

COMPARAISON D'ESTIMATEURS SPECTRAUX APPLIQUÉS À LA SPECTROSCOPIE RMN ¹³C

Norbert HENZEL, Marc TOMCZAK,
David BRIE, Daniel NICOLE*

Centre de Recherche en Automatique de Nancy (CNRS URA 821) - *Laboratoire d'Étude des Systèmes Organiques et Colloïdaux (CNRS URA 406) - Université de Nancy I, B.P. 239, 54506 Vandoeuvre les Nancy.

RÉSUMÉ

Dans le cadre de la spectroscopie RMN, les performances d'estimateurs spectraux récents restent généralement inférieures à celles de la transformée de Fourier (TF), et leur utilisation soulève des problèmes importants (choix de l'ordre, etc.). A travers des comparaisons effectuées sur signaux réels, nous évoquons ces problèmes et proposons une solution fondée sur un algorithme de matrice instrumentale surdéterminée, associé à la méthode de Prony, et pour lequel une méthode d'estimation de l'ordre de la partie AR d'un modèle ARMA est présentée. Enfin, nous montrons, qu'après décomposition en sous-bandes, notre approche permet d'obtenir des résultats comparables à ceux de la TF en ce qui concerne l'estimation des amplitudes, mais supérieurs sur le plan résolution et précision fréquentielle.

ABSTRACT

In the case of NMR spectroscopy, the Fourier transform generally performs better than modern spectral estimators. Furthermore, the latter give rise to important problems (order estimation, etc.). These problems are discussed through some comparisons on real signals. Then, a scheme using Prony's method in association with an overdetermined instrumental variable algorithm is proposed. For this algorithm, an order estimation procedure for the AR part of an ARMA model is set out. Finally, it is shown that a subband decomposition approach allows to achieve results comparable to, and even better in the case of frequency resolution accuracy, than those of Fourier based estimators.

1. INTRODUCTION

L'analyse spectrale par transformée de Fourier (TF) est utilisée classiquement en spectroscopie RMN. L'application, dans ce cas de figure, d'estimateurs spectraux plus modernes, a fait l'objet, depuis quelques années, d'une littérature conséquente [1]-[6]. Il s'agit essentiellement des estimateurs autorégressifs (AR), de la méthode du maximum d'entropie, et de la méthode de Prony et Prony/décomposition en valeurs singulières (SVD-Prony). Les résultats présentés sont souvent satisfaisants, mais ils ne concernent en général que des simulations ou des cas expérimentaux peu complexes. Ainsi, J. Cadzow [7] a souligné les difficultés rencontrées avec la méthode de Prony lorsque certaines raies sont particulièrement proches. De plus, tous les auteurs s'accordent à reconnaître que le choix de l'ordre (et/ou de la dimension du sous-espace signal) s'avère un problème délicat en pratique. Enfin, nous n'avons pas connaissance de résultats probants sur le cas particulier de l'analyse dite quantitative (cf. paragraphe 2). Dans cet article, nous comparons les performances de quelques estimateurs modernes et de la TF, sur deux exemples choisis pour leur relative complexité. Après un bref exposé du problème, nous présentons rapidement les estimateurs sélectionnés, puis nous illustrons et discutons les nombreuses difficultés rencontrées. Nous proposons alors une méthode permettant de s'affranchir de ces problèmes dans les cas simples. Celle-ci repose sur un algorithme de matrice instrumentale surdéterminée, récursif en temps et sur l'ordre (LORIV), associé à la méthode de Prony et à une procédure d'estimation de l'ordre de la partie AR d'un modèle ARMA. Enfin, nous montrons que dans les cas plus

complexes, une décomposition préalable en sous-bandes permet à notre méthode d'égaliser les performances de la TF, voire de les surpasser en ce qui concerne la résolution et la précision fréquentielle.

2. POSITION DU PROBLÈME

La spectroscopie RMN ¹³C constitue un excellent outil d'analyse chimique de mélanges complexes. L'identification des substances présentes dans le mélange (analyse qualitative), et la détermination du pourcentage de chacune (analyse quantitative), s'effectuent par analyse spectrale du signal numérique délivré par le spectromètre. Celui-ci, constitué d'une grande quantité d'échantillons (N≈128k), correspond à un certain nombre d'accumulations de la même mesure, ceci afin d'améliorer le rapport signal sur bruit (RSB). Il peut être modélisé comme une superposition de sinusoides amorties exponentiellement, plus un terme représentatif du bruit de mesure :

$$x_n = \sum_{i=1}^p A_i \exp((\alpha_i + 2\pi j f_i) n T + j \theta_i) + e_n \quad (1)$$

L'analyse qualitative repose essentiellement sur une estimation précise de la fréquence f_i de chaque sinusoides, quant à l'analyse quantitative, elle dépend d'une bonne estimation des amplitudes A_i , c'est-à-dire de la surface sous chaque raie du spectre TF. Il faut noter que dans la pratique, le nombre p de raies peut atteindre plusieurs centaines. D'autre part, chaque accumulation demande 0,5 à 5 mn d'expérimentation [8]. Les limitations de l'approche TF se situent au niveau du pouvoir de résolution; en effet, toute non-détection fausse totalement



l'interprétation. Par ailleurs, il arrive souvent que des raies soient trop proches pour un calcul direct de leur surface, il faut alors procéder à leur désenchevêtrement, ce qui requiert l'utilisation d'algorithmes d'optimisation non-linéaire [9]. Dans ce contexte, il était naturel d'envisager d'utiliser d'autres estimateurs spectraux, connus pour leur grand pouvoir de résolution, bien que la présence de bruit et le nombre élevé d'observations, ne soient pas des éléments de nature à les favoriser par rapport à la TF.

3. MÉTHODES

Parmi les estimateurs susceptibles de fournir des résultats intéressants, nous avons retenu la méthode de Prony, la méthode de Lagunas, et la méthode MUSIC [10]. La méthode de Prony comporte des avantages certains, puisqu'elle permet l'estimation directe des paramètres utiles (A_i , f_i , α_i , θ_i). Les fréquences et les facteurs d'amortissement α_i sont calculés comme les racines du polynôme $\sum_{i=0}^p a_i z_i$, où $z_i = \exp(\alpha_i + 2\pi j f_i)$ et $a_0 = 1$. Les a_i sont les coefficients de prédiction linéaire du signal, usuellement estimés par la méthode de covariance. Pour des RSB faibles, cette approche se traduit par un biais important en fréquence et sur les coefficients d'amortissement. Une solution consiste à utiliser des ordres p très supérieurs au nombre réel de raies. Dans ce cas, la connaissance des coefficients de prédiction arrière permet de mieux séparer les racines réelles de celles dues au bruit (critère du cercle) [11]. Enfin, l'application de la technique SVD améliore encore les performances de la méthode. En effet, l'influence du bruit peut alors être atténuée en ne considérant qu'une approximation de rang réduit de la matrice des données, correspondant aux valeurs singulières les plus élevées [12]. Dans les cas favorables, leur nombre fournit une bonne estimation de l'ordre.

Bien que non adaptée au problème de l'analyse quantitative, la méthode MUSIC a été retenue pour son grand pouvoir de résolution dans le cas de signaux à bande étroite dans du bruit blanc. Elle repose sur l'hypothèse (vérifiée en l'absence de bruit) que les p - m vecteurs propres $\mathbf{v}_{m+1}, \dots, \mathbf{v}_p$ du sous-espace bruit de la matrice des données, qui possède au total p vecteurs propres et m vecteurs propres principaux (ceux associés aux plus grandes valeurs propres), sont orthogonaux aux vecteurs $\mathbf{e}(f_i)$ où $\mathbf{e}(f) = [1 \exp(2\pi j f T) \dots \exp(2\pi j f m T)]^T$ et $i = 1, \dots, m$. L'estimateur MUSIC est alors donné par :

$$P_M(f) = (\mathbf{e}^H(f) (\sum_{i=m+1}^p \mathbf{v}_i \mathbf{v}_i^H) \mathbf{e}(f))^{-1} \quad (2)$$

Le choix des paramètres m et p s'effectue dans les mêmes conditions que pour SVD-Prony.

Bien que d'un pouvoir de résolution légèrement inférieur aux méthodes AR, la méthode de Lagunas est connue pour sa meilleure résolution en amplitude. Elle constitue une extension de la méthode de Capon, qui calcule la densité spectrale à une fréquence f_i comme l'énergie minimale de sortie d'un filtre RIF de coefficients $[a_0^i, \dots, a_p^i] = \mathbf{a}_i^T$, sous la contrainte $\mathbf{e}^H(f_i) \mathbf{a}_i = 1$. Lagunas impose une contrainte supplémentaire de douceur autour de f_i , d'où l'expression de l'estimateur:

$$P_L(f) = (\mathbf{e}^H(f) \mathbf{R}_p^{-1} \mathbf{e}(f)) ((\mathbf{e}^H(f) \mathbf{R}_p^{-2} \mathbf{e}(f)))^{-1} \quad (3)$$

4. RÉSULTATS EXPÉRIMENTAUX

Le spectre de la figure 1 a été obtenu par TF d'un signal de $N=128k$ échantillons, accumulé 1000 fois, correspondant à de l'éthylbenzène additionné de solvant. La raie centrale est en fait composée de 3 raies très proches d'amplitude presque égale. Ceci représente la principale difficulté de cet exemple, le nombre de raies étant particulièrement faible (9 raies, soit 18 sur $f/f_c \in [-0,5, 0,5]$). Pour des raisons évidentes d'encombrement mémoire et de temps de calcul, il n'est pas possible de traiter l'ensemble du signal avec les méthodes haute résolution. De plus, l'expérience montre que les performances ne s'améliorent pas lorsque N augmente au delà d'un certain seuil, ce qui s'explique en partie par le fait que le RSB décroît avec le temps. Enfin, l'augmentation de N oblige à utiliser des ordres de plus en plus élevés pour conserver une résolution suffisante, cette surparamétrisation s'accompagne d'un surcroît de complexité numérique néfaste à la qualité de l'estimation. Nous avons effectué un grand nombre d'essais avec différentes valeurs des paramètres N , p , et m , en nous limitant à des temps de calcul raisonnables ($N_{\max}=4096$, $p_{\max}=200$, $m_{\max}=25$). Pour Lagunas ou AR-Prony, l'utilisation de critères d'ordre du type AIC est possible, mais, en raison du bruit, de la nature du signal et de la proximité de certaines raies, ils s'avèrent peu fiables. Dans le cas de MUSIC et SVD-Prony, une fois p fixé à une valeur suffisante, m peut être déterminé par l'examen des valeurs singulières (fig. 2a). Le tableau 1 présente les résultats offrant le meilleur compromis f_i/A_i , les cases grisées indiquent les raies non-détectées. Aucune des 3 méthodes ne détecte toutes les raies, la moins bonne résolution correspond bien sûr à la méthode de Lagunas (3 raies manquantes). SVD-Prony offre la meilleure précision en fréquence et en amplitude. Comme attendu, les amplitudes données par MUSIC ne correspondent en rien à la réalité, mais, de façon plus surprenante, ce phénomène affecte également la méthode de Lagunas. De façon générale, SVD-Prony fournit les résultats les plus constants dans une large gamme d'ordres, et est la seule méthode digne d'intérêt pour l'analyse quantitative. Elle présente toutefois un coût de calcul important. Dans le cas particulier où p est choisi entre $6N/8$ et $7N/8$ ($N \approx 4k-8k$), l'effort de calcul peut être réduit [5], mais au delà de $p \approx 1000$, les racines sont mal estimées. Une variante, approximative en présence de bruit, permet de s'affranchir du calcul des racines [5]. Avec cette méthode, et $N=4096$, $p=3584$, $m=18$, les 3 raies centrales sont séparées, mais au détriment du rapport des amplitudes qui est nettement plus mauvais qu'avec $p=175$.

Dans le deuxième exemple, on s'intéresse à une partie très enchevêtrée comprenant 5 raies (fig. 3), d'un spectre RMN de 27×2 raies au total. Les résultats (éloquentes) fournis par SVD-Prony sont indiqués dans le tableau 2. Avec $N=8k$ et $p=7680$, une deuxième composante est détectée, mais de nombreuses composantes parasites apparaissent.

Les deux exemples précédents sont représentatifs des nombreux essais que nous avons effectués tant sur des signaux RMN différents, qu'avec d'autres méthodes. Rappelons qu'un spectre RMN standard comprend 200 ou 300 raies et nécessite l'emploi

d'ordres élevés. Dans de tels cas, les problèmes de lourdeur numérique, de choix de l'ordre, de composantes parasites, etc., deviennent insurmontables et affectent l'estimation. A titre d'illustration, la figure 2b montre les valeurs singulières obtenues pour un signal à 200 raies. On constate l'inefficacité du critère, due en partie au bruit et à la présence de composantes fortement amorties.

5. LA MÉTHODE LORIV-PRONY

Un des principaux obstacles rencontrés précédemment, reste le niveau élevé du bruit d'observation, qui met en défaut les méthodes précédentes. Un modèle ARMA(p,q) doit permettre une meilleure prise en compte de ce bruit, ce qui aura pour effet de réduire le nombre de paramètres nécessaires. Afin de réduire encore la charge de calcul, il est possible d'estimer uniquement la partie AR correspondant au signal non-bruité. La méthode de la matrice instrumentale à observations retardées assure une estimation non biaisée des paramètres AR. Pour améliorer ses performances dans le cas de signaux à bande étroite, nous avons choisi la forme surdéterminée (OIV), qui s'apparente à l'approche Yule-Walker modifiée surdéterminée [13]. L'algorithme ORIV présenté dans [13] est une forme récursive d'OIV, mais il souffre de sévères problèmes numériques. Bien que ce dernier problème ait été résolu [14], nous avons jugé intéressant d'en développer une version en treillis, très stable numériquement, récursive en temps et sur l'ordre et de complexité numérique moindre (LORIV) [15]. Par manque de place, les détails de cet algorithme sont omis. La propriété de récursivité sur l'ordre permet de simplifier grandement l'estimation de l'ordre. Ce point a été résolu en utilisant une approche fondée sur les statistiques des tests locaux (LTS), proposée en [16], qui permet de construire une méthode d'estimation de l'ordre dès lors que l'algorithme d'identification est choisi. Son intérêt majeur est qu'elle évite l'identification au sens du maximum de vraisemblance du modèle ARMA, nécessaire pour le calcul de critères du type AIC ou SDD [10]. Avec la méthode LTS, l'ordre p est estimé par : $\hat{p} = \text{argmin}(S(p) + pf(N))$, où $p \leq p_{\max}$ préfixé, et où $S(p)$ est la statistique du test local associé à l'algorithme ORIV, et $pf(N)$ une fonction de pénalisation destinée à éviter une surparamétrisation. Nous avons utilisé 2 fonctions : $pf(N)=2p$ (type AIC) et $pf(N)=2\log(N)$ (type SDD).

Dans le cas de l'exemple 1 et pour $N=16k$, la méthode LORIV utilisée avec un ARMA(100,200) et un nombre d'équations prises en compte $C=2400$, détecte parfaitement les 9 raies et ceci sans introduire de raies parasites. Cette approche se caractérise par une très bonne robustesse aux changements sur p et q , qui peuvent aisément être fixés par tâtonnement. La valeur du facteur de surdétermination dépend de N .

La partie AR estimée par LORIV peut être injectée dans la méthode de Prony (LORIV-Prony). Ici aussi, le critère du cercle s'applique. Les résultats observés sur les 2 exemples précédents sont indiqués dans les tableaux correspondants. Dans le premier cas, les 9 raies sont détectées très précisément, et le rapport d'amplitude est à peu près conservé comparativement à la TF. Dans le deuxième cas, seule une raie supplémentaire par rapport à SVD-Prony est obtenue. Toutefois, sur l'ensemble des

27 raies, LORIV-Prony s'avère mieux respecter le rapport d'amplitude. On note également que d'une manière générale, notre approche implique des ordres p sensiblement inférieurs à ceux imposés par les autres méthodes.

6. DÉCOMPOSITION EN SOUS-BANDES

En définitive, dans les cas un tant soit peu complexes, aucune des méthodes évoquées ne peut rivaliser avec la TF. Le grand nombre d'échantillons disponibles suggère une décomposition du problème en sous-bandes. Une telle approche comporte un double avantage. D'une part, chaque sous-problème étant plus simple (nombre de raies réduit), les estimations, tant de l'ordre que des paramètres, sont effectuées dans de meilleures conditions, donc de meilleure qualité. D'autre part, la complexité numérique globale peut être réduite dans la mesure où l'on traite moins d'échantillons, avec des ordres donc nettement plus faibles, sur chaque sous-bande [17]. On considère à nouveau l'exemple 2, et plus particulièrement la bande enchevêtrée. Après translation fréquentielle, filtrage, et décimation adéquats, les 1311 échantillons restants sont traités par LORIV-Prony. L'ordre p est déterminé par LTS, la figure 6 montre les courbes obtenues avec les fonctions de pénalisation de type AIC (trait plein) et SDD (pointillé). Dans ce dernier cas, l'ordre indiqué est $p=29$. Les résultats correspondants sont spécifiés dans le tableau 2, le spectre est représenté sur la figure 5. La précision et la résolution fréquentielles sont identiques à celles de la TF, les amplitudes estimées sont quelque peu différentes de celles obtenues par TF et désenchevêtrement mais restent tout de même cohérentes. La figure 4 montre que l'absence de surdétermination affecte fortement les performances, car plusieurs raies parasites font leur apparition.

7. CONCLUSIONS

Nous avons montré que les estimateurs retenus dans notre étude ne peuvent égaler les performances de la TF en spectroscopie RMN, aussi bien du point de vue temps de calcul que sur le plan résolution fréquentielle et en amplitude. De plus, ils nécessitent le choix difficile de certains paramètres. Toutefois, la méthode que nous proposons (LORIV-Prony), associée à la procédure d'estimation de l'ordre LTS, offre des performances comparables à celles de la TF dans les cas peu complexes. Enfin, nous avons évoqué les avantages d'une décomposition préalable en sous-bandes et montré que dans ce contexte, notre méthode parvient à rivaliser avec la TF dans un cas de figure plus proche des cas standards. Toutefois, la faisabilité d'une analyse quantitative, à l'aide de cette méthode, doit être étudiée de façon plus approfondie.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Tang J., Norris J.R., *J. Magn. Reson.*, **78**, 23-30, 1988.
- [2] Laue E.D. et al., *J. Magn. Reson.*, **62**, 437-452, 1985.
- [3] Barkhuijsen H. et al., *J. Magn. Reson.*, **61**, 465-481, 1985.
- [4] Barkhuijsen H. et al., *J. Magn. Reson.*, **73**, 553, 1987.
- [5] Stephenson D.S., *Progr. in NMR Spectr.*, **20**, 515-626, 1988.
- [6] Van Huffel S. et al., *Proc. of the 6th EUSIPCO*, 1721-1724, Brussels, aug. 24-27, 1992.
- [7] Cadzow J.A., Wu M.-M., *IEE Proc.*, **134**, Pt F, 1, 69-78, 1987.
- [8] Ernst R.R. et al.: *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford University Press, 1987.



[9] Matlenjiewicz M., Henzel N. et al., *Analyst*, 117, 387-393, march 1992.
 [10] Kay S.M.: *Modern spectral estimation. Theory and application.* Prentice-Hall, 1988.
 [11] Kumaresan R., Tufts D.W., *IEEE Trans. ASSP*, 30, 6, 833-840, 1982.
 [12] Tufts D.W., Kumaresan R., *IEEE Proc.*, 70, 9, 975-989, 1982.
 [13] Friedlander B., *IEEE Trans. Aut. Cont.*, 29, 4, 353-356, 1984.
 [14] Porat B., Friedlander B., *IEEE Trans. Aut. Cont.*, 34, 6, 656-658, 1989.

[15] Tomczak M., Brie D. et al.: *LORIV, un algorithme de matrice instrumentale récursif en temps et sur l'ordre pour l'analyse spectrale de signaux à bande étroite bruités.* Rapport interne CRAN en préparation.
 [16] Brie D., Tomczak M. et al., *Proc. of IMACS/IFAC MIM'S2*, 402-406, Brussels, april 12-16, 1993.
 [17] Caspary O., Tomczak M. et al., *Proc. of the IEEE Instr. Measur. Techn. Conf.*, 128-131, Irvine (USA), may 18-20, 1993.

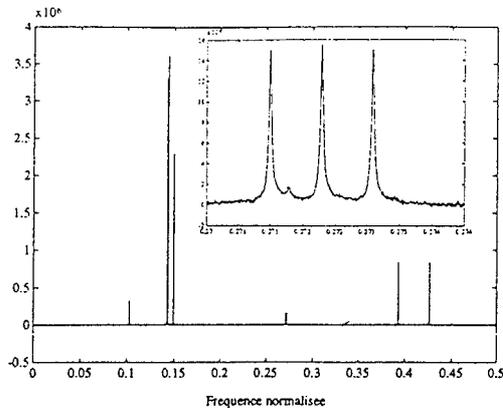


Fig. 1: Exemple 1, Transformée de Fourier

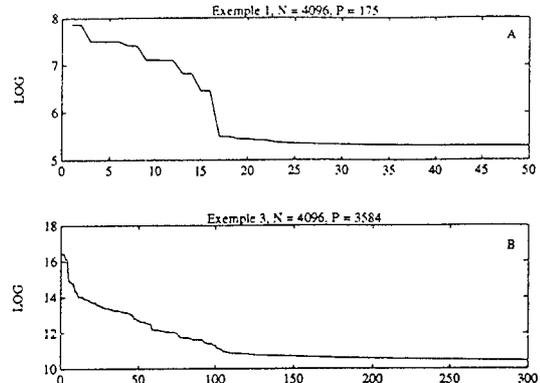


Fig. 2: Valeurs Singulières

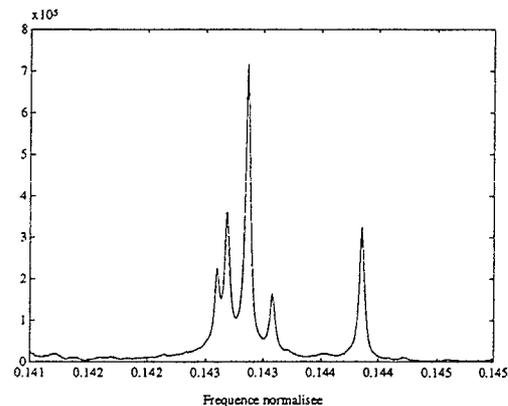


Fig. 3: Exemple 2, Transformée de Fourier, partie achevée

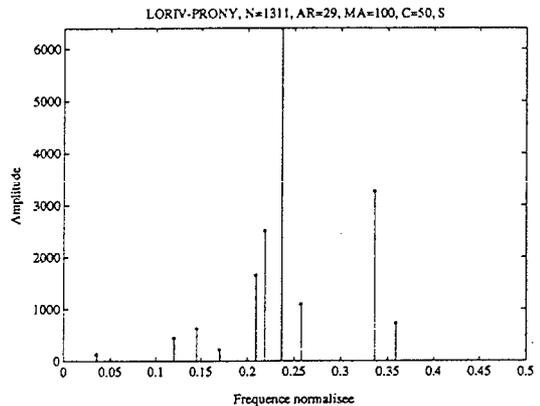


Fig. 4: Résultat LORIV-PRONY en sous-bandes et sans surdétermination

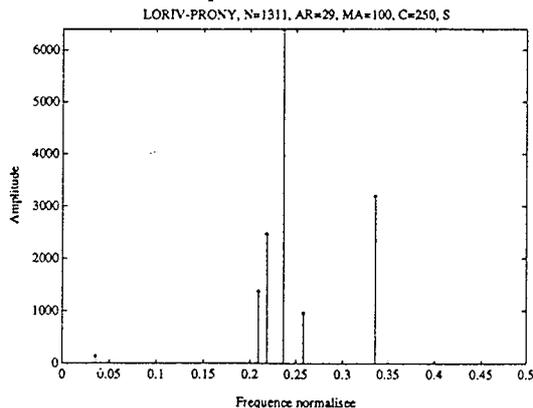


Fig. 5: Résultat LORIV-PRONY en sous-bandes

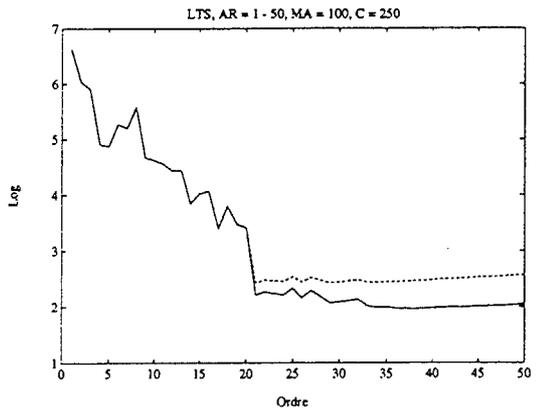


Fig. 6: Choix de l'ordre par LTS

	FFT (Spectre entier) N=65536		SVD-PRONY (Spectre entier) N=4096, P=175 M=18		MUSIC (Spectre entier) N=4096, P=175 M=18		LAGUNAS (Spectre entier) N=4096, P=22		LORIV-PRONY (Spectre entier) N=8192, AR=150 MA=500, C=2500	
	f	A	f	A	f	A	f	A	f	A
1	0.10269	0.12528	0.10269	0.11643	0.10273	0.37661	0.10237	0.07154	0.10269	0.12169
2	0.14276	1.0	0.14277	0.97465	0.14290	0.76369	0.14351	0.47504	0.14391	0.96824
3	0.14391	0.8678	0.14391	1.0	0.144	0.974	0.14351	0.47504	0.14391	0.96824
4	0.14960	0.51936	0.14960	0.52120	0.14974	0.41986	0.14839	0.0890	0.14960	0.54995
5	0.27096	0.18696	0.27117	0.08132	0.27121	0.60355	0.27182	0.43754	0.27182	0.17344
6	0.27179	0.18361	0.27117	0.08132	0.27121	0.60355	0.27182	0.43754	0.27182	0.17344
7	0.27259	0.17531	0.27243	0.08762	0.27256	0.64082	0.27260	0.16444	0.27260	0.16444
8	0.39284	0.53060	0.39283	0.43732	0.39293	0.59339	0.39293	0.92107	0.39265	0.46400
9	0.42628	0.71474	0.42628	0.54170	0.42638	1.0	0.42626	1.0	0.42627	0.67975

Tableau 1

	FFT (Spectre entier) N = 65536		SVD-PRONY (Spectre entier) N=4096, P=3584 M = 54		LORIV-PRONY (Spectre entier) N=16386, AR=800 MA=2000, C=1300		LORIV-PRONY (Sous-bande) N=1311, AR=29, MA=50, C=250	
	f	A	f	A	f	A	f	A
1	0.14260	0.28207	0.14260	0.28207	0.14260	0.28207	0.14260	0.21713
2	0.14269	0.47954	0.14269	0.47954	0.14268	1.0	0.14269	0.38702
3	0.14286	1.0	0.14286	1.0	0.14287	1.0	0.14287	1.0
4	0.14308	0.20931	0.1431	1.0	0.14308	0.15096	0.14308	0.15096
5	0.14385	0.44238	0.14386	0.19195	0.14385	0.50015	0.14385	0.50015

Tableau 2