Structural tomographic approaches for urban area analysis using high resolution SAR tomography : TomoSAR

Clément Rambour

université PARIS-SACLAY

NNT : 2019SACLT007

Thèse de doctorat

Approches structurelles pour la tomographie SAR haute résolution en milieu urbain dense

Thèse de doctorat de l'Université Paris-Saclay
préparée à Telecom ParisTech

Ecole doctorale n°580 Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat : Traitement du Signal et des Images

Thèse présentée et soutenue à Paris, le 18/02/2019, par

Clément Rambour

Composition du Jury :

Charles Soussen Professeur, CentralSupélec (L2S)	Président
Pascal Monasse Professeur, École des Ponts ParisTech (LIGM)	Rapporteur
Antoine Roueff Maitre de Conférence, Centrale Marseille	Rapporteur
Gilda Schirinzi Professeur, Dipartimento di Ingegneria, Université de Naples Parthénope	Examinatrice
Jean-François Giovannelli Professeur, Université de Bordeaux (IMS Lab)	Examinateur
Hélène Oriot Ingénieur de recherche, ONERA	Directrice de thèse
Florence Tupin Professeur, Télécom ParisTech (LTCI)	Directrice de thèse
Loïc Denis Maitre de Conférence, Laboratoire Hubert Curien, Université de Saint-Etienne	Co-directeur de thèse

Remerciements

Je tiens tout d'abord à remercier les rapporteurs qui m'ont permis d'améliorer ce manuscrit et de préparer ma soutenance. Je remercie plus généralement l'ensemble des membres du jury qui ont assisté à ma soutenance et tout particulièrement ceux venant de loin.

Je remercie les membres du DEMR qui m'ont accueilli au début de ma thèse et n'ont jamais été avares de conseils ou propositions. Parmi eux Frédéric et Abigael, j'ai apprécié que vous m'intégriez dans vos (multiples) pauses café ainsi que vos coups de main quand j'ai pu en avoir besoin. Hélène, l'intérêt que tu portais à mes travaux ainsi que tes encouragements ont été très motivants tout au long de ma thèse.

La liste des gens à citer à Telecom est longueet je crains d'en oublier certains. Bien sûr, Florence, je te suis très reconnaissant de ton investissement. Ta rigueur et ton exigence m'ont beaucoup appris. Mais ce ne serait pas te faire justice que de citer ces seules qualités car j'ai aussi apprécié les valeurs que tu portes (du moins celle que j'ai pu te voir défendre notamment en conseil de département). Jean-Marie, sans toi je ne sais pas sur quoi j'aurais pu travailler ! Ta facilité d'accès et les échanges que nous avons pu avoir sur de nombreux sujets m'ont été plus qu'utiles. Enfin, Pietro, Jean-Marc, Alasdair, Samy et Said, c'était un plaisir de pouvoir plaisanter au détour d'un couloir ou autour d'une bière.

Chez les doctorants, je suis très reconnaissant à ceux qui ont su si bien m'accueillir. Flora, Sylvain (2D), Hélène et Guillaume notamment, sans vous ça n'aurait définitivement pas été pareil ! Sonia, Paul et Antoine H., parce que vous n'étiez pas toujours à Telecom, nous n'avons pas pu autant nous voir mais je garde des bons souvenirs de nos échanges. Chenguang et Weiying, je garderai en mémoire notre partage de bureau et les moments que nous y avons partagé ! Alessio, on a fait notre thèse en parallèle et je n'aurais pas pu demander meilleur compagnon de route. Peut être qu'un jour je vais vraiment m'inscrire à la boxe avec toi qui sait. Je sais que le flambeau s'est transmis car Nicolas, Antoine P, Nicolas, Xiangli et Arthur, je suis très content que vous soyez dans les parages ! Je vous souhaite d'aller au bout de votre thèse et d'être fiers de vous. Enfin, Bastien si j'ai pu te prendre pour une nuisance intra-bureau, je n'aurais pas pu plus me tromper ! Ta bonne humeur et ton ouverture d'esprit font de toi le co-bureau parfait. Sylvain (3D), Adrien, Malik, Jeremy et les autres, vous étiez bizarres avec votre dimension en plus mais quand on arrivait à se croiser ça faisait plaisir.

Laurent, ça a toujours été un plaisir de partager avec toi que ce soit sur des questions scientifiques ou générales. J'ai aussi trouvé agréable que, bien que tu connaisses beaucoup de monde, tu restes très inclusif en conférence. C'est bien sûr très apprécié quand on est jeune chercheur. Yue, ta disponibilité et ta bonne humeur ont également été des gros plus dans cette collaboration.

Loïc, je sais que je te dois beaucoup dans cette thèse. Tu réunis des qualités qui font de toi un excellent encadrant : clarté, efficacité, disponibilité et gentillesse. Enfin, tu es toujours prêts à t'intéresser à une idée quand bien même elle pourrait être bancale ce

qui est très stimulant et encourageant.

Rémy, en plus d'être un super post-doc, tu étais bienveillant et de précieux conseil. J'espère pouvoir travailler avec toi à l'avenir.

Mélissa, merci pour ton soutien et ton amour sans lesquels, il aurait été bien moins simple d'aller au bout de ces trois ans.

Mes parents, merci de croire en moi, de m'encourager et de vous intéresser à ce que je fais.

La "Mifa", vous êtes des champions, merci à vous aussi de votre soutien sans faille.

Les anciens de prépas qui sont entrés dans Ph.D. game, c'est un peu grâce à vous que je me suis lancé la dedans. Donc, spéciale dédicace comme on dit !

Les Pandor-Margulis, merci de m'inclure autant et de votre amitié.

Les anciens collocs, j'ai fait tout ça pour vous impressionner, j'espère que ça a marché.

Enfin, merci Florence Besnard, Salimatou Yansane et Marie-Laure Chauveaux pour votre aide et votre patience (notamment vis à vis de mon temps de réponse).

À R.P.M. 1^{er}
et les suivants

J'en ai vu des vertes et des pas mûres. Je le sais que tout irait sur des roulettes, s'il y avait des roulettes. Mais il n'y a pas de roulettes. A l'endroit où il devrait y avoir des roulettes il y a des boulons.

- *Un roi sans divertissement*, Jean Giono

Contents

Synthèse des travaux

Introduction

Le suivi de la terre par images aériennes ou spatiales est depuis longtemps un sujet d'intérêt dans une grande variété de disciplines allant de l'urbanisme à la séismologie en passant par l'étude de la biodiversité. La possibilité d'observer certaines scènes directement en trois dimensions apporte une information qui peut être très utile à la compréhension de certains milieux. Ainsi, de même qu'observer des organes en 3-D plutôt que de simples coupes peut être déterminant en imagerie médicale, pouvoir exploiter différents angles de vues de scènes complexes telles que des zones densément peuplées ou certains reliefs naturels trouve des intérêts dans de nombreuses applications. Ainsi l'étude des activités humaines ou le suivi de déformation du sol ou des bâtiments peuvent grandement bénéficier de ces techniques.

De nombreuses modalités existent pour obtenir une visualisation 3-D d'une scène. Utiliser plusieurs images du même objet prise sous un point de vue différent est une approche classique. De bons résultats sont d'ailleurs obtenus de cette façon à partir d'images naturelles. Ainsi des algorithmes permettant d'obtenir des modèles 3-D de zones urbaines avec une résolution centimetrique sont maintenant couramment utilisés. Une technique d'imagerie qui a fait ses preuves depuis de nombreuses années est l'imagerie par RAdio Detection and Ranging (RADAR) à synthèse d'ouverture ou SAR. Bien que la résolution atteignable est généralement en deçà de ce qui peut être obtenu en optique, cette technique d'acquisition présente d'autres avantages. Ainsi, le RADAR étant actif, des images peuvent être prises indépendamment de la luminosité ambiante. De plus les ondes émises ne sont pas stoppées par la couverture nuageuse ce qui peut se révéler utile lorsque la zone imagée est soumise à des intempéries ou pour observer un volcan actif. L'imagerie SAR peut donc être un instrument de prédilection pour observer des zones à l'épreuve des éléments.

Les images SAR correspondent à l'addition cohérente des ondes réfléchies par les différents réflecteurs présents au sol. Chaque pixel possède une information de phase qui peut être liée à la distribution volumique des réflecteurs qui se projettent dans ce pixel. La tomographie SAR exploite la différence de phase entre plusieurs images recalées pour extraire la localisation des objets au sol. Cette technique a été utilisée avec succès pour reconstruire des volume de végétation (Reigber and Moreira, 2000) (Huang *et al.*, 2010) ou de glace (Tebaldini *et al.*, 2016). Des paysages plus complexes comme des

zones urbaines denses ont également pu être reconstruits avec cette technique. Toutefois l'intensité très variable des signaux réfléchis par les bâtiments ainsi que la quasi omniprésence de phénomènes de layovers (différents réflecteurs se projetant dans la même case) rend l'utilisation de la tomographie SAR plus compliquée.

Contributions Si les zones urbaines densément construites peuvent être difficiles à analyser, elles présentent toutefois la propriété d'être très structurées. Ainsi, il est raisonnable de s'attendre à ce que des murs soient droits, des toits et des rues plats et horizontaux. Cette structuration peut être essentielle pour améliorer les reconstructions tomographiques. Deux approches peuvent être envisagées pour en tirer partie :

- Analyser les redondances dans les images SAR et se servir des pixels les plus simmilaires pour améliorer l'estimation de signal reprojeté.
- Inverser le signal SAR en s'assurant que la reconstruction réponde à des contraintes structurelles.

Le premier point a été exploité pour la première contribution présentée dans ces travaux :

Contribution (1): Critère de similarité pour large pile d'images SAR

L'algorithme Non Local - SAR (NL-SAR) est une méthode efficace pour débruiter des images SAR ainsi que des piles d'images interférométriques et/ou paramétriques. La matrice de covariance du signal est fondamentale pour la plupart des estimateurs utilisés en tomographie SAR . Une meilleure estimation de celle-ce pourrait donc aboutir à une meilleur reconstruction tomographique. Lorsque la dimension devient trop importante, les performances de NL-SAR ont cependant tendancent à diminuer. Ici nous proposons un nouveau critère robuste à l'augmentation de la dimension en modifiant le critère de similarité entre pixel utilisé par l'algorithme.

Ce nouveau critère est effectivement plus robuste à l'augmentation de la dimension et permet une sélection prenant explicitement en compte les réflecteurs les plus puissants pour chaque pixel. Cette approche reste toutefois limité car les échantillons similaires peuvent s'avérer soit trop peu nombreux soit contraints à un ensemble depoints réparti de façon isométrique autour du pixel d'intérêt. Les résultats tomographique ne sont donc pas sensiblement améliorés.

Le second axe de recherche envisagé peut sembler plus prometteur pour les paysages densément construits. En effet, l'approche Compressive Sensing (CS) (Budillon *et al.*, 2011)(Zhu and Bamler, 2010a) permettant d'inverser chaque pixel sous contrainte de parcimonie s'est imposée ces dernières années comme l'un des meilleurs estimateurs pour la reconstruction de milieu urbain par tomographie SAR. CS s'inscrit donc com-

plètement dans la deuxième approche. De plus, comme il ne dépend pas de la matrice de covariance du signal il est adapté à des milieux hétérogènes.

Contribution (2): Estimation tomographique sous contraintes exprimées en géométrie sol

L'approche CS classique en tomographie SAR n'utilise pas d'autres *a priori* que la parcimonie. Nous proposons dans ces travaux d'aller plus loin en introduisant une régularisation géométrique du volume reconstruit. Pour cela, l'estimation est faite directement sur un cube *i.e.* un ensemble de voxels et la fonction objectif à minimiser fait intervenir la norme quadratique du gradient dans les directions horizontales et verticale.

Cette extension naturelle de l'approche CS classique permet de favoriser des distributions de réflectivité qui soient structurées et de réduire les erreurs de reconstruction. Le cadre proposé dans ce nouvel algorithme pour l'introduction de contrainte spatiales est relativement simple et peut être repris pour ajouter d'autres *a priori*.

L'analyse de reconstruction 3-D par tomographie SAR n'est pas toujours facile. Ainsi, même après avoir projeté la scène dans le repère sol, la forte dynamique des voxels, la présence de lobes secondaires et/ou d'erreurs rend l'interprétation difficile. Pouvoir situer la surface urbaine *i.e.* l'ensemble des objets (murs, toits, routes, etc) ayant réfléchi l'onde incidente n'est donc pas exactement une tâche triviale.

Contribution (3): Segmentation de surface urbaine par graph-cut

La troisième contribution présentée dans ces travaux est un algorithme de segmentation de surface urbaine dans des cubes tomographiques. Il est attendu que la surface réponde à plusieurs contraintes :

- La surface est proche des voxels de forte intensité
- la surface est lisse et composée de zones planes selon les axes horizontaux et verticaux
- Chaque rayon partant du capteur vers la scène intersecte la surface exactement une fois
- La surface n'intersecte pas plus d'une fois chaque ligne verticale

La fonction de coût correspondant à ces contraintes est non-convexe et difficile à optimiser. Toutefois ce problème après discrétisation peut s'écrire comme la recherche d'une coupe minimale dans un graphe bien construit.

L'algorithme de segmentation proposé est très général au sens où n'importe quelle estimation proposant une mesure physique de la réflectivité (amplitude ou intensité) peut être utilisée en entrée. Associée à une approche parcimonieuse type CS, les résultats obtenus sont particulièrement prometteurs.

L'information apportée par la surface est, de plus, utilisable pour régler les paramètres de certains algorithme. CS est en effet sensible au réglage du paramètre de parcimonie.

Contribution (4): Reconstruction 3-D et segmentation de surface alternées

La quatrième contribution de cette thèse est l'association naturelle des deux dernières. Nous proposons ainsi un algorithme itératif où après chaque reconstruction 3-D, la surface estimée est utlisée pour raffiner le réglage du paramètre de parcimonie. Ainsi, la parcimonie est calculée comme une fonction croissante de la distance à la surface. De cette façon, plus on est loin de la surface, plus il est coûteux de trouver de la matière.

Peu d'itérations suffisent pour corriger les tomogrammes d'erreurs résiduelles et pour diminuer les artefacts générés par la minimisation de la variation totale.

Conclusion

Au moins quatre contributions originales sont présentées dans ces travaux. Si le premier axe de recherche n'a pas pu aboutir à une réelle amélioration des techniques d'inversion tomographique, le deuxième axe a été plus fructueux. Ainsi, les approches développées basées sur une régularisation structurelle exprimée en géométrie sol ont permis d'une part d'obtenir des résultats équivalants ou au dessus de l'état de l'art et de poser un cadre permettant l'utilisation de contraintes 3-D qui soient naturelles aux objets observés d'autre part. Ainsi, ces travaux peuvent ouvrir sur de nombreuses perspectives portant soit sur une amélioration des contraintes structurelles utilisées et leur optimisation, soit sur l'analyse de pile d'images SAR à partir du volume reconstruit.

Chapter 1

Introduction

1.1 Context

Exploiting multiple views of the same object to obtain a 3-D interpretation has been a research topic in many fields and with quantities of applications. When the observed object is, in fact, the earth, many different imaging techniques can be consider depending on the soil, the vegetation, the relief, the luminosity... 3-D reconstruction obtained by exploiting multiple optical images can be computed with a great precision providing centimeter resolution. Complex landscapes such as, for instance, dense urban areas can then be processed to retrieve the 3-D distribution of the scene. The 3-D rendering of urban areas has many various applications such as city management, architecture, crisis management, building and ground deformation monitoring or demography expansion. SAR images are an other well known technique to observe the earth from the sky or space. The resolution of SAR images is generally worse than what can be obtained with optical sensors although modern sensors can provide images of few centimeters resolution. The SAR imaging technique however provides other advantages. The sensor being active, images can be taken even by night or through clouds. Depending on the operating bandwidth, the wave emitted by the RADAR can penetrate some environments (vegetation, ice) and provide information about their structure or the presence of underneath objects (Reigber and Moreira, 2000) (Huang *et al.*, 2010)(Tebaldini *et al.*, 2016). Finally, as the obtained images are composed of the coherent summation of back-scattered signals, the pixels have a phase information that can be linked to their 3-D distribution.

Exploiting phase differences between SAR images to obtain the height of the scatterers is an important research field that backs to the 1980. SAR interferometry consists in estimating the height of the main back-scattering object for each pixel exploiting the phase difference between a pair of images. As more data with higher resolution became available, using more than two images became possible with the advantages to avoid fringes unwrapping, drastically lowering ambiguities or even suppressing them. Of course with more accurate information, not only one but several scatterers height could be evaluated. Depending on the imaged scene even a 3-D reflectivity distribution

could be observed.

SAR tomography is the technique consisting in evaluating the scatterers reflectivity in the 3-D space using a stack of co-registered SAR images. Many results have been shown on uniform areas with a volume distribution of scatterers such as ice or forest (Reigber and Moreira, 2000) (Lombardini and Reigber, 2003)(Guillaso and Reigber, 2005) (Huang *et al.*, 2010)(Tebaldini and Rocca, 2012). Results on urban areas also present accurate scatterers localization. Dense urban scene are however hard to process due to the very heterogeneous back-scattering mechanisms and the importance of layover phenomena. However, urban areas are very structured and geometrical priors can be used to enhance the tomographic reconstruction. In the past years algorithms exploiting the sparsity of the signal such as MUSIC (Schmidt, 1986) or the well-known (Zhu and Bamler, 2010a)(Budillon *et al.*, 2011) have led to promising results with many scatterers being retrieved. Nevertheless, the redundancy of the buildings and geometrical shapes were never used to perform the tomographic inversion. The subject of this PhD is to explore different approaches to exploit this structural information for the SAR tomographic reconstruction of urban areas.

1.2 Contributions

At least two strategies can be tried to exploit the particular geometry of urban areas:

- Analyze the redundancies in the SAR images to select pixels presenting the same behavior.
- Perform the tomographic inversion under some priors on the reconstructed result.

The first approach is motivated by the good performances of non-local denoising algorithms such as NL-SAR (Deledalle *et al.*, 2015). By exploiting similarities between pixels and patches in the images, these approaches manage to obtain a good estimation of the covariance matrix. This parameter is central for most of the estimators used in SAR tomography

Contribution (1): Similarity criterion for large stack of SAR images

We present a similarity criterion robust to the increase of SAR images in the tomographic stack. This new criterion is based on a decomposition of the signal in a deterministic, stable part and a stochastic one.
This criterion is tested over different configurations of dense urban areas.

The second approach may be actually more suited to complex landscapes such as dense urban areas. Indeed the well-known (Zhu and Bamler, 2010a)(Budillon *et al.*, 2011) approaches that estimate the reflectivity for each pixel as the sparsest projected

vector lead to correct results for urban areas. This approach may be consider as one of the top state-of-the-art method for urban landscape and may be the best suited for dense concentration of buildings. Contrary to most of the other methods used in SAR tomography, Compressive Sensing (CS) doesn't need any estimation of the covariance matrix and thus is not impacted by inaccurate parameter estimation. The main issue to add geometrical priors to the reconstruction is to deal with the distortions induced by the active side-looking sensor. Once it is done, different priors and optimization strategies can be consider.

Contribution (2): Ground coordinate based geometrical priors for SAR tomography

We present a new algorithm allowing to perform the tomographic reconstruction of a scene under structural priors. Its spirit is very close to CS as it also takes benefit from the sparsity. However, we go one step further by constructing the projection operator as a mapping from ground to radar coordinates. This allows us to express geometrical priors in the scene frame and favor smooth vertical and horizontal distribution of scatterers.

Having a framework in which the estimated reflectivity is described in its natural coordinate system opens the door to many applications. From the 3-D reflectivity one may want to retrieve the actual urban surface on which are located the scatterers. This information can moreover be used to refine the quality of the tomographic inversion.

Contribution (3): Graph-Cut based urban surface segmentation from SAR tomographic reconstruction

We present a segmentation algorithm to extract the urban surface from 3-D tomographic reconstruction expressed in ground geometry. The priors on the surface are more sophisticated than previously and allow to select shapes corresponding to the following considerations:

- The surface is located near high intensity voxels.
- The buildings are described by smooth, uniform structures. The urban surface is then expected to correspond to a Total Variation (TV) minimization along the vertical and horizontal directions.
- The scatterers on buildings are supposed to be opaque for the electromagnetic wave. The surface is then expected to go through each ray from the sensor to the scene exactly once.
- To avoid distortions in the vertical structures, it is impossible for the surface to intersect multiple time a vertical line

The cost function associated to these hypotheses is highly non-convex. However by discretizing the problem, it can be formulated as minimum cut / maximum flow search on a well constructed graph. The method proves to be efficient on real data and performs particularly well when associated to CS like tomographic reconstructions.

Contribution (4): Alternated 3-D reconstruction and surface segmentation

The estimated urban surface can be used to tune more accurately the parameters of the estimators used to perform the inversion. This strategy is used to design a new algorithm that performs iteratively the reconstruction in ground coordinates and segment the urban surface from it.

1.3 Organization of the manuscript

This manuscript is divided in two parts. First basic concepts of SAR imaging and state-of-the-art methods for SAR tomography are presented. In chapter 2, the SAR images acquisition is summarized. In chapter 3, the different statistical models used in SAR tomography are presented. Finally, the chapter 4 details different algorithms used in SAR tomography.

The second part is centered on the different developed methods during this PhD. Chapter 5 is a small introduction summarizing the presented research axes in this part. Chapter 7 is focused on covariance matrix estimation. Chapter 8 presents the proposed

tomographic reconstruction method integrating 3-D geometrical priors (submitted in (Rambour *et al.*, 2018a)). Finally, chapter 9 corresponds details the graph-cut based urban surface segmentation and the AlteRnatEd 3-D REconstruction and Surface Segmentation (REDRESS) algorithm (submitted in (Rambour *et al.*, 2018b)).

Part I

SAR tomography

Chapter 2

Synthetic Aperture Radar

2.1 Introduction

This chapter is an introduction to the SAR techniques. First, the bases of the SAR images acquisition are described and the principle of the aperture synthesis are given. Then, simple multibaseline phase model and foundation of SAR tomography are described.

2.2 Acquisition

SAR is a coherent, microwave imaging technique that produces an image of the spatially distributed complex reflectivity of the scene with a resolution depending on the RADAR characteristics. As many other active imaging techniques, the principle is to transmit an electromagnetic wave in a direction called the Line of Sight (LOS) to an object of interest. The time from the transmission to the reception of the wave gives an indication on the distance between the object and the transmitter. In SAR imagery, the sensor is embedded on a moving platform (a plane or a satellite) and illuminates the ground according to an incidence angle θ and a squint angle α. The geometry of acquisition for a single antenna is shown in Fig. 2.1.

Range resolution: After its transmission, the wave reaches a back-scattering object at a given time delay, it is then received after propagating back to the sensor. The time between the transmission and the reception is then:

$$\Delta t = \frac{2d}{c} \tag{2.1}$$

with d the distance from the RADAR to the ground and c the speed of light. The RADAR transmits short pulses towards the scene that then back-scatters part of that incoming signal depending on its composition and geometry. To be able to separate two scatterers in the range direction, their echoes must be temporally separated by more than the duration τ between two pulses. The slant range resolution is then defined as:

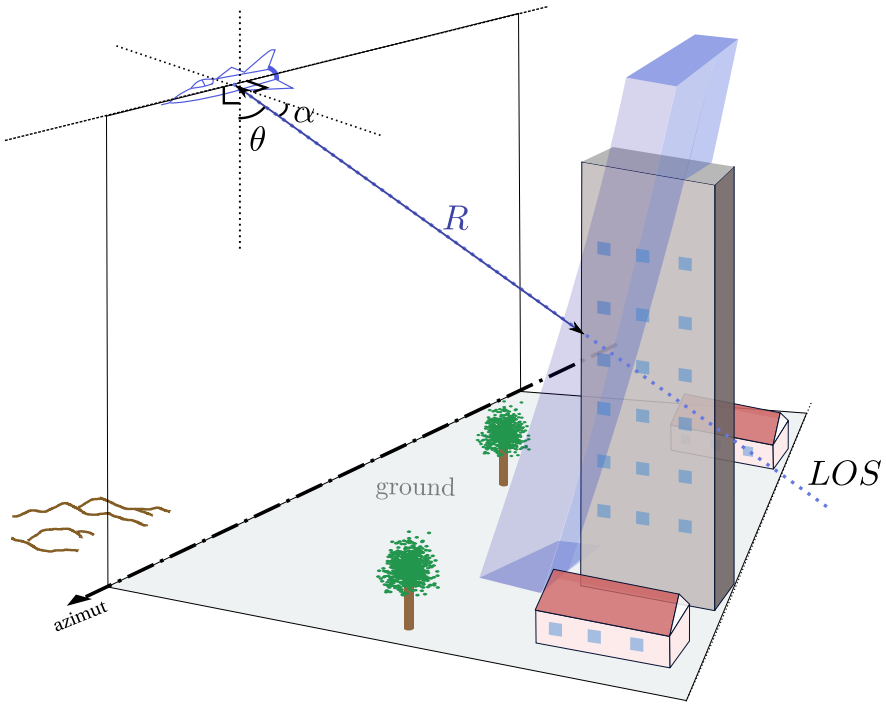


Figure 2.1 – Geometry of acquisition of a SAR image.

$$\delta_{r0} = \frac{c\tau}{2} \tag{2.2}$$

To obtain a good resolution in range one should then transmit close pulses. However, to obtain a good Signal-to-noise ratio (SNR), the pulses should be long to transmit more energy. To still achieve a good resolution, a frequency modulated pulse is transmitted. With a chirp of central frequency f_0 and frequency excursion $[f_0 - \frac{B_c}{2}, f_0 + \frac{B_c}{2}]$, the resolution is, after signal processing (see section 2.3):

$$\delta_r = \frac{c}{2B_c} \tag{2.3}$$

B_c here is the chirp bandwidth defined as $B_c = KT$ with T the transmission duration and K the rate of frequency change or chirpyness. Therefore, a high bandwidth leads to a well resolved signal. Due to the incidence angle of the sensor, the ground range resolution is (under the assumption that the squint angle α is negligible and the ground flat):

$$\delta_{gr} = \frac{c}{2B_c \sin(\theta)} \tag{2.4}$$

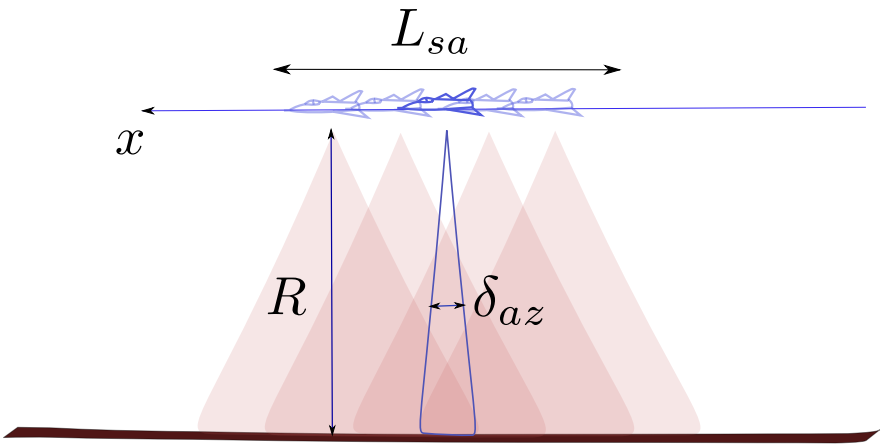


Figure 2.2 – Construction of the synthetic antenna. A point seen multiple times during the sensor motion can be localized precisely along the x axis by combining all the received echoes.

Azimuth resolution: The angular spread given by a conventional RADAR is related to the size of the antenna L and the wavelength λ by:

$$\gamma = \frac{\lambda}{L}, \tag{2.5}$$

The resolution in azimuth for a scene located at a distance R from the antenna is then:

$$\delta_{az_0} = \gamma R = \frac{\lambda R}{L}, \tag{2.6}$$

For SAR system in space, the order of magnitude of R can be $10^5 m$ and for X-band systems λ is a few centimeters which leads to an order of magnitude of the numerator in (2.6) of $10^3 m$ to achieve a metric resolution. Of course, such a dimension for an antenna mounted on a satellite is not realistic. To improve this low resolution, the SAR technique consists of coherently combine the echoes of the scatterers as they are seen from successive azimuth positions of the RADAR. This creates a synthetic antenna of length L_{sa} as illustrated in Fig. 2.2. The new angular spread is:

$$\gamma_{sa} = \frac{\lambda}{2L_{sa}}, \tag{2.7}$$

where the factor 2 in the denominator comes from the return trip of the transmitted wave. The length of the synthetic antenna correspond to the displacement of the sensor for which a given point back-scatters a signal to the antenna. L_{sa} is then given by the size of the antenna footprint $L_{sa} = \delta_{az_0}$. This gives the achievable resolution in azimuth:

$$\delta_{az} = \gamma_{sa} R = \frac{L}{2} \tag{2.8}$$

It is worth to notice that the resolution in azimuth does not depend on the distance

between the RADAR and the ground. Moreover the smaller the antenna the larger its footprint and then the better the resolution.

2.3 Basics of SAR raw data processing

Range compression: Let's consider that the emitted signal consists in a pulse defined as:

$$\forall t \in [-\frac{T}{2}, \frac{T}{2}], \underline{s}_e(t) = \exp\big(2\underline{j}\pi(f_0 t + \frac{K}{2}t^2)\big) \tag{2.9}$$

To better identify real-valued and complex-valued variables which will be useful in the next chapters, all complex-valued variables are underline. The received echo from a scatterer located at a distance R to the sensor is then

$$\underline{s}_r = \underline{s}_e\left(t - \frac{2R}{c}\right) \tag{2.10}$$

To exploit the phase modulation and achieve the resolution stated in (2.3) a matched filter is applied to the signal. This matched filter is nothing else than the transmitted pulse itself and one can verify that:

$$\underline{s}_e^*(-t) \star \underline{s}_e(t) \simeq \mathrm{sinc}_{B_c}(t)\exp\big(2\underline{j}\pi f_0 t\big) \tag{2.11}$$

where $\mathrm{sinc}_{B_c}(x) = \frac{\sin(\pi B_c t)}{\pi B_c t}$ with $B_c = KT$. After convolution by the pulse replica, the received signal becomes then:

$$\underline{s}_r = \mathrm{sinc}_{B_c}\big(t - \frac{2R}{c}\big)\exp\big(-2\underline{j}\pi(\frac{2R}{\lambda} - f_0 t)\big) \tag{2.12}$$

After centering of the spectrum, the received signal is then:

$$\underline{s}_r = \mathrm{sinc}_{B_c}\left(t - \frac{2R}{c}\right)\exp\left(\frac{-4\underline{j}\pi R}{\lambda}\right) \tag{2.13}$$

The achieved resolution is given by the width of the *sinc* function and corresponds then to the one stated in (2.3).

Azimuth compression: The azimuth compression step is analog to the previous one. The phase modulation here is introduced by the motion of the sensor. From 2.3, the distance R from the antenna at position x and a point on the ground is given by:

$$R = \sqrt{R_0^2 + (x - x_0)^2} \tag{2.14}$$

R_0 is the distance from a reference point along the track to the target. As the distance from the sensor to the target is generally higher than the synthetic antenna, a second

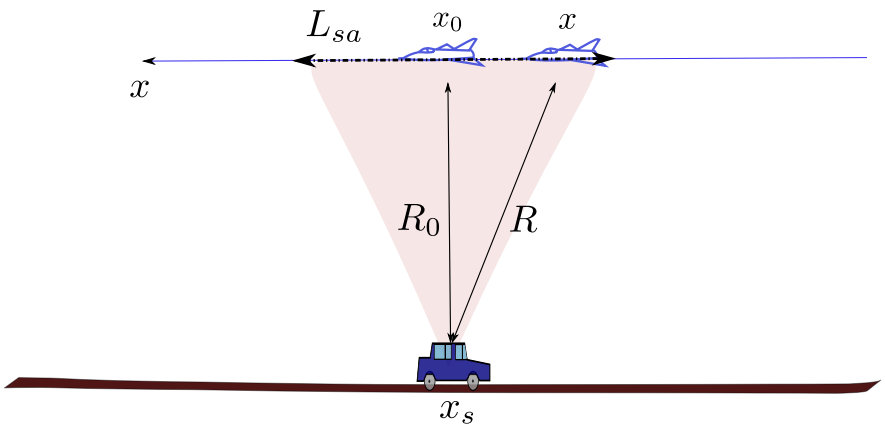


Figure 2.3 – Construction of the synthetic antenna. The phase variation induced by the motion of the sensor is exploited for the azimuth compression.

order approximation gives:

$$R = R_0 + \frac{(x - x_0)^2}{2R_0} \tag{2.15}$$

The received signal from the antenna in x is then (by a change of origin):

$$\underline{s}_r = \text{sinc}_{B_c}\left(t - \frac{2R}{c}\right) \exp\left(\frac{-4\underline{j}\pi R_0}{\lambda}\right) \exp\left(\underline{j}\pi K_x x^2\right) \tag{2.16}$$
$$\text{with } K_x = \frac{2}{\lambda R_0}$$

From the relation (2.14), the convolution of the signal by its replica gives then:

$$\underline{s}_r = \text{sinc}_{B_c}\left(t - \frac{2R}{c}\right) \text{sinc}_{B_x}(x) \exp\left(\frac{-4\underline{j}\pi R_0}{\lambda}\right) \tag{2.17}$$

The bandwidth is defined as $B_x = K_x L_{sa}$. The azimuth resolution is then:

$$\delta_{az} = \frac{1}{B_x} = \frac{\lambda R_0}{2L_{sa}} = \frac{L}{2} \tag{2.18}$$

These two processes describe the basic of the range and azimuth compression for a point like signal. In practice advanced techniques have to be employed to treat the SAR images depending on the acquisition system.

2.4 SLC SAR images

The complex value $\underline{v}(x, r)$ obtained at pixel coordinates (x, r) after SAR synthesis of the SAR image corresponds to the convolution of the complex 3-D reflectivity $\underline{u}(x, y, z)$ with the Point Spread Function (PSF) of the sensor (Reigber and Moreira, 2000; Fornaro *et*

al., 2003):

$$\underline{v}(x,r) = \iiint \underline{f}(x-x', r-\rho_{y,z})\underline{u}(x,y,z). \\ \exp\left(-\frac{4\underline{j}\pi}{\lambda}\rho_{y,z} + \underline{j}a_n(x,y,z)\right)\mathrm{d}x'\mathrm{d}y\mathrm{d}z + \underline{\epsilon} \tag{2.19}$$

To avoid complicated notations x and azimuth axis coincide. Here x and r stand for the coordinates of the focused data, $\rho_{y,z}$ is the distance between the sensor and a scatterer at a position (y,z) for a given x. $a_n(x,y,z)$ is the phase shift corresponding to the Atmospheric Phase Screen (APS) contribution. $\underline{f}$ corresponds to the PSF and depends of the sensors and the processing of the data. $\underline{\epsilon}$ is a white additive Gaussian noise and models the thermal noise. The formula (2.19) is illustrated Fig. 2.4.

After apodisation, the PSF is approximated by a boxcar function that may have a near zero width. The equation 2.19 is then written:

$$\underline{v}(x,r) = \int_{(y,z)\in\Delta_r} \underline{u}(x,y,z)\exp\left(-\frac{4\underline{j}\pi}{\lambda}\rho_{y,z} + \underline{j}a(x,y,z)\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z + \underline{\epsilon} \tag{2.20}$$

The r-th radar resolution cell is defined by $\Delta_r = \{(y,z)\,|\,r-\delta_r/2 \leq \rho_{y,z} \leq r+\delta_r/2 \text{ and } z \leq z_{\max}(y)\}$ with δ_r the step in range and $z_{\max}(y)$ the maximum height for a point at (x,y) location to be illuminated by the radar. x, y and z stand for the coordinates of a reflector in 3-D space and x, r and h its coordinates in RADAR space.

The SAR range imaging system is the cause of geometrical distortions in the observed images. Indeed, due to the incidence angle of the RADAR, the structured objects presenting an angle β between their normal and the LOS inferior to 0° are stretched toward the sensor whereas the ones for which $\beta > 0°$ are compressed in the image. These effects are illustrated in Fig. 2.5. Objects presenting an angle $\beta > 0°$ are then projected backwards in the range direction as illustrated in Fig. 2.6 where a building is projected in decreasing range positions as its height increases.

As the emitted electromagnetic waves are coherent, the total backscattering amplitude is affected by a speckle effect. This phenomenon is produced by the coherent summation of all the contributions in the radar cells that may be constructive or destructive. Depending on the application and the resolution, the speckle may be considered as a noise. In this document, it is rather described as a phenomenon corresponding to complex back-scattering mechanisms.

2.5 Phase model and calibration

The topography of a scene can be obtained through the analysis of the phase difference in a stack of N co-registered SAR images. Even if theoretically the 3-D location of the scatterers could be obtained directly from the different $\rho_{n;y,z}$ for $n \in \{1,\cdots,N\}$, it is generally more convenient to work with the optical paths differences *ie.* the in-

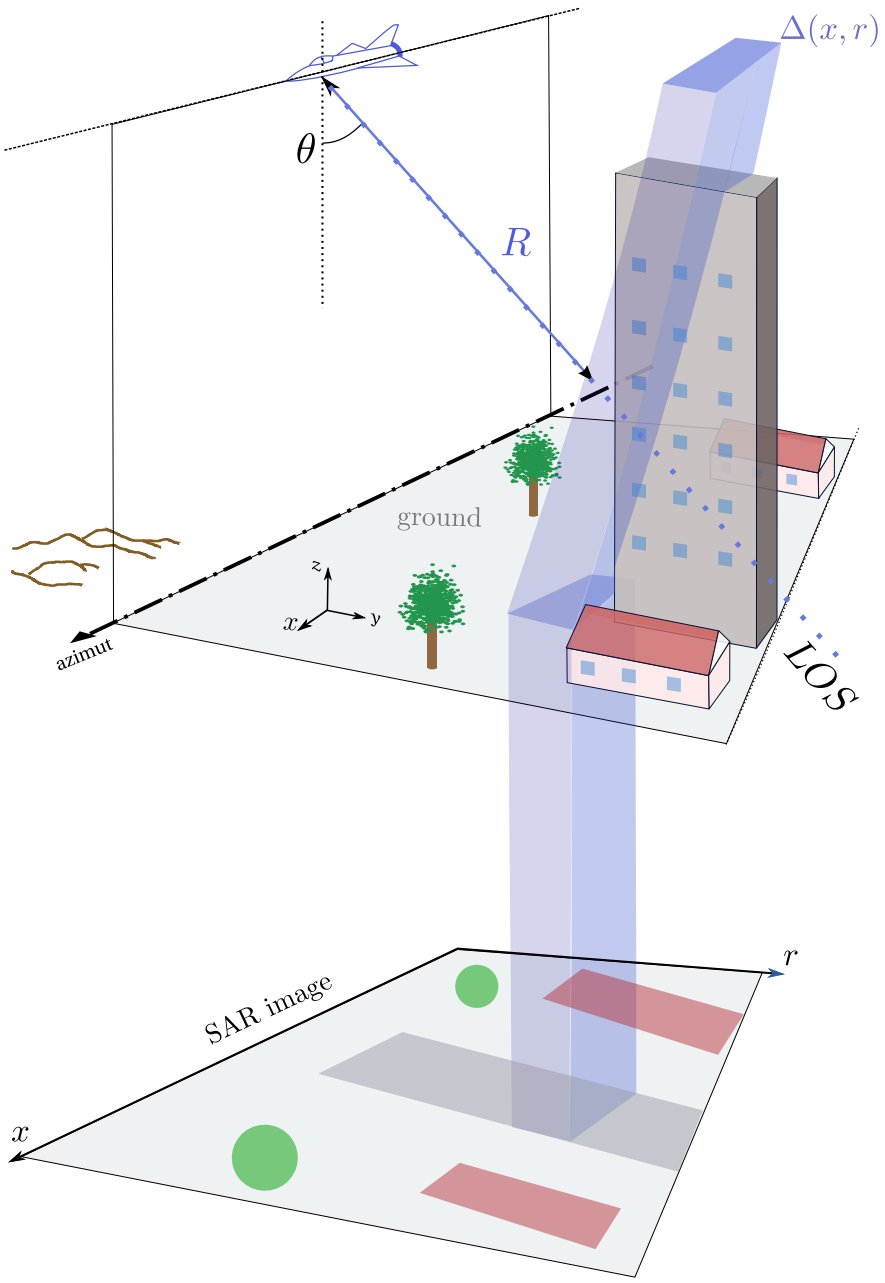


Figure 2.4 – The value of each pixel in the Single-Look Complex (SLC) SAR image corresponds to the coherent summation of the back-scattered signals inside a given radar cell of resolution.

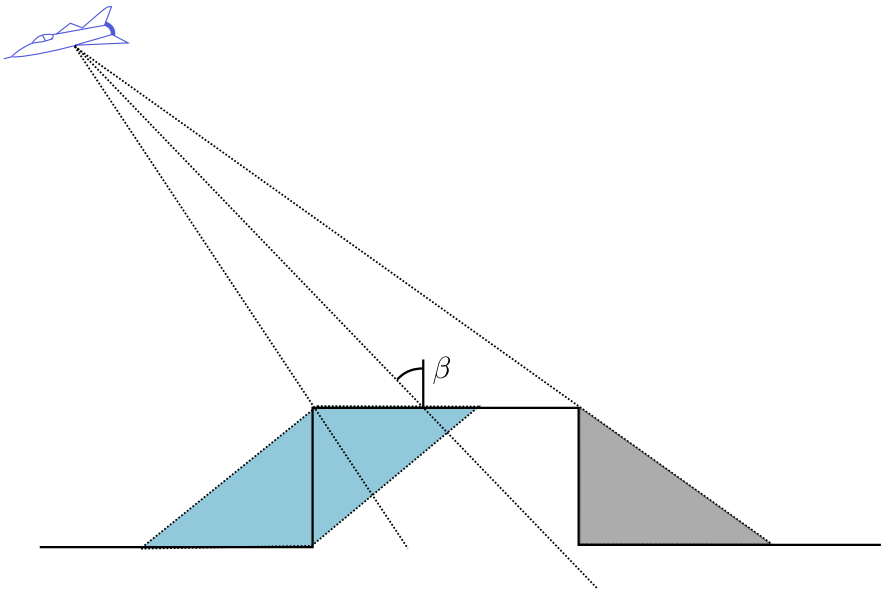
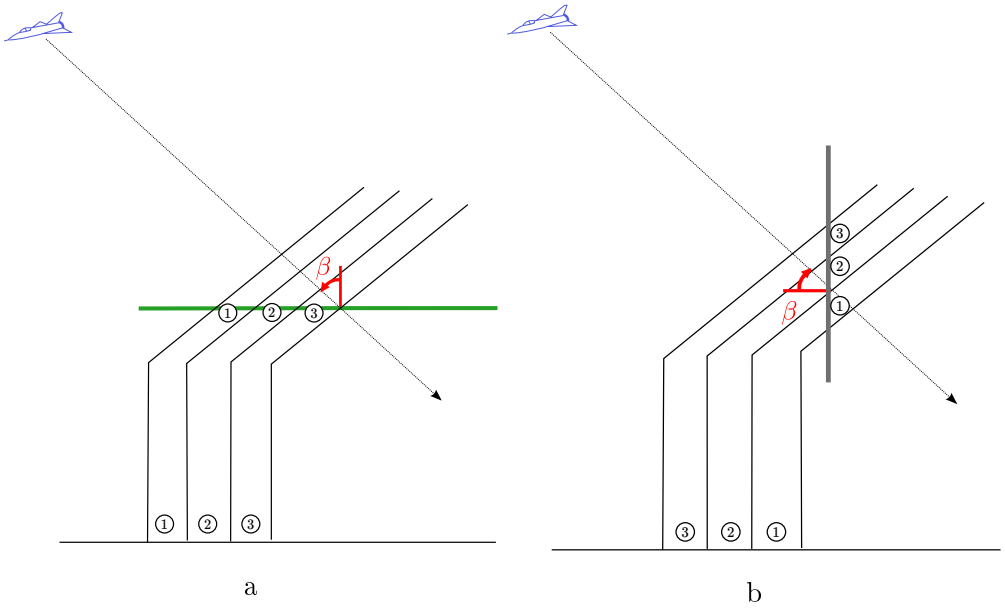


Figure 2.5 – The gray areas cannot be seen by the sensor as it is hidden by the stair step structure. In the blue areas different elements are projected in the same radar resolution cells producing a layover effect.



a b

Figure 2.6 – Distortions induced by the side-looking sensor. When $\beta > 0°$ the order is preserved (a) whereas it is inverted when $\beta > 0°$ (b).

terferograms. This way, the height or elevation of scatterers can be seen as directly proportional to the phase. To be able to use classical phase calibration algorithms and be in the same framework as the other state-of-the-art tomographic estimators, the SAR images are used as interferometric pile through this manuscript. Starting from the set of SAR images, the interferometric SAR stack is built by removing the phase of the master image to every SAR image. For simplicity the first image of the stack is chosen as the master one. The value of a pixel $\underline{v}_n(x,r)$ from the n^{th} is then:

$$\underline{v}_n(x,r) = \int_{(y,z)\in\Delta_r} \underline{u}(x,y,z)\exp\left(-\underline{j}\varphi_{n;x,y,z}\right)\mathrm{d}y\mathrm{d}z + \underline{\epsilon} \tag{2.21}$$

where the interferometric phase is defined as:

$$\varphi_{n;x,y,z} = -\frac{4\underline{j}\pi}{\lambda}(\rho_{n;y,z} - \rho_{1;y,z}) + \underline{j}(a_n(x,y,z) - a_1(x,y,z)) \tag{2.22}$$

The optical path difference is generally split into two terms relative to the elevation $h(y,z)$ and the range variation of the scatterer $\Delta\rho_{1;y,z}$ from the center of the RADAR cell (D'Aria *et al.*, 2010):

$$\rho_{n;y,z} - \rho_{1;y,z} = \frac{b_n h(y,z)}{r} + \frac{b_n \Delta\rho_{1;y,z}}{r\tan(\theta)} \tag{2.23}$$

The distance difference in (2.23) is obtained using a first order approximation whereas a second order one was used for the azimuth compression in (2.15). This means that a phase factor depending on the square of the distance variation will corrupt the estimation of $\underline{u}(x,y,z)$ and needs to be compensated via post-processing. However, here as in the vast majority of the tomographic applications, we are mostly interested in the amplitude of the reflectivity and the second order term is left in the phase of the reflectivity.

The analysis of interferometric data has a longer history than SAR tomography and many phase models have been proposed for various applications and the scene. Due to the high concentration of dihedral and trihedral structures occurring in urban areas, many bright point-like echoes appear on the corresponding SAR images. These points are generally very stable between the acquisitions and the corresponding pixels are well described by the PS model (D'Aria *et al.*, 2010). The PS interferometric phase model for a point located in (y,z) is built as a linear combination of its elevation, its potential motion along the LOS direction and linear perturbation along the azimuth and range direction:

$$\varphi_{n;x,y,z} = a_0 + k_x^{(n)}x + k_r^{(n)}r + \xi_n h(y,z) + t_n v + e \tag{2.24}$$

where a_0 is a constant phase offset for all the images, $k_x^{(n)}$ and $k_r^{(n)}$ are the unknown coefficients of the linear phase perturbation across the image, $\frac{\xi_n}{2\pi} = \frac{2b_n}{\lambda r}$ is the spatial elevation frequency associated to the sampling and b_n is the baseline n as illustrated

in Fig. 2.7, t_n the time between the first acquisition and the current one and v is the unknown scatterer slant range velocity. This phase model makes the assumption that only one scatterer is present in each cell. This strong hypothesis is true if the pixel stability is high enough so that almost all the received echoes come from the same scatterer. The system (2.24) is composed of six unknowns and has to be estimated for the whole stack of SAR images. As there is no knowledge of the unwrapped phase or on the number of scatterers in the PS-like cells, the problem formulated in 2.24 is highly non-linear and cannot be inverted from a set of L bright stable pixels. Different algorithms have been proposed these last two decades to estimate the phase model coefficients and address the problem of optimizing the phase model given by:

$$(\hat{a}_0, \hat{k}_x^{(n)}, \hat{k}_r^{(n)}, \hat{h}, \hat{t}_n, \hat{e}) = \text{argmax} \left| \sum_n^N \sum_{l=1}^L \underline{v}_n(x_l, r_l) \exp(-\underline{j}\varphi_{n;x,y,z}) \right| \tag{2.25}$$

where $\varphi_{n;x,y,z}$ is given in (2.24). The spaceborne data presented in this manuscript were processed using a similar approach as the method described in (D'Aria *et al.*, 2010) and the optimization was performed alternatively on the linear APS components and the elevation. Moreover, even if this hypothesis will not always hold, the scatterers are supposed to remain still between the acquisitions. After calibration of the data, the complex value of a pixel in the SAR interferometric images is defined by:

$$\underline{v}_n(x, r) = \int_{(y,z)\in\Delta_r} \underline{u}_n(x, y, z) \exp\big(-\underline{j}\xi_n h(y, z)\big) \mathrm{d}y\mathrm{d}z + \underline{\epsilon} \tag{2.26}$$

2.6 SAR tomography

SAR tomography is the extension of the 2-D SAR imaging to three dimensions. As conventional 2-D SAR imaging uses a synthetic aperture in the azimuth direction, 3-D SAR imaging is performed by a synthetic aperture in the elevation direction by collecting several images from parallel tracks. Using multibaseline interferometry techniques on the well-calibrated SAR images stack, it is possible to retrieve the localization of the scatterers in the third dimension. This approach allows separating scatterers mapped in the same resolution cell, which is likely to happen on the dense urban area due to the layover phenomenon.

A SAR tomographic stack consists in N SAR SLC images perfectly co-registered. Each SAR image of the stack corresponds to a slightly different trajectory of the sensor over the scene. We consider all images to have been co-registered with respect to a master image in a preprocessing step. Each image is acquired from a slightly different angle at each pass of the sensor. This angular diversity induces a different distance $\rho_{n;y,z}$ to each antenna thus a different phase shift which can be exploited to retrieve the 3-D location of the scatterers. After phase calibration and under a far-range approximation, it is generally the different baselines b_n that are used to characterize the received echoes

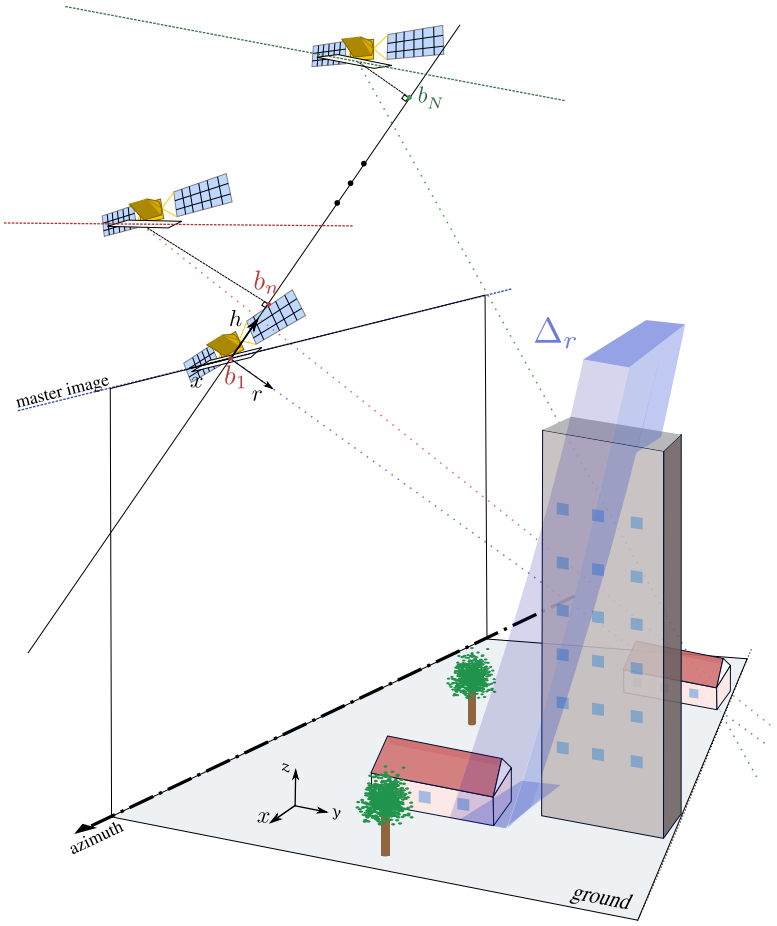


Figure 2.7 – SAR tomographic geometry of acquisition. The scene on the ground is seen multiple times from a slightly different angle by the sensor at each pass. The phase difference between the images is then exploited to retrieve the 3-D distribtution of the reflectivity of the scene.

rather than the difference in the wave propagation.

Through the equation of the SAR interferometric pixels (2.26), it can be seen that the value of $\underline{v}_n(x,r)$ is given by the Fourier transform of the 3-D reflectivity function. A common way to retrieve $\underline{u}(x,y,z)$ is then to apply a Discrete Inverse Fourier Transform (DIFT) to the vector $\underline{\boldsymbol{v}}(x,r) = \Big(\underline{v}_1(x,r), \cdots, \underline{v}_N(x,r)\Big)^T$ collecting the SLC values for all the pixels located at the position (x,r) in the different images. However, doing so, the obtained estimation of the reflectivity is described in the RADAR reference frame (x,r,h). The interpolation of the reflectivity function is then given by:

$$\hat{\underline{u}}(x,r,h) = \underline{\boldsymbol{a}}(r,h)^H \underline{\boldsymbol{v}}(x,r) \tag{2.27}$$

where the vector $\underline{\boldsymbol{a}}(x,r)$ is the steering vector associated to location (r,h) and defined as:

$$\underline{\boldsymbol{a}}(r,h) = \Big(\exp(-\underline{j}\xi_1 h), \cdots, \exp(-\underline{j}\xi_N h)\Big)^T \tag{2.28}$$

This approach only leads to poor estimation of the reflectivity as the resolution δ_h is given by the maximal orthogonal baseline Δb:

$$\delta_h = \frac{\lambda R}{2\Delta b} \tag{2.29}$$

The resolution associated to a maximal orthogonal baseline of $100m$ with a wavelength $\lambda = 0.0311m$ would then be around $55m$ in a spaceborn configuration ($R \simeq 10^5 m$).

Chapter 3

Multibaseline signal models

This chapter presents the signal statistical models used in SAR tomography. The first model corresponding to distributed sources such as rugged surfaces (floor,street,rooftops) or volumes (forest,ice) is generic whereas the second one is a derivation of the first one for strong stable scatterers.

3.1 Sensor array signal model

From now on to the end of the first part of this manuscript when no spatial regularization is involved, the subscripts identifying the location of the cells are dropped and all the equations are relative to one radar cell. In the case where different pixels are involved as in the empirical covariance matrix computation, they will receive a discrete subscript but no 2-D or 3-D spatial information.

Under the hypothesis that there is a finite number D of point like scatterers in the observed cell, we define the vector $\boldsymbol{h} = \left(h_1 \cdots h_D\right)^T$ containing all the elevations of the different scatterers in the observed cell. For $1 \leq d \leq D$, the vector $\underline{\boldsymbol{u}}_d = \left(\underline{u}_{1,d} \cdots \underline{u}_{N,d}\right)^T$ collects the complex reflectivity of one scatterer for each acquisition. After discretization of (2.21), we can express $\underline{v}_n$, the SLC value of the pixel corresponding to the n^{th} track as the sum of the complex signals back-scattered by each of the D scatterers:

$$\underline{v}_n = \sum_d \underline{u}_{n,d} \exp(-\underline{j}\xi_n h_d) + \underline{\epsilon}\,, \tag{3.1}$$

Two models can be considered depending on the behavior of the scatterers reflectivity from one image to another. The first one is general and does not make any assumption on the scatterers correlation. The second one is adapted to very bright stable scatterers or campaign where the acquisitions were close enough in time to provide coherent images with no temporal decorrelation. The different models and the type of scatterer distribution they are considering are illustrated in Fig. 3.1.

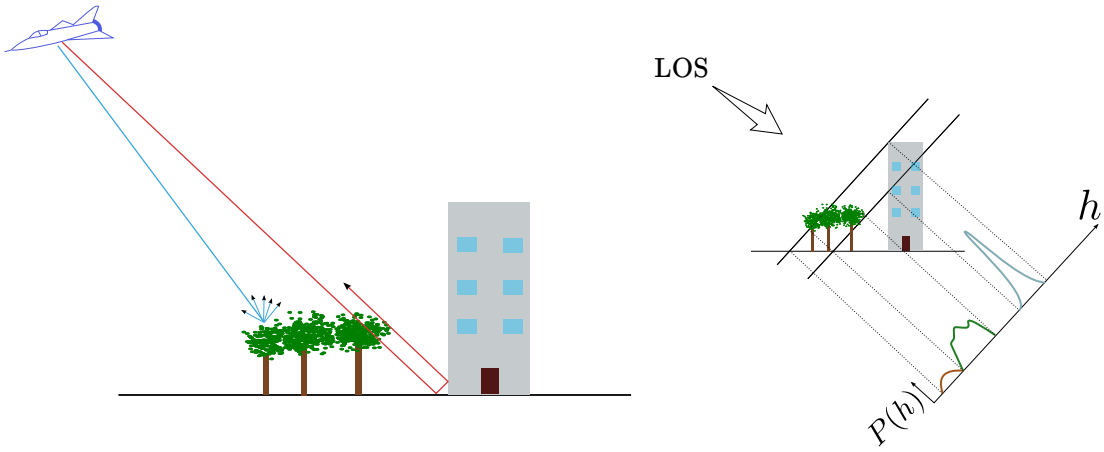


Figure 3.1 – Models depending on the structure of the back-scattering elements: distributed scatterers like vegetation or rough objects back-scatter the wave in all the directions. The signal from these scatterers obeys the unconditional model where no stability assumption can be made between the acquisitions. Dihedral or trihedral objects reflect most of the wave in the sensor direction. These scatterers follow the conditional model. The signal retrieved from distributed scatterers is generally of low intensity and poorly localized compared to conditional ones.

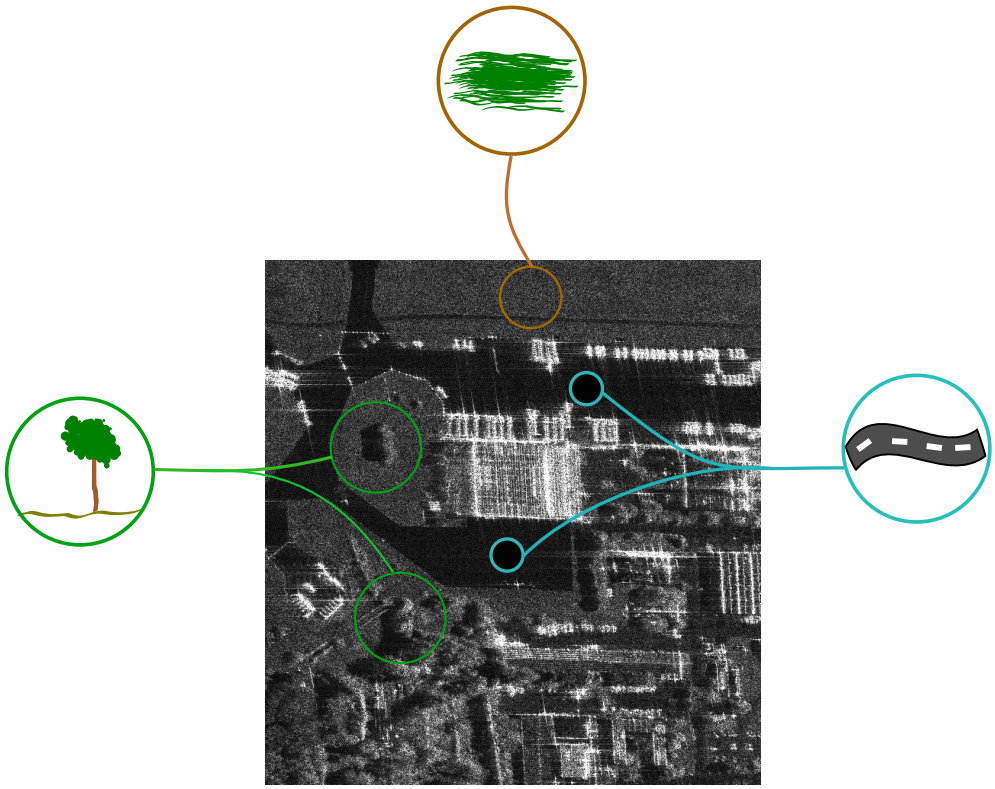


Figure 3.2 – Example of objects back-scattering a signal following the unconditional model in a SAR image obtained from the sensor SETHI of the ONERA. This model is well suited for natural rough elements such as trees or grass. Depending on their granularity and the used bandwidth, roads may back-scatter some signal following this model.

3.2 Unconditional (distributed) signal model

The first model named unconditional or stochastic signal model corresponds to elements presenting a varying reflectivity across the images. Such signals are generally reflected by rough flat surfaces or volumic scatterer distributions as seen in Fig. 3.2. Without any assumption on the scatterers correlation, the vector $\underline{\boldsymbol{v}}$ collecting the signals received by all the antennas is written:

$$\underline{\boldsymbol{v}} = \sum_d \underline{\boldsymbol{u}}_d \odot \underline{\boldsymbol{a}}(h_d) + \underline{\boldsymbol{\epsilon}} \tag{3.2}$$

with $\odot$ being the Schur-Hadamard product (elementwise multiplication). $\underline{\boldsymbol{a}}(h_d)$ is the steering vector for the elevation h_d:

$$\underline{\boldsymbol{a}}(h_d) = \begin{pmatrix} \exp(-\underline{j}\xi_1 h_d) & \cdots & \exp(-\underline{j}\xi_N h_d) \end{pmatrix}^T \tag{3.3}$$

The noise $\underline{\boldsymbol{\epsilon}}$ is supposed to be a stationary, independent, white Gaussian noise with power σ_n^2 and covariance matrix $\sigma_n^2 \underline{\boldsymbol{I}}_N \in \mathbb{C}^N$, ($\boldsymbol{I}_N$ being the matrix identity of size N). The vectors $\underline{\boldsymbol{u}}_d$ are modeled as the product of a random vector $\boldsymbol{x}_d$ times their amplitude:

$$\underline{\boldsymbol{u}}_d = \sqrt{\tau_d}\underline{\boldsymbol{x}}_d\,, \tag{3.4}$$

τ_d stand for the d^{th}-scatterer intensity and $\underline{\boldsymbol{x}}$ is the speckle effect associated to this scatterer. $\underline{\boldsymbol{x}}_d$ is modeled as a stationary, circular Gaussian random variable with zero mean and covariance matrix $\underline{\boldsymbol{C}}_d = E\{\underline{\boldsymbol{x}}_d \underline{\boldsymbol{x}}_d^H\}$ and $[C_d]_{ii} = 1,\ \forall 1 \leq i \leq N$. The covariance matrix of the signal is then given by:

$$\underline{\boldsymbol{R}} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^h\} = \sum_d \tau_d \underline{\boldsymbol{C}}_d \odot \underline{\boldsymbol{a}}(h_d)\underline{\boldsymbol{a}}(h_d)^H + \sigma_n^2 \underline{\boldsymbol{I}}_N \tag{3.5}$$

Another form of the previous equation without the Schur-Hadamard product is:

$$\underline{\boldsymbol{R}} = \sum_d \tau_d \underline{\boldsymbol{L}}_d \underline{\boldsymbol{C}}_d \underline{\boldsymbol{L}}_d^h + \sigma_n^2 \underline{\boldsymbol{I}}_N \tag{3.6}$$

where the matrix $\underline{\boldsymbol{L}}_d \in \mathbb{C}^N$ is the diagonal matrix associated to $\underline{\boldsymbol{a}}(h_d)$.

3.3 Conditional (determinist) signal model

The second model is called conditional or deterministic by simplification even though it is not associated to a fully deterministic model since the additive noise is still present.

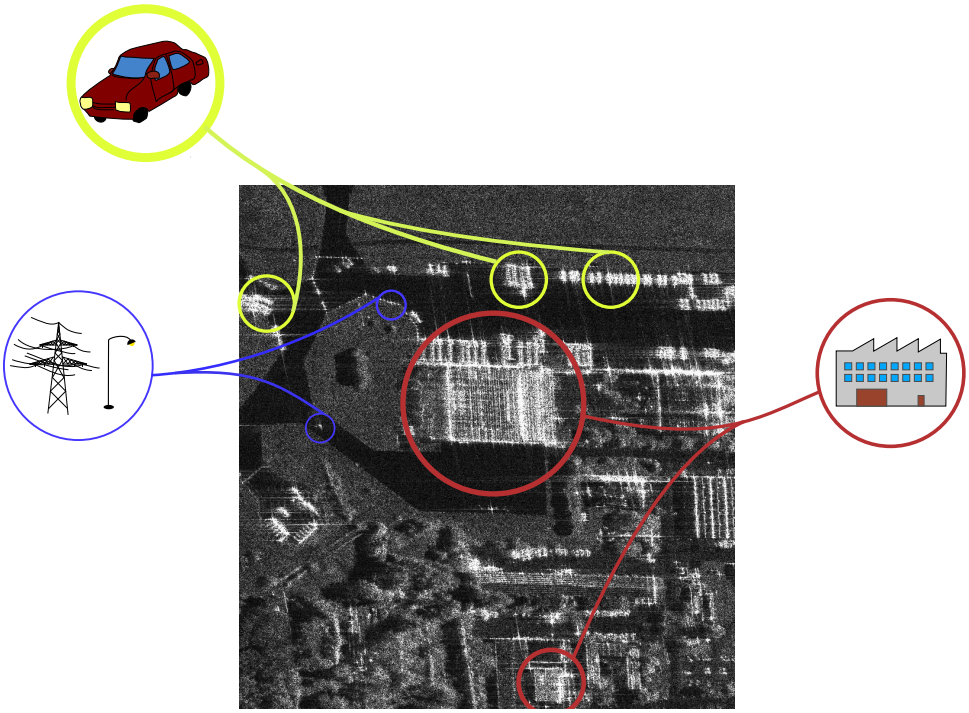


Figure 3.3 – Example of objects back-scattering a signal following the conditional model in a SAR image obtained from the sensor SETHI of the ONERA. Most of these strong bright scatterers come from man-made objects such as buildings, cars or pylons.

The conditional model is used to describe fully coherent signals presenting no fluctuation from one image to another. For repeat pass acquisitions, only point-like scatterers may be seen as constant in time and are generally formed of dihedral or trihedral elements such as corners, walls, windows... This kind of scatterer is generally found on artificial structures and are characteristic of urban areas as illustrated in Fig. 3.3.

As the vectors $\underline{\boldsymbol{u}}_d$ are now considered constant, the former equation (3.2) can be written as:

$$\underline{\boldsymbol{v}} = \sum_d \underline{u}_d \, \underline{\boldsymbol{a}}(h_d) + \underline{\boldsymbol{\epsilon}} = \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}} + \underline{\boldsymbol{\epsilon}} \tag{3.7}$$

$\underline{\boldsymbol{A}}(\boldsymbol{h}) \in \mathbb{C}^{N \times D}$ is the so-called steering matrix and its d-th column $[\underline{\boldsymbol{A}}(\boldsymbol{h})]_d$ corresponds to the steering vector $\underline{\boldsymbol{a}}(h_d)$ associated to the elevation of the d-th scatterer:

$$[\underline{\boldsymbol{A}}(\boldsymbol{h})]_d = \underline{\boldsymbol{a}}(h_d) = \begin{pmatrix} \exp(-\underline{j}\xi_1 h_d) & \cdots & \exp(-\underline{j}\xi_N h_d) \end{pmatrix}^T \tag{3.8}$$

$\underline{\boldsymbol{u}} = \begin{pmatrix} \underline{u}_1 \cdots \underline{u}_D \end{pmatrix}^T$ is the vector collecting the scatterers reflectivity. As the only non-deterministic part of equation (3.7) is due to the additive gaussian noise, the covariance matrix is then $\sigma_n^2 \boldsymbol{I}_N$. Many spectral estimators however are defined on the non-central

Figure 3.4 – Example of objects back-scattering a signal following the hybrid model in a SAR image obtained from the sensor SETHI of the ONERA.

covariance matrix as the unknown parameter $\boldsymbol{h}$ is contained in the mean. The non-central covariance matrix is then:

$$\underline{\boldsymbol{R}} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H\} = \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}(\boldsymbol{h})^H + \sigma_n^2 \boldsymbol{I}_N \tag{3.9}$$

with $\boldsymbol{S} = \underline{\boldsymbol{u}}\underline{\boldsymbol{u}}^H \in \mathbb{C}^{D\times D}$ being the signal covariance matrix.

3.4 Hybrid signal model

The hybrid model is the combination of the two preceding ones corresponding to a mixture of D_d fully coherent and D_u decorrelated signals:

$$\underline{\boldsymbol{v}} = \sum_{d=1}^{D_u} \underline{\boldsymbol{u}}_d \odot \underline{\boldsymbol{a}}(\boldsymbol{h}_d) + \sum_{d=D_u+1}^{D_u+D_d} \underline{u}_d\, \underline{\boldsymbol{a}}(\boldsymbol{h}_d) + \underline{\boldsymbol{\epsilon}} \tag{3.10}$$

The non-central covariance matrix is then given by:

$$\underline{\boldsymbol{R}} = \sum_{d=1}^{D_u} \underline{\boldsymbol{\tau}}_d\, \underline{\boldsymbol{C}}_d \odot \underline{\boldsymbol{a}}(\boldsymbol{h}_d)\underline{\boldsymbol{a}}(\boldsymbol{h}_d)^H + \sum_{d=D_u+1}^{D_u+D_d} \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{S}}\,\underline{\boldsymbol{A}}(\boldsymbol{h}) + \underline{\boldsymbol{\epsilon}} \tag{3.11}$$

3.5 Phase Model

It is possible to link the conditional model to the phase model given in section 2.5 when the modulus of the scatterers is temporally stable. Under this consideration, the equation (3.2) can then be written as in the conditional model with the phase model being included in the steering vectors. As stated in the section 2.5, phase models on urban areas may include potential movement of the scatterers. In this case the signal model is then:

$$\underline{v}_n = \sum_d \underline{u}_d \exp(-\underline{j}\xi_n h_d + \eta_n v_d) + \underline{\epsilon} \tag{3.12}$$

$$\text{And then, } \quad \underline{\boldsymbol{v}} = \underline{\boldsymbol{A}}(\boldsymbol{h}, \boldsymbol{v}) + \underline{\boldsymbol{\epsilon}} \tag{3.13}$$

where $\eta_n = \frac{4\pi t_n}{\lambda}$ is the so-called "velocity frequency" and t_n is the time interval between the acquisitions of the master and the n^{th} image. The scatterers slant range velocities correspond to the vector $\boldsymbol{v} = \begin{pmatrix} v_1 \cdots v_D \end{pmatrix}^T$. The matrix $\underline{\boldsymbol{A}}(\boldsymbol{h}, \boldsymbol{v})$ is again define as the concatenation of the steering vectors:

$$[\underline{\boldsymbol{A}}(\boldsymbol{h}, \boldsymbol{v})]_d = \underline{\boldsymbol{a}}(h_d, v_d) = \begin{pmatrix} \exp(-\underline{j}(\xi_1 h_d + \eta_1 v_d)) & \cdots & \exp(-\underline{j}\xi_N h_d + \eta_N v_d) \end{pmatrix}^T \tag{3.14}$$

Here, only one extension of the phase model to include the displacement of the objects is considered. Other models have been proposed including for instance the temperature (Budillon *et al.*, 2017b)(Weissgerber and Nicolas, 2016) or the clutter decorrelation (Aghababaee *et al.*, 2018). As stated before, only the non-moving scatterers model is used in this manuscript. The extension described in (3.12) is conceptually easy to add but may enlarge an already high number of unknowns when working on many pixels at once.

3.6 Conclusion on urban signal models

Urban areas are, by definition, filled with many man-made structures presenting many straight walls, metallic objects, corners, etc. The resulting SAR images present a high level of bright point-like scatterers. Many of those scatterers can be considered as stable in time and are correctly described by the conditional signal model. This model or the hybrid one are then generally used for describing back-scattering mechanisms over dense urban areas.

Chapter 4

SAR tomographic methods

4.1 Introduction

The aim of SAR tomography is to retrieve an estimation of the 3-D reflectivity of a given scene. After co-registration of the SAR tomographic stack this corresponds to the inversion of equation (2.26):

$$\underline{v}_n(x,r) = \int_{(y,z)\in\Delta_r} \underline{u}_n(x,y,z)\exp\big(-\underline{j}\xi_n h(y,z)\big)\mathrm{d}y\mathrm{d}z + \underline{\epsilon} \tag{2.26}$$

If the phases are correctly calibrated and the APS removed, in the ideal case of equispaced trajectories, focusing in the direction orthogonal to the line of sight can be simply performed by application of the inverse discrete Fourier transform:

$$\underline{\hat{\boldsymbol{u}}} = \underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})^H \underline{\boldsymbol{v}} \tag{4.1}$$

where the vector $\widetilde{\boldsymbol{h}}$ stands for the sampling along the elevation direction. The matrix $\underline{\boldsymbol{A}}(\tilde{\boldsymbol{h}})$ is then the inverse Discrete Fourier Transform (DFT) matrix. The resolution of this focusing is inversely proportional to the maximal orthogonal baseline. Increasing the number of tracks within this maximal orthogonal baseline improves the sampling in Fourier domain, hence it reduces height ambiguities. However, the vertical resolution is generally much worse compared to the resolution in azimuth and range directions. Moreover, the baselines are generally irregularly distributed which produces side-lobes higher than expected and degrades the interpretation of the reconstructed volume. Several spectral super-resolution techniques have been introduced to overcome these phenomena. This chapter details the two main categories of approaches used in SAR tomography. The first category is composed of the spectral analysis estimators coming mainly from the Direction Of Arrival (DOA) literature, they exploit the covariance matrix of the received signal. The second category corresponds to the more recent CS approaches which achieve super-resolution without resorting to an estimation of the covariance matrix and are more suited to sparse and heterogeneous areas. In

the previous section two signals models were described. Only the conditional model is considered in the following tomographic approaches as it is well suited to urban areas. Some of them are nevertheless robust to decorrelation phenomena.

Except for the unconditional Maximum Likelihood, all the presented method are used in the presented works. The methods are illustrated on simulations and on a real slice. The simulations are built following the conditional model with the scatterer having their amplitude fixed for all the images. The SNR corresponding to the additive noise is of 1.4 dB. The amplitude of the scatterers is set equal to 1 and the PSF function is taken as a *sinc* function inducing a mixture between adjacent cell. Simulated slice presents scatterers well separated in the range direction. Real data are however hard to model accurately due to various decorrelation mechanisms. It is then useful to look at results obtained from both a simple and a real scenario.

4.2 Spectral Analysis Techniques

Several estimation algorithm from the spectral analysis field that are used for SAR tomography are presented in this section. They allow estimating the parameters of a multibaseline SAR signal even if it is corrupted by a speckle effect.

4.2.1 Beamforming techniques

Beamforming is an important technique in signal array processing to estimate a DOA. This approach has been used in many different applications such as RADAR, SONAR, wireless telecommunications or medical echography. Beamforming techniques consist in finding a Finite Impulse Response (FIR) filter whose output is maximal when the received signal originates from a source located in a direction of interest.

Beamforming filters are the first algorithms used in SAR images to perform the unmixing of scatterers (Homer *et al.*, 1996). Many beamforming estimators are derived in the DOA literature and the ones used in SAR tomography are designed to suppress the noise under a directional constraint. The problem is then to find the filter $\underline{\boldsymbol{f}}(h)$ that maximizes the signal to noise ratio *i.e.*, that minimizes the noise level for a given output level:

$$\min_{\underline{\boldsymbol{f}}} E\{|\underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{\epsilon}}|^2\} \quad s.t \quad \underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{a}}(h) = 1 \tag{4.2}$$

The solution $\underline{\boldsymbol{f}}$ of the directional beamforming problem are also called Minimum Variance Distortionless Response (MVDR) filters. The resolution of this linearly constrained quadratic minimization uses the following results (P. Stoica, 1997):

Theorem 4.1 *Let $\underline{\boldsymbol{D}} \in \mathbb{C}^n$ be a positive definite matrix, $\underline{\boldsymbol{X}} \in \mathbb{C}^{n\times m}, \underline{\boldsymbol{G}} \in \mathbb{C}^{n\times k}$ and $\underline{\boldsymbol{C}} \in \mathbb{C}^{m\times k}$ with $k \leq n$. Then the following minimization problem:*

$$\min_{\underline{\boldsymbol{X}}} \underline{\boldsymbol{X}}^H \underline{\boldsymbol{D}} \underline{\boldsymbol{X}} \quad s.t \quad \underline{\boldsymbol{X}}^H \underline{\boldsymbol{G}} = \underline{\boldsymbol{C}}$$

has a unique solution given by: $\underline{\boldsymbol{X}}_0 = \underline{\boldsymbol{D}}^{-1}\underline{\boldsymbol{G}}\left(\underline{\boldsymbol{G}}^H \underline{\boldsymbol{D}}^{-1}\underline{\boldsymbol{G}}\right)^{-1}\underline{\boldsymbol{C}}^H$

Conventional Beamforming: According to equation (3.9), the term to minimize in the equation (4.2) can be rewriten using the noise covariance matrix $\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}_N$:

$$E\{|\underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{\epsilon}}|^2\} = \underline{\boldsymbol{f}}(h)^H \boldsymbol{R}_n \underline{\boldsymbol{f}}(h) = \sigma_n^2 \underline{\boldsymbol{f}}^H \underline{\boldsymbol{f}} \tag{4.3}$$

The previous optimization problem then becomes:

$$\min_{\underline{\boldsymbol{f}}} \underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{f}}(h) \quad s.t \quad \underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{a}}(h) = 1 \tag{4.4}$$

From the Theorem 4.1, the filter solution of the conventional beamforming problem is then:

$$\underline{\boldsymbol{f}}_{BF}(h) = \frac{\underline{\boldsymbol{a}}(h)}{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{a}}(h)} = \frac{\underline{\boldsymbol{a}}(h)}{N} \tag{4.5}$$

The estimation of the power for the elevation h is given by the filtered signal:

$$P_{BF}(h) = E\{|\underline{\boldsymbol{f}}_{BF}(h)^H \underline{\boldsymbol{v}}|^2\} = \frac{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{R}} \underline{\boldsymbol{a}}(h)}{N^2} \tag{4.6}$$

with $\underline{\boldsymbol{R}}$ the non-centered covariance matrix of the signal $\underline{\boldsymbol{R}} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H\}$. An "image" interpretation of the conventional beamforming for an additive white gaussian noise can be expressed through the empirical covariance matrix:

$$\underline{\hat{\boldsymbol{R}}} = \frac{1}{L}\sum_{l=1}^{L} \underline{\boldsymbol{v}}_l \underline{\boldsymbol{v}}_l^H \tag{4.7}$$

where L corresponds to the number of samples used for the estimation. The expression of the estimated power using this covariance matrix estimation is then:

$$P_{BF}(h) = \frac{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{R}}\, \underline{\boldsymbol{a}}(h)}{N^2} = \frac{1}{LN^2}\sum_{l=1}^{L} |\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{v}}_l|^2 \tag{4.8}$$

When using the empirical covariance matrix the power of the filtered signal can be seen as an averaging of the inverse DFT of the signal for different pixels. Fig 4.1 illustrates conventional Beamforming performances for scatterer localization and on real data.

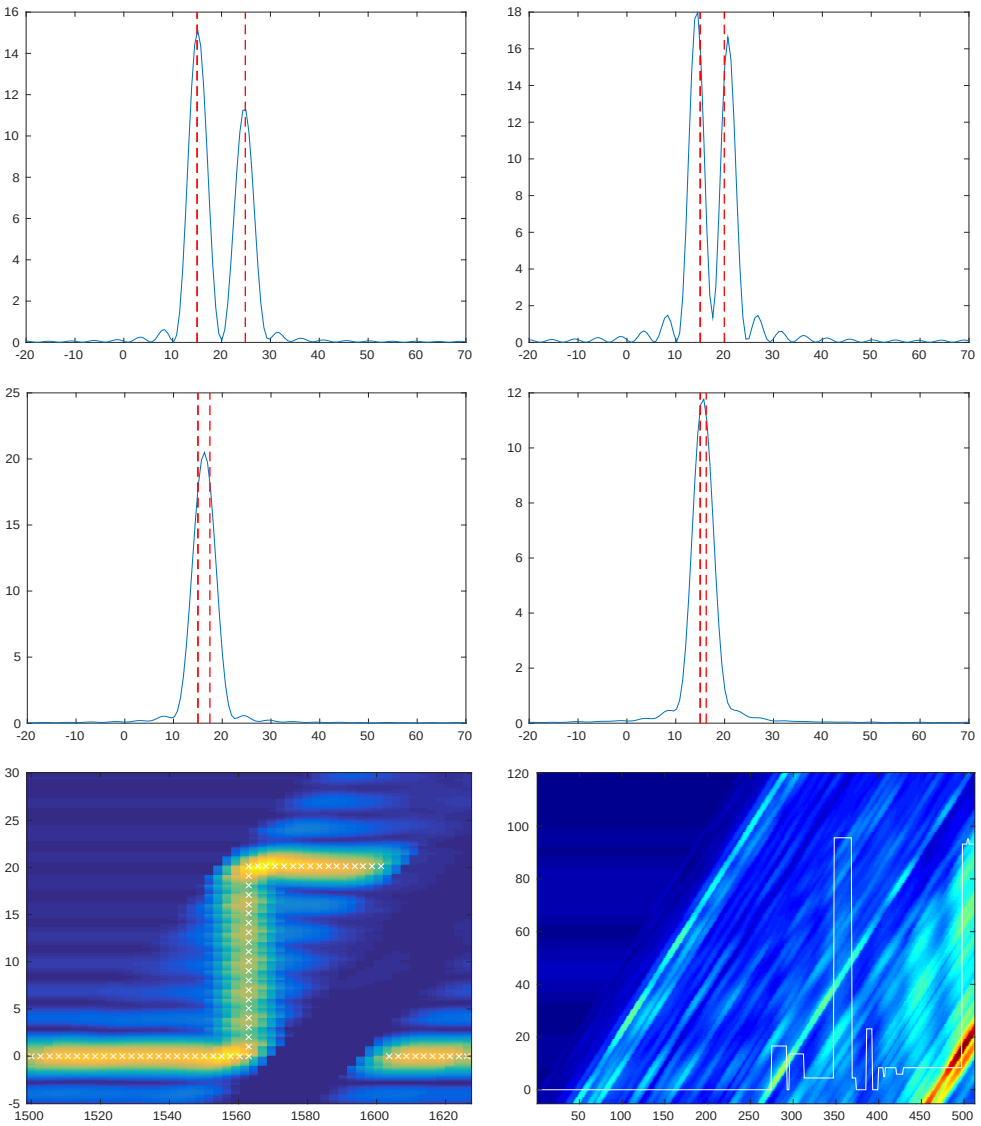


Figure 4.1 – Unmixing power of the conventional beamforming. The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue curves is the estimated intensity P_{BF}. The last row illustrates the estimated signal intensity for a simulation where the scatterers are well separated and a real urban slice. The ground truth is indicated in white.

Capon Beamforming: The conventional beamforming filter is derived from a model of the noise distribution. However when dealing with SAR images stacks, specially when the dimension N becomes high, or when the echoes are produced by distributed scatterers, the unconditional model is more appropriate and decorrelation must be taken into account. When no prior on the noise is available, the beamforming filter can be designed to attenuate the power from every direction other than h under a unit gain constraint (Capon, 1969):

$$\min_{\underline{\boldsymbol{f}}} \underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{R}} \underline{\boldsymbol{f}}(h) \quad s.t \quad \underline{\boldsymbol{f}}(h)^H \underline{\boldsymbol{a}}(h) = 1 \tag{4.9}$$

From the theorem 4.1, the solution is given by:

$$\underline{\boldsymbol{f}}_C(h) = \frac{\underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{a}}(h)}{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{a}}(h)} \tag{4.10}$$

The estimated power for the elevation h is then:

$$P_C(h) = E\{|\underline{\boldsymbol{f}}_C(h)^H \underline{\boldsymbol{v}}|^2\} = \frac{1}{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{a}}(h)} \tag{4.11}$$

This method has empirically shown better resolution and lobe suppression than conventional beamforming (P. Stoica, 1997)(Gini and Lombardini, 2005). When the covariance matrix is poorly estimated this estimator may however presents a more hieratic behavior than conventional beamforming due to its dependency with $1/\hat{\underline{\boldsymbol{R}}}^{-1}$.

Fig 4.2 illustrates conventional Beamforming performances for scatterer localization and on real data.

4.2.2 MUSIC

The MUSIC estimator (Schmidt, 1986) is designed to retrieve a finite number of DOA for a signal corrupted by an additive white noise. This estimator is easy to compute and is one of the first to produce a sparse representation of the estimated signal. The MUSIC estimator derivation comes from the analysis of the non-centered covariance matrix subspaces.

The signal $\underline{\boldsymbol{v}}$ consisting of the echoes produced by $D < N$ scatterers following the conditional model has a covariance matrix structure given by the equation (3.9):

$$\underline{\boldsymbol{R}} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H\} = \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{S}}\underline{\boldsymbol{A}}(\boldsymbol{h})^H + \sigma_n^2 \boldsymbol{I}_N \tag{3.9}$$

As $\underline{\boldsymbol{R}}$ is Hermitian, there exists an orthogonal basis where this matrix is diagonal. The eigenvalues and eigenvectors of $\underline{\boldsymbol{R}}$ are denoted by $\lambda_N \geq \cdots \geq \lambda_1$ and $\underline{\boldsymbol{e}}_N, \cdots, \underline{\boldsymbol{e}}_1$. Under the assumption that the kernel of $\underline{\boldsymbol{R}}$ is empty, the last $N - D$ eigenvalues are

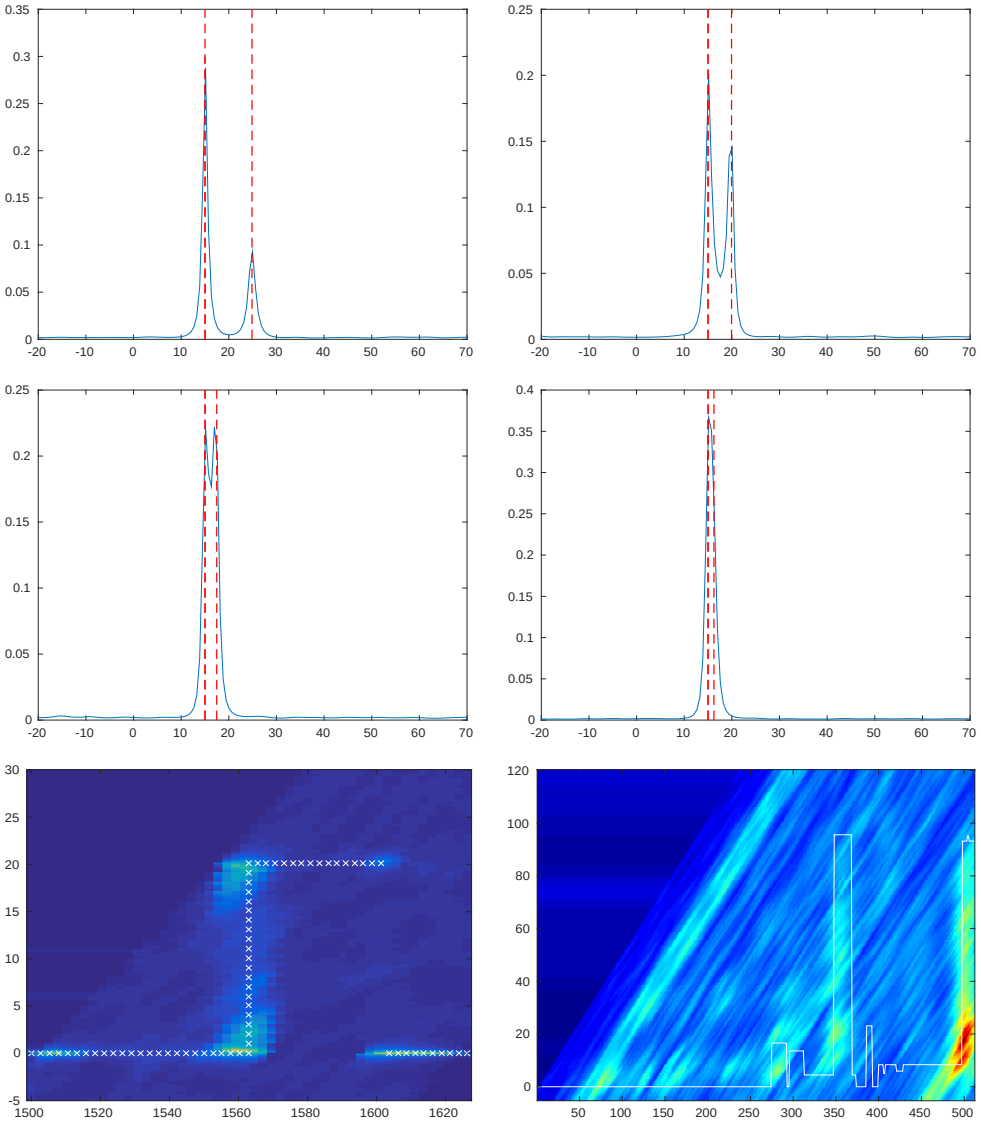


Figure 4.2 – Unmixing power of the Capon beamforming. The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue curves is the estimated intensity P_C. The last row illustrates the estimated signal intensity for a simulation where the scatterers are well separated and a real urban slice. The ground truth is indicated in white..

equal to σ_n^2. Let us divide the subspaces of $\underline{\boldsymbol{R}}$ in two groups:

- The signal subspace:

$$\underline{\boldsymbol{E}}_s = \left[\underline{\boldsymbol{e}}_1, \cdots, \underline{\boldsymbol{e}}_D\right] \quad \text{and} \quad \boldsymbol{\lambda}_s = \left[\lambda_1, \cdots, \lambda_D\right] \tag{4.12}$$

- The complementary subspace or so-called "noise subspace":

$$\underline{\boldsymbol{E}}_n = \left[\underline{\boldsymbol{e}}_{N-D}, \cdots, \underline{\boldsymbol{e}}_N\right] \quad \text{and} \quad \boldsymbol{\lambda}_n = \left[\lambda_{N-D}, \cdots, \lambda_N\right] = \left[\sigma_n^2, \cdots, \sigma_n^2\right] \tag{4.13}$$

Let $\boldsymbol{D}_s$ and $\boldsymbol{D}_n$ be the diagonal matrices built from $\boldsymbol{\lambda}_s$ and $\boldsymbol{\lambda}_n$. The covariance matrix $\underline{\boldsymbol{R}}$ can now be written as:

$$\underline{\boldsymbol{R}} = \underline{\boldsymbol{E}}_s \boldsymbol{D}_s \underline{\boldsymbol{E}}_s^H + \underline{\boldsymbol{E}}_n \boldsymbol{D}_n \underline{\boldsymbol{E}}_n^H \tag{4.14}$$

As the subspaces are orthogonal, multiplying the previous equation by the matrix $\underline{\boldsymbol{E}}_n$ from the right side leads to

$$\underline{\boldsymbol{R}}\,\underline{\boldsymbol{E}}_n = \underline{\boldsymbol{E}}_n \boldsymbol{D}_n = \sigma_n^2 \underline{\boldsymbol{E}}_n \tag{4.15}$$

From equation (3.9) the same matrix can also be described as:

$$\underline{\boldsymbol{R}}\,\underline{\boldsymbol{E}}_n = \sigma_n^2 \underline{\boldsymbol{E}}_n = \underline{\boldsymbol{A}}(\boldsymbol{h}) \underline{\boldsymbol{S}}\, \underline{\boldsymbol{A}}(\boldsymbol{h})^H \underline{\boldsymbol{E}}_n + \sigma_n^2 \underline{\boldsymbol{E}}_n \tag{4.16}$$

Combining equations (4.15) and (4.16) and supposing that $\underline{\boldsymbol{S}}$ is full rank leads to the following relation from which MUSIC is derived:

$$\underline{\boldsymbol{A}}(\boldsymbol{h})^H \underline{\boldsymbol{E}}_n = \boldsymbol{0} \tag{4.17}$$

The equation (4.17) states that the steering vectors associated to the back-scatterers are orthogonal to the noise subspace. The position of the scatterers can then be found by minimizing the projection:

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{h}}{\operatorname{argmin}} \underline{\boldsymbol{A}}(\boldsymbol{h})^H \underline{\boldsymbol{E}}_n \underline{\boldsymbol{E}}_n^H \underline{\boldsymbol{A}}(\boldsymbol{h})^H \tag{4.18}$$

The problem (4.18) is non-convex and depends on the unknown number of scatterers D. To avoid these difficulties $\hat{\boldsymbol{h}}$ is generally estimated by looking at the maximums of the MUSIC Pseudo Spectrum (MUSIC-PS) defined as:

$$P_{MU}(h) = \frac{1}{\underline{\boldsymbol{a}}(h)^H \underline{\boldsymbol{E}}_n \underline{\boldsymbol{E}}_n^H \underline{\boldsymbol{a}}(h)^H} \tag{4.19}$$

For discrete scatterer distributions, the MUSIC algorithm performs better than beamforming estimators. Moreover, it is also robust to multiplicative noise for a small number of scatterers as the orthogonality between the steering vectors and the noise

subspace is likely to be preserved. When the back-scattered echoes are highly correlated, the kernel of $\underline{\boldsymbol{R}}$ can be non-empty which produces a degeneracy as the orthogonality relation is not exclusive to the signal steering vectors.

Fig 4.3 illustrates MUSIC performances for scatterer localization and on real data.

4.2.3 WSF

The WSF methods are in the same spirit as MUSIC but introduce the empirical covariance matrix subspace distribution in the estimators. At least two different WSF estimators can be derived. First, it can be observed that the orthogonality relation used in the MUSIC derivation implies that, if the rank of $\underline{\boldsymbol{S}}$ is equal to D, then the range space of $\underline{\boldsymbol{A}}(\boldsymbol{h})$ coincides with the one of $\underline{\boldsymbol{E}}_s$. Then it can be stated that there is an unknown linear transformation $\underline{\boldsymbol{T}}$ such that:

$$\underline{\boldsymbol{E}}_s = \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{T}} \tag{4.20}$$

This linear relation and the orthogonal one defined in equation (4.18) can be used to derived two estimators based on the distance minimization between the matrix $\underline{\boldsymbol{A}}(\boldsymbol{h})$ and the weighted subspace of the empirical covariance matrix $\underline{\hat{\boldsymbol{R}}}$:

- The Noise Subspace Fitting (NSF) estimator obtained by minimizing the following criterion:
$$||\underline{\boldsymbol{E}}_n^H \underline{\boldsymbol{A}}(\boldsymbol{h})||^2_{F_{\underline{\boldsymbol{W}}}} \tag{4.21}$$
- The Signal Subspace Fitting (SSF) estimator obtained by minimizing the following criterion:
$$||\underline{\boldsymbol{E}}_s - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{T}}||^2_{F_{\underline{\boldsymbol{W}}}} \tag{4.22}$$

Where $||\boldsymbol{X}||^2_{F_{\underline{\boldsymbol{M}}}} = \mathrm{tr}(\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T)$ is the weighted Froebenius norm and $\underline{\boldsymbol{W}}$ is an Hermitian positive semidefinite weighting matrix. Consistent estimates of $\underline{\boldsymbol{W}}$ are based on the empirical covariance matrix subspace distribution and allow to asymptotically reach the Cramer-Rao lower bound. They are given by (Huang *et al.*, 2016; Viberg and Ottersten, 1991; Stoica and Sharman, 1990):

$$\underline{\boldsymbol{W}}_{SSF} = \left(\underline{\boldsymbol{D}}_s - \sigma_n^2 \boldsymbol{I}_N\right)^2 \boldsymbol{D}_s^{-1} \tag{4.23}$$
$$\underline{\boldsymbol{W}}_{NSF} = \left(\underline{\boldsymbol{A}}(\boldsymbol{h})^H \underline{\boldsymbol{E}}_s \underline{\boldsymbol{W}}_{SSF}^{-1} \underline{\boldsymbol{E}}_s^H \underline{\boldsymbol{A}}(\boldsymbol{h})\right)^{-1} \tag{4.24}$$

WSF techniques are supposed to provide high elevation resolution in SAR tomography. However, the cost functions are non-convex and multimodal and thus hard to optimize. A way proposed in (Viberg *et al.*, 1991) is to choose the result given by a suboptimal minimization criterion such as MUSIC as an initialization.

Fig 4.4 illustrates WSF performances for scatterer localization and on real data.

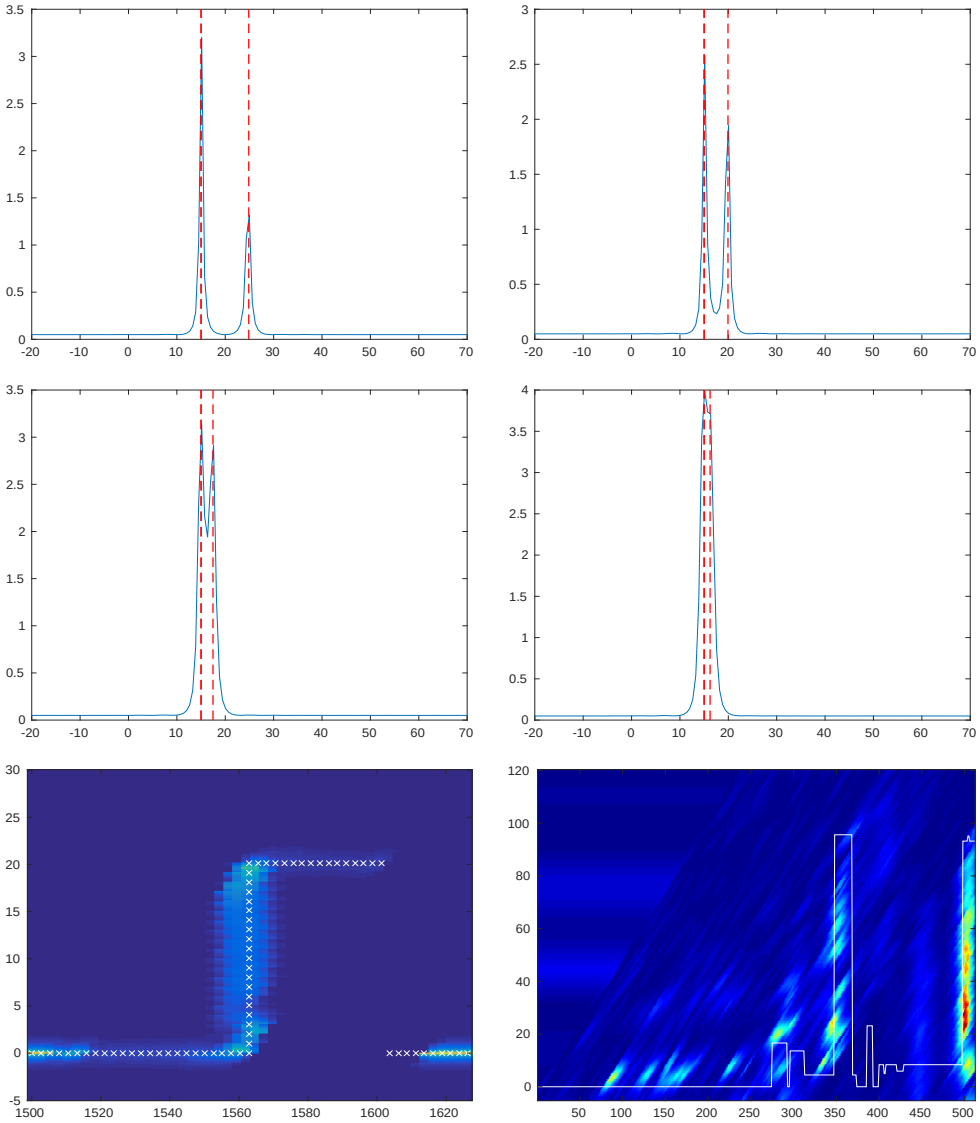


Figure 4.3 – Unmixing power of MUSIC through its pseudo-spectrum. The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue curves is the pseudo-spectrum P_{MU} value. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice. The ground truth is indicated in white.

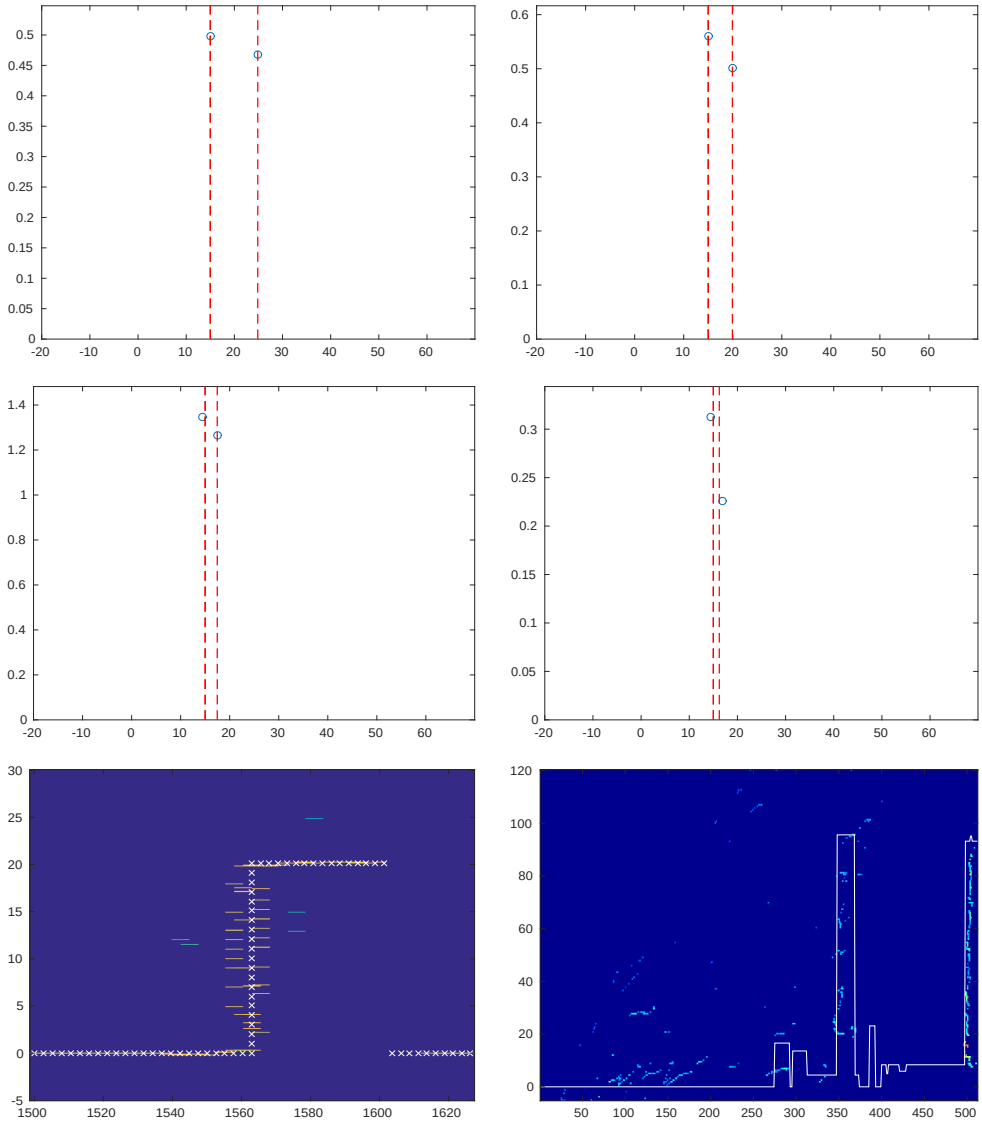


Figure 4.4 – Unmixing power of WSF (here SSF). The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue dot the estimated scatterers position and intensity. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice.

4.2.4 ML

The ML techniques could refer to different approaches depending on the signal model used. For a signal following the conditional model (3.7), the negative log-likelihood is given by (up to a constant):

$$\mathcal{L}(\boldsymbol{h}, \underline{\boldsymbol{u}}) = ||\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}||_2^2 \tag{4.25}$$

When no assumption can be made on the signal covariance matrix, the negative log-likelihood becomes then:

$$\mathcal{L}(\boldsymbol{h}, \underline{\boldsymbol{u}}) = \left(\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}\right)^H \underline{\boldsymbol{R}}^{-1} \left(\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}\right) = ||\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}||^2_{\underline{\boldsymbol{R}}^{-1}} \tag{4.26}$$

where $||\boldsymbol{x}||^2_{\boldsymbol{M}} = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}$ is the weighted ℓ_2 norm. The minimization of $\mathcal{L}$ in (4.25) and (4.26) with respect to both $\boldsymbol{h}$ and $\underline{\boldsymbol{u}}$ can be hard to achieve due to the phase dependency in $\boldsymbol{h}$ and the unknown number of scatterers. For given elevations, the conditional ML coincides with the Least-Square (LS) estimate of the signal. A LS approximation of the reflectivity $\hat{\underline{\boldsymbol{u}}}$ can thus be computed after using a spatial component estimator such as MUSIC or WSF .

When multiple independent and identically distributed samples are available, the negative log-likelihood for the conditional model can be written using the empirical covariance matrix (Stoica and Sharman, 1990):

$$\mathcal{L}(\boldsymbol{h}) = \mathrm{tr}\left(\boldsymbol{I}_N - \underline{\boldsymbol{A}}(\boldsymbol{h}) \left(\underline{\boldsymbol{A}}(\boldsymbol{h}) \underline{\boldsymbol{A}}(\boldsymbol{h})^H \right)^{-1} \underline{\boldsymbol{A}}(\boldsymbol{h}) \hat{\underline{\boldsymbol{R}}} \right) \tag{4.27}$$

This function is highly nonlinear, multimodal and hard to minimize in a reasonable amount of time. Moreover, the conditional ML has been proven statistically less efficient for big number of samples than WSF techniques as it does not achieve Cramer-Rao lower bound (Stoica and Sharman, 1990).

Fig 4.5 illustrates ML performances for scatterer localization and on real data.

4.2.5 M-RELAX

Multilook-RELAXation spectral estimator (M-RELAX) is an iterative algorithm that can be used to minimize (4.25) or (4.27) when multiple samples sharing the same scatterer elevation distribution are available (Li and Stoica, 1996) (Gini *et al.*, 2002). At each step, the previously estimated sources are refined to account the newly detected one. When the number of scatterers is high, the correction step may be time consuming as it must be run multiple time until convergence. This procedure is described in the following algorithm:

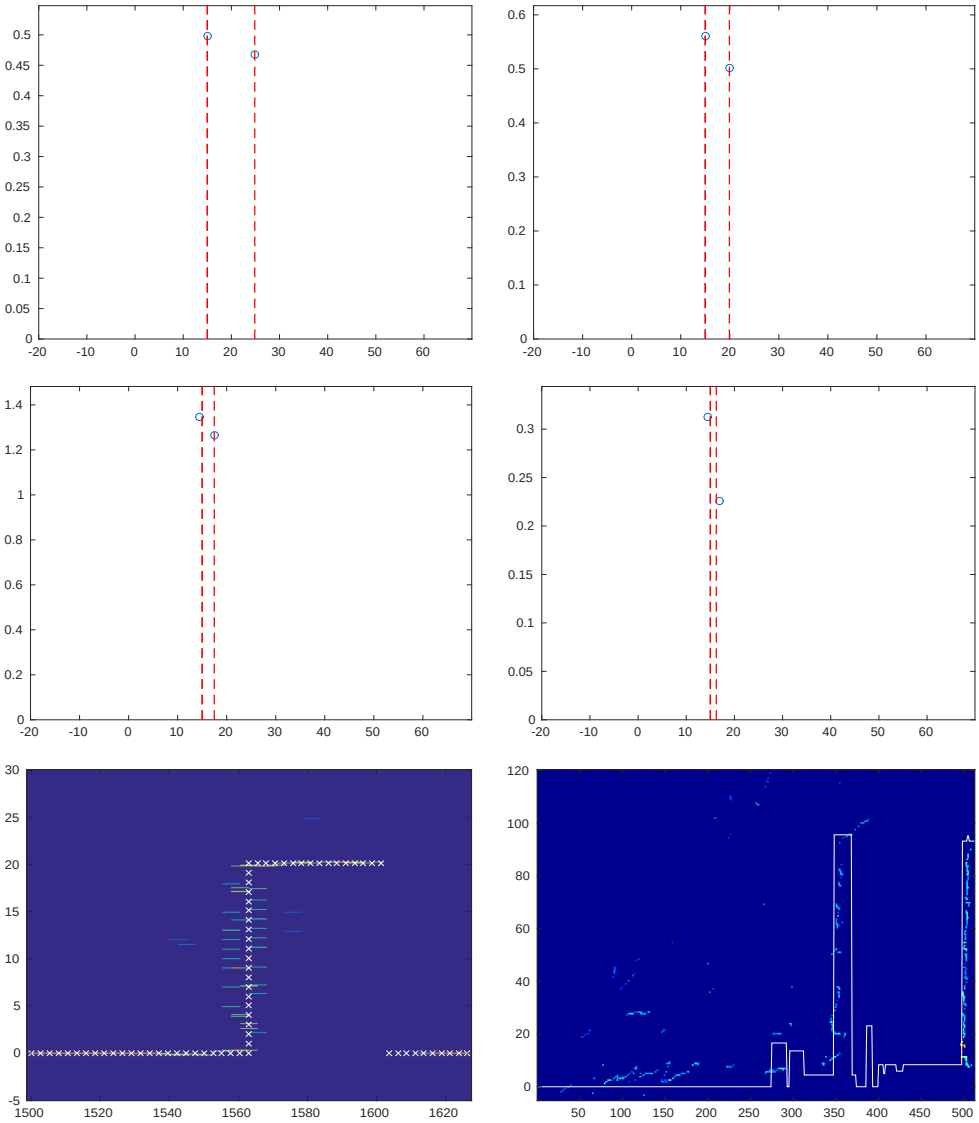


Figure 4.5 – Unmixing power of ML according to formula 4.27. The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue dot the estimated scatterers position and intensity. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice.

Algorithm 1 M-RELAX

Input: $\{\underline{\boldsymbol{v}}_l\}_{l=1}^{L}$ (L *i.i.d.* samples) D (maximum number of scatterers)
κ (precision)
Output: $\{\underline{\hat{\boldsymbol{u}}}_l\}_{l=1}^{L} \in \mathbb{C}^D$ (discrete complex vector of reflectivity for each sample)
$\hat{\boldsymbol{h}} \in \mathbb{C}^D$ (discrete complex vector of elevation)

Initialization :

$\underline{\hat{\boldsymbol{R}}} \leftarrow \frac{1}{L}\sum_l \underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H$

$\hat{h}_1 \leftarrow \underset{h}{\operatorname{argmax}}\ \underline{\boldsymbol{a}}(h)^H \underline{\hat{\boldsymbol{R}}}\underline{\boldsymbol{a}}(h)$

$\underline{\hat{u}}_{1,l} \leftarrow \frac{\underline{\boldsymbol{a}}(\hat{h}_1)^H \underline{\boldsymbol{v}}_l}{N}$

$d \leftarrow 2$

while $d < D$ **do**
  while $\hat{\kappa} < \kappa$ **do**
    $k \leftarrow d$
    while $k > 0$ **do**
      $\underline{\hat{\boldsymbol{v}}}_l^{(k)} \leftarrow \underline{\hat{\boldsymbol{v}}}_l - \sum_{\substack{i=1 \\ i\neq k}}^{d} \hat{u}_{i,l}\,\underline{\boldsymbol{a}}(h_i)$
      $\underline{\hat{\boldsymbol{R}}}^{(k)} \leftarrow \frac{1}{L}\sum_l \underline{\hat{\boldsymbol{v}}}_l^{(k)} \underline{\hat{\boldsymbol{v}}}_l^{(k)H}$
      $\hat{h}_k \leftarrow \underset{h}{\operatorname{argmax}}\ \underline{\boldsymbol{a}}(h)^H \underline{\hat{\boldsymbol{R}}}^{(k)} \underline{\boldsymbol{a}}(h)$
      $\underline{\hat{u}}_{k,l} \leftarrow \frac{\underline{\boldsymbol{a}}(\hat{h}_k)^H \underline{\hat{\boldsymbol{v}}}_l^{(k)}}{N}$
      $k \leftarrow k - 1$
    end while
    $\hat{\kappa} \leftarrow \frac{1}{L}\sum_l ||\underline{\boldsymbol{v}}_l - \sum_{i=1}^{d} \underline{\hat{u}}_l^{(i)} \underline{\boldsymbol{a}}(h^{(i)})||_2^2$
  end while
  $d \leftarrow d + 1$
end while
return $\underline{\hat{\boldsymbol{u}}}$

M-RELAX is the extension of the single look algorithm RELAX for which the elevation search steps (2) and (9) are replaced by:

$$\hat{h}^{(k)} \leftarrow \underset{h}{\operatorname{argmax}}\ \underline{\boldsymbol{a}}^H(h)\underline{\boldsymbol{v}}$$

The RELAX and M-RELAX algorithms without the correction process are forms of the CLEAN algorithm which presents no guaranty to converge to the global minimum but is much faster to optimize.

Fig 4.6 illustrates M-RELAX performances for scatterer localization and on real data.

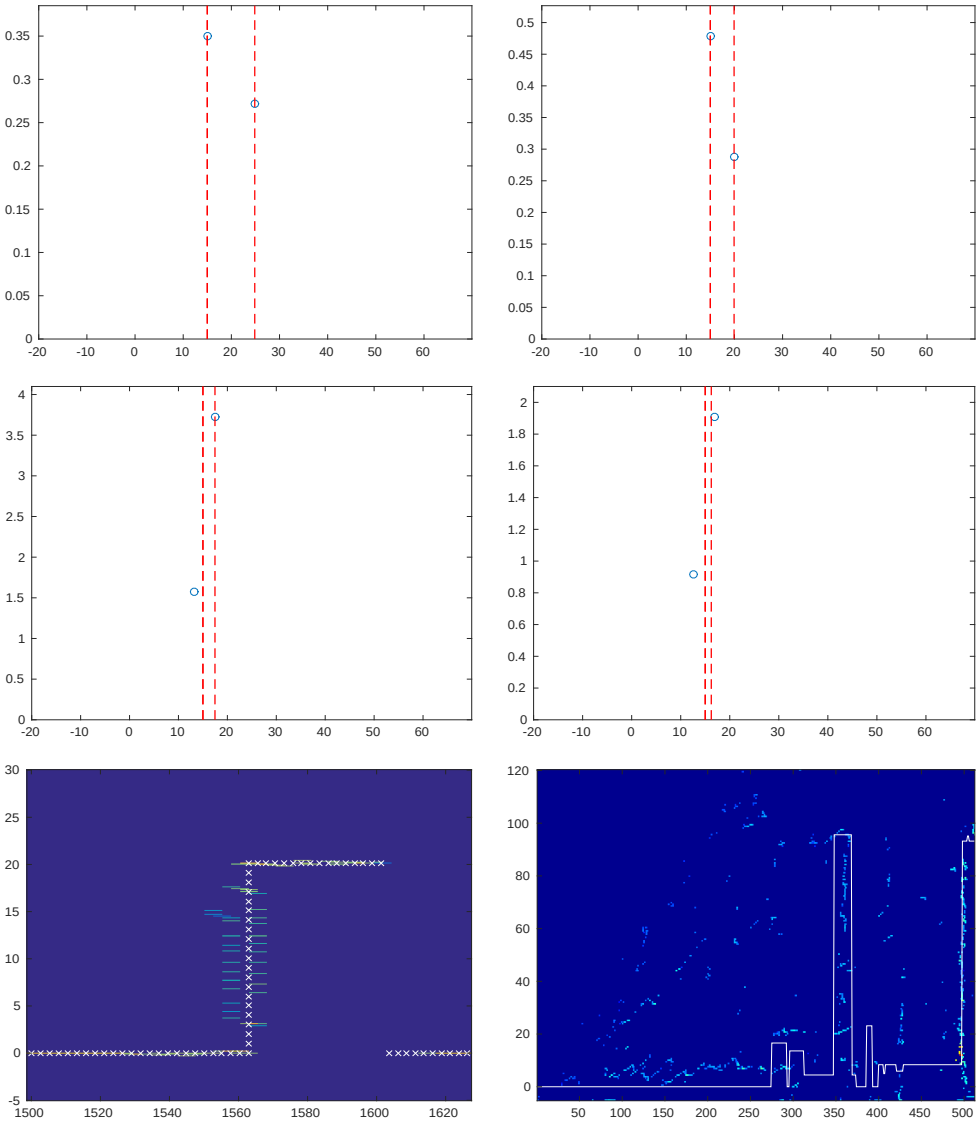


Figure 4.6 – Unmixing power of M-RELAX . The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the blue dot the estimated scatterers position and intensity. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice. The ground truth is indicated in white.

4.2.6 SPICE

The recent SPICE method (Stoica *et al.*, 2011) is a fully non-parametric sparse algorithm based on the minimization of the covariance fitting criteria:

$$f = ||\underline{\boldsymbol{R}}^{-1/2}\big(\underline{\hat{\boldsymbol{R}}} - \underline{\boldsymbol{R}}\big)\underline{\hat{\boldsymbol{R}}}^{-1/2}||_F^2 \tag{4.28}$$

The matrix $\underline{\boldsymbol{R}}$ is structured as stated in the conditional model (3.9) except that sources are fully incoherent ($\underline{\boldsymbol{S}}$ is diagonal) and the noise may be colored. The covariance matrix has then the form:

$$\underline{\boldsymbol{R}} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H\} = \sum_{d=1}^{D} \tau_d\, \underline{\boldsymbol{a}}(h_d)\underline{\boldsymbol{a}}^H(h_d) + \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix} \tag{4.29}$$

The previous equation can be rewrite into a product of matrices:

$$\underline{\boldsymbol{R}} = \underline{\boldsymbol{E}}\boldsymbol{D}\underline{\boldsymbol{E}} \tag{4.30}$$

$$\text{with } \underline{\boldsymbol{E}} = \Big[\underline{\boldsymbol{a}}(h_1), \cdots, \underline{\boldsymbol{a}}(h_D), \boldsymbol{I}_N\Big] = \Big[\underline{\boldsymbol{e}}_1, \cdots, \underline{\boldsymbol{e}}_D, \underline{\boldsymbol{e}}_{D+1}, \cdots, \underline{\boldsymbol{e}}_{D+N}\Big]$$

$$\text{and } \underline{\boldsymbol{D}} = \begin{bmatrix} \tau_1 & & & & & \\ & \ddots & & & & \\ & & \tau_D & & & \\ & & & \sigma_1^2 & & \\ & & & & \ddots & \\ & & & & & \sigma_N^2 \end{bmatrix} = \begin{bmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_D & & & \\ & & & d_{D+1} & & \\ & & & & \ddots & \\ & & & & & d_{N+D} \end{bmatrix}$$

Under this covariance matrix structure assumption, a consistent estimate of the solution of the minimization of f is given by the following optimization problem:

$$\min_{\tau_d \geqslant 0} \text{tr}\big(\underline{\hat{\boldsymbol{R}}}^{\frac{1}{2}}\underline{\boldsymbol{R}}^{-1}\underline{\hat{\boldsymbol{R}}}^{\frac{1}{2}}\big) \qquad \text{s.t.} \qquad \sum_{d=1}^{D+N} w_k d_k \tag{4.31}$$

$$\text{with } w_k = \frac{\underline{\boldsymbol{e}}_k^H \underline{\hat{\boldsymbol{R}}}^{-1} \underline{\boldsymbol{e}}_k}{N} \tag{4.32}$$

The minimization algorithm is presented in detail in (Stoica *et al.*, 2011). This algorithm achieves a good performance when the covariance matrix is correctly estimated and when the uncorrelated conditional model is respected. This method presents no parameter tuning and is relatively easy to optimize which makes it very promising. The signal model used however is even more restrictive than the conditional model.

Fig 4.7 illustrates M-RELAX performances for scatterer localization and on real data.

4.3 Compressive Sensing

Except for conventional beamforming, the estimators detailed in the previous section may all achieve super resolution under some hypotheses. However, they all need an estimation of the covariance matrix $\underline{\boldsymbol{R}}$ and/or the number of scatterers D. Finding D is still a hard task as the true signal distribution may be hard to model. Over uniform areas, $\underline{\boldsymbol{R}}$ can be efficiently estimated locally and the presented spectral analysis techniques may be used to obtain the reflectivity distribution along the elevation. When the scene is heterogeneous, more complex approaches are needed to estimate $\underline{\boldsymbol{R}}$ (see Chapter 7). When considering large stacks of SAR images and when the scene is very heterogeneous, the estimation of $\underline{\boldsymbol{R}}$ becomes very challenging. The CS approach introduced more recently (Zhu and Bamler, 2010a)(Budillon *et al.*, 2011) than most of the previous estimators is an efficient way to overcome this difficulty as it uses directly the back-projection of $\underline{\boldsymbol{v}}$ as stated in (4.1).

4.3.1 Projection and prior

The use of CS for SAR tomography is relatively recent. The estimation of the reflectivity profile $\underline{\boldsymbol{u}}$ along the height direction, for a given SAR resolution cell, is obtained by solving the following optimization problem:

$$\min_{\underline{\boldsymbol{u}}} ||\underline{\boldsymbol{u}}||_0 \quad \text{s.t.} \quad \underline{\boldsymbol{v}} = \underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})\underline{\boldsymbol{u}} \tag{4.33}$$

Whereas many spectral estimators are designed to retrieve a discrete set of signal parameters, the CS approach tries to retrieve the sparsest reflectivity profile depending on the elevation sampling $\widetilde{\boldsymbol{h}} \in \mathbb{R}^{N_h}$. The CS theory insures the existence of an exact solution for the problem (4.33) if the matrix $\underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})$ satisfies some conditions. Chronologically, the first one is the Restrictive Isometry Property (RIP) condition:

Definition 4.1 *A matrix $\underline{\boldsymbol{M}}$ satisfies the RIP property of order k is there exists a $\delta_k \in [0,1]$ such as*

$$(1-\delta_k)||\underline{\boldsymbol{s}}||_2^2 \leq ||\underline{\boldsymbol{M}}\underline{\boldsymbol{s}}||_2^2 \leq (1+\delta_k)||\underline{\boldsymbol{s}}||_2^2$$

where $\underline{\boldsymbol{s}}$ is any vector of sparsity at most k.

This property can be understood as any subset of at most k columns of $\underline{\boldsymbol{M}}$ must be as close to orthogonality as possible. A matrix satisfying the RIP of order $2k$ can then be seen as preserving approximately the distance between k-sparse vectors. From the RIP order of a matrix, one can compute a bound on the signal sparsity to ensure an exact reconstruction. In practice, it is however very hard to verify that a given matrix

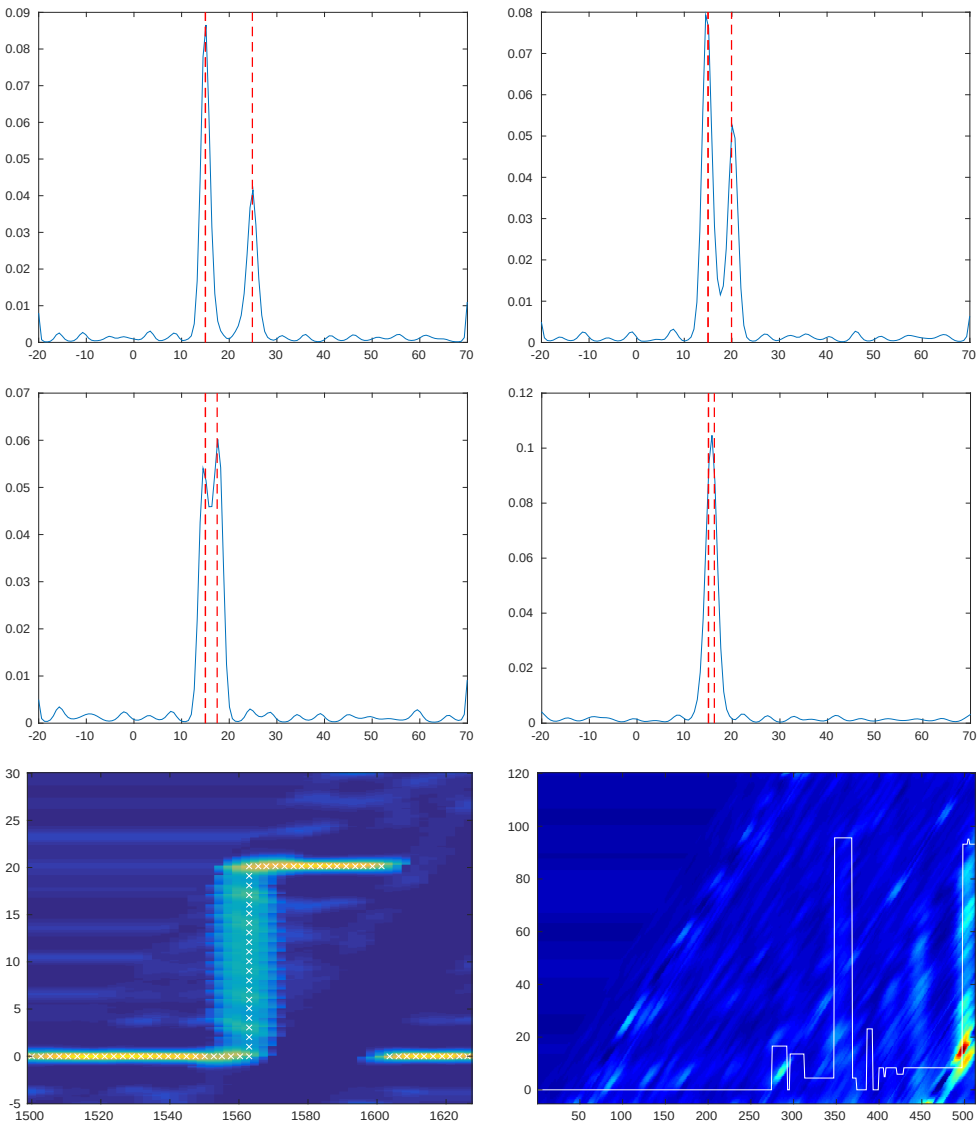


Figure 4.7 – Unmixing power of SPICE according to formula 4.27. The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines and the curve is the estimated intensity. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice.

satisfy the RIP or to calculate the corresponding constant δ_k as it requires a search over all $\binom{n}{k}$ submatrices. The coherence is a more intuitive and more easily computable matrix descriptor defined by:

$$\text{Let } \underline{\boldsymbol{M}} = \left[\underline{\boldsymbol{c}}_1, \cdots, \underline{\boldsymbol{c}}_m\right] \in \mathbb{C}^{n\times m}, \tag{4.34}$$
$$\mu(\underline{\boldsymbol{M}}) = \max_{1\leq i<j\leq m} \frac{|\underline{\boldsymbol{c}}_i^H \underline{\boldsymbol{c}}_j|}{||\underline{\boldsymbol{c}}_i||_2 ||\underline{\boldsymbol{c}}_j||_2}$$

The coherence basically indicates how much the columns of $\underline{\boldsymbol{M}}$ are correlated. The more they are the harder it becomes to retrieve the exact sparse signal. Again, the coherence is an indication on how orthogonal are the columns of a matrix. In the presence of additive noise, bound errors in the reconstruction with k-sparse vector can be computed from the coherence (Ben-Haim *et al.*, 2010). Other conditions that may be more adapted to sensing matrix consisting of an oversampled DFT and thus to SAR tomography may be found.

Going back to the SAR tomography, when a high oversampling is applied along the elevation axis, the columns of $\underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})$ are almost fully correlated with their neighboring ones. This means that under too much oversampling, it is almost impossible to know exactly from which direction a signal is received which corresponds to a finite maximum resolution. Avoiding oversampling is also not an option as the distribution of the signal may be continuous in space.

The solution to the combinatorial problem (4.33) can be approximated using the classical convex relaxation of the ℓ_0 pseudo-norm into an ℓ_1 norm:

$$\min_{\underline{\boldsymbol{u}}} ||\underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}||_2 + \mu_1 ||\underline{\boldsymbol{u}}||_1 \tag{4.35}$$

Different algorithms have been proposed to solve this problem such as Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996), basis pursuit (Chen *et al.*, 1998) or greedy algorithms as matching pursuit (Mallat and Zhang, 1993).

4.3.2 Drawbacks

The matrix $\underline{\boldsymbol{A}}(\widetilde{\boldsymbol{h}})$ is over-complete and does not guarantee to satisfy either the RIP or low coherence. The sparse reconstruction obtained through the resolution of (4.35) has nevertheless led to successful reconstructions of sparse urban scenes. However, artifacts can generally be found in those results. For instance, small spurious impulses far from the true localization of the objects or spreading of the scatterers to adjacent lines due to the oversampling. The parameter μ_1 is also generally hard to tune globally on the image because of the high dynamic of SAR images: a large value of μ_1 leads to the suppression of low intensity structures whereas a low value of it cannot allow outliers suppression.

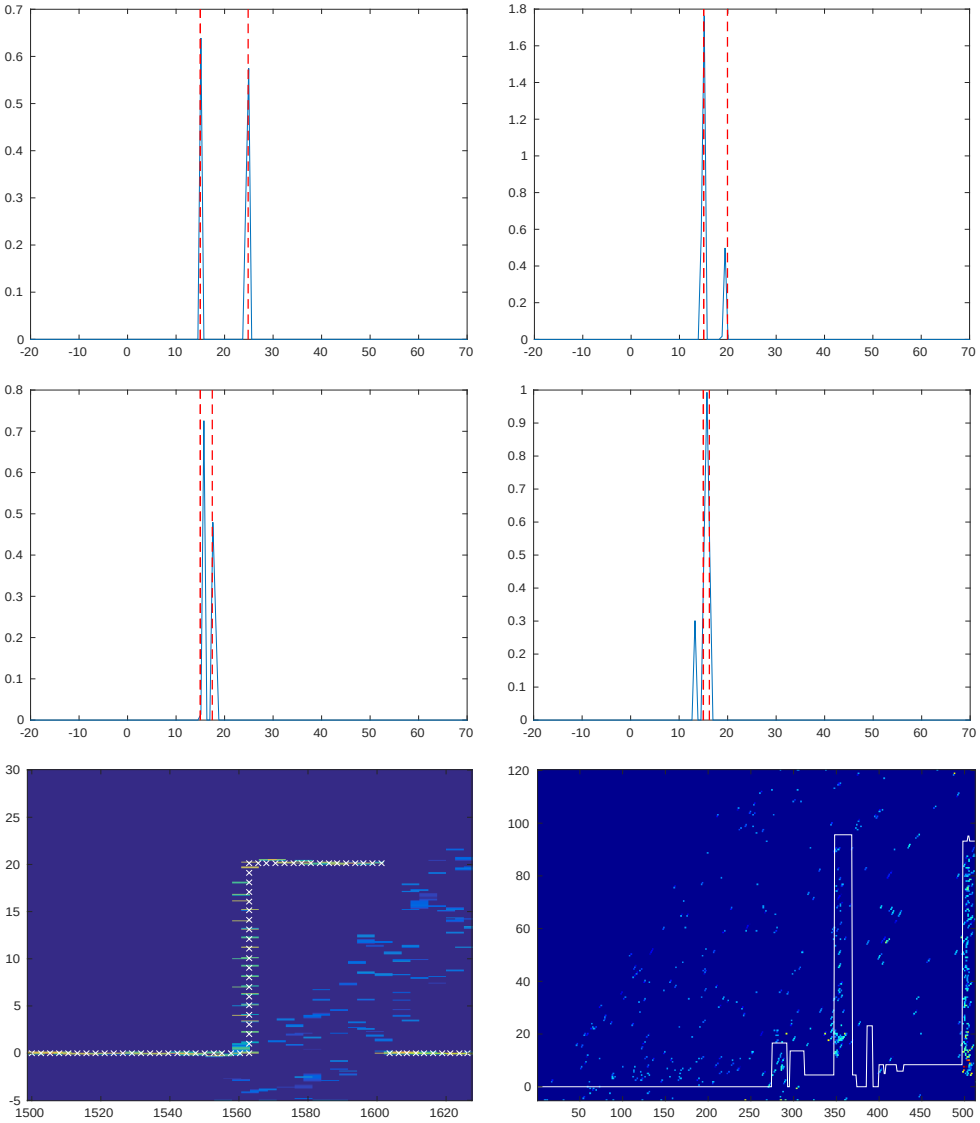


Figure 4.8 – Unmixing power of CS . The first 4 images present the estimation of two scatterers separated by 2, 1, 0.5 and 0.25 times the ambiguity height. The position of the scatterers is indicated by the dashed red lines. The last row illustrates the estimated signal for a simulation where the scatterer are well separated and a real urban slice. The ground truth is indicated in white. For the simulation the SNR corresponding to the additive white noise is 0.3 dB.

Fig 4.8 illustrates CS performances for scatterer localization on simulations and on real data.

4.3.3 Scatterers selection

To refine the results, the volume reconstructed with CS is generally post-processed by estimating the number of scatterers in order to select only the most significant points. The result is then a set of discrete points defined by their 3-D localization and complex reflectivity. As the distribution of the reflectivity in dense urban configurations is hard to model, this approach often fails to select points with low reflectivity.

The Scale-down by ℓ_1 norm Minimization, Model selection, and Estimation Reconstruction (SL1MMER) algorithm (Zhu and Bamler, 2012a) estimates the number of non-zero points $\hat{D}$ in the cell using a Model Order Selection (MOS) technique. The MOS approaches essentially consist in finding the number of scatterers that minimize the penalized log-likelihood of the data:

$$\hat{D} = \underset{D}{\operatorname{argmin}} - \log \mathrm{p}(\underline{\boldsymbol{v}}|\hat{\boldsymbol{h}}(D)) + C(D) \tag{4.36}$$

Under conditional model hypothesis, the last equation becomes:

$$\hat{D} = \underset{D}{\operatorname{argmin}}\Big\{\frac{||\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\hat{\boldsymbol{h}}(D))||_2^2}{2\sigma_n^2} + C(D)\Big\} \tag{4.37}$$

where $\hat{\boldsymbol{h}}(D)$ is composed of the elevations of the D most powerful pixels in the cell. Different penalties can be used for $C(D)$ such as the Bayesian Information Criterion (BIC) or Akaike criterion (Burnham and Anderson, 2004)(Stoica and Selen, 2004).

Close to CS , the recent FAST-SUP-GLRT detector (Budillon *et al.*, 2017a) avoids post processing selection by applying a sub-optimal statistical test taking into account the distribution of the data based on an approximated ℓ_0 norm minimization. Even if it does not take into account the geometry of the scene, the statistic of the data can be more accurately represented than with conventional CS.

As the distribution of the reflectivity in dense urban configurations is hard to model, this approach may fail to select points with low reflectivity.

4.4 Summary

Non-parametric estimators like conventional beamforming or Capon beamforming are easy to implement, fast to compute and adapted to continuous reflectivity distributions. They can be used to give a global overview of an urban scene. Parametric spectral estimators based on a sparse description of the data like MUSIC or WSF perform well on urban areas that are mostly composed of point-like scatterers. Maximum-Likelihood is computationally costly but may lead to the best reflectivity estimate under

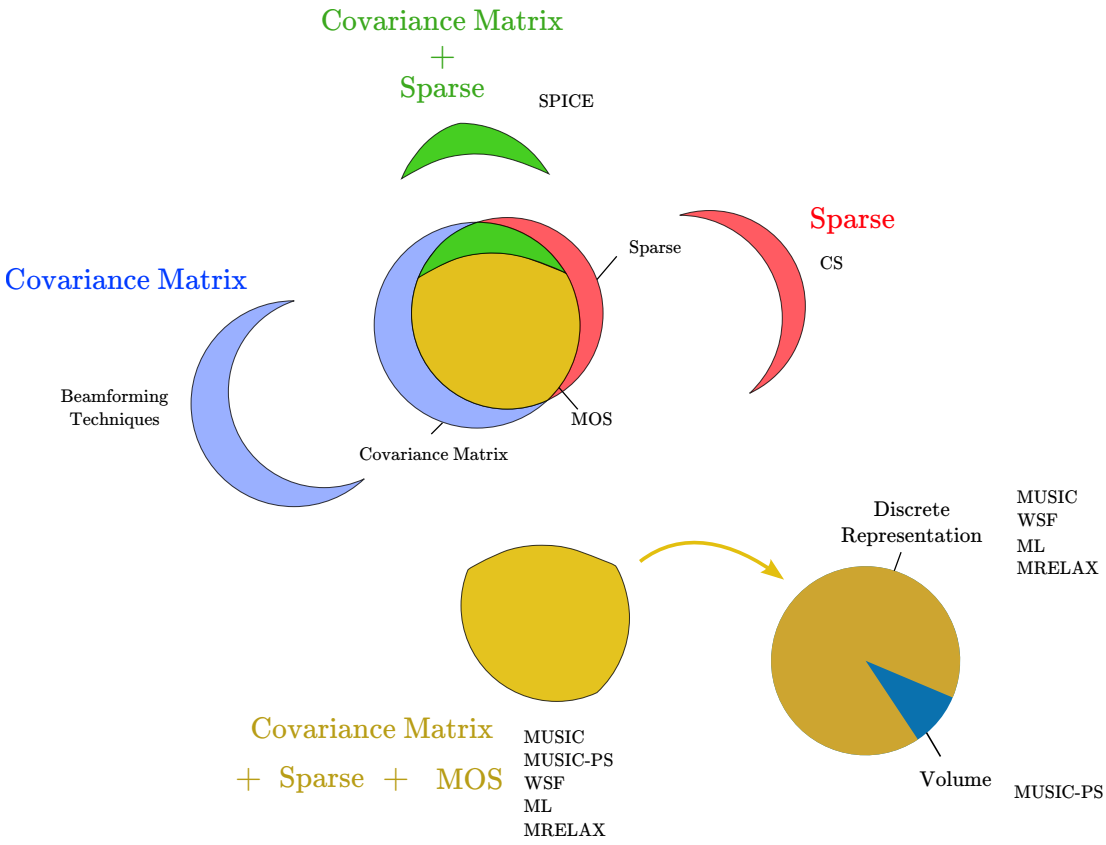


Figure 4.9 – Diagram representing the estimators presented in this part around three axes: sparsity, MOS and covariance based.

the assumption that the signal distribution is correctly modeled. M-RELAX can be used to maximize sequentially the likelihood but does not insure global convergence. SPICE is a recent non-parametric algorithm optimizing the reflectivity of the scatterers to fit the estimated covariance matrix to its model given by (3.9). Shared priors, input parameters and output representation between estimators are illustrated in the diagram Fig. 4.9.

Part II

Contributions : Introducing Spatial Regularization in SAR tomography

Chapter 5

On the road toward structural tomography

Dense urban areas are filled with layover areas where multiple strong intensity backscatterers resulting from walls and corners are projected into the same radar resolution cells. In addition to these very bright points other echoes coming from flat smooth objects such as ground or roof element are also present. The reconstructions obtained from this mixed signals generally only present high intensity scatterers. Walls and corner shape elements are then more or less well represented depending on the used estimator whereas distributed scatterers are mostly invisible.

It is almost certain that reflectivity fluctuations should arise in big temporal pile. These changes can have multiple origins such as interferometric decorrelation, weather, movement of the scatterers or phase errors. The obtained reconstructions can then be distorted due to phase mis-modeling. Finally, decorrelations and sidelobes are likely to cause outliers with amplitudes of the same order of magnitude as the retrieved scatterers. Figure 5.1 shows an example of reconstructed profiles illustrating the decorrelation, distortions and outliers issues. In this example, a building facing the sensor (the Mirabeau Tower) sends back a strong signal whereas most of the other scatterers on the profile are relatively low intensity. The reconstruction obtained with conventional beamforming is then highly impacted with high side lobes propagating in the radar cells were the building is projected. When looking at the conventional beamforming spectrum as well as the MUSIC-PS , it appears the signal estimated is determined by the brightest scatterers in it. This side lobe phenomena combined with the main lobe size may also produce outliers even with sparse approaches such as MUSIC or CS .

In the first part, state-of-the-art estimators for SAR tomography have been described. As stated in the preamble of chapter 3, with these methods, the estimation of the scatterers height and reflectivity is done pixel wise. 3-D SAR tomography is then seen as a concatenation of local estimations. Dense urban areas are however very structured as most of scatterers are distributed along roughly vertical or horizontal plans. The inherent difficulty induced by dense city landscape can then be reduced by taking into account the geometrical behavior of the scene. Urban 3-D reconstruction should

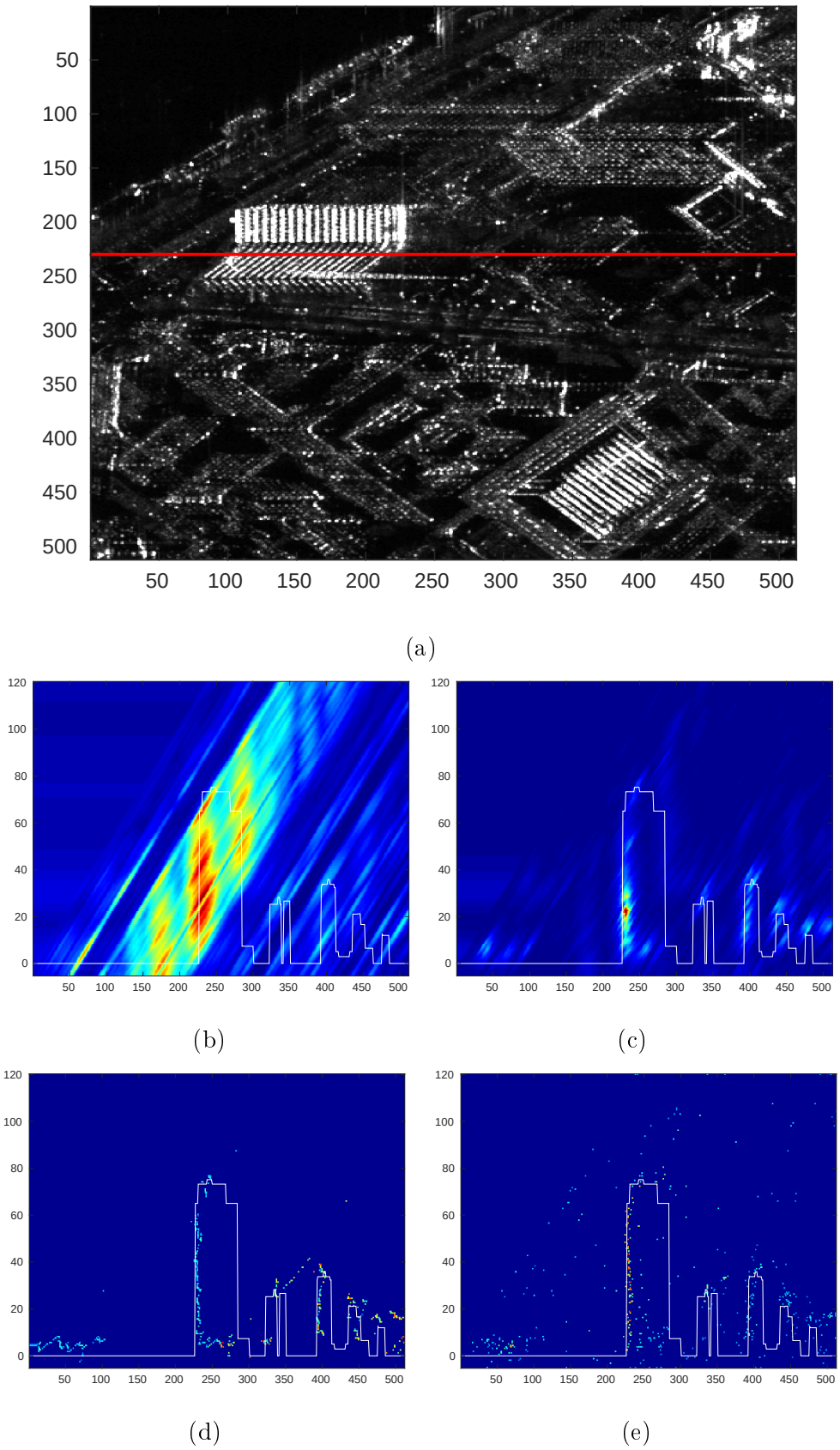


Figure 5.1 – Tomographic slice obtained from the red profile shown in the SAR image in (a). The reconstructions are obtained using conventional beamforming (b), MUSIC-PS (c), MUSIC (d), CS (e). For each slice the ground truth is printed in white.

also be sparse as most of the objects are punctual in ar least one direction: a wall or a roof are seen by the sensor as a surface located inside a cube. To use these priors at least two strategies may be considered: exploit the redundancies to have the best representation of the data or inverse the data and regularize the estimation.

The first concept consists in cleverly use similar data to estimate their parameters. This idea is at the origin of the patch based restoration methods where the neighborhood of the pixels is used to characterize them. Similar pixels, *i.e.* described by akin patches, are regrouped to estimate their parameters. For SAR tomography the desired parameter is generally the covariance matrix as the spectral estimators are function of it. The prior are then used in the date space to define which and how patches are alike.

The second concept is closer to CS as no local or non-local average is used and the SLC vectors are directly inverted. The priors are then used in the estimation space to favor the reconstruction satisfying the desired behavior. With conventional CS , sparsity is the only structural prior used whereas in the following work, geometrical constraints are also proposed.

After presenting the data and the evaluation protocol, the first strategy is addressed through chapter 7. The second approach is detailed in chapter 8 and chapter 9 for two variation of geometrical priors.

Chapter 6

Evaluation of tomographic reconstructions in urban environment

6.1 Geometric model for urban areas

Urban areas present very structured man made objects. The observed buildings can roughly be seen as the composition of plane objects either along the horizontal plane or along the vertical direction. Of course, a detailed 3-D representation would present many variations such as rooftop inclinations, balconies, various window shapes, crenelations, ornament, etc. Nevertheless these features are not expected to be retrieved with the available resolution of modern space-borne sensors.

Artificial structures present many dihedral or trihedral back scatterers (conditional model). In practice, most of the observed signal comes from building walls which are more likely to present corner shaped elements. Other smooth surface objects such as ground parts or rooftops back-scatter the wave in every direction and produce much weaker signals more likely to be subject to decorrelation mechanisms (unconditional model).

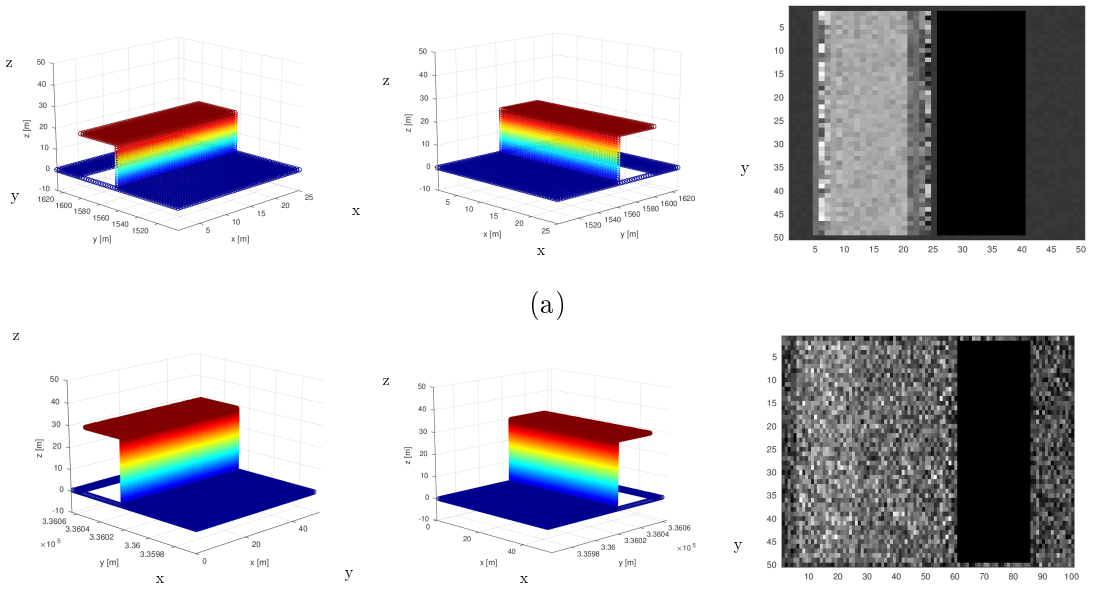
These last considerations are summed up in three hypotheses:

1. The structures illuminated by the sensor can be seen as plane smooth objects

2. The natural elongation directions of the urban structures are the horizontal plan and the vertical direction

3. The back-scattering signal is sparse in the 3-D space

These hypotheses will guide the construction of the algorithm we developed to retrieve the SAR tomographic signal.



(a)

(b)

Figure 6.1 – Two simulations of a simple 3-D scene: (a) considering a regular sampling of trajectories along the elevation axis (`Reg` case); (b) using the trajectories of TerraSAR-X satellites (`TSX` case). For each scene, two 3-D views are presented (left and middle) as well as the temporal mean intensity image (right).

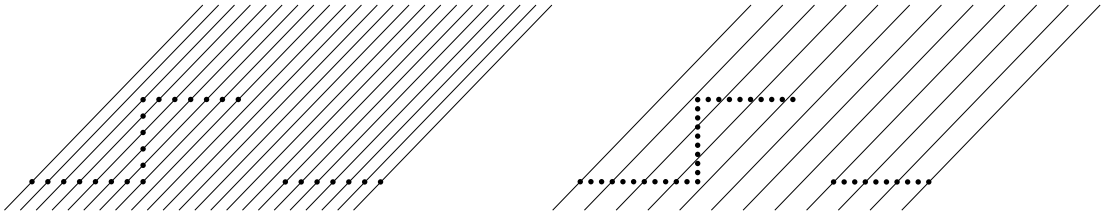


Figure 6.2 – Illustration of the diffent elevation samplings in the `Reg` case (left) and `TSX` case (right).

6.2 Presentation of the datasets

6.2.1 Simulations

Two stacks of images are simulated in two different configurations of trajectories: a regular sampling along the elevation axis h (referred to in the following as `Reg`), and the actual trajectories of TerraSAR-X given in Fig. 6.5 (referred to in the following as `TSX`). As in (Zhu and Bamler, 2010a), the decorrelation effect is introduced by adding a Gaussian white noise and the random scatterer phase is chosen uniform between $-\pi$ and π. The reflectivity distribution is then constituted of discrete point-like scatterers following the conditional signal model (3.7). The reflectivity corresponding to a scatterer located at the position (x, y, z) has the following expression:

$$\underline{u}(x,y,z) = \sqrt{\tau}\exp(\underline{j}\varphi) \quad \text{with } \varphi \sim \mathcal{U}(-\pi,\pi) \tag{6.1}$$

Fig. 6.1 shows the theoretical distribution of the scatterers. Both scenes are composed of a ground at a constant altitude, a wall and a roof, resulting in a large layover area in the SAR images. The simulated building is higher than the estimated elevation resolution (as given by Fourier inversion) in the `TSX` case. The reflectivity of the scatterers is set constant on all the scene. The SNR for the additive noise level is 1.4 dB and is the same for both experiments. In the `Reg` case, the scatterers are well separated in the azimuth and range directions. In the `TSX` experiment, the density of scatterers is larger in each cell, resulting in clusters of neighboring scatterers being projected in the same radar cell, as illustrated in the Fig 6.2. The distance between the scatterers is set as a third of the cell resolution. All the `TSX` images are thus corrupted with a speckle effect induced by the coherent addition of the back-scatterered signals. Due to the relatively strong additive noise and the geometrical decorrelation the average coherence is 0.63 for the `Reg` images and 0.68 for the `TSX` ones. Finally as the resolution is not the same in the two experiments the `TSX` structure is taller (20m in the `Reg` case against 30m in the `TSX` one) and the resulting images have a larger size in range.

6.2.2 Real data

The dataset is composed of 40 TerraSAR-X images of the south west of Paris. The corresponding temporal mean intensity image and the optical view of the scene are presented in Fig. 6.3 and 6.4. The images were acquired using the spotlight mode of the sensor in which the varying squint angle allows to improve the size of the synthetic azimuth antenna. The wavelength λ is of 3.11 cm. The images have a size 2048×2048 pixels with a resolution of 45 cm in range and 87 cm in azimuth. The spatial and temporal baselines are presented in Fig. 6.5. The total spatial baseline Δb span is more than 775 m and the total temporal baseline more than 5 years with a large gap of almost two years. The theoretical resolutions in h and z are given by :

$$\delta_h = \frac{\lambda R}{2\Delta b} = 1.75m \tag{6.2}$$

$$\delta_z = \delta_h \sin(\theta) = 6.99m \tag{6.3}$$

with an incidence angle $\theta = 0.6$ rad, a wavelength $\lambda = 0.0311$ m and a distance $R = 6.1510^5 m$. The characteristics of the sensor and of the scene are summarized in the table 6.1.

The scene is very heterogeneous with different back-scattering mechanisms. Most of the scene is composed of buildings producing very bright point-like echoes. The majority of the constructions are composed of either 20 m to 40 m height buildings or

Sensor	TerraSAR-X
Dates	3-4 years
Site	Paris
Composition	Mostly Buildings and water
Number of tracks	40
Wavelength	3.11 cm
Slant-range resolution	0.45 m
azimuth resolution	0.87 m

Table 6.1 – Global characteristics of the sensor and the observed scene

80 m to 110 m skyscrapers. Other very typical Parisian monuments can also be observed such as the *Eiffel Tower* in the upper right corner or the *Maison de la Radio* in the left side of the image. The Seine river with many bridges crossing it flows from the top right to the bottom left corner. As the water is very smooth, almost all the signal is scattered in the specular direction which creates these typical very low intensity areas in the image. Many vegetation areas are visible in the optical image but they produce very low intensity and low coherency signal in the SAR images. Moreover most of the trees answers are projected in the same radar cell as neighboring buildings (layover phenomena) and are thus hidden by brighter scatterers. Vegetation is still visible in some places such as on the island (*l'ile aux Cygnes*) where no buildings are present.

6.3 Accuracy and Completeness

In (D'Hondt *et al.*, 2018), the authors present two metrics to compare the different SAR tomographic results: the accuracy and the completeness. These two errors give a complementary evaluation of an estimated a point cloud.

Accuracy For a given discrete reconstruction $\hat{\mathcal{P}}$, the accuracy represents the mean distance from each point in $\hat{\mathcal{P}}$ to the ground truth $\mathcal{P}$.

$$A(\hat{\mathcal{P}}, \mathcal{P}) = \frac{1}{N_{\hat{p}}} \sum_{j=1}^{N_{\hat{p}}} \min_{k} \text{dist}\,(\hat{p}_j - p_k) \tag{6.4}$$

where $\hat{p}_j \in \hat{\mathcal{P}}$ is the j^{th} point of the estimated point cloud $\hat{\mathcal{P}}$ and $p_k \in \mathcal{P}$ is the k^{th} point of the ground truth. $N_{\hat{p}}$ is the number of points in the estimated reconstruction. The function *dist* is the distance used to compute the evaluation. Accuracy indicates whether reconstructed points are correctly located.

Completeness The completeness corresponds to the mean distance from each point of the ground truth to the points in $\hat{\mathcal{P}}$:

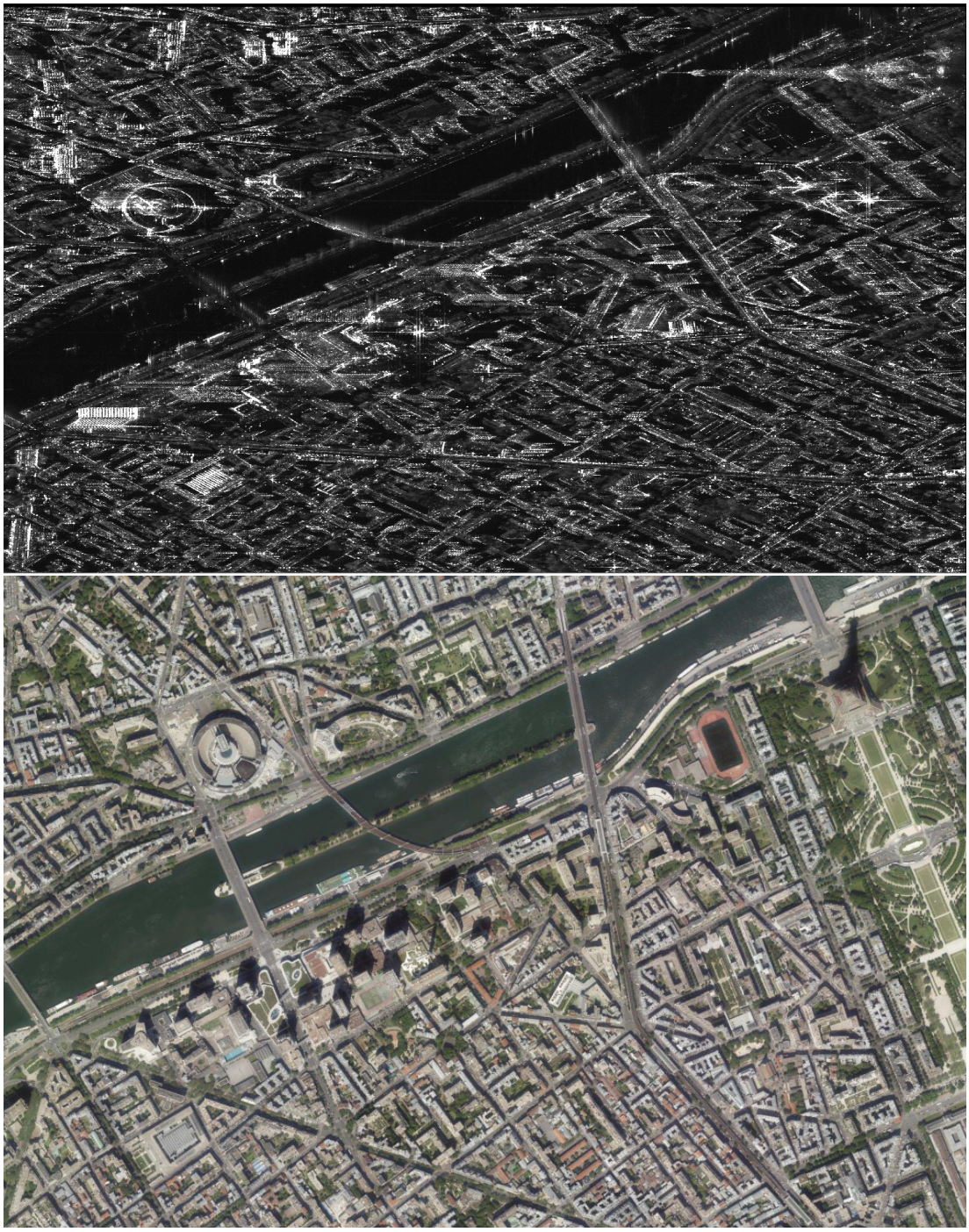


Figure 6.3 – Up: temporal intensity average of the stack of SAR images, bottom: corresponding optical view of the same zone.

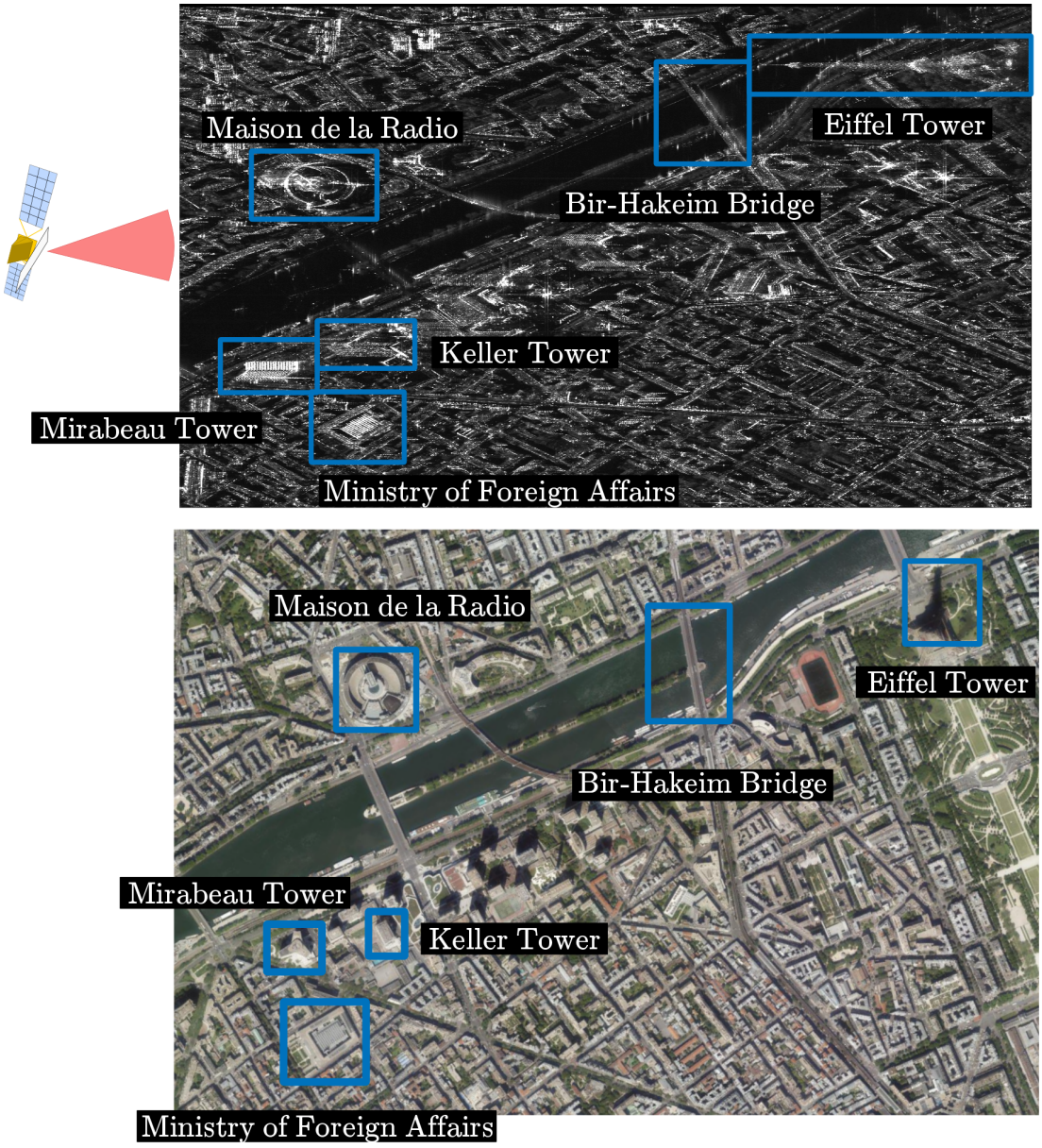


Figure 6.4 – Annotated SAR and optical images of the observed scene. The position of the sensors with respect to the scene is indicated for both images. The typical distortions induce by the SAR ranging acquisition system are well visible as the tall structures such as the Eiffel Tower or the Mirabeau Tower are projected along the range direction.

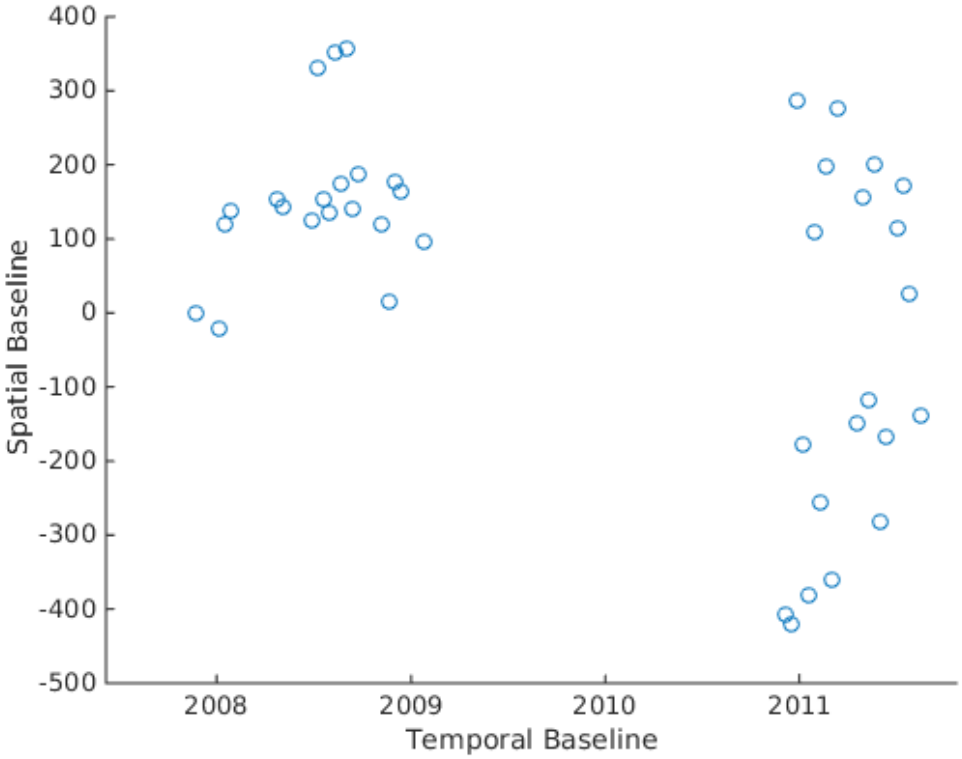


Figure 6.5 – Temporal and spatial baselines

$$C(\hat{\mathcal{P}}, \mathcal{P}) = \frac{1}{N_p} \sum_{k=1}^{N_p} \min_j \text{dist}\left(\hat{p}_j - p_k\right) , \tag{6.5}$$

with N_p the number of points in the ground truth. Completeness indicates whether the ground truth is well represented by the set of points in the reconstruction.

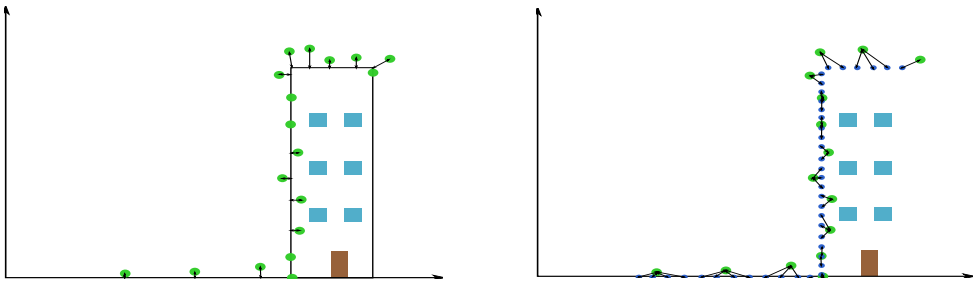


Figure 6.6 – Accuracy (a) et completeness (b) criteria. The accuracy computes the mean of the error for each estimated point. The completeness gives an indication on the proportion of holes in the rendering.

In the presented work we chose to use the Euclidean distance for the evaluation. The two metrics are illustrated in Fig. 6.6 where green dots represent estimated points $\hat{p}_j$ and blue squares the ground truth points p_k. Accuracy and completeness provide

complementary information: accuracy improves when reconstructed points are close to actual points from the ground truth, but does not indicate when points are missing (holes in the reconstruction). For instance, if we could retrieve a single point of the scene (the strongest permanent scatterer, for example) with a location very close to the ground truth, the accuracy would be excellent ($A(\hat{\mathcal{P}}, \mathcal{P})$ near zero) while the completeness $C(\hat{\mathcal{P}}, \mathcal{P})$ would be large, indicating that much of the scene is missing in the reconstruction. Conversely, if we loosely select points (many points for each resolution cell), we would obtain a dense volume, thus a good completeness (low completeness value $C(\hat{\mathcal{P}}, \mathcal{P})$), but erroneously selected points lying far from the true surfaces would lead to a poor accuracy score (large accuracy value $A(\hat{\mathcal{P}}, \mathcal{P})$).

6.4 Evaluation protocol

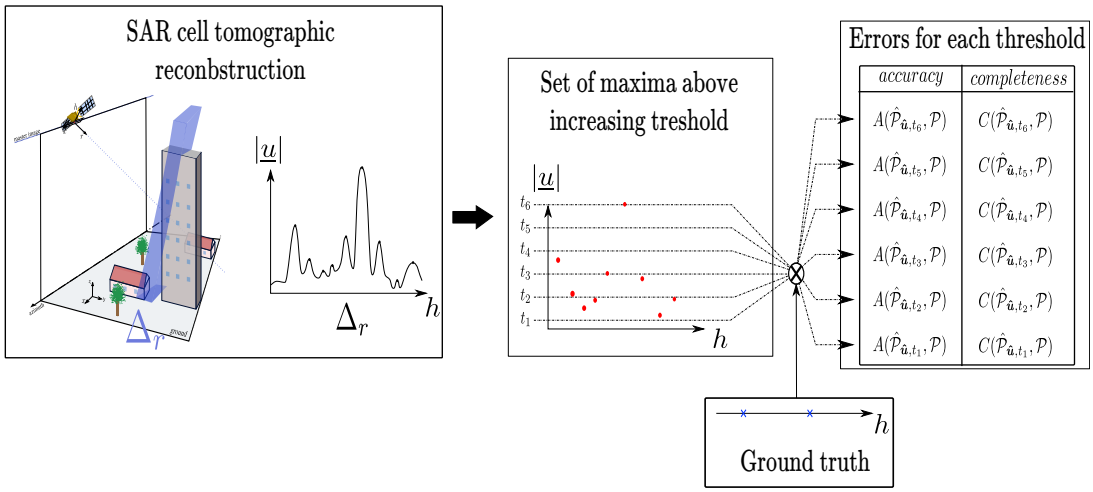


Figure 6.7 – **Evaluation protocol:** each RADAR cell is described by the set of its maxima. Different thresholds are applied to further select the points which are then compared to the ground truth. This operation is done for all the cells of the image.

Tomographic SAR volume reconstructions are generally not quantitatively evaluated. The works presenting error evaluations are mostly done on point clouds extracted from the tomograms. This implies that the evaluation depends both on the signal estimation technique as well as the method to estimate the number of scatterer. However the works presented in this document focus on the tomographic reconstruction and does not propose a point extraction step. Moreover, as stated in the section 4.3.3, the estimation of the number of scatterers is not a trivial task and is still an ongoing research topic. Using classical MOS methods, the number of reflectors may be underestimated as only the conditional model is used to model the signal. Here, we are more interested in how good a tomographic estimation potentially is, *i.e.* if the evaluation is done with the best number of scatterers estimation. The used strategy is then to describe a volume reconstruction a family of sets of 3-D points.

Beforehand, 3-D points must be extracted from the volume of voxels obtained by the

tomographic reconstruction algorithms. This extraction step is performed by finding the local maxima in the height direction for each radar resolution cell. This produces a first point-based representation of the data. Then, to reduce the sensitivity to noise and side-lobes, we apply a threshold on the obtained point cloud. Starting from a reconstructed volume $\underline{\hat{\boldsymbol{u}}}$, we obtain the collection of 3-D points $\hat{\mathcal{P}}_{\underline{\hat{\boldsymbol{u}}},t}$:

$$\hat{\mathcal{P}}_{\underline{\hat{\boldsymbol{u}}},t} = \Big\{ p_j = (x_j, y_j, z_j) \in \mathbb{R}^3, \text{ such that} \tag{6.6}$$

$$|\underline{\hat{\boldsymbol{u}}}_{x_j,y_j,z_j}| \geq |\underline{\hat{\boldsymbol{u}}}_{x_j,\nu_y,\nu_z}| \text{ and } |\underline{\hat{\boldsymbol{u}}}_{x_j,y_j,z_j}| > t \Big\} \tag{6.7}$$

where ν_y and ν_z are the coordinates of the direct neighboring voxels restricted to the same radar resolution cell and t is a threshold.

Evaluating tomographic estimators can be done for different regimes (accurate reconstruction of the strongest scatterers, dense reconstruction of most scatterers). To capture these different cases, rather than considering a single reconstruction obtained with a giving threshold, the accuracy score $A(\hat{\mathcal{P}}_{\underline{\hat{\boldsymbol{u}}},t}, \mathcal{P})$ is represented as a function of the completeness $C(\hat{\mathcal{P}}_{\underline{\hat{\boldsymbol{u}}},t}, \mathcal{P})$. Applications that focus on the reconstruction of permanent scatterers will favor algorithms that achieve the highest accuracy values (even if the completeness is poor). If the reconstruction of surfaces is the aim, completeness should be favored, even if this degrades the accuracy. The evaluation protocol is summarized in the diagram shown in Fig. 6.7.

Chapter 7

Covariance Matrix Estimation

The spectral estimators used in SAR tomography are almost all based on the analysis of the covariance matrix. When the observed area is very heterogeneous with a high variety of scatterers height, material and orientation the estimation of the local covariance matrix may be very challenging. Different strategies have been proposed to perform this estimation. Local estimation techniques exploit only the information close to the pixel of interest whereas non-local ones are designed to exploit self similarity in the image.

7.1 Local Approaches

The estimation of the covariance matrix in SAR tomography is often done by local averaging using an isotropic kernel such as a boxcar or a gausscar filter. The value of the covariance matrix is then given by:

$$\underline{\hat{\boldsymbol{R}}} = \sum_{p} w_p \underline{\boldsymbol{v}}_p \underline{\boldsymbol{v}}_p^H \tag{7.1}$$

where w_p are the coefficients of the isotropic filter and the subscript p corresponds to a 2-D shift $p \in [-H, H]^2$ indicating the support of the filter. The number of samples is then $L = (2 \times H + 1)^2$. Although these filters are fast and easy to compute, they lack of spatial adaptivity and induce resolution loss in the images. Moreover, as SAR images have a very high dynamic range, bright pixels may have amplitudes several orders of magnitude larger than the background. These basic local filters then produce blurry estimation of point-like bright pixels.

Other local methods like the Lee filter (Lee, 1981) are designed to adapt locally on the content inside the support of the function. The Lee filter is built to reduce the speckle phenomena while preserving sharp structures in the image. This filter is designed to minimize the mean square error for a linear speckle noise model. The Lee filter gives then as output for each channel n of the tomographic image *i.e.* each 2-D SAR image:

$$\hat{I}_n = \overline{I}_n + k_{\text{Lee}} \left(I_n - \overline{I}_n \right) \tag{7.2}$$

where $I_n = |\underline{\boldsymbol{v}}(n)|^2$ is the intensity of the noisy pixel, $\overline{I}_n$ the mean intensity within the filter window and $\hat{I}_n$ the estimated intensity. The coefficient k_{Lee} adjusts the local smoothing and is computed as:

$$k_{\text{Lee}} = 1 - \frac{\gamma_S^2}{\gamma_I^2} \tag{7.3}$$

with γ_S^2 standing for the theoretical coefficient of variation of the speckle: $\gamma_S^2 = 1/\sqrt{L}$ and γ_I^2 coefficient of variation of the current window. When no structures appear inside the window, γ_I^2 should tend toward γ_S^2 and k_{Lee} is then equal to 0. The intensity estimation is then given by the local averaging. When bright structures are present, the local variance is expected to increase resulting in a coefficient tending to 1. When k is near 1, the subimage is too heterogeneous and intensity is left unchanged.

To improve the denoising, the smoothing can be combined with oriented windows allowing to preserve straight edges more efficiently. The windows are composed of a rectangular mask and 8 edge-aligned oriented ones. Although continuous smooth edges can be well preserved with this method, sudden discontinuities and orientation changes produce artifacts. Moreover the limited number of orientations in the mask limits the denoising power for complex structures and textures.

7.2 Non-Local SAR algorithm

The NL-SAR algorithm (Deledalle *et al.*, 2015) became in the recent years one of the top denoising algorithm for SAR images. It can be used to restore multichannel data and has shown very good results for SLC images as well as interferometric and polarimetric data. Due to its performances, this approach may be seen as very promising for SAR tomography. In the following the algorithm is summed up and results for increasing number of channels are studied.

The estimation of the pixel value can be extended to non-connected neighborhoods. The pixels used to perform the estimation are then selected based on their similarity. This allows to use far apart information when the local neighborhood is not sufficient to perform a satisfactory denoising. This led to Lee's sigma filter (Lee, 1983) where the restoration of the image is done based on a similarity between the pixels.

Rather than using only a pixel wise similarity, using the similarity between the local neighborhoods (*i.e.* patches) of the pixels allows to use the structural information to select the pixels. This idea has been popular in various image processing algorithms that followed the seminal "NL-means" approach (Buades *et al.*, 2005). The NL-SAR algorithm is an extension of this method for the SAR images restoration.

The similarity used in NL-SAR is derived from the fully unconditional model where the N-dimensional complex pixels follow a circular complex Gaussian distribution:

$$\mathrm{p}(\underline{\boldsymbol{v}}|\underline{\boldsymbol{R}}) = \frac{1}{\pi^N |\underline{\boldsymbol{R}}|} \exp\left(-\underline{\boldsymbol{v}}^H \underline{\boldsymbol{R}}^{-1} \underline{\boldsymbol{v}}\right) \tag{7.4}$$

where $\underline{\boldsymbol{R}} \in \mathbb{C}^{N\times N} = E\{\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^H\}$ is the complex covariance matrix, and $|\underline{\boldsymbol{R}}|$ its determinant. The empirical covariance matrix corresponds to the averaging of the L elements inside a local window:

$$\hat{\underline{\boldsymbol{R}}} = \frac{1}{L} \sum_{l=1}^{L} \underline{\boldsymbol{v}}_l \underline{\boldsymbol{v}}_l^H \tag{7.5}$$

L is referred as the equivalent number of looks. When $N \leq L$, the empirical covariance matrix follows a complex Wishart distribution given by:

$$\mathrm{p}\left(\hat{\underline{\boldsymbol{R}}} \,|\, \underline{\boldsymbol{R}}\right) = \frac{L^{LN} |\hat{\underline{\boldsymbol{R}}}|^{L-D}}{\Gamma_N(L) |\underline{\boldsymbol{R}}|^L} \exp\left(-L\mathrm{tr}(\underline{\boldsymbol{R}}^{-1}\hat{\underline{\boldsymbol{R}}})\right) \tag{7.6}$$

When $L < N$, the complex empirical covariance matrix is singular. The matrix is then said to have a degenerate distribution.

Similarity Criterion Several methods have been proposed to express a similarity criterion between pixels corrupted by a speckle effect. The one used in the NL-SAR algorithm is based on a hypothesis test on the empirical covariance matrix distribution. For two empirical covariance matrices $\hat{\underline{\boldsymbol{R}}}_1$ and $\hat{\underline{\boldsymbol{R}}}_2$, the test is derived as follows:

$$\begin{cases} \mathcal{H}_0 : \underline{\boldsymbol{R}}_1 = \underline{\boldsymbol{R}}_2 \triangleq \underline{\boldsymbol{R}}_{12} & \text{(7.7a)} \\ \mathcal{H}_1 : \underline{\boldsymbol{R}}_1 \neq \underline{\boldsymbol{R}}_2 & \text{(7.7b)} \end{cases}$$

With independent Wishart distributed empirical covariance matrices $\hat{\underline{\boldsymbol{R}}}_1$ and $\hat{\underline{\boldsymbol{R}}}_2$, the Generalized Likelihood Ratio (GLR) corresponding to the hypothesis test in (7.7a)-(7.7b) is given by (Conradsen *et al.*, 2003):

$$\mathcal{L}_G(\hat{\underline{\boldsymbol{R}}}_1, \hat{\underline{\boldsymbol{R}}}_2) = \frac{|\hat{\underline{\boldsymbol{R}}}_1|^L |\hat{\underline{\boldsymbol{R}}}_2|^L}{|\frac{1}{2}\left(\hat{\underline{\boldsymbol{R}}}_1 + \hat{\underline{\boldsymbol{R}}}_2\right)|^{2L}} \tag{7.8}$$

The closer the empirical covariances matrix $\hat{\underline{\boldsymbol{R}}}_1$ and $\hat{\underline{\boldsymbol{R}}}_2$, the larger $\mathcal{L}_G$ is. For identical matrices, it is equal to 1 whereas it tends toward 0 for very different ones. Note that the GLR is only defined if the number of samples used in the empirical covariance matrices computation is larger than the number of channels N. The patches are then compared

by computing the similarity between all the pixels. For two patches centered at two positions i and i', the similarity Δ is then:

$$\Delta(i,i') = \sum_{p} -\log \mathcal{L}_G \left[\hat{\underline{\boldsymbol{R}}}(i'+p), \hat{\underline{\boldsymbol{R}}}(i+p)\right] \tag{7.9}$$

As in equation (7.1), $p \in [-H, H]$ indicates a local 2-D shift for a window of size $L = (2 \times H + 1)^2$. The quantity Δ is then a dissimilarity indication between patches. As all the patches are compared, Δ represents how much the structures surrounding the pixels of interests are not-alike.

Weights computation A common way to define the weights from the (dis)similarity measure is to use an exponential kernel $w(i,i') = \exp\left[-\Delta(i,i')/h\right]$, with $h > 0$ a filtering parameter. The NL-SAR algorithm proposes a framework that makes the computation of the weights independent of the parameters such as the search window and the patch size, the number of samples used for the estimation of the empirical covariance matrices or the number of channels. To do so the exponential kernel is composed with two other functions:

- $\mathbb{F}$ the cumulative distribution function of Δ under the hypothesis $\mathcal{H}_0$
- $\mathbb{G}^{-1}$ the reciprocal of the χ^2 distribution.

The weight for a pixel i' respectively to a pixel i is then defined as:

$$w(i,i') = \exp\left[-\frac{\mathbb{G}^{-1} \circ \mathbb{F}\left[\Delta(i,i')\right]}{h}\right] \tag{7.10}$$

The χ^2 function corresponds to the distribution of the dissimilarity Δ under a Gaussian additive noise where the exponential kernel has proven to be efficient for denoising (Buades *et al.*, 2005). The NL-SAR weights computation corresponds then to a mapping to the SAR image statistics to an additive Gaussian scenario. Finally to avoid having any pixel having a higher weight than the central one and favor pixels following the same distribution, the kernel is modified as follow:

$$w(i,i') = \begin{cases} \exp\left[-\frac{\mathbb{G}^{-1} \circ \mathbb{F}\left[\Delta(i,i') - c\right]}{h}\right] & \text{if } i \neq i' \\ 1 & \text{otherwise} \end{cases} \tag{7.11}$$

with $c = E\left\{\mathbb{G}^{-1} \circ \mathbb{F}\left[\Delta(i,i')\right] | \mathcal{H}_0\right\}$.

The non-local estimation of the covariance matrix is then given by:

$$\hat{\underline{\boldsymbol{R}}}^{\mathrm{NL}}(i) = \frac{\sum_{i'} w(i,i') \underline{\boldsymbol{v}}_{i'} \underline{\boldsymbol{v}}_{i'}^H}{\sum_{i'} w(i,i')} \tag{7.12}$$

Bias reduction To reduce the spreading of bright structures induced by the empirical covariance matrix pre-estimation step, a bias correction step is applied to the non-local estimate. This step is similar to the Lee filter strategy and the corrected Non-Local Bias Reduced estimation is then:

$$\hat{\underline{\boldsymbol{R}}}^{\mathrm{NLRB}}(i) = \hat{\underline{\boldsymbol{R}}}^{\mathrm{NL}}(i) + \alpha\left[\underline{\boldsymbol{v}}_i \underline{\boldsymbol{v}}_i^H - \hat{\underline{\boldsymbol{R}}}^{\mathrm{NL}}(i)\right] \tag{7.13}$$

The parameter α computation is detailed in (Deledalle *et al.*, 2015) and depends on the variance of the non-local estimation.

Unsupervised local adaptation The NL-SAR algorithm present many parameters: the size of the patches, the size of the search window, the number of samples used for the local estimation of the empirical covariance matrices or the scale parameter h. However, one of the strength of the method is its unsupervised adaptation to the local context. The quality of the estimation is given in terms of variance reduction and the set of parameters is set locally to have the best reduction. As the variance decreases monotonously with the number of samples (or looks) used for the estimation, the best set of parameters is the one maximizing the number of looks. After non-local estimation and bias-reduction, the equivalent number of looks is given by (Deledalle *et al.*, 2015):

$$\hat{L}^{NLRB}(i) = \frac{\hat{L}^{NL}(i)}{(1-\alpha)^2 + \left(\alpha^2 + \frac{2\alpha(1-\alpha)}{\sum_{i'} w(i,i')}\right)} \tag{7.14}$$

The set of parameters maximizing this quantity is then used for the estimation. Examples of denoising using the presented local and non-local algorithms are presented in Fig. 7.1 and 7.2.

7.3 Proposed pixel similarity for large tomographic stacks of urban areas

In urban areas where the amplitudes of bright pixels are significantly higher than their surrounding background and where the SAR images are very heterogeneous, finding similar patches may be a difficult task. When the dimension of the SAR stack increases patches may then appear more and more unique within the search window. From the covariance matrix structure detailed in the MUSIC algorithm, it is possible to get an intuition justifying the reduced efficiency of the Wishart GLR :

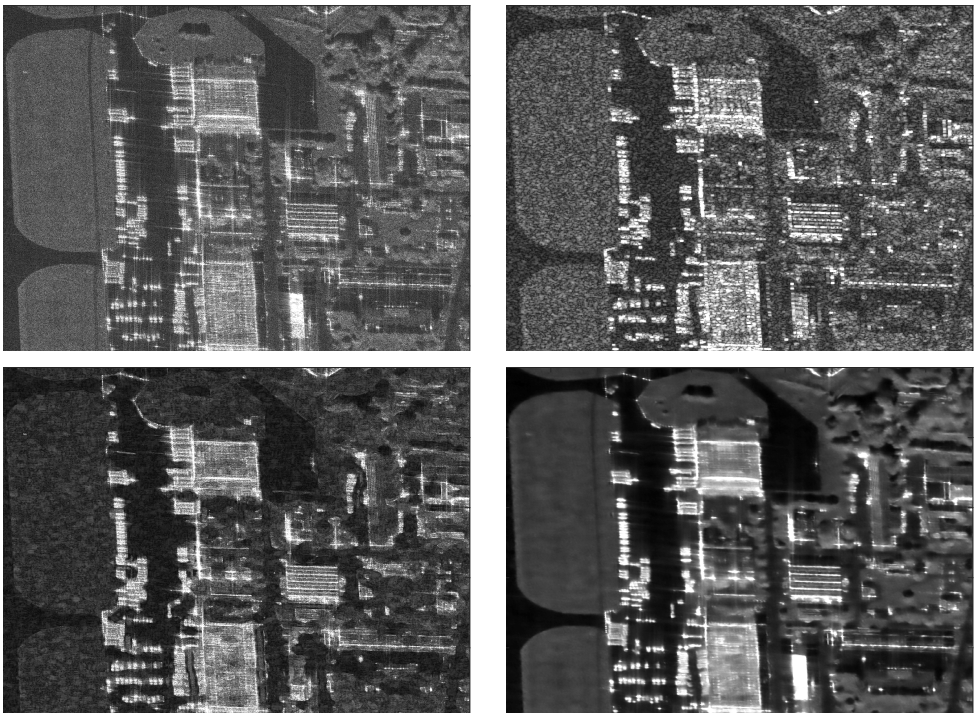


Figure 7.1 – Images of buidings near Nimes-Garon airport obtained by the Sethi airborn sensor - ONERA. From left to right : noisy SLC SAR image and its denoised version using a gaussian kernel, the Lee filter or NL-SAR algorithm.

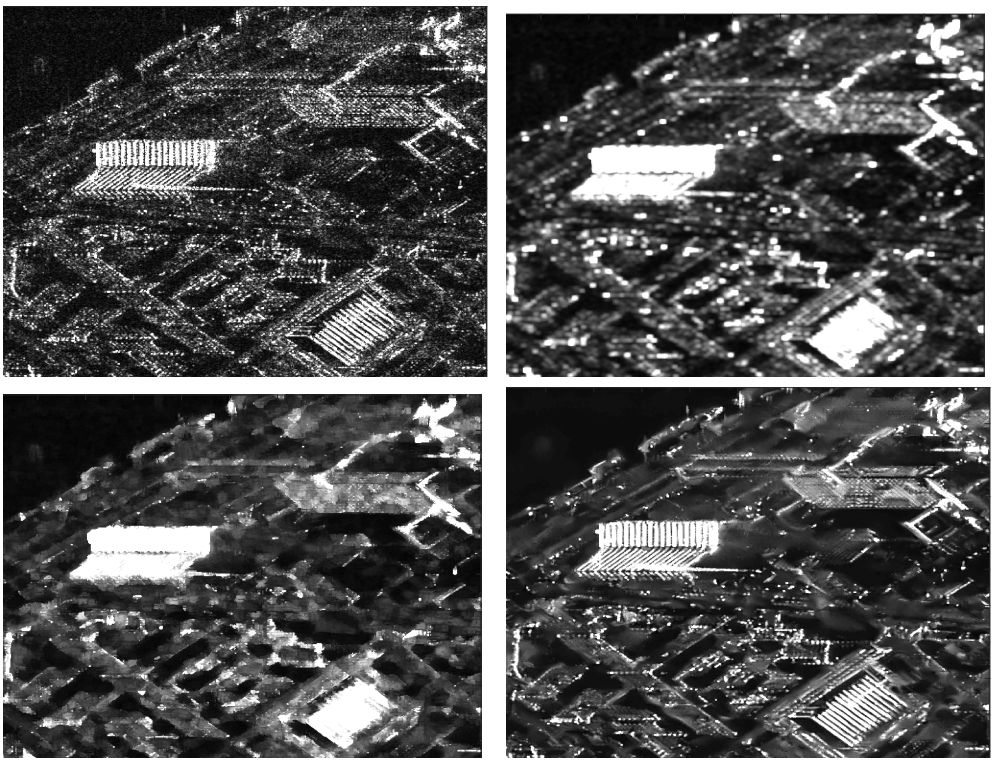


Figure 7.2 – Images of buildings in the south-west of Paris obtained by the TerraSAR-X spaceborn sensor - DLR. From left to right : noisy SLC SAR image and its denoised version using a gaussian kernel, the Lee filter or NL-SAR algorithm.

$$\underline{\boldsymbol{R}}_1 = \underline{\boldsymbol{V}}_1 \begin{bmatrix} \mu_1 & & & & \\ & \ddots & & & \\ & & \mu_{d_1} & & \\ & & & \sigma_n^2 & \\ & & & & \ddots \\ & & & & & \sigma_n^2 \end{bmatrix} \underline{\boldsymbol{V}}_1^H, \quad \underline{\boldsymbol{R}}_1 = \underline{\boldsymbol{V}}_2 \begin{bmatrix} \nu_2 & & & & \\ & \ddots & & & \\ & & \nu_{d_2} & & \\ & & & \sigma_n^2 & \\ & & & & \ddots \\ & & & & & \sigma_n^2 \end{bmatrix} \underline{\boldsymbol{V}}_2^H$$

$$\text{and } \underline{\boldsymbol{R}}_{12} = \underline{\boldsymbol{V}}_{12} \begin{bmatrix} \rho_2 & & & & \\ & \ddots & & & \\ & & \rho_{d_{12}} & & \\ & & & \sigma_n^2 & \\ & & & & \ddots \\ & & & & & \sigma_n^2 \end{bmatrix} \underline{\boldsymbol{V}}_{12}^H \tag{7.15}$$

where $\underline{\boldsymbol{V}}_1$, $\underline{\boldsymbol{V}}_2$ and $\underline{\boldsymbol{V}}_{12}$ are composed of the eigenvectors of the corresponding matrices, d_1, d_2 and d_{12} are the number of significant eigenvalues; $\{\mu_n\}_{1\leq n\leq d_1}$, $\{\nu_n\}_{1\leq n\leq d_1}$ and $\{\rho_n\}_{1\leq n\leq d_1}$, the corresponding eigenvalues. σ_n^2 is the additive noise power. Now, let's look at the simple example where $\hat{\underline{\boldsymbol{R}}}_1$ and $\hat{\underline{\boldsymbol{R}}}_2$ follow very similar distributions except for the last subspace. Here $d_1 = d_2 = d_{12} = d$ and

$$\left[\underline{\boldsymbol{V}}_1\right]_k^H \left[\underline{\boldsymbol{V}}_2\right]_k = \begin{cases} 1 \text{ if } k \neq d \\ 0 \text{ otherwise} \end{cases} \quad \text{and} \quad \mu_k = \nu_k = \rho_k \text{ if } k \neq d \tag{7.16}$$

The last assumption states that all the eigenvectors of $\underline{\boldsymbol{V}}_1$ and $\underline{\boldsymbol{V}}_2$ are the same except for the d^{th} ones that are orthogonal. Then the GLR under Wishart distribution for the empirical covariance matrices as defined in (7.8) is:

$$\mathcal{L}_G(\hat{\underline{\boldsymbol{R}}}_1, \hat{\underline{\boldsymbol{R}}}_2) = \frac{|\hat{\underline{\boldsymbol{R}}}_1|^L |\hat{\underline{\boldsymbol{R}}}_2|^L}{|\frac{1}{2}\left(\hat{\underline{\boldsymbol{R}}}_1 + \hat{\underline{\boldsymbol{R}}}_2\right)|^{2L}} = \prod_{i=1}^{d-1} \frac{\mu_i^L \nu_i^L}{\left(\frac{\mu_i+\nu_i}{2}\right)^{2L}} \left(\frac{2^4\sigma_n^2}{\mu_d\nu_d}\right)^{2L} \tag{7.17}$$

As the noise power is generally at least 2 or 3 dB lower than the scatterers power and $\frac{\mu_i\nu_i}{\left(\frac{\mu_i+\nu_i}{2}\right)^2} \leq 1$, $\mathcal{L}_G$ is expected to be smaller than 1. Depending on the ratio $\frac{2^4\sigma_n^2}{\mu_d\nu_d}$, $\mathcal{L}_G$ may be very low. Now if the d^{th} columns of $\underline{\boldsymbol{V}}_1$ and $\underline{\boldsymbol{V}}_2$ are different but not orthogonal, the previous equation becomes:

$$\mathcal{L}_G(\hat{\underline{\boldsymbol{R}}}_1, \hat{\underline{\boldsymbol{R}}}_2) = \frac{|\hat{\underline{\boldsymbol{R}}}_1|^L |\hat{\underline{\boldsymbol{R}}}_2|^L}{|\frac{1}{2}\left(\hat{\underline{\boldsymbol{R}}}_1 + \hat{\underline{\boldsymbol{R}}}_2\right)|^{2L}} = \prod_{i=1}^{d-1} \frac{\mu_i^L \nu_i^L}{\left(\frac{\mu_i+\nu_i}{2}\right)^{2L}} 2^{4L} \left(\frac{2^4\sigma_n^2\mu_d\nu_d}{\rho_d\rho_{d+1}}\right)^{2L} \tag{7.18}$$

The ratio $\frac{\mu_d \nu_d}{\rho_d^2}$ is also smaller than 1 through Weyl's Inequalities (see below). Again the ratio $\frac{2^{4L}\sigma_n^2}{\rho_{d+1}^2}$ may be very low. The more acquisitions, the more likely the back-scattering signals are to differ and the more probable it becomes to find dissimilar subspaces.

Property 7.1 ***Weyl's Inequalities***
Let $\underline{\boldsymbol{M}}$, $\underline{\boldsymbol{N}}$ *and* $\underline{\boldsymbol{P}} \in \mathbb{C}^N$ *with at least two of them being Hermitian and* $\underline{\boldsymbol{M}}+\underline{\boldsymbol{N}} = \underline{\boldsymbol{P}}$. *Let* $\{\mu_n\}_{1\leq n\leq N}$, $\{\nu_n\}_{1\leq n\leq N}$ *and* $\{\rho_n\}_{1\leq n\leq N}$ *be their respective eigenvalues in decreasing order, then for all* $i = 1, \cdots, N$*:*

$$\mu_i + \nu_N \leq \rho_i \leq \mu_i + \nu_1 \qquad \textit{and} \qquad \mu_N + \nu_i \leq \rho_i \leq \mu_1 + \nu_i$$

If the matrices are also positive definite then:

$$\forall i = 1, \cdots, N, \quad \mu_i < \rho_i \qquad \textit{and} \qquad \nu_i < \rho_i$$

It is also noticeable that the orthogonality example is not a theoretic configuration in urban areas. Indeed, let's consider two signals with only one strong contribution. Their covariance matrices are then:

$$\begin{aligned} \underline{\boldsymbol{R}}_1 &= \mu \underline{\boldsymbol{a}}(h_1)\underline{\boldsymbol{a}}_1^H + \sigma_n^2 \boldsymbol{I}_N \\ \underline{\boldsymbol{R}}_2 &= \nu \underline{\boldsymbol{a}}(h_2)\underline{\boldsymbol{a}}_1^H + \sigma_n^2 \boldsymbol{I}_N \end{aligned} \tag{7.19}$$

where h_1 and h_2 stand for the elevation of the scatterer of the first and second pixel. The signal subspaces are orthogonal if:

$$\sum_{i=1}^{N} \underline{\boldsymbol{a}}(h_1)^H \underline{\boldsymbol{a}}(h_2) = \sum_{i=1}^{N} \exp\big(-\underline{j}\xi_i(h_1 - h_2)\big) = 0 \tag{7.20}$$

When the baselines are uniformly distributed with a sampling distance Δb, the last condition is verified if $\Delta h = \frac{\lambda R}{2\Delta b N}$. For the TerraSAR-X sensors, this corresponds to a difference in elevation of 5m or equivalently 3m variation with respect to the vertical axis. For very bright scatterers with high backscattering power, this implies that they should be seen as more dissimilar than a bright point versus pure noise. This overly discriminative power may be softened by adapting the scale parameter and/or applying a stronger smoothing when estimating the empirical covariance matrices. Moreover, when the number of channels is high, empirical covariance matrices need to be estimated with a lot of samples or need to be shrinked to ensure having non singular matrices. Then, due to the reduction of the pre-estimation quality, it is harder to differentiate neighboring pixels. Large scale and smoothing parameters tend also to make the bias reduction parameter closer to 1 resulting in an output image closer to the input one.

Examples of denoising in a dense urban scenario and the similarity maps corresponding to the GLR under Whishart distribution are presented for increasing numbers of

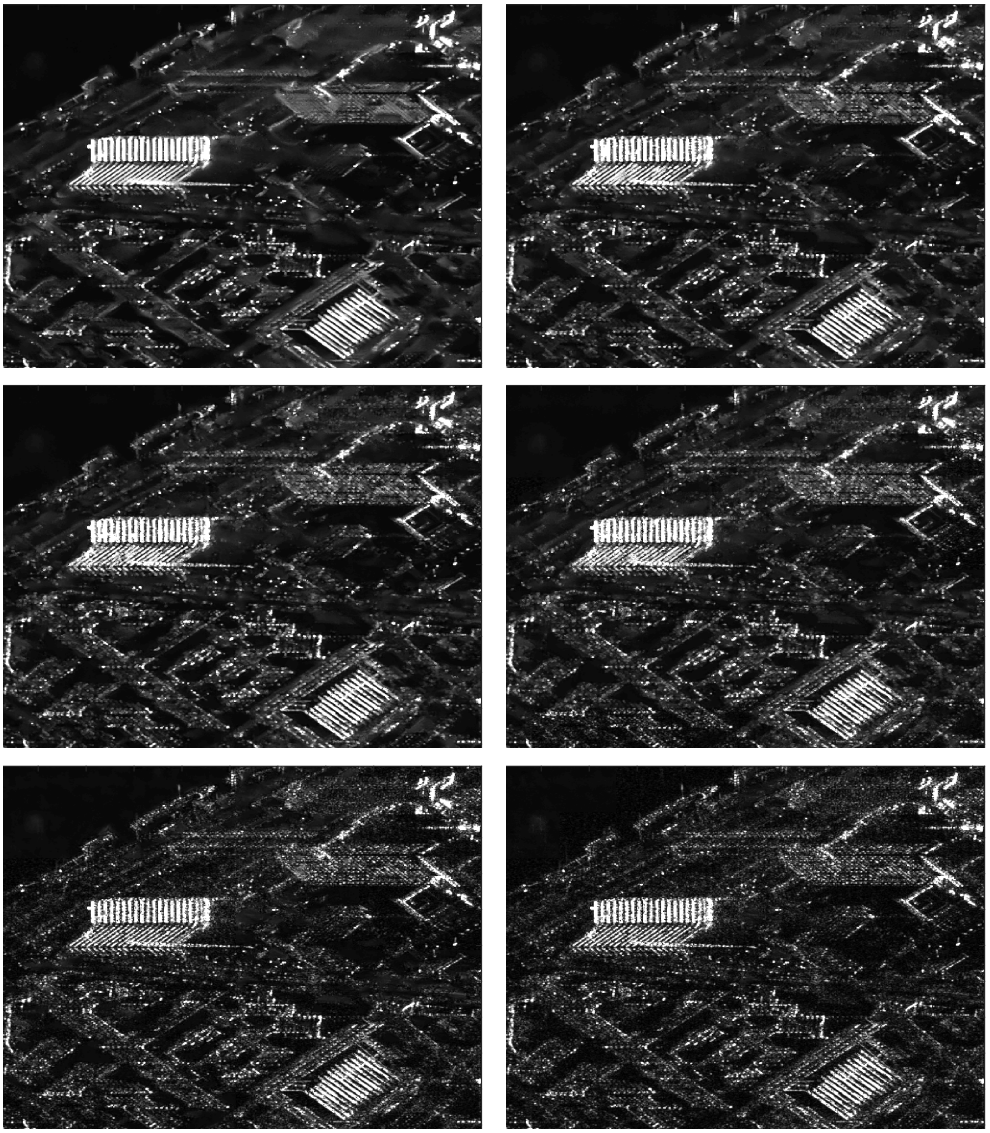


Figure 7.3 – Deterioration of the denoising power of NL-SAR for an increasing number of images. From top left to bottom right the number of channels is 2, 4, 6, 8, 10 and 12.

images in Fig. 7.3 and 7.4.

Non-local approaches and specially NL-SAR algorithm are nevertheless still of interest for SAR tomography. It has proven to improve the tomographic estimation for distributed 3-D backscattering mechanisms such as forest or ice landscapes (Aghababaee *et al.*, 2017). It has also been used successfully for small sets of data of hybrid areas (D'Hondt. and Al., 2017) (D'Hondt *et al.*, 2018). Few strategies have been proposed to extend its good performances to high number of channels. A recently proposed approach is to perform the denoising for each pair of interferometric images and then reconstruct the covariance matrix (Shi *et al.*, 2018). This however does not guaranty a coherent estimation of the weights between pairs or to take into account long-term variation of the data.

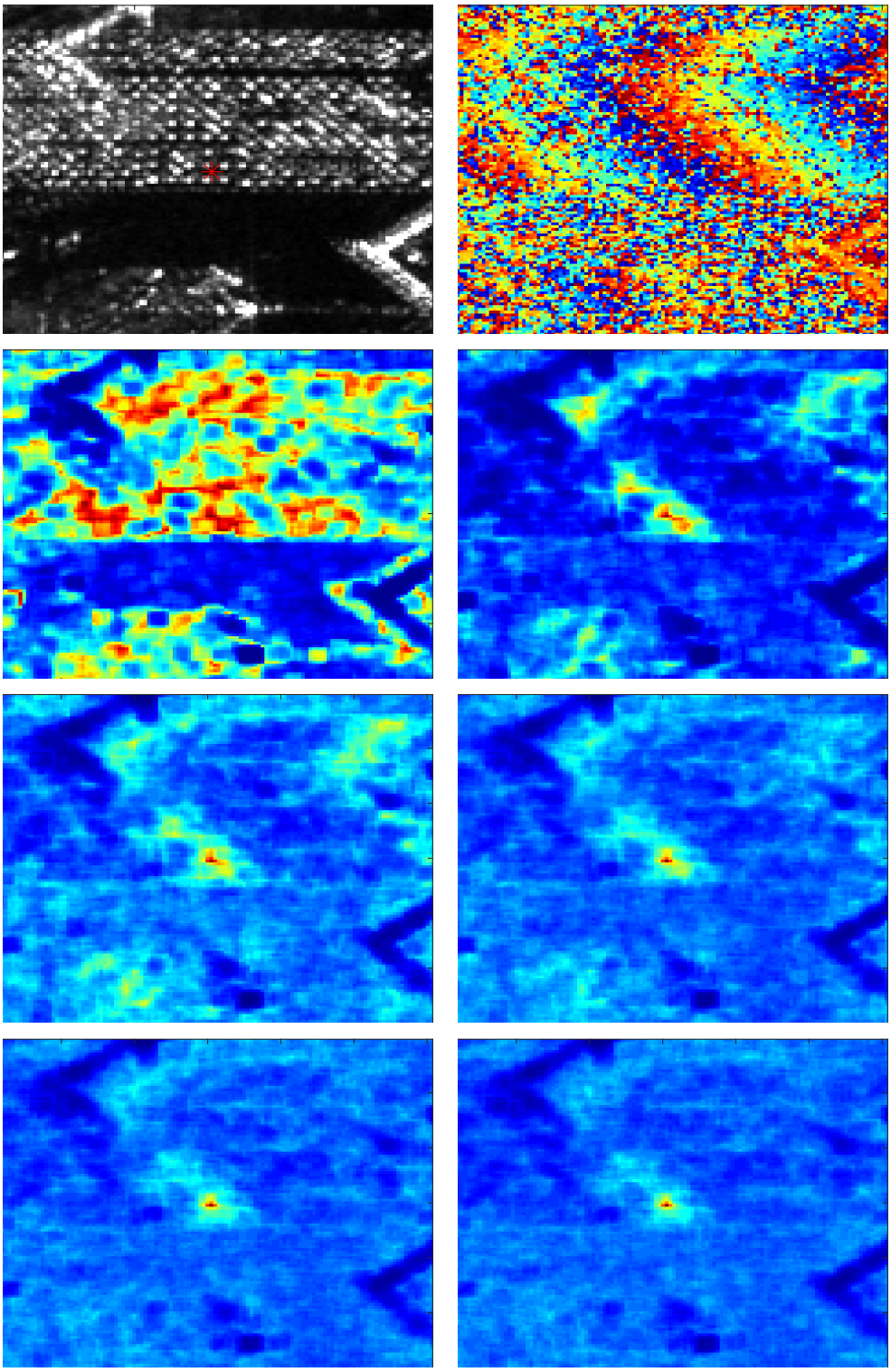


Figure 7.4 – Deterioration of the discriminative power of the GLR for increasing number of images. From top left to bottom right the number of channels is 2, 4, 6, 8, 10 and 12. On top the amplitude image and one interferogram. The central pixel in red in the amplitude image is compared to the rest of the window.

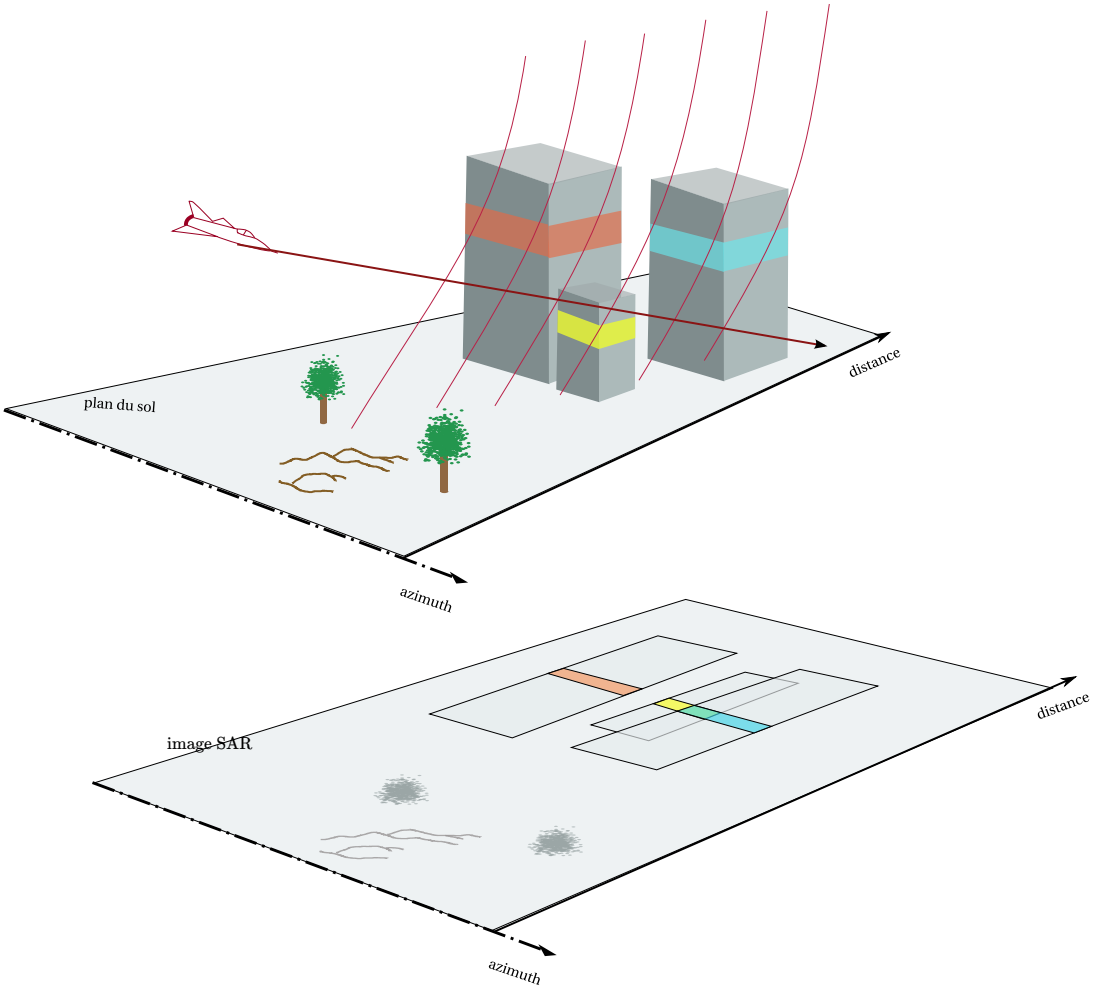


Figure 7.5 – Selection of similar pixels for SAR tomography should be based on scatterers location and reflectivity. In this example, pixels in the window of the same color should be seen as similar as they present the same mixture of scatterers: same height and same buildings.

7.4 Proposed PS and background based similarity

For SAR tomography, the interferometric phase is the most important information to retrieve. The similarity could then be improved in focusing more on the scatterers geometric distribution inside the SAR pixels as illustrated in Fig. 7.5. In urban areas, most of the usable information is carried by the PS that stay highly coherent between the acquisitions. A new similarity function could then be based on the conditional or hybrid model to have an explicit information of the strong scatterers in the pixels distribution. This contribution was presented in (Rambour *et al.*, 2017). Under hybrid model, the pixel $\underline{\boldsymbol{v}}$ follows an non-centered N dimensional circular Gaussian distribution:

$$\mathrm{p}\left(\underline{\boldsymbol{v}} \,|\, \underline{\boldsymbol{R}}, \boldsymbol{h}, \underline{\boldsymbol{u}}\right) = \frac{1}{\pi^N |\underline{\boldsymbol{R}}|} \exp\left[-\left(\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}\right)^H \underline{\boldsymbol{R}}^{-1}\left(\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{h})\underline{\boldsymbol{u}}\right)\right] \tag{7.21}$$

To determine if two pixels $\underline{\boldsymbol{v}}_1$ and $\underline{\boldsymbol{v}}_2$ are similar or not based on their PS like scatterers, we adopt an iterative approach where we jointly extract their PS components, update the mean and covariance of their distribution and test if they are likely to share the same distribution or not. Rather than testing the equality of height and reflectivity of the targets the similarity is based on their likelihood as in equation (7.8). The extraction of the main components of the signal is done using the M-RELAX (Li and Stoica, 1996) scheme to solve the following non-linear least squares:

$$\left(\hat{\underline{\boldsymbol{u}}}, \boldsymbol{z}\right) = \underset{(\underline{\boldsymbol{u}}, \boldsymbol{z})}{\operatorname{argmin}} \; ||\underline{\boldsymbol{v}} - \underline{\boldsymbol{A}}(\boldsymbol{z})\underline{\boldsymbol{u}}||^2_{\underline{\boldsymbol{R}}^{-1}} \tag{7.22}$$

The proposed algorithm iteratively estimates the main signal components from the empirical covariance matrix of the two pixels $\underline{\boldsymbol{v}}_1$ and $\underline{\boldsymbol{v}}_2$. The extracted targets are compared with the ones of the mixed pixel $\underline{\boldsymbol{v}}_{12} = \frac{1}{2}\left(\underline{\boldsymbol{v}}_1 + \underline{\boldsymbol{v}}_2\right)$ to compute the similarity.

After the first iteration of the algorithm, the first components from $\hat{\underline{\boldsymbol{R}}}_1$, $\hat{\underline{\boldsymbol{R}}}_2$ and $\hat{\underline{\boldsymbol{R}}}_{12} = \frac{1}{2}\left(\hat{\underline{\boldsymbol{R}}}_1 + \hat{\underline{\boldsymbol{R}}}_2\right)$ are extracted, respectively $(\hat{\underline{u}}_1^{(1)}, \hat{z}_1^{(1)})$, $(\hat{\underline{u}}_2^{(1)}, \hat{z}_2^{(1)})$ and $(\hat{\underline{u}}_{12}^{(1)}, \hat{z}_{12}^{(1)})$. The estimation of the empirical covariance matrices of $\underline{\boldsymbol{v}}_1$ and $\underline{\boldsymbol{v}}_2$ can then be updated as in the algorithm M-RELAX :

$$\hat{\boldsymbol{v}}_{\boldsymbol{i}}^{(1)} = \boldsymbol{v}_{\boldsymbol{i}} - \hat{\underline{u}}_i^{(1)} \underline{\boldsymbol{a}}(\hat{z}_i^{(1)}) \tag{7.23}$$

$$\hat{\underline{\boldsymbol{R}}}_i^{(1)} = \frac{1}{L}\sum_{l=1}^{L} \bar{\boldsymbol{v}}_{i_l}^{(1)} \bar{\boldsymbol{v}}_{i_l}^{(1)H} \text{ for } i = \{1, 2\}. \tag{7.24}$$

Then, extracted scatterers similarity is derived from the following hypotheses:

$$\begin{cases} \mathcal{H}_0^{(1)} : & \underline{u}_1^{(1)} = \underline{u}_2^{(1)} = \underline{u}_{12}^{(1)} \quad \text{and} \quad z_1^{(1)} = z_2^{(1)} = z_{12}^{(1)} \\ \mathcal{H}_1^{(1)} : & \underline{u}_1^{(1)} \neq \underline{u}_2^{(1)} \quad \text{or} \quad z_1^{(1)} \neq z_2^{(1)} \end{cases} \tag{7.25}$$

and the corresponding likelihood ratio :

$$\mathcal{L}^{(1)} = \frac{\mathrm{p}(\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2 | \mathcal{H}_0, \underline{\boldsymbol{R}}_{12}^{(1)}, \underline{u}_{12}^{(1)}, z_{12}^{(1)})}{\mathrm{p}(\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2 | \mathcal{H}_1, \underline{\boldsymbol{R}}_1^{(1)}, \underline{u}_1^{(1)}, z_1^{(1)}, \underline{\boldsymbol{R}}_2^{(1)}, \underline{u}_2^{(1)}, z_2^{(1)})} \tag{7.26}$$

The log of the generalized likelihood ratio test is computed by injecting the parameters values in (7.26) and takes the following form:

$$\begin{aligned} \log \mathcal{L}_G^{(1)} = \log \frac{|\underline{\boldsymbol{R}}_1^{(1)}||\hat{\underline{\boldsymbol{R}}}^{(1)}|}{|\hat{\underline{\boldsymbol{R}}}_{12}^{(1)}|^2} &+ ||\underline{\boldsymbol{v}}_1 - \underline{\boldsymbol{a}}(\hat{z}_1^{(1)})\hat{\underline{u}}_1^{(1)}||^2_{\hat{\underline{\boldsymbol{R}}}_1^{-1}} \\ &+ ||\underline{\boldsymbol{v}}_2 - \underline{\boldsymbol{a}}(\hat{z}_2^{(1)})\hat{\underline{u}}_2^{(1)}||^2_{\hat{\underline{\boldsymbol{R}}}_2^{-1}} - ||\underline{\boldsymbol{v}}_1 - \underline{\boldsymbol{a}}(\hat{z}_{12}^{(1)})\hat{\underline{u}}_{12}^{(1)}||^2_{\hat{\underline{\boldsymbol{R}}}_{12}^{-1}} \\ &- ||\underline{\boldsymbol{v}}_2 - \underline{\boldsymbol{a}}(\hat{z}_{12}^{(1)})\hat{\underline{u}}_{12}^{(1)}||^2_{\hat{\underline{\boldsymbol{R}}}_{12}^{-1}} \end{aligned} \tag{7.27}$$

where $||\boldsymbol{x}||_M^2 = \boldsymbol{x}^H \boldsymbol{M} \boldsymbol{x}$. The procedure can be repeated for all the scatterers contained in the tested pixels, *e.g* for the k^{th} iteration the following hypothesis are tested:

$$\begin{cases} \mathcal{H}_0^{(k)} : & \underline{u}_1^{(k)} = \underline{u}_2^{(k)} = \underline{u}_{12}^{(k)} \quad \text{and} \quad z_1^{(k)} = z_2^{(k)} = z_{12}^{(k)} \\ \mathcal{H}_1^{(k)} : & \underline{u}_1^{(k)} \neq \underline{u}_2^{(k)} \quad \text{or} \quad z_1^{(k)} \neq z_2^{(k)} \end{cases}$$

where the covariance matrix, the altitudes and amplitudes are updated at each iteration. As just a few strong scatterers are expected to be present in each pixel, only three targets are tested for each pixel. The similarity criterion is then given by the sum of all the log likelihood ratio :

$$\mathcal{L} = \sum_k \log \mathcal{L}_G^{(k)} \tag{7.28}$$

This ensures to take into account the similarity between all the different contributions and avoids to have any parameter to tune. Through the algorithm iterations and the parameters updates, different decompositions of the signal into stable strong scatterers and residual are jointly tested. Moreover as the first contributions are several magnitude higher than the following ones, the value of $\mathcal{L}$ is conditioned by the first elements in the sum in (7.28). Examples of similarity map are presented in Fig. 7.6 and 7.7. The similarity of the central pixel with the rest of the window is drawn using the Wishart GLR and the proposed similarity. The window has a size 120x120 and the 2000 most similar pixels are overlaid on the amplitude image. The number of images here is 20 and 5 iterations are done. In Fig. 7.6 the observed scene corresponds to a skyscraper (the Keller Tower) next to a shadow area. Similar points should then correspond to part of the building at the same altitude as the central pixel. The Wishart GLR doesn't manage to well select the pixels containing iso-height scatterers whereas the proposed similarity detect the correct samples on the building. More important, the shadow area is clearly seen as dissimilar as the building which is not the case with Wishart GLR. In Fig. 7.7,

the observed buildings are lower and closer to each others. In this example, the Wishart GLR manages to select points corresponding to buildings but the similarity image does not show a lot of contrast. The proposed similarity presents a more structured selection with a clear preference for walls. The selected points are roughly concentrated at the same height as the central pixel. Finally, the discriminative power of the proposed similarity is robust to an increasing number of images. This point is illustrated in Fig. 7.8 on the Keller Tower and 7.9 on dense smaller buildings for tomographic stacks of size 10, 20, 30 and 40.

7.5 Limits of Non Local approaches for urban tomography

Although we have shown that the similarity used for non-local selection can be improved, this approach still presents drawbacks for SAR tomography over urban areas. Firstly, the proposed similarity function is time consuming and can then hardly be consider for a non local approach. Secondly, the current debiasing step can lead to singular matrix specially for big dimensions which may completely distort the tomographic estimations. The *max* operator used in the debiasing step of NL-SAR could thus be changed to a less radical one for big dimensions. Moreover, even for dense urban areas, no clear benefit can be notice using a kernel based on the similarity as illustrated in the Fig. 7.10. Finally, for dense urban areas, finding enough pixels containing exactly the same distribution of scatterers may become impossible for a high number of SAR images. All these considerations can be seen as the classical chicken or the egg problem. Indeed to find the most similar samples, one needs to estimate its components. The current non-local approaches for SAR tomography lead then to this mildly unsatisfying conclusion: the more data, the harder it is to restore them.

Rather than trying to differentiate pixels based on their content to refine their components estimation, a straight forward backprojection with priors on the results should give a simpler and more efficient framework. The good results obtained with CS for urban areas are in favor of this strategy. In the next part this option is studied along with the difficulty to add structural priors on results obtained in a RADAR configuration.

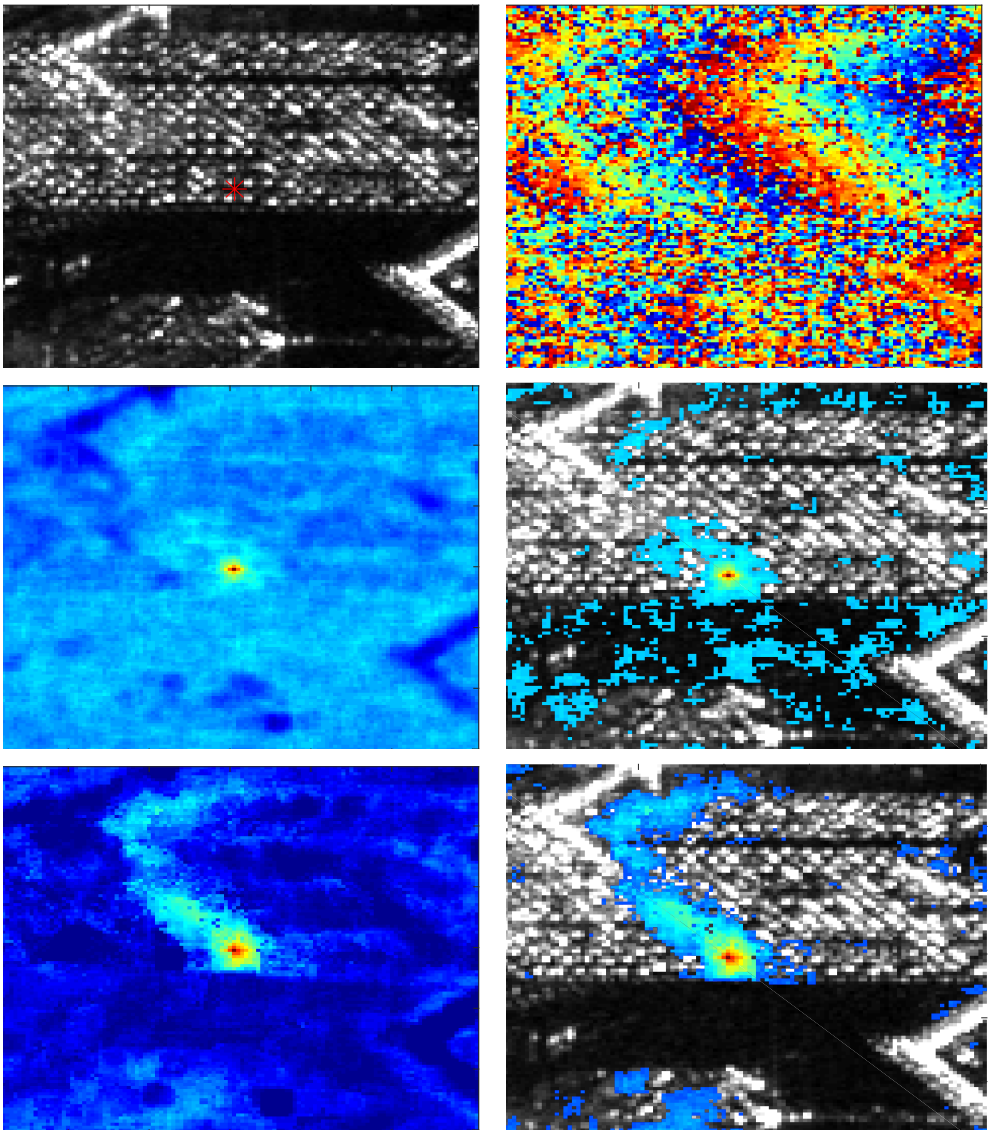


Figure 7.6 – Comparison of the selective power of the proposed similarity and the Wishart GLR. On top the amplitude image and one interferogram. The central pixel indicated in red in the amplitude image is compared to the rest of the window. On the middle the similarity map obtained with Wishart GLR and the 2000 most similar pixels. On the bottom the similarity map obtained with the proposed method and the 2000 most similar pixels.

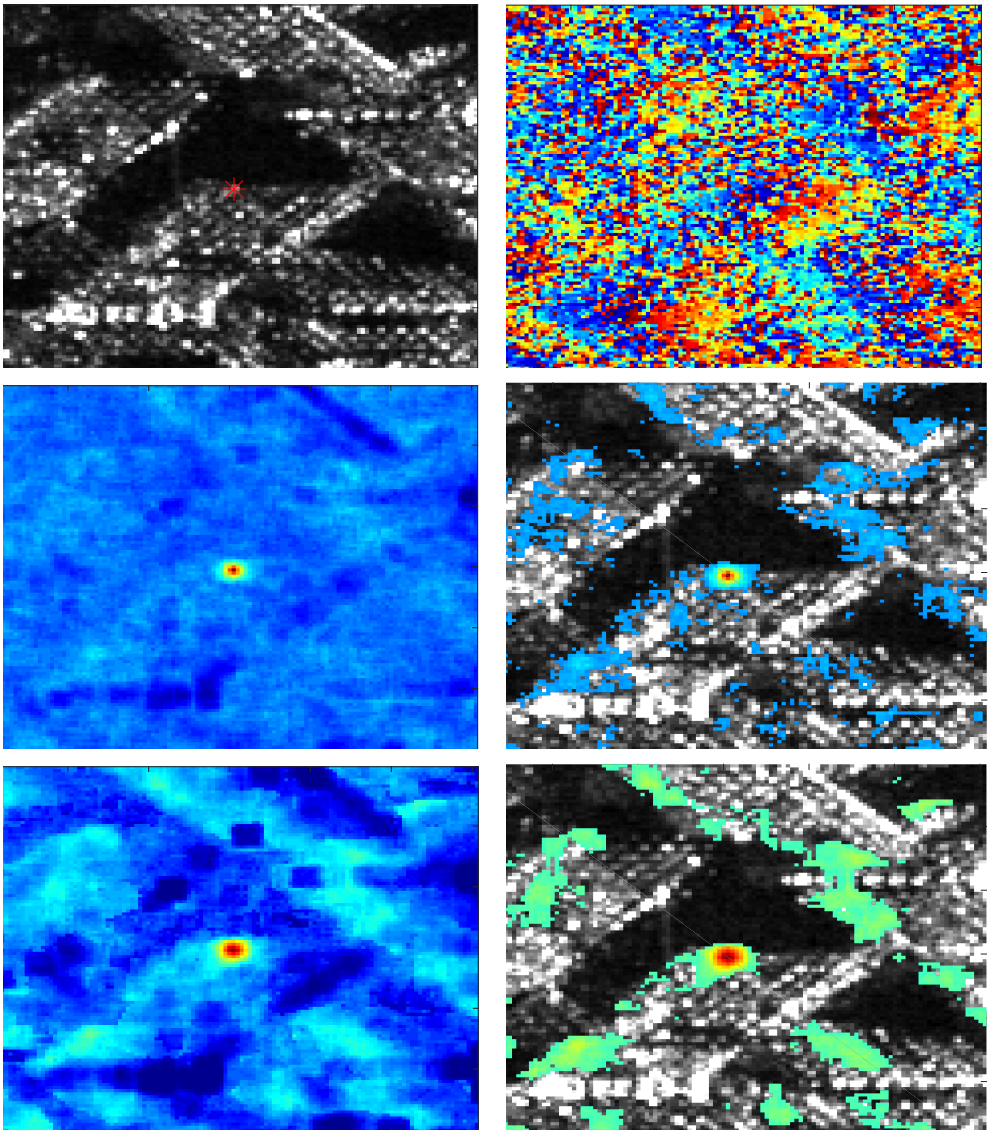


Figure 7.7 – Comparison of the selective power of the proposed similarity and the Wishart GLR. On top the amplitude image and one interferogram. The central pixel indicated in red in the amplitude image is compared to the rest of the window. On the middle the similarity map obtained with Wishart GLR and the 2000 most similar pixels. On the bottom the similarity map obtained with the proposed method and the 2000 most similar pixels.

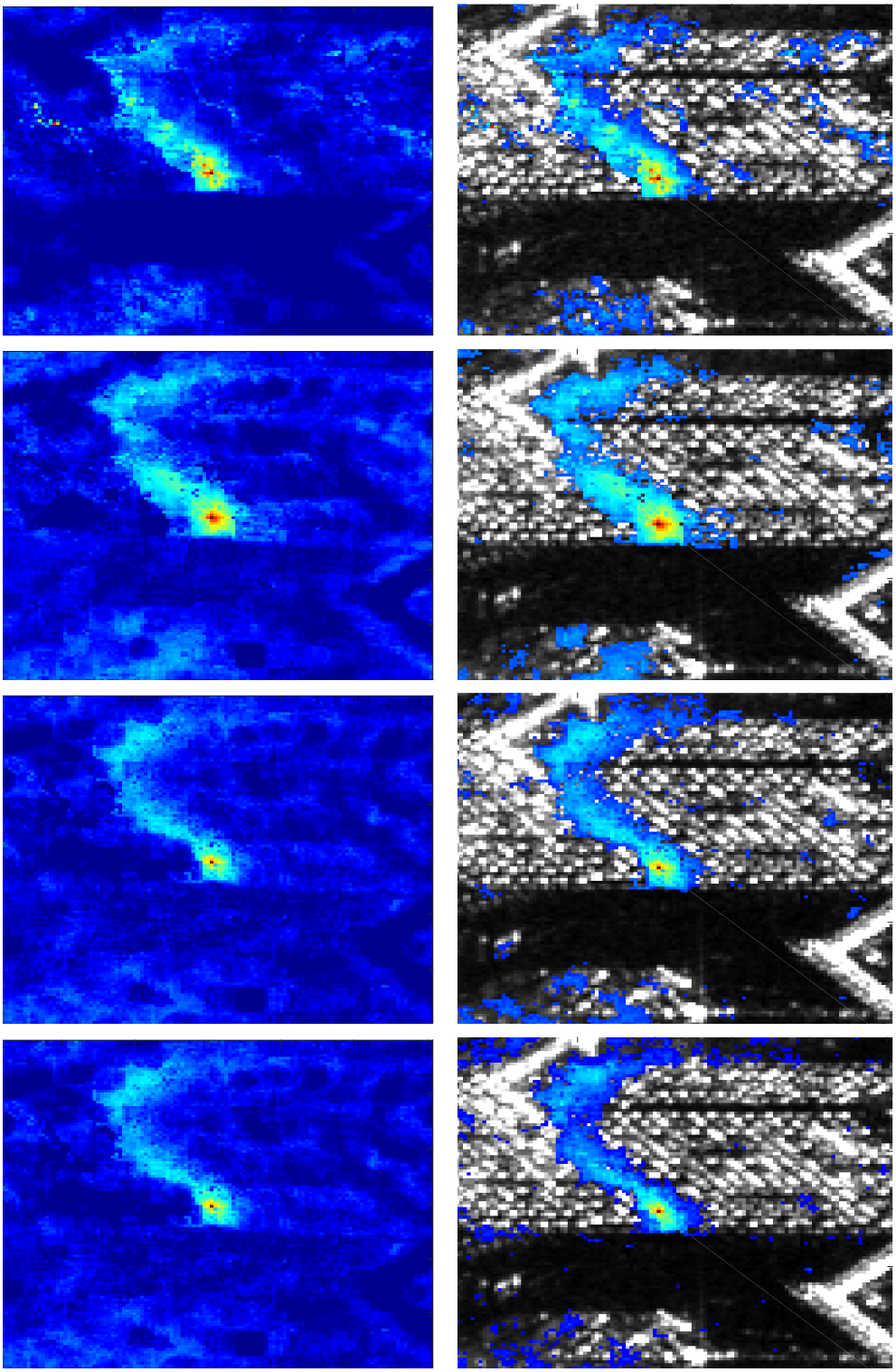


Figure 7.8 – Comparison of the selective power of the proposed similarity for increasing number of SAR images in the tomographic pile. From top left to bottom right the number of images is $N = 10, 20, 30, 40$. The selection of the pixels is robust to the increasing size of the images stack: most of the selected pixels are located at the same height and on the same building.

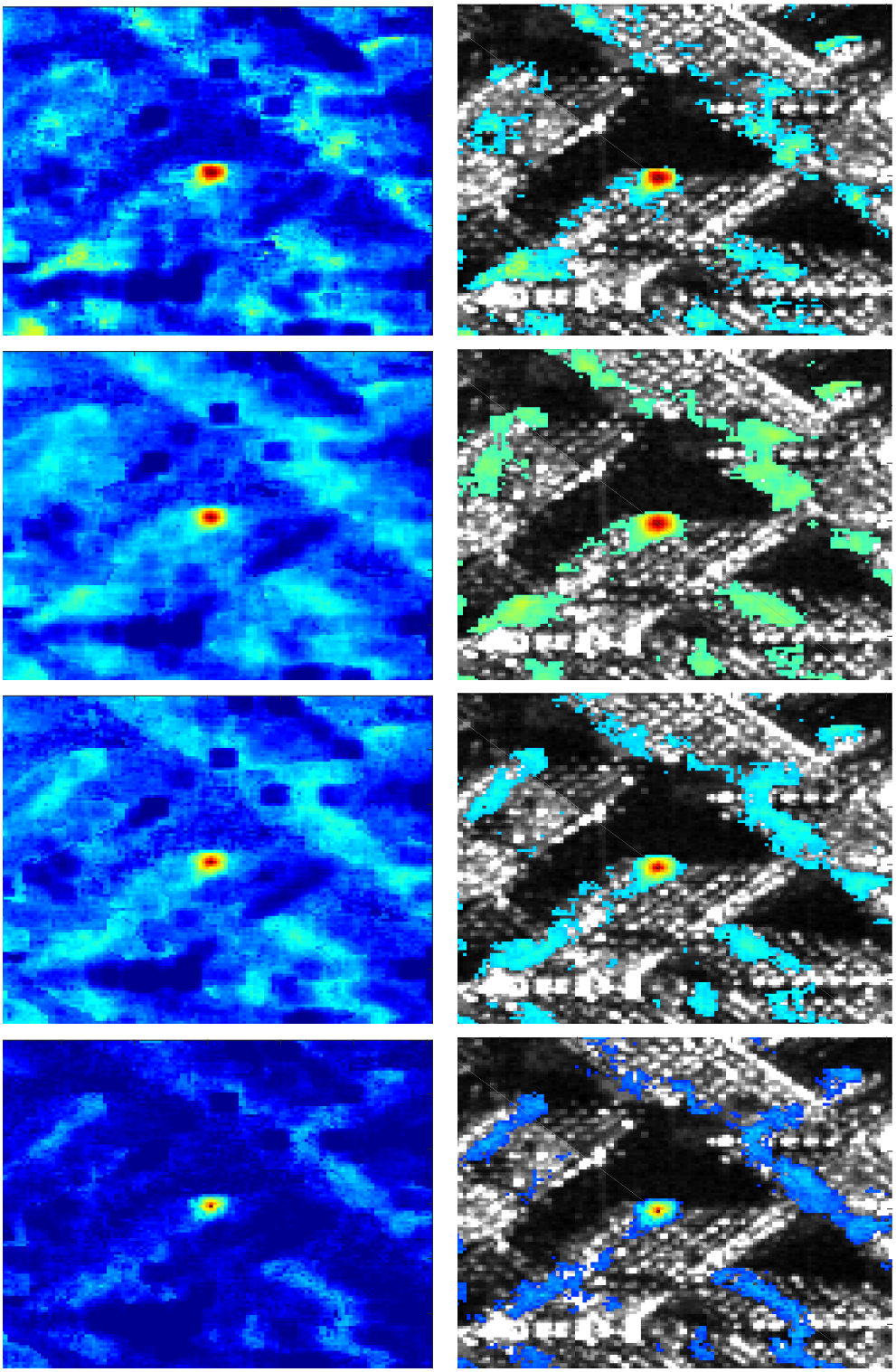


Figure 7.9 – Comparison of the selective power of the proposed similarity for increasing number of SAR images in the tomographic pile for a pixel of dense buildings. From top left to bottom right the number of images is $N = 10, 20, 30, 40$. The selection of the pixels is robust to the increasing size of the images stack: most of the selected pixels are located at the same height and on the same building.

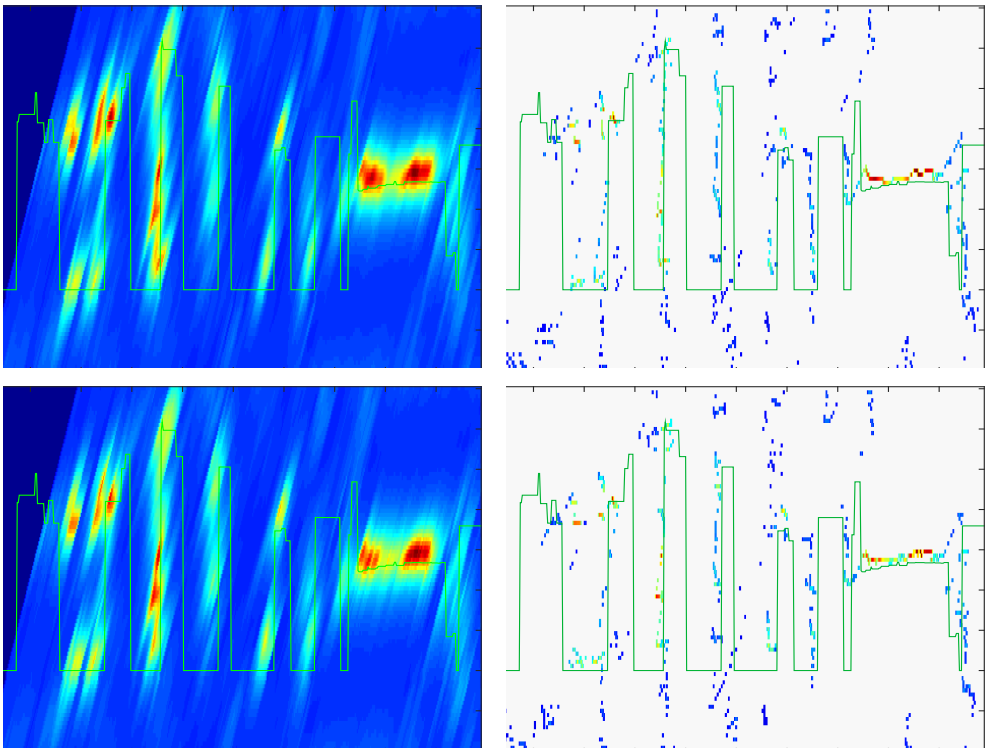


Figure 7.10 – Comparison of the tomographic reconstruction for one profile of consisting of dense buildings. The top row has been obtained using a gausscar kernel whereas the bottom one weighted the samples using the proposed similarity criterion. On the left the results correspond to the MUSIC-PS and on the right to the discrete MUSIC with 3 scatterers for each cell. The ground truth is shown in green.

Chapter 8

Tomographic inversion with spatial regularization

The presented work in this chapter comes mostly from (Rambour *et al.*, 2018a). Some detailed were added in the optimization section.

In order to enforce some spatial smoothness, the tomographic inversion has to be performed globally in ground coordinates. Rather than considering the collection of measurements at a given radar pixel, from now on the notations $\underline{\boldsymbol{u}} \in \mathbb{C}^{N_x.N_y.N_z}$ and $\underline{\boldsymbol{v}} \in \mathbb{C}^{N_x.N_r.N}$ will respectively refer to column vectors obtained by stacking all the values in the 3-D volume, and all the values in the tomographic stack of SAR images. N_x, N_y and N_z are the number of voxels in each direction in ground geometry while N_r is the size of SAR images in the range direction and N is the number of images in the tomographic stack. The linear operator $\underline{\boldsymbol{\Phi}} \in \mathbb{C}^{(N_x.N_r.N)\times(N_x.N_y.N_z)}$ maps the volume of complex reflectivities in 3-D space to the complex amplitudes in the tomographic stack of SAR images, see Fig 8.1. An element of $\underline{\boldsymbol{\Phi}}$ is defined as:

$$\underline{\boldsymbol{\Phi}}_{i,j} = \begin{cases} \exp(-\underline{j}\varphi) & \text{if } x_i = x_j \text{ and} \\ & r_i - \frac{\delta_r}{2} < \rho_{1;y_j,z_j} < r_i + \frac{\delta_r}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{8.1}$$

with $\varphi = \xi_i h(y_j, z_j)$ the phase shift due to the path between voxel j with coordinates (x_j, y_j, z_j) and antenna i. As in the previous equations, $\rho_{1;y_j,z_j}$ corresponds to the distance between antenna 1 (of the master image) and the point with ground coordinates (y_j, z_j). The size of a radar pixel in the range direction is noted δ_r.

The construction of the matrix $\underline{\boldsymbol{\Phi}}$ is illustrated on Fig. 8.2 and $\underline{\boldsymbol{\Phi}}$ is sparse: only a few entries are different from zero, so that products of the form $\underline{\boldsymbol{\Phi}\boldsymbol{u}}$ can be computed efficiently.

The observed SAR tomographic stack $\underline{\boldsymbol{v}}$ can be modeled by the following (complex-valued) linear model:

$$\underline{\boldsymbol{v}} = \underline{\boldsymbol{\Phi}}\,\underline{\boldsymbol{u}} + \underline{\boldsymbol{\epsilon}} \tag{8.2}$$

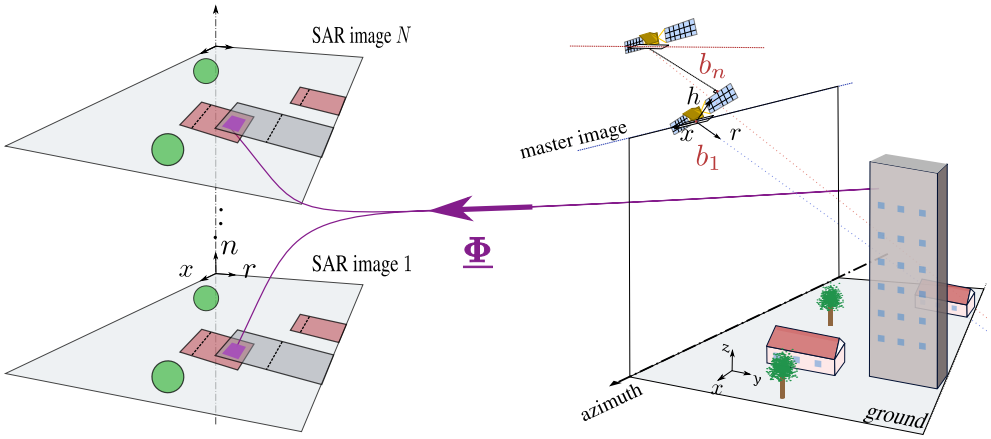


Figure 8.1 – The operator $\underline{\mathbf{\Phi}}$ performs the projection of the voxels from the 3-D scene in ground geometry on the SAR tomographic stack.

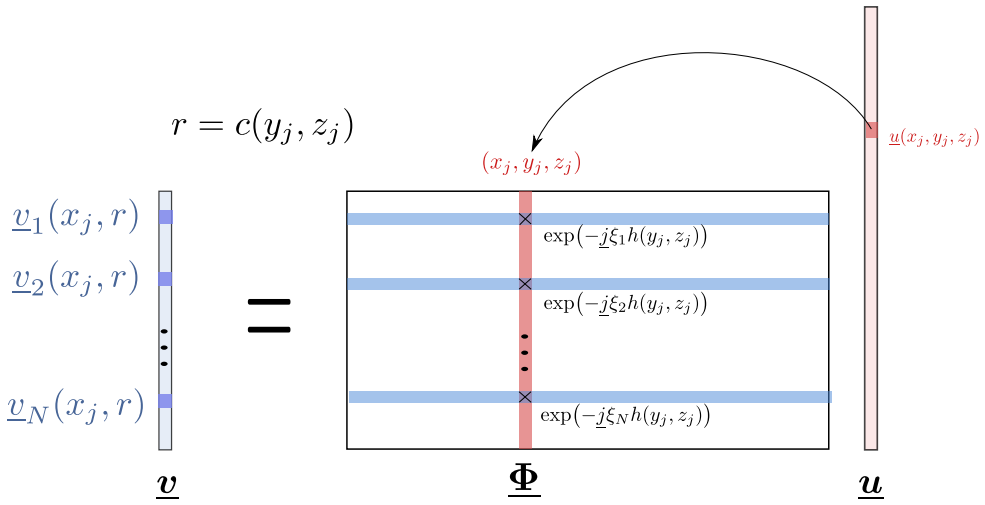


Figure 8.2 – Construction of the matrix $\underline{\mathbf{\Phi}}$. The radar cell associated to each position of the 3-D scene is computed as well as the phase term corresponding to the distance between a point in the 3-D space and a given antenna.

where $\underline{\boldsymbol{\epsilon}}$ stands for the noise. This corresponds to a generalization of the tomographic direct model (4.1) where we additionally consider the geometric transformation from ground geometry to SAR geometry and model at once the measurements for all the pixels. In order to invert this tomographic model, it is necessary to introduce some regularization terms.

8.1 Ground base spatial priors for SAR tomography

It is often desired to reconstruct volumes with a discretization in heights that is finer than the resolution given by the synthetic aperture in the height direction (i.e., super-resolution). The inversion of equation (8.2) is therefore ill-posed (more unknowns than measurements) and requires some regularization. In the following, we denote $\mathcal{R}$ the regularization function. Since the intrinsic phase of a scatterer is typically modeled

as uniformly distributed and independent from one scatterer to another, no specific regularization can be enforced on the phase of our unknown complex reflectivities $\underline{\boldsymbol{u}}$. We define the regularization $\mathcal{R}$ as a function of the modulus of $\underline{\boldsymbol{u}}$ only. The reconstruction of the volume of complex reflectivities $\hat{\underline{\boldsymbol{u}}}$ is thus obtained by solving an optimization problem of the general form:

$$\hat{\underline{\boldsymbol{u}}} = \arg\min_{\underline{\boldsymbol{u}}} \frac{1}{2}||\underline{\boldsymbol{\Phi}}\,\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}||_2^2 + \mathcal{R}(|\underline{\boldsymbol{u}}|) \tag{8.3}$$

To design the regularization function $\mathcal{R}$, we need to select a function that favors volumes of reflectivities $|\underline{\boldsymbol{u}}|$ that are often present in urban environments. Many different such functions could be considered, we selected a function based on the two following remarks:

- The 3-D scene can be represented as a sparse volume. Indeed a good reconstruction should retrieve only the illuminated part of the buildings and of the ground. The estimated volume is then mostly filled with zero intensity voxels.
- The illuminated structures are spatially smooth (continuous surfaces: frontages, rooftops, ground).

These remarks suggest the following regularization function:

$$\begin{aligned}\forall \boldsymbol{w} \in \mathbb{R}^{N_x.N_y.N_z},\ \mathcal{R}(\boldsymbol{w}) = &\frac{\mu_{\mathrm{x}}}{2}\|\boldsymbol{D}_x\boldsymbol{w}\|_2^2 + \frac{\mu_{\mathrm{y}}}{2}\|\boldsymbol{D}_y\boldsymbol{w}\|_2^2 \\ &+ \frac{\mu_{\mathrm{z}}}{2}\|\boldsymbol{D}_z\boldsymbol{w}\|_2^2 + \mu_{\ell_1}\|\boldsymbol{w}\|_1\end{aligned} \tag{8.4}$$

where the matrices $\boldsymbol{D}_x$, $\boldsymbol{D}_y$ and $\boldsymbol{D}_z$ stand for the finite differences operators in the x, y and z directions, and parameters μ_x, μ_y, μ_z and μ_{ℓ_1} weight each term. The ℓ_1 norm favors vectors $\underline{\boldsymbol{w}}$ with many zeros while the terms with the finite difference operators enforce a spatial smoothness. As μ_{ℓ_1} controls the sparsity of the reconstructed volume, it plays a crucial role.

In urban environments, dihedral and trihedral structures produce very strong echoes. When only an ℓ_1 norm is minimized, it is hard to reconstruct at the same time very strong scatterers and weaker scatterers on the ground or rooftops. The role of the first three terms is thus to favor spatial smoothness and hence preserve these scatterers whenever they are close to other scatterers. In urban areas most of the buildings are expected to show straight walls along the z direction. Of course the true orientation of the objects is a priori unknown and it is most likely that they may not always follow the x and y directions. However, when the horizontal smoothing is small compared to the power of the scatterers, the only major cost is to introduce some thickness in vertical objects with the benefit to ensure a global reduction of isolated outliers and retrieve part of flat areas with a lower intensity. When reconstructing images of urban areas, we were unable to find a value of the sparsity parameter μ_{ℓ_1} that would both preserve the weakest scatterers and successfully suppress side-lobes in areas with strong scatterers. To improve the reconstructions, we introduced a spatially variant regularization based

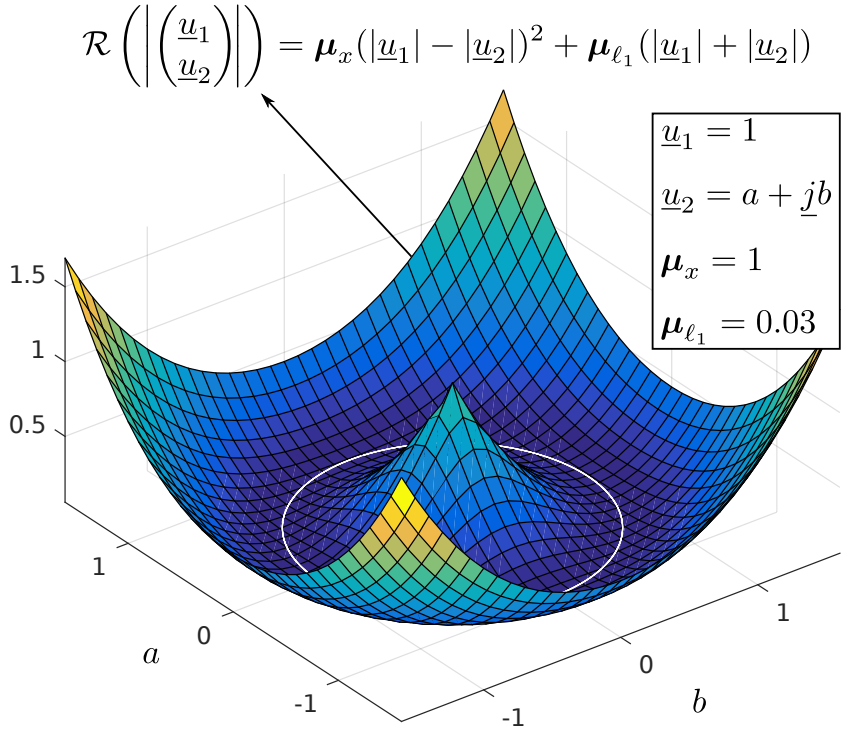


Figure 8.3 – Illustration of the non-convexity of spatial regularizations expressed on the modulus of the complex reflectivities.

on the square-root of the estimated intensity of the master image which can be obtained using a denoising algorithm such as NL-SAR (Deledalle *et al.*, 2015) or simply using the average intensity depending of the configuration of the acquisitions. We define the diagonal matrix $\boldsymbol{D}_{\ell_1}$ whose j-th diagonal entry is equal to the square-root of the estimated intensity at the corresponding azimuth and range coordinates (i.e., such that the range r verifies $r - \frac{\delta_r}{2} < \rho_{1;y_j,z_j} < r + \frac{\delta_r}{2}$). The equation (8.4) is then modified into:

$$\begin{aligned}\forall \boldsymbol{w} \in \mathbb{R}^{N_x.N_y.N_z},\ \mathcal{R}(\boldsymbol{w}) = \frac{\mu_{\mathrm{x}}}{2}\|\boldsymbol{D}_x\boldsymbol{w}\|_2^2 + \frac{\mu_{\mathrm{y}}}{2}\|\boldsymbol{D}_y\boldsymbol{w}\|_2^2 \\ + \frac{\mu_{\mathrm{z}}}{2}\|\boldsymbol{D}_z\boldsymbol{w}\|_2^2 + \mu_{\ell_1}\mathbf{1}^T\boldsymbol{D}_{\ell_1}\boldsymbol{w}\end{aligned} \tag{8.5}$$

where $\mathbf{1}$ is the vector of size $N_x.N_y.N_z$ with each entry equal to 1.

8.2 Optimization algorithm

The minimization problem (8.3) is not easy. It is indeed large scale (millions up to several billions unknowns) and non-convex. The non-convexity is illustrated in a simple case where the vector $\underline{\boldsymbol{u}}$ has only two elements in Fig. 8.3: the spatial smoothness favors vectors such that the modulus of each entry is close. Since the regularization is independent on the phase, the minimum (for a fixed value of $\underline{u}_1$) corresponds to a set of complex values with equal modulus (the white circle drawn on Fig. 8.3).

We have to use an algorithm that can deal with the non-convexity of the problem. To solve the minimization problem (8.3), we use a variable splitting approach in order to break down the problem into a sequence of simpler problems. We introduce two new vectors: $\underline{\boldsymbol{f}}$ (complex-valued: $\underline{\boldsymbol{f}} \in \mathbb{C}^{(N_x.N_y.N_z)}$) and $\boldsymbol{w}$ (real-valued and non-negative: $\boldsymbol{w} \in \mathbb{R}^{+(N_x.N_y.N_z)}$). Problem (8.3) is formally equivalent to the following constrained problem:

$$\hat{\underline{\boldsymbol{u}}} = \arg\min_{\underline{\boldsymbol{u}}} \frac{1}{2}\|\underline{\boldsymbol{\Phi}}\,\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}\|_2^2 + \mathcal{R}(\boldsymbol{w}) \tag{8.6}$$
$$\text{s.t.} \begin{cases} \underline{\boldsymbol{u}} = \underline{\boldsymbol{f}} \\ |\underline{\boldsymbol{f}}| = \boldsymbol{w} \end{cases}$$

Different methods could be consider to find a solution of (8.7). The Lagrangian dual ascent for instance would be one them if it was for the non-convexity introduced by the modulus. To insure convergence we then look at algorithms exploiting the augmented Lagrangian of the previous problem:

$$\begin{aligned}\mathcal{L}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{f}}, \boldsymbol{w}, \underline{\boldsymbol{d}}_1, \boldsymbol{d}_2) =& \frac{1}{2}||\underline{\boldsymbol{\Phi u}} - \underline{\boldsymbol{v}}||_2^2 + \frac{\beta_1}{2}||\underline{\boldsymbol{f}} - \underline{\boldsymbol{u}} + \underline{\boldsymbol{d}}_1||_2^2 \\ &+ \frac{\beta_2}{2}\|\boldsymbol{w} - |\underline{\boldsymbol{f}}| + \boldsymbol{d}_2\|_2^2 + \mathcal{R}(\boldsymbol{w})\end{aligned} \tag{8.7}$$

where $\underline{\boldsymbol{d}}_1 \in \mathbb{C}^{(N_x.N_y.N_z)}$ and $\boldsymbol{d}_2 \in \mathbb{R}^{(N_x.N_y.N_z)}$ are the scaled dual variables and β_1 and β_2 are penalty parameters (relevant only to the optimization method, i.e., impacting the convergence). Alternating Direction Method of Multipliers (ADMM) is a well known augmented Lagrangian method. It backs from works in the 1970s (Glowinski, R., Marroco, A., 1975)(gab, 1976) and is very close to other dual decomposition algorithms such as the Lagrangian Methods of Multipliers (LMM) (Bertsekas, 2014), the Douglas-Rachford splittings (Douglas and Rachford, 1956), Dykstra's alternating projections (Boyle and Dykstra, 1986) or Bregman iterative algorithm for ℓ_1 minimization problems (Yin *et al.*, 2008). ADMM is based on the LMM algorithm which for the current problem consists in the following iterations:

$$\left(\underline{\boldsymbol{u}}^{(k+1)}, \underline{\boldsymbol{f}}^{(k+1)}, \boldsymbol{w}^{(k+1)}\right) = \underset{\underline{\boldsymbol{u}}, \underline{\boldsymbol{f}}, \boldsymbol{w}}{\text{argmin}} \mathcal{L}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{f}}, \boldsymbol{w}, \underline{\boldsymbol{d}}_1^{(k)}, \boldsymbol{d}_2^{(k)}) \tag{8.8}$$

$$\underline{\boldsymbol{d}}_1^{(k+1)} = \underline{\boldsymbol{d}}_1^{(k)} + \frac{\beta_1}{2}\left(\underline{\boldsymbol{f}}^{(k+1)} - \underline{\boldsymbol{u}}^{(k+1)}\right) \tag{8.9}$$

$$\underline{\boldsymbol{d}}_2^{(k+1)} = \underline{\boldsymbol{d}}_2^{(k)} + \frac{\beta_2}{2}\left(\boldsymbol{w}^{(k+1)} - |\underline{\boldsymbol{f}}^{(k+1)}|\right) \tag{8.10}$$

Minimizing the augmented Lagrangian jointly with respect to $\underline{\boldsymbol{u}}, \underline{\boldsymbol{f}}$, and $\boldsymbol{w}$ may be hard to achieve. The ADMM solves the last problem by taking advantages of the dual decomposition and optimizes sequentially the primal variables. For the given problem,

the algorithm is described by iterating the following steps:

$$\underline{\boldsymbol{u}}^{(k+1)} = \underset{\underline{\boldsymbol{u}}}{\operatorname{argmin}} \frac{1}{2}||\underline{\boldsymbol{\Phi}}\,\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}^{(k)}||_2^2 + \frac{\beta_1}{2}||\underline{\boldsymbol{f}}^{(k)} - \underline{\boldsymbol{u}} + \underline{\boldsymbol{d}}_1^{(k)}||_2^2 \tag{8.11}$$

$$\underline{\boldsymbol{d}}_1^{(k+1)} = \underline{\boldsymbol{d}}_1^{(k)} + \frac{\beta_1}{2}\left(\underline{\boldsymbol{f}}^{(k)} - \underline{\boldsymbol{u}}^{(k+1)}\right) \tag{8.12}$$

$$\underline{\boldsymbol{f}}^{(k+1)} = \underset{\underline{\boldsymbol{f}}}{\operatorname{argmin}} \frac{\beta_1}{2}||\underline{\boldsymbol{f}} - \underline{\boldsymbol{u}}^{(k+1)} + \underline{\boldsymbol{d}}_1^{(k+1)}||_2^2 + \frac{\beta_2}{2}||\boldsymbol{w}^{(k)} - |\underline{\boldsymbol{f}}| + \boldsymbol{d}_2^{(k)}||_2^2 \tag{8.13}$$

$$\underline{\boldsymbol{d}}_2^{(k+1)} = \underline{\boldsymbol{d}}_2^{(k)} + \frac{\beta_2}{2}\left(\boldsymbol{w}^{(k)} - |\underline{\boldsymbol{f}}^{(k+1)}|\right) \tag{8.14}$$

$$\boldsymbol{w}^{(k+1)} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \frac{\beta_2}{2}||\boldsymbol{w} - |\underline{\boldsymbol{f}}^{(k+1)}| + \boldsymbol{d}_2^{(k+1)}||_2^2 + \mathcal{R}(\boldsymbol{w}) \tag{8.15}$$

$$\tag{8.16}$$

Different convergence proofs of ADMM exist in the litterature. From (Boyd *et al.*, 2011), the convergence of LMM and ADMM is guaranteed under the following two hypothesis:

- The objective functions are closed, proper and convex
- The unaugmented Lagrangian $\mathcal{L}_0$ has a saddle point:

$$\begin{aligned}\mathcal{L}_0(\underline{\boldsymbol{u}}, \underline{\boldsymbol{f}}, \boldsymbol{w}, \underline{\boldsymbol{\gamma}}_1, \boldsymbol{\gamma}_2) =& \frac{1}{2}||\underline{\boldsymbol{\Phi}}\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}||_2^2 + \mathcal{R}(\boldsymbol{w}) \\ &+ \underline{\boldsymbol{\gamma}}_1^H\left(\underline{\boldsymbol{u}} - \underline{\boldsymbol{f}}\right) + \boldsymbol{\gamma}_2^H\left(|\underline{\boldsymbol{f}}| - \boldsymbol{w}\right)\end{aligned} \tag{8.17}$$

  with $\underline{\boldsymbol{\gamma}}_1$ and $\boldsymbol{\gamma}_2$ the unscale dual variable.

These conditions are very general and allow to use these algorithms for many applications. An other strength of these approaches is that they converge even if the minimization steps are inexact. In our case, the convexity assumption is not satisfied so LMM and ADMM can only insure local convergence and the obtained solution may depend on the initialization.

To solve this constrained optimization problem, we apply a variation of the precedent methods described in (Mourya *et al.*, 2015) and minimize jointly on the variables $\underline{\boldsymbol{u}}$ and $\boldsymbol{w}$ while $\underline{\boldsymbol{f}}$ is substituted with its optimal value $\underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w})$:

$$\underline{\boldsymbol{f}}^* = \arg\min_{\underline{\boldsymbol{f}}} \frac{\beta_1}{2}||\underline{\boldsymbol{f}} - \underline{\boldsymbol{u}} + \underline{\boldsymbol{d}}_1||_2^2 + \frac{\beta_2}{2}\left\|\boldsymbol{w} - |\underline{\boldsymbol{f}}| + \boldsymbol{d}_2\right\|_2^2 \tag{8.18}$$

$$= \left[\frac{\beta_1 \cdot |\underline{\boldsymbol{u}} - \underline{\boldsymbol{d}}_1| + \beta_2 \cdot (\boldsymbol{w} + \boldsymbol{d}_2)}{\beta_1 + \beta_2}\right]^+ \exp\left[\underline{\mathrm{j}} \cdot \arg\left(\underline{\boldsymbol{u}} - \underline{\boldsymbol{d}}_1\right)\right] \tag{8.19}$$

where $.^+$ is the identity on $[0, +\infty[$ and is the constant null function on $]-\infty, 0[$.

Proof: The second term in (8.18) only depends on the modulus of $\underline{\boldsymbol{f}}$. The phase of $\underline{\boldsymbol{f}}^*$ is then driven by the first term and must be chosen equal to that of $\underline{\boldsymbol{u}} - \underline{\boldsymbol{d}}_1$ so as to minimize the cost function. There remains to estimate the modulus of $\underline{\boldsymbol{f}}^*$ which is a

solution of a 1-D quadratic problem

$$\arg\min_{\rho\geq 0} \frac{\beta_1}{2}||\boldsymbol{\rho} - |\underline{\boldsymbol{u}} + \underline{\boldsymbol{d}}_1|||_2^2 + \frac{\beta_2}{2}\left\|\boldsymbol{w} - \boldsymbol{\rho} + \boldsymbol{d}_2\right\|_2^2 \tag{8.20}$$

ρ is either given by the unconstrained solution:

$$\rho^* = (\beta_1 \cdot |\underline{\boldsymbol{u}} - \underline{\boldsymbol{d}}_1| + \beta_2 \cdot (\boldsymbol{w} + \boldsymbol{d}_2))/(\beta_1 + \beta_2),$$

or $\rho = 0$. The optimal value of ρ and of the phase together lead to the expression (8.19). □

This leads to the following algorithm which alternates between the joint minimization with respect to variables $\underline{\boldsymbol{u}}$ and $\boldsymbol{w}$, and the update of dual variables.

Algorithm Tomographic SAR 3-D Inversion

Input: $\underline{\boldsymbol{v}}$ (stack of SLC SAR images)
Output: $\underline{\hat{\boldsymbol{u}}}$ (3-D cube of complex reflectivities)

Initialization :
1: $\underline{\hat{\boldsymbol{d}}}_1 \leftarrow \mathbf{0}$
2: $\hat{\boldsymbol{d}}_2 \leftarrow \mathbf{0}$
3: **while** not converged **do**
4: $\{\underline{\hat{\boldsymbol{u}}}, \hat{\boldsymbol{w}}\} \leftarrow$ `approximate_min`$(\underline{\boldsymbol{v}}, \underline{\hat{\boldsymbol{d}}}_1, \hat{\boldsymbol{d}}_2, \underline{\hat{\boldsymbol{u}}}, \hat{\boldsymbol{w}})$
5: $\hat{\boldsymbol{d}}_2 \leftarrow \hat{\boldsymbol{d}}_2 + \hat{\boldsymbol{w}} - |\underline{\boldsymbol{f}}^*(\underline{\hat{\boldsymbol{u}}}, \hat{\boldsymbol{w}})|$
6: $\underline{\hat{\boldsymbol{d}}}_1 \leftarrow \underline{\hat{\boldsymbol{d}}}_1 + \underline{\boldsymbol{f}}^*(\underline{\hat{\boldsymbol{u}}}, \hat{\boldsymbol{w}}) - \underline{\hat{\boldsymbol{u}}}$
7: **end while**
8: **return** $\underline{\hat{\boldsymbol{u}}}$

Compared to ADMM , this hierarchical approach has been shown in (Mourya *et al.*, 2015) to converge faster and to be less sensitive to the tuning of the optimization parameters β_1 and β_2; moreover, we improve the convergence by constraining $\boldsymbol{w}$ to be positive, such a constraint would be costly to enforce with ADMM .

After replacing $\underline{\boldsymbol{f}}$ by its optimal value $\underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w})$ and constraining $\boldsymbol{w}$ to be positive, the objective function is differentiable (since $\boldsymbol{w} \geq \mathbf{0}$, $\|\boldsymbol{w}\|_1 = \mathbf{1}^T\boldsymbol{w}$, which is differentiable; moreover, after substituting $\boldsymbol{f}$ by the optimal value $\boldsymbol{f}^*$ in the quadratic terms, the cost function becomes smooth, see (Mourya *et al.*, 2015)). The cost function can thus be minimized using a limited memory quasi-Newton algorithm that handles positivity constraints, such as L-BFGS-B (Zhu *et al.*, 1997), or the slightly more efficient algorithm VMLM-B (Thiébaut, 2002) that we used, with Eric Thiébaut's freely available implementation[1]. The minimization step described in Procedure `approximate_min` does not need to be performed up to a high precision. A few (e.g., ten) iterations of the quasi-Newton algorithm are sufficient since the algorithm is warm-restarted. In the definition of the gradient of the cost function (lines 2 to 4 of Procedure `approximate_min`), we used the fact that $\frac{\partial \mathscr{C}}{\partial [\boldsymbol{f}^*]_i}\frac{\partial [\boldsymbol{f}^*]_i}{\partial [\boldsymbol{u}]_j} = \frac{\partial \mathscr{C}}{\partial [\boldsymbol{f}^*]_i}\frac{\partial [\boldsymbol{f}^*]_i}{\partial [\boldsymbol{w}]_k} = 0$ for all i, j, k, even though $\boldsymbol{f}^*$ depends on variables $\underline{\boldsymbol{u}}$ and $\boldsymbol{w}$, by applying the chain rule and noting that $\frac{\partial \mathscr{C}}{\partial [\boldsymbol{f}^*]_i} = 0$ for all i.

[1] https://github.com/emmt/OptimPackLegacy

Procedure `approximate_min`

Input: $\underline{\boldsymbol{v}}$ (stack of SLC SAR images)
Input: $\hat{\underline{\boldsymbol{d}}}_1$ (vector of dual variables)
Input: $\hat{\boldsymbol{d}}_2$ (vector of dual variables)
Input: $\hat{\underline{\boldsymbol{u}}}$ (current vector of primal variables)
Input: $\hat{\boldsymbol{w}}$ (current vector of primal variables)
Output: $\{\hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}\}$ (approximate solution)

1: $\mathscr{C}(\underline{\boldsymbol{v}}, \hat{\underline{\boldsymbol{d}}}_1, \hat{\boldsymbol{d}}_2, \hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}) = \frac{1}{2}||\underline{\boldsymbol{\Phi u}} - \underline{\boldsymbol{v}}||_2^2 + \mathcal{R}(\boldsymbol{w}) \quad + \frac{\beta_1}{2}||\underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w}) - \underline{\boldsymbol{u}} + \hat{\underline{\boldsymbol{d}}}_1||_2^2 + \frac{\beta_2}{2}\big\|\boldsymbol{w} - |\underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w})| + \hat{\boldsymbol{d}}_2\big\|_2^2$,

2: $\nabla_{\underline{\boldsymbol{u}}}\mathscr{C} = \underline{\boldsymbol{\Phi}}^H(\underline{\boldsymbol{\Phi u}} - \underline{\boldsymbol{v}}) + \beta_1(\underline{\boldsymbol{u}} - \underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w}) - \hat{\underline{\boldsymbol{d}}}_1)$

3: $\nabla_{\boldsymbol{w}}\mathscr{C} = (\mu_x \boldsymbol{D}_x^T\boldsymbol{D}_x + \mu_y \boldsymbol{D}_y^T\boldsymbol{D}_y + \mu_z \boldsymbol{D}_z^T\boldsymbol{D}_z)\boldsymbol{w} + \mu_{\ell_1}\mathbf{1} + \beta_2(\boldsymbol{w} - \underline{\boldsymbol{f}}^*(\underline{\boldsymbol{u}}, \boldsymbol{w}) + \hat{\underline{\boldsymbol{d}}}_2)$.

4: $\nabla\mathscr{C}(\underline{\boldsymbol{v}}, \hat{\underline{\boldsymbol{d}}}_1, \hat{\boldsymbol{d}}_2, \hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}) = \begin{pmatrix} \nabla_{\underline{\boldsymbol{u}}}\mathscr{C} \\ \nabla_{\boldsymbol{w}}\mathscr{C} \end{pmatrix}$,

Call quasi-Newton minimization algorithm :
5: $\{\hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}\} \leftarrow$ `quasi_Newton`$\big[\mathscr{C}(\underline{\boldsymbol{v}}, \hat{\underline{\boldsymbol{d}}}_1, \hat{\boldsymbol{d}}_2, \hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}), \nabla\mathscr{C}(\underline{\boldsymbol{v}}, \hat{\underline{\boldsymbol{d}}}_1, \hat{\boldsymbol{d}}_2, \hat{\underline{\boldsymbol{u}}}, \hat{\boldsymbol{w}}),$ `constraint:` $\hat{\boldsymbol{w}} \geq 0\big]$

The necessary condition for $\boldsymbol{f}^*$ being a minimizer of the augmented Lagrangian is then respected. As the function we want to minimize is non-convex, only the convergence to a local minimum can be expected. In practice the algorithm can then be stopped after a fixed number of iterations (around 60) or when the distance between two iterations falls below a given threshold.

The priors proposed in this section are designed to be simple and to lead to a function $\mathcal{R}$ that is easy to optimize. In the following experiments, the relevance of this method is shown by comparing to other state of the art estimators on simulated and real data. The proposed framework is very general and can easily be adapted to include other spatial regularizations expressed in ground coordinates.

8.3 Influence of the different parameters

To study the influence of each regularization parameter, we performed different reconstructions of the `Reg` scene for various combinations of regularization values (in particular, with one parameter chosen so as to illustrate the effect of either under-regularization or over-regularization). In these expreriments, the phase noise is not present to show clearly the structural influence of the spatial smoothing. The results are presented in figures 8.4, 8.5, 8.6 and 8.7. On this well-sampled scene with a good spatial resolution, the parameter with the largest influence is μ_{ℓ_1}, associated to the sparsity constraint. Fig. 8.4 illustrates that when μ_{ℓ_1} is too large, there are some holes in the reconstruc-

tion, while a value of μ_{ℓ_1} that is too small leaves side-lobes and outliers. The effect of over-smoothing (values of μ_x, μ_y or μ_z too large) is to expand the structures in the direction of the spatial smoothing. This is visible in particular with structures whose orientation differs from that of the smoothing, see in Fig. 8.6 the widening of the wall due to excessive smoothing in the horizontal direction. The same effect appears on the rooftop in Fig. 8.7. Insufficient smoothing translates into residual fluctuations (i.e., large variance) that are reduced by increasing spatial smoothing, see in particular the corner between the roof and the wall, or the ground and the wall in Fig. 8.6 and 8.7. Outliers located far from the actual surfaces can also be observed when the spatial regularization is too weak.

By combining sparsity and spatial smoothness constraints, our algorithm is very flexible and applicable to the reconstruction of different kinds of areas. The downside of this flexibility is the necessity to tune four regularization parameters. In numerical simulations, the ground-truth can be used to select the set of regularization parameters $\{\mu_x, \mu_y, \mu_z, \mu_{\ell_1}\}$ that offers the best performance, as measured by the Minimum Accuracy / Completeness Trade-off (MACT) (MACT) $\min A(\hat{\mathcal{P}}_{\underline{\hat{\mathbf{u}}},t^*}, \mathcal{P})^2 + C(\hat{\mathcal{P}}_{\underline{\hat{\mathbf{u}}},t^*}, \mathcal{P})^2$. The regularization parameters can then be tuned in order to reach the best possible trade-off. We recommend performing this tuning by order of importance: μ_{ℓ_1}, then μ_z, then μ_x and μ_y. While tuning the sparsity parameter, the other smoothing parameters should be put to 0.

Derivative-free methods can also be used to set all the parameters at once. We compared the described alternating minimization strategy with Nelder-Mead's simplex method by initializing with different set of parameters and found similar values for the best accuracy/completeness trade-off. The differences for μ_{ℓ_1} is of 21% and below 14% for the smoothing parameters. In the absence of ground truth, a simple numerical simulation using the same geometrical configuration and SNR can be generated in order to automatically tune the parameters. If a simulated scene is not an option, the alternating minimization strategy can still be applied in an interactive fashion: the user tunes each parameter, by order of importance, in order to reach a satisfying reconstruction.

To further illustrate the behavior of our algorithm with respect to its regularization parameters, we plotted accuracy as a function of completeness for different sets of parameters. These curves are drawn for our two simulation cases `Reg` and `TSX` respectively in Fig. 8.8 and 8.9. We observe that the best accuracy / completeness trade-off (point of the curve closest to the origin of the axes) is reached for a unique set of parameters that can thus be found for example by binary search.

The optimization parameters β_1 and β_2 have an impact on the convergence speed. We found that, when starting from a volume initialized at zero, using large penalty parameters β_1 and β_2 produces very quickly a sparse reconstruction while lower contrasted structures are correctly reconstructed after many more iterations. Smaller values of the parameters help to reconstruct those structures, at the cost of a slower convergence (i.e., sidelobes suppression) in the brightest areas. Penalty parameters β_1 and β_2 can be set according to methods described in (Boyd *et al.*, 2011;

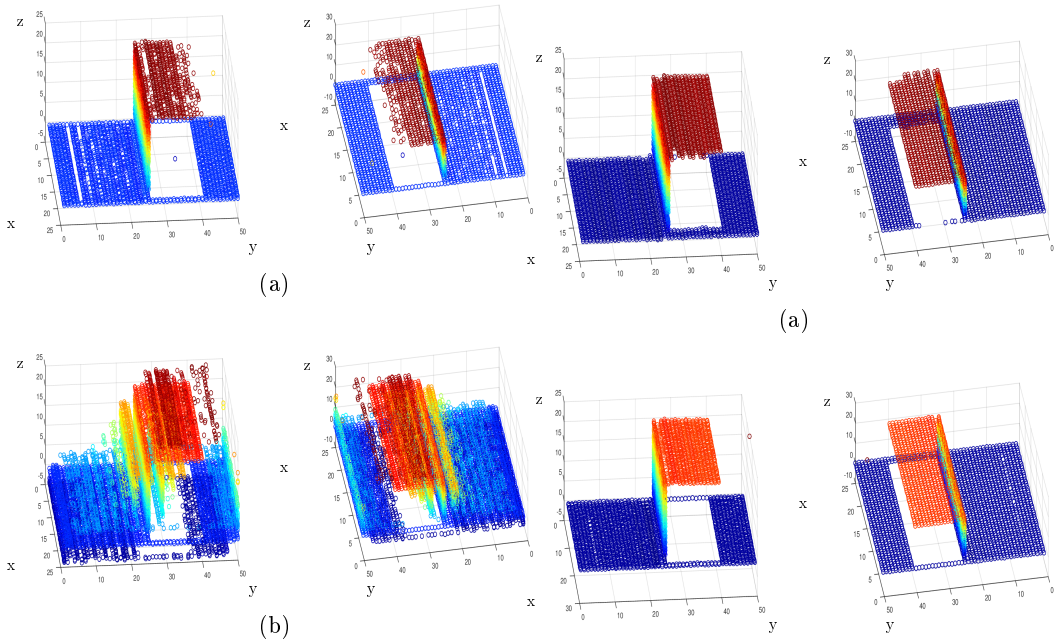


Figure 8.4 – Influence of the parsimony μ_{ℓ_1} parameter. In (a) the parameter is way above its optimal value resulting in a lot of holes in the structure. In (b) we use a small value of μ_{ℓ_1} resulting in a high number of outliers.

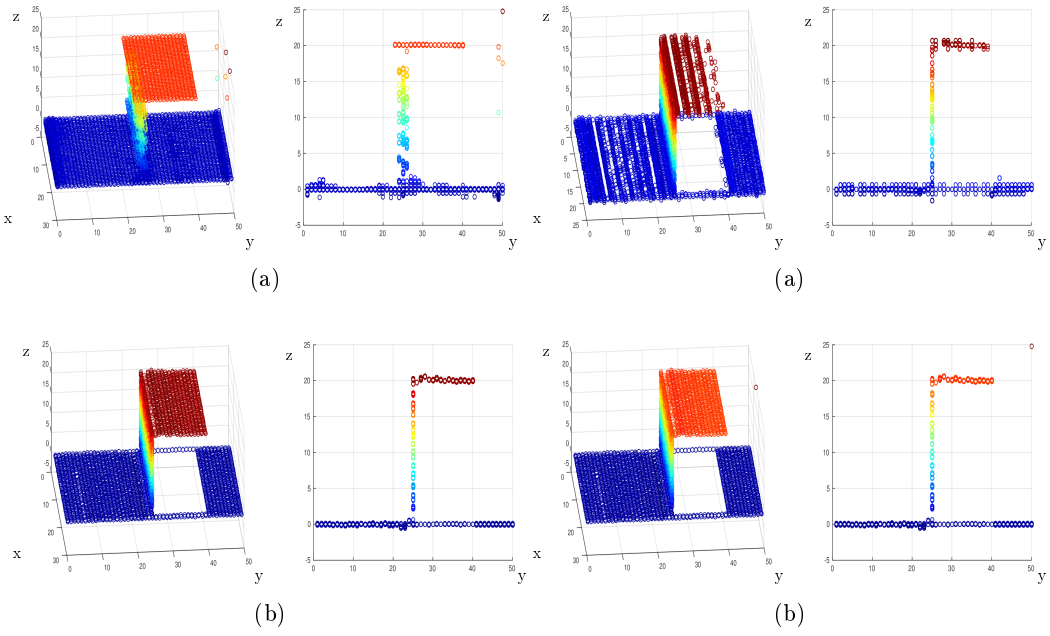
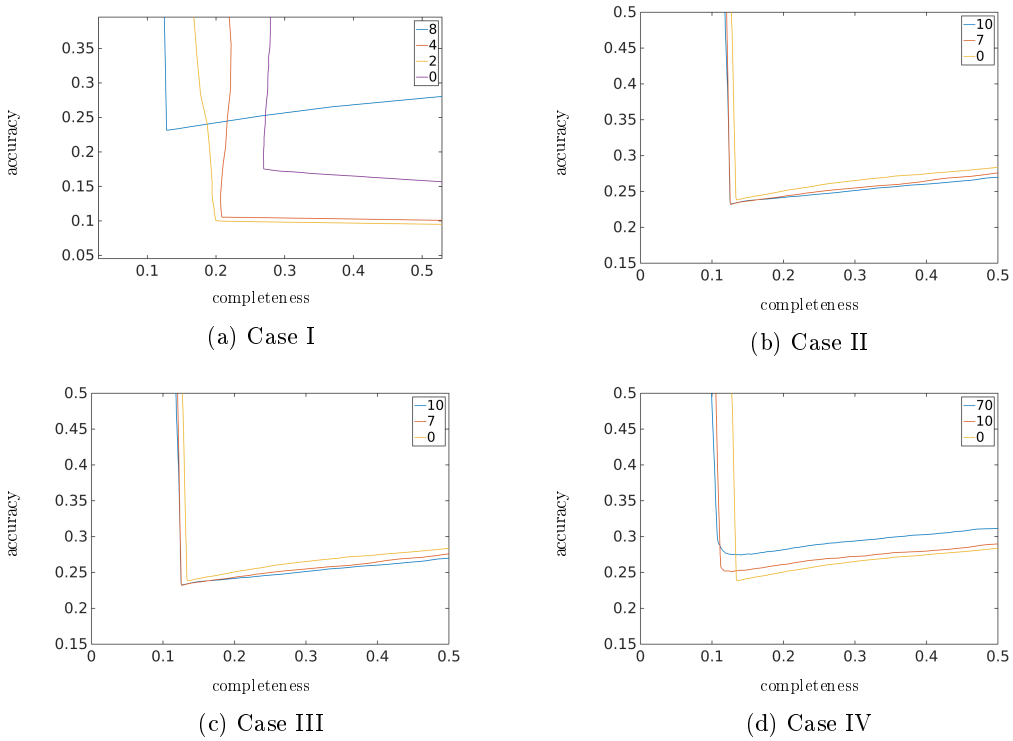


Figure 8.5 – Influence of the μ_x parameter. In (a) the parameter is way above its optimal value. In (b) the value of the parameter is set to zero.

Figure 8.6 – Influence of the μ_y parameter. In (a) the parameter is way above its optimal value. In (b) the value of the parameter is set to zero.

Figure 8.7 – Influence of the μ_z parameter. In (a) the parameter is way above its optimal value. In (b) the value of the parameter is set to zero.



(a) Case I (b) Case II

(c) Case III (d) Case IV

Figure 8.8 – Study of the influence of each parameter on the metrics for the `Reg` scene (cf. Fig. 6.1 (a)). The plots in (a), (b), (c) and (d) correspond respectively to different values of μ_{ℓ_1}, μ_x, μ_y and μ_z with the three other parameters being fixed.

He *et al.*, 2000). Here, we fixed those parameters to 10 after having tested different values.

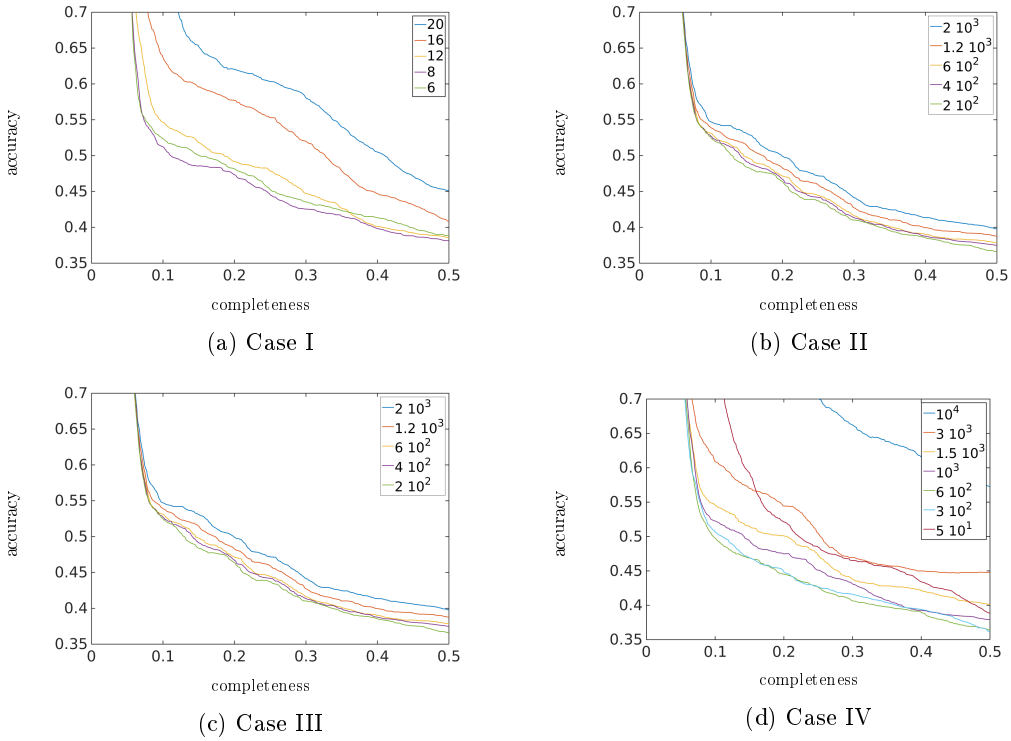
8.4 Comparison with other tomographic estimators

8.4.1 Simulated Data

The figures 8.10, 8.11, 8.12 and 8.13 give the 3-D reconstructions obtained using several state-of-the-art tomographic reconstruction methods and our spatially regularized approach: (a) classical beamforming, (b) Capon beamforming, (c) the parametric estimators MUSIC, (d) Compressed Sensing, and (e) our algorithm. The corresponding curves of accuracy as a function of completeness are presented in figure 8.14 for the two scenes.

In the easy case where the antennas are uniformly distributed, classical beamforming presents good performances with respect to the accuracy-completeness metrics, achieving a smaller error than Capon beamforming or MUSIC. However this method is limited by its available resolution introducing thickness in the wall. In the `TSX` case where the theoretical resolution is worse, very little of the wall is the retrieved, leading to a poor completeness score. A closer inspection of the `TSX` reconstructions indicates the presence of side-lobes (i.e., many outliers).

Both Capon beamforming and MUSIC, which rely on the covariance matrix to



(a) Case I

(b) Case II

(c) Case III

(d) Case IV

Figure 8.9 – Study of the influence of each parameter on the metrics for the `TSX` scene (cf. Fig. 6.1 (b)). The plots in (a), (b), (c) and (d) correspond respectively to different values of μ_{ℓ_1}, μ_x, μ_y and μ_z with the three other parameters being fixed.

achieve super resolution, manage to suppress most of the outliers. However, the averaging introduced for the covariance matrix estimation produces an extension of the ground and walls. For the `Reg` case, this explains the bounded accuracy scores even when the completeness is poor. Capon beamforming manages to reduce the side lobes in the `Reg` case but not in the `TSX` simulation where the irregular sampling produces dramatic biasing effects on the estimation of the position of the scatterers.

In the `TSX` scene, the best accuracy / completeness trade-off from the MUSIC pseudo-spectra corresponds to a sparse scene with no outliers. However the bias present in the estimation of the covariance matrix strongly deforms the wall and introduces some fake discontinuities.

All the spectral estimators presented (classical beamforming, Capon beamforming and MUSIC) suffer from the averaging step needed to estimate the covariance matrix, which produces noticeably distorted walls and/or rooftop in the `TSX` reconstruction, and an expansion of the rooftop and a corresponding reduction of the ground in the `Reg` case.

As expected, CS is one of the top estimators in accuracy and often manages to achieve the most accurate representations. However, the precision comes with the cost of a loss of information which explains the relatively bad scores in completeness.

The proposed 3-D inversion algorithm shows improved performances compared to the CS approach. Some outliers that were present even when imposing an ℓ_1 penalty

are suppressed thanks to the spatial smoothness prior.

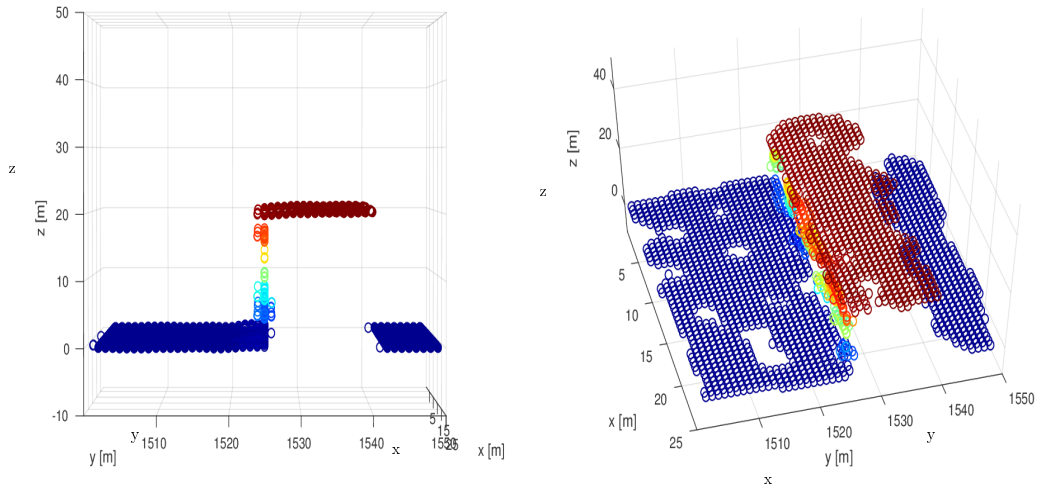
The tuning of the parameter in the simulated experiences has been done by exploiting the knowledge of the scene. Of course, when dealing with real data, this information is not available. Then, to test if the proposed method is able to exhibit better performances when the tuning of the parameter is not ideal, the model parameter are modified with increasing perturbation values. The results are shown in the table 8.1. Up to an isotropic perturbation of 50, the proposed method still shows a better Minimum Accuracy-Completeness Tradeoff (MACT) than the other tested methods. An other concern is whether the parameters need to be adapted to the scatterers power. Real-data experiments conducted over a very heterogeneous scene indicate that the reconstruction doesn't present artifacts due to excessive smoothing in some areas. To confirm this observation with a controlled experiment, reconstruction scores of the `TSX` scene with random scatterers amplitude are also tested. The amplitudes were chosen randomly for each azimuth in a given range. The different experiments correspond increasing ranges. The results of the experiments are shown in table 8.2. The proposed method still performs better, even when the amplitudes of the scatterers varies by 3 orders of magnitude and the regularization parameters are kept constant.

Estimator	**MACT**
Beamforming	0.96
Capon Beamforming	0.98
MUSIC	0.66
CS	0.71
3-D inversion ($\boldsymbol{\mu}^*$)	**0.57**
3-D inversion ($\boldsymbol{\mu}^* - 50\%$)	0.62
3-D inversion ($\boldsymbol{\mu}^* - 100\%$)	0.62
3-D inversion ($\boldsymbol{\mu}^* + 50\%$)	0.59
3-D inversion ($\boldsymbol{\mu}^* + 100\%$)	0.65
3-D inversion ($\boldsymbol{\mu}^* + 200\%$)	0.74

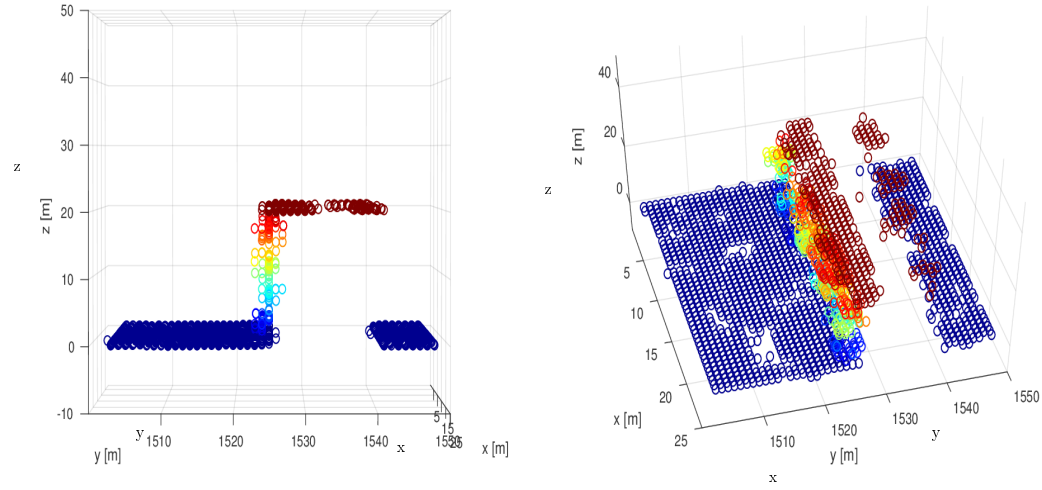
Table 8.1 – Evolution of the Minimum Accuracy-Completeness Tradeoff (MACT) when the parameters are getting far from their optimal value $\boldsymbol{\mu}^* = \{\mu^*_{\ell_1}, \mu^*_x, \mu^*_y, \mu^*_z\}$. The smaller the MACT the better.

Estimator/Amplitudes	$[1, 1]$	$[10^{-1}, 10]$	$[10^{-2}, 10^2]$	$[10^{-3}, 10^3]$
Beamforming	0.96	1.09	1.10	1.25
Capon Beamforming	0.98	1.35	1.38	1.58
MUSIC	0.66	0.66	0.68	0.69
CS	0.71	0.71	0.72	0.72
3-D inversion	**0.57**	**0.58**	**0.60**	**0.60**

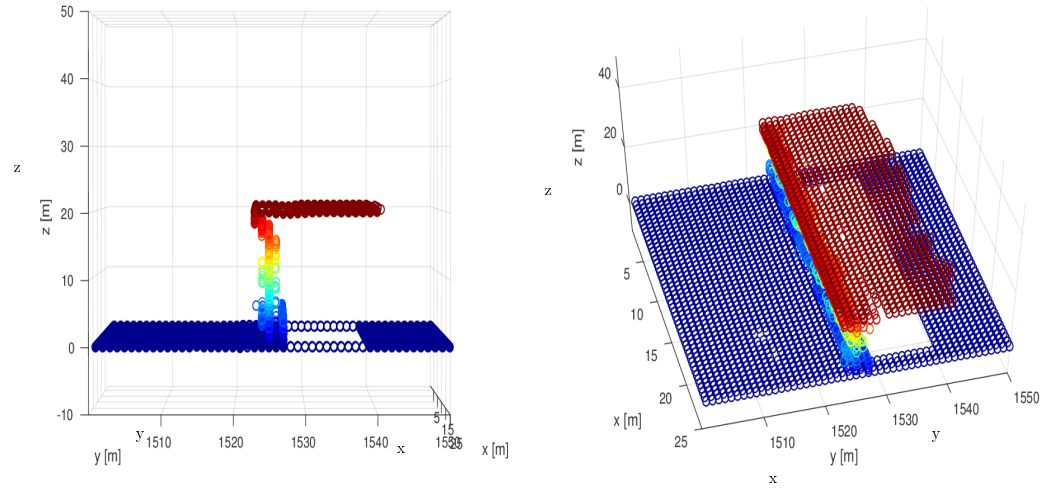
Table 8.2 – Evolution of the MACT value when the scatterer amplitude is fixed uniformly for each azimuth. The experience is repeated for increasing amplitude interval size.



(a)

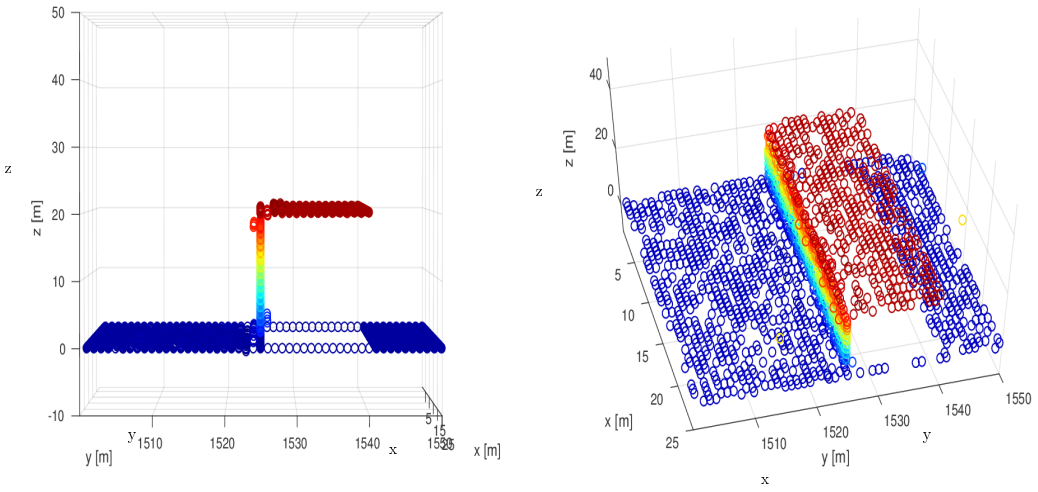


(b)

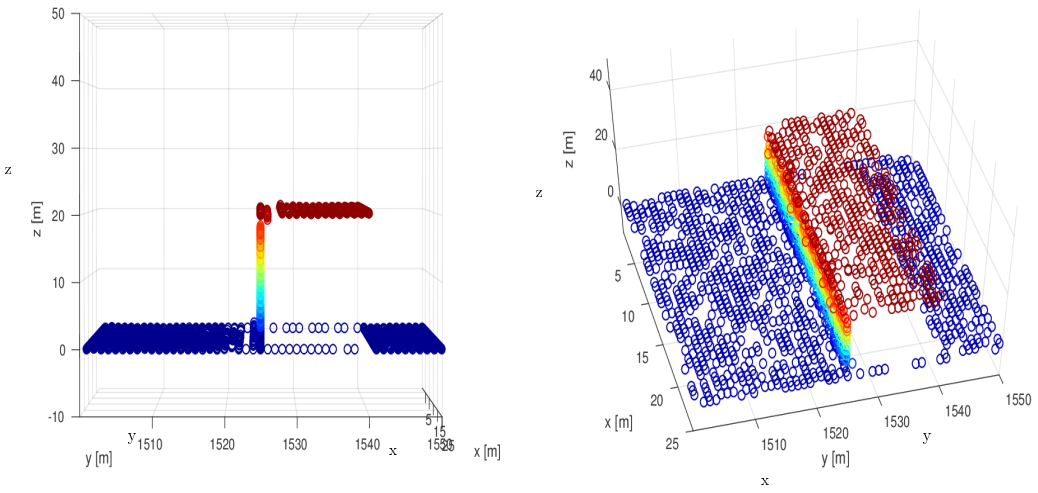


(c)

Figure 8.10 – Best representation on the `Reg` scene using classical beamforming (a), Capon beamforming (b), MUSIC (c), CS (d) and our tomographic 3D inversion (e).

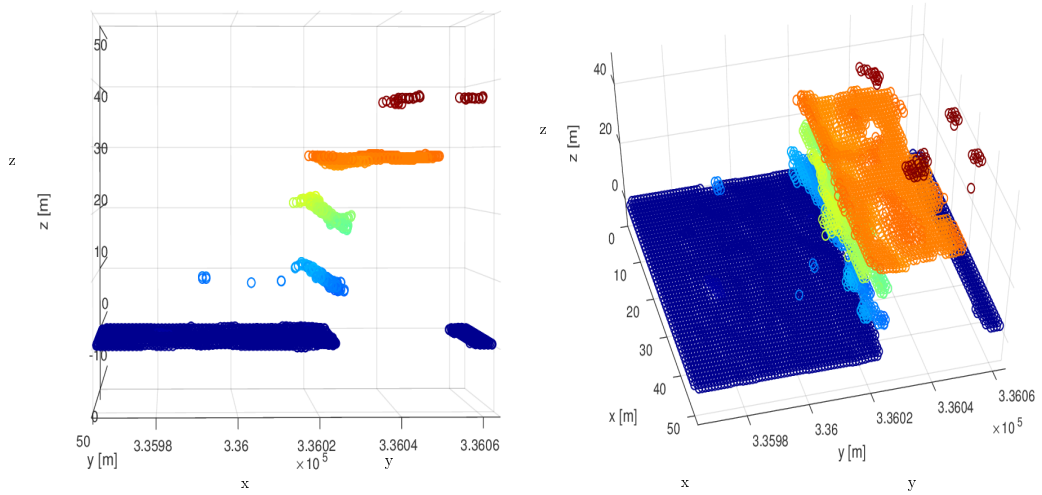


(a)

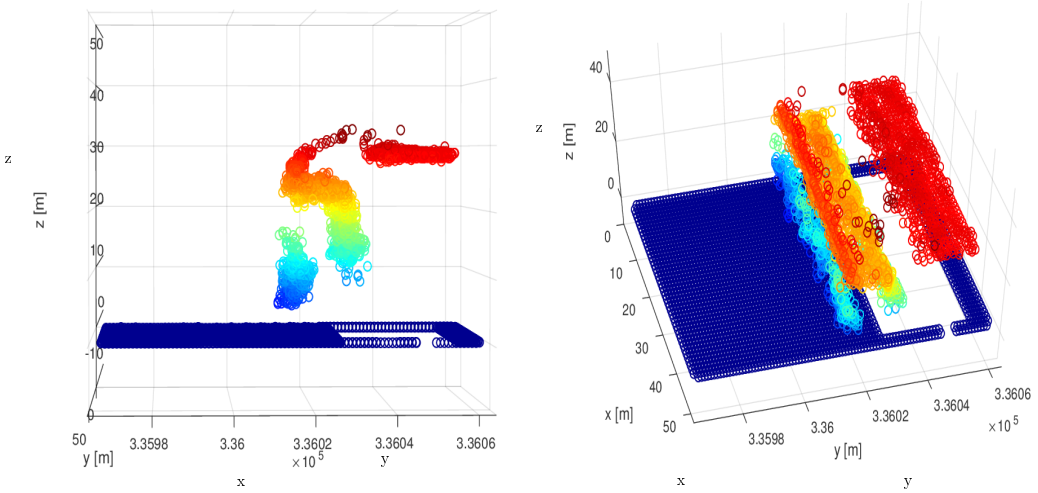


(b)

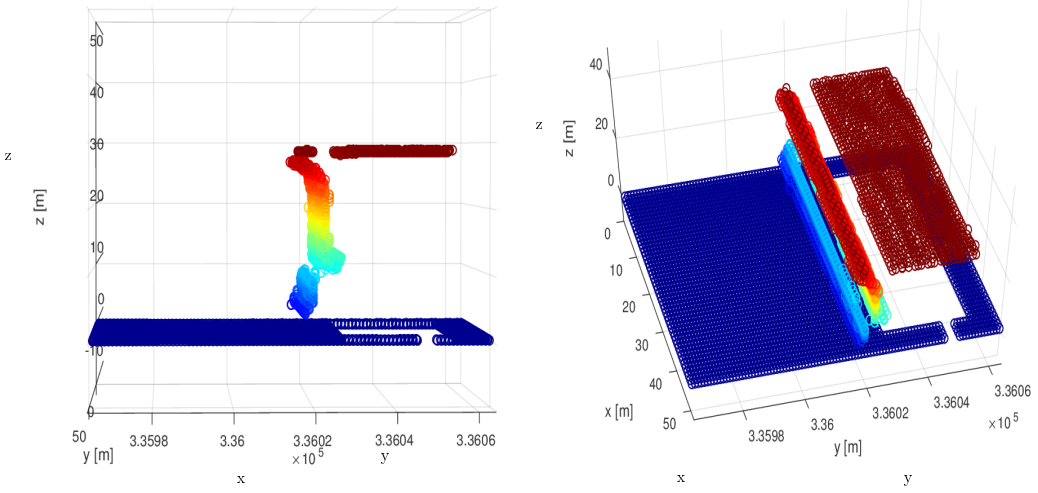
Figure 8.11 – Best representation on the `Reg` scene using classical beamforming (a), Capon beamforming (b), MUSIC (c), CS (d) and our tomographic 3D inversion (e).



(a)

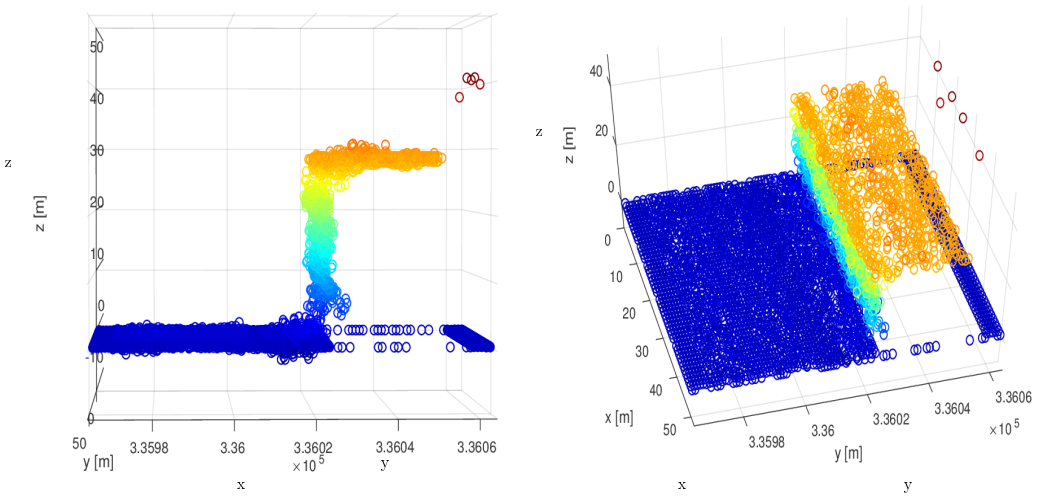


(b)

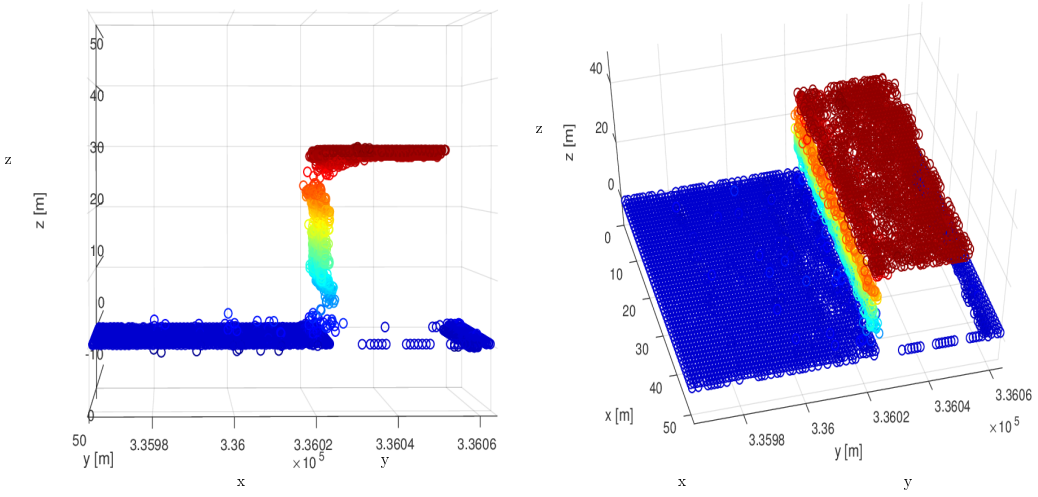


(c)

Figure 8.12 – Best representation on the `Reg` scene using classical beamforming (a), Capon beamforming (b), MUSIC (c), CS (d) and our tomographic 3D inversion (e).

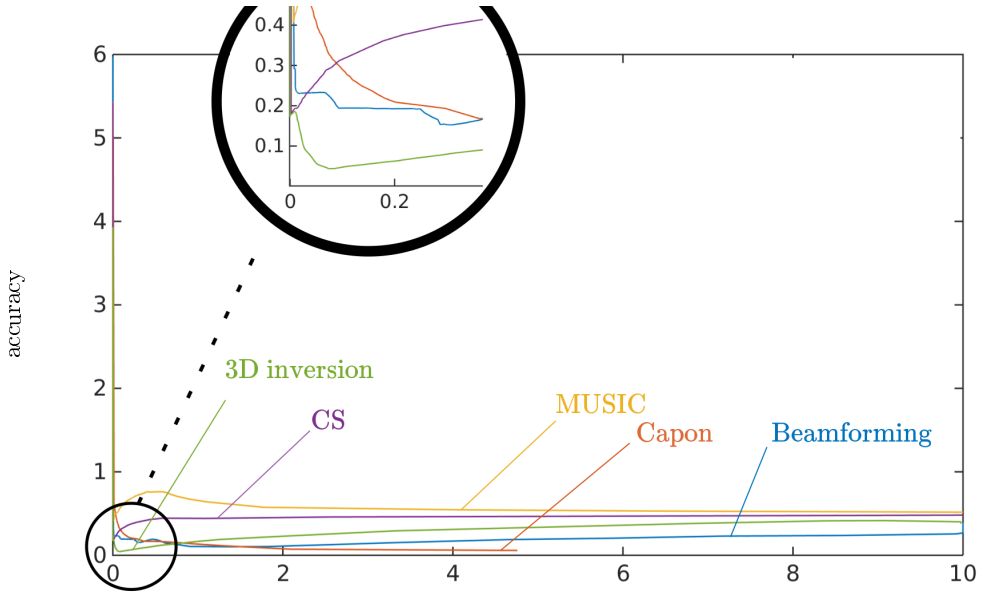


(a)

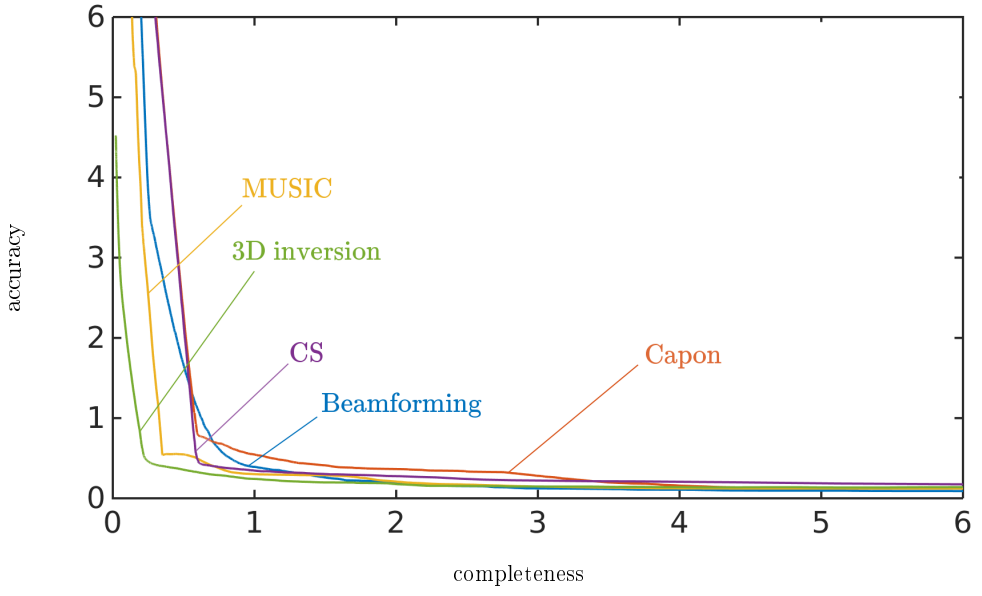


(b)

Figure 8.13 – Best representation on the `TSX` scene using classical beamforming (a), Capon beamforming (b), MUSIC (c), CS (d) and our tomographic 3D inversion (e).



(a)



(b)

Figure 8.14 – Accuracy vs. completeness for classical beamforming, Capon beamforming, MUSIC, Compress Sensing and our 3-D Inversion algorithm. In (a) the evaluation is done on the `Reg` scene and in (b) on the `TSX` scene *cf.* 6.1.

8.4.2 Real Data

We now compare our algorithm to other SAR tomographic estimators on a stack of 40 TerraSAR-X images acquired in spotlight mode over the *front de Seine* in the south-west of Paris, France. The slant-range resolution is 0.45 m and the azimuth resolution 0.87 m. The observed scene is presented in Fig. 6.3 and 6.4 in 6. The spatial and temporal baselines are shown in Fig. 6.5. The total spatial baseline span Δb is more than 775 m and the total temporal baseline more than 5 years with a large gap of almost two years. The theoretical resolutions in h and z are given by :

$$\delta_z = \delta_h \sin(\theta) = \frac{\lambda r \sin(\theta)}{2\Delta b} = 6.99m \tag{8.21}$$

with an incidence angle $\theta = 0.6$ rad and wavelength $\lambda = 0.0311$ m. Sparse reconstructions in urban SAR tomography have been shown to significantly improve this resolution (Zhu and Bamler, 2010b).

To evaluate the different tomographic estimators we use a rough ground truth of the scene. The ground truth surface is shown in Fig. 8.15[2]. The big structures such as the skyscrapers or the tall buildings are well represented by smooth polygons. However, the vegetation and the small structures on the docks, bridges or streets are not represented. We therefore performed the evaluation only on some areas where we found the ground truth to be accurate enough. The results of the evaluation are given in Fig. 8.15.

It is noteworthy that MUSIC performs particularly well on areas containing a lot of ground surface. In the first test, the sub-scene is composed of a mixture of tall buildings, small structures near the ground level and smooth flat areas (streets and docks). As the diffuse signal back-scattered by flat surfaces is much weaker than the one reflected by the dihedral or trihedral structures, it is generally not taken into account by sparse representations of the scene. CS is then unable to correctly represent ground areas. Taking more points into account results only in more outliers generally due to the sidelobes of the brightest points. The averaging step used to estimate the covariance matrix allows MUSIC to retrieve part of the ground points or to extend the signal coming from one punctual target close to the ground to neighboring pixels. Unlike the non-parametric beamforming methods, MUSIC is designed to retrieve a sparse scene which removes most of the side lobes and outliers. MUSIC is then able to outperform CS according to our evaluation method that includes scatterers on the ground. On the second tested area corresponding to the red rectangle in Fig. 8.15, the performances of MUSIC and CS in terms of accuracy are very similar. MUSIC seems to perform a little bit better than CS. However, this may be due to its ability to retrieve more points on the ground. The small difference between the two methods and the limited accuracy of the ground truth makes it difficult to draw a clear conclusion between the two methods.

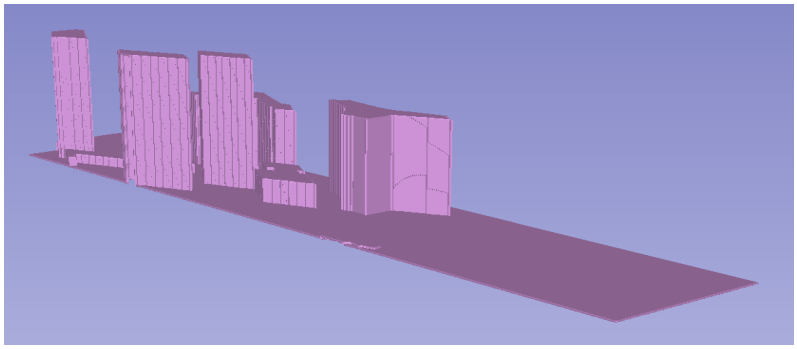
In both experiments, the presented algorithm achieves the best scores in term of accuracy and completeness. Moreover, the analysis of the metrics on real data shows

[2]Ground truth on Paris complimentary provided by the IGN, France.

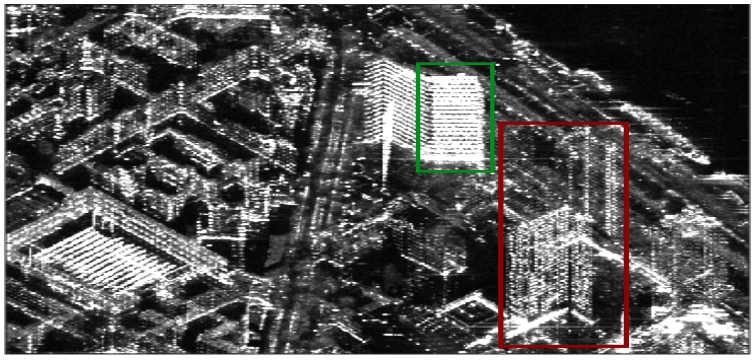
that our method dominates the other approaches for all accuracy/completeness trade-offs.

The 3-D plain representation of the scene obtained by our approach is presented in Fig. 8.18. The heat color is chosen to be proportional to the intensity of the voxels normalized by the averaged intensity of the corresponding radar cell. This normalization is used only to help the visualization as the dynamic range of SAR images is very high. In Fig. 8.19, the same image is shown except that the near zero intensity voxels are made fully transparent and the ground truth is superimposed. The results obtained on the same scene with conventional CS are shown in Fig. 8.16 and Fig. 8.17. The colormap are the same for both representations and are chosen to have the dark blue color associated with intensities below 10^{-4} times the maximum voxel amplitude. This visually sets to zero dark voxels that still have a non-zero intensity as the algorithm may have not fully converged. An other equivalent way to have a fully sparse representation of the modulus of the signal is to look directly at the image $\boldsymbol{w}$ as an output of the algorithm. In the 3-D representation, most of the isolated outliers are suppressed thanks to the spatial smoothing and the buildings where most of the high reflectivity voxels are massed are retrieved. The smoothing denoising however comes with the cost of slightly blurred scatterers. We can see that most of the very bright buildings present in the SAR images in Fig. 6.3 are well reconstructed. The two towers that were mostly missing in the 2D intensity image have a point-like representation but are fairly visible in this 3-D rendering.

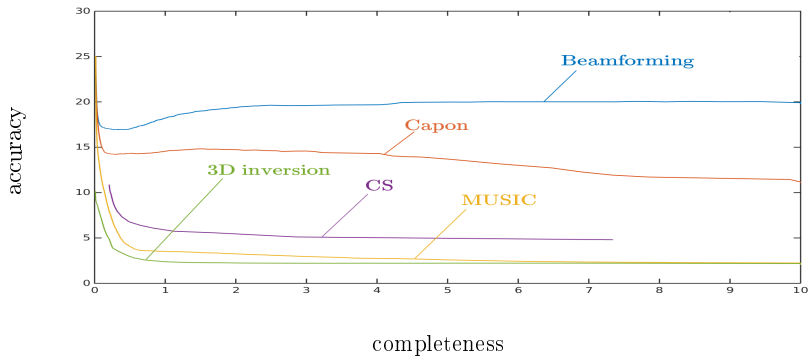
By using the georeferencing of the voxels, we can project the corresponding point cloud into Google Earth ©. The results are presented in Fig. 8.20. This step allows us to have more details in the visualization of the scene. We can now see that some points above the Mirabeau Tower are relevant: they correspond to the structure of its rooftop. Our reconstruction method also correctly identified parts of the structures on the dock and several buildings hidden behind the Mirabeau tower.



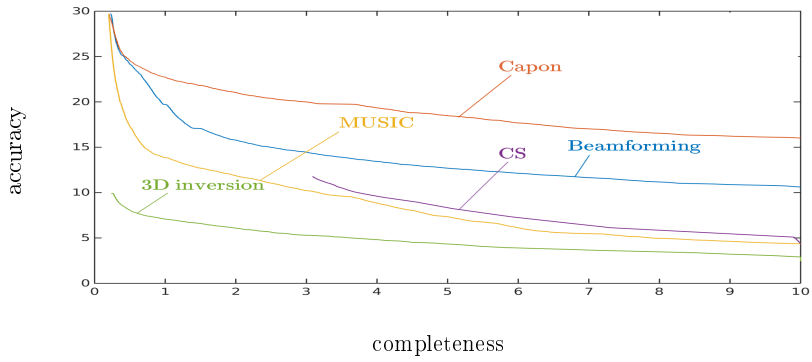
(a)



(b)

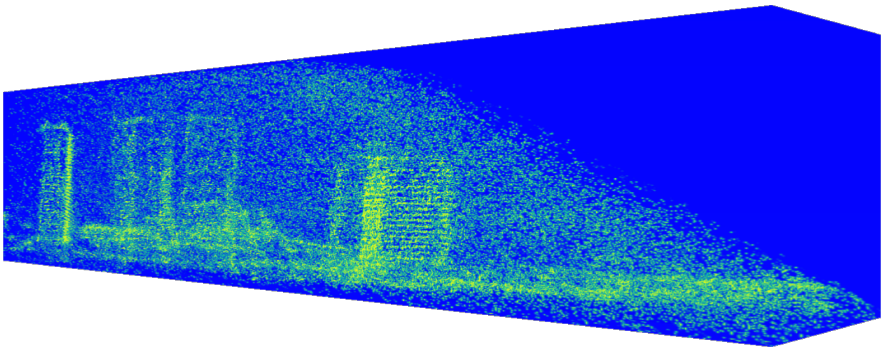


(c)

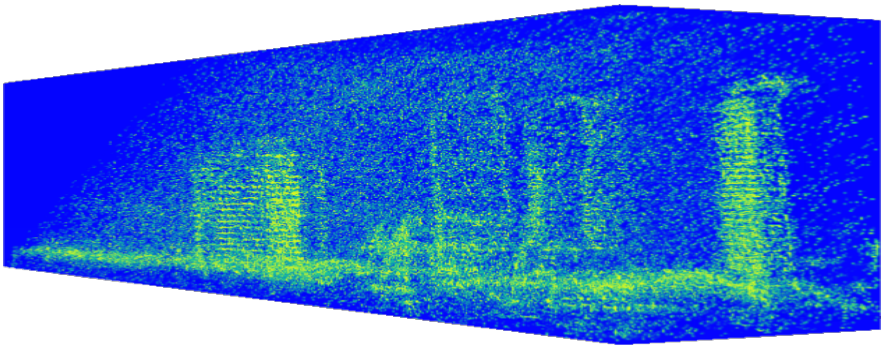


(d)

Figure 8.15 – Accuracy as a function of completeness, computed based on the ground truth displayed in (a), for the following tomographic estimators: classical beamforming, Capon beamforming, MUSIC, CS and the proposed 3D inversion. In (c) the evaluation is done over the area corresponding to the red rectangle in (b). In (d) the test area is inside the green rectangle.

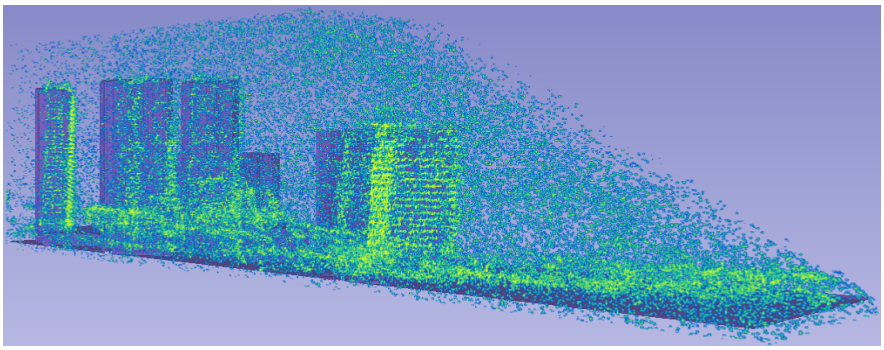


(a)

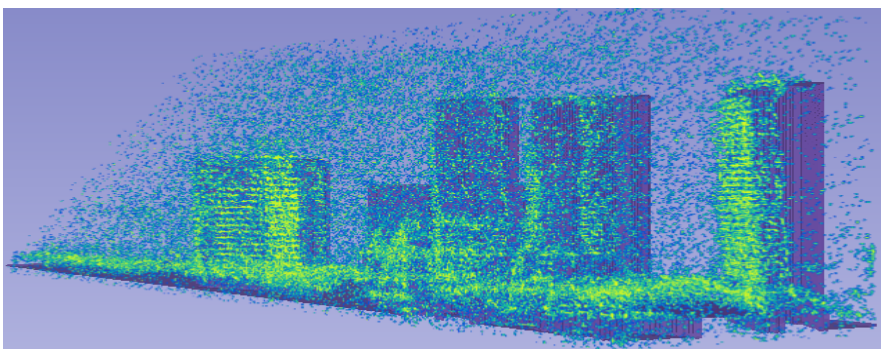


(b)

Figure 8.16 – Visualization of the module of the reconstructed cube (Andriy Fedorov et al., 2012) from a stack of 40 TerraSAR-X images of Paris (cf. Fig.6.3) using the conventional CS algorithm.

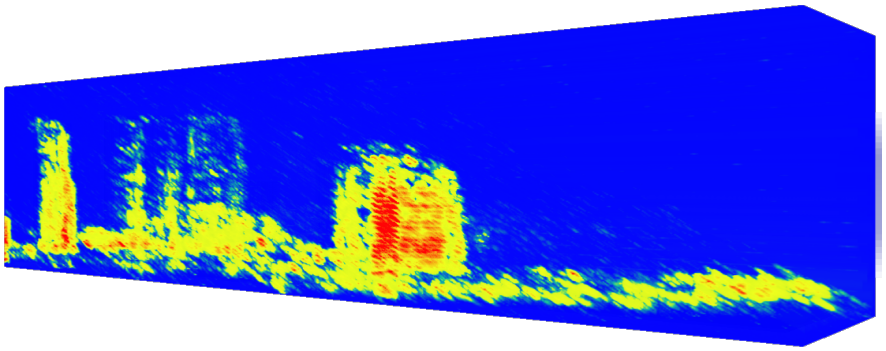


(a)

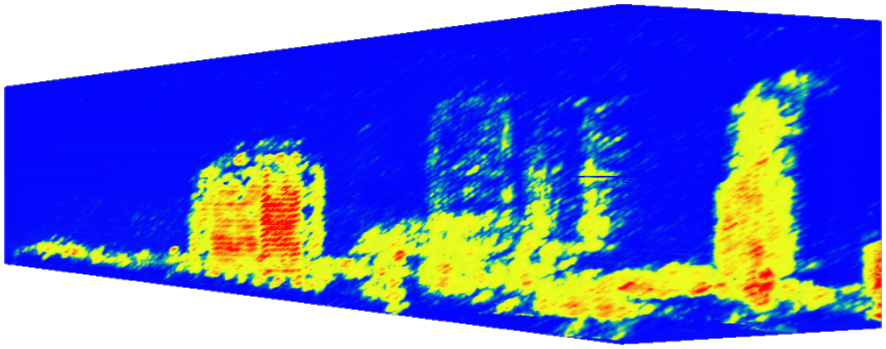


(b)

Figure 8.17 – Visualization of the thresholded module of the reconstructed cube (Andriy Fedorov et al., 2012) using the conventional CS algorithm with the ground truth.

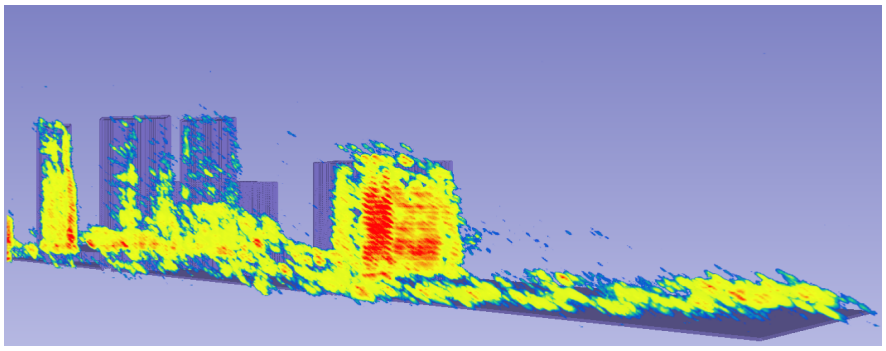


(a)

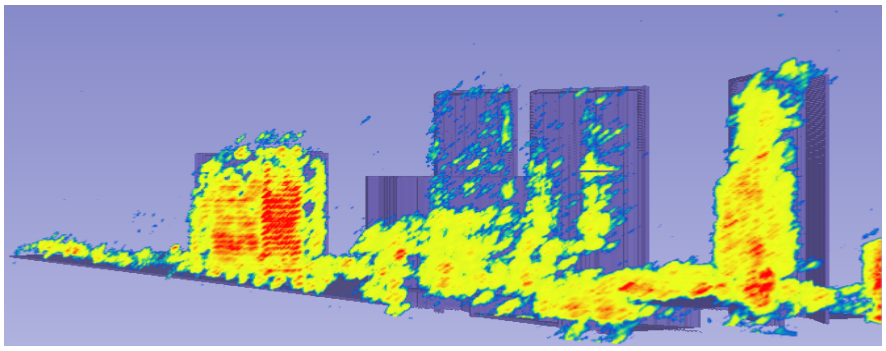


(b)

Figure 8.18 – Visualization of the module of the reconstructed cube (Andriy Fedorov et al., 2012) from a stack of 40 TerraSAR-X images of Paris (cf. Fig.6.3) using the 3D inversion algorithm.

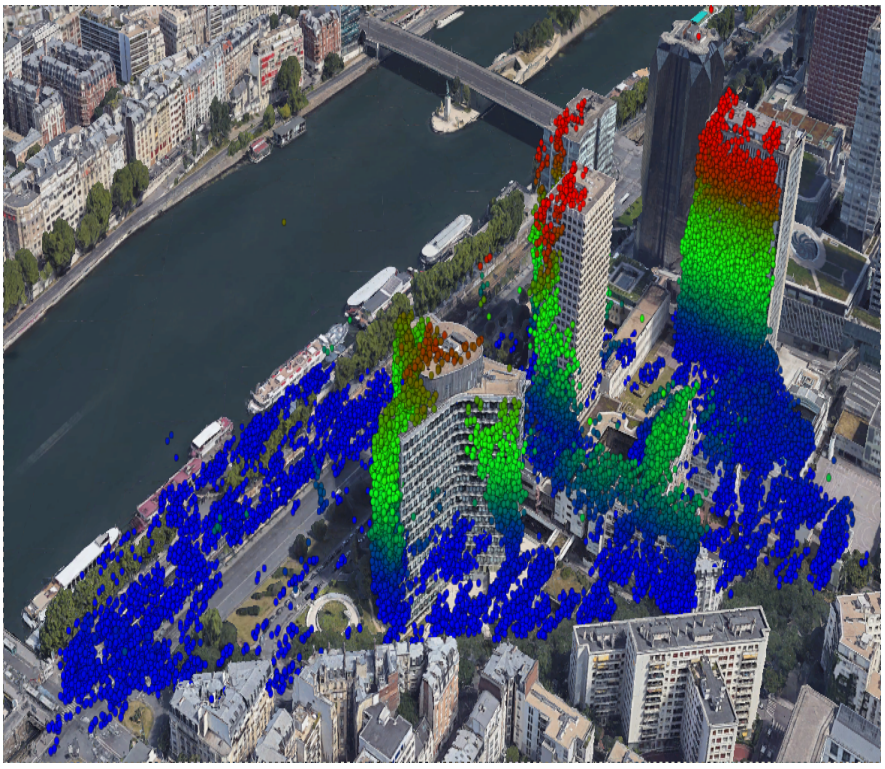


(a)

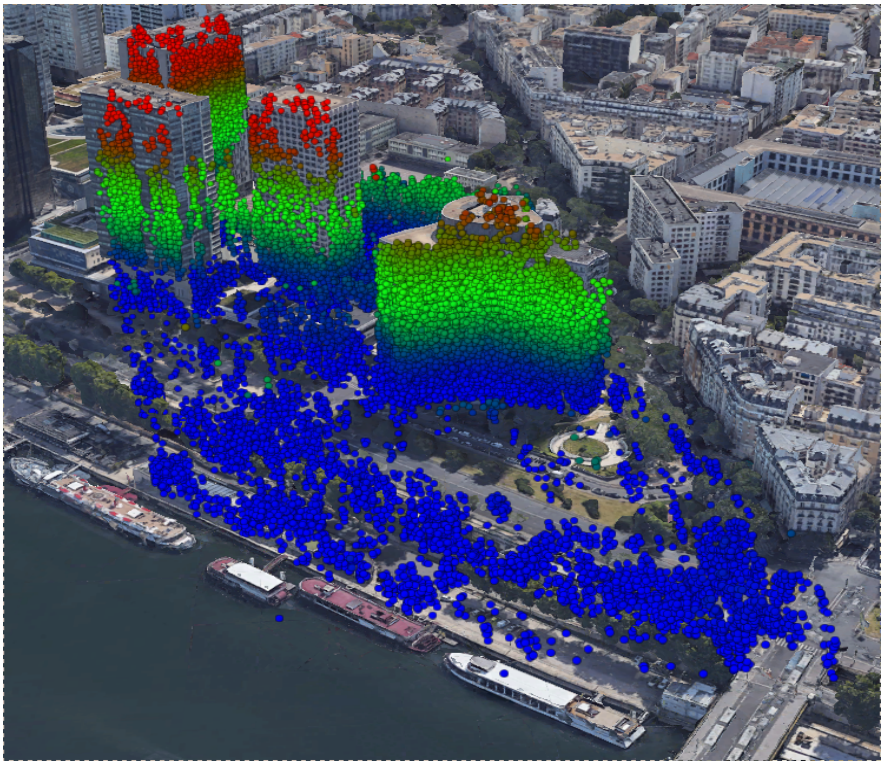


(b)

Figure 8.19 – Visualization of the thresholded module of the reconstructed cube (Andriy Fedorov et al., 2012) super-imposed with the ground truth.



(a)



(b)

Figure 8.20 – Visualization of the point cloud obtained from the reconstructed cube and projected in Google Earth © .

Chapter 9

Segmentation in urban SAR tomography

The presented work in this chapter comes mostly from (Rambour *et al.*, 2018b). Some results were added in the experimental part.

9.1 Graph-cut based surface segmentation

Starting from the tomographic reconstruction (a 3-D volume $\underline{\boldsymbol{u}}$) obtained with one of the methods described in the previous chapters (see chapters 4 and 8) we aim to recover the urban surfaces (ground, building facades, roofs). Following a typical approach in computer vision for surface reconstruction, we formulate the problem as an energy minimization problem (Kolmogorov *et al.*, 2014). We seek a surface $\mathscr{S}$ corresponding to an elevation map: $(x, y) \mapsto z = \mathscr{E}(x, y)$ that both fits well the reconstructed tomographic volume and that is smooth. We first formulate a cost function that captures these two properties, then we describe an efficient graph-based algorithm to perform the minimization of the cost function.

9.1.1 Definition of the cost function

The first component of the cost function favors surfaces that are faithful to the reconstructed tomographic volume. We seek surfaces such that, when considering a given ray direction in 3-D space, the scatterer encountered along the ray falls close to the ray-surface intersection, see Fig. 9.1. The reflectivity profile along the ray may display several local maxima due to residual sidelobes after the tomographic inversion. Rather than detecting these maxima and deciding for the most meaningful maximum, we consider that a satisfying location of the surface is a location such that the reflectivity profile is split into two well-balanced halves. We define the cumulative reflectivity $C^{-}(r_s)$ from the antenna to the surface $\mathscr{S}$ and the cumulative reflectivity $C^{+}(r_s)$ from

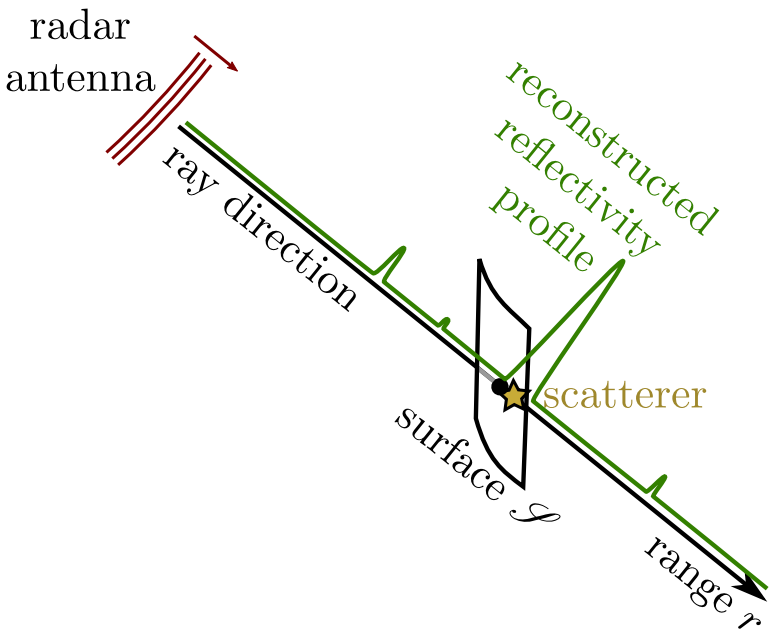


Figure 9.1 – We seek a surface $\mathscr{S}$ that, for each ray, is close the scatterer(s) found along the ray.

the surface to the maximum range:

$$C^-(r_s) = \int_{r_{\min}}^{r_s} |u(r)| \, \mathrm{d}r, \tag{9.1}$$

$$C^+(r_s) = \int_{r_s}^{r_{\max}} |u(r)| \, \mathrm{d}r \,, \tag{9.2}$$

where r_s is the range of the surface, i.e., the distance from the radar to the surface, in the direction of the ray. If the surface is such that $C^-(r_s) < C^+(r_s)$, then it is too close to the radar: most of the reflectivity of the scatterers encountered along the ray is located beyond the surface. Conversely, if $C^-(r_s) > C^+(r_s)$, the surface is too far from the radar: scatterers accounting for most of the reflectivity are located before the surface. The imbalance $C^-(r_s) - C^+(r_s)$ is therefore an indication of bad surface localization. In order to favor surfaces that are located close to the position of equilibrium, we define the penalty:

$$\begin{aligned} D(r) = &\int_{r_{\min}}^{r} \left[C^-(r_s) - C^+(r_s)\right]_+ \mathrm{d}r_s \\ &+ \int_{r}^{r_{\max}} \left[C^+(r_s) - C^-(r_s)\right]_+ \mathrm{d}r_s \,, \end{aligned} \tag{9.3}$$

where the notation $[\cdot]_+$ denotes the positive part: $\forall w$, $[w]_+ = \max(w, 0)$. The term $[C^-(r_s) - C^+(r_s)]_+$ in the first integral of equation (9.3) is non-zero only if the distance r_s is larger than the distance of equilibrium r_{equi} (where r_{equi} is such that $C^+(r_{\text{equi}}) =$

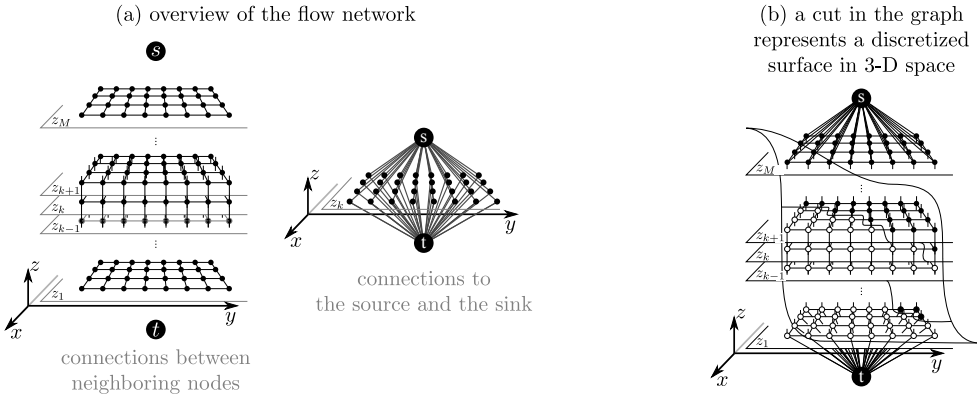


Figure 9.2 – Representation of the topology of the flow network: (a) a node represents a 3-D (x, y, z) location in ground geometry, each node is connected to its 6 closest neighbors and also to the source s and to the sink t; (b) a cut separates the graph into two disconnected sub-graphs, it represents a discretized version of the segmented surface $\mathscr{S}$.

$C^-(r_{\text{equi}})$). Then, if $r > r_{\text{equi}}$, the first integral equals $\int_{r_{\text{equi}}}^{r} (C^-(r_s) - C^+(r_s))\,\mathrm{d}r_s$. Conversely, the second integral in (9.3) is non-zero only if the distance r_s is smaller than the distance of equilibrium r_{equi}. It is then equal to $\int_{r}^{r_{\text{equi}}} (C^+(r_s) - C^-(r_s))\,\mathrm{d}r_s$. $D(r)$ is thus a function that *monotonically increases with the distance* $|r - r_{\text{equi}}|$ and that is minimal and equal to zero when $r = r_{\text{equi}}$.

The second component of the cost function guarantees that the segmented surface be smooth. To prevent the surface from oscillating in order to pass through the position of equilibrium r_{equi} for each ray, we penalize the area $\mathscr{A}(\mathscr{S})$ of the surface. In order to favor surfaces with horizontal or vertical parts, we suggest measuring the area with respect to the ℓ_1 distance (i.e., Manhattan distance $\|\boldsymbol{p}\|_1 = |p_x| + |p_y| + |p_z|$).

To summarize, we suggest defining the segmentation as the surface $\mathscr{S}$ that is a solution to the following variational problem:

$$\min_{\mathscr{S}} \int_{\text{ray}\in\mathscr{R}} D_{\text{ray}}(r_{\text{ray}\to\mathscr{S}})\,\mathrm{d}\mathscr{R} \;+\; \beta\,\mathscr{A}(\mathscr{S})\,, \tag{9.4}$$

where $\mathscr{S}$ is required to be representable as an elevation map $\mathscr{E}(x, y)$ (formally, there exist a function $\mathscr{E} : (x, y) \mapsto \mathscr{E}(x, y)$ such that $\mathscr{S}$ be the boundary of the epigraph of $\mathscr{E}$). To prevent from introducing too many notations, we denote 'ray' for the generic definition of a ray in an adequate parameterization (a line in 3-D space), $\mathscr{R}$ represents the set of all rays, $r_{\text{ray}\to\mathscr{S}}$ is the distance from the radar to the surface $\mathscr{S}$ along the direction defined by 'ray', D_{ray} is the penalty defined by equation (9.3) for the direction specified by 'ray'. Finally, β is a parameter that balances the fidelity to the tomographic reconstruction and the spatial smoothness of the surface.

9.1.2 Graph-cut algorithm for minimization

The variational problem (9.4) is very challenging to solve. We show in this paragraph that, after discretization of the surface and of the set of rays, it can be transformed into a minimum cut problem on a particular graph. By computing the minimum cut using available efficient graph-cut libraries, we obtain a fast method to solve the surface segmentation problem.

The surface is represented by an elevation map $\mathscr{E} : (x, y) \mapsto z = \mathscr{E}(x, y)$ (which guarantees that it is representable as an elevation map). The horizontal location (x, y) and the elevation z are discretized. To make an easier connection between the elevation map and the surface it defines, we consider the layer cake decomposition of the elevation. With this decomposition, a discrete elevation map corresponds to a binary volume (a discrete version of the epigraph of $\mathscr{E}$) and the boundary in that volume defines the discrete surface.

We build a graph as depicted in figure 9.2, with a node to represent each voxel of the binary volume of $\mathscr{E}$. Two special nodes, called the source (denoted 's') and the sink (denoted 't') are added in order to simulate a flow from the source to the sink. Nodes are connected together by directed edges with specific capacities and a flow is said to be admissible if and only if the flow along each edge is non negative and smaller or equal to the edge capacity, and there is no flow accumulation/creation at nodes (except at the source and at the sink). By the max-flow min-cut theorem, algorithms that identify the maximum admissible flow on the graph can also identify the minimum cost cut among all possible cuts in the graph[1], see for example (Boykov and Kolmogorov, 2004). During the graph construction, by creating edges with well-chosen capacities, we can make the cost of any cut exactly match the cost of the corresponding surface in the variational formulation (9.4).

To represent the first term in equation (9.4), we substitute D_{ray} with its definition in equation (9.3):

$$\begin{aligned}\int_{\text{ray}\in\mathscr{R}} D_{\text{ray}}(r_{\text{ray}\to\mathscr{S}})\,\mathrm{d}\mathscr{R} = &\int_{\text{ray}\in\mathscr{R}}\int_{r_{\min}}^{r_{\text{ray}\to\mathscr{S}}} \left[C^-_{\text{ray}}(r_s) - C^+_{\text{ray}}(r_s)\right]_+ \mathrm{d}r_s\,\mathrm{d}\mathscr{R} \\ &+ \int_{\text{ray}\in\mathscr{R}}\int_{r_{\text{ray}\to\mathscr{S}}}^{r_{\max}} \left[C^+_{\text{ray}}(r_s) - C^-_{\text{ray}}(r_s)\right]_+ \mathrm{d}r_s\,\mathrm{d}\mathscr{R}\,. \end{aligned} \tag{9.5}$$

Each of the two terms corresponds to summations over a half-space whose boundary is $\mathscr{S}$: the half-space that contains the radar and the half-space with the farther ranges, respectively. We add an edge directed from the source to node i, the node that represents the 3-D position (x_i, y_i, z_i) and that is located at the distance r_i from the radar antenna. The capacity[2] of this edge is set to $\left[C^-_i(r_i) - C^+_i(r_i)\right]_+$, where C^-_i and C^+_i are the cumulative reflectivities computed along the ray directed from the radar through the

[1] the cost of a cut is the sum of the capacities of all edges cut that are directed from a node in the source partition to a node in the sink partition

[2] note that an edge with zero capacity can be suppressed because it carries no flow and has no contribution to the cost of the cuts

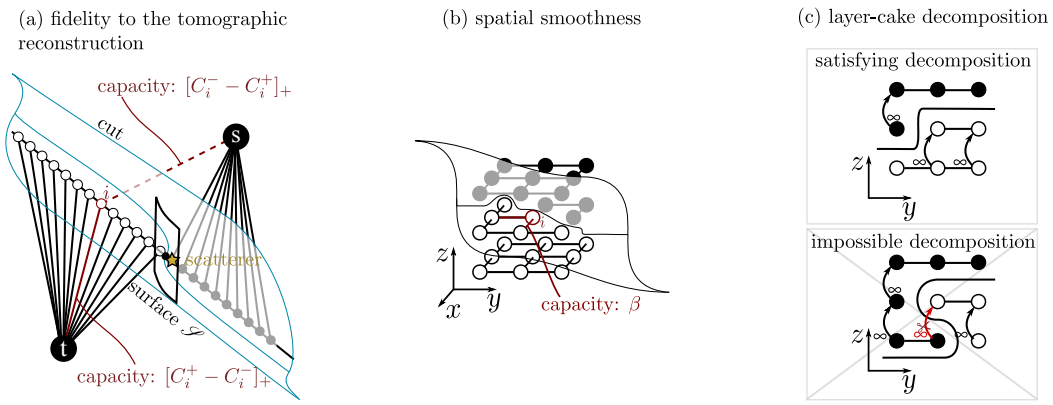


Figure 9.3 – The capacities of the edges are chosen so that the cost of the cut corresponds to the energy of the surface. (a) the fidelity to the tomographic reconstruction is enforced via edges originating from the source or leading to the sink. (b) the spatial smoothness of the surface is obtained by adding bi-directional edges between neighboring nodes in the x and y directions. (c) to prevent the cut from severing twice a column of nodes along the z direction, ascending edges with infinity capacity are added. These edges are counted in the total cost of the cut only when they go down-stream: from the partition containing the source to the partition containing the sink.

point of coordinates (x_i, y_i, z_i). Another directed edge is added from node i to the sink, with capacity $\left[C_i^+(r_i) - C_i^-(r_i)\right]_+$. To separate the graph into two parts by a cut, some edges must be severed (unless the cut passes through the distance of equilibrium r_{equi}) and the sum of the capacities of those edges corresponds to a discretization of equation (9.5), see Fig. 9.3(a).

Additional edges are created to account for the regularization term $\beta\mathscr{A}(\mathscr{S})$: bi-direction edges between pairs of nodes that are direct neighbors in the x or y directions, with capacity β, see Fig. 9.3(b). Finally, ascending edges with infinite capacity are included between neighboring nodes in the z direction. These edges are necessary to guarantee that the cut defines a surface that is representable by an elevation map, see 9.3(c). Similar edges are added in Ishikawa's graph construction that is also based on the layer-cake decomposition (Ishikawa, 2003).

In our implementation, we computed efficiently the summations along the rays by resampling the reconstructed tomographic volume in ray geometry so that sums could be carried out along columns in this new geometry. For the construction of the graph and the computation of the minimum cut, we used the graph-cuts library by Boykov and Kolmogorov (Boykov and Kolmogorov, 2004).

9.2 Joint reconstruction and surface segmentation

The knowledge provided by the segmented urban surfaces can help to improve the inversion described in chapter 8. The reconstruction algorithm that can most readily be extended to include segmented surfaces is the 3-D inversion method described in

equation (8.3). Under the assumption that the signal retrieved over urban areas is mainly constituted of punctual bright points, sparsity may be an efficient enough prior to obtain clean tomograms. Nonetheless, this implies that the tuning of the sparsity parameter be done locally according to the position of the scatterers. In CS for SAR tomography, the sparsity constraint is generally set locally in the range and azimuth direction but constant for each radar cell. Here we propose to use the 3-D information provided by the estimated surface to go one step further and perform a spatially varying penalization of the sparsity.

When applying CS or the 3D inversion, the sparsity parameter μ is set proportional to the level of spurious elements in the reconstruction. Generally μ is set according to the noise level (Zhu and Bamler, 2012b), but as decorrelation mechanisms and side-lobes should also be discarded, the knowledge of the sensor thermal noise may not be enough. Many SAR tomographic algorithms propose to estimate the number of backscattering elements in order to extract the largest scatterers in each radar cell. This step cleans the estimated tomograms from residual outliers, but is also a challenging task for large multitemporal stacks in dense environments. Moreover the CS approach may then lose one of its asset with respect to MUSIC or WSF if it also needs an estimation of the number of targets.

Under the assumption that the location of the urban surface is known, the sparsity parameter μ can be spatially tuned to lead to refined tomograms. Even when the surface is roughly known, it provides information on where the reconstructed signal should be located. In the proposed iterative algorithm, μ is computed as a function of the distance to the surface in the 3-D space and the number of iterations:

$$\mu_k(\boldsymbol{p}, \mathscr{S}) = \mu_0 + \frac{b}{(n-1)^2}\left(\frac{k}{n-k}\, d(\boldsymbol{p}, \mathscr{S})\right)^2 \tag{9.6}$$

where $d(\boldsymbol{p}, \mathscr{S})$ is the Euclidean distance from the point $\boldsymbol{p} = (x, y, z)^T$ to the estimated surface, k is the current iteration and n the total number of iterations. We define by $\boldsymbol{\mu}(\mathscr{S}) \in \mathbb{R}^{N_x.N_y.N_z}$ the 3-D sparsity parameter map. As the surface location estimation may be subject to errors in the first iterations, it is important to avoid over-penalizing points moderately close to the surface during the first reconstructions. This why we use the square of the distance d multiplied b a factor smaller than 1. As the number of iterations increases, the reconstruction and thus the surface estimation should be more accurate (and better in match) which suggests an increase of the penalization of the distance from a reconstructed voxel to the surface. $\mu_0 + b$ is then the desired minimal sparsity that need to be applied to voxel not on the surface.

The proposed iterative reconstruction and surface segmentation is summarized in the following algorithm:

Procedure REDRESS

Input: $\underline{\boldsymbol{v}}$ (stack of SLC SAR images)
Output: $\underline{\hat{\boldsymbol{u}}}$ (3-D cube of complex reflectivities)

$\mathscr{S}$ (urban surface)

Initialization :
1: $k \leftarrow 0$
2: **while** $k < n$ **do**
3: $\underline{\hat{\boldsymbol{u}}} \leftarrow$ `3-D Inversion`$(\underline{\boldsymbol{v}}, \boldsymbol{\mu}(\mathscr{S}))$
4: $\mathscr{S} \leftarrow$ `graph cut`$(\underline{\hat{\boldsymbol{u}}})$
5: $k \leftarrow k + 1$
6: **end while**
7: **return** $\underline{\hat{\boldsymbol{u}}}, \hat{\mathscr{S}}$

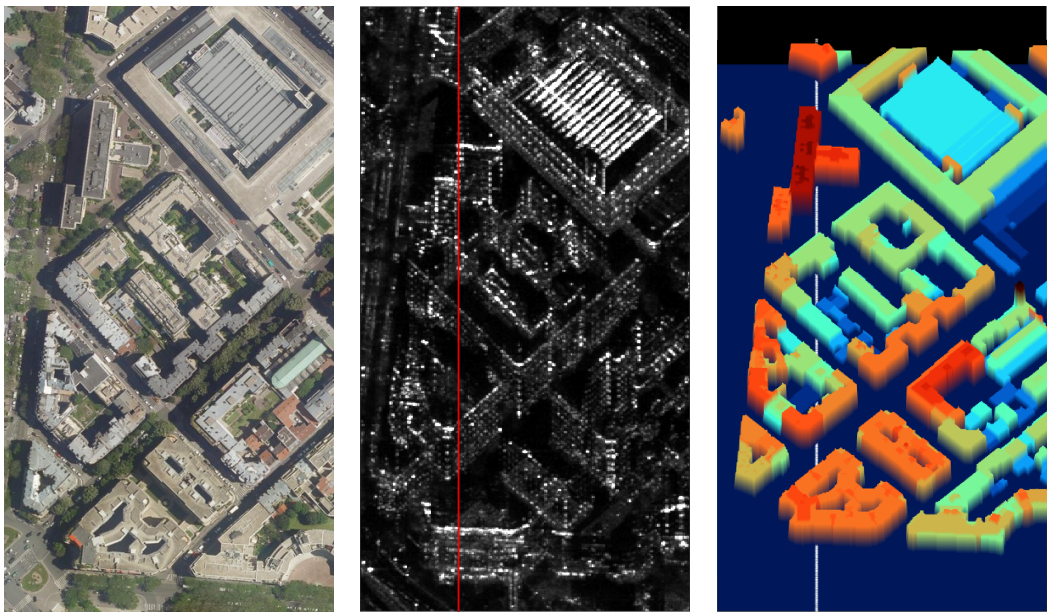
The refined tuning of the sparsity according to the surface allows to considerably improve the scatterers localization and main lobe reduction. In some cases, however, the segmented surface follows the lobe main extension direction and is not as localized as would be expected for a collection of point-like scatterers. In the global reconstruction of the scene, most of the artifacts due to the TV penalization are suppressed after 10 iterations. The obtained surface is then very close to the ground truth and provides the lowest error according to table 9.1.

9.3 Experiments

To validate both the generality of the segmentation method and its efficiency on real data we present different experiments performed on a set of 40 TerraSAR-X images of a part of Paris, France. The selected area corresponds to the French Ministry of Foreign Affairs and buildings in its neighborhood in the south west of the city. The optical view of the scene is presented Fig. 9.4 side to side with the temporal average of the SAR intensities. To evaluate our results we use a ground truth map extracted from Google Earth©. Different SAR tomographic reconstruction methods introduced in section 2 (Capon Beamforming, MUSIC , WSF , SPICE , CS and the 3-D inversion) are first applied on the slice represented by the red line in fig. 9.4(b), then on the entire data set. The segmentation by graph-cut is then performed in both cases. The obtained surfaces are compared with the ground truth for each tomographic estimator.

The results are shown in Fig 9.5 for the reconstructions of the slice, and in Fig. 9.6 for the reconstruction of the whole scene. With the first experiment the behavior of each estimator and the resulting surface can be observed in greater details. In the reconstructions, the areas where the surface is occluding itself are detected as shadow areas and removed. The resulting gaps introduced are filled according to the height of the first point outside it.

The second experiment illustrates the role of the 3-D smoothing. The surface is shown as seen from the sensor point of view. Since some tomographic estimators provide an estimate of the reflectivities, those reflectivities can be plotted to illustrate the



(a) (b) (c)

Figure 9.4 – Observed urban areas : optical image (a), temporal mean of the corresponding SAR image (b), and the 3-D model from Google Earth used as a ground truth. The red line in (b) and white one in (c) correspond to the slice shown in Fig. 9.5

distribution of scatterers on the reconstructed surfaces.

A third experiment presents the evolution of the reconstructed slice as the REDRESS algorithm iterates *cf.* Fig. 9.8. It can be observed that the distribution of reflectivities becomes much sharper after a few iterations.

To estimate the covariance matrix at each point, we used a 7×7 Gaussian filter. For MUSIC and WSF , the number of scatterers is set constant and equal to 2 to avoid selecting too many outliers while allowing multiple scatterers within each radar resolution cell. For these two estimators, the reflectivity is estimated by mean square minimization, to keep a physical interpretation of the tomograms. As the scene is very heterogeneous with a lot of layover, this step introduces some undesired mixing of the information in the image. The surfaces estimated from tomographic reconstructions using spectral analysis techniques present noticeable artifacts in the dense areas. Some structures are too extended, partially filling streets or the building atrium. Meanwhile, the averaging step makes the tomographic estimation smoother in homogenous areas for the fully sparse approaches MUSIC and WSF .

The CS technique performed on all the data set presents results that seem visually the closest to the ground truth. Many details can be observed in this reconstruction: most of the rooftops and visible streets are well segmented and the buildings atrium are also retrieved.

For all the previous estimators, the TV minimization produces some building elongation resulting in phantom structure in low intensity signal area. This can be seen for instance in the bottom right part for Fig. 9.6 or around the position 450 for Fig. 9.5.

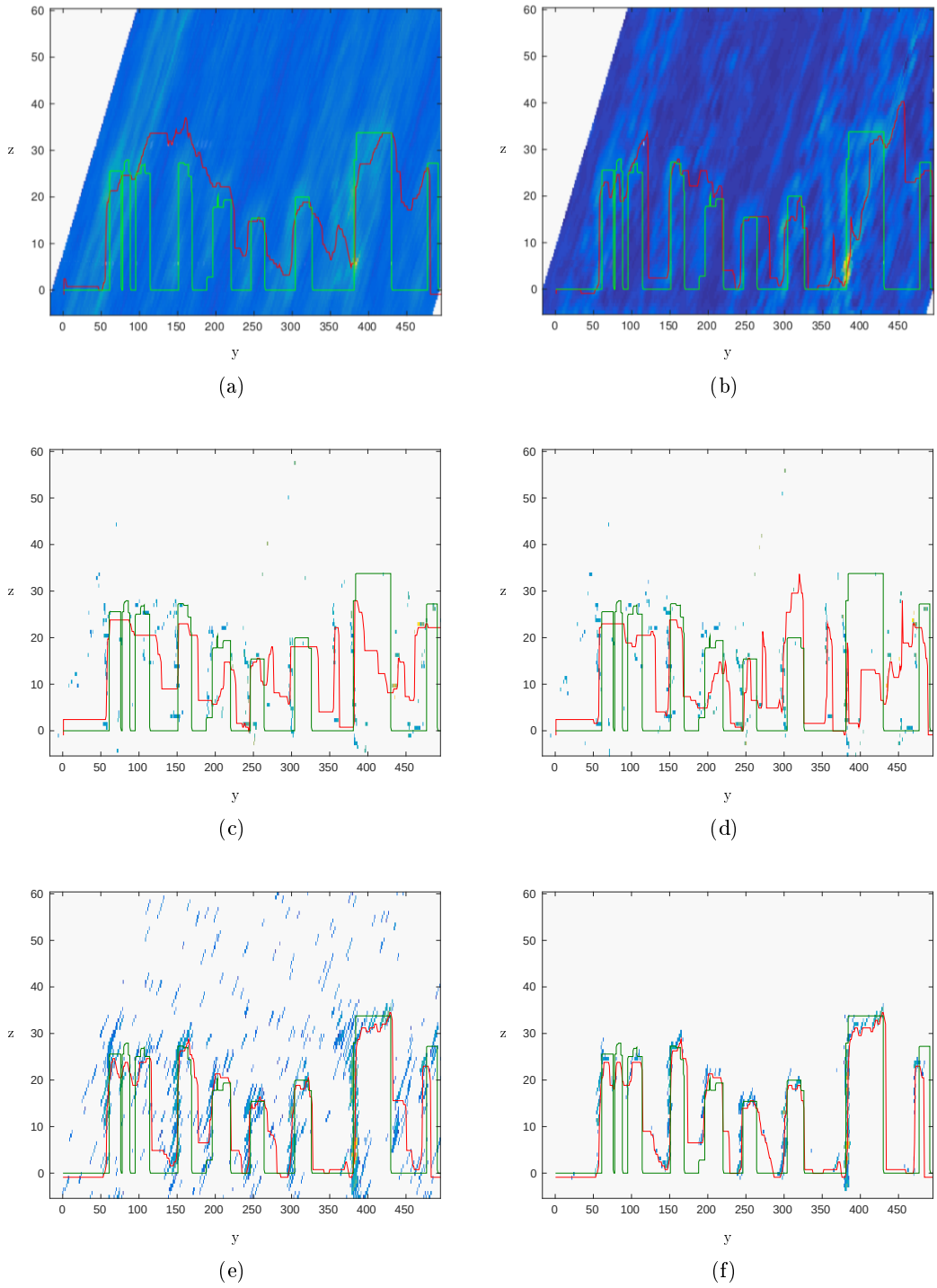


Figure 9.5 – Urban surface estimation using graph-cut segmentation of the tomograms, as described in section 3. The estimated surface corresponds to the red profile. The ground truth for the given slice is shown in green. The tomograms are obtained using Capon beamforming (a), SPICE (b), MUSIC (c), WSF (d), the 3-D inversion approach (e) and REDRESS (f).

Image of the backscattering power of the surface are presented in Fig 9.7. As expected corners such as bottom of buildings or windows send back a high intensity signal.

Finally, the reconstructions obtained using the REDRESS algorithm presents the closest results to the ground truth with far less TV artifacts. 3-D representation of the tomographic reconstruction using the REDRESS algorithm are shown in Fig 9.9 and 9.10. The accuracy of the results is also illustrated by plotting the non-zero voxels into Google Earth © *cf.* fig. 9.11.

To conduct a quantitative comparison of the segmentation results, we report the mean error for each estimated surface to the ground truth *cf.* 9.1. The TV parameter β is set, for each method, as the one minimizing this error.

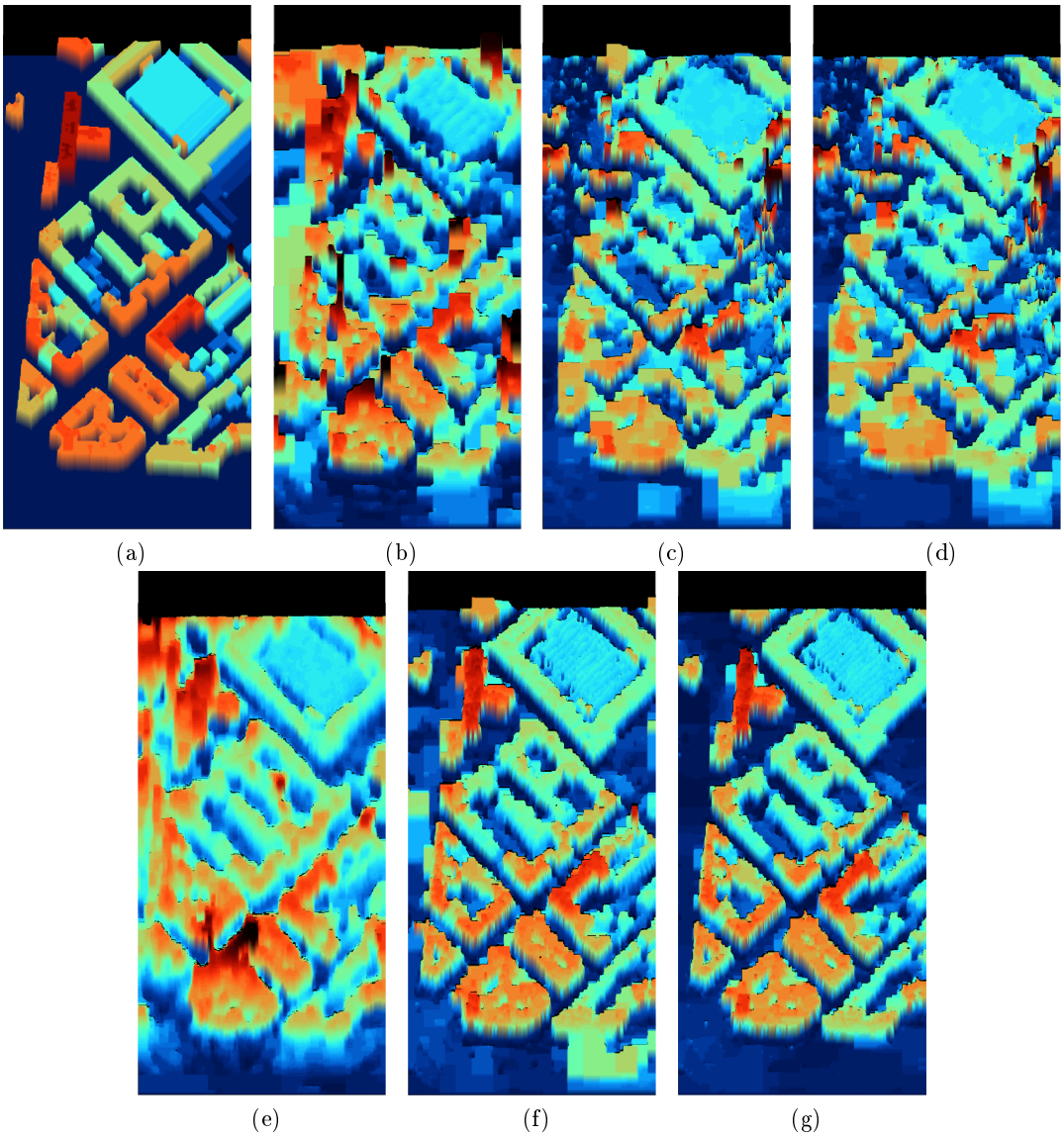


Figure 9.6 – Ground truth height (a), scene surface estimation using SPICE (b), MUSIC (c), WSF (d), Capon beamforming (e), the 3-D inversion (f) and REDRESS (g). For each results, the image shows the surface colored according to its height.

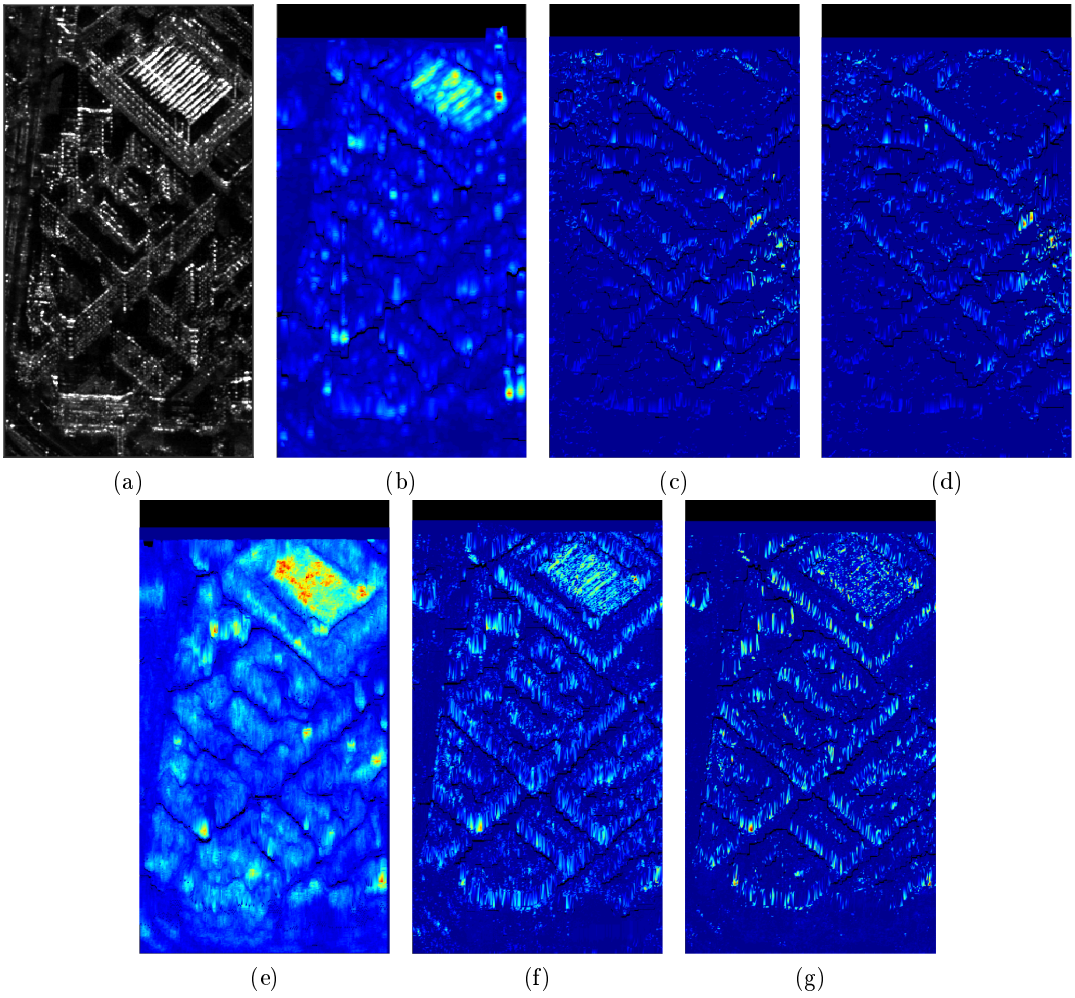
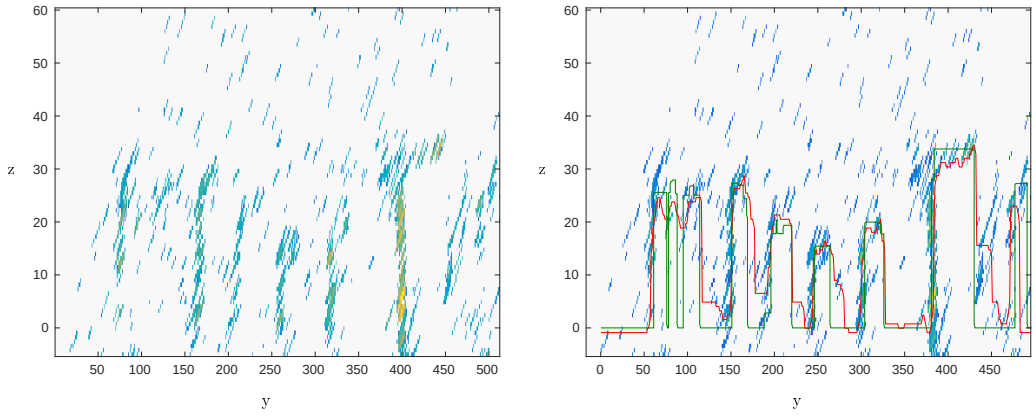


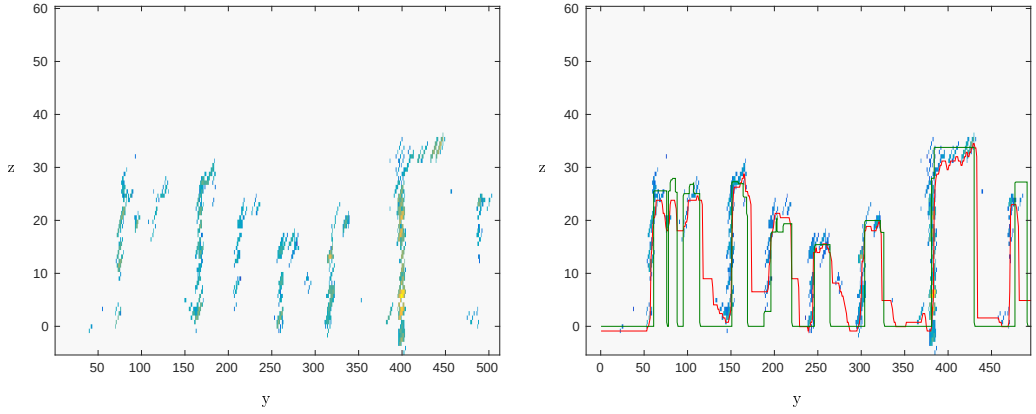
Figure 9.7 – Average amplitude SAR image (a), surface amplitude estimation using SPICE (b), MUSIC (c), WSF (d), Capon beamforming (e), the 3-D inversion (f) and REDRESS (g). For each results, the image shows the surface colored according to its amplitude.

9.4 Conclusion

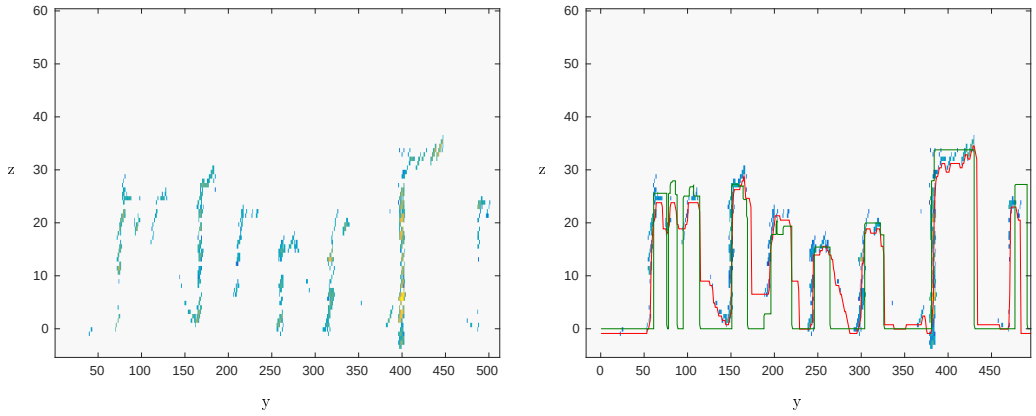
In this chapter we introduced a graph-cut based segmentation algorithm to estimate the urban surfaces from a SAR tomographic reconstruction. The proposed approach is very general and can be used in combination with many different tomographic algorithms. Experiments done on a set of 40 TerraSAR-X images of Paris show good results for different tomographic estimators (Capon beamforming, MUSIC , WSF , SPICE , CS and 3-D inversion). As the 3-D inversion algorithm is designed to use 3-D priors, we also present an algorithm that alternatively reconstructs the 3-D distribution of reflectivities, segments the urban surfaces from the volume of reflectivities and updates the regularization so as to improve the subsequent 3-D reconstruction. While the non-iterative 3-D inversion algorithm fails in some cases to reduce the main lobes of the strong scatterers, the alternating scheme achieves a much sharper estimation of the distribution of reflectivities.



(a)



(b)

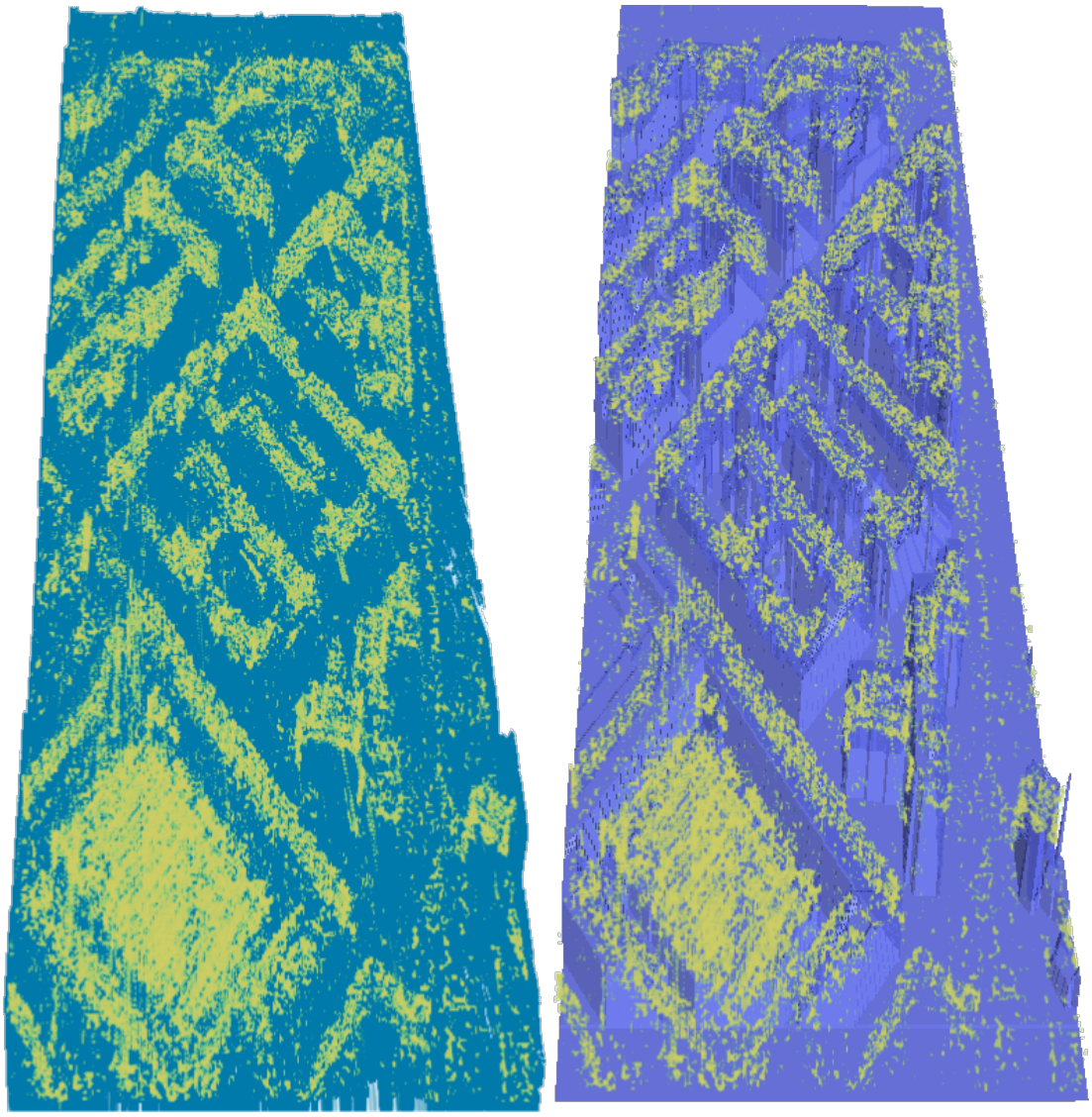


(c)

Figure 9.8 – Three different iteration steps from the alternate reconstruction algorithm. On the left column, the estimated reflectivities are shown for the profile presented in Fig 9.4. On the right, the estimated surface (red) and the ground truth (green) are superimposed in addition to the estimated reflectivities. Rows (a), (b) and (c) correspond respectively to the first, third and fifth iterations (last one).

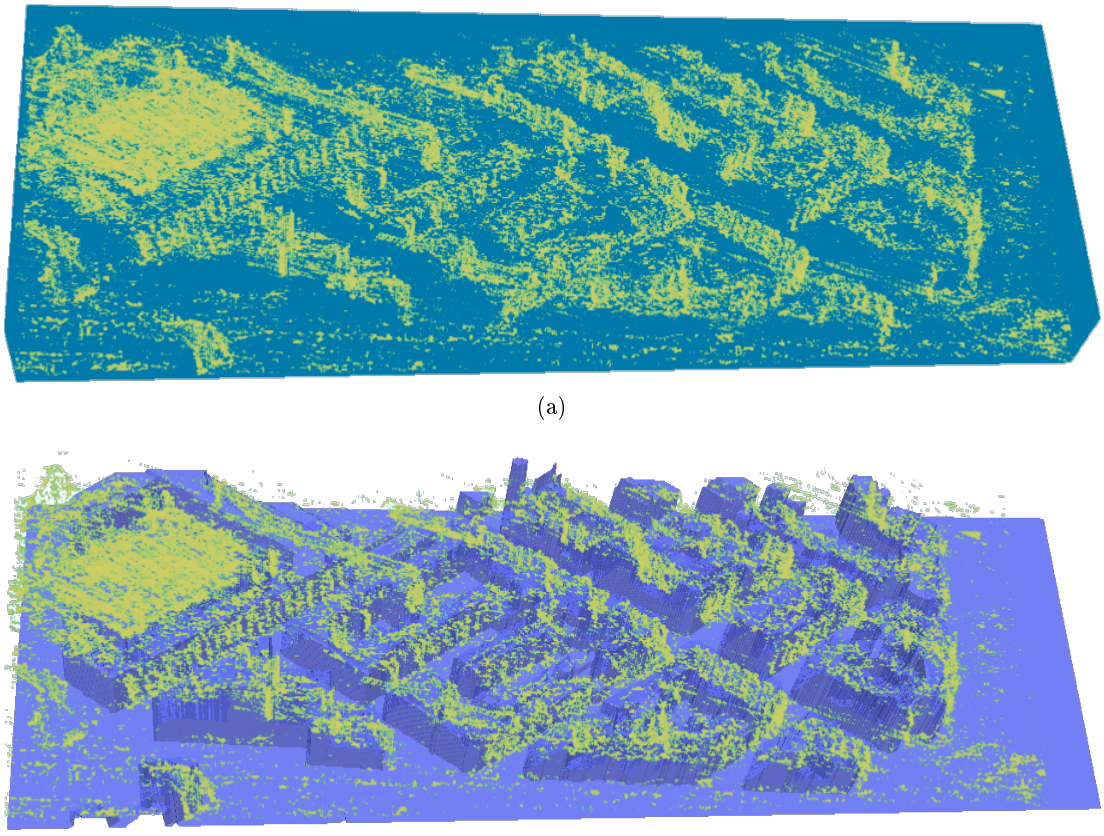
Estimator	**Mean Error**	$\boldsymbol{\beta}$
Capon Beamforming	4.58 m	1.5
MUSIC	3.23 m	1.3
WSF	3.12 m	1.6
SPICE	4.24 m	12.6
3-D inversion	2.50 m	2.0
REDRESS	**1.60 m**	2.0

Table 9.1 – Mean errors between the estimated surfaces and the ground truth, last column: optimal β values used for the surface segmentation.



(a) (b)

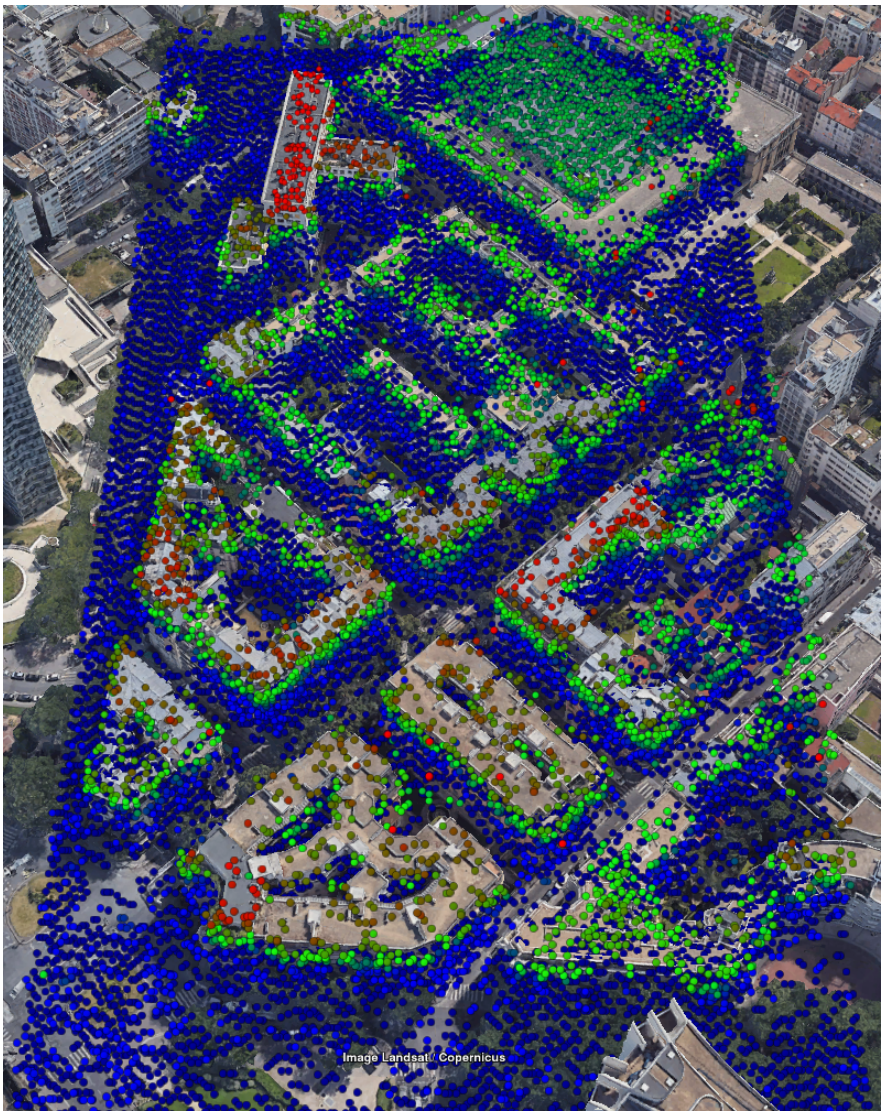
Figure 9.9 – Visualization of the module of the reconstructed cube REDRESS algorithm (left), super-imposed with the ground truth (right) (Andriy Fedorov et al., 2012).



(a)

(b)

Figure 9.10 – Visualization of the module of the reconstructed cube using the REDRESS algorithm (left), super-imposed with the ground truth (right) (Andriy Fedorov et al., 2012).



(a)

Figure 9.11 – Plotting of the non-zero intensity scatterers in Google Earth ©. For computational purpose 1 voxel out 10 is drawed.

Conclusion and perspectives

Conclusion

The main objective of this PhD was to provide methods to use the structural information to perform the SAR tomographic estimation of the 3-D reflectivity for dense urban scenes. Two different strategies were explored:

- Exploit the redundancy in the SAR images
- Regularize the estimation

The first method was based on the non-local estimation of the covariance matrix. This approach proved to be challenging and didn't offer guaranties to outperform basics local covariance matrix estimation approaches.

The second approach was expected to be relatively successful as it was close to the well known CS approach which may be consider as one of the top tomographic estimator for urban areas. The proposed algorithm allowed to perform the tomographic 3-D inversion of a scene under geometrical priors described in the ground coordinate system. This allowed to reduce the number of outliers and favor structured distribution of scatterers.

Tomograms and volume representation of the reflectivity may be hard to analyze, specially with close buildings of comparable height. We proposed a post-processing graph-cut based segmentation to extract the urban surfaces of a tomographic reconstruction. Due to the very generic design of the method, it can used in association to any estimator providing a measure of the scene reflectivity (amplitude or power). The surface is expected to be retrieved were high intensity voxels are found. The geometrical priors consist of a TV minimization under constraints avoiding the structures to be transparent for the electromagnetic wave and crenelation like distortions.

Finally, we linked the 3-D reconstruction with the surface segmentation, to design an iterative algorithm that refine the reflectivity estimation based on the estimated surface.

All the proposed method were tested on a stack of TerraSAR-X images and the results were compared to a ground truth. The REDRESS algorithm lead to a average error less than 2m and preserves many details such as buildings courtyard or some roof reliefs.

Part of the PhD was also dedicated to the study and phase calibration to a set of airborne SAR images obtained with the sensor SETHI from the ONERA. This work is however still in progress and the data are not usable for SAR tomography yet.

Perspectives

SAR tomography is still a relatively recent research field. Its application to dense urban areas is a trending topic and new algorithms are proposed either to perform the tomographic inversion or to apply some post-processing regularization. The framework presented here to perform the tomographic inversion in ground coordinates allows to used many different regularization strategies in a plug-and-play fashion. Some of alternative to the presented priors could then be tried:

- Going back to the smoothing presented in chapter 8, strategies allowing to know the local orientation of the scatterers distribution could be used to locally tune the gradient parameters. This could be done first as an improve REDRESS algorithm.
- Rather than using the quadratic norm of the gradient, minimizing the TV for the tomographic reconstruction should also be tried.

Besides trying sophisticated priors, the 3-D projection operator could be used to retrieve a SAR image associated to a given sensor track. This could be useful for several applications:

- The proposed methods may introduce an increase of the number of parameters. We discussed how to set these parameters in chapter 8 but this remain an open question. This consideration is also true for conventional CS . A tuning strategy could be to find the set of parameters that allows to best preserve the input SAR images.
- Being able to exploit a stack of SAR images to generate a new one could be used for changes detection or image restoration.

When dealing with a new set of SAR images, phase calibration can be a challenging task. An idea could be to perform iteratively the calibration and the tomographic reconstruction. At each step the linear contribution of the APS and the scatterers height could then be estimated using the information from the previous iteration to enhance the results.

Acronyms

ADMM Alternating Direction Method of Multipliers. 89–91

APS Atmospheric Phase Screen. 14, 18, 27, 125

BIC Bayesian Information Criterion. 46

CS Compressive Sensing. VIII–X, 3, 4, 27, 42, 45, 46, 51–53, 78, 114–116, 119, 124, 125

DFT Discrete Fourier Transform. 27, 29, 44

DIFT Discrete Inverse Fourier Transform. 20

DOA Direction Of Arrival. 27, 28, 31

FIR Finite Impulse Response. 28

GLR Generalized Likelihood Ratio. 67, 69, 71, 72, 74, 77–80

LASSO Least Absolute Shrinkage and Selection Operator. 44

LMM Lagrangian Methods of Multipliers. 89, 90

LOS Line of Sight. 9, 14, 17

LS Least-Square. 37

M-RELAX Multilook-RELAXation spectral estimator. 37, 39, 40, 42, 76

MACT Minimum Accuracy / Completeness Trade-off (MACT). 93

ML Maximum Likelihood. III, 37, 38

MOS Model Order Selection. 46, 47, 62

MUSIC MUltiple SIgnal Classification. III, 2, 31, 33–35, 37, 46, 51, 52, 69, 83, 114–117, 119

MUSIC-PS MUSIC Pseudo Spectrum. 33, 51, 52, 83

MVDR Minimum Variance Distortionless Response. 28

NL-SAR Non Local - SAR. VIII, 2, 66–70, 73, 78, 88

NSF Noise Subspace Fitting. 34

PS Permanent Scatterer. IV, 17, 18, 76, 77

PSF Point Spread Function. 14, 28

RADAR RAdio Detection and Ranging. VII, 1, 9, 11, 12, 14, 17, 20, 28, 62, 78

REDRESS Alte**R**nat**E**d 3-**D** **RE**construction and **S**urface **S**egmentation. 5, 115–118, 121, 122, 124, 125

RIP Restrictive Isometry Property. 42, 44

SAR Synthetic Aperture Radar. IV, VII–X, 1–4, 9–11, 13–15, 17–22, 24, 25, 28, 31, 34, 42, 44, 51–53, 55, 57–60, 62, 65, 66, 68–70, 73, 75, 76, 78, 81, 82, 85–87, 91, 92, 103, 104, 119, 124, 125

SL1MMER Scale-down by ℓ_1 norm Minimization, Model selection, and Estimation Reconstruction. 46

SLC Single-Look Complex. 15, 18, 20, 21, 53, 66, 70

SNR Signal-to-noise ratio. 10, 28, 45, 57

SPICE SParse Iterative Covariance-based Estimation. IV, 41, 43, 48, 115, 117, 119

SSF Signal Subspace Fitting. 34, 36

TV Total Variation. 4, 115, 116, 124, 125

WSF Weighted Subspace Fitting. III, 34, 36, 37, 46, 114–117, 119

Bibliography

H. Aghababaee, A. Budillon, G. Ferraioli, V. Pascazio, and G. Schirinzi. On the role of non-local filtering in forest vertical structure characterization using sar tomography. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6032–6035, July 2017. 73

Hossein Aghababaee, Alessandra Budillon, Giampaolo Ferraioli, and Angel Caroline Johnsy andVito Pascazio andGilda Schirinzi. Multiple Scatterers Detection Based on Signal Correlation Exploitation In Urban SAR Tomography. In *2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2018. 26

Andriy Fedorov et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9):1323 – 1341, 2012. Quantitative Imaging in Cancer. 106, 107, 121, 122

Z. Ben-Haim, Y. C. Eldar, and M. Elad. Coherence-Based Performance Guarantees for Estimating a Sparse Vector Under Random Noise. *IEEE Transactions on Signal Processing*, 58:5030–5043, October 2010. 44

Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014. 89

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. 90, 95

Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1124–1137, 2004. 112, 113

James P. Boyle and Richard L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces", booktitle="advances in order restricted statistical inference. pages 28–47, New York, NY, 1986. Springer New York. 89

Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *SIAM Journal on Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 4(2):490–530, 2005. 66, 68

A. Budillon, A. Evangelista, and G. Schirinzi. Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):488–499, Jan 2011. XIV, 2, 42

A. Budillon, A. C. Johnsy, and G. Schirinzi. A Fast Support Detector for Super-resolution Localization of Multiple Scatterers in SAR Tomography. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6):2768–2779, June 2017. 46

Alessandra Budillon, Angel Caroline Johnsy, and Gilda Schirinzi. Extension of a Fast GLRT Algorithm to 5D SAR Tomography of Urban Areas. *Remote Sensing*, 9(8):844, 2017. 26

Kenneth P. Burnham and David R. Anderson. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004. 46

J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969. 31

S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998. 44

K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver. A test statistic in the complex wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(1):4–19, jan 2003. 67

D. D'Aria, A. Ferretti, A. Monti Guarnieri, and S. Tebaldini. SAR Calibration Aided by Permanent Scatterers. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):2076–2086, April 2010. 17, 18

C. A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger. NL-SAR: A Unified Nonlocal Framework for Resolution-Preserving (Pol)(In)SAR Denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2021–2038, April 2015. 2, 66, 69, 88

O. D'Hondt, C. López-Martínez, S. Guillaso, and O. Hellwich. Nonlocal Filtering Applied to 3-D Reconstruction of Tomographic SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 56(1):272–285, Jan 2018. 58, 73

Jim Douglas, Jr. and H. H. Rachford, Jr. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956. 89

Olivier D'Hondt. and Al. Impact of non-local filtering on 3D reconstruction from tomographic SAR data. In *2017 IEEE International Geoscience and Remote Sensing Symposium*, 2017. 73

G. Fornaro, F. Serafino, and F. Soldovieri. Three-dimensional focusing with multipass SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(3):507–517, March 2003. 14

A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17 – 40, 1976. 89

F. Gini and F. Lombardini. Multibaseline cross-track sar interferometry: a signal processing perspective. *IEEE Aerospace and Electronic Systems Magazine*, 20(8):71–93, Aug 2005. 31

F. Gini, F. Lombardini, and M. Montanari. Layover solution in multibaseline sar interferometry. *IEEE Transactions on Aerospace and Electronic Systems*, 38(4):1344–1356, Oct 2002. 37

Glowinski, R., Marroco, A. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975. 89

S. Guillaso and A. Reigber. Scatterer characterisation using polarimetric sar tomography. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, volume 4, pages 2685–2688, July 2005. 2

B. S. He, H. Yang, and S. L. Wang. Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356, Aug 2000. 95

J. Homer, I. D. Longstaff, and G. Callaghan. High resolution 3-d sar via multi-baseline interferometry. In *IGARSS '96. 1996 International Geoscience and Remote Sensing Symposium*, volume 1, pages 796–798 vol.1, May 1996. 28

Y. Huang, L. Ferro-Famil, and A. Reigber. Under foliage object imaging using SAR tomography and polarimetric spectral estimators. In *8th European Conference on Synthetic Aperture Radar*, pages 1–4, June 2010. XIII, 1, 2

Y. Huang, J. Levy-Vehel, L. Ferro-Famil, and A. Reigber. 3D imaging for underfoliage targets using L-band Multi-Baseline PolInSAR Data and sparse estimation methods. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1460–1463, July 2016. 34

H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, Oct 2003. 113

Vladimir Kolmogorov, Pascal Monasse, and Pauline Tan. Kolmogorov and Zabih's Graph Cuts Stereo Matching Algorithm. *Image Processing On Line*, 4:220–251, 2014. 109

Jong-Sen Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17(1):24 – 32, 1981. 65

Jong-Sen Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24(2):255 – 269, 1983. 66

Jian Li and P. Stoica. Efficient mixed-spectrum estimation with applications to target feature extraction. *IEEE Transactions on Signal Processing*, 44(2):281–295, Feb 1996. 37, 76

F. Lombardini and A. Reigber. Adaptive spectral estimation for multibaseline sar tomography with airborne l-band data. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, volume 3, pages 2014–2016, July 2003. 2

S. G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, Dec 1993. 44

Rahul Mourya, Loïc Denis, Jean-Marie Becker, and Éric Thiébaut. Augmented Lagrangian without alternating directions: Practical algorithms for inverse problems in imaging. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 1205–1209. IEEE, 2015. 90, 91

R. Moses P. Stoica. In *Introduction to Spectral Analysis*. Prentice Hall, 1997. 28, 31

Clément Rambour, Loïc Denis, Florence Tupin, Jean-Marie Nicolas, Hélène Oriot, Laurent Ferro-Famil, and Charles Deledalle. Similarity Criterion for SAR Tomography over Dense Urban Area. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2017. 76

Clément Rambour, Loïc Denis, Florence Tupin, and Hélène Oriot. Introducing spatial regularization in SAR tomography reconstruction. working paper or preprint, May 2018. 5, 85

Clément Rambour, Loïc Denis, Florence Tupin, Hélène Oriot, Yue Huang, and Laurent Ferro-Famil. URBAN SURFACE RECONSTRUCTION IN SAR TOMOGRAPHY BY GRAPH-CUTS. working paper or preprint, 2018. 5, 109

A. Reigber and A. Moreira. First demonstration of airborne SAR tomography using multibaseline L-band data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2142–2152, Sep 2000. XIII, 1, 2, 14

R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, Mar 1986. 2, 31

Y. Shi, X. X. Zhu, and R. Bamler. Non-Local Compressive Sensing Based SAR Tomography. *ArXiv e-prints*, November 2018. 73

P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, July 2004. 46

P. Stoica and K. C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(7):1132–1143, July 1990. 34, 37

P. Stoica, P. Babu, and J. Li. SPICE: A Sparse Covariance-Based Estimation Method for Array Processing. *IEEE Transactions on Signal Processing*, 59(2):629–638, Feb 2011. 41

S. Tebaldini and F. Rocca. Multibaseline Polarimetric SAR Tomography of a Boreal Forest at P- and L-Bands. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1):232–246, Jan 2012. 2

S. Tebaldini, T. Nagler, H. Rott, and A. Heilig. Imaging the internal structure of an alpine glacier via l-band airborne sar tomography. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7197–7209, Dec 2016. XIII, 1

Eric Thiébaut. Optimization issues in blind deconvolution algorithms. In *Astronomical Data Analysis II*, volume 4847, pages 174–184. International Society for Optics and Photonics, 2002. 91

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. 44

M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Transactions on Signal Processing*, 39(5):1110–1121, May 1991. 34

M. Viberg, B. Ottersten, and T. Kailath. Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Transactions on Signal Processing*, 39(11):2436–2449, Nov 1991. 34

Flora Weissgerber and Jean-Marie Nicolas. Joint measurement of height and deformation by radar Interferometry: the example of the Eiffel Tower. In *ONERA-DLR Aerospace Symposium*, Oberpfaffenhofen, Germany, June 2016. 26

W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008. 89

X. X. Zhu and R. Bamler. Tomographic SAR Inversion by L_1 -Norm Regularization ; The Compressive Sensing Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3839–3846, Oct 2010. XIV, 2, 42, 56

X. X. Zhu and R. Bamler. Very High Resolution Spaceborne SAR Tomography in Urban Environment. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12):4296–4308, Dec 2010. 103

X. X. Zhu and R. Bamler. Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic sar. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1):247–258, Jan 2012. 46

X. X. Zhu and R. Bamler. Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation With Application to Spaceborne Tomographic SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1):247–258, Jan 2012. 114

Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997. 91

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Approches structurelles pour la tomographie SAR haute résolution en milieu urbain dense

Mots clés: SAR, Tomographie, Problèmes inverses, parcimonie, graph-cut

Résumé: La tomographie SAR exploite plusieurs acquisitions d'une même zone acquises d'un point de vue légerement différent pour reconstruire la densité complexe de réflectivité au sol. Cette technique d'imagerie s'appuyant sur l'émission et la réception d'ondes électromagnétiques cohérentes, les données analysées sont complexes et l'information spatiale manquante (selon la verticale) est codée dans la phase. De nombreuse méthodes ont pu être proposées pour retrouver cette information. L'utilisation des redondances naturelles à certains milieux n'est toutefois généralement pas exploitée pour améliorer l'estimation tomographique. Cette thèse propose d'utiliser l'information structurelle propre aux structures urbaines pour régulariser les densités de réflecteurs obtenues par cette technique.

Title: Structural Approaches for SAR tomography over dense urban areas

Keywords: SAR, Tomography, Inverse problems, sparsity, graph-cut

Abstract: SAR tomography consists in exploiting multiple images from a same area acquired from a slightly different angle to retrieve the 3-D distribution of the complex reflectivity on the ground. As the transmitted waves are coherent, the desired spatial information (along the vertical axis) is coded in the phase of the pixels. Many methods have been proposed to retrieve this information in the past years. However, the natural redundancies of the scene are generally not exploited to improve the tomographic estimation step. This Ph.D. presents new approaches to regularize the estimated reflectivity density obtained through SAR tomography by exploiting the urban geometrical structures.