A FILTER LENGTH CRITERION FOR MINIMUM ENTROPY DECONVOLUTION

T.J. ULRYCH†, C.J.WALKER† and A.J. JURKEVICS*

ABSTRACT

The Minimum Entropy Deconvolution (MED) technique is an important advance in deconvolution techniques. The usual assumption of minimum phase when the disturbing function is not known a priori is obviated in this approach. This paper presents a simple criterion for determining the length of the MED operator which is often a sensitive parameter in the deconvolution. We present examples of the application of this criterion to both scalar and complex time series.

1. INTRODUCTION

Recently, *Wiggins (1977)* has introduced a method of deconvolution which shows considerable promise in diverse applications. This method, which is applicable to processes which may be considered as the convolution of a band-limited function with a series of impulses, is called the Minimum Entropy Deconvolution (MED) technique. MED has the considerable advantage over conventional approaches in that it does not require any phase assumptions about the disturbing function, nor does it require the assumption of the randomness of the impulse series.

The efficiency of MED, particularly when the signal to noise ratio is small, can be improved, as shown by *Ooe and Ulrych (1978)*, by modifying the norm used by Wiggins by incorporating an exponential transformation into the algorithm. Both of these approaches require a specification of the length of the deconvolution or MED operator. As it turns out, the deconvolved output is rather sensitive to this parameter, and we present in this paper a very simple criterion which we have found useful in determining the length of the MED operator. The examples in this paper illustrate our approach by applying the criterion to the MED and the exponential MED (MEDEX) algorithms. A final example illustrates the extension of the Wiggins technique to the complex case which has potentially important application in studies of the earth's polar motion [*Smylie and Mansinha (1968)*] and holographic deconvolution [*Farr (1970); Treitel (1974)*].

2. THEORETICAL CONSIDERATIONS

2.1 The varimax norm

Consider a set of input signals

$$x_{it} \quad i=1,2,\ldots,N_s \ ; \ t=1,2,\ldots,N_t$$

where N_s and N_t are the number of traces and number of input samples respectively.

The x_{it} are modelled as the convolution of a time invariant wavelet, w_t, with impulsive signals, q_{it}.

Thus,

$$x_{it} = \sum_{m=1}^{n} w_m q_{i,t-m+1} + n_{it} \qquad (1)$$

where n_{it} represents additive noise.

†Dept. Geophysics and Astronomy, University of British Columbia, Vancouver, Canada.
*DelMar Technical Associates, DelMar, California.

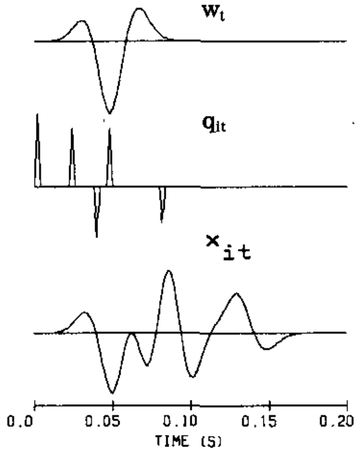

Fig. 1. The input to the scalar example.

The deconvolution operation is expressed in the form

$$y_{it} = \sum_{\ell=1}^{L} f_\ell \, x_{i,t-\ell+1} \qquad (2)$$

where y_{it} is the estimate of q_{it} and f_t is the deconvolution operator. The conventional least squares estimate of f_t is based on the assumption that the q_{it} are white noise series and that w_t, and hence f_t, have the minimum phase property. The MED technique of *Wiggins (1977)* does not require these restrictive assumptions. Instead, the approach insists that each signal, q_{it}, consists of a few large spikes of unknown and arbitrary amplitude and location, separated by values close to zero. In other words, we follow the principle of parsimony and insist on q_{it} having a simple structure.

A norm which measures simple structure has been used for years by statisticians in the field of factor analysis. This norm, which is called the varimax norm, can be understood in terns of the fourth standardized moment, α_4, or the kurtosis of the probability distribution of the q_{it}. In terms of the cumulants, K_n, of the probability distribution, the kurtosis is

$$\alpha_4 = K_4/K_2^2 \qquad (3)$$

The value of α_4 is dependent on the shape of the hump and on the length and height of the tails of a single-humped distribution. Thus, maximizing the kurtosis has the effect of increasing the relative frequencies of the large and near zero spikes compared to the spikes of intermediate size.

In keeping with the definition of α_4, the varimax norm which is maximized is given by

$$V = \sum_{i=1}^{N_s} V_i = \sum_{i=1}^{N_s} \left\{ \sum_{j=1}^{N_t} y_{ij}^4 \Big/ \left[\sum_{j=1}^{N_t} (y_{ij})^2 \right]^2 \right\} \qquad (4)$$

This, then, is a mathematical measure of parsimony or simple structure applied by *Wiggins (1977)*. A discussion of this and other norms and their relationship to factor analysis is presented by *Ooe and Ulrych (1978)*.

Now a word about the name of the technique which we are describing. Inasmuch as the principle of maximum entropy with which we have recently become familiar [*Burg (1975); Ulrych and Bishop (1975)*] describes randomness or uncertainty, so minimum entropy is an expression of simplicity and certainty. Since the problem which *Wiggins (1977)* addresses is the deconvolution of an impulse series which consists of a few large spikes, this simple structure led him to adopt the description of minimum entropy. It should be realized, however, that the simple structure is achieved by the maximization of the varimax and not the minimization of the entropy.

2.2 The normal equations

We consider, briefly, the equations which arise from maximizing the varimax norm given by (4) with respect to the coefficients, f_k.

We have

$$\frac{\partial V}{\partial f_k} = 0 = \sum_i \frac{\partial V_i}{\partial f_k}$$

$$= \sum_i \left(4V_i u_i^{-1} \sum_j y_{ij} \frac{\partial y_{ij}}{\partial f_k} - 4u_i^{-2} \sum_j y_{ij}^3 \frac{\partial y_{ij}}{\partial f_k} \right) \qquad (5)$$

where $u_i = \sum_j y_{ij}^2$ is the sample variance.

Since, from (2), we have $\frac{\partial y_{ij}}{\partial f_k} = x_{i,j-k}$

we can write (5) as

$$\sum_\ell f_\ell \sum_i v_i u_i^{-1} \sum_j x_{i,j-\ell+1}\, x_{i,j-k+i}$$

$$= \sum_i u_i^{-2} \sum_j y_{ij}^3\, x_{i,j-k+1} \qquad (6)$$

In matrix form, (6) becomes

$$\underline{R}\ \underline{f} = \underline{g} \qquad (7)$$

where $\underline{R}$ is a Toeplitz autocorrelation matrix consisting of weighted input autocorrelations, $\underline{g}$ is a column vector of weighted cross correlations of the cube of the outputs with the inputs and $\underline{f}$ is the column vector of the filter coefficients.

It is clear from (6) that the MED filter is designed to shape x_{it} into y^3_{it}. It is not surprising, therefore, that in MED processing, large amplitude spikes are enhanced compared to smaller amplitude ones. Although this is a desirable characteristic from the point of view of noise suppression, it is restrictive as far as the identification of small reflection coefficients is concerned. In order to achieve a balance in the deconvolution process, *Ooe and Ulrych (1978)* have incorporated an exponential transformation in the MED algorithm, defined as

$$z_{it} = 1 - \exp(-y_{it}^2 / 2s^2)$$

where s is a constant.

The new varimax norm which incorporates this transformation is defined as

$$V_e = \sum_i \left\{ \sum_j z_{ij}^2 \;/\; \left(\sum_j z_{ij}\right)^2 \right\}$$

The logic behind this transformation, the resulting normal equations and its application in the case of input signals with noise are discussed by *Ooe and Ulrych (1978)*. In this paper we will only compare the results of the Ve and Vnorms.

The scalar case described above has recently been extended to the complex case by *Ulrych, Ooe and Walker (1978)*. Here we simply mention that the normal equations can be written as

$$\underline{R}\ \underline{f}^* = \underline{g} \qquad (8)$$

where * indicates the complex conjugate and the quantities in (8) are defined in (9) which is the complex equivalent of (7).

$$\sum_\ell f_\ell^* \sum_i v_i u_i^{-1} \sum_j x^*_{i,j-\ell}\, x_{i,j-k}$$

$$= \sum_i u_i^{-2} \sum_j y_{ij} (y^*_{ij})^2 x_{i,j-k} \qquad (9)$$

where $u_i = \sum_j y_{ij} y^*_{ij}$

2.3 Computational procedure

Although (7) and (8) are highly non-linear their solution may be conveniently obtained in an iterative manner. A considerable simplification in this procedure is due to the fact that in both cases the matrix $\underline{R}$ is Toeplitz. In the complex case it is in fact of the block-Toeplitz type. Consequently the Levinson recursion [*Robinson (1967)*] can be used in the solution of both equations. The iterative solution consists of assuming an initial filter $\underline{f}^o$ which in this work was

$$\underline{f}^o = (0,0,\ldots,0,1,0,\ldots,0,0)^T$$

computing $\underline{R}$ and $\underline{g}$, solving for the updated filter, recomputing $\underline{R}$ and $\underline{g}$ and proceeding in this manner until convergence, as expressed by a constant varimax (*see Fig. 2*) has been achieved.

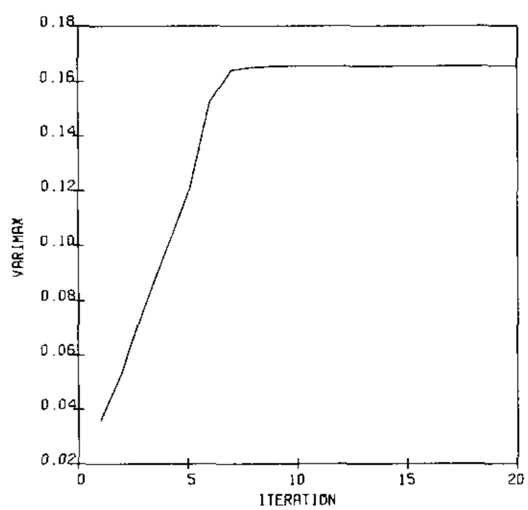

Fig. 2. Typical behaviour of the varimax norm as a function of iteration number.

2.4 A simple MED operator length criterion

Experience with MED shows that the deconvolved output is highly dependent on the choice of L, the length of the MED operator [equation (2)]. We have found that L, particularly for a single trace input, is dependent on the source wavelet and on the position of the spikes. For overly long filter lengths, the output may be unduly simplified. Too short a filter length may not resolve the output spikes.

A criterion which we have found useful and which we illustrate with examples is very simple. If $\underline{y}_i(L_1)$ and $\underline{y}_i(L_2)$ are the column vectors of the MED outputs determined from (2) for two lengths L_1 and L_2, we define a squared error to be

$$E(L) = \sum_i \left\{ \left[\underline{y}_i(L_1) - \underline{y}_i(L_2) \right]^T \cdot \left[\underline{y}_i(L_1) - \underline{y}_i(L_2) \right] \right\}$$

we take the correct length, L, to be that length which corresponds to [E(L)] min. Our assumption is that the change in the deconvolved output for perturbations in the filter length will be least when centered around the correct filter length. Fig. 3 illustrates a fairly typical variation of log [E(L)] with filter length. We remark that for large values of L the deconvolved output will have a very simple structure and E(L) will be very small. However, this condition is easily recognised and no ambiguity exists. In some cases, particularly when long filters are required, an ambiguity does exist in the sense that a number of minima in E(L) occur. In this case the choice of the correct length must be determined by examining the corresponding outputs.

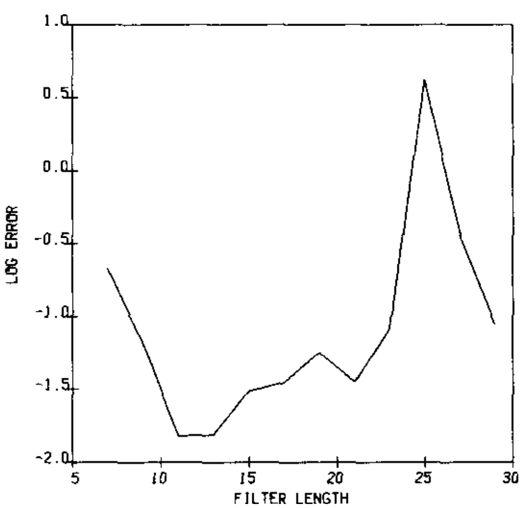

Fig. 3. Variation of log[E(L)] with filter length; MED, *noise free case.*

A different but complimentary approach to the choice of the correct filter length is illustrated in Figures 4 and 5 which show

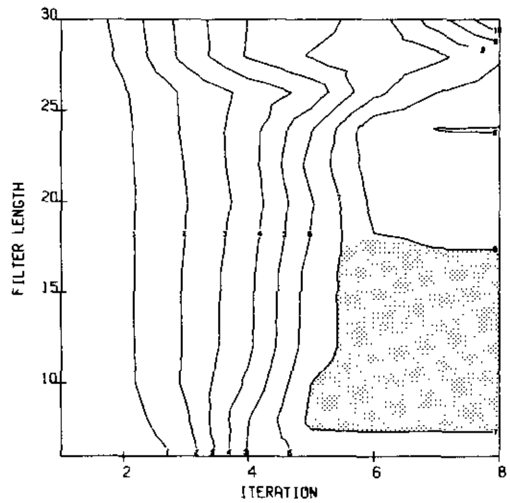

Fig. 4. Variation of varimax as a function of filter length and number of iterations; MED, noise free case.

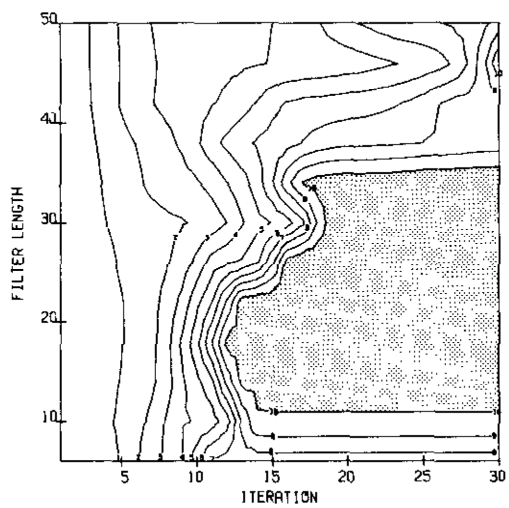

Fig. 5. Variation of varimax as a function of filter length and number of iterations: MEDEX, $s = y_{max}/2$, *noise free case.*

contours of varimax as a function of filter length and number of iterations. We have found that the correct length is indicated by regions such as the stippled areas shown in these figures, where the varimax remains constant for variations in L and the number of iterations.

3. *Numerical Examples*

Various examples of MED processing for both real and synthetic data have been presented by *Wiggins (1977)*. *Ooe and Ulrych (1978)* have compared the MED and MEDEX techniques for both single and multi-trace inputs. In this paper we are primarily concerned with illustrating the filter length criterion which we have discussed. We do this by considering both scalar and complex time series as inputs to the MED and MEDEX algorithms.

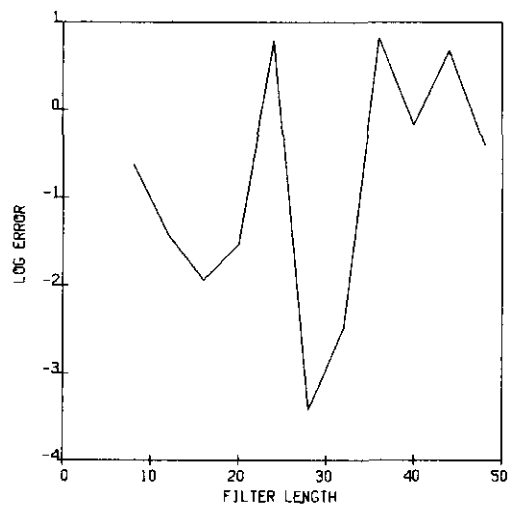

Fig. 6. Variation of log[E(L)] with filter length; MEDEX, s = $y_{max}/2$, noise free case.

3.1 *Scalar time series input*

The input, illustrated in Fig. 1, consists of the convolution of a mixed delay wavelet with a sequence of five impulses. The deconvolution is first considered when the input is noise free. A typical varimax diagram is shown in Fig. 2, and the behaviour of the criterion E(L) is shown in Fig. 3 and Fig. 6 for MED and MEDEX algorithms. Figures 7 and 8 illustrate the deconvolved outputs using MED and MEDEX for various filter lengths, as well as for the one given by [E(L)] min. These figures demonstrate the usefulness of the E(L) criterion, as well as the reduction in noise level of the deconvolved output when MEDEX is used. The effect of overly long filter lengths is also shown in these figures. Figures 9 to 12 illustrate the deconvolution when the signal to noise ratio is 32 dB. the value of s in the MEDEX algorithm was chosen on the basis of past experience to be s = $y_{max}/2$. The MED

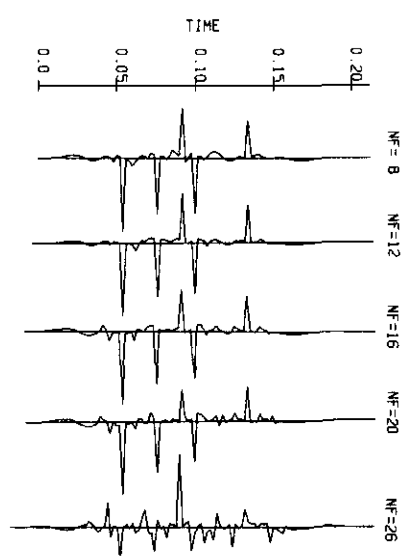

Fig. 7. MED output for various filter lengths (NF); noise free case.

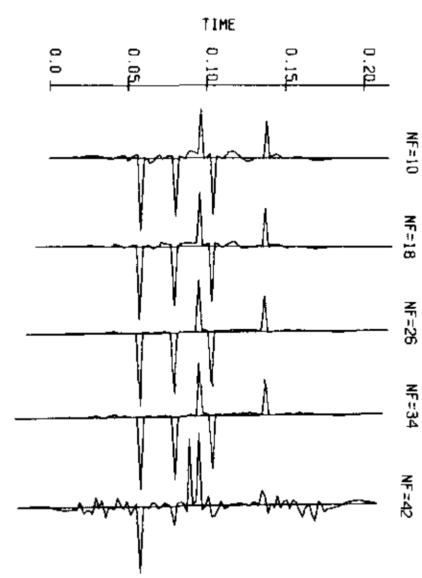

Fig. 8. MEDEX output, s = $y_{max}/2$, noise free case.

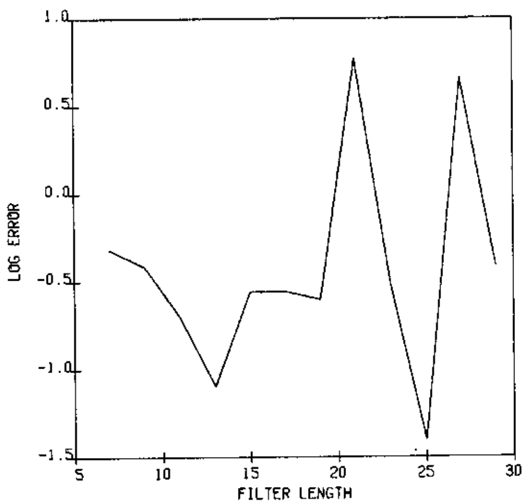

Fig. 9. Variation of log[E(L)] with filter length; MED, 2.5% noise.

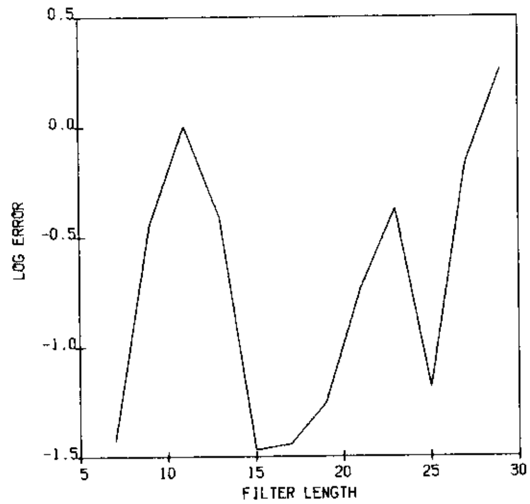

Fig. 11. Variation of log[E(L)] with filter length; MEDEX 2.5% noise.

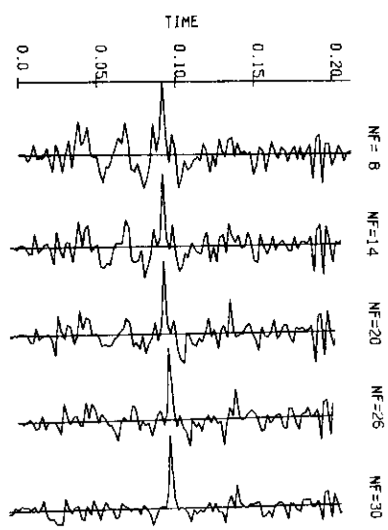

Fig. 10. MED output, 2.5% noise.

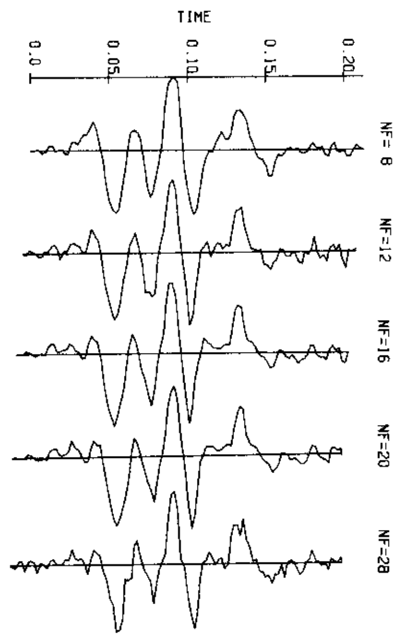

Fig. 12. MEDEX output, s = $y_{max}/2$, 2.5% noise.

output is very noisy in this case, and it is difficult to evaluate the effectiveness of E(L). The improvement in the deconvolution using MEDEX is illustrated in Fig. 12. Although the output does not change very much with L, $E(L)_{min}$ appears to indicate the optimum length correctly. The effectiveness of E(L) also applies to multi-trace inputs with which we have experimented.

3.2 Complex time series input

The polar motion of the earth is an excellent example of a complex time series in geophysics. One theory of the Chandler wobble [*Smylie and Mansinha (1968)*] is to model it as the response of a damped harmonic oscillator to excitation by large earthquakes which can be represented as a random walk process. The object, then, is to deconvolve the polar motion and obtain the excitation signature [*Smylie et. al. (1970)*].

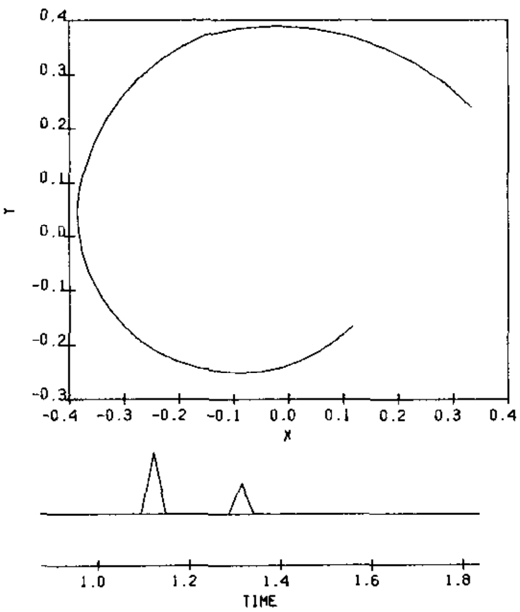

Fig. 13. One year of synthetic Chandler wobble data with Q = 50. The derivative of the step function excitation is shown below.

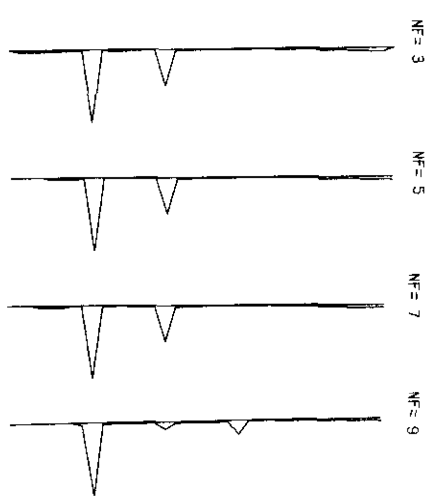

Fig. 14. MED output for various filter lengths (NF); noise free case.

Fig. 13 shows a synthetic wobble with a frequency of 0.82 cycles per year and a Q of 50. The derivative of the step function input is also shown in this figure. We have deconvolved this input, both in the absence and in the presence of additive noise, by first of all differencing it and then applying the complex version of MED described by equation (9). The outputs, for various filter lengths, are shown in Figures 14 and 15. The optimal filter length as indicated by our criterion was 5 in both instances. It can be seen that even when the input series is short (we have only used about one year of synthesized data) the outputs compare favourably with the input excitation.

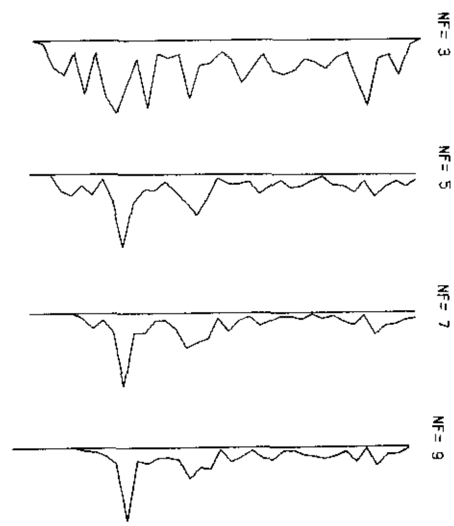

Fig. 15. MED output for various filter lengths (NF); added noise is 20% of the step amplitude.

4. Conclusions

The MED technique of *Wiggins (1977)* is a new and important approach to the problem of deconvolution. The output of the MED process is sensitive to the choice of the length of the MED operator, particularly in the case of a single input trace. This paper presents a simple criterion for estimating the correct length and illustrates the procedure with examples of scalar and complex input series. The extension of the MED technique to the complex case is important in the study of the excitation mechanism of the earth's polar motion. The improvement in the signal to noise ratio of the deconvolved output as a result of the exponential transformation introduced into the MED algorithm by *Ooe and Ulrych (1978)* is evident in examples presented in this paper.

References

Burg, J.P., 1975, Maximun Entropy Spectral Analysis, Ph.D. dissertation: Stanford University, Stanford, CA.

Farr, J.B., 1970, *Acoustical holography* experiments using digital processing: Acoustical Holography, v. 2, chap. 16, p. 225-242.

Ooe, M. and Ulrych, T.J., 1978, Minimum entropy deconvolution with an exponential transformation. Submitted to Geophysical Prospecting.

Robinson, E.A., 1967, Multichannel Time Series Analysis with Digital Computer Programs: Holden Day, San Francisco.

Smylie, D.E., Clarke, G.K.C., and Mansinha, L., 1970, Deconvolution of the Pole Path in Earthquake Displacement Fields and the Rotation of the Earth. (L. Mansinha, D.E. Smylie and A.E. Becks eds.) p. 136: D. Reidel.

Smylie, D.E. and Mansinha, L., 1968, Earthquakes and the Observed Motion of the Rotation Pole: J. Geophys. Res. *73*, p. 7661-7673.

Treitel, S., 1974, The complex Wiener filter: Geophysics, Vol. 39, No. 2.

Ulrych, T.J., Ooe, M. and Walker, C.J., 1978, In preparation.

Ulrych, T.J. and Bishop, T. 1975, Maximum entropy spectral analysis and autoregressive decomposition: Reviews of Geophysics, *13*, p. 183-200.

Wiggins, R.A., 1977, Minimum entropy deconvolution: Proc. Int. Symp. Computer Aided Seismic Analysis and Discrimination, June 9 & 10, Falmouth, Mass., IEEE Computer Society, p. 7-14.