}_k$ are not very important if one thinks to the oversampling of the signals and the difficulty one encounters in setting the true instants at which jumps occur and their amplitudes. Compare *figures n° 5 and 7* with *figures n° 6 and 8*.

Nevertheless, when a first step towards a classification of the detected failures is of interest, the estimates given by the Kalman filter may be key parameters (for example, $\hat{\mu}_k$, $\hat{\sigma}_k^2$), and a slope noise variance q^2 may be useful (of order of magnitude 10^{-5} in this case).

The size M of the window used for searching the estimate $\hat{\theta}_k$ partly depends upon the number of steps in which most of the failures occur, and could not be taken here less than 20. On the other hand, when using M=30 and M=40, we observed that $k-\hat{\theta}_k$ was nearly always between 15 and 20. So this size (M=20) was chosen.

As was explained before, choosing the threshold ε for the first GLR algorithm is a difficult taks; so no "rule of thumb" will be given for it.

Therefore, let us now consider the setting of the parameters p, ν_m, λ involved in the new version. The size p of the window used when smoothing the successive estimates $\hat{\nu}_k$ also obviously depends upon the number of steps necessary for actual jumps; p=15 seemed to be more appropriate than p=10 (for M=20). Setting of ν_m and λ depends upon the size of the jumps one is interested in. In our example, both small and large jumps are interesting and meaningful for the geologist; but, on the other hand, it is not desired to obtain too much false alarms in very noisy areas. Keeping in mind the form of our test statistic, which involves the variance of the smoothed estimate $\bar{\nu}_k$ as a multiplicative component of the real threshold for the difference $|\bar{\nu}_k-\nu_m|$ which is of interest, we decided to associate high thresholds for low minimum amplitudes of jumps, and much lower thresholds for higher minimum amplitudes. For example, (ν_m,λ) = (25,1000) and (40,100) appeared to be adequate values. Furthermore,

it seems that these two choices for the pair (ν_m, λ) allows a first step towards a hierarchy among the detected failures, (for example see *figures n° (5,9), (10,11)*) which is obviously not possible with the first GLR algorithm.

Finally, a comment, concerning the reactualization procedure of the Kalman filter after a detected failure, may help characterization and classification of the alarms given by GLR detectors. To take into account the fact that the slopes before and after a failure are scarcely the same, a useful tool consists in over-initializing the error variance c_k on $\hat{\mu}_k$, and the error covariance b_k on $(\hat{x}_k, \hat{\mu}_k)$, simply by multiplying the reactualized values proposed by A.S. WILLSKY by constant factors.

As a first conclusion, GLR algorithms are powerful tools for failure detection and estimation, essentially because they estimate, at each time, the amplitude of a possible failure, and, on the other hand, because they allow processing of multidimensional signals. But, even for scalar data, they involve a large computing burden, and therefore it is necessary to compare them to less complicated algorithms in order to evaluate which gain is obtained with this increased complexity and to give a guide for solving one of the tradeoffs we mentioned in the Introduction: the tradeoff in complexity versus performance of such detectors.

III. "FILTERED DERIVATIVES" DETECTORS AND CUSUM TESTS.

This section is devoted to the application of some simple failure detection algorithms to the particular problem we presented in the Introduction. Most of the algorithms have already been studied, both via simulation analysis and from a theoretical point of view; the interested reader is referred to [2] for this study, and to [1] for an application of these techniques to edge detection in digitized pictures.

The behaviour and limitations of these algorithms will be analyzed, and a new algorithm, which is a mixing of Willsky's GLR detector and Hinkley's cumulative sum test, will be presented. This algorithm will be shown to be far less time consuming than GLR algorithm, and still efficient and rather robust.

III.1. "Filtered derivatives" detectors.

These simple well-known detectors first filter the analysed signal before comparing its derivative to a threshold λ. The two simple filters considered in [2] were the integrating filter and a filter, called "triangular", the impulse response of which is of the form:

$$(34) \quad a\, t\, 1_{\{0\le t\le \ell\}} - a(2\ell - t)\, 1_{\{\ell\le t\le 2\ell\}}, \quad \text{with } a\ell^2 = 1.$$

After one of these filtering operations and derivation (in discrete time), the obtained observations are:

$$(35) \quad z_n = \frac{y_{n+\ell} - y_{n-\ell}}{2\ell}$$

in the first case, and:

$$(36) \quad z_n = \frac{(y_{n+1}+y_{n+2}+\ldots+y_{n+\ell}) - (y_{n-1}+y_{n-2}+\ldots+y_{n-\ell})}{\ell^2}$$

in the second.

With the aid of these two detectors, we mentioned, in [2], two ways of setting the alarms: the first one, called "rough", consisted in setting the alarm when $|z_n| \geq \lambda$ and forgetting the 2ℓ observations following this instant, in order to avoid successive detections in the neighborhood of a jump. The second method, called "with counter", consisted in counting the number of instants at which the new signal $|z_n|$ exceeds the threshold λ and which occur within the 2ℓ instants following the first overcrossing, and in setting the alarm only if this number of crossings exceeds a fixed value N_c; the instant of alarm was then chosen to be the first of these crossing times. The advantage of such a method is obviously to avoid false alarms, the drawback being the risk of nondetection if N_c is too high. It was shown that $N_c=2$ was a good value, in the simulation analysis of simple jumps in mean as well as in the framework of edge detection in noisy pictures (see [1], [2]).

However, when applied to the real data which we consider in this paper, these algorithms appear to be quite inefficient in very noisy areas, as it could be expected for. Thus a more elaborated type of counter has been used: this new counter takes into account z_n itself, and not only its absolute value, and the alarm is now given when the number of crossings of the same type (i.e. overcrossings of the level $+\lambda$ or undercrossings of the level $-\lambda$) exceeds the threshold N_c during the 2ℓ instants following the first one, which is taken as the estimated jump time.

Some examples of application of this new "filtered derivative" algorithm, with $2\ell=16$, $\lambda=320$, $N_c=4$, can be shown on *figures n° 12 to 15*. It can be seen that this algorithm is not robust enough for our purpose (see *figure n° 13*): two many false alarms occur in very noisy areas.

III.2. Hinkley's cumulative sum test.

This cumulative sum test has been introduced by E.S. PAGE [13] in quality control and studied by D.V. HINKLEY [11] for detecting jumps in mean in a sequence of independent random Gaussian variables. It has been theoretically investigated in [2] for the two cases of jump in mean for Gaussian

observations and jump in the parameter in a sequence of independent Bernoulli random variables; new results concerning mean time between false alarms and mean time delay for detection of a jump (under the hypothesis that no false alarm occurred before the jump) were derived.

Let us recall the principle of this algorithm. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of independent identically distributed random variables, of Gaussian laws with common variance σ^2, and the mean of which changes from $m_o>0$ (known) for $n<\theta$ to $m_1=-m_o$ for $n\geq\theta$, where θ is the unknown instant to be estimated. (In the case of known means before and after the jump, we may always assume by translation that the problem is to detect a change of sign). If the means are unknown, the first mean m_o is estimated and the two detectors, one for an increase in mean, the other for a decrease in mean, are adjusted in order to find jumps, the amplitude of which is supposed to be greater than a minimum amplitude $\nu_m = |m_1-m_o|$ fixed a priori.

For a change of sign, the test works as follows: let $S_o=0$ and $S_n = \sum_{i=1}^{n} Y_i$ $(n\geq 1)$, and observe the current maximum of this integrated signal:

$$(37)\qquad M_n = \max_{0\leq k\leq n} S_k .$$

When $M_n-S_n \geq h$, where $h>0$, the alarm is given. See *figure n° 22*. Recall that the drift of S_n is positive before the jump and negative after.

Contrary to the "filtered derivatives" algorithms, this detector has the advantage of memorizing the entire past of the observed process by the use of (M_n), and not only a fixed-size past as all filtering operations with sliding windows do; nevertheless, it does not require complex calculations because:

$$(38)\qquad M_n-S_n = (M_{n-1}-S_{n-1}-Y_n)\ 1_{\{S_n<M_{n-1}\}} .$$

When the means are unknown, the two Hinkley's detectors which work in parallel are as follows: the first one given by:

$$(39)\quad \begin{cases} S_n = \sum_{i=1}^{n} (Y_i - m_o + \frac{\nu_m}{2}) \qquad (n\geq 1) \qquad (S_o=0) \\ M_n = \max_{0\leq k\leq n} S_k \end{cases}$$

is looking for downwards jumps ($m_o \rightarrow m_o - \nu_m$), and the second one given by:

$$(40)\begin{cases} U_n = \sum_{i=1}^{n} (Y_i - m_o - \frac{\nu_m}{2}) & (n \geq 1) \qquad (U_o = 0) \\ m_n = \min_{0 \leq k \leq n} U_k \end{cases}$$

is looking for upwards jumps ($m_o \rightarrow m_o + \nu_m$). The decision concerning the presence of a jump is taken according to the first of these two detectors which gives an alarm.

Two estimates of the instant of jump θ are possible: either the instant T_h of alarm decreased by its theoretical bias, which has been calculated in [2], or the last instant T_M, at which the maximum M_{T_h} has been reached between the instants 0 and T_h. In the sequel, this last estimator of θ will be the only estimator that we shall consider.

For detecting jumps in the mean in signals, the state model of which is of order greater than 1, the procedure obviously consists in using the previous Hinkley's detector for the innovations of the Kalman filter built according to the considered model.

But, in spite of their properties of efficiency and robustness which were shown in [2], Hinkley's algorithms cannot be used without knowledge of the first mean m_o, and, in order not to forget possible jumps which could occur during the estimation of m_o, one usually introduces a filtered derivative detector during this only period where the error variance on the estimated m_o is too large. This has been done for line-by-line edge detection in noisy images (see [1]), and we shall come back to this coupling of algorithms in the next paragraph.

But one can also think of using a threshold h which is a function, not only of the minimum amplitude of jump ν_m and the measurement noise σ^2, but also of the error variance on the estimated m_o. This has been done for our data, and an example can be seen on *figure n° 16*. It immediately appears that this procedure may not be efficient for detecting two successive jumps and thus the use of a "filtered derivative" detector during the estimation of m_o is quite necessary.

Let us first mention another procedure which has been elaborated for detecting two opposite jumps which occur quickly one after the other, in the case where a coupled "filtered derivative" detector is not sufficient to detect the second one (except if one accepts many false alarms): it consists in using, for the second jump, the same detector that has found the first one, but the mean estimated before the first one. This procedure has been used for blood vessels detection in noisy images by J. GASNIER (see [8]).

III.3. Coupling Hinkley's algorithm and "filtered derivative" detector.

Considering the noisy signal which has to be analyzed it is quite clear that a simple integrating filter, which was used for edge detection, is not efficient enough to detect "early" jumps without giving raise to numerous false alarms. Therefore, we used the more complex counting of the crossings which was presented at the end of the first paragraph of this section. See *figures n° 17, 18*. Comparing *figures n° 16* and *17*, one can appreciate how the use of this "filtered derivative" detector, while estimating m_o, can reduce some undesirable behaviour of the Kalman filter after detection of a jump. (Of course, we use here the same state model for the data as was used with GLR detectors).

Let us recall that this drawback of Hinkley's algorithm, with respect to Willsky's GLR algorithm, cannot be overcome: after each detection of a jump, the analysis of the signal "starts again from zero" as at the beginning, and one cannot do anything else because the processor does not obtain any information about the size of the jump when a detection occurs; the minimum amplitude ν_m is in fact a poor information because, with noisy real data, the "symmetrical" situation (corresponding to the change of sign of m_o) is scarcely encountered. Therefore, the "reactualization" procedure used by A.S. WILLSKY to initialize the Kalman filter after each detection of a jump cannot be introduced when Hinkley's cusum tests are used.

In the next paragraph, a new algorithm, which keeps the computational advantage of Hinkley's algorithm and the power of Willsky's GLR detector, will be presented and shown to allow a real improvement.

III.4. A new algorithm mixing GLR and cusum tests.

A large part of the computing time involved by Willsky's GLR algorithm is due to the search of the estimate $\hat{\theta}$ of the failure time which maximizes the likelihood ratio $\ell(k,\theta)$ for $k-M+1 \leq \theta \leq k$. Thus using Hinkley's estimate leads to a noticeable reduction of the computing time, all the more as the estimation of ν with the aid of Willsky's formulas does no longer need computing the functions $F(k,\theta)$, $G(k,\theta)$, $C(k,\theta)$, $d(k,\theta)$ for all θ between $k-M+1$ and k, but only for the time θ given by Hinkley's test. In other words, as long as: $M_k=S_k$ or $m_k=U_k$ (i.e. the random walk S_k increases or U_k decreases) the estimate of θ is taken to be k, and F, G, C, d are computed with the aid of the "initialized" formulas (23b) (25b); and when M_k or m_k remains the same, say M_{k_o} or m_{k_o}, $\hat{\theta}$ is taken to be k_o and F, G, C, d are computed with the aid of the recursion formulas (23) (25).

This new algorithm, which gives an estimate of the amplitude of the failure together with the failure time, can therefore be connected with Willsky's updating procedure for the Kalman filter after each detected failure. It should be noticed than this algorithm is a little less efficient than the modified version of Willsky's GLR algorithm we have presented in section II.2; the main explanation for this behaviour is the fact that the ideal "symmetrical" case for Hinkley's detector is scarcely encountered is noisy areas, and therefore the estimate of the failure time is less accurate: the delay for detection is often larger. This can be seen for example on *figure n° 19* (compare with *figure n° 4*) where the horizontal arrows indicate the delays for detection.

However, as far as one is concerned with the classification of the detected failures, a first short analysis has shown that this new algorithm mixing GLR and cusum tests may be almost as efficient as the modified GLR algorithm.

IV. COMPARATIVE STUDY AND CONCLUSIONS.

Several elements of comparison between the different algorithms have been given all along the former sections. This last section is devoted to a more precise comparative study in terms of quantitative evaluations of the computations, the number of data that have to be memorized, the number of parameters to be adjusted, and, on the other hand, qualitative pieces of information such as sensitivity of the algorithms to the setting of the parameters (model, filter, detector, updator), efficiency (quick detection after actual failures, but not too many "false" alarms), and robustness (essentially with respect to modelling errors, and high non-stationarities).

IV.1. Quantitative evaluations of the complexity of the algorithms.

We consider here only three detectors: Willsky's original GLR algorithm and the modified version we have presented in section II.2, and the new algorithm mixing GLR and cusum detectors presented in section III.4. They will be respectively denoted by W, WM and HKW. The computations are related to the special model (3) (see Introduction).

	Kalman filter	W	WM	HKW
nb of additions at each step	17	$8(M-1)$	$8M+1$	6 or 6+8
nb of multiplications at each step	20	$9(M-1)+2$	$9M+2$	2 or 2+6
nb of parameters to be chosen	3 or 2 $(\tau, q^1, (q^2))$	2 (M, ε)	4 (M, p, ν_m, λ)	2 (ν_m, h)
nb of data to be memorized (apart from the current data y_k)	6 $(\hat{x}_{k-1}, \hat{\mu}_{k-1}, \hat{\sigma}^2_{k-1}, a_{k-1}, b_{k-1}, C_{k-1})$	$4M$ $(f^1, f^2, C, d$ over the window)	$4M+p+1$ $(f^1, f^2, C, d$ over the window and $\bar{\nu}_{k-1}, \ldots, \bar{\nu}_{k-p}, S_{k-1})$	10 (2 cusums, 2 extrema, 2 extremum times, $f^1, f^2, C, d)$
nb of tests (search for extremum) at each step apart from failure test	—	$M-1$	$M+2$	2

For the updator (the same for all the algorithms), 5 additions and 10 multiplications at each alarm. (b_k and c_k being updated and emphasized).

IV.2. Practical versus theoretical point of view for studying efficiency and robustness.

It should be emphasized that we here study sensitivity, efficiency and robustness only from a qualitative point of view, for the following reasons. First, some theoretical studies concerning some of the algorithms have already been done: for the first GLR algorithm, A.S. WILLSKY and H.L. JONES derived the probability of false alarm and the probability of correct detection of a jump of magnitude ν at time θ (see [16]; for Hinkley's cusum test, M. BASSEVILLE derived formulas for the mean time between false alarms and the mean time delay for detection of a jump, under the hypothesis that no false alarm occured before the jump (see [2], where a comparative simulation analysis, concerning Hinkley's and Shiryaev's cusum tests and some "filtered derivatives" detectors, can also be found). The second reason that led us to this qualitative point of view is that theoretical results concerning the properties of an algorithm are scarcely sufficient to give a guide for the setting of its various parameters in practice. Three examples, concerning the real data studied in the present paper, can be given: the probability of correct detection derived in [16] can be used only if the magnitude ν and the failure time θ are known, and furthermore it has been shown in section II.1 that the first GLR algorithm is rather sensitive to the choice of the window size and of the threshold; second, the theoretical results concerning Hinkley's cusum test may be difficult to be used in practice because of the complexity of the formulas and, especially, because the "symmetrical situation" for detection is scarcely encountered (see section III.2), all the more as the underlying model which is used is a poor one with respect to the actual structure of the data (think of the very noisy areas of the signals); and, third, standard choices of thresholds coming from theoretical results may be very different from the choices which appear to be interesting in practice: for example, the likelihood ratio test used with the smoothed estimate $\overline{\nu}$ in section II.2 is usually set with a threshold of about 2 (at the level 5 %), but thresholds of orders of magnitude 100 and even 1000 appeared to give interesting results. So it seemed to us that qualitative informations coming from implementation of the algorithms in real situations (and not ideal ones) could be of interest.

IV.3. Concluding remarks.

Let us first notice that the causality underlying the recursive processing of the data has to be actually taken into account: all the studied failure detection algorithms differently work according to the processing direction. For instance, compare the *figures n° 2* and *4* (where the signal is processed from right to left) respectively to the *figures n° 20* and *21* (where the signal is processed from left to right). This can be explained by the differences that generally exist between the measurement noises before and after the failures: it is obviously easier to detect transitions between a smooth signal and a noisy one than in the opposite case; and, in our special application, this fact is due to the "geological" causality: the way in which a layer has been formed depends upon the previous ones (and thus downwards ones, or right to left on the figures).

Second, let us indicate that, because of the good results which have been obtained especially with the modified GLR algorithm, it does not seem that non sequential treatments by sliding blocks could improve the global results if one think of the tradeoff between complexity and efficiency; but this point has to be further investigated.

Finally, let us emphasize the main conclusions coming from the comparative study which has been described in this paper. In spite of the large number (4) of parameters to be chosen for the WM detector, it has been shown that this algorithm is less sensitive to such a choice than HKW algorithm for which only 2 parameters have to be set up. (See section V.1). Furthermore, the main advantages of this WM algorithm are the relative independence of the window sizes M and p with respect to the threshold λ (see section II.2) and the simple interpretation of the two key parameters ν_m and λ, which are more meaningful for the user than the threshold ε on the likelihood ratio of the original test (see (17) - section II.1)). It has to be noticed that the minimal amplitude of jump ν_m, which is involved in the WM algorithm and in all the algorithms based upon Hinkley's detector, has not exactly the same meaning in these two cases because Hinkley's algorithm (and HKW also) allows detection of failures the actual amplitudes of which are less than the fixed ν_m (this fact is due to the non symmetrical case often encountered in practice. See section III.2).

Another consequence of this situation is the delay for detection which is larger with Hinkley's detector than with WM or W algorithm (compare *figures n° 4* and *19*).

A last feature, common to GLR algorithms and cusum tests, has to be emphasized: the robustness of these algorithms with respect to the hypothesis "white measurement noise" is obvious, especially in the serrate areas. Let us only recall that, first, this fact has already been observed in [1] and [8] for Hinkley's detector, and, second, the choice of the estimate of the variance of the measurement noise contributes to this robustness (see section II.1).

APPENDIX

Theoretical behaviour of the GLR.

The purpose of this appendix is the study of Willsky's GLR detector in "stationary" operating mode, in other words after the Kalman filter has converged, in order to analyze the influence of the various model parameters (18) upon the test function and to give a guide for setting the threshold ε (see (17)).

Assuming that the variance σ^2 of the measurement noise is known, we derive from (19) that, when $k \to +\infty$, the Kalman filter converges to the following values:

for the gain:

$$\begin{cases} \alpha = \dfrac{v}{v+\sigma^2} \\ \beta = \dfrac{\tau b + c}{\sigma^2 + v} \end{cases}$$

where $v = \tau^2 a + 2\,\tau b + c + q^1$,

and for the estimation error covariance matrix:

$$\begin{cases} a = (1-\alpha)\, v \\ b = -\beta v + \tau b + c \\ \quad = (1-\alpha)\,(\tau b + c) \\ c = -\beta(\tau b + c) + c + q^2. \end{cases}$$

In the special case where the variance q^2 of the slope noise is zero, we obtain: $\tau b + c = 0$, and thus: $\beta = b = 0$, and $c = 0$. If follows that:

$$v = \frac{q^1}{1-\tau^2+\tau^2\alpha}$$

and $\alpha = \dfrac{v}{v+\sigma^2}$ is the positive root of the equation:

$$\sigma^2\tau^2\alpha^2 + (\sigma^2(1-\tau^2)+q^1)\alpha - q^1 = 0$$

i.e.:

$$\alpha = \frac{-[\sigma^2(1-\tau^2)+q^1] + \sqrt{[\sigma^2(1+\tau)^2+q^1][\sigma^2(1-\tau)^2+q^1]}}{2\,\sigma^2\tau^2}.$$

Again when $k \to +\infty$, the test function $F(k,\theta)$ is such that, if $f^i(n) = \lim_{k\to\infty} f^i(k,k-n)$ $(i=1,2)$, then:

$$f^2(n) = 0 \qquad (\forall\ n\geq 0)$$

and:

$$\begin{cases} f^1(n) = (1-\alpha)\ \tau\ f^1(n-1) + \tau^n\ \alpha & \text{for } n\geq 1 \\ f^1(o) = \alpha \end{cases}$$

(see (23)).

Thus: $f^1(n) = \tau^n\,[\,1-(1-\alpha)^{n+1}\,]$ $\qquad \forall\ n\geq 0,$

and the test function g becomes:

$$g(n) = \tau^n - \tau\ f^1(n)$$

i.e.: $g(n) = \tau^n(1-\alpha)^n$

Therefore, from (25) it comes:

$$C(n) = C(n-1) + \frac{[g(n)]^2}{v+\sigma^2}$$

or, equivalently:

$$C(n) = \frac{1}{v+\sigma^2}\,\frac{1-\tau^{2n+2}(1-\alpha)^{2n+2}}{1-\tau^2(1-\alpha)^2}$$

where: $v = \dfrac{\alpha\sigma^2}{1-\alpha}$,

i.e.: $c(n) = \dfrac{1-\alpha}{\sigma^2}\,\dfrac{1-\tau^{2n+2}(1-\alpha)^{2n+2}}{1-\tau^2(1-\alpha)^2}$.

On the other hand, A.S. WILLSKY and H.L. JONES [16] have shown that, for a general model (4), the GLR $\ell(k,\theta)$ (see (16)) has a non central Chi-squared density with n_f degrees of freedom (n_f = dimension of the "failure" vector ν) and noncentrality parameter $\nu' C(k,\theta)\nu$, where θ and ν are the true values of the failure characteristics. Therefore, for our special model (18), the GLR $\ell(k,\theta)$ is such that, if: $\ell(n) = \lim_{k\to+\infty} \ell(k,k-n)$ (convergence in distribution), then:

$$E[\ell(n)] = 1 + \nu^2 C(n)$$

$$= 1 + \frac{\nu^2}{\sigma^2}(1-\alpha)\frac{1-\tau^{2n+2}(1-\alpha)^{2n+2}}{1-\tau^2(1-\alpha)^2}.$$

The "asymptotical" behaviour of the test functions $\ell(k,\theta)$ within a window of length M $(k-M+1 \leq \theta \leq k)$ can then be deduced from the following remarks:

let us compute:

$$E[\ell(n+1)-\ell(n)] = \frac{\nu^2}{\sigma^2}\tau^{2n+2}(1- \quad)^{2n+3}.$$

If the state noise has variance $q^1=0$, then the Kalman gain $\alpha=0$, and thus $E[\ell(n)]$ is a linear function of n, as has been observed during a simulation study. The threshold ε in this case should linearly increase with the length M of the window.

If the Kalman gain α is near to zero, then:

$$E[\ell(n+1)-\ell(n)] = \frac{\nu^2}{\sigma^2}\tau^{2n+2}[1-(2n+3)\alpha],$$

i.e. $E[\ell(n)]$ is a <u>parabolic</u> function of n. This fact has been observed on real data.

Let us emphasize that this behaviour of the generalized likelihood ratio depends upon the state model.

FIGURE CAPTION

Figure n° 1 : (W) Original Willsky's GLR test (threshold 30).

n° 2 : (W) Original Willsky's GLR test (threshold 40).

n° 3 : (WM) Modified GLR test (threshold 20).

n° 4 : (WM) Modified GLR test (threshold 100).

n° 5 and 7 : Modified GLR test. No slope noise.

n° 6 and 8 : Modified GLR test. With slope noise.

9 ,10, 11 : Modified GLR test. Incidence of (ν_m^f, λ).

12-15 : A "filtered derivative" detector.

16 : Hinkley's detector. Threshold depending upon the error variance before jump.

17-18 : Hinkley's detector coupled to a "filtered derivative" detector. Incidence of ν_m.

19 : (HKW) An algorithm miwing GLR and cusum tests.

20 : W algorithm. Processing from left to right.

21 : WM algorithm. Processing from left to right.

22 : Hinkley's cumulative sum test.

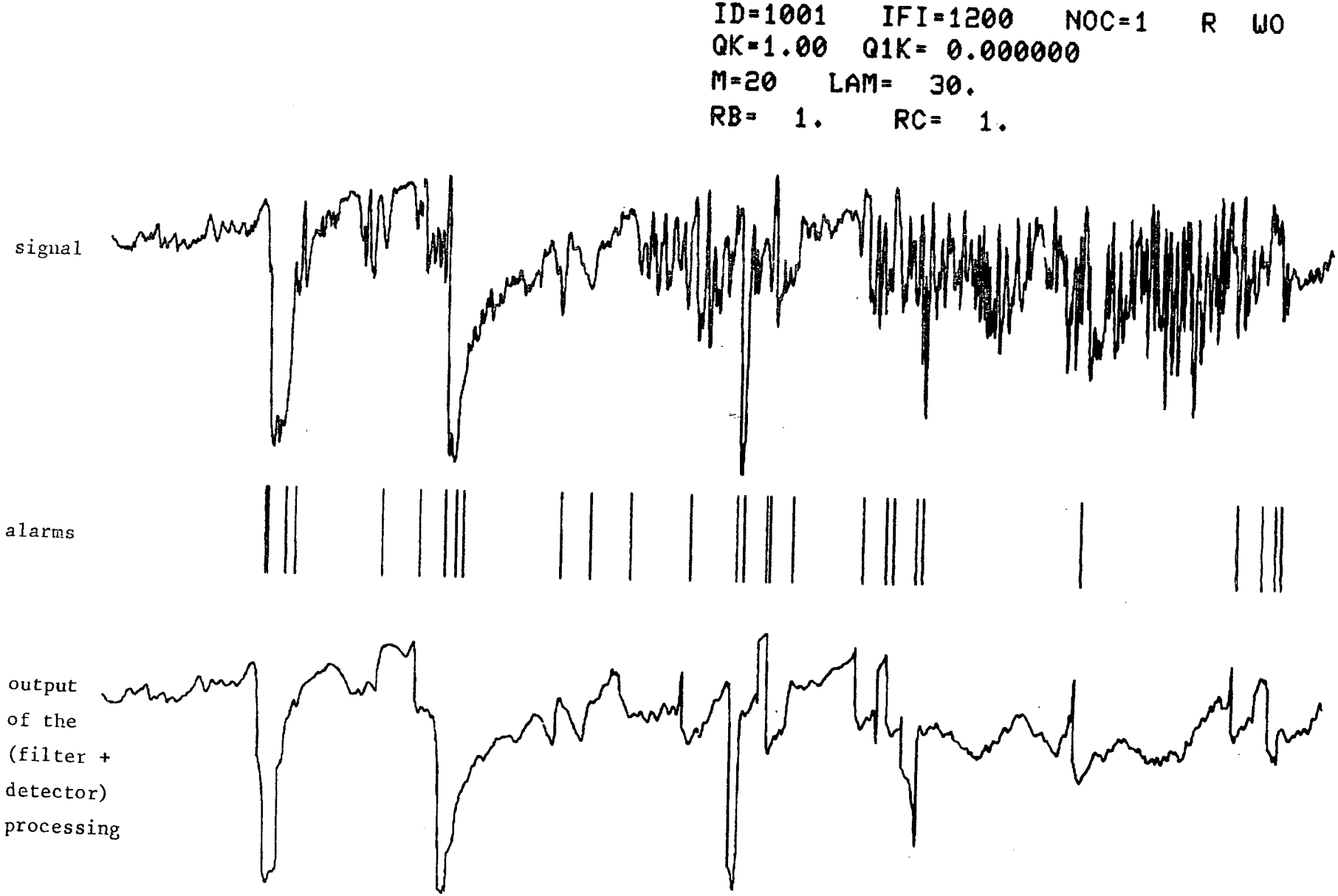
ID=1001 IFI=1200 NOC=1 R WO
QK=1.00 Q1K= 0.000000
M=20 LAM= 30.
RB= 1. RC= 1.
signal
alarms
output
of the
(filter +
detector)
processing

Figure n° 1

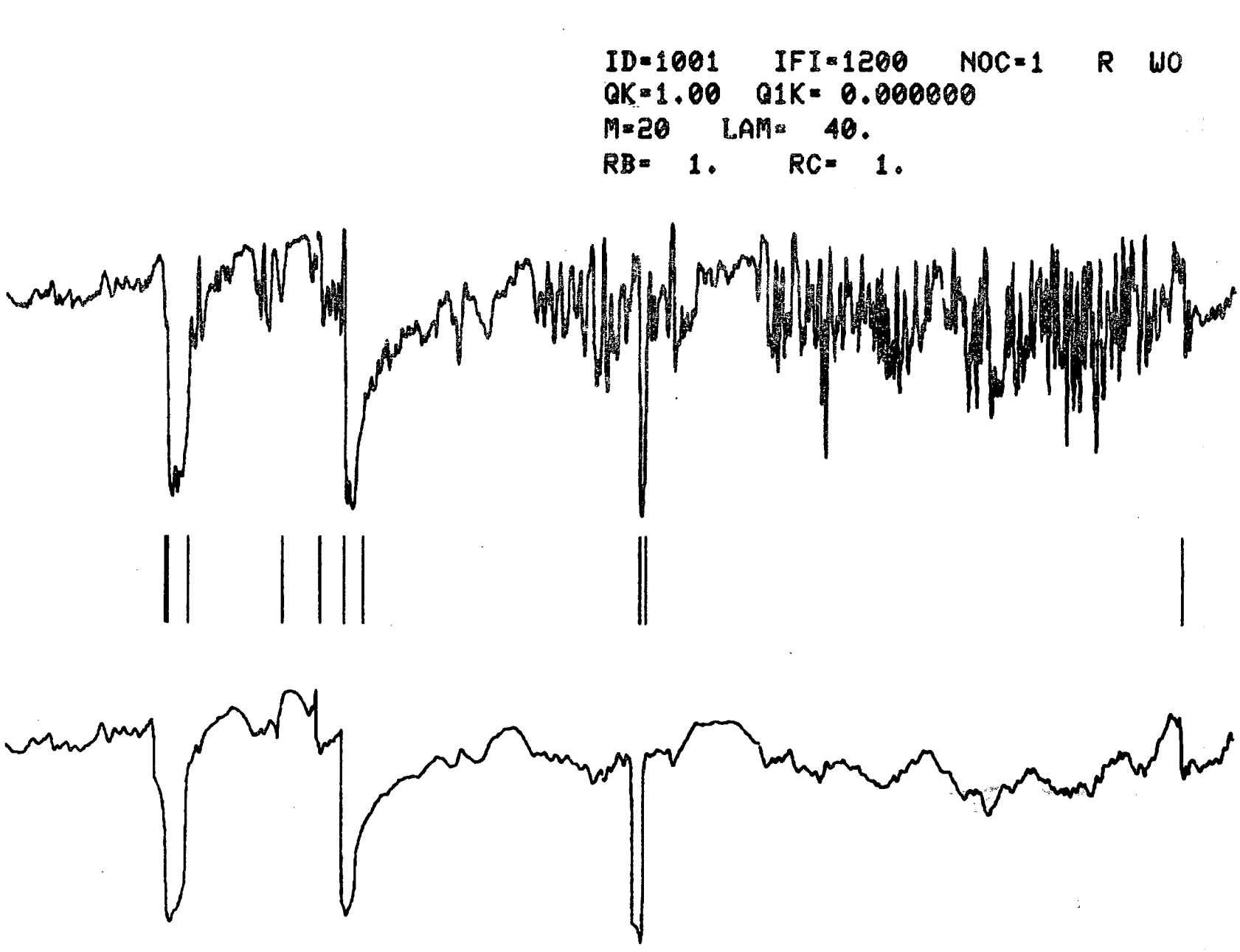

Figure n° 2

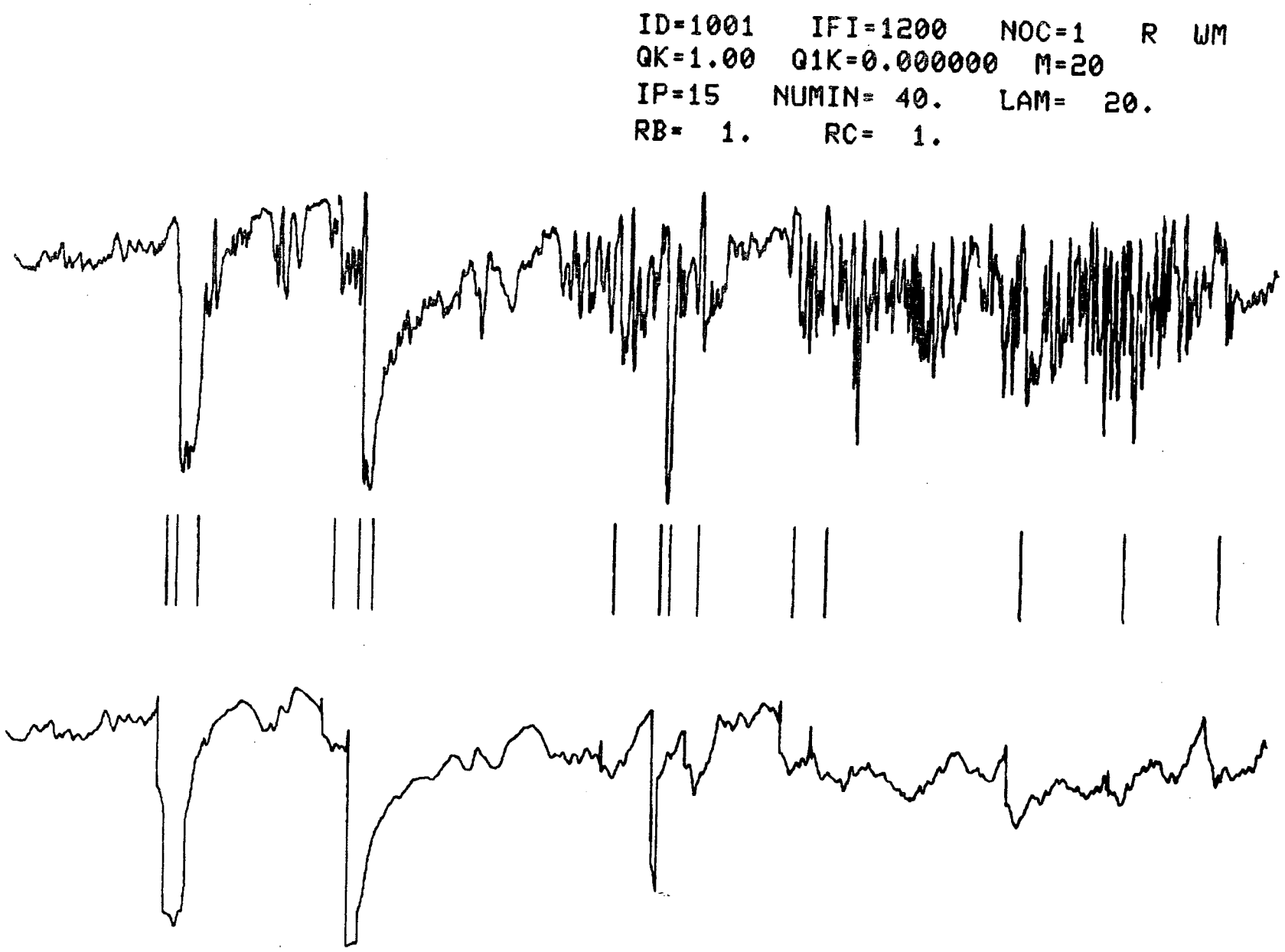
ID=1001 IFI=1200 NOC=1 R WM
QK=1.00 Q1K=0.000000 M=20
IP=15 NUMIN= 40. LAM= 20.
RB= 1. RC= 1.

Figure n° 3

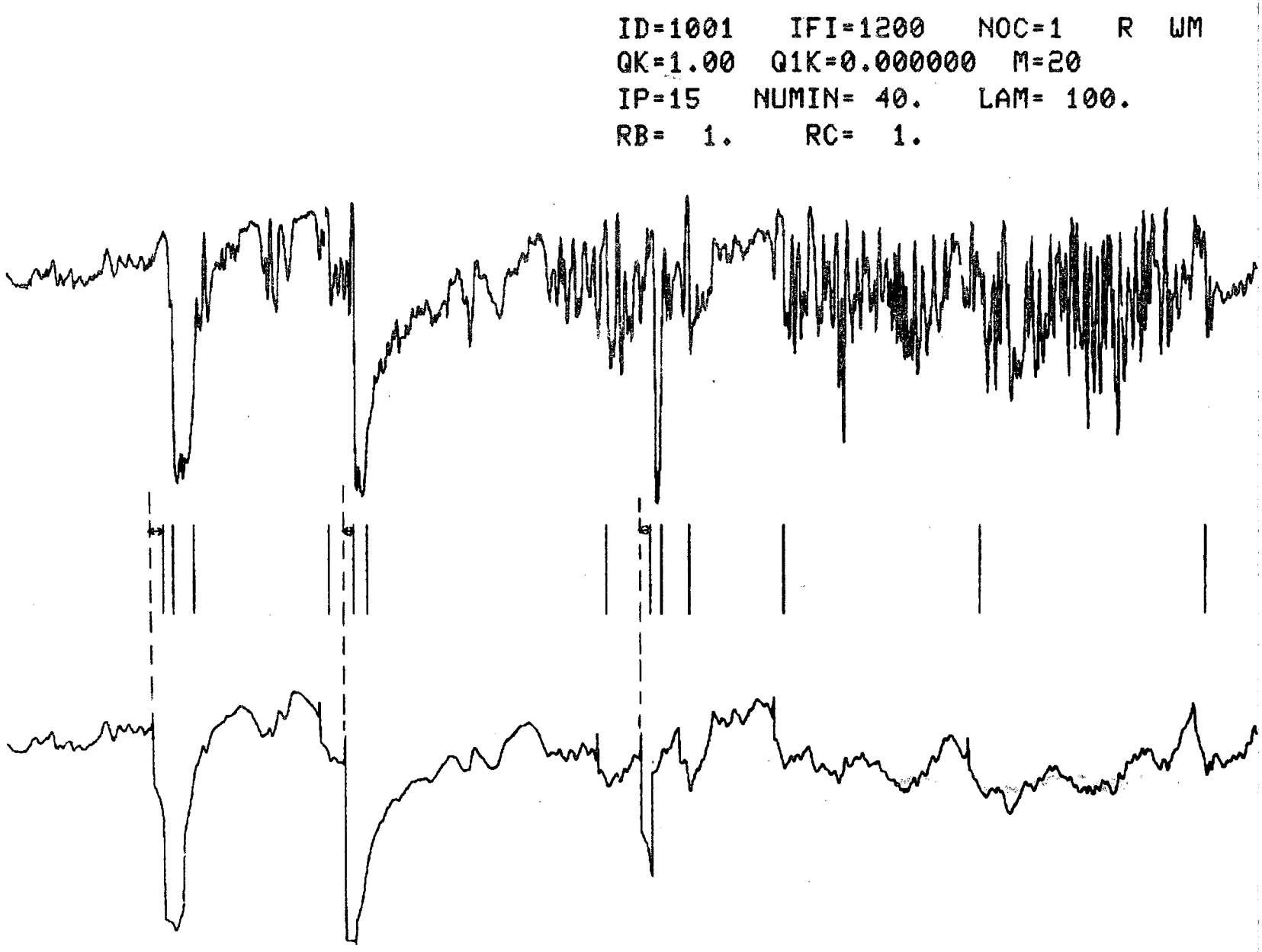
ID=1001 IFI=1200 NOC=1 R WM
QK=1.00 Q1K=0.000000 M=20
IP=15 NUMIN= 40. LAM= 100.
RB= 1. RC= 1.

Figure n° 4

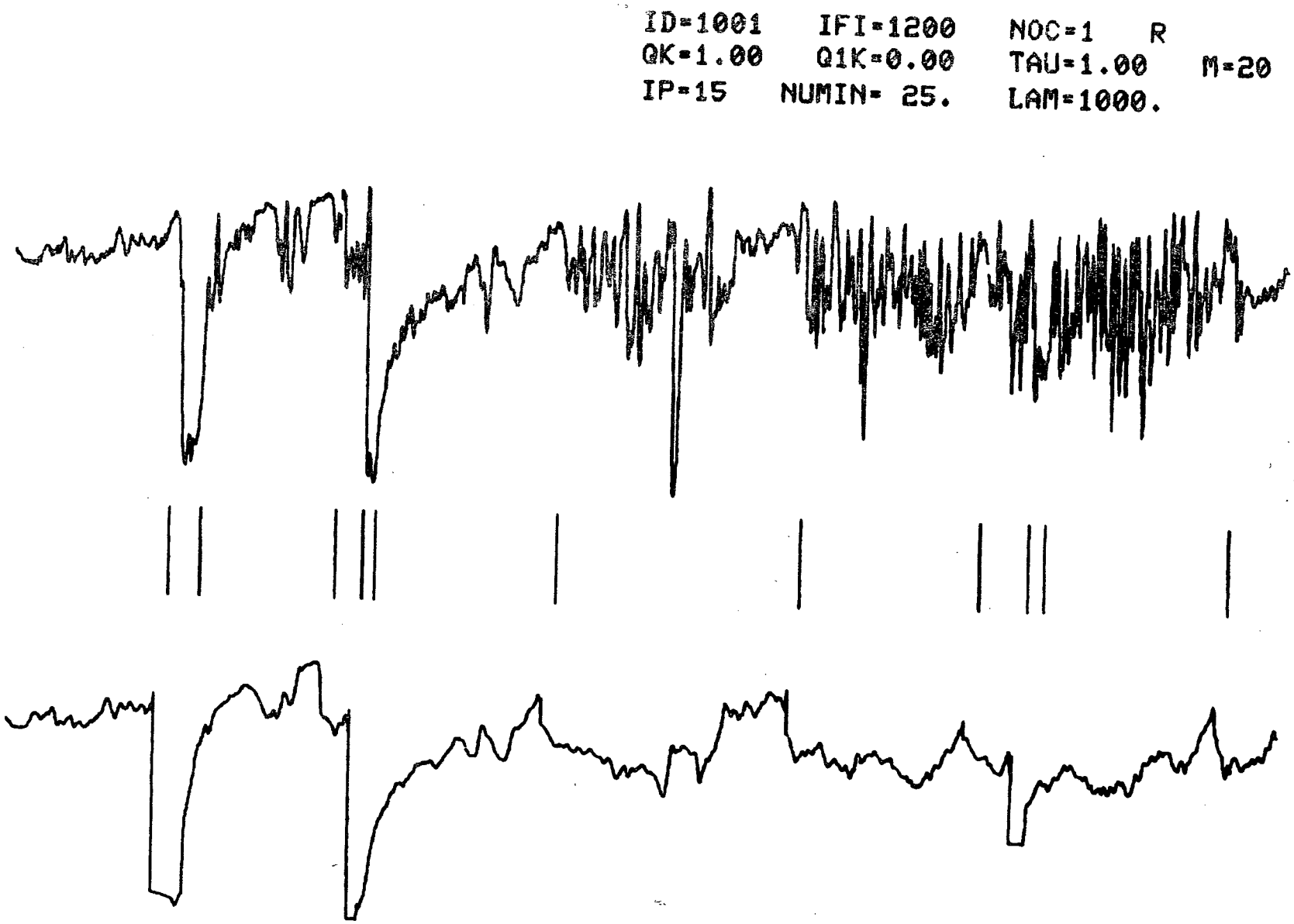
ID=1001 IFI=1200 NOC=1 R
QK=1.00 Q1K=0.00 TAU=1.00 M=20
IP=15 NUMIN= 25. LAM=1000.

Figure n° 5

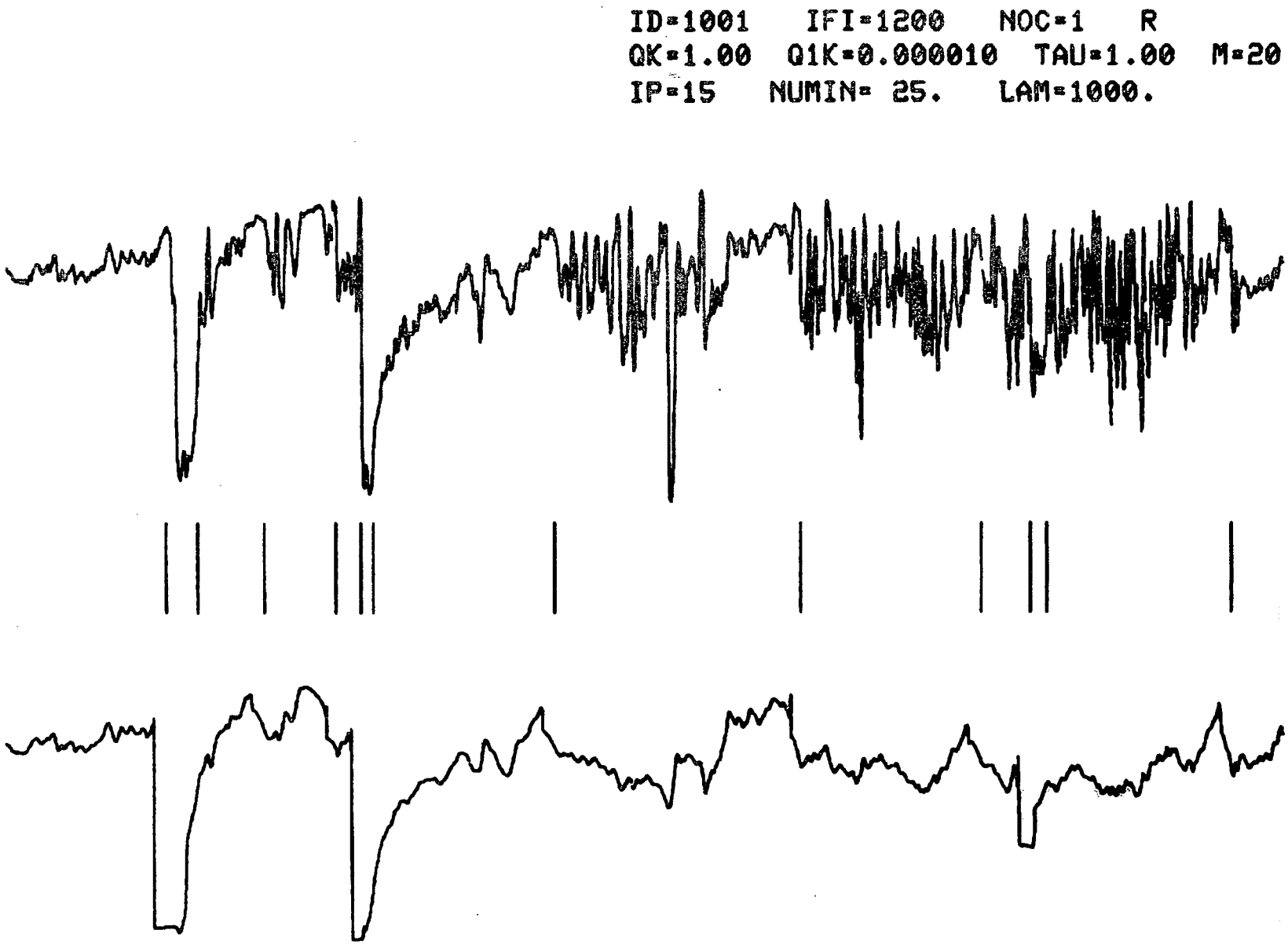
ID=1001 IFI=1200 NOC=1 R
QK=1.00 Q1K=0.000010 TAU=1.00 M=20
IP=15 NUMIN= 25. LAM=1000.

Figure n° 6

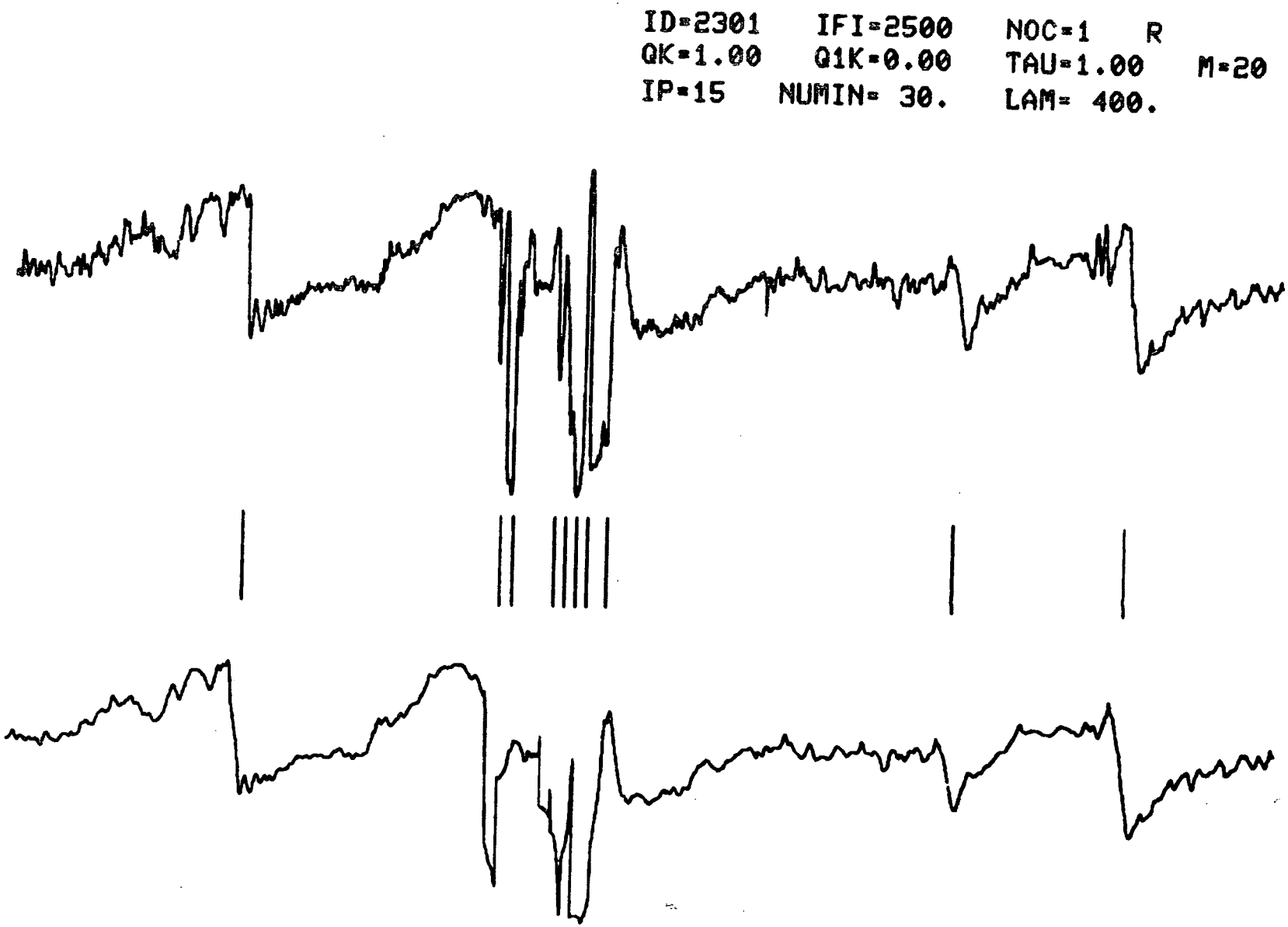
ID=2301 IFI=2500 NOC=1 R
QK=1.00 Q1K=0.00 TAU=1.00 M=20
IP=15 NUMIN= 30. LAM= 400.

Figure n° 7

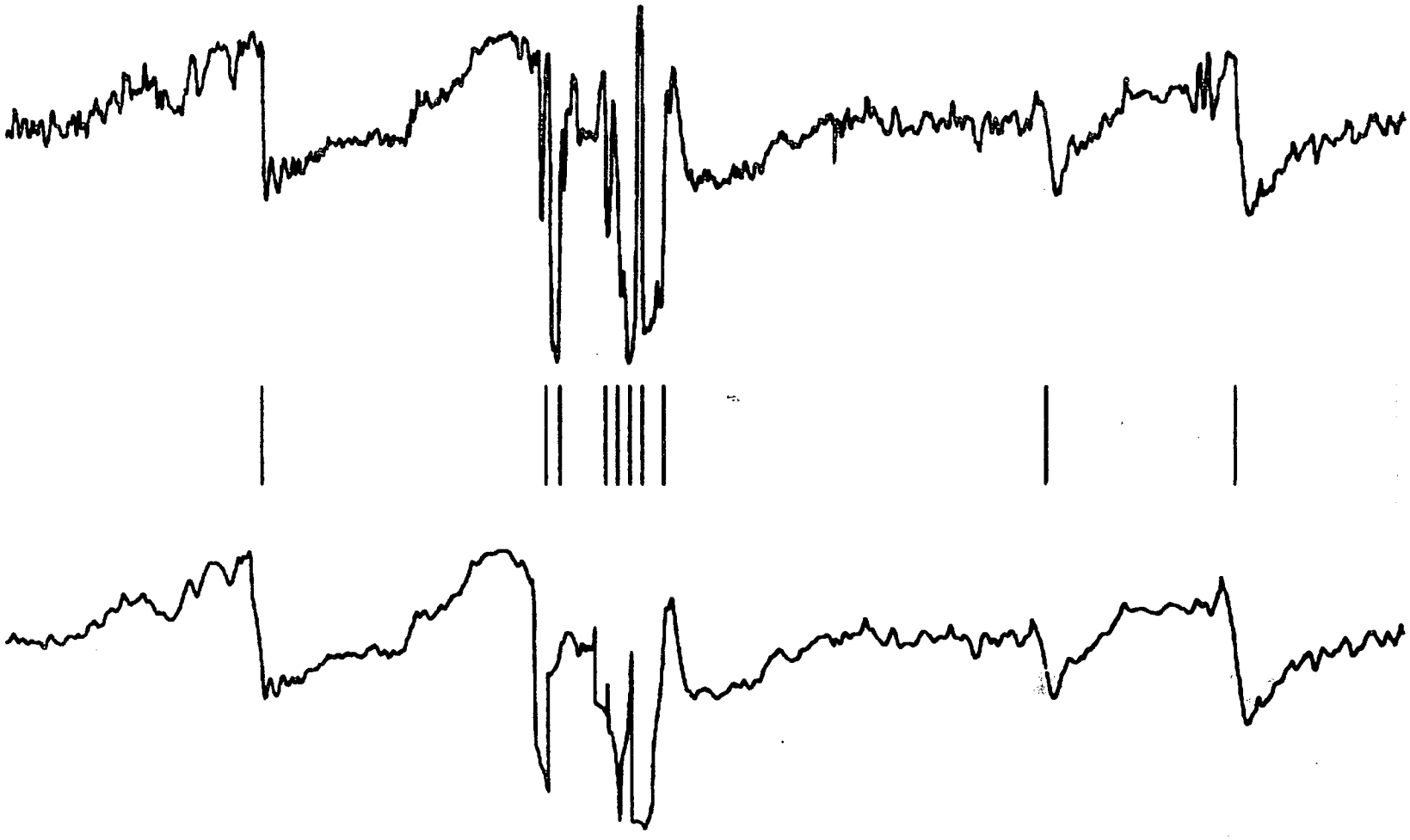
ID=2301 IFI=2500 NOC=1 R
QK=1.00 Q1K=0.000001 TAU=1.00 M=20
IP=15 NUMIN= 30. LAM= 400.

Figure n° 8

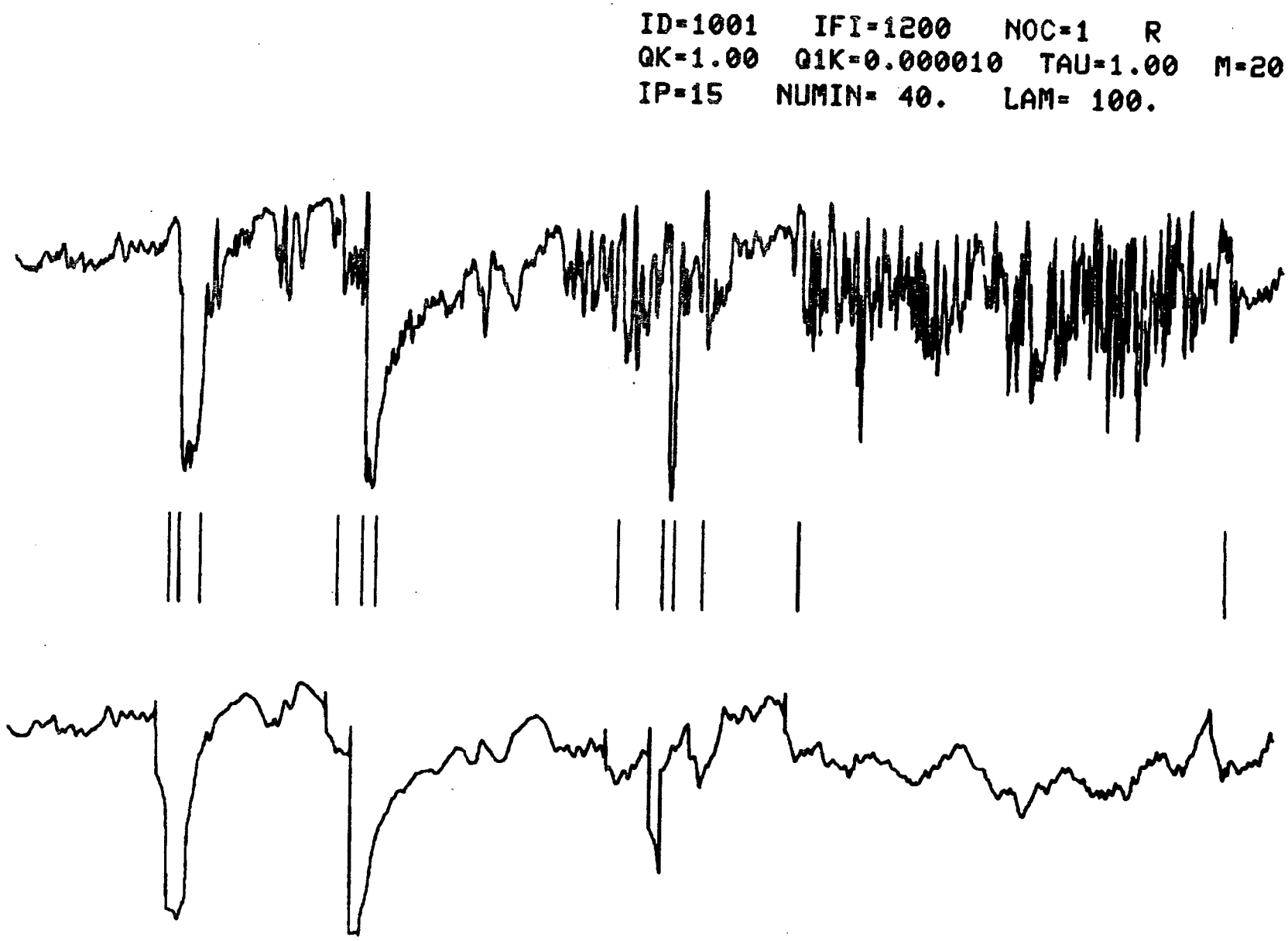
ID=1001 IFI=1200 NOC=1 R
QK=1.00 Q1K=0.000010 TAU=1.00 M=20
IP=15 NUMIN= 40. LAM= 100.

Figure n° 9

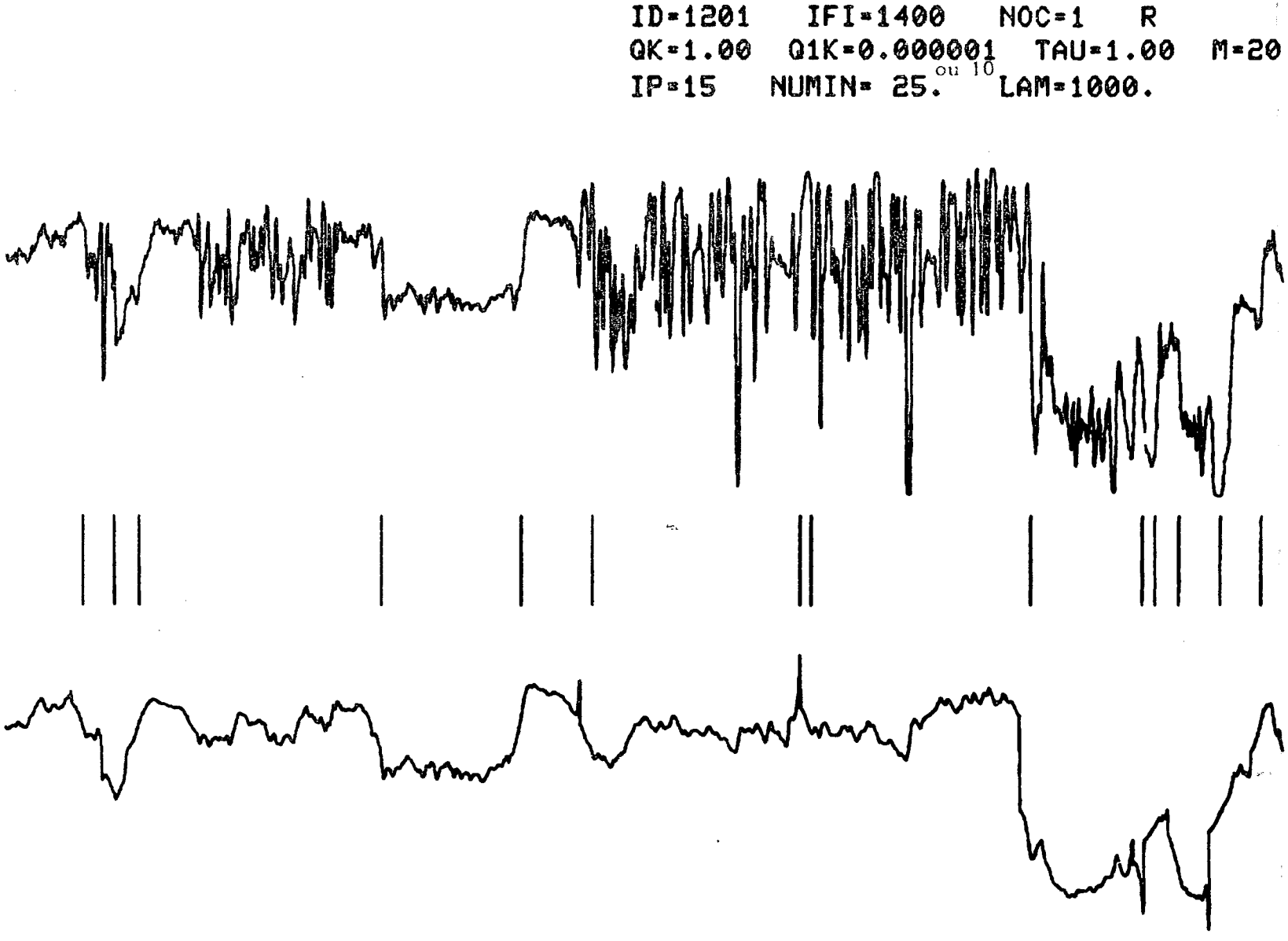
ID=1201 IFI=1400 NOC=1 R
QK=1.00 Q1K=0.000001 TAU=1.00 M=20
IP=15 NUMIN= 25. ou 10 LAM=1000.

Figure n° 10

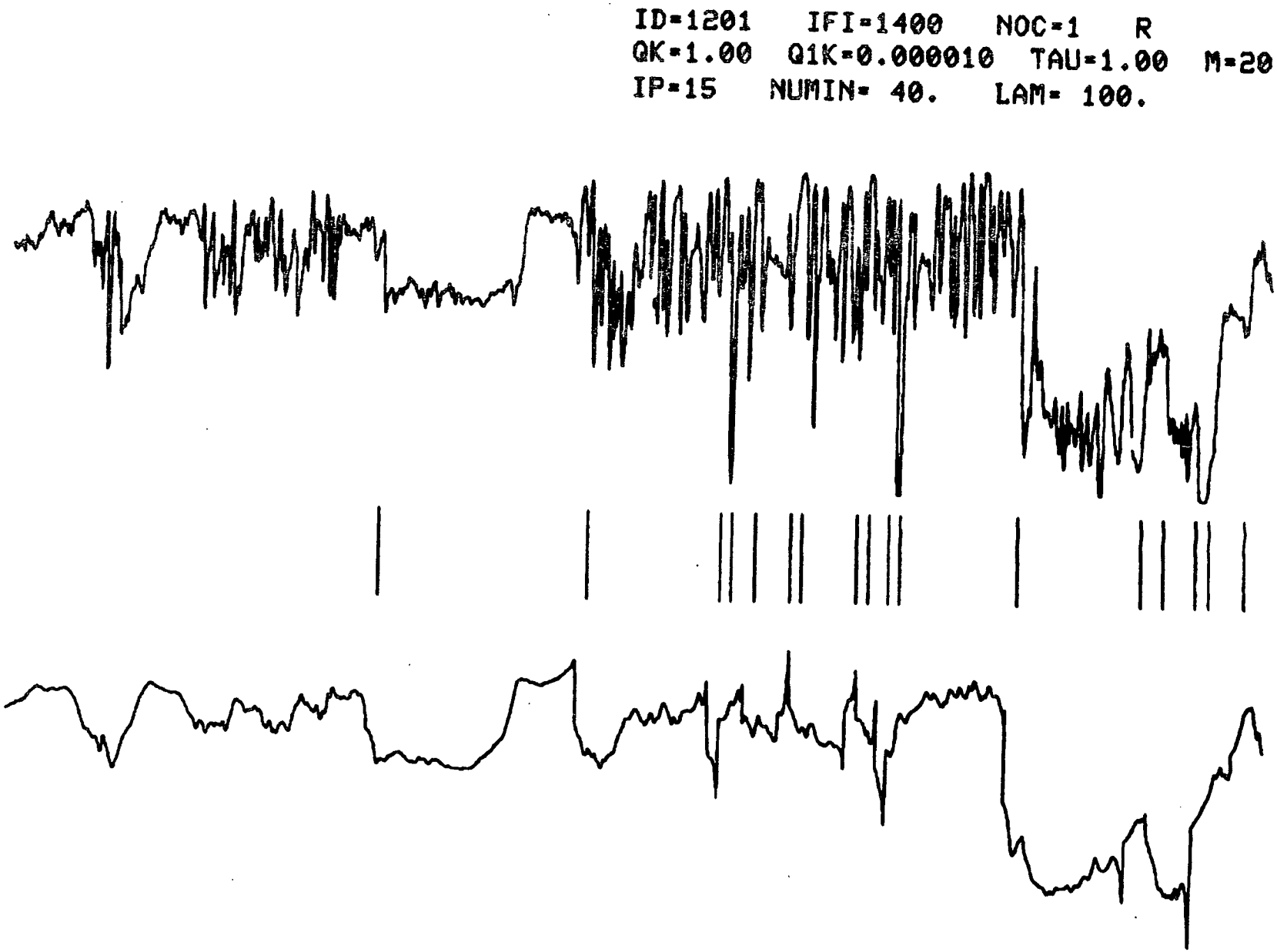
ID=1201 IFI=1400 NOC=1 R
QK=1.00 Q1K=0.000010 TAU=1.00 M=20
IP=15 NUMIN= 40. LAM= 100.

Figure n° 11

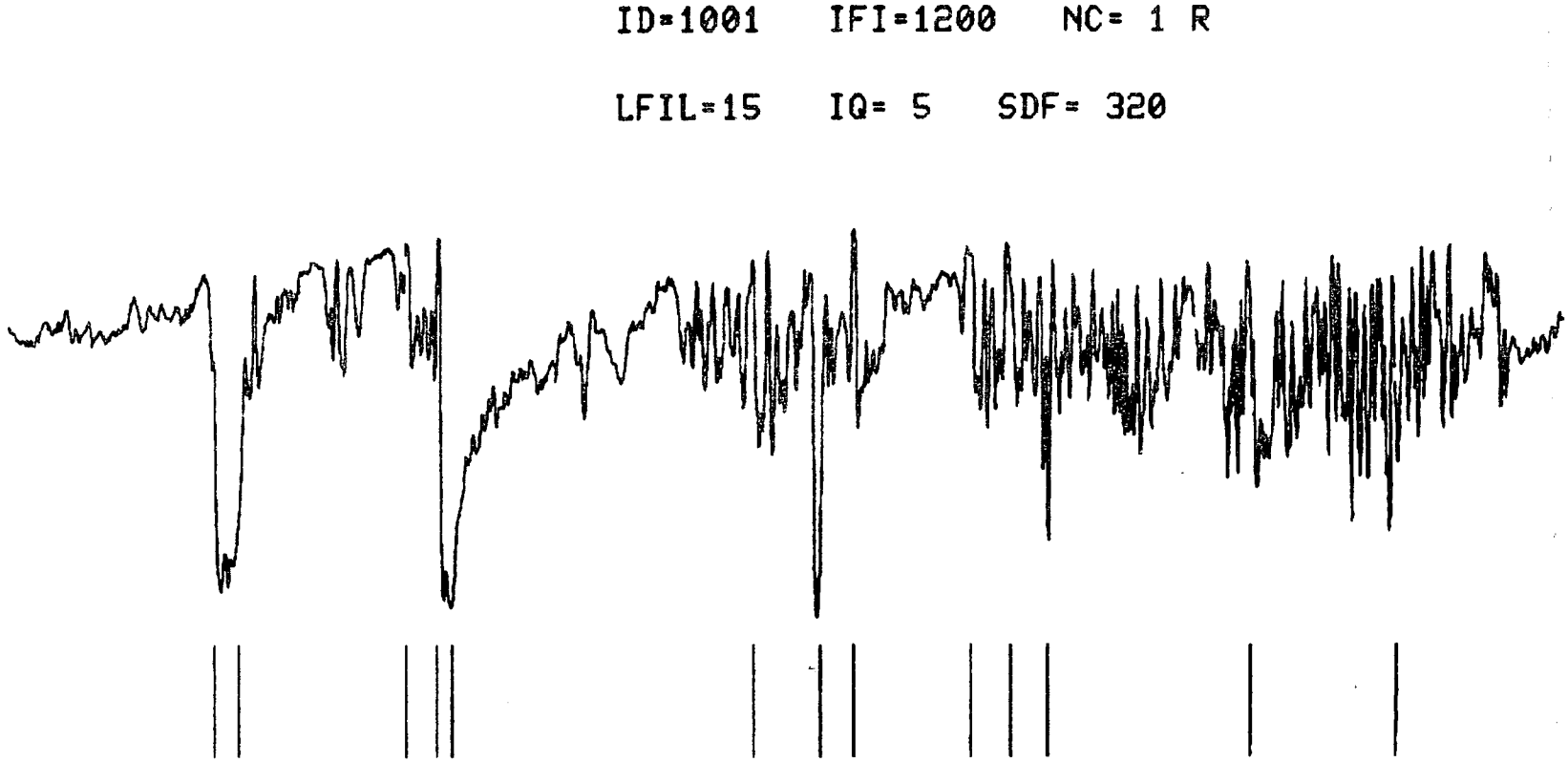

Figure n° 12

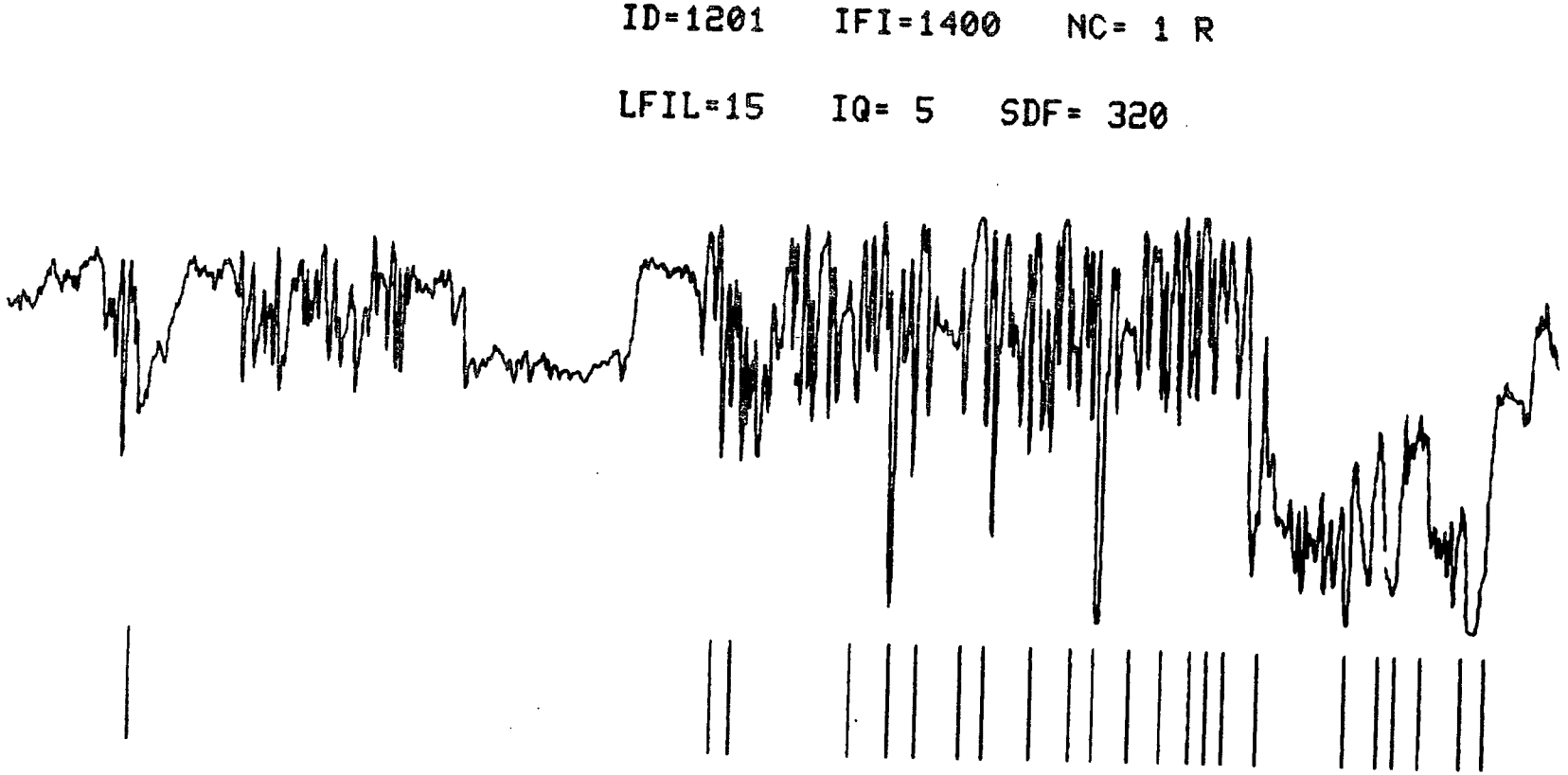

Figure n° 13

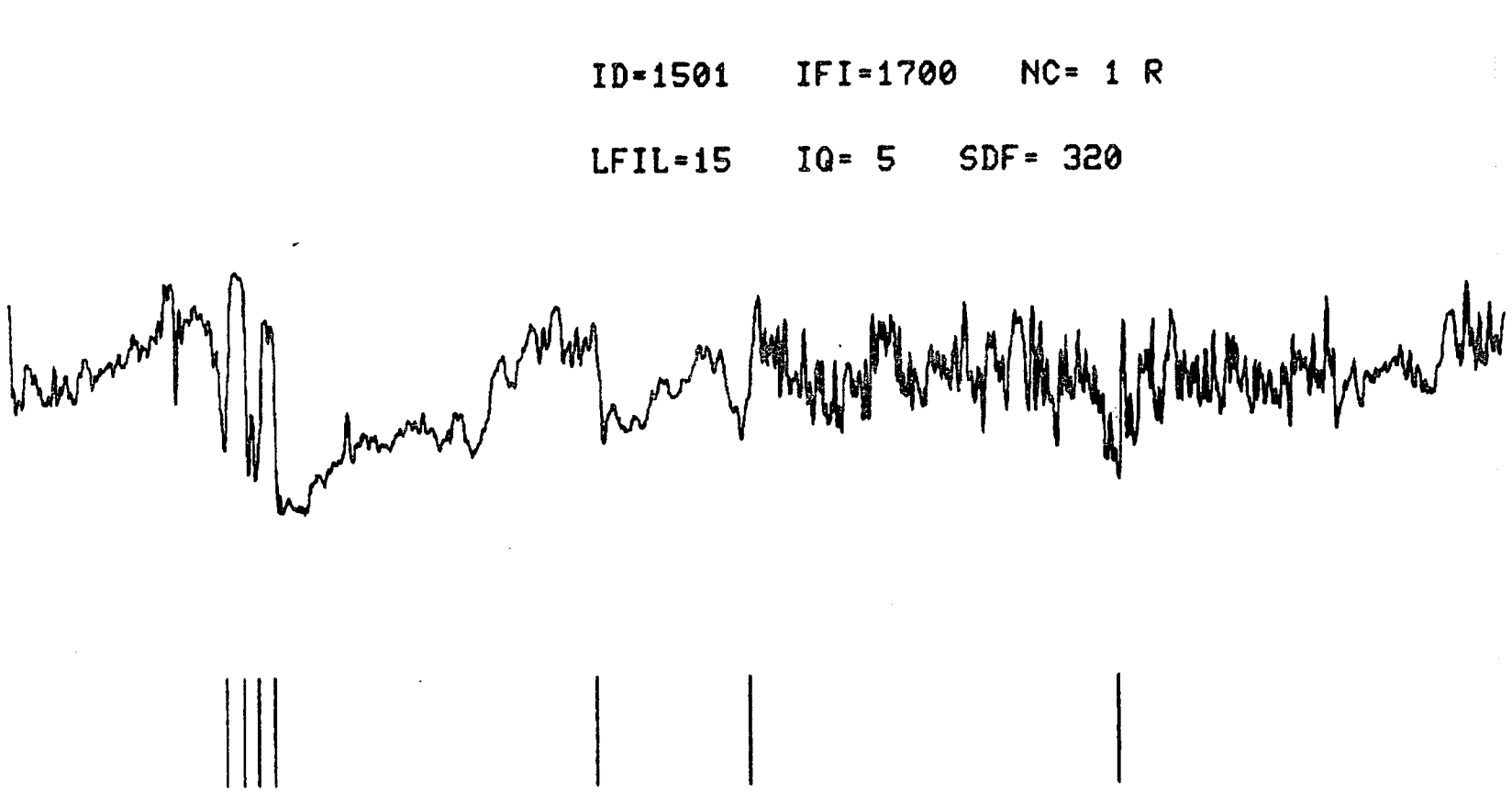

Figure n° 14

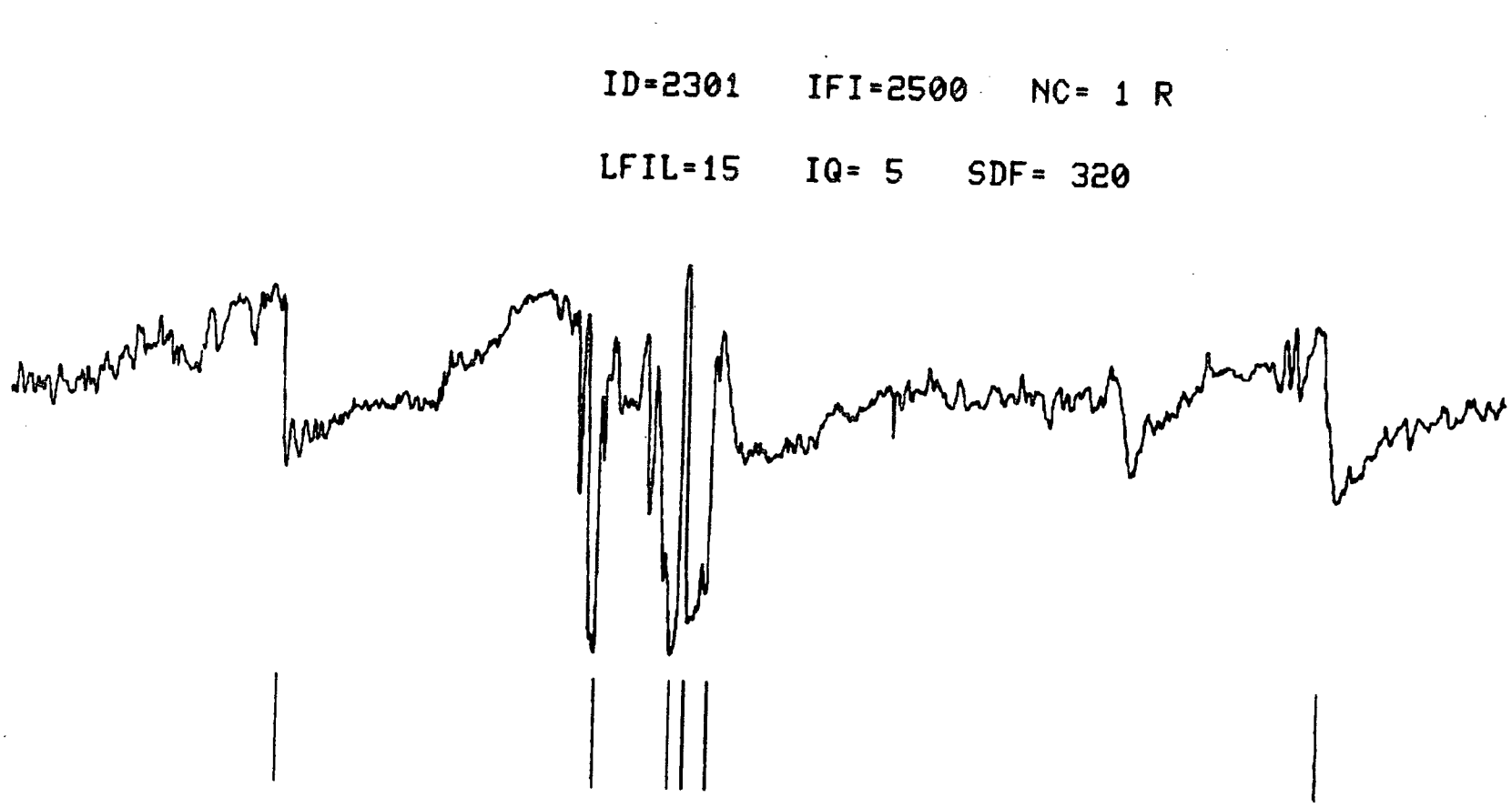

Figure n° 15

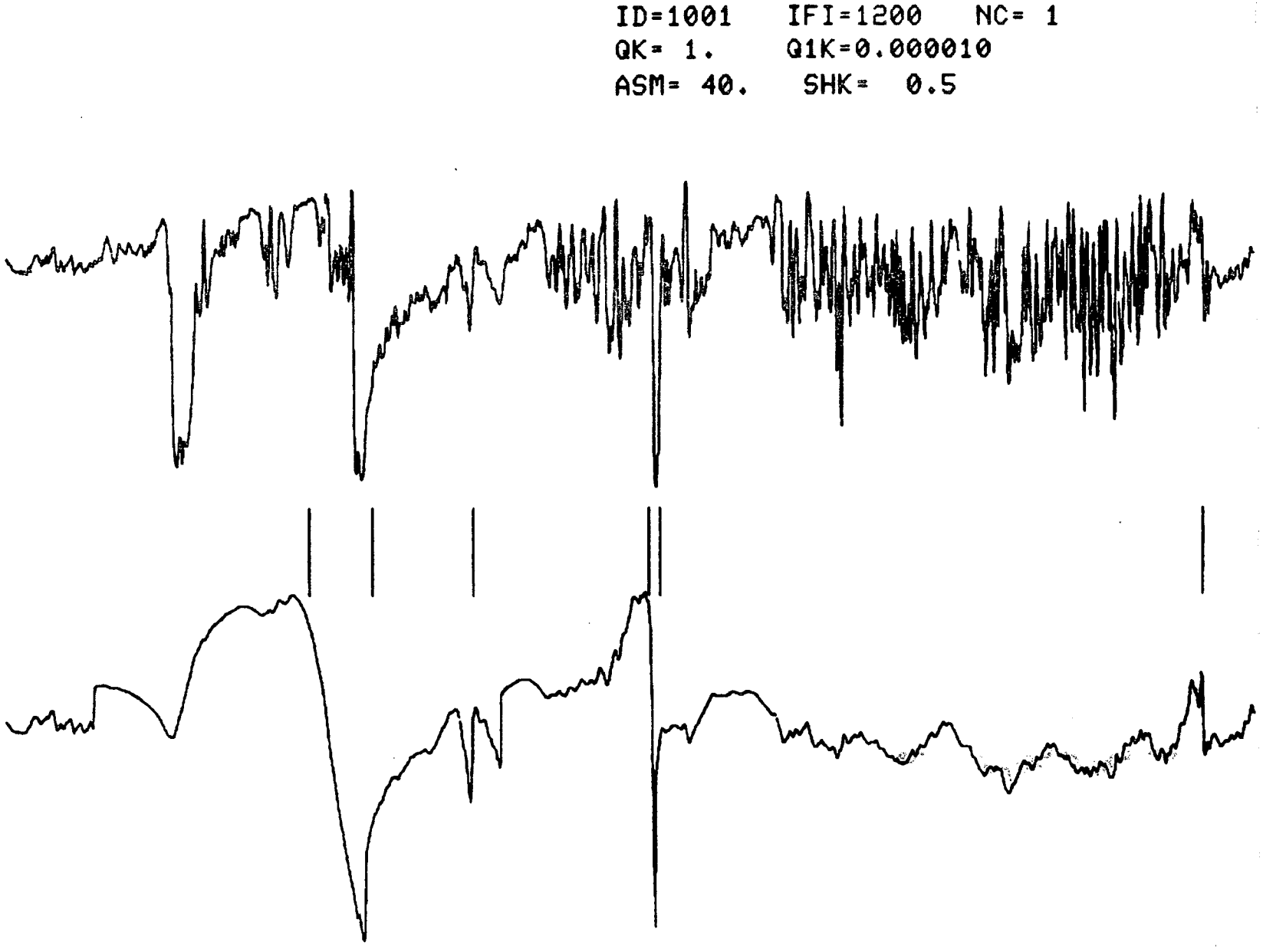

Figure n° 16

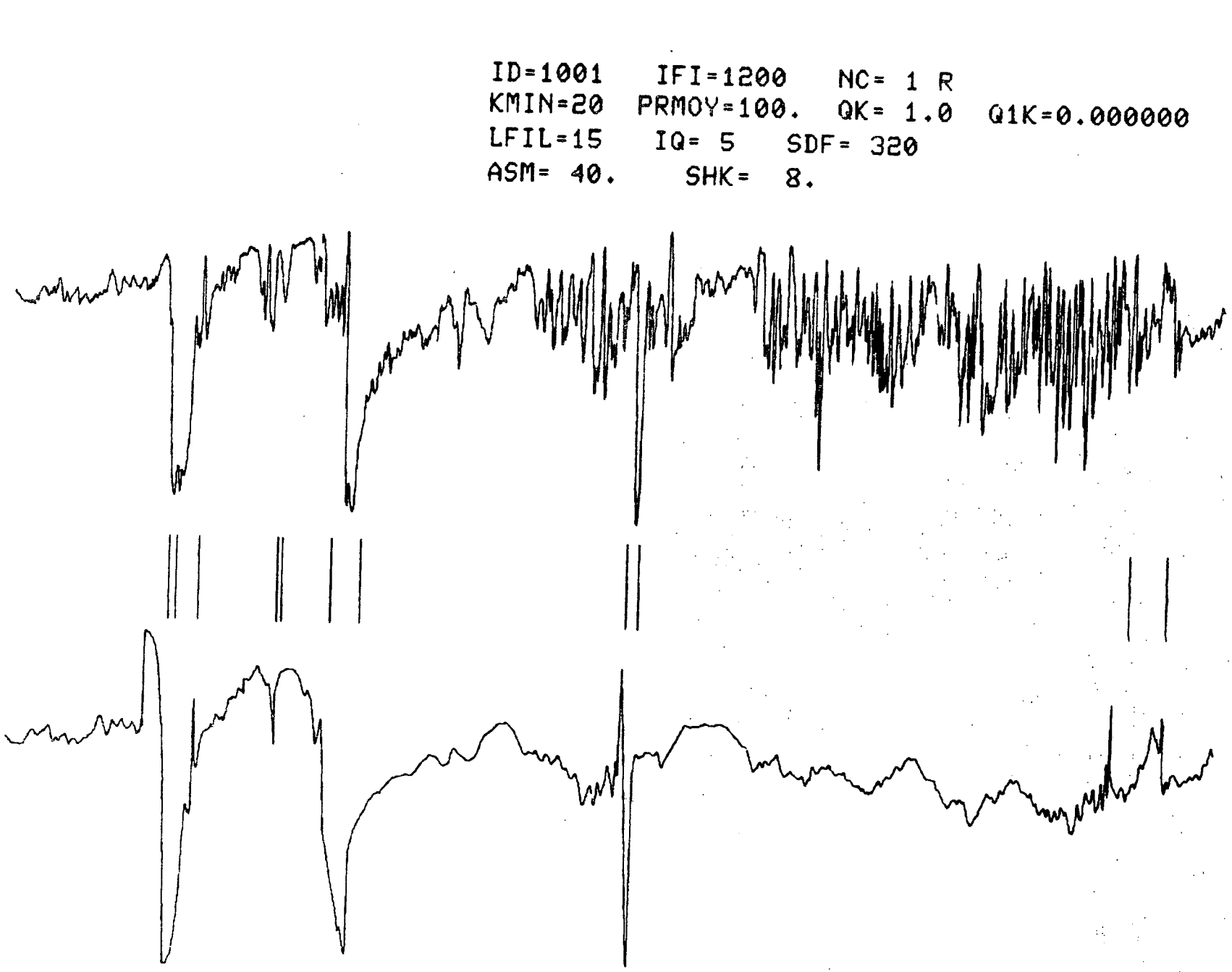
ID=1001 IFI=1200 NC= 1 R
KMIN=20 PRMOY=100. QK= 1.0 Q1K=0.000000
LFIL=15 IQ= 5 SDF= 320
ASM= 40. SHK= 8.

Figure n° 17

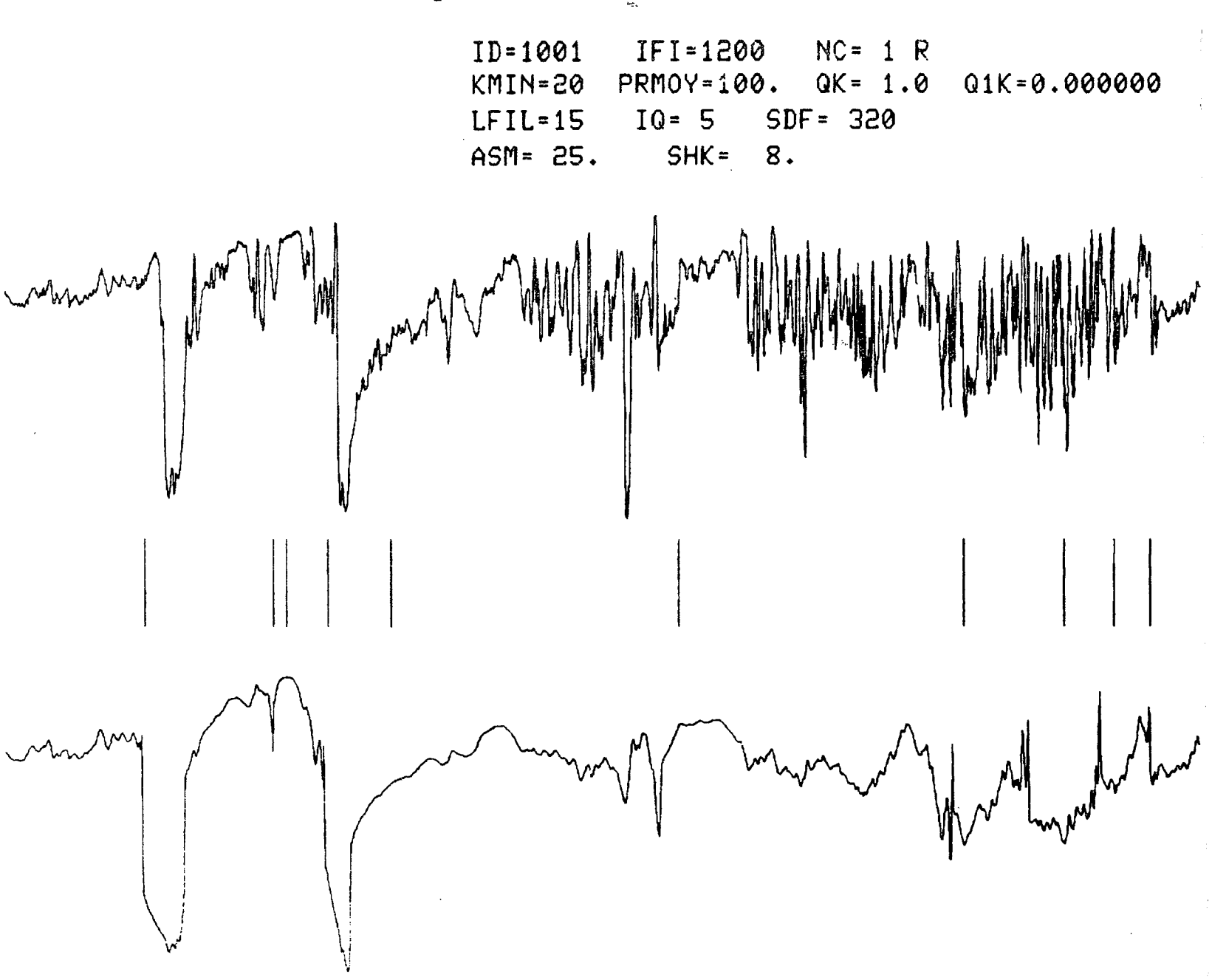
ID=1001 IFI=1200 NC= 1 R
KMIN=20 PRMOY=100. QK= 1.0 Q1K=0.000000
LFIL=15 IQ= 5 SDF= 320
ASM= 25. SHK= 8.

Figure n° 18

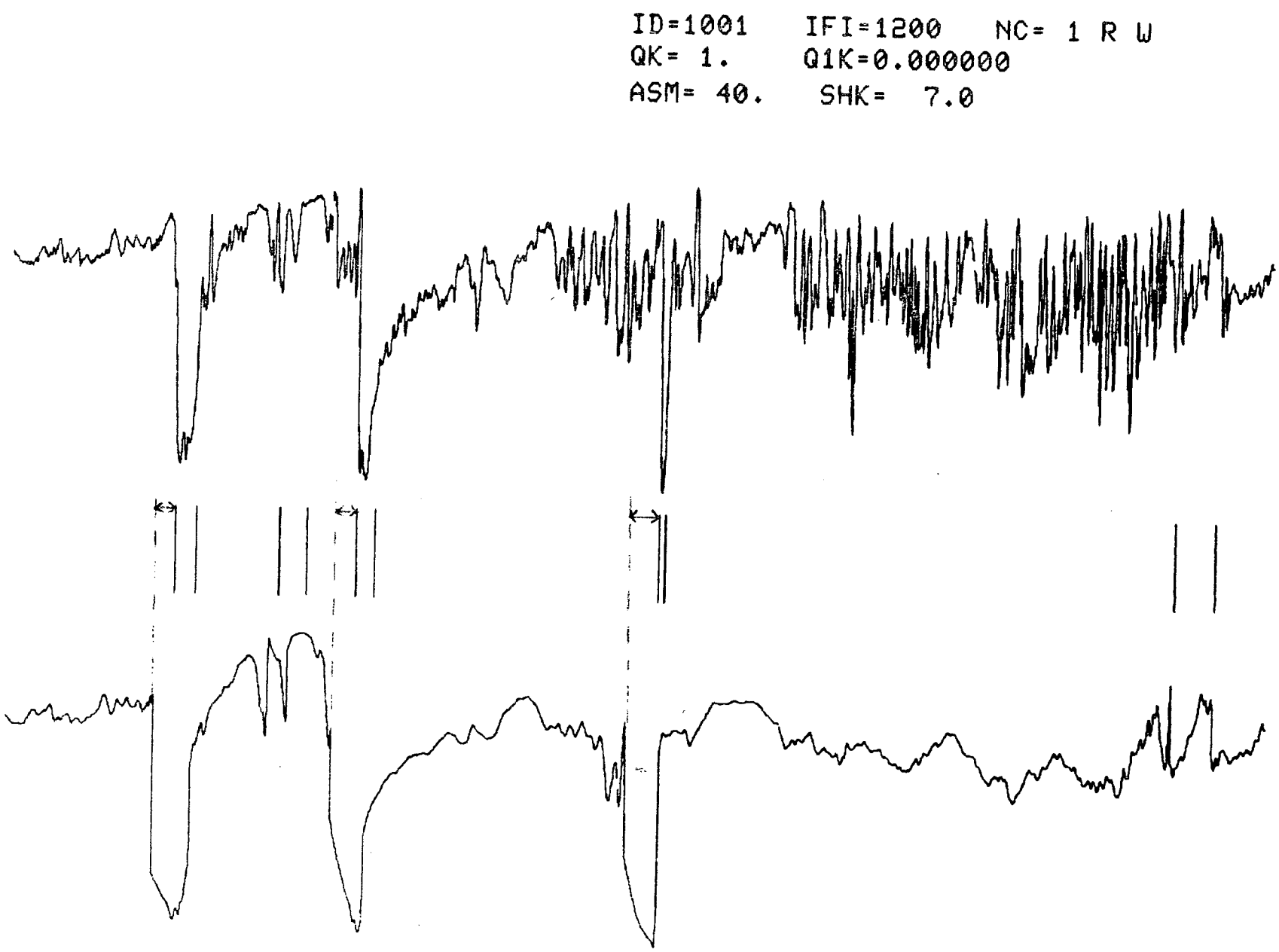

Figure n° 19

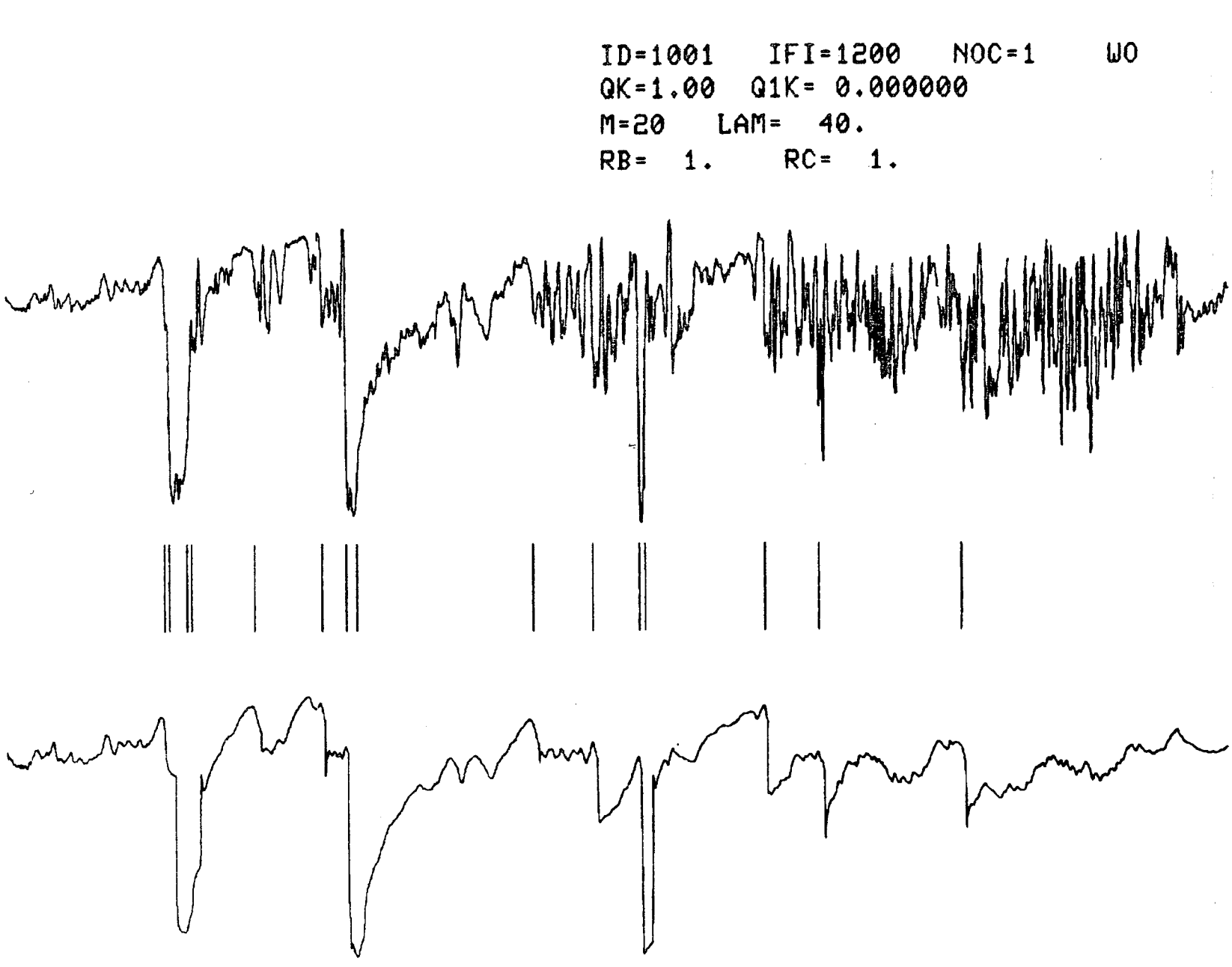

Figure n° 20

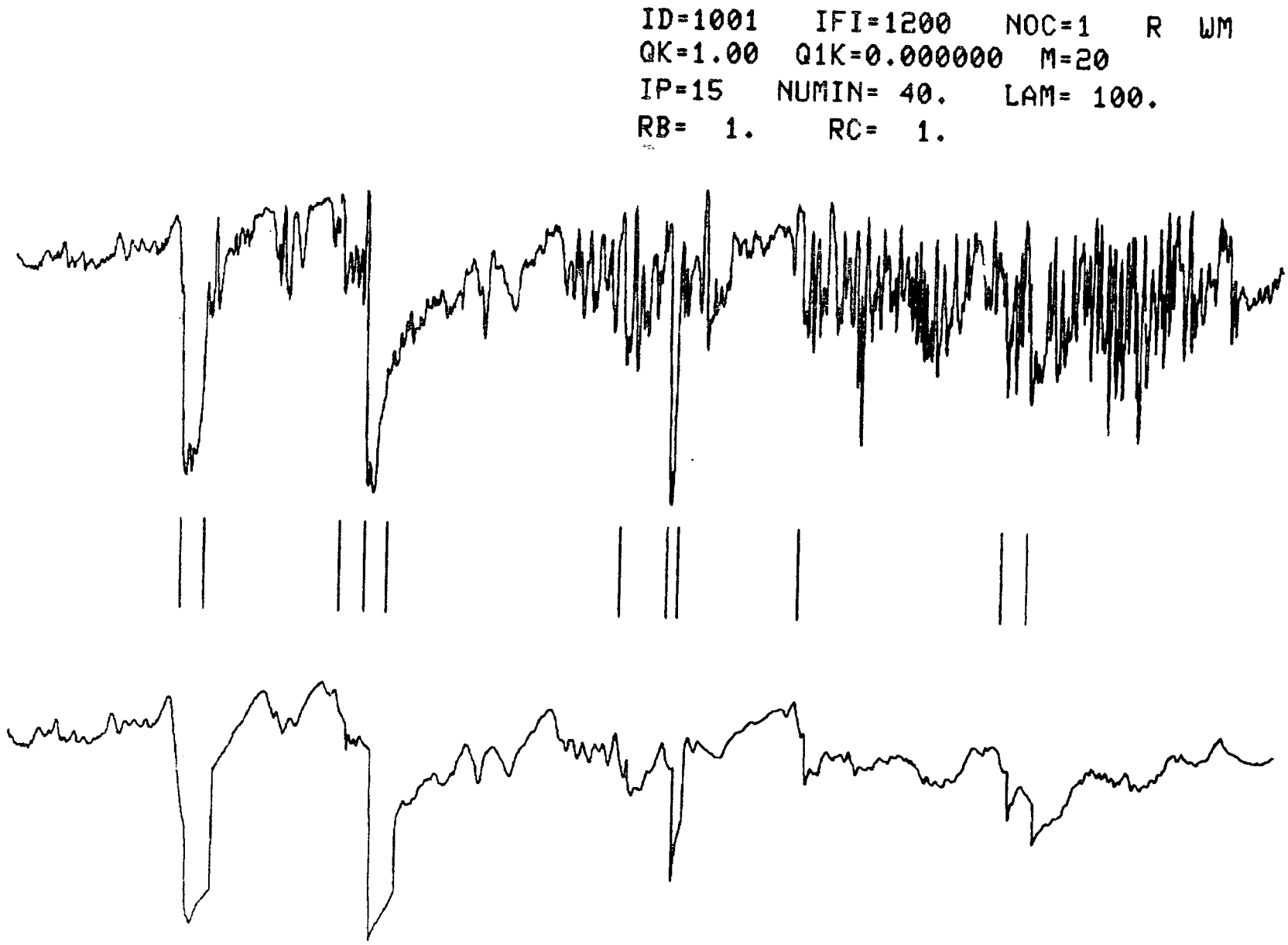
ID=1001 IFI=1200 NOC=1 R WM
QK=1.00 Q1K=0.000000 M=20
IP=15 NUMIN= 40. LAM= 100.
RB= 1. RC= 1.

Figure n° 21

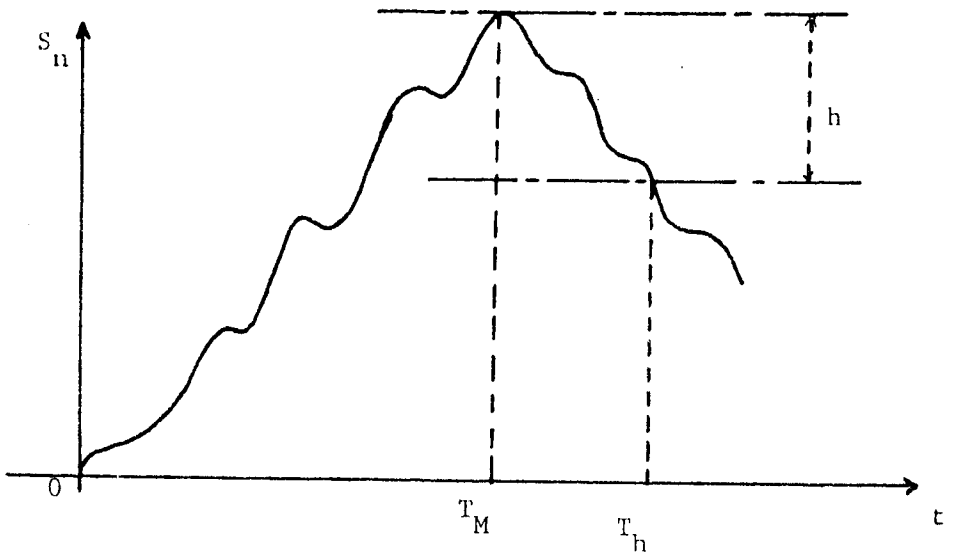

Figure n° 22 : Hinkley's cumulative sum test

ACKNOWLEDGEMENT

The authors wish to thank Dr. G. RUCKEBUSH (Schlumberger-Ridgefield) for an introduction to the special problem considered in this paper, Pr. A.S. WILLSKY (M.I.T.) for very fruitful discussions about failure detection, and C. CLAVIER and J. HARRY (Schlumberger-EPS, Clamart, France) for discussions concerning the geological features of the problem.

BIBLIOGRAPHY

[1] M. BASSEVILLE, B. ESPIAU, J. GASNIER (1981):
"Edge Detection Using Sequential Methods for Change in Level - Part I: A Sequential Edge Detection Algorithm".
IEEE Trans. on Acoustics, Speech and Signal Processing. Vol. ASSP-29, N°1, pp. 24-31. Feb. 81.

[2] M. BASSEVILLE (1981):
"Edge Detection Using Sequential Methods for Change in Level - Part II: Sequential Detection of Change in Mean".
IEEE Trans. on ASSP. Vol. ASSP-29, N°1, pp. 32-50. Feb. 81.

[3] M. BASSEVILLE (1981):
"Changes in statistical models: various approaches in Automatic Control and Statistics".
Submitted to publication to IEEE Trans. on A.C.

[4] R. BUENO, E.Y. CHOW, K.P. DUNN, S.B. GERSHWIN, A.S. WILLSKY (1976):
"Status Report on the Generalized Likelihood ratio failure detection technique, with application to the F8-aircraft".
Proc. IEEE Conf. on Decision and Control. Clearwater. Florida. Dec. 76; pp. 38-47.

[5] C.B. CHANG, K.P. DUNN (1978):
"A Recursive Generalized Likelihood Ratio Test Algorithm for Detecting Sudden Changes in Linear Discrete Systems".
Proc. IEEE Conf. on Decision and Control. San Diego. California. 1978; pp. 731-736.

[6] C.B. CHANG, K.P. DUNN (1979):
"On GLR Detection and Estimation for Unexpected Inputs in Linear Discrete Systems".
IEEE Trans. on Automatic Control. Vol. AC-24, N°3, pp. 499-501. June 79.

[7] E.Y. CHOW (1980):
"Failure Detection System Design Methodology".
Thesis. Laboratory for Information and Decision Systems. M.I.T. Cambridge. Massachussetts. Oct. 80.

[8] J. GASNIER (1980):
"Détection et Suivi de Contours: Etude Comparative de Méthodes Séquentielles et Application à des Radiographies et des Images T.V.".
Thèse 3ème cycle. Univ. of Rennes. France. Oct. 80.

[9] D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER (1978):
"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part I: Identification of Persistent Rhythms".
IEEE Trans. on Biomedical Engineering. Vol. BME 25, N°4, pp. 344-353. July 78.

[10] D.E. GUSTAFSON, A.S. WILLSKY, J.Y. WANG, M.C. LANCASTER, J.H. TRIEBWASSER (1978):
"ECG/VCG Rhythm Diagnosis using Statistical Signal Analysis. Part II: Identification of Transient Rhythms".
IEEE Trans. on Biomedical Engineering. Vol. BME 25, N°4, pp. 353-361.

[11] D.V. HINKLEY (1971):
"Inference about the change-point from cumulative sum-tests".
Biometrika. Vol. 58, N°3, pp. 509-523.

[12] R.K. MEHRA (1970):
"On the Identification of Variances and Adaptive Kalman Filtering".
IEEE Trans. on A.C. Vol. AC-15, N°2, pp. 175-184. April 70.

[13] E.S. PAGE (1954):
"Continuous inspection schemes".
Biometrika. Vol. 41; pp. 100-114.

[14] A.S. WILLSKY (1976):
"A survey of Design Methods for Failure Detection in Dynamic Systems".
Automatica. Vol. 12; pp. 601-611.

[15] A.S WILLSKY, H.L. JONES (1974):
"A Generalized Likelihood Ratio Approach to State Estimation in Linear Systems Subject to Abrupt Changes".
Proc. IEEE Conf. on Decision and Control. Phoenix. Arizona. Nov. 74. pp. 846-853.

[16] A.S. WILLSKY, H.L. JONES (1976):
"A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems".
IEEE Trans. on A.C. Vol. AC-21, N°1, pp. 108-112. Feb. 76.

[17] A.S. WILLSKY, E.Y. CHOW, S.B. GERSHWIN, C.S. GREENE, P.K. HOUPT, A.L. KURKJIAN (1980):
"Dynamic Model-Based Techniques for the Detection of Incidents on Freeways".
IEEE Trans. on A.C. Vol. AC-25, N°3, pp. 347-360. June 80.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique