

DECONVOLUTION DE PROCESSUS MULTI-IMPULSIONNELS PAR ALGORITHMES RAPIDES: APPROCHES AR ET MA.

Yves GOUSSARD (*), Guy DEMOMENT (*), Yves GRENIER (**)

(*) Laboratoire des Signaux et Systèmes (CNRS/ESE)
Plateau du Moulon, 91190 Gif-sur-Yvette, France

(**) Département Systèmes et Communications (ENST)
46, rue Barrault, 75634 Paris Cédex 13, France

(*) et (**) Gréco SARTA "Systèmes Adaptatifs".

RESUME.

Ce texte présente deux approches du problème de la déconvolution de signaux de type multi-impulsionnel ou Bernoulli-Gaussien, filtrés linéairement. La première approche suppose que le filtre est à réponse impulsionnelle finie (filtre MA) tandis que la seconde utilise un modèle de type autorégressif (filtre AR).

ABSTRACT.

This paper describes two approaches to the problem of deconvolution of multipulse or Bernoulli-Gaussian signals, filtered linearly. The first solution is based upon a finite impulse response (moving-average or MA) filter while the second one uses an autoregressive model (AR filter).

INTRODUCTION.

Nous nous intéressons ici à la déconvolution de signaux multi-impulsionnels. Ce problème se rencontre dans une grande variété de domaines: sismique, codage de la parole, contrôle non-destructif, échographie. On se limite au cas monodimensionnel.

On considère que le signal est dégradé par un processus linéaire de type convolution. La résolution complète du problème nécessite plusieurs opérations: identification du système linéaire ou ondelette, estimation des hyperparamètres du problème, déconvolution du processus multi-impulsionnel lui-même. Pour la résolution complète du problème, ces opérations sont exécutées parfois en parallèle, le plus souvent séquentiellement.

Une des méthodes disponibles (Mendel, 1983) est fondée sur une représentation ARMA des signaux observés et fournit une solution élégante et complète, au prix cependant d'une complexité fort importante. Le but de cette communication est de présenter les résultats de deux approches différentes visant à réduire sensiblement cette complexité tout en conservant une qualité comparable du traitement. Ces approches utilisent des représentations respectivement AR et MA des signaux observés. La non-minimalité de ces représentations est compensée par l'emploi d'algorithmes rapides. Les hypothèses particulières à chacune des approches sont détaillées plus bas.

APPROCHE MA.

Dans cette approche, nous adoptons une représentation d'état dégénérée:

$$\begin{aligned} x_{k+1} &= x_k = x \\ z(k) &= h_k^T x \\ h_{k+1} &= Dh_k \end{aligned}$$

où x , vecteur d'état, contient l'ensemble des échantillons du signal à estimer, $z(k)$, h_k et D représentent respectivement l'observation à l'instant k , l'ondelette à l'instant k et un opérateur de décalage permettant de passer de l'observation à l'instant courant vers l'observation à l'instant suivant.

La particularité de cette représentation tient à ce que l'équation d'état ne comporte aucune dynamique, celle-ci étant entièrement reportée dans l'équation d'observation par l'intermédiaire de l'opérateur de décalage D . On a ainsi la possibilité d'obtenir en ligne une estimée lissée du signal, avantage qui s'ajoute à la robustesse commune aux représentations non-minimales. L'emploi d'algorithmes rapides permet d'éviter une trop forte augmentation du volume des calculs liée à la grande dimension du vecteur d'état x .

L'emploi de cette représentation a conduit à des résultats satisfaisants lorsque x est supposé gaussien (Demoment et al., 1984, Demoment et Reynaud, 1985). Mais un modèle gaussien est mal adapté à la représentation de signaux à caractère impulsionnel, et la méthode a été étendue aux signaux de type Bernoulli-gaussien (B-G):

$$x(k) = \gamma(k)w(k)$$

où w est un processus blanc gaussien centré de variance σ^2 et γ une séquence de Bernoulli définie par:

$$\begin{aligned} p[\gamma(k)=1] &= \lambda \\ p[\gamma(k)=0] &= 1-\lambda \end{aligned}$$

Par rapport au cas gaussien, l'emploi de ce modèle complique la méthode puisqu'une procédure de détection-estimation est nécessaire au lieu d'une simple opération d'estimation. Par ailleurs, on remarque que la loi de x conditionnée à γ est normale. L'espace étant limité, on se contentera ici de décrire qualitativement les principes de la méthode et de sa mise en oeuvre par algorithmes rapides. Pour plus de détails, se reporter à (Goussard et Demoment, 1987).

Principe de la méthode: Afin de préserver la cohérence de la méthode, il est fait appel à un critère unique pour les deux étapes de la déconvolution. Celui-ci est du type vraisemblance a posteriori (VAP):

$$V[\gamma, w/z] = p[z/\gamma] p[\gamma] p[z]$$

et, d'après la propriété d'invariance du maximum de vraisemblance, l'estimée de x sera obtenue comme le produit des estimées $\hat{\gamma}$ et \hat{w} de γ et w . La maximisation de V est en général un problème difficile, mais sous nos hypothèses, il peut se simplifier. En effet, (Kormylo et Mendel, 1982) ont montré que la maximisation de V est équivalente à la maximisation successive de V_d et V_e définis par:

$$\begin{aligned} V_d &= p[w, z/\gamma] p[\gamma] \\ V_e &= p[z/\gamma, w] p[w] \end{aligned}$$

L'intérêt de ce résultat vient de ce que V_d ne dépend pas



de w et peut être exprimé en fonction de z uniquement. Par conséquent, le problème de déconvolution peut être scindé en deux étapes successives: une étape de détection qui correspond à la maximisation de V_d , puis une étape d'estimation correspondant à la maximisation de V_e . De plus, V_e est la vraisemblance a posteriori de w en supposant la séquence de Bernoulli connue, et nous sommes alors ramenés au cas gaussien dans lequel les critères de maximum de vraisemblance, de variance minimale et de moindres carrés sont équivalents. Il est donc possible de calculer w par filtrage de Kalman sous l'hypothèse $\hat{\gamma} = \gamma$.

Une remarque importante pour la suite doit être faite sur le fonctionnement d'un tel filtre de Kalman. Comme le processus x est non corrélé, le filtre n'utilise à sa récursion courante k que les informations relatives au passé. En particulier, il n'est nécessaire de connaître γ que pour les indices compris entre 1 et k . Un de nos objectifs étant d'obtenir une procédure globalement réursive, il suffira pour y parvenir de montrer que l'on peut effectuer la détection de $\gamma(k)$ à partir des quantités disponibles à l'instant $k-1$.

La détection se fait par maximisation du critère V_d . Son expression exacte, calculée dans (Mendel, 1983) dépend de z et de la loi de probabilité de γ , facile à obtenir. La procédure de détection optimale consisterait à calculer la valeur de V_d pour toutes les séquences γ possibles et à choisir celle qui maximise V_d . Or il y a 2^N possibilités pour γ , N , nombre d'échantillons, étant généralement de l'ordre de plusieurs centaines. Ceci rend la détection optimale du processus de Bernoulli inenvisageable en raison du gigantesque volume de calcul qu'elle nécessiterait. Il faut donc maximiser V_d de manière sous-optimale, et en fait, les méthodes de déconvolution de processus B-G par maximum de vraisemblance se différencient par le type de sous-optimalité retenu.

Comme dans (Kormylo et al., 1982), V_d est approché par Υ , VAP de γ connaissant les observations:

$$\Upsilon = p[z/\gamma] p[\gamma]$$

Au lieu de maximiser Υ globalement et de manière sous-optimale, comme l'ont proposé (Kormylo et Mendel, 1982) avec le détecteur SMLR, nous préférons maximiser séquentiellement les critères $\Upsilon_{k,n}$ définis par:

$$\begin{aligned} \Upsilon_{k,n}[\gamma(k)/z^{k+n}] &= \text{VAP}[\gamma(k)/z^{k+n}] \\ &= p[z^{k+n}/\gamma(k)] p[\gamma(k)] \end{aligned}$$

où z^{k+n} désigne le vecteur des observations jusqu'à l'instant $k+n$, n étant un retard de lissage destiné à améliorer la détection. L'utilisation des critères $\Upsilon_{k,n}$ permet de préserver la structure globalement réursive de l'algorithme. Ils peuvent être mis sous la forme (Goussard et Demoment, 1987):

$$\Upsilon_{k,n} = p[z^{k,k+n}/z^{k-1}, \gamma(k)] p[\gamma(k)]$$

où $z^{k,k+n}$ désigne le vecteur des observations entre les instants k et $k+1$. La difficulté de la maximisation de $\Upsilon_{k,n}$ provient de ce que $p[z^{k,k+n}/z^{k-1}, \gamma(k)]$ dépend non seulement des quantités fournies par le filtre de Kalman à la récursion $k-1$, mais aussi de $\gamma(k), \dots, \gamma(k+n)$. En toute rigueur, son calcul nécessite l'inversion de 2^{n+1} matrices $(n+1, n+1)$ et le calcul des déterminants correspondants.

Cette procédure, coûteuse en temps de calcul, peut être simplifiée en approximant la loi de $[z^{k,k+n}/z^{k-1}, \gamma(k)]$ par une loi normale de même moyenne et de même matrice de covariance. L'expression approchée de $\Upsilon_{k,n}$ ne dépend plus alors que de $\gamma(k)$ et des quantités fournies par le filtre de Kalman à la récursion $k-1$, et son évaluation sous les deux hypothèses $\gamma(k)=0$ et $\gamma(k)=1$

ne nécessite qu'une inversion de matrice et un calcul de déterminant.

Mise en oeuvre par algorithmes rapides: La procédure proposée permet d'effectuer la déconvolution d'une séquence B-G en une seule passe et de manière réursive. Mais, avec la représentation d'état dégénérée que nous avons adoptée, le vecteur d'état x est de très grandes dimensions, et la remise à jour à chaque récursion de la matrice de covariance d'erreur de prédiction nécessite des calculs très lourds qui rendent la méthode numériquement peu attrayante.

Il est possible de réduire le volume des calculs de manière significative en remplaçant les équations de Riccati du filtre de Kalman standard par des équations de Chandrasekhar. Leur principe consiste à s'intéresser aux incréments des covariances et du gain plutôt qu'à ces quantités elles-mêmes. Le modèle d'état dégénéré étant invariant par décalage et non pas simplement invariant, la définition des incréments est modifiée en conséquence. Pour la dérivation précise des équations du filtre, voir (Goussard et Demoment, 1987).

Les équations de Chandrasekhar et le filtre de Kalman standard ne calculant pas exactement les mêmes grandeurs, il est nécessaire de vérifier qu'il est toujours possible d'effectuer le calcul du critère de détection, et de factoriser l'incrément initial de la matrice de covariance d'erreur de prédiction, bien que le processus γ ne soit pas connu à l'initialisation. Le premier point est résolu en remarquant que l'expression approchée du critère $\Upsilon_{k,n}$ peut être calculée en propageant une innovation vectorielle de dimension $n+1$ et les intermédiaires de calcul correspondants, dont seule la première composante sera utilisée lors de la phase d'estimation.

Quant au deuxième point, on démontre que, bien que la factorisation initiale exacte soit en effet impossible sans la connaissance a priori de la séquence de Bernoulli, la forme de cette factorisation ne dépend pas de la réalisation particulière de γ et que les matrices correspondantes peuvent être apprises au fur et à mesure du traitement des données. Il s'agit donc en quelque sorte d'une factorisation initiale aveugle pour laquelle seul doit être fixé a priori un rang de déplacement maximal, c'est à dire le nombre maximal d'impulsions présent dans le signal. On aboutit bien ainsi à une procédure globalement réursive et rapide de déconvolution de signaux B-G.

APPROCHE AR.

Le modèle employé ici est du type autorégressif ou tout-pôle, avec une entrée de type B-G. Soit $z(k)$ le signal observé, le modèle autorégressif est caractérisé par ses coefficients a_i ou plutôt $a_i(k)$ pour insister sur la non-stationnarité possible du modèle. L'entrée est $x(k)$ et le modèle s'écrit:

$$z(k) + a_1(k-1)z(k-1) + \dots + a_p(k-p)z(k-p) = x(k)$$

L'entrée $x(k)$, de type B-G, sera supposée contenir M impulsions, avec $M = \lambda N$. Comme souligné précédemment, la procédure de détection et estimation ne peut être que sous-optimale. La procédure retenue, calquée sur celle proposée par (Atal, 1982), procède de manière globale sur l'ensemble des échantillons $z(1), \dots, z(n)$, en détectant et estimant récursivement les impulsions $x(k_1), \dots, x(k_M)$, par minimisation d'un critère de moindres carrés.

La contrainte la plus importante sera la non-stationnarité du modèle, due à celle du signal: ainsi dans le traitement des signaux sismiques, le modèle se voulant une représentation de l'ondelette, il faudra prendre en compte les diverses atténuations qui affectent les échos successifs: les échos les plus lointains auront une amplitude plus faible, à cause du plus grand nombre

d'interfaces traversées, et des absorptions par les couches qu'ils ont parcourues, le même phénomène se rencontrant également

en échographie. Mais ceci n'est pas le plus gênant, car un contrôle de gain permet de réduire ces difficultés. C'est plutôt l'absorption différente suivant les fréquences, plus forte en haute fréquence qu'en basse fréquence, qui posera le plus de problème. Il va donc falloir utiliser un modèle générateur de l'ondelette (et de chaque écho) qui soit non-stationnaire, ce qui implique une méthodologie spécifique pour la modélisation. Celle retenue ici est une approche par modèles évolutifs (Grenier, 1980 et 1986) qui suppose les coefficients du modèle élaborés sur une base de fonctions du temps choisies à l'avance, $f_0(k) \dots f_m(k)$.

Une telle approche est issue de travaux de Mendel, Rao, Liporace, Hall ... Une revue de ces travaux pourra être trouvée en (Grenier, 1980). Le décalage des indices temps dans les coefficients $a_i(k)$ devenus $a_i(k-i)$, a permis de montrer que l'identification du modèle AR (autorégressif) scalaire non-stationnaire, était remplacée par celle d'un modèle AR vectoriel à coefficients invariants (stationnaires) sur le signal $Z(k)$ vectoriel constitué à partir des produits du signal scalaire $z(k)$ par les fonctions de base:

$$Z(k) = \begin{bmatrix} f_0(k)z(k) \\ \vdots \\ f_m(k)z(k) \end{bmatrix}$$

Si on suppose que les coefficients $a_i(k)$ du modèle s'écrivent:

$$a_i(k) = \sum_{j=0}^m a_{ij} f_j(k)$$

le modèle autorégressif se réécrit alors comme:

$$z(k) + \begin{bmatrix} Z^T(k-1) & \dots & Z^T(k-p) \end{bmatrix} \theta = x(k)$$

$$\theta = \begin{bmatrix} a_{10} & \dots & a_{1m} & a_{20} & \dots & a_{pm} \end{bmatrix}^T$$

L'équation devient ainsi une régression linéaire à coefficients constants du signal vectoriel $Z(k)$ sur le signal scalaire $z(k)$, ce qui permet de réaliser l'estimation de θ au sens des moindres carrés (Grenier, 1980, et 1986).

Le problème de l'estimation de l'entrée (B-G) $x(k)$ est résolu en recherchant les instants $k_1 \dots k_M$ où $\gamma(k)=1$, et les amplitudes $w(k)$ des impulsions.

Nous pouvons retenir comme critère d'erreur l'énergie de la différence entre le signal réel et le signal synthétisé par le modèle excité par son entrée multi-impulsionnelle. Le modèle est non-stationnaire, il a une réponse impulsionnelle bi-indice $h_{k,n}$ telle que:

$$z(k) = \sum_{n \leq k} h_{k,n} x(n)$$

On minimise l'erreur quadratique totale:

$$E = \sum_k \left[z(k) - \sum_{n \leq k} h_{k,n} \gamma(n) w(n) \right]^2$$

Le principe de cette méthode tient en un placement itératif des impulsions. Une première impulsion est placée: c'est parmi les T impulsions possibles celle qui minimise l'erreur E , puis on place une deuxième impulsion sans changer la précédente, et ainsi de suite, la n -ième impulsion sera placée après les $n-1$ premières, à la position où est minimisée l'erreur résiduelle. Soit $z_n(k)$ le signal synthétisé avec les n premières

impulsions:

$$z_0(k) = 0$$

$$z_n(k) = \sum_{i=1}^n h_{k,i} w(k_i)$$

La solution optimale pour le placement de l'impulsion suivante est: prendre k_{n+1} maximisant le rapport $\frac{r_n(k)}{\zeta(k)}$ où $r_n(k)$ est l'intercorrélacion entre $z_n(k)$ et la réponse impulsionnelle $h_{i,k}$, démarrant à l'instant k , $\zeta(k)$ étant l'autocorrélacion de cette même réponse impulsionnelle.

On peut remarquer que cette approche ne permet pas de remettre en cause les positions et les amplitudes des impulsions localisées les premières. Une variante proposée par (Atal et Remde, 1982) consiste à réactualiser l'ensemble des amplitudes estimées, et ceci à chaque positionnement d'une nouvelle impulsion. Ceci est cependant très coûteux, la résolution nécessitant $O(n^3)$ opérations, l'ensemble des calculs jusque $n=M$ représente un coût de $O(M^4)$ opérations, ce qui est énorme: dans l'application à la parole, M vaut au moins 32, pour les signaux sismiques, il y a au minimum 100 à 200 impulsions par trace. En pratique la différence de qualité entre la méthode la plus simple et cette variante est trop faible pour justifier ce coût accru.

Notons aussi que la procédure telle qu'elle vient d'être décrite n'est pas optimisée en temps de calcul. Nous avons pu montrer que l'on pouvait étendre dans le cas non-stationnaire la technique suivante: la séquence d'intercorrélacion s'obtient à partir du signal résiduel $z(k)-z_n(k)$ par un filtrage récursif, ou à réponse impulsionnelle infinie, mais en sens inverse du temps, au moyen du modèle autorégressif estimé. La simplification que cela apporte à l'algorithme est nette: à chaque instant, le calcul de $r_n(k)$ ne nécessite plus que $O(p)$ opérations, au lieu des opérations de convolution avec la réponse impulsionnelle, beaucoup plus coûteuses, mais le gain ne s'arrête pas là.

En effet, après avoir positionné l'impulsion, il faudrait recalculer le signal résiduel suivant, $z(k)-z_{n+1}(k)$. Cette opération serait suivie d'un nouveau filtrage en sens inverse du temps pour calculer l'intercorrélacion suivante. On peut se passer de ces deux opérations, en calculant directement $r_{n+1}(k)$ à partir de $r_n(k)$, on observe en effet que la correction à appliquer à $r_n(k)$ est simplement de soustraire l'autocorrélacion de $h_{i,k_{n+1}}$ centrée sur k_{n+1} avec une amplitude $w(k_{n+1})$.

Dans le cas non-stationnaire, il faut noter cependant deux détails, tout d'abord le filtrage autorégressif, en sens inverse du temps, pour obtenir $r_{n+1}(k)$ n'utilise pas les coefficients $a_i(k-i)$ mais $a_i(k)$, d'autre part l'autocorrélacion de la réponse impulsionnelle n'est plus une véritable autocorrélacion, mais s'obtient comme résultat du filtrage autorégressif en sens inverse du temps de la réponse commençant à l'instant k_n , ce qui conduit à une pseudo-corrélacion non symétrique.

CONCLUSION.

Nous avons présenté deux méthodes sous-optimales mais construites autour d'algorithmes rapides pour la déconvolution de signaux Bernouilli-Gaussiens. Elles se différencient par la forme du modèle décrivant la convolution (AR ou MA), par la méthode de détection, (séquentielle sur le temps, ou récursive sur la contribution décroissante à la reconstruction du signal), et enfin par le critère d'estimation, (maximum de vraisemblance dans un cas, variance minimale dans l'autre). Les figures 1 à 3 montrent un exemple de déconvolution (approche MA) sur une trace unique, la figure 4 un exemple (approche AR) sur plusieurs traces.



REFERENCES.

Atal, B.S. Remde, J.R. (1982), "A new model of LPC excitation for producing natural-sounding speech at low bit rates", IEEE ICASSP-82, pp 614-617.

Demoment, G., Reynaud, R., and Herment, A. (1984). "Range resolution improvement by a fast deconvolution method." Ultrasonic Imaging, Vol.6, pp 435-451.

Demoment, G., and Reynaud, R. (1985). "Fast minimum variance deconvolution." IEEE Trans. Acoustics, Speech & Signal Processing, ASSP-33, pp 1324-1326.

Goussard, Y. and Demoment, G. (1987). "Détection-estimation récursive rapide de séquences Bernoulli-gaussiennes." Traitement du Signal, à paraître.

Grenier, Y., (1980), "Time-dependent ARMA modelling of nonstationary signals", IEEE Trans. Acoustics, Speech & Signal Processing, ASSP-31, pp 899-911.

Grenier, Y., (1986), "Modèles ARMA à coefficients dépendant du temps: estimateurs et applications", Traitement du Signal, vol 3, pp 219-233.

Kormylo, J., and Mendel, J.M. (1982). "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes." IEEE Trans. Inf. Theory, IT-28, pp 482-488.

Mendel, J.M. (1983). "Optimal seismic deconvolution." Academic Press, New-York.

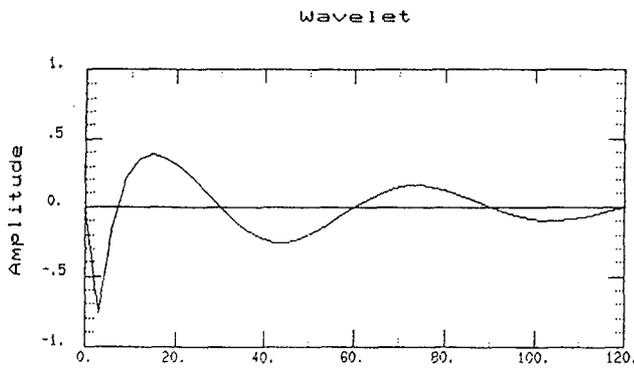


Figure 1. Ondelette utilisée pour le test, approche MA.

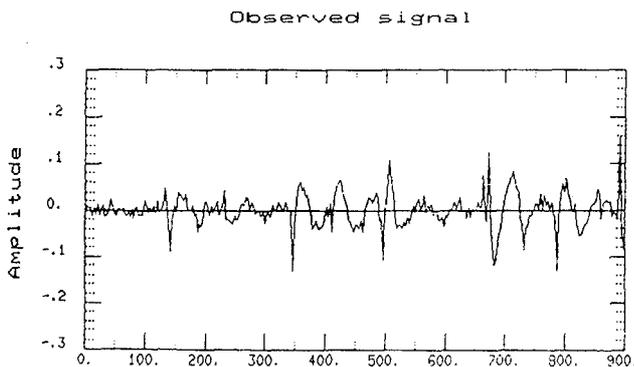


Figure 2. Signal synthétisé.

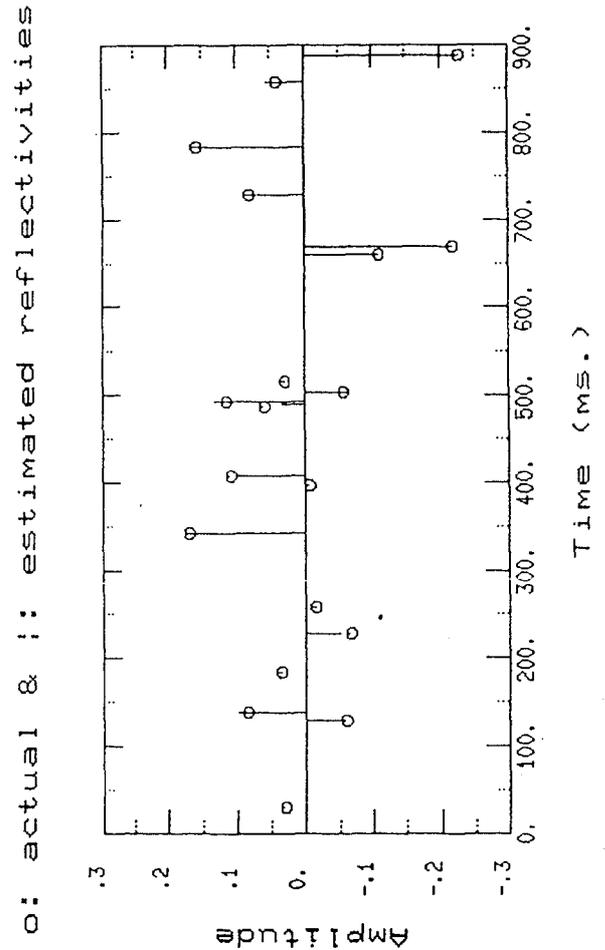


Figure 3. Résultat de déconvolution par approche MA, les cercles donnent les valeurs exactes de $w(k)$, les traits en donnent les estimations.

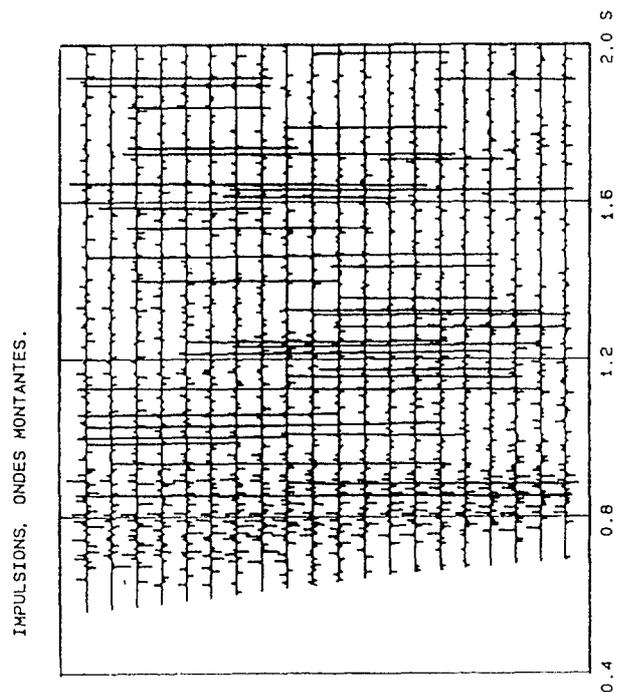


Figure 4. Résultat de déconvolution par approche AR, pour un enregistrement de PSV (profil sismique vertical), sur 20 traces.