



Développement d'une nouvelle technique de pointé automatique pour les données de sismique réfraction

Amin Khalaf

► **To cite this version:**

Amin Khalaf. Développement d'une nouvelle technique de pointé automatique pour les données de sismique réfraction. Géophysique [physics.geo-ph]. Université Pierre et Marie Curie - Paris VI, 2016. Français. <NNT : 2016PA066028>. <tel-01359287>

HAL Id: tel-01359287

<https://tel.archives-ouvertes.fr/tel-01359287>

Submitted on 2 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PIERRE ET MARIE CURIE

Spécialité : **Géophysique Appliquée**

Ecole Doctorale de Géosciences, Ressources Naturelles et Environnement

présentée par

Amin KHALAF

Pour obtenir le grade de :

Docteur de l'Université Pierre et Marie Curie

**Développement d'une nouvelle technique de pointé
automatique pour les données de sismique réfraction**

Soutenue le 15 février 2016 devant le jury composé de :

M ^{me} . Frédérique MOREAU	MC Université de Rennes 1	Rapporteur
M ^{me} . Ginette SARACCO	DR à CNRS - CEREGE	Rapporteur
M. Denis JONGMANS	Professeur à l'ISTerre	Examineur
M. Loïc LABROUSSE	Professeur à l'UPMC	Président
M. Dominique ROUSSET	MC Université de Pau	Examineur
M. Nicolas FLORSCH	Professeur à l'UPMC	Directeur de thèse
M. Christian CAMERLYNCK	MC à l'UPMC	Co-directeur de thèse
M. Christian HERISSON	DR à Cathie Associates	Invité
M. Pierre-Yves GALIBERT	MC à l'UPMC	Invité



To my small family, Ana and petite S.

Remerciements

Je tiens, en premier lieu, à remercier l'École Doctorale Géosciences et Ressources Naturelles (ED 398) qui m'a permis de finir ma thèse malgré le retard hors volonté. Merci en particulier à M. François BAUDIN pour son humanité et compréhension. Merci également à Mme Dovy TRISTANI de m'avoir toujours poussé à finir ma thèse et pour sa gentillesse.

Je désire exprimer ma gratitude aux personnes qui d'une façon ou de l'autre m'ont aidé et encouragé durant mes travaux de recherche. Particulièrement, je remercie mes directeurs de thèse, Nicolas FLORSCH et Christian CAMERLYNCK, pour l'aide et le soutien qu'ils ont manifesté à mon égard durant la rédaction de ce manuscrit. Merci à vous deux pour vos conseils, votre patience et pour la liberté de travail que vous m'avez accordée. Une partie substantielle des travaux effectués dans le cadre de cette thèse a pu être réalisée grâce à un financement du Ministère de l'Enseignement Supérieur syrien et de l'Université de Tichreen que je remercie également.

J'exprime ma gratitude envers les membres du jury qui ont fait de cette soutenance une discussion très intéressante. J'exprime toute ma reconnaissance à Frédérique MOREAU ainsi qu'à Ginette SARACCO d'avoir accepté de consacrer de leur temps pour être les rapporteurs de cette thèse. La version finale de ce mémoire a bénéficié de leur lecture très attentive et de leurs remarques précieuses. Je tiens à remercier également Denis JONGMANS et Dominique ROUSSET d'avoir accepté les places d'examineurs de ma thèse. En effet, leurs critiques et leurs suggestions ont largement contribué à améliorer la qualité de ce mémoire, et je leur en suis très reconnaissant. Christian HERISSON et Pierre-Yves GALIBERT, je vous remercie pour l'intérêt que vous avez manifesté en participant en qualité de membres invités à ce jury. Merci enfin à Loïc LABROUSSE de m'avoir fait l'honneur de présider ce jury.

Merci à l'ensemble des metisiens (ou ex-sisyphiens) que j'ai côtoyés, de près, de loin, voire pas du tout, qui font du labo un contexte de travail si agréable. Je veux aussi remercier mes non-encadrants, mais qui ont contribué à ma thèse : en spécial Antoine CHEVALIER, pour son aide avec mon code et la continuation de thèse ; merci Antoine pour les *golden idées* pour la rédaction; Ludovic BODET, pour les discussions devant le café et son aide à AGU ; Philippe MARTY, pour son aide avec le réseau, ordinateur, sports, mais aussi pour son amitié (et barbecue de week-ends). Merci à tous qui ont fait partie de cette thèse, d'une manière ou de l'autre, administrativement, personnellement, ou même sans savoir, m'ont aidé à continuer : Nadine BOST, Valérie GIRARD, Jean-Marie MOUCHEL, Feras ABDULSAMAD, Dmitri SORO, Sylvain PASQUET, Marine DANGEARD, Nora ROGER, les voisins Christophe BONNET et Bénédicte SABATIER, Fatiha BENTAHAR, Nejla GROUZ, Raphaël FAUCHES, les stagiaires Koné, Ibrahim, Adbullah, et tous les autres que malheureusement j'ai oublié d'écrire, mais sont gardés dans mon mémoire.

Je voudrais exprimer ma profonde gratitude à mes parents, qui m'ont toujours soutenu et sans eux je ne serai pas là. Je veux remercier aussi mes frères, mes sœurs et mes amis, qui même loin sont toujours là pour partager les joies et difficultés.

Enfin, je veux remercier Ana, pour être présente non seulement comme femme, mais comme amie, pour la compréhension quand j'ai dû travailler tous les week-ends et soirs (vu matins), et son soutien. Je veux aussi la remercier pour me donner le plus important cadeau de ma vie ! Ma petite '*Sousou*' Qui n'est pas encore arrivée, mais c'est la raison de ma vie, qui m'a fait continuer quand j'avais envie d'arrêter, à qui je dédie cette thèse.

Amin KHALAF

Paris, février 2016

Résumé

Un pointé précis des temps de première arrivée sismique joue un rôle important dans de nombreuses études d'imagerie de vitesse sismique, ou d'évolution du milieu avec notamment la surveillance en écoute passive de réservoirs ou d'aquifères. En général, les premières arrivées se manifestent par une variation notable de l'amplitude, du contenu fréquentiel, ainsi que des propriétés statistiques de la trace sismique dans les domaines temporel et/ou fréquentiel. Depuis ces dernières décennies, de nombreuses techniques de pointé automatique ou semi-automatique ont été développées. Cependant, la plupart d'entre elles ont été adaptées pour des objectifs sismologiques, telles celles basées sur des rapports LTA/STA. En outre, elles ne sont pas suffisamment performantes sur des données issues de la prospection de proche surface en raison de la complexité des structures du sous-sol, des variations rapides du champ de vitesse et d'un rapport signal/bruit parfois insuffisant.

Une nouvelle stratégie a été développée à partir d'un algorithme adaptif combinant plusieurs approches associant l'utilisation de fenêtres multiples imbriquées, l'estimation des propriétés statistiques d'ordre supérieur et le critère d'information d'Akaike. L'algorithme développé exploite les avantages d'intégrer plusieurs propriétés (l'énergie, la « gaussianité », et la stationnarité) dévoilant la présence des premières arrivées. Cette stratégie mime le pointé manuel, lorsque une tendance globale de la localisation des premières arrivées est définie, pour qu'ensuite les instants précis des premières arrivées soient localisés au voisinage de cette tendance. Tandis que les incertitudes estimées sont, dans certains cas, du même ordre que pour le pointé automatique (notamment en tomographie sismique, ou même pour certains traitements *a posteriori*), l'algorithme fournit aussi automatiquement une estimation sur l'incertitude des pointés. La précision et la fiabilité de cet algorithme ont été évaluées par la comparaison des résultats avec ceux issus d'un pointé manuel ainsi qu'avec d'autres pointeurs automatiques, dans certaines conditions de bruit. Tous les paramètres nécessaires à cet algorithme sont auto-adaptatifs grâce à l'intégration, en série, des résultats de chaque sous-algorithme. Il est simple à mettre en œuvre et ne nécessite pas de grandes performances informatiques.

La présence du bruit sismique dans l'enregistrement détériore souvent la performance d'algorithmes de pointé automatique, particulièrement dans le cas de petits événements. Le signal et le bruit sismique partagent souvent les mêmes bandes de fréquences, ce qui rend le filtrage dans le domaine fréquentiel peu efficace. La sommation peut être une alternative exploitant la redondance des données sismiques de proche surface. Nous proposons une nouvelle procédure fondée sur la double sommation dans le domaine temporel, ce afin d'améliorer la détectabilité des premières arrivées. Cette procédure est fondée sur le principe clé de la ressemblance locale entre les traces « stackées ». Un dé-bruitage par transformée en ondelettes peut être effectué, comme

processus supplémentaire, avant la deuxième sommation, pour les données très bruitées. Une étude comparative de la performance du pointé automatique avec d'autres méthodes a été réalisée sur des jeux de données réelles bruitées. Les résultats ont montré l'intérêt qu'il y a à appliquer la double sommation avant le pointé automatique.

Mots-clés : pointé automatique, algorithme adaptatif, kurtosis, Akaike Information Criterion, tomographie des premières arrivées sismiques, double sommation.

Abstract

Accurate picking of first arrival times plays an important role in many seismic studies, particularly in seismic tomography and reservoirs or aquifers monitoring. The first arrivals do manifest themselves as a noticeable variation in amplitude, frequency content, statistical properties, and/or polarization properties in time and/or frequency domain. Many techniques have been developed for picking first arrivals automatically or semi-automatically, but most of them were developed for seismological purposes which does not reach the accuracy objectives required in the applied seismic applications due to the complexity of near surface structures, and to usual low signal-to-noise ratio.

A new adaptive algorithm has been developed based on combining three picking methods (Multi-Nested Windows, Higher Order Statistics and Akaike Information Criterion). It exploits the benefits of integrating three properties (energy, gaussianity, and stationarity), which reveal the presence of first arrivals. This strategy mimics the human first-break picking, where at the beginning the global trend is defined. Then the exact first-breaks are searched in the vicinity of the now defined trend. Since time uncertainties estimating is of crucial importance for seismic tomography, the developed algorithm provides automatically also the associated errors of picked arrival times. The comparison of resulting P-wave arrival times with those picked manually, and other algorithms of automatic picking, demonstrated the reliable performance of the new algorithm under certain noise conditions. All parameters of our multi-method algorithm are auto-adaptive thanks to the integration in series of each sub-algorithm results in the flow. Hence, it is nearly a parameter-free algorithm, which is straightforward to implement and demands low computational resources.

The high noise level in the seismic records can significantly decline the efficiency of the automatic picking algorithms, particularly in the case of weak arrivals. The problems related to the traditional methods (spectral filtering) arise when the signal and noise share the same band frequency. In the near surface surveys, the stacking addresses these problems thanks to data redundancy. To improve the signal-to-noise ratio of first arrivals, and thereby increase their detectability, an alternative has been proposed based on double stacking in the time domain. This approach stacking is based on the key principle of the local similarity of stacked traces. In cases of very noisy data, denoising using wavelet transform was carried out as an auxiliary procedure before the second stack. A comparative study of automatic picking performance was achieved, using adaptive multi-method algorithm on noisy real data. The results demonstrate the feasibility of applying the double stacking before the automatic picking.

Keywords: automatic picking, adaptive algorithm, multi-nested window, kurtosis, Akaike Information Criterion, first arrival seismic tomography, double stacking.

Sommaire

1. Introduction	1
1.1 Généralités.....	1
1.2 Importance des pointés d'arrivées sismiques et intérêt des pointés automatiques.....	2
1.3 Motivations et objectifs	4
1.4 Organisation de la thèse	7
2 Revue de la littérature	9
2.1 Définitions générales.....	9
2.1.1 Stratégies de pointé manuel	10
2.2 Techniques de pointé automatique.....	12
2.3 Méthodes reposant sur les calculs d'énergie dans le domaine temporel	14
2.4 Les méthodes d'analyse dans le domaine fréquentiel ou temps-fréquence (échelle).....	19
2.5 Les méthodes d'analyses basées sur des propriétés statistiques	19
2.6 Méthodes basées sur les approches autorégressives (AR).....	21
2.7 Les méthodes basées sur l'analyse du mouvement des particules (polarité de l'onde sismique).....	22
2.8 Méthodes alternatives.....	23
3 Développement de l'algorithme adaptatif	25
3.1 Introduction.....	28
3.2 Methodology	32
3.2.1 Principe and methods.....	32
3.2.2 Algorithms implementation.....	38
3.2.3 Entire-shot record algorithm picking.....	43
3.3 Testing and results	46
3.3.1 Test on synthetic data.....	46
3.4 Conclusions.....	51
3.5 Acknowledgements	51
4 Evaluation la performance de l'algorithme développé	53
4.1 Introduction.....	55
4.2 Deep seismic data (petroleum-scale exploration)	56
4.2.1 Modified Coppens' Method (MCM)	56
4.2.2 Fractal-Dimension Method (FDM)	57
4.2.3 Data sets.....	57
4.3 Shallow seismic data (groundwater-scale prospection).....	61
4.4 Near surface seismic data (environment-scale prospection)	63
4.4.1 Modified Energy Ratio (MER).....	63

4.4.2	Data sets.....	64
4.5	Automatic control editing (quality control)	68
4.6	Conclusions and suggestions.....	72
5	Renforcement local des arrivées sismiques par double sommation dans le domaine temporel	75
5.1	Introduction.....	78
5.2	Definitions and general concepts used in stacking	79
5.3	First stacking using local similarity.....	80
5.4	Reference trace	81
5.5	Second stacking using cross correlation.....	83
5.6	Denoising by wavelet transform (WT).....	85
5.7	Applications.....	88
5.8	Conclusions.....	90
6	Conclusions générales et perspectives	91
	Liste des notations.....	95
	Bibliographie	97
	Annexes	111
	Annexe A.– Résumé étendu présenté aux journées scientifiques d’AGAP 2015.....	111
	Annexe B. – Supplementary Material for Chapter 5.....	117
	Table des illustrations	125
	Table des tableaux	131

Introduction

1.1 Généralités

L'objectif des méthodes sismiques est d'acquérir, de traiter, puis d'inverser les champs d'ondes sismiques afin d'obtenir une image interprétable du sous-sol. Cette méthode peut être brièvement résumée comme suit: une perturbation mécanique est émise dans le sol par une source d'énergie contrôlée. Cette énergie se propage dans le sous-sol et va se réfléchir et se réfracter en présence des interfaces géologiques montrant un contraste d'impédance élastique. Les informations sont enregistrées à l'aide d'un ensemble de récepteurs, puis traitées afin d'améliorer leurs qualités, avant d'être utilisées pour produire une image représentative du sous-sol via des méthodes d'inversion. L'image finale est alors utilisée comme un point de départ pour l'interprétation géophysique et la compréhension des structures géologiques.

Les méthodes de prospection sismique sont fréquemment employées pour caractériser la géométrie de la sub-surface. On distingue notamment la sismique réflexion (Deidda et al., 2012; Francese et al., 2002; Yilmaz, 2001), la sismique réfraction (Lanz et al., 1998; Zelt et al., 2006), les méthodes utilisant les ondes de surface (Dal Moro et al., 2015; Socco et al., 2010), ou la combinaison de deux ou plusieurs méthodes (Kim and Kim, 2008; Pasquet et al., 2015b). Parmi ces méthodes, la sismique réfraction de surface est la plus facile à mettre en place, et elle offre également l'avantage d'être relativement peu coûteuse. Grâce aux importantes améliorations en modélisation directe et en algorithmes d'inversion, la tomographie sismique est devenue une méthode privilégiée pour explorer la Terre, à quelques échelles que ce soit. Elle fournit des images bidimensionnelles (2D) ou tridimensionnelles (3D) du champ de vitesses des ondes se propageant dans le sous-sol, selon la manière de déployer les couples sources-récepteurs (appelé par la suite le dispositif). Par exemple, une prospection avec un dispositif en ligne fournit des images 2D du sous-sol, tandis que l'utilisation de dispositifs croisés donne une image 3D des structures de sub-surface (Sheriff and Geldart, 1995, p. 244). De plus, des dispositifs dits hybrides, tel qu'avec un déport latéral en T ou L (*T- or L-shape*), sont également adaptés pour analyser les variations en azimuth de la vitesse sismique (Valois, 2011).

De manière générale, la résolution latérale et la profondeur d'investigation des méthodes de sismiques varient selon l'espacement entre les récepteurs et la dimension du dispositif d'acquisition utilisé (Knödel et al., 2007, p. 353; Milsom and Eriksen, 2011, p. 209). En revanche, leur résolution verticale et leur précision dépendent de la longueur d'onde dominante (Sheriff and Geldart, 1995, p. 174; Yilmaz, 2001, p. 468) et également, pour la sismique réfraction, de la fidélité du pointé des premières arrivées (Lehmann, 2007, p. 37; Röhm et al., 2000; Senkaya and Karsh, 2014). Notons que la résolution de la tomographie des premières arrivées sismiques est en relation inverse

également du nombre d'onde maximal se propageant dans le milieu (Galibert et al., 2014; Spetzler and Snieder, 2004).

1.2 Importance des pointés d'arrivées sismiques et intérêt des pointés automatiques.

L'identification des temps des arrivées sismiques est très importante dans divers domaines (par exemple en sismologie, sismique appliquée, micro-sismicité, ..., etc.). Du point de vue sismologique, ces temps permettent d'obtenir des informations importantes, telles que la localisation d'un séisme (Billings et al., 1994; Stewart, 1977), la structure de l'intérieur de la Terre (par ex. Diehl, 2008; Moriya, 2009), le contexte géotectonique (Ross and Ben-Zion, 2014), la sismicité d'une région (Langet et al., 2014) et l'évaluation du risque sismique. Les applications des premières arrivées, dans le domaine de la sismicité et micro-sismicité, concernent l'évaluation et la surveillance de réservoirs ou d'aquifères, et la tomographie de sismique passive (par ex. Tselentis et al., 2011; Zhang et al., 2009). La sismique réfraction repose entièrement sur l'analyse des temps des premières arrivées. Ainsi, la détermination des temps des premières arrivées est la première étape de flux du traitement des données. En sismique réflexion, les premières arrivées sont nécessaires comme outil auxiliaire, notamment pour la correction différentielle des couches altérées en surface (*static corrections*) des profils sismiques (par ex. Coppens, 1985; Lawton, 1989; Spagnolini, 1991).

La connaissance des vitesses de cisaillement (V_s), combinées aux vitesses de compression (V_p), permet d'évaluer certaines propriétés des roches (par exemple : porosité, saturation) (Sheriff and Geldart, 1995, p. 271). A partir des premières arrivées des ondes de cisaillement générées par des sources sollicitant horizontalement le sol, la vitesse V_s peut être aussi déterminée grâce aux méthodes de réfraction. Plusieurs études de sismique de proche surface sont fondées partiellement sur cette tomographie en ondes S (Bièvre et al., 2012; Valois, 2011). Pasquet et al. (2015a) ont également utilisé les pointés des ondes S afin de comparer et valider les modèles de V_s obtenus par l'analyse de la dispersion des ondes de surface.

Contrairement aux méthodes classiques qui utilisent seulement les temps des premières arrivées pour déduire les modèles de vitesse du sous-sol, la méthode d'inversion des formes d'ondes (FWI : *Full Wave Inversion*) produit une image haute résolution du terrain en exploitant l'ensemble du champ d'ondes à travers des approches d'optimisation (Romdhane, 2010). Dans les dernières années, une version réduite de FWI est apparue, qui inverse une partie de champs d'ondes (les arrivées précoces). Elle est appelée EWI pour « *Early arrival Wavefront Inversion* » (Hanafy et al., 2013; Sheng et al., 2006). L'intégration de premières arrivées précises dans cette approche pourrait aider à la fois à extraire la vraie signature de la source et à estimer le facteur de qualité intrinsèque Q , qui sont des paramètres essentiels pour ce type

d'inversion. De fait, la tomographie des temps des premières arrivées est souvent utilisée pour la reconstruction d'un modèle initial des FWI et EWI (Romdhane, 2010).

Traditionnellement, les premières arrivées ont été souvent pointées manuellement par des analystes humains, suffisamment familiarisés avec différentes sortes de formes d'ondes et de débuts de phases sismiques à identifier. Actuellement, les gros volumes de données numériques acquis rendent cette technique du pointé inefficace, relativement coûteuse en temps et en main-d'œuvre. La subjectivité de pointeurs humains et des autres facteurs concernant leur état physiologique et/ou psychologique (la fatigue, le stress de travail, etc.) (Saragiotis et al., 2002) rendent les résultats issus des opérateurs humains souvent hétérogènes et non-consistants (Cuenot, 2003). D'autres inconvénients du pointé manuel sont fonction de l'application attendue de ces premières arrivées (par exemple, toutes les opérations en temps réel ou quasi-réel ne sont pas applicables pour le pointé manuel, notamment en raison de la grande taille des jeux de données). Le seul avantage pertinent du pointé manuel est toutefois la possibilité de pointer, avec un certain degré de précision, les événements sismiques même si les phases sont masquées par un bruit, ou sont mélangées avec autres phases sismiques moins intéressantes. Cependant, l'accomplissement d'une telle tâche par l'inspection visuelle demande une grande expérience.

Parmi les avantages potentiels du pointé automatique, la rapidité des analyses avec des jeux de données considérables, est associée à des résultats homogènes et consistants (Cuenot, 2003). De plus, le pointé automatique constitue un outil précieux en réduisant la subjectivité de l'œil de l'analyste.

En sismologie, le pointé automatique est devenue indispensable du fait de l'augmentation régulière des réseaux sismologiques permanents ou temporaires (Küperkoch, 2010; Stewart, 1977). L'automatisation permet de calculer les coordonnées du séisme et d'analyser ses signaux en temps quasi-réel, et de provoquer l'alerte précoce en cas de tremblement de terre majeur (Lomax et al., 2012; Takanami and Kitagawa, 1988). En ingénierie des réservoirs (en particulier, pour les réservoirs non conventionnels), l'analyse en temps réel des signaux micro-sismiques (dus aux processus de fracturation hydraulique) requiert un pointé automatique précis des événements sismiques afin de surveiller le développement et la production du réservoir, et de comprendre les géomécanisme de fracturation (Oye and Roth, 2003; Zeng et al., 2014). En génie civil et pour les études environnementales, de nombreuses applications (construction de barrages, caractérisation de la géométrie des aquifères, surveillance temporelle des sites miniers) demandent une image plus détaillée des caractéristiques mécaniques du sous-sol et de la répartition spatiale d'éventuels contrastes géotechniques. Pour toutes ces applications, les méthodes d'imagerie sismique connaissent un essor considérable (Shtivelman, 1996). Des séries de profils sismiques, suffisamment denses, avec espacement de récepteurs plus serrés sont prérequis. Ces méthodes se trouvent immédiatement confrontées à un problème crucial : un volume

considérable de traces à pointer, et par conséquent, un coût élevé. C'est pourquoi, il est incontournable de développer des algorithmes du pointé automatique des premières arrivées.

Dans le cadre de cette thèse, tous les développements abordés concernent le pointé automatique des premières arrivées de sismique réfraction de proche surface.

1.3 Motivations et objectifs

Les temps des premières arrivées sont des données essentielles d'entrée de l'algorithme d'inversion tomographique dans lequel un modèle discret de la vitesse du milieu est mis à jour à travers un processus itératif. Le modèle final de vitesse est celui qui converge vers des écarts minimaux entre les temps des arrivées calculés et ceux observés (Lehmann, 2007, p. 27). Toutefois, la fiabilité du modèle résultant (inversé) dépend directement de la paramétrisation du modèle, de la vitesse de l'encaissant (Lanz et al., 1998; Zelt and Ellis, 1988), du solveur de problème direct (Rawlinson et al., 2014, p. 23), du modèle initial et de la fiabilité des données (Ramanantoandro and Bernitsas, 1987; Yordkayhun, 2011). Certes, les développements modernes des outils informatiques permettent de discrétiser le modèle en mailles fines et les solveurs modernes fondés sur le tracé de rais (Cerveny, 2005) ou la prise en compte du volume de Fresnel (*Wave Eikonal Traveltime* « WET » : Schuster and Quintus-Bosz, 1993) permettent une modélisation directe (et par conséquent une inversion) avec une précision considérable. Toutefois, la fiabilité et la consistance des premières arrivées restent une des déterminations majeures de la qualité de l'inversion tomographique. De plus, aucun type de traitement ne peut corriger les erreurs pouvant affecter les données d'entrée. Par conséquent, des pointés précis et cohérents constituent une exigence fondamentale pour une inversion tomographique fiable.

Quelle que soit la méthode sismique utilisée en prospection de proche surface, la profondeur d'investigation est, en pratique, liée à l'offset maximal. Les phases des premières arrivées des traces à l'offset lointain présentent un faible rapport signal sur bruit et conduisent en conséquence à des incertitudes élevées. Une problématique double est liée aux pointés dégradés par le bruit. S'ils sont conservés, la qualité d'inversion (ou l'interprétation en couches) va être dégradée. Par contre, s'ils ne sont pas pris en compte, c'est la profondeur d'investigation qui va diminuer. Ainsi ces pointés doivent être affinés, soit manuellement par l'inspection visuelle (Yordkayhun, 2011), soit semi automatiquement via les techniques de corrélation croisée (Akram et al., 2014; Lanz et al., 1998), soit en attribuant une certaine incertitude à chaque pointé avant l'inversion (Diehl et al., 2009; Küperkoch et al., 2010; Zelt et al., 2006). Récemment, l'estimation automatique des incertitudes de temps des premières arrivées a pris, pour les sismologues, une importance équivalente à celle du pointé automatique lui-même (Di Stefano et al., 2006). Aujourd'hui, les pointés associés avec une grande barre d'erreur peuvent être écartés ou pondérés avec de faibles poids avant utilisation.

Malheureusement, tous les logiciels d'inversion tomographique de la sismique appliquée ne prennent pas en compte l'incertitude des temps de premières arrivées.

Le développement des algorithmes de pointé automatique combinant différentes méthodes, dans la communauté des sismologues, a été motivé, d'une part, par la certitude que la combinaison de différents aspects du sismogramme peut améliorer le pointeur automatique (Cuenot, 2003; Nippres et al., 2010) et d'autre part, par la possibilité d'estimer automatiquement l'incertitude des pointés résultants (Diehl et al., 2009). En effet, ces algorithmes sont devenus des standards en sismologie. L'utilisation de ce genre d'algorithmes est encore limitée au pointé automatique s'agissant des données de la sismique appliquée.

Notons qu'un filtrage fréquentiel est souvent appliqué avant le pointé automatique (ou manuel) malgré ses effets négatifs sur les caractéristiques (polarité et temps) des signaux de premières arrivées (Allam et al., 2014; Senkaya and Karsh, 2014). De tels filtres ont été adaptés pour filtrer les signaux ayant une bande de fréquence différente de celle du bruit. Dans certains cas, atténuer les bruits sismiques en utilisant les filtres fréquentiels de façon extravagante et aveugle, change radicalement les caractéristiques des premières arrivées (Aldersons, 2004).

Une alternative peut être d'avoir recours à des méthodes de filtrage ou dé-bruitage (*denoising*) plus sophistiquées qui permettent de transférer les données dans un domaine dans lequel le signal utile et le bruit peuvent être séparés. La représentation temps-fréquence (RTF) est devenue de fait un outil puissant pour analyser le comportement des séries temporelles, notamment les signaux variant en temps et en fréquence tel que le signal sismique (Stockwell et al., 1996). Par exemple, Parolai (2009) et Tselentis et al. (2012) ont proposé de filtrer le sismogramme en utilisant la transformée en S (Stockwell et al., 1996) avant le pointé afin d'augmenter la détectabilité des premières phases sismiques. La transformée en ondelettes (*Wavelet Transform* « WT ») a été également utilisée pour améliorer le sismogramme avant le pointé automatique (par ex. Gaci, 2014), et pour concevoir des techniques plus complexes (par ex. Anant and Dowla, 1997; Bogiatzis and Ishii, 2015; Soma et al., 2005; Zhang et al., 2003). La transformée en S présente un avantage par rapport à la transformée en ondelettes, car elle conserve la phase absolue de chaque composante de fréquence analysée (Pinnegar and Eaton, 2003). Cependant, malgré ses avantages potentiels, la décomposition temps-fréquence en S reste « calculatoirement » un peu coûteuse, par rapport à la transformée en ondelettes. D'autres auteurs ont essayé d'améliorer le rapport signal sur bruit (RSB) dans d'autres domaines, par exemple par réfraction virtuelle (ou interférométrie) (par ex. Mallinson et al., 2011; Mikesell et al., 2012), ou dans le domaine de Radon (ou tau-p) (Mousa and Al-Shuhail, 2012; Sabbione et al., 2015).

Par ailleurs, la redondance des données sismiques issues de la prospection de proche surface (imagerie par la réfraction) offre de nombreuses perspectives d'améliorer leur

qualité. La sommation ou « *stacking* » en couverture multiple, après une correction de temps dite « correction dynamique », est une des méthodes classiquement employée en sismique réflexion pour améliorer le rapport signal sur bruit (RSB). En supposant qu'on somme N traces, l'amplitude des signaux (dits *stackés*) est constructivement proportionnelle à N , tandis que celle des autres signaux, y compris le bruit, n'est que d'un facteur égal à \sqrt{N} . Par conséquent, l'amélioration du RSB est en \sqrt{N} (par ex. [Yilmaz, 2001](#)). En parallèle de la sommation conventionnelle (simple sommation linéaire), la sommation pondérée est communément utilisée afin d'améliorer la qualité des sections sismiques. Les poids de cette sommation sont calculés soit à partir de la corrélation ou ressemblance locale avec une trace référence ([Liu et al., 2009](#); [Sanchis and Hanssen, 2011](#)), du RSB ([Neelamani et al., 2006](#)), ou à partir de la phase instantanée ([Schimmel and Paulssen, 1997](#)). La qualité de la sommation fondée sur la corrélation dépend largement de la qualité de la trace référence ([Sanchis and Hanssen, 2011](#)). L'alignement des ondes réfractées est un enjeu majeur contrôlant la rentabilité de leur sommation, notamment lors de la sommation traditionnelle ([Rashed, 2008](#)). Plusieurs méthodes permettent d'estimer le retard entre deux traces et ainsi de le corriger (i.e. corriger la vitesse d'une onde pour la rendre infinie ou aligner les signaux). L'une d'elles est fondée sur les statistiques d'ordre deux, notamment sur le calcul de la fonction d'inter-corrélation entre un signal référence et ce même signal retardé en temps (par ex. [Lou et al., 2013](#); [VanDecar and Crosson, 1990](#)). L'alignement optimisé est souvent effectué par l'intermédiaire d'un processus itératif utilisant l'inter-corrélation ([Akram et al., 2014](#); [Zelt et al., 1987](#)). Cependant, [Liu et al. \(2009\)](#) ont prouvé que leur sommation fondée sur la ressemblance locale n'est pas influencée par le désalignement des traces *stackées*.

On peut noter que l'existence de plusieurs façons d'organiser les données d'un profil sismique (collection en points de tir communs, collection en points milieu communs (CMP), collection en offsets communs, etc.) permet de disposer de plusieurs types de sommation.

Par ces motivations mentionnées ci-dessus, l'objectif de cette thèse consiste dans les trois points suivants :

- développer un algorithme du **pointé automatique précis** de premières arrivées avec une estimation automatique de leurs **incertitudes**,
- mettre en place une procédure d'**édition automatique** des mauvais pointés,
- dans les cas des données d'une mauvaise qualité, **renforcer localement** les premières arrivées.

1.4 Organisation de la thèse

Le mémoire de thèse est organisé en six chapitres. Le premier est cette introduction. Le deuxième chapitre est consacré à une présentation historique et non exhaustive des techniques utilisées au pointé automatique des premières arrivées sismiques.

Dans le troisième chapitre, je présente une nouvelle approche développée ayant un double but : permettre le pointé automatique avec grande précision des premières arrivées sismiques et associer automatiquement à ce pointé une barre d'erreur. Cette approche est basée sur un algorithme adaptatif combinant trois méthodes de pointé automatique.

Le quatrième chapitre est une évaluation de la performance de l'approche proposée vis-à-vis d'autres stratégies récemment développées sur des données réelles à enjeux spéciaux. Dans un cadre de l'utilisation tomographique des premières arrivées sismiques, je propose une procédure automatique et un contrôle de la qualité des résultats issus de l'algorithme développé au troisième chapitre.

Suite aux conclusions et aux limitations de l'algorithme implémenté, le cinquième chapitre vise à pré-traiter les données sismiques avant le pointé automatique afin d'améliorer le RSB, notamment des premières arrivées sismiques aux offsets lointains. Pour cela, une nouvelle procédure basée sur une double sommation dans une fenêtre glissante a été développée. La performance de cette procédure a été testée sur des données synthétiques et réelles.

En conclusion de ce mémoire (sixième chapitre), je présente une synthèse des résultats et des algorithmes développés, ainsi que des perspectives d'applications pour des données trois composantes.

Revue de la littérature

Sommaire

2.1 Définitions générales.....	9
2.1.1 Stratégies de pointé manuel.....	10
2.2 Techniques de pointé automatique.....	12
2.3 Méthodes reposant sur les calculs d'énergie dans le domaine temporel.....	14
2.4 Les méthodes d'analyse dans le domaine fréquentiel ou temps-fréquence (échelle)	19
2.5 Les méthodes d'analyses basées sur des propriétés statistiques..	19
2.6 Méthodes basées sur les approches autorégressives (AR).....	21
2.7 Les méthodes basées sur l'analyse du mouvement des particules (polarité de l'onde sismique)	22

2.1 Définitions générales

Tout d'abord, il est nécessaire que l'on mette en place une définition satisfaisante de la première arrivée sismique. Classiquement, elle est définie par le premier départ notable (*first kick*) de l'impulsion sismique émergeant du bruit de fond ([Hatherly, 1982](#)).

Dans l'exploration sismique, le pointé des débuts de phases sismiques est une tâche à effectuer le plus précisément possible pour un nombre considérable de traces, souvent supérieur à plusieurs milliers. En général, ces arrivées sont associées à l'énergie de l'onde réfractée le long d'une interface ou d'une onde directe qui se propage directement de la source vers le récepteur.

Lorsqu'il s'agit d'une source explosive, le signal est causal et le temps de première arrivée à estimer correspond au début de l'émergence du signal dans le bruit. Toutefois, pour une source vibratoire, le signal n'est pas causal du fait de la corrélation mise en œuvre dans le pré-traitement, et le temps de première arrivée à estimer est obtenu en pointant le maximum de la première ondelette. La précision de ce pointé est dégradée par la faible impulsivité de la première arrivée.

La qualité du pointé des premières arrivées est liée à la structure proche surface, au type de la source, et au rapport de signal sur bruit ([Sabbione and Velis, 2010](#)). [Diehl \(2008\)](#) a signalé que la forme de l'ondelette sismique peut être aussi affectée par d'autres facteurs comme par exemple : la dispersion, l'atténuation, la diffusion, l'interférence avec d'autres phases, et les caractéristiques du système d'enregistrement.

En cas d'enregistrements (locaux et régionaux) de haute fréquence, la superposition de ces composantes peut conduire à des formes d'ondes très complexes.

2.1.1 Stratégies de pointé manuel

En général, une phase sismique est identifiée par l'une de deux observations visuelles :

- Un changement considérable en amplitude : l'amplitude dépasse le bruit de fond pendant un certain temps. Ceci peut également être désigné comme une modification du rapport signal sur bruit basé sur l'amplitude (ARSB). Sur ce principe, une phase sismique peut être déterminée si son amplitude est supérieure au bruit par au moins un facteur de 1.5 (i.e., $ARSB \geq 1.5$) (Diehl, 2008). La **Figure 2.1a** représente un exemple typique d'une première arrivée caractérisée par un changement de l'ARSB.
- Un changement soudain du contenu fréquentiel : un changement de la fréquence dominante indique l'arrivée d'une phase sismique. Cette observation peut être référée comme un changement dans le rapport signal sur bruit reposant sur la fréquence (FRSB). Contrairement à l'ARSB, il est souvent beaucoup plus difficile à quantifier visuellement. Par ailleurs, les fréquences dominantes du bruit et du signal peuvent parfois être très similaires. Mais le FRSB peut aider à déterminer la position approximative d'une phase sismique, surtout pour des grandes fenêtres temporelles (Diehl, 2008). La **Figure 2.1b** représente un exemple typique d'une première arrivée localisée par un changement du FRSB.

Une fois la phase sismique reconnue par l'un de ces critères (ARSB ou FRSB) ou, plus vraisemblablement, par une combinaison des deux, le début de l'ondelette de la première arrivée doit être pointé le plus précisément possible dans cette zone de transition. Cependant, sa position n'est pas complètement indépendante de l'incertitude de l'observation visuelle (i.e., de la détermination de la zone de transition). En outre, l'identification de cette zone en basant sur l'inspection visuelle de l'ARSB et de la FRSB est à la fois assez subjective et fortement dépendante de la fenêtre temporelle utilisée, ainsi que de la mise à l'échelle de l'amplitude (Diehl, 2008). Dans ce contexte, le temps absolu d'une arrivée sismique est difficile à déterminer. Souvent, il est même impossible de mettre en place une définition robuste d'une arrivée pour un signal sismique à bande de fréquence limitée, et échantillonné en présence de bruit. Par conséquent, un pointé cohérent de ce début ne peut être réalisé que d'un point de vue probabiliste (**Figure 2.1**).

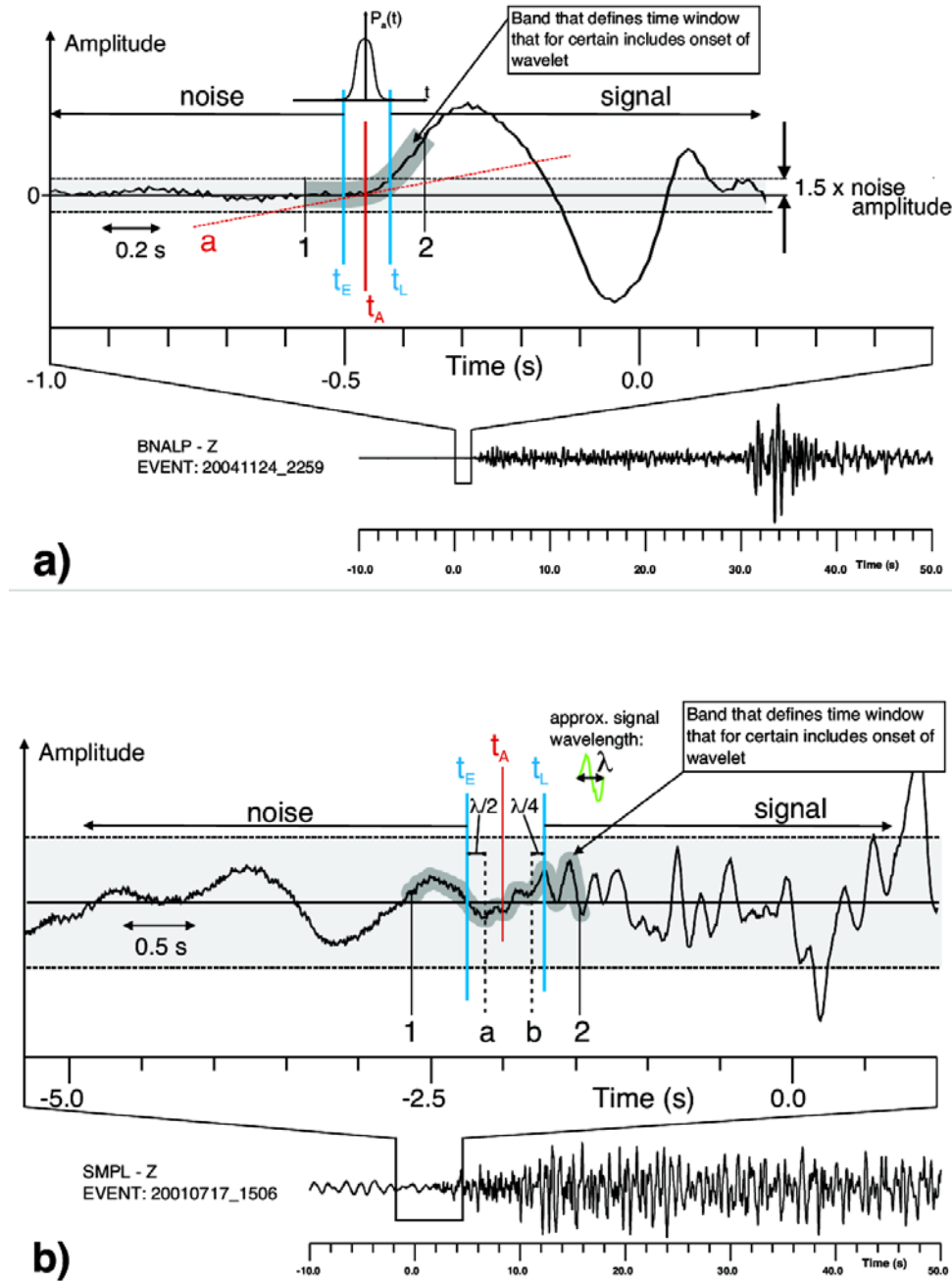


Figure 2.1. Approche du pointé d'un point de vue probabiliste (d'après Diehl, 2008): le plus tôt choix possible correspond à t_E , le plus tard choix possible correspond à t_L . Le temps de la première arrivée la plus probable se trouve dans cet intervalle (i.e., compte tenu de l'apparition d'une phase sismique en fonction probabiliste P_a , le temps d'arrivée est exprimé par le temps le plus probable t_A , c.-à-d. avec $P_a(t_A) = \max(P_a)$. D'autre part, la probabilité d'apparition $P_a(t_E)$ et $P_a(t_L)$ se rapproche de zéro) (Diehl, 2008). (a). Stratégie du pointé fondée sur l'attribut de l'amplitude (ARSB). (b). Stratégie du pointé fondée sur l'attribut de la fréquence (FRSB).

D'une part, au point de vue physique, le mouvement des particules du sous-sol augmente progressivement lorsque la perturbation des ondes sismiques arrive (Ricker, 1953), et le bruit de fond peut submerger la première partie d'ondelette sismique et retarder son apparition (Hatherly, 1982). D'autre part, d'un point de vue pratique, pour un tir sismique d'une source active, les pics/creux ultérieurs dans le même train d'onde de première arrivée sont plausiblement plus forts après un certain offset. Par conséquent, il est parfois utile d'inspecter le signal rétrospectivement pour estimer le temps précis de la première arrivée (Milsom and Eriksen, 2011, p. 201). Cependant, parce que les hautes fréquences sont sélectivement absorbées par le sous-sol (Futterman, 1962), l'intervalle temporel entre la première arrivée et un pic/creux ultérieur augmente progressivement avec l'offset.

2.2 Techniques de pointé automatique

Avant le développement des outils informatiques tel qu'on les connaît aujourd'hui, les premières arrivées des ondes sismiques (sismique réfraction et sismologie globale) étaient pointées manuellement. Ce travail manuel nécessitait un important investissement humain et un temps considérable surtout si le volume de donnée était important. Avec le développement des outils informatiques et la puissance de calcul des machines dans les années 70, un grand nombre d'algorithmes de pointé automatique des premières arrivées sismiques a été développé. Ces algorithmes ont remplacé le travail manuel du fait qu'ils ont l'avantage d'être rapides. Toutefois, ceux-ci ont été développés pour résoudre certains problèmes particuliers (tomographie, monitoring en temps réel, etc.) ou pour un certain jeu de données. Par ailleurs, la variabilité de la forme et de la phase des ondes sismiques et surtout en présence de bruit, le pointé automatique reste encore un défi. Il est alors difficile de faire une chronologie des approches et des techniques développées. En conséquence, il est impossible de prétendre à l'exhaustivité d'une revue historique.

Au cours de ces dernières décennies, plusieurs algorithmes de pointé automatique, notamment pour pointer la première arrivée des ondes P, ont été développés. La plupart d'entre eux ont été développés pour des applications sismologiques. Cependant, les techniques développées pour la sismique exploratoire (sismique appliquée à une source active ou passive, micro sismique) sont souvent les mêmes algorithmes qu'en sismologie avec toutefois certaines modifications de leurs paramètres selon la nature des signaux.

D'un point de vue sismologique, et en premier lieu, pour localiser un séisme, on distingue les méthodes de la détection et celles de pointé d'une phase intéressante (Allen, 1982; Tselentis et al., 2012). La plupart des méthodes de détection automatique sont basées sur l'analyse du rapport signal sur bruit (RSB) en fonction du temps. L'astuce commune consiste à calculer un attribut caractéristique qui est élaboré pour renforcer les changements dans le signal. Par contre, le pointé automatique de cette

phase déterminée par les méthodes de détection est une étape *a posteriori*, et est basée sur un critère d'augmentation ou de diminution brutale et locale d'une fonction attribut calculée à partir de la trace sismique. Autrement dit, un pointeur incorpore la fonction de détecteur. La différence la plus importante entre pointeur et détecteur de phase est la précision du pointé au moment de la première arrivée.

Cependant, les techniques de pointé automatique peuvent être classifiées selon le nombre de composantes ou de traces prises en compte pendant l'opération de pointé ; (i) méthodes adaptées pour pointer une trace unique (ou le sismogramme d'une station mono composante) (par ex. [Baer and Kradofer, 1987](#); [Coppens, 1985](#); [Earle and Shearer, 1994](#)). (ii) méthodes basées sur l'analyse de la polarisation des ondes qui sont généralement développées pour pointer les données trois composantes (par ex. [Cichowicz, 1993](#); [Reading et al., 2001](#); [Lois et al., 2013](#)). (iii) méthodes adaptées pour un groupe de traces profitant des quelques propriétés qui révèlent la présence des phases des premières arrivées sur l'enregistrement (par ex. [Blais, 2012](#); [Gelchinsky and Shtivelman, 1983](#); [Sabbione and Velis, 2010](#)). Elles sont souvent utilisées pour le pointé des données de sismique appliquée ou de micro-sismique.

D'autre part, il est possible de classer les différentes approches du pointé automatique selon les critères proposés par Withers et al. (1998). Ils proposent de les classer en quatre catégories selon le domaine dans lequel la fonction caractéristique (FC) est calculée :

- dans le domaine temporel (par ex. [Allen, 1978](#); [Stewart, 1977](#)),
- dans le domaine fréquentiel (par ex. [Goforth and Herrin, 1981](#)),
- à partir des paramètres de polarisation (par ex. [Aster et al., 1990](#)),
- par méthodes adaptatives (c.à.d. dont les paramètres s'adaptent automatiquement aux données) (par ex. [Withers et al., 1999](#)).

Ces méthodes servent souvent de bases aux algorithmes de pointé plus sophistiqués. Bien qu'aujourd'hui l'augmentation des capacités de calcul permette le développement d'algorithmes de plus en plus sophistiqués, Withers et al. (1998) soulignent, dans une comparaison des méthodes précédentes, que le gain apporté par les méthodes les plus complexes est faible par rapport au coût de mise en œuvre. De plus, lorsque ces méthodes doivent gérer simultanément d'importantes quantités d'informations, ces dernières sont, toujours selon Withers et al. (1998), à proscrire au profit des méthodes du domaine temporel en général moins complexes.

Cuenot (2003) a proposé un état de l'art en matière d'algorithmes de détection et de pointage automatique de début de phase sismique. Il a mis en place trois familles d'algorithmes : (i) par l'analyse de l'énergie (AE), (ii) avec les méthodes autorégressives (AR), (iii) au moyens de réseaux de neurones (RN), et (iv) enfin via une combinaison de plusieurs méthodes.

Ici, les méthodes et les techniques automatiques de détection et de pointé des premières arrivées sont celles où la fonction caractéristique (CF) peut être calculée comme décrite dans les paragraphes suivants.

2.3 Méthodes reposant sur les calculs d'énergie dans le domaine temporel

La plupart des algorithmes existants s'appuient sur l'analyse de l'énergie des champs d'onde incidents et de son évolution au cours du temps. Ces algorithmes sont souvent appelés dans la littérature « STA/LTA » du fait qu'ils ont comme principe la comparaison de la moyenne à court terme (STA = *Short Term Average*) avec la moyenne à long terme (LTA = *Long Term Average*). Les STA sont sensibles aux augmentations rapides d'amplitude dans le temps. Elles permettent de mesurer l'amplitude instantanée du signal. Les LTA mesurent l'amplitude locale du bruit de fond. Elles permettent donc de surveiller la moyenne du bruit sismique. Le rapport STA/LTA à un instant donné (t) donné correspond au RSB local. Ce rapport est ainsi comparé à seuil prédéterminé, et si cette valeur est dépassée, un déclenchement est signalé (**Figure 2.2**). Si l'événement satisfait ensuite à quelques critères simples, il est reporté. Les détails de la procédure sont largement décrits par ([Allen, 1982](#); [Baer and Kradolfer, 1987](#); [Trnkoczy, 2012](#)).

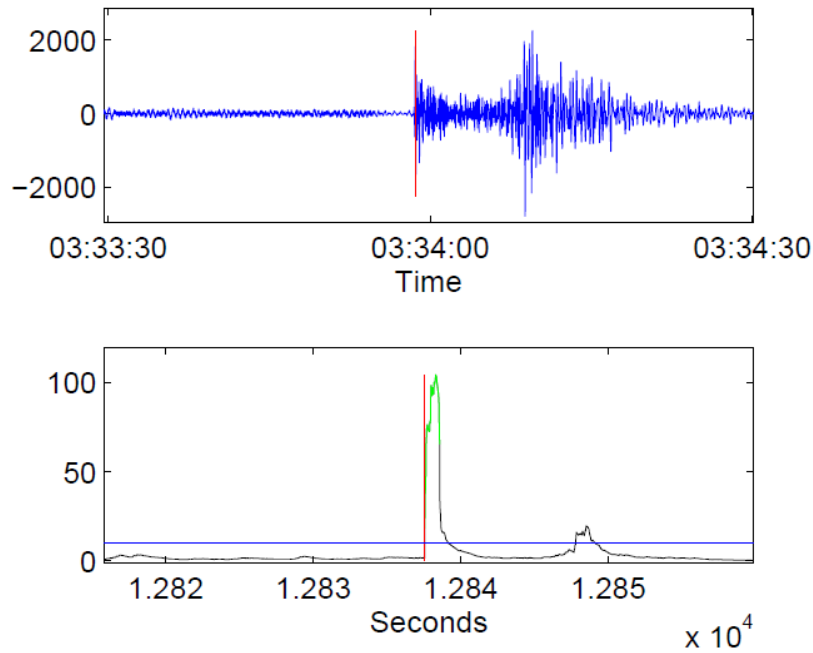


Figure 2.2. Algorithme STA/LTA: détermination et pointé de phase. **(en haut)**. Sismogramme filtré avec un filtre passe-bande (2-10 Hz) et pointé automatique (ligne rouge verticale). **(en bas)**. Fonction STA/LTA calculée par STA = 1 sec et LTA = 20 sec et seuil THR (ligne bleu). Le déclenchement correspond à l'intersection entre la courbe de STA/LTA et le seuil (ligne rouge) et la partie verte de la courbe représente le maximum local (d'après Küperkoch, 2010)

Il existe une vaste gamme d'algorithmes de pointé, allant d'un type simple s'appuyant sur un seuil d'amplitude à ceux très sophistiqués utilisant des méthodes adaptatives (Sharma et al., 2010). Les différentes variantes de l'algorithme STA/LTA dans le domaine temporel proviennent du calcul de la fonction caractéristique (FC). La FC utilisée dans le calcul STA/LTA peut être aussi simple que la valeur absolue de la trace, ou l'énergie (Wong et al., 2009), ou plus complexe.

Swindell and Snell (1977) ont proposé le Z-détecteur. C'est une forme modifiée de détecteur d'énergie conventionnelle (STA/LTA), utilisant la variable normalisée Z (2.1) comme une fonction caractéristique :

$$Z_i = \frac{x_i - \mu}{\sigma}, \quad (2.1)$$

où x_i est considérée comme moyenne à court terme (STA), μ et σ étant respectivement sa moyenne et son écart type. Le Z-détecteur estime l'écart des données sismiques à la valeur-moyenne, exprimée en unité de son écart-type. Il présente également l'avantage de s'adapter automatiquement à la variance du bruit fond (Withers et al., 1998). L'existence d'événements sismiques est déterminée après avoir comparé cette FC par rapport à un seuil prédéfini (pour plus des détails voir Berger and Sax, 1981).

Stewart (1977) a utilisé, à partir des différences de premier ordre du sismogramme, une enveloppe modifiée permettant d'accentuer les variations de pente. La fonction caractéristique résultante correspond à une version filtrée haute-fréquence des données originales. Les avantages de cette méthode sont que (i) le comportement oscillatoire du signal et la polarité de la première ondelette sont conservés, (ii) les fréquences indésirables sont éliminées, et (iii) la probabilité d'identifier des premières arrivées dans des environnements bruyants est augmentée (Stewart, 1977). En outre, la vitesse du calcul est un autre avantage, parce que cette méthode utilise uniquement les opérations élémentaires (l'addition, la soustraction et la comparaison) (Allen, 1982). Un ensemble de critères de détection et de seuils doivent alors être mis en place pour déclarer un événement (Berger and Sax, 1981; Withers et al., 1998).

Allen (1978) a développé un algorithme reposant sur une enveloppe égale au carré des données plus une pondération du carré de la dérivée première (2.2). Cette fonction caractéristique inclut alors des composantes des données non filtrées et des données filtrées avec un filtre passe-haut (Withers et al., 1998).

$$FC(t) = x^2(t) + C_1 x'(t)^2, \quad (2.2)$$

où $x(t)$ est le signal sismique et $x'(t)$ sa dérivée première.

Il est évident que le choix de cette fonction vise à créer une nouvelle série temporelle qui fournit des informations sur la variation du signal à la fois dans le domaine temporel et fréquentiel. La constante C_1 correspond à une fonction poids, sa valeur dépend de la fréquence d'échantillonnage et du bruit fond de chaque station. A partir de la FC (équation (2.2)), les STA et LTA sont calculées par récurrence:

$$\begin{aligned} STA(t) &= STA(t-1) + C_2 (FC(t) - STA(t-1)) \\ LTA(t) &= LTA(t-1) + C_3 (FC(t) - LTA(t-1)), \end{aligned} \quad (2.3)$$

où C_2 et C_3 sont des constantes dépendant des filtres utilisés.

Ce détecteur d'événement est suivi par une série de contrôles logiques et mathématiques afin de reconnaître les phases sismiques et leurs temps d'arrivées. De plus, l'algorithme calcule le contenu fréquentiel du signal à travers la pluralité des zéros (passages par zéro), la forme de la fonction d'enveloppe, la polarité de la première impulsion, puis procède à une estimation de l'erreur des résultats (Allen, 1982; 1978). Cet algorithme peut être utilisé en temps réel parce qu'il analyse les critères échantillon par échantillon grâce à la simplicité du calcul. Néanmoins, il présente des inconvénients importants tels qu'un seuil fixe, l'abondance des paramètres empiriques (les constantes et les longueurs de fenêtres de STA , LTA) qui le rend incapable de détecter des événements des enregistrements où le RSB est très variable.

Baer and Kradolfer (1987) ont proposé une version modifiée de l'algorithme d'Allen afin de surmonter les faiblesses de ce dernier, à la fois pour pointer précisément le début de phase P et pour réduire la charge de calcul de la détection en temps réel. Les

modifications ont été faites sur la FC et aussi sur la procédure de détection en incorporant une valeur-seuil déterminée dynamiquement et sensible au changement du bruit. Küperkoch et al. (2010) ont montré que cette FC n'est pas seulement sensible aux changements en amplitude et fréquence, mais aussi en phase.

Earle and Shearer (1994) ont proposé un algorithme reposant sur la technique de STA/LTA et permettant de faire du pointé de phases. Etant donné que leur algorithme est sensible aux petites variations en amplitude, il est bien adapté à détecter autant de phases possibles dans le sismogramme. Ils ont utilisé l'enveloppe analytique du sismogramme comme une FC plutôt que son énergie comme entrée du filtre STA/LTA. Cette FC peut être définie :

$$E(t) = \sqrt{x^2(t) + \bar{x}^2(t)}, \quad (2.4)$$

où $x(t)$ le sismogramme et $\bar{x}(t)$ la transformé de Hilbert. La série temporelle des valeurs prises par le rapport STA/LTA est convoluée avec une fenêtre de Hanning afin d'atténuer les effets négatifs du bruit impulsionnel. Lorsque la fonction de rapport lissée dépasse le seuil prédéfini, une phase d'arrivée est détectée. Le temps de l'arrivée correspond au point d'inflexion du rapport situé entre le point de déclenchement et le maximum local qui le suit.

Dans la même veine, le nombre des fenêtres et leurs configurations ont encore élargi la variété des algorithmes basés sur STA/LTA. Coppens (1985) a proposé une approche reposant sur les rapports d'énergie de deux fenêtres temporelles imbriquées pour pointer les premières arrivées sismiques (source active) sur la collection d'offset-commun. Cette méthode diffère des précédentes par la largeur des fenêtres où la longueur de la deuxième (LTA) augmente en fonction du temps. Une version modifiée de cet algorithme (appelée MCM = *Modified Coppens Method*) a été proposée par Sabbione and Velis (2010) qui ajoutent une constante de stabilisation permettant de réduire les fluctuations rapides du rapport de l'énergie (voir **Chapitre 4**). La **Figure 2.3** montre un exemple des différentes étapes du pointé automatique par l'algorithme de MCM.

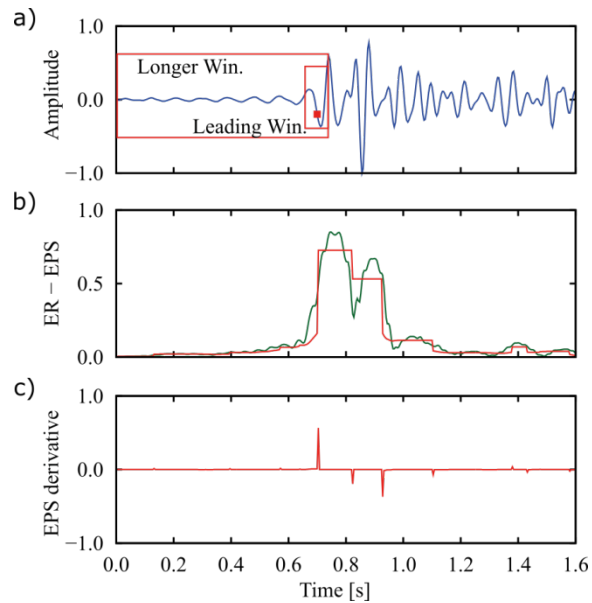


Figure 2.3. Exemple d'événement pointé par la technique MCM: **(a)** configuration d'algorithme (deux fenêtres imbriquées) **(b)** le rapport d'énergie (en vert) et sa version filtrée par le filtre de préservation des bords (en rouge). **(c)** la dérivée de premier ordre où la première arrivée (illustrée par carré rouge sur (a)) correspond au maximum de cette dérivée (modifiée d'après [Sabbione and Velis, 2010](#)).

Une version particulière de l'algorithme STA/LTA a été proposée par Han et al. (2009). Elle consiste à calculer le rapport d'énergie de deux fenêtres égales (avant et arrière). La largeur de la fenêtre doit être deux à trois fois la période dominante de la première arrivée (Han et al., 2010) (pour plus des détails, voir **Chapitre 4**).

La technique ATA/BTA/DTA (Chen and Stewart, 2005) consiste à mesurer la moyenne du rapport signal/bruit sur une durée de temps avant (BTA = *Before Term Average*), après (ATA = *After Term Average*), et après un délai plus long (DTA = *Delayed Term Average*) d'un point de temps instantané. D'après eux, cette méthode est capable de discriminer entre une forte amplitude de faible durée (événement) et une forte amplitude de longue durée (bruit).

Les choix de la longueur des fenêtres ainsi que du critère de détection exigent une bonne connaissance des propriétés des phases visées (amplitude, fréquence, durée). Pour s'affranchir du problème d'un seuil fixe, Akram (2014) a proposé un critère de détection dynamique qui tient compte des propriétés instantanées du signal. D'autre part, des variantes de l'algorithme STA/LTA capables d'adapter dynamiquement la longueur respective des fenêtres en fonction du contenu fréquentiel du signal ont été développées (Tong, 1995). Parmi ces méthodes, l'algorithme STA/LTA simple occupe une place de choix. Mais il faut quand même noter que puisque l'algorithme de STA/LTA est très sensible au niveau de bruit quand il change de façon significative, sa précision est rapidement dégradée.

2.4 Les méthodes d'analyse dans le domaine fréquentiel ou temps-fréquence (échelle)

Le domaine spectral est un autre domaine classique de description du signal. Par exemple, en s'appuyant sur la transformée de Fourier rapide, Shensa (1977) a proposé un détecteur automatique de la densité spectrale de puissance. Un autre détecteur utilisant la transformée de Walsh a été développé par Goforth and Herrin (1981).

Cependant, l'analyse du spectre des signaux n'est pas toujours suffisante pour lever l'ambiguïté de l'identification, du fait de la non-stationnarité des signaux. Une approche plus physique des signaux consiste alors à représenter ceux-ci conjointement en temps et en fréquence. La transformée en ondelettes est l'une des descriptions les plus utilisées en sismologie ces dernières années.

Anant et Dowla (1997) ont utilisé la transformée en ondelettes discrète (DWT) afin de pointer les temps d'arrivée des ondes P et S. Leur algorithme utilise les informations d'amplitude et la polarisation des ondes des données à trois composantes. Zhang et al. (2003) ont utilisé l'analyse multi-échelle par ondelettes avec le critère AIC (*Akaike Information Criteria* (Akaike, 1971)) pour les enregistrements à une seule composante. L'idée de base est fondée sur le fait que les caractéristiques principales d'un signal, telles que les phases sismique, apparaissent sur les coefficients d'ondelettes à différentes gammes de fréquences. Par ailleurs, la transformée en ondelettes continues a été aussi investiguée pour le pointé automatique (Bogiatzis and Ishii, 2015; Karamzadeh et al., 2013).

D'autres domaines de la représentation temps-fréquence ont été adaptés au pointé automatique, par exemple la transformée de Fourier à court terme (Hafez, 2009) et la transformée en S (Stefano Parolai, 2009; Tselentis et al., 2012).

2.5 Les méthodes d'analyses basées sur des propriétés statistiques

Avec l'arrivée d'un signal sismique, les propriétés statistiques du sismogramme changent brusquement. Par conséquent, en mesurant l'une de ces propriétés à l'aide d'une fenêtre mobile pour construire une FC, on identifie l'instant de première arrivée. D'après Küperkoch et al. (2010), étant donnée une variable aléatoire réelle x de n valeurs, les moments centrés d'ordre 2, 3, et 4 (respectivement, la variance ($\hat{\sigma}^2$), le coefficient d'asymétrie, appelé le « *skewness* » (\hat{S}), et le coefficient d'aplatissement, appelé le « *kurtosis* » (\hat{K})) peuvent être calculés avec le moment centré (\hat{m}) d'ordre (k) selon les équations (2.5):

$$\hat{m}_k = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^k, \quad (2.5)$$

$$\hat{\sigma}^2 = \hat{m}_2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2,$$

$$\hat{S} = \frac{\hat{m}_3}{\hat{\sigma}^3},$$

$$\hat{K} = \frac{\hat{m}_4}{\hat{\sigma}^4},$$

où \bar{x} est la moyenne arithmétique, et $\hat{\sigma}$ est l'écart type.

Saragiotis et al. (2004, 2002) ont été parmi les premiers à utiliser les propriétés statistiques d'ordre supérieur (*HOS = Higher Order Statistics*) pour le pointé automatique. Ils ont introduit le pointeur (*PAI-S/K = Phase Arrival Identification-Skewness/Kurtosis*) fondé sur le fait qu'il y a une déviation du caractère Gaussien due à la présence d'une arrivée sismique (Figure 2.4).

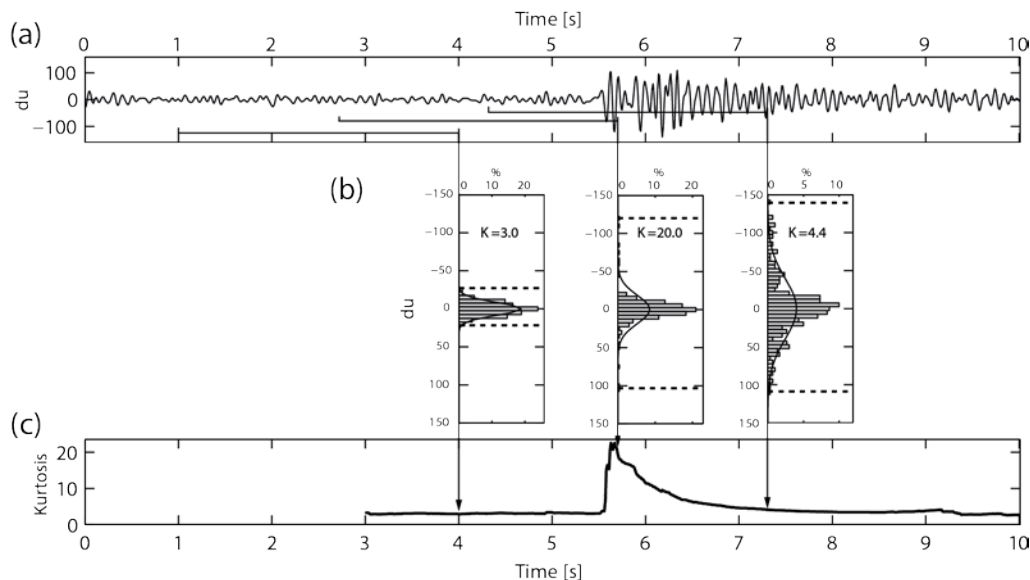


Figure 2.4. Exemple de kurtosis utilisé comme FC: **(a)**. trace sismique filtrée (3-45 Hz), les trois barres horizontales indiquent les fenêtres de calcul de kurtosis. **(b)**. Kurtosis et histogrammes de distribution pour les trois fenêtres, les lignes solide représentent le meilleur ajustement avec une distribution Gaussienne. **(c)**. La FC résultant basée sur le kurtosis (d'après Baillard et al., 2014).

Gentili and Michellini (2006) ont utilisé le kurtosis, le skewness, et une combinaison des deux avec leurs dérivés temporelles comme des entrées d'un pointeur utilisant un réseau de neurones artificiels. Küperkoch et al. (2010) ont ajouté l'AIC à la méthode de Saragiotis et al. (2002) pour optimiser l'identification du début de la phase P. Des séries de transformations non-linéaires ont été effectuées par Baillard et al. (2014) sur l'attribut de kurtosis afin d'isoler le début de l'émergence de FC (où cette dernière caractérise l'arrivée sismique).

2.6 Méthodes basées sur les approches autorégressives (AR)

Ces méthodes sont principalement utilisées pour le pointé de phase. Elles nécessitent d'avoir une estimation initiale des arrivées et doivent par conséquent être mises en œuvre en commun avec d'autres techniques. Une série temporelle $x(t)$ peut être modélisée de façon statistique par une méthode autorégressive :

$$x(t) = \sum_{k=1}^M a_k x(t-k) + \xi(t), \quad (2.6)$$

où M est l'ordre du modèle AR, a_k les coefficients ($k = 1, \dots, M$), et ξ représente le bruit blanc.

Les pointeurs AR sont basés sur l'idée que la non stationnarité de la trace sismique peut être estimée en la divisant en de petits segments localement stationnaires, chacun modélisé comme un processus AR (Takanami and Kitagawa, 1988). Les paramètres des modèles AR (l'ordre et/ou la valeur des coefficients) changent considérablement lorsque la caractéristique du segment en cours de sismogramme est différente du précédent (arrivée sismique) (Sleeman and van Eck, 1999). Le critère AIC (*Akaike Information Criteria* (Akaike, 1971)) est un critère mathématique permettant de qualifier l'accord entre des données et le modèle au sens des moindres carrés (Zhang et al., 2003), et est communément utilisé pour déterminer l'ordre du processus AR. Cependant, lorsque l'ordre d'AR est fixe, ce critère mesure la représentativité du modèle.

La théorie et la mise en œuvre des différents algorithmes d'AR-AIC sont décrits en détails par exemple, par Leonard and Kennett (1999). Récemment, Küperkoch et al. (2012) ont montré l'avantage à utiliser la prédiction d'erreur AR (*AR prediction error*), par rapport aux coefficients AR qui ne contiennent que les informations de contenu fréquentiel, pour détecter les changements de l'amplitude, la fréquence, la phase, ainsi que de la polarisation.

La **Figure 2.5** montre un exemple de pointé d'un événement par le pointeur AR-AIC de Sleeman and van Eck (1999).

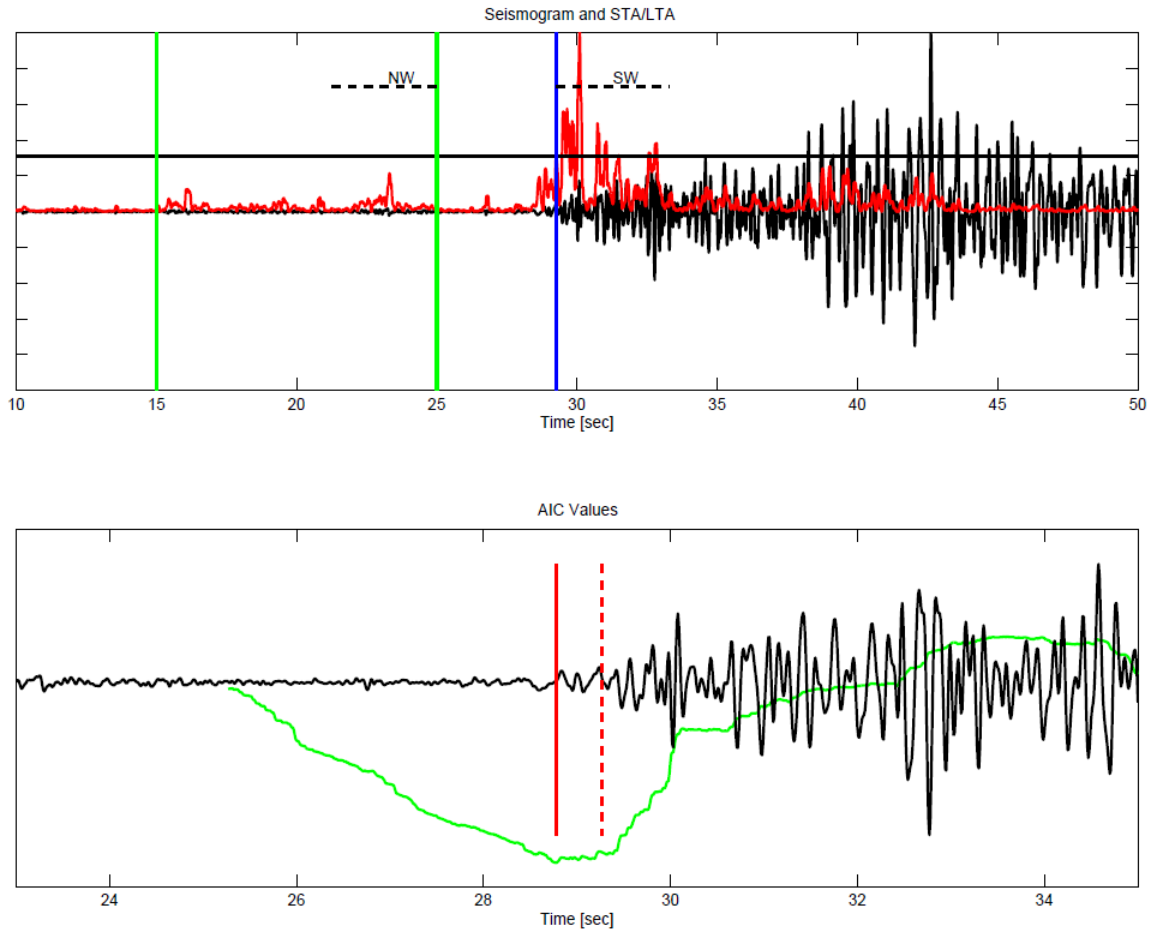


Figure 2.5. Exemple du pointé par AR-AIC: **(en haut)**. Sismogramme (noir) et pointé initial (bleu) détecté par le rapport STA/LTA (rouge). Le niveau du bruit a été estimé à partir du segment entre les deux lignes vertes. NW et SW sont les fenêtres de détermination des coefficients du modèle AR pour, respectivement, le bruit et le signal. **(en bas)**. Le sismogramme zoomé (noir) autour du pointé de STA/LTA (pointillé rouge) et le pointé automatique par AR-AIC (rouge). La fonction d'AIC (verte), (d'après Küperkoch, 2010).

Enfin, formulé en sismologie par Maeda (1985), le AIC peut être calculé directement à partir du sismogramme sans effectuer le processus AR. Cette méthode présente l'avantage d'être rapide en terme du temps de calcul et facile à mettre en œuvre.

2.7 Les méthodes basées sur l'analyse du mouvement des particules (polarité de l'onde sismique)

Ces techniques utilisent l'analyse de la polarisation de l'onde et des paramètres afférents ou des valeurs propres (*principal component analysis*). Les ondes de volume (P et S) présentent un haut degré de polarisation rectiligne. Cette dernière caractérise le mouvement d'une particule du sol au passage d'une onde sismique. La plupart des

développements a porté sur les données à 3 composantes du fait que la propriété de polarisation n'est accessible que pour les enregistrements des multi-composantes.

Montalbetti and Kanasewich (1970) ont modifié le filtre de polarisation de Flinn (1965) pour améliorer le RSB des enregistrements, et par conséquent la détectabilité des phases télé-sismiques. Cichowicz (1993) a assemblé des caractéristiques importantes d'une arrivée de l'onde S dans une fonction caractéristique qui se compose d'un produit de différents filtres de polarisation tels que l'angle de déflexion, le degré de polarisation, et le rapport entre les énergies transversale et totale. Cet algorithme peut être considéré comme un socle pour la détection des phases S, et fait souvent partie des algorithmes hybrides récents.

2.8 Méthodes alternatives

Citons également les méthodes de traitement relevant du traitement d'image (Mousa et al., 2011), et les méthodes fondées sur l'analyse de l'entropie (Sabbione and Velis, 2010) ou de la dimension fractale (Boschetti et al., 1996; Jiao and Moon, 2000; Sabbione and Velis, 2010) (voir **Chapitre 4**). D'autres méthodes utilisent l'inter-corrélation des phases sismiques (Gelchinsky and Shtivelman, 1983; Peraldi and Clement, 1972). Ces algorithmes ne sont souvent pas entièrement automatiques. Ces techniques sont souvent utilisées comme procédure semi-automatique pour raffiner/optimiser des pointés automatiques, voire même les pointés manuels dans une fenêtre temporelle contenant les premières arrivées (Rawlinson and Kennett, 2004; Senkaya and Karsh, 2014). On trouve également des techniques faisant usage des réseaux de neurones artificiels (Gentili and Michelini, 2006; McCormack et al., 1993), développées dans un premier temps dans le domaine pétrolier. Ce dernier type de méthodes nécessite une étape importante d'apprentissage (méthodes dites par apprentissage).

Enfin, des techniques hybrides ont fréquemment été proposées pour pointer automatiquement les premières arrivées, en particulier les phases des ondes S à partir des trois composantes du mouvement (par ex. Diehl et al. 2009). Ils combinent le détecteur de Baer and Kradolfer (1987) avec le détecteur de la polarité proposé par Cichowicz (1993). Les informations fournies par ces deux détecteurs sont utilisées pour guider le pointeur AR-AIC (Takanami and Kitagawa, 1988).

Nippres et al. (2010) ont introduit le concept de pointeurs automatiques, en tandem, en combinant soit le STA/LTA, soit la méthode d'estimation de la période prédominante (T_{pd}) (Hildyard et al., 2008) avec la FC de kurtosis développée par (Saragiotis et al., 2002).

Akram (2014) a développé un pointeur reposant sur le rapport commun de l'énergie ($JER = Joint Energy Ratio$) en combinant le STA/LTA avec le rapport des valeurs

propres ($PER = Peak\ Eigenvalue\ Ratio$). Ces valeurs propres ont été estimées par la décomposition en valeurs singulières de l'enregistrement des trois composantes.

Par ailleurs, les techniques automatique de la détection et du pointé ont été examinées régulièrement par certaines études méthodologiques, de façon détaillée (par ex. [Akram, 2014](#); [Munro, 2005](#); [Sabbione and Velis, 2010, 2013](#)).

Le choix de l'algorithme du pointé automatique dépend fortement, d'une part, des propriétés des signaux en question et de leur clarté (impulsivité) sur l'enregistrement, et du nombre de composantes, d'autre part, du meilleur compromis entre fiabilité et rapidité.

Développement de l'algorithme adaptatif

Sommaire

3.1 Introduction.....	28
3.2 Methodology	32
3.2.1 Principe and methods.....	32
3.2.2 Algorithms implementation.....	38
3.2.2.1 Single-trace algorithm picking	39
3.2.2.2 Error and quality assessment	40
3.2.3 Entire-shot record algorithm picking	43
3.3 Testing and results	46
3.3.1 Test on synthetic data.....	46
3.4 Conclusions	51
3.5 Acknowledgements	51

Ce chapitre correspond à un article soumis à *Near surface Geophysics* dans lequel nous proposons une nouvelle stratégie basée sur un algorithme adaptatif combinant plusieurs méthodes afin de mettre en place un pointeur automatique précis des données sismiques de proche surface.

A travers l'introduction, nous présentons une brève revue de la littérature qui permet de rendre compte de la diversité et du nombre important de méthodes déjà proposées pour le pointé automatique des premières arrivées sismiques. En effet, nous rappelons les avantages et les inconvénients de chacune d'elles, notamment en insistant sur leurs principaux défauts pour des données sismiques de proche surface.

La méthodologie proposée met en place une stratégie mimant le pointé manuel : lorsque une tendance globale de la localisation des premières arrivées est définie, afin que les pointés précis soient identifiés au voisinage de cette tendance. Basé sur cette stratégie, un algorithme adaptatif a été développé associant l'utilisation de fenêtres multiples imbriquées, l'estimation des propriétés statistiques d'ordre supérieur et le critère d'information d'Akaike.

La validation de la méthodologie proposée a été effectuée en deux étapes sur des données sismiques synthétiques, d'un modèle de vitesse multicouches (i) d'abord non-bruitées pour valider les paramètres internes de l'algorithme développé, (ii) puis associées à un bruit pour évaluer la performance de cet algorithme via des scénarios quasi-réels. Les résultats obtenus nous permettent de valider la robustesse et la précision de la méthodologie proposée.

Enfin, les conclusions mettent particulièrement l'accent sur les avantages apportés par l'intégration de plusieurs propriétés (l'énergie, la gaussianité et la stationnarité),

pour à la fois réaliser un pointeur efficace et robuste des ondes P, et définir ainsi une estimation de l'incertitude du pointé des temps d'arrivées (barre d'erreur).

Development of an adaptive multi-method algorithm for automatic picking of first arrival times: application to near surface seismic data

Amin Khalaf (1,2), Christian Camerlynck (1), Nicolas Florsch (1,3) and Ana Schneider (1).

(1) *Sorbonne Université, UPMC Univ. Paris 06, CNRS, UMR 7619 Metis, 4 place Jussieu, 75005 Paris, France.*

(2) *Tishreen University, Faculty of Sciences, Departement of Geology, Lattakia, Syria.*

(3) *Sorbonne Université, UPMC Univ. Paris 06, UMI 209 Ummisco, 4 place Jussieu, 75005 Paris, France.*

Corresponding author:

Amin Khalaf

UPMC, METIS.

Case 105, 4 place Jussieu, 75005 Paris, France.

Phone : +33 (0)1 44 27 48 85

Fax : +33 (0)1 44 27 45 88

E-mail address: amin.khalaf@upmc.fr

Submitted for Publication in Near Surface Geophysics

Abstract

Accurate picking of first arrival times plays an important role in many seismic studies, particularly in seismic tomography and reservoirs or aquifers monitoring. Many techniques have been developed for picking first arrivals automatically or semi-automatically. Most of them have been developed for seismological purposes which do not attain the accuracy objectives required in shallow seismic due to the complexity of near surface structures, and low signal-to-noise ratio. We propose a new adaptive algorithm for near surface data combining three picking methods: Multi-Nested Windows, Higher Order Statistics, and Akaike Information Criterion. They benefit from combining many properties, revealing the presence of first arrivals, finally providing an efficient and robust first arrivals picking. This strategy mimics the human first-break picking, where at the beginning the global trend is defined. Then the exact first-breaks are sought in the vicinity of the point the observed trend suggests. In a multistage algorithm, three successive phases are combined, where each of them characterize a specific signal property. Within each phase, the potential picks and their error range are automatically assessed, and then used sequentially as prior in the following phase picking. Since having realistic estimates of the error in picked traveltimes is crucial for seismic tomography, our adaptive algorithm provides automatically the associated time uncertainties of picked arrival times. The accuracy and robustness of the implemented algorithm are successfully proved on synthetics which have special challenges for automatic pickers. The comparison of resulting P-wave arrival times with those picked manually demonstrated a reliable performance of the new scheme under different noisy conditions. All parameters of our multi-method algorithm are auto-adaptive thanks to the integration in series of each sub-algorithm results in the flow. Hence, it is nearly a parameter-free algorithm, which is straightforward to implement and demands low computational resources.

3.1 Introduction

Identifying and picking seismic phase's times are very important processes in several fields (e.g. seismology, seismicity, applied seismic, among others). From a seismological point of view, up to now, important seismological information can be obtained such as location of earthquakes, interior structure of the Earth, geotectonic context of a region, and seismic risk evaluation. The applications of the arrival times in the field of micro- and seismicity are found for assessment and monitoring reservoirs or aquifers, including the passive seismic tomography. Regarding to the seismic prospection, first arrival times determination is found in the head of seismic refraction (layered and tomography) processing flow, and as auxiliary process in seismic reflection processing, particularly for static (weathering layer) corrections. The refraction interpretation quality of the inverted models (velocities and thickness) is influenced by the accuracy

of the first arrival times (Senkaya and Karsh, 2014). Although the error in arrival times may be small, a non-negligible error can result in velocity and depth calculations (Ramanantoandro and Bernitsas, 1987).

The first events can be classically defined as the first remarkable onset of seismic energy, i.e. “the first noticeable departure of the seismic pulse from a background signal” (Hatherly, 1982). However, they do manifest themselves as a noticeable variation in amplitude, frequency content, statistical properties, and/or polarization properties in time and/or frequency domain. First event quality depends on the source type and near-surface conditions (Yilmaz, 2001). The ambient noise may over-shadow the early part of the seismic phase and delay its apparent occurrence.

Traditionally, the first arrival times are manually picked by human operators who are familiar with various kinds of the waveforms and the considered seismic phase onsets. Currently, the huge data volume (especially in active-source seismic applications) makes this visual handpicking inefficient, relatively time- and labor-consuming. On the other hand, the subjectivity in human-eye picking (Hatherly, 1982; Jiao and Moon, 2000) and other factors relative to their physiological and/or psychological state make the manual picking results biased or even inconsistent (Saragiotis et al., 2002; Sabbione and Velis, 2010). The advantage of the manual method is the possibility to pick, with good accuracy, complicated events or those covered by the noise or by other less significant waves. However, such cases require an experienced analyst.

During the last decades, with advances in knowledge and technology, the automation of a procedure such as picking has become popular. Among the potential benefits of automatic picking, it allows rapid analysis of large data sets, with homogeneous and consistent results. Moreover, it is a valuable tool since it is more objective, but it may also provide false picks (Mousa et al., 2011). Picking techniques by computer (i.e. by algorithms) can be divided into semi-automatic and full-automatic. The semi-automatic algorithms are often based on applying an automatic procedure either to initial picks (e.g. VanDecar and Crosson, 1990) or guiding/refining its results by the user interaction (e.g. Lou et al., 2013). With poor quality data, the picking procedure using a semi-automatic algorithm with interactive correction by user can take up to 20-30% of the total processing time (Sabbione and Velis, 2010).

Many algorithms have been suggested for the full-automatic picking of seismic phase onsets. These algorithms can be classified into two main approaches: single-trace and multi-trace. Single-trace algorithm can be a “detector”, to recognize a seismic phase arrival, or a “picker”, to determine a precise onset time of a detected phase (Allen, 1982). Multi-trace algorithms can be seen as an extension of single-trace algorithm, working on several traces at once. They are usually used for picking the active-source seismic data, benefiting from the similarity of waveforms from nearby events. Recently, the number of developed techniques has increased. It is difficult to track different approaches and developments, and categorize all available methods. Withers et al.

(1998) sorted previously used methods into four classes: i) time domain (STA/LTA, Z-statistic), ii) frequency domain, iii) particle motion, and iv) adaptive window length processing. Recently, Tselentis et al. (2012) provided an excellent classification of the common event-detection and phase-picking methodologies as proposed in the literature. Among these categories, the most reported for single-trace (P-, S-phase, or both) picking are energy ratio criteria (STA/LTA) (Allen, 1982, 1978; Baer and Kradolfer, 1987; Earle and Shearer, 1994), autoregressive-Akaike information criteria (AIC) (Maeda, 1985; Sleeman and van Eck, 1999; Takanami and Kitagawa, 1988), higher order statistics (HOS) (Saragiotis et al., 2004, 2002), polarization information (Cichowicz, 1993), and wavelet transform (WT) (Anant and Dowla, 1997; Zhang et al., 2003). Each of these methods has advantages and drawbacks according to the emergence of the seismic phase of interest depending on seismograms and noise background. Most of the methods mentioned above were mainly implemented for seismological purposes.

A handful of automatic techniques for picking the first breaks of refraction records are reported in the literature by comparing to those for seismological applications. An algorithm based on cross correlation of the seismic trace with a reference “model” trace is proposed by Peraldi and Clement (1972). This picking method fails to accurately pick the first arrivals because they are based on incorrect assumption, which is that the waveform shape does not change markedly from trace to trace (Gelchinsky and Shtivelman, 1983). Hatherly (1982) proposed an algorithm combining the correlation method with a linear least-squares prediction method. This algorithm is non-robust against random noise. Gelchinsky and Shtivelman (1983) used a method based on spatial correlation properties of the refracted signals. To prevent an apparent cycle skipping problem in cross correlation approach, Zelt et al. (1987) implemented a Monte Carlo technique, in which the cross correlation function is transformed into a probability distribution, with semblance optimization to pick low amplitude arrivals in crustal refraction data. However, accuracy of the methods based on cross-correlation may be affected significantly when the data has low signal-to-noise (S/N) ratio.

Data pre-processing can help for improving the detectability of first onsets and further for increasing the accuracy of automatic picking (Mousa et al., 2011; Mousa and Al-Shuhail, 2012). The related methods, which are most often based on detecting sudden changes in some attributes of the seismic trace, generally provide quite stable arrival times with moderate signal-to-noise ratios (Blais, 2012). But their accuracy can be reduced with noisy scenarios, requiring tuning of some of their parameters, especially regarding the calculating windows size. Among these approaches, energy ratio method and its variants (Coppens, 1985; Sabbione and Velis, 2010; Wong et al., 2009), entropy and fractal dimension methods (Boschetti et al., 1996; Sabbione and Velis, 2010), are well-known and more reported in the automatic picking literature. Non-traditional methods based on artificial neural networks (Gentili and Michelini, 2006; McCormack et al., 1993), which is trained to recognize and pick phase onsets, are

used also for picking the refracted arrivals. Their shortcoming is that they need an adequate training set and expertise. However, some of them are well integrated into open-source or commercial software packages.

The common drawback of the aforementioned automatic methods results from data noise. In strong heterogeneous velocity fields, the first arrivals do not necessarily correspond to the most energetic wavefields (Geoltrain and Brac, 1993). Therefore, the performance of automatic techniques of weak events can be considerably degraded by seismic noise (Tselentis et al., 2012). It is the case of refracted arrivals at far offset where its SNR ratio decreases gradually with offset. One way of addressing this problem is to filter the data before automatic picking. But problems arise when the signal arrivals and noise share the same frequency band. Many authors have reported that signal distortion due to filtering could cause picking quality deterioration. Allam et al. (2014) showed the potential effects of causal and non-causal filters on head wave picking quality (non-causal filters can introduce small artifacts ringing to the front of high-amplitude arrivals, and causal filters are a little better, but they also produce a phase shift). In addition, the troublesome effects inherent in choosing corner frequencies and slopes of a filter have been discussed by Aldersons (2004) and Senkaya and Karsh (2014).

Combining multi-methods was also proposed to increase the automatic technique accuracy. These hybrid algorithms are often addressed to pick the complicated phase and particularly S-phases (Ross and Ben-Zion, 2014; Diehl et al., 2009; Küperkoch et al., 2012). Since each method is based on identifying a specific property of that part of the trace where the first arrival occurs, the hybrid algorithm combines the advantageous properties of each individual method.

In near surface seismic tomography, a dual problem relative to the bad quality picks at far offsets impacts the refraction inversion results; keeping them in the inversion process leads to decrease the inversion results (or layer interpretation) quality. The investigation depth decreases by excluding them before inversion, while including the first arrival times with their uncertainty allows balancing the tradeoff between depth investigation and quality inversion. First arrival time as physical measurement needs an estimation about its own uncertainty (Di Stefano et al., 2006). Estimating picking uncertainties is the second most important goal in automatic picking algorithm (Aldersons, 2004). Di Stefano et al. (2006) estimated automatically the uncertainties of arrivals by using quality weighting algorithm which takes into account the waveform sampling rate, the spectral density analysis and the SNR. Diehl et al. (2009) obtained a robust estimate of error interval for automatic S-arrivals by combining picking information from different methods (STA/LTA detector, polarization detector, and autoregressive picker). Küperkoch et al. (2010) introduced an automatic quality estimation of picked P-onset from local slope and SNR of the CF used in picking procedure. Their weighting scheme calibrated with a hand-picked data leads to classify the arrival patterns into four classes (0-4).

Although using combined algorithms becomes a standard approach for picking the seismological data, it has limited use for near surface seismic data. No automatic picking algorithm can be characterized as the best for all traces in seismic records. However, when combined they can complement each other weaknesses; where one algorithm fails for traces, the other can be accurate, and vice-versa.

Here we propose a new adaptive algorithm based on the combination of three picking methods (MNW, HOS, and AIC) in order to provide robust and precise picking including time uncertainty estimation. This work focuses on overcoming the most of shortcomings inherent when using one picking method, and on providing the required information for near surface tomography (accurate P-wave arrival times and their errors). Exploiting all sub-algorithms results is a fruitful way to avoid the filtering, and makes this new scheme being an auto-adaptive and nearly parameter-free algorithm. Furthermore, since all algorithms are applying on raw single component records and since all calculations are performed in the time domain, the implemented algorithm does not require high computational resources. The efficiency of our algorithm is reported on synthetic signals which involve special challenges for automatic picking algorithms.

3.2 Methodology

3.2.1 Principe and methods

The most automatic techniques are based on the fact that each noise and signal have different properties in time or/and frequency domain. The separating point can be identified by using an algorithm that enhances and detects an attribute discontinuity linked to this transition. Our proposed strategy is based upon an adaptive algorithm by combining three picking methods: MNW, HOS, and AIC respectively (Multi-Nested Windows, Higher Order Statistics, and Akaike Information Criterion). The implemented algorithm imitates the human first-break picking, where at the beginning the global trend is defined, and then the precise first arrival times are searched in the vicinity of this trend. The picking is performed by a multistage algorithm (three successive stages) that combines the three sub-algorithms to work in series. In this section, we describe the mathematical background of three methods and their implementation in the flow of adaptive algorithm for the two cases (trace-by-trace and shot-gather picking). To better illustrate and facilitate the comparison, each trace is normalized to its maximum amplitude before any process.

Multi-Nested Windows (MNW) method

The motion of particles increases gradually when the disturbance of the seismic waves arrives, so no sudden take-off in the trace at the first break can be noticed (Ricker, 1953). Therefore, the early advent of the first arrival (here, P-wave) is usual low, and the later one would be more energetic. In near surface prospection, the first

arrivals could be covered by the noise, particularly at the far offsets due to the spherical divergence of the front wave, the attenuation and the interference with background noise (Mallinson et al., 2011).

Let $a(t)$ represent the amplitudes of trace at a time t ($t \in [1, N]$, of N samples). The average energy in the before, after, and delayed windows (respectively *BEA*, *AEA*, and *DEA*) are defined by equation (3.1):

$$\begin{aligned} BEA(t) &= \frac{1}{L_b} \int_{t-L_b}^t a^2(\tau) d\tau \\ AEA(t) &= \frac{1}{L_a} \int_t^{t+L_a} a^2(\tau) d\tau \\ DEA(t) &= \frac{1}{L_d} \int_{t+d}^{t+d+L_d} a^2(\tau) d\tau, \end{aligned} \quad (3.1)$$

where L_b, L_a and L_d are the lengths of corresponding windows in term of the dominant period (T_d) of the first arrival wave (here, P-wave) (in samples); d is the time delay for shifting the *DEA* window. Trnkoczy (2012) quantified the tradeoffs between the windows size and trigger sensitivity of the short-term average/long-term average (STA/LTA; Allen, 1978) algorithm, which is the most widely used in weak-motion seismology. It is claimed that the preceding window should be a few times longer than later window for microseismic data picking (Han et al., 2009; Wong et al., 2009). In this work, the lengths of the windows were chosen to be $L_b = 4T_d$, $L_a = T_d$, $L_d = (1-d)T_d$ and $d = 0.6T_d$. The energy ratios ER_1 and ER_2 are defined by equation (3.2).

$$\begin{aligned} ER_1(t) &= \frac{AEA(t)}{BEA(t) + \beta} \\ ER_2(t) &= \frac{DEA(t)}{BEA(t) + \beta}, \end{aligned} \quad (3.2)$$

where β is a constant used to avoid numerical instability in the computation of energy ratios. In this study, β was adopted as 0.005 being suitable for the data sets used. It is noticed that ER_1 is slightly equivalent to energy ratio ($ER=STA/LTA$) methods (e.g. Han et al., 2009). However, when ER exceeds a pre-defined threshold, the first arrival is considered as detected. But the onset picking may need further processing. Wong et al. (2009) showed that the onset of the first arrival is very close to the maximum derivative of ER . In case of noisy traces, the first-break times derived from ER technique are somewhat later (Gaci, 2014). To overcome the problems due to weakness of first arrivals, we propose a new characteristic function (CF_{mnw}) by summing the two energy ratios (ER_1, ER_2), calculated in the nested windows shown in **Figure 3.1a**, as follows.

$$CF_{mnw(t)} = ER_1(t) + ER_2(t) \quad (3.3)$$

Figure 3.1b demonstrates the ability of incorporated energy ratios to enhance the presence of the first arrival onset, where it is shown that the new CF is greater in value and has more sharpness than traditional STA/LTA at the first arrival time. Thereby, the accuracy in picking should be enhanced by comparing to traditional energy ratio methods.

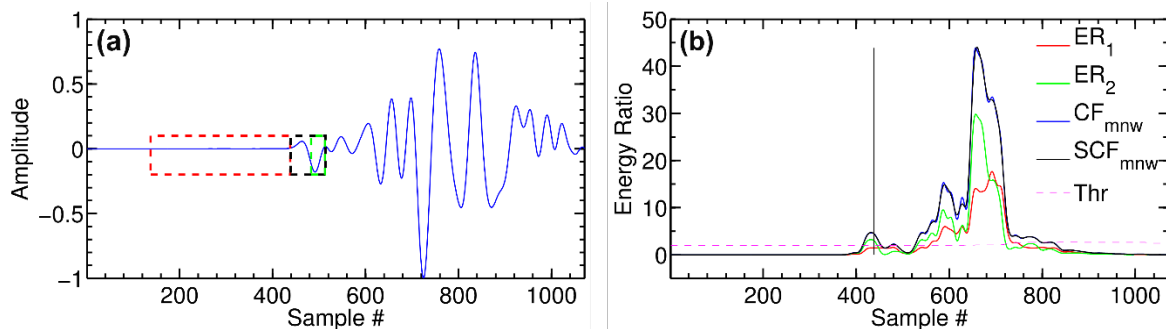


Figure 3.1. Illustration of MNW method: (a). normalized free-noisy trace and the three windows (red = before instant t ; black = just after instant t ; green = after a delay instant t) used on the calculation of CF_{mnw} . (b). Energy ratios (ER_1 = energy ratio calculated between the before window and after window; ER_2 = calculated energy ratio between delayed window and before window), CF_{mnw} = Characteristic Function of the incorporated energy ratio, and Thr = theoretical threshold.

Beside the drawbacks concerning windows size, choosing a threshold is another crucial problem of energy ratios methods (Akram and Eaton, 2012; Akram, 2014, p. 48). Assuming that the noise average doesn't depend on time, i.e. the seismic noise satisfies local stationary in short observation-scale (Ait Laasri et al., 2014). Consequently, when all windows were found before the first arrival (i.e. in noise part), ER_1 and ER_2 (given by equation (3.2) are approximately 1, resulting in a CF_{mnw} is around 2 (equation (3.3)). Based on these considerations, the theoretical threshold is given by equation (3.4):

$$Thr(t) = 2 + 3\sigma(t), \quad (3.4)$$

where $\sigma(t)$ is the standard deviation computed within the window preceding the instant t . By adding the σ term in equation (3.4), the fluctuations of most of the noises are avoided, making this threshold slightly time dependent. However, when the later and delayed windows are occupied by a part of the seismic phases, the CF_{mnw} becomes greater than 2. The beginning of the potential zone (BPZ) (P-wavetrain) is determined once the CF_{mnw} exceeds $thr(t)$. In order to minimize spike fluctuations on the picking procedure, a smoothing process of CF_{mnw} was performed via local regression tools (using for example *loess MatLab function*) within a window of half dominant period ($0.5T_d$). This function and its parameter have been chosen according to the noise trace situation. Since CF_{mnw} characterizes the waveform change of the P-wavetrain, the potential first arrival times can be allocated to the two first local maxima of the smoothed CF_{mnw} (SCF_{mnw}) after BPZ within a predefined searching window. Therefore,

one of the two initial picks, which corresponds to the maximum SNR, represents the first break (tP_1), as shown in **Figure 3.1**. The approximated interval error (tE_1) can be assessed by $tE_1 = \max(e'_1, e''_1)$ where e'_1 is the absolute difference between BPZ and the first local maximum, and e''_1 is the absolute difference between the two maxima if they exist.

The results of MNW method can be summarized by considering three parameters: the early begging of the potential zone (BPZ), the potential first arrival time (tP_1), and the approximated interval error (tE_1).

Higher Order Statistics (HOS) method

The seismic noise follows closely a Gaussian probability distribution (Ait Laasri et al., 2014). In contrast, the seismic signal is non-Gaussian (Persson, 2003; Saragiotis et al., 2002). Higher moments, or statistics (as skewness and kurtosis which are normalized central moments of order three and four, respectively) are well suited to check on the gaussianity distribution of time series (Giannakis and Tsatsanis, 1994). Moving skewness and kurtosis reflect (in different degrees) the changing in distribution shape around a chosen position of the time series (Press et al., 1992). However, the ability of HOS to pick first arrivals was successfully tested by Saragiotis et al. (2002) on real seismic applications. Furthermore, the superior efficiency of kurtosis over other metrics (such as skewness) to detect and pick seismic phases has been shown by many authors (e.g., Küperkoch, 2010; Nippres et al., 2010; Tselentis et al., 2011).

In this study, the fourth order central moments (kurtosis attribute) was used to characterize the amplitude distribution. The CF based on kurtosis (CF_k) is calculated using equation (3.5) with n_k -samples rolling window over the seismic trace.

$$CF_k(t) = \frac{1}{n_k} \sum_{i=t-n_k+1}^t \left(\frac{a_i - \bar{a}_t}{\sigma_t} \right)^4, \quad (3.5)$$

where $CF_k(t)$ is the kurtosis values, a_i represents a vector of seismic amplitudes inside a window of size n_k , \bar{a}_t and σ_t are respectively the mean and the standard deviation over n_k samples. The kurtosis function value for each window were ascribed to the end sample of the window, as shown in **Figure 3.2a**. When the moving window integrates the first part of the seismic phase, the CF_k increases due to the transition from Gaussian to non-Gaussian behavior, as shown in **Figure 3.2b**. In case of complicated or emergent onsets, the kurtosis increases progressively before reaching its maximum value. The onset time, which corresponds to the very beginning of this increase, matches well the human-operator picking (Baillard et al., 2014). The onset time of the considered phase can be determined using the maximum CF_k derivative (Langet et al., 2014; Lois et al., 2013). For this purpose, the succession of transformations proposed by Baillard et al. (2014) was applied to CF_k in order to isolate the P-phase onset. Thus, the first arrival time (tP_2) is defined by the minimum of the smoothed Baillard function (SCF_{Bk}) (**Figure 3.2b**). When the non-impulsive arrivals are submerged by the noise,

the kurtosis local maxima identifying this arrivals shift up in time. It was assumed that the approximated error (tE_2) of this method can be estimated by the time interval between the maximum of the CF_k function and the picked onset. For the clear and impulsive phases this time interval is very small, while in case of noisy and non-impulsive phases this interval becomes significantly larger. In addition, to mitigate the adverse effects of rapid fluctuations of CF_k on the automatic picking procedure, a low pass filtering is performed to smooth CF_k . The same function and parameters used in previous method were applied in this method.

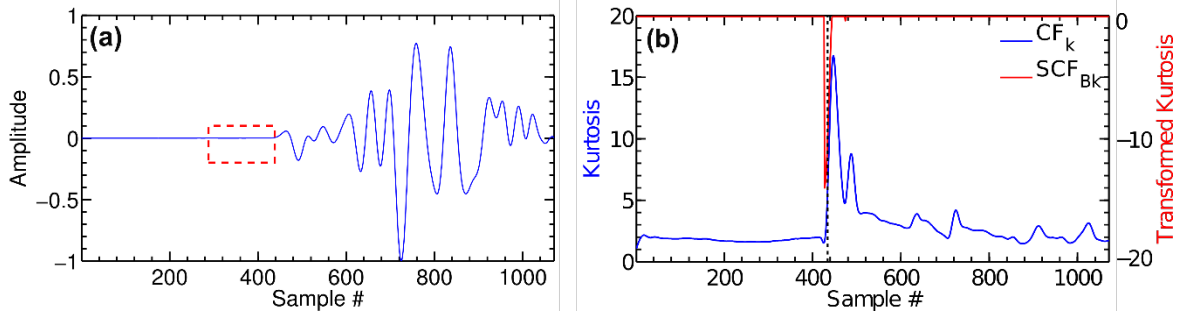


Figure 3.2. Illustration of HOS based kurtosis method: (a). normalized free-noisy trace and windows (red rectangle) used for calculating CF_k . (b). the resulting kurtosis CF (blue solid line), and transformed kurtosis based on Baillard procedure (red solid line). The black dashed line indicates the manual pick of first arrival of P-wave.

The results of HOS method can be resumed in two parameters: the potential first arrival time (tP_2) and the time sensitivity range (tE_2).

Akaike Information Criterion (AIC) method

AIC (Akaike, 1974), which points out the degree of the reliability of the model fitting, is widely used in the autoregressive (AR) techniques (e.g. Leonard and Kennett, 1999; Leonard, 2000; Takanami and Kitagawa, 1988). These approaches are based on the assumption that the seismic trace can be split into locally stationary segments, where each of them is satisfying a different autoregressive model. Hence, the point where the AIC is minimized indicates the optimal separation of two time series (noise and signal plus noise) and this is interpreted as the first break arrival (Sleeman and van Eck, 1999). In this study, instead of using AR coefficients for calculating AIC function, the CF based AIC (CF_{aic}) is calculated via Maeda's version (1985). Let $a(t)$ a seismic trace of length N , the CF_{aic} is computed by equation (3.6) directly from amplitudes sequences splitting in two parts, as shown in Figure 3.3a.

$$CF_{aic}(t) = t \log(\sigma^2(a)_{1\sim t}) + (N - t + 1) \log(\sigma^2(a)_{t+1\sim N}), \quad (3.6)$$

where $\sigma^2(a)_{1\sim t}$ and $\sigma^2(a)_{t+1\sim N}$ are the variance of a in the two segments.

Under ideal conditions, the point which is associated to the minimum of the CF_{aic} is searched for, and its corresponding time is interpreted as first arrival time. However, “with real field data the life seems be harder”, in certain circumstances that is not

fulfilled due to interference the noise and other later wavefields. Furthermore, the seismic trace contains of many significant features and the AIC picker defines the arrival of the stronger phase as a global minimum. It is necessary to apply AIC picker on time limited window (*searching window*) containing only the interesting phase (Zhang et al., 2003). Taking into account these points for improving the accuracy of AIC picking, we implemented the multi-model inference approach, which was developed to pick the transmitted ultrasound signals for clinical tomography device (Li et al., 2009). This approach incorporates the entire set of models in the searching window to compute a weighted average model, based on normalized Akaike weights as following:

- Define the minimum of the CF_{aic} in the searching window ($CF_{aic_{min}}$).
- Calculate the differences (Δ_i) between CF_{aic} values and $CF_{aic_{min}}$:

$$\Delta_i = CF_{aic_i} - CF_{aic_{min}}, i = 1, \dots, N \quad (3.7)$$

- Compute the normalized Akaike weights (AW):

$$AW_i = \frac{\exp(-\Delta_i/2)}{\sum_{r=1}^N \exp(-\Delta_r/2)} \quad (3.8)$$

- Set all values of (AW) out the searching window to zero.
- The potential first break time (tP_3) is obtained by using the weighted average:

$$tP_3 = \sum_{i=1}^N AW_i t_i, \quad (3.9)$$

where t_i is the time corresponding to each sample in the time series.

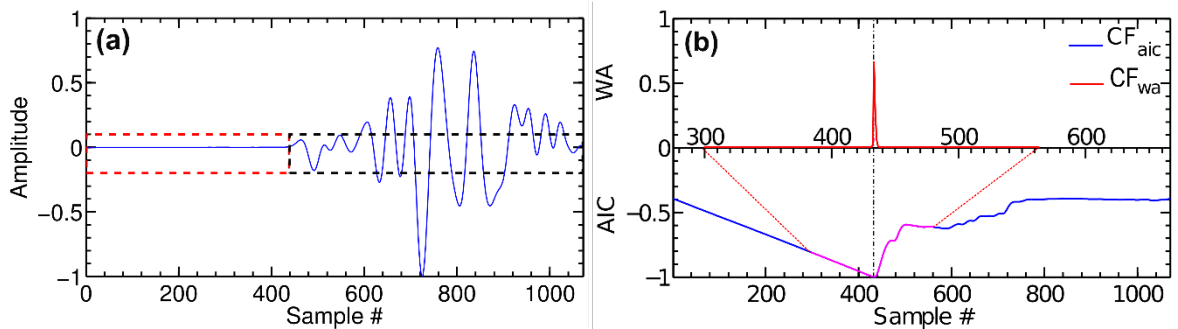


Figure 3.3. Illustration of AIC picker based upon Akaike Weights: **(a)**. normalized free-noisy trace and two variable length windows before and after (red and black dashed rectangles, respectively) **(b)**. The CFs based on AIC (blue solid line) and on AW (red solid line). The part of CF_{aic} used in calculation CF_{wa} is highlighted by purple line. The black vertical dashed line represents the manual pick of the first arrival.

For traces showing high SNR , the difference between the improved AIC picker based on weighted model averaging scheme and ordinary AIC picker based on the best model (corresponding to the minimum AIC value) is negligible (**Figure 3.3b**). On the other

hand, when the traces show low SNR (where the global minimum of AIC function is not very sharp), the new AIC picker is more accurate since it integrates the other candidate models based on its relative importance estimated by Akaike weights (Li et al., 2009).

The onset clarity reflects usually itself by the sharpness of the AIC minimum (Diehl et al., 2009; Toomey et al., 1994). A prominent AIC minimum is typically produced by an impulsive event, whereas the broader local minima are related to non-impulsive or noisy events. We estimate the time uncertainty of AIC picker (tE_3) from normalized Akaike weights function within searching window. The lower and upper limit of error interval correspond with, respectively, to the first and last sample where AW function (equation (3.8)) is greater than an appropriate threshold, which can be derived empirically from AW maximum. We found that a threshold of 10% of the maximum AW function, gives satisfying estimation of the uncertainty interval.

The results of AIC method can be summarized also by two parameters: the potential first-break of P-wave (tP_3), and the error range (tE_3).

Finally, in free-noisy traces one method is enough to pick correctly the first arrivals. Additionally, applying independently many algorithms should yield approximately the same results. Whereas in low SNR data, each method has its own CF that characterizes a specific trace property. Then, their results may be significantly different. In such cases, using multi-method algorithm helps to fill gaps in each used methods while maintaining good performance, and further, to guide picking algorithm toward the interesting zone.

3.2.2 Algorithms implementation

The main idea of implemented strategy is that the integration of multi-algorithms should enhance the capability of picking the weak events and reducing error picking for different noise levels. In the proposed strategy, the picking implementation uses a multistage algorithm (three successive phases), wherein each phase has its own CF. The information provided by each sub-algorithm are used either to orient the following phase picking (define target window), or to setup the essential parameters in the next phase, or the both. The integration in series of each phase results in the implemented flow provides auto adaptive parameters estimation. Therefore, one *a priori* parameter is needed only for the first phase picking (in MNW). Subsequently, all potential picks derived from sub-algorithms were combined by statistical tools to provide the first arrival times and their relative errors. In the following subsections we describe the flow chart of the implemented picking algorithm and its parameters for a signal trace (trace-by-trace algorithm) and for a set of traces (entire-shot gather algorithm).

3.2.2.1 Single-trace algorithm picking

A vertical component of single trace, extracted from a record simulated by finite differences modeling, was chosen to illustrate the related topics in internal stages of the algorithm picking (input-output parameters and its own CFs). A significant level noise (random and bandlimited noise) is added to selected trace, as shown in **Figure 3.4a**. This trace presents a double serious challenge to be picked automatically since the first arrival has weak onset and is embedded completely by the add noise. The first arrival was picked manually on clean trace before adding the noises and marked by black vertical solid line (**Figure 3.4a**).

Phase 1 (**MNW sub-algorithm**): the main goal of this phase is to enhance the presence of weak event and to determine the very beginning of the considering seismic phase. The CF characterizes the variations of energy as a function of time. The first period length of the P-wave (T_d) is the only *a priori* parameter which needs fine-tune by user. However, it can be obtained directly from the input data trace. For this testing trace, we estimated T_d to 75 samples. The picking procedures, as described previously, are applied on smoothed CF (SCF_{mnw}). The searching window (SW1) is located just after the beginning of potential zone (BPZ). We find its length equals to one and half period ($1.5T_d$) is suited to pick the two local minima for which one of them is potentially the first arrival (tP_1). **Figure 3.4b** shows the SCF_{mnw} and the related results.

Phase 2 (**HOS sub-algorithm**): the size of the rolling window (N_k) used in calculating the CF based kurtosis (CF_K) can be estimated automatically from the approximated sensitivity interval of the previous sub-algorithm (tE_1). In the most tested traces we found N_k of twice the (tE_1) provides satisfactory results. In order to avoid the under- or over-estimating of CF_K , we added a security condition in which when estimated N_k less than $0.5T_d$ or greater than $2T_d$, the algorithm setup the N_k to one period (T_d). The searching window ($SW2 = tE_1 + T_d$) is setup automatically around the potential first arrival derived from MNW sub-algorithm (tP_1); tE_1 before tP_1 and T_d after. The potential first arrival (tP_2) and associated time uncertainty (tE_2) are also determined according to the descriptions in the related section (HOS method). The related CFs and results of this sub-algorithm are illustrated on the **Figure 3.4c**.

It is worth mentioning that it is sufficient to compute the kurtosis attribute only for the part of the trace corresponding to the searching window. Thus, that makes the process picking speeding up.

Phase 3 (**AIC sub-algorithm**): Since the performance of AIC methods depends strongly upon the searching window configuration, the results of previous sub-algorithms (MNW and HOS based kurtosis) were jointly used to lead the AIC picker, and further to setup its search window size. We define the size of this window and its location (respectively, SW_3 and C_{aic}) according to equation (3.10).

$$\begin{aligned} SW_3 &= 2 \max(tE_1, tE_2) \\ C_{aic} &= \text{mean}(tP_1, tP_2) \end{aligned} \quad (3.10)$$

Finally, the potential first arrival (tP_3) and its time uncertainty (tE_3) are also obtained according to the related explications in AIC method section. The related attribute functions and the results are illustrated on the **Figure 3.4d**.

All potential first arrival onsets and their associated errors picked by the sub-algorithms are shown on the **Figure 3.4e**. The input/outputs parameters of each sub-algorithm can be summarized in the following simplified flows:

$$\begin{aligned} T_d \text{ (predefined by user)} &\xrightarrow{\text{MNW sub-algo}} \text{BPZ, } tP_1, \text{ and } tE_1 \\ tP_1 \text{ and } tE_1 \text{ (from MNW sub-algo)} &\xrightarrow{\text{HOS sub-algo}} tP_2 \text{ and } tE_2 \\ SW_3 \text{ and } C_{aic} \text{ (from MNW sub-algo \& HOS sub-algo)} &\xrightarrow{\text{HOS AIC-algo}} tP_3 \text{ and } tE_3 \end{aligned}$$

Final picking: the three sub-algorithms results can be accounted for obtaining the final first-break time of P-wave. In analogy to the multi-window and weighting scheme proposed by Lois et al. (2013), the final pick time (tP_{final}) is obtained by the weighted mean of the picks set produced by the previous pickers ($tP_i; i = 1,2,3$) according to

$$tP_{final} = \frac{\sum_{i=1}^3 Q_i tP_i}{\sum_{i=1}^3 Q_i}, \quad (3.11)$$

where Q_i is a quality factor of the picked phase calculated into a predefined time windows and is given in dB (Küperkoch et al., 2010):

$$Q = 20 \log\left(\frac{A_s}{A_n}\right), \quad (3.12)$$

where A_n is the RMS-amplitude of the noise calculated within a time window of length ($nois_wind = 3T_d$) before the determined tP_i . The A_s is the RMS-amplitude of the P-wave evaluated on the signal window ($sig_wind = 1T_d$) starting immediately after tP_i .

However, Q values can be considered as an accepted automatic estimator of quality pick. Where if the pick is near to correct one, Q estimator has high and positive value. On the other hand, indeed Q estimator having value close to zero or negative indicates, respectively, very early picking or late picking. In final picking procedure, all picks corresponding with negative Q values were excluded out the final picking process (*i.e.* before the weighting means approach).

3.2.2.2 Error and quality assessment

While an ultimate goal of any automatic picker is to realize the highest accuracy possible, an associated uncertainty estimation is of great importance in *a posteriori* processing of first arrival time (*e.g.* in tomographic inversion). For these purposes, we suggest an approximated estimation of final pick error by using a simple statistical tool. From a practical and physical point view, the standard deviation can be

considered a good error estimator of repeated measurements of a physical parameter (here, the arrival time). Therefore, the final time uncertainty (tE_{final}) is produced by the standard deviation of three potential picks resulting from the aforementioned sub-algorithms.

$$tE_{final} = std(tP_1, tP_2, tP_3) \quad (3.13)$$

Additionally, the Q-factor calculated by equation (3.12) for the final pick could be another qualitative estimator for determining pick quality. Since the first arrival onset is completely embedded by the random noise, the final pick is slightly shifted (**Figure 3.4e**). Therefore, that demonstrates the high sensibility of the three picking methods to the higher amplitudes. Although when the added noise amount is considerable, the correct first arrival time is still in the range error of the implement algorithm (**Figure 3.4e**), with a difference between theoretical and picked arrival time about 25 samples (i.e. 12.5 ms for sampling rate 0.5 ms).

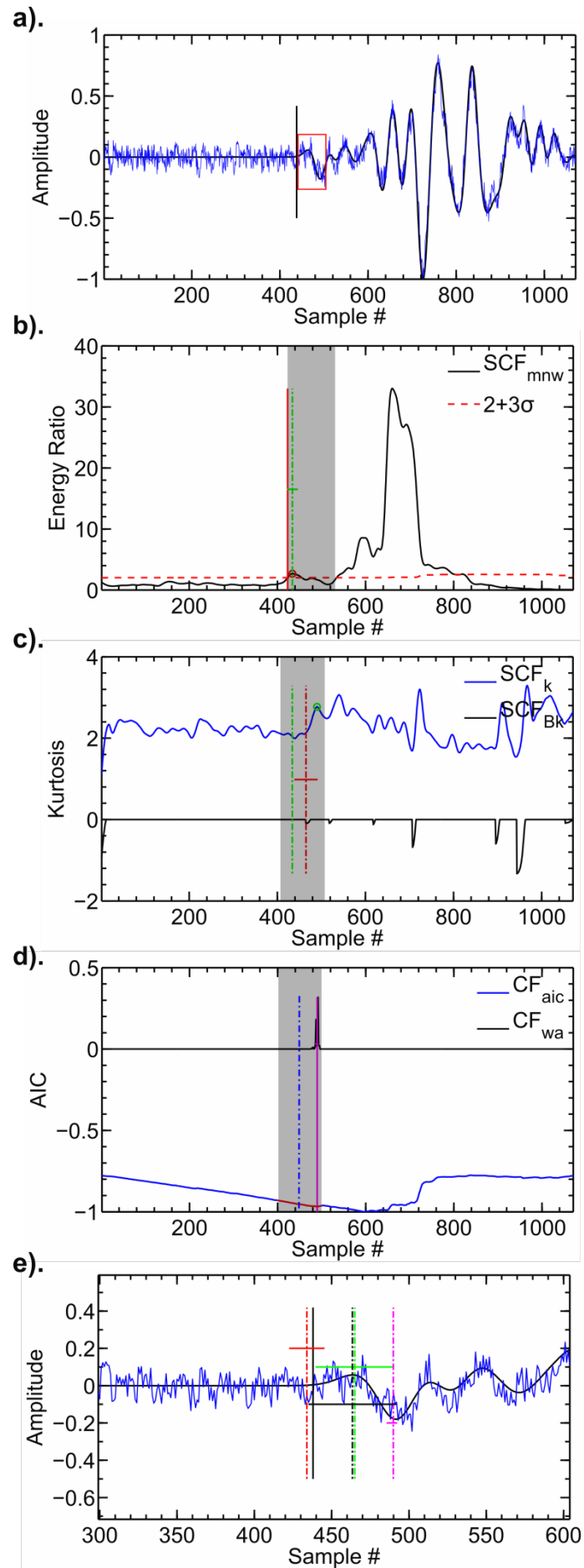


Figure 3.4. Illustration of the Single-trace algorithm applied to a synthetic trace: **(a)**. Noise-free synthetic trace (black), after add noise (blue), manual first arrival (vertical black line). The first period of P-wave is framed by the red rectangle **(b)**. MNW sub-algorithm topics: threshold (dashed red line), BPZ (vertical red line), picked potential first arrival and the associated error (respectively, vertical pointed and horizontal green line). **(c)**. HOS sub-algorithm topics: the guide picker (vertical pointed green line), maximum of kurtosis (small green circle), picked potential first arrival and associated error (respectively, vertical pointed and horizontal red line). **(d)**. AIC sub-algorithm topics: the picker guide (vertical pointed blue line), the red part of CF_{aic} is considered for calculating CF_{wa} , the picked potential first arrival and associated error (respectively, vertical pointed and horizontal red purple). The gray zones in b, c, and d represent the searching windows SW1, SW2, and SW3, respectively. **(e)**. Zoom in of the trace in (a) around the first arrival (vertical solid line). The final pick and its error are marked by vertical pointed and horizontal solid lines, respectively. The potential first arrivals and their associated errors picked by the three sub-algorithms are re-delineated on the noisy trace.

A disadvantage of this algorithm is that when the MNW fails to define the potential zone of the P-wavetrain, the others sub-algorithms neither succeed for following it (because they are inter-connected together by auto adaptive parameters). As a result, this may lead to false picking.

High-fidelity filtering (e.g. [Aldersons, 2004](#)) or denoising (e.g. [Parolai, 2009](#); [Tselentis et al., 2012](#); [Gaci, 2014](#)) could reduce these effects improving the accuracy picking. In near surface seismic surveys, exploitation of the redundant data and their events coherency is another solution of such related problems, as we proposed and described in the next section.

3.2.3 Entire-shot record algorithm picking

The basic idea of this algorithm is to integrate all picks provided from preceding sub-algorithms by exploitation the properties of coherency of the first arrivals and its consistency over all offsets. This algorithm is an extension of the single-trace algorithm, and it is adapted for full automatically picking of shot-gather records. Two main additions, which are addressed to correctly and under different noise conditions determine the trend (BPZ) of P-wavetrians, were added in MNW sub-algorithm.

Using a fixed MNW-threshold might reduce the accuracy of picking the first arrivals mainly caused by high variability of the noise and the P-waveforms. In order to elevate this drawback, we proposed a variable-threshold engine combined with an iterative process. The latter is a look-like Monte Carlo approach. Through this process, an optimal solution is built from the database picks producing by the MNW sub-algorithm via variable-threshold. Here, we call the solution is the set of picks (BPZ picks) that build the trend of the first-breaks over the whole shot gather. On other words, this

routine consists of two subroutines. In the first, a matrix of all possible solutions is generated by multi threshold engine picking. The second one is an iterative system consisting of : (1) randomly building a solution, (2) calculating the corresponding cost function, (3) repeating the process for a predetermined number of iterations, (4) choosing the best solution that is corresponding to maximum cost function. The latter combines three main properties (or cost functions) that bring to light the presence of the first arrivals of shot gather:

Energy cost function ($Cost_{En}$): we defined $Cost_{En}$ (equation (3.14)) as a function of the after energy ratio (AER) that is the average of the CF_{mnw} within limited time window just after the built solution. The AER was modulated by quality factor (Q-factor; estimated according to equation (3.12)) of the related picks, and normalized by an error ($Error_{ER}$). We assumed that error is given by the standard deviation of the solutions matrix over threshold dimension (Std_{gr}).

$$Cost_{En} = \sum \left(\frac{AER \times Q}{2 Error_{ER}} \right)^2 \quad (3.14)$$

$$Error_{ER} = Std_{gr}$$

This cost function has to maximize over the whole shot record.

Smooth cost function ($Cost_{smooth}$): By definition, a curve is smooth when all the derivatives exist and are continuous. The simplicity and efficiency of centered 2-order derivative ($deriv2$) make it an attractive candidate for analyzing smoothness curve. The second order derivative of the travel time shape has to minimize (*i.e.* the solution is hold as close as possible to a smooth curve). The shape of the trend is normalized by an attributed error ($Error_{smooth}$), which can be defined as the standard deviation of the selected solution (Std_{sol}). The $Cost_{smooth}$ is then given through the equation (3.15):

$$Cost_{smooth} = \sum \left(\frac{|deriv2|}{2 Error_{smooth}} \right)^2 \quad (3.15)$$

$$Error_{smooth} = Std_{sol}$$

SNR cost function ($Cost_{snr}$): The SNR of raw traces, calculated in predetermined windows around the built solution, has to be maximized. The $Cost_{snr}$ is defined according the equation (3.16).

$$Cost_{snr} = \sum \left(\frac{Snr}{2 Error_{snr}} \right)^2 \quad (3.16)$$

$$Error_{snr} = Std_{gr} + Std_{sol}$$

where $Error_{snr}$ is the associated error of SNR function.

The total cost function is the sum of the normalized costs functions, and can be expressed as follows:

$$Cost_{total} = \alpha \cdot Cost_{En} + \frac{\beta}{Cost_{smooth}} + \gamma \cdot Cost_{snr} \quad (3.17)$$

where α, β and γ are weights between $[0, 1]$, which allow controlling the importance of individual objective functions. In this study, we give an equal contribution of the three cost functions ($\alpha = \beta = \gamma = 1$) for converting the system toward the best solution. The trend of the first arrivals is then to find the optimal solution (sol^*) which maximize the total cost function ($Cost_{total}$). The problem can be simply formulated as,

$$sol^* = \underset{sol}{arg \max} Cost_{total} \quad (3.18)$$

The general flow of our iterative system is somewhat similar to Monte Carlo approach (randomizing selection and then minimizing a likelihood function). In our approach, the random selection is limited or somehow guided, and the likelihood function is absent (i.e. full directly searching).

Taking in consideration that the first arrivals are consistent and coherent over all offsets, its trend should be as smoothed as possible. In order to eliminate outliers and high first order derivative points, the final solution should be passed through a smoothing process. We used *rloess matlab* function, which is based on local regression using linear least squares, where it assigns lower weight to outliers in regression. The smoothing window length is chosen depending on the complexity of subsurface and on the expected form of travel time curve (in percent of total traces number). In present work, smoothing window of 50% of traces number is used for the synthetic data.

In very noisy case, the correct first arrivals could be found in the shadow of strong ambient noise. Hence, they might be blinded for the variable-threshold engine. In such case, it is desirable to rebuild the solutions matrix, but within a tolerance window centered on the smoothed solution. The tolerance window length of two time dominant period ($2T_d$) is suitable for this purpose. The new candidates' matrix are processed again though the iterative system, and the final smoothed solution can then be interpreted as the trend of the potential zone of the P-wavetrain arrivals.

As a whole, the rest of the algorithm is the same flow of single-trace algorithm. In particular, some modifications to the input parameters of each sub-algorithm (especially the searching windows size) were adjusted by using statistical tools. The abundance of each sub-algorithms outputs permits also to set forwardly the SW location for the next sub-algorithm (i.e. guiding mode).

Since the trend (BPZ) might be delayed due to use of multi-threshold and/or smoothing process in MWN sub-algorithm, it is appropriate to add a confidence interval before the estimating smoothed trend. Therefore, the P-wavetrain zone is located around the smoothed trend rather than just after it (like as in single-trace algorithm). By conserving the SW1 size (i.e. $1.5T_d$ like as in single-trace algorithm), shifting the BPZ up to the half of the dominated period ($0.5T_d$) before the smoothed trend is adequate for avoiding late phases picking.

The parameters setting of the other sub-algorithms are recapitulated in the **Table 3.1** that presents a comparative summary of the parameters algorithms considered here.

Table 3.1 : Comparative summary of the modification between the single-trace and entire-shot algorithms parameters (~: no parameters, ==: the same parameters).

	Single-trace Algorithm			Entire-shot Algorithm		
	Sub-algorithms			Sub-algorithms		
Parameters	MNW	HOS	AIC	MNW	HOS	AIC
Calculating windows	$L_b = 4T_d$ $L_a = 1T_d$ $L_d = (1-d)T_d$ $d = 0.6T_d$	$[0.5-2]T_d$ $2tE_1$	~	==	$[0.5-2] T_d$ $2\text{median}(tE_1)$	~
Guide	BPZ	tP_1	$\text{mean}(tP_1, tP_2)$	Trend	==	==
Searching window	$1.5T_d$	$tE_1 + T_d$	$2\max(tE_1, tE_2)$	$1.5T_d$	$\max(tE_1)$	$w1=\text{median}(tE_1)$ $w2=\text{median}(tE_2)$ $2\max(w1,w2)$
Location	After	Around	Around	Around	Around	Around

3.3 Testing and results

In this chapter, only synthetic data are considered to assess the performance of the implemented algorithm. These data show many serious challenges for the first arrivals picking, especially for an automatic picker.

3.3.1 Test on synthetic data

Synthetic data sets were produced by finite-difference (FD) modelling in the time-space domain using *FDELMODC* program developed by Thorbecke and Draganov (2011). The geometry of acquisition system consists of an array of 72 receivers with a spacing of 4 m in flat topography. The seismic sources are simulated by an explosive Ricker wavelet with a maximum-frequency of 90Hz, spreading at each 4 m between the receivers. The geological model used to create wavefield data is assumed to be an elastic media incorporating realistic subsurface structures. This model is composed of many layers (C1-C3) and a substratum (S) fractured by a normal fault (F); three layers at the left side of the fault and two layers on the right one, as shown in **Figure 3.5a**. All velocities and densities are assumed to be constant in each layer, except the fault zone with a linear velocity gradient about 12.5 m/s per m, with $v_0 = 625$ m/s. Densities of each layer (ρ) are set using the Gardner's rule for P-wave velocities ($\rho = 0.31V_p^{0.25}$, V_p in m/s and ρ in g/cm³). The details about geometrical parameters model are shown in **Figure 3.5a** and **Table 3.2**. Since the FD is one of the direct methods, which require to grid the model into a finite number of cells (Carcione et al., 2002), the model (500x150 m) was discretized into 1000x300 grids (i.e. the space is equally sampled in x and z at 0.5 m). The Perfected Matched Layer (PML) absorbing boundary with 120 grid points was implemented at three sides, with a free surface. We generate many inline shot gathers; in each one, a total of 0.6 s with a time increment of 0.5 ms was modelled for a vertical point force located at the free surface. **Figure 3.5b** shows an example end-on spread shot-gather (the source at 198 m) of the vertical

component of the wavefield simulated, in which the theoretical P-wave arrival times are denoted in small green circles.

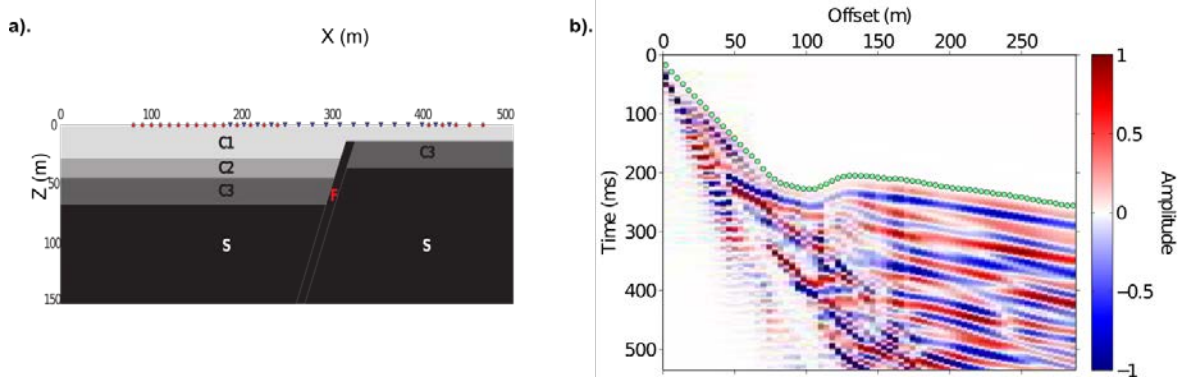


Figure 3.5. Numerical simulation: (a). Schematic view of the multi-layer model (layers C1-C3, substratum S, and F page of normal fault) used for generating the synthetic data. The positions of the sources and the receivers are highlighted by red stars and blue triangles, respectively. The thicknesses and velocities are given in the **Table 3.2**. (b). Normalized vertical component of the wavefield simulated by a source located at 198 m from the left side of the model. The small green circles represent the manual picks.

Table 3.2 : Thickness and velocities of the layers model used in the simulation

Layer	Thickness (m)		Velocity P-wave V_p (m/s)	Velocity S-wave V_s (m/s)
	Left	Right		
C1	28.5	14	400	230
C2	16.5	0	1000	580
C3	22.5	22.5	1800	1050
F	10.5		625 - 2500	365 - 1500
S	∞		2500	1500

In addition to the accumulated difficulties inherent of weak events, complex changes in the waveforms, and nonlinear travel-time curve due to the presence of fault, we added another classic challenge represented by the noise adding to the simulated wavefields. Two types of added noise (random zero-mean Gaussian and band-limited noises) were computed using the function *suaddnoise* of the Seismic Unix (SU) package (Cohen and Stockwell, 2012). The band-limited noise has approximately the double frequency range of the simulated source (Ricker wavelet).

We applied the Entire-shot picking algorithm, which is described previously, to pick the first arrival times in the shot gather shown in **Figure 3.5b**. The first dominant period of the P-waves is approximated to 75 samples. The automatic picks resulting from adaptive algorithm are compared to the theoretical picks calculated from the model shown in **Figure 3.5a**. The automatic process was repeated for the same shot-gather in **Figure 3.6a** but with additive random Gaussian noise 20%, limited-band noise with [40-180] Hz, and the two previous noises, respectively in **Figure 3.6b**, **Figure 3.6c**, and **Figure 3.6d**. Despite the significant added noises, the picks resulting from automatic algorithm have a good consistence with the theoretical picks.

The problem of subsurface structures complexity of and variation of the waveform of the first arrivals has completely overcome. Although automatic picking performed satisfactorily overall, an over-estimating of the arrival times is noticed after adding the noise. This shift-down of the picks is increasing with the offset and it is remarkable at the events corresponding to the fault, as expected due to their weak onsets.

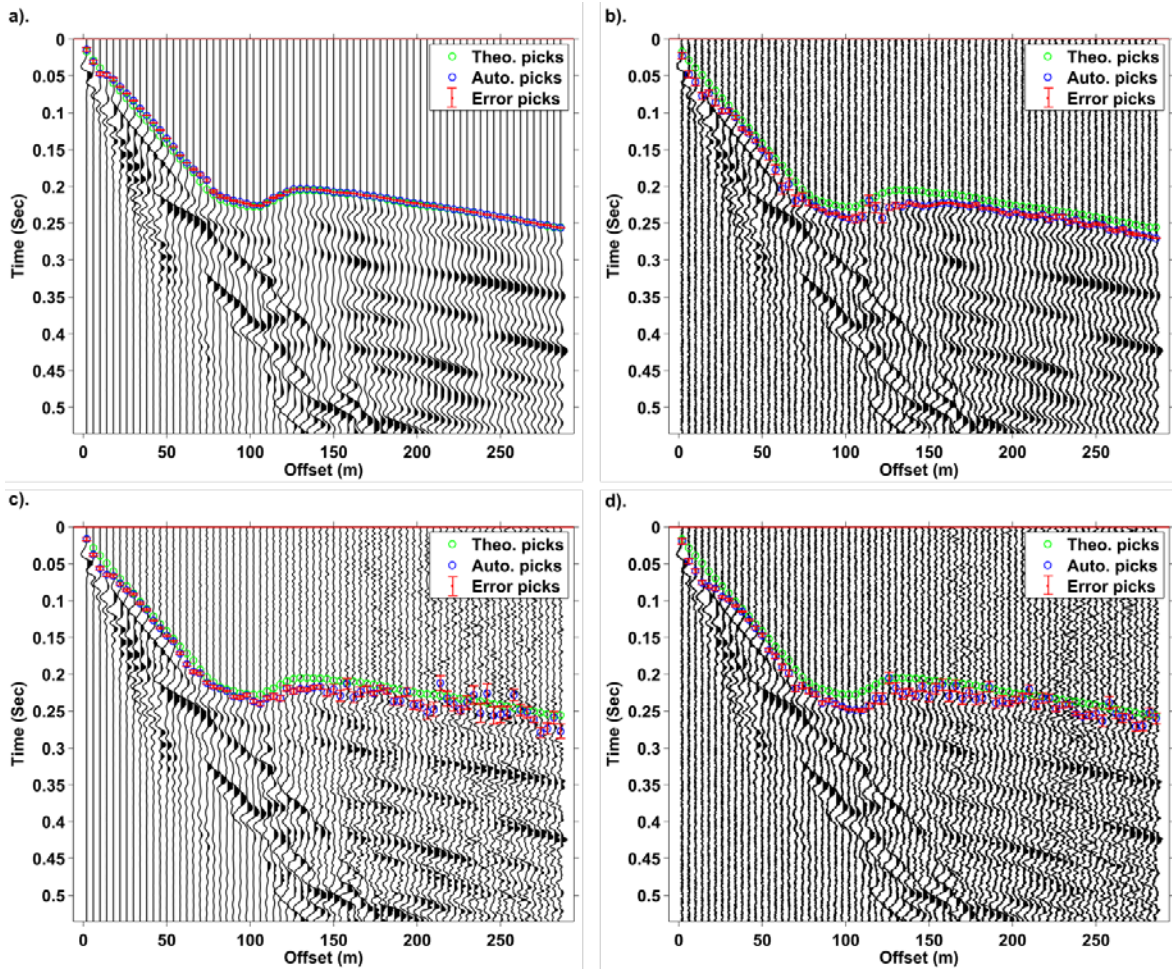


Figure 3.6. A synthetic vertical components record with the first arrivals times obtained theoretically (in green) and automatically with their interval errors using the proposed adaptive algorithm (in blue and the error barre in red): (a).Noise-free shot-gather. (b). With 20% random white noise. (c). With random band-limited noise [40-180] Hz. (d). With the noise in (a) and in (b) (random and band-limited noise).

To assess the effectiveness of the automatic adaptive algorithm, we compare the picks with those that are determined theoretically for two cases: noise-free and noisy synthetic records. In the case of clear shot, sixty percent of the automatic picks are within -2 ms of the actual picks, and 90% are within -10 ms (**Figure 3.7a**). This analysis can be considered as the calibration step of the algorithm parameter setting. Despite the small deviation from the actual picks, the median of differences between the automatic and the actual picks is still about -1 ms, and the second and third quartile (50%) of the differences are included in the [0, -4] ms interval (**Figure 3.7c**).

A possible explanation of this deviation is the first dominant period variation of the P-wave with the offset. Therefore, this statistic result analysis is validated parameters choosing of our algorithm. For noisy data (random and band-limited noises), about 35% of the differences are found at 10 ms, and 75% are within [0, 20] ms interval. **Figure 3.7b** shows a statistical analysis in each quality class, which are determined in analogy to the empirical uncertainty categories (A-D) of [Lois et al. \(2013\)](#) classified according the SNR (Q ; equation (3.12)) (**Figure 3.8**), through the boxplot graphs. Clearly, all distributions of the four classes are found on the positive side of the difference between automatic and actual picks. Visual inspections of individual sub-algorithm results and corresponding waveform shows that delay time picking is mainly due to their high sensibility to the high amplitudes in presence of high level noise. The dispersion symmetry around the median and the absence of outliers shows again good consistence of the two picking types.

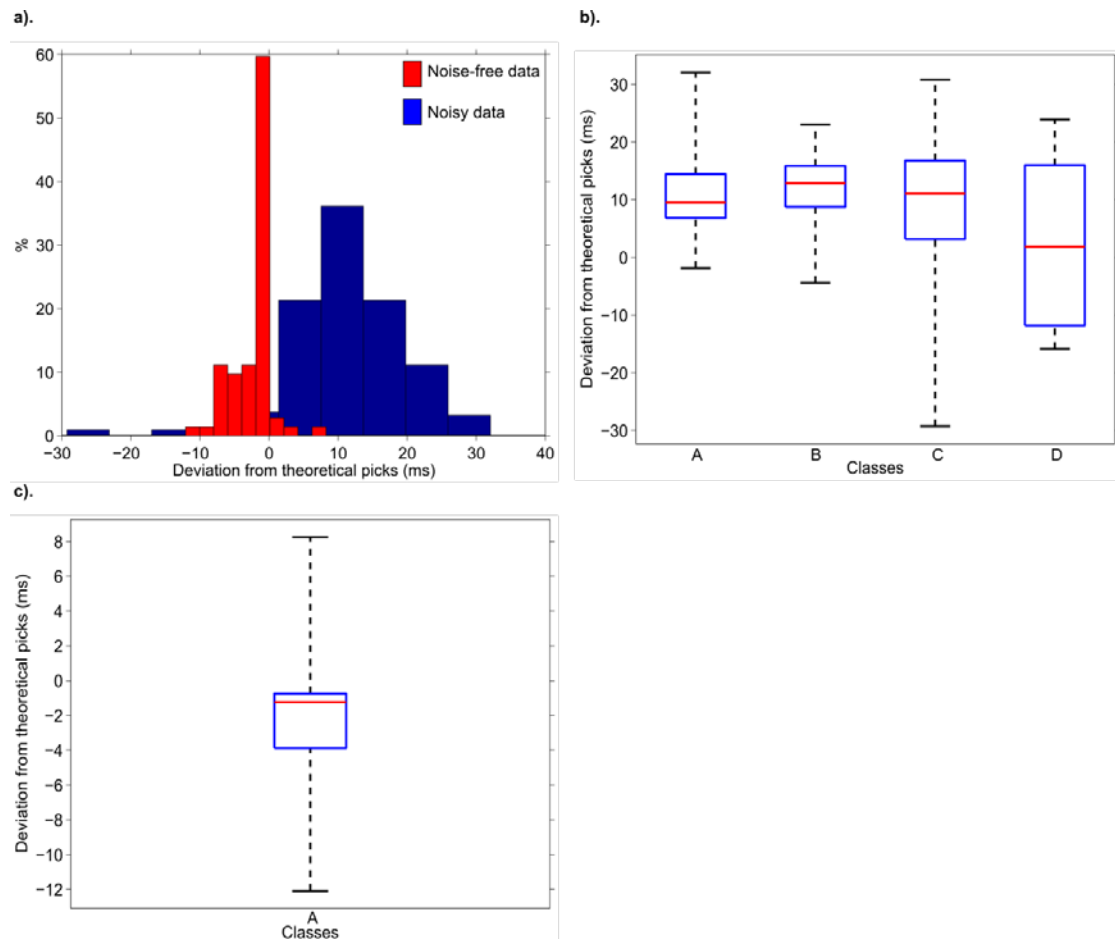


Figure 3.7. Statistical analysis: **(a)**. Histograms of differences between automatic and actual picks for noise-free data (in red) and noisy data (blue). **(b)**. Boxplot graphs of the differences for noisy data within the quality classes (A, B, C, and D). **(c)**. the same but for noise-free data.

The quality assessment of automatic picks is shown in **Figure 3.8**. By adding the noise the first arrivals quality was degraded by about 1/5 of their level at the noise-free

gather (i.e. from 50 dB to less than 10 dB, as shown in **Figure 3.8**; see red pointed curves evaluation comparing to the red solid ones). The accuracy of our adaptive algorithm versus the quality of the arrivals is presented in the tables including in the **Figure 3.8**. The RMS of the differences in case of zero noise is 3.6 ms, and the maximum RMS in noisy cases is 16 ms which is normally corresponding with C and D quality classes. About 50% of the noisy traces are related to the B and C classes. However, in spite of the decreasing of quality arrival onsets due to additive noise, the implemented algorithm kept a good performance, where the maximum RMS in the very noisy case did not exceed the acceptable error (20 ms).

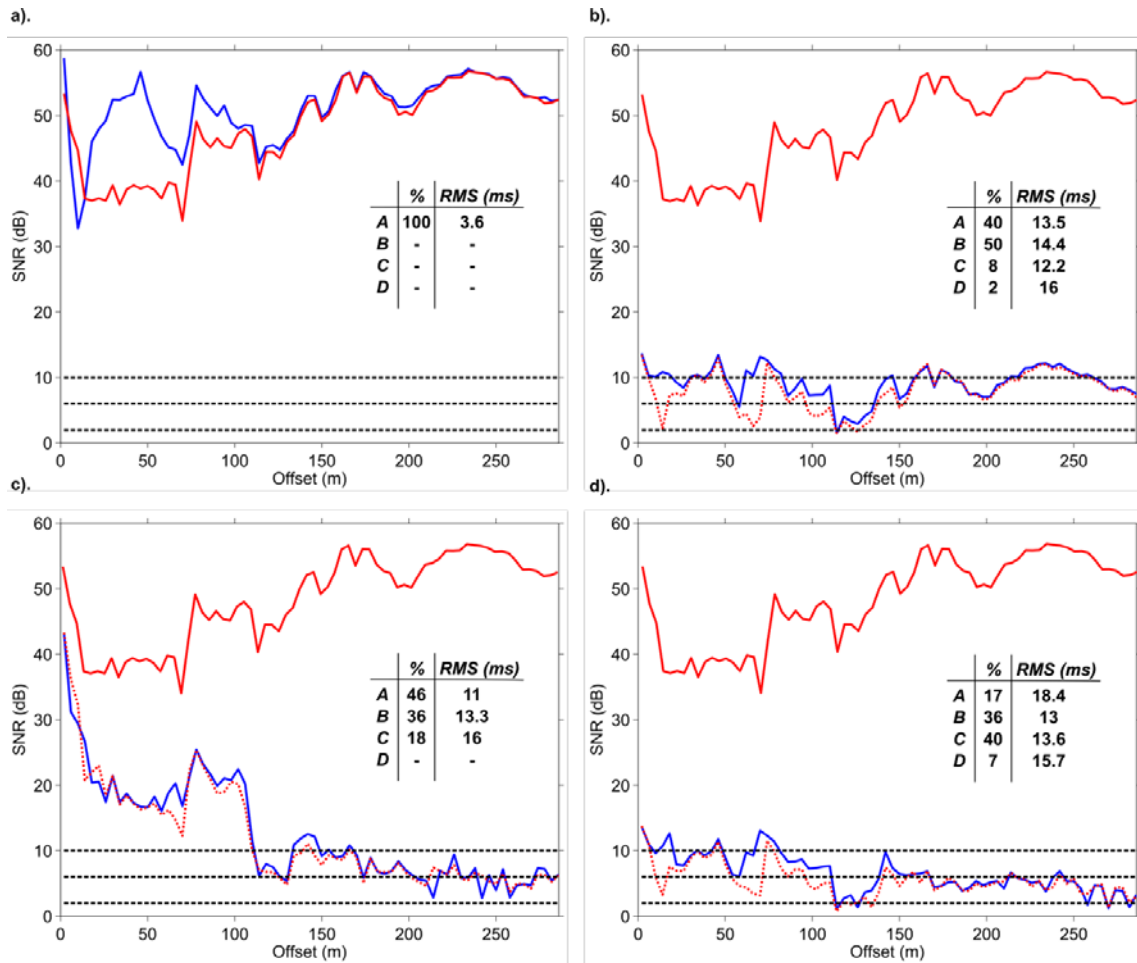


Figure 3.8. Quality assessment of automatic and actual picks, respectively in blue and red curves for zero noise (a), 20% random noise added (b), band-limited noise added (c), and random and band-limited noise added (d). The red solid curve represent the quality factor of the actual picks for the noise-free gather, the red pointed curve the same but after adding the noise. The tables in a, b, c, and d show the percent of the traces in each quality class and the related RMS of the differences between the automatic and real picks. The dashed black lines represent the limits of quality categories.

3.4 Conclusions

We have introduced an adaptive multi-method algorithm to pick automatically the first arrival times of near surface seismic data. Using MNW method combined with iterative look-like Monte Carlo approach supplied a good localization of the first arrivals trend, even if the arrival onset embedded by a high noise level. Integrating the three method results (MNW, HOS, and AIC) as a combination of the first arrival properties (energy, gaussianity, and stationarity) yielded fruitfully as a precise, nearly parameter-free, and full automatic picking algorithm.

The adaptive algorithm's implementation on synthetic data provided sufficiently good results in comparison with the calculated theoretically picks from a realistic earth model. Moreover, the noise robustness tests, with no added noise and with additive random Gaussian and limited-band noises, demonstrated the reliability of proposed algorithm, resulting in maximum RMS of 16 ms in the case of exaggerated noise. In addition to the first arrival times, the associated uncertainties with quality assessment were determined automatically. This information can be utilized in a posterior processing such as in tomographic inversion.

The number of user-tuning parameters was kept as low as possible as an attempt to reduce parameters errors and make it more user friendly. In fact, the main period of the first arrivals is the only parameter required for performing the automatic picking. All the others parameters are auto-estimated and adaptive across the processing flow. In addition, the inherent problems filtering were completely avoided and no a priori band-pass filtering is therefore needed.

Finally, in cases of very weak events with high noise presence, local enhancement of the first arrival signals helped overcoming the only limitation (a bit delay picking) of our algorithm by improving its accuracy. For instance, like as this enhancement may be achieved by local stacking or high fidelity filtering.

3.5 Acknowledgements

We are grateful to M. Antoine Chevalier, for his support and advice concerning to Monte Carlo approach.

Evaluation la performance de l’algorithme développé

Sommaire

4.1 Introduction.....	55
4.2 Deep seismic data (petroleum-scale exploration)	56
4.2.1 Modified Coppers’ Method (MCM).....	56
4.2.2 Fractal-Dimension Method (FDM)	57
4.2.3 Data sets.....	57
4.3 Shallow seismic data (groundwater-scale prospection).....	61
4.4 Near surface seismic data (environment-scale prospection)	63
4.4.1 Modified Energy Ratio (MER)	63
4.4.2 Data sets.....	64
4.4.2.1 Profile 1 (good survey conditions and high SNR)	64
4.4.2.2 Profile 2 (bad survey conditions and very low SNR)	66
4.5 Automatic control editing (quality control)	68
4.6 Conclusions and suggestions.....	72

Dans le chapitre précédent, nous avons développé un algorithme adaptatif automatique (AAA) permettant de pointer automatiquement les temps des premières arrivées ainsi que leurs incertitudes. L’application de cet algorithme sur des données synthétiques a fourni des résultats cohérents, en termes de précision, avec ceux pointés manuellement. Ce chapitre a pour objectif de valider la méthode proposée sur différentes séries de données réelles, possédant des rapports signal/bruit différents et pour différentes échelles de prospection [(i) sismique profonde ou « pétrolière » (ii) sismique de sub-surface (jusqu’à 150 m) et (iii) sismique de proche surface]. Cette validation est effectuée en comparant notre algorithme avec d’autres méthodes du pointé automatique [méthode modifiée de Coppers (MCM) ([Sabbione and Velis, 2010](#)), de la dimension fractale (FDM) ([Sabbione and Velis, 2010](#)) et du rapport d’énergie modifiée (MER) ([Han et al., 2009](#))].

La qualité de l’impetus des premières arrivées dépend fortement, non seulement du rapport signal/bruit, mais aussi de la forme du signal source utilisé. La précision de l’algorithme développé a été principalement évaluée via deux types de source considérés dans le cas (ii): sources impulsionnelles (dynamite) et sources vibratoires (Vibrois). Elle a aussi été évaluée avec d’autres types de source souvent utilisées dans les cas (ii) et (iii) : sources quasi- impulsionnelles (chute de poids et marteau).

Un facteur empirique de contrôle complémentaire est mis au point, à l’aide d’une fonction poids « qualité » et de l’incertitude des temps d’arrivée, afin de vérifier la

cohérence des pointés acquis. En conséquence, les pointés imprécis ou de piètre qualité sont ainsi automatiquement écartés.

Résultats: A l'exception des sources vibratoires où l'algorithme de pointé automatique basé sur les méthodes FDM et MCM reste légèrement meilleur, l'algorithme AAA reste le plus performant pour des sources impulsionnelles ou quasi-impulsionnelles pour toutes les conditions du bruit. Il n'y a pas vraiment de challenge dans la précision et la détermination du pointé automatique des premières arrivées, excepté aux grands offsets.

La discussion objective des résultats permet de définir les conditions dans lesquelles la technique proposé est la plus efficace ou lorsque les résultats sont moins satisfaisants.

Efficiency comparison of the proposed method versus some other strategies

4.1 Introduction

The aim of this chapter is to evaluate the developed algorithm performance as applied to real seismic data sets (with different SNRs), and to assess the impact of ambient noise and source type on our algorithm robustness. The quality of first arrival onsets depends on the near-surface structure, source type, and signal-to-noise ratio (SNR) (Sabbione and Velis, 2010; Yilmaz, 2001). Since the earth acts as a minimum phase low pass filter (Aki and Richards, 1980), the phase and shape of the first arrival signal change for different parts of the shot gather. Therefore, the automated picking of first arrivals can be a very difficult task for data acquired in a complex near-surface background. Additionally, if the source wavelet is zero-phase as when Vibroseis sources are used, the sweep correlation often produces side-lobes that arrive before the first onset, thus making the automatic picking even more tricky.

Exploration-scale active-source seismic datasets were chosen here to validate the implemented algorithm on real data. These data were acquired by using different source types for land seismic applications (i.e. different survey-scales: (i) deep or up to 3000 m - petroleum scale, (ii) shallow or up 150m - groundwater prospecting scale, and (iii) very near surface or up 30m - environment scale). The performance assessment is achieved by comparing the algorithm result with manual and other automatic picking techniques. We used the manual picks (where they supposed to be correct) as a reference in most of the benchmarking cases.

In the following sections, three existing automatic picking methods (MCM, FDM, and MER) are compared versus our proposed algorithm (AAA: Automatic Adaptive Algorithm) in terms of performance accordingly to different source and environment surveys. Table 4.1 presents a summary of the picking algorithms considered here. Each of the automatic picking methods has at least one parameter which may affect the accuracy of the picks and the results of each can be data dependent.

Wrong picks editing is generally achieved manually. Here, we propose a semi-automatic quality control (Q-control) for discarding the erroneous picks. This technique is based on analyzing simultaneously the picked arrivals quality (SNR or Q-factor) and the time uncertainty derived from our algorithm picking.

Table 4.1 : Overview of automatic picking algorithm considered

Algorithm	Key reference(s)	Parameters and recommended values
Modified Copen's Method (MCM)	Sabbione and Velis (2010)	Window size = 1.5 times arrival wave's dominant period (T_d) Stabilization constant $\beta = 0.2$
Fractal Dimension Method (FDM)	Sabbione and Velis (2010)	Window size = kT_d , k the lowest integer that gives the size window $\geq 48 + T_d/2$
Modified Energy Ratio (MER)	Han et al. (2009) Wong et al. (2009)	Window size = 2-3 times arrival wave's dominant period (T_d)
Automatic Adaptive Algorithm (AAA)	Chapter 3 of this study	Window size = arrival wave's dominant period (T_d)

The Douglas' concept (1997) "the less filtering, the better picking algorithm" has been applied, where no any data preprocessing, including the filtering, was performed before the automatic picking process.

4.2 Deep seismic data (petroleum-scale exploration)

The following methods were compared:

4.2.1 Modified Coppers' Method (MCM)

Sabbione and Velis (2010) present a modified version of Coppers' method (MCM) (Coppers, 1985). As for other energy-ratio algorithms, energy is computed in two windows. However, the size windows and their location setting differ from the other energy ratio techniques (e.g. STA/LTA; Chen and Stewart, 2005; Munro, 2005; Saari, 1991 or MER; Han, 2010; Wong et al., 2009). They calculated the energy of the seismic trace $s(t)$ within two nested windows

$$E_1(t) = \sum_{i=t-n_l+1}^t s_i^2, \quad (4.1)$$

$$E_2(t) = \sum_{i=1}^t s_i^2,$$

where n_l is the length of the leading window. The second window length increases with time. The energy ratio is then calculated as

$$ER(t) = \frac{E_1(t)}{(E_2(t) + \beta)}, \quad (4.2)$$

where β is called "stabilization constant". Sabbione and Velis (2010) recommend using $n_l = T_d$ (T_d the dominant period of first arrival waves) and $\beta = 0.2$. They applied also an edge-preserving smoothing (EPS) filtering on the resulting energy-ratio attribute. Finally, the first arrival time is assigned to the sample that corresponds with the highest derivative of the filtered attribute.

4.2.2 Fractal-Dimension Method (FDM)

Sabbione and Velis (2010) estimated the fractal dimension of a seismic trace $s(t)$ (with noise-tuned added by user) based on its variogram $V(h, t)$ calculated within a sample-by-sample moving window (n_f), for four lag distances $h = 1-4$ samples.

$$V(h, t) = \frac{1}{n_f - h} \sum_{i=t-n_f+1}^{t-h} (s_{i+h} - s_i)^2 \quad (4.3)$$

Then the fractal dimension D is given by the slope b of the straight line fitting in the Mandelbrot-Richardson plot (equation (4.4)), based on the fact that the power law between the variogram and the lag distance ($V(h) \sim h^{4-2D}$):

$$D = 2 - b/2. \quad (4.4)$$

However, most fractal-based picking algorithms are based on the fact that there is a significant difference of D between the seismic signal and the background noise (Boschetti, 1996; Jiao and Moon, 2000; Liao et al., 2010). The random noise shows often a higher fractal dimension (about 2) than the signal (around one) (Sabbione and Velis, 2010).

Sabbione and Velis (2010) recommend $n_f = kT_d$, where k is the lowest integer that gives $n_f \geq 48 + T_d/2$ and they tuned manually the amount of added noise accordingly to the background noise of the seismic traces. They applied also EPS on the resulting fractal-dimension attribute to enhance the variation transition between the noise and signal. At the end, the first arrival time is assigned to the sample that corresponds with the lowest derivative of the filtered attribute.

Sabbione and Velis (2010) combined each of the two above methods with a mispick-correcting technique, which is implemented via a five steps procedure (refer to Sabbione and Velis (2010) for more details), exploiting the benefits of the data in the entire shot record. The alignment of all picks over the shot gather was performed by piecewise (of a hypothetical refraction model) linear regression. Though this process the picks associated with bad or dead traces, which are corresponded with high standard deviation, can be discard.

Here, the same parameters of Sabbione's algorithms were applied for comparison purpose. Additionally, a linear interpolation of the mispicks traces has been performed.

4.2.3 Data sets

In order to compare the Automatic Adaptive Algorithm (AAA) performance with the strategies of Sabbione and Velis (2010), we choose two shot gathers from the seismic records in the book of Yilmaz (2001), available at <http://www.cwp.mines.edu/data/oz.original/>.

Shot gather 6 is shown in **Figure 4.1a**. The data consists in 48 traces with sampling rate ($dt = 4$ ms) and distance interval ($dx = 100$ m). The shot gather was acquired in the Far East with a dynamite source. The dynamite source often provides impulsive

first arrivals (Sopher et al., 2014). The first arrival times picked manually and automatically by three methods are shown in **Figure 4.1a** and **b**, whereas the quality of the arrivals corresponding with them are shown in **Figure 4.1c** and **d**. The quality (SNR) of the resulting picks is computed using the ratio of RMS amplitudes in specified signal and noise windows (see **Chapter 3**). Apart the FDM quality picks, the quality of first arrivals corresponding with two others pickers is comparable with those obtained by manual picking.

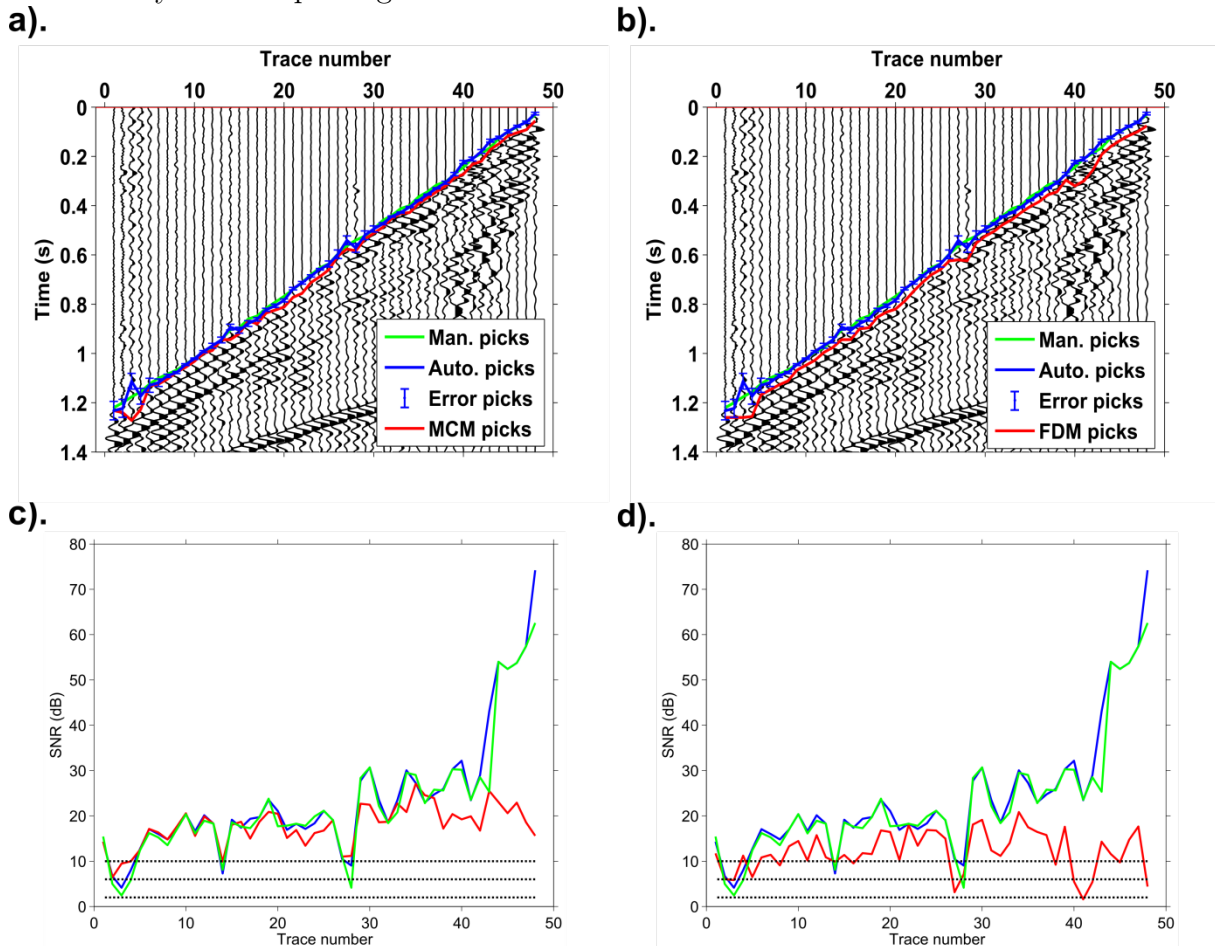


Figure 4.1. Performance comparison on a dynamite shot gather: **(a) and (b)**. Dynamite shot record number 6 from Yilmaz (2001) with the first arrival picks determined manually (green), adaptive algorithm (blue), and MCM/FDM (red), respectively on a and b. **(c) and (d)**. The quality of the first arrival signals corresponding with the picks shown in a and b, respectively.

Shot gather 23 is shown in **Figure 4.2a**. The data consists in 48 traces with sampling rate ($dt = 2 \text{ ms}$) and trace spacing ($dx = 67 \text{ m}$). The shot gather was acquired in the San Joaquin Basin with a Vibroseis source. The Vibroseis source often produces precursor energy lobes before the first arrivals due to its zeros phase wavelet signature. Therefore, locating where is the accurate first arrival onsets, in the most cases, can be difficult even for a qualified human operator. The first arrival picks (picked manually and automatically) with corresponding quality estimation are shown

in **Figure 4.2**. Since this shot is issued from Vibroseis source, the first arrival onsets are not easy to identify on the seismic record even for the manual picking.

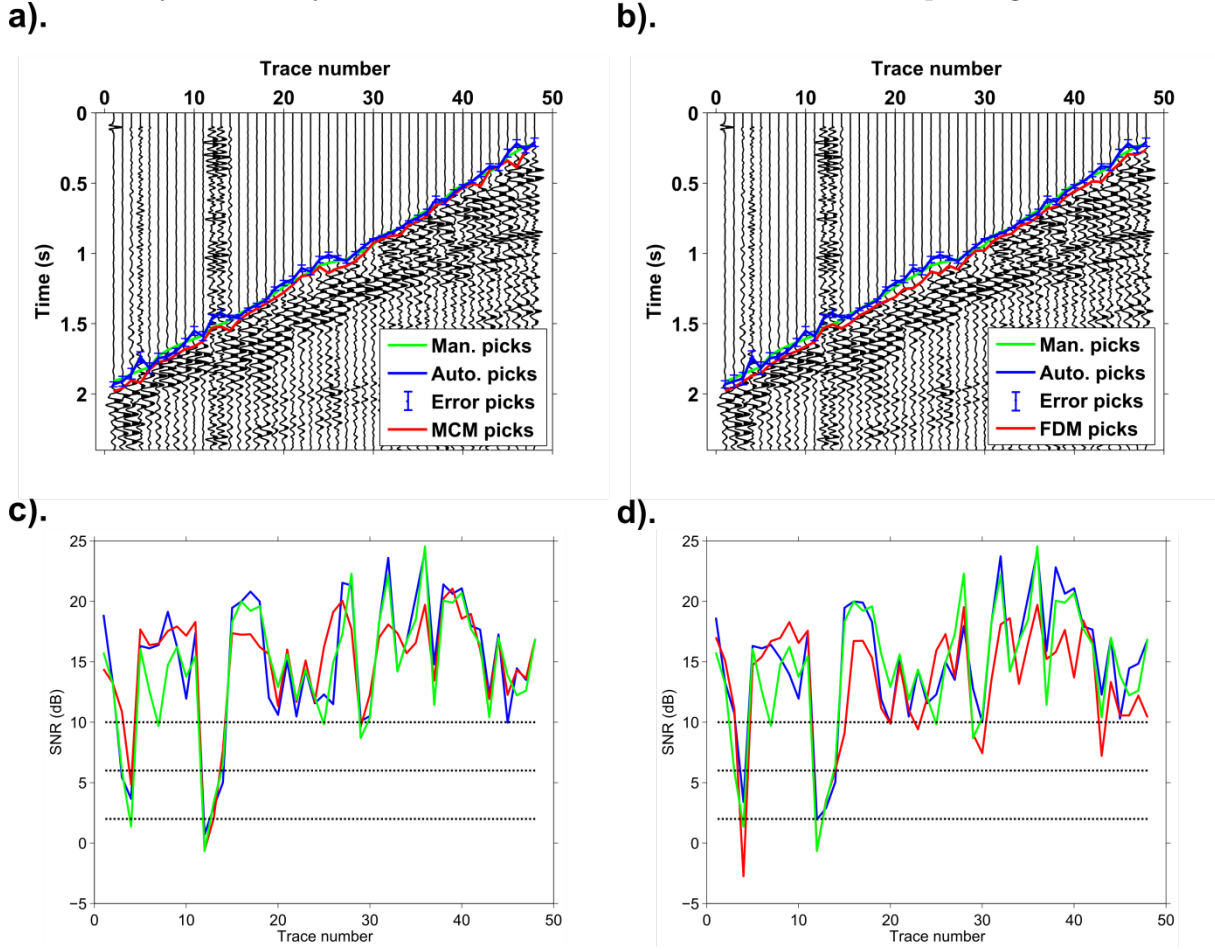


Figure 4.2. Performance comparison on a Vibroseis shot gather: **(a) and (b)**. Vibroseis shot record number 23 from [Yilmaz \(2001\)](#) with the first arrival picks determined manually (green), adaptive algorithm (blue), and MCM/FDM (red), respectively on **a** and **b**. **(c) and (d)**. The quality of the first arrival signals corresponding with the picks shown in **a** and **b**, respectively.

The manual picks, which are determined by an experienced geophysicist, are used as reference for benchmarking the developed algorithm results against those of MCM and FDM algorithms. By visual inspections on the MCM and FDM results, it is found that MCM and FDM methods picked the first-breaks with a delay about the half main period of the first arrivals (i.e. they picked the first considerable maximum of the P-wavetrains). For this reason, I performed a correction for the picks issued from MCM and FDM method by shifting them up $0.5 T_d$.

Figure 4.3 shows the comparative performance of AAA, MCM and FDM picking algorithms for dynamite shot gather. For P-wave arrival picking, AAA, MCM and FDM give overall accepted residuals with respect to manual time picks (50% of picks for each method are within the ± 20 ms error margin with respect to the referred picks). The manual picking error is assumed to be 20 ms (i.e. 5 samples for $dt = 4$ ms).

The results are considered as satisfactory when only half of the channels has an uncertainty in the range of what is considered appropriate for manual picking (Oye and Roth, 2003). The RMS and percentage of picked traces with error margin (20 ms) of the absolute time deviation from manual picks are shown in **Table 4.2**, highlighting stability of the algorithms. Although all algorithms perform similarly, AAA has the smallest RMS (15 ms) and the biggest percentage of picked traces (87.5%) in the considered error margin on the dynamite shot record.

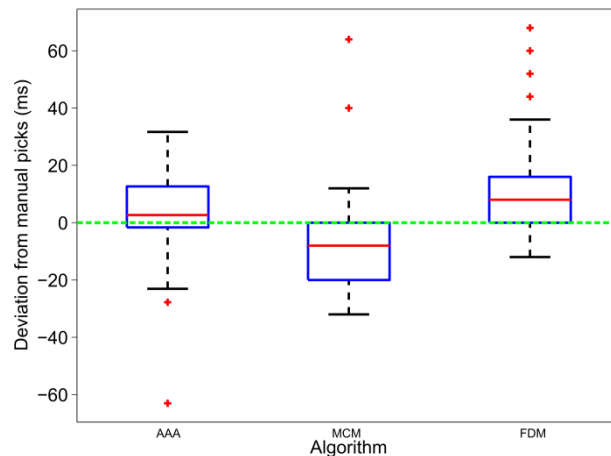


Figure 4.3. Boxplots of the differences between manually and automatically derived P-first arrivals using AAA, MCM, and FDM on the dynamite shot gather shown in **Figure 4.1**.

The same strategy used for the dynamite shot gather was followed for evaluating time picking algorithms on Vibroseis record. **Figure 4.4** shows the distribution of the differences between the manual picks and those picked by of AAA, MCM and FDM automatic algorithms. Each of AAA and MCM results has a distribution error over 20 ms , despite of their median matches well with zeros difference with respect to the manual picks. In contrary the FDM show acceptable distribution of the deviations from the referred picks but about 75% of distribution is found in the positive side due to the late picking of the first arrival times. For this dataset, the AAA shows bad picking results comparing to the strategies proposed by Sabbione and Velis (2010) where less than 50% of traces were picked in the error margin of the manual picking. On other hand, FDM reveals better performance in comparison with the AAA and MCM in picking this seismic shot. **Table 4.2** confirms these results, where 62.5% of the picks are found within the accepted error margin ($\pm 20\text{ ms}$), and the RMS is equals to 22 ms . However, FDM results might be improved by adding different noise-tuned amounts. Although the FDM algorithm gives a consistent picks over shot gather, it tends to pick very lately the first arrivals (**Figure 4.4**).

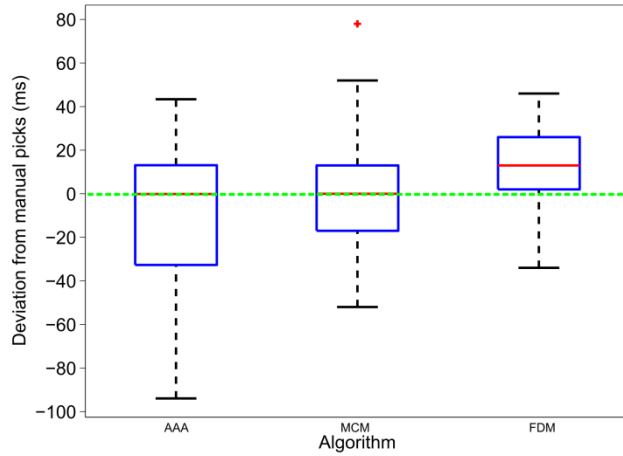


Figure 4.4. Boxplots of the differences between manually and automatically derived P-first arrivals using AAA, MCM, and FDM on the Vibroseis shot gather shown in **Figure 4.2**.

Table 4.2. Assessment of the RMS of the differences and the picking errors within a 20 ms tolerance for the real data shown in **Figure 4.1** and **Figure 4.2**. Here, the errors are considered to be between the manual and automatic picking using AAA, MCM, and FDM picking algorithms. (Abs. diff = absolute difference)

Source-type	Statistical properties	AAA	MCM	FDM
Dynamite	RMS (ms)	15	18	21
	Abs. diff. $\leq 20\text{ms}$ (%)	87.5	81.25	83.33
Vibroseis	RMS (ms)	33	27	22
	Abs. diff. $\leq 20\text{ms}$ (%)	49.92	64.58	62.5

In conclusion, the AAA outperformed over the MCM and FDM Sabbione and Velis’s strategies for impulsive onset data record (dynamite shot gather), but it failed to pick correctly the non-impulsive seismic record (Vibroseis shot gather). The AAA performance on Vibroseis data might improve by pre-processing the seismic records before automatic picking process (e.g. compressing the arrival waves by deconvolution).

4.3 Shallow seismic data (groundwater-scale prospection)

These data were acquired during a near surface survey in the Chaux forest in the East of France. An accelerated 40 kg weight (a hitch-mounted in front of a tractor), dropped onto a steel plate, is used as a seismic source for generating the energy, with a 4 m source interval along the profile. The recorder system consists of 72 14-Hz vertical component geophones, with 4 m receiver spacing along a maximum offset about 282 m. The recording time was 1 s with a delay of -0.2 s and sampled with a 1 ms sampling rate.

In general, near-source traces exhibit well-defined and prominent events onset and thus does not present any challenges for automatic picker. But the difficulties of first arrivals picking arise at far offsets, where the SNR of the first arrivals are often gradually decreased due to interference of ambient noise. Finding the correct first arrivals under such conditions of noise is one of the persisting grand challenges of automatic picking techniques. In such circumstances, the amount of rejected mispicked traces becomes however considerable and therefore, it reduces the investigation depth by seismic methods.

A full offset shot (end-on spread gather) has been chosen to illustrate algorithm's performance, for which a comparative analysis between the automatic and manual picking results has been done. It is worth mentioning that no prior filtering process took place. Each trace is scaled to its maximum amplitude. The dominant period of the P-waves is estimated directly from the whole waveforms of the profile data and approximated to be 35 samples (generalized for all traces). Since the travel time curve seems to be linear with the offset, a window about 75% of total trace shot is suited for smoothing purpose in the like-look Monte Carlo routine. The other required parameters are setup automatically (auto adaptive parameters) as described in the previous chapter.

Figure 4.5 shows an example of the picking results by our adaptive algorithm (blue), beside those picked manually (green). It is shown that the AAA accuracy is deteriorated at far offsets, as it is expected with such as bad quality of the first arrival signals.

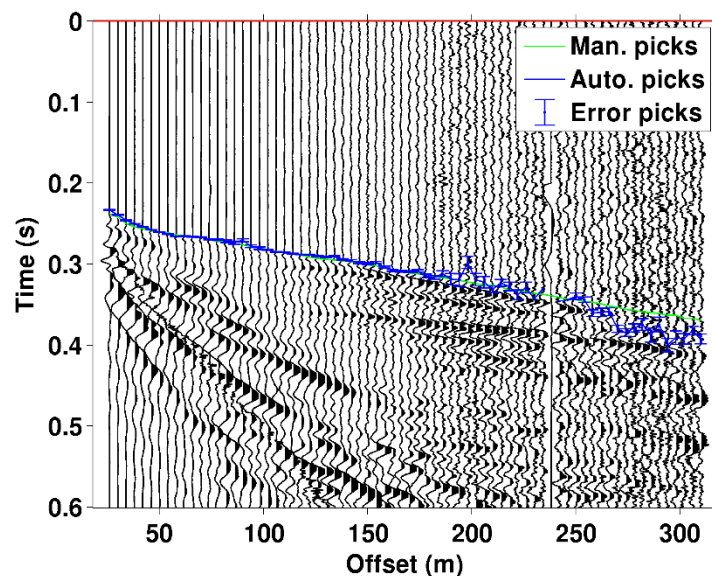


Figure 4.5. An example of automatic picking results by AAA on weight-drop-source shot gather.

The results of AAA applied on about 20 shot gathers along this profile are summarized by histogram and boxplot diagrams as shown in **Figure 4.6**. The picks

were classified within empirical quality classes (1-4) or uncertainty classes (A-D) (Lois et al., 2013). We defined these categories by the manual pick onsets quality determined by the SNR traces (4.6). The traces percent are 71.5%, 9.2%, 9%, and 10.3 %, respectively, within each of 1, 2, 3, and 4 category.

$$Q\text{-class} = \begin{cases} 1, & \text{if } SNR > 10 \\ 2, & \text{if } 10 \geq SNR > 6 \\ 3, & \text{if } 6 \geq SNR > 2 \\ 4, & \text{if } \geq SNR > 2 \end{cases} \quad (4.5)$$

The RMS of the differences is about 10 ms, and 73.4% of picked traces fall within ± 5 ms (i.e. margin error only 5 samples) around the manual picks. The algorithm performance is still limited at far offsets traces where the arrival onsets are completely submerged by noise, this is confirmed by the boxplot of the last quality class (4) (Figure 4.6 right). Enhance these trace onsets is an indispensable task, or at least they should be excluded out before any posterior processing.

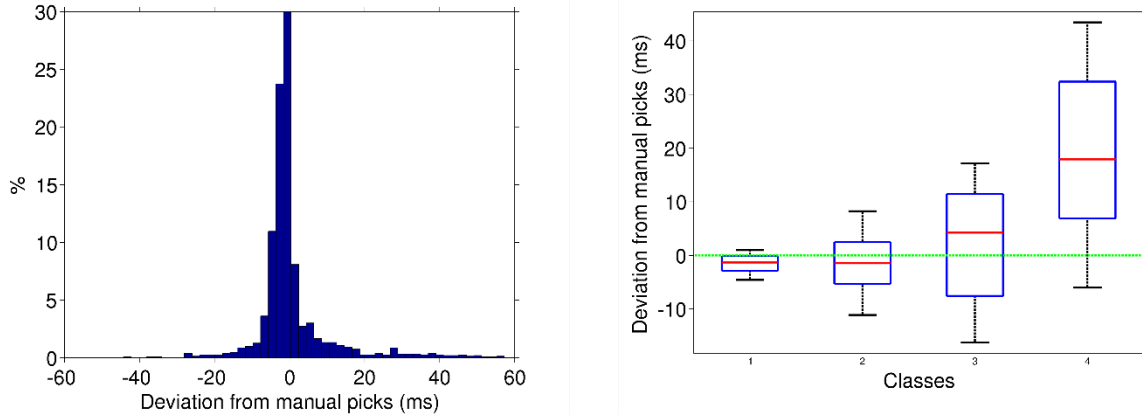


Figure 4.6. Histogram of AAA picks deviations from manual picks (**left**). Boxplot diagrams of the differences between manually and automatically picks within Q- classes (1-4) classified according to SNR of manual picks (**right**).

4.4 Near surface seismic data (environment-scale prospection)

The following methods were considered:

4.4.1 Modified Energy Ratio (MER)

The MER algorithm, proposed by Han et al. (2009), is an extension of the standard STA/LTA with both pre- and post-sample windows of equal size (N_e). Assume a seismic trace $s(t)$, in which the time index $t = 1, 2, \dots, N$. The energy ratio (ER) given by

$$ER(t) = \frac{\sum_{i=t}^{i=t+N_e} s_i^2}{\sum_{i=t-N_e}^{i=t} s_i^2} \quad (4.6)$$

Given (4.6), the modified energy ratio, denoted as MER, can be calculated as

$$MER(t) = (ER(t) \cdot abs(s(t)))^3. \quad (4.7)$$

The first arrival time of the trace $s(t)$ is indicated by the sample of $MER(t)$ that gives the largest amplitude (Wong et al., 2009). The recommend length of window (N_e) is two to three times the source wavelet dominant period ($2-3 T_d$) of the related seismic arrival wave (Akram, 2014, p. 54; Han, 2010, p. 39; K. Liu et al., 2015). The MER method has been shown to more accurately pick noisy first breaks than other energy ratio methods (e.g., (STA/LTA) in Han et al., 2010). However, for low SNRs (3-1.5), Gaci (2014) showed that the STA/LTA method gives more accurate results than the MER due to the presence of more than one peak on the MER attribute. Furthermore, he claimed that the denoised STA/LTA version by wavelet transform outperforms the MER in term of accuracy at these levels of SNRs ($3 - 1.5$).

4.4.2 Data sets

A lot of shot gathers from two near surface seismic profiles have been chosen to compare the Automatic Adaptive Algorithm (AAA) performance with the MER method proposed by Han et al. (2010). It is worth mention, here, that the two compared algorithms require the same parameter (the main period of the first arrival). The performance of the two algorithms (AAA and MER) was compared with the manual picks, which are considered as reference for benchmarking process.

4.4.2.1 Profile 1 (good survey conditions and high SNR)

Several geophysical data were originally collected for hydro-geophysical investigation of Durance's silt and gravels (Merle domain / Coussouls de Crau area in South of France) when introducing Master students (Fieldtrip training in October 2010) to geophysical prospection. Among the various methods used in this study, a 2D seismic refraction profile with a 200 m length has been also recorded for identifying the water table at an experimental plot. The seismic source consisted of an aluminum plate hit vertically by a 5 kg sledgehammer. The recorded run 1 s after each shot with a delay of -20 ms, sampling at 0.25 ms on 48 14 Hz vertical-component geophones. Five shots at each location were averaged to increase the signal-to-noise ratio. Receivers were placed every two meters and sources every four meters. The profile topography was perfectly horizontal and the acquisition conditions were ideal with a good SNR of the first arrivals.

The picking of first arrivals that are required for comparison purpose was done manually using the Pickwin (version 4.2.0.0) module of SeisImager computer program from GEOMETRICS. For the automatic picking and display purposes the picking window was set in a way that only the first 200 ms of the record were taken in consideration.

Figure 4.7 shows two examples demonstrating the AAA robustness against MER in presence a moderate ambient noise.

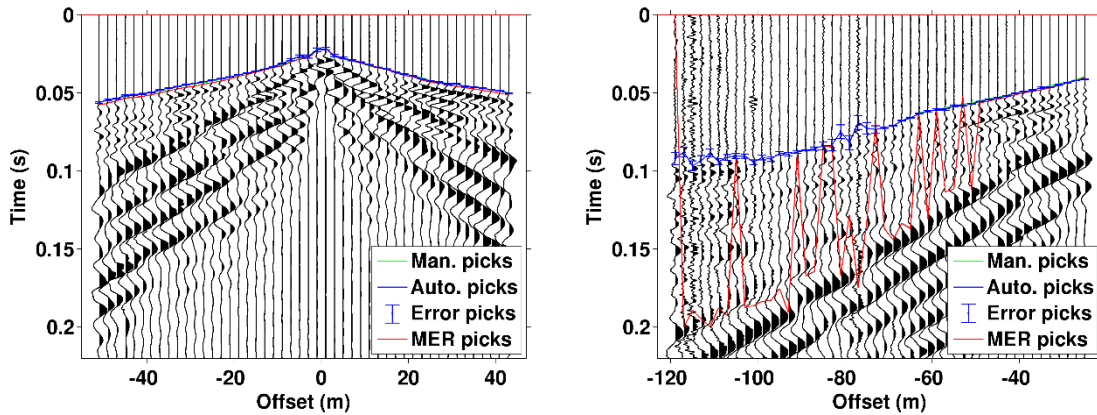


Figure 4.7. Examples of AAA (bleu curves) robustness versus MER (red curves) in presence a moderate noise: **(Left)**. Very high quality shot gather. **(Right)**. A moderate quality shot gather.

64 shot gathers along the profile were picked by the two algorithms. The performance of each of them was evaluated against the manual picking. The histograms of the deviation from the latter (**Figure 4.8**) show the consistence of AAA results where its distribution is approximately normal, whereas the MER results have abnormal distribution. That reflects some secondary effect due to surface waves.

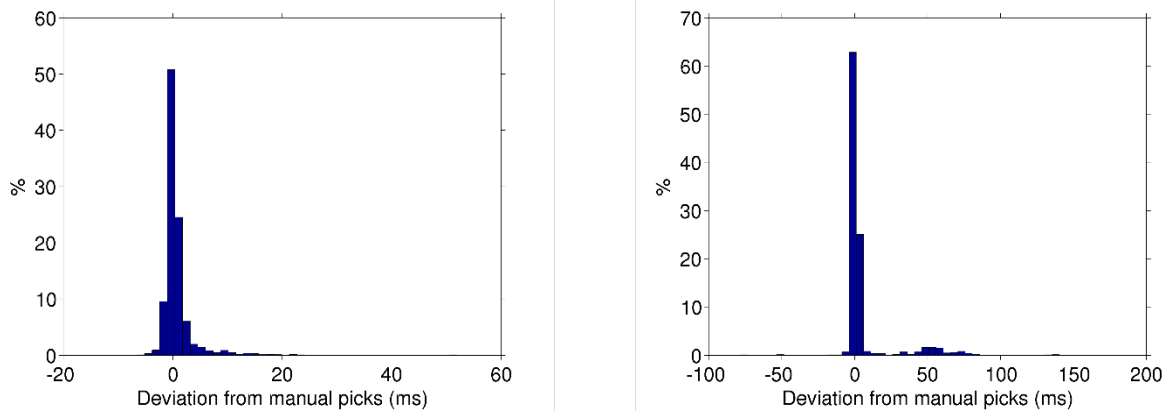


Figure 4.8. Histograms of differences between automatic and manual picking times using the proposed AAA **(left)** and MER algorithm **(right)**.

Boxplots of the differences are also shown on **Figure 4.9**. Viewing the data in this way offers insights about the accuracy and consistency of the proposed AAA algorithm. The median of the AAA deviation from the manual picking is very close to zero (boxplot 1 in **Figure 4.9**) and with RMS of differences equal to 3.5 ms, while 90% of the differences between the MER and the manual picks are positive (boxplot 2 in **Figure 4.9**) with RMS about 19.3 ms. That explains the late picking by MER. However, the median of MER distribution results is about 1 ms and 50% of the

difference is less than 2 ms for the two picking algorithms. We use a customized boxplot where the lower and upper whiskers are set to be 10% and 90% of distribution, respectively, and all outliers were deleted.

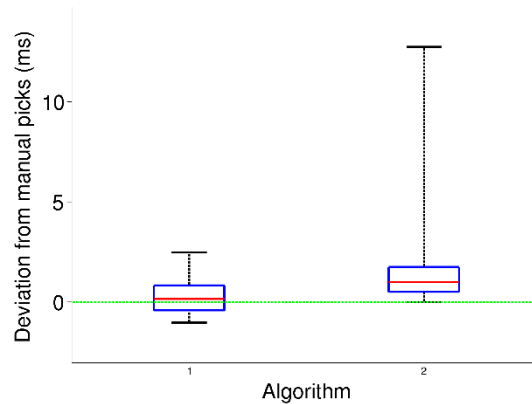


Figure 4.9. Customized boxplots of the deviations automatic picks (1 = AAA) and (2 = MER) from the manual picks, the pointed green line is the baseline though the difference equals zero.

To emphasize the robustness of using AAA compared to the MER, we supposed that 2 ms (i.e. 8 samples for 0.25 sampling rate) is the manual picking error for this data set, then we calculated the percent of the picked traces in the error range. 85% of traces were picked by AAA in this margin error versus only 77.5% of traces picked by MER.

4.4.2.2 Profile 2 (bad survey conditions and very low SNR)

During the field-training course for Master students (Octobre 2009) in the Vanoise National Parc in the northern French Alps, several geophysical methods were used for two objectives (1) understanding and practicing the geophysical methods and measurements in the field (2) investigating and characterizing permafrost in the Plan du Lac rock glaciers and therefore, detecting the groundwater flow at the end of summer. A seismic profile has been achieved in order to survey the substratum depth and layers stratigraphy in this studied area, which has been previously surveyed using georadar prospection by Monnier et al. (2008), and then to correlate the results of both methods. The seismic source consisted of an aluminum plate hit vertically by a 5 kg sledgehammer, with five to seven hits at each source location is used to increase the signal-to-noise ratio. The recorded was 0.5 s after each shot with a delay of -20 ms, sampling at 0.25 ms on 48 14 Hz vertical-component geophones. Receivers were placed every 5 meters and sources at every 5 meters also. The profile topography was very rugged and the acquisition conditions were difficult with very low SNR of the first arrivals (mountain surveys conditions).

Figure 4.10 shows two examples of picking algorithm results for two noise conditions (left: moderate noise and right: high noise level). In both cases, the proposed algorithms fail to pick accurately the far offset traces. But at near offset traces, the AAA still keep its performance with low SNR of the first arrival signals (**Figure 4.10** right).

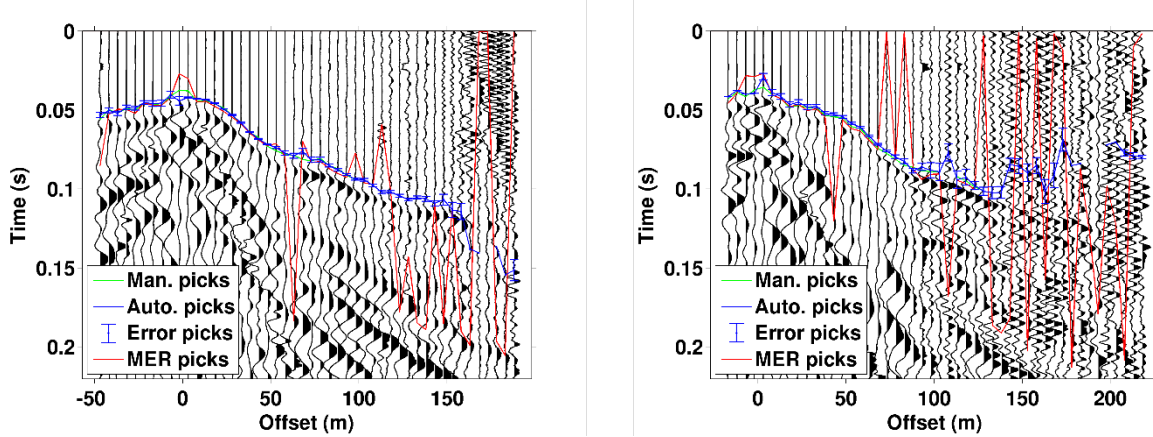


Figure 4.10. Examples of picked shot gather by automatic algorithms (AAA = bleu curves) and (MER =red curves) with highly correlated ambient noise: **(Left)**. Moderate quality shot gather. **(Right)**. Very bad quality shot gather.

The shot gathers, which are well picked manually by master students (on 20 shots), were taken in account for statistical results analysis. The picking results undergo the same way followed above for comparing different automatic algorithms applied to this data set. The histograms of the differences between the results derived from automatic picking algorithm and those picked manually (reference picks) are shown in **Figure 4.11**.

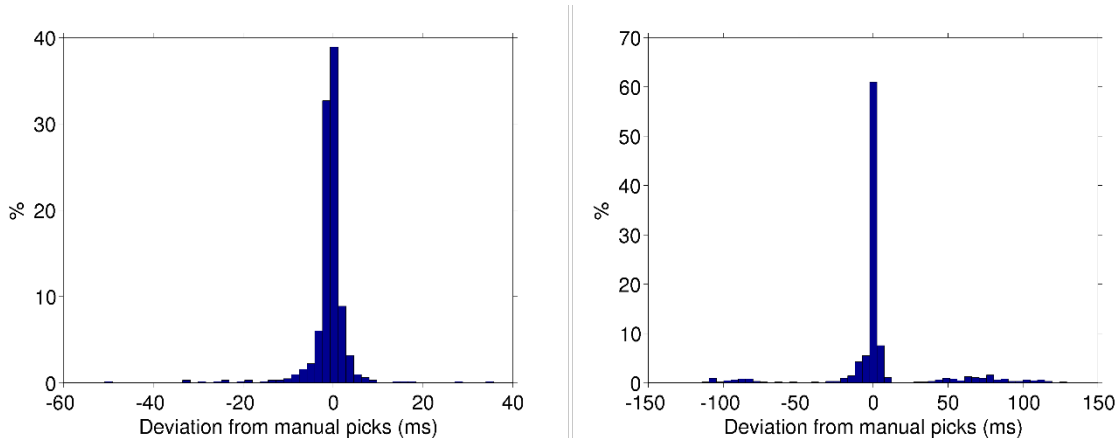


Figure 4.11. Histograms of differences between manually and automatically derived first arrival times (P-onsets) by AAA (left) and MER (right).

Figure 4.12 show the boxplots (1=AAA, and 2=MER) of the differences between the automatic and manual picks. It is noticed that the AAA difference results shows a Gaussian distribution with about 80% in $[-3, +3]$ ms time interval and the median of

the distribution is about -0.5 ms (**Figure 4.11-left** and **Figure 4.12**). On the other hand, only 50% of the MER differences are found at the same interval (i.e. [-3, 3] ms), with abnormal distribution (**Figure 4.11-right** and **Figure 4.12**). The RMSs of the differences are 4.96 and 34.35 ms, respectively, for AAA and MER. That again reflects the high sensitivity of MER by the high energy of the surface waves. By assuming the manual picking error of ± 2 ms (in analogy to vertical seismic profiling (VSP) studies for P-wave picking by Stewart et al. (1984)), we found that 73. % of the traces picked by AAA were in this margin error, whilst only 53.7 % of picked traces were in the margin error with MER picking method.

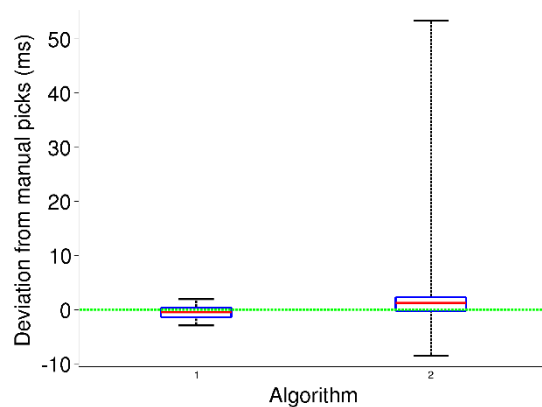


Figure 4.12. Customized boxplots of the deviations automatic picks (1 = AAA) and (2 = MER) from the manual picks, the pointed green line is the baseline though the difference equals zero.

4.5 Automatic control editing (quality control)

Interference with other events (later phases), imperfect picking processing and/or ambient noise, when the latter is present in any experiment, produce some uncertainties in the picked arrival times. Therefore, manual adjustment and/or editing of the picked first arrival times are necessary from time to time to take into consideration the inconsistent shape of the first arrival wavelet from trace to trace, which is typical for many near-surface settings (Ivanov and Miller, 2004). Random noise can be another source of wavelet-shape alteration (Ivanov and Miller, 2004). Proper weighting of automatic picker results has been discussed by many authors (e.g. Di Stefano et al., 2006; Diehl et al., 2009; Küperkoch et al., 2010; Lois et al., 2013). Mateeva (1999) proposed an estimated error in picked traveltimes (4.8) via the normalized quality factor “or P-ratio” for a neural network picking algorithm. The quality factor is, in term of neural network picking, the labels of the features on trace between 1 “first arrival” and 0 “not first arrival”. Mateeva supposed that the P-ratio distribution is Gaussian, which is consistent with the Gaussian-error assumption in

seismic tomography inversion. Thus, when the first arrival is in the searching time window, its time uncertainty τ is given by:

$$\tau = \sqrt{\frac{-0.125}{\ln(1-P^2)}} T_d, \quad (4.8)$$

where P is the normalized quality factor and T_d is the dominated first arrival period. Taking in consideration the weighting scheme proposed by Lois et al. (2013) and with analogy to Mateeva's formula (equation (4.8)), we defined our automatic editing/quality control to get rid of incorrect picks, as follows:

- All picks corresponding with SNR (Q-factor) equal or less than 2 dB (which are shaded with a transparent red window on **Figure 4.13**) are rejected.
- All picks corresponding with Q-factor equal or greater than 10 dB (which are shaded with a transparent green window on **Figure 4.13**) are accepted.
- The picks corresponding with $2 \text{ dB} < \text{SNR} < 10 \text{ dB}$ (which are shaded with a transparent yellow window on **Figure 4.13**) should be passed through following quality control:
- We adapted the input arguments of equation (4.8) according to our AAA outputs, where we defined P as the normalized Q-factor (SNR) by the median of SNRs over a predefined threshold SNR (Q-threshold). This threshold is defined according to the first arrivals quality over whole shot gather. We found that a threshold of 6-10 dB is suitable for most of the testing datasets. We modified T_d to be the related error of AAA picks. This error was estimated as the standard deviation of all sub-algorithms picks within a moving spatial-window of 3-5 traces after removing the local trend of the final picks in these window-traces. **Figure 4.14**-left shows the SNR and the corresponding pick errors, which are used for calculating the modified tau (τ in ms) (**Figure 4.14**-right).
- Then, all picks corresponding with the τ greater than a predefined-user threshold-error are rejected (**Figure 4.14**-right).
- Finally, we defined also a mispick gap control; where all picks after a gap of 3-5 mispick traces could be removed for keeping the consistency over the whole shot gather.

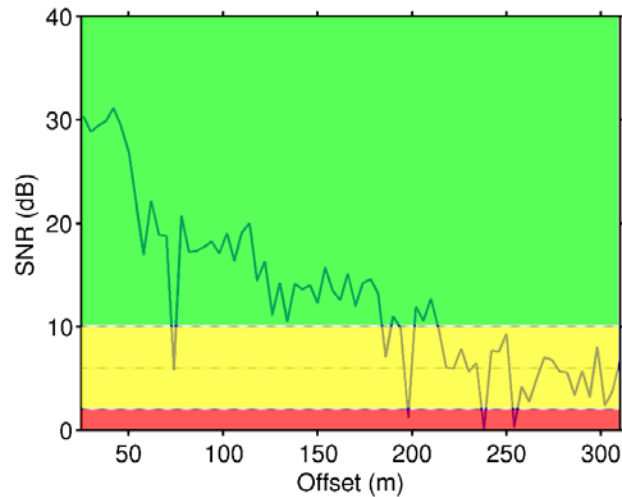


Figure 4.13. An illustrative example of quality control-1 based on SNR of the picked arrivals; picks in the transparent red window are to be rejected, picks in the transparent green window are to be accepted, and those in yellow one are to examine by the quality control-2 based on adapted Mateeva’s formula, equation (4.8).

However, the tradeoff between the picked arrivals quality and their consistency over the offset is governed by the two input functions (Q-factor and errors). On the other hand, both thresholds (Q-threshold and threshold-error), which react inversely to each other, control the amount-reject of the erroneous and inconsistent picks. Furthermore, the outliers picks will fortunately be rejected by most of inversion schemes due to their traveltimes are too inconsistent with the rest of the data (Mateeva, 1999).

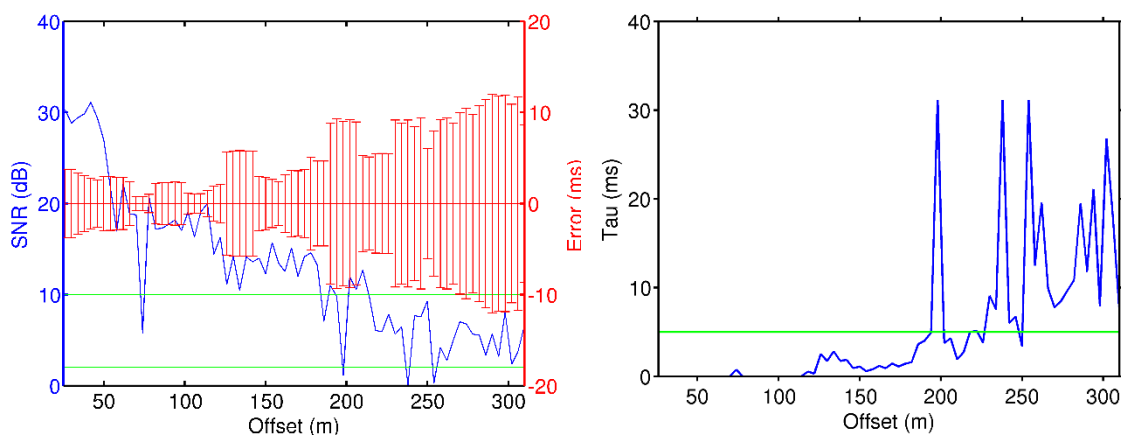


Figure 4.14. An illustrative example of quality control-2 based on modified Mateeva’s equation (4.8): **(Left)**. Input arguments functions of equation (4.8) (Q-factor in blue curve and pick errors in the red error bars), the two horizontal green lines represent the SNR at the 2 dB and 10 dB. **(Right)**. The resulting tau (τ) function (in blue curve) and the threshold-error at the level of 5 ms (in horizontal green line).

Figure 4.15 and **Figure 4.16** show two examples of the Q-control/editing process applied to two shot gathers recorded by two different sources, respectively, accelerated weight-drop and sledgehammer. The Q-control parameters for the first example are set to be Q-threshold = 10 dB, threshold-error = 5 ms, and the mispicks gap = 5 traces. It is noticed that, all edited picks on **Figure 4.15**-left are too consistent with the hand-picking data, and the automatic picks corresponding with bad arrivals quality or those picked with high uncertainty are excluded out. On other hand, the last pick (**Figure 4.15**-left) is rejected also by the mispicks gap control (**Figure 4.15**-right).

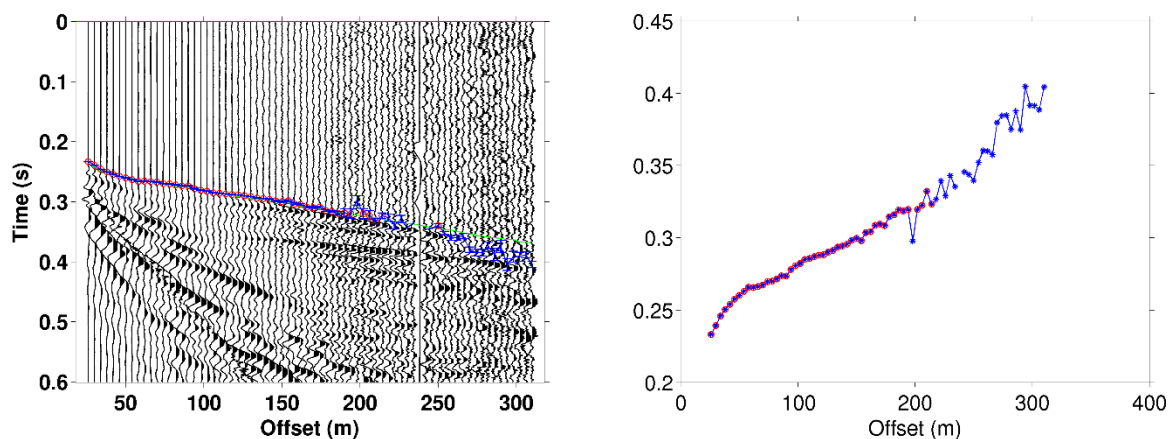


Figure 4.15. An example of Q-control/editing application on a field-data record: **(Left)**. A weight-drop source shot gather, blue asterisks and error bars are the outputs of AAA picking, green line represent the interpolated manual picks, and the edited picks are highlighted in small red circles. **(Right)**. Blue asterisks are the same automatic picks shown in left figure and small red circles are the final output of editing process (including the mispicks gap control).

Since the second example (sledgehammer source record) has low SNR of the first arrivals, we applied Q-control procedure with different parameters than those of the first example. We found these parameters (Q-threshold = 6 dB, threshold-error = 10 ms, and the mispicks gap = 3 traces) are efficient for keeping only the picks with high SNR. It is noticed that edited traces are the same those are picked manually (**Figure 4.16**).

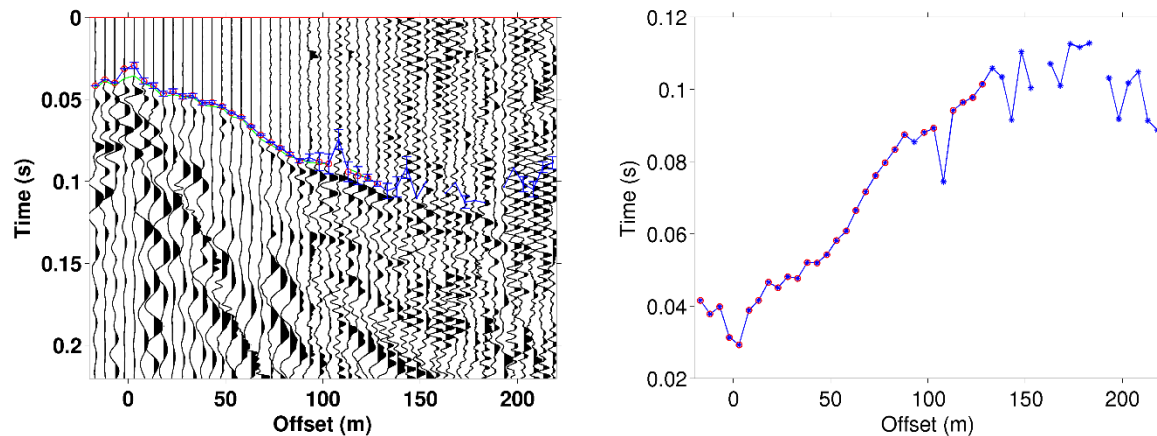


Figure 4.16. Another example of Q-control/editing application on a field-data record: **(Left)**. A sledgehammer-source shot gather, blue asterisks and error bars are the outputs of AAA picking, green line represent the interpolated manual picks, and the edited picks are highlighted in small red circles. **(Right)**. Blue asterisks are the same automatic picks shown in left figure and small red circles are the final output of editing process (including the mispicks gap control).

In practice, configuration parameters of the proposed Q-control/editing can be adjusted by comparison with some selected shot gathers picked manually, and then applied it on all shot gathers along profile.

4.6 Conclusions and suggestions

The testing results on different shot gathers confirm our implemented algorithm performance against other picker (MCM, FDM, and MER) for the most testing data sets, particularly for moderate to high noise conditions. The AAA shows high accuracy picking on impulsive seismic data (generated using a dynamite, a weight drop, or a sledgehammer). Overall, more than 73% of the impulsive-arrival traces were picked in the margin error of manual picking. However, it is less efficient for non-impulsive source (Vibroseis) records where FDM outperformed the other methods (including our algorithm). Unfortunately, the AAA performance is degraded for the far offset traces, which is characterized by very weak onsets correlated with high ambient noise, as the other picking algorithms. Therefore, the efficiency of automatic picking (including our algorithm) should be improved by enhancing the weak onsets traces before applying the automatic process (e.g. by using one of smart stacking approaches).

On other hand, depending on first arrivals quality, some of the automatic picks derived from AAA picking might be wrong (or inconsistent), particularly at the far offset. Additionally, there might be some dead traces where their picks are naturally wrong and must be discarded (Sabbione and Velis, 2010). The proposed Q-control procedure can achieved semi-automatically a high-fidelity editing of the automatic picks resulting from AAA. It takes in account the two essential factors, which are

considered when the manual editing is done, the arrival signal quality and consistency over the whole shot gather. Testing on real data records demonstrates a comparable performance of the proposed Q-control to manual editing. Finally, Q-control parameters settings can be tuned for a specific shot gather and globalized along entire seismic profile.

Renforcement local des arrivées sismiques par double sommation dans le domaine temporel

Sommaire

5.1	Introduction.....	78
5.2	Definitions and general concepts used in stacking	79
5.3	First stacking using local similarity.....	80
5.4	Reference trace	81
5.5	Second stacking using cross correlation.....	83
5.6	Denoising by wavelet transform (WT).....	85
5.7	Applications.....	88
5.8	Conclusions.....	90

Suite aux limitations de l'algorithme développé sur des données réelles, notamment aux offsets lointains, ce chapitre, correspond à un article en préparation pour une soumission à *Geophysical Prospecting*. Il concerne le traitement préliminaire utilisant la redondance d'informations des données sismiques, à l'intérieur d'une fenêtre glissante. Ce traitement peut être effectué avant le pointé des premières arrivées afin d'améliorer le rapport signal/bruit.

Une brève introduction rappelle le problème posé et les méthodes classiquement utilisées. La méthodologie proposée est fondée sur une double sommation dans le domaine temporel. La première sommation est effectuée en recherchant une similarité locale entre traces d'offset communs pour différent tirs. La deuxième sommation consiste à effectuer une corrélation croisée sur traces adjacentes, pour le même tir. Cela est valable dans le cas de prospection sismique de proche surface, où la distance entre les récepteurs est faible. En prévision d'une diminution du rapport signal/bruit des données sismiques, à grand offset d'un côté, et à hautes-fréquences de l'autre, il peut être nécessaire d'effectuer un filtrage préliminaire par transformée en ondelettes, adaptée au « débruitage » de signaux non-stationnaires. L'analyse multi-résolution/multi-échelle utilisant une base orthogonale d'ondelettes est alors proposée.

Finalement, les applications et les conclusions sur des jeux de données réelles montrent les avantages apportés, pour le pointé automatique, lors de l'application cette double sommation dans le cas de données sont trop fortement bruitées.

Chapitre 5. Renforcement local des arrivées sismiques par double sommation dans le domaine temporel

Local enhancement of first arrivals embedded in background noise by double stacking: application to automatic picking

Amin Khalaf (1,2), Christian Camerlynck (1) and Nicolas Florsch (1,3).

(1) *Sorbonne Université, UPMC Univ. Paris 06, CNRS, UMR 7619 Metis, 4 place Jussieu, 75005 Paris, France.*

(2) *Tishreen University, Faculty of Sciences, Departement of Geology, Lattakia, Syria.*

(3) *Sorbonne Université, UPMC Univ. Paris 06, UMI 209 Ummisco, 4 place Jussieu, 75005 Paris, France.*

Corresponding author:

Amin Khalaf

UPMC, METIS.

Case 105, 4 place Jussieu, 75005 Paris, France.

Phone : +33 (0)1 44 27 48 85

Fax : +33 (0)1 44 27 45 88

E-mail address: amin.khalaf@upmc.fr

Intended for Publication in Geophysical Prospecting

Abstract

The high noise level in the seismic records can significantly decline the efficiency of the automatic picking algorithms, particularly in the case of weak arrivals. The traditional methods used for improving the quality of arrivals are commonly based on pure spectral filtering. The problems related to the later arise when the signal and noise share the same band frequency. Furthermore, some of these filters may distort the key parameters of automatic picking (onset time and polarity). In the near surface surveys, the stacking addresses these problems thanks to data redundancy. To improve the signal-to-noise ratio of first arrivals, and thereby increase their detectability, we propose an optimal alternative: the double stacking in the time domain. The first weighted stack is based on local similarity of the common offsets traces for a few adjacent shots. This stack approach is more tolerant to significant misalignment signals, and locally emphasizes the buried signals. The second stack is applied to a few neighboring traces (within moving spatial window), in the same shot gather, after correcting the delay times via cross correlation tools. In cases of very noisy data, the waveform stacking and the cross-correlation analysis are less efficient at higher frequencies, due to the limited coherence of short-period waveform. De-noising using wavelet transform was carried out as an auxiliary procedure before the second stack. To validate our approach, a synthetic profile was generated from a realistic model via finite difference modeling, and considerable noise levels were added to the simulated data. The weak events are remarkably enhanced, and are more identifiable on the stacked sections. We perform additionally a comparative study of the performance of automatic picking methods, using adaptive multi-method algorithm on noisy real data. The results demonstrate the feasibility of applying the double stacking before the automatic picking.

5.1 Introduction

Many real-world signals are often corrupted by additive noise due to imperfect acquisition and transmission. The seismic noise sources can be separated into two categories. The first category of noise comes from experimental errors. These errors comprise any unexpected perturbation of the recording environment during data acquisition. A geophone can have malfunctions or the recording systems can have coupling with electrical signals creating erratic noise in the seismic record. The second category of noise comes from outside factors, such as human activity nearby seismic survey area might also contaminate seismic records. Wind motion or cable vibrations can generate random noise. Another type of the noise has origin of how the wave propagates in the ground. For instance, the surface waves (commonly termed ground roll) is often discarded as unwanted signal in seismic reflection. Another example, multiple reflections are a main source of coherent energy in the seismic record. These

noise sources create more coherent energy on the data and can be misinterpreted as true signal. In the reflection method, surface waves and multiple reflections are regarded as noise and should be eliminated in the data processing. However, they are considered as useful waves for another type of seismic processing (e.g., MASW: Multichannel Analysis of Surface Waves), which are typically otherwise removed from seismic data.

Generally, the noise (coherent and incoherent) introduces uncertainty as it masks the refraction and direct arrival events. Therefore, its presence in the seismic records declines the efficiency of the automatic picking algorithms, particularly in the case of weak arrivals.

Improving the quality of signal embedded in background noise is an important task in seismic data processing. The quality of the seismic section can be improved by means of different methods. Stacking as one of the three primary stages (deconvolution, stacking, and migration) in seismic data processing (Yilmaz, 2001, p. 42) is a major process to enhance signal-to-noise ratio in seismic data.

The main reason for the endurance and the development of stacking is due to the fact that it is the only process capable of separating signal from noise even when they both have the same frequency. Another reason is that other major seismic data processing steps, such as velocity analysis and migration, some way or another involve the use of stacking concept.

5.2 Definitions and general concepts used in stacking

Shot gather: Recorded data are sorted in such a way that all traces have the same source position.

Common mid-point (CMP) gather: Recorded data are sorted into gather where all traces have common midpoint (i.e., the position on the surface midway between the source and the receiver) or have common depth-point (CDP) located in depth directly beneath the corresponding CMP position.

Stacking: In a conventional sense, stacking consists of the simple average of CDP traces (x_i) that contribute equally and constructively to the stacked trace (\bar{x}) as following:

$$\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t) \quad (5.1)$$

This method increases signal-to-noise ratio by factor \sqrt{N} (Mayne, 1962; Sheriff and Geldart, 1995; Yilmaz, 2001) where N is the number of CDP traces summed.

Stacking enhances the coherent signals and attenuates the random ones. But the conventional stacking is optimal only if there are no amplitude variations between the trace (correlated traces), and if the noise components are stationary (perfectly distributed random noise) (Neelamani et al., 2006). Fortunately, there are many non-linear stacking methods, where every sample entering the summation is weighted according to some suitable criterion.

Chapitre 5. Renforcement local des arrivées sismiques par double sommation dans le domaine temporel

In these cases, weighting traces appropriately before stacking can enhance the discriminatory effect of the conventional stack, where the weights can be derived according to different optimality criteria. If w_i is the weight applied to the i^{th} trace, then the weighted stacked trace is given by:

$$\bar{x}(t) = \left(\sum_{i=1}^N w_i \right)^{-1} \sum_{i=1}^N w_i \cdot x_i(t) \quad (5.2)$$

However, a more exhaustive description and historical review of CMP stacking is given by Rückemann (2012) and Rashed (2014).

A possible approach to the non-linear summation might be the phase weighted stack (PWS) designed by Schimmel and Paulssen (1997) and further developed in Schimmel (1999) and Schimmel and Gallart (2003). It is based on the assumption that the signal and the noise have different instantaneous phases.

Here, we propose hybrid stacking technique based on two stacking methods, called double stacking. It is designed for the roll along technique acquisition (**Figure 5.1**). This technique can be considered a suitable sources-receivers array for the processing of both refraction and reflection data (Kim and Kim, 2008; Orlovsky et al., 1998).

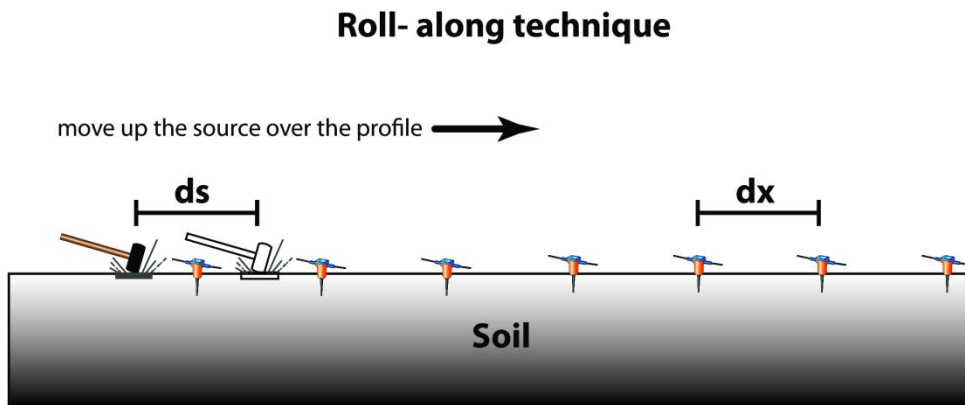


Figure 5.1. Simplified schema of roll-along technique acquisition.

5.3 First stacking using local similarity

Local correlation is a measure of the similarity between two signals (Liu et al., 2009). Based on this property, Liu et al. (2009) proposed a nonlinear stacking, where its weights represent a measure of the local correlation between each trace of the CMP and a conventional stacking reference trace. All mathematical equations governing this technique stacking were explained in the **Annexe B.1**.

If a trace is locally very similar to the reference trace, the local correlation coefficient is high and should be given a positive contribution to stacked trace. Lower weight will be corresponding with a trace that locally differs from the reference trace. This stack

was developed for overcoming the problems relating to CMP stacking in seismic reflection. The idea is very appealing because it overcomes many problems such as signal misalignment. Here, we adapted it to seismic refraction.

By analogy to CMP gathers, we apply this stacking method on the common-offset gather of a few adjacent shots along the profile. Furthermore, instead of using the linear stacking result as reference trace, the latter was built by another non-linear stacking, as described in the next section.

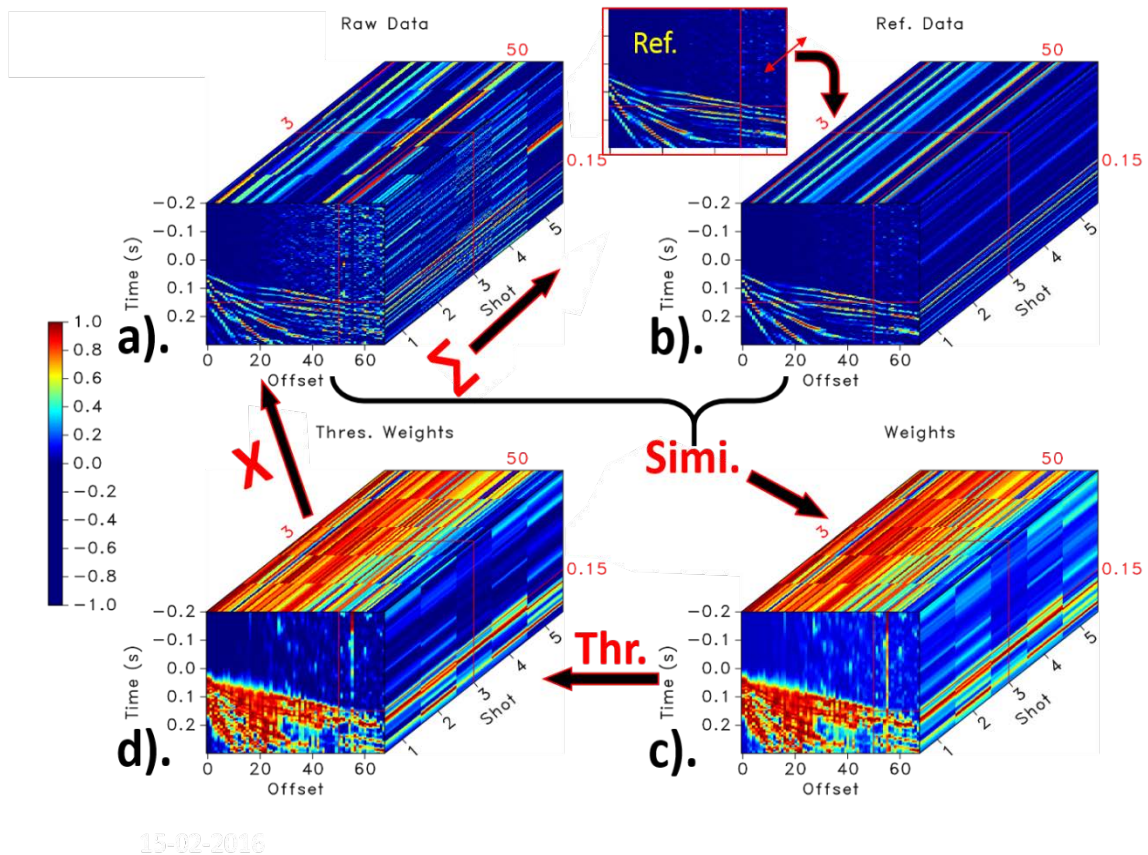


Figure 5.2. Pictorial description of local similarity stacking: (a). Cube of 5 adjacent real shot gathers. (b). Cube of repeated reference trace. (c). Local similarity coefficients cube of the 5 shot gathers. (d). Weights used in local similarity weighted stacking (local similarity cube after applying a thresholding).

5.4 Reference trace

The quality of the final stacked data is governed by the quality of the reference trace (Sanchis and Hanssen, 2011). Sanchis and Hanssen (2011) proposed an enhanced local correlation stacking method to overcome the reference trace quality problem. In their study, they used Kalman filter and S/N ratio-based stack to produce the reference trace and also proposed a new weight normalization scheme to be used with the local correlation weighting functions proposed by Liu et al. (2009). Although the performance of the Kalman reference method for the most of noise conditions (Sanchis

and Hanssen, 2011), it has a longer processing time by comparing with the other methods (Rashed, 2014).

As for coherency, the most important parameter is the ratio between receiver spacing and wavelength of the seismic waves. In shallow seismic prospection, where the geophone spacing is in order of meters, traces from neighboring geophones can be safely considered coherent in offsets larger than several tens of meters (Kim et al., 2012). Kim and Kim (2008) proved that P-wave component can be emphasized by the proper arrangement of source positions (called P-beam shot gathers). Consequently, we made the simplifying assumption that the common-offset gather traces within a short spatial window are approximately in-phase, where the linear (i.e. refraction travel-times) moveout correction can be omitted.

Admittedly, this assumption is not entirely exact with high complex subsurface or with complicated topography surface. But, the refracted waves at far offset traces, which are often corresponding to the deepest high velocities structures, tend to be horizontally aligned or at least in-phase.

Motivated by these facts, we introduce a new reference trace based on phase-weighted stack (PWS) (Schimmel and Paulssen, 1997) for Liu' method (Liu et al., 2009). The PWS is a nonlinear stacking technique where each sample of a linear stack is weighted by an amplitude-unbiased coherence measure (see the governing equations in **Annexe B.2**). It has been successful in detection weak coherent signal in seismograms (e.g. Frassetto et al., 2010). The advantage of PWS is that the coherency function is explicitly independent of amplitude (Schimmel and Paulssen, 1997). However, the disadvantage of this stack is that it does not necessarily preserve the waveform. Also, a stack is dominated by the most coherent frequencies present in the signal.

This type of stack might produce an optimal pilot traces that honor data phase and in second degree the frequency. For these reasons, we choose it to be the reference trace for the stacking based on the local similarity stacking method. The whole flow of the latter is summarized in the **Figure 5.2**.

Figure 5.3 shows an example of local similarity stacking on real data over 5 shot points. It is remarkable that all random noise was removed out and the correlated signals (e.g., refracted waves) are enhanced. On other hand, the refracted signals on the far offset remain far from being improved.

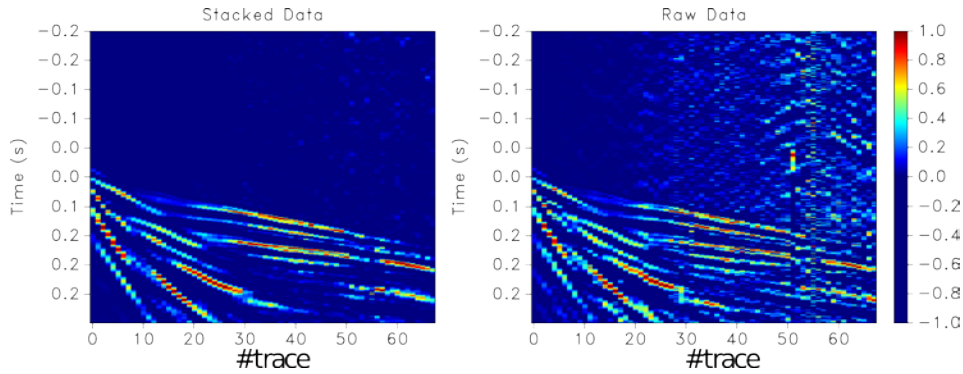


Figure 5.3. Example of enhanced local-similarity stacking with using PWS-trace as the reference trace: **(Left)**. Stacked shot gather. **(Right)**. Raw shot gather.

5.5 Second stacking using cross correlation

While the alignment of CMP-reflection traces is rather simple via dynamic correction or Normal Moveout (NMO) correction, this process is not easy in refraction gathers. This is due to the fact that wavefields of multi-refractors are not separable on the seismic record like as for the reflector signals. The similarity of waveform in adjacent traces leads naturally to the use of correlation methods. But this similarity is also a function of distance between the traces. For enough spacing, the event coherence is degraded due to the waveform distortion induced by multi-pathing appears as uncorrelated noise (VanDecar and Crosson, 1990). Fortunately, in shallow seismic surveys the spacing receivers is small. Therefore, waveforms remain suitably coherent for cross-correlation analysis. Based on aforementioned facts, if we shift neighboring records accordingly to the travel time curve of the desired waves (here refracted waves) and sum them, the desired wave should be amplified whilst the others weakened. In other words, this stacking procedure can be achieved by doing a slowness correction (time-shift corrections) within short time-spatial window, moving along the shot gather followed by linear stacking (**Figure 5.4**).

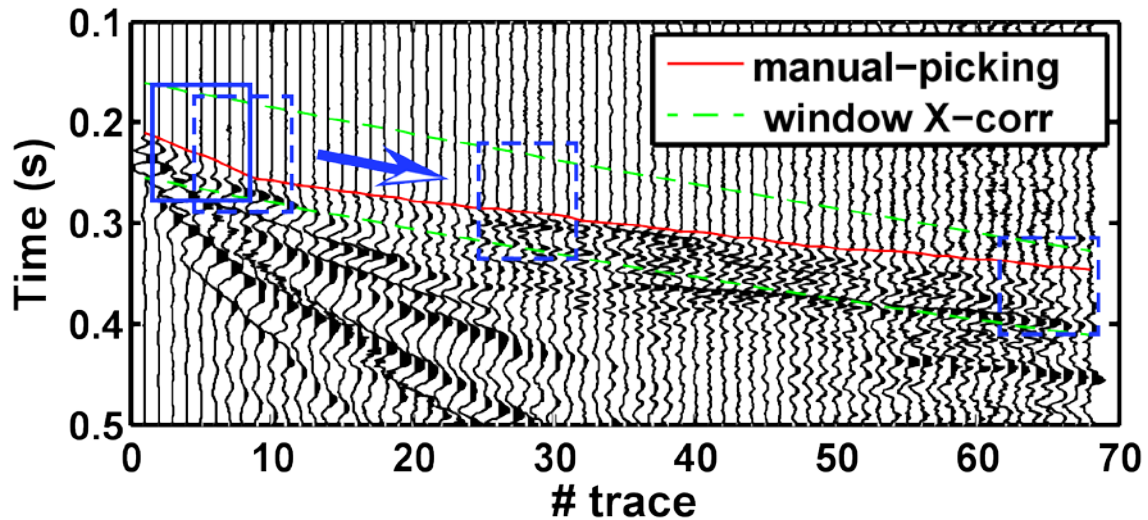


Figure 5.4. The second stacking using cross-correlation in limited time-spatial window moving (blue rectangle) over the shot gather.

A simple and effective method of obtaining the time lag between two traces containing similar waveforms is to locate the offset of the maximum of their cross-correlation function. For more than two traces, the time-shifts can be calculated by choosing one of them as standard “pilot trace” and then cross-correlate it with the remaining traces. This technique may be less robust due to the dependence on pilot trace and even interference of random noise (VanDecar and Crosson, 1990). Cycle skipping is also another potential problem when locating the cross correlation maximum between signals pairs (Zelt et al., 1987).

Recently, in order to reduce this sensitivity, Lou et al. (2013) proposed to evaluate the delay times by iterative process of an initial times within limited time window. The basic idea behind this approach is that if all stacking traces are aligned with no relative shift, then the resulting stacked trace has the highest power. They cross-correlate each trace against a stacked beam trace. The traces are adjusted according to the new the resultant relative delay times. The process is repeated until the relative shift times corrections converge to zeros (i.e., the difference between the two last stacked beam traces is close to zero). The simplicity of the first approach makes us to choose it for time shift estimation in our moving stacking approach. This stacking technique is applied on the shot gathers resulting from the first stacking. The whole chart flow of the developed stacking technique is illustrated on **Figure 5.5**.

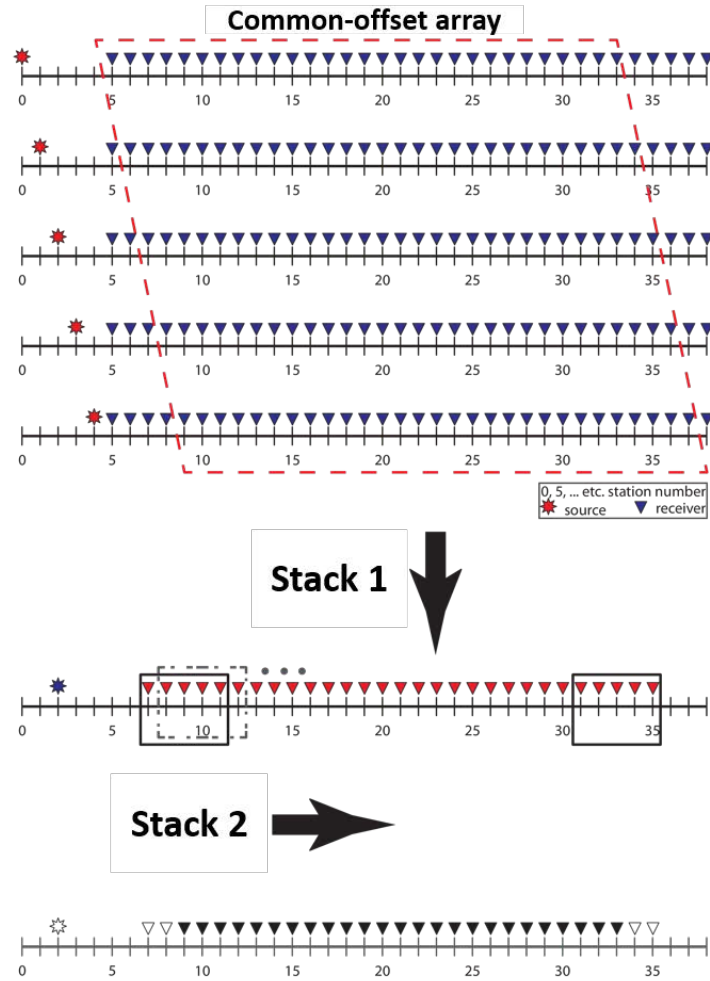


Figure 5.5. Array chart describes the double stacking flow on conventional roll-along acquisition technique.

However, with noisy data by high frequencies the cross-correlation analysis loses its efficiency for estimation delay time between the traces, due to the limited coherence of short-period waveform. In these cases, de-noising the traces before the second stacking is an inevitable process.

5.6 Denoising by wavelet transform (WT)

Roughly speaking, the goal of the wavelet transform is to see both the *forest* and the *trees* (Graps, 1995). “If we look at a signal with a large window, we would notice gross features (the *forest*). Similarity, if we look at a signal with a small window, we would notice small features (the *trees*)”

The use of WT has proved its efficiency in the signal theory, particularly in signal denoising, since 30 years ago. Wavelet capability to give detail time-frequency information is the main reason for this investigation, where the wavelets can be localized in both frequency/scale (via dilations) and in time (via translations). This

Chapitre 5. Renforcement local des arrivées sismiques par double sommation dans le domaine temporel

property promises a possibility for better discrimination between the noise and the real data. Indeed, the WT concentrates the energies of most signals of interest in a few coefficients, while the power of noise is uniformly spread throughout all the coefficients (Cai and Silverman, 2001; Donoho and Johnstone, 1994; Liu, 2013).

The advantage of using a wavelet filter instead of a conventional band pass filter, at first, is the lack of the requirement of stationarity of the signal, and is the adaptability of the analysis function (mother wavelet). Indeed, there exist many wavelet bases, among which the most suitable one can be chosen for the problem at hand.

The most straightforward way of distinguishing information from noise in the wavelet domain consists of thresholding the wavelet coefficients. In other words, the denoising procedure proceeds in three steps: (i) choose a wavelet to compute the wavelet decomposition of signal at N levels; (ii) then for each level from 1 to N , select a threshold and apply soft threshold to the detail coefficients; (iii) finally, compute wavelet reconstruction based on the original approximation coefficients and the modified detail coefficients.

An important part of the time-frequency/scale denoising is how it determines if coefficient at a given frequency/scale is to be attenuated or not. In other words, the choice of threshold is a crucial for denoising process. Hard or soft thresholding functions (Donoho and Johnstone, 1994) is the most used with WT denoising.

Mathematically, the WTs are broadly divided into two classes: continuous (CWT) and discrete (DWT). The CWT can be defined as the inner product of a signal $x(t)$ with dilated and translated version of a mother wavelet $w(t)$ (Grossmann and Morlet, 1984)

$$CWT(b, a) = \langle x, w_{b,a} \rangle = \int_{-\infty}^{+\infty} x(t)w_{b,a}^*(t)dt, \quad (5.3)$$

$$w_{b,a}(t) = \frac{1}{\sqrt{a}}w\left(\frac{t-b}{a}\right),$$

where a and b are the scale and the translation parameters, respectively, and $*$ indicates the complex conjugate operation.

In some practical application, the DWT is preferred because signals are recorded as discrete time samples.

By choosing the set of parameters $\{a = 2^j, b_{j,k} = 2^j k\}$, with integers j, k , the WT is defined only at discrete “dyadic” scales (a) and discrete times (b). This way of obtaining a DWT is referred to as the *Mallat’s algorithm* or multiresolution analysis (MRA) (Mallat, 1989) in the literature (see the **Annexe B.3** for more mathematical details about DWT). The MRA separates the signal into details (D_i) at different scales ($i = 1, \dots, n$), the remaining part being a coarse approximation (A_n) (see **Figure 5.6**). The lower levels give the details corresponding to the high frequency components and the higher levels provide information of the low frequency components (**Figure 5.6**).

In this paper, we used MRA with a quadratic B-Spline mother wavelet. Since this wavelet have similar shape to the first arrival waveforms, it can be suitable for analysis the seismic trace (Anant and Dowla, 1997). The denoising procedure is performed by incorporating neighbor coefficients (Cai and Silverman, 2001) and Zerotrees of wavelet coefficients (Shapiro, 1993). This method is based on the inter- and intra-scale correlation of neighboring wavelet coefficients to select those that deviate from baseline noise. The first approach is to decide whether each coefficient should be kept or not based not only on its value but also on the value of its closest neighbors in the same scale. The second one is to also use information from the decomposition at different scales. Combining these two approaches make the denoising fully automatic without needing to setup a threshold. This approach was implemented for analyzing clinical diagnostic signals (e.g., Liu et al., 2015). More details with corresponding mathematical equations can be found in the **Annexe B.3**.

Figure 5.6 demonstrates the main steps of denoising by WT on real seismic trace.

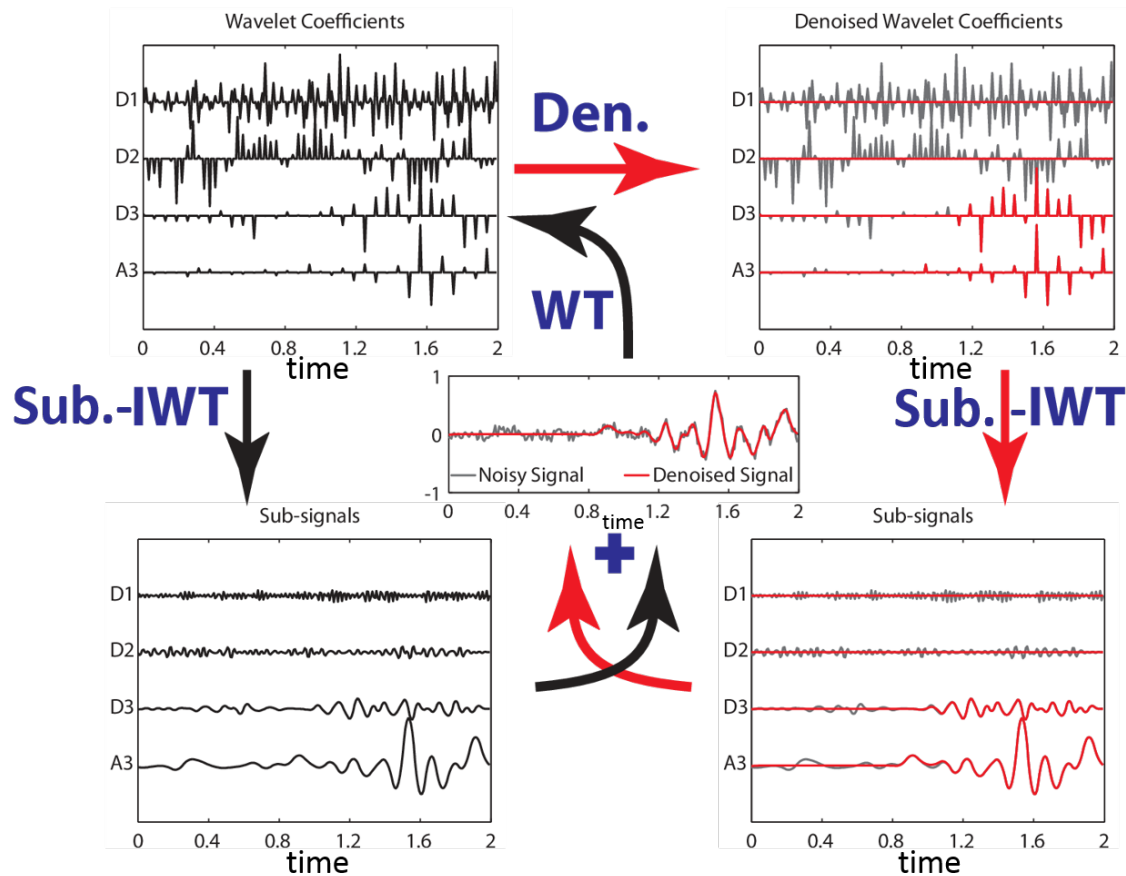


Figure 5.6. De-noising by wavelet transform (precisely with multiresolution decomposition): **WT** is the dyadic wavelet transform for N level (here, 3 levels), then the results are the details (D1, D2 and D3) and the approximation A3. **Sub-IWT** is the inverse wavelet transform of each level. **Den.** indicates to the denoising or thresholding process. The flow with black arrows shows the reversibility property of WT and generating sub-band signals from the WT coefficients. The flow with red

arrows is the same of the black ones, but with denoising procedure.

5.7 Applications

To quantify double stacking performance, we applied the automatic picking algorithm developed in the previous chapters on a real seismic profile. The spacing between geophones (72 geophones) was 4 m and the shot spacing was also 4 m. The energy of the source, an accelerated 40 kg weight dropped onto a steel plate, was stacked 5-8 times. Shot gathers from two different shot positions along the profile have been selected to validate the efficiency of double stacking before automatic picking.

The first record (**Figure 5.7**) is characterized by good signal to noise ratio of the first arrival signals for all offsets. In contrast, the second record (**Figure 5.8**) is characterized by variable signal to noise ratio, where the first arrival waves are completely embedded by the noise at the far offsets.

- Double stacking parameters: the first stack was performed over 5 adjacent shot points, the phase weights used in PWS are smoothed within a time window of 25 ms. They are also emphasized by an arbitrary exponent (2) defining the strength of weighting. The similarity weights are smoothed by a threshold of 80 %. The second stack was performed by a moving window of 5 traces along the resulting shot gather. The delay time corrections between the group traces are calculated by cross-correlation within time window of three times the dominated period of the refracted waves. This window is centered on theoretical travel time curve assessed from the similarity weights.
- Automatic picking parameters: the only parameter needed for the adaptive picking algorithm (Khalaf et al., 2015) is the dominant period of first arrivals (here, we estimated it to be 75 samples).

The root mean square (RMS) error of the differences between the picks derived automatically and those picked manually on these shot gathers is used for estimating the feasibility of the double stacking approach. To quantify the improvement, the RMS was calculated in variable length spatial-windows along the offset according to the quality of the first arrival signals.

Figure 5.7 presents the results of the automatic picking of shot gathers before and after applying the double stacking (only the raw data presented). These data have relatively good quality so that refracted event can be recognized over the whole offset range. Thus, the accuracy of automatic picking results is slightly improved by applying the double stacking. Whereas, the results on **Figure 5.8** clearly illustrates the good performance of enhancement based on double stacking.

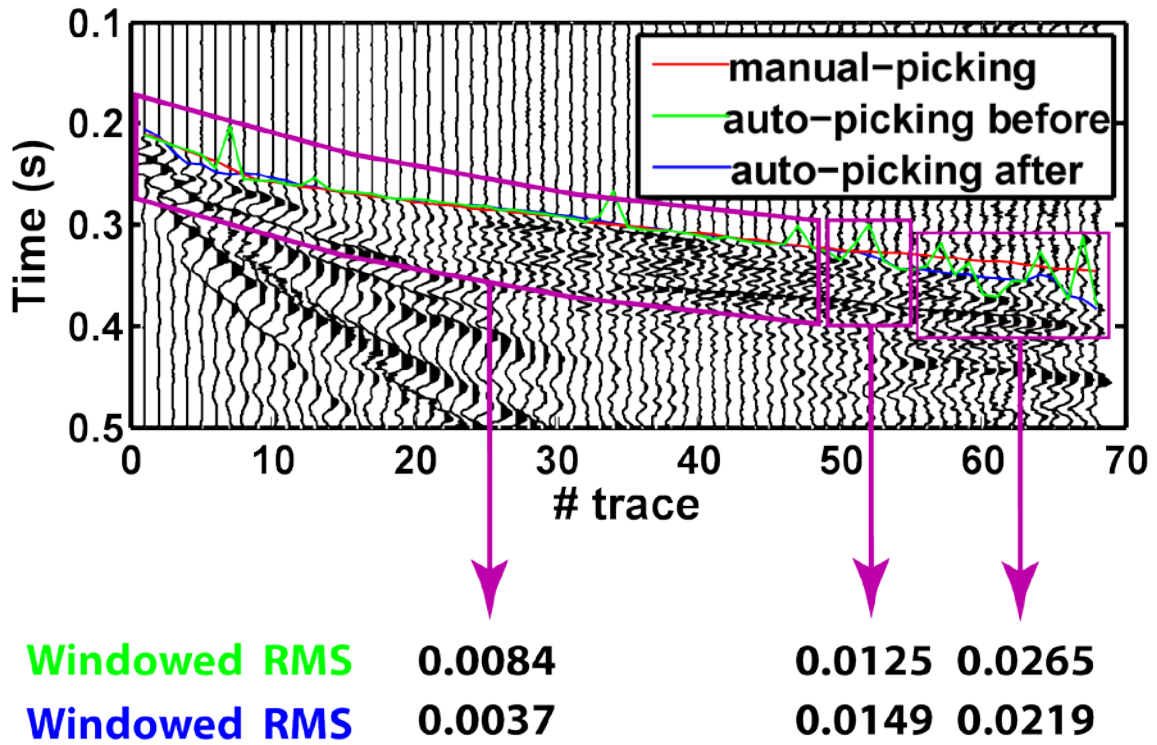


Figure 5.7. Comparison of automatic picking before and after applying double stacking on good quality data.

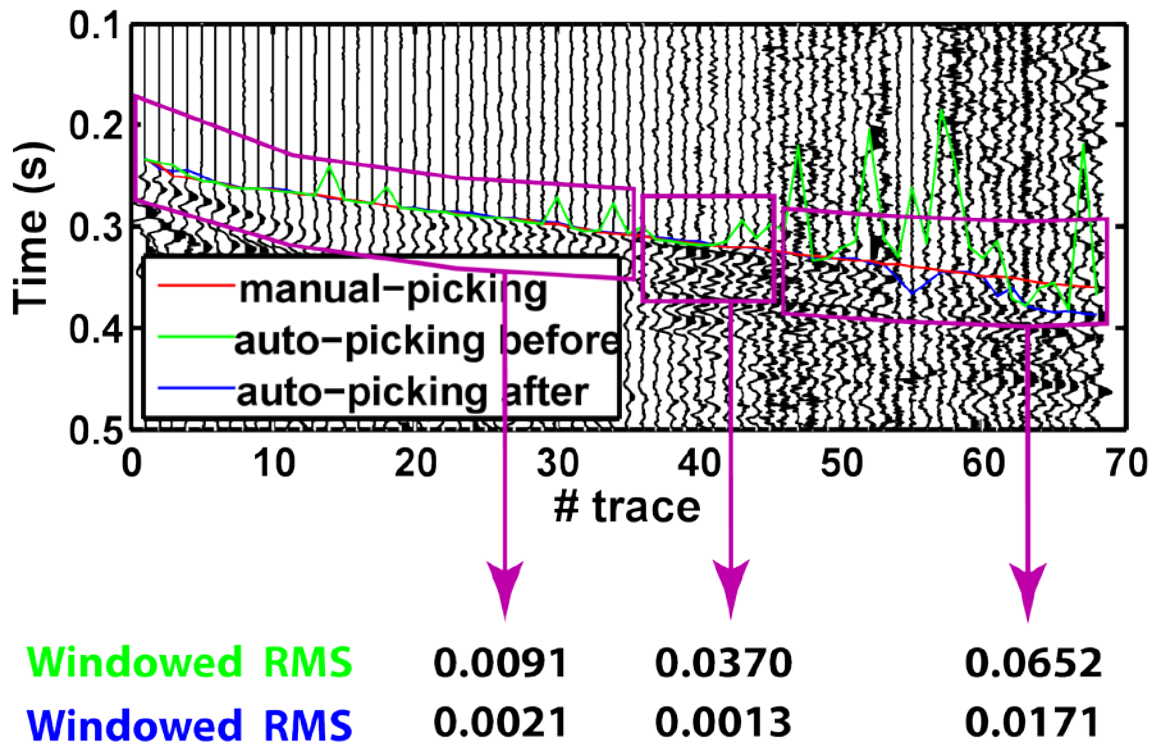


Figure 5.8. Comparison of automatic picking before and after applying double stacking on bad quality data.

5.8 Conclusions

In this paper, we have proposed and demonstrated a new stacking approach that combine two types of stacking. Exploiting the data redundancy was resulting in two stacks achieving over offset, based on the local similarity between groups of traces from adjacent shot points (stack1), and within shot gather itself (stack2). The two stacks are based on the assumption that the shape and the phase of mean features in the stacked trace do not change significantly from trace to another. The denoising by WT has an important role in the second stack for high level of noise.

The ability to emphasize the embedded events by applying the double stacking has been shown on real data. Furthermore, these results confirm the double stacking as a valuable alternative for improving the quality of the seismic gathers, particularly for applications required a good visibility of certain features, such as automatic picking. The automatic picking results were improved by about 3.8 times against those without double stacking.

However, this method has the major drawback of not preserving physically meaningful amplitudes because of the weight normalization. But for automatic picking, that does not pose any problems.

To alleviate the reference trace (resulting from PWS) dependence on similarity and misalignment traces, we suggest aligning the common offset traces via cross-correction before applying the PWS. In this way, the double stacking should be efficient no matter how the sub-surface structure complexity is being.

To validate this proposed approach, we suggest applying it to numerical model or on real data with high complex velocity structures or at least with complicated topographic surface profile, and further comparing it with existing ones.

Although the potential pitfalls of this sophisticated stacking technique, it is turned out to be very promising if we take in consideration the trace alignment before each stack.

Conclusions générales et perspectives

Ce travail de thèse s'inscrit dans une perspective diachronique où demeure une problématique relative à la précision du pointé automatique des données sismiques. En effet, l'identification correcte et précise des temps des arrivées sismiques est très importante dans divers domaines (par exemple en sismologie, sismique appliquée, micro-sismicité).

Au cours des années, de nombreuses approches différentes du pointé automatique ont été développées. Mais le développement d'un algorithme robuste, précis et satisfaisant pour tous les scénarii (type de source et de récepteur, niveau de bruit, domaine d'application, d'entre autres) n'est pas encore disponible. La diversité de type de données et l'ambition d'avoir un pointé précis ont conduit à une grande variété des algorithmes du pointé automatique. J'ai présenté une revue historique et non exhaustive de la plupart des techniques du pointé automatique citées dans la littérature (cf. **Chapitre 2**).

Comme la tomographie sismique en réfraction nécessite de connaître précisément les temps de premières arrivées, elle a joué un rôle moteur, en parallèle avec la sismologie globale, pour motiver le développement des algorithmes de plus en plus sophistiqués.

L'objectif principal de la thèse fut de mettre en place un algorithme de pointé précis pour la tomographie sismique, notamment pour la sismique proche surface, et a comporté trois axes. Les conclusions que l'on peut tirer à partir de ces axes sont :

Axe 1. Pointé automatique précis : nous avons développé un algorithme adaptatif, conçu pour les données sismiques de proche surface. Ce travail nous a amené à intégrer trois propriétés (l'énergie, la « gaussianité », et la stationnarité) indiquant la première arrivée, à la fois pour le pointé précis et l'estimation automatique de l'incertitude des pointés acquis. Partant du principe que la précision de pointé est, avant tout, limitée par la présence de bruits et par la qualité de l'*impetus* des premières arrivées, la stratégie mise en place nous a permis de valoriser l'apport de l'association de plusieurs algorithmes. L'algorithme implémenté est simple à mettre en œuvre et ne demande pas de grandes performances informatiques, grâce à l'intégration, en série, des résultats de chaque sous-algorithme (MNW, HOS, et AIC). En fait, deux algorithmes ont été développés ; le premier est conçu pour les traces individuelles (ou les sismogrammes). Ses applications sont limitées pour les données de sismique active, mais il peut être utilisé dans les applications sismologiques. La deuxième est une extension du premier et adapté pour les enregistrements de sismique active. Le diagramme sur la figure suivant résume l'organigramme de l'implémentation de cet ensemble.

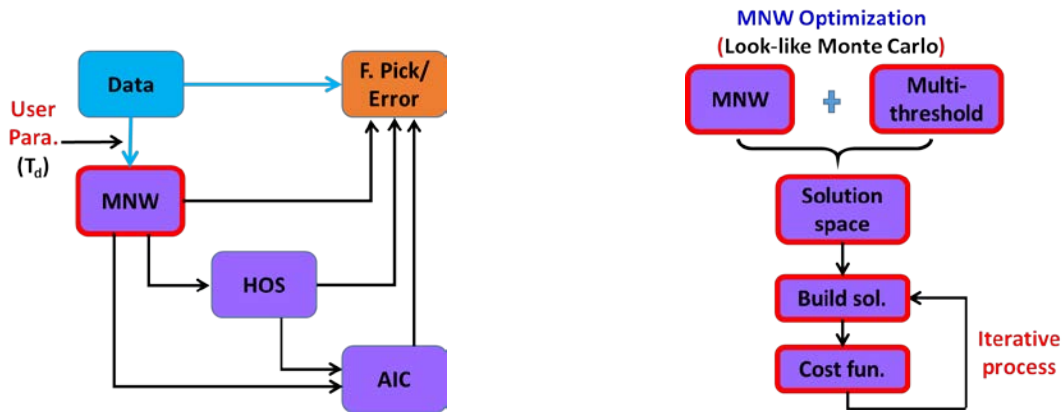


Diagramme des algorithmes implémentés: **(gauche)**. Trace-par-trace. **(droite)**.
extension pour les tirs sismiques.

La précision et la fiabilité de cet algorithme ont été évaluées par la comparaison des résultats à ceux issus d'un pointé manuel et à d'autres pointeurs automatiques sur des données synthétiques (cf. **Chapitre 3**) et réelles (cf. **Chapitre 4**), présentant des défis spéciaux. En plus de permettre un pointé automatique d'une qualité quasi-comparable à un pointé manuel, l'algorithme nous fournit aussi automatiquement une estimation sur l'incertitude des pointés grâce de la redondance de résultats des trois méthodes. Cependant, la faiblesse de l'algorithme est apparue aux offsets lointains. Dans ces cas, les incertitudes estimées ont une importance du même ordre que du pointé automatique.

Axe 2. Edition automatique des pointés acquis : en nous fondant sur l'idée que « mieux vaut ne pointer aucune trace que de la pointer incorrectement », nous avons mis au point une relation empirique pour estimer l'incertitude sur la première arrivée via un algorithme adaptatif. Cette relation permet d'attribuer à chaque trace pointée, un poids dépendant de la qualité de détermination du temps d'arrivée, et par suite, d'écartier les mauvais pointés avant les traitements en aval (par exemple, avant l'inversion tomographique) (cf. **Chapitre 4**).

Axe 3. Renforcement local des signaux des premières arrivées faibles : Le rapport signal sur bruit des événements de premières arrivées se dégradant rapidement avec l'offset, ceux-ci deviennent difficiles à pointer, notamment par le pointé automatique. Suite aux limitations du pointeur développé, nous avons mis en place une technique de sommation afin d'améliorer cette détection. Cette technique s'appuie sur la redondance des données sismiques, notamment celles acquises par la technique du *roll-along*. Elle combine deux types de sommation fondés sur le principe clé de la ressemblance locale entre les traces sommées. Il est préférable d'effectuer préalablement un dé-bruitage par transformée en ondelettes lorsque les données sont très bruitées, car la sommation basée sur l'inter-corrélation est moins efficace aux fréquences plus élevées. La performance de double sommation a été validée sur de jeu de données réelles (cf. **Chapitre 5**).

Perspectives:

Sur le pointé automatique : La précision peut encore être améliorée en intégrant des attributs de polarisation d'ondes, ce qui serait notamment utile lorsque le début du signal est submergé par le bruit, et pourrait apporter une information supplémentaire sur la présence des premières arrivées.

Sur la double sommation : L'efficacité de la double sommation pourrait être optimale si le désalignement des traces a été pris en compte, avant, la sommation. C'est-à-dire, il faut s'affranchir de toutes les hypothèses sur l'homogénéité du sous-sol par alignement des traces via l'inter-corrélation.

Liste des notations

2D	Bidimensionnelle
3D	Tridimensionnelle
AAA	Automatic Adaptive Algorithm
AE	Analyse de l'Energie
AIC	Akaike Information Criteria
API-S/K	Phase Arrival Identification-Skewness/Kurtosis
AR	Autorégressives
ARSB	Rapport Signal sur Bruit basé sur l'Amplitude
ATA	After Term Average
AW	Akaike weights
BPZ	Beginning of the Potential Zone
BTA	Before Term Average
CDP	Common Depth-Point
CF	Characteristic Function
CMP	Points Milieu Communs / Common Mid-Points
CWT	Continuous Wavelet Transform
DTA	Delayed Term Average
DWT	Transformée en Ondelettes Discrète/Discrete Wavelet Transform
ER	Energy Ratio
EWI	Early arrival Wavefront Inversion
EWT	Early arrival Waveform Tomography
FC	Fonction Caractéristique
FD	Finite-Difference
FDM	Fractal-Dimension Method
FRSB	Rapport Signal sur Bruit basé sur la Fréquence
FWI	Full Wave Inversion
HOS	Higher Order Statistics
IWT	Invers Wavelet Transform
JER	Joint Energy Ratio
LTA	Long Term Average
MASW	Multichannel Analysis of Surface Waves
MCM	Modified Coppens' Method
MER	Modified Energy Ratio
MNW	Multi-Nested Windows
MRA	Multiresolution Analysis
NMO	Normal Moveout
NW	Noise Window
PER	Peak Eigenvalue Ratio
PML	Perfected Matched Layer

PWS	Phase Weighted Stack
RMS	Root Mean Square
RN	moyens de Réseaux de Neurones
RSB	Rapport Signal sur Bruit
RTF	Représentation Temps-Fréquence
S/N	Signal-to-Noise
SCF	Smoothed Characteristic Function
SNR	Signal-to-Noise Ratio
STA	Short Term Average
SU	Seismic Unix
SW	Signal Window
THR	Threshold
WET	Wave Eikonal Traveltime
WT	Wavelet Transform / Transformée en Ondelette

Bibliographie

- Ahmadi, M., Quian Quiroga, R., 2013. Automatic denoising of single-trial evoked potentials. *Neuroimage* 66, 672–80. doi:10.1016/j.neuroimage.2012.10.062
- Ait Laasri, E.H., Akhouayri, E.S., Agliz, D., Atmani, A., 2014. Automatic detection and picking of P-wave arrival in locally stationary noise using cross-correlation. *Digit. Signal Process. A Rev. J.* 26, 87–100. doi:10.1016/j.dsp.2013.12.009
- Akaike, H., 1974. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Stat. Math.* 26, 363–387. doi:10.1007/BF02479833
- Akaike, H., 1971. Autoregressive model fitting for control. *Ann. Inst. Stat. Math.* 23, 163–180. doi:10.1007/BF02479221
- Akram, J., 2014. Downhole microseismic monitoring: processing, algorithms and error analysis. Calgary University, Alberta, Canada.
- Akram, J., Eaton, D., 2012. Adaptive microseismic event detection and automatic time picking, in: *GeoConvention 2012*. pp. 1–5.
- Akram, J., Eaton, D.W., Consortium, M.I., 2014. Refinement of arrival-time picks using an iterative, cross-correlation based workflow, in: *GeoConvention 2014: FOCUS*. pp. 1–6.
- Aldersons, F., 2004. Toward a Three-Dimensional Crustal Structure of the Dead Sea region from Local Earthquake Tomography. PhD thesis. Tel-Aviv University.
- Allam, A.A., Ben-Zion, Y., Peng, Z., 2014. Seismic Imaging of a Bimaterial Interface Along the Hayward Fault, CA, with Fault Zone Head Waves and Direct P Arrivals. *Pure Appl. Geophys.* 171, 2993–3011. doi:10.1007/s00024-014-0784-0
- Allen, R., 1982. Automatic phase pickers: Their present use and future prospects. *Bull. Seismol. Soc. Am.* 72, S225–242.
- Allen, R., 1978. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* 68, 1521–1532.
- Anant, K.S., Dowla, F.U., 1997. Wavelet transform methods for phase identification in three-component seismograms. *Bull. Seismol. Soc. Am.* 87, 1598–1612.
- Aster, R.C., Shearer, P.M., Berger, J., 1990. Quantitative measurements of shear wave polarizations at the Anza Seismic Network, southern California: Implications for shear wave splitting and earthquake prediction. *J. Geophys. Res.* 95, 12449. doi:10.1029/JB095iB08p12449
- Baer, M., Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events. *Bull. Seismol. Soc. Am.* 77, 1437–1445.

- Baillard, C., Crawford, W.C., Ballu, V., Hibert, C., Mangeney, A., 2014. An Automatic Kurtosis-Based P- and S-Phase Picker Designed for Local Seismic Networks. *Bull. Seismol. Soc. Am.* 104, 394–409. doi:10.1785/0120120347
- Berger, J., Sax, R.L., 1981. *Seismic Detectors: The State-of-the-Art*, Technical report, SSR-R-80-4588.
- Bièvre, G., Jongmans, D., Winiarski, T., Zumbo, V., 2012. Application of geophysical measurements for assessing the role of fissures in water infiltration within a clay landslide (Trièves area, French Alps). *Hydrol. Process.* 26, 2128–2142. doi:10.1002/hyp.7986
- Billings, S.D., Sambridge, M.S., Kennett, B.L.N., 1994. Errors in hypocenter location: Picking, model, and magnitude dependence. *Bull. Seismol. Soc. Am.* 84, 1978–1990.
- Blias, E., 2012. Optimization approach to automatic first arrival picking for three-component three-dimensional vertical seismic profiling data. *Geophys. Prospect.* 60, 1024–1029. doi:10.1111/j.1365-2478.2011.01014.x
- Bogiatzis, P., Ishii, M., 2015. Continuous Wavelet Decomposition Algorithms for Automatic Detection of Compressional- and Shear-Wave Arrival Times. *Bull. Seismol. Soc. Am.* 105, 1628–1641. doi:10.1785/0120140267
- Boschetti, F., Dentith, M.D., List, R.D., 1996. A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics* 61, 1095–1102. doi:10.1190/1.1444030
- Cai, T., Silverman, B., 2001. Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā Indian J. Stat.* 2.
- Carcione, J.M., Herman, G.C., ten Kroode, A.P.E., 2002. Seismic modeling. *Geophysics* 67, 1304–1325. doi:10.1190/1.1500393
- Cerveny, V., 2005. *Seismic Ray Theory*. Cambridge University Press.
- Chen, Z., Stewart, R., 2005. Multi-window algorithm for detecting seismic first arrivals, in: *CSEG Annual Convention*. pp. 1–4.
- Cichowicz, A., 1993. An automatic S-phase picker. *Bull. Seismol. Soc. Am.* 83, 180–189.
- Cohen, A., Daubechies, I., Feauveau, J.-C., 1992. Biorthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* 45, 485–560. doi:10.1002/cpa.3160450502
- Cohen, J.K., Stockwell, J.J.W., 2012. CWP/SU: Seismic Un*x Release No. 43R1: an open source software package for seismic research and processing. *Cent. Wave Phenomena, Color. Sch. Mines*.
- Coppens, F., 1985. First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophys. Prospect.* 33, 1212–1231.

doi:10.1111/j.1365-2478.1985.tb01360.x

- Cuenot, O., 2003. Les algorithmes de detection automatique d'ondes sismiques. LTA 1.
- Dal Moro, G., Ponta, R., Mauro, R., 2015. Unconventional optimized surface wave acquisition and analysis: Comparative tests in a perilagoon area. *J. Appl. Geophys.* 114, 158–167. doi:10.1016/j.jappgeo.2014.12.016
- Deidda, G.P., Battaglia, E., Heilmann, Z., 2012. Common-reflection-surface imaging of shallow and ultrashallow reflectors. *Geophysics* 77, B177–B185. doi:10.1190/geo2011-0401.1
- Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P., Chiarabba, C., Giardini, D., 2006. Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity. *Geophys. J. Int.* 165, 121–134. doi:10.1111/j.1365-246X.2005.02799.x
- Diehl, T., 2008. 3-D Seismic Velocity Models of the Alpine Crust from Local Earthquake Tomography. University of Karlsruhe, Germany.
- Diehl, T., Deichmann, N., Kissling, E., Husen, S., 2009. Automatic S-wave picker for local earthquake tomography. *Bull. Seismol. Soc. Am.* 99, 1906–1920. doi:10.1785/0120080019
- Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455. doi:10.1093/biomet/81.3.425
- Douglas, A., 1997. Bandpass filtering to reduce noise on seismograms: Is there a better way? *Bull. Seismol. Soc. Am.* 87, 770–777.
- Earle, P.S., Shearer, P.M., 1994. Characterization of Global Seismograms Using an Automatic-Picking Algorithm. *Bull. Seismol. Soc. Am.* 84, 366–376.
- Fomel, S., 2007a. Local seismic attributes. *Geophysics* 72, 29–33. doi:10.1190/1.2437573
- Fomel, S., 2007b. Shaping regularization in geophysical-estimation problems. *Geophysics* 72, R29–R36. doi:10.1190/1.2433716
- Francese, R.G., Hajnal, Z., Prugger, A., 2002. High-resolution images of shallow aquifers—A challenge in near-surface seismology. *Geophysics* 67, 177–187. doi:10.1190/1.1451490
- Frassetto, a., Zandt, G., Gilbert, H., Owens, T.J., Jones, C.H., 2010. Improved imaging with phase-weighted common conversion point stacks of receiver functions. *Geophys. J. Int.* 182, 368–374. doi:10.1111/j.1365-246X.2010.04617.x
- Futterman, W.I., 1962. Dispersive body waves. *J. Geophys. Res.* 67, 5279–5291. doi:10.1029/JZ067i013p05279
- Gaci, S., 2014. The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans. Geosci. Remote Sens.* 52, 4558–

4563. doi:10.1109/TGRS.2013.2282422
- Galibert, P.-Y., Valois, R., Mendes, M., Guérin, R., 2014. Seismic study of the low-permeability volume in southern France karst systems. *Geophysics* 79, EN1–EN13. doi:10.1190/geo2013-0120.1
- Gelchinsky, B., Shtivelman, V., 1983. Automatic picking of first arrivals and parameterization of traveltimes curves*. *Geophys. Prospect.* 31, 915–928. doi:10.1111/j.1365-2478.1983.tb01097.x
- Gentili, S., Michelini, A., 2006. Automatic picking of P and S phases using a neural tree. *J. Seismol.* 10, 39–63. doi:10.1007/s10950-006-2296-6
- Geoltrain, S., Brac, J., 1993. Can we image complex structures with first-arrival traveltimes? *Geophysics* 58, 564–575. doi:10.1190/1.1443439
- Giannakis, G.B., Tsatsanis, M.K., 1994. Time-domain tests for Gaussianity and time-reversibility. *IEEE Trans. Signal Process.* 42, 3460–3472. doi:10.1109/78.340780
- Goforth, T., Herrin, E., 1981. An automatic seismic signal detection algorithm based on the Walsh transform. *Bull. Seismol. Soc. Am.* 71, 1351–1360.
- Graps, A., 1995. An introduction to wavelets. *IEEE Comput. Sci. Eng.* 2, 50–61. doi:10.1109/99.388960
- Grossmann, A., Morlet, J., 1984. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. Math. Anal.* 15, 723–736. doi:10.1137/0515056
- Han, L., 2010. Microseismic monitoring and hypocenter location. Calgary University, Alberta, Canada.
- Han, L., Wong, J., Bancroft, J., 2010. Time picking on noisy microseismograms. *Proc. GeoCanada 2010 Conv.*
- Han, L., Wong, J., Bancroft, J., 2009. Time picking and random noise reduction on microseismic data: CREWES Research Report.
- Hanafy, S.M., Yu, H., Abdullah, K., Kaust, T., 2013. Early arrival waveform inversion of shallow seismic land data, in: 2013 SEG Annual Meeting. Society of Exploration Geophysicists, pp. 1738–1742.
- Hatherly, P.J., 1982. A computer method for determining seismic first arrival times. *Geophysics* 47, 1431–1436. doi:10.1190/1.1441291
- Hildyard, M.W., Nippres, S.E.J., Rietbrock, A., 2008. Event detection and phase picking using a time-domain estimate of predominate period T_{pd} . *Bull. Seismol. Soc. Am.* 98, 3025–3032. doi:10.1785/0120070272
- Ivanov, J., Miller, R.D., 2004. Semi-Automatic Picking Of First Arrivals Through Cross Correlation Using Spline Interpolation Applied To Near-Surface Seismic

- Surveys, in: 17th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems.
- Jiao, L., Moon, W.M., 2000. Detection of seismic refraction signals using a variance fractal dimension technique. *Geophysics* 65, 286–292. doi:10.1190/1.1444719
- Karamzadeh, N., Javan Doloei, G., Reza, A.M., 2013. Automatic Earthquake Signal Onset Picking Based on the Continuous Wavelet Transform. *IEEE Trans. Geosci. Remote Sens.* 51, 2666–2674. doi:10.1109/TGRS.2012.2213824
- Khalaf, A., Camerlynck, C.M., Schneider, A.C., Florsch, N., 2015. Development of an Adaptive Multi-Method Algorithm for Automatic Picking of First Arrival Times: Application to Near Surface Seismic Data, in: AGU Fall Meeting. AGU, San Francisco, California, USA.
- Kim, H.-S., Kim, J.-Y., 2008. High-resolution profiling of alluvial aquifer in potential riverbank filtration site by use of combining CMP refraction and reflection seismic methods. *J. Appl. Geophys.* 66, 1–14. doi:10.1016/j.jappgeo.2008.08.003
- Kim, H.-S., Kim, W.-K., Min, D.-J., Ha, I.-S., Kim, J.-Y., Kim, Y.-S., 2012. Feasibility of the P- and S-beam methods for stability estimation of the Daegok dam in Korea. *Geosci. J.* 16, 77–84. doi:10.1007/s12303-012-0002-y
- Knödel, K., Lange, G., Voigt, H.-J., 2007. *Environmental Geology: Handbook of Field Methods and Case Studies*. Springer Science & Business Media.
- Küperkoch, L., 2010. Automated Recognition, Phase Arrival Time Estimation, and Location of Local and Regional Earthquakes. Ruhr-Universität Bochum, Germany.
- Küperkoch, L., Meier, T., Brüstle, A., Lee, J., Friederich, W., 2012. Automated determination of S-phase arrival times using autoregressive prediction: application to local and regional distances. *Geophys. J. Int.* 188, 687–702. doi:10.1111/j.1365-246X.2011.05292.x
- Küperkoch, L., Meier, T., Lee, J., Friederich, W., Working Group, E., 2010. Automated determination of P -phase arrival times at regional and local distances using higher order statistics. *Geophys. J. Int.* 181, 1159–1170. doi:10.1111/j.1365-246X.2010.04570.x
- Langet, N., Maggi, A., Michelini, A., Brenguier, F., 2014. Continuous Kurtosis-Based Migration for Seismic Event Detection and Location, with Application to Piton de la Fournaise Volcano, La Reunion. *Bull. Seismol. Soc. Am.* 104, 229–246. doi:10.1785/0120130107
- Lanz, E., Maurer, H., Green, A., 1998. Refraction tomography over a buried waste disposal site. *Geophysics* 63, 1414–1433. doi:10.1190/1.1444443
- Lawton, D.C., 1989. Computation of refraction static corrections using first-break traveltimes differences. *Geophysics* 54, 1289–1296. doi:10.1190/1.1442588

- Lehmann, B., 2007. Seismic Traveltime Tomography for Engineering and Exploration Applications. EAGE publications.
- Leonard, M., 2000. Comparison of Manual and Automatic Onset Time Picking. *Bull. Seismol. Soc. Am.* 90, 1384–1390. doi:10.1785/0120000026
- Leonard, M., Kennett, B.L.N., 1999. Multi-component autoregressive techniques for the analysis of seismograms. *Phys. Earth Planet. Inter.* 113, 247–263. doi:10.1016/S0031-9201(99)00054-0
- Li, C., Huang, L., Duric, N., Zhang, H., Rowe, C., 2009. An improved automatic time-of-flight picker for medical ultrasound tomography. *Ultrasonics* 49, 61–72. doi:10.1016/j.ultras.2008.05.005
- Liu, G., Fomel, S., Jin, L., Chen, X., 2009. Stacking seismic data using local correlation. *Geophysics* 74, V43–V48. doi:10.1190/1.3085643
- Liu, K., Fear, E., Potter, M., 2015. Antenna Aperture Localization for Arrival Time Correction Using First-Break. *Prog. Electromagn. Res. B* 62, 105–120. doi:10.2528/PIERB14121908
- Liu, X., 2013. Time-Arrival Location of Seismic P-Wave based on Wavelet Transform Modulus Maxima. *J. Multimed.* 8, 32–39. doi:10.4304/jmm.8.1.32-39
- Liu, X., Wan, H., Shang, Z., Shi, L., 2015. Automatic extracellular spike denoising using wavelet neighbor coefficients and level dependency. *Neurocomputing* 149, 1407–1414. doi:10.1016/j.neucom.2014.08.055
- Lois, A., Sokos, E., Martakis, N., Paraskevopoulos, P., Tselentis, G.-A., 2013. A new automatic S-onset detection technique: Application in local earthquake data. *Geophysics* 78, KS1–KS11. doi:10.1190/geo2012-0050.1
- Lomax, A., Satriano, C., Vassallo, M., 2012. Automatic Picker Developments and Optimization: FilterPicker--a Robust, Broadband Picker for Real-Time Seismic Monitoring and Earthquake Early Warning. *Seismol. Res. Lett.* 83, 531–540. doi:10.1785/gssrl.83.3.531
- Lou, X., van der Lee, S., Lloyd, S., 2013. AIMBAT: A Python/Matplotlib Tool for Measuring Teleseismic Arrival Times. *Seismol. Res. Lett.* 84, 85–93. doi:10.1785/0220120033
- Maeda, N., 1985. A method for reading and checking phase times in auto-processing system of seismic wave data. *Zisin = Jishin* 38, 365–380.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 674–693. doi:10.1109/34.192463
- Mallinson, I., Bharadwaj, P., Schuster, G., Jakubowicz, H., 2011. Enhanced refractor

- imaging by supervirtual interferometry. *Lead. Edge* 30, 546–550. doi:10.1190/1.3589113
- Mateeva, A., 1999. Estimating errors in picked traveltimes. Colorado School of Mines.
- Mayne, W.H., 1962. Common reflection point horizontal data stacking techniques. *Geophysics* 27, 927–938. doi:10.1190/1.1439118
- McCormack, M.D., Zaucha, D.E., Dushek, D.W., 1993. First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics* 58, 67–78. doi:10.1190/1.1443352
- Meada, N., 1985. A Method for Reading and Checking Phase Time in Auto-Processing System of Seismic Wave Data. *Zisin (Journal Seismol. Soc. Japan. 2nd ser.)* 365.
- Mikesell, T.D., van Wijk, K., Ruigrok, E., Lamb, A., Blum, T.E., 2012. A modified delay-time method for statics estimation with the virtual refraction. *Geophysics* 77, A29–A33. doi:10.1190/geo2012-0111.1
- Milsom, J., Eriksen, A., 2011. *Field Geophysics*. John Wiley & Sons.
- Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J., 1997. *Wavelet toolbox: Matlab User's Guide (R2015b)*. The MathWorks, Inc.
- Monnier, S., Camerlynck, C., Rejiba, F., 2008. Ground penetrating radar survey and stratigraphic interpretation of the Plan du Lac rock glaciers, Vanoise Massif, northern French Alps. *Permafr. Periglac. Process.* 19, 19–30. doi:10.1002/ppp.610
- Montalbetti, J.F., Kanasewich, E.R., 1970. Enhancement of Teleseismic Body Phases with a Polarization Filter. *Geophys. J. Int.* 21, 119–129. doi:10.1111/j.1365-246X.1970.tb01771.x
- Moriya, H., 2009. Spectral matrix analysis for detection of polarized wave arrivals and its application to seismic reflection studies using local earthquake data. *Earth, Planets Sp.* 61, 1287–1295. doi:10.1186/BF03352982
- Mousa, W.A., Al-Shuhail, A.A., 2012. Enhancement of first arrivals using the τ - p transform on energy-ratio seismic shot records. *Geophysics* 77, V101–V111. doi:10.1190/geo2010-0331.1
- Mousa, W.A., Al-Shuhail, A.A., Al-Lehyani, A., 2011. A new technique for first-arrival picking of refracted seismic data based on digital image segmentation. *Geophysics* 76, V79–V89. doi:10.1190/geo2010-0322.1
- Munro, K., 2005. Analysis of microseismic event picking with applications to landslide and oil-field monitoring settings. University of Calgary, Alberta, Canada.
- Neelamani, R., Dickens, T.A., Deffenbaugh, M., 2006. Stack-and-Denoise: A New Method to Stack Seismic Datasets, in: 2006 SEG Annual Meeting. Society of Exploration Geophysicists.

- Nippres, S.E.J., Rietbrock, A., Heath, a. E., 2010. Optimized automatic pickers: application to the ANCORP data set. *Geophys. J. Int.* 181, 911–925. doi:10.1111/j.1365-246X.2010.04531.x
- Orlowsky, D., Rüter, H., Dresen, L., 1998. Combination of common-midpoint-refraction seismics with the generalized reciprocal method. *J. Appl. Geophys.* 39, 221–235. doi:10.1016/S0926-9851(98)00013-5
- Oye, V., Roth, M., 2003. Automated seismic event location for hydrocarbon reservoirs. *Comput. Geosci.* 29, 851–863. doi:10.1016/S0098-3004(03)00088-8
- Parolai, S., 2009. Determination of dispersive phase velocities by complex seismic trace analysis of surface waves (CASW). *Soil Dyn. Earthq. Eng.*
- Parolai, S., 2009. Denoising of Seismograms Using the S Transform. *Bull. Seismol. Soc. Am.* 99, 226–234. doi:10.1785/0120080001
- Pasquet, S., Bodet, L., Dhemaied, A., Mouhri, A., Vitale, Q., Rejiba, F., Flipo, N., Guérin, R., 2015a. Detecting different water table levels in a shallow aquifer with combined P-, surface and SH-wave surveys: Insights from VP/V_S or Poisson's ratios. *J. Appl. Geophys.* 113, 38–50. doi:10.1016/j.jappgeo.2014.12.005
- Pasquet, S., Bodet, L., Longuevergne, L., Dhemaied, A., Camerlynck, C., Rejiba, F., Guérin, R., 2015b. 2D characterization of near-surface V P/V_S: surface-wave dispersion inversion versus refraction tomography. *Near Surf. Geophys.* 13, 315–331. doi:10.3997/1873-0604.2015028
- Peraldi, R., Clement, A., 1972. Digital processing of refraction data study of first arrivals*. *Geophys. Prospect.* 20, 529–548. doi:10.1111/j.1365-2478.1972.tb00653.x
- Persson, L., 2003. Statistical tests for regional seismic phase characterizations. *J. Seismol.* 7, 19–33. doi:10.1023/A:1021216313892
- Pinnegar, C.R., Eaton, D.W., 2003. Application of the S transform to prestack noise attenuation filtering. *J. Geophys. Res.* 108, 1–10. doi:10.1029/2002JB002258
- Polikar, R., 1996. The engineer's ultimate guide to wavelet analysis-the wavelet tutorial [WWW Document]. ...
<http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. *Numerical recipes in C: the art of scientific computing*, Second Edi. ed. doi:10.1016/0898-1221(90)90201-T
- Ramanantoandro, R., Bernitsas, N., 1987. A computer algorithm for automatic picking of refraction first-arrival time. *Geoexploration* 24, 147–151. doi:10.1016/0016-7142(87)90088-3
- Rashed, M., 2014. Fifty years of stacking. *Acta Geophys.* 62, 505–528.

doi:10.2478/s11600-013-0191-4

- Rashed, M.A., 2008. Coordinated by Jeff Deere Smart stacking: A new CMP stacking technique for seismic data. *Lead. Edge* 27, 462–467. doi:10.1190/1.2907176
- Rawlinson, N., Fichtner, A., Sambridge, M., Young, M.K., 2014. Chapter One – Seismic Tomography and the Assessment of Uncertainty. *Adv. Geophys.* 55, 1–76. doi:10.1016/bs.agph.2014.08.001
- Rawlinson, N., Kennett, B., 2004. Rapid estimation of relative and absolute delay times across a network by adaptive stacking. *Geophys. J. Int.* 157, 332–340. doi:10.1111/j.1365-246X.2004.02188.x
- Reading, A.M., Mao, W., Gubbins, D., 2001. Polarization filtering for automatic picking of seismic data and improved converted phase detection. *Geophys. J. Int.* 147, 227–234. doi:10.1046/j.1365-246X.2001.00501.x
- Ricker, N., 1953. The form and laws of propagation of seismic wavelets. *Geophysics* 18, 10–40. doi:10.1190/1.1437843
- Röhm, A.H.E., Bijwaard, H., Spakman, W., Trampert, J., 2000. Effects of arrival time errors on travelttime tomography. *Geophys. J. Int.* 142, 270–276. doi:10.1046/j.1365-246x.2000.00145.x
- Romdhane, A., 2010. Apport de l'inversion des formes d'onde sismique - approximation 2D élastique - à la caractérisation des milieux de la proche surface: cavités souterraines, glissements de terrain. Université Pierre et Marie Curie - Paris VI.
- Ross, Z.E., Ben-Zion, Y., 2014. Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophys. J. Int.* 199, 368–381. doi:10.1093/gji/ggu267
- Rückemann, C., 2012. Comparison of Stacking Methods Regarding Processing and Computing of Geoscientific Depth Data, in: *In Proc. Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012)*. pp. 35–40.
- Sabbione, J.I., Sacchi, M.D., Velis, D.R., 2015. Radon transform-based microseismic event detection and signal-to-noise ratio enhancement. *J. Appl. Geophys.* 113, 51–63. doi:10.1016/j.jappgeo.2014.12.008
- Sabbione, J.I., Velis, D., 2010. Automatic first-breaks picking: New strategies and algorithms. *Geophysics* 75, V67–V76. doi:10.1190/1.3463703
- Sabbione, J.I., Velis, D.R., 2013. A robust method for microseismic event detection based on automatic phase pickers. *J. Appl. Geophys.* 99, 42–50. doi:10.1016/j.jappgeo.2013.07.011
- Sanchis, C., Hanssen, A., 2011. Enhanced local correlation stacking method. *Geophysics* 76, V33–V45. doi:10.1190/1.3552687

- Saragiotis, C.D., Hadjileontiadis, L.J., Panas, S.M., 2002. PAI-S/K: A robust automatic seismic P phase arrival identification scheme. *IEEE Trans. Geosci. Remote Sens.* 40, 1395–1404. doi:10.1109/TGRS.2002.800438
- Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T., Panas, S.M., 2004. Automatic P Phase Picking Using Maximum Kurtosis and κ -Statistics Criteria. *IEEE Geosci. Remote Sens. Lett.* 1, 147–151. doi:10.1109/LGRS.2004.828915
- Schimmel, M., 1999. Phase cross-correlations: Design, comparisons, and applications. *Bull. Seismol. Soc. Am.* 89, 1366–1378.
- Schimmel, M., Gallart, J., 2003. The use of instantaneous polarization attributes for seismic signal detection and image enhancement. *Geophys. J. Int.* 155, 653–668. doi:10.1046/j.1365-246X.2003.02077.x
- Schimmel, M., Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophys. J. Int.* 130, 497–505. doi:10.1111/j.1365-246X.1997.tb05664.x
- Schuster, G.T., Quintus-Bosz, A., 1993. Wavepath eikonal traveltime inversion: Theory. *Geophysics* 58, 1314–1323. doi:10.1190/1.1443514
- Senkaya, M., Karsh, H., 2014. A semi-automatic approach to identify first arrival time: the Cross-Correlation Technique. *Earth Sci. Res. J.* 18, 107–113. doi:10.15446/esrj.v18n2.35887
- Shapiro, J.M.J., 1993. embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Process.* 41, 3445–3462. doi:10.1109/78.258085
- Sharma, B.K., Kumar, A., Murthy, V.M., 2010. Evaluation of seismic events detection algorithms. *J. Geol. Soc. India* 75, 533–538. doi:10.1007/s12594-010-0042-8
- Sheng, J., Leeds, A., Buddensiek, M., Schuster, G.T., 2006. Early arrival waveform tomography on near-surface refraction data. *Geophysics* 71, U47–U57. doi:10.1190/1.2210969
- Shensa, M., 1977. The Deflection Detector-Its Theory and Evaluation on Short-Period Seismic Data, TR-77-03, Texas Instruments.
- Sheriff, R.E., Geldart, L.P., 1995. *Exploration Seismology*. Cambridge University Press.
- Shtivelman, V., 1996. Kinematic inversion of first arrivals of refracted waves—A combined approach. *Geophysics* 61, 509–519. doi:10.1190/1.1443977
- Sleeman, R., van Eck, T., 1999. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Phys. Earth Planet. Inter.* 113, 265–275. doi:10.1016/S0031-9201(99)00007-2
- Socco, L.V., Foti, S., Boiero, D., 2010. Surface-wave analysis for building near-surface

- velocity models — Established approaches and new perspectives. *Geophysics* 75, 75A83–75A102. doi:10.1190/1.3479491
- Soma, N., Takehara, T., Asanuma, H., Niitsuma, H., Baria, R., Michelet, S., Wyborn, D., 2005. Automatic Wave Picking Technique for Multi-Component Microseismicity and It's Practical Application to on Site Analysis in HDR Development, in: World Geothermal Congress. pp. 24–29.
- Sopher, D., Juhlin, C., Huang, F., Ivandic, M., Lueth, S., 2014. Quantitative assessment of seismic source performance: Feasibility of small and affordable seismic sources for long term monitoring at the Ketzin CO2 storage site, Germany. *J. Appl. Geophys.* 107, 171–186. doi:10.1016/j.jappgeo.2014.05.016
- Spagnolini, U., 1991. Adaptive picking of refracted first arrivals. *Geophys. Prospect.* 39, 293–312. doi:10.1111/j.1365-2478.1991.tb00314.x
- Spetzler, J., Snieder, R., 2004. The Fresnel volume and transmitted waves. *Geophysics* 69, 653–663. doi:10.1190/1.1759451
- Stewart, R.R., Huddleston, P.D., Kan, T.K., 1984. Seismic versus sonic velocities: A vertical seismic profiling study. *Geophysics* 49, 1153–1168. doi:10.1190/1.1441745
- Stewart, W.S., 1977. Real-time detection and location of local seismic events in central california. *Bull. Seismol. Soc. Am.* 67, 433–452.
- Stockwell, R., Mansinha, L., Lowe, R.P., 1996. Localization of the complex spectrum: the S transform. *Signal Process. IEEE Trans.* 44, 998–1001. doi:10.1109/78.492555
- Swindell, W., Snell, N., 1977. Station processor automatic signal detection system, phase I: Final report, station processor software development, Texas Instruments Report No. ALEX (01)-FR-77-01, AFTAC Contract Number FO8606-76-C-0025, Texas Instruments Incorporated, Dallas, Texas.
- Takanami, T., Kitagawa, G., 1988. A new efficient procedure for the estimation of onset times of seismic waves. *J. Phys. Earth* 36, 267–290. doi:10.4294/jpe1952.36.267
- Taner, M.T.M., Schuelke, J.S., O'Doherty, R., Baysal, E., 1994. Seismic Attributes Revisited, in: 1994 SEG Annual Meeting. Society of Exploration Geophysicists, pp. 1104–1106.
- Thorbecke, J.W., Draganov, D., 2011. Finite-difference modeling experiments for seismic interferometry. *Geophysics* 76, H1–H18. doi:10.1190/geo2010-0039.1
- Tong, C., 1995. Characterization of seismic phases-an automatic analyser for seismograms. *Geophys. J. Int.* 123, 937–947. doi:10.1111/j.1365-246X.1995.tb06900.x
- Toomey, D.R., Solomon, S.C., Purdy, G.M., 1994. Tomographic imaging of the shallow

- crustal structure of the East Pacific Rise at 9°30'N. *J. Geophys. Res.* 99, 24135–24157. doi:10.1029/94JB01942
- Trnkoczy, A., 2012. Understanding and parameter setting of STA / LTA trigger algorithm 1 Introduction. *New Man. Seismol. Obs. Pract.* 2 2, 1–20. doi:10.2312/GFZ.NMSOP - 2 _IS_8.1
- Tselentis, G., Paraskevopoulos, P., Sokos, E., Martakis, N., Lois, A., Enterprises, L., 2011. A method for microseismic event detection and P-phase picking. *Time* 1638–1642.
- Tselentis, G.-A., Martakis, N., Paraskevopoulos, P., Lois, A., 2011. High-resolution passive seismic tomography for 3D velocity, Poisson's ratio ν , and P-wave quality Q_P in the Delvina hydrocarbon field, southern Albania. *Geophysics* 76, B89–B112. doi:10.1190/1.3560016
- Tselentis, G.-A., Martakis, N., Paraskevopoulos, P., Lois, A., Sokos, E., 2012. Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics. *Geophysics* 77, KS43–KS54. doi:10.1190/geo2011-0301.1
- Valois, R., 2011. Caractérisation structurale de morphologies karstiques superficielles et suivi temporel de l'infiltration à l'aide des méthodes électriques et sismiques. Université Pierre et Marie Curie, Paris, France.
- VanDecar, J., Crosson, R., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares. *Bull. Seismol. Soc. Am.* 80, 150–169.
- Withers, M., Aster, R., Young, C., 1999. An automated local and regional seismic event detection and location system using waveform correlation. *Bull. Seismol. Soc. Am.* 89, 657–669.
- Withers, M., Aster, R., Young, C., Beiriger, J., Moore, S., Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull. Seismol. Soc. Am.* 88, 95–106.
- Wong, J., Han, L., Bancroft, J.C., Stewart, R., 2009. Automatic time-picking of first arrivals on noisy microseismic data, in: CSEG. CREWES, pp. 1–6.
- Yilmaz, Ö., 2001. *Seismic data analysis: Processing, Inversion, and Interpretation of Seismic Data (Investigations in Geophysics, No. 10)*. Society of Exploration Geophysicists.
- Yordkayhun, S., 2011. Detecting near-surface objects with seismic traveltime tomography: Experimentation at a test site. *Sonklanakaran J. Sci. Technol.* 33, 477–485.
- Zelt, C., Ellis, R., 1988. Practical and efficient ray tracing in two-dimensional media for

- rapid traveltimes and amplitude forward modeling. *Can. J. Explor. Geophys.* 24, 16–31.
- Zelt, C.A., Azaria, A., Levander, A., 2006. 3D seismic refraction traveltimes tomography at a groundwater contamination site. *Geophysics* 71, H67–H78.
- Zelt, C.A., Drew, J.J., Yedlin, M.J., Ellis, R.M., 1987. Picking noisy refraction data using semblance supplemented by a Monte Carlo procedure and spectral balancing. *Bull. Seismol. Soc. Am.* 77, 942–957.
- Zeng, X., Zhang, H., Zhang, X., Wang, H., Zhang, Y., Liu, Q., 2014. Surface Microseismic Monitoring of Hydraulic Fracturing of a Shale-Gas Reservoir Using Short-Period and Broadband Seismic Sensors. *Seismol. Res. Lett.* 85, 668–677. doi:10.1785/0220130197
- Zhang, H., Sarkar, S., Toksöz, M.N., Kuleli, H.S., Al-Kindy, F., 2009. Passive seismic tomography using induced seismicity at a petroleum field in Oman. *Geophysics* 74, WCB57–WCB69. doi:10.1190/1.3253059
- Zhang, H., Thurber, C., Rowe, C., 2003. Automatic P-Wave Arrival Detection and Picking with Multiscale Wavelet Analysis for Single-Component Recordings. *Bull. Seismol. Soc. Am.* 93, 1904–1912. doi:10.1785/0120020241

Annexes

Annexe A.– Résumé étendu présenté aux journées scientifiques d'AGAP 2015

Développement d'un algorithme multi-approche adaptatif pour le pointé automatique de données sismiques de proche surface

Amin Khalaf¹, Christian Camerlynck¹, Nicolas Florsch^{2,1}

¹ UPMC - UMR 7619 Métis, case 105, 4 place. Jussieu - 75252 Paris cedex 05,

²UPMC – UMI 209 UMMISCO, case 105, 4 place. Jussieu - 75252 Paris cedex 05, France

amin.khalaf@upmc.fr

Abstract

The use of several algorithms in series becomes a standard approach in seismic studies. Here we propose a new adaptive algorithm based on three picking methods in order to provide robust and precise picking with time uncertainty estimation. This new scheme is nearly a parameter-free algorithm, which is straightforward to implement and demands low computational resources.

Introduction

Un pointé précis des temps de première arrivée sismique joue un rôle important dans de nombreuses études d'imagerie de vitesse sismique, ou d'évolution du milieu avec notamment la surveillance en écoute passive de réservoirs ou d'aquifères. Actuellement, les gros volumes de données numériques acquises et la grande performance des outils informatiques motivent la mise en œuvre d'algorithmes plus sophistiqués pour optimiser le pointé automatique des premières arrivées sismiques.

En général, les premières arrivées se manifestent par une variation notable de l'amplitude, du contenu fréquentiel, ainsi que des propriétés statistiques de la trace sismique dans les domaines temporel ou fréquentiel. Depuis ces dernières décennies, de nombreuses techniques de pointé automatique ou semi-automatique ont été développées. Cependant, la plupart d'entre elles ont été adaptées pour des objectifs sismologiques, telles celles basées sur des rapports LTA/STA. En outre, l'application d'un seul algorithme sur des données issues de la prospection de proche surface n'est pas suffisamment performante en raison de la complexité des structures de la proche surface, des variations rapides du champ de vitesse et d'un rapport signal/bruit insuffisant. Le taux de rejet reste important et la faible fiabilité des pointés nécessite un contrôle manuel coûteux. Pour ces raisons, l'application conjointe d'un ensemble d'algorithmes peut permettre de surmonter ces obstacles en utilisant la totalité de l'information contenue dans le film sismique.

Nous proposons une nouvelle stratégie basée sur un algorithme adaptif combinant plusieurs approches associant l'utilisation de fenêtres multiples imbriquées (MNW = Multi-Nested Windows), l'estimation des propriétés statistiques d'ordre supérieur (HOS = Higher Order Statistics) et le critère d'information d'Akaike (AIC = Akaike Information Criterion ; Akaike 1974). Cet algorithme développé a un double but : d'une part, permettre le pointé automatique avec grande précision des premières arrivées (ondes P) de données sismiques de proche surface ; d'autre part, associer automatiquement à ce pointé une barre d'erreur.

Principe et méthodes

Le mouvement des particules de sous-sol augmente progressivement lorsque la perturbation des ondes sismique arrive. Par conséquent, la première manifestation de la première phase sismique (P-phase) est généralement faible. Par contre, la deuxième partie est plus énergétique. Dans la prospection sismique de proche surface, les premières arrivées sont souvent noyées dans le bruit. L'utilisation des rapports d'énergie dans trois fenêtres imbriquées (MNW) renforce localement la présence des premières arrivées et améliore par suite la précision de pointé automatique.

D'un point de vue pratique, le bruit sismique suit une distribution quasi Gaussienne, en revanche le signal sismique tend vers une distribution non-Gaussienne. Cette différence peut être utilisée pour signaler l'arrivée sismique. La déviation de la gaussienité causée par l'arrivée sismique peut être détectée par les attributs statistiques d'ordre supérieur (HOS) (Saragiotis et al., 2004), notamment par les méthodes basées sur le calcul du kurtosis au travers d'une fenêtre mobile (e.g. Küperkoch et al. 2012).

Du point de vue statistique, la trace sismique peut être divisée en segments, dites localement stationnaires, modélisés par un processus autorégressif (AR). Le début d'une phase sismique peut être identifié en analysant la variation des coefficients de AR et/ou son ordre (e.g. Sleeman and van Eck 1999). Le critère AIC est un outil mathématique permettant de qualifier le modèle AR. Le minimum d'AIC représente le point de séparation idéal entre les deux modèles (bruit et signal) et, par conséquent, la première arrivée. Dans cette étude, l'attribut d'AIC a été calculé directement de la trace sismique selon la version modifiée par Meada (1985).

Implémentation d'algorithmes

Le pointeur implémenté fonctionne suivant un algorithme avec trois étapes successives (MNW, HOS et AIC), où chacune joue son propre rôle. Cette stratégie mime le pointé manuel, lorsque une tendance globale de la localisation des premières arrivées est définie, pour qu'ensuite les instants précis des premières arrivées soient localisés au voisinage de cette tendance.

L'idée principale est fondée sur le fait que les premières arrivées construisent le premier barrage énergétique significatif sur le film sismique (shot gather). Ce bloc cohérent de l'énergie peut être renforcé et localisé par l'algorithme de MNW ; Une importante collection de modèles (représentant cette tendance) est générée par un multi-seuil sur l'attribut de MNW. La solution optimale est sélectionnée par un système itératif (look-like Monte Carlo) en maximisant une fonction de coût contenant des informations dévoilant la présence des premières arrivées. Les

paramètres obtenus (tendance et premières arrivées avec ses erreurs associées) lors de cette étape vont servir comme des outils et guides importants pour le deuxième algorithme (HOS-kurtosis).

Les arrivées potentielles vont être déterminées par l'algorithme de HOS-kurtosis avec les incertitudes associées dans la zone potentielle définie par MNW. Les résultats obtenus des deux premiers pointeurs sont intégrés pour orienter l'algorithme d'AIC en dernière étape. Les pointés finaux sont définis via la moyenne pondérée des trois arrivées potentielles obtenues précédemment (avec les poids du facteur de qualité de chaque pointé). Les barres d'erreurs finales sont également évaluées automatiquement par l'écart type de toutes les pointés déterminés, dans une fenêtre spatiale, après la soustraction la tendance locale.

Tous les paramètres nécessaires à cet algorithme sont auto-adaptatifs. La période dominante de la phase P est le seul paramètre que l'utilisateur doit ajuster pour l'algorithme de MNW. En outre, aucun traitement préliminaire n'est nécessaire en dehors du centrage moyen des traces. L'algorithme mis en place permet d'exploiter les avantages qu'il y a à intégrer plusieurs propriétés, afin d'améliorer l'identification des premières arrivées, et de réaliser un pointeur efficace et robuste des ondes P avec ses incertitudes.

Application et résultats

Deux types de données (synthétiques et réelles) ont été mis à disposition pour valider la robustesse du nouvel algorithme développé. Ces données présentent des enjeux spéciaux à les pointer, notamment avec un pointeur automatique. Les données synthétiques ont été créées à partir d'un modèle réaliste par l'approche en différence finies (figure 1a). Ces données simulent une campagne sismique de proche surface, et présentent des challenges accumulés à les pointer automatiquement (variations de la forme d'onde avec l'offsets, temps du parcours non-linéaire, bruit blanc gaussien et d'une bande limitée ajouté) (figure 1b).

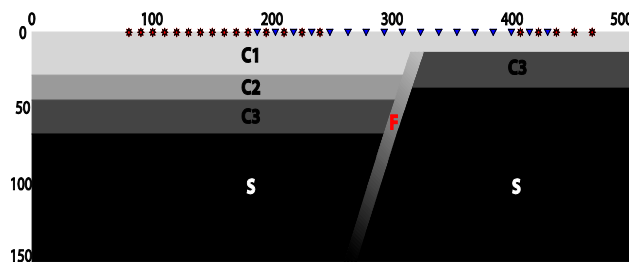


Figure 1 : Schéma du modèle, multicouche (C1 à C3 et S) avec une faille normale (F), utilisé pour la génération de données synthétiques. Les positions de sources et des récepteurs sont schématisées respectivement par des étoiles rouges et triangles bleus

Les données réelles acquises pour des applications opérationnelles en proche surface, avec une source sismique traditionnellement utilisée (chute de poids), montrent généralement une difficulté pour les grands offsets, où le rapport signal-à-bruit (RSB) des premières arrivées se dégrade (figure 3a).

La figure 2a présente les pointés obtenus par le pointeur automatique développé, avec leur barre d'erreur associée. La comparaison entre les pointés issus du pointeur développé et ceux issus d'un pointé manuel (avant ajouter les bruits) s'est fait par le calcul de l'erreur quadratique moyenne (RMS) (Figure 3b). Ce dernier est calculé dans quatre catégories

empiriques (A, B, C et D) classifiées selon le RSB associé. Comme on peut le voir, il y a une bonne consistance entre les deux types des pointeurs malgré le taux de bruit ajouté, mais on constate que les pointés obtenus par pointeur automatique sont décalés d'environ 30 échantillons (pour les trois premières classes). Cela résulte de la sensibilité élevée des trois sous-algorithmes à la forte énergie.

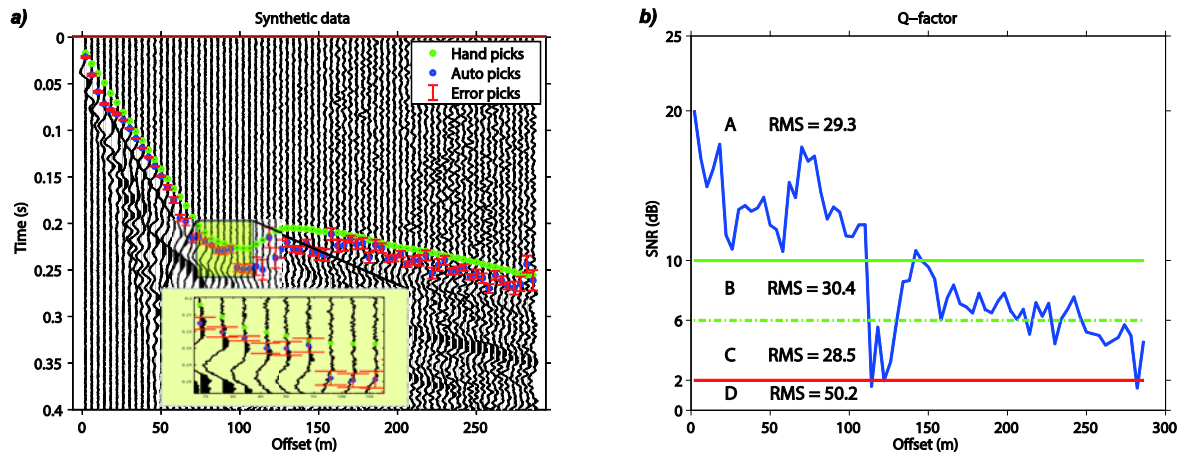


Figure 2 : a) données synthétiques simulées du modèle en figure (1), les cercles en bleu et en vert : pointés obtenus respectivement de pointé automatique et manuel, les barres rouges : erreurs liées au dernier. b) la RMS du facteur de qualité selon quatre classes définies empiriquement, versus l'offsets (en bleu) calculé d'après RSB.

L'application du nouvel pointeur sur des données réelles montre sa robustesse et sa consistance (Figure, 3a). Le problème de mauvaise qualité des premières arrivées à l'offset lointain a quasiment disparu. La RMS dans toutes les catégories n'a pas excédé les 25 échantillons (Figure, 3b).

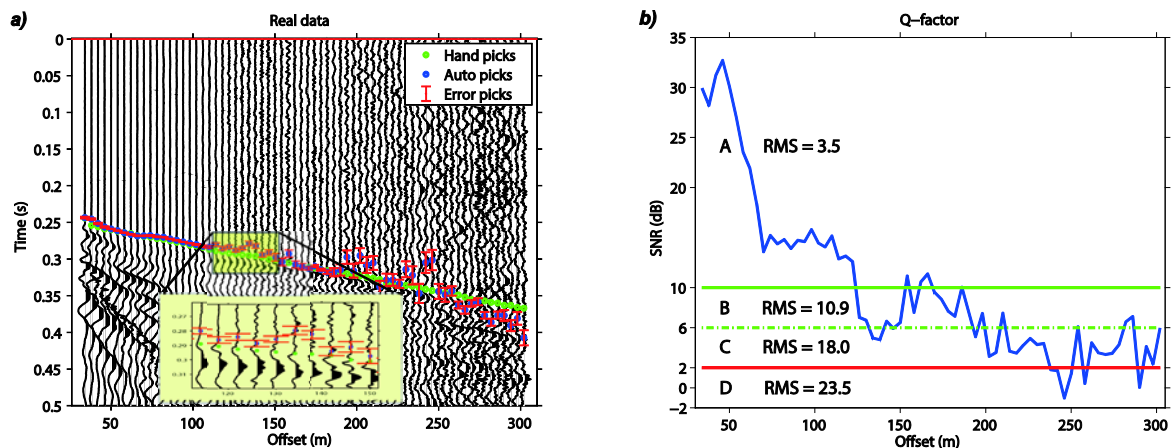


Figure 3 : a) les résultats d'algorithme développé appliqué sur des données réelles (en proche surface), les cercles en bleu et en vert : pointés obtenus respectivement de pointé automatique et manuel, les barres rouges : erreurs liées au dernier. b) la RMS de pointés triés en quatre classes définies empiriquement versus le facteur de qualité avec l'offset (en bleu) calculé d'après RSB.

La figure 4 représente une analyse statistique des résultats obtenus grâce à cet algorithme adaptatif appliqué aux données réelles d'un profil sismique de proche surface. Environ 40% des

différences entre les pointés automatique et celle du pointé manuel se trouvent très proche de zéro (Figure 4a). Une étude statistique approfondie dans chaque catégorie de qualité est montrée sur la figure 4b. La boîte à moustache permet de voir aisément la dispersion des résultats autour de la médiane. Ces dernières, pour toutes les classes, sont quasi nulles. Les 2ème et 3ème quartiles (50%) des différences sont compris dans l'intervalle $[-0.025, +0.05]$ sec. Manifestement, les grandes distributions correspondant normalement aux classes de faible qualité se déploient largement du côté positif de la médiane. L'asymétrie de dispersion de deux dernières classes montre encore fois une remarquable sensibilité de l'algorithme implémenté aux phases énergétiques en présence de bruit. Les valeurs extrêmes sont souvent liées aux traces dont les premières arrivées sont difficiles à identifier. Dans ces cas, les incertitudes estimées ont d'importance équivalente à celles du de pointé automatique. Alors, les pointés associés à une grande erreur pourraient être écartés ou pour le moins pondérés par de faible poids avant les traitements a posteriori des premières arrivées (par exemple avant l'inversion tomographique).

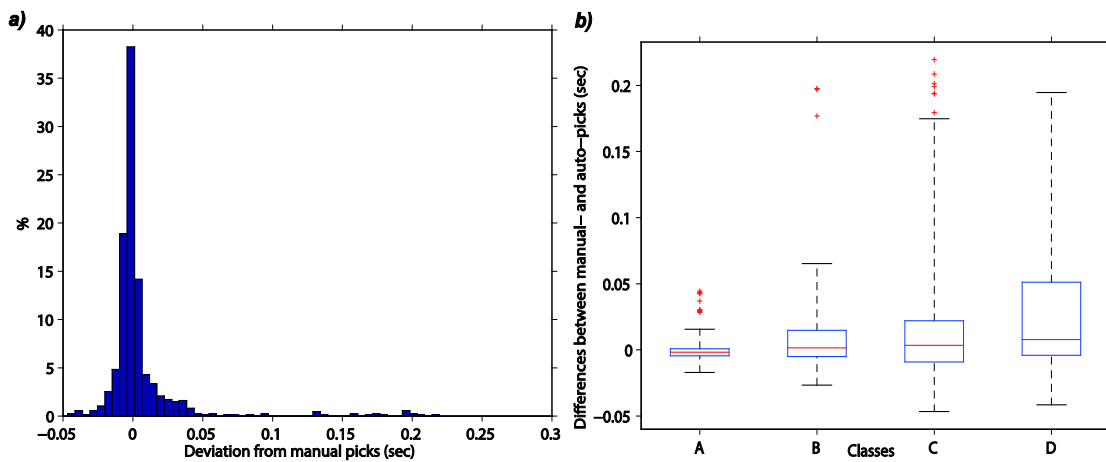


Figure 4 : analyse statistique de performance du schéma développé appliqué sur des données d'un profil sismique de proche surface. a) Histogrammes (en %) de la déviation de pointé manuel (en seconde). b) diagramme en boîte des différences entre le pointé automatique et manuel (en seconde) dans les quatre classes (A, B, C, et D).

Conclusions et perspectives

Partant du principe que la précision de pointé est, avant tout, limitée par la présence de bruits, et par l'impulsivité des premières arrivées, la stratégie mise en place nous a permis de valoriser l'apport de l'association de plusieurs algorithmes. L'algorithme implémenté est simple à mettre en œuvre et ne demande pas de grandes performances informatiques, grâce à l'intégration, en série, des résultats de chaque sous-algorithme. La précision et la fiabilité de cet algorithme ont été évaluées par la comparaison des résultats à ceux issus d'un pointé manuel des données contenant des défis spécieux. La précision peut encore être améliorée en intégrant des attributs de polarisation d'ondes (pour des données en trois composantes), ce qui serait notamment utile lorsque le début du signal est submergé dans le bruit, et pourrait apporter une information supplémentaire des arrivées dans certaines zones, comme les événements non-impulsifs au-dessus de la faille.

Références

- Akaike, Hirotugu. 1974. "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes." *Annals of the Institute of Statistical Mathematics* 26 (1): 363–87. doi:10.1007/BF02479833.
- Küperkoch, L., T. Meier, a. Brüstle, J. Lee, and W. Friederich. 2012. "Automated Determination of S-Phase Arrival Times Using Autoregressive Prediction: Application to Local and Regional Distances." *Geophysical Journal International* 188 (2): 687–702. doi:10.1111/j.1365-246X.2011.05292.x.
- Meada, N. 1985. "A Method for Reading and Checking Phase Time in Auto-Processing System of Seismic Wave Data." *Zisin (Journal of the Seismological Society of Japan. 2nd Ser.)*, March, 365. https://www.jstage.jst.go.jp/article/zisin1948/38/3/38_3_365/_article.
- Saragiotis, C.D., L.J. Hadjileontiadis, I.T. Rekanos, and S.M. Panas. 2004. "Automatic P Phase Picking Using Maximum Kurtosis And κ -Statistics Criteria." *IEEE Geoscience and Remote Sensing Letters* 1 (3). IEEE: 147–51. doi:10.1109/LGRS.2004.828915.
- Sleeman, Reinoud, and Torild van Eck. 1999. "Robust Automatic P-Phase Picking: An on-Line Implementation in the Analysis of Broadband Seismogram Recordings." *Physics of the Earth and Planetary Interiors* 113 (1-4): 265–75. doi:10.1016/S0031-9201(99)00007-2.

Mots-clés

Pointé automatique, sismique, algorithmes adaptatifs, multifenêtre, kurtosis, Akaike Information Criterion (AIC), look-like Monte Carlo.

Annexe B. – Supplementary Material for Chapter 5

B.1. Sets of equations for first stacking using local similarity

Seismic attribute is defined as a measurement derived from seismic data (Taner et al., 1994). Local similarity is a local attribute measured between two signals (Fomel, 2007a), and it, in turn, can be defined as the local correlation between the two signals (Liu et al., 2009).

Let x_t and y_t are two time series of N samples, the global correlation coefficient (γ) between them can be defined as follows:

$$\gamma = \sum_{t=1}^N x_t y_t \left(\sqrt{\sum_{t=1}^N x_t^2 \sum_{t=1}^N y_t^2} \right)^{-1} \quad (4)$$

To locally measure the similarity between two time series, the conventional way is to compute the correlation coefficient in sliding window (w) between them:

$$\gamma_w(i) = \sum_{t=i-w/2}^{i+w/2} x_t y_t \left(\sqrt{\sum_{t=i-w/2}^{i+w/2} x_t^2 \sum_{t=i-w/2}^{i+w/2} y_t^2} \right)^{-1} \quad (5)$$

In more elegant way and based on linear algebra, Fomel (2007) defined the local similarity (γ_t) as the product of two vectors p_t and q_t that are the solution of the following least-squares estimation problems:

$$\begin{aligned} \gamma_t &= \sqrt{p_t q_t}, \\ p_t &= \arg \min_{p_t} \left(\sum_t (x_t - p_t y_t)^2 + R[p_t] \right), \\ q_t &= \arg \min_{q_t} \left(\sum_t (x_t - q_t y_t)^2 + R[q_t] \right), \end{aligned} \quad (6)$$

where R is a regularization operator designed to constrain the solution in a desired behavior, such as locality and smoothness. Shaping regularization can conveniently be applied to solving inverse problems in equations (6) iteratively and replaces the regularization operator R with a shaping (smoothing) operator (Fomel, 2007b). In this study the shaping operator is chosen to be Gaussian smoothing with an adjustable radius. Then, the product of smoothed p_t and q_t defines the local correlation coefficients and, as result, the local similarity attribute.

The non-linear (weighted) stacking of a CMP-gather, containing N traces, based on the local correlation coefficients can be defined according Liu et al. (2009) as follows:

- Compute reference trace by linear (equal-weight) stacking of the NMO-corrected CMP-gather (assuming that stacked trace is close to the desired trace).
- Estimate the local correlation coefficients (γ_i) between the reference trace and each (i th) NMO-corrected trace (x_i) as discussed earlier (Equations (6)).
- Apply soft thresholding (7) to all local correlation coefficients.

$$w_i(t) = \begin{cases} \gamma_i(t) - \varepsilon, & \gamma_i > \varepsilon \\ 0, & \gamma_i \leq \varepsilon \end{cases}, \quad (7)$$

where ε is the threshold value, w_i resulting weights of the i th trace.

- Apply weighted stacking according to following equations:

$$\begin{aligned}\bar{s}(t) &= \frac{1}{K H(t)} \sum_{i=1}^N w_i(t) \cdot x_i(t) \\ K &= \sum_{t=0}^t \sum_{i=0}^N w_i(t),\end{aligned}\tag{8}$$

where normalization K is the sum of weights over CMP-gather. Normalization $H(t)$ is the number of samples with $w_i(t) \cdot s_i(t) \neq 0$ at a given t time.

B.2. Sets of equations used in reference trace calculation based on Phase-Weighted Stack (PWS)

The PWS (Schimmel and Paulssen, 1997) is a nonlinear stacking technique designed for the purpose of noise reduction. To allow phase information to enhance the signal in the linear stack a separate stack is computed for the phase. At first, the analytical signal or complex trace $S(t)$ is computed from the seismic trace $s(t)$:

$$S(t) = s(t) + iH(s(t)) = A(t) e^{i\vartheta(t)},\tag{9}$$

where $H(s(t))$ is the Hilbert transform of $s(t)$, defined by the convolution $\frac{1}{\pi t} * s(t)$. $A(t)$ is the envelope of $s(t)$, and $\vartheta(t)$ its instantaneous phase.

For N traces stacking, the phase stack $c(t)$ is obtained by the absolute value of the sum of instantaneous phase of the traces

$$c(t) = \frac{1}{N} \left| \sum_{j=1}^N e^{i\vartheta_j(t)} \right|.\tag{10}$$

It varies in amplitude between 0 and 1 and is a measure of the coherence between the different traces. It is noticed that no amplitude information are explicitly involved in the phase stack.

The phase stack $c(t)$ can be smoothed, similarly to other measures, by averaging over a time window centered at time t with width $2T$ (Schimmel and Paulssen, 1997):

$$\tilde{c}(t) = \frac{1}{2T + 1} \sum_{\tau=t-T/2}^{t+T/2} c(\tau),\tag{11}$$

where $\tilde{c}(t)$ is smoothed phase stack.

The final PWS is obtained by multiplication with the linear stack:

$$\bar{s}(t) = \tilde{c}^v(t) \frac{1}{N} \sum_{j=1}^N s_j(t),\tag{12}$$

where v is a parameter to tune the transition between coherent and less coherent signal summation. With $v = 0$ the linear stack is retrieved. The degree of distortion of the waveform is determined by the stationarity of the trace and the value of v (Schimmel and Paulssen, 1997).

Consequently, with PWS, each sample of linear stack is weighted by an amplitude-unbiased coherence measure (Equation (12)).

B.3. Sets of equations used for denoising by wavelet transform (WT)

A wavelet is a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero (i.e. it is a waveform of effectively limited time duration).

In contrast to Fourier Transform (FT), wavelet transform is used for non-stationary signals with transitory phenomena (e.g. [Anant and Dowla, 1997](#)). The frequency response of these signals varies in time (i.e. these signals have quick local variations, for example the seismic signals). By the FT analysis, a signal is decomposed into sine waves of various frequencies. While in the wavelet analysis, a signal is decomposed into scaled and shifted versions of the mother wavelet ([Misiti et al., 1997](#)).

The Multiresolution analysis (MRA) of a signal is produced by convolution of the signal with a scalable modulated window. In other word, MRA employs a method in which the signal is analyzed at different frequencies with different resolutions. Of course there is still some trade-off: at high frequencies, MRA has a good time resolution and a poor frequency resolution. On the contrary, at low frequencies, it has a good frequency resolution and a low time resolution. This trade-off is mostly suitable to majority naturally occurring signals which tend to have low frequency components for long duration (so a good time resolution is not necessary) and high frequency components of short duration (a good time resolution is provided).

Suppose that $\psi(t)$ is the analyzing wavelet function (also known as mother wavelet). The term “mother” reflects the fact, other wavelets belonging to a given wavelet family $\psi_{b,a}$ (called also daughter wavelets) can be derived from ψ by scaling and translation as follows

$$\psi_{b,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \text{ with } L^2 \text{ norm} \quad (13)$$

In equation (13), the translation $b \in R$ represents the location of a certain scale or dilation $a \in R^{+*}$, which is shifted along the signal in time.

For any signal $x(t)$ (assuming $x \in L^2(R)$) with finite energy, the continuous wavelet transform (CWT) is defined ([Grossmann and Morlet, 1984](#)):

$$CWT_x^\psi(b, a) = \langle x, \psi_{b,a} \rangle = \int_{-\infty}^{+\infty} x(t)\psi_{b,a}^*(t)dt, \quad (14)$$

where the angle brackets denote the inner product, and the asterisk symbol $*$ denotes the complex conjugate. The integral in (14) can be interpreted as a scalar operation of the signal and a dilated band-pass filter (a basis function $\psi(t)$). For each shift, the resulting wavelet coefficient can be considered as a measure of similarity between the wavelet and the corresponding segments of the signal. Here, the similarity of frequency content is only explicitly involved in consideration ([Polikar, 1996](#)). The scale or dilation parameter a represents frequency information (a is inversely proportional to radian frequency). As a result, low frequencies correspond to high scales and vice versa.

If $\psi(t)$ satisfies admissible condition (i.e. $\int_{-\infty}^{+\infty} \frac{\hat{\psi}(t)}{|\omega|} d\omega < +\infty$, where $\hat{\psi}(t)$ is the FT of $\psi(t)$) and the signal $x(t)$ satisfies the finite energy condition (i.e. $\int_{-\infty}^{+\infty} x(t)dt < \infty$), the $x(t)$ can be reconstructed without loss any information as follows

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} CWT_x^\psi(b, a) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \frac{da db}{a}, \quad (15)$$

where C_ψ is a constant depending on the kind of the mother wavelet $\psi(t)$

$$C_\psi = 2\pi \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega. \quad (16)$$

Theoretically, a and b are continuous. But in fact, they are sampled with a small enough step size due to digital computers use. Although that they can still be considered continuous.

One drawback of the CWT is that the decomposition of the signal is often redundant, due to a and b are continuous with real numbers. It is therefore costly in terms of computation time and resources. If a and b are chosen to be discrete then analysis will be much easier and will not generate the huge data. On other hand, the original signal can be completely reconstructed by a sample version of $CWT_x^\psi(b, a)$. One way to discretize the scale and the translation parameters is

$$a = a_0^j, \quad b = kb_0 a_0^j, \quad j, k \in \mathbf{Z} \quad (17)$$

with a fixed values of $a_0 > 1$ and $b_0 > 0$ (a_0 is the resolution step).

Typically, we sample the $\psi_{b,a}$ in dyadic grid (i.e. $a_0 = 2$ and $b_0 = 1$). By considering these values in (17) and therefore in (13), the corresponding discrete wavelet family is expressed by

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right). \quad (18)$$

By this dyadic discretization of mother wavelet parameters, the WT is defined only at discrete scales and times (Eq. (18), and called the dyadic WT (DWT).

$$\begin{aligned} DWT_x^\psi(j, k) &= \langle x, \psi_{j,k} \rangle \\ &= 2^{-\frac{j}{2}} \sum_{t=-\infty}^{\infty} x(t) \cdot \psi^*(2^{-j}t - k) \end{aligned} \quad (19)$$

In practice, the DWT is obtained by passing a signal successively through low-pass (*LP*) and high-pass filters (*HP*), followed by a decimation (subsampling) procedure (Mallat, 1989). This computation algorithm is referred as the Mallat's algorithm in the literature. The low and high pass filters used are determined according to the used mother wavelet (Misiti et al., 1997). The outputs of the low-pass and high-pass filters are called approximation (a) and detail (d) coefficients, respectively. Here, it should be noticed that the pair filters should be quadrature mirror. **Figure B.9** demonstrates the Mallat's pyramid tree algorithm for 3-levels decomposition of a signal.

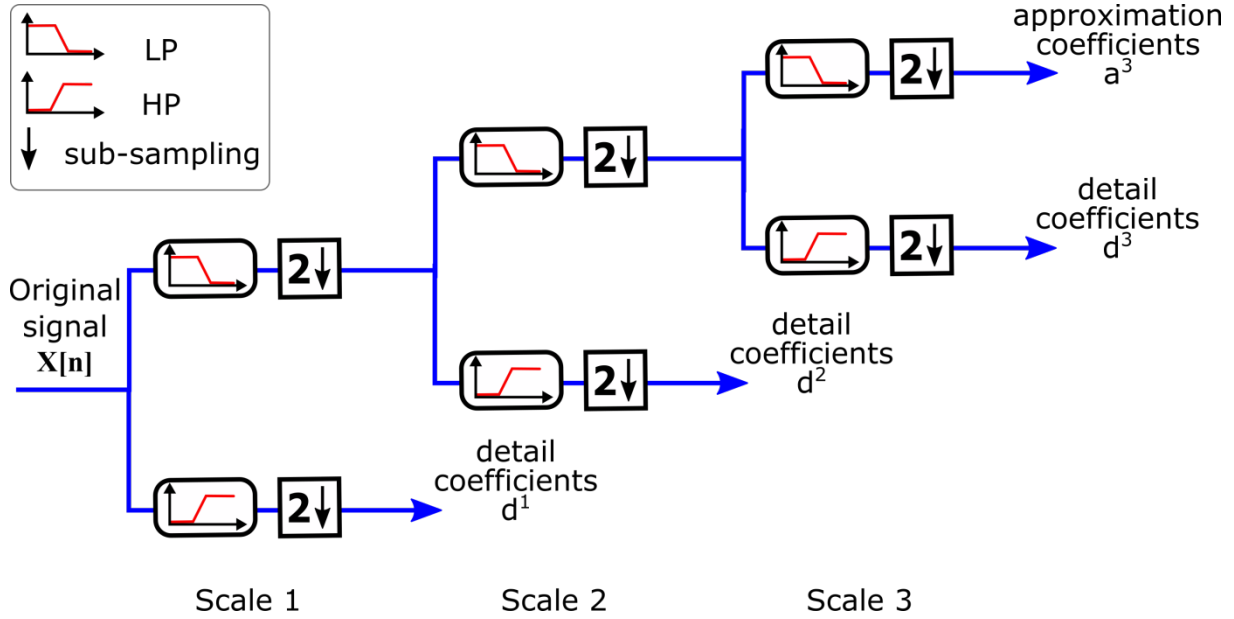


Figure B.9. Mallat's algorithm demonstration for 3-levels decomposition of a signal (adapted from [Anant and Dowla, 1997](#)).

Mathematically, let $x[n]$ is a discrete signal of N samples. At each j scale, the outputs of filtering and sub-sampling can be expressed by the two following equations ([Cohen et al., 1992](#)):

$$\begin{aligned} a^{j+1}[n] &= \sum_{k \in \mathbb{Z}} LP[k - 2n] a^j[k], \\ d^{j+1}[n] &= \sum_{k \in \mathbb{Z}} HP[k - 2n] a^j[k], \end{aligned} \quad (20)$$

with $a^0 = x[n]$, i.e. the original signal (**Figure B.9**). The output approximation coefficients (low filtering outputs) become input to the next level in the decomposition.

The impulse responses of both filters (LP, HP) are dependent on each other by ([Polikar, 1996](#)):

$$HP[L - 1 - n] = (-1)^n \cdot LP[n], \quad (21)$$

where L is the filter length. In terms of DWT, the LP, HP are known as the scaling function and wavelet function, respectively.

Since the (LP, HP) is a pair of conjugate mirror filters, the signal can be perfectly reconstructed by using another pair conjugate mirror filter ($\widetilde{LP}, \widetilde{HP}$). For each level, the reconstruction is given by:

$$a^j[n] = \sum_{k \in \mathbb{Z}} \widetilde{LP}[n - 2k] a^{j+1}[k] + \widetilde{HP}[n - 2k] d^{j+1}[k]. \quad (22)$$

In case of orthogonal wavelets the $\widetilde{LP}[n] = LP[-n]$ and $\widetilde{HP}[n] = HP[-n]$ ([Misiti et al., 1997](#)).

Whereas the wavelet decomposition (equations (20) involves filtering and sub-sampling, the wavelet reconstruction (equations (22) consists of up-sampling (by inserting zeros between samples) and filtering.

Wavelet-based denoising

The idea behind denoising with wavelets is the fact that less significant (small-valued) coefficients in the wavelet decomposition would correspond to noise. Thus, by zeroing them out, the noise is also removed.

Conventionally, wavelet coefficients, in each scale-level, ($C_{j,k}$) are modifying by using either hard or soft thresholding (Donoho and Johnstone, 1994) given respectively by:

$$C_{den}^h(j, k) = \begin{cases} C(j, k), & \text{if } |C_{j,k}| > \lambda_j \\ 0, & \text{if } |C_{j,k}| \leq \lambda_j \end{cases}, \quad (23)$$

$$C_{den}^s(j, k) = \begin{cases} \text{sign}(C(j, k))(|C(j, k)| - \lambda_j), & \text{if } |C_{j,k}| > \lambda_j \\ 0, & \text{if } |C_{j,k}| \leq \lambda_j \end{cases}, \quad (24)$$

The threshold-value (λ_j) is chosen according to the signal energy and the noise variance. It can be determined as universal threshold (Donoho and Johnstone, 1994), for each scale j , given by:

$$\begin{aligned} \lambda_j &= \sigma_j \sqrt{2 \log_e N_j}, \\ \sigma_j &= \text{std}(C_{j,k}) \\ &= \text{median} \{ |C_{j,1} - \bar{C}_j|, |C_{j,2} - \bar{C}_j|, \dots, |C_{j,k} - \bar{C}_j| \} / 0.6745, \end{aligned} \quad (25)$$

where N_j is the number of coefficient in each scale j .

In this study, we follow the automatic denoising method proposed by Ahmadi and Quian Quiroga (2013) for clinical diagnostic signals denoising. This method is based on the inter- and intra-scale variability of the wavelet coefficients and their deviations from baseline values.

Two denoising steps involved in this denoising scheme, that is neighboring coefficients denoising (Cai and Silverman, 2001) and zerotrees denoising (Shapiro, 1993). The first denoising incorporates information of neighboring coefficients in the thresholding process, for each scale j , as follows:

$$C_{den}^1(j, k) = \begin{cases} C(j, k), & \text{if } C_{j,k-1}^2 + C_{j,k}^2 + C_{j,k+1}^2 > \lambda_j^2 \\ 0, & \text{if } C_{j,k-1}^2 + C_{j,k}^2 + C_{j,k+1}^2 \leq \lambda_j^2 \end{cases}, \quad (26)$$

where λ_j is estimated as in equation (25).

At the second denoising step (zerotrees denoising), the output coefficients (C_{den}^1) from the first step are kept or deleted according to a logical control, in which the dependency of the current coefficient to whom in the higher scale is examined. In other words, “if a parent coefficient is removed, all its children should be removed as well” (Ahmadi and Quian Quiroga, 2013).

For each scale, this denoising process is obtained recursively starting from the last level by the formula (27). The approximation coefficients should be excluded out this process.

$$C_{den}^2(j, k) = \begin{cases} C_{den}^1(j, k), & \text{if } C_{den}^1(j+1, k) \neq 0 \\ 0, & \text{if } C_{den}^1(j+1, k) = 0 \end{cases}. \quad (27)$$

References

- Ahmadi, M., Quian Quiroga, R., 2013. Automatic denoising of single-trial evoked potentials. *Neuroimage* 66, 672–80. doi:10.1016/j.neuroimage.2012.10.062
- Anant, K.S., Dowla, F.U., 1997. Wavelet transform methods for phase identification in three-component seismograms. *Bull. Seismol. Soc. Am.* 87, 1598–1612.
- Cai, T., Silverman, B., 2001. Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā Indian J. Stat.* 2.
- Cohen, A., Daubechies, I., Feauveau, J.-C., 1992. Biorthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* 45, 485–560. doi:10.1002/cpa.3160450502
- Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455. doi:10.1093/biomet/81.3.425
- Fomel, S., 2007a. Local seismic attributes. *Geophysics* 72, 29–33. doi:10.1190/1.2437573
- Fomel, S., 2007b. Shaping regularization in geophysical-estimation problems. *Geophysics* 72, R29–R36. doi:10.1190/1.2433716
- Grossmann, A., Morlet, J., 1984. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. Math. Anal.* 15, 723–736. doi:10.1137/0515056
- Liu, G., Fomel, S., Jin, L., Chen, X., 2009. Stacking seismic data using local correlation. *Geophysics* 74, V43–V48. doi:10.1190/1.3085643
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 674–693. doi:10.1109/34.192463
- Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J., 1997. *Wavelet toolbox: Matlab User's Guide (R2015b)*. The MathWorks, Inc.
- Polikar, R., 1996. The engineer's ultimate guide to wavelet analysis-the wavelet tutorial [WWW Document]. ... <http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>.
- Schimmel, M., Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophys. J. Int.* 130, 497–505. doi:10.1111/j.1365-246X.1997.tb05664.x
- Shapiro, J.M.J., 1993. embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Process.* 41, 3445–3462. doi:10.1109/78.258085
- Taner, M.T.M., Schuelke, J.S., O'Doherty, R., Baysal, E., 1994. Seismic Attributes Revisited, in: 1994 SEG Annual Meeting. Society of Exploration Geophysicists, pp. 1104–1106.

Table des illustrations

- Figure 2.1.** Approche du pointé d'un point de vue probabiliste (d'après Diehl, 2008): le plus tôt choix possible correspond à tE , le plus tard choix possible correspond à tL . Le temps de la première arrivée la plus probable se trouve dans cet intervalle (i.e., compte tenu de l'apparition d'une phase sismique en fonction probabiliste Pa , le temps d'arrivée est exprimé par le temps le plus probable tA , c.-à-d. avec $PatA = \max(Pa)$. D'autre part, la probabilité d'apparition $PatE$ et $PatL$ se rapproche de zéro) (Diehl, 2008). (a). Stratégie du pointé fondée sur l'attribut de l'amplitude (ARSB). (b). Stratégie du pointé fondée sur l'attribut de la fréquence (FRSB).11
- Figure 2.2.** Algorithme STA/LTA: détermination et pointé de phase. **(en haut)**. Sismogramme filtré avec un filtre passe-bande (2-10 Hz) et pointé automatique (ligne rouge verticale). **(en bas)**. Fonction STA/LTA calculée par $STA = 1$ sec et $LTA = 20$ sec et seuil THR (ligne bleue). Le déclenchement correspond à l'intersection entre la courbe de STA/LTA et le seuil (ligne rouge) et la partie verte de la courbe représente le maximum local (d'après Küperkoch, 2010)15
- Figure 2.3.** Exemple d'événement pointé par la technique MCM: **(a)** configuration d'algorithme (deux fenêtres imbriquées) **(b)** le rapport d'énergie (en vert) et sa version filtrée par le filtre de préservation des bords (en rouge). **(c)** la dérivée de premier ordre où la première arrivée (illustrée par carré rouge sur (a)) correspond au maximum de cette dérivée (modifiée d'après Sabbione and Velis, 2010).18
- Figure 2.4.** Exemple de kurtosis utilisé comme FC: **(a)**. trace sismique filtrée (3-45 Hz), les trois barres horizontales indiquent les fenêtres de calcul de kurtosis. **(b)**. Kurtosis et histogrammes de distribution pour les trois fenêtres, les lignes solide représentent le meilleur adjusment avec une distribution Gaussienne. **(c)**. La FC résultant basée sur le kurtosis (d'après Baillard et al., 2014).20
- Figure 2.5.** Exemple du pointé par AR-AIC: **(en haut)**. Sismogramme (noir) et pointé initial (bleu) détecté par le rapport STA/LTA (rouge). Le niveau du bruit a été estimé à partir du segment entre les deux lignes vertes. NW et SW sont les fenêtres de détermination des coefficients du modèle AR pour, respectivement, le bruit et le signal. **(en bas)**. Le sismogramme zoomé (noir) autour du pointé de STA/LTA (pointillé rouge) et le pointé automatique par AR-AIC (rouge). La fonction d'AIC (verte), (d'après Küperkoch, 2010).22
- Figure 3.1.** Illustration of MNW method: **(a)**. normalized free-noisy trace and the three windows (red = before instant t; black = just after instant t; green = after a delay instant t) used on the calculation of CF_{mnw} . **(b)**. Energy ratios ($ER1 =$ energy ratio calculated between the before window and after window; $ER2 =$ calculated energy ratio between delayed window and before window), $CF_{mnw} =$ Characteristic Function of the incorporated energy ratio, and $Thr =$ theoretical threshold.34

Figure 3.2. Illustration of HOS based kurtosis method: **(a)**. normalized free-noisy trace and windows (red rectangle) used for calculating CF_k . **(b)**. the resulting kurtosis CF (blue solid line), and transformed kurtosis based on Baillard procedure (red solid line). The black dashed line indicates the manual pick of first arrival of P-wave.....36

Figure 3.3. Illustration of AIC picker based upon Akaike Weights: **(a)**. normalized free-noisy trace and two variable length windows before and after (red and black dashed rectangles, respectively) **(b)**. The CFs based on AIC (blue solid line) and on AW (red solid line). The part of CF_{aic} used in calculation CF_{wa} is highlighted by purple line. The black vertical dashed line represents the manual pick of the first arrival.....37

Figure 3.4. Illustration of the Single-trace algorithm applied to a synthetic trace: **(a)**. Noise-free synthetic trace (black), after add noise (blue), manual first arrival (vertical black line). The first period of P-wave is framed by the red rectangle **(b)**. MNW sub-algorithm topics: threshold (dashed red line), BPZ (vertical red line), picked potential first arrival and the associated error (respectively, vertical pointed and horizontal green line). **(c)**. HOS sub-algorithm topics: the guide picker (vertical pointed green line), maximum of kurtosis (small green circle), picked potential first arrival and associated error (respectively, vertical pointed and horizontal red line). **(d)**. AIC sub-algorithm topics: the picker guide (vertical pointed blue line), the red part of CF_{aic} is considered for calculating CF_{wa} , the picked potential first arrival and associated error (respectively, vertical pointed and horizontal red purple). The gray zones in b, c, and d represent the searching windows SW1, SW2, and SW3, respectively. **(e)**. Zoom in of the trace in (a) around the first arrival (vertical solid line). The final pick and its error are marked by vertical pointed and horizontal solid lines, respectively. The potential first arrivals and their associated errors picked by the three sub-algorithms are re-delineated on the noisy trace.....43

Figure 3.5. Numerical simulation: **(a)**. Schematic view of the multi-layer model (layers C1-C3, substratum S, and F page of normal fault) used for generating the synthetic data. The positions of the sources and the receivers are highlighted by red stars and blue triangles, respectively. The thicknesses and velocities are given in the **Table 3.2**. **(b)**. Normalized vertical component of the wavefield simulated by a source located at 198 m from the left side of the model. The small green circles represent the manual picks.....47

Figure 3.6. A synthetic vertical components record with the first arrivals times obtained theoretically (in green) and automatically with their interval errors using the proposed adaptive algorithm (in blue and the error barre in red): **(a)**.Noise-free shot-gather. **(b)**. With 20% random white noise. **(c)**. With random band-limited noise [40-180] Hz. **(c)**. With the noise in (a) and in (b) (random and band-limited noise).48

Figure 3.7. Statistical analysis: **(a)**. Histograms of differences between automatic and actual picks for noise-free data (in red) and noisy data (blue). **(b)**. Boxplot graphs of the differences for noisy data within the quality classes (A, B, C, and D). **(c)**. the same but for noise-free data.49

Figure 3.8. Quality assessment of automatic and actual picks, respectively in blue and red curves for zero noise **(a)**, 20% random noise added **(b)**, band-limited noise added **(c)**, and random and band-limited noise added **(d)**. The red solid curve represent the quality factor of the actual picks for the noise-free gather, the red pointed curve the same but after adding the noise. The tables in a, b, c, and d show the percent of the traces in each quality class and the related RMS of the differences between the automatic and real picks. The dashed black lines represent the limits of quality categories.....50

Figure 4.1. Performance comparison on a dynamite shot gather: **(a) and (b)**. Dynamite shot record number 6 from Yilmaz (2001) with the first arrival picks determined manually (green), adaptive algorithm (blue), and MCM/FDM (red), respectively on a and b. **(c) and (d)**. The quality of the first arrival signals corresponding with the picks shown in **a** and **b**, respectively.58

Figure 4.2. Performance comparison on a Vibroseis shot gather: **(a) and (b)**. Vibroseis shot record number 23 from Yilmaz (2001) with the first arrival picks determined manually (green), adaptive algorithm (blue), and MCM/FDM (red), respectively on **a** and **b**. **(c) and (d)**. The quality of the first arrival signals corresponding with the picks shown in **a** and **b**, respectively.59

Figure 4.3. Boxplots of the differences between manually and automatically derived P-first arrivals using AAA, MCM, and FDM on the dynamite shot gather shown in **Figure 4.1**.60

Figure 4.4. Boxplots of the differences between manually and automatically derived P-first arrivals using AAA, MCM, and FDM on the Vibroseis shot gather shown in **Figure 4.2**.61

Figure 4.5. An example of automatic picking results by AAA on weight-drop-source shot gather.62

Figure 4.6. Histogram of AAA picks deviations from manual picks **(left)**. Boxplot diagrams of the differences between manually and automatically picks within Q-classes (1-4) classified according to SNR of manual picks **(right)**.....63

Figure 4.7. Examples of AAA (bleu curves) robustness versus MER (red curves) in presence a moderate noise: **(Left)**. Very high quality shot gather. **(Right)**. A moderate quality shot gather.65

Figure 4.8. Histograms of differences between automatic and manual picking times using the proposed AAA **(left)** and MER algorithm **(right)**.65

Figure 4.9. Customized boxplots of the deviations automatic picks (1 = AAA) and (2 = MER) form the manual picks, the pointed green line is the baseline though the difference equals zero.66

Figure 4.10. Examples of picked shot gather by automatic algorithms (AAA = bleu curves) and (MER =red curves) with highly correlated ambient noise: **(Left)**. Moderate quality shot gather. **(Right)**. Very bad quality shot gather.67

Figure 4.11. Histograms of differences between manually and automatically derived first arrival times (P-onsets) by AAA (left) and MER (right).67

Figure 4.12. Customized boxplots of the deviations automatic picks (1 = AAA) and (2 = MER) form the manual picks, the pointed green line is the baseline though the difference equals zero.68

Figure 4.13. An illustrative example of quality control-1 based on SNR of the picked arrivals; picks in the transparent red window are to be rejected, picks in the transparent green window are to be accepted, and those in yellow one are to examine by the quality control-2 based on adapted Mateeva’s formula, equation (4.8).70

Figure 4.14. An illustrative example of quality control-2 based on modified Mateeva’s equation (4.8): **(Left)**. Input arguments functions of equation (4.8) (Q-factor in blue curve and pick errors in the red error bars), the two horizontal green lines represent the SNR at the 2 dB and 10 dB. **(Right)**. The resulting tau (τ) function (in blue curve) and the threshold-error at the level of 5 ms (in horizontal green line).70

Figure 4.15. An example of Q-control/editing application on a field-data record: **(Left)**. A weight-drop source shot gather, blue asterisks and error bars are the outputs of AAA picking, green line represent the interpolated manual picks, and the edited picks are highlighted in small red circles. **(Right)**. Blue asterisks are the same automatic picks shown in left figure and small red circles are the final output of editing process (including the misspicks gap control).71

Figure 4.16. Another example of Q-control/editing application on a field-data record: **(Left)**. A sledgehammer-source shot gather, blue asterisks and error bars are the outputs of AAA picking, green line represent the interpolated manual picks, and the edited picks are highlighted in small red circles. **(Right)**. Blue asterisks are the same automatic picks shown in left figure and small red circles are the final output of editing process (including the misspicks gap control).72

Figure 5.1. Simplified schema of roll-along technique acquisition.80

Figure 5.2. Pictorial description of local similarity stacking: **(a)**. Cube of 5 adjacent real shot gathers. **(b)**. Cube of repeated reference trace. **(c)**. Local similarity coefficients cube of the 5 shot gathers. **(d)**. Weights used in local similarity weighted stacking (local similarity cube after applying a thresholding).81

Figure 5.3. Example of enhanced local-similarity stacking with using PWS-trace as the reference trace: **(Left)**. Stacked shot gather. **(Right)**. Raw shot gather.83

Figure 5.4. The second stacking using cross-correlation in limited time-spatial window moving (blue rectangle) over the shot gather.84

Figure 5.5. Array chart describes the double stacking flow on conventional roll-along acquisition technique.	85
Figure 5.6. De-noising by wavelet transform (precisely with multiresolution decomposition): WT is the dyadic wavelet transform for N level (here, 3 levels), then the results are the details (D1, D2 and D3) and the approximation A3. Sub-IWT is the inverse wavelet transform of each level. Den. indicates to the denoising or thresholding process. The flow with black arrows shows the reversibility property of WT and generating sub-band signals from the WT coefficients. The flow with red arrows is the same of the black ones, but with denoising procedure.	87
Figure 5.7. Comparison of automatic picking before and after applying double stacking on good quality data.....	89
Figure 5.8. Comparison of automatic picking before and after applying double stacking on bad quality data.	89
Figure B.9. Mallat's algorithm demonstration for 3-levels decomposition of a signal (adapted from Anant and Dowla, 1997).....	121

Table des tableaux

Table 3.1 : Comparative summary of the modification between the single-trace and entire-shot algorithms parameters (~: no parameters, ==: the same parameters).
46

Table 3.2 : Thickness and velocities of the layers model used in the simulation.....47

Table 4.1 : Overview of automatic picking algorithm considered56

Table 4.2. Assessment of the RMS of the differences and the picking errors within a *20 ms* tolerance for the real data shown in **Figure 4.1** and **Figure 4.2**. Here, the errors are considered to be between the manual and automatic picking using AAA, MCM, and FDM picking algorithms. (Abs. diff = absolute difference).....61

Résumé :

Un pointé précis des temps de première arrivée sismique joue un rôle important dans de nombreuses études d'imagerie de vitesse sismique, ou d'évolution du milieu avec notamment la surveillance en écoute passive de réservoirs ou d'aquifères. En général, les premières arrivées se manifestent par une variation notable de l'amplitude, du contenu fréquentiel, ainsi que des propriétés statistiques de la trace sismique dans les domaines temporel et/ou fréquentiel. Depuis ces dernières décennies, de nombreuses techniques de pointé automatique ou semi-automatique ont été développées. Cependant, la plupart d'entre elles ont été adaptées pour des objectifs sismologiques, telles celles basées sur des rapports LTA/STA. En outre, elles ne sont pas suffisamment performantes sur des données issues de la prospection de proche surface en raison de la complexité des structures du sous-sol, des variations rapides du champ de vitesse et d'un rapport signal/bruit parfois insuffisant.

Une nouvelle stratégie a été développée à partir d'un algorithme adaptatif combinant plusieurs approches associant l'utilisation de fenêtres multiples imbriquées, l'estimation des propriétés statistiques d'ordre supérieur et le critère d'information d'Akaike. L'algorithme développé exploite les avantages d'intégrer plusieurs propriétés (l'énergie, la « gaussianité », et la stationnarité) dévoilant la présence des premières arrivées. Cette stratégie mime le pointé manuel, lorsque une tendance globale de la localisation des premières arrivées est définie, pour qu'ensuite les instants précis des premières arrivées soient localisés au voisinage de cette tendance. Tandis que les incertitudes estimées sont, dans certains cas, du même ordre que pour le pointé automatique (notamment en tomographie sismique, ou même pour certains traitements *a posteriori*), l'algorithme fournit aussi automatiquement une estimation sur l'incertitude des pointés. La précision et la fiabilité de cet algorithme ont été évaluées par la comparaison des résultats avec ceux issus d'un pointé manuel ainsi qu'avec d'autres pointeurs automatiques, dans certaines conditions de bruit. Tous les paramètres nécessaires à cet algorithme sont auto-adaptatifs grâce à l'intégration, en série, des résultats de chaque sous-algorithme. Il est simple à mettre en œuvre et ne nécessite pas de grandes performances informatiques.

La présence du bruit sismique dans l'enregistrement détériore souvent la performance d'algorithmes de pointé automatique, particulièrement dans le cas de petits événements. Le signal et le bruit sismique partagent souvent les mêmes bandes de fréquences, ce qui rend le filtrage dans le domaine fréquentiel peu efficace. La sommation peut être une alternative exploitant la redondance des données sismiques de proche surface. Nous proposons une nouvelle procédure fondée sur la double sommation dans le domaine temporel, ce afin d'améliorer la détectabilité des premières arrivées. Cette procédure est fondée sur le principe clé de la ressemblance locale entre les traces « stackées ». Un dé-bruitage par transformée en ondelettes peut être effectué, comme processus supplémentaire, avant la deuxième sommation, pour les données très bruitées. Une étude comparative de la performance du pointé automatique avec d'autres méthodes a été réalisée sur des jeux de données réelles bruitées. Les résultats ont montré l'intérêt qu'il y a à appliquer la double sommation avant le pointé automatique.

Mots-clés : pointé automatique, algorithme adaptatif, kurtosis, Akaike Information Criterion, tomographie des premières arrivées sismiques, double sommation.