Parameter Estimation and Modeling of High Resolution Synthetic Aperutre Radar Data

Matteo Soccorsi

TELECOM ParisTech

Thèse

Présentée pour obtenir le grade de docteur
de l'École Nationale Supérieure des Télécommunications

Spécialité : **Signal et Images**

Matteo SOCCORSI

Sujet :

PARAMETER ESTIMATION AND MODELING OF HIGH RESOLUTION SYNTHETIC APERTURE RADAR DATA

SOUTENUE LE 12 JANVIER 2010 DEVANT LE JURIE COMPOSÉ DE :

Mme.	JARABO AMORES Maria Pilar	Présidente
M.	TROUVÉ Emmanuel	Rapporteur
M.	HINZ Stefan	Rapporteur
M.	NICOLAS Jean-Marie	Examinateur
M.	SOUYRIS Jean-Claude	Examinateur
M.	DATCU Mihai	Directeur de thèse
M.	GLEICH Dusan	Co-directeur de thèse

Tell me where is fancy bred,
Or in the heart or in the head ?
How begot, how nourished ?
Reply, reply.
It is engender'd in the eyes,
With gazing fed ; and fancy dies
In the cradle, where it lies.
Let us all ring fancy's knell ;
I'll begin it - Ding, dong, bell.
Ding, dong, bell.

W. Shakespeare -*The Merchant of Venice*

Abstract

New generations of Synthetic Aperture Radar (SAR) sensors, e.g. TerraSAR-X, provide high resolution images of the Earth. Differently from low resolution images the images are reach of details, i.e. trains, vehicles, cars, etc. never seen before. Furthermore, the speckle is not fully developed when the number of scatterers in the resolution cell decreases and the properties of the system are strongly visible in case of single dominant reflectors. New methods of analysis are required in order to extract the information content in high resolution SAR images.

The thesis is approaching this problematic by statistical modeling and Bayesian inference for complex SAR image analysis.
The Tikhonov regularization method is applied for image restoration because it allows to reformulate the ill-posed image estimation problem into a well-posed problem by the selection of a convex function. It allows to use the required image and prior models and to find the Maximum *A Posteriori* (MAP) estimate solution, exploiting the connection to the Bayesian framework.
Furthermore it allows the optimization to be performed on complex-valued data and to include the system impulse response which has to be included to correctly model the SAR image.

The use of the Rate Distortion for model selection is possible because of the connection between the mutual information and the Occam factor which permits the model selection in the first level of Bayesian inference.
The model selection is applied in order to optimize the parameters of the Model Based Despeckling (MBD) algorithm for image denoising and feature extraction : the optimal average analyzing window and the optimal average model order. The method is a global approach and suits in case of large data sets because of its simplicity and fastness.
The Rate Distortion based model selection is appropriate for the design of image information mining systems.

The Tikhonov regularization shows to be a powerful method for the regularization of complex-valued images. It is recommended in applications where the phase is required, e.g. interferometry, target analysis, because it provides an estimation of the image reflectivity while preserving the phase of the signal.
The use of parametric prior models, e.g. Gauss-Markov Random Field (GMRF), in the optimization function may enable the extraction of texture parameters. The application of Tikhonov approach may lead to the generation of spatial and radiometrically enhanced product, as well as to the specification of spatial descriptors for labeling and classification of the image content.

The results are provided on simulated SAR data and actual TerraSAR-X data.

Résumé (...en français)

Le premier Radar à Synthèse d'Ouverture (RSO) pour l'Observation de la Terre (OT) a été lancé par la NASA en 1978 à bord du satellite SEASAT. Il a cessé de fonctionner quelques mois plus tard mais la technologie RSO et les développements ont continué depuis. Des satellites équipés de capteurs avec une meilleure résolution ont été lancés et de nouvelles missions sont prévues. Le principal avantage du capteur d'imagerie RSO est qu'il est indépendant des conditions météorologiques (par exemple les nuages) et de la lumière du jour parce que c'est un capteur actif. Une vue d'un satellite RSO est montré Figure A. Il s'agit du satellite allemand TerraSAR-X lancé en Juin 2007. D'autres types de capteurs sont également consacrés à l'imagerie de la Terre depuis l'espace, par exemple les capteurs optiques mais ils ont comme inconvénient majeur la limitation due à la couverture nuageuse et la lumière du soleil parce qu'ils sont des capteurs passifs. Les satellites permettent de couvrir la Terre avec un temps de revisite qui dépend de l'orbite. Cela permet d'acquérir des données sur les mêmes régions et d'analyser les séquences d'images. Les variations de la surface de la Terre, les déformations des structures de l'ordre de quelques millimètres peuvent être ainsi détectées et des modèles 3D de la surface de la Terre peuvent être générés grâce aux acquisitions RSO. De 1978 à nos jours, beaucoup de missions et de nouveaux capteurs ont été développés pour l'OT, voir Figure B, comme TanDEM-X dont le lancement est prévu en 2010. Ceci augmente considérablement la quantité de données RSO disponibles et utlisables dans de nombreuses applications différentes et rend la recherche et l'extraction d'informations des archives ardue. En parallèle, l'amélioration de la résolution décuple le contenu informatif. Il rend l'interprétation automatique des images plus difficiles parce que le niveau de détail augmente fortement. L'analyse des données RSO est compliquée parce que l'image cohérente est corrompue par le bruit de speckle, voir Figure C. Afin de faire face à ce speckle, est nécessaire de trouver un modèle correct des données non corrompues et de le supprimer. Une estimation précise est également requise pour le système d'imagerie RSO qui doit être correctement modélisé. La thèse propose une nouvelle méthode pour analyser les données RSO Haute Résolution (HR) afin de faciliter la compréhension, l'interprétation et l'indexation de ces images.

Progrès en modélisation RSO et extraction d'informations

L'analyse d'images RSO en terme d'estimation, de restauration et de débruitage de l'image a été appliquée dans le passé sur des données détectées. Les filtres adaptatifs de Lee (1980), Frost et al. (1982) et Kuan et al. (1985) ont été développés pour estimer la réflectivité de l'image tout en préservant ses caractéristiques. Toutefois, l'estimation de la réflectivité et l'extraction de caractéristiques de l'image (les contours) se sont avérés

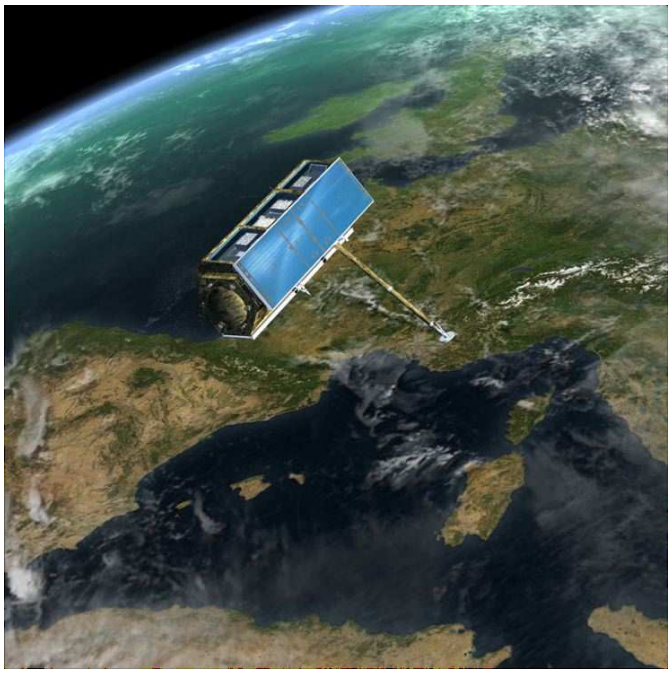


FIGURE A – Vue du satellite TerraSAR-X lancé en Juin 2007. Il fournit des données haute résolution jusqu'à 1 m de résolution de la Terre.

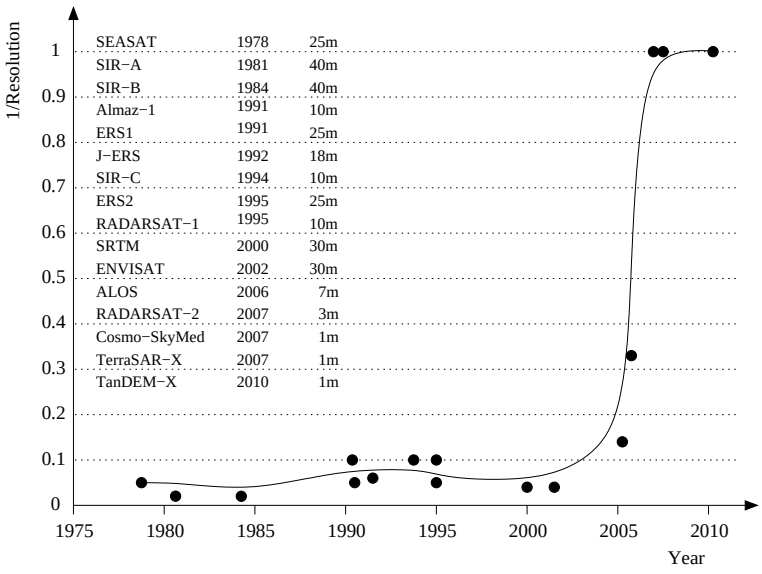


FIGURE B – Résolution spatiale du RSO grandissante avec le temps. Avec l'augmentation de la résolution, la compréhension automatisée des images devient une tâche difficile. Les informations contenues augmentent énormément avec la résolution. La thèse porte sur l'élaboration de nouvelles méthodes pour exploiter pleinement les informations contenues dans ces données haute résolution.

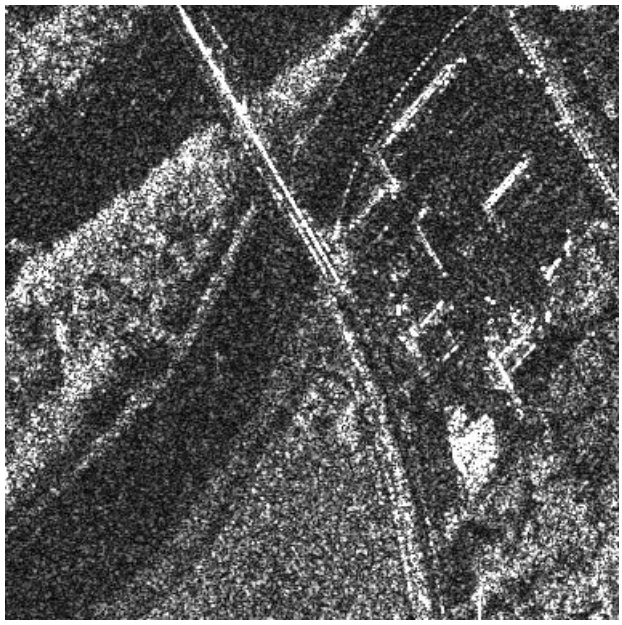


FIGURE C – Exemple d'image RSO Haute Résolution. Des bâtiments, des structures construites par l'homme, de la végétation et de l'eau sont présents dans l'image mais ne sont pas clairement différenciés. La compréhension de l'image vise à faciliter son interprétation.

plus efficaces avec des données complexes (Fjortoft et al., 1999; Fjortoft & Lópes, 2001). À l'exception de quelques contributions (Jakeman & Pusey, 1976; Szajnowski, 1977)
(Sekine & Mao, 1990; Tison et al., 2004), l'effort principal a été de trouver et d'ajuster les modèles statistiques sur les données détectées (amplitude ou intensité) en négligeant la phase, après la transformation non linéaire de la réflectivité complexe du système de coordonnées cartésien au système de coordonnées polaires. Toutefois, la modélisation de données complexes a plusieurs avantages :

- Exploitation de toutes les informations disponibles à pleine résolution.
- Simplicité du modèle de données.
- Comportement du modèle qui représente la signature déterministique ou l'incertitude dans la complexité de la scène.
- C'est la seule façon d'avoir une modélisation précise.

L'inconvénient des données complexes est le faible rapport signal sur bruit. De nombreux modèles statistiques, empiriques ou théoriques, ont été proposés pour traiter le bruit de speckle. La distribution Gamma a été présentée comme un modèle pour une fonction de densité de probabilité de l'intensité RSO multilookée où l'amplitude correspond à une distribution de Nakagami. Ils sont une généralisation des exponentielles négatives et des distributions de Rayleigh respectivement. La distribution K (Jakeman & Pusey, 1976, 1978; Oliver, 1984; Jao, 1984) est obtenue pour l'intensité du signal en supposant que la population de diffuseurs dans la cellule de résolution peut être contrôlée par un processus de migration naissance-mort. Dans l'hypothèse que le bruit et le signal suivent une distribution Gamma (Oliver, 1991; Oliver & Quegan, 2004a), le modèle produit se ramène à une distribution K pour l'intensité. La distribution de Nakagami-Rice a été proposée pour modéliser les statistiques RSO, en présence d'un réflecteur unique et

fort dans le clutter homogène (Dana & Knepp, 1986; Tison et al., 2004). Les distributions Gaussiennes Inverses ont aussi été utilisées pour modéliser les statistiques d'amplitude dans Frery et al. (1997), Muller & Pac (1999) et Eltoft (2003). Dans Kuruoglu & Zerubia (2004), un modèle de Rayleigh *heavy-tailed* est présenté sous l'hypothèse que les parties réelles et imaginaires du signal rétrodiffusé sont des variables aléatoires conjointement symétrique-α-stable. Des distributions gaussiennes généralisées sont supposées pour les parties réelles et imaginaires dans Moser et al. (2006). La distribution $\mathcal{G}$ pour l'amplitude, présentée dans Frery et al. (1997), est le résultat d'un modèle multiplicatif assumant une distribution Nakagami pour le bruit de speckle et une distribution gaussienne inverse généralisée pour le signal. Dans Muller & Pac (1999), un cas particulier du modèle $\mathcal{G}$, appelée *branche harmonique* $\mathcal{G}^h$, est proposé, tandis que la distribution $\mathcal{G}^0$ est prouvée être équivalente à une pdf de Fisher dans Tison et al. (2004).

En plus des modèles *théoriques* ou en partie théoriques mentionnés ci-dessus, plusieurs modèles empiriques ont été utilisés pour caractériser les statistiques de l'amplitude (ou intensité) des données RSO, tels que le log-normal (Szajnowski, 1977), Weibull (Sekine & Mao, 1990) et Pearson (Delignon et al., 1997).

La famille des champs aléatoires de Gauss-Markov (Chelappa et al., 1985) a été utilisée avec succès dans la vision par ordinateur pour la génération de texture et la compression. Dans Walessa & Datcu (2000), ils ont été utilisés pour l'estimation du Maximum ou À Postériori (MAP) comme prior dans le cadre de l'inférence bayésienne avec la distribution Gamma.

Le problème direct du calcul de la réponse du système d'imagerie à partir d'une image donnée est souvent supposé connu et bien posé. Le modèle habituel est une convolution par un noyau donné ou fonction d'étalement du point qui, dans la plupart des cas, implique que le problème inverse du calcul de l'image réelle à partir des observations est mal posé. Un principe général pour faire face à l'instabilité du problème inverse est la régularisation, qui consiste principalement à restreindre l'ensemble des solutions admissibles et à inclure des informations a priori (non négativité, la régularité, l'existence de contours, etc) dans la formulation du problème. Une modélisation précise du système d'imagerie et un bon choix de la régularisation sont essentiels pour un processus de restauration d'images satisfaisant. Tikhonov (Tikhonov & Arsenin, 1977) a introduit des méthodes de régularisation pour les problèmes déterministiques en introduisant des fonctions stabilisantes qui jouent un rôle analogue a la distribution log prior de l'estimation du MAP. La régularisation des problèmes mal posés a été l'objet de nombreuses recherches, par exemple Horn & Rhunck (1981) pour déterminer les flux optiques, Ikeuchi & Horn (1981) pour la reconstruction de la forme à partir de l'ombre, Torre & Poggio (1986) pour la détection des contours et Marroquin et al. (1987) pour la vision par ordinateur. Lorsque la fonction n'est pas convexe, il est nécessaire d'adopter une méthode de recuit simulé pour la minimiser. La fonction que sous considérons est convexe, ainsi les méthodes de descente de gradient peuvent être appliquées.

En Çetin & Karl (2001), la régularisation est utilisée afin de développer une méthode de formation d'image améliorée pour SpotLight (SL) pour le RSO aéroporté. En exploitant un mode d'imagerie similaire Çetin & Karl (2001) et Pan & Reeves (2006), nous introduisons une vraisemblance complexe, qui permet de prendre en compte la nature complexe de la fonction de transfert du système. Cette approche est différente de celle suggérée dans Çetin & Karl (2001), où les auteurs proposent un modèle pour la formation des images. Nous utilisons comme prior la fonction Huber-Markov qui se révèle être un puissant modèle pour la préservation des contours dans Pan & Reeves (2006).

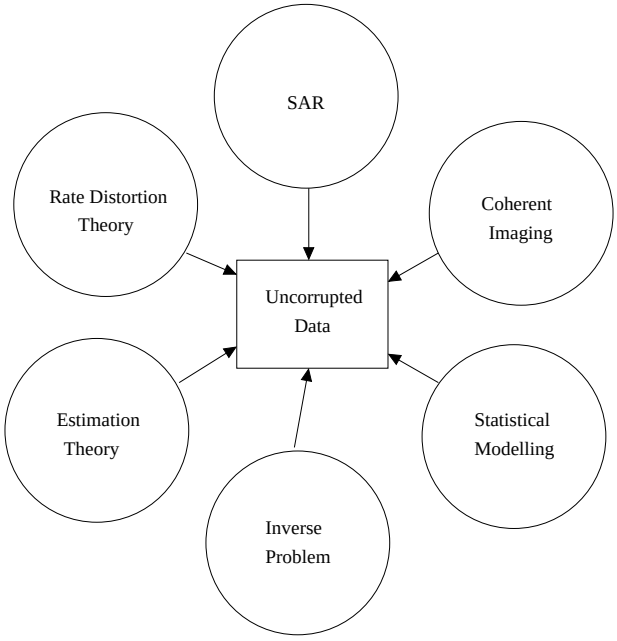


FIGURE D – Les données non corrompues sont estimées en considérant le problème comme un problème inverse par modélisation statistique du système RSO et le processus de formation d'image cohérente. Le taux de distorsion, dans le cadre de la théorie de l'information, est utilisé pour la sélection du modèle.

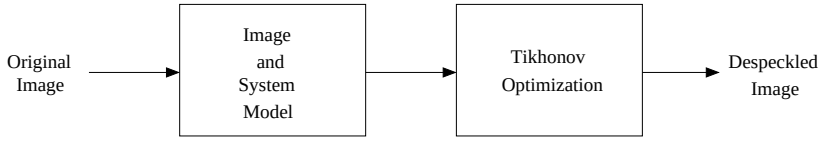


FIGURE E – Schéma de la méthode d'optimisation de Tikhonov. Le modèle de l'image et du système sont choisis, et l'optimisation est effectuée. La sortie est l'image dont le speckle a été filtré.

Le problème *mal posé* de la restauration d'image devient un problème *bien posé* en choisissant un prior convexe qui garantit que la solution existe, qu'elle est unique et dépend continûment des données. La dernière condition, appelée stabilité, garantit que les petites perturbations dans les données ne provoquent pas de changements dramatiques dans la solution. Eventuellement une sélection de modèles est effectuée afin d'estimer la fenêtre moyenne optimale d'analyse et l'ordre du processus autorégressif (AR) par le biais du taux de distorsion. Auparavant, un travail similaire n'a été effectué que par Pesaresi (1996) pour le RSO basse résolution et par Soccorsi et al. (2006) pour les données optiques.

Contribution de la thèse

La thèse porte sur l'extraction d'informations et l'amélioration des données RSO de un mètre de résolution visant à fournir des meilleurs descripteurs de contenu pour la compréhension des scènes et la reconnaissance de cibles, pour des produits améliorés ra-

diométriquement et spatialement. Pour atteindre cet objectif, la thèse approche le problème de la modélisation des images RSO et propose une nouvelle solution fondée sur l'estimation du problème inverse pour l'extraction d'information (Figure D). Le problème de la sélection du modèle est géré par le taux de distorsion, en raison de sa correspondance avec le cadre de l'inférence bayésienne. Nous commençons l'analyse avec l'extension de la famille de champs aléatoires de Gauss-Markov linéaires a des données à valeurs complexes, qui s'applique aux variables aléatoires à valeurs complexes : la distribution normale à plusieurs variables complexes et le modèle paramétriques des champs aléatoires de Gauss-Markov en cas de variables aléatoires correctes et incorrectes. Le second modèle est utilisé pour l'estimation des paramètres dans le deuxième niveau d'inférence bayésienne. Ainsi, nous étendons l'analyse à l'approche bayésienne complète, y compris les premier et deuxième niveaux d'inférence bayésienne, c'est-à-dire à la sélection du modèle et à l'estimation des paramètres. Ce résultat est obtenu par la méthode d'estimation du MAP qui a été appliquée pour supprimer le speckle et extraire des caractéristiques par Walessa & Datcu (2000). La méthode d'estimation MAP existe dans le domaine complexe principalement dans les applications en tomographie (Pascazio & Ferraiuolo, 2003). Cette approche permet de reconstruire les parties réelle et imaginaire, mais traite avec du bruit additif et, par conséquent, elle ne supprime pas le speckle dans le cas des images RSO.

La méthode proposée (Soccorsi et al., 2009) est une régularisation de Tikhonov dans le domaine complexe, voir le diagramme Figure E. Le speckle est traité comme un processus aléatoire à valeurs réelles. L'approche dans le domaine complexe permet de gérer la formation de l'image cohérente comme information ou comme incertitude dans le cas de structures ou de textures de la scène. La nouveauté repose dans l'utilisation des données complexes ce qui permet d'inclure la fonction de transfert du système. La méthode est équivalente à l'estimation bayésienne MAP. La méthode, qui fonctionne avec des données à valeur complexe, est démontrée être meilleure que les méthodes classiques de suppression du speckle (par exemple, le filtre de Lee, MBD) sur les données détectées. Selon le prior adopté, elle permet de débruiter l'image et d'estimer les paramètres de texture.

Dans le contexte de l'optimisation des paramètres pour l'extraction de caractéristiques, la fenêtre d'analyse optimale (moyenne) (Pesaresi, 1996) et l'ordre optimal (moyen) du processus d'auto-régression sont estimés à l'aide du taux de distorsion (Soccorsi & Datcu, 2008). Cela confirme que le taux de distorsion est une bonne méthode basée sur l'entropie pour la sélection de modèle.

Les résultats peuvent être utilisés dans les systèmes d'exploration d'information d'images pour la compréhension de l'image.

Dans le chapitre 1, les bases pertinantes du système d'imagerie RSO sont présentées. Au début, le radar RSO est décrit ainsi que l'acquisition des données et la formation des images. Ensuite, la représentation de l'information et les statistiques des données sont décrites. Le chapitre se termine par quelques considérations sur le speckle, sa simulation et les modèles d'intensité pour sa réduction.

Le problème est de trouver un modèle pour la réponse impulsionnelle du système et pour l'image. Ceci est adressé comme un problème de l'estimation et la sélection de modèles qui est présenté dans le chapitre 2. Il contient les bases théoriques appliquées par la suite à l'estimation de l'image. Il y a une introduction qui s'étend des processus stochastiques aux MRF en passant par les chaînes de Markov. Le chapitre 2 contient un modèle de champs aléatoires de Gauss-Markov, les méthodes d'estimation des paramètres déterministiques et stochastiques, l'inférence bayésienne et il se termine par une

nouvelle méthode théorique de sélection du modèle réalisée par le taux de distorsion.

Ce chapitre présente les modèles d'estimation du MAP et régularisation de Tikhonov. Il commence par un aperçu du MBD, qui a été notre méthode de référence pour supprimer le speckle de l'image et pour extraire des caractéristiques. Il se poursuit par une extension de l'estimation du MAP dans le domaine complexe qui fournit les cannaux complexes reconstruits mais pas une image sans speckle.

Ensuite, la famille de modèles de champs aléatoires de Gauss-Markov étendue au domaine complexe est présentée. Le modèle linéaire à valeurs complexes intègre la fonction de transfert du système, ainsi il est possible de modéliser les paramètres et de distinguer les textures corrélées.

L'estimation MAP de Tikhonov a été développée et testée avec un modèle linéaire à valeurs complexes, en tant que modèle de données, et le modèle de Huber-Markov comme fonction prior.

Ce chapitre présente les résultats expérimentaux obtenus avec les analyses et les modèles présentés au chapitre 3. La sélection de modèle par le taux de distorsion appliqué à la taille de la fenêtre d'analyse et à l'ordre du modèle est présentée en premier. Ensuite, les résultats pour la régularisation statistique d'images complexes et la modélisation de données à valeur complexe de champs aléatoires de Gauss-Markov sont fournis, suivis par la présentation des principaux résultats de l'optimisation de type Tikhonov pour la suppression du speckle. La méthode proposée est comparée avec les filtres adaptatifs les plus connus.

Résultats

Estimation de la taille moyenne optimale de la fenêtre d'analyse

L'algorithme MBD (Walessa & Datcu, 2000) a été exécuté pour six différentes tailles de la fenêtre d'analyse de 11×11 à 61×61 pixels avec un pas de dix pixels de chaque côté. L'ordre du modèle a été fixé à quatre dans cette expérience. L'ordre d'entrée des modèles pour l'analyse est trois ce qui mène à un espace de cardinalité huit avec : six paramètres spatiaux, la norme des paramètres et la variance du modèle. Ensuite, l'espace des carctéristiques est mis en cluster de 2 à 128 classes, ce qui signifie 7 étapes pour la version dyadique de k-means. Le nombre d'itérations a été fixé expérimentalement à 30. La distorsion globale a été calculée à chaque itération. Les courbes de la distorsion en fonction du nombre de clusters sont présentées dans la Figure F : ils ont un comportement asymptotique et les courbes sont décalées vers le bas quand la taille de la fenêtre augmente.

Les valeurs de la distorsion sont indiquées dans le tableau A ainsi que le nombre équivalent de vues de l'image avec le speckle filtré qui est proportionnel à l'inverse du bruit. Les courbes de distorsion montrent un minimum pour la fenêtre d'analyse de taille 61×61 pixels. Par conséquent, comme les courbes se rapprochent à partir d'une taille de 41 × 41 pixels, ce choix de taille est un bon compromis entre la complexité et la distorsion globale. Le nombre équivalent de vues de l'image avec le speckle filtré indiqué dans le tableau A confirme que la suppression optimale du speckle est atteinte pour une fenêtre d'analyse de taille 41×41. La méthode, appliquée sur les données RSO, montre un comportement différent des courbes de distorsion dans la gamme des tailles de fenêtre considérée dans l'analyse, par rapport à Soccorsi et al. (2006) où elle a été appliquée avec succès pour évaluer la taille moyenne optimale de la fenêtre d'analyse sur des données optiques. Cette

différence pourrait dépendre de plusieurs raisons : soit l'extraction de caractéristiques ne code pas correctement la texture des données, soit le minimum de distorsion est atteint pour une fenêtre d'analyse de plus grande taille. Toutefois, le choix d'une fenêtre plus grande rend le temps de calcul beaucoup trop long, donc il ne peut pas être considéré comme une solution raisonnable. Une analyse plus approfondie est nécessaire afin de comprendre le comportement asymptotique de la distorsion qui à un moment est censée augmenter avec la taille de la fenêtre d'analyse.

Estimation de l'ordre du modèle moyen optimal

Pour l'expérience suivante, le MBD a été exécuté avec différents ordres du modèle et la distorsion a été calculée en tenant également compte de la distorsion introduite par le codage source. Les courbes de distorsion sont montrées figure G et les valeurs de distorsion dans le tableau B. Le minimum de la courbe de distorsion est atteint pour les modèles d'ordre 4, 5 et 6 et le modèle d'ordre 4 est le meilleur compromis en terme de complexité. En outre, le nombre équivalent de vues, montré dans le tableau B, atteint son maximum pour le modèle d'ordre 4, qui correspond à la meilleure élimination du bruit dans l'image.
Ce résultat global est comparé avec la carte de l'ordre du modèle optimal obtenue par sélection de modèle par le biais l'information de Fisher. Bien que le moyennage de la carte de l'ordre du modèle n'ait pas de sens, il est permis mathématiquement et on peut ainsi obtenir une estimation de la valeur attendue pour le modèle à savoir 4,2. La valeur est compatible avec l'analyse globale effectuée par le taux de distorsion et il est une nouvelle confirmation de la validité de la méthode.

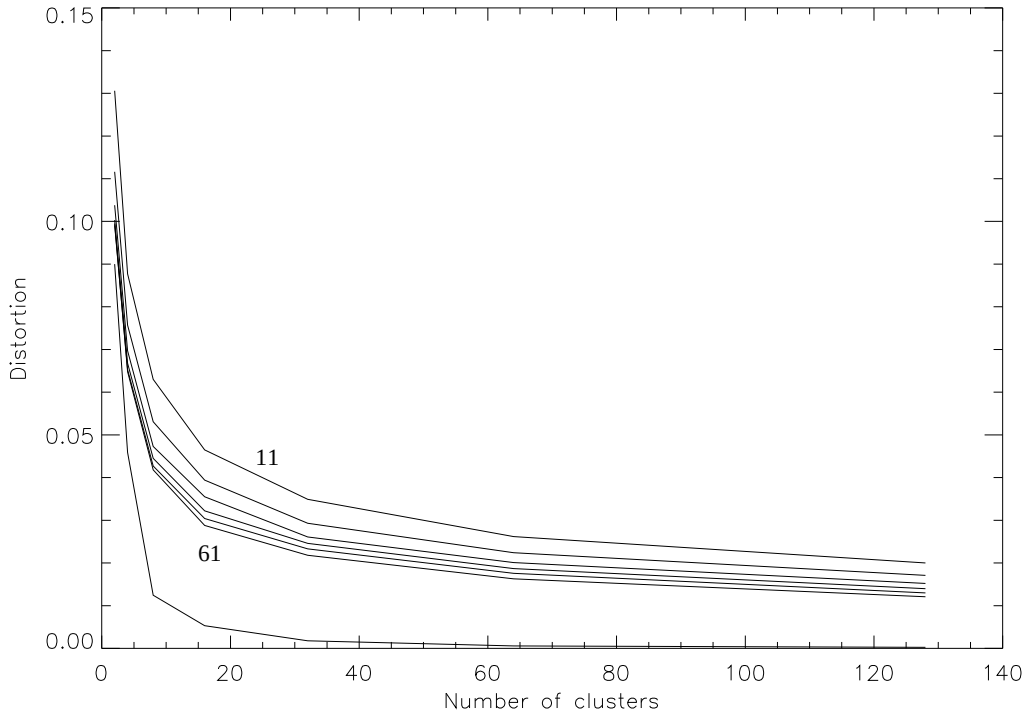


FIGURE F – Courbes de distorsion pour les fenêtres de tailles 11×11, 21×21, 31×31, 41×41, 51×51 and 61×61 respectivement, de haut en bas.

AW size	Number of clusters 2	4	8	16	32	64	128	ENL
11×11	13.06	8.77	6.30	4.65	3.49	2.62	2.00	67.94
21×21	11.16	7.57	5.31	3.94	2.93	2.24	1.71	97.90
31×31	10.38	6.97	4.73	3.55	2.61	2.01	1.52	101.2
41×41	10.03	6.68	4.45	3.22	2.46	1.87	1.40	104.7
51×51	9.91	6.53	4.27	3.04	2.33	1.76	1.30	99.53
61×61	9.95	6.48	4.18	2.88	2.18	1.63	1.21	92.78

TABLE A – Valeurs de distorsion ($\times 10^{-2}$) et nombre équivalent de vues de l'image dont le speckle a été filtré par des fenêtres d'analyse de 6 tailles différentes. Le meilleur compromis entre la complexité et la distorsion est de choisir un fenêtre de taille 41×41 pixels (en rouge) qui a le maximum ENL.

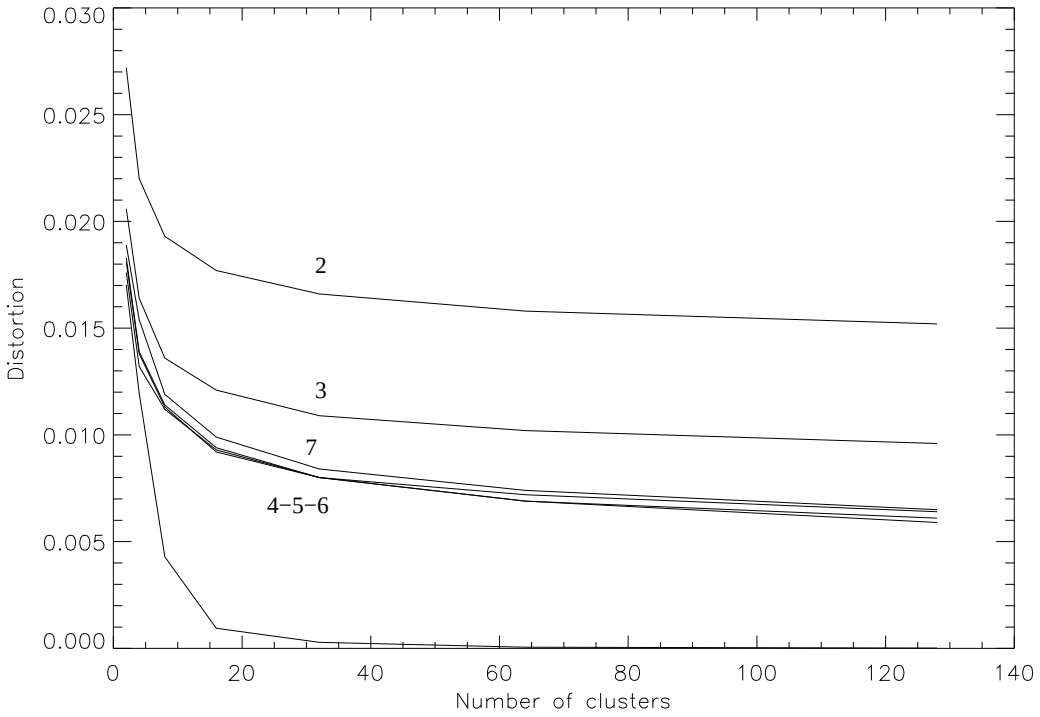


FIGURE G – Courbes de distorsion pour les modèles d'ordre 2, 3, 4, 5, 6 et 7. Les courbes atteignent un minimum pour les modèles d'ordre 4, 5 et 6.

Model Order	Number of clusters 2	4	8	16	32	64	128	ENL
2	2.72	2.20	1.93	1.77	1.66	1.58	1.52	73.03
3	2.06	1.64	1.36	1.21	1.09	1.02	0.96	101.9
4	1.81	1.38	1.13	0.92	0.80	0.72	0.64	104.6
5	1.76	1.32	1.12	0.93	0.80	0.69	0.61	47.95
6	1.83	1.39	1.14	0.94	0.80	0.69	0.59	39.74
7	1.89	1.54	1.19	0.99	0.84	0.74	0.65	47.96

TABLE B – Valeurs de distorsion ($\times 10^{-2}$) et nombre équivalent de vues de l'image dont le speckle a été filtré pour les différents ordres de modèles. Les modèles d'ordre 4, 5 et 6 ont des valeurs de distorsion comparable. Le meilleur choix en terme de complexité est le modèle d'ordre 4 (en rouge).

Suppression du speckle basé sur la régularisation de Tikhonov

La méthode de régularisation de Tikhonov avec le prior des champs aléatoires de Huber-Markov HMRF a été comparée aux filtres adaptatifs les plus connus : les filtres améliorés de Lee, de Frost, Kuan, Gamma et le MBD. La première expérience a été réalisée sur quatre textures synthétiques GMRF, la seconde sur quatre images de texture Brodatz, la troisième expérience a été réalisée avec une image optique Quick Bird QB et la dernière, avec des données TerraSAR-X haute-résolution.
Les méthodes appliquées pour la comparaison des filtres sont d'ordre qualitatif et quan-

titatif. Les mesures comparatives prises en compte sont les suivantes

- préservation de la valeur moyenne.
- erreur quadratique moyenne $\mathrm{E}\{|\hat{\mathbf{x}} - \mathbf{x}|^2\}$
- indice de similarité structurelle (Zhou Wang & Simoncelli, n.d.) $\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}})$.
- nombre équivalent de vues (ENL) de l'image avec les speckle filtré $\hat{\mathbf{x}}$.
- valeur moyenne de l'intensité du bruit $\mathrm{E}\{\mathbf{y}/\hat{\mathbf{x}}\}$.
- nombre équivalent de vues du rapport d'images $\mathbf{y}/\hat{\mathbf{x}}$.
- comparaison visuelle du rapport d'images $\mathbf{y}/\hat{\mathbf{x}}$.

Toutes les méthodes ne sont pas applicables à toutes les expériences. Par exemple, la définition du nombre équivalent de vues au cas des textures n'a pas de sens et certaines méthodes ne sont pas utilisables sans image de référence. Ainsi, ellse ne sont pas adaptées au cas des expériences sur les données RSO réelles.

Le filtre avec MSE le plus bas est préférable. D'autre part, le MSE montre de faibles performances pour les images (Wang & Bovik, 2009), donc le SSIM est également utilisé. Il s'agit d'une mesure de la qualité de l'image par rapport à une autre qui est censée avoir une qualité parfaite. Il est calculé comme suit (Wang et al., 2004; Channappayya et al., 2008)

$$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + c_2)} \tag{1}$$

où $\mathbf{x}$ est l'image de référence et $\hat{\mathbf{x}}$ est l'image avec le speckle filtré, μ et σ sont respectivement la moyenne et la variance, c_1 et c_2 sont deux constantes introduites afin d'éviter les problèmes numériques lorsque les valeurs de la moyenne et de la variance sont proches de zéro. Le SSIM tend vers un lorsque l'image sélectionnée se rapproche de celle de référence. Par conséquent, le filtre dont le SSIM est le plus proche de un est préféré.

La valeur moyenne de l'image doit être préservée par le filtre, car l'espérance du signal ne doit pas être modifiée par le traitement par le filtre.

Le nombre équivalent de vues de l'image avec le speckle filtré est calculé en utilisant

$$ENL = \frac{\mathrm{E}\{\mathrm{x}\}}{\mathrm{E}(\mathrm{x} - \mathrm{Ex})^2} \tag{2}$$

où x représente l'intensité du signal. C'est également une méthode classique pour la comparaison des filtres, mais elle ne peut être appliquée que si une zone uniforme suffisamment large est disponible dans les données.

L'espérance du rapport d'images doit être un pour une intensité du bruit distribuée selon une exponentielle négative avec une variance unitaire, donc le filtre qui se rapproche le plus de cette valeur a de meilleures performances que les autres.

Le nombre équivalent de vues du rapport d'images doit être égal à un dans le cas idéal.

Le rapport d'images est montré pour une comparaison visuelle qualitative. Dans le cas où les contours ne sont pas bien conservés le rapport montre des motifs dus à la procédure de filtrage non optimale.

La condition de convergence est choisie comme étant le minimum MSE atteint et la valeur initiale du paramètre k est fixée à 0,9. Le seuil de τ de la fonction de Huber a été fixé comme mode de l'histogramme des données. Le nombre d'itérations est choisi comme condition de convergence dans le cas de données RSO réelles, car l'image de référence pour le calcul du MSE n'est pas disponible.

Le choix de traiter et de tester des images avec ENL = 1 est un point critique, car il met à l'épreuve les performances des filtres, mais il permet de comparer le filtre développé

qui travaille sur les données complexes avec les autres filtres qui fonctionnent sur des données détectées. La différence sur les données repose sur la transformation non linéaire des coordonnées cartésiennes aux polaires. Par la suite, les termes rapport d'images et image speckle sont utilisés comme termes alternatifs. Par souci de concision, seuls les résultats sur les textures Brodatz et sur les images RSO seront effectivement présentés.

Les textures Brodatz

La deuxième expérience a été réalisée sur des images optiques appartenant à l'archive des textures Brodatz (Randen, 1997). Les images sélectionnées, les originales et celles qui sont corrompues par le speckle sont montrées figure H1. Elles ont été choisies afin d'avoir une sélection de différents types de textures. La texture 1 représente une structure linéaire, réalisée par des traits, la texture 2 est une structure régulière de chevrons, la troisième est une structure ronde formée par des pierres et la dernière est une texture de marbre. Les images Brodatz sont numérotées en partant d'en bas à gauche dans le sens inverse des aiguilles d'une montre. Chaque image a une taille de 256×256 pixels. Les images avec le speckle filtré sont montrées dans la figure H. Les rapports d'images pour une comparaison qualitative sont présentés figure I, alors que les mesures pour la comparaison des filtres sont présentées dans le tableau 12.
En comparant les images avec le speckle filtré aux images originales, il semble que les filtres améliorés de Lee et de Frost, le MBD et la méthode proposée donnent des résultats similaires mais meilleurs que les filtres Kuan et Gamma. La méthode proposée est à la deuxième place, après le MBD, en comparant les motifs des rapport d'images, voir les figures I5 et I6. En se référant au tableau 12, tous les filtres ont raisonnablement préservé la valeur moyenne. La méthode proposée présente le meilleur MSE pour la quatrième image, alors qu'elle obtient des résultats similaires au filtre de Lee amélioré pour les images deux et trois. Le filtre HMRF est en troisième position après les filtres améliorés de Lee et de Frost dans le cas de l'image un. Ils surpassent les autres filtres en terme de SSIM pour la première image, tandis que le filtre Kuan montre le meilleur SSIM mesure de l'indice dans le cas des images deux et trois. Le filtre amélioré Frost donne de meilleurs résultats en terme d'indice SSIM dans le cas de l'image quatre. La méthode proposée a le deuxième meilleur indice SSIM dans le cas de l'image deux. Le filtre MBD a de bons résultats pour la moyenne de l'image de speckle, qui est proche de un pour toutes les images. Le filtre amélioré de Lee a des performances légèrement meilleures seulement dans le cas de l'image deux. La méthode proposée présente une valeur moyenne acceptable pour le rapport d'images mais elle tend à surestimer l'espérance statistique du speckle dans toutes les images. Le nombre équivalent de vues est meilleur pour le filtrage amélioré de Lee, sauf pour l'image quatre où le filtre Gamma a des performances légèrement meilleures.
La méthode proposée converge après deux itérations dans le cas de l'image une, deux et trois et après quatre itérations pour la quatrième image.

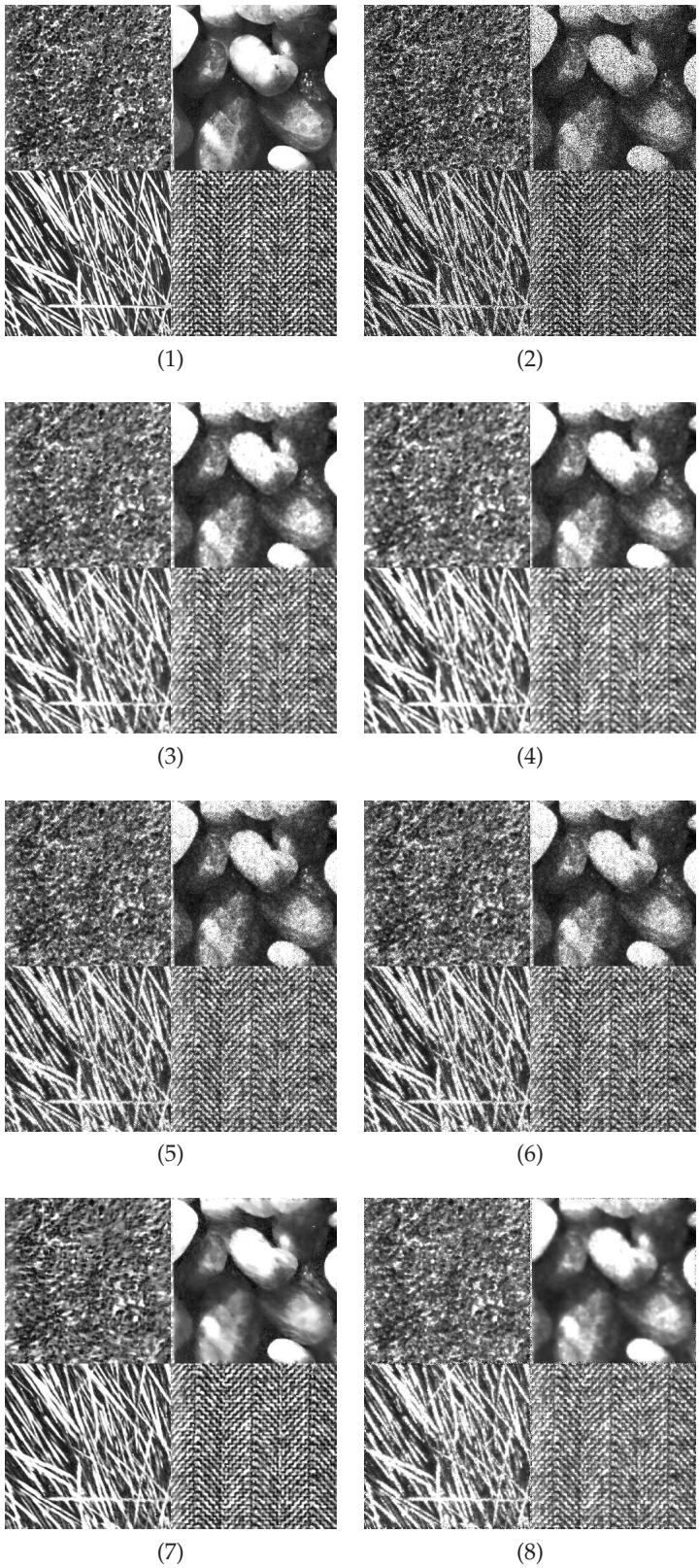


FIGURE H – Expériences avec les textures Brodatz corrompues par le speckle avec ENL = 1. Images originales (1), images bruitées (2), avec speckle filtré par les filtres améliorés de Lee (3), de Frost (4), Kuan (5), Gamma (6), MBD (7) et la méthode proposée (8).

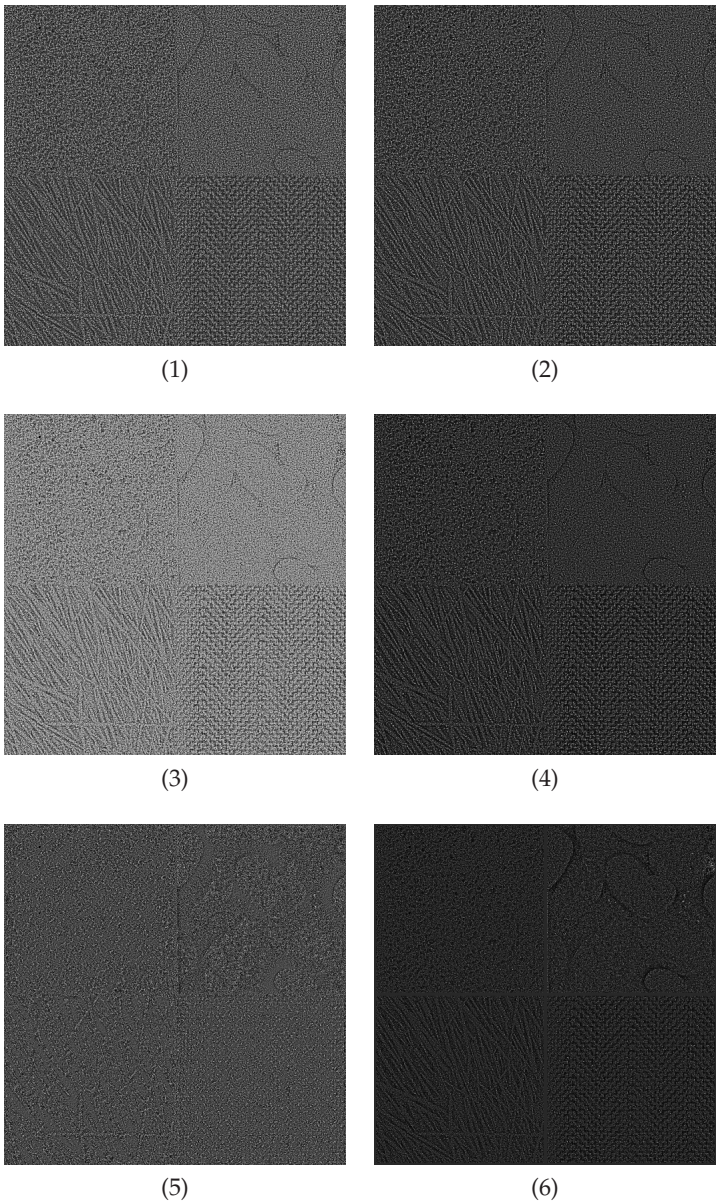


FIGURE I – Expériences avec les textures Brodatz corrompues par le bruit de speckle avec ENL = 1. Rapports d'images : les filtres améliorés de Lee (1), de Frost (2), Kuan (3), Gamma (4), MBD (5) et la méthode proposée (6).

B1 $\mu = 127.032$	E$\{\hat{\mathbf{x}}\}$	mse	SSMI	E$\{\mathbf{y}/\hat{\mathbf{x}}\}$	ENL$\{\mathbf{y}/\hat{\mathbf{x}}\}$
e. Lee	125.31	1611.66	0.55	1.08	1.07
e. Frost	126.35	1696.21	0.57	1.23	0.67
Kuan	128.08	2705.41	0.45	0.88	2.50
Gamma	128.45	2033.89	0.51	1.27	0.56
MBD	131.04	2057.95	0.30	1.04	1.32
HMRF	128.05	1940.59	0.33	1.18	0.82
B2 $\mu = 109.550$					
e. Lee	192.07	1498.40	0.27	1.03	1.03
e. Frost	107.47	1580.76	0.25	1.25	0.58
Kuan	109.02	1896.36	0.55	0.86	2.16
Gamma	109.64	1961.16	0.18	1.32	0.46
MBD	108.96	1703.58	0.31	1.05	1.27
HMRF	108.99	1532.28	0.38	1.11	0.65
B3 $\mu = 90.2339$					
e. Lee	88.91	884.361	0.39	1.11	1.03
e. Frost	89.20	923.286	0.53	1.27	0.68
Kuan	89.65	1223.63	0.55	0.90	2.64
Gamma	89.82	1035.55	0.48	1.31	0.57
MBD	91.23	1068.50	0.29	1.04	1.23
HMRF	89.61	895.82	0.35	1.14	0.84
B4 $\mu = 104.012$					
e. Lee	100.828	828.234	0.63	1.20	1.13
e. Frost	100.905	816.941	0.65	1.26	0.95
Kuan	101.024	1738.90	0.45	0.94	3.32
Gamma	101.071	832.590	0.62	1.2	0.91
MBD	106.331	1097.71	0.55	1.04	1.23
HMRF	100.91	720.26	0.23	1.24	0.86

TABLE C – Mesures pour la comparaison des textures Brodatz avec speckle filtré. En partant d'en bas à gauche dans le sens contraire de celui des aiguilles d'une montre, en référence à la Figure H1 : B1, B2, B3, B4.

Images RSO

La dernière expérience a été effectuée sur des données réelles RSO. Concrètement les analyses ont été faites sur une acquisition TerraSAR-X mode HR SL, dont la résolution est $\sim$ 1.5 m en range et $\sim$ 1,1 m en azimuth. La région traitée a une taille de 1024 $\times$ 1024 pixels et est montrée figure J. Les images avec speckle filtré sont montrées figure K et les raport d'images figure L. Les mesures pour la comparaison quantitative des filtres sont présentées dans le tableau 14. Dans l'image réelle le speckle est corrélé, ce qui n'était pas le cas dans l'image simulée.

Dans la figure K, on remarque que la performance des filtres améliorés de Lee, de Frost et des filtres Gamma est similaire. Or le Kuan ne supprime pas suffisament le bruit. Le résultats du MBD et de la méthode proposée montrent une meilleure élimination du bruit.

En ce qui concerne les rapports d'images, toutes montrent des motifs, ceux-ci étant

légèrement inférieurs pour le MBD. Dans le cas des filtres améliorés de Lee et de Frost et Gamma des performances similaires sont confirmées, ainsi que la mauvaise performance du filtre de Kuan. Néanmoins le MBD montre des artefacts dans la zone uniforme et le long des bords. La méthode proposée montre la région la plus uniforme, mais l'effet de flou des contours est aussi visible.

Les seules mesures quantitatives comparatives possibles sont celles qui ne demandent pas d'image de référence, puisque dans ce cas elle n'est pas disponible. En conséquence, on ne peut comparer que la valeur moyenne de l'image correspondante, de l'ENL du speckle et de l'ENL de l'image avec speckle filtré.

La meilleure moyenne des raports d'image est obtenue parle le MBD, suivi par le Kuan, les filtres améliorés de Lee et de Frost, le Gamma et, finalement, la méthode proposée.

Le meilleur ENL du bruit de speckle est donnée par le filtre Gamma, le filtre de Frost amélior, le HMRF, le Lee amélioré et, enfin, le MBD. Le filtre de Kuan a notablement surestimé.

Le meilleur ENL sur l'image avec speckle filtré a été obtenu par la méthode proposée après 5 itérations. Le MBD est à la deuxième place. Le Frost amélioré et Kuan ont une performance similaire, suivis par le Lee amélioré. La pire performance est obtenue par le filtre Gamma.

	$E\{\mathbf{y}/\hat{\mathbf{x}}\}$	$ENL\{\mathbf{y}/\hat{\mathbf{x}}\}$	$ENL\{\hat{\mathbf{x}}\}$
e. Lee	1.17	1.26	7.91
e. Frost	1.21	1.08	8.07
Kuan	0.94	3.56	8.12
Gamma	1.22	0.99	2.60
MBD	1.02	1.52	17.16
HMRF	1.29	0.75	21.71

TABLE D – Mesures pour la comparaison quantitative des filtres sur l'image RSO réelle présentée dans la figure J.

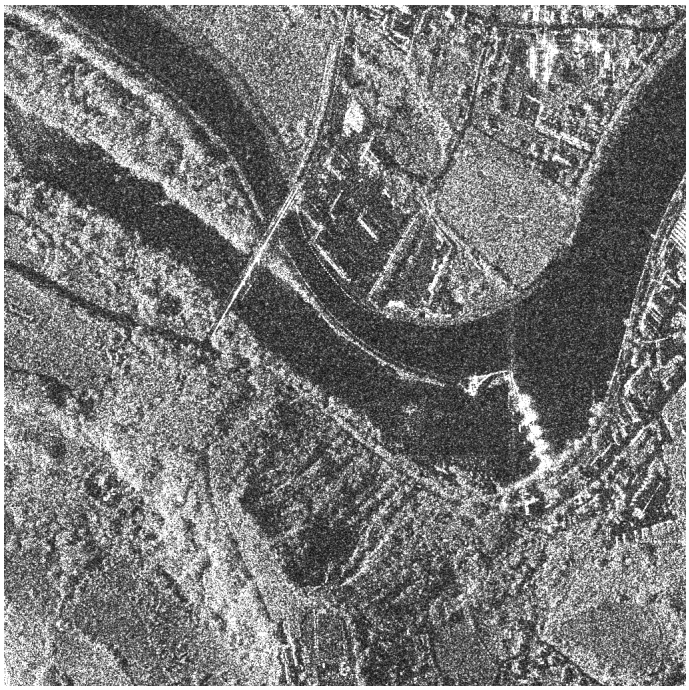
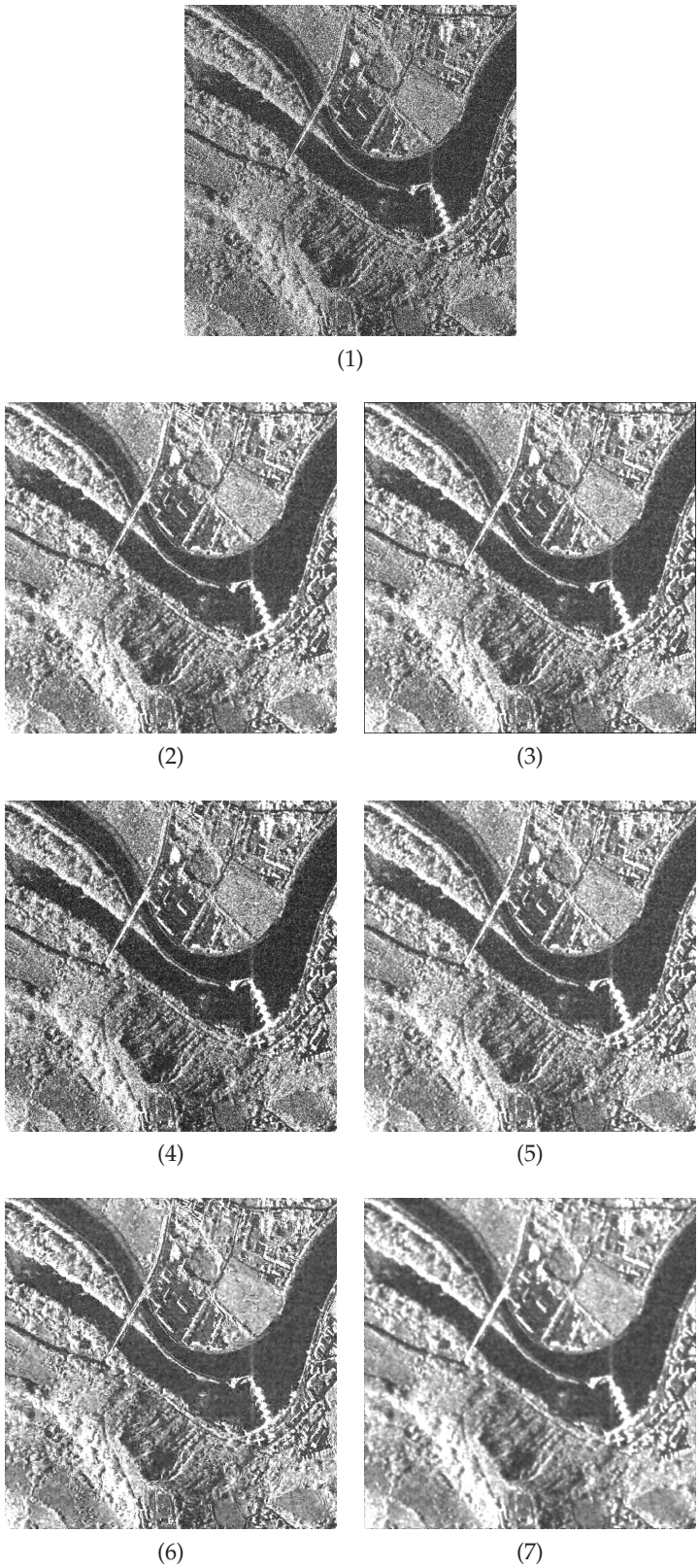


FIGURE J – Expérience avec une image RSO réelle TerraSAR-X HR SL. Polarisation HH, orbite descendante, résolution au sol $\sim$ 1.5 m, résolution azimut $\sim$ 1.1 m. Maribor, en Slovénie, le 29 Octobre 2008, 5 :01 :06 UTC.



(1) (2) (3) (4) (5) (6) (7)

FIGURE K – Expérience avec une réelle RSO image. Image originale (1), avec speckle filtré par les filtres améliorés de Lee (2), de Frost (3), Kuan (4), Gamma (5), le MBD (6) et la méthode proposée (7).

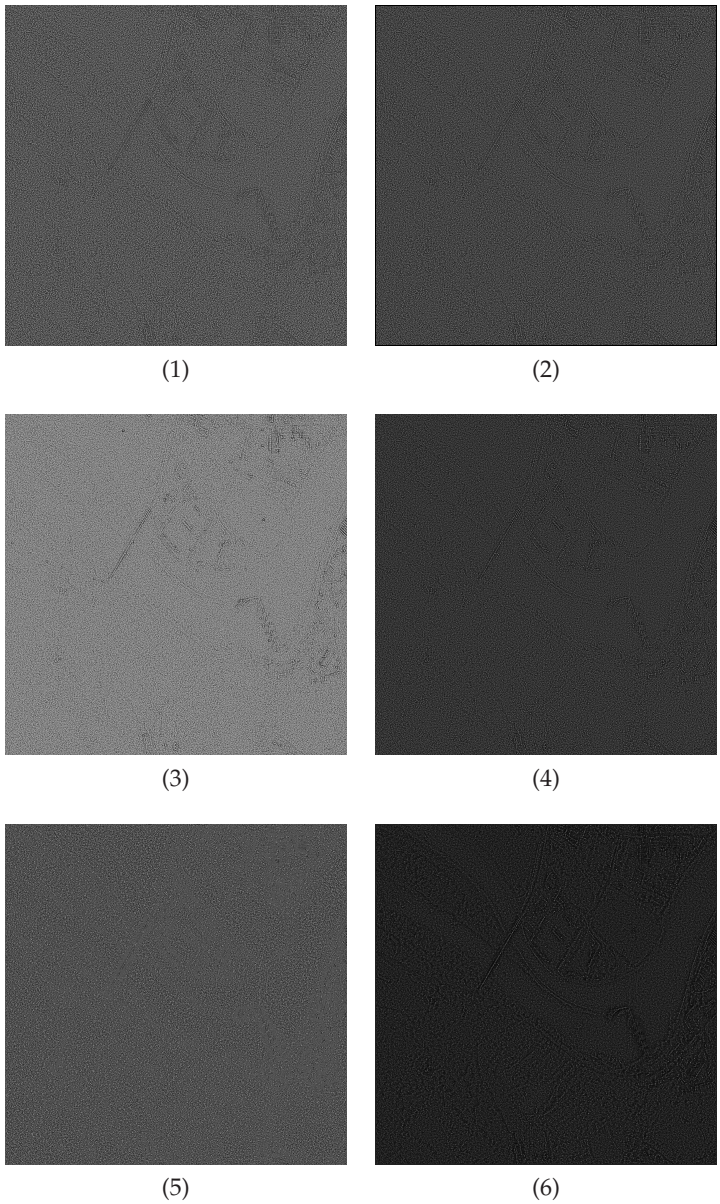


FIGURE L – Expérience avec une image RSO réelle. Rapports d'images : Lee amélioré (1), Frost amélioré (2), Kuan (3), Gamma (4), MBD (5) et la méthode proposée (6).

Observations et discussion

Étant donné le lien avec le cadre de sélection de modèle bayésien, l'estimation optimale des paramètres est effectuée avec succès par la courbe de taux de distorsion. La méthode est une approche globale qui peut être utile dans les applications d'exploration de données pour l'interprétation du contenu. L'optimisation de la taille de la Fenêtre d'Analyse et de l'Ordre du Modèle sont utiles pour l'extraction optimale de caractéristiques visant au développement d'un Système d'Exploration des Données sur des Images.

La régularisation statistique permet de modéliser le signal complexe par l'estimation

des paramètres et de lisser les canaux réel et imaginaire. Il peut être utile dans la reconstruction 3D du signal, par example, dans des applications de tomographie. Le modèle de données en nombres complexes GMRF peut ainsi modéliser le signal complexe par l'estimation des paramètres du modèle complexe. Concretement, dû a sa capacité pour modéliser les patrons de phase, il peut être appliqué pour la reconnaissance des cibles mais il ne s'applique pas directement aux zones de texture stationnaires.

La régularisation de Tikhonov montre une meilleure performance que les autres filtres au niveau de la mesure du MSE. En outre, une inspection visuelle des images avec le speckle filtré et des rapports d'images indique des résultats comparables entre eux. Cependant, un modèle a priori qui s'adapte mieux à la texture synthétique générée pourrait améliorer les résultats. Le modèle GMRF est prévu pour être utilisé comme modèle prior dans des expériences futures.

Dans le cas du test avec la texture Brodatz, la méthode proposée ne surpasse pas les autres filtres. Toutefois elle se comporte de manière satisfaisante avec la vraie texture. Le test avec l'image RSO simulée avec la méthode proposée montre un bon filtrage du speckle. De plus, toutes les mesures sont cohérentes. Finalement, les meilleurs résultats en termes de ENL sur l'image avec speckle filtré sur les données réelles RSO sont obtenus avec la méthode proposée.

La complexité des méthodes, qui on été programmées en C++, a été mesurée à travers le temps d'exécution sur une machine avec un processeur Intel core II 2.0 GHz. Pour tous les filtres adaptatifs une fenêtre de taille 5×5 pixels a été utilisée. Pour le MBD, une fenêtre de 41×41 pixels a été utilisée car il a été démontré que c'était la fenêtre optimale de moyennage. L'image RSO de la figure J de taille 1024×1024 pixels a été utilisée pour cela. Les temps d'exécution sont les suivants : la méthode MBD 250 s, Lee amélioré 1,5 s, Frost amélioré 1,8 s, Kuan 1,1 s, Gamma 1,3 s et la méthode proposée 18 s. La méthode la plus demandante au niveau computationnel est le MBD, puisque les paramètres du GMRL doivent être estimés.

Application sur la Classification

Les caractéristiques extraites ont été validées par classification non-supervisée K-means des paramètres du modèle de Champs Aléatoires de Gauss-Markov (Gauss-Markov Random Field, GMRF). L'algorithme de Lloyd généralisé (Generalized Lloyd Algorithm, GLA), également connu en tant que *K-means* dans la littérature de clustering, est la généralisation dans un espace multidimensionnel de l'algorithme de Lloyd (LA) pour la conception d'un quantificateur scalaire (Gersho & Gray, 1991).

L'algorithme est basé sur l'utilisation itérative de l'opération de modification du livre de codes. Il est basé sur les étapes suivantes :

1. Commencer avec un livre de codes initial $\mathcal{C}_m = \{\boldsymbol{y}_i; i = 1, \ldots, N\}$.
2. Étant donné le livre de codes $\mathcal{C}_m = \{\boldsymbol{y}_i; i = 1, \ldots, N\}$, trouver la partition optimale dans des cellules de quantification de sorte que l'assignement soit faite au code le plus proche. Autrement dit, former des cellules du plus proche voisin (condition du voisin plus proche) :

$$R_i = \{\boldsymbol{x} : d(\boldsymbol{x}, \boldsymbol{y}_i) < d(\boldsymbol{x}, \boldsymbol{y}_j);\ \text{all } j \neq i\}. \tag{3}$$

   Si $d(\boldsymbol{x}, \boldsymbol{y}_i) = d(\boldsymbol{x}, \boldsymbol{y}_j)$ pour un ou plusieurs $j = i$, attribuer $\boldsymbol{x}$ à l'ensemble R_j pour lequel j est le plus petit.

3. Étant donné l'ensemble de cellules qui viennent d'être calculées, trouver l'alphabet de reproduction optimal, autrement dit, le livre de codes $\mathcal{C}_{m+1} = \{cent(R_i); i = 1, \ldots, N\}$, où $cent(\cdot)$ est le centre de la cellule (condition de centroide, Centroid Condition).
4. Calculer la distorsion moyenne pour $\mathcal{C}_{m+1}$. Si le changement a été suffisamment petit par rapport a l'itération préalable, arrêter l'algorithme. Sinon, mettre $m + 1 = m$ et passer à l'étape 2.

Chaque application des étapes deux et trois (itération de Lloyd) doit réduire ou ne pas modifier la distorsion moyenne.

Généralement les critères d'arrêt sont les suivants :
- le nombre maximal d'itérations ;
- la position des centres ne doit pas changer (ou la distorsion ne doit pas diminuer) beaucoup d'une itération à la suivante ;
- l'erreur de distorsion, $\varepsilon = |D_{i+1} - D_i|$, est inférieure ou égale au seuil fixé.

Ils peuvent être utilisés individuellement ou en combinaison afin d'obtenir le mode d'interruption souhaité.

Une scène ESAR-X aéroportée acquise sur la ville de Dresden a été analysée (voir figure M1). Les caractéristiques du modèle ont été extraites et ensuite classifiées par l'algorithme k-means non-supervisé.

Les résultats de la classification avec les algorithmes GMRF à valeurs complexes et MBD sont indiqués dans les figures M2 et M3 respectivement.

Étant donné le contenu de l'image, cinq classes ont été choisies pour la classification :
- noir : l'eau, les ombres et les zones sombres ;
- bleu : végétation ;
- vert : zone résidentielle ;
- orange : les bâtiments ;
- jaune : diffuseurs très forts.

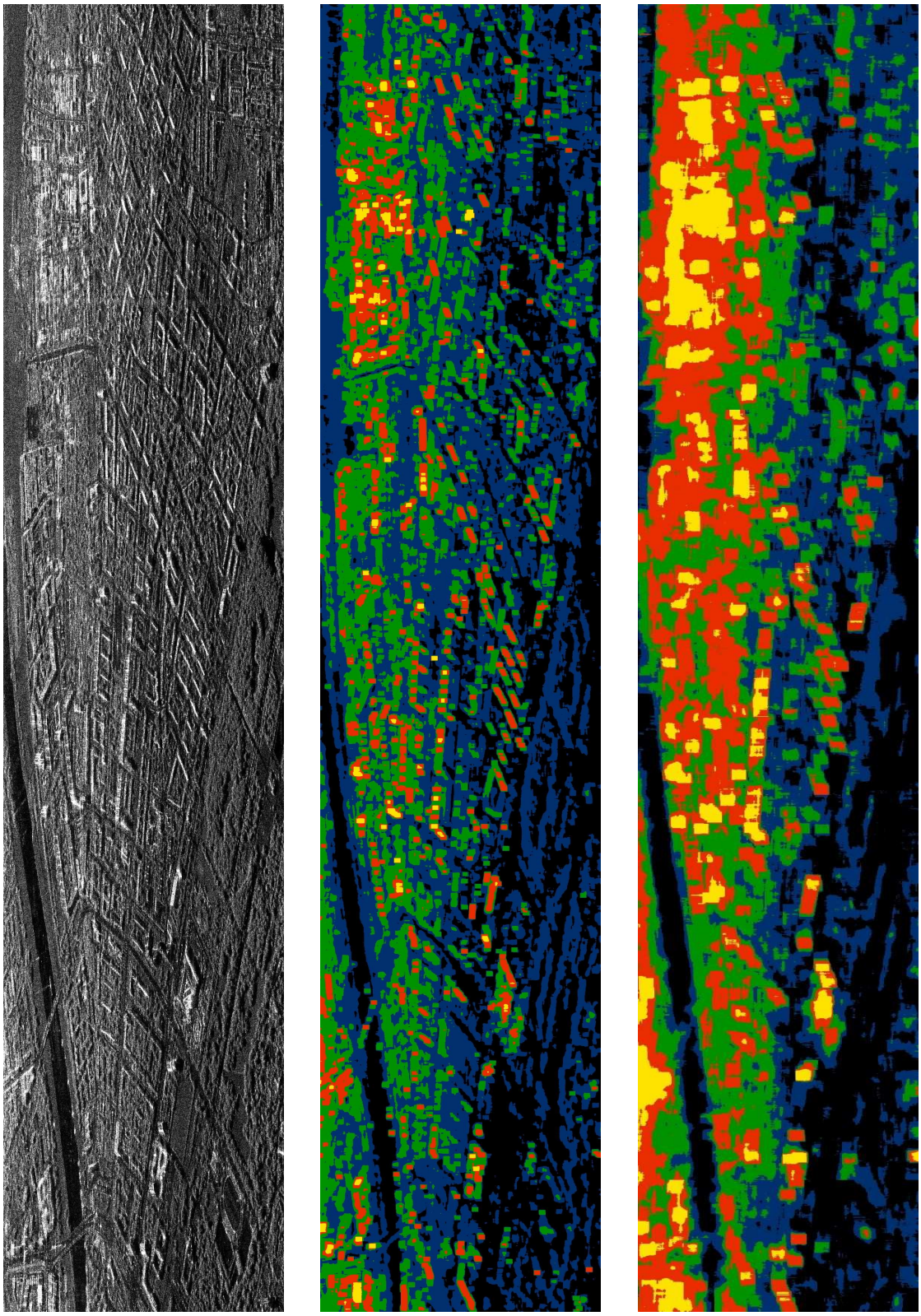
Une comparaison visuelle des figures M2 et M3 montre que la résolution de la classification faite avec le GMRF à valeurs complexes est meilleure.

Les matrices de confusion pour la classification avec le GMRF et avec le MBD sont indiquées dans les tableaux E et F respectivement. Dans le cas du MBD, les classes ne sont pas bien séparées : l'eau et les zones sombres sont classées comme végétation. En même temps celle-ci n'est pas séparée de la zone résidentielle. La performance pour les bâtiments et les diffuseurs forts est mauvaise. La matrice de confusion pour le modèle GMRF montre de meilleurs résultats spécialement pour les bâtiments et les diffuseurs forts. L'eau est aussi beaucoup mieux séparée, tandis que la végétation, contrairement au MBD, est fusionnée à la zone résidentielle.

Conclusions

La taux de distorsion a été appliqué à l'algorithme déjà existant du MBD pour l'optimisation des paramètres d'extraction de caractéristiques. Puisque l'algorithme est implémenté dans un système d'exploration des données sur des images, le problème de l'extraction optimale de caractéristiques se pose.

L'analyse doit être abordée en prenant en compte la grande quantité de données devant être traitées par le système. En conséquence une méthode globale, simple et rapide



(1) (2) (3)

FIGURE M – K-means résultats de classification à cinq classes : noir (eau), bleu (végétation), vert (quartier résidentiel), orange (bâtiment) et jaune (diffuseurs forte). (1) Image d'amplitude, (2) classification GMRF à valeurs complexes et (3) classification MBD.

%	Eau	Végétation	Quartier résidentiel	Bâtiments	Diffuseurs forts
Eau	67	28	5	-	-
Végétation	29	63	8	-	-
Quartier résidentiel	32	9	59	-	-
Bâtiments	11	11	-	78	-
Diffuseurs forts	-	-	9	18	73

TABLE E – Matrice de confusion de classification non-supervisée K-means pour les champs aléatoires de Gauss-Markov aux valeurs complexes.

%	Eau	Végétation	Quartier résidentiel	Bâtiments	Diffuseurs forts
Eau	28	61	11	-	-
Végétation	27	73	-	-	-
Quartier résidentiel	18	41	41	-	-
Bâtiments	24	29	12	35	-
Diffuseurs forts	16	11	11	17	44

TABLE F – Matrice de confusion de classification-non supervisée K-means pour MBD.

est nécessaire. La solution a été trouvée dans le cadre de la théorie de l'information et du traitement, en mesurant l'erreur de codage d'une source de données.

La compression avec perte de données est appliquée par K-means dyadique sous l'hypothèse de distributions gaussiennes mélangées. La taille moyenne optimale de la fenêtre de l'analyse permet une estimation robuste des paramètres pour la description des images avec des caractéristiques. La moyenne optimale de l'ordre du modèle (Model Order, MO) permet d'éviter la sélection d'un modèle qui fasse du surapprentissage sur les données.

En outre, le résultat dépend de la diversité des données et le nombre de classes codées par les clusters. L'optimisation est effectuée sous l'hypothèse de données stationnaires dans la fenêtre d'analyse. Cette hypothèse n'est pas toujours respectée. Pour surmonter ce problème, une fenêtre d'analyse adaptative peut être utilisée.

Le résultat de la sélection de l'ordre de modèle est comparé avec la carte locale d'ordre du modèle choisie par l'information de Fisher. La carte montre comment l'ordre u modèle est liée au contenu de l'image. Les zones uniformes ou celles qui présentent une faible complexité sont mieux représentées par un ordre de modèle bas, tandis que les zones texturées et celles qui montrent de fortes variations sont mieux représentées par un modèle d'ordre élevé.

La Sélection de Modèle globale effectuée par taux de distorsion semble être une solution raisonnable parce que la sélection de MO adaptative locale est computationalement chèr.

L'analyse des données avec valeurs complexes commence par l'extension dans le domaine complexe de l'estimation par maximum a posteriori (MAP) pour la régularisation statistique de la partie réelle et la partie imaginaire. Le modèle d'image linéaire est régularisé sous la contrainte d'un terme de probabilité a priori. Celui-ci est donné par une distribution de Gibbs, qui est une distribution exponentielle dont l'exposant, connu comme la fonction de l'énergie, caractérise les familles des champs aléatoires de Markov (MRF)

différentes.

Les paramètres de ce terme probabilistique (a priori) sont estimés à partir des données incomplètes par une procédure de maximisation de l'espérance (Expectation maximization, EM). Les images des paramètres estimés sont utilisées pour trouver la solution du problème mal posé de l'estimation de l'image à partir des données avec du bruit.

Les Champs Aléatoires de Gauss-Markov (Gauss Markov Random Field, GMRF), caractérisés par une fonction quadratique d'énergie, sont capables de décrire les caractéristiques locales des images. Bien que les champs aléatoires de Gauss-Markov ne soient pas le meilleur choix pour la reconstruction de profils avec des fortes discontinuités, ils sont intéressants parce que le posteriori est convexe et gaussien. En fait, le gradient peux être calculé analytiquement et le fonctionnel minimisé sans tomber dans des minimaux locaux.

L'image paramètre estimée semble vraiment similaire, parce que les canaux réel et imaginaire ne sont pas corrélés mais pas indépendants. Les paramètres donnent une mesure de la variation qui a lieu dans le signal. Donc elle pourrait être utilisée pour étiqueter/classifier le contenu de l'image. L'image d'amplitude est floue, mais aucun artefact est généré par le filtre.

Les famille élue des modèles de champs aléatoires de Gauss-Markov est isotrope. D'autres modèles, caractérisés par un vecteur de paramètres au lieu d'un scalaire, sont en mesure d'estimer des paramètres directionnels qui peuvent capturer des structures à différentes échelles et orientation.

C'est le cas du modèle à valeurs complexes de champs aléatoires de Gauss-Markov développé par l'extension de la définition classique des champs aléatoires de Gauss-Markov au domaine complexe. Les paramètres estimés sont capables de modéliser des patrons différents dans la phase complexe de l'image.

Dans une image RSO seule, ces modèles sont essentiellement visibles dans le voisinage des diffuseurs forts en raison de la réponse impulsionnelle du système. On suppose que c'est possible de caractériser la texture locale par la modélisation du patron des phases, mais l'hypothèse n'a pas été confirmée.

En réalité, ces motifs sont vraiment rares et pas évidents. En outre, les valeurs des paramètres sont similaires pour des textures différentes. Par la modélisation des données avec les champs aléatoires de Gauss-Markov à valeurs complexes, la variance du modèle pourrait distinguer des texture avec des intensités différentes.

Le modèle d'image linéaire dans la méthode de Tikhonov permet d'inclure la réponse impulsive du système et la contrainte du modèle a priori, permettent ainsi de traiter les zones uniformes et de préserver les contours grâce à la variation totale et la fonction Huber-Markov respectivement. La fonction convexe assure que la solution existe et est unique et elle a été minimiseé par la méthode itérative de Newton.

Les paramètres de filtre permettent un réglage optimal de l'optimisation. Initialement, ils ont été choisis expérimentalement, en trouvant qu'ils contrôlent le degré d'efficacité des modèles antérieurs, à savoir le degré de lissage par rapport au degré de préservation des contours.

Le problème a été résolu comme un problème de sélection de modèle avec un cadre bayésien de sélection de modèle. Le paramètre norme définit la forme de la fonction de pondération, par exemple une distribution de Gauss ou de Laplace. Il doit être initialisé

expérimentalement dans la première itération de l'algorithme et il est ensuite estimé. Le pas numérique, qui contrôle la vitesse de convergence et la précision de la solution, doit aussi être choisi expérimentalement.

Quatre expériences ont été réalisées afin de donner des résultats aussi complets que possible. Le filtre montre une performance supérieure pour le débruitage de textures synthétiques et des données RSO réelles. Dans le cas de la texture Brodatz et des données RSO simulées, la performance du filtre est similaire à celle du filtre de Lee amélioré et du MBD.

La nouveauté de la démarche repose sur la possibilité d'inclure la fonction de transfert du système, le modèle a priori et l'utilisation de la pleine résolution et des informations contenues dans les données à valeurs complexes.

Finalement l'application des méthodes proposées est présentée pour la classification non supervisée et supervisée des modèles. La classification K-means est comparée avec le modèle MBD.

Bien que les résultats soient loin d'être optimaux, les méthodes proposées semblent être prometteuses. Une amélioration des résultats pourraient être obtenue grâce au prétraitement des données afin d'en extraire l'information des bords pour une meilleure séparation des classes et/ou combiner différemment les paramètres. En outre, d'autres modèles paramétriques peuvent être envisagés.

Table des matières

Acknowledgments

First of all, I want to thank Mihai Datcu for having supervised my thesis with enthusiasm and scientific rigor, and Dusan Gleich for having co-supervised my thesis with encouragement and availability during these years.

I also thank the members of the jury for honoring me with the participation to the defense : Maria Pilar Jarabo Amores, president of the jury, Emmanuel Trouvé and Stefan Hinz reviewers, Jean-Marie Nicolas and Jean-Claude Souyris examiners.

I also thank Roger Fjørtoft for his contribution as reviewer, all the members of the Center of Competence and the department of Photogrammetry and Image Analysis of DLR.
I want to thank also people with whom I spent good and bad moments and I discussed about science and life during these years : Marc, Alessandro, Antonio, Inés, Franz, Marco, Houda, Daniele, Daniela, Nestor, Marie, Fernando, Elisa, Cristian, Sebastián, Jaime, Stephan, Peter, Marine.
A special thanks goes to Marie and Fernando for the French translation.

Ringrazio tutte le persone che mi sono state vicine in questi anni di dottorato e che mi hanno supportato, aiutato e sopportato in questa difficile scelta. Le persone che ho conosciuto e incontrato qui a Monaco, la polizia monacense che non mi ha mai fatto sentire ignorato. I miei amici di Roma : Daniele, Roberta e Francesco. I miei amici dell'Università di Tor Vergata : Chiara, Cosimo, Fabio, Andrea, Giorgio, Emanuele, Lino e tutti gli altri ragazzi dell'EOlab. Un ringraziamento particolare a Saverio per la sua amicizia e il suo supporto logistico durante i miei rientri a Roma. Ringrazio la mia famiglia, i miei genitori, per aver sempre appoggiato e supportato la mia scelta, e mia sorella per il suo aiuto e il suo sostegno. Infine ringrazio la mia ragazza, Amaia, per avermi sopportato, per essermi stata vicino e per il suo indispensabile aiuto.

Table des figures

Liste des tableaux

List of Algorithms

Acronyms and Abbreviations

AD Anisotropic Diffusion
AMI Aritmetic Mean Intensity
AR Auto Regressive
AW Analyzig Window
BIC Bayesian Information Criterion
DFT Discrete Fourier Transform
EM Expectation-Maximization
ENL Equivalent Number of Look
EO Earth Observation
EPOS Edge Preserving Optimized Speckle
FFT Fast Fourier Transform
GGMAP Gamma-Gamma MAP
GMRF Gauss-Markov Random Field
GRF Gibbs Random Field
GUI Graphic User Interface
HMRF Huber-Markov Random Field
HR High Resolution
IDFT Inverse Discrete Fourier Transform
iid independent identically distributed
IIM Image Information Mining
IIR Infinitive Impulse Response
ITU International Telecommunication Union
LMSE Linear Mean Square Estimation
LSE Least Squares Error
MAP Maximum *A Posteriori*
MBD Model Based Despeckling
MDL Minimum Description Length
MGD Multi-Look Ground-range Detected
ML Maximum Likelihood
MMSE Minimum Mean-Square Error

MO Model Order
MRF Markov Random Field
MSE Mean-Squared Error
MTD Moving Target Detector
NMSE Normalized Mean-Square Error
PDE Partial Differential Equation
pdf probability density function
PMSE Peak Mean-Square Error
QAM Quadratic Amplitude Modulation
QB Quick Bird
RAR Real Aperture Radar
RCS Radar Cross Section
RMA Range Migration Algorithm
RMS Root Mean Square
SαS Symmetric-α-Stable
SAR Synthetic Aperture Radar
SLAR Side Looking Aperture Radar
SNR Signal to Noise Ratio
SL Spot Light
SSE Semantic Search Engine
SSIM Structural SIMilarity
SVM Support Vector Machine
SWF Spatial Whitening Filter
TV Total Variation
UTC Coordinated Universal Time
UTM Universal Tranversal Mercator
WLMSE Widely Linear Mean Square Estimation

Introduction

The first Synthetic Aperture Radar (SAR) sensor for Earth Observation (EO) was launched by NASA in 1978 on board of the satellite SEASAT. It stopped working few months later but the SAR technology and development has continued ever since. Satellites equipped with sensors with better resolution have been launched and new missions are planned. The major advantage of the SAR imaging sensor is that it is independent from weather conditions (e.g. clouds) and daylight because it is an active sensor. A view of a SAR satellite is shown in Figure 1. It is the German satellite TerraSAR-X launched in June 2007. Other kinds of sensors also are devoted to Earth imaging from space, e.g. optical sensors, but they have as major drawback the limitation due to the cloud coverage and the sunlight because they are passive sensors. Satellites allow to have an Earth coverage with a revisit time which depends on the orbit. It permits to acquire data on the same areas and to analyse the image sequences. Changes in the Earth surfaces, structure deformations of the order of millimeters may be detected and 3D models of the Earth surface may be generated by SAR acquisitions. From 1978 up to today, many missions and new sensors have been developed for EO, see Figure 2, where the launch of TanDEM-X is planned in the beginning of 2010. This enormously increases the availability of SAR data to be used in many different applications making the search and retrieval of information from the archives arduous. In parallel, the improvement of resolution hugely increases the information content. It makes the automatic interpretation of the images more difficult because the level of detail highly grows. The analysis of SAR data is made complicated because the coherent image is corrupted by speckle noise, see Figure 3. In order to deal with speckle we need to find a correct model of the uncorrupted data and to infer them. An accurate estimation accounts also the SAR imaging system which has to be correctly modelled. The thesis provides a new methodology to analyze High Resolution (HR) SAR data to facilitate image understanding, interpretation and indexing.

Progresses in SAR Modeling and Information Extraction

The analysis of SAR images in term of image estimation, restoration and denoising has been applied in the past on detected data. The well-known adaptive filter Lee (1980), Frost et al. (1982) and Kuan et al. (1985) have been developed to estimate the image reflectivity while preserving image features. However the capability of estimating reflectivity and extracting image features (e.g. edges) has been demonstrated to be more effective with complex valued data (Fjortoft et al., 1999; Fjortoft & Lópes, 2001). Except of few contributions in the literature (Jakeman & Pusey, 1976; Szajnowski, 1977; Sekine & Mao, 1990; Tison et al., 2004) the main effort has been to find and fit statistical models on detected data (amplitude or intensity) neglecting the phase, after non linear transformation of

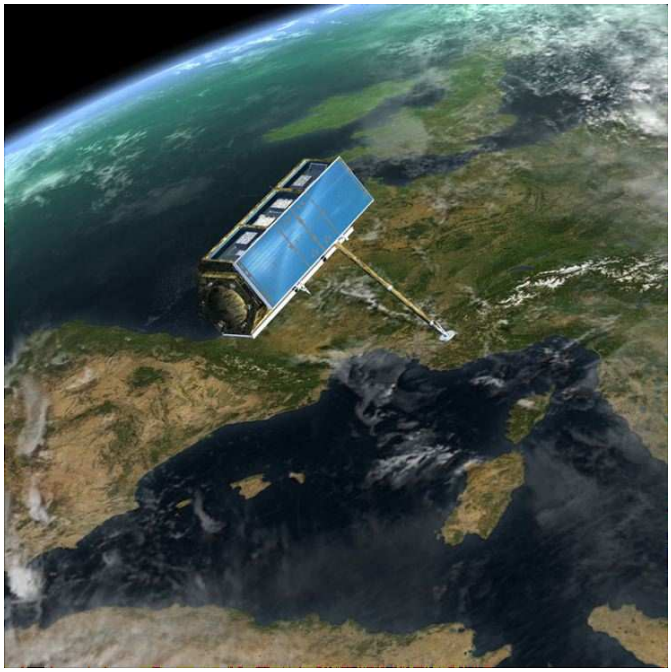


FIGURE 1 – View of TerraSAR-X satellite. It was launched in June 2007. It is providing HR data of the Earth surface up to 1 m resolution.

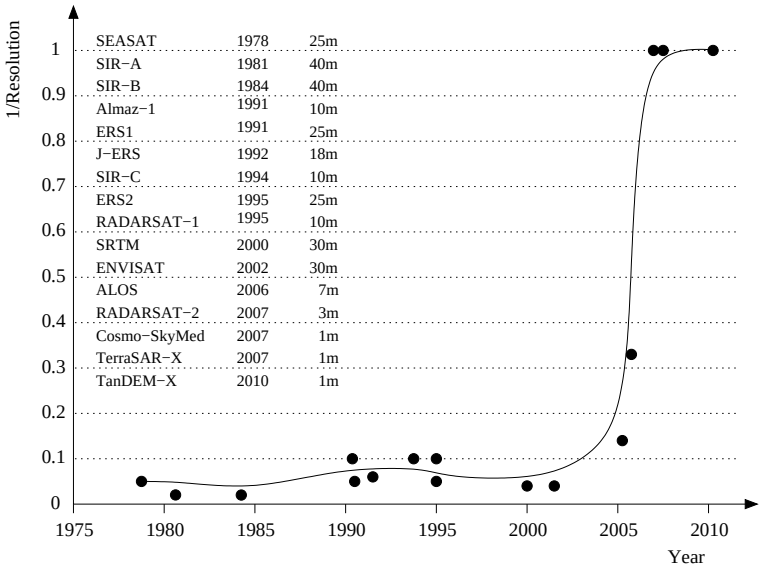


FIGURE 2 – Growing spaceborne SAR resolution with time. With increasing resolution, automated image understanding becomes a hard task. The information content increases enormously with the resolution. The thesis addresses the development of new methods to fully exploit the information in HR data.

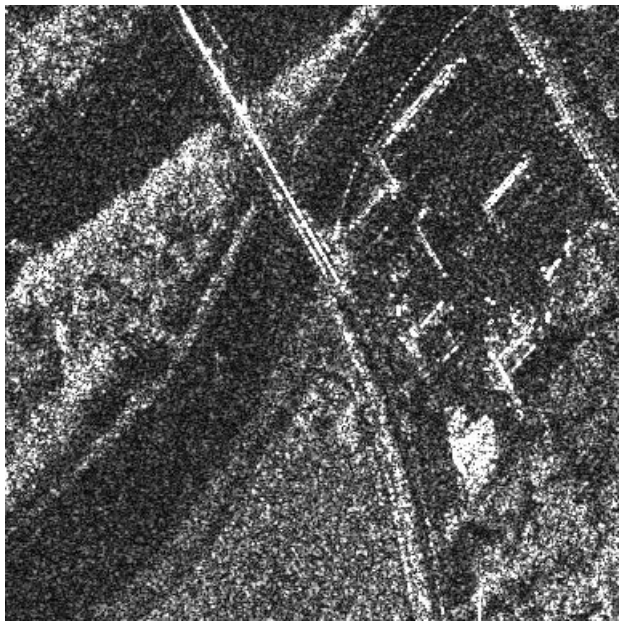


FIGURE 3 – Example of HR SAR image. Buildings, man made structures, vegetation and water are present in the image but are not clearly distinguishable. Image understanding aims to facilitate image interpretation.

the complex reflectivity from Cartesian to polar coordinate. However, modeling complex-valued data has several advantages :

- Exploiting the full available information at full resolution.
- Simplicity of the data model.
- Behavior of the model to represent deterministic signature or incertitude in the complexity of the scene.
- It is the only way to have accurate system modeling.

The drawback of comple-valued data is the low signal-to-noise-ratio. Many statistical models, with empirical or theoretic basis, have been proposed to handle speckle. The Gamma distribution has been introduced as a model for a multilook SAR intensity probability density function (pdf) where the correspondent amplitude has a Nakagami distribution. They are a generalization of negative exponential and Rayleigh distributions respectively. The K distribution (Jakeman & Pusey, 1976, 1978; Oliver, 1984; Jao, 1984) is obtained for the signal intensity assuming the population of scatterers in the resolution cell to be controlled by a birth-death-migration process. In the hypothesis of Gamma distributed noise and Gamma distributed signal (Oliver, 1991; Oliver & Quegan, 2004a), the product model results in a K distribution for the intensity. The Nakagami-Rice distribution has been proposed to model the SAR statistics in the presence of a single strong reflector in homogeneous clutter (Dana & Knepp, 1986; Tison et al., 2004). Inverse Gaussian distributions have also been employed to model the amplitude statistics in Frery et al. (1997), Muller & Pac (1999) and Eltoft (2003). In Kuruoglu & Zerubia (2004) a *heavy-tailed* Rayleigh model is presented under the hypothesis that the real and imaginary parts of the backscattered signal are jointly Symmetric-α-Stable (SαS) random variables. Generalized Gaussian distributions are assumed for real and imaginary parts in Moser et al. (2006). The $\mathcal{G}$ distribution for the amplitude, presented in Frery et al. (1997), is the result

of a multiplicative model assuming a Nakagami distribution for the speckle noise and a generalized inverse Gaussian distribution for the signal. In Muller & Pac (1999), a particular case of the $\mathcal{G}$ model, called *harmonic branch* $\mathcal{G}^h$, is proposed, while the $\mathcal{G}^0$ distribution is proved to be equivalent to a Fisher pdf in Tison et al. (2004).
In addition to the above mentioned *theoretical* or partially theoretical models, several empirical models have been used to characterize the statistics of SAR amplitude (or intensity) data, such as log-normal (Szajnowski, 1977), Weibull (Sekine & Mao, 1990) and Pearson (Delignon et al., 1997).
The Gauss-Markov Random Field (GMRF) family of models (Chelappa et al., 1985) has been successfully used in computer vision for texture generation and compression. In Walessa & Datcu (2000) it has been used in MAP estimation as prior in the framework of Bayesian inference together with the Gamma distribution.
The direct problem of computing the imaging system response from a given image is often assumed to be known and well-posed. The usual model for it is the convolution by a given kernel or point spread function which, in most of the cases, implies that the inverse problem of computing the true image from the observations is an ill-posed problem. A general principle for dealing with the instability of the inverse problem is that of regularization, which mainly consists in restricting the set of admissible solutions and including some a priori information (non negativity, smoothness, existence of edges, etc.) in the formulation of the problem. Both the accurate modeling of the imaging system and the choice of regularization will be essential for a satisfactory image restoration process. Tikhonov (Tikhonov & Arsenin, 1977) has introduced methods for regularizing deterministic problems by introducing stabilizing functionals which play a role analogous to the log prior distribution of MAP estimation. The regularization of ill-posed problems has been the subject of much research, e.g. Horn & Rhunck (1981) for optical flow determination, Ikeuchi & Horn (1981) for the shape reconstruction from shading, Torre & Poggio (1986) for edge detection and Marroquin et al. (1987) for computational vision. When the functional is not convex it is necessary to adopt a simulated annealing method to minimize it. The functional we consider is convex, thus gradient descent methods can be applied.
In Çetin & Karl (2001), the regularization is used in order to develop an enhanced image formation method for Spot Light (SL) airborne SAR. By exploiting a similar image mode, starting from Çetin & Karl (2001) and Pan & Reeves (2006), we introduce a complex-valued likelihood which allows to take into account the complex-valued nature of the system transfer function. This approach is different from the one suggested in Çetin & Karl (2001) where the authors suggest a model for image formation. We use as prior the Huber-Markov function which is shown to be a powerful model for the preservation of the edges in Pan & Reeves (2006).
The *ill-posed* image restoration problem becomes a *well-posed* problem by choosing a convex prior, which ensures that the solution exists, is unique and depends continuously on the data. The last condition, called stability, ensures that small perturbations in the data do not cause dramatic change in the solution. Eventually, a model selection is performed in order to estimate the optimal average analyzing window and the order of the Auto Regressive (AR) process through Rate Distortion. Previously, a similar work has been only done by Pesaresi (1996) for low resolution SAR and and Soccorsi et al. (2006) for optical data.

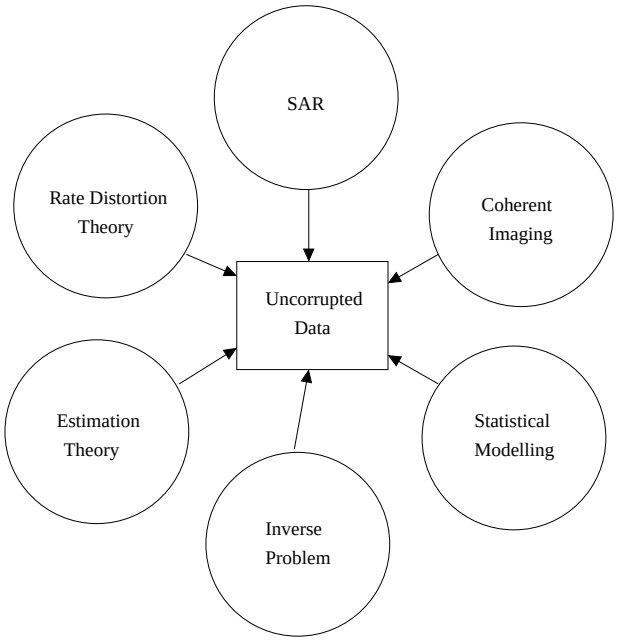


FIGURE 4 – The uncorrupted data are estimated considering the problem as an inverse problem by statistical modeling of the SAR system and the coherent image formation process. Rate Distortion, in the framework of Information Theory, is used for Model Selection.

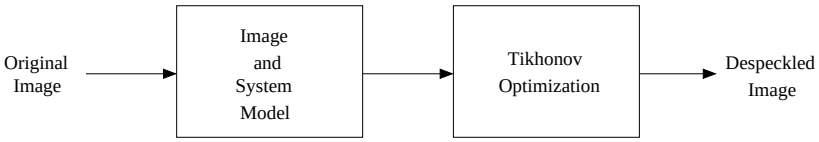


FIGURE 5 – Diagram of the Tikhonov optimization method. The model of the image and of the system are chosen, and the optimization is performed. The output is the despeckled image.

Contribution of the Thesis

The thesis deals with information extraction and enhancement of meter resolution SAR data aiming to provide better content descriptors for further scene understanding and target recognition, as well as radiometrically and spatially enhanced products. To achieve this goal the thesis approach the problem of modeling SAR images and propose novel solution based on estimation of inverse problem for information extraction (Figure 4). The problem of model selection is managed by Rate Distortion, because of the correspondence with the Bayesian evidence framework. We start the analysis with the extension of the linear Gauss-Markov Random Field (GMRF) family of models to complex-valued data, which applies to complex-valued random variables : the multivariate complex Normal distribution and the parametric GMRF model in case of proper and improper random variables. The latter model is used for parameter estimation in the second level of Bayesian inference. Thus, we extend the analysis to the full Bayesian approach, including first and second level Bayesian inference, i.e. model selection and parameter estimation. This is achieved by the MAP estimation method which has been applied for despeckling and feature extraction by Walessa & Datcu (2000). The MAP estimation method exists in the complex domain mainly in tomography applications (Pascazio & Ferraiuolo, 2003). The approach allows the reconstruction of the real and imaginary part but deals with additive noise and thus, it does not remove speckle in case of SAR images.
The proposed method (Soccorsi et al., 2009) is a Tikhonov regularization approach in the complex domain, see diagram in Figure 5. Speckle is handled as a random process with real-valued data. The complex-domain approach allows to handle the coherent image formation as information or as incertitude in case of structures or textures in the scene. The novelty relies in the use of complex-valued data which allows to include the system transfer function. The method is equivalent to the Bayesian MAP estimate. The method, which works with complex valued data, is demonstrated to be better than the classical despeckling methods (e.g. Lee filter, MBD) on detected data. Depending on the adopted prior it allows to denoise the image and to estimate texture parameters.
In the context of parameter optimization for feature extraction, the optimal (average) analyzing window (Pesaresi, 1996) and the optimal (average) order of the AR process are estimated with the use of Rate Distortion (Soccorsi & Datcu, 2008). It confirms that Rate Distortion is a successful entropy-based method for model selection.
The results can be used in Image Information Mining (IIM) systems for image understanding facilities. The thesis is divided in four parts. Chapter one presents the SAR system and the image formation process. It describes the information contained in the acquired data and different ways to represent them. Indeed, the information may be represented in alternative ways with different statistical properties. Image transformations allow to use multiplicative and additive noise models. The physical origin of speckle and the classical methods to estimate the mean reflectivity are summarized. In chapter two, the theoretical basis of estimation theory is presented. The estimation problem is addressed in the Bayesian framework and the outline of an estimator is presented. Furthermore, Rate Distortion theory is presented after introducing the basic concept of information theory. In the chapter the correspondence between estimation theory and Bayesian frame is delineated. Furthermore, the Rate Distortion theory is linked with the evidence framework in the context of Bayesian model selection. Chapter three presents the methodology adopted for image and system modeling for the estimation of the uncorrupted data. The extension of the GMRF model in complex-valued domain is presented as well as the complex-image

regularization method. It is adopted for the regularization of the real and imaginary parts of complex-valued data. Afterwards the Tikhonov optimization method is described. The results are presented in Chapter four. The Tikhonov optimization is applied on simulated, synthetic and actual SAR images and the model selection is performed by Rate Distortion for the optimal parameter estimation for feature extraction.

Chapitre 1

SAR System : Data Acquisition and Image Formation

In this chapter the relevant bases of the SAR imaging system are presented. In the beginning the SAR sensor is described together with the data acquisition and the image formation. Then, the representation of the information and the data statistics are delineated. The chapter ends with some consideration about speckle, speckel simulation and the intensity models for speckle reduction.

1.1 SAR System

The Radar is an instrument which allows exploiting radio wave propagation to sense the presence of an object and its distance. The ancestor of the Radar device was invented by Christian Hülsmeyer in 1904. The word, now entered in the common language, is actually the acronym of Radio Detection and Ranging. The basic functionalities of the instruments are localization and distance measure. With the World War II it had a strong development which let to the modern Radar. Simplifying, it is constituted by an antenna which is used in transmission and reception in case of monostatic configuration. The sensor is fixed to a mobile platform and sends pulses which illuminate the scene. The electromagnetic wave, diffracted with an angle equal to the incident ray, goes back to the transmitting antenna, is received and collected. The microwave frequency bands used are listed in Table 1 : the Radar nomenclature (IEEE-AESS, 2003) is consistent with the International Telecommunication Union (ITU) nomenclature (NTIA, 2008). The metric wavelength includes P-band radars and the frequencies of the order of THz are not considered in the nomenclature in Table 1. The wavelengths used in EO applications are limited due to the atmospheric attenuation. The major attenuation contributions are given by the water vapour (22.2 GHz and 183.3 GHz) and oxygen (61.2 GHz and 118.8 GHz).
In addition to the basic functionalities, new applications have been found for Radar sensors in the last decades. In particular, its characteristics of active sensor independent from day light makes it interesting for Earth Observation and remote sensing applications.
A spaceborne or a airborne SAR is equipped with the sensor in Side Looking Aperture Radar (SLAR) configuration : Real Aperture Radar (RAR) and SAR. The first is a classical Radar while the second allows to overcome the limit of resolution of the standard Radar exploiting the prolonged illumination time of a point in the flight direction.

Radar nomenclature					ITU nomenclature	
Radar letter designation	Frequency range		Wavelength range		Adjectival band designation	Corresponding wave designation
HF	3-30	MHz	100-10	m	HF	Dekametric
VHF	30-300	MHz	10-1	m	VHF	Metric
UHF	300-1000	MHz	1-0.3	m	UHF	Decimetric
L	1-2	GHz	30-15	cm		
S	2-4	GHz	15-7.5	cm		
C	4-8	GHz	7.5-3.75	cm	SHF	Centimetric
X	8-12	GHz	3.75-2.5	cm		
K_u	12-18	GHz	2.5-1.67	cm		
K	18-27	GHz	1.67-1.11	cm		
K_a	27-40	GHz	11.1-7.5	mm	EHF	Millimetric
V	40-75	GHz	7.5-4.0	mm		
W	75-110	GHz	4.0-2.7	mm		
mm	110-300	GHz	2.7-1.0	mm		

TABLE 1 – Radar frequency nomenclature (IEEE-AESS, 2003) and ITU nomenclature (NTIA, 2008). The ITU defines no specific service for radar, and the frequency assignments listed are derived from those radio services that use radiolocation. The world is divided into regions and a different band, in the frequency range, is assigned to each one according to the standard IEEE-AESS (2003).

1.1.1 SAR Geometry and Resolution

The SAR sidelooking acquisition geometry is shown in Figure 1. The satellite flights along its orbit at height h from the Earth which is approximated as flat. The sensor looks at right or left side with a variable incidence angle θ. The footprint corresponds with the -3 dB antenna main lobe which illuminates the ground. The width in the range direction is called swath. The width of the swath can be of the order of 1 up to 100 Km. The geometry is a simplification because the curvature of the orbit and the curvature of the Earth should be considered. Resolution includes the two directions range (or across-track) and azimuth (or along-track). In a RAR the resolution in azimuth is given by the antenna beamwidth (footprint) at -3 dB, while the discrimination of two different targets in distance is given by the time length of the impulse, before correlating with the chirp. In the SAR configuration the discrimination in distance corresponds to the resolution in range. The pulse has to satisfy two requirements :

1. it has to be *short* to better discriminate close targets ;
2. it has to be *powerful* in order to ensure that the echos have enough energy.

The two requirements stated above are difficult to realize from a technical point of view. TerraSAR-X, for example, has a chirp of duration 30.7μs (Stangl et al., 2006), which yields to a resolution of 4.6 Km. The obtained range resolution is not acceptable for the required application and it is much lower than the real resolution obtained with the sensor, which is of the order of meter (DLR, 2008).
The footprint in azimuth depends on the width θ_a of the main lobe of the antenna in

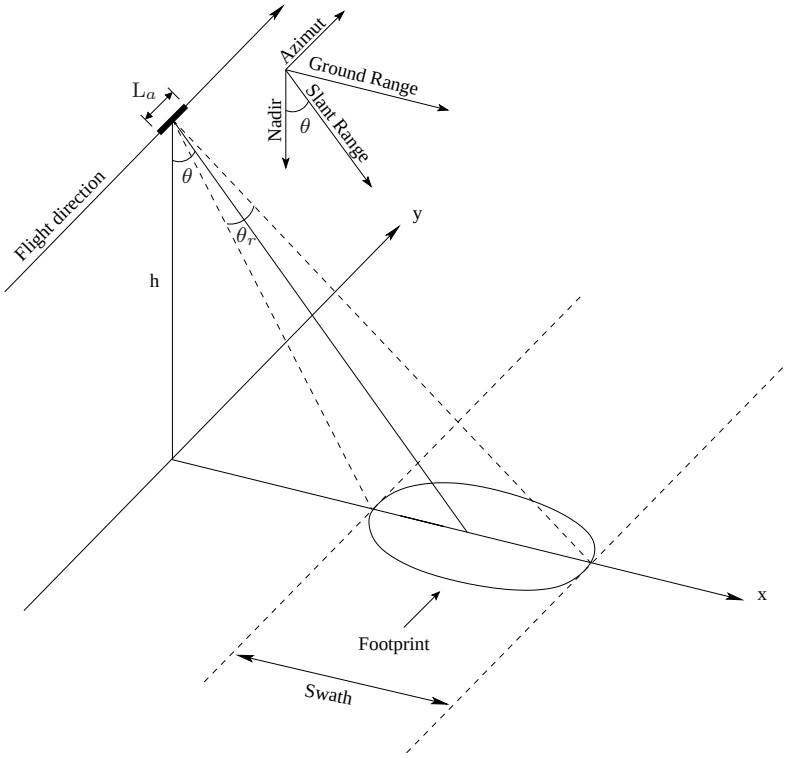


FIGURE 1 – SAR acquisition geometry.

azimuth. For a generic antenna it can be approximated by the formula

$$\theta_a \approx 65\frac{\lambda}{L_a} \qquad [\text{degrees}] \tag{1.1}$$

where λ is the wavelength and L_a is the dimension of the antenna in the azimuth direction.
In the case of TerraSAR-X, which has a centre frequency of 9.65 GHz and an antenna of 4.8 m in azimuth direction, the angle is of about 0.42° using Equation 1.1. Considering that the satellite flies with an orbit altitude of 514 Km, the equivalent footprint is about 5.3 Km for a side-looking incident angle of 45°. Again the resolution is not acceptable for the application and it is far from the operational resolution value.
The dimension of the swath depends on the beamwidth in elevation : for TerraSAR-X, depending on the operation mode, it is up to 500 Km (Stangl et al., 2006).

1.1.1.1 Range resolution

The geometry of the SAR in range is showed in Figure 2. The resolution depends on the pulse duration and is given by

$$\delta_{sr} = \frac{c\tau}{2} \tag{1.2}$$

where c is the velocity of the light and τ is the pulse duration. The factor 2 accounts for the two-ways, forward and backward, wave propagation.

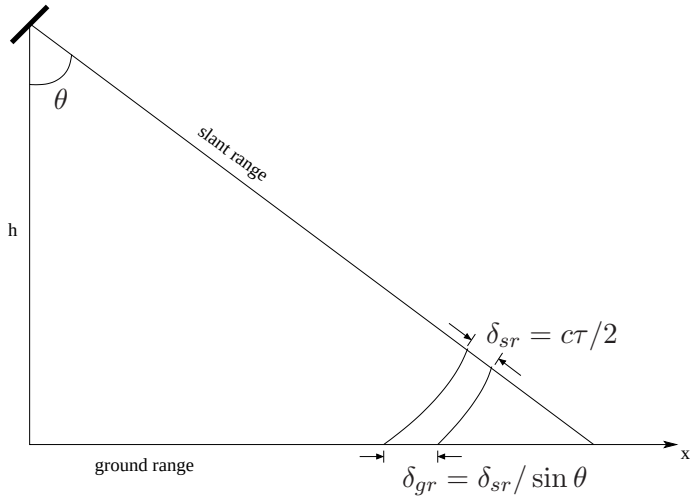


FIGURE 2 – Geometry for the SAR resolution in range.

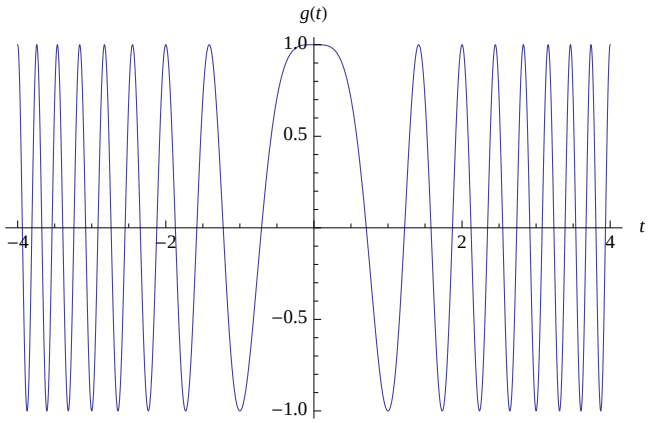


FIGURE 3 – Linear chirp (time in [μS]). Example of a chirp of duration 8 μS.

Equation 1.2 expresses the resolution in *slant range* direction. The equivalent resolution in *ground range* is given by the projection

$$\delta_{gr} = \frac{c\tau}{2\sin\theta}. \tag{1.3}$$

Thus, the *ground range* resolution depends on the incidence angle : for high incidence angle $\delta_{gr} \sim \delta_{sr}$, for low incident angle $\delta_{gr} \gg \delta_{sr}$.

Since for electronic reasons it is not possible to design a transmitter with a short pulse, a quadratic phase modulation, a so-called linear chirp (Figure 3), is considered in order to reach the desired resolution. The equation of a generic chirp with a carrier of frequency f_0 has the following form

$$g(t) \doteq \mathrm{Re}\{\mathrm{rect}_T(t) \cdot \mathrm{e}^{j[2\pi f_0 t + \pi k t^2]}\} = \mathrm{rect}_T(t) \cdot \cos(2\pi f_0 t + \pi k t^2) \tag{1.4}$$

The term *linear* comes from the fact that the equivalent frequency modulation is linear, Figure 4. The instantaneous chirp frequency in MHz is shown in Figure 4 as a function of

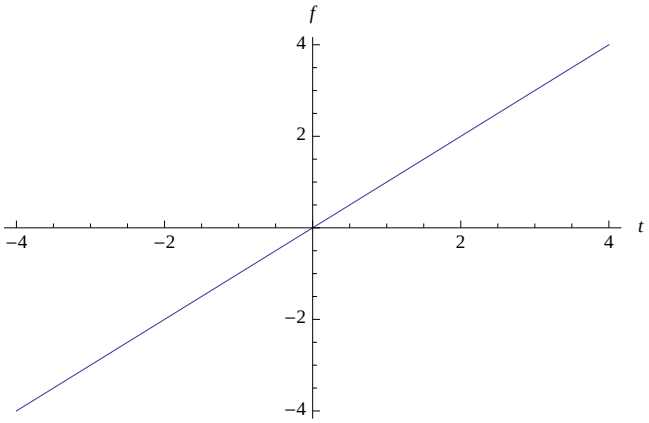


FIGURE 4 – Chirp frequency (the instantaneous frequency $[\mathrm{MHz}]$ is shown as function of time $[\mu\mathrm{s}]$). The duration time is 8 μs while the bandwidth is 8 MHz.

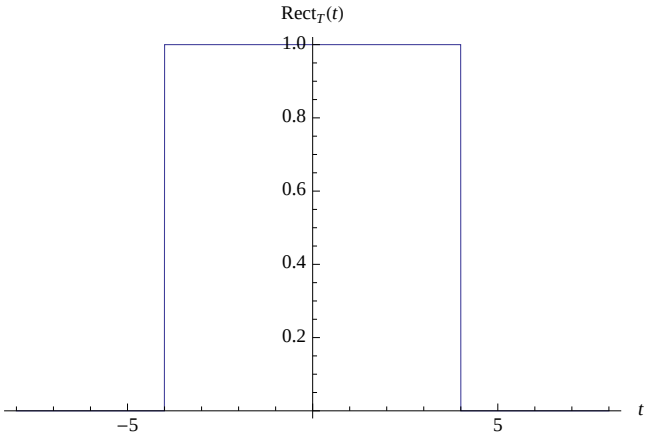


FIGURE 5 – Chirp window. The signal, shown in Figure 3, is obtained by the multiplication of the infinitive chirp with a $\mathrm{rect}_T(t)$ function of duration $T = 8$ μs.

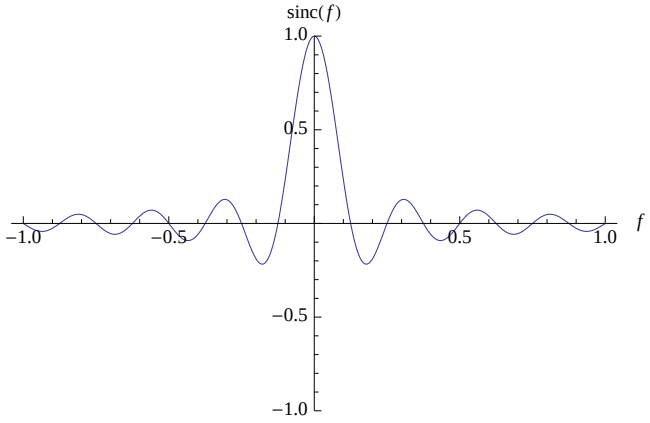


FIGURE 6 – The normalized $\mathrm{sinc}(f) = \sin(\pi f)/\pi f$ function corresponds to the Fourier transform of the $\mathrm{rect}_T(t)$ shown in Figure 5. The first zero is at $f = 1/8\ [\mu s]^{-1} = 0.125$ MHz.

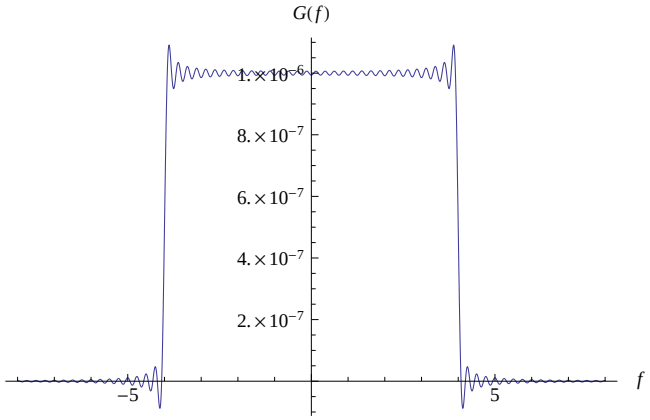


FIGURE 7 – The spectrum $G(f)$ of the chirp, shown in Figure 3, is given by the convolution of the $\mathrm{sinc}(f)$ of Figure 6 and the Fourier transform of the chirp function which results in a constant function. The spectrum $G(f)$ is approximately a $\mathrm{rect}_B(f)$ where $B \approx 8\ MHz$ and the height is $\approx \sqrt{T/B} = 10^{-6}$s. The approximation with a $\mathrm{rect}_B(f)$ is better when the product $T \cdot B$, which is called *compression factor*, is high because the Gibbs oscillation effect decreases.

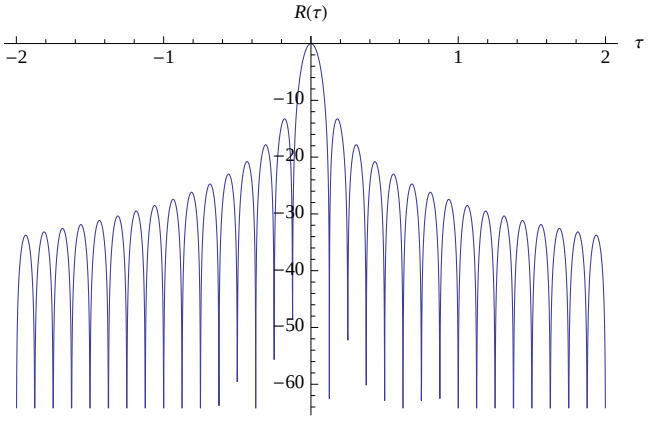


FIGURE 8 – Chirp correlation function corresponding to the chirp shown in Figure 3. The time is in $[\mu s]$ and the dependent variable is in logarithmic scale. The energy is concentrated in the centre, this results in a better resolution. The main lobe has a duration $\tau \approx 1/B = 0.125\ \mu s$. The side lobes are lower than $-13\ \mathrm{dB}$.

time, measured in μs : the duration and the bandwidth are the main characteristics of the chirp signal which define the slope of the line $k = B/T$.
The spectrum $G(f)$ of the chirp signal can be computed with the help of the stationary phase (Hein, 2004). Intuitively it is composed by an ensemble of frequencies, thus the spectrum has a constant amplitude in the bandwidth. On the other hand, the constant is convolved with the $\mathrm{sinc}(f)$ function, shown in Figure 6, obtained by the Fourier transform of the $\mathrm{rect}_T(t)$. It results in the spectrum shown in Figure 7 which corresponds approximately to a $rect_B(f)$.
The autocorrelation function $R(\tau)$, shown in Figure 8, corresponds to the output of the matched filter and gives the resolution, which is equal to

$$\delta_{sr}^{\mathrm{chirp}} = \frac{c}{2B} \tag{1.5}$$

where B is the bandwidth of the chirp. The example in Figure 8 shows that the energy is concentrated in a time of $\sim 0.125\ \mu$s, thus, considering the chirp in Figure 3 of duration $8\ \mu$s and bandwidth $8\ \mathrm{MHz}$, the resolution increases from $\sim 1\ \mathrm{Km}$, for a non-modulated pulse, to less than $20\ \mathrm{m}$, when the linear chirp is employed.
In case of TerraSAR-X, the chirp has a duration of 30.7 μs and an available bandwidth up to $300\ \mathrm{MHz}$ (Stangl et al., 2006), thus, according to Equation 1.5 the resolution is up to 1 m.

1.1.1.2 Azimuth resolution

The motion of the platform is used in order to improve the resolution in the along track direction by generating a synthetic aperture more extended than the physical one. The synthetic aperture is obtained by observing that a generic point P is illuminated by the antenna beam at different orbit positions, thus several pulses are collected and coherently summed. The geometry of the SAR in azimuth is shown in Figure 9.
The azimuthal extension of the illuminated area is

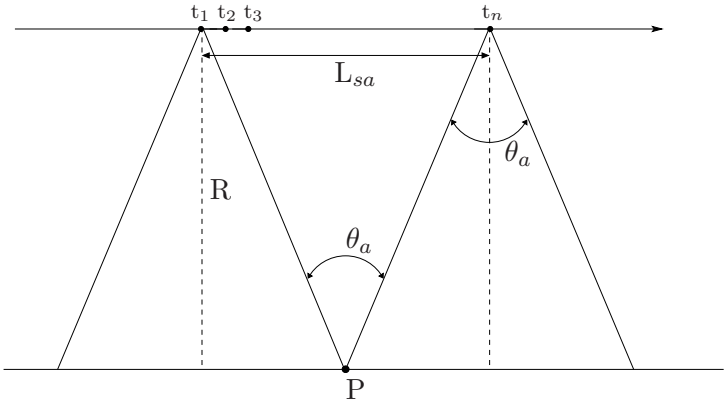


FIGURE 9 – Geometry for SAR resolution in azimuth or cross-range direction.

$$\mathrm{L}_{sa} = R \cdot \theta_a \tag{1.6}$$

thus the resulting azimuth resolution is

$$\delta_{az} = \mathrm{R}\frac{\lambda}{2 \cdot \mathrm{L}_{sa}} = \mathrm{R}\frac{\lambda}{2 \cdot R \cdot \theta_a} \approx \frac{\lambda}{2 \cdot \frac{\lambda}{\mathrm{L}_a}} = \frac{\mathrm{L}_a}{2} \tag{1.7}$$

where in the first equality the factor $1/2$ takes into account the two-way path, the second equality is obtained by replacing Equation 1.6 in Equation 1.7 and the approximation is obtained using Equation 1.1. The resolution results approximately equal to half of the size of the physical antenna.
The phase of the signal is

$$\begin{aligned} \phi(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0) = -\frac{4\pi}{\lambda}\mathrm{R}(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0) &= -\frac{4\pi}{\lambda}\sqrt{\mathrm{r}_0^2 + \mathrm{v}^2(\mathrm{t}-\mathrm{t}_0)^2} && (1.8)\\ &\approx -\frac{4\pi}{\lambda}\left(\mathrm{r}_0 + \frac{\mathrm{v}^2}{2\mathrm{r}_0}(\mathrm{t}-\mathrm{t}_0)^2\right) && (1.9) \end{aligned}$$

where r_0 is the slant range distance and the target is considered to be at location of minimum distance at azimuth time $\mathrm{t} = \mathrm{t}_0$. The last equation is obtained by considering a parabolic approximation for the distance function $\mathrm{R}(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0)$. The signal results modulated with a frequency

$$\begin{aligned} \mathrm{f_D}(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0) &= \frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{dt}}\phi(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0) = -\frac{2}{\lambda}\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{R}(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0)\\ &= -\frac{2}{\lambda}\frac{\mathrm{v}^2}{\mathrm{R}(\mathrm{t}-\mathrm{t}_0;\mathrm{r}_0)}(\mathrm{t}-\mathrm{t}_0) \approx -\frac{2}{\lambda}\frac{\mathrm{v}^2}{\mathrm{r}_0}(\mathrm{t}-\mathrm{t}_0) \end{aligned} \tag{1.10}$$

which is a function of the slant range distance r_0 and of the azimuth time position $(\mathrm{t}-\mathrm{t}_0)$. $\mathrm{f_D}$ is also known as *Doppler frequency*.
In other words, the Doppler effect modulates the signal in azimuth with a chirp exploiting the fact that the distance between the sensor and the target changes in time. According to Equation 1.10, the time $\mathrm{t} = \mathrm{t}_0$, i.e. when the distance between the target and the sensor is minimum, is called zero Doppler frequency. In case of TerraSAR-X (stripmap mode), which has an antenna of size $\mathrm{L}_a = 4.8$ m the resolution in azimuth is up to 2.4 m, according to Equation 1.7.

1.1.2 Radar equation

The radar equation has the following form

$$\mathrm{P_R}(\tau) = \mathrm{P_T}(\tau - 2\mathrm{R}/\mathrm{c}) \cdot \sigma \cdot \frac{\lambda^2 \cdot |\varrho_\theta(\theta)|^2}{(4\pi)^3 \cdot \mathrm{R}^4} \cdot |\varrho_{\theta_a}(\theta_a)|^2 \tag{1.11}$$

where $\mathrm{P_R}$ is the received power, $\mathrm{P_T}$ is the transmitted power, λ is the wavelength, σ is the Radar Cross Section (RCS), $|\varrho_\theta(\theta)|^2$ is the two-way antenna gain in range, $|\varrho_{\theta_a}(\theta_a)|$ is the two-way antenna gain in azimuth and R is the distance of the illuminated target.
In order to have an idea of the order of magnitude of the transmitted/received powers : the transmitted power is of the order of KW, while the received power is of the order of mW because of the term R^{-4}, for real aperture radar.

1.1.3 SAR Impulse Response

The SAR system impulse response can be approximated by the following expression

$$\mathrm{h}_a(\tau, \mathrm{t}; \mathrm{r}_0) = C(\mathrm{r}_0) \cdot \varrho_{\theta_a}\left(\frac{\mathrm{vt}}{\mathrm{r}_0}\right) \cdot g\left(\tau - \frac{2\mathrm{R}(\mathrm{t}; \mathrm{r}_0)}{\mathrm{c}}\right) \cdot \exp\{\mathrm{j}\pi \mathrm{f_D} \mathrm{t}^2\} \tag{1.12}$$

where t is respect to the time of closest approach t_0,

$$C(\mathrm{r_o}) = \lambda \cdot \varrho_\theta(\theta) \cdot \varrho_{rec}/(\mathrm{R}^2 \cdot (4\pi)^{3/2}) \tag{1.13}$$

includes the gain of the receiving chain ϱ_{rec} and the one-way terms in the radar equation 1.11. The term $\varrho_{\theta_a}(\mathrm{v}t/r_0)$ takes into account the time dependency antenna pattern in azimuth direction, because the same point (target) is seen with different antenna gain in time. The delayed sent pulse is $g(\tau - 2\mathrm{R}(\mathrm{t}; \mathrm{r}_0)/\mathrm{c})$ and $\exp\{\mathrm{j}\pi \mathrm{f_D} \mathrm{t}^2\}$ is the azimuth phase modulation term of Equation 1.8. The latter term is normalized to respect the distance of closest approach r_0 by multiplication by $\exp\{\mathrm{j}4\pi \mathrm{r}_0/\lambda\}$ and it is obtained by the approximation of parabolic distance in Equation 1.9.

1.2 SAR radiometry and geometry

Several basic backscattering mechanisms depend on micro- and macroscopic properties of the scatterers. The reflection mechanisms (Woodhouse, 2006) are shown in Figure 10 and are listed below :

- Reflection from smooth surfaces, i.e. mirror effect, Figure 10a ;
- Double bouncing, Figure 10b ;
- Reflection from rough surfaces (diffraction), Figure 10c ;
- Volume scattering, Figure 10d.

The effect shown in Figures 10a and 10c depends on the roughness of the surface. The roughness of a surface, in electromagnetism, depends on the wavelength and on the incident angle. A surface can be considered to be *rough*, fulfilling the following expression

$$\sigma_h > \frac{\lambda}{8 \sin \psi} \tag{1.14}$$

where σ_h is the Root Mean Square (RMS) value of the surface elevation, λ is the wavelength and ψ is the grazing angle (complementary to the incident angle). It is worth

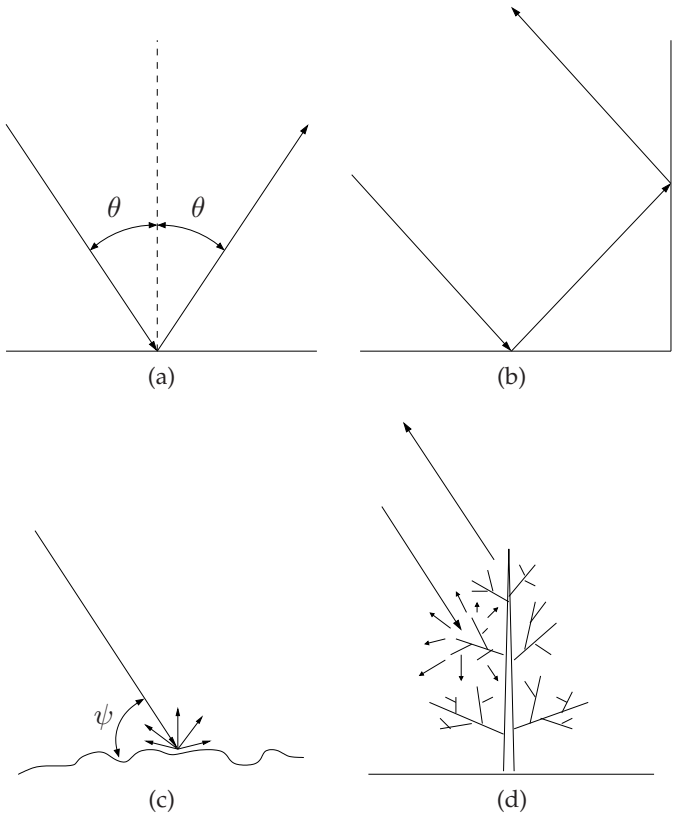


FIGURE 10 – Backscattering mechanisms : reflection (a) ; double bouncing (b) ; diffraction (c) ; volume scattering (d).

noting that by decreasing the grazing angle ψ the same surfaces can go from *rough* to *smooth*. In radar systems the aim of the measurement is to estimate the normalized RCS, σ_0. The operation to be done in order to go from the sensor measurement to the desired physical parameter is called *calibration*.
The factors which affect the calibration are the following : atmosphere, antenna, electronic, SAR processor, platform and downlink. Each of the previous factors plays a different role in the calibration which can be distinguished in *internal* and *external* calibration.
The internal calibration monitors the transmitted power and the antenna gain in order to keep their product constant. The external calibration is used to estimate the gain of the antenna by the use of extended distributed scatterers with known RCS.
The radiometric representation which allows parameter retrieval independent of the projection geometry is the normalized backscatter σ_0. Thus, the radar brightness β_0 (beta nought), which is a sensor dependent measure and is represented by the Digital Number pixel values DN, denoting the amplitude, has to be converted into σ_0 as follow (DLR, 2008)

$$\sigma_0 = (k_s \langle |DN|^2 \rangle - NEBN) \cdot \sin(\theta_i) \tag{1.15}$$

where θ_i is the local incident angle, k_s is the calibration and processor scaling factor for SAR signals annotated in each products and $NEBN$ is the noise equivalent β_0, which is only to be used for uncorrected products and derived from the noise profiles. For flat terrain or see surfaces, the annotated incidence angles are sufficiently accurate for this conversion - otherwise the local slopes from a terrain model have to be taken into account.

1.2.1 Geometric distortion effects

The images are affected by distortion effects due to the topography of the ground, which is not flat, the radar system, which is basically an instruments to measure distances, and the acquisition geometry, which depends on the incidence angle.
The topography of the terrain induces some geometric distortions. The effects are shown in Figure 11. The geocoding (Schreier, 1993) is a procedure to minimize geometrical distortions and resampling the image to a homogeneous, predefined map grid, e.g. to the Universal Tranversal Mercator (UTM) grid.

1.3 SAR Data Acquisition and Image Formation

The SAR data acquisition and image formation system is shown in Figure 12, together with the end-to-end SAR system. The raw data are obtained by

$$\begin{aligned} \mathrm{d}(\tau, \mathrm{t}) &= \iint_{-\infty}^{+\infty} \gamma^0(\mathrm{r}, \mathrm{t}') \cdot \mathrm{h_a}(\tau - 2\mathrm{r}/\mathrm{c}, \mathrm{t} - \mathrm{t}'; \mathrm{r}) \mathrm{dr} \mathrm{dt}' \\ &= \int_{-\infty}^{+\infty} \gamma^0(\mathrm{r}, \mathrm{t}) *_{\mathrm{t}} \mathrm{h_a}(\tau - 2\mathrm{r}/\mathrm{c}, \mathrm{t}; \mathrm{r}) \mathrm{dr} \end{aligned} \tag{1.16}$$

where the symbol $*_t$ denotes the convolution in the azimuth direction. In the hypothesis of a narrow swath, the approximation of a range invariant impulse response can be done, thus $\mathrm{h_a}(\tau - 2\mathrm{r}/\mathrm{c}, \mathrm{t}; \mathrm{r}) \approx \mathrm{h_a}(\tau - 2\mathrm{r}/\mathrm{c}, \mathrm{t}; \mathrm{r_0})$ and the integral in Equation 1.16 can be approximated as

$$\mathrm{d}(\tau, \mathrm{t}) = \frac{\mathrm{c}}{2} \gamma^0(\tau \mathrm{c}/2, \mathrm{t}) *_\tau *_{\mathrm{t}} \mathrm{h_a}(\tau, \mathrm{t}; \mathrm{r_0}) \tag{1.17}$$

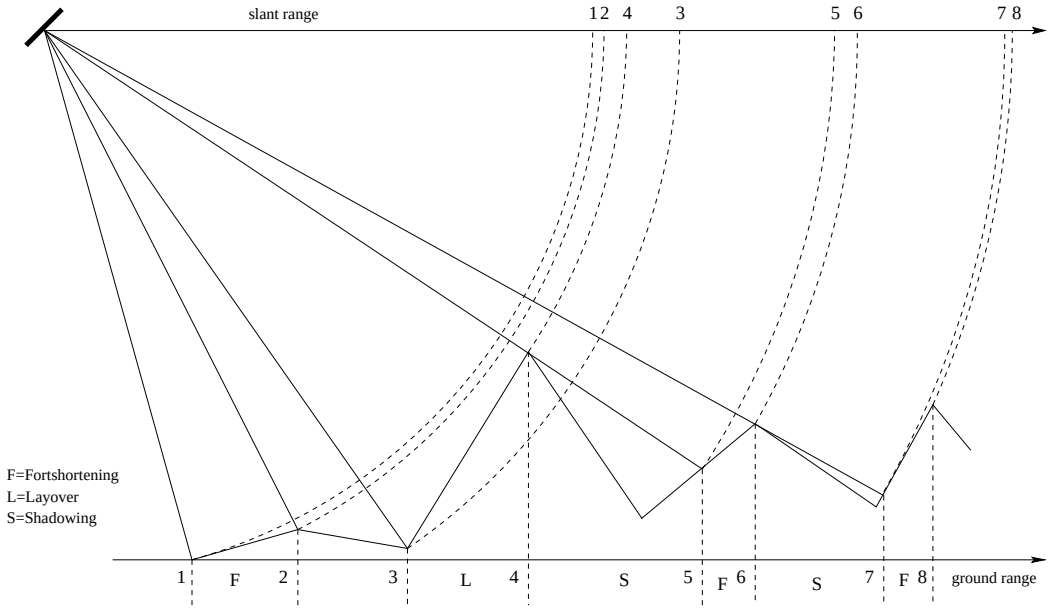


FIGURE 11 – Image geometric distortions due to the ground topography and the incident angle of the radar observation system. There are three distortion effects : fortshortening (F), layover (L) and shadowing (S), which affect the slant range projection.

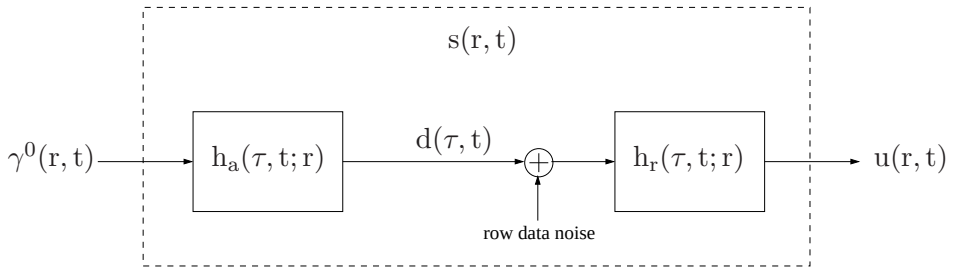


FIGURE 12 – End-to-end SAR system (Bamler & Schättler, 1993). $\gamma_0(\mathrm{r},\mathrm{t})$ is the complex reflectivity function, $\mathrm{h_a}(\tau,\mathrm{t};\mathrm{r})$ is the data acquisition system impulse response, $\mathrm{d}(\tau,\mathrm{t})$ represents the raw data, $\mathrm{h_r}(\tau,\mathrm{t};\mathrm{r})$ is the image formation system impulse response, $\mathrm{u}(\mathrm{r},\mathrm{t})$ is the complex image and $\mathrm{s}(\mathrm{r},\mathrm{t})$ is the equivalent end-to-end SAR system impulse response.

which corresponds to a range invariant convolution.
According to Figure 12 the complex image is obtained by

$$\mathrm{u}(\tau,\mathrm{t}) = \iint_{-\infty}^{+\infty} \mathrm{d}(\tau,\mathrm{t}') \cdot \mathrm{h_r}(\tau - 2\mathrm{r}/\mathrm{c}, \mathrm{t} - \mathrm{t}'; \mathrm{r})\mathrm{d}\mathrm{r}\mathrm{d}\mathrm{t}'. \tag{1.18}$$

Equation 1.18 represents the SAR focusing process. Although, it looks simple, it has some peculiarities which can make it arduous to solve

- The support of $\mathrm{h_r}(\cdot)$ can be as large as a hundred range samples (due to range migration) and several thousand azimuth samples, which doesn't allow the direct (time domain) implementation of Equation 1.18 in most cases.
- Equation 1.18 is *range-variant*, i.e. an implementation via a two-dimensional Fast Fourier Transform (FFT) and a single spectral filter multiplication is only possible within a narrow range segment.
- Due to range migration $\mathrm{h_r}(\cdot)$ is inherently two-dimensional and non-separable. Hence, the range-variance cannot be accounted for by simply using range dependent one-dimensional azimuth correlation kernels.

Thus, SAR image formation requires a range-variant two-dimensional linear filter operation. The common approaches to data focusing are three (Carrara et al., 1995) :

1. Chirp scaling (Raney, 1992; Raney et al., 1994).
2. Polar reformatting (Walker, 1980).
3. Range Migration Algorithm (RMA) (Cafforio et al., 1991).

Each method has advantages and disadvantages. The first method is implemented for TerraSAR-X.

A commonly adopted inversion strategy is correlation or matched filtering, for which an inverted and delayed replica of the signal is used

$$\mathrm{h_r}(\tau,\mathrm{t};\mathrm{r}) \propto \mathrm{h_a^*}(-\tau,-\mathrm{t};\mathrm{r}) \tag{1.19}$$

where the notation $\cdot^*$ denotes the complex conjugate. Thus, Equation 1.18 becomes

$$\begin{aligned} \mathrm{u}(\tau,\mathrm{t}) &\propto \iint_{-\infty}^{+\infty} \mathrm{d}(\tau,\mathrm{t}') \cdot \mathrm{h_a^*}(2\mathrm{r}/\mathrm{c} - \tau, \mathrm{t}' - \mathrm{t}; \mathrm{r})\mathrm{d}\tau\mathrm{d}\mathrm{t}' \\ &= \int_{-\infty}^{+\infty} \mathrm{d}(\tau,\mathrm{t}) \otimes_{\mathrm{t}} \mathrm{h_a}(2\mathrm{r}/\mathrm{c} - \tau, \mathrm{t}; \mathrm{r})\mathrm{d}\tau \end{aligned} \tag{1.20}$$

where the symbol $\otimes_t$ denotes the correlation operation between $\mathrm{d}(\tau,\mathrm{t})$ and $\mathrm{h_a}(2\mathrm{r}/\mathrm{c} - \tau, \mathrm{t}; \mathrm{r})$ in the azimuth direction. In the hypothesis of narrow swath, the approximation of range invariance is valid, thus 1.20 can be rewritten as follow

$$\begin{aligned} \mathrm{u}(\tau\mathrm{c}/2,\mathrm{t}) &\approx \mathrm{d}(\tau,\mathrm{t}) *_\tau *_{\mathrm{t}} \mathrm{h_r}(\tau,\mathrm{t};\mathrm{r}_0) \\ &\propto \mathrm{d}(\tau,\mathrm{t}) \otimes_\tau \otimes_{\mathrm{t}} \mathrm{h_r}(\tau,\mathrm{t};\mathrm{r}_0). \end{aligned} \tag{1.21}$$

Considering the dashed block in Figure12, the SAR end-to-end impulse response $\mathrm{s}(\mathrm{r},\mathrm{t})$ is defined as

$$\mathrm{s}(\mathrm{r},\mathrm{t}) \propto \mathrm{spline}\left(\frac{2\mathrm{v}}{\mathrm{L}}\mathrm{t}\right) \cdot \mathrm{sinc}\left(\frac{2\mathrm{B}}{\mathrm{c}}\mathrm{r}\right) \mathrm{e}^{\mathrm{j}2\pi\mathrm{f_{DC}}\mathrm{t}} \tag{1.22}$$

where $\mathrm{f_{DC}}$ is the frequency where the spectrum in azimuth has the maximum, i.e. Doppler centroid, and the $\mathrm{spline}(\cdot)$ is defined as

$$\mathrm{spline}(x) = \begin{cases} 2/3 - x^2 + |x|^3/2 & \text{for} \quad |x| \leq 1 \\ 4/3 - 2|x| + x^2 - |x|^3/6 & \text{for} \quad 1 < |x| \leq 2 \\ 0 & \text{else.} \end{cases}$$

and it has been obtained by considering the antenna pattern $\varrho_{\theta_a} \propto \mathrm{sinc}(\cdot)^2$. The end-to-end impulse response $\mathrm{s(r,t)}$ is range invariant.

1.4 Representation of the Information

The SAR data are stored in the I and Q channels for real and imaginary part respectively : an example is shown in Figures 13a and 13b. Evidently this way of visualization is not really easy to interpret and the content of the image is not easily understandable. Alternative ways of image representation are shown in Figures 13c, 13e and 13f. The phase of the image is shown in Figure 13d for completeness. Different statistics of the data correspond to each representation as shown in Section 1.5. An alternative, but equivalent representation is in frequency domain. The range and azimuth profile of the 2D-spectrum are shown in Figures 14a and 14b.

1.4.1 Image Transformations

The image can be transformed in different ways in order to obtain the information in the desired form. Some transformation methods are listed below.

- Cartesian to polar ;
- Linear transformation ;
- Logarithm ;
- Wavelet ;
- etc.

The methods are shortly presented and described in the following paragraphs.

1.4.1.1 Transformation from Cartesian to polar coordinates

The transformation from Cartesian to polar coordinates in a 2-dimensional space is given by the following equations

$$A = \sqrt{\mathrm{x}_1^2 + \mathrm{x}_2^2} \qquad 0 \geq A < \infty \tag{1.23}$$

$$\theta = \arctan \frac{\mathrm{x}_2}{\mathrm{x}_1} \qquad -\pi \geq \theta < \pi \tag{1.24}$$

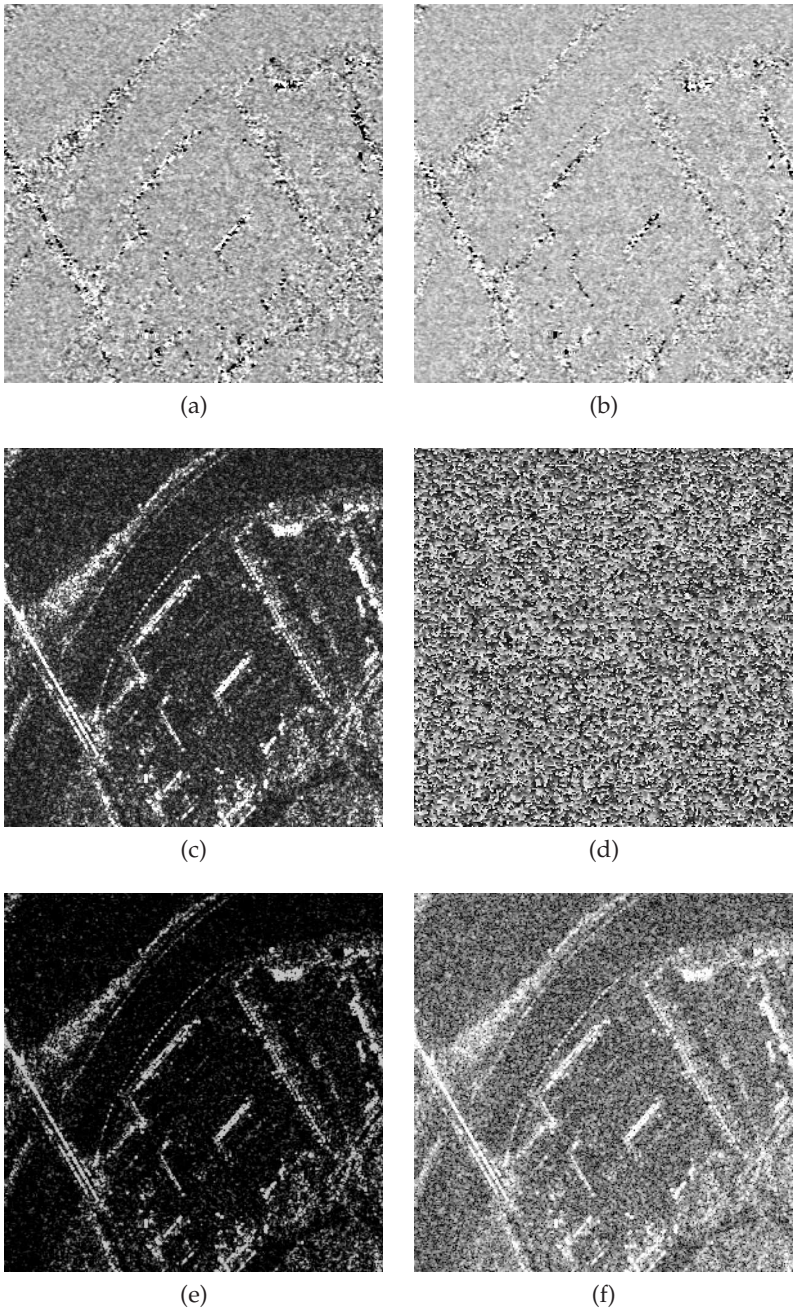
The inverse transformation is

$$\mathrm{x}_1 = A\cos\theta \tag{1.25}$$

$$\mathrm{x}_2 = A\sin\theta. \tag{1.26}$$

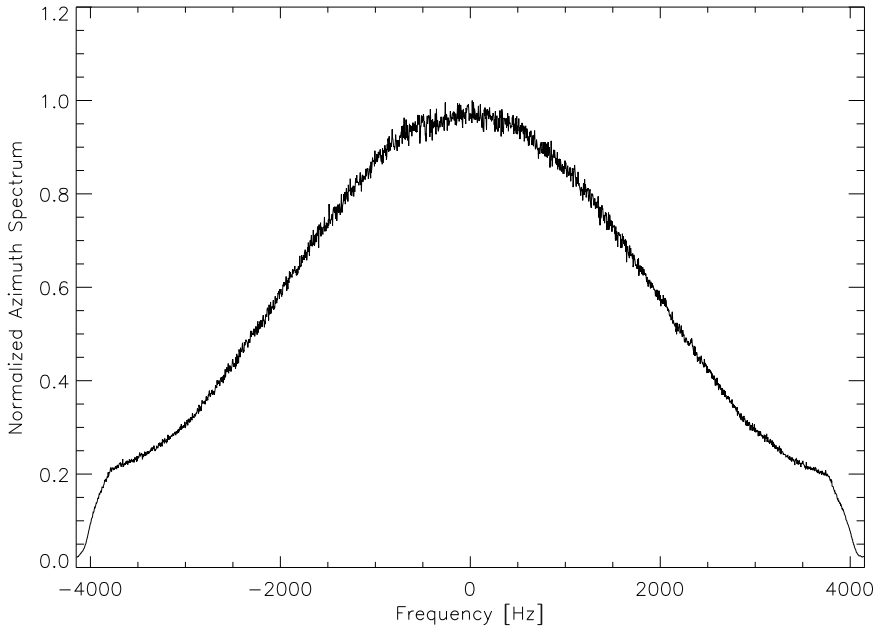
The Jacobian is needed for transforming the pdf

$$\det \frac{\partial(\mathrm{x}_1, \mathrm{x}_2)}{\partial(A, \theta)} = \begin{vmatrix} \cos\theta & -A\sin\theta \\ \sin\theta & A\cos\theta \end{vmatrix} = A\cos^2\theta + A\sin^2\theta = A. \tag{1.27}$$

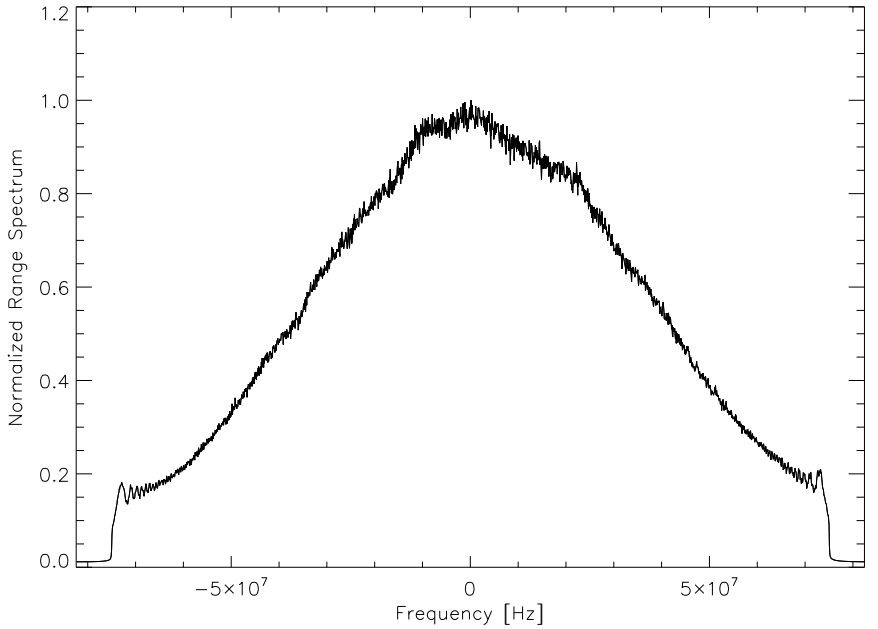


(a) (b) (c) (d) (e) (f)

FIGURE 13 – Possible representation of SAR data image : (a) real part, (b) imaginary part, (c) amplitude, (d) phase, (e) intensity and (f) log-intensity.



(a) Azimuth



(b) Range

FIGURE 14 – Normalized amplitude spectrum in azimuth (a) and range (b). The frequency range respects the real bandwidth of a SL TerraSAR-X image. In this case is 8.3 KHz the azimuth bandwidth and 164.8 MHz the range bandwidth.

Given the pdf of the random variables in Cartesian coordinates $p(\mathrm{x}_1, \mathrm{x}_2)$ one finds the pdf in polar coordinates

$$p(A, \theta) = A \cdot p(\mathrm{x}_1, \mathrm{x}_2) = A \cdot p(A\cos\theta, A\sin\theta) \tag{1.28}$$

with

$$A > 0 \qquad 0 \geq \theta > 2\pi \tag{1.29}$$

Further we consider an important case : x_1 and x_2 are independent random variables with normal pdf of zero mean and the same variance σ^2

$$\begin{aligned} p(\mathrm{x}_1, \mathrm{x}_2) &= p(\mathrm{x}_1)p(\mathrm{x}_2) & (1.30)\\ &= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\mathrm{x}_1^2 + \mathrm{x}_2^2}{2\sigma^2}\right) & (1.31)\end{aligned}$$

The real and imaginary parts of a SAR image together with their pdf are shown in Figures 13a,13b and Figures 16a, 16b, respectively.
In polar coordinates the pdf has the following form

$$p(A, \theta) = \frac{A}{2\pi\sigma^2}\exp\left(-\frac{A^2}{2\sigma^2}\right) \tag{1.32}$$

where

$$p(A) = \frac{A}{\sigma^2}\exp\left(-\frac{A^2}{2\sigma^2}\right) \tag{1.33}$$

is a Rayleigh distribution and

$$p(\theta) = \frac{1}{2\pi} \tag{1.34}$$

is a uniform distribution. An example of *A-image* and *θ-image*, are the amplitude and phase image shown in Figures 13c and 13d, with the respective density functions shown in Figures 16c and 16d.

1.4.1.2 Linear Transformation

We consider the vectors of random variables $\mathrm{X} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]$ and $\mathrm{Y} = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_n]$. Without loss of generality we consider the random variables to have zero mean. A linear transformation is given by

$$\mathrm{Y} = \mathrm{AX} \tag{1.35}$$

where A with $n \times n$ elements is the matrix characterizing the transformation. With simple computation we can obtain

$$\begin{aligned} \mathrm{E}\{\mathrm{Y}\} &= \mathrm{AE}\{\mathrm{X}\} & (1.36)\\ \mathrm{C_Y} &= \mathrm{E}\{\mathrm{YY}^T\} = \mathrm{AE}\{\mathrm{XX}^T\}\mathrm{A}^T = \mathrm{AC_X A}^T & (1.37)\\ \mathrm{C_{XY}} &= \mathrm{E}\{\mathrm{XY}^T\} = \mathrm{E}\{\mathrm{XX}^T\}\mathrm{A}^T = \mathrm{C_X A}^T & (1.38)\\ \mathrm{C_{YX}} &= \mathrm{C}_{\mathrm{XY}}^T & (1.39)\end{aligned}$$

An important transformation is the Karhunen-Loève transform. We can write

$$\mathrm{C_X} = \mathrm{M}\Lambda\mathrm{M}^T \tag{1.40}$$

where Λ is the diagonal matrix of eigenvalues of $\mathrm{C_X}$, and M is the othogonal matrix having the eigenvectors of $\mathrm{C_X}$ as columns. It follows

$$\Lambda = \mathrm{M}^T \mathrm{C_X} \mathrm{M}. \tag{1.41}$$

The Karhunen-Loève transform has the matrix

$$\mathrm{A} = \mathrm{M}^T \tag{1.42}$$

and transforms X in Y having a diagonal matrix $\mathrm{C_Y} = \Lambda$. The Karhunen-Loève transform uncorrelates the components of the random variable vector X.

1.4.1.3 Logarithm Transformation

The Logarithm transforms multiplicative noise $\mathrm{x} \cdot \mathrm{n}$ in additive noise $\log(\mathrm{x}) + \log(\mathrm{n})$. After the transformation, a linear estimator can be applied in order to retrieve the quantity $\log(x)$
An example can be given for speckle which is successfully modelled as multiplicative noise. In the following, the statistics of speckle are investigated after logarithmic transformation. By applying the square function to transform a Rayleigh distributed random variable, an exponential distribution is obtained

$$p(\mathrm{y}|\mathrm{x}) = \frac{1}{\mathrm{x}} \exp\left\{-\frac{\mathrm{y}}{\mathrm{x}}\right\} \qquad \mathrm{y} \geq 0 \tag{1.43}$$

$$p(\mathrm{y}|\mathrm{x}) = 0 \qquad \mathrm{y} < 0 \tag{1.44}$$

The pdf $p(\mathrm{y}|\mathrm{x})$ is the likelihood expressing the incertitude introduced by the noise : the speckle process. The latter can be successfully modeled as multiplicative noise

$$\mathrm{y} = \mathrm{x} \cdot \mathrm{n} \tag{1.45}$$

Thus the speckle noise is described by

$$p(\mathrm{n}) = \exp\{-\mathrm{n}\} \qquad n \geq 0 \tag{1.46}$$

$$p(\mathrm{n}) = 0 \qquad n < 0$$

Using the transformation

$$f(\cdot) = \log(\cdot) \tag{1.47}$$

we obtain the pdf of the process

$$\mathrm{n}' = \log \mathrm{n} \tag{1.48}$$

$$p(\mathrm{n}') = \exp(\mathrm{n}' - \mathrm{e}^{\mathrm{n}'}) \tag{1.49}$$

where the last equation is obtained by applying the Jacobian of the transformation 1.48 to equation 1.46 and corresponds to the Fisher-Tippet distribution.
Another method to obtain an additive signal dependent noise from the product model 1.45 is the following (Gleich & Datcu, 2007a)

$$\mathrm{y} = \mathrm{x} + \mathrm{x}(\mathrm{n} - 1) = \mathrm{x} + \mathrm{n}' \tag{1.50}$$

where equation 1.50 is obtained by summing and subtracting the signal x.

1.5 SAR statistics

The SAR images are dominated by the speckle effect, which is recognizable as the occurrence of bright and dark pixels in uniform backscattering areas.
The phenomenon has a physical origin and belongs to coherent imaging systems, e.g. SAR, laser, etc. It is a deterministic phenomenon, thus if all the conditions were reproducible, two images would have the same appearance in term of speckle. This event is exceptional because it would require the same orbit for the satellite, the same look angle of the sensor, no changes on the scene and so on. Thus, the speckle effect is successfully modelled as a stochastic process because of the actual non reproducibility of the same event. Thus, the pixel values are random variables, affected by multiplicative noise, which allow the estimation of the RCS.

1.5.1 Physical origin of speckle

The physical origin of the speckle is due to the coherent summation of complex valued components. The process is described in Figure 15. In Figure 15a the model of a resolution cell characterized by some scatterers is shown. The difference on the path length depends on the relative distance of the scatterers on the ground and the sensor incidence angle. Thus, $\delta = l \cdot \sin\theta$, where l is the distance in the ground of the two scatterers of Figure 15a and θ is the incidence angle (see Figure 2). The phase difference corresponds to

$$\Delta\varphi = \frac{4\pi}{\lambda}\delta \tag{1.51}$$

where $2\pi/\lambda$ is the wave number and the factor 2 comes from the two-way path. The reflected components are summed up according to

$$u = A\mathrm{e}^{j\varphi} = \sum_{k=1}^{n} A_k \mathrm{e}^{j\varphi_k} \tag{1.52}$$

where $u = u_r + ju_i$, $A = |u|$ and $\varphi = \arctan(u_y/u_x)$. An example is shown in Figure 15b for $n = 8$ components. The process corresponds to a random walk in the complex plane. A constructive or destructive interference, which causes high reflectivity or low reflectivity respectively, can occur because of the coherent summation.
The speckle is said to be fully developed if

1. The number of scatterers n per resolution cell is high ;
2. The amplitude and the phase of scatterer k are independent of those of the $n-1$ other scatterers ;
3. The amplitude and phase of each scatterer are independent ;
4. The amplitudes are identically distributed for all scatterers ;
5. The phases of the scatterers are uniformly distributed.

Under the above assumptions the single look distributions can easily be derived.

1.5.2 Data statistics

The estimated distributions corresponding to the images shown in Figure 13 are shown in figure 16. The real and imaginary part, shown in Figures 16a and 16b, follow both a

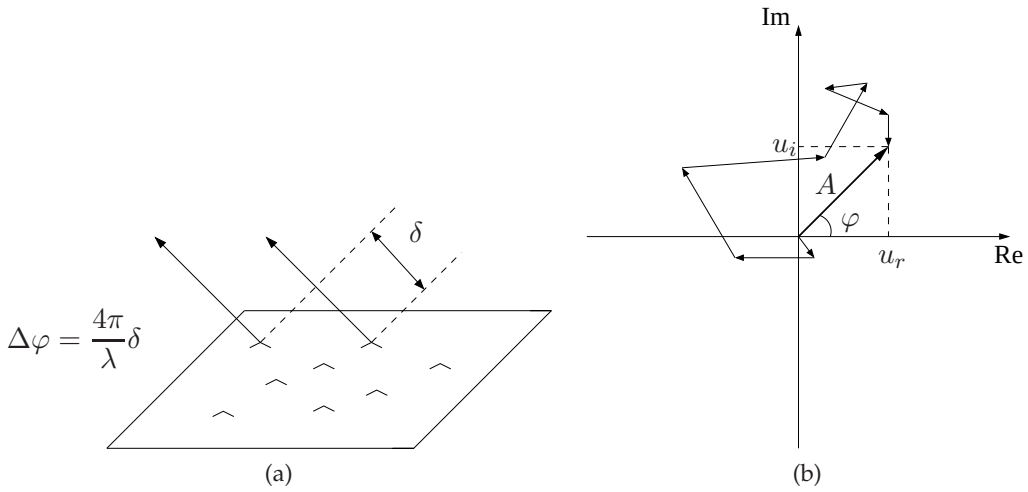


FIGURE 15 – Speckle physical genesis. (a) the model of a resolution cell with scatterers. (b) an example of a coherent vector summation in the complex plane.

zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$ according to the central limit theorem, assuming the number of scatterers n to be large. We observe that $2\sigma^2$ is equivalent to the radar reflectivity, which, neglecting the sensor noise, is proportional to the RCS. According to the Quadratic Amplitude Modulation (QAM), the real and imaginary channels are uncorrelated and are realizations of a proper complex random process (Schreier & Scharf, 2003). The non-linear tranformation from Cartesian to polar coordinates yields to the following pdf for amplitude

$$p(A) = \frac{A}{\sigma^2} \exp\left(-\frac{A^2}{2\sigma^2}\right) \qquad A \geq 0 \tag{1.53}$$

which is a Rayleigh distribution and phase

$$p(\theta) = \frac{1}{2\pi} \qquad 0 \leq \theta < 2\pi \tag{1.54}$$

which is a uniform distribution. The distribution of the intensity $I = A^2$ is the exponential distribution

$$p(I) = \frac{1}{\sigma^2} \exp\left(-\frac{I}{2\sigma^2}\right) \qquad I \geq 0 \tag{1.55}$$

which can be obtained from 1.53 applying the transformation $A = \sqrt{I}$. The estimated Rayleigh and uniform distribution for amplitude and phase are shown in Figures 16c and 16d, respectively. The estimated intensity is shown in Figure 16e). The logarithmic transformation of the intensity yields to a Fisher-Tippet distribution (Figure 16f).

1.5.3 Scatterers data statistics

In presence of a strong scatter, locally, the mean of the Gaussian distributions is not zero anymore. Thus the observed amplitude can be modelled by a Rice distribution which has the following form

$$p(A) = \frac{A}{\sigma^2} \exp\left(-\frac{A^2 + A_T^2}{2\sigma^2}\right) I_0\left(\frac{A_T A}{\sigma^2}\right) \qquad A \geq 0 \tag{1.56}$$

where $A_T \geq 0$ is the amplitude of the target and $I_0(\cdot)$ is the modified Bessel function of first kind of order 0

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{x\cos\theta}\mathrm{d}\theta \tag{1.57}$$

Considering the number of scatterers n in the resolution cell as a random variable, the resulting signal intensity has a K distribution (Jakeman & Pusey, 1976, 1978; Oliver, 1984; Jao, 1984) of the form

$$p(I) = \frac{2}{a\Gamma(\nu)}\left(\frac{I}{2a}\right)^{\nu} K_{\nu-1}\left(\frac{I}{a}\right) \qquad I > 0 \tag{1.58}$$

where a is the scale parameter and $\nu > 0$ is the shape parameter, $\Gamma(\cdot)$ is the Gamma function

$$\Gamma(\mathrm{x}) = \int_0^{\infty} t^{x-1}\mathrm{e}^{-t}\mathrm{d}t \tag{1.59}$$

and $K_n(\cdot)$ is the modified Bessel function of second kind of order n

$$K_n(x) = \frac{(2x)^n}{\sqrt{\pi}}\Gamma\left(n+\frac{1}{2}\right)\int_0^{\infty}\frac{\cos t}{(t^2+x^2)^{n+1/2}}\mathrm{d}t \tag{1.60}$$

The K distribution can also be used to model the observed amplitude. This can be verified by performing the transformation $I = A^2$ in Equation 1.58 which leads again to a K distribution. An alternative method to obtain the K distribution is to consider σ modelled by a χ distribution in the Rayleigh distribution of Equation 1.53. Thus, the marginalization with respect to σ leads again to a K-distribution for the observed intensity. This method is successfully applied to model see clutter (Watts et al., 2006). In Lee et al. (1994) the intensity and phase statistics of multilook polarimetric and interferometric SAR imagery are shown and investigated. More complicated statistical models, than the ones presented above, with empirical or theoretic basis, have been proposed to handle speckle, e.g. log-normal, Weibull, Fisher, Gamma, K, Rice, Nakagami, generalized Gaussian, inverse Gaussian distributions, *heavy-tailed* Rayleigh.

For example, the Nakagami-Rice distribution has been proposed to model the SAR statistics in the presence of a single strong reflector in homogeneous clutter (Tison et al., 2004; Dana & Knepp, 1986). The Gamma distribution has been introduced as a model for a multilook SAR intensity pdf where the correspondent amplitude has a Nakagami distribution. They are a generalization of negative exponential and Rayleigh distributions respectively. In Kuruoglu & Zerubia (2004) a *heavy-tailed* Rayleigh model is presented under the hypothesis that the real and imaginary parts of the backscattered signal are jointly SαS random variables. The K distribution (Jakeman & Pusey, 1976, 1978; Oliver, 1984; Jao, 1984) is obtained for the signal intensity assuming the population of scatterers in the resolution cell to be controlled by a birth-death-migration process. In the hypothesis of Gamma distributed noise and Gamma distributed signal (Oliver, 1991; Oliver & Quegan, 2004a), the product model results in a K distribution for the intensity. Generalized Gaussian distributions are assumed for real and imaginary parts in Moser et al. (2006), while inverse Gaussian distributions have also been employed to model the amplitude statistics in Frery et al. (1997), Muller & Pac (1999) and Eltoft (2003). The $\mathcal{G}$ distribution for the amplitude, presented in Frery et al. (1997), is the result of a multiplicative model assuming a Nakagami distribution for the speckle noise and a generalized inverse Gaussian distribution for the signal. In Tison et al. (2004), the $\mathcal{G}^0$ distribution is proved to be equivalent to a Fisher pdf, while, in Muller & Pac (1999), a further particular case of the $\mathcal{G}$

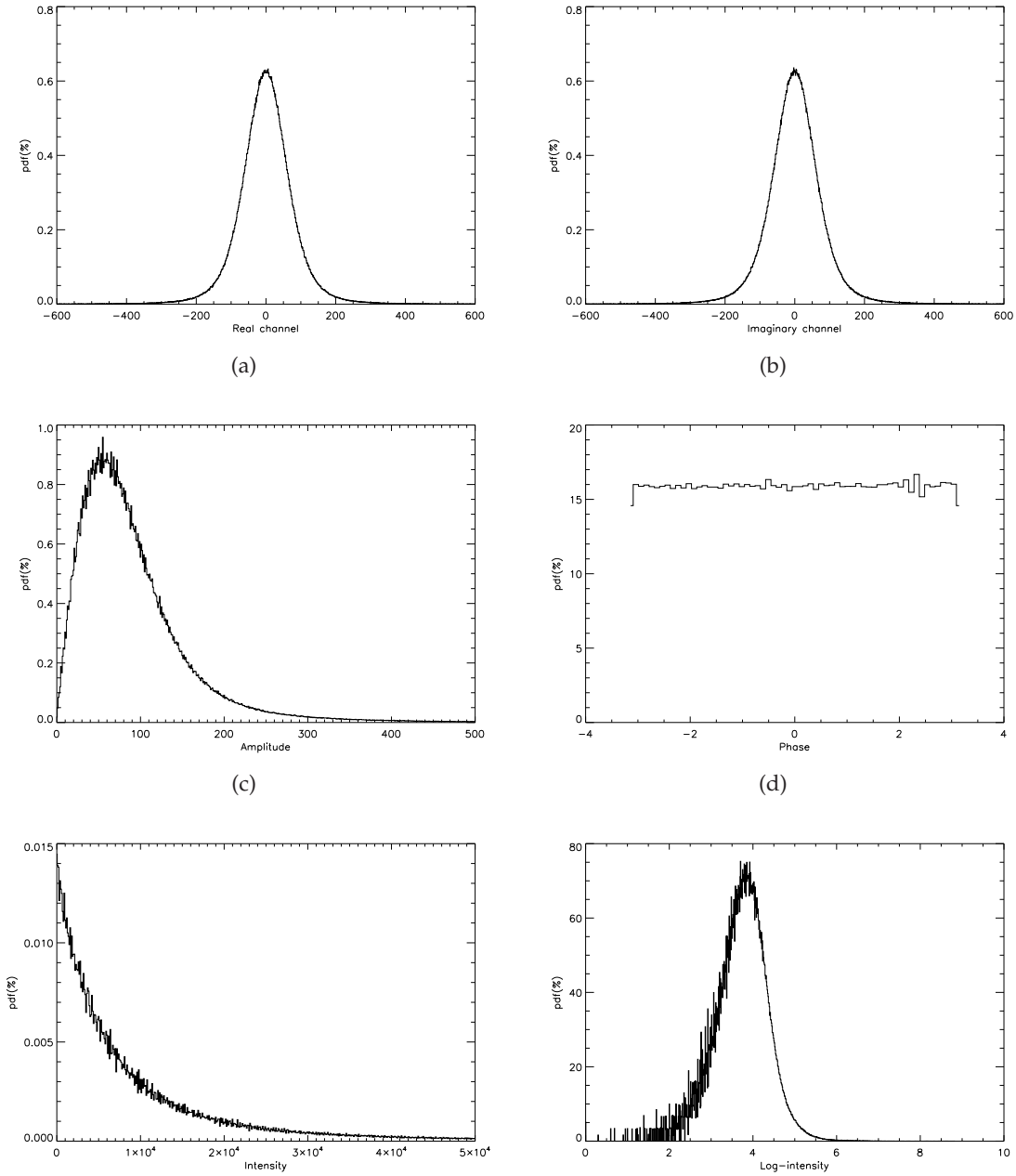


FIGURE 16 – SAR image data statistics, normalized estimated histograms : (a) real part (Gaussian), (b) imaginary part (Gaussian), (c) amplitude (Rayleigh), (d) phase (Uniform), (e) intensity (Exponential) and (f) log-intensity (Fisher-Tippet).

model, called *harmonic branch* $\mathcal{G}^h$, is proposed.
In addition to the above mentioned *theoretical* or partially theoretical models, several empirical models have been used to characterize the statistics of SAR amplitude (or intensity) data, such as Weibull (Sekine & Mao, 1990), log-normal (Szajnowski, 1977) and Pearson (Delignon et al., 1997).

1.5.4 Multilooking

The method of multilooking is a way to reduce speckle at the cost of resolution. In practice it consists of band-pass filtering the spectrum of the complex data in order to obtain L independent looks (realizations) of the image scene which are later summed incoherently. The procedure is basically the following :

1. Discrete Fourier Transform (DFT) of the complex image ;
2. Band pass filtering in order to obtain L parts of the spectrum ;
3. Inverse Discrete Fourier Transform (IDFT) of each look ;
4. Square law detection and incoherent summation of the looks.

The speckle is reduced by a factor of L. The statistics of the multi-look intensity data results in a Gamma distributed signal

$$p(I) = \frac{L^L I^{L-1}}{\Gamma(L)\sigma^{2L}} \exp\left(-\frac{LI}{2\sigma^2}\right) \tag{1.61}$$

where $\Gamma(\cdot)$ is the Gamma function defined in 1.59. Many areas of natural clutter are successfully modeled by a Gamma distributed radar reflectivity σ^2 (Oliver & Quegan, 2004a). Considering a multiplicative model, see section 1.7.1, with the speckle noise and the signal both modelled by a Gamma distribution, the observed intensity is K-distributed.
The principle of the multi-look is exploited by the so called azimuth splitting analysis. The method consists in splitting the azimuth spectra. If we refer to Figure 9 it means to split the antenna and to watch the target with different look angles with respect to the nadir. Referring to the illumination time : the first half time the target is illuminated with positive angles respect to the nadir, while for the second half of the time it is illuminated with negative angles. The angles in time correspond to the Doppler history, thus, in term of spectral analysis it means to discriminate variation in the Doppler frequency (e.g. target motion).
Depending on the value of L, the resolution can be drastically reduced. In order to preserve resolution the looks can be partially overlapped by appropriate weighting of the filters which maintain low the dependency of each look. In this case the actual intensity statistics will be a Goodman distribution (Bruniquel & Lopès, 1998) which, anyway, can be well approximated by the Gamma distribution by properly tuning the parameter L. In Figure 17 a possible configuration of the band-pass filter-bank is shown. Three Gaussian filters partially overlapped are shown together with the antenna weight. The spectrum (14a) has to be unweighted before filtering ; in azimuth the antenna pattern and the existing window have to be balanced in order to allow each look to have the same energy.
The windowing method (Oppenheim & Schafer, 1975) is often used to reduce the side lobe effect. Common used windows are Rectangular, Hamming, Hanning, Blackman, Bartlett, to which different heights of the side-lobes correspond : -13, -25, -31, -41,

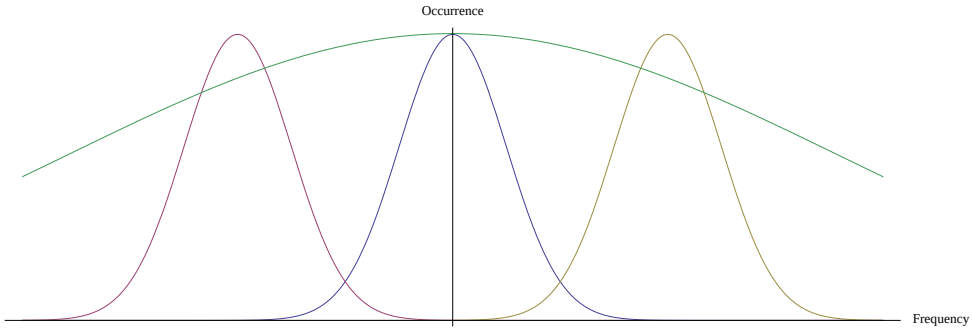


FIGURE 17 – Multi-look band pass filtering example. Three Gaussian filter bank partially overlapped. The curve above corresponds to the antenna weight.

-57 dB, respectively. On the other hand, the reduction of the side-lobes is paid in term of a larger main lobe which causes a blurring of the transitions. In time domain the multi-look may be done through spatial averaging of the observed intensity.

1.6 Speckle Reduction

The state-of-the-art of the despeckling of detected vs. complex data starts with the diagram shown in Figure 18. For each despeckling approach the main reference is provided. It is worth noticing the lack of investigation for complex-valued data in comparison with the real-valued image.

In Section 1.6.2, it is shown how the Aritmetic Mean Intensity (AMI) estimator is only a particular case of the Spatial Whitening Filter (SWF). Thus, the complex data has the full available information where the phase information is lost on detected data. Another interesting work on edge detection and localization on complex SAR imagery is presented by Fjortoft et al. (1999).

For texture modeling, it is interesting to notice that in Lopès et al. (1990a) the authors justify the use of detected images by saying that by taking a Gaussian distribution as underlying texture pdf model, one allows implicitly σ_0 to be negative with no physical significance. Thus, the authors introduce a MAP estimator for Gamma distributed scenes.

The use of the complex data is restricted to particular applications e.g. interferometry, Moving Target Detector (MTD), but not in despeckling applications or texture exploitations (Oliver & Quegan, 2004b). An automated approach in order to exploit the statistical simplicity of the complex data and the advantages introduced by the use of positive non-symmetric pdf, which is the case of transformed data, would be nice, but no algorithm has currently reached this level of sophistication.

1.6.1 Spatial averaging

Multilooking can be obtained in time domain by spatial averaging. We can consider a set of N complex samples, corresponding to N adjacent pixels in a complex radar image, where $\mathbf{Z}$ is a signal vector containing th complex amplitudes $Z_1, Z_2, \ldots, Z_n$. If the speckle is fully developed, the probability density function of the signal vector is a circulant

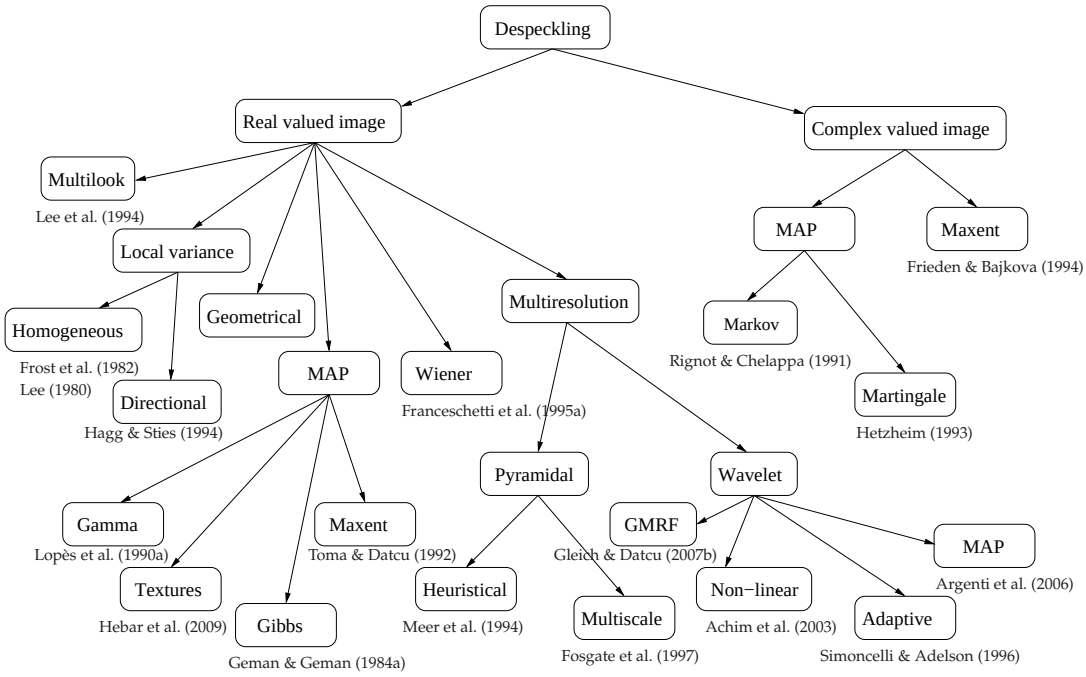


FIGURE 18 – Despeckling state-of-the-art.

complex Gaussian distribution

$$p(\mathbf{Z}) = \frac{1}{\pi^N |\mathbf{C_Z}|} \exp(-\mathbf{Z}^H \mathbf{C_Z}^{-1} \mathbf{Z}) \tag{1.62}$$

where $\mathbf{C_Z}$ is the $N \times N$ complex covariance matrix corresponding to signal vector $\mathbf{Z}$. If furthermore, we suppose that the underlying reflectivity R is constant, $\mathbf{Z} = \sqrt{R}\mathbf{S}$, where $\mathbf{S}$ is the speckle vector, the covariance matrix of the signal vector $\mathbf{Z}$ is given by (Lopès et al., 1993)

$$\mathbf{C_Z} = R \cdot \mathbf{C_S} \tag{1.63}$$

where $\mathbf{C_S}$ represents the covariance matrix of the speckle vector $\mathbf{S}$. The elements of $\mathbf{C_S}$ are the spatial correlation coefficients $\rho_S(\Delta x, \Delta y)$ of the speckle. The spatial correlation only depends on sensor and processor parameters. In Oliver & Quegan (2004c) a study on correlated texture is presented in case of approximated Gaussian autocorrelation function.

1.6.2 Estimators of the Mean Reflectivity

The Maximum Likelihood (ML) estimator of the radar reflectivity is the SWF given by

$$\hat{R} = \frac{1}{N} \mathbf{Z}^H \mathbf{C_S}^{-1} \mathbf{Z} \tag{1.64}$$

which, in case of $\mathbf{C_S} = \mathbf{I}$, i.e. uncorrelated speckle, becomes the AMI estimator

$$\hat{I} = \frac{1}{N} \mathbf{Z}^H \mathbf{Z} = \frac{1}{N} \sum_{k=1}^{N} I_k \tag{1.65}$$

If $\mathbf{C_S}$ is correctly computed or perfectly estimated, $\hat{R}$ is unbiased (Lopès & Séry, 1997). The variance of $\hat{R}$ computed on N samples is N times lower than that of the observed intensity, and $\hat{R}$ is Gamma distributed.

1.7 Despeckling Filters

The speckle is actually a deterministic process belonging to the SAR coherent image formation system. On the other hand, it is successfully modelled as a stochastic model due to the impossibility to repeat the acquisition conditions. Changes in the illuminated scene as well as in the satellite orbit make each scene appearing different even if all the other acquisition parameters (e.g. sensor mode, acquisition angle, etc.) remain the same. Neglecting the thermal noise, as in Figure 12, we can consider the signal $\mathrm{a(r,t)} = \gamma_0\mathrm{(r,t)} \cdot \mathrm{u(r,t)}$ where the fully developed speckle is modeled as a white zero-mean complex Gaussian process u that modulates the scene complex reflectivity $\gamma_0\mathrm{(r,t)}$, at the 2-D spatial position $\mathrm{(r,t)}$, to form the input signal to the linear system. This leads to the following expression for the detected power

$$\mathrm{I(r,t)} = |(\gamma_0\mathrm{(r,t)} \cdot \mathrm{u(r,t)}) * \mathrm{s(r,t)}|^2 \tag{1.66}$$

where $\mathrm{s(r,t)}$ is the system impulse response with reference to Figure 12.

1.7.1 Multiplicative Model

The main objective of speckle filtering is to retrieve the *unspeckled* scene radar backscatter from the observed image. This requires the use of a model that relates the two entities, at each pixel, as a function of speckle noise. The most commonly used model is the multiplicative speckle noise model that expresses the observed intensity as the product of the scene signal intensity and speckle noise intensity

$$\mathrm{y} = \mathrm{x} \cdot \mathrm{n} \tag{1.67}$$

where $\mathrm{y} = \mathrm{I(r,t)}$ is the observed intensity of the pixel located at $\mathrm{(r,t)}$, $\mathrm{x} = |\gamma_0\mathrm{(r,t)}|^2$ is the terrain reflectivity, and n is the intensity of fully developed speckle noise, which is modeled as a unit mean Gamma distribution (see Equation 1.46). The approximate intensity expression 1.67 might be deduced from the exact intensity expression 1.66 in various ways, leading to different expressions for the named *multiplicative speckle model*.
The most well known speckle models are

1. *Saleh and Goldfinger's model* with correlated speckle noise and uncorrelated scene signal (Goldfinger, 1982; Saleh & Rabbani, 1980). The approximate intensity y is given by
$$\mathrm{I}_m\mathrm{(r,t)} = |\gamma_0\mathrm{(r,t)}|^2 \mathrm{R_s}(0,0) \cdot \mathrm{u'(r,t)} \tag{1.68}$$
where m denotes the multiplicative model, $\mathrm{R}(\cdot)$ is the auto-correlation function and $\mathrm{u'} = |\mathrm{u(r,t)} * \mathrm{s(r,t)}|^2/\mathrm{R_s}(0,0)$ is the speckle-noise correlated process distributed along a unit mean gamma, and $\mathrm{s(r,t)}$ is the system impulse response 12.
2. *Ulaby's model* with correlated speckle noise and uncorrelated scene signal (Ulaby et al., 1986)
$$\mathrm{I}_m\mathrm{(r,t)} = \frac{|\gamma_0\mathrm{(r,t)}|^2}{\mathrm{E}\{|\gamma_0\mathrm{(r,t)}|^2\}} \left[\mathrm{E}\{|\gamma_0\mathrm{(r,t)}|^2\} * |\mathrm{s(r,t)}|^2\right] \cdot |\mathrm{u(r,t)} * \mathrm{s(r,t)}|^2 \tag{1.69}$$

3. *Kuan et al.'s model* with correlated speckle noise and correlated scene signal (Kuan et al., 1985)
$$\mathrm{I}_m(\mathrm{r},\mathrm{t}) = \left[|\gamma_0(\mathrm{r},\mathrm{t})|^2 * |\mathrm{s}(\mathrm{r},\mathrm{t})|^2\right] \cdot \mathrm{u}'(\mathrm{r},\mathrm{t}) \tag{1.70}$$
4. *Frost et al.'s model* with white speckle noise (Frost et al., 1982). This model might be better adapted to SAR systems using the following expression
$$\mathrm{I}_m(\mathrm{r},\mathrm{t}) = \left[|\gamma_0(\mathrm{r},\mathrm{t})|^2 \cdot n\right] * |\mathrm{s}(\mathrm{r},\mathrm{t})|^2 \tag{1.71}$$
where $n = |\mathrm{u}(\mathrm{r},\mathrm{t})|^2/\mathrm{E}\{|\mathrm{u}(\mathrm{r},\mathrm{t})|^2\}$ is the unit mean gamma distributed white process.
5. *Lee's model* with uncorrelated speckle noise uncorrelated scene signal (Lee, 1980)
$$\mathrm{I}_m(\mathrm{r},\mathrm{t}) = |\gamma_0(\mathrm{r},\mathrm{t})|^2 \cdot \mathrm{n} \tag{1.72}$$

In Section 1.7.3 some common filters based on the presented model are described.

1.7.2 Product Model

Under the assumption that the multiplicative speckle model of 1.67 is satisfied at each pixel position, the product model defines the expression of the unconditional pdf of the observed intensity as (Jakeman, 1980; Lewinski, 1983)

$$\mathrm{p}(\mathrm{y}) = \int_0^{+\infty} \mathrm{p}_\mathrm{n}(\mathrm{y}|\mathrm{x})\mathrm{p}_\mathrm{x}(\mathrm{x})\mathrm{d}\mathrm{x} \tag{1.73}$$

where the fully developed speckle of χ^2 pdf is assumed to be non-stationary in intensity mean, with an intensity mean $\mathrm{E}\{\mathrm{n}\}$ that varies spatially from one pixel to another according to the distribution p_x. The product model implicitly assigns scene mean variation to speckle intensity mean variations, and as a result the radar reflectivity $|\gamma_0(\mathrm{r},\mathrm{t})|^2$ is taken as the ensemble average of speckle intensity at the position $(\mathrm{r},\mathrm{t}) : \mathrm{E}\{\mathrm{n}\} = |\gamma_0(\mathrm{r},\mathrm{t})|^2$. The spatial averaging of the conditional speckle distribution leads to the unconditional distribution of the stationary mean $\bar{\mathrm{x}} = \langle \mathrm{E}\{\mathrm{y}|\mathrm{x}\}\rangle_t = \langle \mathrm{x}\rangle_t$. This supposes that the limit $\bar{\mathrm{x}}$ exists and that the speckle mean variation process x is ergodic and stationary such that its spatial average converges to its ensemble average $\mathrm{E}\{\mathrm{x}\} = \langle \mathrm{x}\rangle_t = \bar{\mathrm{x}}$.
The extension of the multiplicative model 1.73 to multi-dimensional random variables is possible. The method enables, by the use of Bayes rules, the MAP estimator which is the most complete model-based estimator. It will be presented together with other estimation methods in the next chapter.

1.7.3 Overview of Existing Approaches

The estimator of the mean reflectivity presented in Section 1.6.2 is a simple method for speckle reduction. Many other methods exists, some of them based on the models described in Section 1.7. Other simple methods which do not rely on any specisfic knowledge about the statistics of the noise are the median filter (Rees & Satchell, 1997) and the geometric filter (Crimmins, 1985). The Kuan (Kuan et al., 1985, 1987) and Lee (Lee, 1980) adaptive filters belong to the family of the statistical filters as well as, the Edge Preserving Optimized Speckle (EPOS) filter (Hagg & Sties, 1994). The Frost filter (Frost et al., 1982) and the homomorphic Wiener (Franceschetti et al., 1995b) belong to the class of Wiener filters. Filters based on wavelet transforms (Gleich & Datcu, 2007b, 2006; Achim et al.,

2003) adopt a multi-scale approach, while other methods of interest are the ones based on Partial Differential Equation (PDE)s and calculus of variation in the context of scale space theory (Yu & Acton, 2002; Kim et al., 2002). The Bayesian filters are another class of important filters. We can mention the Gamma-Gamma MAP (GGMAP) filter (Lopès et al., 1990a, 1993, 1990b), the MBD filter (Walessa & Datcu, 2000), the texture preserving filter proposed in Li et al. (2007). A homomorphic transformation in the Bayesian frame is adopted in Achim et al. (2006).
The rest of this chapter presents the most known filters.

1.7.3.1 Simple filters

Median filter The median filter is window- or neighborhood-based. A considered pixel is replaced by the median of all gray-values within the estimation window. As a result, the median filter does not introduce any new gray-value in the image, differently to the mean filter, but preserving the edges much better (Rees & Satchell, 1997). However, fine details, e.g. isolated point-scatterers, are filtered out. The median filter is a non-linear filter and does not use any explicit data or noise model.

Geometric filter The morphological filters have a completely different approach than the mean or median filters. The original value of a center pixel is replaced by a non-linear combination of pixels from a neighborhood system. The filter introduced by Crimmins (1985) belongs to this class. It is an iterative filter based on the concept of *dilatation* and *erosion* where the first is used to smooth small dark regions and the second to smooth small light regions. As for the mean and median filters, no assumptions about the noise are made. However, the one by Crimmins (1985) has a better chances to preserve edges and fine details. The smoothness of the filtered image and the loss of information is determined by the applied number of iterations.

1.7.3.2 Statistical filters

Statistical filters are window-based filters driven by the local statistics, mean and variance, of the data.This kind of filters can be interpreted as a locally varying convolution kernel applied to the image. The basic properties of the noise are captured by mean and variance. Furthermore, they can take into account the features of the image, i.e. reduced smoothing or no smoothing is applied in areas where the locally estimated coefficient of variation does not correspond to the known noise statistics. The additional roughness is recognized to be caused by image structures instead of noise. Hence, the assumption of stationary mean and variance is relaxed. Improved versions of most statistical filters exist which take into account additional structural information to allow a better filtering along edges.

Kuan filter The Kuan filter (Kuan et al., 1985, 1987) belongs to the class of Minimum Mean-Square Error (MMSE) filters, i.e. $\mathrm{E}\{(\hat{\mathrm{x}}-\mathrm{x})^2\}$ has to be minimized, where $\hat{\mathrm{x}}$ is the estimate of X. The filter has been deduced by transforming the observed signal $\mathrm{y}=\mathrm{x}+(\mathrm{n}-1)\mathrm{x}$ into x and an additive signal dependent noise term. It can be considered optimal if both x and y are Gaussian distributed. With $\mathrm{E}\{N\}=1$ for intensity image, the filter equation is given by

$$\hat{\mathrm{x}}=k\mathrm{y}+(1-k)\mu_{\mathrm{y}} \tag{1.74}$$

where

$$k = \frac{\sigma_{\mathrm{x}}^2}{\sigma_{\mathrm{x}}^2 + (\mu_{\mathrm{y}}^2 + \sigma_{\mathrm{x}}^2)/L} \tag{1.75}$$

The variance of the radar reflectivity is derived by $\sigma_{\mathrm{x}}^2 = (L\sigma_{\mathrm{y}}^2 - \mu_{\mathrm{y}}^2)/(L+1)$ where L denotes the equivalent number of looks of the speckle noise.
The Kuan filter is an adaptive filter based on a test of the local coefficient of variation. Compared to the filter presented before, this one represents an important improvement, which is directly visible in the filtering results. However, the filter only computes a weighted sum of noisy and mean-filtered pixels. It does neither exploit the full knowledge of the noise distribution, nor does it include additional assumptions about the noise-free data.

Lee filter The well-known Lee filter (Lee, 1980) is a special case of the Kuan filter. It differs from the latter only in the weighting factor k because of a linear approximation made for the multiplicative noise model. For the Lee filter k can be found to be

$$k = \frac{\sigma_{\mathrm{x}}^2}{\sigma_{\mathrm{x}}^2 + \mu_{\mathrm{y}}^2/L} \tag{1.76}$$

Due to its more accurate modelling of the multiplicative noise behavior, the Kuan filter is to be preferred. However, the visual appearance of images filtered with Lee's is identical.

EPOS filter The EPOS filter (Hagg & Sties, 1994) also relies on the analysis of mean and variance. The estimation window is divided into eight triangular areas to guarantee improved edge-preserving capabilities. The mean of the most homogeneous areas is taken as estimate of $\hat{\mathrm{x}}$. Borders are extremely well preserved and the filter smooths right up to the edges, unlike the Kuan filter.However the EPOS filter does not allow smoothly varying cross-sections. The filtering results are composed of areas of almost constant cross-section separated by sharp edges. Texture is absolutely not preserved by this kind of approach.

1.7.3.3 Wiener filter

The Wiener filter (Frost et al., 1982; Franceschetti et al., 1995b) is the optimal linear filter in the sense of MMSE for stationary signal corrupted by additive noise. It requires the knowledge of the power spectra of the noise and of the noise free scene, or equivalently, of their auto-correlation functions. Hence, the full Wiener filter is able to directly deal with correlated additive noise. In order to obtain optimal results with a Wiener filter both the noise and the noise-free signal must be Gaussian distributed, which is not the case for SAR signal.

Frost filter The Frost filter (Frost et al., 1982) is an adaptive Wiener filter resulting in a locally changing convolution kernel for the noisy image depending on scene homogeneity. It was derived directly for multiplicative noise under the assumption of locally stationary image data. After a good number of simplifications the impulse response is given by

$$h(r) = K_1 \exp\left\{-K\frac{\sigma_{\mathrm{y}}^2}{\mu_{\mathrm{y}}^2}|r|\right\} \tag{1.77}$$

where K is the filter parameter, which determines the strength of the filtering, K_1 is a normalization constant and $|r|$ is the radial distance from the center pixel to be filtered. The convolution kernel $h(r)$ is valid under the assumption of a scene reflectivity X obeying an auto-regressive process with an exponentially decreasing isotropic auto-correlation function.
However, these approximations only result in a performance comparable to most other statistical filters. Unlike a full Wiener filter, the Frost filter does not rely on the computation of covariances or power-spectra to profit for a better image description.

Homomorphic Wiener filter This approach is based on a homomorphic processing of the SAR image in combination with full Wiener filtering (Franceschetti et al., 1995b). The homomorphic logarithmic transformation is used to convert the multiplicative speckle into additive noise. The required power spectra can be calculated analytically for the noise and by an iterative procedure for the noise-free image signal. However, convergence of the latter estimate to the correct power spectrum of the cross-section is not ensured. In contrast to the Frost filter, the local estimation of the power spectrum allows a much better restoration especially of linear structures. On the other hand, artifacts may be introduced if the power spectrum is not correctly estimated or if the estimation window lies over highly non-stationary areas. Since rather large windows are required, typically between 8×8 and 16×16 pixels, this filter is not well suited for images with high variations in scene content.
Because of the information contained in the estimated power spectra, the results are of a satisfactorily quality for stationary textured areas. Note that the filtering and the estimation in the Fourier domain make this approach several magnitude slower than statistical filters.

1.7.3.4 Multi-scale and scale-space approaches for despeckling

Multi-scale approaches rely on the analysis of the image at different resolutions. The image is presented by a pyramidal decomposition in order to profit from inter-scale dependencies. It allows to separately analyze lower and higher details and to exploit correlations at lower scale.
The scale-space representation aims to embed the original signal into a one-parameter family of derived signals where fine scale structures are successively suppressed. A crucial requirement is that structures at coarse scales in the multi-scale representation should constitute simplifications of corresponding structures at finer scales. They should not be accidental phenomena created by the method for suppressing fine-scale structures.

Filters based on wavelet transform Wavelet-based filters work on the wavelet transformed image. Since these approaches are applied under the assumption of additive Gaussian noise, the image is subject to a homomorphic transform (in case of SAR the logarithmic transform) before the wavelet decomposition is computed. The wavelet coefficients are then shrunk according to various methods in order to reduce the noise energy in different sub-bands. The inverse-transformed image still exhibits a lot of details, while the noise has been reduced. However, strong noise reduction is accompanied by the introduction of wavelet artifacts, which can be as disturbing as the speckle noise itself. Better wavelet shrinkage methods might solve this problem but the main drawback remains the additive Gaussian noise approximation, which is not valid for low values of L. In

addition, a model for the noise-free image is not included, and the number of possible wavelets to apply is another free parameter. Examples of wavelet filters are available in Gleich & Datcu (2007b), Gleich & Datcu (2006) and Achim et al. (2003).

Anisotropic Diffusion The filter proposed in Perona & Malik (1990) is based on the computation of the PDE. It is based on the following equation

$$\frac{\partial}{\partial t}\mathrm{x} = \mathrm{div}(c(||\nabla \mathrm{x}||)\nabla \mathrm{x}) \tag{1.78}$$

where $c(||\nabla \mathrm{x}||)$ is a non-linear function of the gradient. Equation 1.79 is interpreted as an evolution in time which results in the following iterative update scheme

$$\mathrm{x}^{t+1} = \mathrm{x}^{t} + \lambda\langle c(||\nabla \mathrm{x}||), \nabla \mathrm{x}\rangle \tag{1.79}$$

where $\langle \cdot, \cdot \rangle$ is the scalar product and λ is a constant.
In Perona & Malik (1990) are proposed two alternative expressions for the $c(\cdot)$ function : $c(||\nabla \mathrm{x}||) = \exp\{-(||\nabla \mathrm{x}||/K)^2\}$ and $c(||\nabla \mathrm{x}||) = 1/(||\nabla \mathrm{x}||/K)^2$. The filter is controlled by the diffusion coefficient $c(||\nabla \mathrm{x}||)$ which is close to one inside the regions enabling the smoothing, and which is close to zero when the gradient is high, disabling the smoothing. The filter based on anisotropic diffusion is really effective and preserves edges but, on the other hand, the smoothness is controlled by the number of iterations. Moreover the method is demonstrated to be equivalent to the Lee and Frost filters (Yu & Acton, 2002).

1.7.3.5 Bayesian filters

The Bayesian approach is the only one that correctly models the speckle noise statistics in form of prior function. The quality of the estimate strongly depends on the prior assumptions about the noise-free data. Thus, we can expect an improved quality of the filtered image. The Bayesian filters refers to the MAP estimate. The GGMAP filter (Lopès et al., 1990a, 1993, 1990b) uses a Gamma distribution for both prior and likelihood functions. A more complex parametric model is introduced in Walessa & Datcu (2000) where a Gauss-Markov prior is used in order to describe spatial structures. The model, described in Section 3.1, improves the quality of the despeckled image and allows the estimation of textural parameters.
The method proposed in this thesis, based on Tikhonov regularization, is also classified as a Bayesian filter due to their equivalence. The model is described in detail in Section 3.4. It should be emphasized that the approach can be equivalent to the anisotropic diffusion method.

1.8 Speckle simulation

According to Raney & Wessels (1988), the simulation of SAR imagery from a sensor with a specified bandwidth and coherence properties goes through the following steps :

1. Image file
   a. From a source of ideal imagery, the reflectivity map, create one unspeckled image by convolving the source against the (desired) two-dimensional impulse response function.
   b. Sum additive noise to the image in case it has to be included.

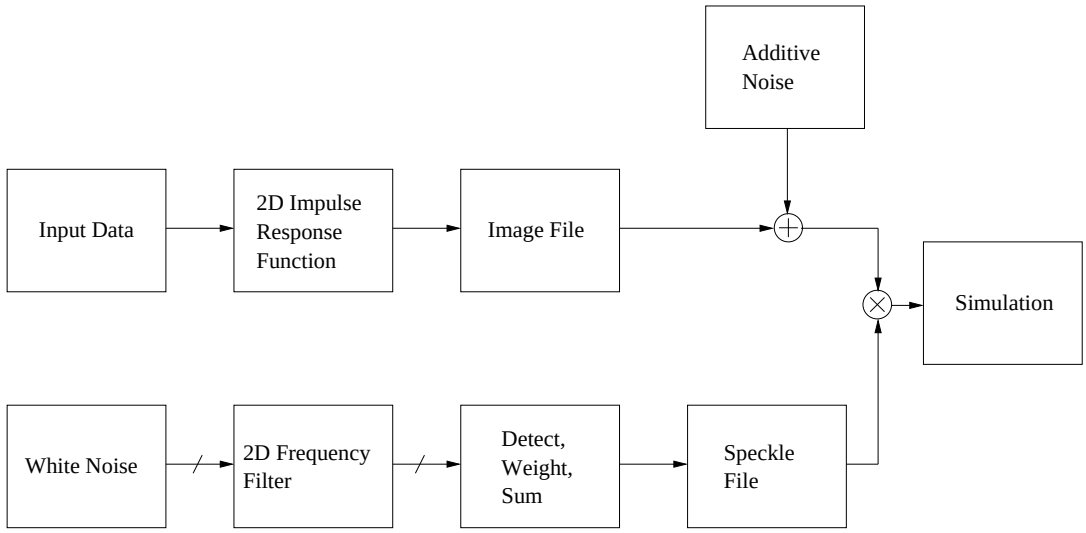


FIGURE 19 – Block diagram of the speckle simulation process (Raney & Wessels, 1988).

2. Speckle file

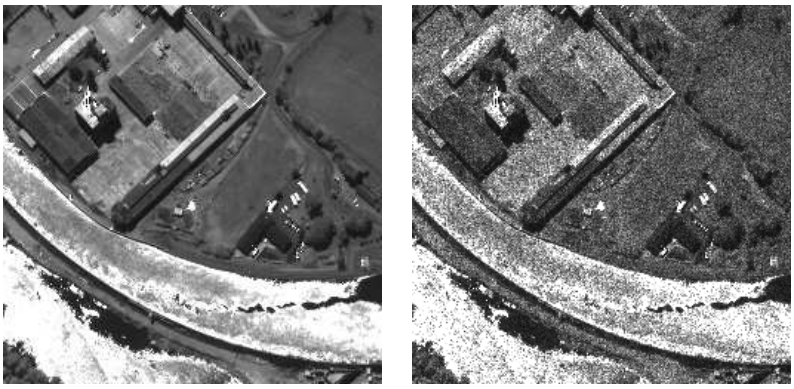
   a. Prepare N files each of which is a complex Gaussian pseudo-random field, essentially a *white noise* source. Adjacent samples should be statistically independent.
   b. Bandpass filter each file with the two-dimensional frequency spectra corresponding to the radar and processor to be simulated. Each filter should be weighted and overlapped as per the described system.
   c. Square law detect the filter outputs, and sum, again using any weighting representative of the system. Normalize.
   d. Store the resulting real variates as a *speckle file*. This is of course also in two dimensions.

3. Simulation

   a. Subsample the image file and the speckle file to match the desired pixel spacing.
   b. Pixel by pixel, multiply the two files together to create the final speckled image file.

The block diagram of the simulation procedure is presented in Figure 19. An example is provide in Figure 20 where a noise 3-looks SAR image has been simulated from the original Quick Bird (QB) optical image input. While, for the detected image, we use the term of speckle to describe the salt and pepper effect visible in the image and modelled as multiplicative noise ; we use the generic term *noise* in case of complex-valued signals. The simulation of complex-valued image noise goes through the definition of the Signal to Noise Ratio (SNR) and has to be dimensioned according to the desired value of SNR. It is defined as follows

$$\mathrm{SNR} = -10 \log \xi \tag{1.80}$$



(a) (b)

FIGURE 20 – Optical image with simulated 3-look speckle noise : original (a) and noisy image (b).

where ξ is an error measure. Common error measures are, for example, the Normalized Mean-Square Error (NMSE), defined as

$$\xi_{\text{NMSE}} = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}|F(j,k) - \hat{F}(j,k)|^2}{\sum_{j=1}^{J}\sum_{k=1}^{K}|F(j,k)|^2} \tag{1.81}$$

or the Peak Mean-Square Error (PMSE), which has the following form

$$\xi_{\text{PMSE}} = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}|F(j,k) - \hat{F}(j,k)|^2}{[\max\{F(j,k)\}]^2} \tag{1.82}$$

where $F(\cdot,\cdot)$ is a reference image and $\hat{F}(\cdot,\cdot)$ is a second image. The sum is over the whole image of size (J,K) and $\max\{\cdot\}$ represents the maximum value assumed by the function.

1.9 Summary of the chapter

In this chapter, the basic radar concepts have been presented from the SAR sensor geometry of acquisition to the image formation. Furthermore, the statistics of the data and the methods for speckle reduction, with the models and the most known filters, have been presented. The problem is to find a model for the system impulse response and the image. It is addressed as an estimation and model selection problem, thus in the next chapter the theoretical bases of Bayesian inference and estimation theory together with basic concepts of information and distortion theories are presented.

Chapitre 2

Statistical Modelling and Estimation

The Chapter contains the theoretical bases later applied for image estimation. There is an introduction which spans from stochastic processes to Markov Random Field (MRF) through Markov chain. The GMRF model is presented. The Chapter includes the estimation methods for deterministic and stochastic parameters. Then, the two levels of Bayesian inference are described and the information theory concepts are introduced in order to link the information content with the parameter estimation. The Chapter ends with the model selection carried out via rate distortion, which is a novel approach.

2.1 Stocastic modeling

A collection of random variables indexed by a parameter such as time or space is known as *stochastic process*.
In applied statistics, after the collection of empirical data, a theoretical probability distribution is fitted in order to extract more information from the data. If the fit is good, the properties of the set of data can be approximated by the properties of the theoretical distribution.
Statistic is an applied science which allows to develop methods to extract information from the observed data in order to understand the phenomena which generates the data and to take decisions. Figure 1 shows the process spanning from the data to the information through the statistical modeling, whereas the decision process is not shown. A statistical model is a family of probability distributions defined on the space of observations.
Considering a random variable X, we denote a realization of the random variable with the notation $X = x$, which represents a value in the space of observations $\mathcal{X}$. The proba-

FIGURE 1 – The diagram shows the flow from the data to the information through the statistical modeling. In the scheme the decision problem is not shown.

bility density function is a transformation given by

$$\begin{aligned} p(\mathrm{X}) : \mathcal{X} &\rightarrow [0,1] \in \mathbb{R}_0^+ \\ \mathrm{x} &\rightarrow p_{\mathrm{X}}(\mathrm{X}=\mathrm{x}) \\ \int p_{\mathrm{X}}(\mathrm{X}=\mathrm{x})\mathrm{dx} &= 1 \end{aligned} \tag{2.1}$$

where we use the short notation $p_{\mathrm{X}}(\mathrm{X}=\mathrm{x}) = p(\mathrm{x})$.
A family of distributions is characterized by an ensemble of parameters $\boldsymbol{\theta}$, e.g. the Gaussian is a two-parameter exponential family distribution : the variance σ and the mean μ. Fitting the model means to choose an estimator for the parameters. The dependency of a distribution to a parameter vector $\boldsymbol{\theta}$ is denoted by $p(\mathrm{x}|\boldsymbol{\theta})$ and it is called *conditioning* in statistical language. A basic rule for statistical conditioning is given by the following formula

$$p(\mathrm{x}|\boldsymbol{\theta}) = \frac{p(\mathrm{x},\boldsymbol{\theta})}{p(\boldsymbol{\theta})} \tag{2.2}$$

Equation 2.2 means that the conditional distribution $p(\mathrm{x}|\boldsymbol{\theta})$ equals the ratio between the joint probability distribution $p(\mathrm{x},\boldsymbol{\theta})$ and the marginal parameter distribution $p(\boldsymbol{\theta})$. In case of statistical independence, the product rule allows to write the joint distribution as $p(\mathrm{x},\boldsymbol{\theta}) = p(\mathrm{x})p(\boldsymbol{\theta})$ and, thus, Equation 2.2 becomes $p(\mathrm{x}|\boldsymbol{\theta}) = p(\mathrm{x})$. It means that the realization of the random process X is independent from the occurrence of the event $\boldsymbol{\theta}$.
We can refer to $\boldsymbol{\theta}$ as a random variable or as a parameter vector. In the latter case, it defines a *parametric* model. If the structure of the model is not specified explicitly by a parameter, but it is determined from the data, the model is *non-parametric*. However, *non-parametric* is not meant to imply that such models completely lack parameters but that the number and the nature of the parameters are flexible and not fixed in advance, e.g. a histogram is a simple non-parametric estimate of a probability distribution.

2.2 Stochastic processes

If we consider a collection of realizations x assumed by a random process X in the observation space $\mathcal{X}$, a stochastic process is defined as

$$\mathrm{X} = \{\mathrm{X_t} : \mathrm{t} \in \mathcal{T}\} \tag{2.3}$$

where t is a time variable. If the process in 2.3 assumes values continuously in $\mathcal{T}$, it is a continuous-time process, whereas a discrete-time process assumes values only for discrete value in $\mathcal{T}$.
The discrete-time process can be represented by the following notation

$$\mathrm{X} = \{\mathrm{X_n} : 1 \leq i \leq n\} \tag{2.4}$$

which describes a process of cardinality n. We are interested on independent identically distributed (iid) processes and Markovian processes in the hypothesis of stationarity and ergodicity. The statistical independence of random variables belonging to the process X can be mathematically formalized as follows

$$p(\mathrm{X}) = p(\{\mathrm{X}_t : t \in \mathcal{T}\}) = \prod_{t \in \mathcal{T}} p(\mathrm{X}_t) \tag{2.5}$$

which means that the joint probability density function equals the product of the marginal distributions.
The process is said identically distributed when the random variables which belong to the process have the same distribution, thus

$$\forall\ t, u \in \mathcal{T} \qquad p(\mathrm{X}_t) = p(\mathrm{X}_u) \tag{2.6}$$

which asserts the identity of any pair of distribution of the realizations.

2.3 Markovian Process

We suppose the measurements to be a function of time t, $\mathrm{X}_n = \mathrm{X}(t_n)$ with $t_n \geq t_{n-1}$. The data ordered with respect to time correspond to a time series. If the measurement depends on the preceding ones the process is called Markovian, where the length of the backwards linkage is the order of the chain. A Markov chain of order m-th is expressed by

$$p(\mathrm{X}_n; t_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_1; t_{n-1}, \ldots, t_1) = p(\mathrm{X}_n; t_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_{n-m}; t_{n-1}, \ldots, t_{n-m}) \tag{2.7}$$

with $m \leq n$, where we adopt the following equivalent notation

$$p(\mathrm{X}_n; t_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_1; t_{n-1}, \ldots, t_1) = p(\mathrm{X}_n; t_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_1) \tag{2.8}$$

The statistical independence is denoted by the chain of order zero

$$p(\mathrm{X}_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_1) = p(\mathrm{X}_n). \tag{2.9}$$

The term Markov chain is most frequently used to denote a first order Brownian motion-like process

$$p(\mathrm{X}_n | \mathrm{X}_{n-1}, \ldots, \mathrm{X}_1) = p(\mathrm{X}_n | \mathrm{X}_{n-1}) \tag{2.10}$$

where the dependency is restricted only to the previous sample.
The joint pdf of a first order Markov chain can be written, applying recursively Equation 2.2, as follows

$$p(\mathrm{X}_1, \ldots, \mathrm{X}_n) = p(\mathrm{X}_1) \prod_{k=2}^{n} p(\mathrm{X}_k | \mathrm{X}_{k-1}) \tag{2.11}$$

where, in case of $n = 2$, the process is characterized by a 2-dimensional joint pdf.

2.4 Gibbs-Markov Random Fields

A random field is an ensemble of random variables belonging to a multi-dimensional space. Thus, the concept allows the extension of the Markov process from the one-dimensional to multidimensional case through the definition of the concepts of neighborhood and cliques. It finds application in image processing (2-D signals), where the pixels of an image are the nodes of a regular lattice $\mathcal{L}$. The neighborhood system $\mathcal{N}$ is defined as follows

$$\mathcal{N} = \{\mathcal{N}_r : \forall r \in \mathcal{L}\} \tag{2.12}$$

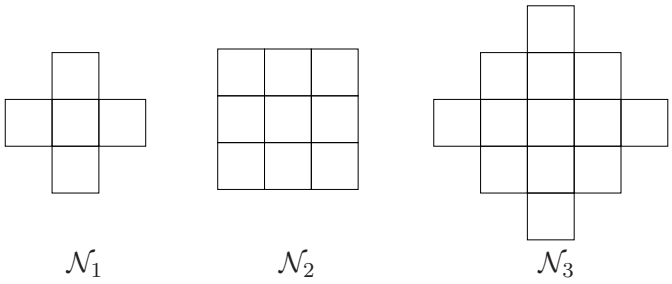


FIGURE 2 – Example of neighborhood system for increasing model order.

where $\mathcal{N}_r$ is a set of neighbor sites. Three examples of neighborhood system are shown in Figure 2
Thus a Markov Random Field is defined as follows

$$p(\mathrm{X}_s|\mathrm{X}_r : r \in \mathcal{N}, r \neq s) = p(\mathrm{X}_s|\mathrm{X}_r : r \in \mathcal{N}_s) \tag{2.13}$$

where $\mathrm{p}(\mathrm{X}) \geq 0$ and r and s are two sites in $\mathcal{L}$. The local relationship has the following properties

1. A site is not neighboring to itself : $i \notin \mathcal{N}_i$

2. The neighboring relationship is mutual : $r \in \mathcal{N}_s \Leftrightarrow s \in \mathcal{N}_r$

The Markovianity is a local characterization of the random fields. The pair $(\mathcal{L}, \mathcal{N})$ defines a graph, where $\mathcal{L}$ contains the nodes and $\mathcal{N}$ specifies the link.
The vicinity interaction can be also represented by a set of cliques $\mathcal{C} = \{c_k\}$ which belong to $\mathcal{N}$. The concept of clique allows to link an energy function to a realization x of a random field, as follows

$$U(\mathrm{x}|\theta) = \sum_{c \in \mathcal{C}} V_c(\mathrm{x}|\theta) \tag{2.14}$$

where the energy function $U(\mathrm{x}|\theta)$ is given by the sum of the potential function $V_c(\mathrm{x}|\theta)$ defined on each clique. Thus, a Gibbs Random Field (GRF) follows a distribution of the form

$$\mathrm{p}(\mathrm{X}) = Z^{-1} \cdot \exp\{-U(\mathrm{X}|\theta)/T\} \tag{2.15}$$

where

$$Z = \sum_{x \in X} \exp\{-U(\mathrm{x}|\theta)/T\} \tag{2.16}$$

is the partition function which has a normalization purpose.
The Hammersley-Clifford theorem (Spitzer, 1971) establishes the equivalence of the representation of a random field by a Gibbs distribution or a Markov process. The local characterization of a Gibbs Random Fields can be formalized, according to the Hammersley-Clifford theorem, as follows

$$p(\mathrm{x}_s|\mathrm{x}_r : r \in \mathcal{N}, r \neq s, \theta) = \frac{\exp\{-T^{-1} \sum_{c \in \mathcal{N}} V_c(\mathrm{x}, \theta)\}}{\sum_{\{\mathrm{x}_r | r \neq s\}} \exp\{-T^{-1} \sum_{c \in \mathcal{N}} V_c(\mathrm{x}_r, \theta)\}} \tag{2.17}$$

where the denominator corresponds to the partition function and the numerator depends only on the neighborhood. Gibbs distributions and Markov Random Fields have been applied for the first time to image processing and computer vision by Geman & Geman (1984b).

2.4.1 Gauss-Markov Random Field

An example of random field is given by the Gaussian family, called Gauss-Markov Random Field (GMRF) and introduced by Chelappa & Kashyap (1983). It is characterized by the following potential function

$$U(\mathrm{x}_s|\mathrm{x}_r : r \in \mathcal{N}_s, \boldsymbol{\theta}) = -\frac{|\mathrm{x}_s - \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r})|^2}{2\sigma^2} \tag{2.18}$$

where $\mathcal{N}$ is characterized by an odd symmetry $\theta_r = \theta_{-r}$. Thus, the conditional pdf takes the form

$$p(\mathrm{x}_s|\mathrm{x}_r, r \in \mathcal{N}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{|\mathrm{x}_s - \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r})|^2}{2\sigma^2}\right\} \tag{2.19}$$

The Gaussian process is a special case of GMRF whose Gibbs energy consists only of single site clique potentials. It has no contextual interaction.
Thus, the realization takes the form of a multivariate Gaussian distribution

$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}_\mathrm{X}|^{1/2}} \exp\{-\frac{1}{2}\mathbf{x}^T\mathbf{C}_\mathrm{X}^{-1}\mathbf{x}\} \tag{2.20}$$

where the notation is lower case bold symbols for vectors and the upper case bold symbols for matrices, $|\cdot|$ is the determinant operator, N is the cardinality of the lattice $\mathcal{L}$ and $\mathbf{C}_\mathrm{X} = \mathbf{C}(\boldsymbol{\theta})$ is the definite positive covariance matrix.
In case that the covariance matrix $\mathbf{C}_\mathrm{X}$ is factorisable, i.e. $\mathbf{C}_\mathrm{X} = A^T A$, the model corresponds to a simultaneous Auto Regressive (AR) process

$$\mathrm{x}_s = \frac{1}{2}\sum_{r\in\mathcal{N}} \theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r}) + \mathrm{e}_s \tag{2.21}$$

where $\mathrm{e}_s = \mathcal{N}(0, \sigma^2)$ is a zero-mean Gaussian process and has an autocorrelation function given by

$$\mathrm{E}\{\mathrm{e}_s\mathrm{e}_{s+r}\} = \begin{cases} -\theta_r\sigma^2 & \text{if } r \in \mathcal{N} \\ \sigma^2 & \text{if } r = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.22}$$

The model has been applied for texture synthesis and texture parameter estimation (Chelappa et al., 1985). We extend the GMRF model to the complex domain and used it for analysis and synthesis, i.e. retrieval and forward modeling, of complex-valued images.

2.5 Parameter estimation

Estimation theory is a branch of statistics and signal processing that deals with estimating the values of parameters based on measured/empirical data. An estimator $\hat{\theta} = f(\mathrm{x}_1, \ldots, \mathrm{x}_n) = f(\mathbf{x})$ is a function of the observations $\mathbf{x} = \{\mathrm{x}_1, \ldots, \mathrm{x}_n\}$ of a random variable X. The parameter can be either a scalar θ or a vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ and for the estimation the likelihood function is used

$$p(\mathbf{x}|\theta) = p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\theta) = \prod_{i=1}^{n} p(\mathrm{x}_n|\theta) \tag{2.23}$$

which is the conditional pdf for independent observations given the parameter.

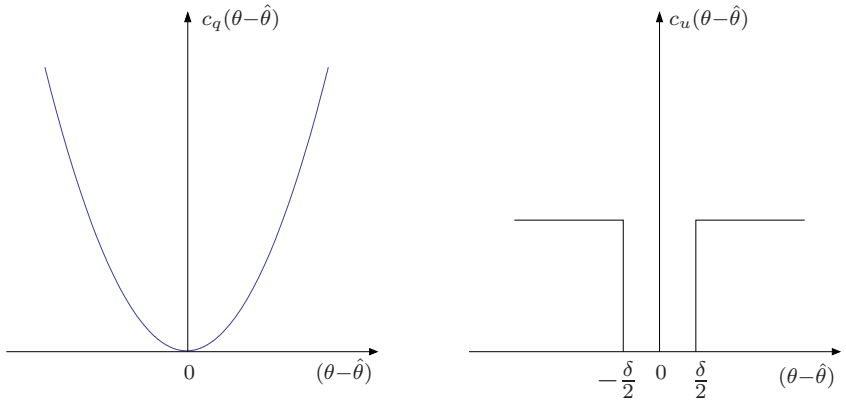


FIGURE 3 – Example of quadratic and uniform cost functions.

2.5.1 Bayes Risk and Bayesian estimators

A Bayesian estimator considers the parameter to be estimated as a random parameter. Through the definition of a cost function $c(\epsilon_\theta)$ which depends on the estimation error $\epsilon_\theta = \theta - \hat{\theta}$ it is possible to define the Bayes risk given by the expectation of the cost function

$$R = \mathrm{E}\{c(\epsilon_\theta)\} = \iint c(\epsilon_\theta) p(\mathbf{x}, \theta) \ \mathrm{d}\mathbf{x}\mathrm{d}\theta \tag{2.24}$$

where, here and in the following, the integrals are evaluated on the whole domain of definition for the variable, if not specified explicitly. Thus, the parameter estimation problem will be stated as a minimization of the Bayes risk. Common used cost functions are the quadratic defined as

$$c_q(\epsilon_\theta) = (\theta - \hat{\theta})^2 \tag{2.25}$$

and the uniform cost function

$$c_u(\epsilon_\theta) = \begin{cases} 0 & \text{if } |\theta - \hat{\theta}| \leq \delta/2 \\ 1 & \text{if } |\theta - \hat{\theta}| > \delta/2 \end{cases} \tag{2.26}$$

The plots of the cost functions are shown in Figure 3.

2.5.1.1 Minimum Mean Square Error (MMSE) Estimator

The use of the quadratic cost function 2.25 leads to the MMSE. Thus, by replacing 2.25 in 2.24 and applying the rule for conditional probability 2.2, the Bayes risk can be rewritten as

$$R_q = \int p(\mathbf{x}) \int (\theta - \hat{\theta})^2 p(\theta|\mathbf{x}) \ \mathrm{d}\theta \mathrm{d}\mathbf{x}. \tag{2.27}$$

The minimization of the Bayes risk 2.27, with respect to θ, is equivalent to the minimization of the integrand

$$I(\hat{\theta}, \mathbf{x}) = \int (\theta - \hat{\theta})^2 p(\theta|\mathbf{x}) \ \mathrm{d}\theta \tag{2.28}$$

because the term $\int p(\mathbf{x}) \ \mathrm{d}\mathbf{x}$ does not depend on $\hat{\theta}$ and both integrals in 2.27 are positive. Thus

$$\frac{\partial}{\partial \hat{\theta}} I(\hat{\theta}, \mathbf{x}) = 2\hat{\theta} \int p(\theta|\mathbf{x}) \ \mathrm{d}\theta - 2 \int \theta p(\theta|\mathbf{x}) \ \mathrm{d}\theta \tag{2.29}$$

which equaling to zero and remembering that $\int p(\theta|\mathbf{x})\mathrm{d}\theta = 1$ leads to the equation of the MMSE estimator given by

$$\hat{\theta}_{\mathrm{MMSE}}(\mathbf{x}) = \int \theta p(\theta|\mathbf{x}) \ \mathrm{d}\theta \tag{2.30}$$

which corresponds to the conditional mean. As shown in 2.30 the MMSE estimator is a function of the observation vector $\mathbf{x}$.

2.5.1.2 Maximum A Posteriori (MAP) Estimator

The MAP estimator is obtained by using the uniform cost function 2.26. The resulting risk function to minimize has the following form

$$\begin{aligned} R_u &= \int p(\mathbf{x}) \int c_u(\theta - \hat{\theta}) p(\theta|\mathbf{x}) \ \mathrm{d}\theta \mathrm{d}\mathbf{x}. \\ &= \int p(\mathbf{x}) \left[1 - \int_{\hat{\theta}-\frac{\delta}{2}}^{\hat{\theta}+\frac{\delta}{2}} p(\theta|\mathbf{x}) \ \mathrm{d}\theta \right] \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.31}$$

where the minimization of the risk R_u requires the maximization of the integral

$$I(\hat{\theta}, \mathbf{x}) = \int_{\hat{\theta}-\frac{\delta}{2}}^{\hat{\theta}+\frac{\delta}{2}} p(\theta|\mathbf{x}) \ \mathrm{d}\theta. \tag{2.32}$$

We observe that in the limit case for $\delta \to 0$ the integral $I(\hat{\theta}, \mathbf{x})$ becomes equal to $\delta p(\hat{\theta}|\mathbf{x})$. Thus, the maximization of I is obtained by the maximization of the posterior density $p(\theta|\mathbf{x})$, which in formula can be written as follows

$$\hat{\theta}_{\mathrm{MAP}} = \arg\max_{\theta} p(\theta|\mathbf{x}). \tag{2.33}$$

As a first comparison, if the mode of the posterior distribution is equal to the mean, i.e. the posterior is symmetric, the MAP estimator equals the MMSE estimator.

2.5.2 Maximum Likelihood (ML) Estimator

If the parameter is not a random variable but is deterministic, the prior distribution becomes a Dirac distribution, thus the risk of Bayes has no more sense. In this case the likelihood function 2.23 helps and leads to the ML estimator. It is defined as follows

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} p(\mathbf{x}|\theta). \tag{2.34}$$

The ML estimates the mode of the likelihood pdf. Often the log-likelihood is used when the distribution belongs to an exponential family.
We notice that the ML estimator is equal to the MAP estimator in case of a uniform a priori distribution. For this reason, even if the ML estimator is not based on the Bayes risk we will still consider it as a Bayes estimator.

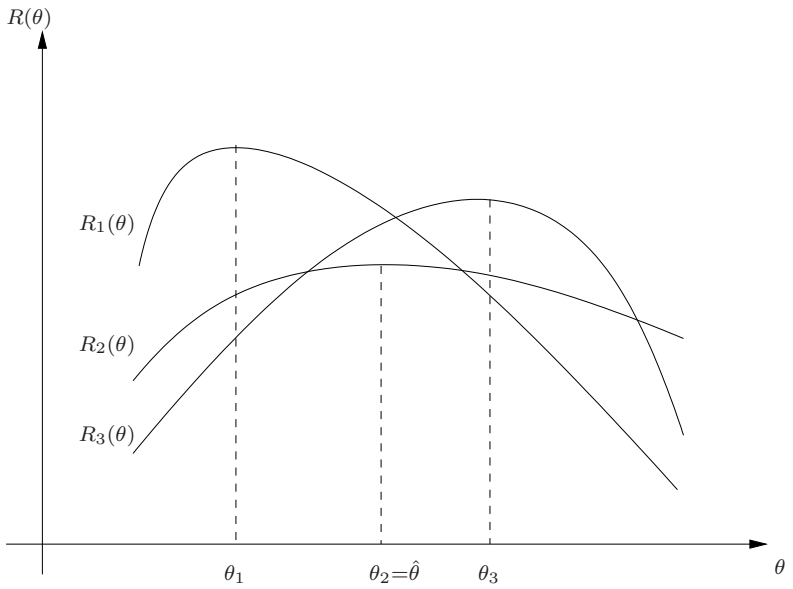


FIGURE 4 – Example of MinMax criterion. The minimum value in the set of maxima $\max\{R(\theta)\}$ of the risk functions is chosen as estimate of $\hat{\theta}$.

2.5.3 MinMax Criterion

The minmax criterion consists of another method to minimize the risk function. It can be formalized as described in the following. Through 2.2 we can rewrite the Bayes risk as follows

$$R = \mathrm{E}\{c(\epsilon_\theta)\} = \int p(\theta) \int c(\epsilon_\theta) p(\mathbf{x}|\theta) \ \mathrm{d}\mathbf{x}\mathrm{d}\theta \tag{2.35}$$

and separate the quantity

$$R(\theta) = \int c(\epsilon_\theta) p(\mathbf{x}|\theta) \ \mathrm{d}\mathbf{x} \tag{2.36}$$

the minmax criterion can be written as

$$\min_{\hat{\theta}} \max_{\theta} R(\theta) \tag{2.37}$$

Differently from the Bayes risk 2.24 that chooses the parameter which minimizes the average risk, the minmax criterion 2.37 chooses the parameter which minimizes the maximum of the risk function $R(\theta)$ in the parameter space. An example is shown in Figure 4. Thus, the criterion assumes that the worse case, corresponding to the maximum of the risk function, occurs for each estimation strategy. On the other hand, when different risk functions are compared, the minmax criterion can lead to not choosing the best strategy, because it does not consider the probability of occurrence of the state, which depends on the observation space.

2.5.4 Bound of an Estimator

An estimator is said to be unbiased if its expected value is equal to the parameter, i.e. $\mathrm{E}\{\hat{\theta}\} = \theta$. If an estimator is unbiased, it is possible to establish an inferior limit for its

variance, called Cramér-Rao lower bound.
Given a random variable X characterized by the pdf $p_{\mathrm{X}}(\mathrm{X}=\mathrm{x}|\theta)=p(\mathrm{x}|\theta)$ dependent on the parameter θ the information on θ contained in X is

$$I(\theta)=\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\right]^2\right\} \tag{2.38}$$

which is called Fisher information. If we consider n iid random variables $\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{n}$ the process is described by the likelihood function 2.23. In 2.38 appears the logarithm of the pdf, thus we rewrite 2.23 as log-likelihood as follows

$$\ln p(\mathbf{x}|\theta)=\ln\prod_{i=1}^{n}p(\mathrm{x}_i|\theta)=\sum_{i=1}^{n}\ln p(\mathrm{x}_i|\theta) \tag{2.39}$$

Thus, the information contained in the samples is

$$\begin{aligned}
I_n(\theta) &= \mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n}p(\mathrm{x}_i|\theta)\right]^2\right\} && (2.40)\\
&= \mathrm{E}\left\{\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln p(\mathrm{x}_i|\theta)\right]^2\right\} && (2.41)\\
&= \mathrm{E}\left\{\sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\ln p(\mathrm{x}_i|\theta)\right]^2\right\} && (2.42)\\
&= \sum_{i=1}^{n}\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln p_{\mathrm{X}}(\mathrm{x}_i|\theta)\right]^2\right\} && (2.43)\\
&= nI(\theta) && (2.44)
\end{aligned}$$

where for the step from 2.40 to 2.41 has been used 2.39, for the step from 2.41 to 2.42 has been applied the property of the orthogonal scalar product, and eventually for the step from 2.42 to 2.43 the linear property of the integrals has been applied. Consequently the Fisher information of n iid random samples is n times the individual Fisher information. The Cramér-Rao inequality states that the mean squared error of any unbiased estimator $\hat{\theta}=f(\mathrm{x}_1,\ldots,\mathrm{x}_n)=f(\mathbf{x})$ of the parameter θ is lower bounded by the reciprocal of the Fisher information

$$\sigma^2_{\hat{\theta}}\geq\frac{1}{nI(\theta)}. \tag{2.45}$$

In case that a parameter vector has to be estimated, 2.45 can be generalized and becomes the following matrix inequality

$$\mathbf{C}_{\boldsymbol{\theta}}\geq\mathbf{I}_n^{-1}(\boldsymbol{\theta}) \tag{2.46}$$

where $\boldsymbol{\theta}=\{\theta_1,\ldots,\theta_m\}$ and the Fisher information matrix $\mathbf{I}_n(\boldsymbol{\theta})$ is calculated on n iid samples. The inequality in 2.46 has to be interpreted in the sense that the difference $\mathbf{C}_{\boldsymbol{\theta}}-\mathbf{I}_n^{-1}(\boldsymbol{\theta})$ results in a non-negative definite matrix.
An element of the fisher information matrix has the following form

$$I_{n,ij}(\boldsymbol{\theta})=\int\frac{\partial}{\partial\theta_i}\ln p(\mathbf{x}|\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\ln p(\mathbf{x}|\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})\mathrm{d}\mathbf{x}. \tag{2.47}$$

As already observed, the Fisher information is a measure of the amount of information about θ that is present in the data. It is worth to point out the relationship between the Fisher information and the entropy. Here we only say that the Fisher information is related to the surface area of the space of observations while the entropy is related to the volume of the space of observations. Their relationship is formalized by the de Bruijn identity (Cover & Thomas, 1991a).

2.6 Bayesian Inference

The cases previously presented are useful when the family of the parametric model is known and fixed. Now we consider the case when we have an ensemble of families of parametric models to model a stochastic process X. We refer to this set of models with the notation $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$. A parameter vector Θ_i is associated with each model $\mathcal{M}_i$.
Two levels of inference can often be distinguished in the process of data modelling. At first level of inference, we assume that a particular model $\mathcal{M}_i$ is true, and we fit that model to the data, i.e. we estimate the model parameter vector Θ_i. The results of this inference are the most probable parameter vector values. This analysis is repeated for each model.

The Bayes' rule can be obtained by applying 2.2 and considering that $p(\mathrm{x}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}, \mathrm{x})$

$$p(\boldsymbol{\theta}|\mathrm{x}) = \frac{p(\mathrm{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(\mathrm{x})} \tag{2.48}$$

Equation 2.48 describes the relation between the marginal and the conditional distributions.

The second level of inference is the task of model comparison. In this case we wish to compare the models and assign a preference or rank to them.
Bayesian inference is distinct from classical decision theory. The goal of inference is, given a defined hypothesis space and a particular data set, to assign probabilities to the hypotheses. Decision theory typically chooses between alternative actions on the basis of these probabilities so as to minimize the expectation of a cost function. Thus, ideal Bayesian predictions do not involve a choice between models ; rather, predictions are made by summing over all the alternative models, weighted by their probabilities.
The first level of Bayesian methods differs from the classical statistics by the inclusion of subjective priors. On the other hand, depending on the prior, a Bayesian result will often differ little from the outcome of a classical approach. What makes Bayesian methods attractive is the second level of inference which involves model comparison.
Model comparison is a difficult task because it is not possible simply to choose the model that fits the data best : more complex models can always fit the data better, so the maximum likelihood model choice would lead us inevitably to implausible, over-parametrized models, which generalize poorly. In this context the Occam razor principle comes in our help. In essence the principle, attributed to 14th-century English logician and Franciscan friar, William of Ockham, states that : when competing hypotheses are equal, the principle recommends selection of the hypothesis that introduces the fewest assumptions and postulates the fewest entities while still sufficiently answering the question. The second level of Bayesian inference embeds the Occam factor enabling model comparison.

2.6.1 Level I : Model Fitting

The first level of inference assumes that a model $\mathcal{M}_i$ is true. The task consists in fitting the model to the data in order to infer the most probable parameter Θ_i. The posterior probability distribution of the model parameter vector $\Theta_i = \boldsymbol{\theta}$ given the data X and the chosen model $\mathcal{M}_i$ is given by

$$p(\boldsymbol{\theta}|\mathrm{X}, \mathcal{M}_i) = \frac{p(\mathrm{X}|\boldsymbol{\theta}, \mathcal{M}_i)p(\boldsymbol{\theta}|\mathcal{M}_i)}{p(\mathrm{X}|\mathcal{M}_i)} \tag{2.49}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \tag{2.50}$$

where $p(\boldsymbol{\theta}|\mathcal{M}_i)$ is the prior of the parameter and $p(\mathrm{X}|\theta, \mathcal{M}_i)$ is the likelihood of the data given the parameter and the model. The probability $p(\mathrm{X}|\mathcal{M}_i)$ is the evidence of the model $\mathcal{M}_i$. Through the first level of Bayesian inference, we obtain the MAP estimator $\hat{\boldsymbol{\theta}}$ of the vector parameter $\boldsymbol{\theta}$, defined as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta}\{p(\mathrm{X}|\boldsymbol{\theta}, \mathcal{M}_i)p(\boldsymbol{\theta}|\mathcal{M}_i)\}. \tag{2.51}$$

In equation 2.51 the evidence term does not appear because it does not depend on the parameter vector, thus the first derivative of $p(\mathrm{X}|\mathcal{M}_i)$ with respect to $\boldsymbol{\theta}$, which has to be computed in 2.51 to maximize 2.49, is zero. Thus, at this level of inference, the evidence is a constant factor and can be neglected.

2.6.2 Level II : Model Selection

The task of the second level of Bayesian inference is to find the most plausible model explaining the data. The posterior probability of a model $\mathcal{M}_i$ is

$$p(\mathcal{M}_i|\mathrm{X}) = \frac{p(\mathrm{X}|\mathcal{M}_i)p(\mathcal{M}_i)}{p(\mathrm{X})} \propto p(\mathrm{X}|\mathcal{M}_i)p(\mathcal{M}_i) \tag{2.52}$$

where $p(\mathrm{X}|\mathcal{M}_i)$ is the model evidence, which is the probability of the data, given the model $\mathcal{M}_i$, and $p(\mathcal{M}_i)$ is the prior probability of the model.
Assuming that we choose to assign equal priors $p(\mathcal{M}_i)$ to the alternative models, models $\mathcal{M}_i$ are ranked by evaluating the evidence.
It can be obtained by marginalization

$$p(\mathrm{X}|\mathcal{M}_i) = \int p(\mathrm{X}|\boldsymbol{\theta}, \mathcal{M}_i)p(\boldsymbol{\theta}|\mathcal{M}_i)\mathrm{d}\boldsymbol{\theta} \tag{2.53}$$

where the integral is evaluated on the parameter space, $p(\mathrm{X}|\boldsymbol{\theta}, \mathcal{M}_i)$ is the likelihood and $p(\boldsymbol{\theta}|\mathcal{M}_i)$ is the prior (see 2.49).
We compute the evidence in order to assess the quality of the model because it is a measure of the degree of belief of the model fitting the data. The marginalization over the parameter space turns into a measure independent from the robustness of the estimation of the parameter, thus it depends only on the data and the model itself.

2.6.3 Evidence Evaluation and Occam Razor

The evidence integral 2.53 sometimes can be calculated in a closed form, but other times it has a complicate form which cannot be solved analytically, but has to be approximated or numerically solved.
A common way to approximate it is to use the Laplace method (or saddle point approximation Butler (2007)) which exploits the Taylor expansion of the integrand around the peak (Gaussian approximation) $\hat{\boldsymbol{\theta}}$. Equation 2.51 ensures that the integrand has a maximum around $\hat{\boldsymbol{\theta}}$, thus the evidence can be approximated as

$$p(X|\mathcal{M}_i) \approx p(X|\boldsymbol{\theta}_{MAP}, \mathcal{M}_i) \underbrace{p(\boldsymbol{\theta}_{MAP}|\mathcal{M}_i) \det\left(\mathbf{H}/2\pi\right)^{-\frac{1}{2}}}_{\text{Occam factor}} \tag{2.54}$$

where $\mathbf{H} = -\nabla^2 \ln p(\boldsymbol{\theta}|X, \mathcal{M}_i)$ is proportional to the Hessian matrix. Reminding the definition of the Fisher information 2.38 the similarity between $\mathbf{H}$ and $\mathbf{I}(\boldsymbol{\theta})$ can be noted : the Fisher information is the expectation of the Hessian matrix.
The logarithm of the Occam factor can be written as

$$\log p(\boldsymbol{\theta}_{MAP}|\mathcal{M}_i) - \frac{k}{2} \log \frac{n}{2\pi} - \log \det\left(\mathbf{I}(\boldsymbol{\theta})\right) \tag{2.55}$$

where n is the number of samples and k is the degree of freedom of the process.
In summary the evidence is obtained by multiplying the best fit likelihood by the Occam factor.
In case of linear model

$$\mathbf{x} = \mathbf{G}\boldsymbol{\theta} + \mathbf{e} \tag{2.56}$$

where $\mathbf{G}$ is a linear transformation and $\mathbf{e}$ is the realization of a iid Gaussian process, the evidence is found to be (Ruanaidh & Fitzgerald, 1996)

$$p(\mathbf{x}|\mathbf{G}, \mathbf{e}) \approx \frac{\pi^{-n/2} \Gamma\left(\frac{q}{2}\right) \Gamma\left(\frac{n-q}{2}\right) \det\left(\mathbf{G}^T\mathbf{G}\right)^{-1/2}}{4 R_\delta R_\sigma (\hat{\boldsymbol{\theta}}^T \hat{\boldsymbol{\theta}})^{q/2} \hat{\sigma}^{n-q}} \tag{2.57}$$

where n is the cardinality of the vector $\mathbf{x}$, q is the cardinality of the vector $\boldsymbol{\theta}$, $\Gamma(\cdot)$ is the Gamma function, R_δ and R_σ are normalization constants, $\hat{\boldsymbol{\theta}}$ is the Least Squares Error (LSE) of the parameter vector $\boldsymbol{\theta}$ and $\hat{\sigma}^2$ is the estimated model variance

$$\hat{\boldsymbol{\theta}} = (\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}^T\mathbf{x} \tag{2.58}$$

$$\hat{\sigma}^2 = \mathbf{x}^T\mathbf{x} - (\mathbf{G}\hat{\boldsymbol{\theta}})^T(\mathbf{G}\hat{\boldsymbol{\theta}}) \tag{2.59}$$

where $(\mathbf{G}\mathbf{G}^T)^{-1}$ is the pseudo-inverse matrix.
The Bayes theorem has been applied for model fitting and model comparison, the Occam factor is included in the second level of inference and it is also possible to link it to the mutual information in the context of information theory as we will see in the following section.

2.7 Elements of Information Theory

In this section some basic concept of the information theory are introduced spanning from Shannon Entropy to Rate Distortion and giving the definition of the Kullback-Leibler distance and the mutual information. The distortion measure will be used for model selection, thus the connection with the Bayesian estimation will be provided.

2.7.1 Measure of Information and Entropy

The entropy, introduced by Shannon (1948), is a measure of the average information of the realization of a random variable X. If the random variable is defined in a discrete observation space $\mathcal{X}$ the entropy is defined as follows

$$H(\mathrm{X}) = -\sum_{\mathrm{x}\in\mathcal{X}} p(\mathrm{x}) \log p(\mathrm{x}) \tag{2.60}$$

which represents the expectation of the random variable $-\log p(\mathrm{x})$. When the incertitude is high the entropy is also high, while the entropy of a certain event is zero. The quantity $-\log p(\mathrm{x})$ represent the information contained in the random variable x and in case the logarithm is base 2, then the information is measured in bit. If for example the logarithm has natural base the information is measured in nat. From the definition of the entropy Shannon derived the source coding theorem. If at each realization x a code c_{x} of length $\mathcal{L}(c_{\mathrm{x}})$ is assigned in order to ensure a unique decoding, the following inequality, for the average length, is respected

$$\sum_{\mathrm{x}\in\mathcal{X}} p(\mathrm{x})\mathcal{L}(c_{\mathrm{x}}) \geq H(\mathrm{X}) \tag{2.61}$$

where the equality is verified when $\mathcal{L}(c_{\mathrm{x}}) = -\log p(\mathrm{x})$. A way to ensure that the code has a unique decoding is to avoid that a codeword is a prefix of a longer one. A binary code which satisfies the prefix rule exists if and only if the following inequality is respected

$$\sum_{\mathrm{x}\in\mathcal{X}} 2^{-\mathcal{L}(c_{\mathrm{x}})} \leq 1 \tag{2.62}$$

which is called Kraft inequality (Kraft, 1949). In the case that $\mathcal{L}(c_{\mathrm{x}}) = \lceil -\log p(\mathrm{x}) \rceil$, the average length $\sum_{\mathrm{x}\in\mathcal{X}} p(\mathrm{x})\mathcal{L}(c_{\mathrm{x}})$ is inferior to $H(\mathrm{X}) + 1$ (Huffman, 1952). The Shannon theorem says that it is possible to approach the limit established for the entropy. On the other hand, we have to point out that the theorem is valid for sources whose realizations are an infinite series of symbols and it is strongly dependent on the source distribution.

2.7.2 Kullback-Leibler Divergence and Mutual Information

The Kullback-Leibler divergence was introduced by Kullback & Leibler (1951) in the framework of the statistical problem of discrimination. The divergence is a measure of the inefficiency of assuming that the distribution of a random variables is $q(\mathrm{X})$ when the true distribution is $p(\mathrm{X})$. It can be interpreted as a version of the maximum entropy principle (Jaynes, 1957). The Kullback-Leibler divergence is defined as

$$D_{KL}(p||q) = \sum_{\mathrm{x}\in\mathcal{X}} p(\mathrm{x}) \log \frac{p(\mathrm{x})}{q(\mathrm{x})} \tag{2.63}$$

From the point of view of coding theory it means that if we have a random source $\mathcal{X}$ with a probability $p(\mathrm{X})$ we build a code using $q(\mathrm{X})$ as distribution, i.e. we need $H(p) + D(p||q)$ bits on the average to describe the random variable. The divergence is zero when $q(\mathrm{x}) = p(\mathrm{x})$. It is worth to notice that it is not symmetric.
The extension of the divergence measure to couples of random variable can be done

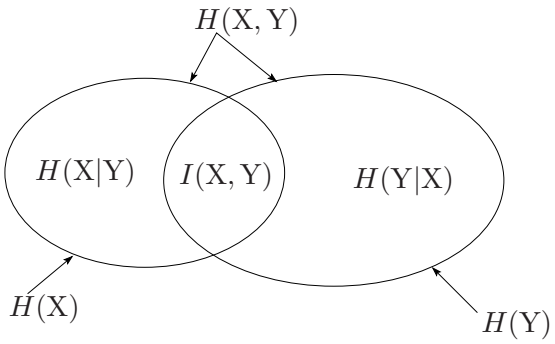


FIGURE 5 – Venn diagram illustrates the relationships among Shannon's basic measures of information : entropy, joint entropy, conditional entropy and mutual information.

considering two random variables X and Y with joint probability density $p(\mathrm{X},\mathrm{Y})$, thus the mutual information is defined as

$$I(\mathrm{X},\mathrm{Y}) = D(p(\mathrm{X},\mathrm{Y})||p(\mathrm{X})p(\mathrm{Y})) = \sum_{\mathrm{x}\in\mathcal{X}}\sum_{\mathrm{y}\in\mathcal{Y}} p(\mathrm{x},\mathrm{y})\log\frac{p(\mathrm{x},\mathrm{y})}{p(\mathrm{x})p(\mathrm{y})}. \tag{2.64}$$

It measures the mutual dependence of two random variables or, in other words, it represents the reduction in the uncertainty of X due to the knowledge of Y. Alternative definitions for the mutual information are the following

$$\begin{aligned} I(\mathrm{X},\mathrm{Y}) &= H(\mathrm{X}) + H(\mathrm{Y}) - H(\mathrm{X},\mathrm{Y}) && (2.65)\\ &= H(\mathrm{X}) - H(\mathrm{X}|\mathrm{Y}) && (2.66)\\ &= H(\mathrm{Y}) - H(\mathrm{Y}|\mathrm{X}) && (2.67)\end{aligned}$$

where the mutual entropy $H(\mathrm{X},\mathrm{Y})$ and the conditional entropies $H(\mathrm{X}|\mathrm{Y})$ and $H(\mathrm{Y}|\mathrm{X})$ are defined as follows

$$\begin{aligned} H(\mathrm{X},\mathrm{Y}) &= -\sum_{\mathrm{x}\in\mathcal{X}}\sum_{\mathrm{y}\in\mathcal{Y}} p(\mathrm{x},\mathrm{y})\log p(\mathrm{x},\mathrm{y}) && (2.68)\\ H(\mathrm{X}|\mathrm{Y}) &= -\sum_{\mathrm{x}\in\mathcal{X}}\sum_{\mathrm{y}\in\mathcal{Y}} p(\mathrm{x},\mathrm{y})\log p(\mathrm{x}|\mathrm{y}) && (2.69)\\ H(\mathrm{Y}|\mathrm{X}) &= -\sum_{\mathrm{x}\in\mathcal{X}}\sum_{\mathrm{y}\in\mathcal{Y}} p(\mathrm{x},\mathrm{y})\log p(\mathrm{y}|\mathrm{x}) && (2.70)\end{aligned}$$

The quantities in 2.65 are shown graphically in the Venn diagram in Figure 5. Many other information measures exist (Taneja, 2001), but we use the one described above because it allows us to define the theory for lossy compression.

2.7.3 Rate Distortion Theory

The rate distortion theory gives a quantitative answer to the goodness of representing the information of a given source, formalizing the number of bits necessary for coding the realizations of a random process X given the fact that losses are allowed during coding. We suppose that a distortion measure $d(\mathrm{x},\hat{\mathrm{x}})$ defined between the realization x and its reconstructed version $\hat{\mathrm{x}}$ is given. Moreover we have R bits available to represent the

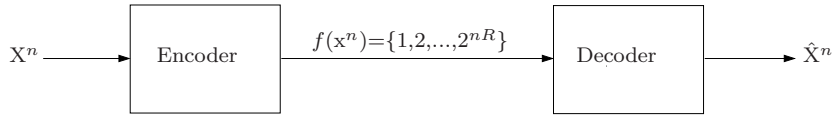


FIGURE 6 – Rate distortion encoder and decoder.

information. The functions of coding and decoding of n consecutive realizations are as follows

$$f_n \quad : \quad \mathcal{X}^n \rightarrow \{1, 2, \ldots, 2^{nR}\} \tag{2.71}$$

$$g_n \quad : \quad \{1, 2, \ldots, 2^{nR}\} \rightarrow \hat{\mathcal{X}}^n \tag{2.72}$$

A scheme of a rate distortion encoder-decoder is shown in Figure 6. The distortion between sequences X^n and $\hat{\mathrm{X}}^n$ is defined by

$$D = \mathrm{E}\{d(\mathrm{X}^n, \hat{\mathrm{X}}^n)\} = \frac{1}{n}\sum_{i=1}^{n} d(\mathrm{X}_i, g_n(f_n(\mathrm{X}_i))) \tag{2.73}$$

So the distortion for a sequence is the average of the per symbol distortion of the element of the sequence.
Shannon (1959) defined the achievable coding region as the ensemble of the pair (R,D) obtainable by a set of coding-decoding functions when n tends to infinitive. The bound of the region is given by the distortion rate function $R(D)$, see Figure 6, for which the inequality $R \geq R(D)$ is verified. On the other hand, the rate distortion function as inferior lower bound can be reformulated as a minimization problem

$$R(D) = \min_{p(\mathrm{x}|\hat{\mathrm{x}}):\mathrm{E}_{(\mathrm{x},\hat{\mathrm{x}})}\{d(\mathrm{x},\hat{\mathrm{x}})\}\leq D} I(\mathrm{X}, \hat{\mathrm{X}}) \tag{2.74}$$

where the minimization is over the conditional pdf $p(\mathrm{x}|\hat{\mathrm{x}})$ which represents the probability of having the output $\hat{\mathrm{x}}$ given as input the symbol x. The distortion rate function is a concave decreasing function, see Figure 7, and represents the rate vs. the distortion : for low rate the losses are high.

2.8 Rate Distortion and Model Selection

The rate distortion problem, presented in the previous section, is addressed as a minimization problem. This can be exploited in order to perform model selection by choosing the model with the minimum distortion.
In order to explain model selection by rate distortion, we present the mutual information of a Gaussian channel and we show that it is the inverse of the Occam factor, thus the minimization of the mutual information (rate distortion) means the maximization of the Occam factor.
On the other hand, the example provided by the Gaussian source is general, because a distribution can be approximated as a Gaussian close to the maximum exploiting the Laplace method (Butler, 2007).

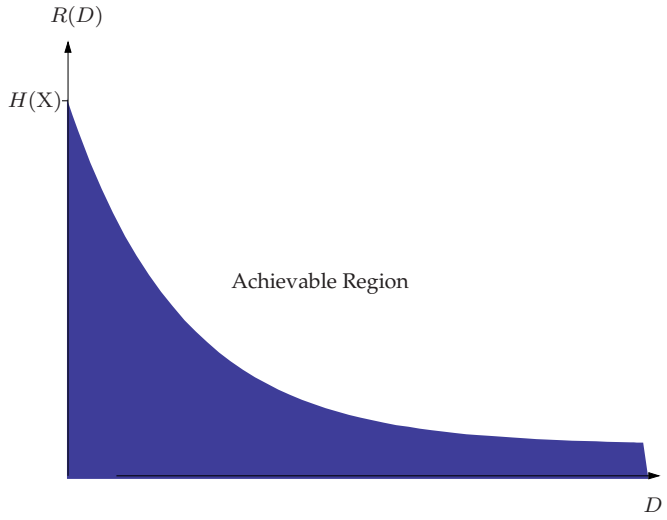


FIGURE 7 – Example of Rate distortion curve.

The mutual information of a Gaussian source $\mathcal{N}(0, \sigma^2)$ is (Cover & Thomas, 1991b)

$$\begin{aligned} I(\mathrm{X}, \hat{\mathrm{X}}) &= H(\mathrm{X}) - H(\mathrm{X}|\hat{\mathrm{X}}) && (2.75)\\ &= \frac{1}{2}\log(2\pi e)\sigma^2 - H(\mathrm{X}-\hat{\mathrm{X}}|\hat{\mathrm{X}}) && (2.76)\\ &\geq \frac{1}{2}\log(2\pi e)\sigma^2 - H(\mathrm{X}-\hat{\mathrm{X}}) && (2.77)\\ &\geq \frac{1}{2}\log(2\pi e)\sigma^2 - H(\mathcal{N}(0, \mathrm{E}\{(\mathrm{X}-\hat{\mathrm{X}})^2\})) && (2.78)\\ &\geq \frac{1}{2}\log(2\pi e)\sigma^2 - \frac{1}{2}\log(2\pi e)\mathrm{E}\{(\mathrm{X}-\hat{\mathrm{X}})^2\} && (2.79)\\ &\geq \frac{1}{2}\log(2\pi e)\sigma^2 - \frac{1}{2}\log(2\pi e)D && (2.80)\\ &= \frac{1}{2}\log\frac{\sigma^2}{D} && (2.81) \end{aligned}$$

where to write $\mathrm{E}\{(\mathrm{X}-\hat{\mathrm{X}})^2\} = D$ we considered a squared-error distortion measure. Thus, referring to 2.54 we can write for the one dimensional case

$$\text{Occam factor} = p(\theta_{MAP}|\mathcal{M}_i)\sigma_{\theta|\mathbf{x}} \tag{2.82}$$

which, considering a non-informative prior Jeffreys distribution $p(\theta_{MAP}|\mathcal{M}_i) = 1/\sigma_\theta$ can be rewritten

$$\text{Occam factor} = \frac{\sigma_{\theta|\mathbf{x}}}{\sigma_\theta} \tag{2.83}$$

which correspond to the ratio of the posterior accessible volume of $\mathcal{M}_i$ parameter space to the prior accessible volume. Thus, by 2.75 and 2.83 the mutual information can be written as

$$I(\mathrm{X}, \hat{\mathrm{X}}) = \log\sqrt{\frac{\text{Volume of the Prior}}{\text{Volume of the Posterior}}} = \log\frac{1}{\text{Occam factor}}. \tag{2.84}$$

Selecting the model which minimizes the mutual information $I(\mathrm{X}, \hat{\mathrm{X}})$ is equivalent to selecting the model which maximizes the Occam factor. This states the equivalence of using rate distortion and Bayesian model selection. We are going to apply the rate distortion for model selection in order to estimate optimal model for feature extraction.

2.9 Summary of the chapter

In this chapter the theoretical background of the statistical framework has been presented. The definition of stochastic process and Markovian process has been provided. The equality of Markovian processes and Gibbs distributions has been presented and the GMRF family has been described.
Furthermore, the bases of parameter estimation have been delineated with emphasis on the Bayesian framework. The concept of entropy from information theory has been linked with parameter estimation through the Cramér-Rao inequality. Last but not least, the rate distortion has been presented at the end of the chapter focusing the attention to its link with the Bayesian model selection.
The basic theory presented will be used in the next chapter in order to design the Tikhonov like restoration filter, in the frame of Bayesian MAP estimate. The Rate Distortion will be used to provide results for model selection.

Chapitre 3

Image Restoration

This chapter presents the models for Maximum A Posteriori *(MAP) estimation and Tikhonov regularization. It starts with an overview of Model Based Despeckling (MBD) which has been our reference method for image despeckling and feature extraction. It follows with an extension of the MAP estimate in the complex domain which provides the reconstructed complex channels but not a despeckled image.*
Then, the Gauss-Markov Random Field (GMRF) family of models extended to the complex-valued domain is presented. The complex-valued linear model embeds the system transfer function, thus the parameters are able to model and separate correlated textures.
The Tikhonov MAP estimate has been developed and tested with the complex-valued linear model, as data model, and the Huber-Markov model as prior function. Then, the prior has been replaced with the GMRF family of models in order to enable parameter estimation.
The methods have been compared with classical estimation filters and with MBD for feature comparison.

3.1 Model Based Despekling and Feature Extraction

The MBD has been developed by Walessa & Datcu (2000). The radar cross section $|\gamma_0|^2$ is the noiseless image we want to estimate from an observed intensity SAR image $\mathrm{y} = \mathrm{I}_m$, see Section 1.7.1. The microwave propagation and the SAR processing (focusing) are modeled as a linear system. The noisy intensity image y is approximated as the reflectivity $|\gamma_0|^2$ convolved with the incoherent system point spread function $|\mathrm{s}|^2$, where s is the point spread function of the coherent system.
The speckle effect is modeled as a multiplicative Gamma distributed noise n, correlated by the SAR end-to-end system (Equation 12). Hence, the degraded image multiplicative model may be written as :

$$y = \left[|\gamma_0|^2 * |\mathrm{s}|^2\right] \cdot \mathrm{n} = \mathrm{x} \cdot \mathrm{n} \tag{3.1}$$

The problem to be solved is the reconstruction of the original radar cross section from y (inverse problem). Unlike conventional techniques that only remove speckle noise by estimating x, an inverse approach takes into account the SAR image formation for image reconstruction. The problem is formulated as a Bayesian model fitting and solved by using a scene-understanding paradigm. The algorithm performs a model selection based

on evidence calculation and considers the approximation $\mathrm{x} = |\gamma_0|^2$. It induces over the class of GMRFs a partition in a family of models centered on fixed values of the parameter vector $\boldsymbol{\theta}$. The algorithm is used for the estimation of parameters from textured images in the presence of noise. The model is locally chosen according to the evidence (second level of Bayesian inference).
The likelihood function of the observed square root intensity, presented in Section 1.5.2, is a Gamma distribution, given by

$$p(\mathrm{y}|\mathrm{x}) = 2\left(\frac{\mathrm{y}}{\mathrm{x}}\right)^{2L-1}\frac{L^L}{\mathrm{x}\Gamma(L)}\exp\left(-L\left(\frac{\mathrm{y}}{\mathrm{x}}\right)^2\right) \tag{3.2}$$

where L denotes the Equivalent Number of Look (ENL). The prior function is a GMRF model

$$p(\mathrm{x}|\mathrm{x}_r, r\in\mathcal{N},\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\mathrm{x}-\sum_{r\in\mathcal{N}}\theta_r\mathrm{x}_r)^2}{2\sigma^2}\right\} \tag{3.3}$$

where $\boldsymbol{\theta}$ and σ are model parameters describing textural information and the prediction uncertainty of the model, respectively. The prior 3.3 is equivalent to 2.19, i.e. GMRF, for a symmetric neighborhood system.
The first derivative of the log-posterior approximated by the product of the likelihood 3.2 and the prior 3.3 leads to the MAP estimation of the noise free image x, given by

$$\frac{\partial}{\partial\mathrm{x}}\log p(\mathrm{x}|\mathrm{y},\mathrm{x}_r, r\in\mathcal{N},\boldsymbol{\theta}) = -\frac{2L}{\mathrm{x}} + \frac{2L\mathrm{y}^2}{\mathrm{x}^3} - \frac{\mathrm{x}-\sum_{r\in\mathcal{N}}\theta_r\mathrm{x}_r}{\sigma^2}. \tag{3.4}$$

Equation 3.4 yields a fourth order polynomial where the solution depends on the model parameter vector $\boldsymbol{\theta}$.
In the general case, an arbitrary number of models $p_i(\mathrm{x}|\boldsymbol{\theta}_i)$ characterized by their potential functions and having a different number of parameters are considered. However, for simplicity, we denote the used models by $p(\mathrm{x}|\boldsymbol{\theta})$. To select the best model and take into account the noise, the algorithm has to maximize the evidence

$$p(\mathrm{y}|\boldsymbol{\theta}) = \int_{\mathrm{x}} p(\mathrm{y}|\mathrm{x},\boldsymbol{\theta})p(\mathrm{x}|\boldsymbol{\theta})\mathrm{d}\mathrm{x} \tag{3.5}$$

$$\approx \frac{(2\pi)^{\frac{N^2}{2}}}{\sqrt{|\mathbf{H}|}}\prod_{i=1}^{N^2} p(\mathrm{y}_i|\mathrm{x}_{\mathrm{MAP},i})p(\mathrm{x}_{\mathrm{MAP},i}|\mathrm{x}_{\mathrm{MAP},r}, r\in\mathcal{N}_i,\boldsymbol{\theta}) \tag{3.6}$$

as a function of $\boldsymbol{\theta}$. The evidence integral is approximated with the Laplace method. By further approximation of the Hessian matrix with the elements of its main diagonal, the determinant takes the form $|\mathbf{H}| = \prod_{i=1}^{N^2} h_{ii}$. The latter approximation is consistent with the hypothesis of statistical independence which is done in general in order to write the likelihood as the product of the densities. Thus, the log-evidence can be written as

$$\log p(\mathrm{y}|\boldsymbol{\theta}) = \sum_{i=1}^{N^2}\frac{1}{2}(\log 2\pi - \log h_{ii}) + \log p(\mathrm{y}_i|\mathrm{x}_{\mathrm{i,MAP}}) + \log p(\mathrm{x}_{\mathrm{MAP},i}|\mathrm{x}_{\mathrm{MAP},r}, r\in\mathcal{N}_i,\boldsymbol{\theta}) \tag{3.7}$$

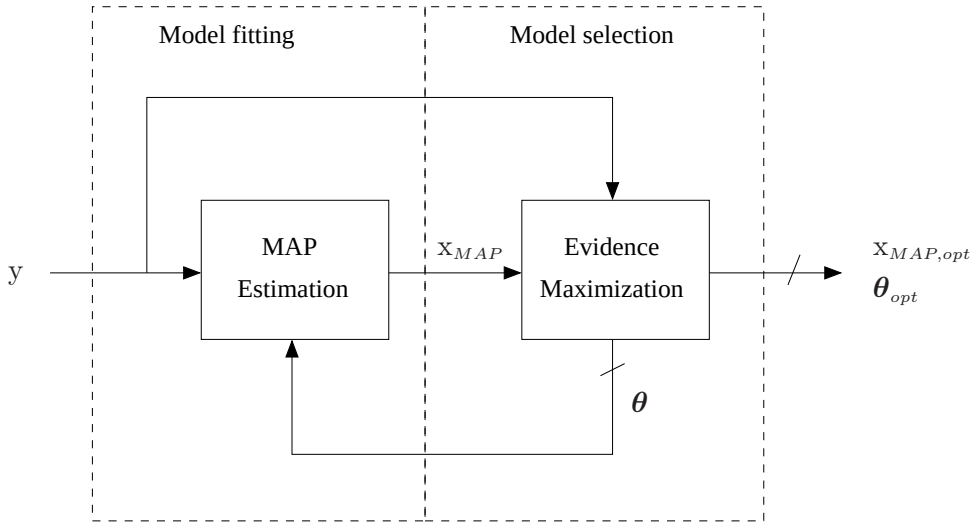


FIGURE 1 – Flowchart of the MBD method.

where h_{ii} are the elements of the main diagonal of the Hessian matrix

$$h_{ii} = -\frac{\partial^2}{\partial x_i^2}\left[\sum_{j=1}^{N^2}\log p(x_j|y_j, x_r, r\in\mathcal{N}_j, \boldsymbol{\theta})\right] \tag{3.8}$$

$$= \frac{6Ly_i^2}{x_{i,MAP}^4} - \frac{2L}{x_{i,MAP}^2} + \frac{1}{\sigma^2}\left(1+\sum_{r\in\mathcal{N}_i}\theta_r^2\right). \tag{3.9}$$

Being able to approximately compute the evidence, the final step for parameter estimation consists in finding the maximizing parameter vector. To achieve this, the evidence is maximized according to the following iterative algorithm

1. Choose an initial guess for $\boldsymbol{\theta}$.
2. Calculate the MAP estimate 3.4 of x using the current parameter values of $\boldsymbol{\theta}$.
3. Compute the evidence 3.5 with $\boldsymbol{\theta}$ and x_{MAP}.
4. Keeping x_{MAP} fixed, a new $\boldsymbol{\theta}$ is iteratively chosen, in a new loop, to maximize the evidence and perform model selection.
5. This procedure is repeated from step one with the new $\boldsymbol{\theta}$ until convergence is reached.

Thus, the procedure allows not only to estimate the best parameter explaining the model (model fitting or first level of Bayesian inference) but also to compare the evidence of different models and to chose the one with the highest evidence exploiting the model selection capability of the second level of Bayesian inference.

3.2 MAP Estimation and Feature Extraction in Complex Domain

The MBD model works fine on detected data, thus we want to extend the MAP method to the complex domain investigating models which fit with complex-valued images.

One model proposed by Pascazio & Ferraiuolo (2003) for statistical regularization of tomographic images has the following energy function

$$U(\mathbf{x}) = \frac{||\mathbf{y} - \mathbf{A}\mathbf{x}||^2}{2\sigma^2} + \frac{\sum_{r\in\mathcal{N}_s}(\mathbf{x}_{R,s} - \mathbf{x}_{R,r})^2}{2\sigma_R^2} + \frac{\sum_{r\in\mathcal{N}_s}(\mathbf{x}_{I,s} - \mathbf{x}_{I,r})^2}{2\sigma_I^2} \tag{3.10}$$

where the first term represents the likelihood function and the second and third term represent the prior GMRF model applied to the real and imaginary part respectively.
The notation refers to stationary random fields. It can be extended to non-stationary data exploiting the local Markov property which leads to the sum of the energy functions. In term of computation, the non-stationarity is faced by fixing boundary conditions of continuity.
Even if the prior models separately real and imaginary part, the posterior jointly depend on them because of the likelihood term. In 3.10 an isotropic GMRF prior is considered to model the real and the imaginary part. The advantage of using the likelihood presented in 3.10 is that it refers to the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{3.11}$$

where the matrix $\mathbf{A}$ can embed the system transfer function as well as the image formation model (Çetin & Karl, 2001). Nevertheless, the posterior combines the two channels through the likelihood, the prior smooths the real and imaginary parts because the isotropic GMRF model is equivalent to the Total Variation (TV) (Pérez, 1998). As a consequence, the speckle, modeled as multiplicative noise in the detect the image, is not removed.
Differently from the MBD approach, the parameters are estimated in advance from the incomplete data by an EM algorithm. The complex image is thus regularized through a MAP estimate.

3.2.1 Expectation-Maximization parameter estimation

The steps of the EM algorithm (Dempster et al., 1977) are :

1. **Expectation**. It consists of the evaluation of the quantity

$$M(\boldsymbol{\sigma} = \boldsymbol{\sigma}(t)) = \mathrm{E}\{\ln f_{\mathbf{X}}(\mathbf{X}, \boldsymbol{\sigma})\} \tag{3.12}$$

   where $f_{\mathbf{X}}(\mathbf{X}, \boldsymbol{\sigma})$ is the referred likelihood function and the expectation allows the estimation of the log-likelihood from the current available data.

2. **Maximization**. It consists on the evaluation of

$$\boldsymbol{\sigma}(t+1) = \arg\max_{\boldsymbol{\sigma}} M(\boldsymbol{\sigma} = \boldsymbol{\sigma}(t)) \tag{3.13}$$

   The updated parameters are iteratively replaced in the Expectation step until convergence is reached.

In our case, the likelihood $f_{\mathbf{X}}(\mathbf{X}, \boldsymbol{\sigma})$ used for the estimation of the parameters involves the prior model which is the second and the third term on the right hand side of 3.10. Thus it takes the form from the product of the Gibbs distributions

$$f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\sigma}) = \prod_{s=1}^{N^2} \frac{1}{Z_R} \exp\{-U(\mathrm{x}_{R,s})\} \frac{1}{Z_I} \exp\{-U(\mathrm{x}_{I,s})\} \tag{3.14}$$

where

$$U(\mathrm{x}_{R,s}) = \frac{\sum_{r\in\mathcal{N}_s}(\mathrm{x}_{R,s}-\mathrm{x}_{R,r})^2}{2\sigma_R^2} \tag{3.15}$$

$$U(\mathrm{x}_{I,s}) = \frac{\sum_{r\in\mathcal{N}_s}(\mathrm{x}_{I,s}-\mathrm{x}_{I,r})^2}{2\sigma_I^2} \tag{3.16}$$

and $Z_R = z_R\sigma_R^{N^2}$ and $Z_I = z_I\sigma_I^{N^2}$ are the partition functions with z_R and z_I constant factors not depending on $\boldsymbol{\sigma}$.
The Maximization step of the log-likelihood 3.14 leads to the following expressions (Saquib et al., 1998) for the estimators

$$\sigma_R^2(t+1) = \frac{1}{N^2}\mathrm{E}\left\{\sum_{r\in\mathcal{N}_s}(\mathrm{x}_{R,s}-\mathrm{x}_{R,r})^2\right\} \tag{3.17}$$

$$\sigma_I^2(t+1) = \frac{1}{N^2}\mathrm{E}\left\{\sum_{r\in\mathcal{N}_s}(\mathrm{x}_{I,s}-\mathrm{x}_{I,r})^2\right\} \tag{3.18}$$

The Expectation step can be evaluated exploiting the ergodicity of the process approximating the ensemble expected-value by time averaging. The latter has to be performed on the realization of the *a posteriori* distribution which has the form of a Gibbs distribution with 3.10 as energy function. It can be obtained by the use of a Gibbs sampler considering the local *a posteriori* distribution to be Gaussian

$$p_{\mathrm{X}_{R,k}}(\mathrm{x}_{R,k}|\mathrm{x}_{R,r}, r\in\mathcal{N}_k, \mathbf{Y}=\mathbf{y}) = \frac{1}{\sqrt{2\pi}\sigma_{R,k}}\exp\left\{-\frac{(\mathrm{x}_{R,k}-\mu_{R,k})^2}{2\sigma_{R,k}^2}\right\} \tag{3.19}$$

where the mean $\mu_{R,k}$ and the variance $\sigma_{R,k}^2$ are (Pascazio & Ferraiuolo, 2003)

$$\mu_{R,k} = \frac{\displaystyle\sum_{r\in\mathcal{N}_k}\frac{\mathrm{x_{R,r}}}{2\sigma_{R,kr}^2} + \sum_{h=1}^{N^2}\frac{|a_{hk}|^2\mathrm{Re}\left\{\frac{\mathrm{y}_h-\sum_{l\neq k}a_{hl}\mathrm{x}_l - ja_{hk}\mathrm{x_{I,k}}}{a_{hk}}\right\}}{4\sigma^2}}{\displaystyle\sum_{r\in\mathcal{N}_k}\frac{1}{2\sigma_{R,kr}^2} + \sum_{h=1}^{N^2}\frac{|a_{hk}|^2}{4\sigma^2}} \tag{3.20}$$

$$\sigma_{R,k}^2 = \left[\sum_{r\in\mathcal{N}_k}\frac{2}{\sigma_{R,kr}^2} + \sum_{h=1}^{N^2}\frac{|a_{hk}|^2}{\sigma^2}\right]^{-1} \tag{3.21}$$

where a_{ij} are the elements of the matrix $\mathbf{A}$, which causes a mutual coupling of the I and Q channels. Thus, Equation 3.19 is used to generate samples for the real part which depends on the imaginary part and on the whole data set. The imaginary part has the same expression as 3.19 where the sub-index $\cdot_R$ is replaced with $\cdot_I$.
The convergence criteria can be the desired number of iterations or a measure distance $d(\boldsymbol{\sigma}(t+1)-\boldsymbol{\sigma}(t)) \leq \delta$ that equals or is less than a desired value $\delta > 0$: a common distance measure is the squared error $||\boldsymbol{\sigma}(t+1)-\boldsymbol{\sigma}(t))||^2$.
The pseudo code of the EM algorithm is shown in Algorithm 1.

Input: Data y, Data size $N \times N$, Noise variance σ_n^2
Output: Parameter θ

$\theta_{new} \leftarrow$ `ParameterEstimation` (y)
$\theta_{old} \leftarrow 0$
$\mathrm{x} \leftarrow \mathrm{y}$
while $|\theta_{old} - \theta_{new}| > 10^{-6}$ **do**
  for $i \leftarrow 1$ **to** N **do**
    for $j \leftarrow 1$ **to** N **do**
      $\mu \leftarrow$ `MeanValueEstimation` $(\mathrm{x}, \theta_{new}, \sigma_n^2)$
      $\sigma^2 \leftarrow$ `VarianceEstimation` $(\theta_{new}, \sigma_n^2)$
      $\mathrm{x}[i,j] \leftarrow \mathcal{N}(\mu, \sigma^2)$
    end
  end
  $\theta_{old} \leftarrow \theta_{new}$
  $\theta_{new} \leftarrow$ `ParameterEstimation` (x)
end
$\theta \leftarrow \theta_{new}$

Algorithm 1: EM algorithm for parameter estimation. The function `ParameterEstimation()` refers to the differential parameter of the GMRF prior model. The `MeanValueEstimation()` and `VarianceEstimation()` functions refer to Equations 3.20 and 3.21, respectively. For the sake of simplicity the code refers to parameter estimation of the real or imaginary part.

3.2.2 MAP image regularization

Equation 3.10 defines a quadratic functional which can be minimized by using a steepest descent technique (Shewchuk, 1994) to find the MAP solution. The complex gradient is given by

$$\nabla U(\mathbf{x}) = -2\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}) + 2\frac{\sum_{r\in\mathcal{N}_s}(\mathbf{x}_{R,s} - \mathbf{x}_{R,r})}{2\sigma_R} - 2j\frac{\sum_{r\in\mathcal{N}_s}(\mathbf{x}_{I,s} - \mathbf{x}_{I,r})}{2\sigma_I} \tag{3.22}$$

Equation 3.22 is the gradient of the functional 3.10. On the other hand, the likelihood is the ℓ^2-norm of a complex-valued linear model and the prior can be rewritten in a compact form using the notation of the complex Gaussian distribution.
The iterative algorithm is given by

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha\nabla U(\mathbf{x}_i). \tag{3.23}$$

where alpha is chosen experimentally, with the compromise that a low value increases the time of convergence, while a high value reduces the precision of the estimation. Methods of optimization of 3.23 exist, e.g. the orthogonality principle adopted in conjugate gradient techniques. It allows to select the optimal parameter α and the optimal minimum search direction for the gradient $\nabla U(\mathbf{x})$. The described techniques work only in case of regular functions which means quadratic functionals with global minima. In order to minimize a functional with local minima, other techniques have to be adopted, e.g. simulated annealing.
The pseudo code of the steepest descent algorithm used for the MAP estimation is shown

Input: Noisy Data y, Parameter θ
Output: Regularized image x

$\mathrm{x}_{old} \leftarrow \mathrm{y}$
$\alpha \leftarrow 10^{-5}$
$k \leftarrow$ *Desired number of iterations*
while $k \neq 0$ **do**
  $\nabla U \leftarrow$ `GradientCalculation` $(\mathrm{x}_{old}, \theta)$
  $\mathrm{x}_{new} \leftarrow \mathrm{x}_{old} + \alpha \cdot \nabla U$
  $\mathrm{x}_{old} \leftarrow \mathrm{x}_{new}$
  $k \leftarrow k - 1$
end
$\mathrm{x} \leftarrow \mathrm{x}_{new}$

Algorithm 2: Steepest descent algorithm for MAP estimation. The function `GradientCalculation()` computes the gradient according to Equation 3.22. The convergence condition is fixed by the desired number of iterations. The value of α has to be chosen according to the values of the input data.

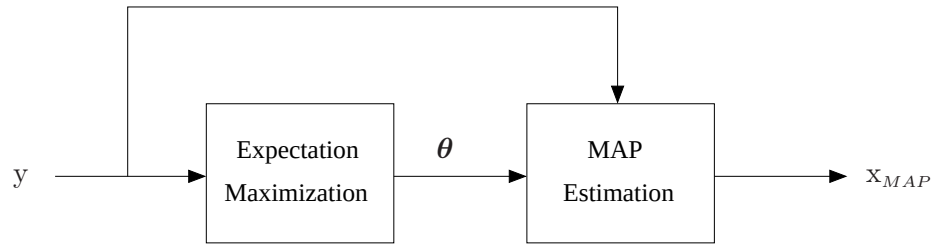


FIGURE 2 – Flowchart of the MAP estimate with Expectation Maximization algorithm.

in Algorithm 2. The described MAP method for statistical regularization adopts a different strategy than MBD. The latter exploits the two levels of Bayesian inference and performs the model selection by evidence maximization at each step of MAP minimization (model fitting). The former method *first* estimates the parameters from incomplete data by the EM and *then* performs the MAP estimation without updating the parameters which are fixed.

3.3 Complex Gauss-Markov Random Field

The complex-valued Gauss-Markov Random Field (GMRF) is an extension of the real one. In the next section we shortly present some considerations on complex random variables and the normal distribution in the general case of improper complex random variables. Thus the complex-valued GMRF is presented.

3.3.1 Normal Distribution of Improper Complex Random Variables

A complex random process $\mathbf{X}$ whose realization is the complex-valued random variable $\mathrm{x} = \mathrm{x}_R + j\mathrm{x}_I$ can always be represented through a pair of real vectors. In most cases, the theory of complex random variables is supposed to be similar to the real one,

as long as the transpose operator is replaced with the Hermitian operator in the definition of the covariance matrix. Most of the time this is justified, but it can happen that x and its conjugate x^* are correlated. Then, the covariance matrix $\mathbf{C}_\mathrm{X}$ no longer completely describes the second-order behavior of x and another quantity, which is known as the complementary covariance or relation function $\mathbf{R}_\mathrm{X}$ has to be taken into account. $\mathbf{C}_\mathrm{X}$ and $\mathbf{R}_\mathrm{X}$ are defined as

$$\mathbf{C}_\mathrm{X} = \mathrm{E}\{\mathbf{x}\mathbf{x}^H\} \tag{3.24}$$
$$\mathbf{R}_\mathrm{X} = \mathrm{E}\{\mathbf{x}\mathbf{x}^T\} \tag{3.25}$$

Vectors that have a vanishing relation function $\mathbf{R}_\mathrm{X}$ are called *proper* (Neeser & Massey, 1993). Proper complex random variables behave very similarly to real random variables and the Gaussian pdf takes the familiar form.
Let $\mathrm{x}_i = \mathrm{x}_{R,i} + j\mathrm{x}_{I,i}$ with $i = 1, \ldots, N$ be the realizations of a complex stochastic process $\mathbf{x}$ with $\mathrm{x}_{R,i}, \mathrm{x}_{I,i} \in \mathbb{R}$ and normally distributed. The joint Gaussian pdf is

$$p(\mathbf{x}) = \frac{1}{\pi^N \det(\mathbf{C}_\mathrm{X})} \exp\{-\mathbf{x}^H \mathbf{C}_\mathrm{X}^{-1} \mathbf{x}\}. \tag{3.26}$$

Considering the real vector $\mathbf{v} \in \mathbb{R}^{2n}$

$$\mathbf{v} = [\mathbf{x}_R^T, \mathbf{x}_I^T]^T = [\mathrm{x}_{R,1}, \ldots, \mathrm{x}_{R,N}, \mathrm{x}_{I,1}, \ldots, \mathrm{x}_{I,N}]^T \tag{3.27}$$

it allows the representation equivalent to 3.26 by a real Gaussian pdf with covariance matrix C_V

$$p(\mathbf{v}) = \frac{1}{(2\pi)^N \det(\mathbf{C}_\mathrm{V})^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{v}^T \mathbf{C}_\mathrm{V}^{-1} \mathbf{v}\right\}. \tag{3.28}$$

Both 3.26 and 3.28 include only the information embedded in the covariance matrix 3.24 but not the information in the 3.25 : they are valid only in case of *proper* complex random processes (Picinbono, 1996).
In contrast to *proper* random variables, the theory of *improper* random variables is quite different from what we are used to and requires special attention.
We want to find a way to exploit the information of both 3.26 and 3.28 and to include them in the pdf. Thus, we consider the following linear transformation (van den Bos, 1995; Picinbono, 1996)

$$\mathbf{w} = \mathbf{M}\mathbf{v} \tag{3.29}$$

where $\mathbf{M}$ has the following form

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}. \tag{3.30}$$

Equation 3.29 transforms the vector 3.27 in

$$\mathbf{w} = [\mathbf{x}^T, \mathbf{x}^H]^T = [\mathrm{x}_1, \ldots, \mathrm{x}_N, \mathrm{x}_1^*, \ldots, \mathrm{x}_N^*]^T. \tag{3.31}$$

The pdf of the random variable $\mathbf{w}$ is

$$p(\mathbf{w}) = \frac{1}{\pi^N \det(\mathbf{C}_\mathrm{W})^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{w}^H \mathbf{C}_\mathrm{W}^{-1} \mathbf{w}\right\}. \tag{3.32}$$

where the matrix $\mathbf{C}_{\mathrm{W}}$ has the following structure

$$\mathbf{C}_{\mathrm{W}} = \begin{bmatrix} \mathbf{C}_{\mathrm{X}} & \mathbf{R}_{\mathrm{X}} \\ \mathbf{R}_{\mathrm{X}}^{H} & \mathbf{C}_{\mathrm{X}}^{*} \end{bmatrix}. \tag{3.33}$$

Thus, 3.32 can be rewritten as

$$p(\mathbf{x}, \mathbf{x}^{*}) = \frac{1}{\pi^{N}[\det(\mathbf{C}_{\mathrm{X}})\det(\mathbf{P}_{\mathrm{X}})]^{1/2}} \exp\left\{ -\frac{1}{2}[\mathbf{x}^{H}, \mathbf{x}^{T}] \begin{bmatrix} \mathbf{C}_{\mathrm{X}} & \mathbf{R}_{\mathrm{X}} \\ \mathbf{R}_{\mathrm{X}}^{H} & \mathbf{C}_{\mathrm{X}}^{*} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^{*} \end{bmatrix} \right\}. \tag{3.34}$$

where $\mathbf{P}_{\mathrm{X}} = \mathbf{C}_{\mathrm{X}}^{*} - \mathbf{R}_{\mathrm{X}}^{H}\mathbf{C}_{\mathrm{X}}^{-1}\mathbf{R}_{\mathrm{X}}$. Thus, considering $p(\mathbf{x}, \mathbf{x}^{*}) = p'(\mathbf{x}_R, \mathbf{x}_I)$, 3.34 embeds the information of the covariance matrix 3.24 and of the relation function 3.25 and it is the general pdf of an *improper* normal distributed complex random process. If the relation function 3.25 vanishes, i.e. $R_{\mathrm{X}} = 0$, Equation 3.34 becomes equivalent to Equation 3.26.

3.3.2 Complex GMRF model

The complex-valued GMRF model is defined as an extension of the classical real-valued GMRF presented in 2.4.1 (Chelappa et al., 1985). The potential function is written as

$$U(\mathrm{x}_s|\mathrm{x}_r : r \in \mathcal{N}_s, \boldsymbol{\theta}) = -\frac{|\mathrm{x}_s - \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r})|^2}{\sigma^2} \tag{3.35}$$

where $\sigma \in \mathbb{R}$, $\mathrm{x}, \theta \in \mathbb{C}$, with $\mathrm{x} = \mathrm{x}_R + j\mathrm{x}_I$ and $\theta = \theta_R + j\theta_I$. The conditional distribution associated to the model is

$$p(\mathrm{x}_s|\mathrm{x}_r, r \in \mathcal{N}, \boldsymbol{\theta}) = \frac{1}{\pi\sigma^2}\exp\left\{ -\frac{|\mathrm{x}_s - \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r})|^2}{\sigma^2} \right\} \tag{3.36}$$

The associated complex AR process has the following form Picinbono & Bouvet (1984)

$$\mathrm{x}_s = \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r} + \mathrm{x}_{s-r}) + \mathrm{e}_s \tag{3.37}$$

where input of the complex AR filter is a complex white generative process e_s and also the coefficients θ_r are complex. Because of the isomorphism between $\mathbb{C}$ and $\mathbb{R}^2$, Equation 3.37 can be rewritten in the following vectorial form

$$\begin{bmatrix} \mathrm{x}_{R,s} \\ \mathrm{x}_{I,s} \end{bmatrix} = \frac{1}{2}\sum_{r\in\mathcal{N}} \begin{bmatrix} \theta_{R,r} & -\theta_{I,r} \\ \theta_{I,r} & \theta_{R,r} \end{bmatrix} \begin{bmatrix} \mathrm{x}_{R,s+r} + \mathrm{x}_{R,s-r} \\ \mathrm{x}_{I,s+r} + \mathrm{x}_{I,s-r} \end{bmatrix} + \begin{bmatrix} \mathrm{e}_{R,s} \\ \mathrm{e}_{I,s} \end{bmatrix} \tag{3.38}$$

The structure of the parameter matrix is given by the real representation of complex random variables. If it is not respected, the output is not an AR complex signal. Moreover, the equivalence between Equations 3.37 and 3.38 is ensured only if the noise e_s has a circular complex normal Gaussian distribution, therefore it is a proper random process. The conditions for the noise are

$$\mathrm{E}\{\mathrm{e}_s\mathrm{e}_{s+r}^{*}\} = \begin{cases} -2\theta_r^{*}\sigma^2 & \text{if } r \in \mathcal{N} \\ 2\sigma^2 & \text{if } r = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.39}$$

where $\mathrm{E}\{\mathrm{e}_{I,s}\mathrm{e}_{R,s+r}\} = -\mathrm{E}\{\mathrm{e}_{R,s}\mathrm{e}_{I,s+r}\} = \theta_{I,r}\sigma^2$ which ensures the structure of the parameter matrix in 3.38.

3.3.3 Proper and Improper White Complex Gaussian Noise

The hypothesis on the noise e_s in 3.37 is not necessary to be so restrictive and it can be also an improper random process with white spectrum, i.e. unitary covariance matrix, but with a non-zero relation function. In the latter case, the model 3.37 cannot be written as 3.38 and the noise in 3.38 does not correspond to the noise in 3.37. In case that the noise in 3.37 is an improper random process, we can simulate it by prefiltering a circular Gaussian white noise.
In case that the noise is not completely white, i.e. improper random process, it can be still predicted, which means that e_s is not the completely unpredictable part of x_s. As a consequence of this fact we can conclude that the parameter vector $\hat{\boldsymbol{\theta}}$ is not the best linear predictor of x, because the prediction error can still be reduced by using another filter.
A generalization of the AR model 3.37 in case of improper Gaussian noise takes the form of a widely linear model (Picinbono & Bondon, 1997)

$$\mathrm{x}_s = \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r(\mathrm{x}_{s+r}+\mathrm{x}_{s-r}) + \frac{1}{2}\sum_{r\in\mathcal{N}_s}\kappa_r(\mathrm{x}_{s+r}+\mathrm{x}_{s-r})^* \tag{3.40}$$

which in general is non-linear

$$f(\mathrm{x}_1+\mathrm{x}_2) = f(\mathrm{x}_1)+f(\mathrm{x}_2)f(\lambda\mathrm{x}) \neq \lambda f(\mathrm{x})$$

where the first expression is respected, but the second, which is a necessary property of classical linear filters, is in general not respected for $\lambda \in \mathbb{C}$. Actually, for a complex function $f(\mathrm{x}) = \mathrm{x}^*$, which returns the conjugate of the complex variable, and a complex number $\lambda \in \mathbb{C}$, it is $\lambda\mathrm{x}^* \neq (\lambda\mathrm{x})^*$, which demonstrates the validity of the second statement in 3.41 and thus the non-linearity.
On the other hand, the widely linear system 3.40 is demonstrated to be more effective for processing complex-valued signals than the results of a classical approach (Picinbono & Chevalier, 1995, 1996). A Widely Linear Mean Square Estimation (WLMSE) results in an error which is inferior than a classical Linear Mean Square Estimation (LMSE) (Picinbono & Chevalier, 1995).
The fully developed speckle is a proper complex Gaussian process.

3.3.4 Synthesis and Analysis

Similarly to the real case discussed in Chelappa et al. (1985), we follow the same approach for parameter retrieval and forward modeling.
Equation 3.37 corresponds to an Infinitive Impulse Response (IIR) filter which can be implemented in the frequency domain exploiting the fact that the θ parameters characterize the autocorrelation function. The Wiener-Khintchine theorem allows to compute the power spectrum from the autocorrelation function by the Fourier transfom, which can be approximated by the DFT and computed with the FFT algorithm.
The autocorrelation function of Equation 3.37 has the following form

$$R_{\mathrm{X}}(k) = \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r R_{\mathrm{X}}(r+k) + \sigma^2\delta(k) \tag{3.41}$$

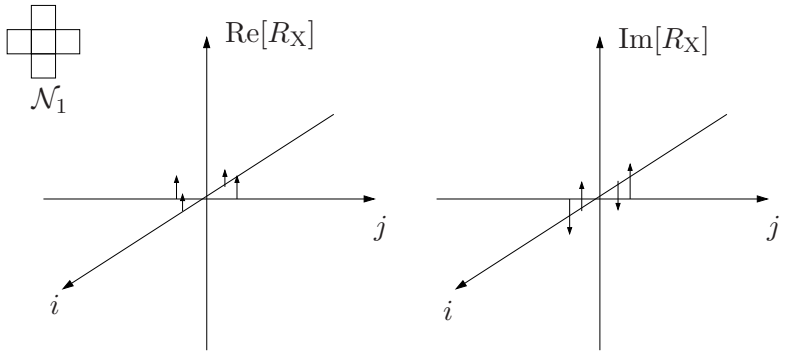


FIGURE 3 – Example of auto-correlation function for model order 1. The real and the imaginary part are even and odd functions respectively.

the Fourier transform of 3.41 is the power spectrum, which has the form

$$\Phi_{\mathrm{X}}(\omega_1,\omega_2) = \frac{\sigma^2}{\mathrm{Re}\left[1 - \frac{1}{2}\sum_{r\in\mathcal{N}_s}\theta_r \exp\left\{j\left(\frac{2\pi}{M}r_1\omega_1 + \frac{2\pi}{N}r_2\omega_2\right)\right\}\right]} \tag{3.42}$$

where ω_1 and ω_2 are the frequencies in the 2-D lattice. The $\mathrm{Re}[\cdot]$ fixes the relationship of the power spectrum 3.42 with the AR function. The auto-correlation 3.41 is a complex-valued function and it has to respect the Hermitian condition $R_{\mathrm{X}}(k)^* = R_{\mathrm{X}}(-k)$, which is fulfilled when the real part is even (symmetric) and the imaginary part is odd (anti-symmetric). An example of an auto-correlation function, for model order 1, is shown in Figure 3. Considering a generic Hermitian function $\mathrm{x}(n) = \mathrm{x}^*(-n)$, the Fourier transform is $\mathrm{X}(\mathrm{e}^{i\omega}) = \mathrm{X}^*(\mathrm{e}^{i\omega})$, but the Fourier transform can only be equal if and only if the imaginary part is zero. Thus, the Fourier transform of an Hermitian signal is a real-valued signal. Furthermore the condition for the power spectrum Equation 3.42 $\Phi_{\mathrm{X}}(\omega) \geq 0$ is guaranteed by the values of the parameter vector. The sum of the parameters has to respect the condition given by the property of the Markov chains : the sum of each line of the transition matrix has to be equal to one (Lakshmanan & Derin, 1993). In Figure 4, the block diagram for the synthesis of GMRF realization, for a given driven circular complex Gaussian noise e and auto-correlation function 3.41, is shown.

An alternative implementation in time domain can be obtained through the Gibbs sampler (Walsh, 2004), introduced in the context of image processing by Geman & Geman (1984b), which is a special case of Metropolis-Hastings sampling. The latter generates a Markov chain from the conditional distribution $p(\mathrm{x}|\mathrm{y})$ whose equilibrium density is the candidate joint distribution $p(\mathrm{x},\mathrm{y})$. A pseudo code of the Gibbs sampler for the synthesis of complex valued GMRF is shown in Algorithm 3. The number of iterations has to be enough to reach the equilibrium density and the support $M \times M$ has to be large enough, with respect to the neighborhood, in order to avoid an undesired border effect : the pixels at the boundary do not belong to a complete neighborhood.
An example of a stationary zero mean complex-value GMRF is shown in Figure 5, where realizations with different parameter values for model order one and two are visible.

The estimation of the parameters is computed considering the linear model 3.37 asso-

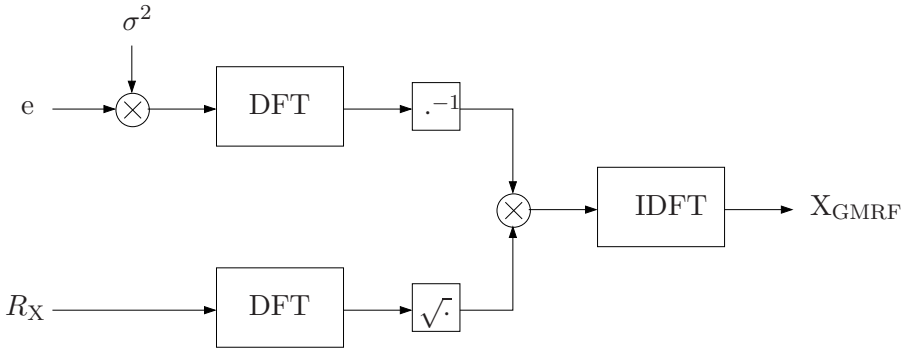


FIGURE 4 – Synthesis of GMRF : R_X autocorrelation function 3.41, e driven noise with power σ^2.

Input: Parameter neighborhood h, Power σ^2, Image size $N \times N$
Output: Complex valued $GMRF$

$t0 \leftarrow$ Complex Array of $N \times N$ elements
$tm \leftarrow t0$
$k \leftarrow 10000$
while $k \neq 0$ **do**
  $\mu \leftarrow$ `2D-Convolution` $(t0, h)$
  for $i \leftarrow 1$ **to** M **do**
    for $j \leftarrow 1$ **to** M **do**
      $tm[i,j] \leftarrow$ `Complex` $(\mathcal{N}(0,1), \mathcal{N}(0,1))$+$\mu[i,j]$
    end
  end
  $t0 \leftarrow tm$
  $k \leftarrow k - 1$
end
$GMRF \leftarrow \sigma \cdot tm$

Algorithm 3: Gibbs sampler for complex GMRF generation.

Texture	Model Order	(1,0)		(0,1)		(1,1)		(1,-1)	
		θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	θ_3	$\hat{\theta}_3$	θ_4	$\hat{\theta}_4$
T1	1	0.18-j0.0	0.25-j0.0	0.28-j0.4	0.35-j0.05	-		-	
T2	1	0.24-j0.0	0.36-j0.1	0.0-j0.25	0.01-j0.37	-		-	
T3	1	0.08-j0.34	0.10-j0.42	0.0-j0.08	0.0-j0.10	-		-	
T4	2	0.0-j0.0	0.01-j0.03	0.07-j0.08	0.08-j0.1	0.0-j0.0	-0.01-j0.02	0.25-j0.1	0.28-j0.12
T5	2	0.0-j0.0	0.0-j0.01	0.0-j0.07	0.0-j0.12	0.38-j0.05	0.48-j0.07	0.0-j0.0	0.0-j0.0

TABLE 1 – Parameters θ corresponding to textures shown in Fig. 5 and estimated parameters $\hat{\theta}$. The textures T1... T5 correspond to row 1... 5 in Fig.5, respectively. The notation $(\cdot, \cdot)$ refers to the coordinate of the parameter with respect to the central pixel.

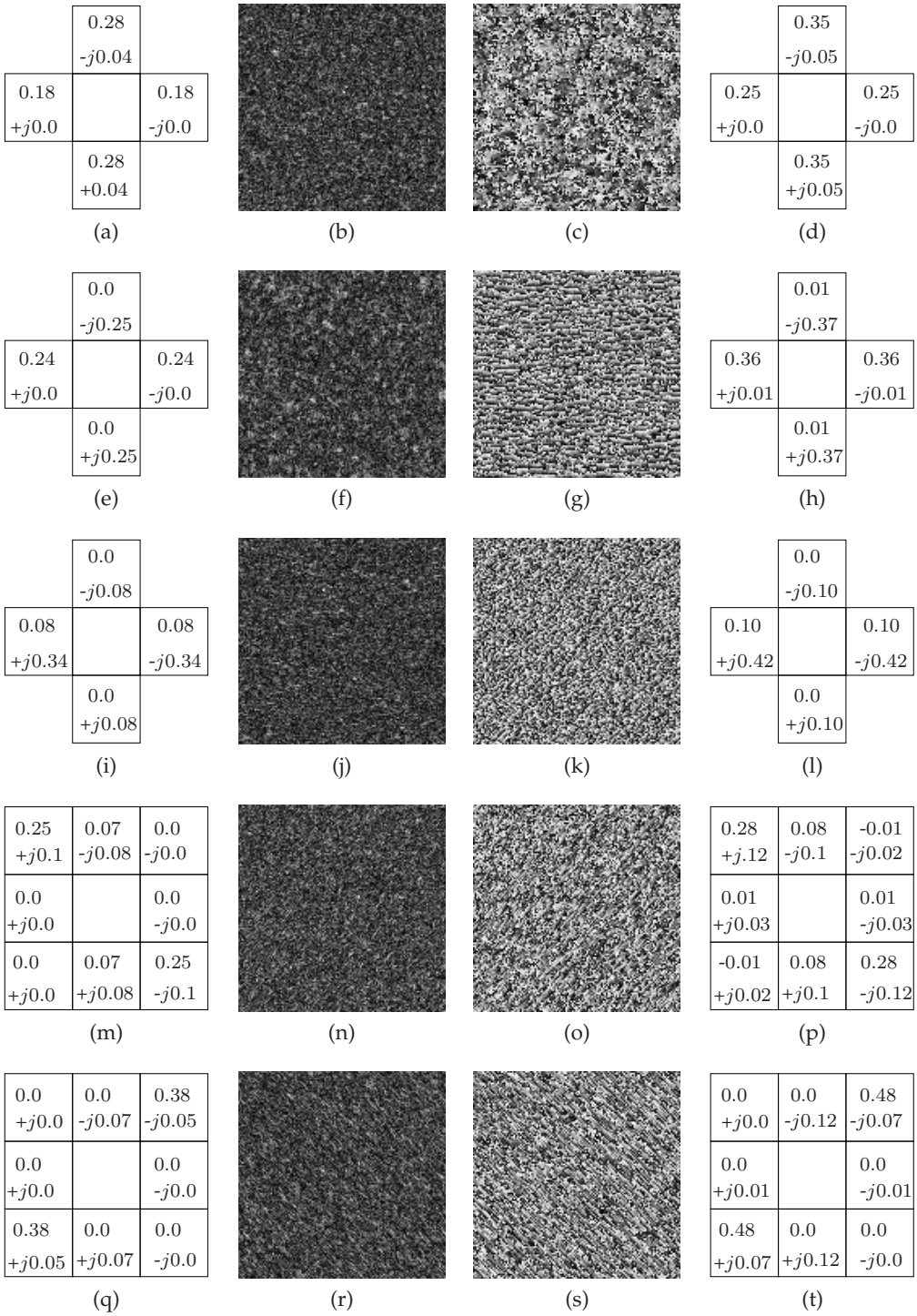


FIGURE 5 – Simulation of stationary zero mean complex-valued GMRF processes : model order and parameters (first column), amplitude (second column), phase (third column) and estimated parameters (fourth column).

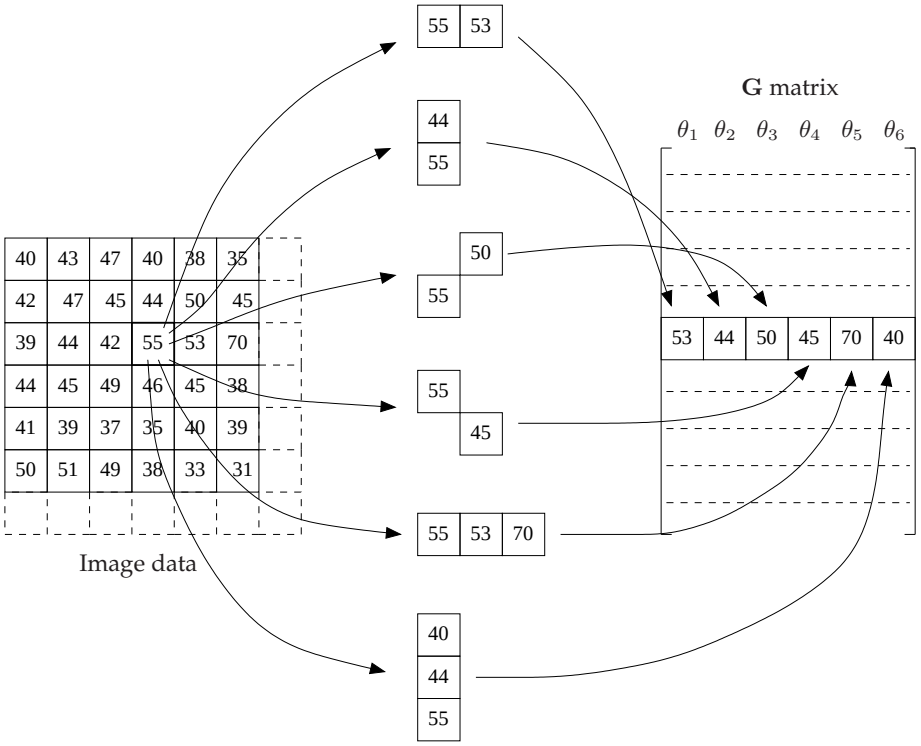


FIGURE 6 – Example of **G** matrix for model order 3.

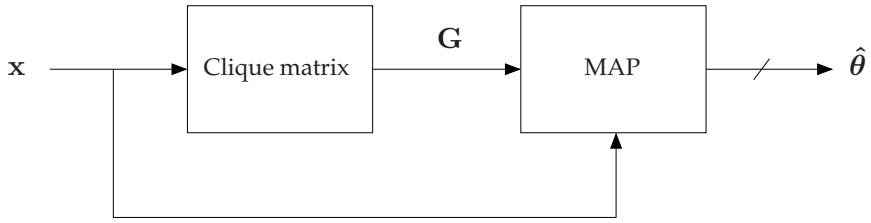


FIGURE 7 – Parameter estimation block diagram.

ciated to the joint pdf which, exploiting the matrix formalism, can be written as follows

$$\mathbf{x} = \mathbf{G}\boldsymbol{\theta} + \mathbf{e} \tag{3.43}$$

where $\mathbf{G}$ is the matrix of the cliques, $\boldsymbol{\theta}$ is the parameter vector, $\mathbf{x}$ is the original image and $\mathbf{e}$ is the circular complex Gaussian noise with variance σ^2.
We perform local computation in an analyzing window of $N \times N$ pixels. On the other hand, we take into account only the pixels belonging to a complete Neighborhood, thus, the model samples are $P < N^2$ because of the border effect. If Q is the cardinality of the parameter vector $\boldsymbol{\theta}$, $\mathbf{G}$ is a $P \times Q$ matrix. The task is to build the $\mathbf{G}$ matrix (Datcu et al., 2004) which is a function of the image pixels : $\mathbf{G} = \mathbf{G}(\mathbf{x})$. An example on how to build the $\mathbf{G}$ matrix is shown in Figure 6. The MAP estimation of the model parameters is equivalent to the LSE estimator because the Gaussian pdf is symmetric and the prior is uniform and equation 3.43 results in an over-determined linear problem. The parameters are estimated by

$$\hat{\boldsymbol{\theta}}_{LSE} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{x} \tag{3.44}$$

where $\hat{\boldsymbol{\theta}}_{LSE}$ represents the estimated parameter vector and $(\mathbf{G}^H\mathbf{G})^{-1}$ is the pseudo-inverse matrix.
The variance of the model is computed as follows

$$\sigma^2 = \frac{1}{P}\sum_{s=1}^{P}|\mathbf{x}_s - \mathbf{G}_s\hat{\boldsymbol{\theta}}|^2 \tag{3.45}$$

where the quantity $\mathbf{G}\hat{\boldsymbol{\theta}}$ represents the best fit of the data. In addition the evidence of the model can be computed according to 2.57 (Ruanaidh & Fitzgerald, 1996), in order to compute model selection.

3.4 Tikhonov regularization for complex image restoration

Tikhonov regularization is a powerful method to restore images affected by degradation processes. In particular, while computing the imaging system response, e.g. blurring, from a given image is a direct and well-posed problem, the inverse problem of computing the true image from the observations is an ill-posed problem. A general principle for dealing with the instability of the inverse problem is that of regularization, which mainly consists in restricting the set of admissible solutions including some *a priori* information (e.g. non negativity, smoothness, existence of edges, etc.) in the formulation of the problem. In this framework we focus on Tikhonov regularization which is a direct regularization method. The formulation of the problem in its unconstrained form is the following (Chan & Mulet, 1995)

$$\min_{\mathbf{x}} f(\mathbf{x}) \equiv \frac{1}{2}||\mathbf{S}\mathbf{x} - \mathbf{y}||_2^2 + \alpha R(\mathbf{x}) \tag{3.46}$$

where α controls the trade off between a good fit to the data and an irregular solution. The term $R(\mathbf{x})$ is the regularization functional. It is often in the quadratic form $R(\mathbf{x}) = ||\mathbf{T}\mathbf{x}||_2^2$ where $\mathbf{T} = \mathbf{I}$ or $\mathbf{T} = \nabla$, the identity matrix or the differential operator, respectively. The latter regularization functional proposed by Rudin et al. (1992) is the Total Variation (TV), which is defined as

$$\mathrm{TV(u)} = \int_\Omega |\nabla \mathrm{u}| = \int_\Omega \sqrt{\mathrm{u}_x^2 + \mathrm{u}_y^2}\ \mathrm{d}x\mathrm{d}y \tag{3.47}$$

and adapts better to discontinuities. Intuitively Equation 3.47 gives a global measure of the variation of the function with respect to its derivative. The Tikhonov regularization with a TV term is equivalent to the Anisotropic Diffusion (AD) (Chan & Mulet, 1995; Perona & Malik, 1990; Snyder et al., 1995). The AD is demonstrated to be equivalent to the Lee (1980) and Frost et al. (1982) adaptive filters in Yu & Acton (2002), where a method for speckle reduction using AD is proposed.
On the other hand, Tikhonov regularization can be expressed in the Bayesian framework

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{3.48}$$

where the likelihood

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left\{-\frac{1}{2}||\mathbf{S}\mathbf{x} - \mathbf{y}||_2^2\right\} \tag{3.49}$$

and the prior

$$p(\mathbf{x}) \propto \exp\{-\alpha R(\mathbf{x})\} \tag{3.50}$$

are proportional to the first and the second term of 3.46. Thus, the optimization of the functional as the MAP estimate of the posterior is equivalent to the Bayesian approach. With these premises the Tikhonov regularization has the advantages of the Bayesian method, e.g. different prior models can be compared (Bouman & Sauer, 1993).

3.4.1 Image Restoration with Huber-Markov prior

The Bayesian frame opens the possibility to use a combination of prior models as done in Çetin & Karl (2001) where two functionals are adopted with the identity matrix and with the differential operator giving a Tikhonov regularization of the form

$$J(\mathbf{x}) = ||\mathbf{y} - \mathbf{S}\mathbf{x}||^2 + \lambda_1^2||\mathbf{x}||^k + \lambda_2^2||\nabla|\mathbf{x}|||^k \tag{3.51}$$

where $\mathbf{y}$ is the noisy data vector, $\mathbf{x}$ is the true image, $\mathbf{S}$ is the SAR system transformation matrix, λ_1, λ_2 are regularization parameters, $|\mathbf{x}|$ is the magnitude of the complex vector $\mathbf{x}$, ∇ is a discrete approximation to the gradient and $||\cdot||^k$ denotes the ℓ^k-norm.
The second term $||\mathbf{x}||$ is proposed for the enhancement of the point-based feature, while the third term $||\nabla|\mathbf{x}|||$ is for the enhancement of the region-based feature (Çetin & Karl, 2001). They both represent a prior information of the image model. The parameter k represents the shape parameter of Generalized Gaussian pdf. If $k = 2$ the Gaussian prior is considered, if $k = 1$ the Laplacian prior is considered.
I extend the model proposed by Çetin & Karl (2001) replacing the first term with the following model

$$P(\mathrm{x}) = \frac{1}{Z}\exp\left\{ -\frac{1}{\tau}\sum_{c\in\mathcal{C}} \rho_T\big(\mathbf{d}_c^t(\mathrm{x})\big)\right\} \tag{3.52}$$

called Huber-Markov and belonging to the Markov Random Field family, where $\rho(\cdot)$ is the Huber function, $\mathbf{d}_c(\mathrm{x})$ is a coefficient vector for the clique c, τ is the *temperature*, and the partition function Z is a normalization constant. The superscript notation $(\cdot)^t$ in (3.52) represents the transpose operator. The Huber function ρ_T is given by

$$\rho_T(\mathrm{x}) = \begin{cases} \mathrm{x}^2 & |\mathrm{x}| \le T \\ T^2 + 2T(|\mathrm{x}| - T) & |\mathrm{x}| > T \end{cases} \tag{3.53}$$

where T is a threshold which defines the quadratic or linear behavior of the function. The Huber function is shown in Figure 8
Thus, the Huber-Markov Random Field (HMRF) kernel in (3.52) becomes

$$\Omega(\mathbf{x}, T) = \frac{1}{\tau}\sum_{i}^{N}\sum_{j}^{N}\sum_{m=0}^{3}\rho_T(d_{i,j,m}) \tag{3.54}$$

The coefficient $d_{i,j,m}$ belongs to the vector $\mathbf{d}_c^t$, which is an approximate rotationally symmetric operator within a 3×3 grid as defined in Brady & Horn (1983). The vector $\mathbf{x}$ corresponds to the image pixels that are inside a window with a size of $N \times N$ pixels. The image roughness, measured at pixel $\mathrm{x}_{i,j}$, is defined using second order derivatives, which

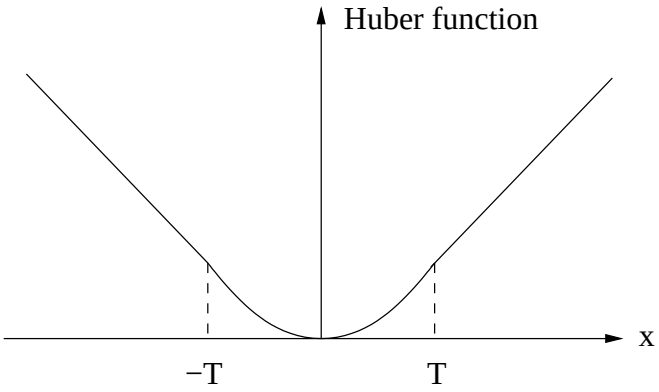


FIGURE 8 – Huber function.

are given by

$$\begin{aligned} d_{i,j,0} &= \mathrm{x}_{i,j+1} - 2\mathrm{x}_{i,j} + \mathrm{x}_{i,j-1} \\ d_{i,j,1} &= \frac{1}{2}(\mathrm{x}_{i-1,j+1} - 2\mathrm{x}_{i,j} + \mathrm{x}_{i+1,j-1}) \\ d_{i,j,2} &= \mathrm{x}_{i-1,j} - 2\mathrm{x}_{i,j} + \mathrm{x}_{i+1,j} \\ d_{i,j,3} &= \frac{1}{2}(\mathrm{x}_{i-1,j-1} - 2\mathrm{x}_{i,j} + \mathrm{x}_{i+1,j+1}) \end{aligned} \tag{3.55}$$

The functions in (3.55) define a kernel proportional to the Laplacian but they are squared and summed at each pixel location according to (3.54).

3.4.2 Restoration algorithm

The cost function using non-quadratic regularization and HMRF is given by

$$\begin{aligned} J(\mathbf{x}) &= \|\mathbf{y} - \mathbf{S}\mathbf{x}\|^2 \\ &\quad +\lambda_1^2 \frac{1}{\tau} \sum_i^N \sum_j^N \left(\Big| \sum_{m=0}^{3} \rho_T(d_{i,j,m}) \Big| + \epsilon \right)^{k/2} \\ &\quad +\lambda_2^2 \sum_i^{N\times N} (|(\nabla|\mathbf{x}|)_i|^2 + \epsilon)^{k/2} \end{aligned} \tag{3.56}$$

where $\epsilon \geq 0$ is a small constant. The variable ϵ is introduced in order to avoid numerical problems (e.g. division by 0). The minimization of expression (3.56) is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} J(\mathbf{x}) \tag{3.57}$$

which gives the solution for the noise free image.

Many different solutions for the cost function, as defined in (3.56), are proposed in Pan & Reeves (2006) and Çetin & Karl (2001). The Hessian update scheme is the most appropriate solution, because it operates using the complex-valued data, as proposed in Çetin & Karl (2001). The solution of (3.57) can be computed iteratively using Hessian

approximation $\mathbf{E}(\mathbf{x})$

$$\begin{aligned}\nabla J(\mathbf{x}) &= -2\mathbf{S}^H\mathbf{y} + 2\mathbf{S}^H\mathbf{S}\mathbf{x} \\ &\quad -k\lambda_1^2\frac{1}{\tau}\mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x},T))\mathbf{\Lambda}_3(\mathbf{d}_c^t(\mathbf{x})) \\ &\quad +k\lambda_2^2\mathbf{\Phi}^H(\mathbf{x})\nabla^t\mathbf{\Lambda}_2(\mathbf{x})\nabla\mathbf{\Phi}(\mathbf{x})\mathbf{x} \end{aligned} \tag{3.58}$$

where

$$\begin{aligned}\mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x},T)) &= \text{diag}\left\{\frac{1}{(|\mathbf{\Omega}_i(\mathbf{x},T)|+\epsilon)^{1-k/2}}\right\} \\ \mathbf{\Lambda}_2(\mathbf{x}) &= \text{diag}\left\{\frac{1}{(|(\nabla|\mathbf{x}|)_i|^2+\epsilon)^{1-k/2}}\right\} \\ \mathbf{\Lambda}_3(\mathbf{d}_c^t(\mathbf{x})) &= \text{diag}\left\{2d_{i,j,0}+d_{i,j,1}+2d_{i,j,2}+d_{i,j,3}\right\} \qquad (3.59) \\ \mathbf{\Phi}(\mathbf{x}) &= \text{diag}\left\{\exp(-j\phi(\mathbf{x}_i))\right\} \qquad (3.60)\end{aligned}$$

$$\begin{aligned}\mathbf{E}(\mathbf{x}) &= 2\mathbf{S}^H\mathbf{S} - k\lambda_1^2\frac{10}{\tau}\mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x},T)) \\ &\quad +k\lambda_2^2\mathbf{\Phi}^H(\mathbf{x})\nabla^t\mathbf{\Lambda_2}(\mathbf{x})\nabla\mathbf{\Phi}(\mathbf{x}) \end{aligned} \tag{3.61}$$

where $\mathbf{\Omega}(\mathbf{x},T)$ represents the Huber-Markov Random Field, $(\cdot)^H$ the Hermitian of the matrix, $(\cdot)^t$ the transpose operator, $\phi(\mathbf{x}_i)$ denotes the phase of the complex number $\mathbf{x}_i$, and diag$(\cdot)$ denotes the diagonal matrix, where the i-th element is given by the expression inside the brackets.

The iterative procedure using Hessian approximation $\mathbf{E}(\mathbf{x})$ and the gradient $\nabla J(\mathbf{x})$ is given by

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} - \gamma\left[\mathbf{E}\big(\hat{\mathbf{x}}^{(n)}\big)\right]^{-1}\nabla J(\hat{\mathbf{x}}^{(n)}) \tag{3.62}$$

where n denotes the current iteration, γ is the step size, and $\nabla J(\hat{\mathbf{x}}^{(n)})$ is the gradient. The convergence constrain is $\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2/\|\mathbf{x}^{(n)}\|^2 < \delta$ where $\delta > 0$ is a constant.

3.4.3 Parameter Estimation

Parameters λ_1, λ_2 and k can be estimated using the evidence framework as first proposed in MacKay (1999) and recently used in Li et al. (2007). The posterior for λ_1 is given by

$$p(\lambda_1|\mathbf{y},\lambda_2,k) \propto p(\mathbf{y}|\lambda_1,\lambda_2,k)p(\lambda_1) \tag{3.63}$$

where $p(\mathbf{y}|\lambda_1,\lambda_2,k)$ is called evidence and $p(\lambda_1)$ is the prior. The evidence is given by marginalization

$$p(\mathbf{y}|\lambda_1,\lambda_2,k) = \int p(\mathbf{y}|\mathbf{x},\lambda_1,\lambda_2,k)p(\mathbf{x}|\lambda_1,\lambda_2,k)d\mathbf{x} \tag{3.64}$$

On the other hand, using the Laplace method, the integral can be approximated around its maximum exploiting the Taylor expansion

$$p(\mathbf{y}|\lambda_1,\lambda_2,k) \simeq p(\mathbf{y}|\mathbf{x},\lambda_1,\lambda_2,k)p(\mathbf{x}|\lambda_1,\lambda_2,k)\sqrt{\frac{(2\pi)^{N^2}}{|\mathbf{E}|}} \tag{3.65}$$

where

$$p(\mathbf{y}|\mathbf{x},\lambda_1,\lambda_2,k)p(\mathbf{x}|\lambda_1,\lambda_2,k) = \frac{1}{Z(\lambda_1)Z(\lambda_2)}\exp\left(-\frac{1}{2}M(\mathbf{x},\mathbf{y}|\lambda_1,\lambda_2)\right) \tag{3.66}$$

where

$$\begin{aligned} M(\mathbf{x},\mathbf{y}|\lambda_1,\lambda_2,k) &= ||\mathbf{y}-\mathbf{S}\mathbf{x}||^2 + \lambda_1^2 Q_1 + \lambda_2^2 Q_2 \\ Q_1 &= \frac{1}{\tau}\sum_i\sum_j\left(\Big|\sum_{m=0}^{3}\rho_T(d^t_{i,j,m})\Big| + \epsilon\right)^{k/2} \\ Q_2 &= \sum_i (|(\nabla|\mathbf{x}|)_i|^2+\epsilon)^{k/2} \\ Z(\lambda_1) &= (2\pi/\lambda_1)^{N^2/2} \\ Z(\lambda_2) &= (2\pi/\lambda_2)^{N^2/2} \end{aligned} \tag{3.67}$$

Thus, considering 3.66 and 3.65, we can approximate the log-evidence from 3.64 as follows

$$\begin{aligned} \ln p(\mathbf{y}|\lambda_1,\lambda_2,k) \simeq\ & -\frac{1}{2}\|\mathbf{y}-\mathbf{S}\mathbf{x}\|^2 \\ & -\frac{1}{2}\lambda_1^2 Q_1 - \frac{1}{2}\lambda_2^2 Q_2 \\ & +\frac{N^2}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{E}| \end{aligned} \tag{3.68}$$

Considering the Jeffrey's prior for the parameter λ_1, λ_2 and k, respectively

$$\begin{aligned} p(\lambda_1) &\propto 1/\lambda_1 \\ p(\lambda_2) &\propto 1/\lambda_2 \\ p(k) &\propto 1/k \end{aligned} \tag{3.69}$$

the posterior is obtained by multiplying 3.68 and each of the priors in 3.69. The maximization of the posterior leads to

$$\frac{\partial}{\partial\lambda_1}\ln p(\lambda_1|\mathbf{y}) \simeq -\lambda_1 Q_1 - \frac{1}{2}\frac{\partial}{\partial\lambda_1}\ln|\mathbf{E}| - \frac{1}{\lambda_1} \tag{3.70}$$

$$\frac{\partial}{\partial\lambda_2}\ln p(\lambda_2|\mathbf{y}) \simeq -\lambda_2 Q_2 - \frac{1}{2}\frac{\partial}{\partial\lambda_2}\ln|\mathbf{E}| - \frac{1}{\lambda_2} \tag{3.71}$$

$$\begin{aligned} \frac{\partial}{\partial k}\ln p(k|\mathbf{y}) \simeq\ & -\frac{1}{4}\lambda_1 Q_1 \ln\tilde{Q}_1 - \frac{1}{4}\lambda_2 Q_2 \ln\tilde{Q}_2 + \\ & -\frac{1}{2}\frac{\partial}{\partial k}\ln|\mathbf{E}| - \frac{1}{k} \end{aligned} \tag{3.72}$$

where the derivative of the Hessian with respect to the each parameter is

$$\frac{\partial}{\partial\lambda_1}\ln|\mathbf{E}| = \frac{-k\lambda_1\frac{20}{\tau}\mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x},T))}{\mathbf{E}(\mathbf{x})} \tag{3.73}$$

$$\frac{\partial}{\partial \lambda_2} \ln |\mathbf{E}| = \frac{2k\lambda_2 \mathbf{\Phi}^H(\mathbf{x}) \nabla^t \mathbf{\Lambda}_2(\mathbf{x}) \nabla \mathbf{\Phi}(\mathbf{x})}{\mathbf{E}(\mathbf{x})} \tag{3.74}$$

$$\begin{aligned} \frac{\partial}{\partial k} \ln |\mathbf{E}| \quad = \quad & \frac{1}{\mathbf{E}(\mathbf{x})} \Big(- \lambda_1^2 \frac{10}{\tau} \mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x}, T)) + \\ & -k\lambda_1^2 \frac{10}{\tau} \mathbf{\Lambda}_1'(\mathbf{\Omega}(\mathbf{x}, T)) + \\ & +\lambda_2^2 \mathbf{\Phi}^H(\mathbf{x}) \nabla^t \mathbf{\Lambda}_2(\mathbf{x}) \nabla \mathbf{\Phi}(\mathbf{x}) + \\ & +k\lambda_2^2 \mathbf{\Phi}^H(\mathbf{x}) \nabla^t \mathbf{\Lambda}_2'(\mathbf{x}) \nabla \mathbf{\Phi}(\mathbf{x}) \Big) \end{aligned} \tag{3.75}$$

where

$$\mathbf{\Lambda}_1'(\mathbf{\Omega}(\mathbf{x}, T)) = \frac{\partial}{\partial k} \mathbf{\Lambda}_1(\mathbf{\Omega}(\mathbf{x}, T)) = -\frac{1}{2} \mathbf{\Lambda}_1 \ln \tilde{\mathbf{\Lambda}}_1 \tag{3.76}$$

$$\mathbf{\Lambda}_2'(\mathbf{x}) = \frac{\partial}{\partial k} \mathbf{\Lambda}_2(\mathbf{x}) = -\frac{1}{2} \mathbf{\Lambda}_2 \ln \tilde{\mathbf{\Lambda}}_2 \tag{3.77}$$

$$\tilde{\mathbf{\Lambda}}_1(\mathbf{\Omega}(\mathbf{x}, T)) = \text{diag} \left\{ \frac{1}{(|\mathbf{\Omega}_i(\mathbf{x}, T)| + \epsilon)} \right\} \tag{3.78}$$

$$\tilde{\mathbf{\Lambda}}_2(\mathbf{x}) = \text{diag} \left\{ \frac{1}{(|(\nabla |\mathbf{x}|)_i|^2 + \epsilon)} \right\} \tag{3.79}$$

From Equation 3.70 we obtain for λ_1

$$\begin{aligned} & \lambda_1^4 \frac{20k}{\tau} Q_1 \mathbf{\Lambda}_1 \\ & +\lambda_1^2 \left(\frac{40k}{\tau} \mathbf{\Lambda}_1 - 4Q_1 \mathbf{S}^H \mathbf{S} - 2k\lambda_2^2 Q_1 \mathbf{\Phi}^H \nabla^t \mathbf{\Lambda}_2 \nabla \mathbf{\Phi} \right) \\ & -4\mathbf{S}^H \mathbf{S} - 2k\lambda_2^2 \mathbf{\Phi}^H \nabla^t \mathbf{\Lambda}_2 \nabla \mathbf{\Phi} = 0 \end{aligned} \tag{3.80}$$

Whereas, from Equation 3.71 we obtain for λ_2

$$\begin{aligned} & \lambda_2^4 k Q_2 \mathbf{\Phi}^H \nabla^t \mathbf{\Lambda}_2 \nabla \mathbf{\Phi} \\ & +\lambda_2^2 \Big(2k \mathbf{\Phi}^H \nabla^t \mathbf{\Lambda}_2 \nabla \mathbf{\Phi} + 4Q_2 \mathbf{S}^H \mathbf{S} - \frac{10k}{\tau} \lambda_1^2 Q_2 \mathbf{\Lambda}_1 \Big) \\ & +2\mathbf{S}^H \mathbf{S} - \frac{10k}{\tau} \lambda_1^2 \mathbf{\Lambda}_1 = 0 \end{aligned} \tag{3.81}$$

where the two Equations 3.80 and 3.81 can be solved as quadratic equations after a simple variable replacement. The expression for the parameter k is more complicated and has to be solved numerically.
The Huber-Markov threshold, introduced in 3.53 is set experimentally. In Pan & Reeves (2006) it was set as 10% of the maximal value, therefore we propose to set T as the mode of the amplitude histogram. The computation of parameter k requires an initial parameter k in order to compute the Q_1 and Q_2. We set it to $k = 0.5$. The parameters ϵ, δ and γ were set experimentally to $\epsilon = 1e-3$, $\delta = 1e-3$ and $\gamma = 1$, respectively. The Equations 3.70, 3.71 and 3.72 allows to estimate the parameters which maximize the evidence.

3.5 Summary of the chapter

In this chapter the models for image restoration are presented. The Gamma-GMRF model for despeckling and feature extraction, and the MAP functional with isotropic GMRFs for the real and imaginary parts are described, as well as, the algorithm for the optimization of convex functionals and the EM algorithm for parameter estimation. The GMRF model is extended in complex-valued domain in case of proper and in the less simple case of improper random variables. Thus, the Tikhonov functional with Huber-Markov prior is presented and because of its correspondence with Bayesian inference it is solved as a MAP estimation problem. The optimization is done by a gradient descent algorithm. The filter parameters are estimated through the evidence framework. The method allows restoration of the image while preserving the image features. The choice of different priors can lead to the preservation of other features or, in the case of GMRF, to the estimation of texture parameters.

Chapitre 4

Experimental Results

This chapter presents the experimental results obtained with the analyses and the models presented in Chapter 3. The model selection by Rate Distortion applied to Analyzing Window size and Model Order is presented first. Then, the results on complex image statistical regularization and complex-valued GMRF data modelling are provided, followed by the presentation of the main results on Tikhonov-like optimization for image despeckling. The proposed method is compared with the most known adaptive filters.

4.1 Rate Distortion-based Model Selection

In data mining applications, the image analysis for feature extraction is an important issue. The image characterization for compression and/or content labeling is based on algorithms which are able to extract information from the image. Such information is further embedded in information mining systems which provide the interface between the data and the users. Most of the data analysis is made on analyzing window basis. Therefore, the following problem arises : which is the optimal size of the AW ? In order to give an answer to this question we verify the method proposed in Soccorsi et al. (2006) on SAR images in order to find the optimal average size of the analyzing window for image parameter estimation. The AW has to be large enough to ensure a reliable estimation and small enough to adapt the data variability which in statistical sense means non-stationarity. The basic steps of the procedure are the following, see flowchart in Figure 1 :

1. Extraction of the primitive features.
2. Feature clustering by vector quantization.
3. Computation of Euclidean distances in the clustered feature space.

Each point has an equivalent step under the point of view of information theory :

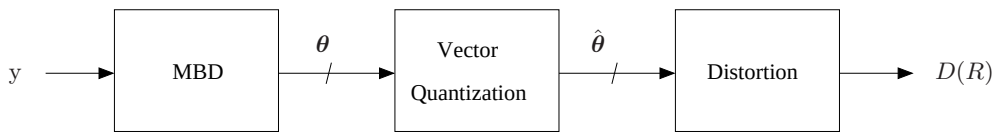


FIGURE 1 – Flowchart for Rate Distortion Model Selection.

1. lossless source coding.
2. lossy data coding compression.
3. global distortion measure evaluation.

In order to extract primitive features, we filtered the image with the MBD algorithm by Walessa & Datcu (2000). The analysis of the image is on window basis which is one of the algorithm input parameters. A large AW makes the estimation more robust in stationary areas but introduces an error when different areas are included. On the other hand, a small AW adapts better to image diversity but does not ensure a sufficient number of samples for a robust estimate. Thus, the optimal average AW has to be found.
The vector quantization was performed with a dyadic k-means which splits locally any cluster in two new clusters and is faster than the classical *generalized Lloyd algorithm* : at iteration l the current number of clusters is 2^l. The drawback of the dyadic k-means algorithm is that only a power of two number of clusters is allowed at each iteration.
Eventually, the global distortion is computed in the clustered feature space

$$D = \frac{1}{n} \sum_{i=1}^{n} |\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i|^2 \tag{4.1}$$

where a squared error distortion measure has been adopted. Further the same method has been applied in order to estimate the optimal model order of the prior GMRF. In order to do this, we have to introduce the distortion of the source, because the hypothesis of lossless data coding drops.
We want to find the optimal fitting but avoiding over-parametrization. In other words, by increasing the number of parameters, the model fits better the data but, on the other hand, we pay in terms of model complexity because the feature space cardinality increases. Thus, the variance of the GMRF model is considered as source distortion term, and added to the global distortion of the lossy compression.
The distortion curves are compared with the theoretical lower bound for a mixture of Gaussian sources, which is the lower distortion curve. It was computed according to (Cover & Thomas, 1991b; So & Paliwal, 2003) :

$$D = \sum_{i=1}^{m} \mu_i \cdot D_i(b_i) \tag{4.2}$$

with

$$D_i(b_i) = n \cdot K \cdot \Lambda_i \cdot 2^{-2\frac{b_i}{n}} \tag{4.3}$$

and

$$\Lambda_i = \left(\prod_{j=1}^{n} \lambda_{ij} \right)^{\frac{1}{n}} \quad \text{for} \quad i = 1, \ldots, m \tag{4.4}$$

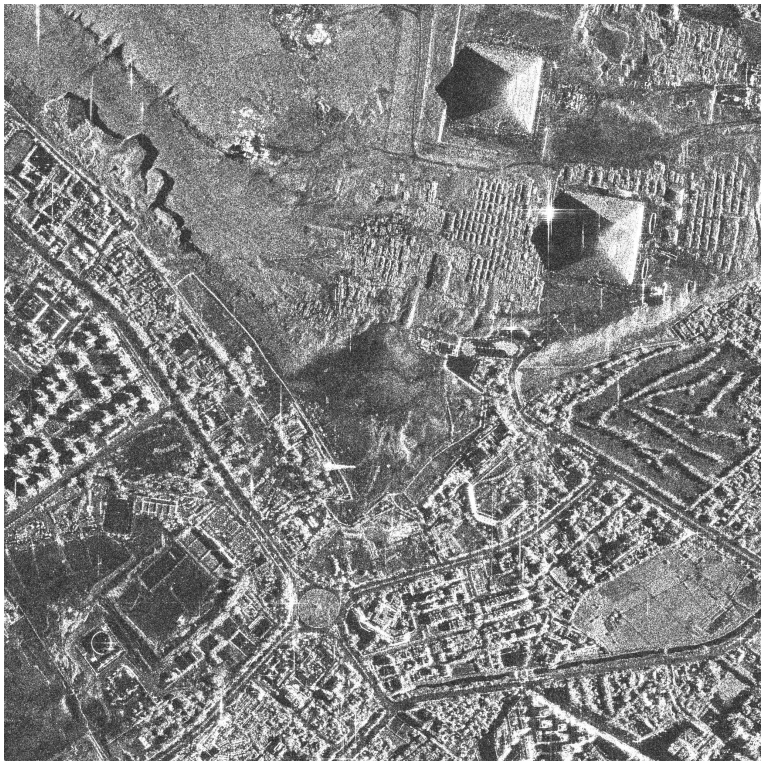
where μ_i is the weight dependent on the number of pixels belonging to the cluster, b_i is the number of bits assigned to i-th cluster, m is the number of clusters, n the space dimension, λ_{ij} the j-th variance of cluster i and K a constant which is approximately $\frac{\pi\sqrt{3}}{2}$ for Gaussian sources.
We assume to represent the number of clusters for each curve with the same number of bits : one bit for two classes, two bits for four classes, ..., seven bits for one hundred twenty eight classes. The lower bound may not be reached because the Gaussian model is

an approximation. The parameter optimization method is validated on the image shown in Figure 2a of size 4096× 4096 pixels. The image has been previously subsampled by a factor of two, thus the processed data size is 2048× 2048 pixels. The image is a TerraSAR-X Multi-Look Ground-range Detected (MGD) HR SL mode. Polarization HH, descending orbit, ground range resolution ∼1.35 m, azimuth resolution ∼1.40 m. Cairo, Egypt 6th of September 2007, 7 :16 :03 UTC.

Estimation of the optimal average size of the Analyzing Window The MBD algorithm (Walessa & Datcu, 2000) has been executed for six different sizes of the analyzing window from 11×11 up to 61×61 pixels with a step of ten pixels per side. The model order was fixed to fourth order in this experiment. The input model order for the analysis is three which leads to a feature space of cardinality eight, with : six spatial parameters, the norm of the parameters and the variance of the model.
Then, the feature space was clustered from 2 up to 128 classes, which means 7 steps for the dyadic version of k-means. The number of iterations was fixed experimentally to 30 iterations. The global distortion was computed at each iteration according to Equation 4.1.
The distortion curves versus the number of clusters are shown in Figure 3 : they have an asymptotic behavior and the curve slides down by increasing the window size.
The values of the distortion are shown in Table 2 together with the ENL of the despeckled image which is proportional to the inverse of the noise.
The distortion curves show a minimum for the analyzing window of size 61×61 pixels. Therefore, because of the curves becoming closer starting from a size of 41×41 pixels this choice of size is a good compromise between complexity and global distortion. The ENL of the despeckled image shown in Table 2 confirms that the optimal despeckling is reached for an analyzing window of size 41×41. The method, applied on SAR data, shows a different behavior of the distortion curves in the range of the considered AW sizes, differently from Soccorsi et al. (2006) where it was successfully applied to assess the optimal average AW size on optical data. This difference could depend on several reasons ; either the feature extraction is not correctly coding the texture of the data, or the minimum of distortion is reached for a larger AW size. However, the choice of a larger analyzing window makes the computation time unacceptable, thus it may not be considered as a reasonable solution. Further analysis is needed in order to understand the asymptotic behavior of the distortion which at some point is expected to increase with increasing size of the AW.

Analyzing Window	Model Order	Computation time
11× 11	4	4h13′
21×21	4	6h39′
31×31	4	10h50′
41×41	4	13h45′
51×51	4	19h30′
61×61	4	23h14′

TABLE 1 – MBD computation time for Model Order 4 and different AW sizes. The algorithm has been executed on a Sun machine with two CPU of 1.6 GHz.



(a)

(b)

FIGURE 2 – (a) Data processed for the Rate Distortion-based estimation of the optimal average AW size and the optimal average model order. (b) The Fisher information-based map of the local optimal model orders is provided. The grey scale corresponds to model order 2 (black) up to model order 7 (white).

Estimation of the optimal average Model Order For the following experiment, MBD has been executed with different model orders and the distortion has been computed taking into acount also the distortion introduced by the source coding. The distortion curves are shown in Figure 4 and the distortion values in Table 3. The minimum of the distortion curve is reached for model orders 4, 5 and 6 and model order 4 is the best compromise in term of complexity. Moreover, the ENL, shown in Table 3, reaches the maximum for model order 4, which corresponds to the best removal of the noise in the image.
This global result is compared with the map of the model order shown in Figure 2b. The map of the local optimal model order has been obtained by model selection through Fisher information. Although the averaging of the model order map is meaningless, mathematically it is permitted and by this operation an expected value for the model order of 4.2 is obtained. The value is consistent with the global analysis performed by Rate Distortion and it is a further confirmation of the validity of the method.

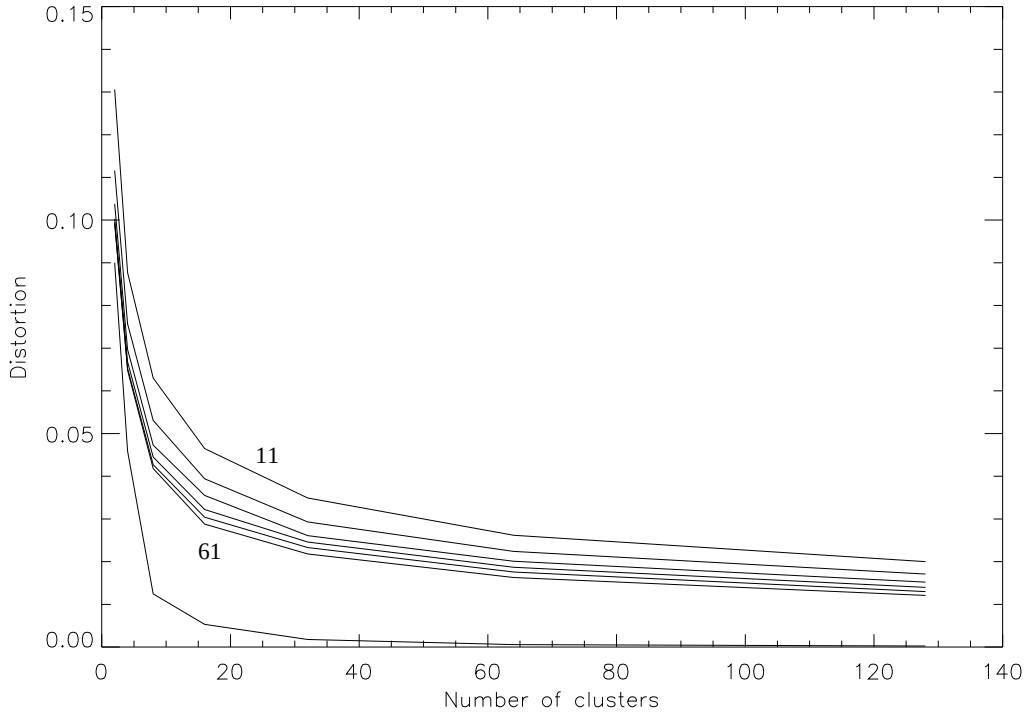


FIGURE 3 – Curves of distortion for windows of sizes 11×11, 21×21, 31×31, 41×41, 51×51 and 61×61 are shown respectively from top to bottom.

AW size	Number of clusters							ENL
	2	4	8	16	32	64	128	
11×11	13.06	8.77	6.30	4.65	3.49	2.62	2.00	67.94
21×21	11.16	7.57	5.31	3.94	2.93	2.24	1.71	97.90
31×31	10.38	6.97	4.73	3.55	2.61	2.01	1.52	101.2
41×41	10.03	6.68	4.45	3.22	2.46	1.87	1.40	104.7
51×51	9.91	6.53	4.27	3.04	2.33	1.76	1.30	99.53
61×61	9.95	6.48	4.18	2.88	2.18	1.63	1.21	92.78

TABLE 2 – Distortion values $(\times 10^{-2})$ and ENL of the despeckled image for AW of 6 different sizes. The best compromise between complexity and distortion is to chose an AW of size 41× 41 pixel (highlighted in red) which has the maximum of ENL.

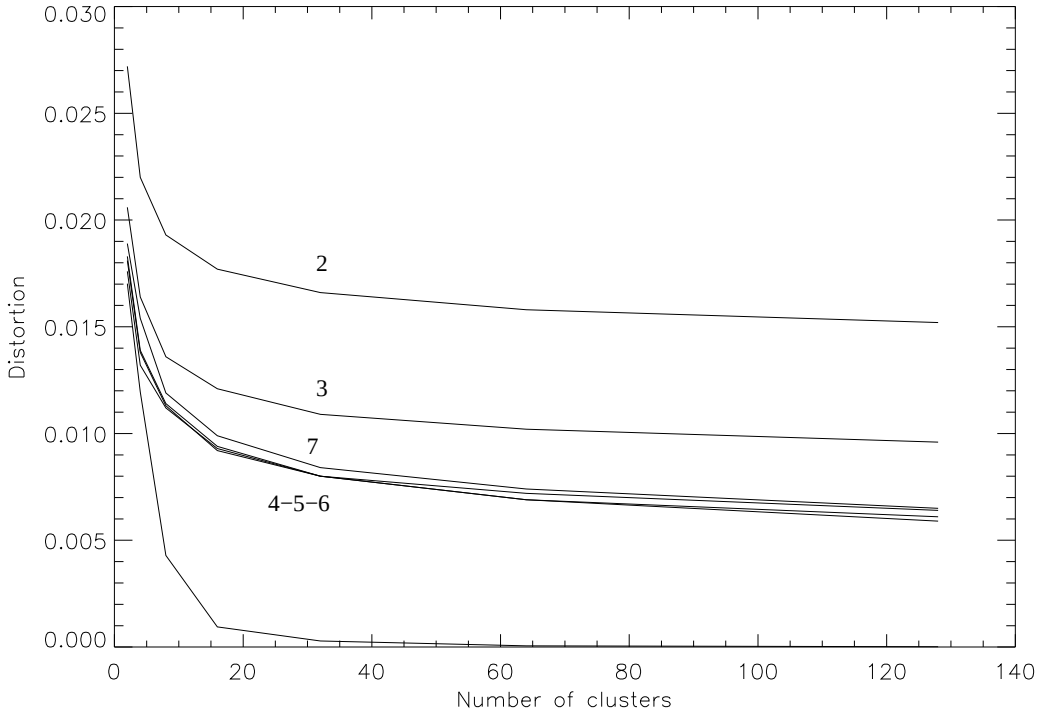


FIGURE 4 – Curves of distortion for model orders 2, 3, 4, 5, 6 and 7. The curve reaches the minimum for model orders 4, 5 and 6.

Model Order	Number of clusters							ENL
	2	4	8	16	32	64	128	
2	2.72	2.20	1.93	1.77	1.66	1.58	1.52	73.03
3	2.06	1.64	1.36	1.21	1.09	1.02	0.96	101.9
4	1.81	1.38	1.13	0.92	0.80	0.72	0.64	104.6
5	1.76	1.32	1.12	0.93	0.80	0.69	0.61	47.95
6	1.83	1.39	1.14	0.94	0.80	0.69	0.59	39.74
7	1.89	1.54	1.19	0.99	0.84	0.74	0.65	47.96

TABLE 3 – Distortion values ($\times 10^{-2}$) and ENL of the despeckled image for different Model Orders. Model orders 4, 5 and 6 have comparable distortion values. The best choice in term of complexity is model order 4 (highlighted in red).

Image model	Equation 3.11
Prior model	Equations 3.16 and 3.16
Optimization/Estimation	Steepest descent/EM

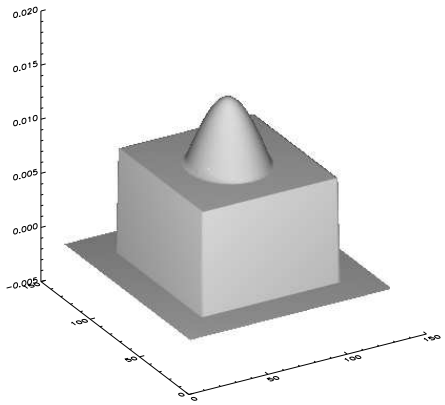
TABLE 4 – Summary of the model equations and optimization/estimation method.

4.2 Preliminary results

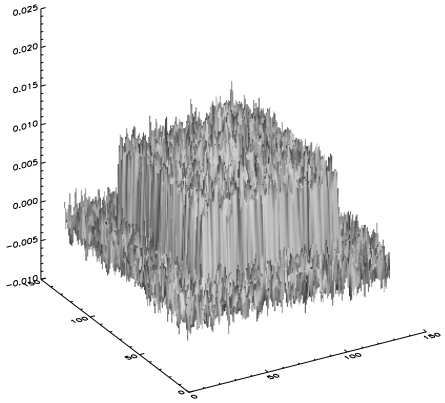
The first results presented are based on a complex-valued MAP statistical regularization and a Complex-valued GMRF linear model. The two methods for parameter estimation are going to be presented with an example of their application.

4.2.1 Complex-valued MAP Statistical Regularization

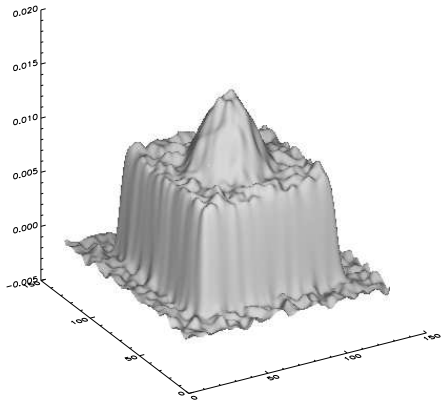
The complex statistical MAP regularization is based on the smoothing of the two I and Q channels. It copes with additive noise. The equation involved are resumed in Table 4. The methods follow the flowchart shown in Figure 2. Before, the parameters are estimated according to the EM algorithm 1. These parameters are used in the MAP estimate to smooth the real and the imaginary channels. Because of the **A** matrix in the energy function of Equation 3.10 the real and imaginary channels are coupled. Assuming, for the test, that the real and imaginary parts have the same profile, only the real part is shown in Figure 5a. The image of 128$\times$ 128 pixels is partitioned in a mosaic of 32 $\times$ 32 pixels. For each tile the parameters are estimated. The parameter σ_R and σ_I estimated from the original data 5a are presented in Table 5a and the correspondent image is shown in Figure 6a. The boundary condition is managed giving to the parameters the mean value of the parameters of the confining regions.
Thus, an additive noise with SNR=10 dB is summed to the image giving the data shown in Figure 5b. The parameters estimated from incomplete data are presented in Table 5b and the respective parameter image is shown in Figure 6b.
The iterations for the estimated MAP are performed according to Equation 3.23 with the parameter $\alpha = 10^{-6}$. The result after 30 iterations is shown in Figure 5c.
The value of the parameter α is chosen experimentally as well as the number of iterations. A higher value of α ensures a faster convergence but a less precise minimization of the functional.
The results are shown on a synthetic complex image profile corrupted by additive noise with SNR=10 dB and setting the value of the $\alpha = 10^{-6}$.



(a)



(b)



(c)

FIGURE 5 – Example of regularization from synthetic real part data. Original profile (a). Noisy profile with SNR=10 dB additive noise (b). Regularized profile after 30 iterations (c).

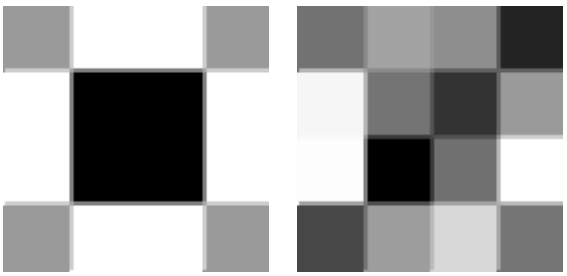
	1	2	3	4
1	4.75	5.82	5.82	4.75
2	5.82	2.31	2.31	5.82
3	5.82	2.31	2.31	5.82
4	4.75	5.82	5.82	4.75

(a)

	1	2	3	4
1	3.51	4.32	4.40	3.59
2	4.32	1.87	1.90	4.37
3	4.37	1.64	1.66	4.34
4	3.59	4.37	4.32	3.57

(b)

TABLE 5 – Estimated parameter σ_R from complete data (a) and incomplete data (b). The values are multiplied by a factor 10^{-3}. The numbers in the first line and column refer to the partition of the image.



(a) (b)

FIGURE 6 – Images of estimated parameters from original data (a) and noisy data (b). They are used in the MAP estimate. In the boundary the mean value of the parameters is computed.

Example on actual SAR data The method has been applied also on real data, shown in Figure 8g. The noise variance has been estimated in a dark area of the data, where low signal is supposed to be, characterized by SNR=78 dB. The parameter is set to $\alpha = 10^{-3}$ and the results are after 10 iterations. The estimated parameters are shown for real and imaginary part in Tables 6a and 6b. The parameter images are shown in Figures 8a and 8b. The 3D profiles of real part and imaginary part before and after denoising are shown in Figure 7. The correspondent images are shown in Figures 8c and 8d for the original and in Figures 8e and 8f after denoising. The resulting amplitude is shown in Figure 8h. As expected the speckle is not removed, because the speckle noise model is multiplicative and not additive. On the other hand, the appearance of the output is a blurred image where some small details appear lost.
The effect of the filter on the signal is plotted in Figure 9 where the normalized real part of one line of the original 8c and of the output signal 8e is shown. The signal appears smoothed, but the strong variations are preserved, depending on the number of iterations.
The method allows to estimate the signal parameters and from them to regularize the noisy signal. The approach permits to smooth real part and imaginary part, taking into account their coupling, but it does not allow to cope with multiplicative noise. Thus, if a despeckled version of the original data is needed other approaches have to be used.

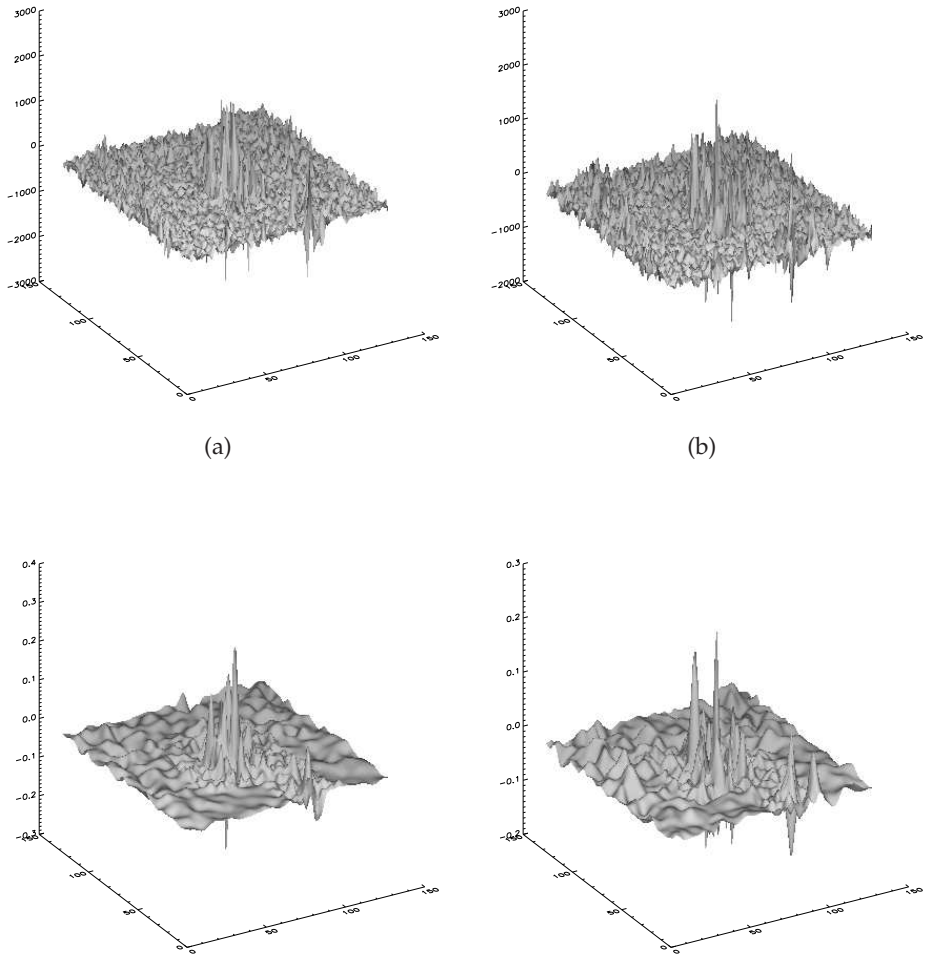
	1	2	3	4
1	0.127	0.116	0.234	0.111
2	0.141	0.314	0.264	0.123
3	0.149	0.214	0.236	0.143
4	0.161	0.138	0.136	0.152

(a)

	1	2	3	4
1	0.125	0.122	0.224	0.119
2	0.144	0.285	0.247	0.113
3	0.155	0.216	0.218	0.144
4	0.164	0.139	0.136	0.159

(b)

TABLE 6 – Estimated parameters from real part (a) and imaginary part (b). The numbers in the first line and column refer to the image partition.



(a) (b)

(c) (d)

FIGURE 7 – 3D plot of real part and imaginary part. Original profiles (a) and (b) and reconstructed profiles (b) and (c). The profiles appear smoothed.

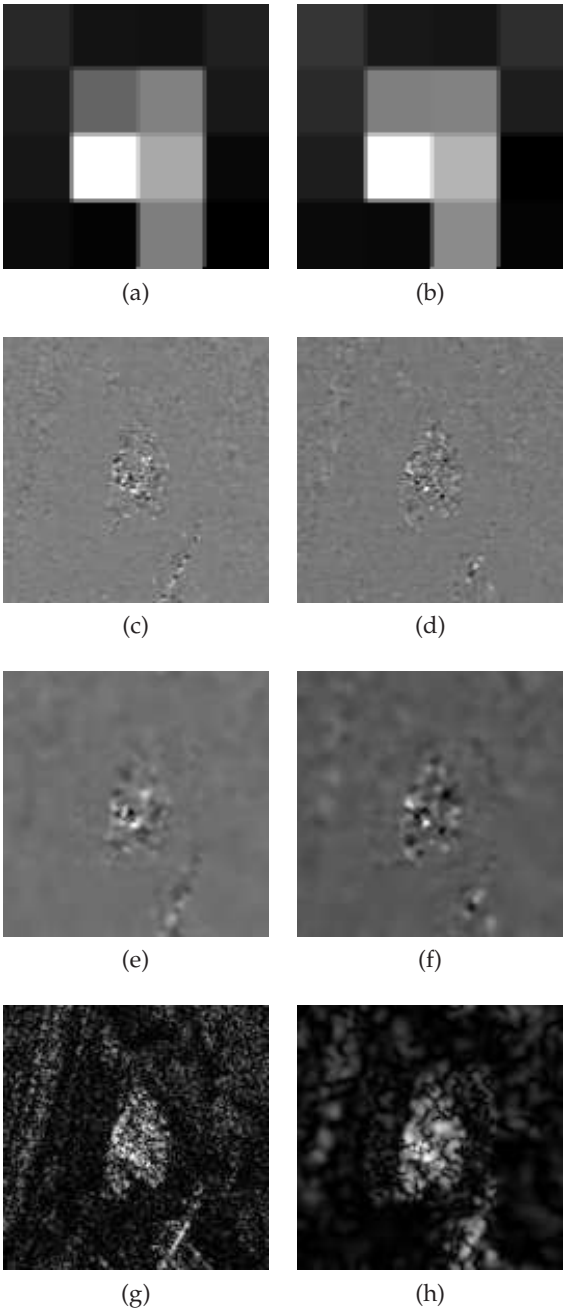


FIGURE 8 – Example of MAP regularization of actual SAR data. Estimated parameter images from real part (a) and imaginary part (b). Original real part (c) and imaginary part (d). Real part and imaginary part after regularization (e) and (f), respectively. Original amplitude image (g) and amplitude image after regularization (h).

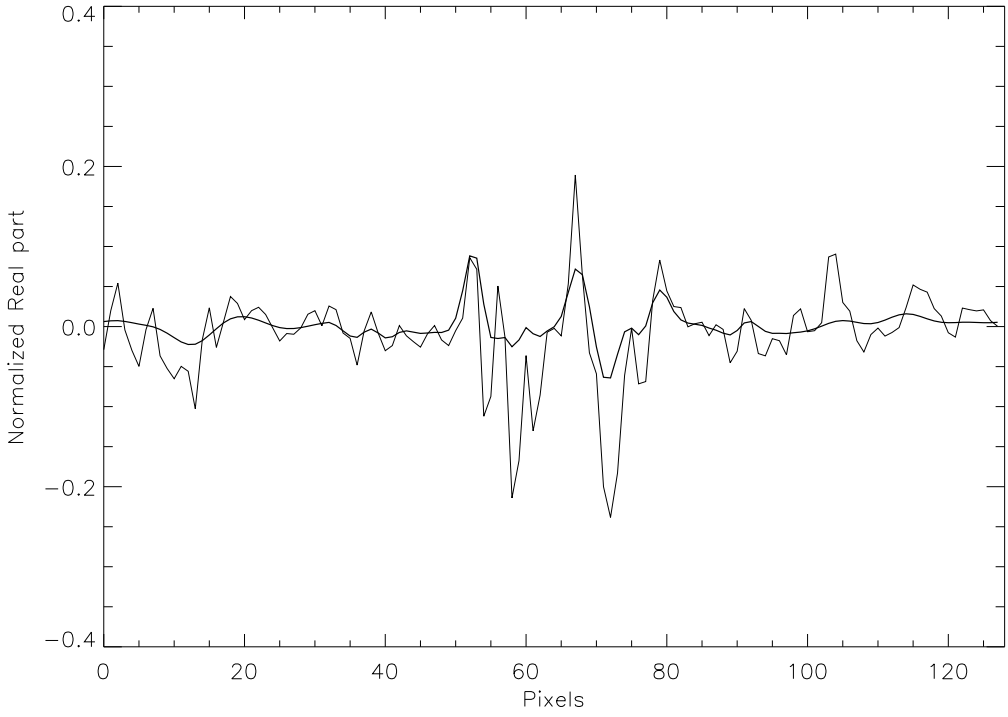


FIGURE 9 – Signal profiles before and after regularization, thin line and thick line respectively. The signal appears smoothed but the peaks are preserved.

Image model	Equation 3.38
Estimation	MAP

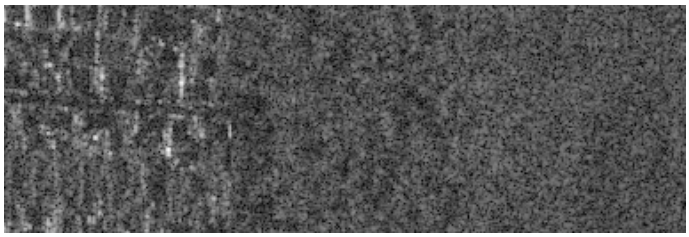
TABLE 7 – Summary of the model equation and estimation method.

4.2.2 Complex-valued GMRF model

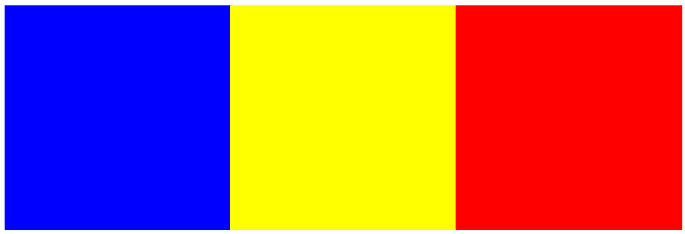
The analysis and the synthesis of the Complex-valued GMRF has been presented in Section 3.3.
It is applied to SAR textures. Three textures have been selected and a mosaic has been build with a residential area from a city, vegetation from a forest and grass from an agricultural field, see Figure 10a.
The mosaic has been analyzed according to the flowchart shown in Figure 7. The parameters have been estimated using Equation 3.44 and the variance Equation 3.45.
The estimated model parameters are presented in Table 8. Except the variances, the values of the parameter vector $\hat{\boldsymbol{\theta}}$ are really similar. However, the unsupervised k-means classification, performed only on the parameter vector $\hat{\boldsymbol{\theta}}$ is able to separate the textures in three different classes, Figure 10b.
After retrieval the forward modeling has been performed according to the block diagram shown in Figure 4. The result is shown in Figure 10c. Even if the Complex-Valued GMRF is able to distinguish the analyzed textures, the forward modeling does not lead to a result very similar to the original texture. This is because the high dynamic of the signal is not well modeled by the Gaussian. Visually the distinction is mainly due to the σ values. Other models, e.g. auto-binomial, have chances to fit better the data (Hebar et al., 2009).

Class	$(1,0)$ $\hat{\theta}_1$	$(0,1)$ $\hat{\theta}_2$	$(1,1)$ $\hat{\theta}_3$	$(-1,1)$ $\hat{\theta}_4$	$\hat{\sigma}$
City	0.20+j0.20	0.21+j20	0.19+j19	0.17+j18	102.0
Forest	0.20+j0.20	0.21+j0.21	0.20+j0.20	0.17+j0.18	40.9
Grass	0.20+j0.21	0.20+j0.19	0.20+j0.20	0.16+j0.16	45.6

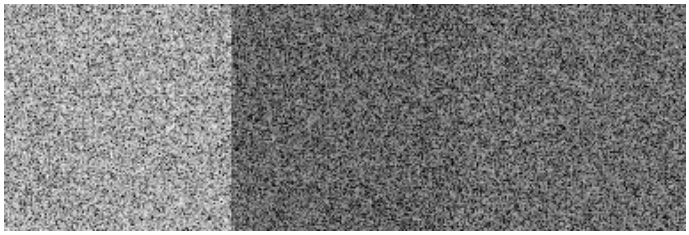
TABLE 8 – Estimated parameter vector $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ from mosaic shown in Figure 10a.



(a) Stationary complex textures.



(b) Classified texture based on GMRF parameter vector shown in Table 8.



(c) Simulated complex textures.

FIGURE 10 – Complex-valued GMRF experiment. Original magnitude texture image (a). Parameter vector $\hat{\boldsymbol{\theta}}$ based k-means unsupervised classification (b). Forward modeling based on the estimated parameters (c).

Image model	Equation 3.11
Prior model	Equation 3.54
Optimization/Estimation	Steepest descent/Evidence maximization

TABLE 9 – Summary of the model equations and optimization/estimation method.

4.3 Tikhonov Regularization

The proposed Tikhonov regularization method with Huber-Markov Random Field (HMRF) prior has been compared with the most well-known adaptive filters : enhanced Lee, enhanced Frost, Kuan, Gamma and MBD. The first experiment has been performed on four GMRF synthetic textures, the second on four Brodatz texture images. The third experiment was carried out with an optical QB image and the last on actual HR TerraSAR-X data.
The methods applied for the comparison of the filter are qualitative and quantitative. The comparative measures taken into account are the following

- Preservation of the mean value.
- Mean-Squared Error (MSE) $\mathrm{E}\{|\hat{\mathbf{x}} - \mathbf{x}|^2\}$
- Structural SIMilarity (SSIM) index (Zhou Wang & Simoncelli, n.d.) $\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}})$.
- Equivalent Number of Look (ENL) of the despeckled image $\hat{\mathbf{x}}$.
- Mean value of the noise intensity $\mathrm{E}\{\mathbf{y}/\hat{\mathbf{x}}\}$.
- ENL of the ratio image $\mathbf{y}/\hat{\mathbf{x}}$.
- Visual comparison of the ratio image $\mathbf{y}/\hat{\mathbf{x}}$.

Not all methods are applicable to all the experiments. For instance the definition of ENL in case of texture has no meaning and some methods are not practical without the reference image. Thus, they are not suitable in case of actual SAR data experiments.
The filter with the lower MSE is preferred. On the other hand, the MSE exhibits weak performance for images (Wang & Bovik, 2009), thus the SSIM is also used. It is a quality measure of an image which is compared to another image which is supposed to have a perfect quality. It is computed as follows (Wang et al., 2004; Channappayya et al., 2008)

$$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + c_2)} \tag{4.5}$$

where $\mathbf{x}$ is the reference image and $\hat{\mathbf{x}}$ is the despeckled image, μ and σ are the mean and the variance, respectively, and c_1 and c_2 are two constants introduced to avoid numerical problems when the values of mean and variance are close to zero. The SSIM index approaches one when the selected image approaches the referenced one. Therefore, the filter whose SSIM index is closer to one is preferred.
The mean value of the image has to be preserved by the filter because the signal expectation doesn't have to be modified by the filter processing.
The ENL of the despeckled image is computed by

$$ENL = \frac{\mathrm{E}\{\mathrm{x}\}}{\mathrm{E}(\mathrm{x} - \mathrm{Ex})^2} \tag{4.6}$$

where x represents the signal intensity. It is also a classical method for filter comparison but it may be applied only if a sufficiently large uniform area is available in the data.
The expectation of the ratio image has to be one for negative exponential distributed

noise intensity with unitary noise variance, thus the filter which approaches more this value has better performance than the others.
The ENL of the ratio image has to be equal to one in the ideal case.
The ratio image is shown for qualitative visual comparison. In case the edges are not well preserved the ratio shows patterns due to the not optimal filtering procedure.
The condition of convergence is chosen as the minimum MSE reached and the intial value for the parameter k is set to 0.9. The threshold τ of the Huber function has been fixed as the mode of the histogram of the data. The number of iterations is chosen as condition of convergence in case of actual SAR data because the reference image for computing the MSE is not available.
The choice to process and test images with ENL=1 is a critical point because it puts on probation the filters performance but it allows to compare the developed filter which works on complex-valued data with the other filters which work on detected data. The difference on the data relies on the non linear transformation from Cartesian to polar coordinates. In the following, the terms ratio image and speckle image are used as alternative terms.

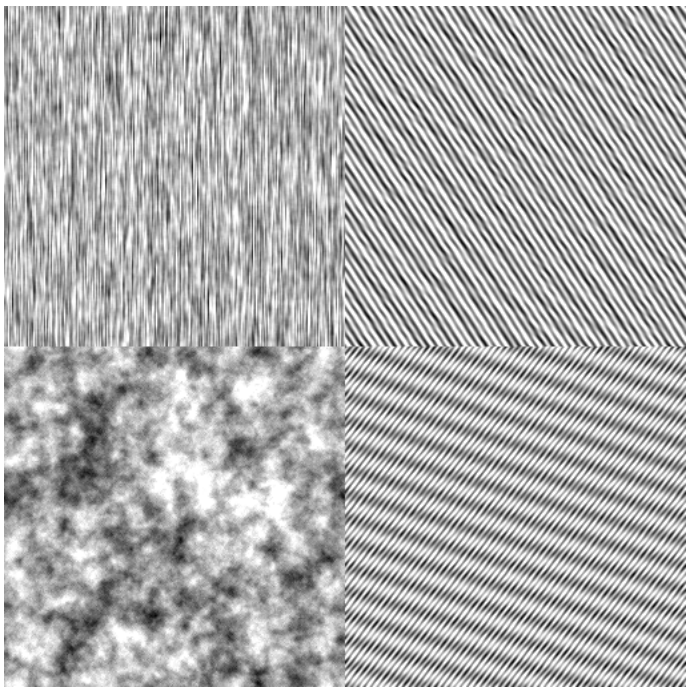
4.3.1 Simulated GMRF texture

The first experiment has been performed on synthetic textural images synthesized by the GMRF model. The images, shown in Figure 11a, have been generated with the parameters shown in Table 10. Each image has a size of 256×256 pixels. The noisy images, corrupted with $\mathrm{ENL} = 1$ speckle noise generated according to Section 1.8, are shown in Figure 11b. The despeckled images are shown in Figure 12 together with the original and the noisy images for comparison. The ratio images are shown in Figure 13. The measures for the filter comparison are available in Table 11.
From inspection of the despeckled image in Figure 12 it is visible that the Lee, Frost, MBD and the proposed method have similar results, while the Kuan and Gamma filters do not remove the noise as well as the others. On the other hand, all the images ratio show patterns especially visible in the second, third and fourth texture, where the numbering starts bottom left counterclockwise. The MBD filter shows an improvement in case of texture three because no pattern is visible in the image ratio in Figure 13e.
The measures in Table 11 show that the mean value is well preserved by all the filters except MBD which is overestimating it. The best MSE is obtained with the proposed method except in case of texture four where the best MSE is reached with MBD and the HMRF with Tikhonov regularization has performance similar to the enhanced Lee filter. In case of SSIM index all the filters do not exhibit excellent performance. A possible explanation is because none of the filters is able to handle with texture except MBD. However the latter does not outperform the others filters in term of SSIM index. In case of texture one the best value is reached by Frost and Gamma filters followed by the enhanced Lee filter. For texture two, three and four the best index is obtained with Kuan filter while the others exhibit poor performance. This is probably due to the fact that a low speckle removal, see Figure 13c, corresponds to a preservation of the texture. If the image is smoothed not only the noise but also the texture is removed. In term of expectation of the ratio images the MBD filter shows always good performances ; for all textures the mean value of the speckle intensity is close to one. The Kuan filter is again at second position. The equivalent number of look of the ratio is good for all the textures in case of enhanced Lee filters, followed by the proposed method and MBD. The Gamma filter and the enhanced

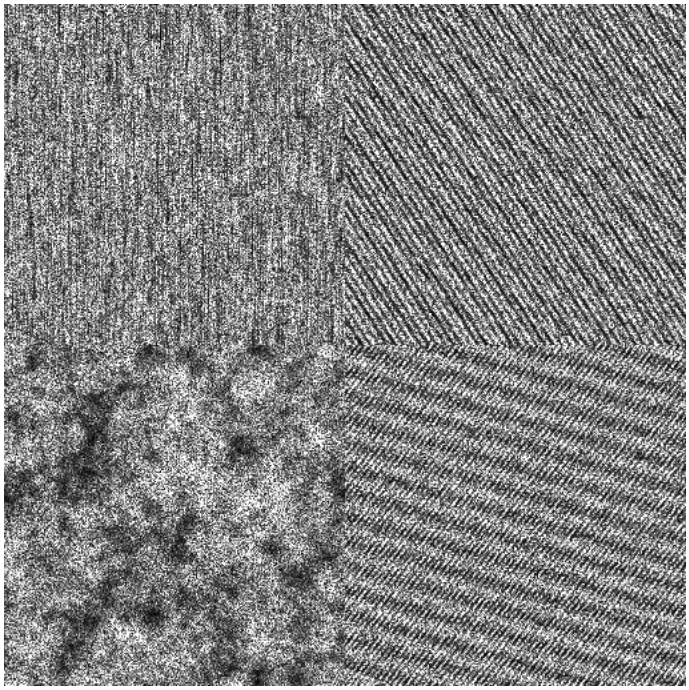
Frost have a good result in case of texture one, but a really poor result in the other cases. The Kuan filter exhibits the worse results because it strongly overestimates the ENL of the ratio images in all cases. The proposed method converges to the minimum MSE after six iterations in case of texture one, after three iterations in case of texture three and after two iterations for texture two and four. This low number of iterations is due to the fact that the removal of speckle in textured images is a compromise between noise reduction and texture preservation which are two opposite requirements.

	$(1,0)$ θ_1	$(0,1)$ θ_2	$(1,1)$ θ_3	$(-1,1)$ θ_4
T1	0.125	0.125	0.125	0.125
T2	0.21	0.15	0.30	-0.12
T3	0.33	0.32	-0.21	0.06
T4	0.3	0.5	-0.15	-0.15

TABLE 10 – Textural parameters of synthetically generated textures shown in Figure 11a. From bottom left counterclockwise : T1, T2, T3 and T4. The coordinate of the parameter is given with respect to the central pixel and considering a symmetric odd neighborhood function.

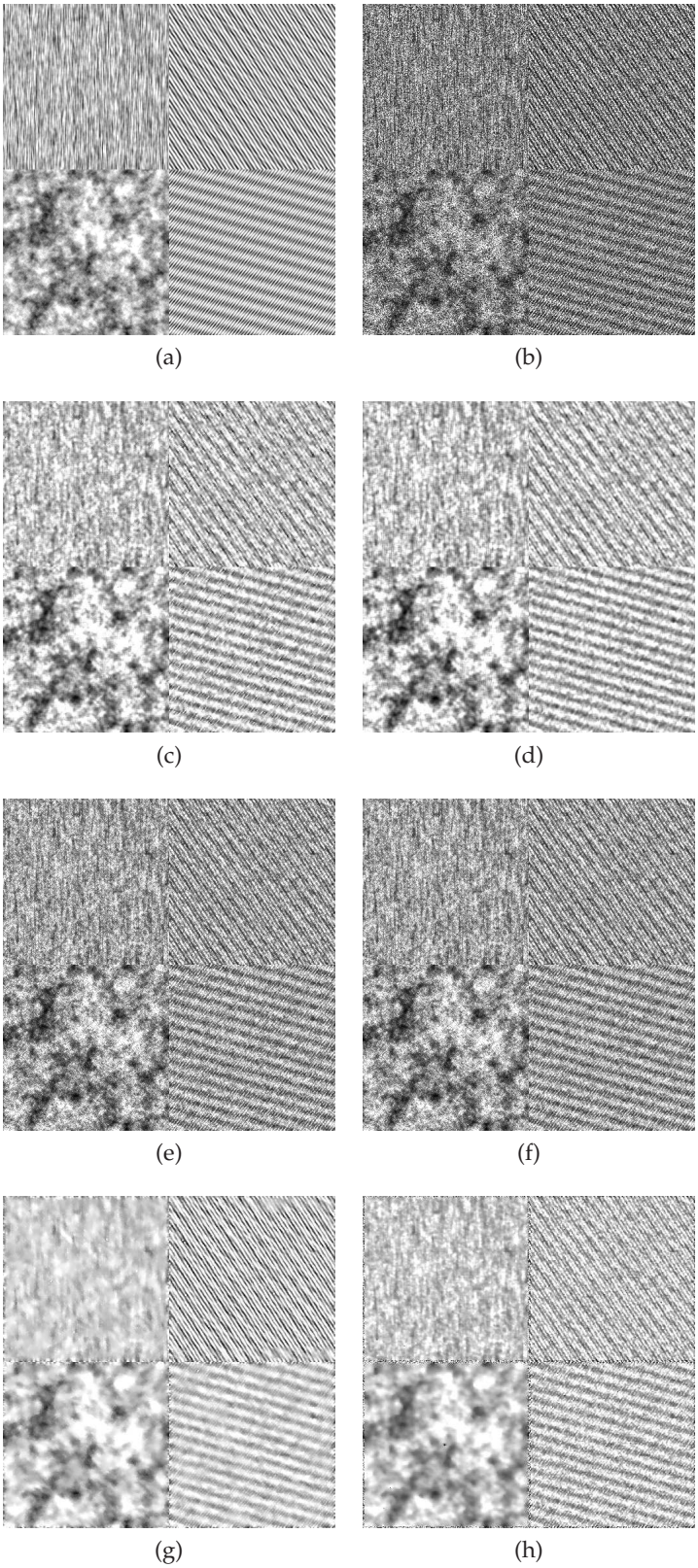


(a)



(b)

FIGURE 11 – Experiment with synthetic texture generated with the parameters shown in Table 10. From bottom left counterclockwise : T1, T2, T3 and T4. Original texture images (a) and corrupted by ENL=1 speckle noise (b).



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 12 – Experiment with GMRF texture corrupted by speckle noise with ENL=1. Original images (a), noisy images (b), despeckled with enhanced Lee (c), enhanced Frost (d), Kuan (e), Gamma (f), MBD (g) and HMRF (h).

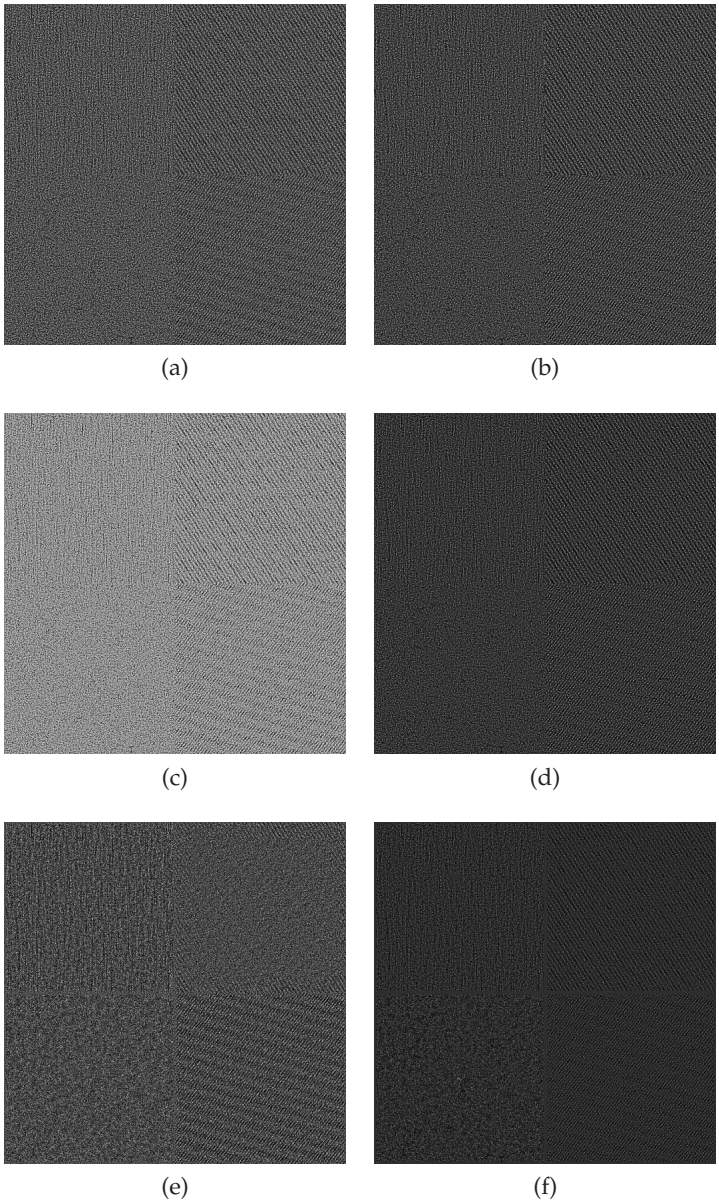


FIGURE 13 – Experiment with GMRF texture corrupted by speckle noise with ENL=1. Ratio images : enhanced Lee (a), enhanced Frost (b), Kuan (c), Gamma (d), MBD (e) and HMRF (f).

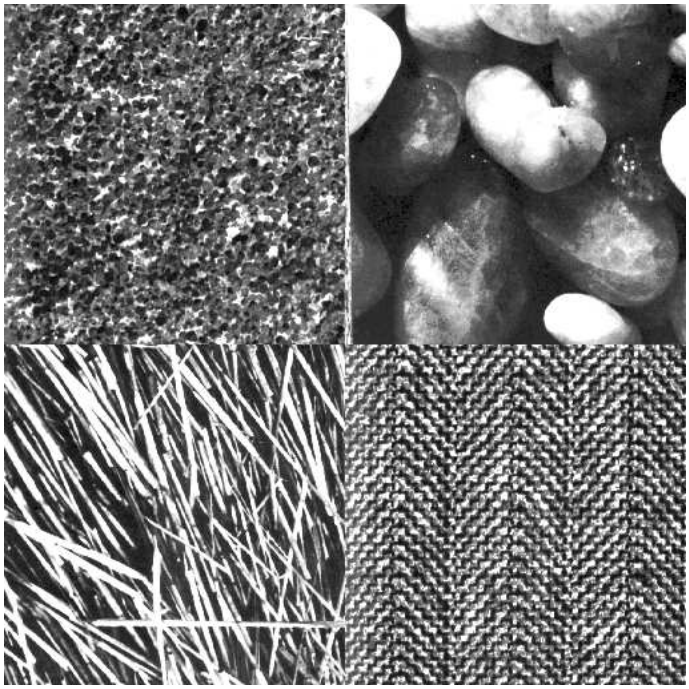
T1 $\mu = 150.253$	E$\{\hat{\mathbf{x}}\}$	MSE	SSMI	E$\{\mathbf{y}/\hat{\mathbf{x}}\}$	ENL$\{\mathbf{y}/\hat{\mathbf{x}}\}$
e. Lee	150.47	1292.92	0.56	1.21	1.16
e. Frost	150.51	1274.96	0.58	1.25	1.02
Kuan	150.53	2727.37	0.32	0.94	3.66
Gamma	150.54	1280.25	0.58	1.26	1.00
MBD	159.67	1803.56	0.31	1.06	1.13
HMRF	150.43	1136.25	0.36	1.25	0.95
T2 $\mu = 157.838$					
e. Lee	157.64	2774.62	0.24	1.177	0.98
e. Frost	157.62	3051.52	0.20	1.36	0.66
Kuan	157.72	3414.99	0.41	0.92	2.57
Gamma	157.74	3183.07	0.15	1.41	0.59
MBD	169.87	3677.60	0.09	1.07	0.82
HMRF	157.71	2673.12	0.16	1.17	0.88
T3 $\mu = 159.017$					
e. Lee	159.30	1927.62	0.24	1.22	1.06
e. Frost	159.31	1957.28	0.24	1.30	0.85
Kuan	159.36	3100.93	0.36	0.94	3.39
Gamma	159.32	1982.22	0.23	1.31	0.81
MBD	170.93	2651.50	0.12	1.08	0.95
HMRF	159.310	1798.09	0.20	1.25	0.91
T4 $\mu = 156.524$					
e. Lee	155.29	2412.20	0.38	1.14	1.00
e. Frost	155.99	2582.23	0.44	1.30	0.69
Kuan	156.55	3301.46	0.48	0.90	2.60
Gamma	156.58	2837.16	0.32	1.35	0.61
MBD	163.90	2275.76	0.40	1.037	1.27
HMRF	156.548	2449.55	0.29	1.14	0.88

TABLE 11 – Measures for the comparison of despeckled GMRF textures. From bottom left counterclockwise, with reference to Figure 11 : T1, T2, T3, T4.

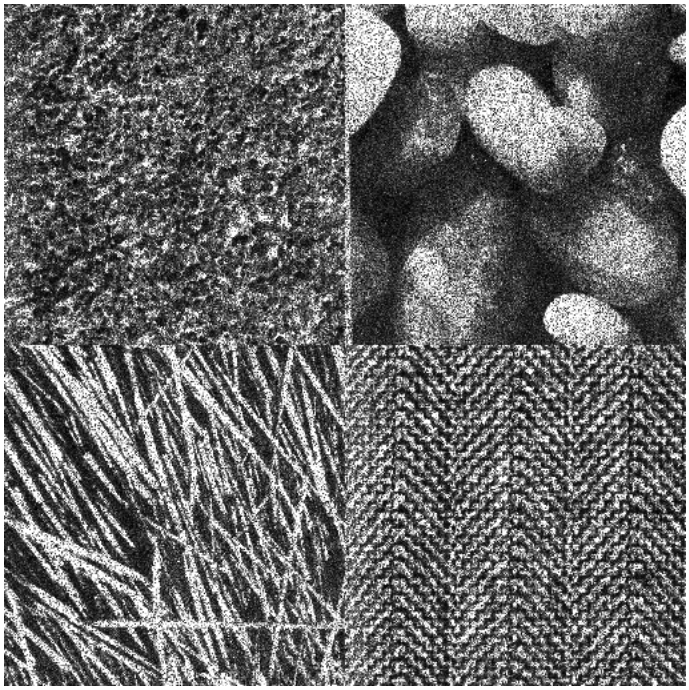
4.3.2 Brodatz textures

The second experiment has been performed on optical images belonging to the Brodatz texture archive (Randen, 1997). The selected images, the originals and the ones corrupted by speckle noise are shown in Figure 14. They have been chosen in order to have a selection of different type of textures. The texture number one represents a linear structure made by strokes; the second texture is a herringbone regular structure, the third is a round structure made by rubbed stones and the last is a marble texture. The Brodatz images are numbered from bottom left counterclockwise. Each image has a size of 256×256 pixels. The noisy images have been generated according to the procedure described in Section 1.8 with ENL=1.
The despeckled images are shown in Figure 15. The ratio images for qualitative comparison are shown in Figure 16, while the measures for the comparison of the filters are presented in Table 12.
The histograms of the amplitude of the noisy image and of the despeckled image with the enhanced Lee filter and with the proposed method are shown in Figure 17a. In Figures

17b and 17c, the histograms of real and imaginary before and after despeckling with the proposed method are shown. The histograms, shown in Figure 17, exhibit interesting behaviors. Referring to Figure 17a, the shape of the enhanced Lee amplitude histogram is similar to one of the noisy image, while the amplitude histogram with the proposed method has a Gaussian-like shape. This is consistent with the histogram of multilook data whose pdf is a Gamma distribution. The real and imaginary part histograms, shown in Figures 17b and 17c, show a symmetric Gamma shape with two modes. This is due to the fact that close to the origin is the speckle noise which is removed.
Comparing the despeckled images with the original images it appears that the enhanced Lee, enhanced Frost, MBD and the proposed method perform similarly and better than the Kuan and the Gamma filters. The proposed method is in the second place, after the MBD, by comparing the ratio image patterns, see Figures 16e and 16f. Referring to Table 12, all the filters preserved reasonably the mean value. The proposed method exhibits the best MSE for the fourth image while it performs similarly to the enhanced Lee filter for images two and three. The HMRF filter is in the third position after the enhanced Lee and the enhanced Frost in case of image one. They outperform the other filters in term of SSIM index for the first image, while the Kuan filter exhibits the best index measure in case of image two and three. The enhanced Frost performs better in term of SSIM index in case of image four. The proposed method has the second best SSIM index in case of image two. The MBD filter performs really good in the expectation of the speckle image, which is close to one for all the images. The enhanced Lee performs slightly better only in case of image two. The proposed method exhibits an acceptable mean value for the ratio image but it tends to overestimated the expectation of the speckle noise in all the images. The equivalent number of look is better for the enhanced Lee filter except for image four where the Gamma filter performs slightly better.
The proposed method converges after two iterations in case of image one, two and three and after four iterations for the fourth image.

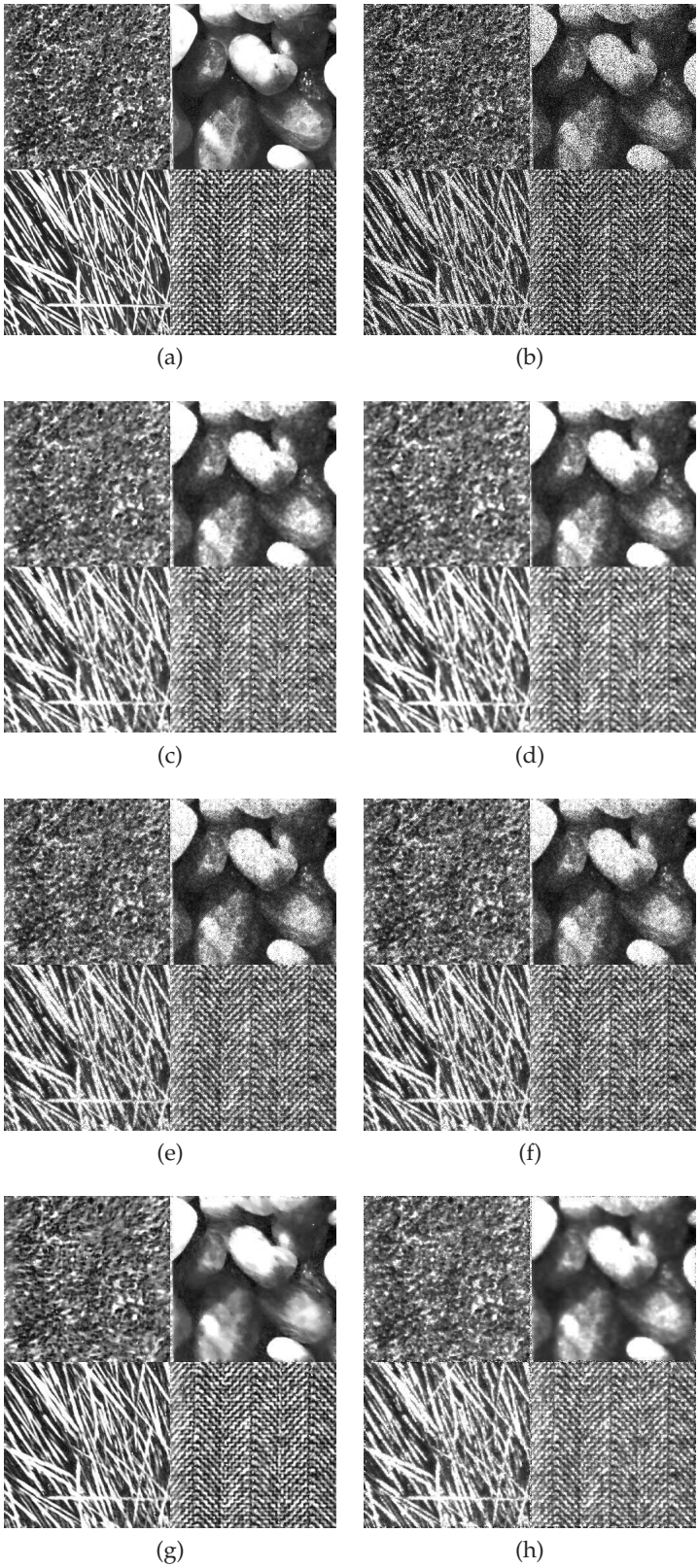


(a)



(b)

FIGURE 14 – Experiment with Brodatz textures. From bottom left counterclockwise B1, B2, B3, B4. Original images (a) and corrupted by speckle noise (b).



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 15 – Experiment with Brodatz texture corrupted by speckle noise with ENL=1 shown in Figure 14b. Original images (a), noisy images (b), despeckled with enhanced Lee (c), enhanced Frost (d), Kuan (e), Gamma (f), MBD (g) and the proposed method (h).

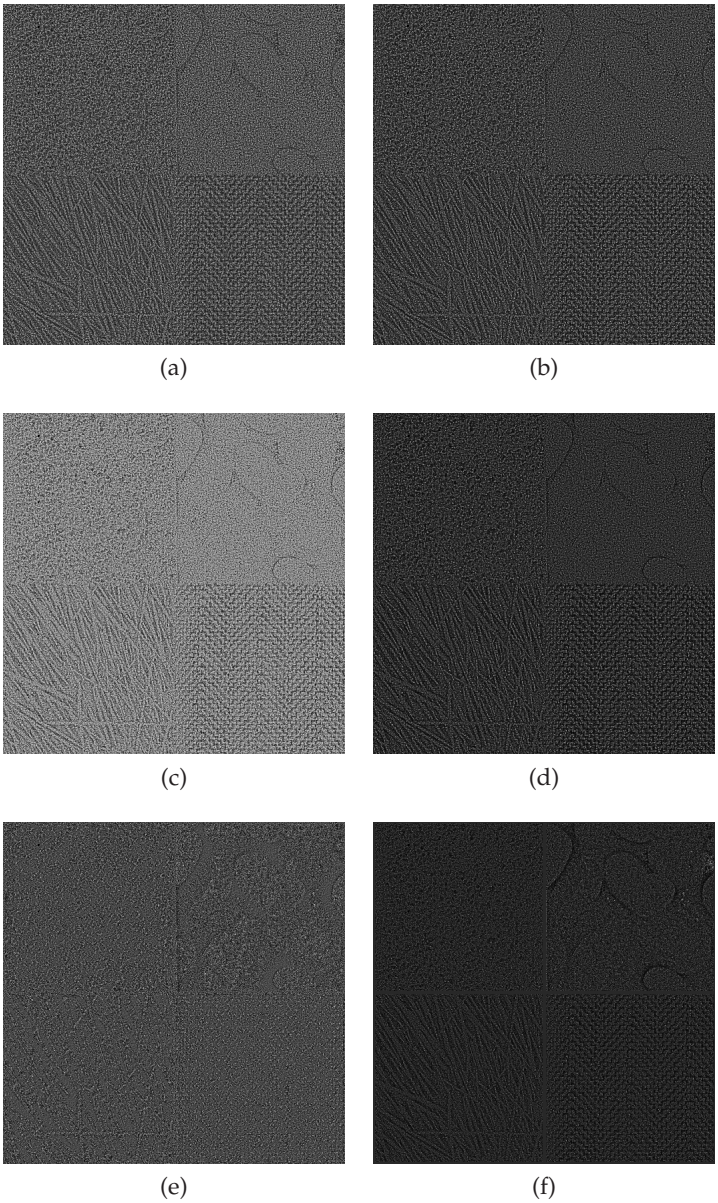
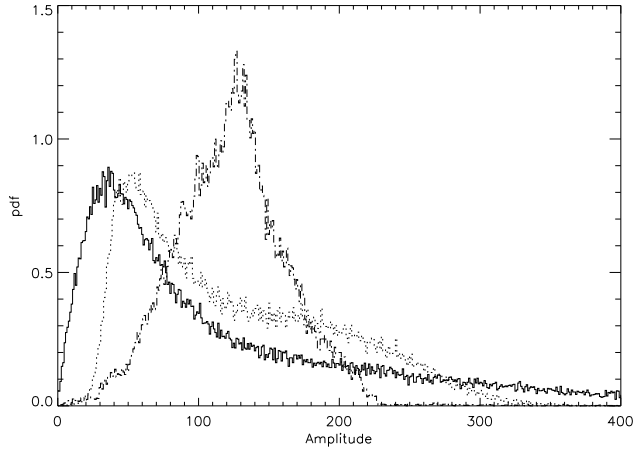


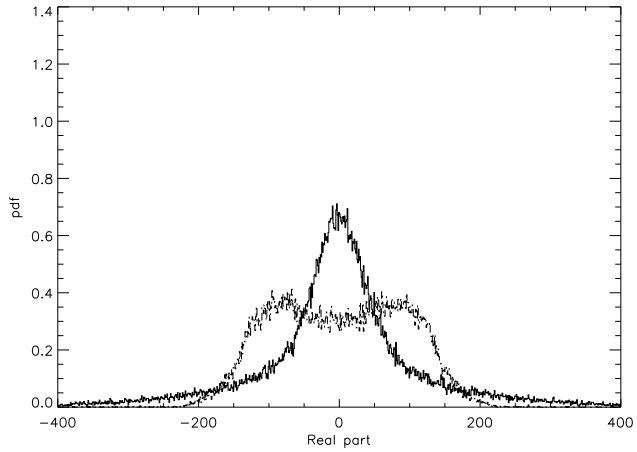
FIGURE 16 – Experiment with Brodatz texture corrupted by speckle noise with ENL=1. Ratio image : enhanced Lee (a), enhanced Frost (b), Kuan (c), Gamma (d), MBD (e) and the proposed method (f).

B1 $\mu = 127.032$	E$\{\hat{\mathbf{x}}\}$	mse	SSMI	E$\{\mathbf{y}/\hat{\mathbf{x}}\}$	ENL$\{\mathbf{y}/\hat{\mathbf{x}}\}$
e. Lee	125.31	1611.66	0.55	1.08	1.07
e. Frost	126.35	1696.21	0.57	1.23	0.67
Kuan	128.08	2705.41	0.45	0.88	2.50
Gamma	128.45	2033.89	0.51	1.27	0.56
MBD	131.04	2057.95	0.30	1.04	1.32
HMRF	128.05	1940.59	0.33	1.18	0.82
B2 $\mu = 109.550$					
e. Lee	192.07	1498.40	0.27	1.03	1.03
e. Frost	107.47	1580.76	0.25	1.25	0.58
Kuan	109.02	1896.36	0.55	0.86	2.16
Gamma	109.64	1961.16	0.18	1.32	0.46
MBD	108.96	1703.58	0.31	1.05	1.27
HMRF	108.99	1532.28	0.38	1.11	0.65
B3 $\mu = 90.2339$					
e. Lee	88.91	884.361	0.39	1.11	1.03
e. Frost	89.20	923.286	0.53	1.27	0.68
Kuan	89.65	1223.63	0.55	0.90	2.64
Gamma	89.82	1035.55	0.48	1.31	0.57
MBD	91.23	1068.50	0.29	1.04	1.23
HMRF	89.61	895.82	0.35	1.14	0.84
B4 $\mu = 104.012$					
e. Lee	100.828	828.234	0.63	1.20	1.13
e. Frost	100.905	816.941	0.65	1.26	0.95
Kuan	101.024	1738.90	0.45	0.94	3.32
Gamma	101.071	832.590	0.62	1.2	0.91
MBD	106.331	1097.71	0.55	1.04	1.23
HMRF	100.91	720.26	0.23	1.24	0.86

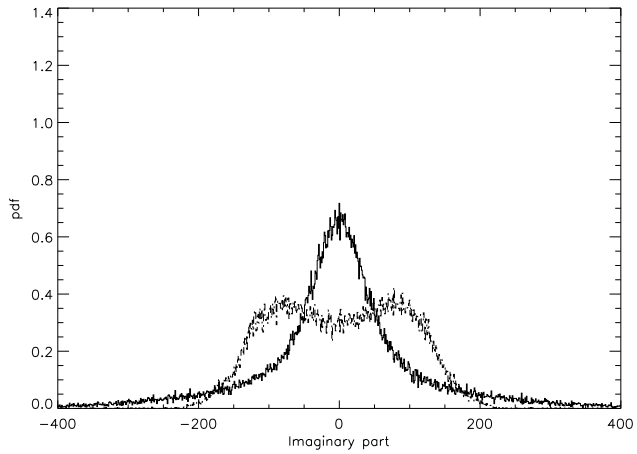
TABLE 12 – Measures for the comparison of despeckled Brodatz textures. From bottom left counterclockwise, with reference to Figure 14 : B1, B2, B3, B4.



(a)



(b)



(c)

FIGURE 17 – Histograms of amplitude images (a) : noisy data (continuous line), enhanced Lee filtered data (dotted line) and HMRF filtered data (dot-dashed line). Histograms of real part (b) and histograms of imaginary part (c) : noisy data (continuous line) and HMRF filtered data (dot-dashed line)

4.3.3 Optical image

The third experiment has been performed on simulated SAR data. The simulated image has been obtained by an optical QB image acquired over Maribor city, Slovenia, on 2nd of August 2007, 16 :21 :39 Coordinated Universal Time (UTC). The panchromatic band with a resolution of 0.61 m has been processed. The tile has been sub-sampled by a factor of 2 in order to obtain the desired resolution, $\sim$1.2 m. The final size of the image is 1024$\times$1024 pixels. It has been corrupted with uncorrelated speckle noise simulated according to Section 1.8 and ENL=1. The original image and the noisy one are shown in Figure 18 (a) and (b) respectively. The despeckled images are shown in Figures 19. The ratio images for visual comparison are shown in Figure 20. The measures for the quantitative filter comparison are presented in Table 13.
Comparing the despeckled image it is visible that the filters perform similarly. The enhanced Lee, enhanced Frost and the Gamma filter results are really similar. The Kuan filter does not remove the speckle as well as the others. The image despeckled with the MBD method appears really similar to the original while the image denoised by the proposed filter appears slightly blurred. On the other hand the blurring effect can be handled reducing the number of iterations. The ratio images confirm the results. The MBD exhibits reduced pattern compared with the other filters which perform similarly. The measurements show that the mean value is well preserved for the enhanced Lee, Kuan, Gamma and HMRF filters, while it is underestimated by the enhanced Frost and overestimated by the MBD filter. The proposed method has the second best MSE after the enhanced Lee filter, followed by the enhanced Frost, Gamma, MBD and Kuan which exhibits the worst value. The MBD shows the best SSIM index followed by the enhanced Lee, the enhanced Frost, Gamma, Kuan and HMRF. The mean value of the speckle image is better in case of MBD filter and Kuan filter. The enhanced Lee is in the third position followed by the proposed method, the enhanced Frost and the Gamma which perform similarly. The equivalent number of looks of the ratio image is better for the enhanced Lee, while the MBD is in the second position. The enhanced Frost, Gamma and HMRF filters underestimate the speckle noise since the number is lower than one. The Kuan filter performs worse compared with the others. The simulated SAR image has uniform areas where it is possible to estimate the equivalent number of looks of the despeckled image as a further measure of comparison. The higher ENL is obtained with the MBD and the proposed method while the Gamma, the enhanced Frost and enhanced Lee perform similarly. The lower ENL is obtained with the Kuan filter. The proposed method reaches the minimum MSE after five iterations.

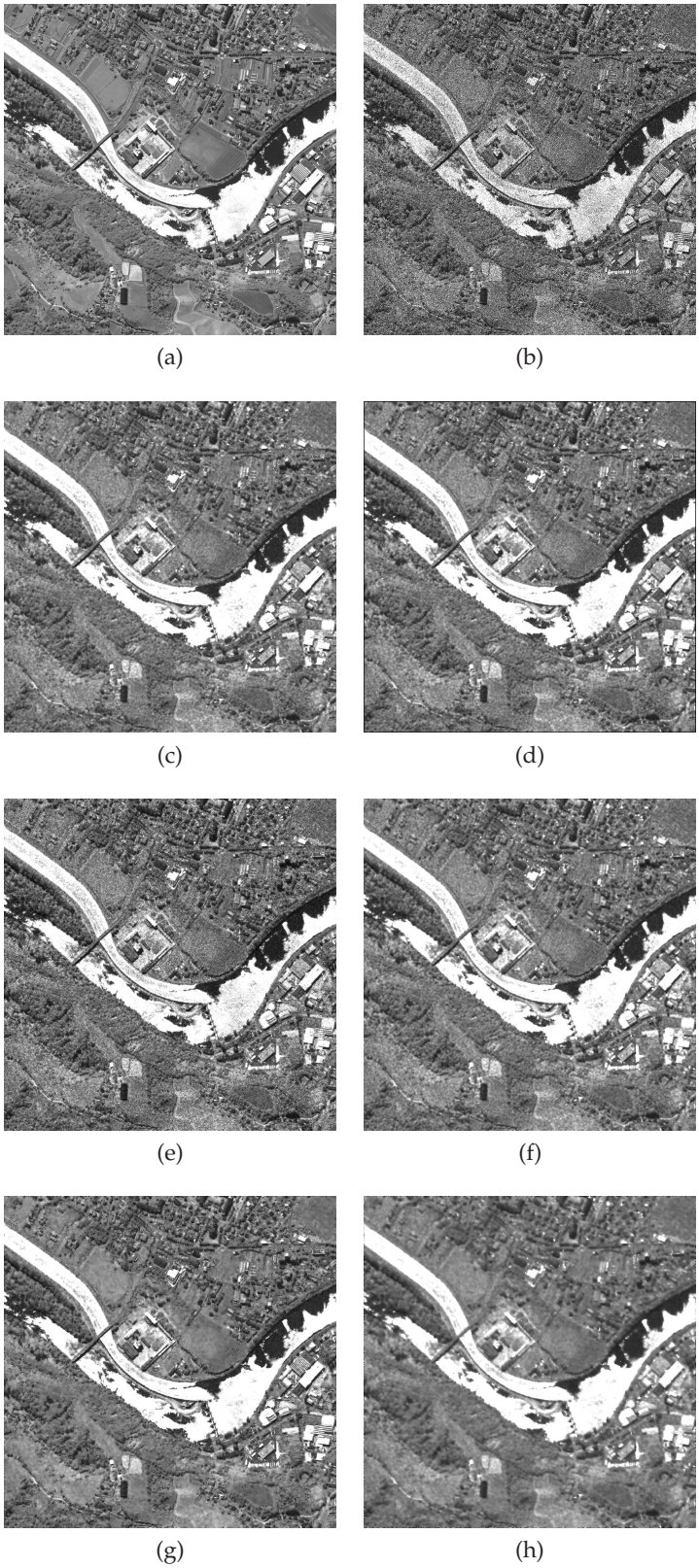
$\mu = 542.645$	E$\{\hat{\mathbf{x}}\}$	MSE	SSMI	E$\{\mathbf{y}/\hat{\mathbf{x}}\}$	ENL$\{\mathbf{y}/\hat{\mathbf{x}}\}$	ENL$\{\hat{\mathbf{x}}\}$
e. Lee	541.03	30019.5	0.879	1.18	1.10403	19.01
e. Frost	537.83	32212.1	0.875	1.27	0.850382	20.09
Kuan	542.75	50653.6	0.855	0.93	3.34667	3.09
Gamma	543.43	32945.3	0.872	1.28	0.771241	20.13
MBD	572.58	41241.5	0.892	1.04	1.18	45.74
HMRF	543.58	30564.1	0.848	1.27	0.66	45.62

TABLE 13 – Measures for the filter comparisons for the simulated SAR image.

(a)

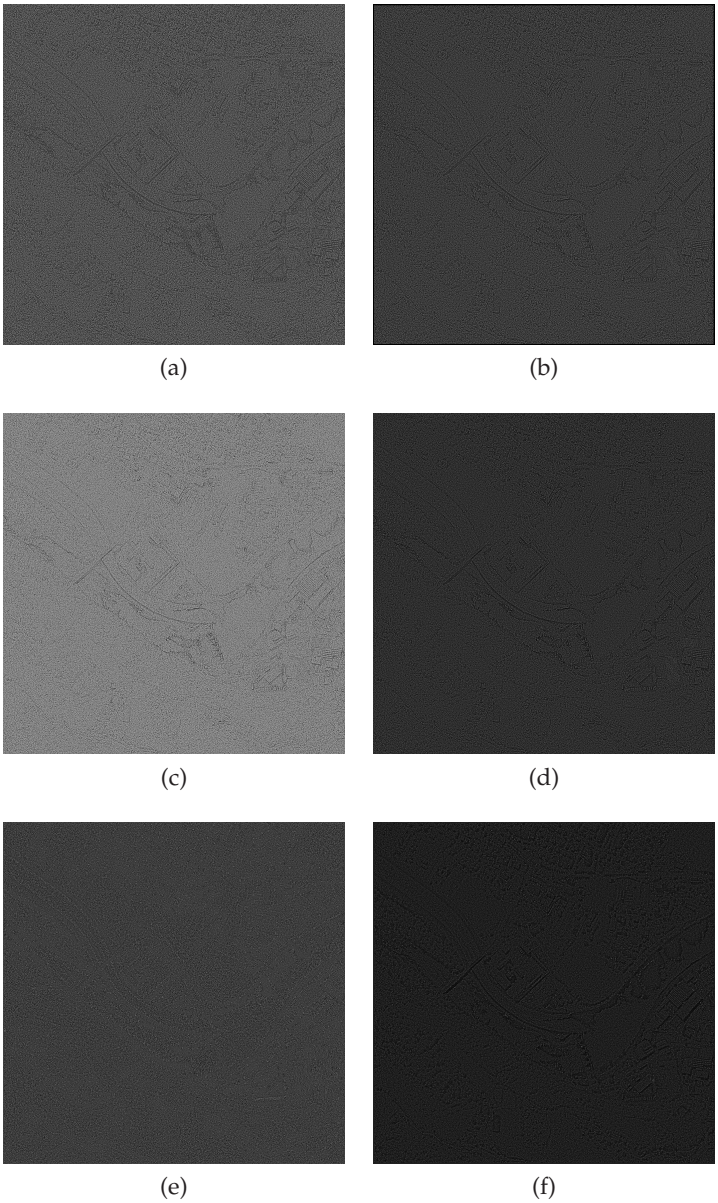
(b)

FIGURE 18 – Experiment with a synthetic SAR image, simulated from QB data. The size of the image is 1024×1024 pixels. Panchromatic band, resolution 0.61 m, Maribor, Slovenia, 2nd of August 2007, 16 :21 :39 UTC. Original image (a) and corrupted with speckle noise (b).



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 19 – Experiment with a synthetic SAR image, simulated from QB data. Original image (a), noisy image (b), despeckled with enhanced Lee (c), enhanced Frost (d), Kuan (e), Gamma (f), MBD (g) and the proposed method (h).



(a) (b) (c) (d) (e) (f)

FIGURE 20 – Experiment with a synthetic SAR image, simulated from QB data. Ratio image : enhanced Lee (a), enhanced Frost (b), Kuan (c), Gamma (d), MBD (e) and the proposed method (f).

4.3.4 SAR image

The last experiment is performed on actual SAR data. It has been performed on a TerraSAR-X HR SL mode image. It has a resolution of ∼1.5 m in range and ∼1.1 m in azimuth. The processed tile has a size of 1024× 1024 pixels and is shown in Figure 21. The same area of the optical image shown in Figure 18a has been selected. The despeckled images are shown in Figure 22. The ratio images are shown in Figure 23 and the measures for quantitative filter comparison are presented in Table 14. Differently from the simulated image the actual SAR image has correlated speckle. A zoomed area of the despeckled and ratio images is shown in Figures 24 and 25 for better visual comparison. In Figure 22 it is visible that the enhanced Lee, enhanced Frost and the Gamma filters perform similarly with exception of the Kuan which does not remove satisfactorily the noise. The MBD and the proposed method show a better removal of the noise. All the ratio images show patterns with a slightly inferior level for the MBD. The zoomed area shows better in detail the results compared with the original image. While in case of enhanced Lee, enhanced Frost and Gamma similar performances are confirmed as well as the poor performance of the Kuan filter, the MBD exhibits artifacts in the uniform area and along the edges. The proposed method shows the smoothest area but the blurred effect of the edges is also visible.
The comparative measurements are reduced to the ones that do not need the reference image which in this case is not available. Thus, it is possible to compare only the mean value of the ratio image, the ENL of the speckle noise and the ENL of the despeckled image. The best expectation of the ratio image is obtained by the MBD followed by the Kuan, the enhanced Lee, the enhanced Frost, the Gamma and the proposed method. The best equivalent number of looks of the speckle noise is given by the Gamma filter, the enhanced Frost, the HMRF, the enhanced Lee and the MBD. The Kuan filter highly overestimated it. The better ENL on the despeckled image was obtained with the proposed method after 5 iterations. The MBD is at second place. The enhanced Frost and Kuan perform similarly followed by the enhanced Lee. The Gamma filter is ranked in the last position.

	$E\{\mathbf{y}/\hat{\mathbf{x}}\}$	$ENL\{\mathbf{y}/\hat{\mathbf{x}}\}$	$ENL\{\hat{\mathbf{x}}\}$
e. Lee	1.17	1.26	7.91
e. Frost	1.21	1.08	8.07
Kuan	0.94	3.56	8.12
Gamma	1.22	0.99	2.60
MBD	1.02	1.52	17.16
HMRF	1.29	0.75	21.71

TABLE 14 – Measures for the filters comparison of the actual SAR image.

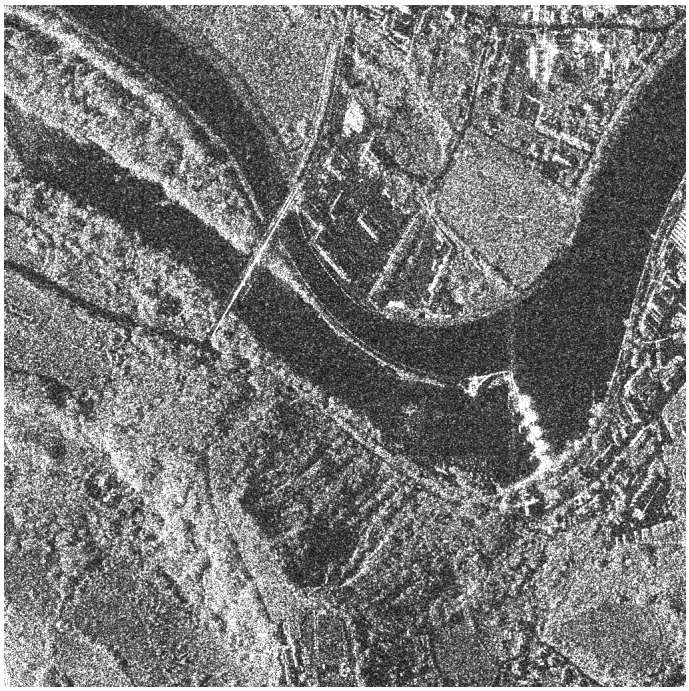
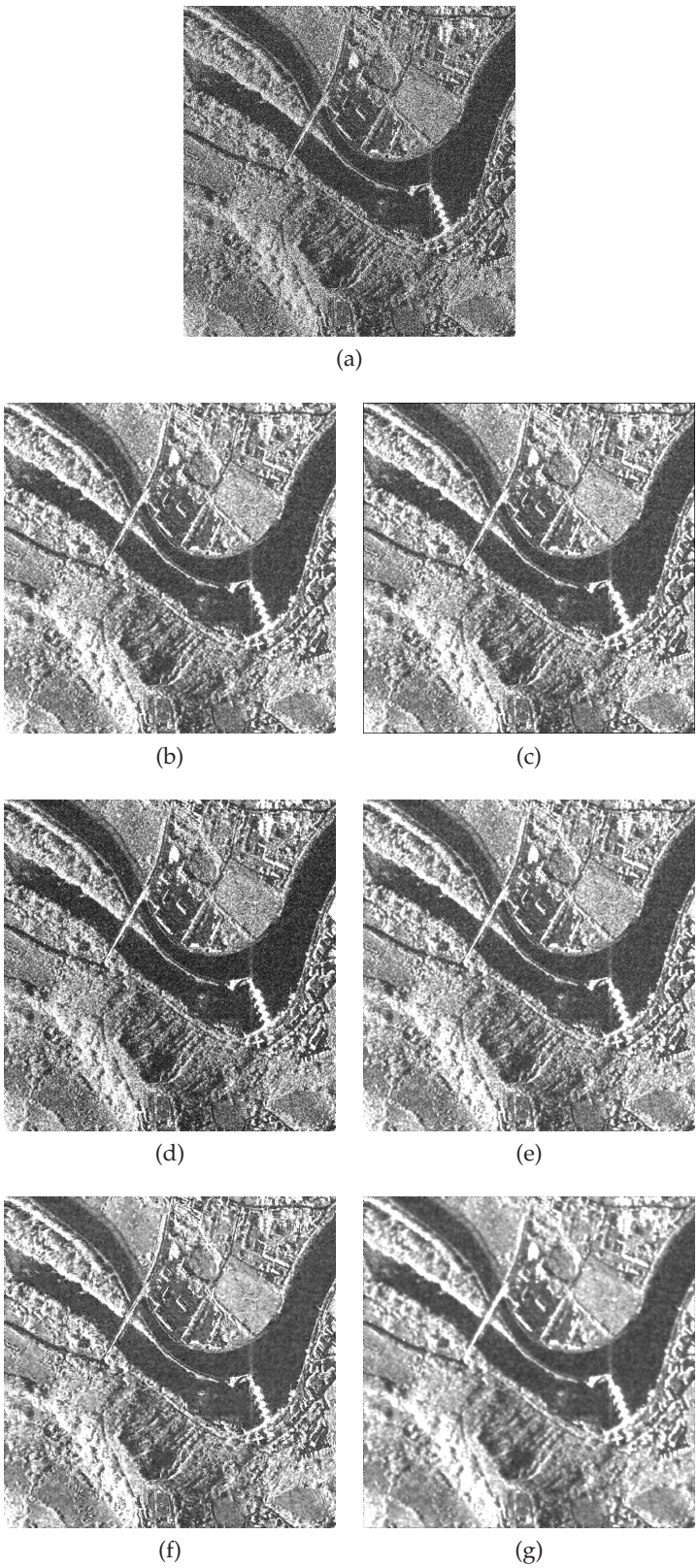
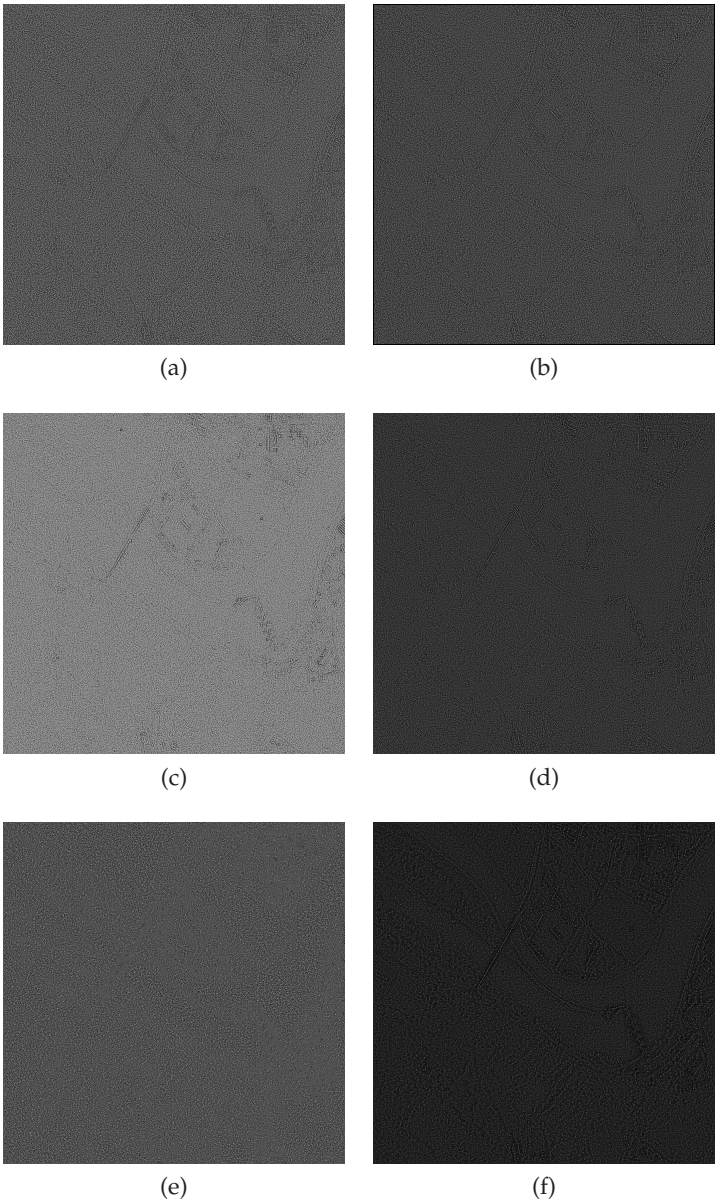


FIGURE 21 – Experiment with an actual SAR image. TerraSAR-X HR SL mode. Polarization HH, descending orbit, ground range resolution ∼1.5 m, azimuth resolution ∼1.1 m. Maribor, Slovenia, 29th of October 2008, 5 :01 :06 UTC.



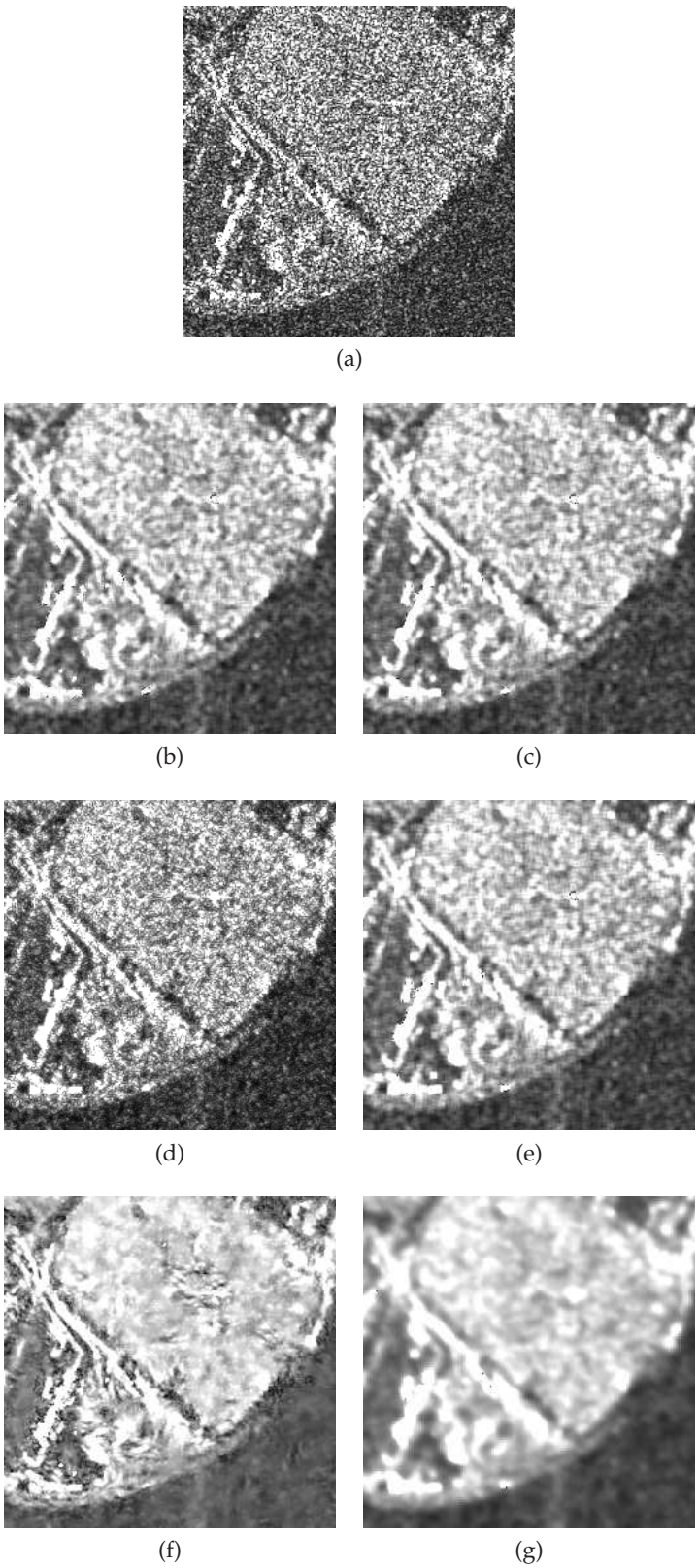
(a) (b) (c) (d) (e) (f) (g)

FIGURE 22 – Experiment with an actual SAR image. Original image (a), despeckled with enhanced Lee (b), enhanced Frost (c), Kuan (d), Gamma (e), MBD (f) and the proposed method (g).



(a) (b) (c) (d) (e) (f)

FIGURE 23 – Experiment with an actual SAR image. Ratio image : enhanced Lee (a), enhanced Frost (b), Kuan (c), Gamma (d), MBD (e) and the proposed method (f).



(a) (b) (c) (d) (e) (f) (g)

FIGURE 24 – Experiment with an actual SAR image. Zoom images. Original image (a), despeckled with enhanced Lee (b), enhanced Frost (c), Kuan (d), Gamma (e), MBD (f) and the proposed method (g).

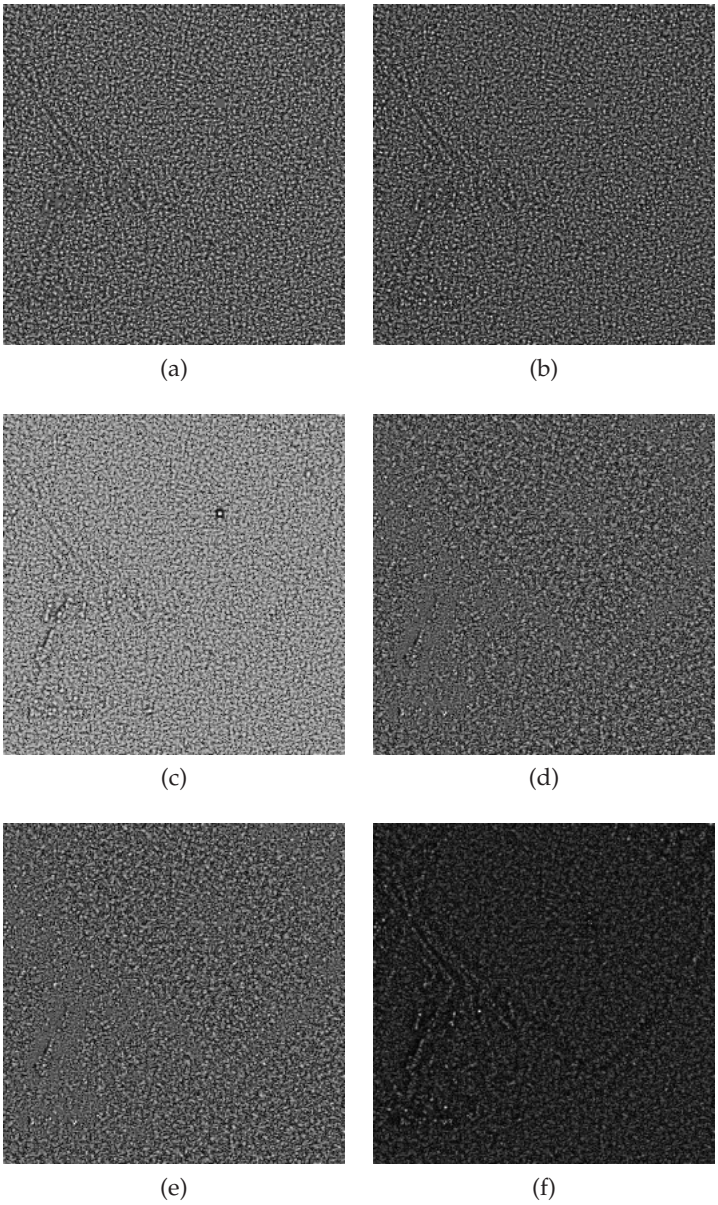


FIGURE 25 – Experiment with actual an SAR image. Zoom of the ratio images : enhanced Lee (a), enhanced Frost (b), Kuan (c), Gamma (d), MBD (e) and the proposed method (f).

4.4 Observations and Discussion

The optimal parameter estimation is successfully performed by the Rate Distortion curve because of the connection with the Bayesian frame model selection. The method is a global approach that may be useful in data mining applications for content interpretation. The optimization of the Analyzing Window size and the Model Order are useful for optimal feature extraction in order to develop Image Information Mining System.
The statistical regularization allows to model the complex signal by parameter estimation and to smooth the real and imaginary channels. It may be useful in 3D signal reconstruction, e.g. in tomographic applications.
The complex-valued GMRF data model may model the complex signal by estimating the complex model parameters. It may be applied for target recognition because of its ability to model phase patterns, but it does not directly apply to stationary textured areas.
The Tikhonov regularization outperforms the other filters in case of MSE measure and has comparable results with the other measures and by visual inspection of the despeckled image and ratio image. On the other hand, a prior model which fits better the synthetic generated texture might improve the results. The GMRF model is planned to be used as a prior in future experiments.
In case of the experiment with Brodatz texture, the proposed method does not outperform the other filters but behaves satisfactorily with the actual texture.
The simulated SAR image is well despeckled by the proposed method and all the measures are consistent.
The best results on the actual SAR data are reached with the proposed method in term of ENL.
The complexity of the methods, which are all programmed in C++, have been measured with the execution time on a machine with an Intel core II 2.0 GHz processor. For all adaptive filters a window a size of 5×5 pixels was used. For MBD a window of 41×41 pixels was adopted because it was demonstrated to be the optimal average AW size. The SAR image, shown in Figure 4.3.4 with 1024×1024 pixels was used for this purpose. The MBD method runs for 250 s, enhanced Lee for 1.5 s, enhanced Frost 1.8 s, Kuan for 1.1 s, Gamma for 1.3 s and the proposed method for 18 s. The most computationally demanding method is the MBD, because it estimates parameters of the GMRF.

4.5 Application : Classification

The extracted features have been validated with different methods and tools. An unsupervised classification of the GMRF model parameters is compared with MBD parameters classification. The image despeckled with the proposed method is classified using a Support Vector Machine (SVM) where the training data was selected using a Semantic Search Engine (SSE) tool.

4.5.1 Unsupervised K-means Classification

The Generalized Lloyd Algorithm (GLA), also known as *K-means* in clustering literature, is the generalization in a multidimensional space of the Lloyd Algorithm (LA) for designing a scalar quantizer (Gersho & Gray, 1991).
The algorithm is based on the iterative use of the code book modification operation and it is based on the following steps :

1. Begin with an initial code book $\mathcal{C}_m = \{\boldsymbol{y}_i; i = 1, \ldots, N\}$.
2. Find for code book $\mathcal{C}_m = \{\boldsymbol{y}_i; i = 1, \ldots, N\}$ the optimal partition into quantization cells to form the nearest neighbor cells (Nearest Neighbor Condition) :

$$R_i = \{\boldsymbol{x} : d(\boldsymbol{x}, \boldsymbol{y}_i) < d(\boldsymbol{x}, \boldsymbol{y}_j);\ \text{all } j \neq i\} \tag{4.7}$$

   if $d(\boldsymbol{x}, \boldsymbol{y}_i) = d(\boldsymbol{x}, \boldsymbol{y}_j)$ for one or more $j = i$, then assign $\boldsymbol{x}$ to the set R_j for which j is smallest.
3. Find $\mathcal{C}_{m+1} = \{cent(R_i); i = 1, \ldots, N\}$, the optimal reproduction alphabet (code book) for the cells just found. Where $cent(\cdot)$ is the center of the cell (Centroid Condition).
4. Compute the average distortion for $\mathcal{C}_{m+1}$. If it has changed by a small enough amount since the last iteration, stop. Otherwise set $m + 1 = m$ and go to Step 2.

Each application of the steps two and three (Lloyd Iteration) must reduce or leave unchanged the average distortion.
The stopping criteria usually are :

- the maximum number of iterations ;
- the center positions do not change (or the distortion does not decrease) significantly from the current iteration to the next one ;
- the distortion error, $\varepsilon = |D_{i+1} - D_i|$, is less or equal to the fixed threshold.

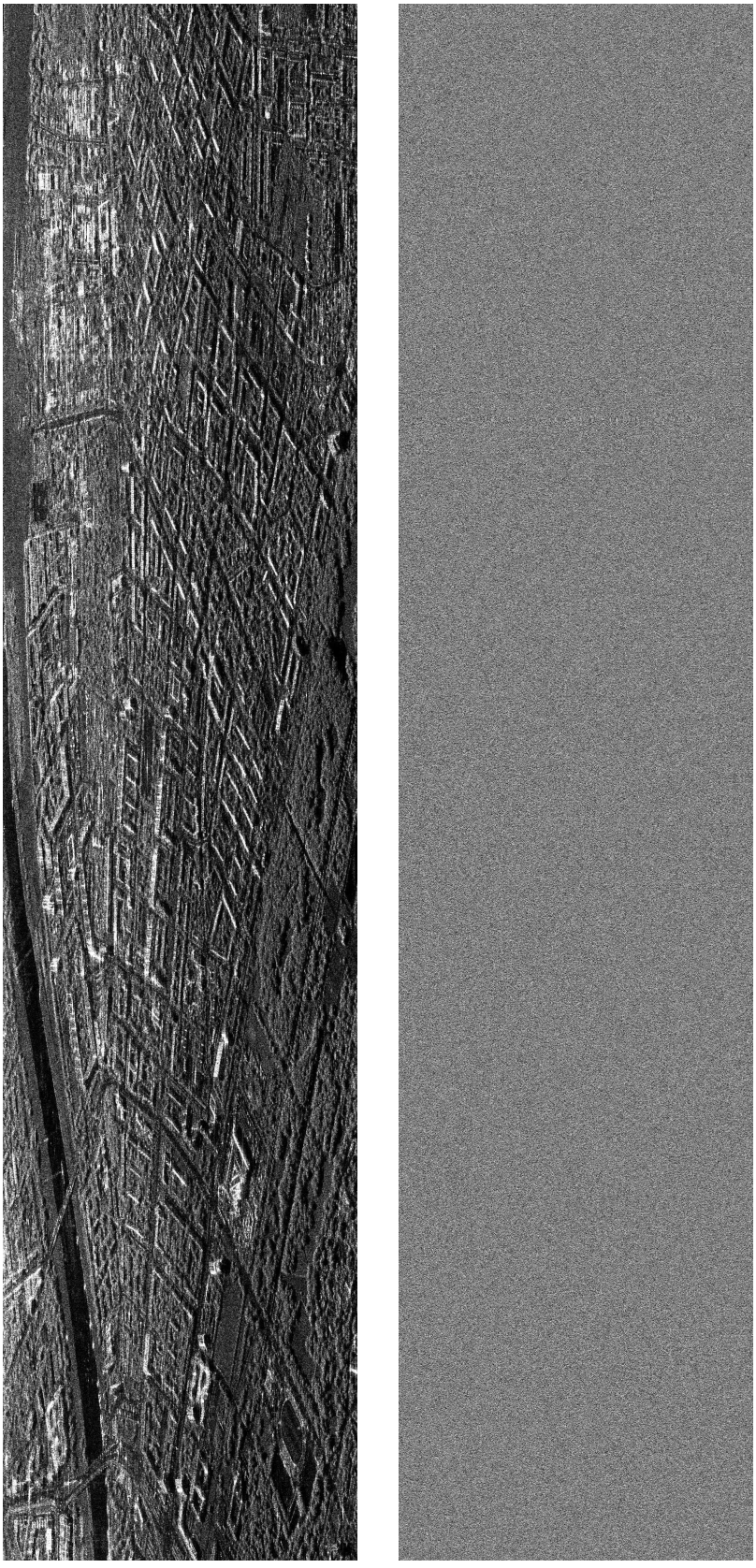
They can be used singularly or in combination in order to obtain the desired interruption mode.
The ESAR-X airborne scene acquired over the city of Dresden, shown in Figure 26a, has been analyzed with the model in Section 3.3.2. The model features have been extracted and then they have been classified by the k-means unsupervised classifier.
The amplitude and the phase of the scene are shown in Figure 26, while the classification results, with the complex-valued GMRF and the MBD algorithm, are shown in Figures 27a and 27b, respectively.
According to the content of the image five classes have been chosen for the classification :

- black : water, shadows and dark areas ;
- blue : vegetation ;
- green : residential area ;
- orange : buildings ;
- yellow : very strong scatterers.

The first comparison of Figures 27a and 27b shows that the resolution of the classification done with the complex-valued GMRF is improved. The confusion matrices for the classification with GMRF and with MBD are shown in Tables 15 and 16, respectively. The classes for MBD case are not well separated : the water and the dark areas are classified as vegetation. The latter is not separated from the residential area. The performance for buildings and strong scatterers are poor. The confusion matrix for the GMRF model gives better results especially for buildings and strong scatterers. Water is also much better separated while vegetation, differently from the MBD case, is merged with the residential area.

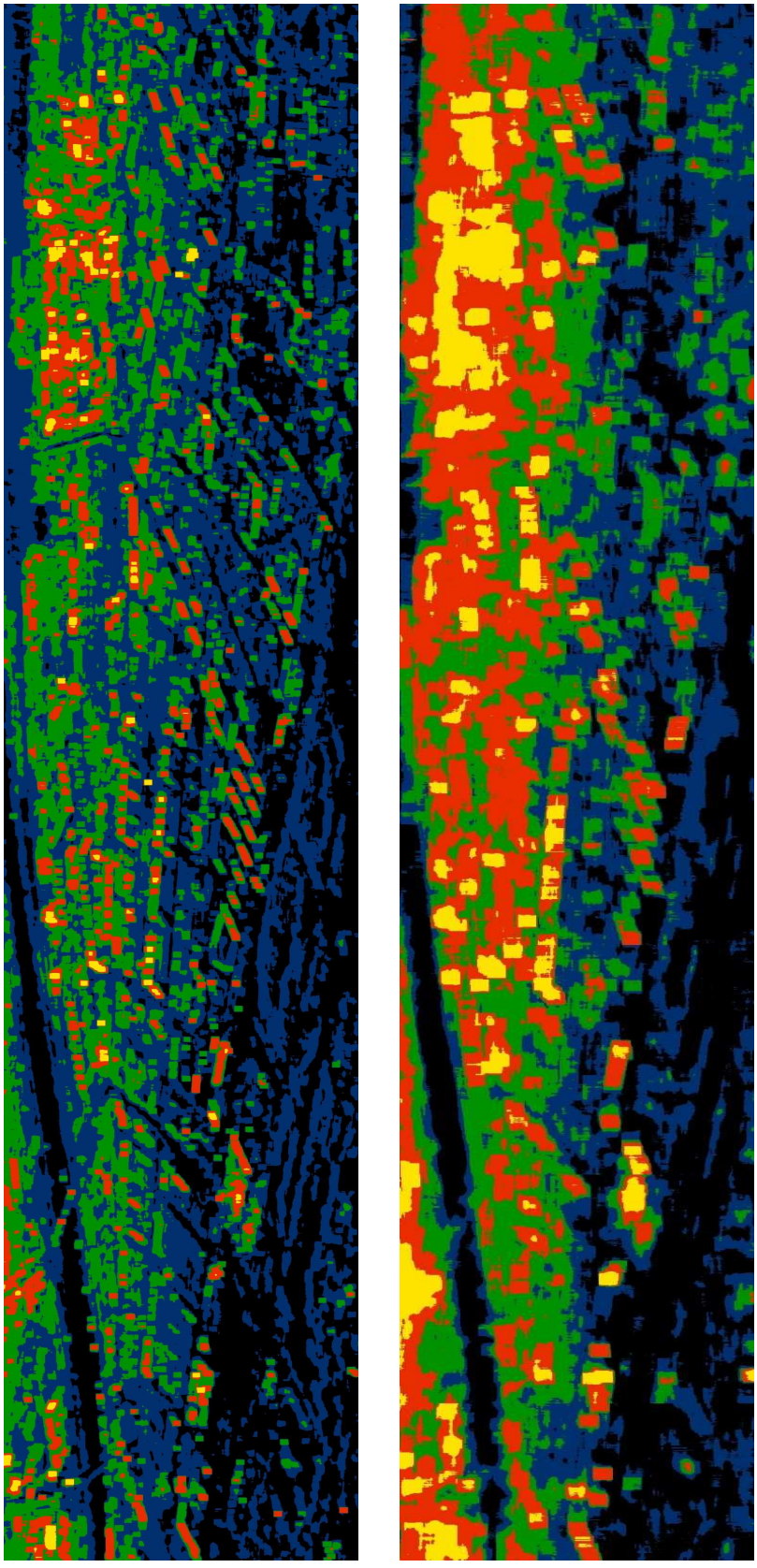
4.5.2 Supervised Support Vector Machine (SVM) Classification

The despeckled image with the Tikhonov regularization algorithm has been included in a Semantic Search Engine (SSE) (Costache & Datcu, 2006; Costache et al., 2006;



(a) (b)

FIGURE 26 – ESAR-X scene acquired on the city of Dresden : (a) amplitude and (b) phase.



(a) (b)

FIGURE 27 – K-means classification results for five classes : black (water), blue (vegetation), green (residential area), orange (building) and yellow (strong scatterers).

%	Water	Vegetation	Residential area	Buildings	Strong scatterers
Water	67	28	5	-	-
Vegetation	29	63	8	-	-
Residential area	32	9	59	-	-
Buildings	11	11	-	78	-
Strong scatterers	-	-	9	18	73

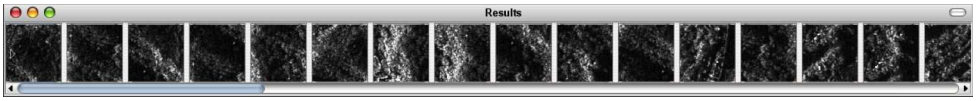
TABLE 15 – K-means unsupervised classification confusion matrix for complex-valued GMRF.

%	Water	Vegetation	Residential area	Buildings	Strong scatterers
Water	28	61	11	-	-
Vegetation	27	73	-	-	-
Residential area	18	41	41	-	-
Buildings	24	29	12	35	-
Strong scatterers	16	11	11	17	44

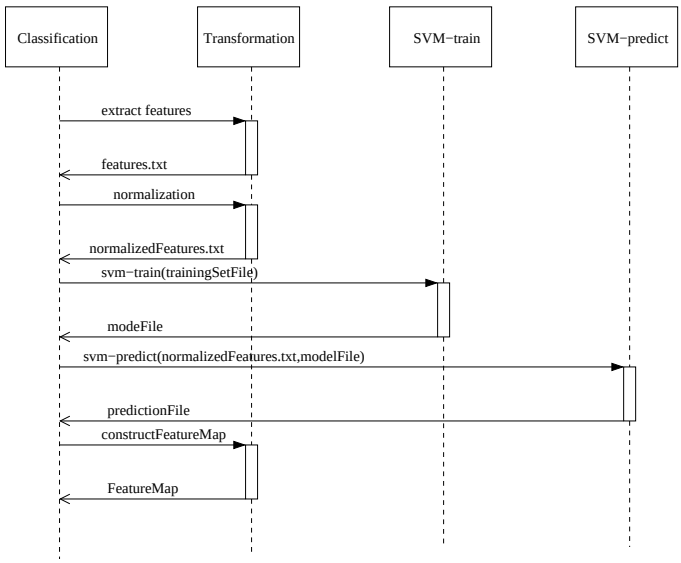
TABLE 16 – K-means unsupervised classification confusion matrix for MBD.

Costache & Datcu, 2007). The used image belongs to a SLC TerraSAR-X scene acquired over the Chinese region of Sichuan affected by the earthquake in May 2008, resolution 1.6m in azimuth and range.
In order to validate the algorithm, both the original and the despeckled image have been separately ingested and classified by a SSE. It is an interactive training/classification tool which can be used for supervised classifications using a Support Vector Machine (SVM). The diagram of the classification is shown in Figure 28c. The tool has a Graphic User Interface (GUI) interface shown in Figures 28a and 28b.
The results of the classification of buildings, fields and roads are shown in Figures 29 for the original and despeckled image. The precision of the classifications is presented in Table 17 for each class.
The classification of the building, shown in Figures 29c and 29d, is improved after despeckling, as shown in Table 17b and 17c. It goes from 27% to 31% while the false negative are decreasing from 5% to 1%.
The improvement of the results for the class fields is similar to the one of the class buildings. It goes from 34% to 49% and the false negative are decreasing from 26% to 11%, as shown in Figures 29e and 29f, and in Tables 17d and 17e.
The classification of the roads is better after despeckling but the false positive percentage is worse. The class roads is merged with fields and buildings as visible in Figure 29h and differently from Figure 29g. This depends on the fact that after despeckling the shadow is much more similar to the streets. On the other hand, in order to extract roads some dedicated software or line detectors are recommended.
The classification with the SSE tool is globally improved after denoising, a significant amelioration depends on the despeckling algorithm but also on the parameters included in the classification. In the specific case the tool uses mean, variance and total variation, but a variable combination of the parameters might be used in order to adapt to the selected class.

(a)



(b)



(c)

FIGURE 28 – SSE tool : GUI (a) and (b) SSE GUI, (c) classification diagram.

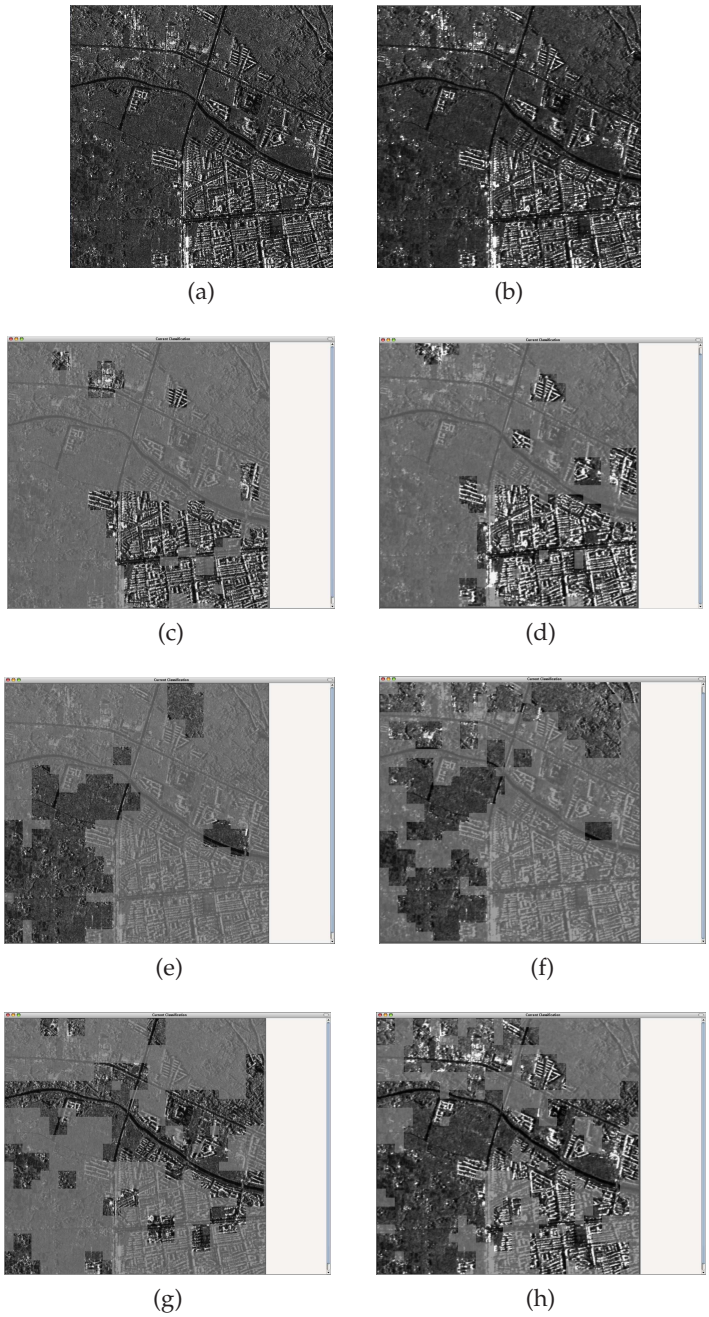


FIGURE 29 – Original image (a), despeckled image (b). Classification results for buildings, fields and roads, second, third and fourth row, respectively, for the original image and for the despeckled image left column and right column respectively.

Confusion matrix

True Positive	False Positive
False Negative	True Negative

(a)

Building
original image

27	1
5	67

(b)

Building
despeckled image

31	1
1	67

(c)

Fields
original image

34	1
26	39

(d)

Fields
despeckled image

49	2
11	38

(e)

Roads
original image

7	21
3	69

(f)

Roads
despeckled image

9	52
1	38

(g)

TABLE 17 – Confusion matrices in % for the SSE classification for the original image (left column) and the despeckled image (right column). First line : building classification (b) and (c) ; second line : fields classification (d) and (e) ; third line : roads classification (f) and (g) .

4.6 Conclusions

The Rate Distortion has been applied on the existing MBD algorithm for despeckling and feature extraction. Since the algorithm is implemented in an IIM system the problem of optimal feature extraction arises. It has to be approached taking into account the large amount of data to be handled by the system, thus a global, simple and fast method is needed. The solution was found in the context of information theory and processing measuring the error in coding a source of data.
The lossy data compression is applied by dyadic k-means in the hypothesis of a mixture of Gaussian source distributions. The optimal average size of the Analyzig Window (AW) allows to have a robust parameter estimation for image feature characterization. The optimal average Model Order (MO) permits to avoid the selection of a model which over parametrized the data.
On the other hand, the result depends on the data diversity and the number of classes coded by the clusters. The optimization is performed in the hypothesis of stationary data inside the analyzing window. This hypothesis is not always respected and to overcome this problem an adaptive Analyzing Window might be used.
The result on MO selection is compared to the local map of model order selected by the Fisher information. The map shows how the MO is related to the image content. Uniform areas or areas which show low complexity are better represented by a low model order, while textured areas and areas which show high variations are better represented by a high model order. The global Model Selection performed by Rate Distortion appears to be a feasible solution because the local adaptive MO selection is computationally demanding.
In Section 4.2, the analysis of complex-valued data starts with the extension in the complex domain of the MAP estimate for the statistical regularization of the real part and imaginary part. The linear image model is regularized under the constrain of the a priori probability term. The latter is given by a Gibbs distribution, which is an exponential distribution whose exponent, the so-called energy function, characterizes the different MRF families. The parameter of the prior are estimated from the incomplete data by an EM procedure, thus the parameter images are used to find the solution of the ill-posed problem of estimating the image from the noisy data.
The GMRF, characterized by a quadratic energy function, is able to describe the local characteristics of the images. Although the GMRF are not the best choice as far as it concerns the reconstruction of profiles with sharp discontinuities, they exhibit appealing properties because the posterior is convex and Gaussian. Actually, it is possible to analytically compute the gradient and to minimize the functional without falling in local minima. The estimated parameter image appears really similar because the real and imaginary channels are uncorrelated but not independent. The parameters give a measure of the variation occurred in the signal, thus they might be used to label the image content. The amplitude image appears blurred, but no artifact is generated by the filter. The chosen GMRF model family is isotropic, others models, characterized not by a scalar but a parameter vector are able to estimate directional parameters which can capture structures at different scale and orientation.
This is the case of the complex-valued GMRF model developed by extending the classical definition of GMRF to complex domain. The estimated parameters are able to model different patterns in the complex image phase. In single SAR image, these patterns are mainly visible in the vicinity of strong scatterers due to the system impulse response.

We suppose to be able to characterize the local texture by modelling the phases pattern but the hypothesis was not confirmed. Actually, the pattern are really seldom and not evident and the values of parameters result to be similar in case of different texture. By modelling the data with the complex-valued GMRF the model variance may distinguish texture with different intensities.
In Section 4.3, describes the Tikhonov-like method for image restoration and feature enhancement. The linear image model allows to include the system impulse response and the prior constrain allow to deal with uniform areas and preservation of edges by the Total Variation and the Huber-Markov function respectively. The convex functional ensures that the solution exists and is unique and it has been minimized by the iterative Newton's method.
The filter parameters allow an optimal tuning of the optimization. In the beginning they have been chosen experimentally finding out that they control the degree of efficiency of the prior models, i.e. the degree of smoothness vs. the degree of edges preservation. The problem has been solved as a problem of model selection by the Bayesian evidence framework. The norm parameter controls the shape of the weighting function, e.g. Gaussian or Laplacian distribution. It has to be initialized experimentally in the first iteration and then it is estimated. The numerical step, which controls the speed of convergence and the precision of the solution, has also to be selected experimentally.
Four experiments have been performed in order to give as much as possible complete results. The filter shows superior performance for denoising synthetic texture and the actual SAR data. In case of the Brodatz texture and the simulated SAR image the filter performance are similar to the enhanced Lee and MBD which also show good performances.
The novelty of the approach relies on the possibility to include the system transfer function, the prior information model and the use of the full resolution/information contained in the complex valued data.
Finally, in Section 4.5 the application of the proposed methods is presented by unsupervised and supervised classification of the models. The k-means classification is compared with the MBD model. The supervised classification is performed to quantify the impact of despeckling in a SSE tool.
Although the results are far to be optimal the proposed methods appear to be promising. An improvement of the results might be obtained by data preprocessing in order to extract edges information for better classes separation and/or combining differently the parameters. Furthermore, other parametric models might be considered.

4.7 Summary of the chapter

In this chapter the experimental results are shown and discussed. The Rate Distortion-based model selection is demonstrated to be a successful method to estimate the optimal average analyzing window and the optimal average model order in the frame of MBD feature extraction. The modeling of the complex images started with the results on complex-valued statistical regularization and the complex-valued GMRF for data modeling. The first method allows to estimate parameters for the restoration of the real and imaginary part but it does not allow to remove speckle. The second method permits to model the image with the complex linear parametric model but it is not inserted in a full MAP estimator for the despeckled image model inversion. The Tikhonov-like regu-

larization with the Huber-Markov prior allows the estimation of the uncorrupted image. The approach is compared with the most known adaptive filters but in contrast to them it works on complex-valued data. The method embeds the SAR system and if needed it may include also the image formation model. Thus, it allows different application and the modeling of a variety of information with the use of different priors. An application of the models is presented by the use of supervised and unsupervised classifications.

Conclusion

The thesis presents a novel model selection framework and complex-valued image analysis with application on remote sensing SAR data.
The model selection method is based on Rate Distortion theory. In order to describe the method an introduction on information theory and Bayesian model selection is presented. The first allows to introduce basic concepts, e.g. entropy, mutual information, distortion measure necessary for presenting the Rate Distortion theory. The Bayesian inference, presented in the framework of estimation theory, allows model selection by the Occam factor. It is embedded in the Bayes formula and quantifies how well the model fits the data. Comparing Occam factors of different models the best model may be selected according with the principle that the simplest explanation or strategy, i.e. model, tends to be the best one. The entropy is inversely proportional to the Occam factor. Since the Rate Disotrion is defined as the lower bound of the mutual information, choosing the model which minimizes the distortion means choosing the model which maximizes the Occam factor. These concepts are presented in Chapter 2, and in particular in Section 2.8, where the connection between the Occam factor and the mutual information is drawn.
The Rate Distortion theory is the theory which deals with the *goodness* of a data *source* distribution representation. The basic problem in Rate Distortion theory can then be stated as follows : given a source distribution and a distortion measure, what is the minimum expected distortion achievable at a particular rate. Equivalently, the problem can be reformulated as follow : what is the minimum rate description required to achieve a particular distortion. The Rate Distortion provides the theoretical fundamentals for providing an answer to the previous question. It finds applications on several fields which span from multimedia and streaming up to vector quantization, clustering and more in general image processing. The feature extraction estimation problem has been recast as a coding problem in the point of view of information theory. The estimation of the optimal average Analyzig Window (AW) size and the optimal average Model Order (MO) has been introduced and gathered as a model selection problem. The results of the estimation are presented in Section 4.1. The advantage of the method relies in its simplicity since it is a global approach and that it can be successfully applied to information mining problems where we have to deal with large amounts of data. The disadvantage is that it cannot be applied in case of adaptive AW or local problems, e.g. local model order selection, where an alternative method has to be used, e.g. evidence framework, Fisher information.
The complex-valued image analysis starts with an overview of the statistical models of SAR data, presented in Section 1.5. The data transformation from Cartesian to polar coordinates is presented as well as the logarithm transformation which is useful to transform the multiplicative noise in additive noise and the linear transformation which can be useful to model the image formation process. In the thesis the linear model is chosen because

of its capability to include the SAR end-to-end-system. Furthermore the statistics of the complex-valued data and the amplitude and phase, also the statistics in case of strong scatterers, low fluctuation of the σ and multilook data have been presented. Beside the multilook method, the spatial averaging technique is described in order to estimate the mean reflectivity. The approximated multiplicative models for speckle reduction are presented together with the most known families of despeckling filters. They are compared to the developed method for image noise reduction in complex-valued domain, although the filters work on detected images, i.e. real-valued domain.
Advanced statistical modeling is presented in Section 3.3 where the GMRF model is extended to the complex-valued domain. The synthesis and the analysis of the complex-valued GMRF is accompanied by the retrieval and forward modeling from actual SAR data. The complex-valued GMRF is used as data model and also as prior model in a full Bayesian estimate to denoise the real channel and the imaginary of the complex-valued image. In the latter case the adopted image model is the linear model. The approach deals with additive noise and allows to estimated isotropic parameters proportional to the gradient and then to use them to regularize the signal. The optimization is implemented with a steepest descent iterative algorithm which is appropriate because the function to be optimized has a regular shape, i.e. it is convex.
The regularization of the real channel and the imaginary channel did not lead to a characterization of the image content through the model parameter because of the randomness of the phase and neither to a removal of the speckle because it is modeled as multiplicative noise. Thus, another approach is required in order to denoise the image and to extract feature parameters. The proposed method is the Tikhonov regularization which is widely used in literature for image restoration and it can be related to the Bayesian framework in its unconstrained form. The connection to the Bayesian inference allows to select prior models and to solve the problem by a MAP estimate. The chosen prior is the Huber-Markov function which belongs to the Markov Random Field family. It has been selected because of its property to preserve non-linearity in the restoration of optical images. The function to optimize has a regular shape and can be iteratively solved through the same method adopted for the complex-valued GMRF MAP, i.e. steepest descent algorithms. The function is characterized by several parameters which have to be tuned. In the beginning the tuning has been done experimentally but then the estimate of the optimal parameters by the evidence framework has been adopted. The theoretical models and the Tikhonov optimization are presented in detail in Section 3.4 while the accurate comparison of the filters is presented in Section 4.3. The results are presented on actual TerraSAR-X data and on simulated images for a more complete description of the filter performances.

Perspective and Future work

The Rate Distortion model selection is demonstrated to be a powerful method which can be applied in cases where the estimation problem can be recast as a model selection problem. The method may be compared to classical model selection methods, i.e. Minimum Description Length (MDL), Bayesian Information Criterion (BIC). On the other hand, its usage has to be preferred in case of a global approach where we have to deal with large databases, e.g. image information mining applications.

Moreover, the theoretical statements linking the model selection methods need further analysis, because they are not always consistent.
Since the speckle is dominating the image content, its removal allows a better definition of the image content at the price of losing detail. The future investigation is focused on the application of the presented Tikhonov regularization in interferometric applications and in the inclusion of parametric models, e.g. GMRF, to enable texture feature extraction.
Although the model of the complex-valued data is simple, the characterization of the information content of the image remains a difficult task. Apart from a signal based approach, the analysis of the topology and the structures has to be included in order to introduce geometric and spatial models. An alternative method is also a multi-scale analysis. The combination of different layers can be used to create a space to characterize the information content of the image.
The analyses of the complex-valued data is usually used in image pairs, i.e. interferometric applications, target analysis, multi-looking, speed retrieval etc. The extension of the analysis of the complex signal to a single image is important to analyze the behavior of the non-stationary signal for modeling and for further application in the previous mentioned fields.

Appendix

Proof of the Cramér-Rao inequality

The proof of 2.45 can be demonstrated by the Cauchy-Schwartz inequality

$$\Big[\mathrm{E}\Big\{\Big(\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)-\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\Big\}\Big)(\hat{\theta}-\mathrm{E}\{\hat{\theta}\})\Big\}\Big]^2 \leq \\ \mathrm{E}\Big\{\Big[\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)-\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\Big\}\Big]^2\Big\}\cdot\mathrm{E}\Big\{[\hat{\theta}-\mathrm{E}\{\hat{\theta}\}]^2\Big\} \tag{4.8}$$

the expectation of $\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)$ which appears in the first and in the second term of the inequality is

$$\begin{aligned}
\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\Big\} &= \int\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.9)\\
&= \int\frac{\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta)}p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.10)\\
&= \int\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.11)\\
&= \frac{\partial}{\partial\theta}\int p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.12)\\
&= \frac{\partial}{\partial\theta}1 && (4.13)\\
&= 0. && (4.14)
\end{aligned}$$

It allows to rewrite the first term in 4.8 as follows

$$\mathrm{E}\Big\{\Big(\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)-\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\Big\}\Big)(\hat{\theta}-\mathrm{E}\{\hat{\theta}\})\Big\}=\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\hat{\theta}\Big\}. \tag{4.15}$$

Thus, with the use of 4.15 and 2.40, 4.8 can be rewritten

$$\Big[\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\hat{\theta}\Big\}\Big]^2 \leq nI(\theta)\sigma_{\hat{\theta}}^2. \tag{4.16}$$

It remains to demonstrate that the first term of the inequality is equal to 1, and this can be done as follows

$$\begin{aligned}
\mathrm{E}\Big\{\frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\hat{\theta}\Big\} &= \int \frac{\partial}{\partial\theta}\ln p(\mathbf{x}|\theta)\hat{\theta}p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.17)\\
&= \int \frac{\frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta)}\hat{\theta}p(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} && (4.18)\\
&= \int \frac{\partial}{\partial\theta}p(\mathbf{x}|\theta)\hat{\theta}\mathrm{d}\mathbf{x} && (4.19)\\
&= \frac{\partial}{\partial\theta}\int p(\mathbf{x}|\theta)\hat{\theta}\mathrm{d}\mathbf{x} && (4.20)\\
&= \frac{\partial}{\partial\theta}\mathrm{E}\{\hat{\theta}\} && (4.21)\\
&= \frac{\partial}{\partial\theta}\theta && (4.22)\\
&= 1 && (4.23)
\end{aligned}$$

where the step from 4.22 to 4.23 is obtained reminding that $\hat{\theta}$ is an unbiased estimator. In Equations 4.13 and 4.21 the operators derivative and integral have been exchanged in the hypothesis that the function respects the Laplace bounded convergence theorem. Thus, the Cramér-Rao inequality 2.45 is demonstrated.

Equivalence of Fisher Information Expressions

The reason why the elements of the Fisher information matrix have the form presented in 4.24 can be better understood by writing the Fisher information as follows

$$I(\theta) = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta^2}\ln p(\mathrm{x}|\theta)\right\}. \tag{4.24}$$

The two expressions 4.24 and 2.38 are equivalent. The demonstration follows

$$\Big(\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\Big)\cdot p(\mathrm{x}|\theta) = \frac{\partial}{\partial\theta}p(\mathrm{x}|\theta). \tag{4.25}$$

By calculating the derivative with respect to θ of 4.25 it is obtained

$$\frac{\partial^2}{\partial\theta^2}p(\mathrm{x}|\theta) = \Big(\frac{\partial^2}{\partial\theta^2}\ln p(\mathrm{x}|\theta)\Big)\cdot p(\mathrm{x}|\theta) + \Big(\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\Big)\cdot\frac{\partial}{\partial\theta}p(\mathrm{x}|\theta) \tag{4.26}$$

Replacing 4.25 in 4.26 it is

$$\frac{\partial^2}{\partial\theta^2}p(\mathrm{x}|\theta) = \Big(\frac{\partial^2}{\partial\theta^2}\ln p(\mathrm{x}|\theta)\Big)\cdot p(\mathrm{x}|\theta) + \Big(\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\Big)^2\cdot p(\mathrm{x}|\theta) \tag{4.27}$$

Integrating 4.27 with respect to x

$$\int\frac{\partial^2}{\partial\theta^2}p(\mathrm{x}|\theta)\mathrm{dx} = \mathrm{E}\Big\{\frac{\partial^2}{\partial\theta^2}\ln p(\mathrm{x}|\theta)\Big\} + \mathrm{E}\Big\{\Big(\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\Big)^2\Big\}. \tag{4.28}$$

The integral on the left side of 4.28 is equal to zero because $\int p(\mathrm{x}|\theta)\mathrm{dx} = 1$, thus

$$\mathrm{E}\Big\{\Big(\frac{\partial}{\partial\theta}\ln p(\mathrm{x}|\theta)\Big)^2\Big\} = -\mathrm{E}\Big\{\frac{\partial^2}{\partial\theta^2}\ln p(\mathrm{x}|\theta)\Big\} \tag{4.29}$$

which demonstrates that the Fisher information can be denoted by the two equivalent expressions 4.24 and 2.38.

Exact Gaussian speckle statistics : Marginal Distribution of Amplitude and Phase

Stating from the bivariate Gaussian distribution

$$p(\mathrm{x},\mathrm{y}) = \frac{1}{2\pi\sigma_\mathrm{x}\sigma_\mathrm{y}}\exp\left\{-\left[\left(\frac{\mathrm{x}-\mu_\mathrm{x}}{2\sigma_\mathrm{x}}\right)^2 + \left(\frac{\mathrm{y}-\mu_\mathrm{y}}{2\sigma_\mathrm{y}}\right)^2\right]\right\} \tag{4.30}$$

we can obtain the marginal distribution of the intensity and the phase by integrating **??** with respect to ϕ and I respectively after transformation of 4.30 from Cartesian to polar coordinates.
In order to obtain the marginal distribution of the intensity, we execute the following rotational transformation Jakeman & Welford (1977), Uozomi & Asakura (1981)

$$\mathrm{u} = \mathrm{x}\cos\delta + \mathrm{y}\sin\delta \tag{4.31}$$
$$\mathrm{v} = \mathrm{y}\cos\delta - \mathrm{x}\sin\delta \tag{4.32}$$

with a rotational angle

$$\tan 2\delta = \frac{2\rho\sigma_\mathrm{x}\sigma_\mathrm{y}}{\sigma_\mathrm{x}^2 - \sigma_\mathrm{y}^2}. \tag{4.33}$$

After the transformation 4.30 becomes

$$p(\mathrm{u},\mathrm{v}) = \frac{1}{2\pi\sigma_\mathrm{u}\sigma_\mathrm{v}}\exp\left\{-\left[\left(\frac{\mathrm{u}-\mu_\mathrm{u}}{2\sigma_\mathrm{u}}\right)^2 + \left(\frac{\mathrm{v}-\mu_\mathrm{v}}{2\sigma_\mathrm{v}}\right)^2\right]\right\} \tag{4.34}$$

where

$$\begin{aligned}
\mu_\mathrm{u} &= \mu_\mathrm{x}\cos\delta + \mu_\mathrm{y}\sin\delta \\
\mu_\mathrm{v} &= \mu_\mathrm{x}\cos\delta - \mu_\mathrm{y}\sin\delta \\
\sigma_\mathrm{u}^2 &= (\sigma_\mathrm{x}^2\cos^2\delta - \sigma_\mathrm{y}^2\sin^2\delta)/\cos 2\delta \\
\sigma_\mathrm{v}^2 &= (\sigma_\mathrm{y}^2\cos^2\delta - \sigma_\mathrm{x}^2\sin^2\delta)/\cos 2\delta
\end{aligned}$$

By expressing 4.34 in polar coordinates and in function of the intensity, after some manipulations and since $I = \mathrm{x}^2 + \mathrm{y}^2 = \mathrm{u}^2 + \mathrm{v}^2$ we obtain

$$\begin{aligned}
p(I,\phi) = \frac{1}{2\pi\sigma_\mathrm{u}\sigma_\mathrm{v}}\exp\Bigg\{-\Bigg[&\left(\frac{\cos^2\phi}{2\sigma_\mathrm{u}^2} + \frac{\sin^2\phi}{2\sigma_\mathrm{v}^2}\right)I + \\
&-\left(\frac{\mu_\mathrm{u}}{\sigma_\mathrm{u}^2}\cos\phi + \frac{\mu_\mathrm{v}}{\sigma_\mathrm{v}^2}\sin\phi\right)\sqrt{I} + \frac{\mu_\mathrm{u}^2}{2\sigma_\mathrm{u}^2} + \frac{\mu_\mathrm{v}^2}{2\sigma_\mathrm{v}^2}\Bigg]\Bigg\}
\end{aligned} \tag{4.35}$$

At this point we have to marginalize with respect to the phase solving the following integral Ohtsubo & Asakura (1977)

$$p(I) = \frac{1}{2\pi\sigma_{\mathrm{u}}\sigma_{\mathrm{v}}} \int_0^{2\pi} \exp\left\{ -\left[\left(\frac{\cos^2\phi}{2\sigma_{\mathrm{u}}^2} + \frac{\sin^2\phi}{2\sigma_{\mathrm{v}}^2} \right) I + \right.\right. \\ \left.\left. - \left(\frac{\mu_{\mathrm{u}}}{\sigma_{\mathrm{u}}^2}\cos\phi + \frac{\mu_{\mathrm{v}}}{\sigma_{\mathrm{v}}^2}\sin\phi \right)\sqrt{I} + \frac{\mu_{\mathrm{u}}^2}{2\sigma_{\mathrm{u}}^2} + \frac{\mu_{\mathrm{v}}^2}{2\sigma_{\mathrm{v}}^2} \right] \right\} \mathrm{d}\phi \tag{4.36}$$

in order to get the marginal distribution of the intensity I. The solution of the integral takes the following form Jakeman & Welford (1977)

$$p(I) = \frac{f_1}{2\sigma_{\mathrm{u}}\sigma_{\mathrm{v}}} \left\{ I_0(f_2) I_0(f_3) + 2\sum_{n=1}^{\infty} I_n(f_2) I_{2n}(f_3) \cos(2n\alpha) \right\} \tag{4.37}$$

where I_n is a modified Bessel function of the first kind and where Jakeman & Welford (1977)

$$\begin{aligned} f_1 &= \exp\left\{ -\left[\frac{1}{4}\left(\frac{1}{\sigma_{\mathrm{u}}^2} + \frac{1}{\sigma_{\mathrm{v}}^2} \right) I + \frac{1}{2}\left(\frac{\mu_{\mathrm{u}}^2}{\sigma_{\mathrm{u}}^2} + \frac{\mu_{\mathrm{v}}^2}{\sigma_{\mathrm{v}}^2} \right) \right] \right\} \\ f_2 &= -\frac{1}{4} I \left(\frac{1}{\sigma_{\mathrm{u}}^2} - \frac{1}{\sigma_{\mathrm{v}}^2} \right) \\ f_3 &= -\sqrt{I} \left(\frac{\mu_{\mathrm{u}}^2}{\sigma_{\mathrm{u}}^4} + \frac{\mu_{\mathrm{v}}^2}{\sigma_{\mathrm{v}}^4} \right)^{1/2} \end{aligned}$$

and

$$\tan\alpha = \sigma_{\mathrm{u}}^2 \mu_{\mathrm{v}} / \sigma_{\mathrm{v}}^2 \mu_{\mathrm{u}}$$

Thus, the marginal probability density function of the intensity 4.37 has a complicated form.

The probability density function of the phase ϕ can be obtained directly by integrating 4.30 with respect to the amplitude. After polar transformation and some manipulation we can write 4.30 as follows

$$p(A, \phi) = dA \exp\{-b(A-c)^2\} \tag{4.38}$$

where

$$b = \frac{\sigma_{\mathrm{x}}^2 \sin^2\phi + \sigma_{\mathrm{y}}^2 \cos^2\phi - \rho\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} \sin 2\phi}{2\sigma_{\mathrm{x}}^2\sigma_{\mathrm{y}}^2(1-\rho^2)} \tag{4.39}$$

$$c = \frac{\sigma_{\mathrm{x}}^2(\mu_{\mathrm{y}}\sigma_{\mathrm{x}}^2 - \mu_{\mathrm{x}}\rho\sigma_{\mathrm{y}}^2)\sin\phi + \sigma_{\mathrm{y}}^2(\mu_{\mathrm{x}}\sigma_{\mathrm{y}}^2 - \mu_{\mathrm{y}}\rho\sigma_{\mathrm{x}}^2)\cos\phi}{\sigma_{\mathrm{x}}^2 \sin^2\phi + \sigma_{\mathrm{y}}^2 \cos^2\phi - \rho\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} \sin 2\phi} \tag{4.40}$$

$$d = \frac{1}{2\pi\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}\sqrt{1-\rho^2}} \exp\left\{ -\frac{(\mu_{\mathrm{x}}\sin\phi - \mu_{\mathrm{y}}\cos\phi)^2}{2(\sigma_{\mathrm{x}}^2 \sin^2\phi + \sigma_{\mathrm{y}}^2 \cos^2\phi - \rho\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} \sin 2\phi)} \right\} \tag{4.41}$$

where we change the notation for the sake of clarity. It is worth to note that b, c and d are independent from A.

Thus, the result of the integral

$$p(\phi) = d \int_0^{\infty} A \exp\{-b(A-c)^2\}\, \mathrm{d}A \tag{4.42}$$

takes the following form

$$p(\phi) = \frac{d}{2b}\left\{\exp\left(-bc^2\right) - \sqrt{\pi}\sqrt{bc^2}(1 - \mathrm{Erf}[\sqrt{bc^2}])\right\} \tag{4.43}$$

where $\mathrm{Erf}[\cdot]$ is the integral of the Gaussian distribution defined as

$$\mathrm{Erf}[\kappa] = \frac{2}{\sqrt{\pi}}\int_0^{\kappa} \mathrm{e}^{-t^2}\,\mathrm{d}t \tag{4.44}$$

and b, c and d come from the notation introduced in 4.38. Equation 4.43 is the pdf of the phase.

List of Publications

Matteo Soccorsi, Dusan Gleich and Mihai Datcu, *GMRF Regularization of Complex SAR Images*, work in progress
Matteo Soccorsi and Mihai Datcu, *Rate Distortion Model Selection for Information Extraction* work in progress
Matteo Soccorsi, Dusan Gleich and Mihai Datcu, *Huber-Markov Model for Complex SAR Image Restoration*, published in Geoscience and Remote Sensing Letter, 2009.
Matteo Soccorsi and Mihai Datcu, *TerraSAR-X a Complex Image Approach for feature Extraction and Modeling*, Proceeding in IGARSS 2008, Boston, US-MA.
Matteo Soccorsi and Mihai Datcu, *Rate Distortion based Optimization of the Analysing Window*, Proceedings in IWAP 2008, University of Technology of Compiègne, France.
Matteo Soccorsi and Mihai Datcu, *TerraSAR-X Data Evidence Maximization-based Feature Extraction and Despeckling* Proceeding in EUSAR 2008, Graf-Zeppelin-Haus, Friedrichshafen, Germany.
Matteo Soccorsi and Mihai Datcu, *TerraSAR-X Data Feature Extraction : a Complex-Valued Data Analysis*, Proceeding in ESA-EUSC 2008, Frascati, Italy.
Matteo Soccorsi and Mihai Datcu, *Phase Characterization of PolSAR Image*, Proceeding in SPIE Europe Remote Sensing 2007, Florence, Italy.
Matteo Soccorsi and Mihai Datcu, *Stochastic Models of SLC HR*, Proc. in IGARSS 2007, Barcelona, Spain.
Matteo Soccorsi, Fabio Del Frate and Mihai Datcu, *Rate Distortion Based Analysis of Image Parameter Estimation for Information Extraction*, Proceeding in IEEE GOLD Remote Sensing Conference 2006, Bari, Italy.
Matteo Soccorsi and Mihai Datcu, *Bayesian Texture based Analysis of HR SLC SAR Images*, Proceeding in ESA-EUSC 2006, Torrejon de Ardoz, Spain.
Matteo Soccorsi and Mihai Datcu, *Space-Variant Model Fitting and Selection for Image Information Extraction*, Proceeding in MaxEnt 2006, CNRS, Paris, France.

Bibliographie

Achim, A., Kuruoglu, E. E. & Zerubia, J. (2006). SAR image filtering based on the heavy-tailed Rayleigh model, *IEEE Transaction on Image Processing* **15**(9) : 2686–2693.

Achim, A., Tsakalides, P. & Bezerianos, A. (2003). SAR image denoising ia bayesian wavelet shrinkage based on heavy-tailed modeling, *IEEE Transaction on Geoscience and Remote Sensing* **41**(8) : 1773–1784.

Argenti, F., Bianchi, T. & Alparone, L. (2006). Multiresolution MAP Despeckling of SAR Images Based on Locally Adaptive Generalized Gaussian pdf Modeling, *IEEE Transaction on Image Processing* **15**(11) : 3385–3399.

Bamler, R. & Schättler, B. (1993). *SAR Data Acquisition and Image Formation*, Wichman, Karlsruhe, Ge, chapter 3, pp. 53–102.

Bouman, C. & Sauer, K. (1993). A generalized Gaussian image model for edge-preserving MAP estimation, *IEEE Trans. on Image Processing* **2**(3) : 296–310.

Brady, M. & Horn, B. K. P. (1983). Rotationally symmetric operators for surface interpolation, *Computer Vision, Graphics and Image Processing* **22**(1) : 70–94.

Bruniquel, J. & Lopès, A. (1998). On the true multilook intensity distribution in sar imagery, *Proceeding in International Geoscience and Remote Sensing Symposium*, Vol. 1, Seattle, US-WA, pp. 1–3.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications*, Cambridge University Press, New York, US-NY.

Cafforio, C., Prati, C. & Rocca, F. (1991). SAR data focusing using seismic migration techniques, *IEEE Transaction on aerospace and Electronic Systems* **27**(2) : 194–207.

Carrara, W. G., Goodman, R. S. & Mejewski, R. M. (1995). *Spotlight Synthetic Aperture Radar : Signal Processing Algorithms*, Artech House, Norwood, US-MA.

Çetin, M. & Karl, W. C. (2001). Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization, *IEEE Transaction on Image Processing* **10**(4) : 623–631.

Chan, T. F. & Mulet, P. (1995). Iterative methods for total variation image restoration, *Proceedings of Winter School on Iterative Methods, Chinese University of Hong Kong*, Spring Verlag, Singapore.

Channappayya, S. S., Bovik, A. C., Jr., R. W. H. & Caramanis, C. (2008). Rate bounds on SSIM index of quantized image DCT coefficients, *Proceeding in Data Compression Conference, (DCC'08)*, Snowbird, US-UT, pp. 352–361.

Chelappa, R. & Kashyap, R. (1983). Estimation and choice of neighbors in spatial interaction models of images, *IEEE Transactions on Information Theory* **29**(1) : 60–73.

Chelappa, R., Chatterjee, S. & Bagdazian, R. (1985). Texture synthesis and compression using Gaussian-Markov random field models, *IEEE Transaction on Systems, Man, and Cybernetics* **SMC-15**(2) : 298–303.

Costache, M. & Datcu, M. (2006). Bayesian enhancement of svm based image search engine, *Proceeding in ESA-EUSC 2006 : Image Information Mining for Security and Intelligence*, Torrejon air base - Madrid (Spain).

Costache, M. & Datcu, M. (2007). Learning-unlearning for mining high resolution eo images, *Proceeding in International Geoscience and Remote Sensing Symposium, (IGARSS'97)*, Barcelona, Spain.

Costache, M., Maitre, H. & Datcu, M. (2006). Categorization based relevance feedback search engine for earth observation images repositories, *Proceeding in International Geoscience and Remote Sensing Symposium, (IGARSS'96)*, Denver, US-CO.

Cover, T. M. & Thomas, J. A. (1991a). *Elements of Information Theory*, Wiley Series in Telecommunication, New York, US-NY.

Cover, T. M. & Thomas, J. A. (1991b). *Elements of Information Theory*, Wiley Series in Telecommunication, New York, US-NY.

Crimmins, T. R. (1985). Geometric filter for speckle reduction, *Applied Optics* **24** : 1438–1443.

Dana, R. A. & Knepp, D. L. (1986). The impact of strong scintillation on space based radar design II : Noncoherent detection, *IEEE Transactions on Aerospace and Electronic Systems* **AES-22** : 34–36.

Datcu, M., Stoichescu, D. A., Seidel, K. & Iorga, C. (2004). Model fitting and model evidence for multiscale image texture analysis, *24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Vol. 735, pp. 35–42.

Delignon, Y., Marzouki, A. & Pieczynski, W. (1997). Estimation of generalized mixtures and its application in image segmentation, *IEEE Transaction on Image Processing* **6** : 1364–1375.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1) : 1–38.

DLR (2008). *TerraSAR-X Ground Segment Basic Product Specification Document*, Website. `http://www.dlr.de/tsx/documentation/SAR_Basic_Products.pdf`.

Eltoft, T. (2003). Speckle modeling and filtering, *Proceeding in IEEE Norwegian Symposium on Signal Processing*, Bergen, Norway.

Fjortoft, R. & Lópes, A. (2001). Estimation of the mean radar reflectivity from a finite number of correlated samples, *IEEE Transaction on Geoscience and Remote Sensing* **39**(1) : 196–199.

Fjortoft, R., Lòpes, A., me Bruniquel, J. & Marthon, P. (1999). Optimal edge detection and edge localization in complex SAR images with correlated speckle, *IEEE Transaction on Geoscience and Remote Sensing* **37**(5) : 2272–2281.

Fosgate, C. H., Krim, H., Irving, W. W., Karl, W. C. & Willsky, A. S. (1997). Multiscale segmentation and anomaly enhancement of sar imagery, *IEEE Transaction on Image Processing* **6** : 7–20.

Franceschetti, G., Pascazio, V. & Schirinzi, G. (1995a). Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging, *Journal of the Optical Society of America A : Optics, Image Science, and Vision* **12** : 686–694.

Franceschetti, G., Pascazio, V. & Schirinzi, G. (1995b). Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging, *Journal of the Optical Society of America* **12**(4) : 686–694.

Frery, A. C., Muller, H.-J., Yanasse, C. C. F. & Sant'Anna, S. (1997). A model for extremely heterogeneous clutter, *IEEE Transaction on Geoscience and Remote Sensing* **35** : 648–659.

Frieden, B. R. & Bajkova, A. T. (1994). Bayesian cross-entropy reconstruction of complex images, *Applied Optics* **33** : 219–226.

Frost, V., Stiles, J., Shanmugan, K. & Holtzman, J. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise, *IEEE Transaction on Pattern Analysis and Machine Learning* **PAMI-4** : 157–165.

Geman, S. & Geman, D. (1984a). Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-9**(6) : 721–741.

Geman, S. & Geman, D. (1984b). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Transacton on Pattern Analysis and Machine Intelligence* **PAMI-6** : 721–741.

Gersho, A. & Gray, R. M. (1991). *Vector Quantization and Signal Compression*, Springer, New York, US-NY.

Gleich, D. & Datcu, M. (2006). Gauss-Markov model for wavelet-based SAR image despeckling, *IEEE Signal Processing Letters* **13**(6) : 365–368.

Gleich, D. & Datcu, M. (2007a). Wavelet-based despeckling of SAR images using Gauss-Markov random fields, *IEEE Transactions on Geoscience and Remote Sensing* **45**(12) : 4127–4143.

Gleich, D. & Datcu, M. (2007b). Wavelet-based despeckling of SAR images using Gauss-Markov random fields, *IEEE Transaction on Geoscience and Remote Sensing* **45**(12) : 4127–4143.

Goldfinger, A. D. (1982). Estimation of spectra from speckled images, *IEEE Transaction on Aerospace and Electronic Systems* **AES-18**(5) : 675–681.

Hagg, W. & Sties, M. (1994). Efficient speckle filtering of SAR images, *Proceeding in International Geoscience and Remote Sensing Symposium. Surface and Atmospheric Remote Sensing : Technologies, Data Analysis and Interpretation, (IGARSS'94).*, Pasadena, US-CA, pp. 2140–2142.

Hebar, M., Gleich, D. & Žarko Čučej (2009). Autobinomial model for SAR image despeckling and information extraction, *IEEE Transaction on Geoscience and Remote Sensing* **47**(8) : 2818–2835.

Hein, A. (2004). *Processing of SAR Data : Fundamentals, Signal Processing, Interferometry*, Springer-Verlag, Berlin, Germany.

Hetzheim, H. (1993). Using martingale representation to adapt models for non-linear filtering, *Proceeding in International Conference in Signal Processing*, Intern. Academic Publ., Beijing, China, pp. 32–35.

Horn, B. K. & Rhunck, B. G. (1981). Determining optical flow, *Artificial Intelligence* **17** : 185–203.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes, *Proceeding in Institute of Radio Engineers*, pp. 1098–1102.

IEEE-AESS (2003). *IEEE Standard for Letter Designations for Radar-Frequency Bands*, The Institute of Electrical and Electronics Engineers, Inc., New York, US-NY. IEEE Std 521^{TM}-2002.

Ikeuchi, K. & Horn, B. K. (1981). Numerical shape from shading and occluding boundaries, *Artificial Intelligence* **17** : 141–183.

Jakeman, E. (1980). On the statistics of K-distributed noise, *Journal of Physics A : Mathematical and General* **13** : 21–48.

Jakeman, E. & Pusey, P. N. (1976). A model for non-Rayleigh sea echo, *IEEE Transaction on Antennas and Propagation* **AP-24** : 806–814.

Jakeman, E. & Pusey, P. N. (1978). Significance of K distributions in scattering experiments, *Physical Review Letters* **40** : 546–550.

Jakeman, E. & Welford, W. T. (1977). Speckle statistics in imaging systems, *Optics Communication* **21**(1) : 72–79.

Jao, J. (1984). Amplitude distribution of composite terrain radar clutter and the K-distribution, *IEEE Transaction on Antennas and Propagation* **AP-32** : 1049–1052.

Jaynes, E. T. (1957). Information theory and statistical mechanics, *The Physical Review* **106** : 620–630.

Kim, J., Tsai, A., Çetin, M. & Willsky, A. S. (2002). A curce evolution-based variational approach to simultaneous image restoration and segmentation, *Proceeding in IEEE International Conference on Image Processing, (ICIP'02)*, Rochester, US-NY.

Kraft, L. G. (1949). *A device for quantizing, grouping, and coding amplitude-modulated pulses*, Massachusetts Institute of Technology.

Kuan, D. T., Sawchuk, A. A., Strand, T. C. & Chavel, P. (1985). Adaptive noise smoothing filter for images with signal-dependent noise, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-7** : 165–177.

Kuan, D. T., Sawchuk, A. A., Strand, T. C. & Chavel, P. (1987). Adaptive restoration of images with speckle, *IEEE Transactions on Acoustics, Speech and Signal Processing* **35** : 373–383.

Kullback, S. & Leibler, R. A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics* **22**(1) : 79–86.

Kuruoglu, E. E. & Zerubia, J. (2004). Modeling SAR images with a generalization of the rayleigh distribution, *IEEE Transactions on Image Processing* **13** : 527–533.

Lakshmanan, S. & Derin, H. (1993). Valid parameter space for 2-D Gaussian Markov random fields, *IEEE Transaction on Information Theory* **39**(2) : 703–709.

Lee, J. (1980). Digital image enhancement and noise filtering by using local statistics, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **PAMI-2** : 165–168.

Lee, J. S., Hoppel, K. W. & Miller, A. R. (1994). Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery, *IEEE Transaction on Geoscience and Remote Sensing* **32** : 1017–1027.

Lewinski, D. J. (1983). Nonstationary probabilistic target and clutter scattering models, *IEEE Transaction on Antennas and Propagation* **AP-31**(3) : 490–498.

Li, H.-C., Hong, W., Wu, Y.-R. & Tai, H.-M. (2007). Texture-preserving despeckling of SAR images using evidence framework, *IEEE Transaction on Geoscience and Remote Sensing Letters* **4**(4) : 537–541.

Lopès, A. & Séry, F. (1997). Optimal speckle reduction for the product model in multilook polarimetric SAR imagery and the wishart distribution, *IEEE Transaction on Geoscience and Remote Sensing* **35**(5) : 632–647.

Lopès, A., Nezry, E., Touzi, R. & Laur, H. (1990a). Maximum a posteriori speckle filtering and first order texture models in SAR images, *Proceeding in International Geoscience and Remote Sensing Symposium. Remote Sensing Science for the Nineties*, Washington, US-DC, pp. 2409–2412.

Lopès, A., Nezry, E., Touzi, R. & Laur, H. (1993). Structure detection and statistical adaptive speckle filtering in SAR images, *International Journal of Remote Sensing* **14**(9) : 1735–1758.

Lopès, A., Touzi, R. & Nezry, E. (1990b). Adaptive speckle filters and scene heterogeneity, *IEEE Transaction on Geoscience and Remote Sensing* **28**(6) : 992–1000.

MacKay, D. J. C. (1999). Comparison of approximate methods for handling hyperparameters, *Neural Computation* **11**(5) : 1035–1068.

Marroquin, J., Mitter, S. & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association* **82** : 76–89.

Meer, P., Park, R.-H. & Cho, K. (1994). Multiresolution adaptive image smoothing, *CVGIP : Graphical Models and Image Processing* **56** : 140–148.

Moser, G., Zerubia, J. & Serpico, S. B. (2006). SAR amplitude probability density function estimation based on a generalized Gaussian model, *IEEE Transaction on Image Processing* **14** : 1429–1442.

Muller, H.-J. & Pac, R. (1999). G-statistics for scaled SAR data, *Proceeding in International Geoscience and Remote Sensing Symposium, (IGARSS'99)*, Hamburg, Germany, pp. 1297–1299.

Neeser, F. D. & Massey, J. L. (1993). Proper complex random process with application to information theory, *IEEE Transaction on Information Theory* **39**(4) : 1293–1302.

NTIA (2008). *Manual of Regulations & Procedures for Federal Radio Frequency Management*, Website. `http://www.ntia.doc.gov/osmhome/redbook/redbook.html`.

Ohtsubo, J. & Asakura, T. (1977). Statistical properties of laser apeckle produced in the diffraction field, *Applied Optic* **16**(6) : 1742–1753.

Oliver, C. & Quegan, S. (2004a). *Understanding Synthetic Aperture Radar Images*, Scitech Publishing Inc., Raleigh, US-NC.

Oliver, C. & Quegan, S. (2004b). *Understanding Synthetic Aperture Radar Images*, Scitech Publishing Inc., Raleigh, US-NC.

Oliver, C. & Quegan, S. (2004c). *Understanding Synthetic Aperture Radar Images*, Scitech Publishing Inc., Raleigh, US-NC.

Oliver, C. J. (1984). A model for non-Rayleigh scattering statistics, *Optica Acta* **31** : 701–722.

Oliver, C. J. (1991). Information from SAR images, *Journal of Physics D : Applied Physics* **24** : 1493–1514.

Oppenheim, A. V. & Schafer, R. W. (1975). *Digital Signal Processing*, Prentice Hall, Inc., Englewood Cliffs, US-NJ.

Pan, R. & Reeves, S. J. (2006). Efficient Huber-Markov edge-preserving image restoration, *IEEE Transaction on Image Processing* **10**(12) : 3728–3735.

Pascazio, V. & Ferraiuolo, G. (2003). Statistical regularization in linearized microwave imaging through MRF-based MAP estimation hyperparameter estimation and image computation, *IEEE Transaction on Image Processing* **12**(5) : 572–582.

Pérez, P. (1998). Markov random fields and images, *CWI Quarterly* **11**(4) : 413–437.

Perona, P. & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **12**(7) : 629–639.

Pesaresi, M. (1996). Textural classification of very high-reslution satellite imagery : Empirical estimation of the interaction between window size and detection accuracy in urban environment, *Proceedings of the International Conference on Image Processing*, Vol. 1, pp. 114–118.

Picinbono, B. (1996). Second-order complex random vectors and normal disrtibutions, *IEEE Transaction on Signal Processing* **44**(10) : 2637–2640.

Picinbono, B. & Bondon, P. (1997). Second-order statistics of complex signals, *IEEE Transactions on Signal Processing* **45** : 411–420.

Picinbono, B. & Bouvet, M. (1984). Complex white noises and autoregressive signals, *IEEE International Conference on Acoustics, Speech, and Signal Processing* **9** : 596–599.

Picinbono, B. & Chevalier, P. (1995). Widely linear estimation with complex data, *IEEE Transactions on Signal Processing* **43**(8) : 2030–2033.

Picinbono, B. & Chevalier, P. (1996). Extensions of the minimum variance method, *Signal Processing* **49** : 1–9.

Randen, T. (1997). *Brodatz Texture Archive*, Website. `http://www.ux.uis.no/~tranden/brodatz.html`.

Raney, R. K. (1992). An exact wide field digital imaging algorithm, *International Journal of Remote Sensing* **13**(5) : 991–998.

Raney, R. K. & Wessels, G. J. (1988). Spatial consideration in SAR speckle simulation, *IEEE Transactions on Geoscience and Remote Sensing* **26**(5) : 666–672.

Raney, R. K., Runge, H., Bamler, R., Cumming, I. G. & Wong, F. H. (1994). Precision sar processing using chirp scaling, *IEEE Transaction on Geoscience and Remote Sensing* **32**(4) : 786–799.

Rees, W. G. & Satchell, M. J. F. (1997). The effect of median filtering on synthetic aperture radar images, *International Journal of Remote Sensing* **18**(13) : 2887–2893.

Rignot, E. & Chelappa, R. (1991). Segmentation of synthetic-aperture-radar complex data, *Journal of the Optical Society of America A : Optics, Image Science, and Vision* **8** : 1499–1509.

Ruanaidh, J. & Fitzgerald, W. (1996). *Numerical Bayesian Methods Applied to Signal Processing*, Springer-Verlag, New York, NY.

Rudin, L. I., Osher, S. & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms, *Physica D* **60** : 259–268.

Saleh, B. E. A. & Rabbani, M. (1980). Linear filtering of speckled images, *Optics Communications* **35**(3) : 327–331.

Saquib, S. S., Bouman, C. A. & Sauer, K. (1998). ML parameter estimation for markov random field with application to Bayesian tomography, *IEEE Transactoin on Image Processing* **7**(7) : 1029–1044.

Schreier, G. (1993). *SAR Geocoding : Data and Systems*, Wichmann, Karlsruhe, GE.

Schreier, P. J. & Scharf, L. L. (2003). Second-order analysis of improper complex random vectors and processes, *IEEE Transaction on Signal Processing* **51**(3) : 714–725.

Sekine, M. & Mao, Y. (1990). *Weibull Radar Clutter*, IEE-UK, London, UK.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell system Technical Journal* **27** : 379–423/623–656.

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion, *IRE Nat. Conv. Rec.*, pp. 142–163.

Shewchuk, J. R. (1994). *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Website. `http://www.cs.cmu.edu/~jrs/jrspapers.html`.

Simoncelli, E. P. & Adelson, E. H. (1996). Noise removal via Bayesian wavelet coring, *Proceeding in IEEE International Conference in Image Processing*, Vol. 1, pp. 379–382.

Snyder, W., Han, Y.-S., Bilbro, G., Whitaker, R. & Pizer, S. (1995). Image relaxation : restoration and feature extraction, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **17**(6) : 620–624.

So, S. & Paliwal, K. K. (2003). Low complexity Gaussian mixture model-based block quantisation of images, *Proceeding in Microelectronic Engineering Research Conference*, Vol. 20, Brisbane, Australia.

Soccorsi, M. & Datcu, M. (2008). Rate distortion based optimization of the analysing window, *International Workshop on Applied Probability, Compiègne, France*.

Soccorsi, M., Datcu, M. & Frate, F. D. (2006). Rate distortion based analysis of image parameter estimation for information mining, *IEEE Gold Remote Sensing conference, Bari, Italy*.

Soccorsi, M., Gleich, D. & Datcu, M. (2009). Huber-Markov model for complex SAR image restoration, *IEEE Geoscience and Remote Sensing Letters*. Accepted for publication.

Spitzer, F. (1971). Markov random field and Gibbs ensembles, *The American Mathematical Monthly* **378**(2) : 142–154.

Stangl, M., Werninghaus, R., Schweizer, B., Fischer, C., Brandfass, M., Mittermayer, J. & Breit, H. (2006). TerraSAR-X technologies and first results, *IEE Proceedings in Radar, Sonar and Navigation* **153**(2) : 86–95.

Szajnowski, W. J. (1977). Estimators of log-normal distribution parameters, *IEEE Transactions on Aerospace and Electronic Systems* **AES-13** : 533–536.

Taneja, I. J. (2001). *Generalized Information Measures and Their Applications*, Website. `http://www.mtm.ufsc.br/~taneja/book/book.html`.

Tikhonov, A. N. & Arsenin, V. Y. (1977). *Solutions of Ill-Posed Problems*, Winston & Sons, New York, US-NY.

Tison, C., Nicolas, J.-M., Tupin, F. & Maitre, H. (2004). A new statistical model for markovian classification of urban areas in high-resolution SAR images, *IEEE Transactions on Geoscience and Remote Sensing* **42** : 2046–2057.

Toma, C. & Datcu, M. (1992). Maximum entropy and minimum cross-entropy methods in image processing, *Proceeding in SPIE'92*, Vol. 1827, Boston, US-MA, pp. 133–134.

Torre, V. & Poggio, T. A. (1986). On edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(2) : 147–163.

Ulaby, F. T., Kouyate, F., Brisco, B. & Williams, T. H. L. (1986). Textural information in SAR images, *IEEE Transaction on Geoscience and Remote Sensing* **GE-24**(2) : 235–245.

Uozomi, J. & Asakura, T. (1981). First-order intensity and phase statistics of Gaussian speckle produced in the diffraction region, *Applied Optics* **20**(8) : 1454–1466.

van den Bos, A. (1995). The multivariate complex normal distribution - a generalization, *IEEE Transaction on Information Theory* **41**(2) : 537–539.

Walessa, M. & Datcu, M. (2000). Model-based despeckling and information extraction from SAR images, *IEEE Transaction on Geoscience and Remote Sensing* **38**(5) : 2258–2269.

Walker, J. L. (1980). Range-doppler imaging of rotating objects, *IEEE Transaction on aerospace and Electronic Systems* **AES-16**(1) : 23–52.

Walsh, B. (2004). *Markov Chain Monte Carlo and Gibbs Sampling*, Website. `http://web.mit.edu/~wingated/www/introductions/mcmc-gibbs-intro.pdf`.

Wang, Z. & Bovik, A. C. (2009). Mean squared error : Love it or leave it ?, *IEEE Signal Processing Magazine* **26**(1) : 98–117.

Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004). Image quality assessment : From error visibility to structural similarity, *IEEE Transaction on Image Processing* **13**(4) : 1–14.

Watts, S., Tough, R. & Ward, K. (2006). *Sea Clutter : Scattering the K-Distribution and Radar Performance*, Institution of Engineering and Technology, London, UK.

Woodhouse, I. H. (2006). *Introduction to Microwave Remote Sensing*, CRC Press, Taylor & Francis Group, Broken Sound Parkway, US-NY.

Yu, Y. & Acton, S. (2002). Speckle reducing anisotropic diffusion, *IEEE Transaction on Image Processing* **11**(11) : 1260–1270.

Zhou Wang, Alan C. Bovik, H. R. S. & Simoncelli, E. P. (n.d.). *The SSIM Index for Image Quality Assessment*, Website. `http://www.ece.uwaterloo.ca/~z70wang/research/ssim/`.