Statistical modeling and processing of high frequency ultrasound images : application to dermatologic oncology

Marcelo Alejandro Pereyra

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Signal, Information et Hypermedia

Présentée et soutenue par :
Marcelo PEREYRA

le : mercredi 4 juillet 2012

Titre :

Statistical modeling and processing of high frequency ultrasound images: application to dermatologic oncology

---

Ecole doctorale :
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :

IRIT

Directeur(s) de Thèse :

Alain AYACHE, PU, Université de Toulouse
Hadj BATATIA, MCF, Université de Toulouse

Rapporteurs :

Jérôme IDIER, DR, IRCCyN Nantes
Xavier DESCOMBES, DR, INRIA Sophia Antipolis

Membre(s) du jury :

Patrice ABRY, DR, ENS Lyon - Examinateur
Steve McLAUGHLIN, PU, Heriot Watt University - Examinateur
Jean-Yves TOURNERET, PU, Université de Toulouse - Examinateur
Petar DJURIC, PU, Stoney Brook University - Invité

Abstract

This thesis studies statistical image processing of high frequency ultrasound imaging, with application to in-vivo exploration of human skin and noninvasive lesion assessment. More precisely, Bayesian methods are considered in order to perform tissue segmentation in ultrasound images of skin. It is established that ultrasound signals backscattered from skin tissues converge to a complex Levy Flight random process with non-Gaussian α-stable statistics. The envelope signal follows a generalized (heavy-tailed) Rayleigh distribution. Based on these results, it is proposed to model the distribution of multiple-tissue ultrasound images as a spatially coherent finite mixture of heavy-tailed Rayleigh distributions. Spatial coherence inherent to biological tissues is modeled by a Potts Markov random field. An original Bayesian algorithm combined with a Markov chain Monte Carlo method is then proposed to jointly estimate the mixture parameters and a label-vector associating each voxel to a tissue. The proposed method is successfully applied to the segmentation of in-vivo skin tumors in high frequency 2D and 3D ultrasound images. This method is subsequently extended by including the estimation of the Potts regularization parameter β within the Markov chain Monte Carlo (MCMC) algorithm. Standard MCMC methods cannot be applied to this problem because the likelihood of β is intractable. This difficulty is addressed by using a likelihood-free Metropolis-Hastings algorithm based on the sufficient statistic of the Potts model. The resulting unsupervised segmentation method is successfully applied to tridimensional ultrasound images. Finally, the problem of computing the Cramer-Rao bound (CRB) of β is studied. The CRB depends on the derivatives of the intractable normalizing constant of the Potts model. This is resolved by proposing an original Monte Carlo algorithm, which is successfully applied to compute the CRB of the Ising and Potts models.

Keywords: ultrasound images, statistical image processing, Markov random field, Bayesian modeling, intractable distributions.

Modélisation et traitement statistiques d'images d'ultrasons de haute fréquence. Application à l'oncologie dermatologique.

Résumé

Cette thèse étudie le traitement statistique des images d ultrasons de haute fréquence, avec application à l exploration in-vivo de la peau humaine et l évaluation non invasive de lésions. Des méthodes Bayésiennes sont considérées pour la segmentation d images échographiques de la peau. On y établit que les ultrasons rétrodiffusés par la peau convergent vers un processus aléatoire complexe de type Levy-Flight, avec des statistiques non Gaussiennes alpha-stables. L enveloppe du signal suit une distribution Rayleigh généralisée à queue lourde. A partir de ces résultats, il est proposé de modéliser l image ultrason de multiples tissus comme un mélange spatialement cohérent de lois Rayleigh à queues lourdes. La cohérence spatiale inhérente aux tissus biologiques est modélisée par un champ aléatoire de Potts-Markov pour représenter la dépendance locale entre les composantes du mélange. Un algorithme Bayésien original combiné à une méthode Monte Carlo par chaine de Markov (MCMC) est proposé pour conjointement estimer les paramètres du modèle et classifier chaque voxel dans un tissu. L approche proposée est appliquée avec succès à la segmentation de tumeurs de la peau in-vivo dans des images d ultrasons de haute fréquence en 2D et 3D. Cette méthode est ensuite étendue en incluant l estimation du paramètre β de régularisation du champ de Potts dans la chaine MCMC. Les méthodes MCMC classiques ne sont pas directement applicables à ce problème car la vraisemblance du champ de Potts ne peut pas être évaluée. Ce problème difficile est traité en adoptant un algorithme Metropolis-Hastings "sans vraisemblance" fondé sur la statistique suffisante du Potts. La méthode de segmentation non supervisée, ainsi développée, est appliquée avec succès à

des images échographiques 3D. Finalement, le problème du calcul de la borne de Cramer-Rao (CRB) du paramètre β est étudié. Cette borne dépend des dérivées de la constante de normalisation du modèle de Potts, dont le calcul est infaisable. Ce problème est résolu en proposant un algorithme Monte Carlo original, qui est appliqué avec succès au calcul de la borne CRB des modèles d Ising et de Potts.

Mots clés : images d ultrasons, traitement statistiques d images, champ aléatoire de Markov, modèles Bayésiens, lois de probabilité non calculables.

Contents

List of Tables

List of Figures

Acknowledgment

Above all, my deepest gratitude is to my *compagne* Anouck, my parents and my brother for their unconditional support and great love, for which my mere expression of thanks does not suffice.

This thesis would not have been possible without the help, support and advice of my scientific supervisor Dr Hadj Batatia, to whom I am extremely grateful. I owe sincere and earnest thankfulness to Prof. Jean-Yves Tourneret and Dr. Nicolas Dobigeon for their time and good advice. I would also like to show my gratitude to Prof. Jérôme Idier and Prof. Xavier Descombes for accepting to review and examine my thesis, and to Prof. Steve McLaughlin, Prof. Patrice Abry and Prof. Petar Djuric for accepting to examine it.

This thesis was developed as part of the CAMM4D project, funded by the French FUI and the Midi-Pyrenees region. I would like to acknowledge the support of the Hospital of Toulouse, Pierre Fabre Laboratories and Magellium for the ultrasound image acquisition, as well as the academic and technical support of the University of Toulouse and its staff. I would particularly like to thank Dr. Nicolas Meyer, Dr. Siham Lourari, Mr. Jérôme Georges, Mrs Armengaud and Mrs Eichen. I am also grateful to the CNES, who provided the SAR and optical images used in Chapter 4.

Last, but by no means least, I thank my friends and family in Argentina, France and elsewhere for their encouragement throughout.

Introduction

L imagerie par ultrasons est une modalité largement répondue avec des applications en diagnostic, examens préventifs, thérapie et chirurgie guidée par l image, entre autres. En oncologie dermatologique, le diagnostic de lésions potentiellement cancéreuses nécessite d effectuer des biopsies - intervention chirurgicale mineure - et de faire une évaluation histologique. Le traitement du cancer de la peau consiste généralement en une seconde opération pour exciser la tumeur. Cependant, les frontières de la lésion sont souvent difficiles à localiser avec précision et des procédures répétitives sont mises en oeuvre pour éradiquer la tumeur entière. Les progrès récents dans les technologies des transducteurs d ultrasons de haute fréquence et des sondes tridimensionnelles apportent l opportunité d examiner la peau d une manière noninvasive et de détourer les tumeurs avant excision sur des images échographiques. Cependant, le changement des pratiques dermatologiques requiert des méthodes de traitement d images qui soient spécifiques pour la dermatologie et pour ces nouvelles technologies. La motivation de cette thèse est de développer des méthodes de traitement d images pour permettre aux dermatologues d utiliser les images échographiques pour explorer la peau humaine in-vivo et d examiner les lésions. Plus précisément, des méthodes bayésiennes sont considérées pour segmenter les tissus biologiques sur des images échographiques. Quatre contributions méthodologiques principales sont proposées :

- Un modèle statistique α-stable des échos d ultrasons rétrodiffusés par la peau [5–7].
- Une méthode Bayésienne de segmentation fondée sur un modèle de mélange α-Rayleigh et un champ aléatoire de Potts-Markov [3, 4, 8, 9].
- Un algorithme MCMC pour estimer le paramètre de régularisation du modèle de Potts dans la méthode de segmentation [10, 11].

- Un algorithme Monte Carlo pour calculer la borne de Cramer-Rao du paramètre d un champ aléatoire de Markov avec application aux modèles d Ising et de Potts.

La thèse est organisée en cinq chapitres. Le chapitre 1 introduit le contexte médical qui motive ce travail, présente les notions générales sur la peau et l imagerie ultrasonore et décrit le type de signaux ultrasons utilisés dans les expérimentations. Les chapitres 2, 3 et 4 ont été présentés comme articles de journaux [4, 5, 10].

Le chapitre 2 étudie la distribution statistique d une région homogène dans une image d ultrasons, représentant un tissu unique et isolé [5]. Partant du modèle largement admis de diffusion ponctuelle, il est établi analytiquement que les signaux ultrasonores (radiofréquences) rétrodiffusés par les tissus de la peau convergent vers à un processus stochastique complexe de type Levy Flight avec des statistiques α-stables nonGaussiennes. L enveloppe du signal (ou mode B) suit une distribution Rayleigh généralisée à queue lourde. Ce modèle généralise le cadre Gaussien classique et fournit une représentation formelle d un cas de statistiques non-Gaussiennes, où le nombre de diffuseurs et la variance de leur section tendent tous deux l infini. En plus, des expressions analytiques sont dérivées pour mettre en relation les paramètres α-stables et les propriétés des diffuseurs. Les résultats de plusieurs expérimentations soutenus par d excellents tests de KS confirment le modèle proposé.

Le Chapitre 3 traite le problème de l estimation conjointe de la distribution statistique et des frontières de multiples tissus dans des images mode B de la peau [4, 8]. En se basant sur les développements du chapitre 2, la distribution de multiples tissus est modélisée comme un mélange spatialement cohérent de distributions Rayleigh à queues lourdes. La cohérence spatiale, inhérente aux tissus biologiques, est modélisée par un champ aléatoire de Potts-Markov représentant la dépendance locale entre les composantes du mélange. Un algorithme Bayésien original combiné à une méthode Monte Carlo par chaine de Markov (MCMC) est proposé pour conjointement estimer les paramètres du modèle ainsi qu un vecteur d étiquettes associant chaque voxel à un tissu. L approche proposée est appliquée

avec succès à la segmentation de tumeurs in-vivo de la peau dans des images d ultrasons de haute fréquence en 2D et 3D.

Les résultats de la segmentation obtenus par la méthode du chapitre 3 dépendent du degré de cohérence spatiale introduite par le champ de Potts. Ce degré est contrôlé par le coefficient de granularité, communément noté β. Ce paramètre important a été fixé empiriquement. Le chapitre 4 étudie l estimation du paramètre du modèle de Potts conjointement avec ceux du modèle Bayésien dans le cadre de l algorithme MCMC [10, 11]. Ce problème a été traité par plusieurs auteurs dans des travaux de recherche récents visant la création d une méthode de segmentation totalement non supervisée. Les méthodes MCMC classiques ne s appliquent pas directement à ce problème du fait que la vraisemblance du champ de Potts ne peut pas être évaluée. Dans la méthode de segmentation proposée, l estimation de β est faite à l aide d une technique de variable auxiliaire récemment présentée dans la littérature statistique. L algorithme développé est appliqué avec succès à des images d ultrasons tridimensionnelles.

Le chapitre 5 traite du calcul de la borne de Cramer-Rao (CRB) des paramètres d un champ aléatoire de Markov. Cette borne représente une limite inférieure de la covariance des estimateurs non biaisés de ces paramètres. Pour ces champs, la borne dépend de la constante de normalisation de la vraisamblance, dont le calcul est infaisable. Ce problème est résolu en proposant un algorithme Monte Carlo qui permet une estimation précise de la borne. L algorithme est appliqué aux modèles d Ising et de Potts. Des comparaisons avec la variance des estimateurs de l état de l art montrent que certains de ces estimateurs sont assez efficaces.

Introduction

Ultrasound imaging is a longstanding medical imaging modality with important applications in diagnosis, preventive examinations, therapy and image-guided surgery. In dermatologic oncology, diagnosis of potentially cancerous skin lesions requires performing a biopsy (a minor surgical intervention) and evaluating it histologically. Treatment for skin cancer generally consists in a second surgery to excise the tumor. However, the exact boundaries of the lesion are not always easy to assess at the time of the surgery and some procedures have to be repeated several times until all the tumor has been removed. Recent advances in high frequency transducers and 3D probes have opened new opportunities to examine skin noninvasively and assess tumor boundaries prior to excision using ultrasound images. However, changing dermatological practices requires developing ultrasound image processing methods that are specific to dermatology and to these new technologies.

The medical motivation of this thesis is to develop image processing methods that aid dermatologists to use in vivo ultrasound imaging to explore the structure of human skin in general and lesions in particular. More precisely, Bayesian methods are considered in order to perform tissue segmentation in ultrasound images. Four main methodological contributions are proposed:

- An α-stable statistical model of ultrasound echoes backscattered from the skin [5–7].
- A Bayesian segmentation method based on an α-Rayleigh mixture model and a Potts-Markov random field [3, 4, 8, 9].

- An MCMC algorithm to estimate the Potts regularization parameter within the segmentation method [10, 11].
- A Monte Carlo algorithm to compute the Cramer-Rao bound of the parameter of an Ising or a Potts Markov random field.

The thesis is organized into five chapters. Chapter I introduces the medical context motivating this work, provides general notions of skin tissues and ultrasound imaging and describes the type of ultrasound signals used in the experiments. Chapters II, III and IV have been presented as journal articles in [4, 5, 10].

Chapter II [5] studies the statistical distribution of single-tissue (i.e., homogeneous) regions in dermatological ultrasound images. Starting from the widely accepted point scattering model and through analytical developments, ultrasound signals backscattered from skin tissues are shown to converge to a complex Levy Flight random process with non-Gaussian α-stable statistics. The envelope or B-mode signal follows a generalized (heavy-tailed) Rayleigh distribution. This model generalizes the Gaussian framework and provides a formal representation for a new case of non-Gaussian statistics, where both the number of scatterers and the variance of their cross-sections tend to infinity. In addition, analytical expressions are derived to relate the α-stable parameters to scatterer properties. Several experimental results supported by excellent goodness-of-fit tests confirm the proposed statistical model.

Chapter III [4, 8] addresses the problem of jointly estimating the statistical distribution and the boundaries of multiple tissues in B-mode ultrasound images of skin. Based on the developments introduced in Chapter II, the distribution of multiple-tissue images is modeled as a spatially coherent finite mixture of heavy-tailed Rayleigh distributions. Spatial coherence inherent to biological tissues is modeled by enforcing local dependence between the mixture components using a Potts Markov random field (MRF). An original Bayesian algorithm combined with a Markov chain Monte Carlo (MCMC) method is then proposed to jointly estimate the mixture parameters and a label-vector associating each voxel to a

tissue. The resulting algorithm is successfully applied to the segmentation of in-vivo skin tumors in high frequency 2D and 3D ultrasound images.

The segmentation results obtained by the method presented in Chapter III depend on the amount of spatial correlation introduced by the Potts model, which is controlled by the granularity coefficient β. This important parameter is set heuristically by cross-validation. Chapter IV [10, 11] studies the estimation of the Potts parameter β jointly with the unknown parameters of a Bayesian model within an MCMC algorithm. This problem has received some attention in the recent image processing literature, as it would lead to unsupervised segmentation algorithms. Standard MCMC methods cannot be applied to this problem because performing inference on β requires computing the intractable normalizing constant of the Potts model. In the proposed segmentation method the estimation of β is conducted using an auxiliary variable technique recently presented in the statistics literature. The resulting algorithm is successfully applied to a 3D ultrasound image.

Chapter V addresses the problem of computing the Cramer-Rao bound of the granularity coefficient β of an Ising or a Potts-Markov random field. This bound provides a lower limit to the variance of estimators of β. For both distributions the bound depends on a normalizing constant that is generally intractable. This problem is addressed by proposing a Monte Carlo algorithm that can estimate the bound efficiently. The proposed algorithm is successfully applied to several Ising and Potts models. These results reveal that some of the state-of-the art algorithms to estimate β are close to being efficient.

Contexte médical

La peau est l organe le plus large de l organisme humain. Elle joue un rôle important dans la protection du corps des facteurs nuisibles exogènes et la régulation de la température. Elle est aussi responsable de la synthèse de la vitamine D. C est aussi l organe qui fournit le sens du touché et de la température. La peau humaine est composée de plusieurs couches stratifiées dont les principales sont l épiderme, le derme et l hypoderme.

L épiderme est la couche externe la plus fine (0.1 à 1.5 mm) représentant une barrière de protection. Il est composé de cellules fabriquées dans la strate basale de l épiderme et migrant progressivement vers la surface pour s assécher et disparaitre. Le derme est la seconde couche de la peau dont la fonction principale est mécanique. Son épaisseur varie entre 1.5 mm et 4 mm. Il est composé principalement de tissus de connexion (fibres d élastine et de collagène) et est traversé par un réseau dense de vaisseaux sanguins et renferme des glandes et les follicules pileux. Le derme consiste en deux strates : le derme papillaire en haut et le derme réticulaire en bas. Le derme papillaire est composé de fibres éparses connectant la strate sous-jacente à l épiderme. Il présente une jonction onduleuse avec l épiderme, dont l intérêt est important en cosmétologie et dermatologie. Le derme réticulaire est plus épais et formé de fibres plus denses d élastine et de collagène, ainsi que d autres structures. L hypoderme est la troisième couche de la peau composée principalement de graisses et de quelques fibres. Il joue un rôle d isolant thermique et permet de lier la peau aux os et aux muscles.

La peau est sujette à plusieurs maladies dont le cancer, la forme la plus fréquente de cancers en Europe en augmentation continue. Le mélanome de la peau est le type le plus rare des cancers de la peau, mais il est le plus dangereux (1000 morts par an en France). La prévention et le diagnostic du mélanome sont reconnus comme un enjeu de santé publique. Le diagnostic de lésions potentiellement malignes requiert des biopsies et de l histologie. Il a été constaté qu un très grand nombre d histologies de ce genre se révèle négatif engendrant néanmoins un coût et un trauma chez les patients. Le traitement des cas malins

se fait par chirurgie et nécessite une localisation assez précise des frontières de la tumeur. Cette précision est très difficile à atteindre et les chirurgiens prennent habituellement des précautions en ajoutant de larges marges.

Visualiser les structures de la peau in-vivo d une manière non invasive a de multiples intérêts en dermatologie. Cela permettrait en particulier de détourer correctement la tumeur avant la chirurgie. Cette visualisation offrirait aussi potentiellement des possibilités de diagnostic, de stadification et de suivi thérapeutique. De plus, c est un moyen de réduire les coûts et les délais de la prise en charge des patients.

Trois technologies complémentaires sont utilisées pour la visualisation de la peau : la microscopie confocale, l imagerie par cohérence optique, et l échographie. La microscopie confocale utilise le laser et des systèmes optiques pour obtenir des images 2D de plans parallèles à la surface de la peau. Ces images sont souvent d une grande résolution, proche de la microscopie classique. C est une modalité d images potentiellement intéressante pour différencier des tissus saints de tissus malins. La tomographie par cohérence optique (OCT) est une technologie d interférométrie capable de générer des images 2D de plans perpendiculaires à la surface de la peau de résolution micrométrique. La résolution verticale de l OCT permet de visualiser la peau à une plus grande profondeur que la microscopie confocale. L imagerie ultrasonore de haute résolution utilise des transducteurs et des sondes d ultrasons pour générer des images 2D de plans perpendiculaires à la surface. Les images obtenues permettent de visualiser la peau jusqu à l hypoderme mais sont souvent utilisées pour imager le derme. Pour le pronostic du mélanome, les images d ultrasons ont un grand intérêt car elles permettent de mettre en évidence le développement vertical de la tumeur dans le derme (c est-à-dire la profondeur de la tumeur). Cet indicateur clinique est un élément clé dans le choix de la méthode thérapeutique. De ce fait, cette thèse se consacre à l étude de ces images et au développement de méthodes de traitement pour aider les dermatologues à explorer la peau et détourer semi-automatiquement les tumeurs.

En imagerie ultrasonore, plusieurs caractéristiques sont importantes pour adapter le traitement des images. L imageur fournit en brut des signaux radiofréquences qui représentent des lignes verticales dans la peau. La juxtaposition de ces lignes donne une image RF en 2D. Afin de faciliter la visualisation des images, l enveloppe du signal complexe des signaux RF est calculée et des filtres sont appliqués donnant les images dites de mode B.

Les images utilisées dans cette thèse ont été acquise avec l imageur Dermcup d ATYS équipé d un élément unique focalisé à 25Mhz avec une sonde 3D de 100 Mhz d un pas latéral de 53 μ m. Les signaux ont été compensés en atténuation et les signaux RF traités dans le chapitre 2 n ont subis aucun autre traitement. Les images traitées dans les chapitres 3, 4 et 5 ont été construites par la transformation d Hilbert des signaux RF et sans aucun filtrage supplémentaire afin de préserver leurs caractéristiques.

L objectif de cette thèse étant d étudier des méthodes de traitement d images ultrasonores de la peau, l utilisation de modèles d observation appropriés est impérative. Le chapitre 2 est dédié à l étude de modèles statistiques pour décrire le modèle d observation.

Chapter 1

Medical Context

1.1 Skin

1.1.1 Skin Anatomy

Skin is the largest vital organ of the human body. It plays a key role in protecting the body from the environment (i.e., bacteria, toxic gases and liquids, ultraviolet radiation, etc.) and by helping regulate water-loss and body temperature. Skin also serves in vitamin D synthesis and participates in the protection of organs from impact damage. In addition, skin is fundamental to the sense of touch and temperature perception. Finally, human skin is composed of three very different layers, the epidermis, the dermis and the hypodermis (see Fig. 1.1).

The epidermis is the most external skin layer, it provides mechanical resistance and constitutes the first barrier against bacteria, toxic chemicals and ultraviolet radiation. It is also the thinnest skin layer, with its thickness generally varying from $0.05 - 0.15$ mm depending on sex, age and region of the body. The epidermis is composed of skin cells that are nourished by diffusion from the lower layers, since the epidermis itself is not directly irrigated by blood vessels. The main type of skin cells composing the epidermis are Merkel cells, keratinocytes, melanocytes and Langerhans cells. Cells are originated by mitoses at

the basal layer of the epidermis, situated at the epidermis-dermis junction, and migrate progressively towards the surface (referred to as stratum corneum) where they die and slough off.

Dermis is the skin layer beneath the epidermis and its principal function is to protect the body from mechanical stress and strain. This layer is approximately 1.5 mm thick in most parts of the body and 4 mm thick in the palms and soles. It is essentially composed of connective tissue (i.e., elastin and collagen fiber bundles), but also harbors blood vessels, a variety of glands, touch and heat receptors and hair follicles. The dermis is structurally divided in two strata, the upper or papillary dermis and the lower or reticular dermis. The papillary dermis is composed by loose fiber bundles that connect the reticular dermis to the epidermis. The upper side of the papillary dermis exhibits a characteristic wave-like or "bumpy" profile that is intertwined with the basal layer, increasing the strength and exibility of the connection between the dermis and the epidermis. It is precisely this profile that produces finger-prints on the fingers surface. The lower side of the papillary dermis merges with the reticular dermis, that provides skin with its characteristic elasticity and exibility. The reticular dermis is much thicker than the papillary dermis and is composed by very dense networks of elastin, collagen and reticular fiber, as well as capillary vessels, sensory receptors and hair follicles.

Lastly, the hypodermis is a layer of fat and loose fibers situated below the dermis. Besides from stocking fat and providing thermal insulation, the hypodermis plays a key role by fixating the dermis to the underlying bone and muscle and supplying it with blood and nerves.

1.1.2 Skin Cancer

Skin may suffer from numerous diseases, the most serious of which is skin cancer. According to the European Cancer Observatory (eu-cancer.iarc.fr), skin cancer is the most common form of human cancer in Europe and its incidence grows every year. There are

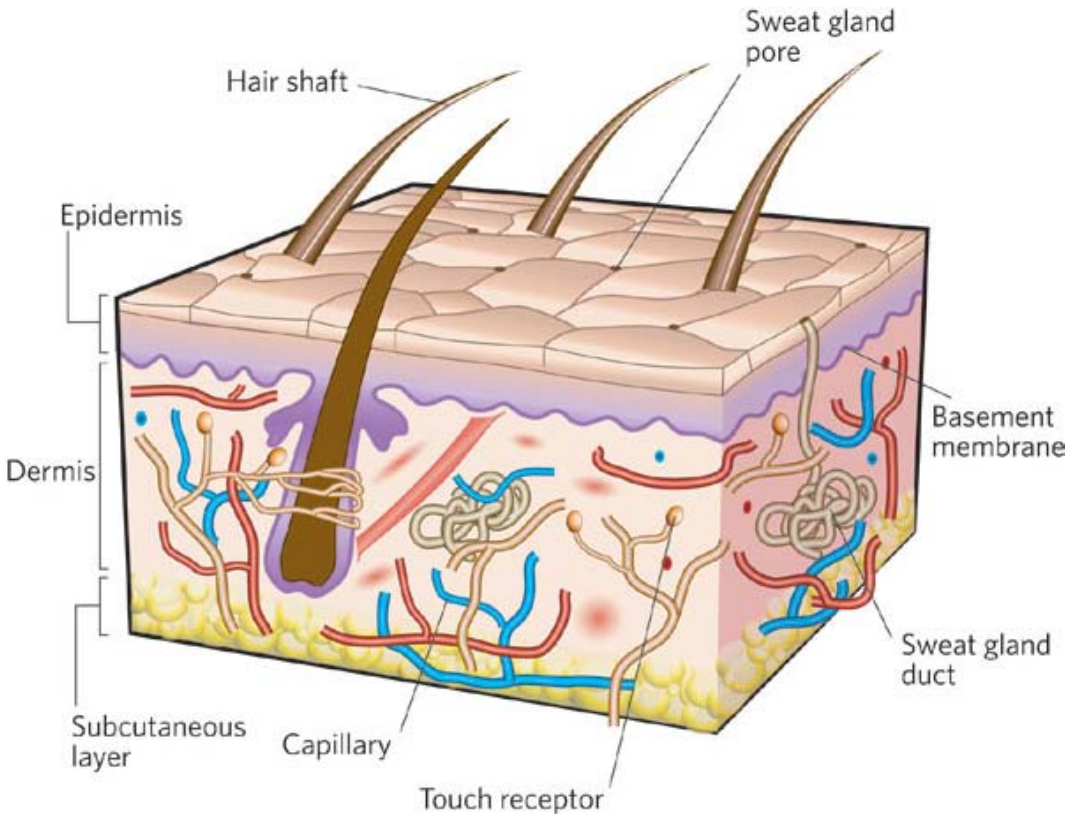


Figure 1.1: Human skin layers. Figure from [2]

three types of skin cancer: basal cell carcinoma, squamous cell carcinoma and melanoma. In France, over 90% of skin cancer patients suffer from either basal or squamous carcinomas, that are rarely fatal because they almost never metastasize (spread onto distant tissues). Unfortunately, skin melanoma does metastasize and is responsible for the death of over 1 000 people in France and 14 000 in Europe every year. In the United States, over 6 500 people die every year from skin melanoma and it is estimated that almost half of the people that live up to age 65 will develop some type of skin cancer at least once. Consequently, accurate diagnosis and treatment of skin cancers in general and skin melanoma in particular are recognized as main public health issues.

Diagnosis of potentially malignant skin lesions requires performing a biopsy and evaluating it histologically. These are minor surgeries that involve tissue fixation, excision, sectioning and staining. This is expensive, time consuming and produces some trauma

to the patient s skin. Biopsy results frequently reveal that the lesion is benign. However, when the presence of a malignant skin lesion is confirmed the patient must be treated for skin cancer. Treatment generally consists in a second surgery to excise the tumor. However, because tumors branch and spread under the skin surface, the exact boundaries of the lesion are not always easy to assess at the time of the surgery. As a result, surgeons take wide precaution margins around the suspected tumor boundaries and verify with histology that the excised tissue contains the entire tumor. Sometimes surgeries have to be repeated several times until all the tumor has been removed; this is expensive and time-consuming.

1.2 Noninvasive skin imaging

Noninvasive imaging of skin can be very valuable to dermatologic oncology in several ways. First of all, it can help skin cancer treatment by allowing surgeons to assess tumor boundaries from images prior to excision. Accurate determination of lesion depth is also very important for diagnosing tumor stage. Noninvasive images provide more reliable geometric information than conventional histology because they preserve tissue integrity, allowing in-vivo measurements and repeated imaging without tissue alteration. On the other hand, measurements based on histological images may be less reliable due to ex-vivo artifacts (i.e., tissue shrinkage), and measurements can only be performed once because histology samples are "destroyed" by serial sectioning (i.e., sample slicing). Images could also provide valuable diagnosis information regarding the nature of the lesion and reduce the need for biopsies. Moreover, in some cases skin cancer is treated using chemotherapy and noninvasive imaging can be used to monitor the tumor s response to the treatment. It is also worth mentioning that noninvasive techniques are generally capable of rapidly imaging skin at different depths without the delay or costs of conventional deeper histology levels. Finally, other potential applications of noninvasive imaging of skin include *"evaluation of non-tumorous skin lesions such as scleroderma, psoriasis, and contact dermatitis,*

determination of the depth of thermal burn injuries, studying the effects of photoaging and studying the effects of nuclear radiation on skin" [12]. Three complementary technologies with important clinical potential for noninvasive imaging of skin are summarized below.

1.2.1 Confocal microscopy:

Confocal microscopy (CM) is a new optical imaging modality with important potential for exploring the epidermis and the epidermis-dermis junction. Confocal microscopes use lasers and optical systems based on the principle of conjugate focal planes to generate 2D images of planes parallel to the skin surface. Modern MC systems can produce 3D images by vertically stacking 2D images acquired on different parallel planes. Because they use visible or nearly visible light, CM systems achieve a sub-micrometer lateral resolution that is comparable to that of standard optical microscopes. For these reasons CM technology is increasingly applied to study in-vivo skin and also as a preliminary step to biopsy in clinical dermatology. Precisely, CM is potentially interesting for *"differentiating between benign and malignant skin lesions, tumor margin mapping, monitoring response to medical or surgical treatments, and pathophysiologic study of inflammatory processes"* [13].

1.2.2 Optical coherence tomography:

Optical coherence tomography (OCT) is an interferometry-based imaging modality that uses near-infrared pulses to generate images of in-vivo skin with micrometer-resolution. Unlike CM, OCT systems generate images of planes oriented perpendicular to the skin surface and produce 3D images by stacking parallel vertical planes. OCT images are characterized by a high vertical resolution and can reach the upper papillary dermis, which is difficult to observe using CM. Already widely spread in ophthalmology, OCT imaging is slowly becoming a dermatology tool to study upper skin morphology (i.e., epidermis thickness and structure). The reader is referred to [14] for more details about the application of OCT to skin imaging.

1.2.3 High-frequency ultrasound:

Ultrasound imaging is widely used in medicine, with important applications in diagnosis, preventive examinations therapy and image-guided surgery. Ultrasound images are also frequently referred to as *echographies* because they are formed by ultrasound echoes. These images are generated by probing tissues with an ultrasound pressure wave focalized to target a specific zone in the body. Recent advances in high-frequency ultrasound technology have enabled the development of ultrasound imaging systems to explore the human skin in-vivo. These systems can produce 2D images of planes perpendicular to the skin surface that span from the epidermis down to the hypodermis, with vertical and horizontal resolutions of the order of the tens of microns (i.e., $30\mu m$) and one hundred microns (i.e., $120\mu m$) respectively. Some systems can also produce 3D images by stacking 2D images acquired on different parallel planes. High-frequency ultrasound systems equipped with 3D probes can be very valuable instruments to perform noninvasive diagnostics and preoperatory explorations of in-vivo dermis and hypodermis. Precisely, they can provided significant aid to surgeons to assess tumor boundaries and in cancer stage diagnosis. It is also worth mentioning that ultrasound images of skin can be very useful to determine the depth of burn injuries, which is a key factor for determining their gravity and the corresponding treatment. Over 500 000 people surfer from thermal burn injuries each year in France and 1000 of these lesions are fatal.

This thesis studies image processing methods to aid dermatologists in the use of high-frequency ultrasound imaging to explore the structure of in-vivo human skin and assess lesions noninvasively. The remainder of this chapter provides basic notions of ultrasound imaging of skin. The reader is invited to consult [15] for thorough details on ultrasound imaging and its medical applications, and [12] for a general introduction to ultrasound imaging of skin.

Basics of ultrasound imaging:

Ultrasound images are generated by emitting high-frequency acoustic pulses (pressure waves) into the body using an ultrasound probe, and subsequently recording the echoes backscattered by the biological tissues as the pulse propagates through them. Echoes are backscattered waves that arise when the emitted pulse encounters a change in the propagation medium (i.e., at the interphase between two mediums with different acoustic properties). In biological tissues, echoes are often produced by microscopic inhomogeneities (i.e., cell nuclei, fibers, etc.) that are acoustically different from their surrounding medium. Ultrasound signal formation models generally refer to these inhomogeneities as "*scatterers*" because they scatter a small portion of the energy of the emitted pulse back to the ultrasound probe. Contrast between tissues in ultrasound images results from differences in their population of scatterers. A medium with few scatterers will be almost transparent to ultrasound waves and appear dark in an ultrasound image. On the contrary, a medium rich in scatterers will appear bright.

Ultrasound pulses are emitted perpendicular to the skin surface and propagate towards the hypodermis following a straight line. As the pulse propagates across the different skin layers it encounters scatterers that generate echoes. An ultrasound signal is then obtained by recording these echoes as they reach the ultrasound probe back on the skin surface. This one-dimensional time-series is denominated the *radio-frequency* (RF) ultrasound signal, and contains information about the tissue contained in the linear path of the pulse. Several RF signals must then be juxtaposed to obtain a 2D or 3D image with information about the tissue within a plane or a volume. In dermatologic ultrasound systems this is achieved by translating laterally the sensor or transducer inside the ultrasound probe over a grid of points on the skin surface and recording a signal at each point.

Ultrasound imaging systems are generally tailored for specific applications, and an important design parameter is the ultrasound frequency. The frequencies normally applied in medical ultrasound systems lie between 3 MHz and 50 MHz depending on the region of the

body to be examined. Frequency is subject to an important trade-off, increasing frequency improves image resolution but reduces penetration into the body. This is because ultrasound pulses are attenuated as they propagate through tissue and the amount of attenuation they undergo increases with their frequency. Medical ultrasound systems are designed to target a specific depth and their frequency is chosen accordingly. Dermatologic ultrasound systems generally target the dermis and use frequencies in the range $16-40$ MHz. They can observe tissue down to approximately 3 mm below the skin surface. There are other important parameters characterizing an ultrasound system, such as the band-width, the focal length, the appodization and the sampling rate (see [15] for details about these parameters).

Ultrasound signals:

Raw ultrasound signals generally undergo two processing steps to put them into an image form. The first step is the compensation of the attenuation that the pulse and the backscattered echoes suffered as they traveled through the tissue. This is achieved by multiplying the RF ultrasound signal with a time-gain-compensation (TGC) function designed to coarsely counteract the effect of attenuation in the different tissues appearing in the image. The second step consists in detecting the envelope of the compensated RF ultrasound signal. In modern ultrasound systems, envelope detection is generally achieved by computing the (complex-valued) analytic extension of the RF signal using the Hilbert transform, and then taking its amplitude (note that envelope detection is equivalent to amplitude demodulation). Fig. 1.2 shows a synthetic high-frequency RF signal in black and its envelope signal in red [1]. These two processing steps produce the so-called *envelope* ultrasound signals, whose juxtaposition defines an envelope image. Envelope images are also known as brightness or *B-mode* images. Finally, envelope images often have a very high dynamic range that cannot be observed by the human eye. To ease their visual interpretation, envelope images are conventionally displayed in logarithmic scale. These steps are summarized in Fig. 1.3.

[1] signals generated using K-wave [16]

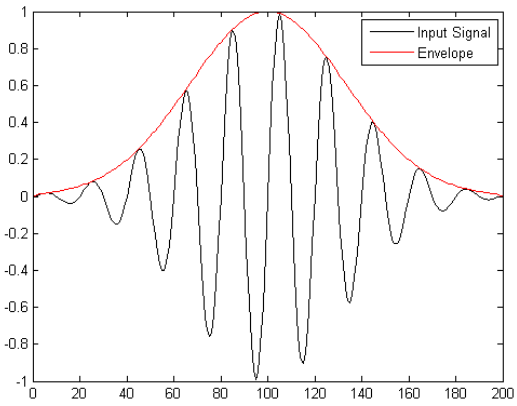


Figure 1.2: A high-frequency signal (black) and its envelope signal(red).

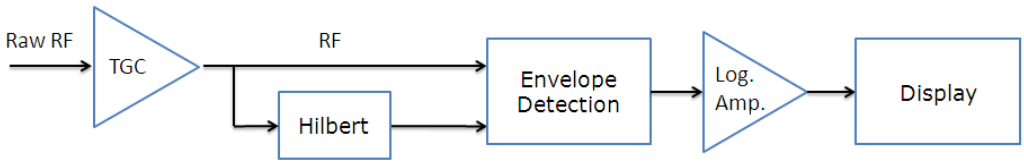
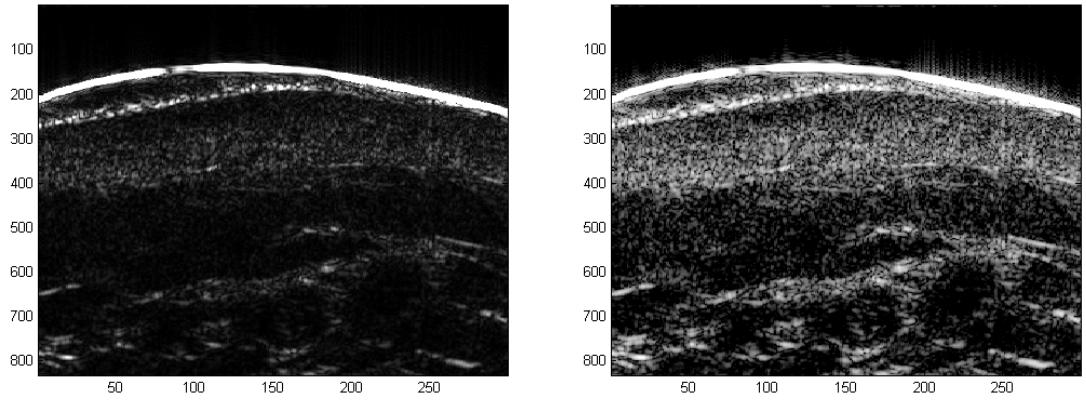


Figure 1.3: Block diagram for ultrasound system back-end.

Data corpus:

The corpus of ultrasound signals and images used in the experiments reported in this thesis have been acquired with an *ATYS dermocup* system [2], equipped with a single-element focalized 25MHz wide-band (40-percent) 3D probe sampled at 100MHz with a 53μm mechanic lateral step. All signals have been automatically compensated for attenuations by the acquisition system. The RF signals used in Chapter II did not undergo any kind of decimation, interpolation, compression or filtering. Similarly, the envelope signals and images used in Chapters II, III and IV have been obtained by Hilbert transform and without any additional processing. However, to ease visualization ultrasound images are always displayed using a logarithmic scale. Fig. 1.4(a) shows one slice of a 3D envelope ultrasound image of skin from the data corpus. The same image is displayed in Fig. 1.4(b) using a logarithmic scale.

[2] Atys Medical, France (http://www.atysmedical.com/pages/produits/dermcup.php)



(a) Linear scale display. (b) Logarithmic scale display.

Figure 1.4: A slice of a 3D envelope ultrasound image of in-vivo human skin.

The aim of this thesis is to study processing methods for skin ultrasound images. Using appropriate observation models to describe the data is fundamental to the success of the proposed methods. Structures in biological tissues being extremely complex, their interaction with ultrasound waves can be well described using stochastic processes and statistical observation models. Chapter II studies a statistical model relating backscattered ultrasound signals with the scattering structures in skin.

Modélisation de signaux ultrasonores cutanés par des processus α-stables symétriques

Les études analytiques de la diffusion d ultrasons dans les milieux aléatoires sont souvent faites dans le cadre de la diffusion ponctuelle. Selon ce modèle, un milieu aléatoire est représenté comme un media de propagation homogène contenant des diffuseurs ponctuels non homogènes, dont les tailles et les positions sont considérées aléatoires. Ces diffuseurs sont trop petits pour être visibles individuellement. Ainsi, le signal rétrodiffusé est modélisé comme l interférence incohérente d un large nombre d ondes diffusées aléatoires, phénomène nommé speckle [17]. Les expressions analytiques de la distribution statistique des signaux rétrodiffusés sont souvent dérivées par le théorème de la limite centrale. En conséquence, le signal converge en distribution à une loi Gaussienne quand le nombre de diffuseurs par cellule de résolution devient grand. Tout aussi, l enveloppe du signal (ou signal mode B) converge à une distribution de Rayleigh.

Cependant, des études ont établi que les statistiques du speckle de tissus biologiques ne sont pas bien représentées par l hypothèse Gaussienne ou Rayleigh [12, 18, 19]. Il a été reporté par exemple que les tissus de la peau et du sein sont mieux décrits par les distributions Gamma généralisée [12] ou Nakagami [20]. Ce chapitre étudie les statistiques du speckle dans les images d ultrasons dermatologiques. Nous montrons analytiquement que ces statistiques ont une distribution limite α-stable non Gaussienne et que le signal ultrason rétrodiffusé constitue un processus stochastique stable Levy Flight. En outre, il est établi que l enveloppe du signal suit une distribution Rayleigh généralisée à queue lourde. Finalement, nous établissons que les sections des diffuseurs ont une loi de puissance et nous dérivons des expressions analytiques pour relier les paramètres α-stables aux propriétés des diffuseurs. Dans le reste de cette thèse, nous considérons uniquement les ultrasons de la peau. Cependant, les résultats de ce chapitre sont généraux et peuvent être appliqués à la modélisation d autres types de données. Le chapitre est organisé comme suit. La

section II introduit le modèle de formation de signal adopté dans cette thèse. La section III établit que les signaux ultrasons convergent à un processus aléatoire complexe de type Levy Flight avec des statistiques α-stables non Gaussiennes, et que son enveloppe suit une distribution Rayleigh généralisée. Les expressions analytiques reliant les paramètres α-stables aux propriétés des diffuseurs sont étudiées dans la section IV. Des simulations et des expériences sur des données in-vivo sont présentées dans les sections V et VI. Des conclusions sont finalement reportées en section VII.

Chapter 2

Modeling Ultrasound Echoes in Skin Tissues using Symmetric α-Stable Processes

Starting from the widely accepted point scattering model, this chapter establishes through analytical developments that ultrasound signals backscattered from skin tissues converge to a complex Levy Flight random process with non-Gaussian α-stable statistics. The envelope signal follows a generalized (heavy-tailed) Rayleigh distribution. It is shown that these signal statistics imply that scatterers have heavy-tailed cross-sections. This model generalizes the Gaussian framework and provides a formal representation for a new case of non-Gaussian statistics, where both the number of scatterers and the variance of their cross-sections tend to infinity. In addition, analytical expressions are derived to relate the α-stable parameters to scatterer properties. Simulations show that these expressions can be used as rigorous interpretation tools for tissue characterization. Several experimental results supported by excellent goodness-of-fit tests confirm the proposed analytical model. Finally, these fundamental results set the basis for new echography processing methods and quantitative ultrasound characterization tools.

2.1 Introduction

Analytical studies of ultrasound scattering in random media are generally conducted within the widely accepted point scattering framework. According to this model, a random medium is represented as a homogenous propagation medium with punctual inhomogeneities denominated scatterers, whose size and position are considered random, and which are too small to be resolved individually. Therefore, the backscattered signal is modeled as the incoherent interference of a large number of randomly scattered waves, a stochastic phenomenon denominated speckle [17].

Analytic expressions for the statistical distribution of the backscattered signal are usually derived by means of the central limit theorem. Accordingly, the backscattered signal converges in distribution to a Gaussian law as the number of scatterers per resolution cell increases. Similarly, the envelope signal, or B-mode signal, converges towards a Rayleigh distribution.

However, previous studies on the statistical distribution of speckle in biological tissues have established that the observed distributions are not always well represented by the Gaussian or Rayleigh assumption [12, 18, 19]. For instance, it has been reported that skin and breast tissues are better described by the Generalized Gamma [12] or the Nakagami distributions [20].

This chapter studies the statistical distribution of speckle in dermatological ultrasound images. We show analytically that speckle statistics have non-Gaussian α-stable limit distributions and establish that the backscattered ultrasound signal constitutes a stable process, namely a *Levy Flight*. Furthermore, the envelope of the backscattered signal is shown to follow a generalized (heavy-tailed) Rayleigh distribution. Finally, we establish that scatterers have heavy-tailed cross-sections and derive analytical expressions to relate the α-stable parameters to scatterer properties. Henceforth, we consider only ultrasound images of skin; however the results in this chapter are general and can be used to model other types of data.

The chapter is organized as follows: The signal formation model used in this work is introduced in Section II. Section III establishes that ultrasound signals converge to a complex Levy Flight random process with non-Gaussian α-stable statistics whose envelope follows a generalized Rayleigh distribution. Analytical expressions to relate the α-stable parameters to scatterer properties are studied in Section IV. Simulations and experiments on in vivo data are presented in Sections V and VI. Conclusions are finally reported in Section VII.

2.2 Problem Statement

Let $T \subset \mathbb{R}^+$ be a region of interest (henceforward ROI) that falls within a single biological tissue. Let $x : T \rightarrow \mathbb{R}$ denote the backscattered RF ultrasound signal received at the transducer. Similarly, let $r : T \rightarrow \mathbb{R}^+$ denote the envelope or B-mode signal $r(t) = |x(t) + jy(t)|$, where $x(t) + jy(t)$ stands for the analytic extension of $x(t)$.

2.2.1 Ultrasound Scattering Model

This chapter represents $x(t)$ in the framework of the widely accepted point scattering model [17][p. 438]. In addition, it is assumed that all scatterers in the region T interact with identical interrogating ultrasound pulses and that there is no multiple-scattering. Then, the backscattered signal can be modeled as the summation of several randomly imposed and scaled replicas of the interrogating pulse $p(t)$, backscattered from a population of point scatterers: [17, p. 438],[21, 22].

$$x(t) = \sum_{i=1}^{M} a_i p(t - \tau_i) \tag{2.1}$$

where $a_i \in (0, 1)$ is the cross-section of the i_{th} scatterer and $\tau_i \in T$ is the relative position of the i_{th} backscattered wave.

We observe that the signal formation model defined in (2.1) describes the backscattered signals as a function of M, $\boldsymbol{\tau} = \{\tau_1, .., \tau_M\}$ and $\mathbf{a} = \{a_1, .., a_M\}$, which are all unknown in biological tissues and cannot be directly observed. In fact, expression (2.1) is used as inverse interpretation models for tissue properties. Accordingly, M, $\boldsymbol{\tau}$ and $\mathbf{a}$ are represented as random variables and samples of the backscattered signal $\{x(t_1), .., x(t_N)\}$ as the realization path of a real-valued process $\{X_t, t \in T\}$ conditional to $\{M, \boldsymbol{\tau}, \mathbf{a}\}$. Then, inference on $\{M, \boldsymbol{\tau}, \mathbf{a}\}$ can be conducted under the hypothesis that $\{X_t\}$ is ergodic (recall that T denotes a single-tissue ROI). Ergocidity in T is assumed because there is only one observation available per resolution cell.

From a methodological point of view inference on tissue properties is performed by first proposing statistical properties for M, $\boldsymbol{\tau}$ and $\mathbf{a}$ and then deriving $f_{X_t}(x)$ (the law of X_t) and comparing it against the empirical distribution. If $\{\tau_1, ..., \tau_M\}$ are *i.i.d.* randomly spread in T, $\{a_1, ..., a_M\}$ are also *i.i.d.* and M is sufficiently large; then in virtue of the central limit theorem it is possible to model X_t as follows [17, 21, 22]:

$$X_t = \eta_t \sqrt{v(t)}; \tag{2.2}$$

where $\{\eta_t, t \in T\}$ is *fully developed speckle* noise (i.e., a standard zero-mean Gaussian process) and $v(t)$ denotes diffuse backscattered power at t.

One important result arises when scatterers are uniformly spread in T and have identical cross-sections. This case is generally referred to as *Gaussian* or *Rayleigh* because $v(t)$ is constant ($v(t) = \sigma^2$) and realization paths $\{x(t_1), .., x(t_N)\}$ exhibit Gaussian statistics $f_{X_t}(x) = \mathcal{N}(0, \sigma^2)$ [22]. Similarly, samples of the envelope or B-mode signal $r(t)$ will follow a Rayleigh distribution [22]. Accordingly, tissues that exhibit Gaussian statistics are characterized by estimation of σ^2, which can be shown to be proportional to $M\langle a_i^2 \rangle$ (where $\langle a_i^2 \rangle$ is the 2nd order moment of the scattering cross-sections). A considerable advantage of modeling with Gaussian laws is linearity and distribution invariance, which guarantees

that a Gaussian sample path will remain Gaussian for linear observation systems (i.e., convolution).

Another important result is the *Rician* case, which extends the *Rayleigh* envelope model by taking into account coherent power backscattered from periodic spatial configurations. Note that the *Rician* envelope case remains an *RF* Gaussian since $f_{X_t}(x) = \mathcal{N}(\varepsilon, \sigma^2)$ (where ε denotes coherent power and σ^2 diffuse power).

However, as previously stated, signals backscattered from biological tissues are often poorly described by *Gaussian* (*Rayleigh-Rice*) statistics. The literature proposes a large variety of generalizations of the *Rayleigh* and *Rice* distributions to model the envelope statistics of *non-Gaussian* tissues. Some important envelope distributions are *K* [18], *Homodyne K* [23], *Nakagami* [19], *Nakagami Gamma* [24], *Generalized Gamma* [25] (also referred to as Generalized Nakagami [26]), *Weibull* [25], *Rician Inverse Gaussian* [27] and *Generalized-Nakagami Inverse Gaussian* [28]. The next section discusses the interpretation of non Gaussian statistics.

2.2.2 Interpretation of Scattering in Human Dermis

The prevailing interpretation states that non-Gaussian statistics represent partially developed speckle. This interpretation relates non-Gaussianity to scattering microstructures at a scale inferior to the resolution-cell. Because a large number of scatterers (at least 10 [29]) is required to enforce the central limit theorem, partially developed speckle can be interpreted as low scatterer density. In fact, since the number of scatterers in tissues is unlikely to be that small, non-Gaussianity is generally interpreted as the result of relatively low scatterer density combined with high scattering cross-section variance and scatterer clustering.

Skin tissues challenge this interpretation because they posses a particularly dense population of inhomogeneities, as they are mainly composed of cells and intertwined networks of collagen and elastin fibres, and yet they exhibit strong non-Gaussian statistics. In view of that, Raju et al. [25] concluded that clustering and cross-section variance in skin tissue

had to be extremely high; otherwise the statistics of the RF backscattered signal would necessarily converge to a Gaussian distribution.

A second important interpretation of non-Gaussian statistics is that the backscattered power $v(t)$ exhibits strong fluctuations in T as a result of scatterer clustering and cross-section heterogeneity at a scale larger than the resolution cell. Therefore, even if scatterers are numerous enough to drive η_t towards Gaussian statistics, observations from different time instants will follow different Gaussian laws $f_{X_t}(x) = \mathcal{N}(0, v(t))$ and the overall statistics will be non-Gaussian. This case has been studied in [30], where the backscattered power is modeled as the realization path $\{v(t_1), \ldots, v(t_N)\}$ of a stationary random process $\{V_t : t \in T\}$ that takes values in $\mathbb{R}^+$ with probability $f_V(v)$, then it is possible to express the probability $f_{X_t}(x)$ as a continuous Gaussian mixture or compound probability distribution:

$$f_{X_t}(x) = \int_0^\infty \mathcal{N}(x|0, v) f_V(v) dv \tag{2.3}$$

Note that (2.3) describes the distribution of the RF backscattered signal; envelope signals require compounding with a Rayleigh distribution (the interested reader is invited to consult [30] for more details about compound distributions and their application to ultrasound echo modeling).

Precisely, it was recently shown that the generalizations of the Rayleigh distribution used for envelope signal modeling can be restated or approximated as compound probability distributions [30]. By doing so, each envelope distribution can be associated to a specific distribution $f_V(v)$. Representing distributions in a compound way has the advantage of giving a clear physical meaning to the empirically observed statistics.

However, we believe that it is still not clear whether this interpretation is closer to the physical reality. In fact, any envelope distribution, which are generalized Rayleigh laws, can be approximated by a sufficiently complex Rayleigh mixture. Dermis, despite having complex collagen and elastin microstructures, is relatively homogeneous at a macro-

structure level. It is therefore unlikely that $v(t)$ undergoes strong fluctuations at a macroscopic level.

Yet another explanation of non-Gaussian statistics is that the signal formation model proposed in (2.1) would be too inaccurate. Non-Gaussian statistics can arise from dependence between a_i and t_i [31] and from multiple scattering [32]. Unfortunately, these assumptions lead to complex models that are often unsuitable for tissue characterization because they depend on medium-specific information which is not available for most biological tissues. The absence of reliable tissue-specific information translates into additional degrees of freedom which in turn reduce the overall confidence on the model.

This chapter presents a new explanation of non-Gaussian statistics that is consistent with the point scattering framework. Instead of exploring distributions that generalize the Gaussian or Rayleigh laws we consider the generalizations of the central limit theorem that could cause X_t to converge towards non-Gaussian distributions.

2.3 An α-Stable Point Scattering Model

In this section we address the case where the point scattering model converges towards non-Gaussian distributions. In particular, we investigate the possible limit distributions for the backscattered signal and discuss their implications. Henceforth all signals will be considered processes and no distinctions will be made between the signal $x(t)$ and its model, the stationary process X_t.

If the distribution of the backscattered signal $x(t)$, as defined by the point scattering model in section 2.2, converges as $M \to \infty$ to a non-Gaussian distribution, then it can be shown that the following three statements are true [1]:

1. The limiting distribution of $x(t)$ must be a symmetric alpha stable distribution with characteristic exponent $\alpha \in (0, 2)$.

[1] part of this chapter has been presented in [6, 7]

2. The limiting distribution of the envelope signal $r(t)$ is a heavy-tailed Rayleigh distribution (denoted $\alpha\mathcal{R}$).

3. The distribution of the scattering cross-sections $f_A(a_i)$ is heavy-tailed (i.e., $\mathrm{Var}(a_i) \rightarrow \infty$).

These results put forward that non-Gaussian distributions may arise as a limit case, where both the number of scatterers per resolution cell and the variance of their cross-sections tend to infinity.

2.3.1 Limit distributions of the RF backscattered signal

At time t, the random variable $X_t = a_1 p(t - t_1) + \ldots + a_M p(t - t_M)$ is defined as a sequence of M *i.i.d.* random variables. A function $\mathcal{L}(X_t)$ is said to be a limit distribution of X_t if for every $M > 0$ there is a pair of values d_M and μ_M such that

$$\lim_{M\to\infty} f\left(\frac{a_1 p(t - t_1) + \ldots + a_M p(t - t_M) - \mu_M}{d_M}\right) = \mathcal{L}(X_t) \tag{2.4}$$

The α-*stable* distributions are the only limit distributions or *domains of attraction* of *i.i.d* random summands [33, 34]. In other terms, all possible limit distributions of the point scattering model belong to the stable family. The celebrated central limit theorem states that the Gaussian distribution is the domain of attraction of finite variance sequences. Random summands whose distribution is symmetric around zero, as it is the case of incoherently backscattered waves, are in the domain of attraction of a special group of α-stable distributions denominated symmetric α-stable (henceforward $S\alpha S$):

$$X_t = \sum_{i=1}^{M} a_i p(t - \tau_i) \underset{M\to\infty}{\overset{d}{\rightarrow}} S\alpha S(\alpha, \gamma) \tag{2.5}$$

where $\xrightarrow[M\to\infty]{d}$ denotes convergence in distribution as M increases. The family of $S\alpha S$ distributions is fully characterized by the characteristic function:

$$E \exp \jmath\theta x = e^{-\gamma^\alpha |\theta|^\alpha} \tag{2.6}$$

where $\alpha \in (0, 2]$ is the characteristic index and $\gamma \in \mathbb{R}^+$ the spread. Note that the upper bound $\alpha = 2$ represents the characteristic function of the Gaussian distribution ($S\alpha S(2, \gamma) \equiv \mathcal{N}(0, \sigma^2 = 2\gamma)$) and $\alpha = 1$ that of the Cauchy distribution. Similarly to the Gaussian case, section 2.4 shows that γ is proportional to the number of scatterers and to the variability of the scattering cross-sections.

Finally, we conclude that if the backscattered signal, as defined in the point scattering framework, converges to non-Gaussian distributions as M increases, then it must converge to a $S\alpha S$ distribution with characteristic exponent $\alpha \in (0, 2)$. We wish to stress the fact that $S\alpha S$ laws represent a non-Gaussian case of *fully developed* speckle since they are limit distributions.

This theoretical result is in accordance with other studies that have empirically observed that signals backscattered from biological tissues were well described by $S\alpha S$ distributions. Stable distributions have been used for quantitative ultrasound lesion assessment [35] and as prior distribution for ultrasound compressive sensing [36]. Stable statistics have been previously explained by interpreting ultrasound echoes as the result of a power-law shot-noise process (PLSN) [37]. The PLSN model is closely related to a point scattering model where the scatterer density follows a Poisson distribution [38]. The PLSN model makes the additional hypothesis that backscattered pulses have a power-law decay $p(t) \propto t^{-\frac{1}{\alpha}}$ and that it is this particular pulse shape that leads to $S\alpha S$ distributions.

Section 2.4 establishes analytically that $S\alpha S$ statistics arise naturally within the point scattering framework when scatterers satisfy a number of properties. These conditions for

convergence provide a formal basis to infer information about scatterers from the empirical statistics.

2.3.2 Limit distributions of the envelope backscattered signal

Let $x(t) + \jmath y(t)$ be the analytic extension of the backscattered signal $x(t)$

$$x(t) + \jmath y(t) = \sum_{i=1}^{M} a_i \left[p(t - \tau_i) + \jmath \tilde{p}(t - \tau_i) \right] \tag{2.7}$$

where $p(t) + \jmath \tilde{p}(t)$ is the analytic extension of the interrogating pulse $p(t)$. Consider a complex-valued process $\{X_t + \jmath Y_t, t \in T\}$ whose realization path is $x(t) + \jmath y(t)$.

Similarly to the Gaussian case, the distribution of the envelope or B-mode signal $r(t) = |x(t) + \jmath y(t)|$ is derived by assuming that $p(t - \tau_i)$ and $\tilde{p}(t - \tau_i)$ are independent (i.e., $|X_t + \jmath Y_t|$ is isotropic). In the context of narrow-band ultrasound imaging this is equivalent to supposing that the pulse s *phase* is uniformly distributed in $(0, 2\pi)$. Noting that (2.7) defines a sequences of *i.i.d.* complex random variables, the limit distributions of $X_t + \jmath Y_t$ belong to the complex $S\alpha S$ family, i.e., as M increases [33]

$$X_t \xrightarrow{d} S\alpha S(\alpha, \gamma), \quad Y_t \xrightarrow{d} S\alpha S(\alpha, \gamma). \tag{2.8}$$

Then, by using polar coordinates ($X_t + \jmath Y_t = R_t \left[\cos(\varphi_t) + \jmath \sin(\varphi_t) \right]$) and marginalizing w.r.t. the phase angle φ_t it can be shown that the limit distribution of the envelope R_t is a generalized (heavy-tailed) Rayleigh distribution [39, p.118]:

$$R_t \xrightarrow{d} \alpha\mathcal{R}ayleigh(\alpha, \gamma) \tag{2.9}$$

where

$$\alpha\mathcal{R}ayleigh(R_t = r|\alpha, \gamma) = \int_0^\infty r\lambda \exp[-(\gamma\lambda)^\alpha] J_0(r\lambda)\, d\lambda \tag{2.10}$$

and where J_0 stands for the 0_{th} order Bessel function of the first kind.

This result is in accordance with other studies that have observed that the tails of the histogram for pre-Rayleigh cases are longer and heavier than those of the K and Nakagami distributions [26]. It is important to model correctly the empirical tails: *"the tails of the density functions and areas occupied by the tails control the type I and type II errors (probabilities of miss and false alarm). Thus, the inadequacy of these distributions in their ability to match their tails to the data histogram may limit their ability to model the statistics of the backscattered echo and, consequently, the ability of their parameters to classify tissues"* [26].

At last, it is not uncommon to represent the point scattering model as a random walk in the complex plane [40], where each backscattered wave constitutes a small random step. Under the Gaussian assumption this random walk is in fact a Brownian motion. The $S\alpha S$ model generalizes this result and represents the backscattered signal as a stable process [33]. More specifically, non-gaussian symmetric stable processes are referred to as *Levy Flights*. Levy ights are a powerful model that extends the Wiener process (Brownian motion) to phenomena with heavy-tailed *i.i.d* increments.

2.3.3 Infinite variance scattering cross-sections

As explained previously, non-Gaussian statistics are generally interpreted as the result of relatively low scatterer density combined with high scattering cross-section variance. However, because skin tissues possess a very high scatterer density, Raju et al. [25] concluded that cross-section variance in skin tissues had to be extremely high. Indeed, we now show that non-Gaussian limit distributions arise when the cross-section variance tends to infinity.

Let $\sum_{i=1}^{M} a_i p(t - \tau_i)$ be a sequence of M *i.i.d.* random variables with a non-Gaussian limit distribution, and $p^* < \infty$ the maximum amplitude of the interrogating pulse (i.e., $p^* = sup(|p(t)|)$). We will assume that the a_i has finite variance and show that this leads to a contradiction.

Suppose that $\mathrm{Var}(a_i) < \infty$. From the definition of the variance, this implies that $\mathrm{E}(a_i) < \infty$. Moreover, the moments $\mathrm{E}\left[p(t-\tau_i)\right]$ and $\mathrm{Var}\left[p(t-\tau_i)\right]$ are known to exist because the support of $p(t-\tau_i)$ is bounded by p^*. Then, the product $a_i p(t-\tau_i)$ has a variance, and it is given by [41]

$$\begin{aligned}\mathrm{Var}\left[a_i p(t-\tau_i)\right] =& \mathrm{Var}\left[p(t-\tau_i)\right]\mathrm{E}(a_i)^2 + \mathrm{E}\left[p(t-\tau_i)\right]^2 \mathrm{Var}(a_i) \\ &+ \mathrm{Var}\left[p(t-\tau_i)\right]\mathrm{Var}(a_i).\end{aligned} \tag{2.11}$$

Clearly $\mathrm{Var}\left[a_i p(t-\tau_i)\right] < \infty$ because all the moments intervening in the r.h.s. of (2.11) are finite. However, in view of the central limit theorem, the sequence $\sum_{i=1}^{M} a_i p(t-\tau_i)$ should then have a Gaussian limit distribution. This contradiction arises from the assumption that $\mathrm{Var}(a_i) < \infty$. We therefore conclude that if the sequence $\sum_{i=1}^{M} a_i p(t-\tau_i)$ has a non-Gaussian limit distribution, then the distribution of the scattering cross-sections must be heavy-tailed, i.e., $\mathrm{Var}(a_i) \to \infty$.

2.4 Physical interpretation of the $S\alpha S$ parameters

The parameters of the $S\alpha S$ and $\alpha\mathcal{R}ayleigh$ distributions have a clear analytical relationship with those of the point scattering model. This section establishes that there are several possible ways to interpret these parameters depending on how we represent scattering structures.

2.4.1 Standard interpretation of α and γ

The standard or first interpretation to α and γ assumes that scatterers are strongly inhomogeneous at a sub-resolution-cell scale. This interpretation is appropriate for complex scattering structures as those found in skin tissues.

Tail parameter α

Let $f_A(a_i)$ and $f_P(p_i)$ denote the probability-density-functions of the random variables a_i and $p_i = p(t - \tau_i)$. Appendix A shows that if X_t has a $S\alpha S$ limit distribution with characteristic exponent $\alpha \in (0, 2)$, then the densities f_A and f_P must verify

$$\lim_{z_i \to \infty} \frac{\int_0^{p^*} f_P(p_i)[\int_{z_i/p_i}^{\infty} f_A(\rho)\, d\rho]\, du_i}{\int_0^{p^*} f_P(p_i)[\int_{z_i/p_i}^{\infty} f_A(l\rho) l\, d\rho]\, du} = l^{\alpha}, \quad \forall l > 0 \tag{2.12}$$

where $z_i = a_i p_i$ and $p^* = sup(p(t))$. This condition is verified for all $f_P(p_i)$ when the scattering cross-sections are distributed according to a Pareto distribution with tail exponent α, i.e.,

$$f_A(a_i) = \alpha \frac{a_m^{\alpha}}{a_i^{\alpha+1}} \tag{2.13}$$

where $a_m > 0$ is the distribution s mode.

Note that (2.13) is closely related to the fact that $f_A(a_i)$ is heavy-tailed (i.e., $\mathrm{Var}(a_i) \to \infty$). Indeed, from (2.13) it can be shown that the statistical moments of a_i verify the following inequality

$$\langle a_i^m \rangle \geq \int_{\xi}^{\infty} L(a_i) a_i^{m-\alpha-1} da_i + o\left(\int_{\xi}^{\infty} L(a_i) a_i^{m-\alpha-1} da_i \right) \tag{2.14}$$

where $\xi \in \mathbb{R}^+$. These moments exist ($\langle a_i^m \rangle < \infty$) only if the improper integral $\int_{\xi}^{\infty} L(a_i) a_i^{m-\alpha-1} da_i$, which dominates (2.14), converges to a finite value $L < \infty$. This is true if and only if $m < \alpha$. However, by definition $\alpha < 2$ thus $m < \alpha < 2$. As a result the distribution of the scattering cross-sections has no statistical moments higher than α, nor a variance.

We conclude that the $S\alpha S$ parameter α is closely related to the statistical properties of the scattering cross-sections. Indeed, the distribution of the scattering cross-sections is heavy-tailed with tail exponent α.

Scale parameter γ

The scale γ generalizes the Gaussian representation of tissue echogenicity. Specifically, appendix B shows that γ is related to the point scattering model as follows

$$\gamma = D^*(\alpha)\sqrt[\alpha]{M}a_m \tag{2.15}$$

where $D^*(\alpha) = \sqrt[\alpha]{\frac{2\pi\langle p_i^\alpha\rangle}{\Gamma(\alpha)\sin(\frac{\pi\alpha}{2})}}$, M is the number of scatterers, $\langle p_i^\alpha\rangle$ is the α-th fractional moment of $p(t-\tau_i)$ and the positive value a_m is given by

$$a_m = \lim_{a_i\to\infty} a_i^\alpha F_A(a_i) \tag{2.16}$$

where $F_A(a_i)$ is the cumulative function of the scattering cross-sections. For Pareto distributions (2.13), a_m (2.16) corresponds to the distribution s mode.

2.4.2 Compound interpretation of α and γ

The second interpretation to α and γ is based on the compound representation that assumes that scatterers are homogenous at a sub-resolution-cell scale but have cross-sections that vary strongly at a larger scale. Recent works have shown that most envelope distributions in the literature could be expressed or approximated as compound density distributions [30]. This representation has the advantage of providing a common model for all distributions, which are described in terms of a modulated distribution (e.g. Rayleigh, Rice) and a modulating distribution (e.g. Gamma, Inverse Gamma, Inverse Gaussian). For completeness we present compounds representations for the $S\alpha S$ and $\alpha\mathcal{R}ayleigh$ distributions and discuss the interpretation of α and γ in this framework.

Univariate and sub-Gaussian multivariate $S\alpha S$ distributions have a well known compound representation [33, p.77]

$$p_{S\alpha S}(x|\alpha,\gamma) = \int_0^\infty \mathcal{N}(x|0,\sigma^2) S_{\frac{\alpha}{2}}(\sigma^2|\gamma\cos\left(\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}},1,0)d\sigma \tag{2.17}$$

where the modulated law $\mathcal{N}(x|0,\sigma^2)$ is a Gaussian distribution and the modulating law $S_{\frac{\alpha}{2}}(\sigma^2|\gamma\cos\left(\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}},1,0)$ is a fully asymmetric positive stable distribution [33, p.77].

It is possible to derive a compound representation for the envelope distribution by using the product decomposition of sub-Gaussian $S\alpha S$ stable vectors [33, p.77]

$$X_t + \jmath Y_t = (U_t + \jmath W_t)\sqrt{A_t}$$

where A_t is a fully asymmetric stable variable and $U_t + \jmath W_t$ is an isotropic complex Gaussian variable independent of A_t. As a result of isotropy the absolute value $\sqrt{U_t^2 + W_t^2}$ is distributed according to a Rayleigh law. Then, by analogy with (2.17) the envelope $R_t \triangleq \sqrt{X_t^2 + Y_t^2} = \sqrt{U_t^2 + W_t^2}\sqrt{A_t}$ admits the following compound representation

$$p_{\alpha\mathcal{R}}(r|\alpha,\gamma) = \int_0^\infty \mathcal{R}(x|\sigma) S_{\frac{\alpha}{2}}(\sigma^2|\gamma\cos\left(\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}},1,0)d\sigma \tag{2.18}$$

where $\mathcal{R}(x|\sigma)$ denotes the Rayleigh distribution.

According to this interpretation each resolution cell backscatters a random amount of power. The statistics of the backscattered power are given by
$S_{\frac{\alpha}{2}}(\sigma^2|\gamma\cos\left(\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}},1,0)$. These statistics would arise in cases where the number of scatterers within each resolution cell is random and follows a heavy-tailed distribution with tail exponent α.

We believe that the standard and the compound interpretations are complementary as they correspond to different types of tissues. Whether to use the one or the other depends on how we model scattering structures. Biological tissues that are composed of large popula-

tions of identifiable punctual microscopic structures (i.e., cell nuclei) are likely to be better described by the compound interpretation. In this case $S\alpha S$ statistics could indicate that those microscopic structures are clustered in a heavy-tailed organization. However, tissues composed of complex scattering structures for which there is no clear decomposition into point scatterers should be interpreted using the standard approach.

2.4.3 PLSN interpretation of α and γ

It should be noted that the above-stated interpretations assume that the α-th fractional moment of the interrogating pulse $\langle p_i^\alpha \rangle$ exists. This is true in particular for pulses with exponential and triangular decay, and more generally for pulses with finite power. If pulses have a power-law decay with exponent $\nu > \frac{1}{2}$ then the $S\alpha S$ parameters should be interpreted using the PLSN model [38]. Accordingly α is related to the pulse shape by $\alpha = \frac{1}{\nu}$ and γ is proportional to the mean scatterer density and to the α-th fractional moment of the scattering cross-sections [37].

2.5 Parameter estimation and density approximations for the $S\alpha S$ and $\alpha\mathcal{R}ayleigh$ laws

2.5.1 Estimation of α and γ

Application of the proposed distributions to ultrasound images requires reliable estimators for the unknown parameters α and γ. Estimation from envelope (B-mode) samples is particularly important since clinical ultrasound systems may not output RF signals. An interesting estimator for the parameters of an $\alpha\mathcal{R}ayleigh$ distribution has been proposed in [42], based on the 1st and 2nd order log-cumulants of the envelope

$$\mathrm{Var}(\log(\mathbf{r})) \approx \frac{\pi^2}{6\alpha^2} \tag{2.19}$$

$$\mathrm{E}(\log(\mathbf{r})) \approx \boldsymbol{C_e}\left(\frac{1}{\alpha} - 1\right) + \frac{\log(\gamma)}{\alpha} + \log(2) \tag{2.20}$$

where $\mathbf{r}$ denotes the envelope sample vector and $\boldsymbol{C_e} = 0.577$ is Euler s constant. We observe that expressions (2.19) and (2.20) depend exclusively on $\log(\mathbf{r})$ and can therefore be easily adapted for log-compressed ultrasound images. This estimator has also been independently derived in [43]. Note that a more accurate estimator based on fractional moments has been presented in [44], however it involves solving a highly non-linear equation.

In cases where the RF signal is available, the parameters α and γ can be estimated either by assuming that the data follows a $S\alpha S$ distribution, or the more general 4-parameter stable distribution. As regards the $S\alpha S$ case, a maximum likelihood estimator based on an optimization algorithm has been proposed in [45]. A fast estimator suitable for real-time applications has been proposed in [46]. For the general 4-parameter case, [47] proposes a maximum likelihood method. Again, a fast although less accurate estimator has been proposed in [48].

2.5.2 Approximation of the probability density function

A shortcoming of stable distributions is that they generally do not have a closed probability density function (recall that stable laws are defined through their characteristic function). However, there are several approximations that can be used for computation purposes.

In particular, $S\alpha S$ can be accurately computed by approximating the following integral [49]

$$p_{S\alpha S}(x|\alpha, \gamma = 1) = \frac{\alpha|x|^{\frac{1}{1-\alpha}}}{\pi|\alpha - 1|} \int_0^{\frac{\pi}{2}} V(\theta, \alpha) \exp\left[-|x|^{\frac{\alpha}{1-\alpha}} V(\theta, \alpha)\right] d\theta \tag{2.21}$$

for $x \neq 0, \alpha \neq 1$ and

$$p_{S\alpha S}(x|\alpha, \gamma = 1) = \frac{\Gamma(1 + \frac{1}{\alpha})}{\pi} \tag{2.22}$$

for $x = 0, \alpha \neq 1$ and where

$$V(\theta, \alpha) = \cos(\alpha)^{\frac{1}{\alpha-1}} \left(\frac{\cos(\theta)}{\sin(\alpha\theta)} \right)^{\frac{\alpha}{1-\alpha}} \frac{\cos(\alpha\theta - \theta)}{\cos(\theta)}.$$

Finally, for $\alpha = 1$ the $S\alpha S$ distribution is equivalent to the symmetric Cauchy distribution

$$p_{S\alpha S}(x|\alpha = 1, \gamma) = \frac{1}{\pi (x^2 + \gamma^2)}.$$

Expressions (2.21) and (2.22) correspond to the standard $S\alpha S$ distribution (i.e., $\gamma = 1$), non standard $S\alpha S$ s can be easily computed by renormalization. The reader is invited to refer to [49] for details about the implementation of these approximations.

Regarding the envelope, evaluating the $\alpha\mathcal{R}ayleigh$ distribution (2.9) requires the computation of the indefinite integral

$$\int_0^\infty \lambda \exp\left[-(\gamma_k \lambda)^{\alpha_k}\right] J_0(r\lambda)\, d\lambda. \tag{2.23}$$

Numerical schemes to approximate this integral are very time-consuming and may fail to converge because (2.23) oscillates an infinite number of times [50]. To circumvent these problems a numerically stable integral equivalent to expression (2.23) has been proposed in [50]. At last, an efficient alternative for $\alpha > 1$ is to approximate (2.9) using the following asymptotic series [39, 43].

$$p_{\alpha\mathcal{R}}(r|\alpha, \gamma) = \sum_{p=0}^{P} a_p r^{2p+1} + o\left(r^{2(P+1)+1}\right) \tag{2.24}$$

as $r \to 0$ and

$$p_{\alpha\mathcal{R}}(r|\alpha, \gamma) = \sum_{p=1}^{P} b_p r^{-\alpha p - 1} + o\left(r^{-\alpha(P+1)-1}\right) \tag{2.25}$$

as $r \to \infty$ where the coefficients a_p and b_p are

$$a_p = \frac{1}{\alpha\gamma}\frac{(-1)^p}{(p!)^2 2^{2p}}\Gamma\left(\frac{2p+2}{\alpha}\right)\gamma^{-2p-1}$$

$$b_p = \frac{(-1)^{p-1}\,2^{p\alpha+1}}{p!\,\pi\gamma}\Gamma^2\left(\frac{p\alpha+2}{2}\right)\sin\left(\frac{p\pi\alpha}{2}\right)\gamma^{p\alpha+1}$$

The choice of the order P and other considerations regarding the implementation of (2.24) and (2.25) have been addressed in [43].

2.6 Experimental Results

This section presents a variety of experiments conducted on synthetic and in vivo data to validate the analytically derived RF (2.5) and envelope (2.9) speckle distributions.

2.6.1 Synthetic Data

The analytical results presented in this study have been derived using limit theorems that supposed an infinite number of scatterers. This section presents simulations conducted to show that the proposed model provides a good approximation even for moderate amounts of scatterers. The simulation parameters were chosen to match those of dermatology ultrasound systems: a single-element focalized 25MHz wide-band (40-percent) probe sampled at 100MHz with a $53\mu m$ mechanic lateral step. The dimensions of the simulated image are $2.3mm \times 6.0mm$ (300 lines composed of 300 samples each) which is approximately equivalent to the region occupied by the dermis in a dermatological echography. Simulations were performed using FUSK [51], a state of the art ultrasound simulator based on the point scattering model. This simulator guarantees a space-invariant interrogating pulse, an assumption of our model that would not be respected by FIELD II [52, 53].

Figures 2.1 and 2.2 present the simulation results obtained for several populations of scatterers. Estimations of α and γ were obtained from the simulated RF signals using the

McCulloch method [2] [48]. Estimates have been evaluated by repeating simulations 250 times. The mean and standard deviation of each estimate are displayed using crosses and error bars respectively.

Figure 2.1 shows the estimates for α for different populations whose scattering cross-sections were simulated using a Pareto distribution with exponent α, as established in Section 2.4.1. The scatterer density for these simulations was of approximately 85 scatterers per resolution cell, which adds up to a total of 8 million scatterers per simulation. We observe that the estimates for α are in good agreement with the values predicted by the expressions in Section 2.4.1. We also observe that the estimation error increases progressively as α decreases; this behavior is consistent with the *McCulloch* estimator used in these experiments [48].

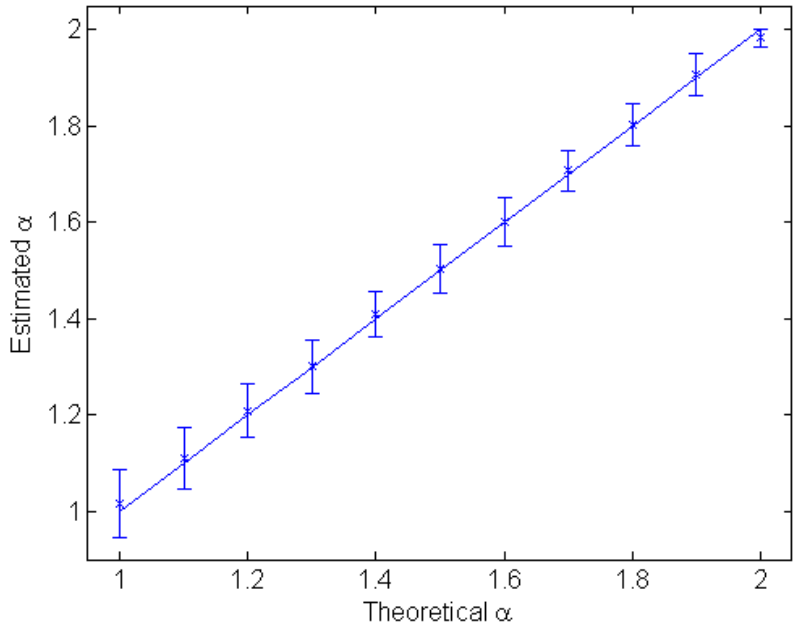
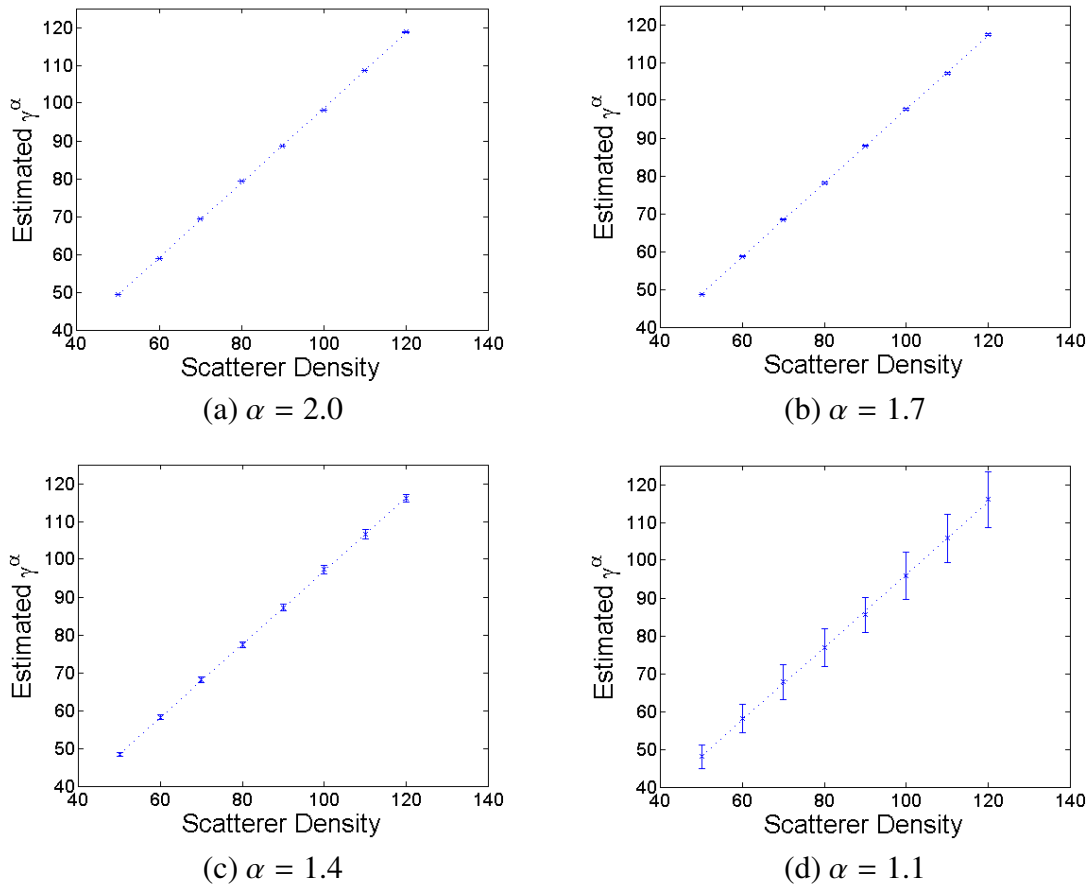


Figure 2.1: Estimated α vs its theoretical prediction (dotted line) 2.4.1. The mean (crosses) and standard deviation (error bars) of each estimate have been computed by repeating simulations 250 times.

Figure 2.2 shows the estimates for γ for several scatterer densities and for four values of α. We observe that these estimates are in good agreement with their theoretical predictions

[2] Code by Mark Veillet available at http://math.bu.edu/people/mveillet/research.html

and put into evidence the linear dependence between γ^α and M established in (2.15). Again, the estimation error increases progressively as alpha decreases. In figures 2.2(a)-(b) the standard deviation is smaller than 2-percent and cannot be assessed visually.



(a) $\alpha = 2.0$ (b) $\alpha = 1.7$

(c) $\alpha = 1.4$ (d) $\alpha = 1.1$

Figure 2.2: Estimated γ^α and its theoretical prediction (dotted line) (2.15) for different scatter densities. The mean (crosses) and standard deviation (error bars) of each estimate have been computed by repeating simulations 250 times.

2.6.2 Application to in vivo data

After validating the proposed model on synthetic data, this section presents experiments conducted on in vivo data. The corpus of ultrasound images used for these experiments has been acquired with an *ATYS dermocup* system, equipped with a single-element focalized 25MHz wide-band (40-percent) probe sampled at 100MHz with a 53μm mechanic lateral

step, subsequently annotated by experts (Fig. 2.3). In order to preserve signal statistics, no filtering, decimation or logarithmic compression were applied.

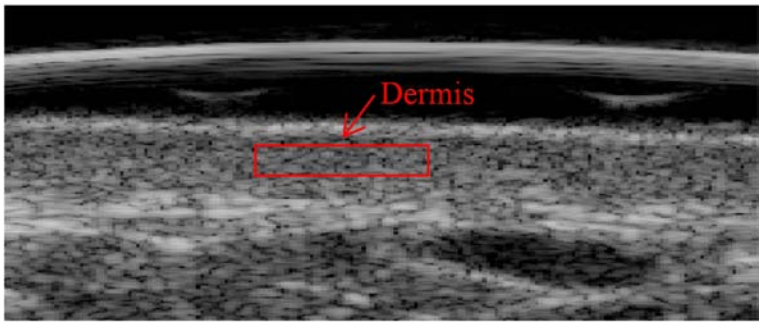


Figure 2.3: High frequency ultrasound image of in-vivo human dermis (forearm).

The first set of results illustrates the capacity of the $S\alpha S$ distribution to describe RF signals backscattered from human skin tissues. In particular, Fig.2.4 depicts a typical probability-density-function obtained from in-vivo forearm dermis, and the estimated $S\alpha S$ distribution. Additionally, to facilitate the analysis of this distribution s characteristic tails, Fig. 2.5 presents the corresponding pdfs in logarithmic scale. For comparison purposes a Gaussian fit has been included as well.

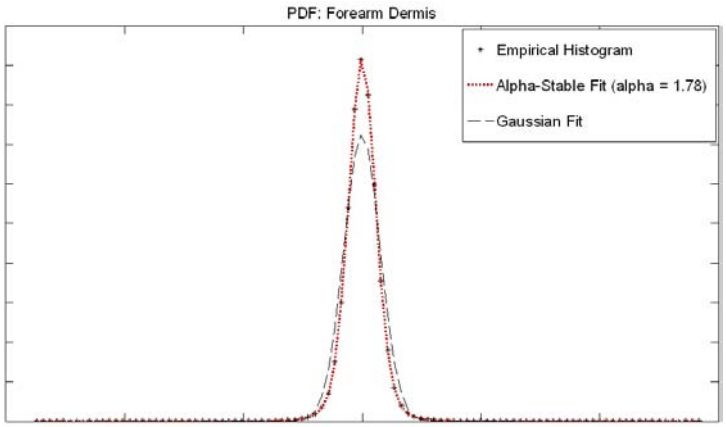


Figure 2.4: Comparison of the empirical pdf obtained from forearm dermis, and the corresponding estimations using the $S\alpha S$ and Gaussian distributions.

We observe (Fig. 2.4) that the proposed distribution accurately fits the empirical data. The tails of the empirical pdf match perfectly those of the estimated $S\alpha S$ distribution (Fig. 2.5).

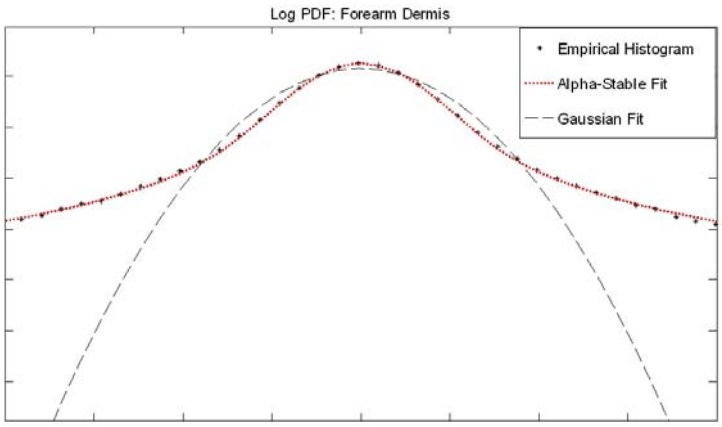


Figure 2.5: Comparison of the tails by means of a logarithmic plot of the pdfs.

Furthermore, the following set of results provides a first comparison between the proposed envelope distribution and previously studied envelope distributions. It has been reported in [25] that the envelope of signals backscattered from dermis tissues were best described by the Generalized Gamma distribution (3 parameters), followed by the Weibull and K distributions (2 parameters). However, the Nakagami distribution, usually encountered in tissues characterization, performed poorly.

Accordingly, we have compared the generalized (heavy-tailed) Rayleigh distribution against the Generalized Gamma and Weibull distributions. Heavy-tailed Rayleigh parameters have been estimated using log-moments [42], whereas the others were estimated using ML methods [25].

Figure 2.6 shows a typical pdf of envelope data obtained from the forearm dermis of a human subject in-vivo. Also, distribution fits are presented, together with their respective goodness-of-fit (KS test). Additionally, in order to better illustrate fitting at the tails, figure 2.7 displays the logarithmic pdfs. For the sake of completeness, the Nakagami distribution has been included as well.

These results (Fig. 2.6) indicate that the proposed envelope distribution provides the best fit, followed by the Generalized Gamma, Weibull and at last the Nakagami distribution. In addition, the tail of the proposed distribution is the only one that accurately de-

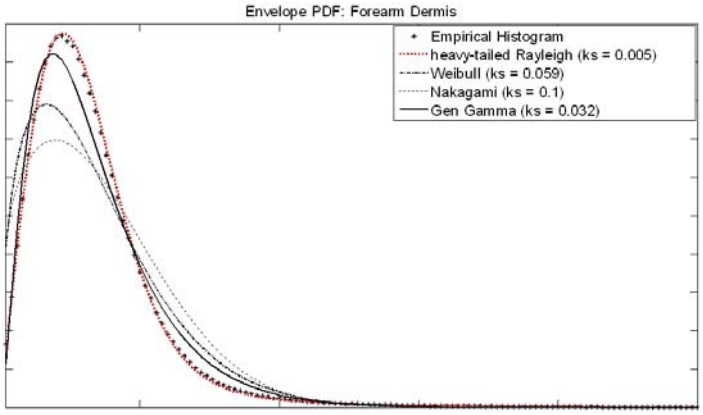


Figure 2.6: Comparison of the empirical envelope pdf obtained from forearm dermis, and the corresponding estimations using the heavy-tailed Rayleigh, Generalized Gamma, Weibull and Nakagami distribution.

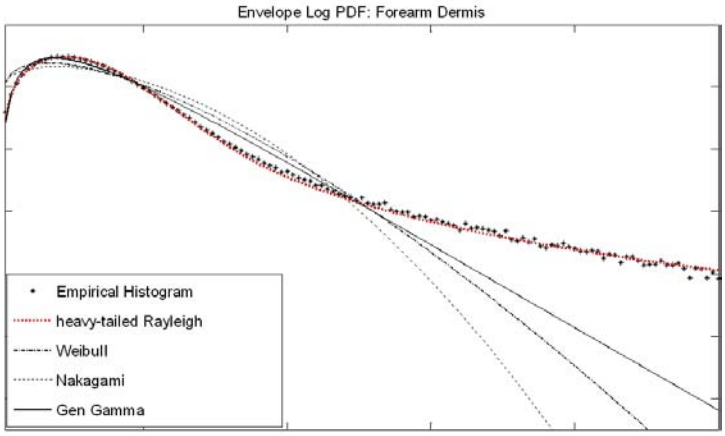


Figure 2.7: Comparison of distributions tails by means of a logarithmic plot of the pdfs.

scribes the outliers found in empirical data (Fig. 2.7), therefore supporting the heavy-tailed assumption.

Moreover, Figure 2.8 depicts an evaluation of the goodness-of-fit of the heavy-tailed Rayleigh, the Generalized Gamma and the Weibull distributions. For the sake of completeness, goodness-of-fit was measured using two complementary methods, the Kolmogorov-Smirnov and the Cramer von Mises tests. Results (Fig. 2.8) are based on 10 ROIs from forearm dermis.

We observe (Fig. 2.8) that the heavy-tailed Rayleigh distribution provided a better fit than the Generalized Gamma distribution in all cases, and for both goodness-of-fit tests. In ad-

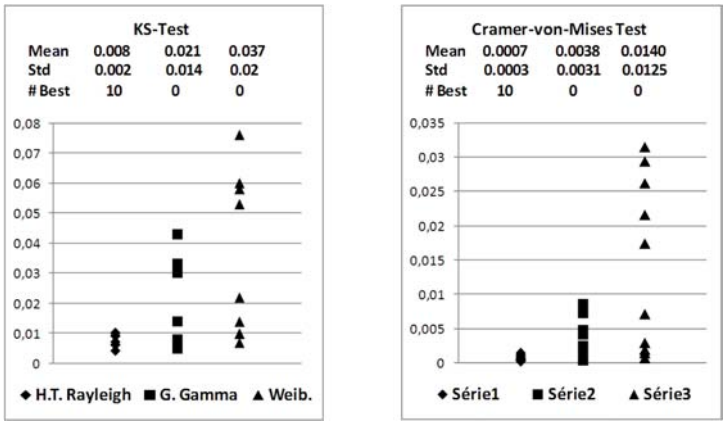


Figure 2.8: Goodness-of-fit measures for the heavy-tailed Rayleigh, Generalized Gamma and Weibull distributions. [Left]: Kolmogorov-Smirnov (KS) test. [Right]: Cramer-von-Mises test.

dition, we wish to stress the fact that the heavy-tailed Rayleigh distribution only has two parameters, while the Generalized Gamma has three.

At last, table 2.1 shows the values of α estimated on four 3D B-mode ultrasound images of forearm dermis using expressions (2.19) and (2.20). Each image corresponds to a different subject. For completeness we illustrate the relationship between the α and the statistical parameters commonly found in the literature: SNR [12, 22], m (Nakagami) [20], ED (K-distribution effective density) [54], and SF (speckle factor) [55]. We observe that parameters are correlated and indicate that the dermis has non-Gaussian speckle statistics.

Table 2.1: Comparison between α and the literature s statistical parameters

α	SNR	m	ED	SF
1.97	1.85	0.96	12.42	1.15
1.95	1.84	0.95	14.35	1.13
1.90	1.72	0.88	3.73	1.53
1.79	1.25	0.64	0.14	14.8

Interpretation of these parameters using classic theories would suggest that the dermis has either a low scatterer density or possibly a high scatterer density coupled with high scattering cross-section variance. However, the information provided by α (supported by excellent goodness-of-fit tests) states that the dermis has heavy-tailed statistics. As a result

the cross-section variance is not defined, nor can the central limit theorem be applied. Stable statistics in dermis arises as a limit case, where both the number of scatterers and the variance of their cross-sections tend to infinity. The other statistical parameters should not be applied on dermis because their estimation involves computing 2nd and 4th order moments which are not defined for heavy-tailed distributions. Their correlation with the tail exponent α can be explained by the fact that α determines their rate of divergence.

2.7 Conclusion

Starting from the widely accepted point scattering model, mathematical developments have shown that speckle in skin tissues follows a Levy Flight and has α-stable statistics. As a result, the envelope signal has a generalized (heavy-tailed) distribution. Also, it has been established that stable statistics imply that scatterers have heavy-tailed cross-sections. Consequently stable distributions model a new case of non-Gaussian statistics, where both the number of scatterers and the variance of their cross-sections tend to infinity. This configuration is believed to arise in tissues with complex scattering structures as dermis. In addition, analytical expressions have been provided to relate the α-stable parameters to scatterer properties. Simulations and experimental results supported by excellent goodness-of-fit tests confirm the proposed analytical results, which provide new insight into non-Gaussian statistics. These fundamental results set the basis for new echography processing methods and quantitative ultrasound characterization tools.

Un modèle de mélange de lois Rayleigh généralisées spatialement corrélées pour la segmentation d'échographies cutanées

L imagerie par ultrasons est une modalité largement répondue avec des applications en diagnostic, examens préventifs, thérapie et chirurgie guidée par l image, entre autres. En oncologie dermatologique, le pronostic du mélanome repose souvent sur l évaluation d indicateurs de surface comme la couleur, la forme et la texture alors que les deux mesures les plus fiables sont la profondeur de la lésion et le nombre de couches de la peau envahies. Actuellement, ces mesures ne peuvent être faites qu après excision. Les progrès récents dans les technologies des transducteurs d ultrasons de hautes fréquences et des sondes tridimensionnelles apportent l opportunité d examiner la peau d une manière non-invasive et de détourer les tumeurs avant excision sur des images échographiques. Cependant, le changement des pratiques dermatologiques requiert le développement d algorithmes robustes de segmentation. Malgré la littérature extensive sur le sujet, la segmentation avec précision d images d ultrasons est toujours une tâche difficile qui entraîne un effort de recherche considérable. Les techniques actuelles sont extrêmement spécifiques aux applications, développées principalement en échocardiographie, examen échographique transrectale de la prostate, maladies intra-vasculaires, rein, et cancer du sein [56]. La segmentation dans les trois premières applications concerne la détection et le suivi des frontières des organes. La délimitation de lésions est un problème significativement difficile du à leurs frontières oues et mal définies. Le fait qu elles soient visuellement différentes des tissus sains sur les images a motivé la création de méthodes de segmentation basées régions [57–59]. Par ailleurs les lésions n ayant pas de formes prédéfinies, elles ne peuvent pas bénéficier de travaux récents sur la modélisation anatomique et probabiliste [60–62]. Les travaux récents en bioinformatique sur la modélisation du développement de tumeurs [63] peuvent potentiellement amener des solutions.

Les premières méthodes de segmentation étaient principalement basées sur le seuillage [64, 65] et ont été remplacées par des techniques utilisant la texture. Madabhushi et al. ont dérivé un contour actif combinant contours et texture [66]. Huang et al. ont proposé un algorithme de segmentation par réseau de neurones utilisant la texture [67]. Des modèles de mélange de Gaussiennes couplés à des champs aléatoires de Markov ont été proposés pour segmenter des lésions à l aide des statistiques des régions [68, 69]. Par ailleurs, depuis le travail original de Dias et al. [70], les mélanges Rayleigh sont devenus un outil puissant pour la segmentation basée région d images d ultrasons. L utilisation de distributions Rayleigh au lieu de Gaussiennes est justifiée par le processus de formation de l image en mode B [21]. On se basant sur l hypothèse que chaque tissu possède ses propres statistiques Rayleigh, la segmentation est obtenue par la séparation des composantes du mélange. Ceci est fait en recherchant le maximum de vraisemblance (ML) ou le maximum a posteriori (MAP) du contour. Dans [70], la méthode d optimisation par programmation dynamique interactive a été utilisée pour estimer le contour MAP et les paramètres du mélange. Des expériences sur des images d échocardiographie pour segmenter avec précision les cavités du coeur ont été reportées.

Avec le développement des modèles déformables, Brusseau et al. ont proposé un contour actif paramétrique [71]. Il s agit d une courbe définie par un ensemble de points dans l image déplacés itérativement pour maximiser la distribution a posteriori de la segmentation [72]. Dans le travail de Brusseau et al., les deux composantes du mélange de Rayleigh sont séparées à l aide d un contour actif dirigé par les statistiques des régions, permettant aussi d estimer le paramètre de chaque distribution. Etant donné que la convergence vers un maximum global n est pas garantie, les auteurs proposent une technique d initialisation ad-hoc. Cette méthode a été améliorée par Cardinal et al. [73] qui lui ont substitué un contour actif basé contour dérivé du travail original d Osher et Sethian [74]. Une autre modification a été l introduction d un algorithme espérance-maximisation pour estimer les paramètres du mélange durant l initialisation, éliminant ainsi le besoin de les estimer itéra-

tivement. Les auteurs ont reporté que leur méthode level-set permettant de résoudre le mélange de Rayleigh donne de meilleurs résultats que l approche classique utilisant le gradient. En outre, Saroul et al. ont appliqué récemment le modèle de mélange de Rayleigh à la segmentation de la prostate dans des images d ultrasons transrectales [75]. Dans ce cas, le level-set a été remplacé par un modèle déformable fondé sur une super ellipse dont l évolution est calculée à l aide d une méthode numérique. Les auteurs ont montré que la régularisation introduite par ce modèle déformable permet de compenser des occlusions partielles.

Les modèles de mélange de Rayleigh ont été étendus à des tissus ayant des statistiques Rayleigh généralisée par Destrempes et al. [76], qui ont proposé une méthode de segmentation de l artère carotide basée sur un mélange de Nakagami et un modèle déformable. Comme dans [73], les paramètres du mélange ont été estimés à l aide d un algorithme EM sous l hypothèse que les observations sont indépendantes. L évolution du modèle déformable a été faite à l aide d un algorithme d optimisation stochastique qui converge vers un optimum global. Cependant, comme les paramètres du mélange sont estimés avec un algorithme EM, la convergence globale n est pas garantie. Une autre contribution importante est la méthode level-set basée région présentée dans [1], qui a adapté la méthode de Chan et Vese [77] aux images d ultrasons avec des statistiques Rayleigh. Cette méthode est capable de segmenter des objets avec des contours lisses dans des conditions de faible rapport signal sur bruit. Ce travail a été généralisé récemment à toutes les distributions de la famille exponentielle [78]. Ces méthodes n ont pas été appliquées pour segmenter des lésions.

Ce chapitre traite le problème d estimer conjointement la distribution statistique et segmenter les lésions dans des image d ultrasons 2D et 3D de haute fréquence de multiples tissus de la peau. A notre connaissance, ce travail est le premier à traiter spécifiquement la segmentation de lésions cutanées dans des images d ultrasons. Nous proposons la mod-

élisation des images de multiples tissus en utilisant un mélange de lois Rayleigh à queue lourde, inspiré par le modèle de l image d un tissu unique étudié en [5].

Le modèle de mélange proposé est complété par un champ aléatoire de Markov pour prendre en considération la corrélation spatiale inhérente aux tissus biologiques. Les champs de Potts-Markov sont particulièrement adaptés pour la segmentation par étiquetage [79–82]. Ces modèles améliorent la segmentation par leur capacité à capturer la corrélation spatiale existant entre les étiquettes des classes voisines [79]. Cette corrélation est engendrée naturellement par l organisation spatiale des tissus biologiques particulièrement importante dans la peau à cause de sa structure en couches. Finalement, d autres modèles plus complexes que le Potts auraient pu être adoptés pour introduire la cohérence spatiale entre les composantes. En particulier, Marroquin et al. [83] ont montré que de meilleurs résultats de segmentation peuvent être obtenus en utilisant un champ caché à deux couches. Ils préconisent de supposer que les étiquettes cachées sont indépendantes et d introduire la corrélation dans la deuxième couche par un champ Markovien vectoriel. De la même façon, Woolrich et al. ont proposé d approximer le champ de Potts en modélisant les poids du mélange par un champ aléatoire de Markov Gaussien. Cependant, ces modèles alternatifs ne sont pas adaptés aux images 3D à cause de leurs coûts de calcul prohibitifs. Ces coûts résultent du fait qu ils introduisent (K+1)N et KN variables cachées respectivement, contre N pour le modèle de Potts (N étant le nombre de voxels et K le nombre de classes). Le problème de segmentation est résolu par un algorithme d optimisation stochastique garantissant une convergence globale, éliminant le besoin d une initialisation ou une méthode supervisée. Le chapitre est organisé comme suit : le modèle statistique du voxel d une image d ultrasons est introduit dans la section II. La section III élabore le modèle Bayésien pour la segmentation des images d ultrasons. Un échantillonneur de Gibbs hybride générant des données asymptotiquement distribuées selon la distribution a posteriori du modèle Bayésien est décrit dans la section IV. Des expériences sur des

données synthétiques et réelles sont présentées dans le section V. Des conclusions sont finalement reportées en section VI.

Chapter 3

Segmentation of skin lesions in 2D and 3D ultrasound images using a spatially coherent generalized Rayleigh mixture model

This chapter addresses the problem of jointly estimating the statistical distribution and segmenting lesions in multiple-tissue high-frequency skin ultrasound images. The distribution of multiple-tissue images is modeled as a spatially coherent finite mixture of heavy-tailed Rayleigh distributions. Spatial coherence inherent to biological tissues is modeled by enforcing local dependence between the mixture components. An original Bayesian algorithm combined with a Markov chain Monte Carlo method is then proposed to jointly estimate the mixture parameters and a label-vector associating each voxel to a tissue. More precisely, a hybrid Metropolis-within-Gibbs sampler is used to draw samples that are asymptotically distributed according to the posterior distribution of the Bayesian model. The Bayesian estimators of the model parameters are then computed from the generated samples. Simulation results are conducted on synthetic data to illustrate the performance of the

proposed estimation strategy. The method is then successfully applied to the segmentation of in-vivo skin tumors in high frequency 2D and 3D ultrasound images.

3.1 Introduction

Ultrasound imaging is a longstanding medical imaging modality with important applications in diagnosis, preventive examinations, therapy and image-guided surgery. In dermatologic oncology, diagnosis relies mainly on surface indicators such as color, shape and texture whereas the two more reliable measures are the depth of the lesion and the number of skin layers that have been invaded. Currently, these can only be evaluated after excision. Recent advances in high frequency transducers and 3D probes have opened new opportunities to perform non-invasive diagnostics using ultrasound images. However, changing dermatological practices requires developing robust segmentation algorithms. Despite the extensive literature on the subject, accurate segmentation of ultrasound images is still a challenging task and a focus of considerable research efforts. Current segmentation techniques are extremely application-specific, developed mainly for echocardiography followed by transrectal prostate examination (TRUS), kidney, breast cancer and (intra) vascular diseases (IVUS) [56]. A survey of the state-of-the-art methods up to 2006 is presented in [56].

Segmentation in echocardiography, TRUS and IVUS is mainly concerned with the detection and tracking of organ boundaries. Lesion delimitation is significantly different and more challenging. On one hand, unlike organs, lesions exhibit soft or “fuzzy" edges that are difficult to capture with boundary detection techniques. On the other, their echogenic and statistical characteristics are visibly different from those of their surrounding tissues. This fact has motivated the development of region-based segmentation techniques as opposed to boundary-based methods, which are still an active research subject in other medical ultrasound domains [57–59]. Similarly, lesions do not have anatomically predefined

shapes as is the case for organs and are unlikely to benefit in the near future from recent works on anatomical or learned statistical shape priors [60–62]. This might change with the improvement of geometric tumor growth models derived from computational biology [63]. Early lesion segmentation methods have focused mainly on thresholding [64, 65] and were superseded by texture-based techniques. Madabhushi *et al.* derived an active contour based on texture and boundary features [66]. Huang *et al.* proposed a texture segmentation technique based on a neural network and a watershed algorithm [67]. In addition, Gaussian mixture models coupled with Markov random fields were proposed to segment lesions based on their region statistics [68, 69]. Moreover, since the seminal work of Dias *et al.* [70], Rayleigh mixtures have become a powerful model for region-based ultrasound image segmentation. The use of Rayleigh instead of Gaussian distributions is strongly justified by the physics of the image formation process that generates B-mode ultrasound images [21]. Based on the assumption that each biological tissue has its proper Rayleigh statistics, tissue segmentation is achieved by separating the mixture components. This is achieved by finding the maximum-likelihood (ML) or maximum-a-posteriori (MAP) estimators of the lesion contours. The optimization problem stemming from the ML and MAP estimators was solved in [70] using an interactive dynamic programming (IDP) algorithm that jointly estimated the MAP contour and the mixture parameters. The authors performed several experiments on real echocardiography images and showed that the proposed method accurately segments heart walls.

With the development of deformable models, Brusseau *et al.* proposed a statistical parametric active contour (AC) [71]. A parametric AC is a regularized curve defined by a set of points in the image domain that can be moved to maximize the segmentation posterior [72]. In the work of Brusseau *et al.*, the 2-mixture components were separated using a statistical region AC which iteratively estimated the Rayleigh parameter of each component and evolved to optimize the segmentation. Also, given that convergence to a global optimum is not guaranteed, the authors proposed an *ad-hoc* automatic initialization technique. This

method was further improved by Cardinal *et al.* [73] who substituted the parametric AC by an edge-based level set (LS) derived from the original work of Osher and Sethian [74]. A second modification was the introduction of an expectation-maximization (EM) algorithm to estimate the mixture parameters during initialization, thus removing the need to estimate them iteratively. The authors reported that the Rayleigh mixture LS method outperforms classic gradient-based level set at intravascular image segmentation. In addition, Saroul *et al.* recently applied the Rayleigh mixture model to prostate segmentation in transrectal ultrasound images [75]. In this case, the LS was replaced by a deformable model based on a super ellipse whose evolution was computed using a variational algorithm. The authors showed that the regularization introduced by this deformable model could compensate partial occlusion.

Rayleigh-mixture models were extended to tissues with generalized Rayleigh statistics by Destrempes *et al.* [76], who proposed a carotid artery segmentation method based on a Nakagami mixture and a deformable model. As in [73], the estimation of the mixture parameters was achieved using an EM algorithm under the assumption that observations are independent. The evolution of the deformable model was computed using exploration/selection, a stochastic optimization algorithm that converges to the global optimum. However, since the mixture parameters are estimated with an EM algorithm, overall global convergence is not guaranteed. One other important contribution is the Rayleigh region-based LS method presented in [1], that adapted the fundamental work of Chan and Vese [77] on ACs without edges to ultrasound images with Rayleigh statistics. These region-based LS should be very appropriate for ultrasound images of lesions as they are able to segment objects with smooth edges under poor signal-to-noise ratio conditions. This work was recently generalized to all the distributions from the exponential family (i.e., Gamma, Rayleigh, Poisson, etc.) in [78]. However, these methods have not yet been applied to lesion segmentation in ultrasound images.

This chapter addresses the problem of jointly estimating the statistical distribution and segmenting lesions in multiple-tissue 2D and 3D high-frequency skin ultrasound images. To our knowledge this is the first ultrasound image segmentation method specific to skin lesions. We propose to model multiple-tissue images using a heavy-tailed Rayleigh mixture, a model that has been inspired by the single-tissue model studied in [5]. The proposed mixture model is equipped with a Markov random field (MRF) that takes into account the spatial correlation inherent to biological tissues. Note that Potts Markov fields are particularly well suited for label-based segmentation as explained in [84] and further studied in [79–82]. Potts Markov models enhance segmentation because of their ability to capture the spatial correlation that exists between neighbor class labels [79]. This correlation arises naturally from the spatial organization of biological tissues and is particularly important in skin because of its layered structure. Finally, while the Potts prior is an effective means to introduce spatial correlation between the class labels, it is interesting to mention that other more complex models could have been used instead. In particular, Marroquin *et al.* [83] have shown that better segmentation results may be obtained by using a two-layer hidden field, where hidden labels are assumed to be independent and correlation is introduced at a deeper layer by a vectorial Markov field. Similarly, Woolrich *et al.* [85] have proposed to approximate the Potts field by modeling mixture weights with a Gauss-Markov random field. However, these alternative models are not well adapted for 3D images because they require significantly more computation and memory resources than the Potts model. These overheads result from the fact that they introduce $(K+1)N$ and KN hidden variables respectively, against only N for the Potts model (N being the number of voxels and K the number of classes). In addition, the segmentation problem is solved using a stochastic optimization algorithm with guaranteed global convergence, removing the need for an initial contour or supervised training. The chapter is organized as follows: The statistical model used for a voxel of an ultrasound image is introduced in Section II. Section III introduces the Bayesian model used for the segmentation of ultrasound images. An hybrid Gibbs sam-

pler generating samples asymptotically distributed according to the posterior distribution of this Bayesian model is described in Section IV. Experiments on synthetic and real data are presented in Section V. Conclusions are finally reported in Section VI.

3.2 Problem Statement

This section describes the mixture model used for ultrasound image voxels[1]. Let $r_n \in \mathbb{R}^+$ denote an observation, or voxel, in an envelope (B-mode) ultrasound image $\boldsymbol{r} = (r_1, \ldots, r_N)^T$ without logarithmic compression. We assume that r_n is defined by means of the widely accepted point scattering model [17]

$$r_n = \left| \sum_{i=1}^{M} a_i \left[p(t_n - \tau_i) + \jmath \tilde{p}(t_n - \tau_i) \right] \right| \tag{3.1}$$

where M is the total number of punctual scatterers, $p(t) + \jmath\tilde{p}(t)$ denotes the analytic extension of the interrogating pulse $p(t)$, $a_i \in (0, 1)$ is the cross-section of the ith scatterer, $\tau_i \in \mathbb{R}^+$ is the time of arrival of the ith backscattered wave and t_n is the sampling time associated with r_n. Recent works on scattering in biological tissues have established that r_n, as defined above, converges in distribution towards an α-Rayleigh distribution as M increases [5]

$$r_n \underset{M\to\infty}{\overset{d}{\to}} \alpha\mathcal{R}(\alpha_n, \gamma_n) \tag{3.2}$$

where $\underset{M\to\infty}{\overset{d}{\to}}$ denotes convergence in distribution, the parameters $\alpha_n \in (0, 2]$ and $\gamma_n \in \mathbb{R}^+$ are the characteristic index and spread associated with the nth voxel.

This chapter considers the case where the ultrasound image $\boldsymbol{r}$ is made up by multiple biological tissues with high scatter density (i.e., $M \to \infty$), each with its own echogenicity and therefore its proper speckle statistics. In view of this spatial configuration, we propose to model $\boldsymbol{r}$ by an α-Rayleigh stationary process with piecewise constant parameters. More

[1] Part of this work has been presented at IEEE ICASSP 11, Prague, Czech Republic, May 2011 [3], and at EUSIPCO 11, Barcelona, Spain, Sep. 2011 [9].

precisely, we assume that there is a set of stationary classes $\{C_1, \ldots, C_K\}$ such that

$$\forall r_n \in C_k, \quad r_n \sim \alpha\mathcal{R}(\alpha_k, \gamma_k) \tag{3.3}$$

where α_k and γ_k are the parameters associated with the class C_k (i.e., the kth biological tissue). As a consequence, it is possible to express the distribution of r_n by means of the following mixture of α-Rayleigh distributions

$$r_n \sim \sum_{k=1}^{K} \omega_k \alpha\mathcal{R}(\alpha_k, \gamma_k) \tag{3.4}$$

where K is the number of classes and ω_k represents the relative weight (or proportion) of the kth class with $\sum_k \omega_k = 1$. Lastly, to take into account the spatial coherence inherent to biological tissues we will consider that the class of a given voxel depends on those of its neighbors.

It should be noted that the proposed α-Rayleigh mixture model is closely related to two other mixture models. On the one hand it generalizes the Rayleigh mixture model, which has been extensively applied to ultrasound image modeling. On the other, it can be shown that before being transformed by acquisition and demodulation, radio frequency ultrasound signals are distributed according to a symmetric α-stable distribution [5]. Hence, the proposed α-Rayleigh mixture model can be interpreted as a transformation of the symmetric α-stable mixture model studied in [86]. In addition, it is interesting to mention that the α-Rayleigh distribution has been used successfully for SAR images in [42, 44]. The methods proposed in [42, 44] have been recently applied to characterize tissues in annotated ultrasound images [5]. This chapter extends those methods by including in the estimation problem the identification of regions in the image with similar α-Rayleigh parameters (each region being associated with a different tissue). This is achieved by proposing a novel Bayesian estimation algorithm based on the α-Rayleigh mixture model (3.4) coupled with a Markov random field prior that captures the spatial coherence inherent to biological

tissues. Finally, akin to [1, 71, 73, 76], note that the model (3.4) uses a simplified image representation based on regions and does not describe the boundaries between tissues explicitly.

The following section addresses the problem of estimating the parameters of the spatially coherent α-Rayleigh mixture model introduced in (3.4) and performing the segmentation of ultrasound images.

3.3 Bayesian Model

A label vector $\boldsymbol{z} = (z_1, \ldots, z_N)^T$ is introduced to map observations $\boldsymbol{r}$ to classes $C_1, \ldots, C_K$ (i.e., $z_n = k$ if and only if $r_n \in C_k$). This label vector will allow each image observation to be characterized and different kinds of tissues to be discriminated. Note that the weights ω_k are directly related to the labels through the probabilities $\mathrm{P}[z_n = k] = w_k$ for $k = 1, \ldots, K$. Consequently, the unknown parameter vector for the mixture (3.4) can be defined as $(\boldsymbol{\theta}, \boldsymbol{z})$ where $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^T$. This section studies a Bayesian model associated with $(\boldsymbol{\theta}, \boldsymbol{z})$. This model requires defining the likelihood and the priors for the unknown parameters.

3.3.1 Likelihood

Assuming that the observations r_n are independent and using the mixture model (3.4), the likelihood of the proposed Bayesian model can be written as

$$p(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}) = \prod_{k=1}^{K} \prod_{\{n|z_n=k\}} p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) \tag{3.5}$$

where $\{n|z_n = k\}$ denotes the subset of indexes $n = 1, \ldots, N$ that verify $z_n = k$,

$$p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) = r_n \int_0^\infty \lambda \exp\left[-(\gamma_k \lambda)^{\alpha_k}\right] J_0(r_n \lambda)\, d\lambda \tag{3.6}$$

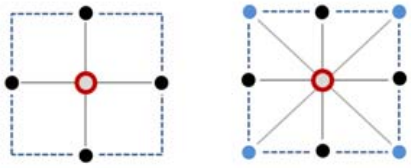


Figure 3.1: 4-pixel (left) and 8-pixel (right) neighborhood structures. The pixel considered appears as a void red circle whereas its neighbors are depicted in full black and blue.

is the probability density function (pdf) of an α-Rayleigh distribution with parameters α_k and γ_k and J_0 is the zeroth order Bessel function of the first kind.

3.3.2 Parameter priors

Labels

It is natural to consider that there is some correlation between the probabilities $\mathrm{P}[z_n = k]$ of a given voxel and those of its neighbors. Since the seminal work of Geman [87], MRFs have become very popular to model neighbor correlation in images. MRFs assume that the distribution of a pixel conditionally to all other pixels of the image equals the distribution of this pixel conditionally to its neighbors. Consequently, it is important to properly define the neighborhood structure. The neighborhood relation between two pixels (or voxels) i and j has to be symmetric: if i is a neighbor of j then j is also a neighbor of i. There are several neighborhood structures that have been used in the literature. In the bidimensional case, neighborhoods defined by the four or eight nearest voxels represented in Fig. 3.1 are the most commonly used. Similarly, in the tridimensional case the most frequently used neighborhoods are defined by the six or fourteen nearest voxels represented in Fig 3.2. In the rest of this chapter 4-pixel neighborhoods will be considered for 2D images and 6-voxel neighborhoods for 3D images. Therefore, the associated set of neighbors, or cliques, can only have vertical, horizontal and depth configurations (see [87, 88] for more details).

Once the neighborhood structure has been established, the MRF can be defined. Let z_n denote the random variable indicating the class of the nth image voxel. In the case of

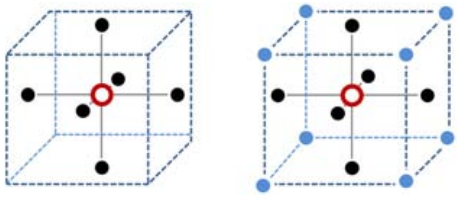


Figure 3.2: 6-voxel (left) and 14-voxel (right) neighborhood structures. The voxel considered appears as a void red circle whereas its neighbors are depicted in full black and blue.

K classes, the random variables $z_1, \ldots, z_N$ take their values in the finite set $\{1, \ldots, K\}$. The whole set of random variables $\boldsymbol{z}$ forms a random field. An MRF is then defined when the conditional distribution of z_n given the other pixels $\boldsymbol{z}_{-n} = (z_1, \ldots, z_{n-1}, z_{n+1}, \ldots, z_N)$ only depends on its neighbors $\boldsymbol{z}_{\mathcal{V}(n)}$, i.e.,

$$\mathrm{P}[z_n|\boldsymbol{z}_{-n}] = \mathrm{P}[z_n|\boldsymbol{z}_{\mathcal{V}(n)}] \tag{3.7}$$

where $\mathcal{V}(n)$ contains the neighbors of z_n according to the neighborhood structure considered.

In this study we will first consider 2D and 3D Potts Markov fields as prior distributions for $\boldsymbol{z}$. More precisely, 2D MRFs are considered for single-slice (2D) ultrasound images whereas 3D MRFs are used for multiple-slice (3D) images. In light of the Hammersley-Clifford theorem, the corresponding prior for $\boldsymbol{z}$ can be expressed as follows:

$$p(\boldsymbol{z}) = \frac{1}{C(\beta)} \exp\left[\sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \beta\delta(z_n - z_{n'})\right] \tag{3.8}$$

where β is the granularity coefficient, $C(\beta)$ is the normalizing constant or partition function [89] and $\delta(\cdot)$ is the Kronecker function. The hyperparameter β tunes the degree of homogeneity of each region in the image. A small value of β induces a noisy image with a large number of regions, contrary to a large value of β that leads to few and large homogeneous regions. In this work, the granularity coefficient β will be fixed a priori. However, it is interesting to mention that the estimation of β has been receiving a lot of attention in the

literature [82, 90–93]. Estimating the granularity coefficient using one of these methods is clearly an interesting problem that will be investigated in future work. Finally, it is interesting to note that despite not knowing $C(\beta)$, drawing labels $\boldsymbol{z} = (z_1, \ldots, z_N)$ from the distribution (3.8) can be easily achieved by using a Gibbs sampler [94].

α-Rayleigh parameters

The prior for each characteristic index α_k ($k = 1, \ldots, K$) is a uniform distribution on $(0, 2]$

$$\alpha_k \sim \mathcal{U}(0, 2). \tag{3.9}$$

This choice is motivated by the fact that the only information available a priori about this parameter, is that it can take values in the interval $(0, 2]$.

The prior for each spread parameter γ_k is an inverse gamma distribution with hyperparameters a_0 and b_0

$$\gamma_k \sim \mathcal{IG}(a_0, b_0), \quad k = 1, \ldots, K. \tag{3.10}$$

This choice is motivated by the fact that the inverse gamma distribution allows either very vague or more specific prior information to be incorporated depending on the choice of the hyperparameters a_0 and b_0 ($a_0 = b_0 = 1$ will be used in our experiments corresponding to a vague prior distribution).

Assuming a priori independence between the parameters α_k and γ_k, the prior for $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\alpha})p(\boldsymbol{\gamma}) = \prod_{k=1}^{K} p(\alpha_k)p(\gamma_k). \tag{3.11}$$

We will also assume that the α-Rayleigh parameters are independent from the labels associated with the image voxels. Thus the joint prior for the unknown parameters $(\boldsymbol{\theta}, \boldsymbol{z})$ can be expressed as

$$p(\boldsymbol{\theta}, \boldsymbol{z}) = p(\boldsymbol{z})\, p(\boldsymbol{\theta}) \tag{3.12}$$

where $p(z)$ has been defined in (3.8) and $p(\boldsymbol{\theta})$ in (3.11).

Figure 3.3 presents the proposed Bayesian model as a directed acyclic graph (DAG) summarizing the relationships between the different parameters and hyperparameters.

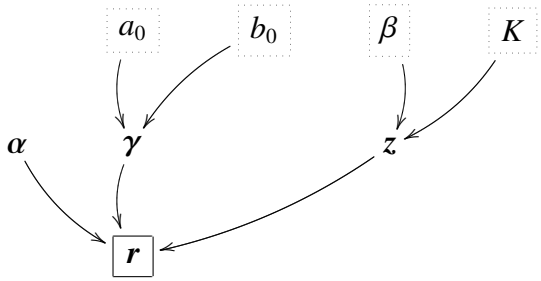


Figure 3.3: Directed acyclic graph (DAG) for the α-Rayleigh mixture model (the fixed nonrandom hyperparameters appear in dashed boxes).

3.3.3 Posterior Distribution of $(\boldsymbol{\theta}, \boldsymbol{z})$

Using Bayes theorem, the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{z}) = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z})$ can be expressed as follows

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{r}) &= \frac{p(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}) p(\boldsymbol{\theta}, \boldsymbol{z})}{p(\boldsymbol{r})} \\ &\propto p(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}) p(\boldsymbol{\theta}, \boldsymbol{z}) \end{aligned} \tag{3.13}$$

where $\propto$ means "proportional to" and the likelihood $p(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z})$ and the joint prior $p(\boldsymbol{\theta}, \boldsymbol{z})$ have been defined in (3.5) and (3.12).

Unfortunately the posterior distribution (3.13) is too complex to derive closed form expressions for the minimum mean square error (MMSE) or MAP estimators of the unknown parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{z}$ [2]. One can think of using the EM algorithm [95] that has received much attention for mixture problems (see [73, 76] for applications to ultrasound images). However, EM algorithms have many known shortcomings. For instance, they suffer from *convergence to local maxima or saddle points of the log-likelihood function and sensitivity*

[2] note that $p(\boldsymbol{\theta}, \boldsymbol{z})$ involves the potential of a Potts Markov field and its intractable partition function $C(\beta)$ and that $p(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z})$ is the product of N indefinite integrals

to starting values [96, p. 259]. Note that analyzing the concavity properties of the logarithm of (3.5) is not easy because the α-Rayleigh distribution does not belong to the exponential family. An interesting alternative is to use a Markov Chain Monte Carlo (MCMC) method generating samples that are asymptotically distributed according to the target distribution (3.13) [94]. The generated samples are then used to approximate the Bayesian estimators. This strategy has been used successfully in many image processing applications [97–101]. One sampling technique allowing the parameters of ultrasound images to be estimated is studied in the next section.

3.4 Hybrid Gibbs Sampler

This section studies a hybrid Metropolis-within-Gibbs sampler for generating samples that are asymptotically distributed according to (3.13). The histogram of the generated samples is guaranteed to converge to the posterior (3.13) [94, p. 269]. One of the most popular methods for generating samples distributed according to a distribution whose pdf or probability masses are known up to a multiplicative constant is the Gibbs sampler. The conventional Gibbs sampler draws samples according to the conditional distributions associated with the distribution of interest (here the posterior (3.13)). When a conditional distribution cannot be sampled easily, one can resort to a Metropolis-Hastings (MH) move, which generates samples according to an appropriate proposal and accept or reject these generated samples with a given probability. The resulting sampler is referred to as Metropolis-within-Gibbs sampler (see [94] for more details about MCMC methods). The sampler investigated in this section is based on the conditional distributions $\mathrm{P}[z|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{r}]$, $p(\boldsymbol{\alpha}|z, \boldsymbol{\gamma}, \boldsymbol{r})$ and $p(\boldsymbol{\gamma}|z, \boldsymbol{\alpha}, \boldsymbol{r})$ that are described in the next paragraphs (see also Algorithm 1 below).

Algorithm 1 Proposed Hybrid Gibbs Sampler

Initialization:

– Sample α_k^0 ($k = \{1, \ldots, K\}$) from the pdf in (3.9).

– Sample γ_k^0 ($k = \{1, \ldots, K\}$) from the pdf in (3.10).

– Generate $z_1^0, z_2^0, \ldots, z_N^0$ with probabilities $P[z_n^0 = k] = \frac{1}{K}$.

for $t = 1, 2, \ldots$ to T **do**

— *Update* $\boldsymbol{\alpha}$ —

for $k = 1, 2, \ldots$ to K **do**

1. Propose $\alpha_k^* \sim \mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}, \sigma_{\alpha,k}^2)$ (see (3.17)).
2. Compute the acceptance ratio using expression (3.18).
3. Draw $u \sim \mathcal{U}(0, 1)$.

if ($u <$ ratio) **then**

4. Set $\alpha_k^{(t)} = \alpha_k^*$.

else

5. Set $\alpha_k^{(t)} = \alpha_k^{(t-1)}$.

end if

end for

— *Update* $\boldsymbol{\gamma}$ —

for $k = 1, 2, \ldots$ to K **do**

6. Propose $\gamma_k^* \sim \mathcal{N}^+(\gamma_k^{(t-1)}, \sigma_{\gamma,k}^2)$ (see (3.19)).
7. Compute the acceptance ratio using expression (3.20).
8. Draw $u \sim \mathcal{U}(0, 1)$.

if ($u <$ ratio) **then**

9. Set $\gamma_k^{(t)} = \gamma_k^*$.

else

10. Set $\gamma_k^{(t)} = \gamma_k^{(t-1)}$.

end if

end for

— *Update* $\boldsymbol{z}$ —

for $n = 1, 2, \ldots$ to N **do**

11. Draw z_n from $\{1, \ldots, K\}$ with probabilities (3.16).

end for

end for

3.4.1 Conditional probability $\mathrm{P}[z|\alpha, \gamma, r]$

The label vector z can be updated coordinate-by-coordinate using Gibbs moves. More precisely, the conditional probabilities $\mathrm{P}[z_n|z_{-n}, r_n, \alpha_k, \gamma_k]$ can be computed using the Bayes rule

$$\mathrm{P}[z_n = k|z_{-n}, r_n, \alpha_k, \gamma_k] \propto p(r_n|z_n = k, \boldsymbol{\alpha}, \boldsymbol{\gamma}) p(z_n|z_{-n}) \tag{3.14}$$

where $k = 1, \ldots, K$ (it is recalled that K is the number of classes) and where z_{-n} is the vector z whose nth element has been removed. These posterior probabilities can be expressed as

$$\begin{aligned} \mathrm{P}[z_n = k|z_{-n}, r_n, \alpha_k, \gamma_k] \propto \pi_{n,k} \triangleq \exp\left[\sum_{n' \in \mathcal{V}(n)} \beta\delta(k - z_{n'})\right] \\ \times r_n \int_0^\infty \lambda \exp\left[-(\gamma_k \lambda)^{\alpha_k}\right] J_0(r_n \lambda)\, d\lambda. \end{aligned} \tag{3.15}$$

The integral $r_n \int_0^\infty \lambda \exp\left[-(\gamma_k \lambda)^{\alpha_k}\right] J_0(r_n \lambda)\, d\lambda$ is evaluated using the approximations presented in paragraph 3.4.4. Once all the quantities $\pi_{n,k}$, $k = 1, \ldots, K$, have been computed, they are normalized to obtain the posterior probabilities $\tilde{\pi}_{n,k} \triangleq \mathrm{P}[z_n = k|z_{-n}, r_n, \alpha_k, \gamma_k]$ as follows

$$\tilde{\pi}_{n,k} = \frac{\pi_{n,k}}{\sum_{k=1}^K \pi_{n,k}}. \tag{3.16}$$

Note that the posterior probabilities of the label vector z in (3.15) and (3.16) define an MRF. Finally, samples z_n are generated by drawing discrete variables from $\{1, \ldots, K\}$ with the respective probabilities $\{\tilde{\pi}_{n,1}, \ldots, \tilde{\pi}_{n,K}\}$. Because of its large dimension, sampling z according to (3.16) is the most computationally intensive step of the proposed hybrid Gibbs sampler. Therefore it is important to chose an efficient implementation for this step. In this work z has been sampled using a parallel chromatic Gibbs sampler [102].

3.4.2 Conditional probability density function $p(\alpha|\gamma, z, r)$

The conditional pdf $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{r})$ can be expressed as follows

$$p(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{r}) \propto p(\boldsymbol{r}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z})p(\boldsymbol{\alpha})$$

where $p(\boldsymbol{r}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z})$ is defined in (3.5) and $p(\boldsymbol{\alpha}) = \prod_{k=1}^{K} p(\alpha_k)$. The generation of samples according to $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{r})$ is not easy to perform. We propose in this chapter to sample $\boldsymbol{\alpha}$ coordinate-by-coordinate using MH moves. In this work, the proposal distribution is a truncated normal distribution centered on the previous value of the chain with variance $\sigma^2_{\alpha,k}$

$$\alpha_k^* \sim \mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}, \sigma^2_{\alpha,k}) \tag{3.17}$$

where α_k^* denotes the proposed value at iteration t and $\alpha_k^{(t-1)}$ is the previous state of the chain. The hyperparameters $\sigma^2_{\alpha,k}$ are adjusted during the burn-in period to ensure an acceptance ratio close to $\frac{1}{3}$, as recommended in [103, p. 316]. This adjustment is performed dynamically by a feedback loop that increases or decreases $\sigma^2_{\alpha,k}$ depending on α_k s acceptance ratio over the last 50 iterations. Note that the proposal (3.17) results from the so-called random walk MH algorithm [94, p. 245]. Finally, since the prior for α_k is uniform, the MH acceptance rate of the proposed move can be expressed as follows

$$\text{ratio} = \min\left\{1, \frac{\mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}|\alpha_k^*, \sigma^2_{\alpha,k})}{\mathcal{N}_{(0,2)}(\alpha_k^*|\alpha_k^{(t-1)}, \sigma^2_{\alpha,k})} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k^{(t-1)}, \gamma_k)}\right\} \tag{3.18}$$

where the likelihoods $p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k)$ and $p_{\alpha\mathcal{R}}(r_n|\alpha_k^{(t-1)})$ have been computed using the approximations described in Section 3.4.4.

3.4.3 Conditional probability density function $p(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r})$

The conditional pdf $p(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r})$ can be expressed as follows

$$p(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r}) \propto p(\boldsymbol{r}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z})p(\boldsymbol{\gamma})$$

where $p(\boldsymbol{r}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z})$ is defined in (3.5) and $p(\boldsymbol{\gamma}) = \prod_{k=1}^{K} p(\gamma_k)$. Again, we propose to sample $\boldsymbol{\gamma}$ coordinate-by-coordinate by using MH moves. The proposal distribution associated with this move is a truncated normal distribution centered on the previous value of the chain with variance $\sigma^2_{\gamma,k}$

$$\gamma_k^* \sim \mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}, \sigma^2_{\gamma,k}\right) \tag{3.19}$$

where γ_k^* denotes the proposed value at iteration t, $\gamma_k^{(t-1)}$ is the previous state of the chain and $\mathcal{N}_{\mathbb{R}^+}$ is the Gaussian distribution truncated on $\mathbb{R}^+$. The acceptance ratio for this move is

$$\text{ratio} = \min\left\{1, \frac{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}|\gamma_k^*, \sigma^2_{\gamma,k}\right)}{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^*|\gamma_k^{(t-1)}, \sigma^2_{\gamma,k}\right)} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^*)p(\gamma_k^*|a_0, b_0)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^{(t-1)})p(\gamma_k^{(t-1)}|a_0, b_0)}\right\} \tag{3.20}$$

where the prior distribution $p(\gamma_k|a_0, b_0)$ has been defined in (3.10). Again, the likelihoods $p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^*)$ and $p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^{(t-1)})$ have been computed using the approximations described in Section 3.4.4.

In the particular case $\alpha_k = 2$, the likelihood simplifies to a Rayleigh distribution for which the prior $p(\gamma_k) = \mathcal{IG}(a_0, b_0)$ is conjugate. As a result the generation of samples from the posterior $p(\gamma_k|\alpha_k, \boldsymbol{z}, \boldsymbol{r})$ reduces to drawing samples from the following inverse gamma distribution

$$\gamma_k^{(t)} \sim \mathcal{IG}\left(a_0 + \sum_{\{n|z_n=k\}}^{N} 1, b_0 + \frac{1}{2}\sum_{\{n|z_n=k\}}^{N} r_n^2\right) \tag{3.21}$$

where we recall that $a_0 = 1$ and $b_0 = 1$.

3.4.4 Approximation of the Likelihood

Evaluating the likelihood function defined in (3.5) involves the computation of the following indefinite integral

$$\int_0^\infty \lambda \exp\left[-(\gamma_k \lambda)^{\alpha_k}\right] J_0(r_n \lambda)\, d\lambda. \tag{3.22}$$

In the case where observations are represented using 8-bit precision (i.e., 256-gray levels) the integral can be pre-computed for each level and stored in a look-up-table. The data used in this work is represented using 32-bit precision and the integral had to be solved numerically. This computation is time-consuming and is required for every observation and at every step of the sampler. An efficient way to alleviate this computational complexity is to use the following asymptotic expansions [39, 43][39, 43][39, 43]

$$p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) = \sum_{p=0}^{P} a_p r_n^{2p+1} + o\left(r_n^{2(P+1)+1}\right) \tag{3.23}$$

as $r_n \to 0$ and

$$p(r_n|\alpha_k, \gamma_k) = \sum_{p=1}^{P} b_p r_n^{-\alpha_k p-1} + o\left(r_n^{-\alpha_k(P+1)-1}\right) \tag{3.24}$$

as $r_n \to \infty$, where the coefficients a_p and b_p are

$$a_p = \frac{1}{\alpha_k \gamma_k} \frac{(-1)^p}{(p!)^2 2^{2p}} \Gamma\left(\frac{2p+2}{\alpha_k}\right) \gamma_k^{-2p-1}$$

$$b_p = \frac{(-1)^{p-1}\, 2^{p\alpha_k+1}}{p!\, \pi \gamma_k} \Gamma^2\left(\frac{p\alpha_k+2}{2}\right) \sin\left(\frac{p\pi\alpha_k}{2}\right) \gamma_k^{p\alpha_k+1}.$$

The decision between using (3.23) or (3.24) for a particular value r_n has been determined by a threshold which has been computed off-line. This threshold and the choice of P have been studied empirically by comparing (3.23) and (3.24) to a numerical solution of the true density (3.5). Appropriate threshold and P values have been selected off-line for different values of α_k and stored in a look-up-table that is used by the proposed algorithm. Other considerations regarding the implementation of (3.23) and (3.24) have been studied in [43].

3.5 Experimental Results

This section presents experimental results conducted on synthetic and real data to assess the performance of the proposed α-Rayleigh mixture model and the associated Bayesian estimation algorithm. In these experiments the algorithm convergence has been assessed using the "between-within variance criterion", initially studied by Gelman and Rubin [104] and often used to monitor convergence [105, p. 33]. This criterion requires running M parallel chains of length L with different starting values and computing the so-called *potential scale reduction factor* (PSRF) that compares the between-sequence and within-sequence variances [104]. A PSRF close to 1 indicates good convergence of the sampler. In our experiments we have observed PSRF values smaller than 1.01 which confirm the good convergence of the sampler (a PSRF bellow 1.2 is recommended in [106, p. 332]). These values were computed using $M = 25$ parallel chains of length $L = 1,000$ whose first 900-steps were discarded.

3.5.1 Synthetic Data

To validate the proposed Bayesian method under controlled ground truth conditions (i.e., known true class labels $\boldsymbol{z}$ and statistical parameters $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$), the algorithm described in Section 3.4 was first applied to the synthetic 3-component α-Rayleigh mixture displayed in Fig. 3.4(a). The parameters associated with the mixture components of the 3 different 2D regions are $\boldsymbol{\alpha} = [1.99, 1.99, 1.8]^T$ and $\boldsymbol{\gamma} = [1, 5, 10]^T$. Figure 3.4(b) shows the resulting observation vector $\boldsymbol{r}$, which is the only input provided to the algorithm. Note that the different observations are clearly spatially correlated. The proposed Gibbs sampler has been run for this example using a two-dimensional random field with a 4-pixel neighborhood structure and a granularity coefficient $\beta = 1$. Figure 3.5 shows histograms of the parameters generated by the proposed Gibbs sampler. These histograms are in good agreement with the actual values of the different parameters. Moreover, the MMSE estimates and

the corresponding standard deviations for the different parameters are reported in Table 3.1. These estimates have been computed from a single Markov chain of 25 000 iterations whose first 100 iterations (burn-in period) have been removed. The MMSE estimates are clearly in good agreement with the actual values of the α-Rayleigh mixture components. Figure 3.4(c) shows the class labels estimated by the MAP rule applied to the last samples of the Markov chain. The three classes are recovered with a few misclassifications due to the complexity of the problem.

In order to illustrate the effect of the granularity parameter, we have considered other values of the parameter β. Fig. 3.4(d) and (e) show the class labels obtained with $\beta = 1.2$ and $\beta = 0.8$. We observe that increasing β from 1.0 to 1.2 reduces significantly the number of isolated misclassifications at the expense of increasing errors at the boundaries between the different classes. Decreasing β from 1 to 0.8 increases the number of misclassifications both at the boundaries and within regions.

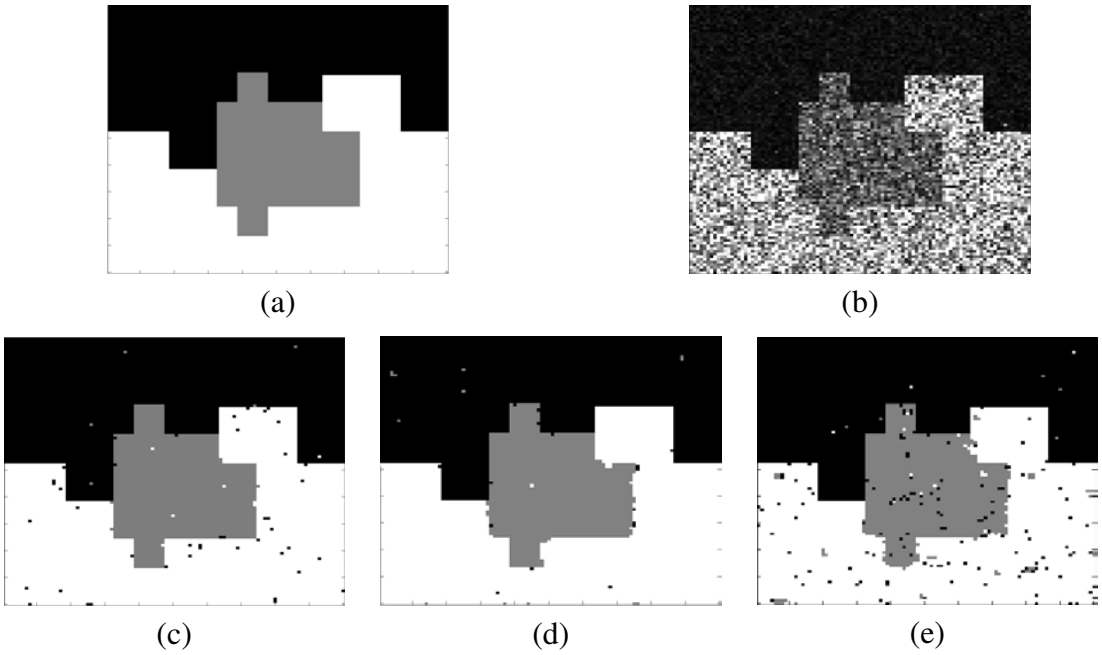


Figure 3.4: (a) True labels, (b) observations, MAP label estimates for (c) $\beta = 1$, (d) $\beta = 1.2$ and (e) $\beta = 0.8$.

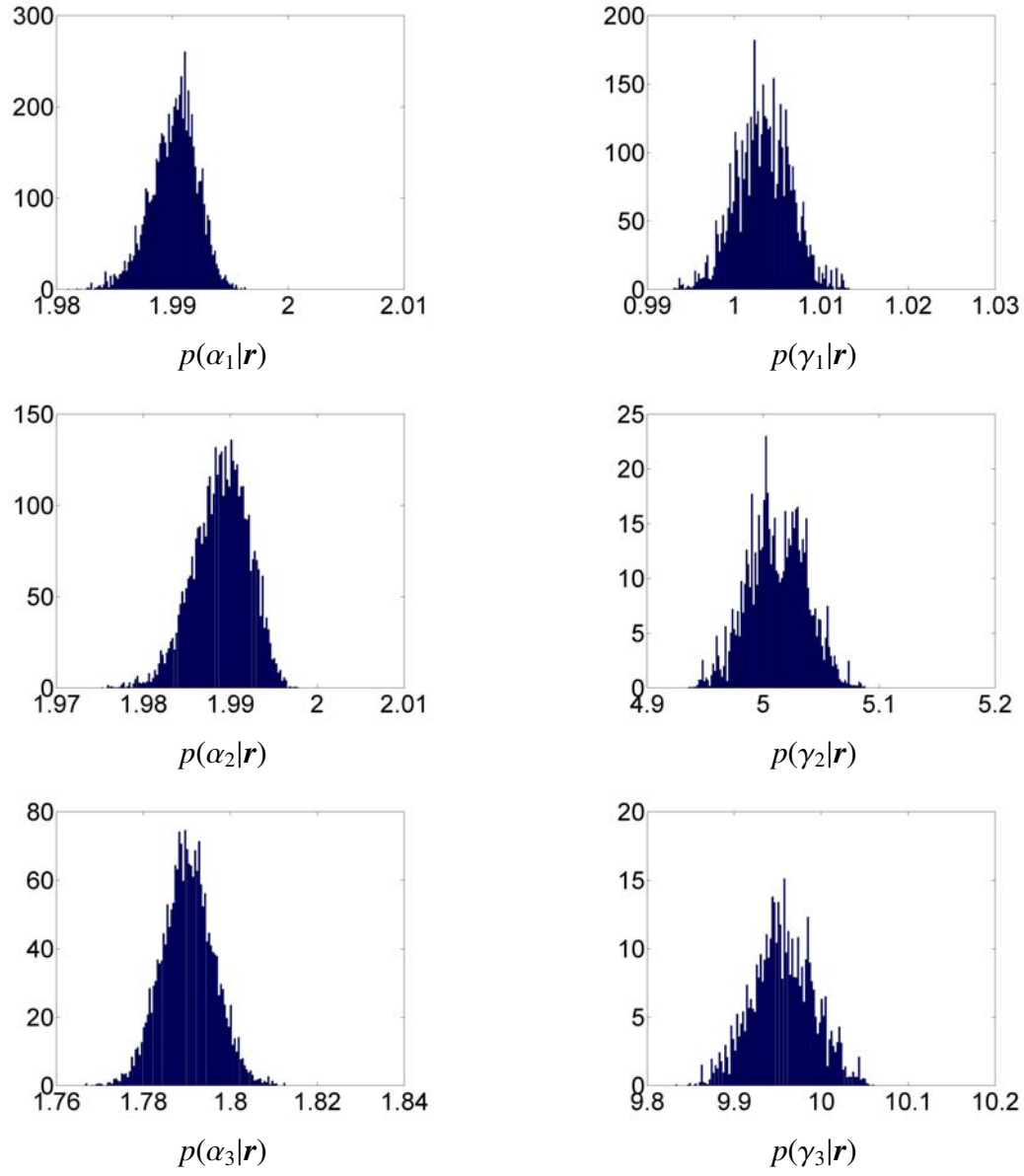


Figure 3.5: Histograms of parameters generated using the proposed Gibbs sampler.

3.5.2 Simulated 3D ultrasound image

The synthetic image studied previously is a toy image that differs from a real ultrasound image in many aspects. These aspects include the spatial organization of skin tissue as well as the different physical phenomena intervening in the formation of ultrasound images (i.e., noise, limited spatial resolution, voxel anisotropy, attenuation, etc.). In order to consider a more realistic scenario, the second set of experiments considers a simulated 3D phantom

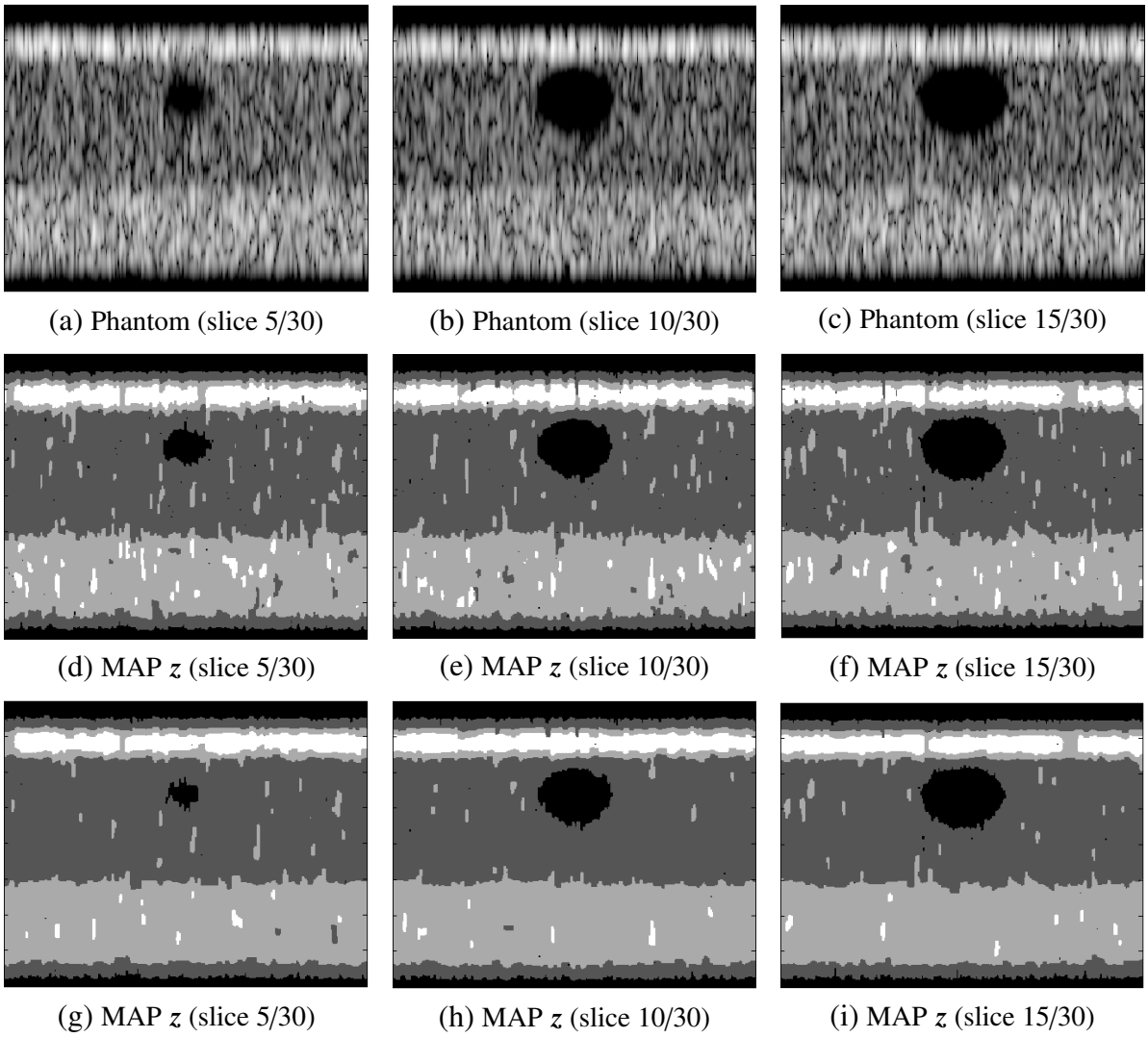
Table 3.1: Parameter estimation

	true value	MMSE estimates	standard deviation
α_1	1.99	1.99	0.002
γ_1	1.00	1.00	0.003
α_2	1.99	1.99	0.003
γ_2	5.00	5.01	0.025
α_3	1.80	1.79	0.006
γ_3	10.00	9.96	0.036

of skin tissue. This 3D phantom image has been simulated using a 3D ultrasound simulator [51], which has been configured with the parameters of the *dermocup* ultrasound system (Atys Medical, France) used in the in-vivo experiments of section 3.5.3. Three slices of the 30-slice 3D phantom are shown in Figs. 3.6(a), 3.6(b) and 3.6(c). The size of each slice is 400×300 pixels. These images are displayed using logarithmic compression; however the proposed algorithm has been applied to B-mode images in linear scale. The 3D skin phantom contains three skin layers (epidermis, papillary dermis and reticular dermis), and one ellipsoidal intra-dermic lesion. Figs. 3.6(d)-(f) and Figs. 3.6(g)-(i) show the corresponding MAP estimated labels obtained with the proposed method using $\beta = 1.0$ and $\beta = 1.2$. We observe that in both cases the skin layers and the lesion are clearly recovered with a few misclassifications due to the complexity of the problem. The number of classes for this experiment has been set to $K = 4$ since there are 3 types of healthy tissue in addition to the lesion. These results were computed using a 3D MRF and a single Markov chain of 1 000 iterations whose first 900 iterations (burn-in period) have been removed.

3.5.3 Application to real data

After validating the proposed Gibbs sampler on synthetic data, this section applies the proposed algorithm to the segmentation of two skin lesions. Experiments were conducted using 3D high frequency B-mode ultrasound images of in-vivo skin tissues. These were acquired with a *dermocup* system (Atys Medical, France), equipped with a single-element



(a) Phantom (slice 5/30) (b) Phantom (slice 10/30) (c) Phantom (slice 15/30)

(d) MAP z (slice 5/30) (e) MAP z (slice 10/30) (f) MAP z (slice 15/30)

(g) MAP z (slice 5/30) (h) MAP z (slice 10/30) (i) MAP z (slice 15/30)

Figure 3.6: Simulated (log-compressed) US images of skin layers with an intradermic lesion and the corresponding estimated labels. Figs. (a)-(c) depict three slices of the 30-slice 3D digital phantom. MAP label estimates for (d)-(f) $\beta = 1$ and (g)-(i) $\beta = 1.2$.

focalized 25MHz wide-band (40-percent) probe sampled at 100MHz with a $53\mu m$ mechanic lateral step. The proposed α-Rayleigh mixture model describes the statistics of envelope (B-mode) ultrasound images without logarithmic compression [5]. Therefore all experiments have been conducted using this type of data. However, to simplify their visual interpretation, results are displayed using logarithmic compression, which is a standard practice in ultrasound imaging [15]. Note that since α-Rayleigh envelope signals arise from symmetric α-stable radio frequency signals [5] it would be possible to apply the pro-

posed method directly to the radio frequency ultrasound image by replacing the α-Rayleigh mixture model (3.5) by a symmetric α-stable mixture model [86].

In this work the number of classes K is assumed to be known a-priori. This important parameter is set by the dermatologist who determines visually the number of tissues within the region to be processed. For skin tissues the number of classes depends on the number of layers contained in that region (i.e., epidermis, papillary (upper) dermis, reticular (lower) dermis, hypodermis) in addition to the lesion. More details regarding the number of classes are available in section 3.5.3.

The Potts granularity coefficient β has been chosen heuristically by testing a few values between 0.5 and 1.5. These tests have suggested that segmentation results best agree with expert annotations for $\beta \in (1, 1.5)$. Finally, β was set to 1 in order to minimize the risk of over-smoothing the segmentation results, which was the main concern of dermatologists. Section 3.5.3 presents segmentation results obtained with other values of β with discussions. Future work will study the estimation of β jointly with the other unknown parameters of the model, as in [99].

Justification of the α-Rayleigh mixture model

the α-Rayleigh mixture model used in this work is based on The assumption that the statistics of single-tissue regions can be well described by an α-Rayleigh distribution. To support this assumption Fig. 3.7 compares the histogram obtained from a B-mode ultrasound image of in-vivo forearm dermis with the corresponding α-Rayleigh, Nakagami and Gamma distribution fits (additional fits are provided in [5]). To better illustrate fitting at the tails, Fig. 3.7 displays the probability density functions in logarithmic scale. We observe that the α-Rayleigh distribution provides the best fit and is the only one to accurately describe the heavy-tail of the histogram.

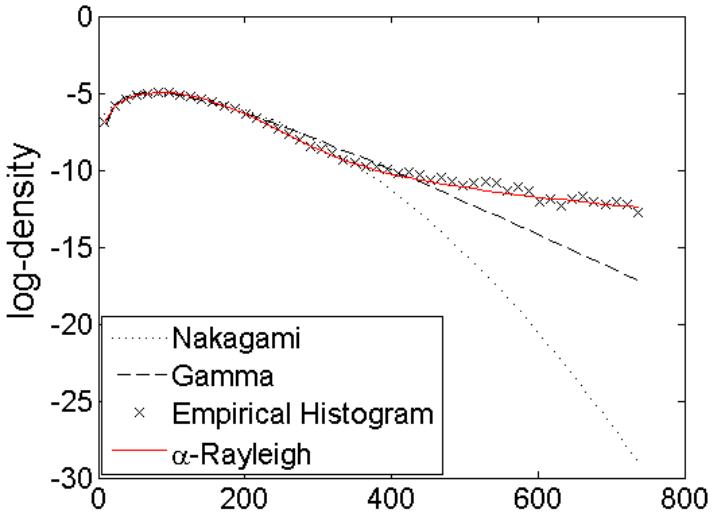


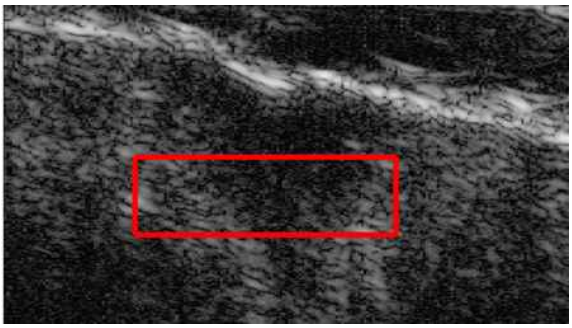
Figure 3.7: Comparison of the B-mode histogram obtained from forearm dermis, and the corresponding estimations using the Nakagami, Gamma and α Rayleigh distributions. Plots presented in logarithmic scale to illustrate fitting at the tails.

Preliminary 2D and 3D experiments

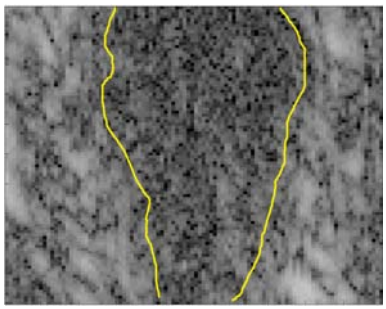
The two following experiments illustrate the importance of introducing spatial correlation between the mixture components. Fig. 3.8(a) shows a skin lesion outlined by the red rectangle. This region is displayed with coarse expert annotations (yellow curve) in Fig. 3.8(b). It should be noted that annotations approximately localize the lesion and do not represent an exact ground truth. The following experiments have been conducted with granularity coefficient $\beta = 1$ and the number of classes $K = 2$ since there are only two types of tissue (i.e., lesion and healthy reticular dermis) within the region of interest (ROI). The results have been computed from a single Markov chain of 1 000 iterations whose first 900 iterations (burn-in period) have been removed.

First, the proposed Bayesian algorithm was used to label each voxel of the ultrasound image as *healthy* or *lesion* tissue. The estimated labels obtained using a bidimensional random field are displayed in Fig. 3.8(c). For comparison purposes, Fig. 3.8(d) shows the estimation results when labels are considered a priori independent, as in [3]. Due to

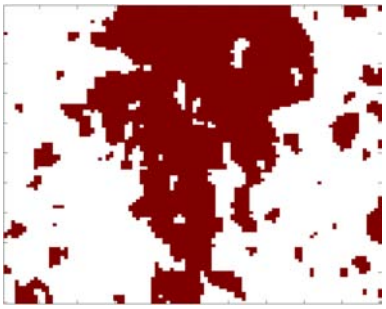
the proposed MRF prior for the labels, the spatial correlations between image voxels are clearly recovered with the proposed segmentation procedure.



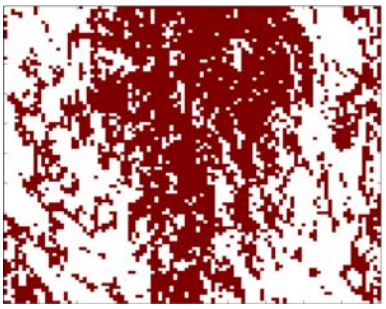
(a) Dermis view with skin lesion (ROI = $100 \times 100 \times 3$).



(b) ROI (slice 2)



(c) MRF Labels z



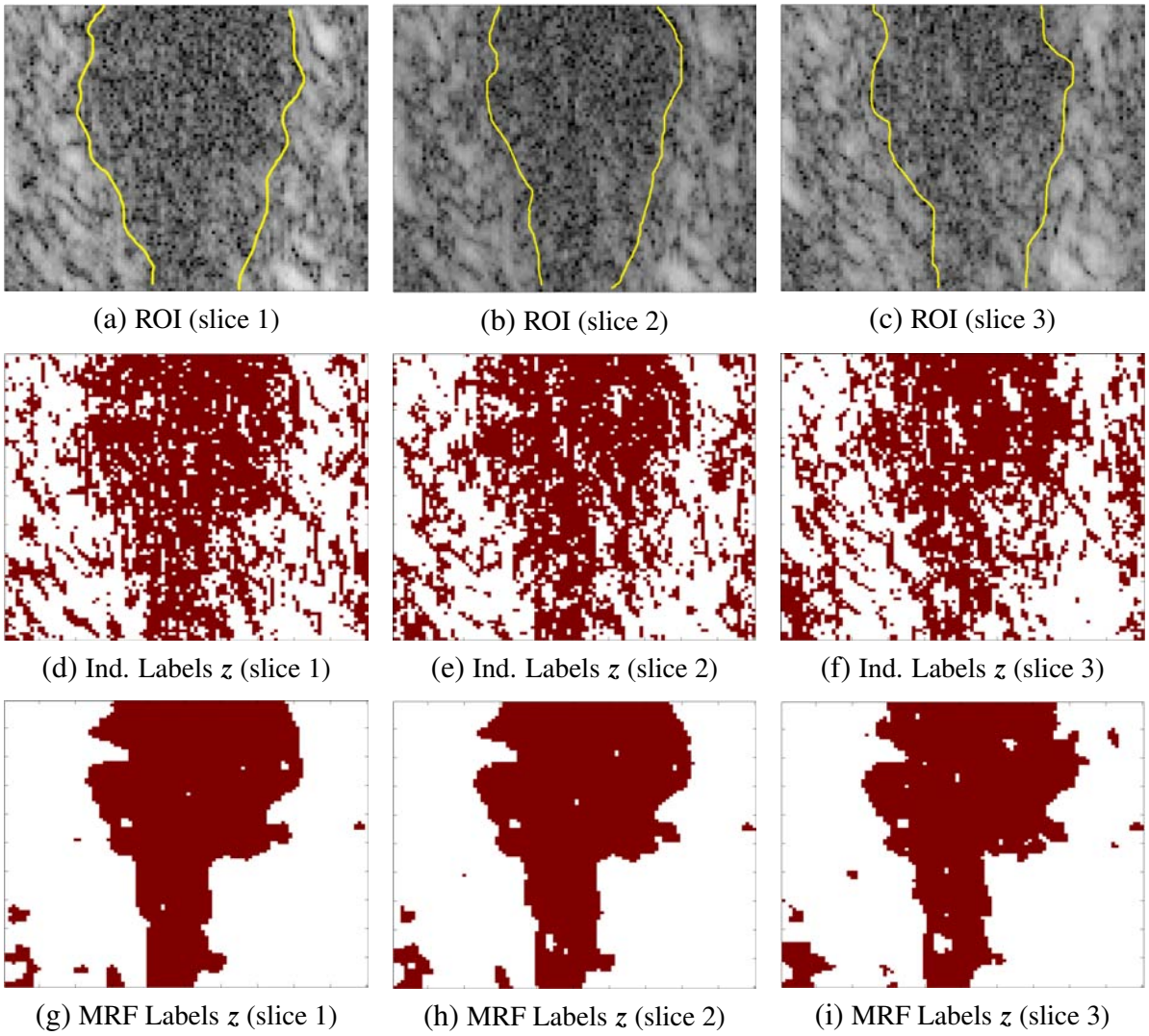
(d) Independent Labels z

Figure 3.8: Log-compressed US images of skin lesion and the corresponding estimated labels (healthy = white, lesion = red) [3])

.

In a second experiment the algorithm was applied in three dimensions using a tridimensional random field. Three slices of the 3D B-mode image associated with the ROI are shown in Figs. 3.9(a), 3.9(b) and 3.9(c). Figs. 3.9(d), 3.9(e) and 3.9(f) show the results obtained when labels are considered a priori independent, as in [3]. The labels estimated with the proposed 3D method are displayed in Figs. 3.9(g), 3.9(h) and 3.9(i) where healthy voxels are represented in white and lesion voxels in red. The size of the 3D images is $100 \times 100 \times 3$ voxels and computing class label estimates using $1,000$ iterations of the proposed algorithm required 43.5 seconds (see Section 3.5.3 for more details about the computational complexity). We observe that most of the MAP labels are in very good agreement with the expert annotations. The improvement obtained when considering correlations in the 3rd dimension can be assessed by comparing Figs. 3.8(c) and 3.9(h), which

have been computed from the same data slice. We observe that using a 3D MRF reduces significantly the number of misclassifications and improves the agreement with the expert annotations.



(a) ROI (slice 1) (b) ROI (slice 2) (c) ROI (slice 3)

(d) Ind. Labels z (slice 1) (e) Ind. Labels z (slice 2) (f) Ind. Labels z (slice 3)

(g) MRF Labels z (slice 1) (h) MRF Labels z (slice 2) (i) MRF Labels z (slice 3)

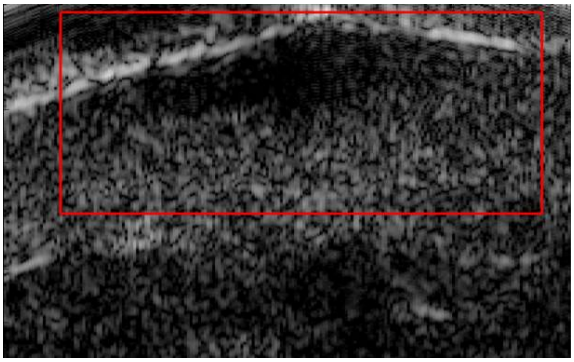
Figure 3.9: Log-compressed US images of skin lesion and the corresponding estimated labels (*healthy* = white, *lesion* = red). Figs. (d)-(f) show the results obtained by considering that voxel labels are independent, as in [3]. Figs. (g)-(i) show the results obtained with the proposed 3D Markov random field (MRF) method.

Comparison with a state of the art method

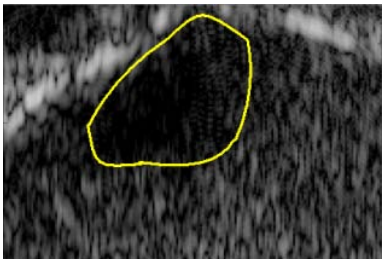
The proposed algorithm has been compared with the state of the art method proposed in [1]. This method considers implicitly that the image is a mixture of two Rayleigh components

and separates them using an LS algorithm. Comparison has been performed with 2D and 3D random fields. The following experiments were conducted with granularity coefficient $\beta = 1$ and number of classes $K = 4$ since there are 3 types of healthy tissue within the ROI in addition to the lesion. The results have been computed from a single Markov chain of 1 000 iterations whose first 900 iterations (burn-in period) have been removed.

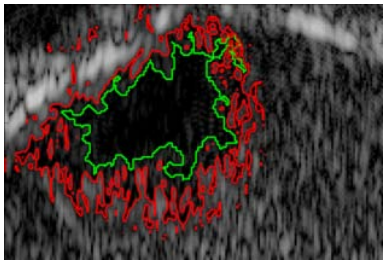
Fig. 3.10(a) shows a skin lesion contained in the ROI outlined by the red rectangle. This region is displayed with coarse expert annotations in Fig. 3.10(b). The proposed 2D Bayesian algorithm was used to label each voxel of the ROI as *healthy* or *lesion* tissue. Then, from the vector of voxels that were labeled as *lesion* we extracted the contour of the largest connected region. The results displayed in Fig. 3.10(c) show the regular shape of the contour obtained by our method, whereas the LS method with strong regularization yields a more irregular contour.



(a) Dermis view with skin lesion (ROI = $160 \times 175 \times 16$).



(b) ROI (slice 7)



(c) 2D Segmentation contour

Figure 3.10: Log-compressed US images of skin melanoma tumor and the corresponding estimated segmentation contours (*proposed* = green, [1] = red). Figure (c) proposed 2D algorithm and level set

.

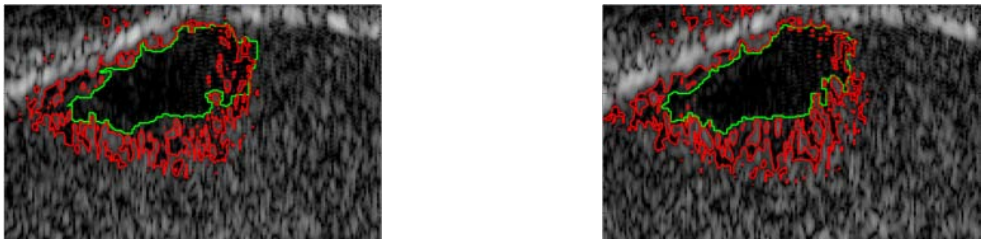
The proposed algorithm was also applied to a 3D B-mode image using a tridimensional random field. The results for eight slices of the image associated with the ROI depicted in Fig. 3.10(a) are shown in Figs. 3.11(a) - 3.11(h). The same color code is used for the contours as in the 2D experiment. The regular shape of the contour obtained by the proposed method is more visible and the recovered lesion fits better the area depicted by the expert. Finally, Fig. 3.12 shows two viewpoints of a 3D reconstruction of the lesion s surface. We observe that the tumor has a semi-ellipsoidal shape which is cut at the upper left by the epidermis-dermis junction. The tumor grows from this junction towards the deeper dermis, which is at the lower right.

Finally, it should be noted that in the in-vivo experiments the proposed algorithm has been applied to regions of interest, as opposed to entire 3D images. This has been motivated by the fact that dermatological ultrasound imaging is used to examine specific regions that have been previously identified by the dermatologist. The method presented in this work should be understood in that clinical context and is not intended to be used in unsupervised applications.

Segmentation results for different values of β

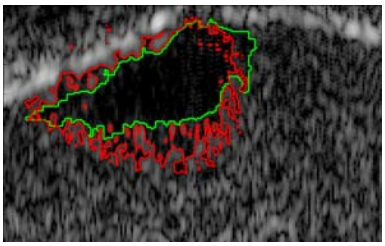
To assess the in uence of the granularity coefficient, this section presents segmentation results obtained by repeating the previous experiment using different values of β. As explained previously, this hyper-parameter tunes the amount of correlation that the Potts Markov field introduces between the class labels. A small value of β defines a weak prior distribution that is very sensitive to noise, contrary to a large value of β that leads to a strong prior that promotes few and large homogeneous regions.

Fig. 3.13(a) shows the 8th slice of the B-mode 3D ROI previously displayed in Figs. 3.10 and 3.11. For visual interpretation this image is displayed in logarithmic scale. Figs. 3.13(b)-(f) show the 8th slice of the 3D MAP class labels obtained with the proposed algorithm for different values of β. These results were computed using $K = 4$ and are displayed

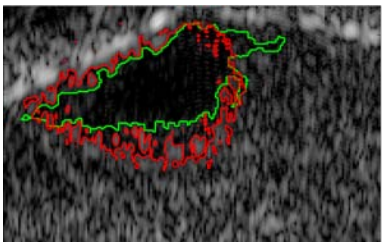


(a) Slice 1

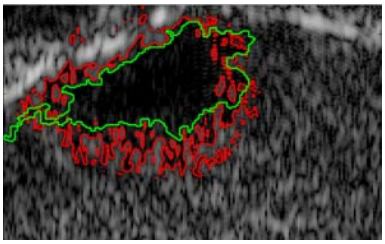
(b) Slice 3



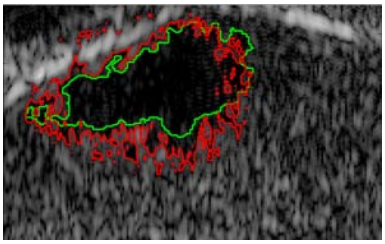
(c) Slice 5



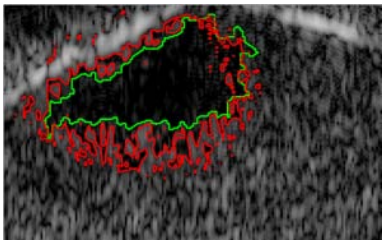
(d) Slice 7



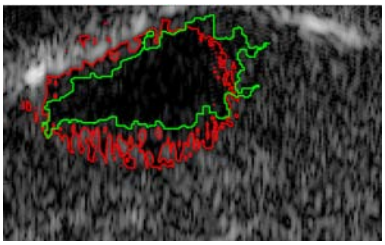
(e) Slice 9



(f) Slice 11



(g) Slice 13



(h) Slice 15

Figure 3.11: 3D segmentation of an 8-slice image.

using the following color code: *lesion* = black, *epidermis* = white, *pap. dermis* = dark gray and *ret. dermis* = light gray. These estimates have been computed from a single Markov chain of 1 000 iterations whose first 900 iterations (burn-in period) have been removed.

We observe that the best results are obtained for $\beta = 1$ and $\beta = 1.25$. The results obtained by fixing the granularity coefficient to a small value ($\beta < 1$) are corrupted by ultrasound speckle noise and fail to capture the different skin layers. On the other hand,

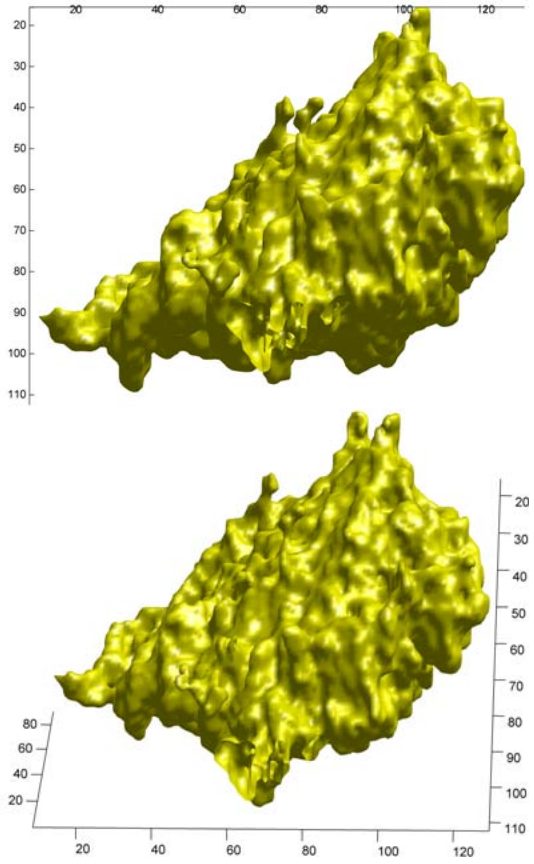
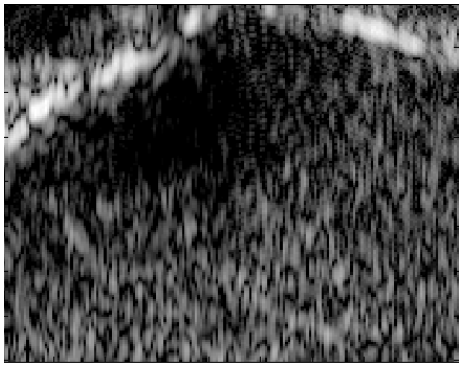


Figure 3.12: 3D reconstruction of the melanoma tumor.

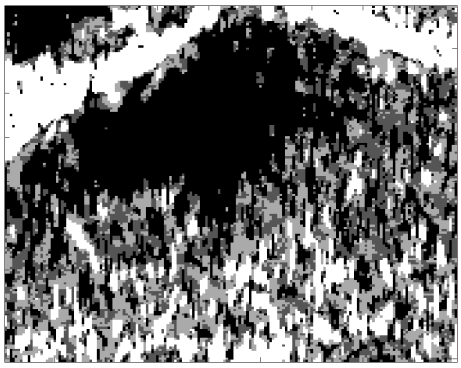
fixing β to a too high value (i.e., $\beta > 1.5$) enforces too much spatial correlation and yields a segmentation with artificially straight boundaries.

Segmentation results for different numbers of classes K

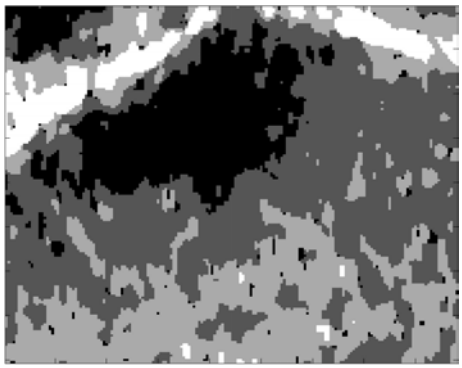
As explained previously, this work assumed that the number of classes K is known a-priori. For skin tissues the number of classes depends on the number of layers contained in that region (i.e., epidermis, papillary (upper) dermis, reticular (lower) dermis, hypodermis) in addition to the lesion. The number of classes will typically vary from $K = 3$ for very small lesions (contained in the upper dermis) to $K = 5$ for lesions that have invaded the lower dermis. In any case the number of classes should be at least equal to the number of tissues that must be identified.



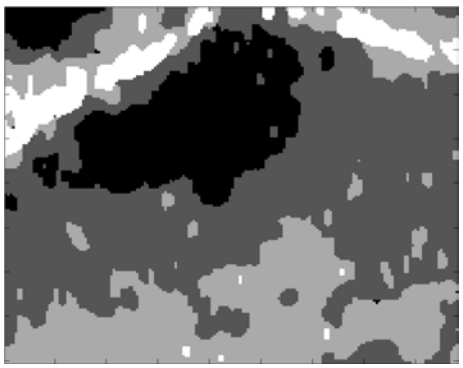
(a) Dermis view with skin lesion (slice 8)



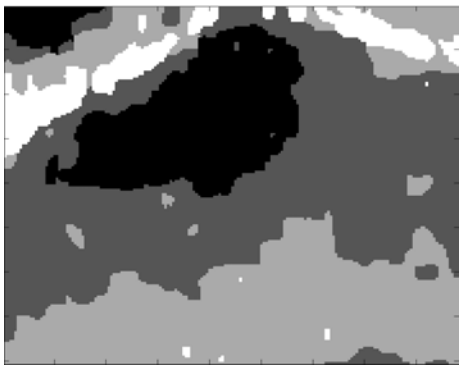
(b) MAP Class Labels (β = 0.5)



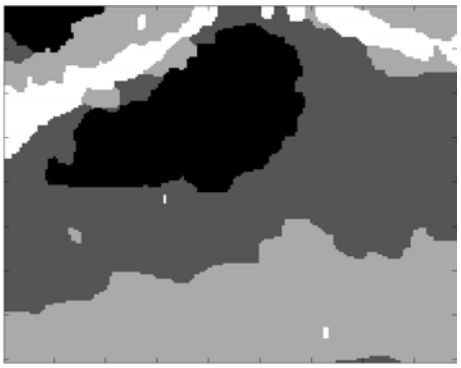
(c) MAP Class Labels (β = 0.75)



(d) MAP Class Labels (β = 1.0)



(e) MAP Class Labels (β = 1.25)



(f) MAP Class Labels (β = 1.5)

Figure 3.13: Log-compressed US images of skin lesion and the corresponding estimated class labels (*lesion* = black, *epidermis* = white, *pap. dermis* = dark gray, *ret. dermis* = light gray)

.

Note that increasing the number of classes beyond the number of tissues will outline differences within a same biological tissue. To illustrate this point the proposed method has been applied to one of our 3D ultrasound images using one additional class. Fig. 3.14(b) shows one slice of the MAP labels estimated using one class per tissue (one class per skin

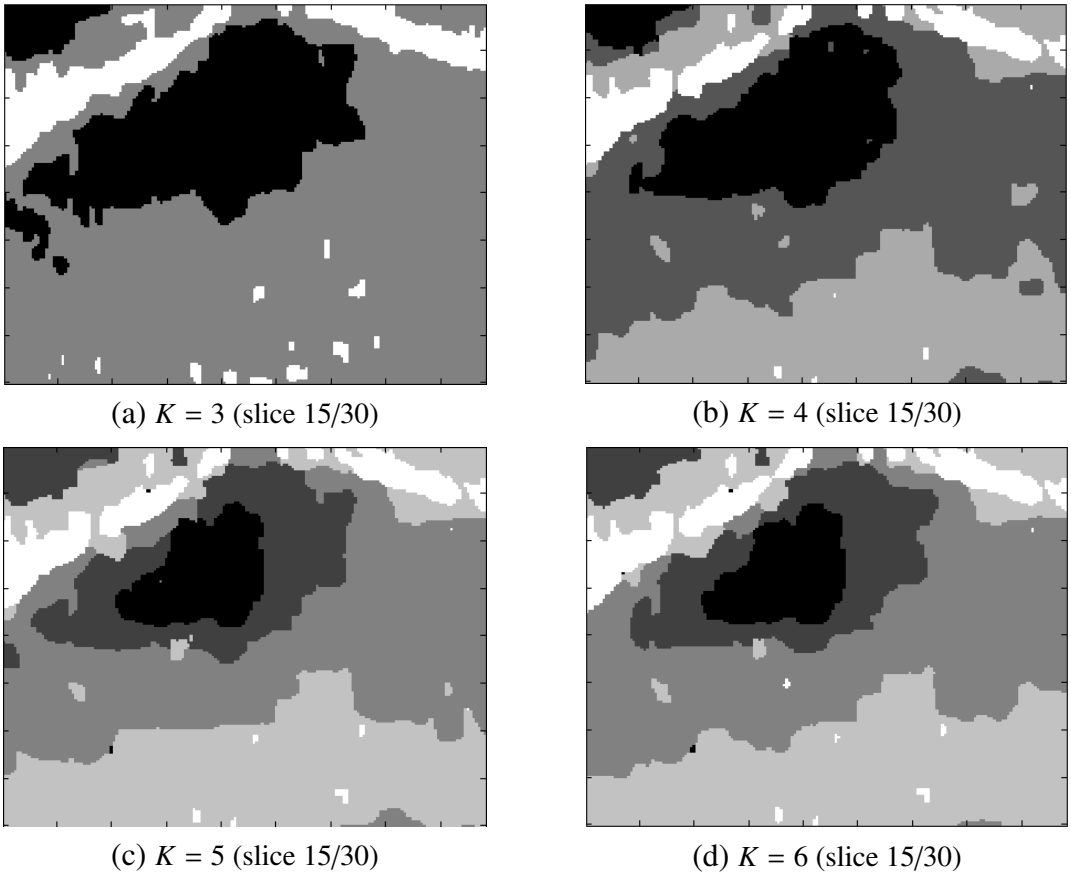
layer plus one class for the lesion), i.e., $K = 4$. Fig. 3.14(c) shows results obtained when considering an additional class, i.e., $K = 5$. We observe that introducing an additional class has not modified significantly the estimation of the lesion boundaries. The proposed method has assigned an additional class to the core of the lesion, which may correspond to necrotic tissue. Moreover, using too many additional classes will result in empty or redundant classes. Fig. 3.14(d) shows one slice of the MAP labels obtained with two additional classes ($K = 6$). We observe that this result is very similar to the one obtained for a single additional class ($K = 5$). In this case the proposed method has assigned one additional class to the core of the lesion and left the other supplementary class unassigned, indicating that $K = 5$ is a more suitable number of classes.

On the other hand, underestimating the number of classes can degrade the performance of the proposed method significantly. Fig. 3.14(a) shows one slice of the segmentation results obtained when there are not enough classes to represent all tissues, i.e., $K = 3$. In this case the papillary dermis and the reticular dermis have been merged into a single class. As a result the estimation of the boundary between the lesion and the papillary dermis is less accurate.

Segmentation of entire 3D B-mode images

In this work the proposed algorithm has been applied to regions of interest, as opposed to entire 3D images. This is motivated by the fact that dermatological ultrasound is used to reexamine carefully regions that have been previously identified using a faster screening modality, typically a dermatoscope (magnifying glass). The method presented in this work should be understood in that clinical context and is not intended for batch processing entire 3D images nor for fully unsupervised applications.

However, for completeness the proposed method has also been applied to an entire B-mode ultrasound image. It should be noted that identifying the hypodermis (the lower part of the image) is difficult because the ultrasound system is calibrated to target the upper



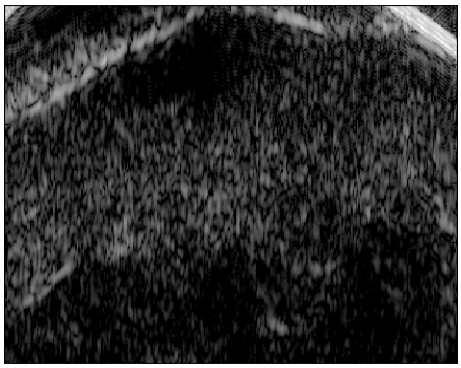
(a) $K = 3$ (slice 15/30) (b) $K = 4$ (slice 15/30)

(c) $K = 5$ (slice 15/30) (d) $K = 6$ (slice 15/30)

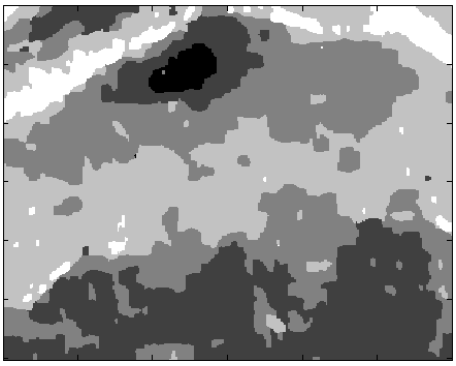
Figure 3.14: Tissue labeling results (central slice) of a 3D ultrasound image containing a lesion. (a) Using too few classes results in a common class for papillary and reticular dermis. (b) Correctly using one class per tissue. (c) The introduction of a supplementary class reveals the core of the lesion. (d) The introduction of two supplementary classes results in an empty class.

dermis, which is where lesions are commonly found. The hypodermis is significantly more affected by attenuation, resolution degradation and noise.

Figure 3.15 shows one slice of the ultrasound image and the corresponding 3D MAP class labels estimated with the proposed algorithm. For visual interpretation Fig. 3.15(a) is displayed using logarithmic scale. The results shown in Fig. 3.15(b) have been computed using $K = 5$ classes because there are 4 types of healthy tissue in addition to the lesion. We observe that the lesion and its core have been accurately detected. Also, the boundaries between the skin layers have also been correctly identified. The hypodermis has been mostly associated to the same class as the lesion, probably because both tissues are very hypoechogenic. Increasing the number of classes did not improve this result.



(a) Log-compressed Ultrasound image (8th slice)



(b) Labels (8th slice)

Figure 3.15: Log-compressed US images of skin lesion and the corresponding estimated class labels computed on the entire image

.

Computational Complexity

Table 3.2 provides averaged execution times for 500 iterations of the proposed algorithm for several image sizes in 2D and 3D and several numbers of classes. The time required to reach convergence can be calculated by multiplying these values by $\frac{9}{5}$, which corresponds to a burn-in period of 900 iterations. These tests have been computed on a workstation equipped with an Intel Core 2 Duo @2.1 GHz processor, 3MB L2 and 3GB of RAM memory. The main loop of the Gibbs sampler has been implemented on MATLAB R2010b (The MathWorks Inc., Natick, MA, 2010). However, C-MEX functions have been used to compute the likelihood and to draw samples of $\boldsymbol{z}$ from (3.15).

Finally, table 3.3 provides the average computing times for the LS method [1] for different image sizes. These results have been computed using the time and space sampling steps indicated in [1] ($\Delta t = 0.1$, $\Delta x = 1$, $\Delta y = 1$ and $\Delta z = 1$). Average estimates were obtained by repeating each experiment 10 times. Note that a comparison between these computing times and those of the proposed method has to be made with some precautions since the LS method has been implemented in [1] using MATLAB whereas the proposed method uses C-MEX functions. .

Table 3.2: Computing times (in seconds) of 500 iterations for different image sizes and number of classes.

	K = 2	K = 4	K = 8	K = 16
64×64	4.9	10.2	18.5	35.0
128×128	8.1	14.6	27.0	51.7
256×256	19.7	36.1	63.5	123.0
512×512	73.7	126.0	223.2	427.0
$64 \times 64 \times 16$	20.8	36.6	68.3	129.5
$128 \times 128 \times 16$	75.1	141.5	254.0	524.4
$256 \times 256 \times 16$	317.7	578.1	1060.5	2300.5
$512 \times 512 \times 16$	1175.7	2225.0	4316.8	9600.0

Table 3.3: Average computing times for the Level Set method [1] for different image sizes.

	250 iterations	time to convergence
64×64	0.77 sec.	0.77 sec. (250 iterations)
128×128	2.38 sec.	9.52 sec. (1000 iterations)
256×256	12.23 sec.	73.35 sec. (1500 iterations)
512×512	22.95 sec.	183.6 sec. (2000 iterations)

3.6 Conclusion

A spatially coherent finite mixture of α-Rayleigh distributions was proposed to represent the statistics of envelope ultrasound images backscattered from multiple tissues. Spatial correlation was introduced into the model by a Markov random field that promotes dependance between neighbor pixels. Based on the proposed model, a Bayesian segmentation method was derived. Bidimensional and tridimensional implementations of this segmentation method were presented using a Markov chain Monte Carlo algorithm that jointly estimates the unknown parameters of the mixture model and classifies voxels into different tissues. The method was successfully applied to several high frequency 3D ultrasound images. Experimental results showed that the proposed technique outperforms a state of the art method in the segmentation of in-vivo lesions. A tridimensional reconstruction of a melanoma tumor suggested that the resulting segmentations can be used to assess lesion

penetration in dermatologic oncology. Future work includes the characterization of the performance of the segmentation algorithm and the study of estimation algorithms for the granularity coefficient defining the Markov random field prior. A comparison with an ML estimator followed by median filtering is also an area of interest for potential future work.

Estimation du coefficient de granularité d'un champ aléatoire de Potts-Markov dans un algorithme MCMC

La modélisation de la corrélation spatiale est fondamentale dans plusieurs applications de traitement d images. Les champs aléatoires de Markov (MRF) ont été reconnus comme un outil efficace pour capturer la cohérence spatiale [87, 107–111]. Le modèle de Potts [84] est un champ qui généralise le modèle d Ising, particulièrement appliqué dans la segmentation Bayésienne. Le degré de cohérence spatiale introduite par un champ de Potts est contrôlé par un coefficient de granularité. Dans la plupart des applications, ce paramètre est fixé empiriquement. Ce chapitre étudie l estimation du paramètre de Potts conjointement avec les autres paramètres d un modèle Bayésien standard de segmentation d images.

Précisément, nous considérons un modèle Bayésien défini par un modèle conditionnel d observation avec des paramètres inconnus et un vecteur d étiquettes discrètes cachées z dont la distribution a priori est un modèle de Potts avec un hyper-paramètre β. Du point de vu méthodologique, faire de l inférence sur β est difficile parce que la distribution $f(z,\beta)$ dépend de la constante de normalisation (notée $C(\beta)$), qui est non calculable. Ce problème a reçu une certaine attention dans la littérature récente du traitement d images afin de créer des algorithmes de segmentation non supervisés [82, 93, 112–114].

Dans ce travail, nous nous intéressons à l estimation de β dans un algorithme Monte Carlo par chaine de Markov (MCMC) qui traite des images 2D et 3D [4, 79, 97, 98, 100, 101]. Les méthodes MCMC sont des outils puissants pour effectuer de l inférence Bayésienne dans des problèmes où le minimum de l erreur aux moindres carrés (MMSE) et le maximum a posteriori (MAP) sont difficiles à établir analytiquement. Les méthodes MCMC génèrent des échantillons asymptotiquement distribués selon la distribution a posteriori jointe des paramètres inconnus. Les échantillons sont alors utilisés pour approcher les estimateurs Bayésiens. Cependant, les méthodes MCMC classiques ne peuvent pas s appliquées directement aux problèmes intégrant un modèle de Potts. Ceci

procède du fait que l inférence sur β requiert le calcul de la constante de normalisation $C(\beta)$. Des algorithmes MCMC spécifiques ont été conçus pour estimer les paramètres de champs de Markov [82, 91, 112, 115–117]. Un algorithme Bayésien variationnel basé sur l approximation de $C(\beta)$ a été proposé récemment dans [93]. L estimation de β par le maximum de vraisemblance à l aide de l algorithme Espérance-maximisation a été étudiée dans [113, 114, 118]. Les stratégies adoptées dans ces travaux pour contourner le problème du calcul de $C(\beta)$ peuvent se diviser en quatre catégories.

Les méthodes dites des estimateurs par pseudo-vraisemblance évitent le calcul de $C(\beta)$ en l éliminant de la distribution a posteriori. Plus précisément, cela revient à définir une distribution a priori de manière à ce que la constante de normalisation s élimine de la distribution a posteriori par simplification de son expression (c est-à-dire $f(\beta) \propto C(\beta)\mathbf{1}_{\mathcal{R}^+}(\beta)$). Ceci donnerait les estimateurs dits de pseudo-vraisemblance [115, 116, 119]. Malgré son attrait analytique, cette approche conduit à une distribution a posteriori souvent mal-adaptée et donne de mauvaises estimations [120]. De plus, comme cela a été noté dans [117], ce type de distribution a priori dépendrait des données ce qui est peu recommandé dans le paradigme Bayésien [121, p. 36]).

Une autre approche pour traiter le problème est d approximer la constante $C(\beta)$. Il existe trois catégories d approximations : i) celles basées sur des développements analytiques, ii) celles basées sur des stratégies d échantillonnage, iii) et celles basées sur la combinaison des deux. Les méthodes mean field, tree-structured mean field, la méthode de Bethe energy, ainsi que deux stratégies d échantillonnage basées sur l algorithme MCMC de Langevin ont été considérées dans [117]. Plus récemment, des expressions récursives exactes ont été proposées pour calculer $C(\beta)$ analytiquement [93, 122]. Cependant, ces méthodes ont été uniquement appliquées à des problèmes de petite taille (champs inférieurs à 40x40) avec un $\beta < 0.5$. Une autre méthode d approximation par échantillonnage consiste à estimer $C(\beta)$ par intégration Monte Carlo [94, Chap. 3], mais avec un coût de calcul élevé et des estimations biaisées. De meilleurs résultats peuvent être obtenus par échantillonnage préférentiel

ou par la méthode d Ogata [123]. Ces méthodes ont été appliquées à l estimation de β dans un algorithme MCMC de traitement d images [91]. Quoique plus précise que l intégration Monte Carlo, l approximation de $C(\beta)$ par ces deux méthodes nécessite un temps de calcul important et devient infaisable pour de larges champs. Ceci a motivé les travaux récents qui réduisent le calcul en combinant l échantillonnage préférentiel avec des approximations analytiques. Plus précisément, des méthodes combinant l échantillonnage préférentiel et des techniques d extrapolation ont été proposées pour le modèle de Potts dans [82] et pour un modèle de Potts à trois états dans [112]. Cependant, nous avons trouvé que ces techniques introduisent un biais significatif (cf. section 4.5.2).

La littérature récente en informatique statistique a montré qu il est possible d éviter le calcul de $C(\beta)$ dans un algorithme MCMC de type Metropolis Hastings [94] en introduisant des variables auxiliaires appropriées [92, 124]. Dans le travail de Moller et al. [92], un vecteur auxiliaire $\boldsymbol{w}$ distribué selon la même distribution que le vecteur d étiquettes $\boldsymbol{z}$) (c est-à-dire $f(\boldsymbol{z}/\beta)$) a été introduit. Des algorithmes Metropolis Hastings qui ne requièrent pas le calcul de $C(\beta)$ ont été proposés pour échantillonner la distribution jointe $f(\beta, \boldsymbol{w}|\boldsymbol{z})$, qui admet la densité a posteriori exacte désirée $f(\beta|\boldsymbol{z})$ comme distribution marginale [92]. Malheureusement, cette méthode souffre d un rapport d acceptation qui se dégrade sévèrement quand la dimension de z augmente. Elle est ainsi inappropriée aux applications de traitement d images (section 4.5.2). Des méthodes de variables auxiliaires, donnant un rapport d acceptation nettement meilleur, ont été proposées dans [124] en utilisant plusieurs vecteurs auxiliaires et des simulateurs Monte Carlo séquentiels [125]. Ces méthodes pourraient être intéressantes pour l estimation du coefficient de Potts, mais ne sont pas considérées dans cette thèse à cause de leur coût de calcul. Une méthode de variable auxiliaire alternative basée sur l échantillonnage simple du rapport $\frac{C(\beta)}{C(\beta^*)}$ (cf. Section 4.3.3) a été proposée dans [126] et améliorée dans [127] en utilisant plusieurs vecteurs auxiliaires et des simulateurs Monte Carlo séquentiels.

Finalement, il est possible d éviter le calcul de la constante de normalisation $C(\beta)$ en utilisant des méthodes MCMC sans vraisemblance [128]. Ses méthodes remplacent l évaluation de la vraisemblance "non calculable" dans un algorithme MCMC de Metropolis Hastings par une méthode de simulation par rejet. Plus précisément, comme pour la méthode de variable auxiliaire [92], un vecteur aléatoire w distribué selon la vraisemblance $f(z|\beta)$ est introduit. Des algorithmes de Metropolis Hastings à deux pas ne nécessitant pas l évaluation de $f(z|\beta)$ ni de $C(\beta)$ peuvent alors générer des échantillons asymptotiquement distribués selon la distribution a posteriori $f(\beta|z)$ [128]. Malgré leur infaisabilité, ces méthodes ont donné naissance à l approche du calcul bayésien approché (ABC) [129, 130]. Cette approche étudie les méthodes sans vraisemblance pour générer des échantillons à partir de densités a posteriori approximatives avec un coût raisonnable. Il semble que ces techniques prometteuses, considérées comme "les plus satisfaisantes pour les problèmes de vraisemblances non calculables" [130], n aient pas encore été appliquées aux problèmes de traitement d images.

La contribution principale de ce chapitre est de proposer un algorithme MCMC par calcul bayésien approché (ABC MCMC) pour estimer conjointement le vecteur d étiquettes z, le paramètre de granularité β et les autres paramètres inconnus du modèle Bayésien. L estimation de β est incluse dans l algorithme MCMC à l aide d une méthode de calcul bayésien approché (ABC) particulièrement adaptée au modèle de Potts et à des données de taille large. On y montre que l estimation de β peut facilement être intégrée à des algorithmes MCMC existants où β était fixé empiriquement. Des applications à des images 2D et 3D de grandes tailles illustrent les performances de la méthode proposée.

Chapter 4

Estimating the Granularity Coefficient of a Potts-Markov Random field within an MCMC algorithm

This chapter addresses the problem of estimating the Potts parameter β jointly with the unknown parameters of a Bayesian model within a Markov chain Monte Carlo (MCMC) algorithm. Standard MCMC methods cannot be applied to this problem because performing inference on β requires computing the intractable normalizing constant of the Potts model. In the proposed MCMC method the estimation of β is conducted using a likelihood-free Metropolis-Hastings algorithm. Experimental results obtained for synthetic data show that estimating β jointly with the other unknown parameters leads to estimation results that are as good as those obtained with the actual value of β. On the other hand, assuming that the value of β is known can degrade estimation performance significantly if this value is incorrect. To illustrate the interest of this method, the proposed algorithm is successfully applied to real bidimensional SAR and tridimensional ultrasound images.

4.1 Introduction

Modeling spatial correlation in images is fundamental in many image processing applications. Markov random fields (MRF) have been recognized as efficient tools for capturing these spatial correlations [87, 107–111]. One particular MRF often used for Bayesian classification and segmentation is the Potts model [84], which generalizes the binary Ising model to arbitrary discrete vectors. The amount of spatial correlation introduced by this model is controlled by the so-called *granularity coefficient* β. In most applications this important parameter is set heuristically by cross-validation.

This chapter studies the problem of estimating the Potts coefficient β jointly with the other unknown parameters of a standard Bayesian image classification or segmentation problem. More precisely, we consider Bayesian models defined by a conditional observation model with unknown parameters and a discrete hidden label vector z whose prior distribution is a Potts model with hyperparameter β (this Bayesian model is defined in Section 4.2). From a methodological perspective, inference on β is challenging because the distribution $f(z,\beta)$ depends on the normalizing constant of the Potts model (hereafter denoted as $C(\beta)$), which is generally intractable. This problem has received some attention in the recent image processing literature, as it would lead to fully unsupervised algorithms [82, 93, 112–114].

In this work we focus on the estimation of β within a Markov chain Monte Carlo (MCMC) algorithm that handles 2D or 3D data sets [4, 79, 97, 98, 100, 101]. MCMC methods are powerful tools to handle Bayesian inference problems for which the minimum mean square error (MMSE) or the maximum a posteriori (MAP) estimators are difficult to derive analytically. MCMC methods generate samples that are asymptotically distributed according to the joint posterior of the unknown parameters. These samples are then used to approximate the Bayesian estimators. However, standard MCMC methods cannot be applied directly to Bayesian problems based on the Potts model. Indeed, inference on β requires computing the normalizing constant of the Potts model $C(\beta)$, which is generally

intractable. Specific MCMC algorithms have been designed to estimate Markov field parameters in [91, 115–117] and more recently in [82, 112]. A variational Bayes algorithm based on an approximation of $C(\beta)$ has also been recently proposed in [93]. Maximum likelihood estimation of β within expectation-maximization (EM) algorithms has been studied in [113, 114, 118]. The strategies involved in these works for avoiding computing the normalizing constant $C(\beta)$ are summarized below.

4.1.1 Pseudo-likelihood estimators

One possibility to avoid the computation of $C(\beta)$ is to eliminate it from the posterior distribution of interest. More precisely, one can think of defining a prior distribution $f(\beta)$ such that the normalizing constant cancels out from the posterior (i.e., $f(\beta) \propto C(\beta)\mathbf{1}_{\mathcal{R}^+}(\beta)$), resulting in the so-called *pseudo-likelihood* estimators [115, 116, 119]. Although analytically convenient this approach generally does not lead to a satisfactory posterior density and results in poor estimation [120]. Also, as noticed in [117] such a prior distribution generally depends on the data since the normalizing constant $C(\beta)$ depends implicitly on the number of observations (priors that depend on the data are not recommended in the Bayesian paradigm [121, p. 36]).

4.1.2 Approximation of $C(\beta)$

Another possibility is to approximate the normalizing constant $C(\beta)$. Existing approximations can be classified into three categories: based on analytical developments, on sampling strategies or on a combination of both. A survey of the state-of-the-art approximation methods up to 2004 has been presented in [117]. The methods considered in [117] are the mean field, the tree-structured mean field and the Bethe energy (loopy Metropolis) approximations, as well as two sampling strategies based on Langevin MCMC algorithms. More recently, exact recursive expressions have been proposed to compute $C(\beta)$ analytically [93, 122]. However, to our knowledge, these recursive methods have only been

successfully applied to small problems (i.e., for MRFs of size smaller than 40×40) with reduced spatial correlation $\beta < 0.5$.

Another sampling-based approximation consists in estimating $C(\beta)$ by Monte Carlo integration [94, Chap. 3], at the expense of very substantial computation and possibly biased estimations (bias arises from the estimation error of $C(\beta)$). Better results can be obtained by using importance or path sampling methods [123]. These methods have been applied to the estimation of β within an MCMC image processing algorithm in [91]. Although more precise than Monte Carlo integration, approximating $C(\beta)$ by importance or path sampling still requires substantial computation and is generally unfeasible for large fields. This has motivated recent works that reduce computation by combining importance sampling with analytical approximations. More precisely, approximation methods that combine importance sampling with extrapolation schemes have been proposed for the Ising model (i.e., a 2-state Potts model) in [82] and for the 3-state Potts model in [112]. However, we have found that this extrapolation technique introduces significant bias (see Section 4.5.2 for more details).

4.1.3 Auxiliary variables and perfect sampling

Recent works from computational statistics have established that it is possible to avoid computing $C(\beta)$ within a Metropolis-Hastings MCMC algorithm [94] by introducing carefully selected auxiliary random variables [92, 124]. In the work of Moller *et. al.* [92], an auxiliary vector $\boldsymbol{w}$ distributed according to the same distribution as the label vector $\boldsymbol{z}$ (i.e., $f(\boldsymbol{z}|\beta)$) is introduced. Metropolis-Hastings algorithms that do not require computing $C(\beta)$ are then proposed to sample the joint distribution $f(\beta, \boldsymbol{w}|\boldsymbol{z})$, which admits the exact desired posterior density $f(\beta|\boldsymbol{z})$ as marginal distribution [92]. Unfortunately this method suffers from a very low acceptance ratio that degrades severely as the dimension of $\boldsymbol{z}$ increases, and is therefore unsuitable for image processing applications (see Section 4.5.2 for more details). Novel auxiliary variable methods with considerably better acceptance ratios have

been proposed in [124] by using several auxiliary vectors and sequential Monte Carlo samplers [125]. These methods could be interesting for estimating the Potts coefficient β. However they will not be considered in this work because they require substantial computation and are generally too costly for image processing applications. An alternative auxiliary variable method based on a one-sample estimator of the ratio $\frac{C(\beta)}{C(\beta^*)}$ has been proposed in [126] and recently been improved by using several auxiliary vectors and sequential Monte Carlo samplers in [127] (the ratio $\frac{C(\beta)}{C(\beta^*)}$ arises in the MCMC algorithm defined in Section 4.3.3). More details on the application of [126] to the estimation of the Potts coefficient β are provided in Section 4.5.2.

4.1.4 Likelihood-free methods

Finally, it is possible to avoid computing the normalizing constant $C(\beta)$ by using likelihood-free MCMC methods [128]. These methods substitute the evaluation of intractable likelihoods within a Metropolis-Hastings algorithm by a simulation-rejection scheme. More precisely, akin to the auxiliary variable method [92], an auxiliary vector $\boldsymbol{w}$ distributed according to the likelihood $f(\boldsymbol{z}|\beta)$ is introduced. Two-step Metropolis-Hastings algorithms that do not require evaluating $f(\boldsymbol{z}|\beta)$ (nor $C(\beta)$) can then be considered to generate samples that are asymptotically distributed according to the exact posterior distribution $f(\beta|\boldsymbol{z})$ [128]. Although generally unfeasible[1], these exact methods have given rise to the *approximative Bayesian computation* (ABC) framework [129, 130], which studies likelihood-free methods to generate samples from approximate posterior densities $f_\epsilon(\beta|\boldsymbol{z}) \approx f(\beta|\boldsymbol{z})$ at a reasonable computational cost. To our knowledge these promising techniques, that are increasingly regarded as "*the most satisfactory approach to intractable likelihood problems*" [130], have not yet been applied to image processing problems.

[1] In spite of being theoretically correct, exact likelihood-free algorithms suffer from several major shortcomings that make them generally impractical (see Section 4.4 for more details).

The main contribution of this chapter is to propose an ABC MCMC algorithm for the joint estimation of the label vector $\boldsymbol{z}$, the granularity coefficient β and the other unknown parameters of a Bayesian model. The estimation of β is included within an MCMC algorithm through an ABC method particularly adapted to the Potts model and to large data sets. It is shown that the estimation of β can be easily integrated to existing MCMC algorithms where β was previously assumed known. Applications to large 2D and 3D images illustrate the performance of the proposed method.

The remainder of the chapter is organized as follows: Bayesian models considered in this work are defined in Section II. Section III describes a generic hybrid Gibbs sampler to generate samples asymptotically distributed according to the approximate posterior distribution of these Bayesian models. The estimation of β using a likelihood-free algorithm is discussed in detail in Section IV. Experiments on synthetic and real data are presented in Sections V and VI respectively. Conclusions are finally reported in Section VI.

4.2 Bayesian Model

Let $r_n \in \mathbb{R}^+$ denote the nth observation, or voxel, in a lexicographically vectorized image $\boldsymbol{r} = (r_1, \ldots, r_N)^T \in \mathbb{R}^N$. We assume that $\boldsymbol{r}$ is made up by multiple regions, characterized by their own statistics. More precisely, $\boldsymbol{r}$ is assumed to be associated with K stationary classes $\{C_1, \ldots, C_K\}$ such that the observations in the kth class are fully described by the following conditional observation model

$$r_n | z_n = k \sim f(r_n | \boldsymbol{\theta}_k) \tag{4.1}$$

where $f(r_n|\boldsymbol{\theta}_k)$ denotes a generic observation model with parameter vector $\boldsymbol{\theta}_k$ characterizing the class C_k. Finally, a label vector $\boldsymbol{z} = (z_1, \ldots, z_N)^T$ is introduced to map observations $\boldsymbol{r}$ to classes $C_1, \ldots, C_K$ (i.e., $z_n = k$ if and only if $r_n \in C_k$).

Several works have established that a Potts model can be used to enhance the fact that the probability $\mathrm{P}[z_n = k]$ of a given voxel is related to the probabilities of its neighbors.

As explained previously, the amount of spatial correlation between adjacent image pixels introduced by the Potts model is controlled by the granularity coefficient β. Existing image classification and segmentation methods have mainly studied the estimation of the class parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_K^T)^T$ and the label vector $\boldsymbol{z}$ conditionally to a known value of β. However, setting β incorrectly can degrade the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ significantly. Moreover, fixing the value of β a priori is difficult because different images can have different spatial organizations. This chapter considers the problem of estimating the unknown parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{z}$ jointly with β. This problem is formulated in a Bayesian framework which requires defining the likelihood of the observation vector $\boldsymbol{r}$ and the priors for the unknown parameters $\boldsymbol{\theta}$, $\boldsymbol{z}$ and β.

4.2.1 Likelihood

Assuming that the observations r_n are independent conditionally to the label vector $\boldsymbol{z}$, the likelihood function associated with the image $\boldsymbol{r}$ is

$$f(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}, \beta) = f(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}) = \prod_{k=1}^{K} \prod_{\{n|z_n=k\}} f(r_n|\boldsymbol{\theta}_k) \tag{4.2}$$

where $f(r_n|\boldsymbol{\theta}_k)$ is the generic probability density function associated with the observation model introduced in (4.1).

4.2.2 Parameter priors

Labels

It is natural to consider that there are some correlations between the characteristics of a given voxel and those of its neighbors. Since the seminal work of Geman [87], MRFs have become very popular to introduce spatial correlation in images. MRFs assume that the distribution of a pixel conditionally to all other pixels of the image equals the distribution of this pixel conditionally to its neighbors. Consequently, it is important to properly define

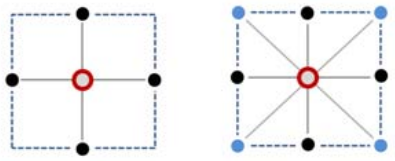


Figure 4.1: 4-pixel (left) and 8-pixel (right) neighborhood structures. The pixel considered appears as a void red circle whereas its neighbors are depicted in full black and blue.

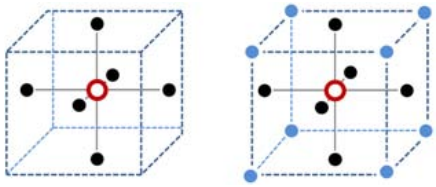


Figure 4.2: 6-voxel (left) and 14-voxel (right) neighborhood structures. The considered voxel appears as a void red circle whereas its neighbors are depicted in full black and blue.

the neighborhood structure. The neighborhood relation between two pixels (or voxels), i and j, has to be symmetric: if i is a neighbor of j then j is also a neighbor of i. There are several neighborhood structures that have been used in the literature. In the bidimensional case, neighborhoods defined by the four or eight nearest voxels represented in Fig. 4.1 are the most commonly used. Similarly, in the tridimensional case the most frequently used neighborhoods are defined by the six or fourteen nearest voxels represented in Fig 4.2. In the rest of this chapter 4-pixel and 6-voxel neighborhoods will be considered for 2D and 3D images, respectively. Therefore, the associated set of neighbors, or cliques, have vertical, horizontal and depth configurations (see [87, 88] for more details).

Once the neighborhood structure has been established, the MRF can be defined. Let z_n denote the random variable indicating the class of the nth image voxel. The whole set of random variables $z_1, z_2, \ldots, z_N$ forms a random field. An MRF is obtained when the conditional distribution of z_n given the other pixels $\boldsymbol{z}_{-n} = (z_1, \ldots, z_{n-1}, z_{n+1}, \ldots, z_N)$ only depends on its neighbors $\boldsymbol{z}_{\mathcal{V}(n)}$, i.e.,

$$f\left(z_n|\boldsymbol{z}_{-n}\right) = f\left(z_n|\boldsymbol{z}_{\mathcal{V}(n)}\right) \tag{4.3}$$

where $\mathcal{V}(n)$ is the index set of the neighbors of the nth voxel, z_{-n} denotes the vector z whose nth element has been removed and $z_{\mathcal{V}(n)}$ is the sub-vector of z composed of the elements whose indexes belong to $\mathcal{V}(n)$.

In the case of K classes, the random variables $z_1, z_2, \ldots, z_N$ take their values in the finite set $\{1, \ldots, K\}$. The resulting MRF (with discrete values) is a Potts-Markov field, which generalizes the binary Ising model to arbitrary discrete vectors. In this study 2D and 3D Potts-Markov fields will be considered as prior distributions for z. More precisely, 2D MRFs are considered for single-slice (2D) images whereas 3D MRFs are investigated for multiple-slice (3D) images. Note that Potts-Markov fields are particularly well suited for label-based segmentation as explained in [84]. By the Hammersley-Clifford theorem the corresponding prior for z can be expressed as follows

$$f(z|\beta) = \frac{1}{C(\beta)} \exp\left[\Phi_\beta(z)\right] \tag{4.4}$$

where

$$\Phi_\beta(z) = \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \beta\delta(z_n - z_{n'}) \tag{4.5}$$

and where $\delta(\cdot)$ is the Kronecker function, β is the granularity coefficient and $C(\beta)$ is the normalizing constant or partition function [89]

$$C(\beta) = \sum_{z \in \{1,\ldots,K\}^n} \exp\left[\Phi_\beta(z)\right]. \tag{4.6}$$

As explained previously, the normalizing constant $C(\beta)$ is generally intractable even for $K = 2$ because the number of summands in (4.6) grows exponentially with the size of z [99]. The hyperparameter β tunes the degree of homogeneity of each region in the image. A small value of β induces a noisy image with a large number of regions, contrary to a large value of β that leads to few and large homogeneous regions. Finally, it is interesting to note

that despite not knowing $C(\beta)$, drawing labels $\boldsymbol{z} = (z_1, \ldots, z_N)^T$ from the distribution (5.12) can be easily achieved by using a Gibbs sampler [94].

It is interesting to mention that while the Potts model is an effective means to introduce spatial correlation between discrete variables, there are other more complex models that could be investigated. In particular, Marroquin *et al.* [83] have shown that in segmentation applications better results may be obtained by using a two-layer hidden field, where hidden labels are assumed to be independent and correlation is introduced at a deeper layer by a vectorial Markov field. Similarly, Woolrich *et al.* [85] have proposed to approximate the Potts field by modeling mixture weights with a Gauss-Markov random field. However, these alternative models are not well adapted for 3D images because they require significantly more computation and memory resources than the Potts model. These overheads result from the fact that they introduce $(K + 1)N$ and KN hidden variables respectively, against only N for the Potts model (N being the number of image pixels and K the number of discrete states of the model).

Parameter vector $\boldsymbol{\theta}$

Assuming a priori independence between the parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$, the joint prior for the parameter vector $\boldsymbol{\theta}$ is

$$f(\boldsymbol{\theta}) = \prod_{k=1}^{K} f(\boldsymbol{\theta}_k) \tag{4.7}$$

where $f(\boldsymbol{\theta}_k)$ is the prior associated with the parameter vector $\boldsymbol{\theta}_k$ which mainly depends on the application considered. Two examples of priors $f(\boldsymbol{\theta})$ will be investigated in Section 4.5.2.

Granularity coefficient β

As explained previously, fixing the value of β a priori can be difficult because different images usually have different spatial organizations. A small value of β will lead to a noisy classification and degrade the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{z}$. Setting β to a too large value will

also degrade the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ by producing over-smoothed classification results. Following a Bayesian approach, this chapter proposes to assign β an appropriate prior distribution and to estimate this coefficient jointly with $(\boldsymbol{\theta}, \boldsymbol{z})$. In this work, the prior for the granularity coefficient β is a uniform distribution on $(0, B)$

$$f(\beta) = \mathcal{U}_{(0,B)}(\beta) \tag{4.8}$$

where $B = 2$ represents the maximum possible value of β. Note that it is unnecessary to consider larger values of B since, for the first order neighborhood structure, "when $\beta = 2$, the Potts-Markov model is almost surely concentrated on single-color images" [131, p. 30].

4.2.3 Posterior Distribution of $(\boldsymbol{\theta}, \boldsymbol{z}, \beta)$

Assuming the unknown parameter vectors $\boldsymbol{\theta}$, $\boldsymbol{z}$, β are a priori independent and using Bayes theorem, the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{z}, \beta)$ can be expressed as follows

$$f(\boldsymbol{\theta}, \boldsymbol{z}, \beta | \boldsymbol{r}) \propto f(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z}) f(\boldsymbol{\theta}) f(\boldsymbol{z}|\beta) f(\beta) \tag{4.9}$$

where $\propto$ means "proportional to" and where the likelihood $f(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{z})$ has been defined in (4.2) and the prior distributions $f(\boldsymbol{\theta})$, $f(\boldsymbol{z})$ and $f(\beta)$ in (4.7), (5.12) and (4.8) respectively. Unfortunately the posterior distribution (4.9) is generally too complex to derive the MMSE or MAP estimators of the unknown parameters $\boldsymbol{\theta}$, $\boldsymbol{z}$ and β. One can think of using the EM algorithm to estimate these parameters. Indeed the EM algorithm has received much attention for mixture problems [95]. However, the shortcomings of the EM algorithm include *convergence to local maxima or saddle points of the log-likelihood function and sensitivity to starting values* [96, p. 259]. An interesting alternative consists in using an MCMC method that generates samples that are asymptotically distributed according to the target distribution (4.9) [94]. The generated samples are then used to approximate the Bayesian estimators. This strategy has been used successfully in several recent image processing ap-

plications (see [97, 98, 100, 132–137] for examples in image filtering, dictionary learning, image reconstruction, fusion and segmentation). Many of these recent MCMC methods have been proposed for Bayesian models that include a Potts MRF [4, 79, 101, 132, 135]. However, these methods only studied the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ conditionally to a known granularity coefficient β. The main contribution of this chapter is to study Bayesian algorithms for the joint estimation of $\boldsymbol{\theta}, \boldsymbol{z}$ and β. The next section studies a hybrid Gibbs sampler that generates samples that are asymptotically distributed according to the posterior (4.9). The samples are then used to estimate the granularity coefficient β, the image labels $\boldsymbol{z}$ and the model parameter vector ϑ. The resulting sampler can be easily adapted to existing MCMC algorithm where β was previously assumed known, and can be applied to large images, both in 2D and in 3D.

4.3 Hybrid Gibbs Sampler

This section studies a hybrid Metropolis-within-Gibbs sampler that generates samples that are asymptotically distributed according to (4.9). The conventional Gibbs sampler successively draws samples according to the full conditional distributions associated with the distribution of interest (here the posterior (4.9)). When a conditional distribution cannot be easily sampled, one can resort to a Metropolis-Hastings (MH) move, which generates samples according to an appropriate proposal and accept or reject these generated samples with a given probability. The resulting sampler is referred to as a Metropolis-within-Gibbs sampler (see [94] for more details about MCMC methods). The sampler investigated in this section is based on the conditional distributions $\mathrm{P}[\boldsymbol{z}|\boldsymbol{\theta}, \beta, \boldsymbol{r}]$, $f(\boldsymbol{\theta}|\boldsymbol{z}, \beta, \boldsymbol{r})$ and $f(\beta|\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{r})$ that are provided in the next paragraphs (see also Algorithm 2 below).

Algorithm 2 Proposed Hybrid Gibbs Sampler

1: Input: initial $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{z}^{(0)}, \beta^{(0)}\}$, number of iterations T.
2: **for** $t = 1$ to T **do**
3: Generate $\boldsymbol{z}^{(t)} \sim \mathrm{P}[\boldsymbol{z}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{z}^{(t-1)}, \beta^{(t-1)}, \boldsymbol{r}]$ according to (4.12)
4: Generate $\boldsymbol{\theta}^{(t)} \sim f(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{z}^{(t)}, \beta^{(t-1)}, \boldsymbol{r})$ according to (4.13)
5: Generate $\beta^{(t)} \sim f(\beta|\boldsymbol{\theta}^{(t)}, \boldsymbol{z}^{(t)}, \beta^{(t-1)}, \boldsymbol{r})$ using Algorithm 4.
6: **end for**

4.3.1 Conditional probability $\mathrm{P}[\boldsymbol{z}|\boldsymbol{\theta}, \beta, \boldsymbol{r}]$

For each voxel $n \in \{1, 2, \ldots, N\}$, the class label z_n is a discrete random variable whose conditional distribution is fully characterized by the probabilities

$$\mathrm{P}\left[z_n = k|\boldsymbol{z}_{-n}, \boldsymbol{\theta}_k, r_n, \beta\right] \propto f(r_n|\boldsymbol{\theta}_k, z_n = k)\mathrm{P}\left[z_n|\boldsymbol{z}_{\mathcal{V}(n)}, \beta\right] \tag{4.10}$$

where $k = 1, \ldots, K$, and where it is recalled that $\mathcal{V}(n)$ is the index set of the neighbors of the nth voxel and K is the number of classes. These probabilities can be expressed as

$$\mathrm{P}\left[z_n = k|\boldsymbol{z}_{\mathcal{V}(n)}, \boldsymbol{\theta}_k, \beta, r_n\right] \propto \pi_{n,k} \triangleq \exp\left[\sum_{n' \in \mathcal{V}(n)} \beta\delta(k - z_{n'})\right] f(r_n|\boldsymbol{\theta}_k, z_n = k). \tag{4.11}$$

Once all the quantities $\pi_{n,k}$, $k = 1, \ldots, K$, have been computed, they are normalized to obtain the probabilities $\tilde{\pi}_{n,k} \triangleq \mathrm{P}\left[z_n = k|\boldsymbol{z}_{\mathcal{V}(n)}, \boldsymbol{\theta}_k, \beta, r_n\right]$ as follows

$$\tilde{\pi}_{n,k} = \frac{\pi_{n,k}}{\sum_{k=1}^{K} \pi_{n,k}}. \tag{4.12}$$

Note that the probabilities of the label vector $\boldsymbol{z}$ in (4.12) define an MRF. Sampling from this conditional distribution can be achieved by using a Gibbs sampler [94] that draws discrete values in the finite set $\{1, \ldots, K\}$ with probabilities (4.12). More precisely, in this work $\boldsymbol{z}$ has been sampled using a 2-color parallel chromatic Gibbs sampler that loops over $n \in \{1, 2, \ldots, N\}$ following the checkerboard sequence [102].

4.3.2 Conditional probability density function $f(\boldsymbol{\theta}|z,\beta,\boldsymbol{r})$

The conditional density $f(\boldsymbol{\theta}|z,\beta,\boldsymbol{r})$ can be expressed as follows

$$f(\boldsymbol{\theta}|z,\beta,\boldsymbol{r}) = f(\boldsymbol{\theta}|z,\boldsymbol{r}) \propto f(\boldsymbol{r}|\boldsymbol{\theta},z)f(\boldsymbol{\theta}) \tag{4.13}$$

where $f(\boldsymbol{r}|\boldsymbol{\theta},z)$ and $f(\boldsymbol{\theta})$ have been defined in (4.2) and (4.7). Generating samples distributed according to (4.13) is strongly problem dependent. Some possibilities will be discussed in Sections 4.5.2 and 4.6. Generally, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T,\ldots,\boldsymbol{\theta}_K^T)^T$ can be sampled coordinate-by-coordinate using the following Gibbs moves

$$\boldsymbol{\theta}_k \sim f(\boldsymbol{\theta}_k|\boldsymbol{r},z) \propto \prod_{\{n|z_n=k\}} f(r_n|\boldsymbol{\theta}_k)f(\boldsymbol{\theta}_k), \quad k=1,\ldots,K. \tag{4.14}$$

In cases where sampling the conditional distribution (4.14) is too difficult, an MH move can be used resulting in a Metropolis-within-Gibbs sampler [94] (Appendices C and D describe the generation of samples $\boldsymbol{\theta}_k$ for the problems studied in Sections 4.5.2 and 4.6).

4.3.3 Conditional probability density function $f(\beta|\boldsymbol{\theta},z,\boldsymbol{r})$

From Bayes rule, the conditional density $f(\beta|\boldsymbol{\theta},z,\boldsymbol{r})$ can be expressed as follows

$$f(\beta|\boldsymbol{\theta},z,\boldsymbol{r}) = f(\beta|z) \propto f(z|\beta)f(\beta) \tag{4.15}$$

where $f(z|\beta)$ and $f(\beta)$ have been defined in (5.12) and (4.8) respectively. The generation of samples according to $f(\beta|\boldsymbol{\theta},z,\boldsymbol{r})$ is not straightforward because $f(z|\beta)$ is defined up to the unknown multiplicative constant $\frac{1}{C(\beta)}$ that depends on β. One could think of sampling β by using an MH move, which requires computing the acceptance ratio

$$\text{ratio} = \min\{1,\xi\} \tag{4.16}$$

with

$$\xi = \frac{f(z|\beta^*)}{f(z|\beta^{(t-1)})} \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})} \tag{4.17}$$

where $\beta^* \sim q(\beta^*|\beta^{(t-1)})$ denotes an appropriate proposal distribution. By replacing (5.12) into (4.17), ξ can be expressed as

$$\xi = \frac{C(\beta^{(t-1)})}{C(\beta^*)} \frac{\exp\left[\Phi_{\beta^*}(z)\right]}{\exp\left[\Phi_{\beta^{(t-1)}}(z)\right]} \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})} \tag{4.18}$$

where β^* denotes the proposed value of β at iteration t and $\beta^{(t-1)}$ is the previous state of the chain. Unfortunately the ratio (4.18) is generally intractable because of the term $\frac{C(\beta^{(t-1)})}{C(\beta^*)}$. The next section presents a likelihood-free MH algorithm that samples β without requiring to evaluate $f(z|\beta)$ and $C(\beta)$.

4.4 Sampling the granularity coefficient

4.4.1 Likelihood-free Metropolis-Hastings

It has been shown in [128] that it is possible to define a valid MH algorithm for posterior distributions with intractable likelihoods by introducing a carefully selected auxiliary variable and a tractable sufficient statistic on the target density. More precisely, consider an auxiliary vector $\boldsymbol{w}$ defined in the discrete state space $\{1, \ldots, K\}^N$ of z generated according to the likelihood $f(z|\beta)$, i.e.,

$$\boldsymbol{w} \sim f(\boldsymbol{w}|\beta) \triangleq \frac{1}{C(\beta)} \exp\left[\Phi_\beta(\boldsymbol{w})\right] \tag{4.19}$$

Also, let $\eta(z)$ be a tractable sufficient statistic of z, i.e., $f(\beta|z) = f[\beta|\eta(z)]$. Then, it is possible to generate samples that are asymptotically distributed according to the exact conditional density $f(\beta|\boldsymbol{\theta}, z, \boldsymbol{r}) = f(\beta|z)$ by introducing an additional rejection step based on $\eta(z)$ into a standard MH move [128] (see Algorithm 3 below).

Algorithm 3 Exact likelihood-free MH step [128]

Input: $\{\beta^{(t-1)}, z^{(t)}\}$
Generate $\beta^* \sim q\left(\beta^* | \beta^{(t-1)}\right)$
Generate an auxiliary variable $\boldsymbol{w} \sim f(\boldsymbol{w}|\beta^*)$
if $\eta(\boldsymbol{w}) = \eta(z^{(t)})$ **then**
    Set ratio $= \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})}$
    Draw $u \sim \mathcal{U}_{(0,1)}$
    if ($u <$ ratio) **then**
        Set $\beta^{(t)} = \beta^*$
    else
        Set $\beta^{(t)} = \beta^{(t-1)}$
    end if
else
    Set $\beta^{(t)} = \beta^{(t-1)}$
end if

Note that the MH acceptance ratio in algorithm 3 is the product of the prior ratio $\frac{f(\beta^*)}{f(\beta^{(t-1)})}$ and the proposal ratio $\frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})}$. The generally intractable likelihood ratio $\frac{f(z|\beta^*)}{f(z|\beta^{(t-1)})}$ has been replaced by the simulation and rejection steps involving the discrete auxiliary vector $\boldsymbol{w}$. Despite not computing $\frac{f(z|\beta^*)}{f(z|\beta^{(t-1)})}$ explicitly, the resulting MH move still accepts candidate values β^* with the correct probability (4.16) [128].

Unfortunately exact likelihood-free MH algorithms have several shortcomings [130]. For instance, their acceptance ratio is generally very low because candidates β^* are only accepted if they lead to an auxiliary vector $\boldsymbol{w}$ that verifies $\eta(z^{(t)}) = \eta(\boldsymbol{w})$. In addition, most Bayesian models do not have known sufficient statistics. These limitations have been addressed in the ABC framework by introducing an approximate likelihood-free MH algorithm (henceforth denoted as ABC-MH) [128]. Precisely, the ABC-MH algorithm does not require the use of a sufficient statistic and is defined by a less restrictive criterion of the form $\rho\left[\eta(z^{(t)}), \eta(\boldsymbol{w})\right] < \epsilon$, where ρ is an arbitrary distance measure and ϵ is a tolerance parameter (note that this criterion can be applied to both discrete and continuous intractable distributions, contrary to algorithm 3 that can only be applied to discrete distributions). The resulting algorithm generates samples that are asymptotically distributed according to an

approximate posterior density [128]

$$f_\epsilon(\beta|z) \approx \sum_{w} f(\beta) f(w|\beta) \mathbf{1}_{[\rho[\eta(z),\eta(w)]<\epsilon]}(w) \tag{4.20}$$

whose accuracy depends on the choice of $\eta(z)$ and ϵ (if $\eta(z)$ is a sufficient statistic and $\epsilon = 0$, then (4.20) corresponds to the exact posterior density).

In addition, note that in the exact likelihood-free MH algorithm, the auxiliary vector w has to be generated using perfect sampling [138, 139]. This constitutes a major limitation, since perfect or exact sampling techniques [138, 139] are too costly for image processing applications where the dimension of z and w can exceed one million pixels. A convenient alternative is to replace perfect simulation by a few Gibbs moves with target density $f(w|\beta^*)$ as proposed in [140]. The accuracy of this second approximation depends on the number of moves and on the initial state of the sampler. An infinite number of moves would clearly lead to perfect simulation regardless of the initialization. Inspired from [141], we propose to use z as initial state to produce a good approximation with a small number of moves. A simple explanation for this choice is that for candidates β^* close to the mode of $f(\beta|z)$, the vector z has a high likelihood $f(z|\beta)$. In other terms, using z as initial state does not lead to perfect sampling but provides a good final approximation of $f(\beta|z)$ around its mode. The accuracy of this approximation can be easily improved by increasing the number of moves at the expense of computing time. However, several simulation results in Section 4.5.2 show that the resulting ABC algorithm approximates $f(\beta|z)$ correctly even for a small number of moves (i.e., one per field component).

4.4.2 Choice of $\eta(z)$, ρ and ϵ

As explained previously, ABC algorithms require defining an appropriate statistic $\eta(z)$, distance function ρ and tolerance level ϵ. The choice of $\eta(z)$ and ρ are fundamental to the success of the approximation, while the value of ϵ is generally less important [130, 142].

Fortunately the Potts MRF, being a Gibbs random field, belongs to the exponential family and has the following one-dimensional sufficient statistic [130, 140]

$$\eta(\boldsymbol{z}) \triangleq \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'}) \tag{4.21}$$

where it is recalled that $\mathcal{V}(n)$ is the index set of the neighbors of the nth voxel. Note that because (4.21) is a sufficient statistic, the approximate posterior $f_\epsilon(\beta|\boldsymbol{z})$ tends to the exact posterior $f(\beta|\boldsymbol{z})$ as $\epsilon \to 0$ [128].

The distance function ρ considered in this work is the one-dimensional Euclidean distance

$$\rho\left[\eta(\boldsymbol{z}), \eta(\boldsymbol{w})\right] = |\eta(\boldsymbol{z}) - \eta(\boldsymbol{w})| \tag{4.22}$$

which is a standard choice in ABC methods [130]. Note from (4.21) and (4.22) that the distance $\rho[\cdot,\cdot]$ between $\eta(\boldsymbol{z})$ and $\eta(\boldsymbol{w})$ reduces to the difference in the number of active cliques in $\boldsymbol{z}$ and $\boldsymbol{w}$. It is then natural to set the tolerance as a fraction of that number, i.e., $\epsilon = \nu\eta(\boldsymbol{z})$ ($\nu = \frac{1}{1000}$ will be used in our experiments). Note that the choice of ν is crucial when the prior density $f(\beta)$ is informative because increasing ν introduces estimation bias by allowing the posterior density to drift towards the prior [143]. However, in this work the choice of ν is less critical because β has been assigned a at prior distribution.

4.4.3 Proposal distribution $q(\beta^*|\beta^{(t-1)})$

Finally, the proposal distribution $q(\beta^*|\beta^{(t-1)})$ used to explore the set $(0, B)$ is chosen as a truncated normal distribution centered on the previous value of the chain with variance s_β^2

$$\beta^* \sim \mathcal{N}_{(0,B)}\left(\beta^{(t-1)}, s_\beta^2\right). \tag{4.23}$$

where the variance s_β^2 is adjusted during the *burn-in* period to ensure an acceptance ratio close to 5%, as recommended in Section 4.5.2. This proposal strategy is referred to as

random walk MH algorithm [94, p. 245]. The choice of this proposal distribution has been motivated by the fact that for medium and large problems (i.e., Markov fields larger than 50×50 pixels) the distribution $f(\beta|z)$ becomes very sharp and can be efficiently explored using a random walk.

The resulting ABC MH method is summarized in Algorithm 4 below. Note that Algorithm 4 corresponds to step 5 in Algorithm 2.

Algorithm 4 ABC likelihood-free MH step [128]

1: Input: $\{\beta^{(t-1)}, z^{(t)}, \nu, s_\beta^2\}$, number of moves M.
2: Generate $\beta^* \sim \mathcal{N}_{(0,B)}\left(\beta^{(t-1)}, s_\beta^2\right)$
3: Generate $\boldsymbol{w} \sim f(\boldsymbol{w}|\beta^*)$ through M Gibbs moves with initial state $z^{(t)}$
4: **if** $|\eta(z^{(t)}) - \eta(\boldsymbol{w})| < \nu\eta(z^{(t)})$ **then**
5: Set ratio $= \frac{f(\beta^*)}{f(\beta^{(t-1)})} \frac{q(\beta^{(t-1)}|\beta^*)}{q(\beta^*|\beta^{(t-1)})}$
6: Draw $u \sim \mathcal{U}_{(0,1)}$
7: **if** ($u <$ ratio) **then**
8: Set $\beta^{(t)} = \beta^*$
9: **else**
10: Set $\beta^{(t)} = \beta^{(t-1)}$
11: **end if**
12: **else**
13: Set $\beta^{(t)} = \beta^{(t-1)}$
14: **end if**

4.5 Experiments

4.5.1 State of the art

This section compares the performance of the proposed ABC-MH method with the *auxiliary variable* [92], the *exchange* [126] and the *extrapolation scheme* (ES) [112] algorithms recently introduced in the literature. Note that the exchange algorithm [126] is also an auxiliary-variable-type technique, whereas the ES algorithm [112] is an approximation method that combines off-line path sampling with an extrapolation scheme. The first set of experiments compares the precision of these estimation methods by considering that the la-

Table 4.1: Estimation of β

True β	Aux. var [92]	Exch. [126]	ES [112]	ABC-MH (Algo. 4)
$\beta = 0.2$	0.20 ± 0.03	0.21 ± 0.03	0.21 ± 0.02	0.20 ± 0.03
$\beta = 0.4$	0.40 ± 0.03	0.41 ± 0.03	0.35 ± 0.03	0.40 ± 0.02
$\beta = 0.6$	0.61 ± 0.03	0.60 ± 0.03	0.45 ± 0.04	0.60 ± 0.02
$\beta = 0.8$	0.80 ± 0.02	0.80 ± 0.02	0.56 ± 0.01	0.80 ± 0.02
$\beta = 1.0$	1.01 ± 0.03	1.00 ± 0.02	0.77 ± 0.05	1.00 ± 0.02
$\beta = 1.2$	1.19 ± 0.03	1.21 ± 0.03	1.21 ± 0.02	1.20 ± 0.03
$\beta = 1.4$	1.37 ± 0.06	1.41 ± 0.04	1.38 ± 0.02	1.41 ± 0.04

bel vector $\boldsymbol{z}$ is known. Precisely, Table 4.1 shows the MMSE estimates of β corresponding to 3-state Potts MRFs of size 50×50 and different granularity coefficients $\beta \in [0.2, 1.4]$. These estimates have been computed using 50 parallel chains of length $T = 10\,250$ iterations whose first 250 iterations (burn-in period) have been removed. Algorithms were initialized randomly, the auxiliary variable method [92] was implemented using the true value of β as auxiliary estimate ($\hat{\beta} = 1$) and the grid step of the ES method was set to $\Delta\beta = 0.1$ as recommended in [112]. To ease interpretation, the best result for each simulation scenario in Table 4.1 is highlighted in red. We observe that the proposed ABC-MH algorithm produced the best results, closely followed by the two auxiliary variable methods. Note that the estimations obtained with the ES algorithm show significant bias.

The second set of experiments compares the ABC-MH method with the *auxiliary variable* [92] and the *exchange* [126] algorithms based on how their acceptance ratios scale with the size of the problem. This scaling is an important characteristics of methods involving auxiliary variables. Table 4.2 shows the mean acceptance ratio of each algorithm for different field dimensions, $K = 4$ and $\beta = 1$. Again, these results have been computed using 50 parallel chains of length $T = 10\,250$ iterations whose first 250 iterations (burn-in period) have been removed. At last, the results in Table 4.2 illustrate the algorithm s relative acceptance probability. These results should not be used as absolute references, since they have been obtained for a very specific setting ($s_\beta^2 = 0.0025$, $K = 4$ and $\beta = 1$).

Table 4.2: Acceptance Ratio

	Aux. var [92]	Exch. [126]	ABC MH (Algo. 4)
10×10	0.36 ± 0.16	0.78 ± 0.03	0.04 ± 1e-4
20×20	0.29 ± 0.14	0.58 ± 0.03	0.02 ± 3e-5
30×30	0.16 ± 0.12	0.43 ± 0.02	0.03 ± 7e-4
40×40	0.08 ± 0.10	0.35 ± 0.02	0.02 ± 3e-5
50×50	0.02 ± 0.05	0.28 ± 0.01	0.03 ± 3e-5
256×256	< 1e-3	0.05 ± 0.01	0.05 ± 5e-5
1024×1024	< 1e-3	0.01 ± 3e-3	0.05 ± 6e-5

Table 4.3: MMSE estimates of E($\beta|z$)

	Aux. var [92]	Exch. [126]	ABC-MH (Algo. 4)
10×10	1.01 ± 0.23	0.96 ± 0.10	1.00 ± 6e-3
20×20	1.00 ± 0.12	0.99 ± 0.05	1.00 ± 3e-3
30×30	1.00 ± 0.10	1.00 ± 0.03	1.00 ± 1e-3
40×40	1.00 ± 0.09	1.00 ± 0.02	1.00 ± 8e-4
50×50	1.00 ± 0.05	1.00 ± 0.01	1.00 ± 4e-4
256×256	not computed	1.00 ± 4e-3	1.00 ± 2e-7
1024×1024	not computed	1.00 ± 8e-4	1.00 ± 1e-7

We observe that the acceptance ratio of the proposed ABC-MH method remains close to 5% for all field sizes, whereas the other ratios decrease rapidly with the dimension of z. The steady acceptance ratio of the ABC-MH method can be explained by the fact that the auxiliary variable $\boldsymbol{w}$ is only used in the sufficient statistic $\eta(z)$ which is a scalar. In addition, we observe that the ratio of the auxiliary variable method [92] degrades significantly faster than the ratio of the exchange algorithm [126] (this result is in agreement with the experiments reported in [126]). Finally, for completeness Table 4.3 shows the MMSE estimates β for the experiments presented in Table 4.2. Again, we observe that the proposed ABC-MH algorithm outperforms the other methods.

4.5.2 Estimation of $f(\boldsymbol{\theta}, \boldsymbol{z}, \beta|\boldsymbol{r})$

This section presents simulation results conducted on synthetic data to assess the importance of estimating the hyperparameter β from data as opposed to fixing it a priori (i.e., the advantage of estimating the posterior $p(\boldsymbol{\theta}, \boldsymbol{z}, \beta|\boldsymbol{r})$ instead of fixing β). Simulations have been performed as follows: label vectors distributed according to a Potts MRF have been generated using different granularity coefficients (in this work bidimensional fields of size 256×256 pixels have been considered). Each label vector has in turn been used to generate an observation vector following the observation model (4.1). Finally, samples distributed according to the posterior distribution of the unknown parameters $(\boldsymbol{\theta}, \boldsymbol{z}, \beta)$ have been estimated from each observation vector using Algorithm 2 coupled with Algorithm 4 (assuming the number of classes K is known). The performance of the proposed algorithm has been assessed by comparing the resulting Bayesian estimates with the true values of the parameters. This chapter presents simulation results obtained using three different mixture models.

Mixture of gamma distributions

The first experiment considers a mixture of gamma distributions. This observation model is frequently used to describe the statistics of pixels in multilook SAR images and has been extensively applied for SAR image segmentation [144]. Accordingly, the conditional observation model (4.1) is defined by a gamma distribution with parameters L and m_k [144]

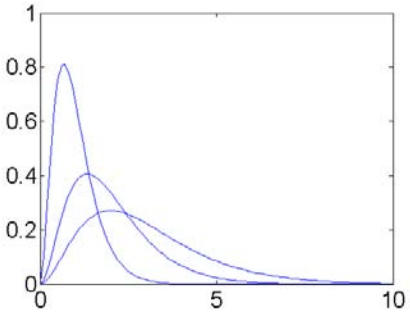
$$r_n|z_n = k \sim f(r_n|\boldsymbol{\theta}_k) = \left(\frac{L}{m_k}\right)^L \frac{r_n^{L-1}}{\Gamma(L)} \exp\left(-\frac{Lr_n}{m_k}\right) \tag{4.24}$$

where $\Gamma(t) = \int_0^{+\infty} u^{t-1}e^{-u}du$ is the standard Gamma function and L (the number of looks) is assumed to be known ($L = 3$ in this chapter). The means m_k ($k = 1, \ldots, K$) are assigned inverse gamma prior distributions as in [144]. The estimation of β, $\boldsymbol{z}$ and $\boldsymbol{\theta} = \boldsymbol{m} = (m_1, \ldots, m_K)^T$ is then achieved by using Algorithm 2. The sampling strategies described in

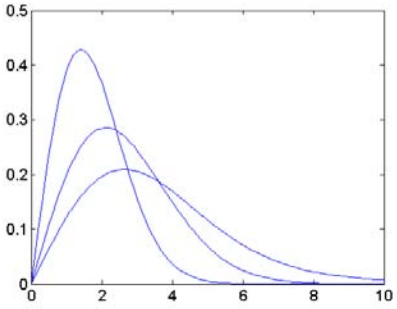
Sections 4.3.1 and 4.4 can be used for the generation of samples according to $P[\boldsymbol{z}|\boldsymbol{m},\beta,\boldsymbol{r}]$ and $f(\beta|\boldsymbol{m},\boldsymbol{z},\boldsymbol{r})$. More details about simulation according to $f(\boldsymbol{m}|\boldsymbol{z},\beta,\boldsymbol{r})$ are provided in Appendix C.

The first results have been obtained for a 3-component gamma mixture with parameters $\boldsymbol{m} = (1;2;3)$. Fig. 4.3(a) shows the densities of the gamma distributions defining the mixture model. Note that there is a significant overlap between the densities making the inference problem very challenging. For each experiment the MAP estimates of the class labels $\boldsymbol{z}$ have been computed from a single Markov chain of $T = 1\,000$ iterations whose first 400 iterations (burn-in period) have been removed. Table 4.4 shows the percentage of MAP class labels correctly estimated. The first column corresponds to labels that were estimated jointly with β whereas the other columns result from fixing β to different a priori values. To ease interpretation, the best and second best results for each simulation scenario in Table 4.4 are highlighted in red and blue. We observe that the proposed method performs as well as if β was perfectly known. On the other hand, setting β to an incorrect value may severely degrade estimation performance. Table 4.5 shows the MMSE estimates of β and $\boldsymbol{m}$ corresponding to the three simulations of the first column of Table 4.4 (proposed method) as well as the standard deviations of the estimates (results are displayed as [mean $\pm$ standard deviation]). We observe that these values are in good agreement with the true values used to generate the observation vectors. Finally, for illustration purposes, Fig. 4.4 shows the MAP estimates of the class labels corresponding to the simulation scenario reported in the last row of Table 4.4. More precisely, Fig. 4.4(a) depicts the class label map, which is a realization of a 3-class Potts MRF with $\beta = 1.2$ and size 256×256 pixels. The corresponding synthetic image is presented in Fig. 4.4(b). Fig. 4.4(c) shows the class labels obtained with the proposed method and Fig. 4.4(d) those obtained when β is perfectly known. Lastly, Figs. 4.4(e)-(h) show the results obtained when β is fixed incorrectly to 0.6, 0.8, 1.0 and 1.4. We observe that the classification produced by the proposed method is

very close to that obtained by fixing β to its true value, whereas fixing β incorrectly results in either noisy or excessively smooth results.



(a) gamma mixture



(b) α-Rayleigh mixture

Figure 4.3: Probability density functions of the distributions mixed for the first set and the second set of experiments.

Table 4.4: Gamma Mixture: Class label estimation ($K = 3$)

			Correct classification with β fixed				
	Proposed method		$\beta = 0.6$	$\beta = 0.8$	$\beta = 1.0$	$\beta = 1.2$	$\beta = 1.4$
True $\beta = 0.8$	$\hat{\beta} = 0.80$	62.2%	61.6%	61.7%	58.8%	41.5%	40.1%
True $\beta = 1.0$	$\hat{\beta} = 1.00$	77.9%	67.3%	73.4%	77.7%	75.9%	74.2%
True $\beta = 1.2$	$\hat{\beta} = 1.18$	95.6%	76.6%	87.8%	94.9%	95.6%	95.5%

Table 4.5: Gamma Mixture: Parameter estimation

	true	MMSE	true	MMSE	true	MMSE
β	0.80	0.80 ± 0.01	1.00	1.00 ± 0.01	1.20	1.18 ± 0.02
m_1	1	0.99 ± 0.02	1	1.00 ± 0.02	1	0.99 ± 0.03
m_2	2	1.99 ± 0.02	2	1.98 ± 0.02	2	1.98 ± 0.07
m_3	3	2.98 ± 0.03	3	2.98 ± 0.04	3	3.01 ± 0.03

Mixture of α-Rayleigh distributions

The second set of experiments has been conducted using a mixture of α-Rayleigh distributions. This observation model has been recently proposed to describe ultrasound images of dermis [5] and has been successfully applied to the segmentation of skin lesions in 3D

ultrasound images [4]. Accordingly, the conditional observation model (4.1) used in the experiments is defined by an α-Rayleigh distribution

$$r_n|z_n = k \sim f(r_n|\boldsymbol{\theta}_k) = p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) \triangleq r_n \int_0^\infty \lambda \exp\left[-(\gamma_k\lambda)^{\alpha_k}\right] J_0(r_n\lambda)\, d\lambda \tag{4.25}$$

where α_k and γ_k are the parameters associated with the kth class and where J_0 is the zeroth order Bessel function of the first kind. Note that this distribution has been also used to model SAR images in [42, 44]. The prior distributions assigned to the parameters α_k and γ_k ($k = 1, \ldots, K$) are uniform and inverse gamma distributions as in [4]. The estimation of β, $\boldsymbol{z}$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T = (\alpha_1, \ldots, \alpha_K, \gamma_1, \ldots, \gamma_K)^T$ is performed by using Algorithm 2. The sampling strategies described in Sections 4.3.1 and 4.4 can be used for the generation of samples according to $P[\boldsymbol{z}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta, \boldsymbol{r}]$ and $f(\beta|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{r})$. More details about simulation according to $f(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{z}, \beta, \boldsymbol{r})$ and $f(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \boldsymbol{z}, \beta, \boldsymbol{r})$ are provided in Appendix D.

The following results have been obtained for a 3-component α-Rayleigh mixture with parameters $\boldsymbol{\alpha} = (1.99; 1.99; 1.80)$ and $\boldsymbol{\gamma} = (1.0; 1.5; 2.0)$. Fig. 4.3(b) shows the densities of the components associated with this α-Rayleigh mixture. Again, note that there is significant overlap between the mixture components making the inference problem very challenging. For each experiment the MAP estimates of the class labels $\boldsymbol{z}$ have been computed from a single Markov chain of $T = 2\,000$ iterations whose first 900 iterations (burn-in period) have been removed. Table 4.6 shows the percentage of MAP class labels correctly estimated. The first column corresponds to labels that were estimated jointly with β whereas the other columns result from fixing β to different a priori values. To ease interpretation, the best and second best results for each simulation scenario in Table 4.6 are highlighted in red and blue. We observe that even if the mixture components are hard to estimate, the proposed method performs similarly to the case of a known coefficient β. Also, setting β incorrectly degrades estimation performance considerably. Table 4.7 shows the MMSE estimates of β, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ corresponding to the three simulations of the first column of Table

4.6 (proposed method). We observe that these values are in good agreement with the true values used to generate the observation vectors. To conclude, Fig. 4.5 shows the MAP estimates of the class labels corresponding to the simulation associated with the scenario reported in the last row of Table 4.6. More precisely, the actual class labels are displayed in Fig. 4.5(a), which shows a realization of a 3-class Potts MRF with $\beta = 1.2$ and size 256×256 pixels. The corresponding observation vector is presented in Fig. 4.5(b). Fig. 4.5(c) and Fig. 4.5(d) show the class labels obtained with the proposed method and with the actual value of β. Lastly, Figs. 4.5(e)-(h) show the results obtained when β is fixed incorrectly to $0.6, 0.8, 1.0$ and 1.4. We observe that the proposed method produces classification results that are very similar to those obtained when β is fixed to its true value. On the other hand, fixing β incorrectly generally leads to very poor results.

Table 4.6: α-Rayleigh Mixture: Class label estimation ($K = 3$)

			Correct classification with β fixed				
	Proposed method		$\beta = 0.6$	$\beta = 0.8$	$\beta = 1.0$	$\beta = 1.2$	$\beta = 1.4$
True $\beta = 0.8$	$\hat{\beta} = 0.82$	56.48%	52.27%	56.33%	44.80%	33.29%	33.43%
True $\beta = 1.0$	$\hat{\beta} = 1.01$	75.49%	61.08%	68.14%	75.53%	54.14%	41.68%
True $\beta = 1.2$	$\hat{\beta} = 1.18$	94.92%	67.71%	83.08%	94.37%	94.80%	69.48%

Table 4.7: α-Rayleigh Mixture: Parameter estimation

	true	MMSE	true	MMSE	true	MMSE
β	0.80	0.81 ± 0.013	1.00	1.01 ± 0.015	1.20	1.18 ± 0.021
α_1	1.99	1.98 ± 0.010	1.99	1.99 ± 0.010	1.99	1.99 ± 0.004
γ_1	1.00	1.00 ± 0.009	1.00	1.00 ± 0.009	1.00	1.00 ± 0.005
α_2	1.99	1.99 ± 0.007	1.99	1.97 ± 0.008	1.99	1.99 ± 0.005
γ_2	1.50	1.47 ± 0.012	1.50	1.49 ± 0.010	1.50	1.50 ± 0.005
α_3	1.80	1.80 ± 0.008	1.80	1.80 ± 0.006	1.80	1.79 ± 0.007
γ_3	2.00	2.02 ± 0.014	2.00	1.97 ± 0.017	2.00	2.00 ± 0.009

Mixture of Gaussian distributions

For completeness, the proposed method was also applied to a mixture of Gaussian distributions. This third set of simulations was conducted using a 4-component Gaussian mixture with means $\mu = (0; 1; 2; 4)$ and variances $\sigma^2 = (1.0; 0.5; 1.5; 0.7)$. See appendix E for details about the inference of μ_k and σ_k^2. The probability density functions of these Gaussian components are displayed in Fig. 4.6. For each experiment the MAP estimates of the class labels z have been computed from a single Markov chain of $T = 2\,000$ iterations whose first 900 iterations (burn-in period) have been removed. Table 4.8 shows the percentage of MAP class labels correctly estimated in each simulation. The first column corresponds to labels that were estimated jointly with β whereas the other columns result from fixing β to different a priori values. Again, to ease interpretation the best and second best results for each simulation scenario in Table 4.4 are highlighted in red and blue.

Table 4.8: Gaussian Mixture: Class label estimation ($K = 4$)

			Classification with β fixed				
	Proposed method		$\beta = 0.6$	$\beta = 0.8$	$\beta = 1.0$	$\beta = 1.2$	$\beta = 1.4$
True $\beta = 0.8$	$\hat{\beta} = 0.80$	67.3%	67.0%	67.3%	64.2%	42.0%	28.0%
True $\beta = 1.0$	$\hat{\beta} = 1.00$	73.7%	69.0%	71.9%	73.6%	68.1%	50.3%
True $\beta = 1.2$	$\hat{\beta} = 1.21$	94.8%	67.7%	83.1%	94.4%	94.8%	69.5%

We observe that the proposed method performs similarly to the case of a known coefficient β. Also, setting β incorrectly degrades estimation performance considerably. Table 4.9 shows the MMSE estimates for β and $\boldsymbol{\theta}$ corresponding to the three simulations of the first column of table 4.8. We observe that these values are in good agreement with the true values used to generate the observation vectors.

Furthermore, for illustration Fig. 4.7 shows the MAP estimates of the class labels corresponding to the last row of table 4.8. More precisely, Fig. 4.7(a) depicts the class label vector, which is a realization of a 3-class Potts MRF with $\beta = 1.2$ and size 256×256 pixels. The corresponding observation vector is presented in Fig. 4.7(b). Fig. 4.7(c) shows

Table 4.9: Gaussian Mixture: Parameter estimation

	true	MMSE	true	MMSE	true	MMSE
β	0.80	0.80 ± 0.01	1.00	1.008 ± 0.004	1.20	1.208 ± 0.004
μ_1	0	0.01 ± 0.01	0	0.00 ± 0.01	0	0.02 ± 0.01
σ_1^2	1	1.01 ± 0.02	1	0.99 ± 0.02	1	1.01 ± 0.01
μ_2	1	1.03 ± 0.01	1	1.01 ± 0.01	1	1.00 ± 0.01
σ_2^2	0.5	0.50 ± 0.01	0.5	0.51 ± 0.01	0.5	0.51 ± 0.01
μ_3	2	1.97 ± 0.02	2	2.00 ± 0.02	2	2.00 ± 0.01
σ_3^2	1.5	1.61 ± 0.06	1.5	1.57 ± 0.03	1.5	1.53 ± 0.02
μ_4	4	3.99 ± 0.01	4	3.99 ± 0.01	4	4.00 ± 0.01
σ_4^2	0.7	0.72 ± 0.02	0.7	0.71 ± 0.01	0.7	0.70 ± 0.01

the class labels obtained with the proposed method and Fig. 4.7(d) those obtained when β is perfectly known. Fig.s 4.7(e)-(h) show the results obtained when β is fixed incorrectly at $0.6, 0.8, 1.0$ and 1.4. We observe that the classification produced by the proposed method is very close to that obtained by fixing β to its true value, whereas the fixing β incorrectly results in either noisy or excessively smooth results.

4.6 Application to real data

After validating the proposed Gibbs sampler on synthetic data, this section presents two applications of the proposed algorithm to real data.

4.6.1 Pixel classification of a 2D SAR image

The proposed method has been applied to the unsupervised classification of a 2D multi-look SAR image acquired over Toulouse, France, depicted in Fig. 4.8(a). This image was acquired by the TerraSAR-X satellite at 1m resolution and results from summing 3 independent SAR images (i.e., $L = 3$). Potts MRFs have been extensively applied to SAR image segmentation using different observations models [118, 145–148]. For simplicity

the observation model chosen in this work is a mixture of gamma distributions (see Section 4.5.2 and Appendix C for more details about the gamma mixture model). The proposed experiments were conducted with a number of classes $K = 4$ (setting $K > 4$ resulted in empty classes). Fig. 4.8(b) shows the results obtained with the proposed method. The MMSE estimate of the granularity coefficient corresponding to this result is $\hat{\beta} = 1.62 \pm 0.05$, which has enforced the appropriate amount of spatial correlation to handle noise and outliers while preserving contours. Fig. 4.8(c) shows the results obtained by fixing $\beta = 1$, as proposed in [146]. These results have been computed from a single Markov chain of $T = 5\,000$ iterations whose first 1 000 iterations (burn-in period) have been removed. Finally, for visual interpretation Fig. 4.8(d) shows the same region observed by an airborne optical sensor. We observe that the classification obtained with the proposed method has clear boundaries and few miss-classifications.

4.6.2 Lesion segmentation in a 3D ultrasound image

The proposed method has also been applied to the segmentation of a skin lesion in a dermatological 3D ultrasound image. Ultrasound-based lesion inspection is an active topic in dermatological oncology, where patient treatment depends mainly on the depth of the lesion and the number of skin layers it has invaded. This problem has been recently addressed using an α-Rayleigh mixture model (D.1) coupled with a tridimensional Potts MRF as prior distribution for the class labels [4]. The algorithm investigated in [4] estimates the label vector and the mixture parameters conditionally to a known value of β that is set heuristically by cross-validation. The proposed method completes this approach by including the estimation of β into the segmentation problem. Some elements of this model are recalled in Appendix D for completeness.

Fig. 4.9(a) shows a 3D B-mode ultrasound image of a skin lesion, acquired at 100MHz with a focalized 25MHz 3D probe (the lesion is contained within the ROI outlined by the red rectangle). Fig. 4.9(b) presents one slice of the 3D MAP label vector obtained with the

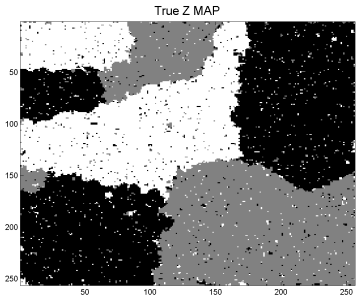
proposed method. The MMSE estimate of the granularity coefficient corresponding to this result is $\hat{\beta} = 1.020 \pm 0.007$. To assess the in uence of β, Figs. 4.9(c)-(g) show the MAP class labels obtained with the algorithm proposed in [4] for different values of β. These results have been computed using $K = 4$ since the region of interest (ROI) contains 3 types of healthy skin layers (epidermis, papillary dermis and reticular dermis) in addition to the lesion. Labels have been computed from a single Markov chain of $T = 12\,000$ iterations whose first 2 000 iterations (burn-in period) have been removed.

We observe that the proposed method produces a very clear segmentation that not only sharply locates the lesion but also provides realistic boundaries for the healthy skin layers within the ROI. This result indicates that the lesion, which is known to have originated at the dermis-epidermis junction, has already invaded the upper half of the papillary dermis. We also observe that the results obtained by fixing β to a small value are corrupted by ultrasound speckle noise and fail to capture the different skin layers. On the other hand, choosing a too large value of β enforces excessive spatial correlation and yields a segmentation with artificially smooth boundaries. Finally, Fig. 4.10 shows a frontal viewpoint of a 3D reconstruction of the lesion surface. We observe that the tumor has a semi-ellipsoidal shape which is cut at the upper left by the epidermis-dermis junction. The tumor grows from this junction towards the deeper dermis, which is at the lower right.

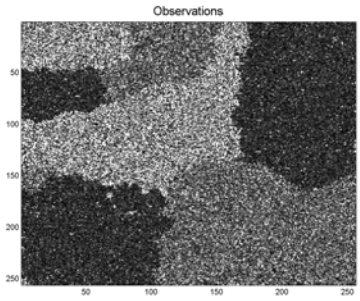
4.7 Concluding Remarks

This chapter presented a hybrid Gibbs sampler for estimating the Potts parameter β jointly with the unknown parameters of a Bayesian model. In most image processing applications this important parameter is set heuristically by cross-validation. Standard MCMC methods cannot be applied to this problem because performing inference on β requires computing the intractable normalizing constant of the Potts model. In this work the estimation of β has been included within an MCMC method using an ABC likelihood-free Metropolis-

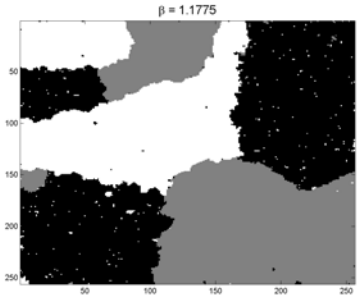
Hastings algorithm, in which intractable terms have been replaced by simulation-rejection schemes. The ABC distance function has been defined using the Potts potential, which is the natural sufficient statistic of the Potts model. The proposed method can be applied to large images both in 2D and in 3D scenarios. Experimental results obtained for synthetic data showed that estimating β jointly with the other unknown parameters leads to estimation results that are as good as those obtained with the actual value of β. On the other hand, choosing an incorrect value of β can degrade the estimation performance significantly. Finally, the proposed algorithm was successfully applied to real bidimensional SAR and tridimensional ultrasound images. This study assumed that the number of classes K is known. Future works could relax this assumption by studying the estimation of β within a reversible jump MCMC algorithm [149, 150] or by considering model choice ABC methods [140]. Other perspectives for future work include the estimation of the total variation regularization parameter in image restoration problems [151] and the estimation of texture descriptors defined through Markov fields [107].



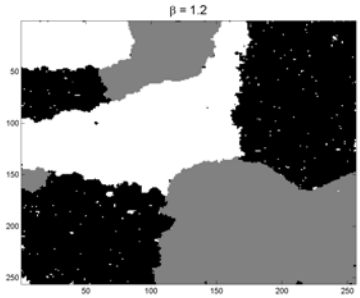
(a) True Labels ($\beta = 1.2$)



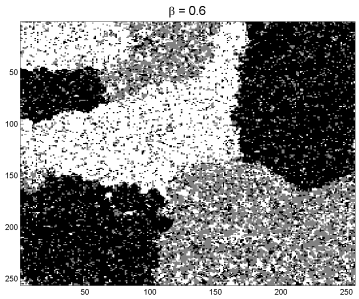
(b) Observations



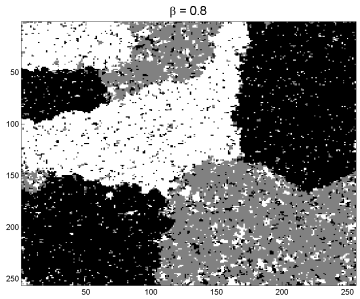
(c) Estimated β



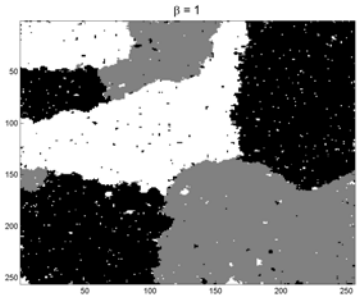
(d) True $\beta = 1.2$



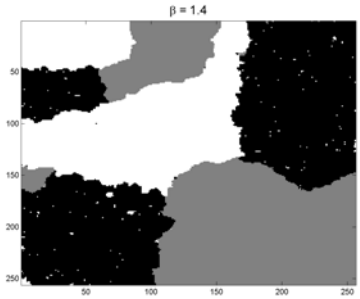
(e) $\beta = 0.6$



(f) $\beta = 0.8$

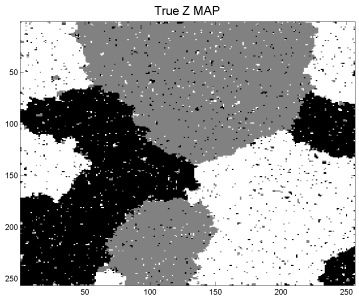


(g) $\beta = 1.0$

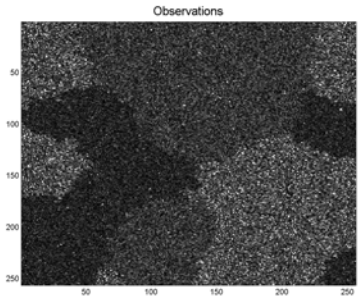


(h) $\beta = 1.4$

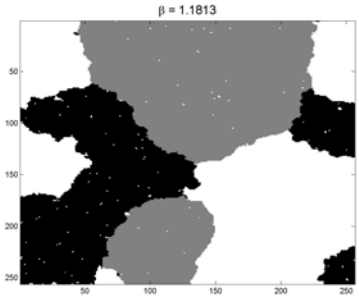
Figure 4.4: Gamma mixture: Estimated labels using the MAP estimators. (a) Ground truth, (b) observations, (c) proposed algorithm (estimated β),(d) true $\beta = 1.2$, (e)-(h) fixed $\beta = (0.6, 0.8, 1.0, 1.2, 1.4)$.



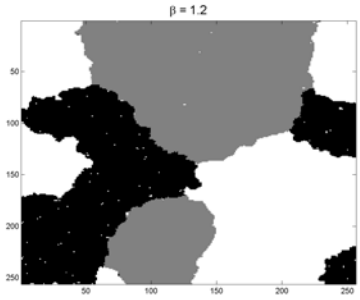
(a) True Labels ($\beta = 1.2$)



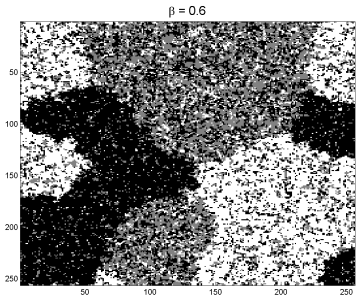
(b) Observations



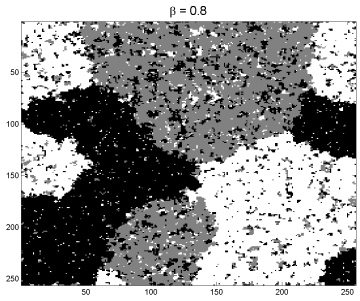
(c) Estimated β



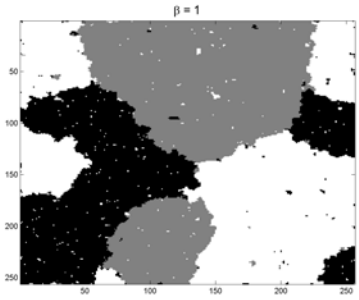
(d) True $\beta = 1.2$



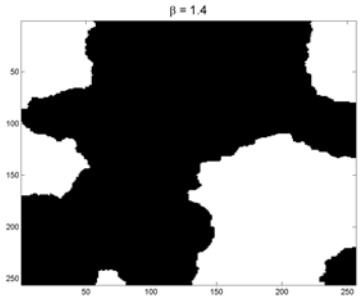
(e) $\beta = 0.6$



(f) $\beta = 0.8$



(g) $\beta = 1.0$



(h) $\beta = 1.4$

Figure 4.5: α-Rayleigh mixture: MAP estimates of the class labels. (a) Ground truth, (b) observations, (c) proposed algorithm (estimated β),(d) true $\beta = 1.2$, (e)-(h) fixed $\beta = (0.6, 0.8, 1.0, 1.2, 1.4)$.

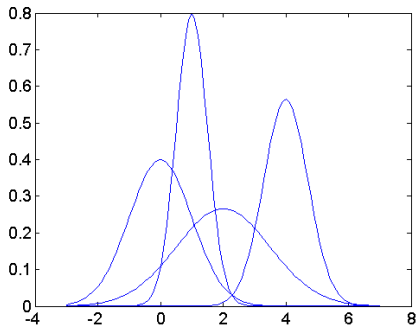
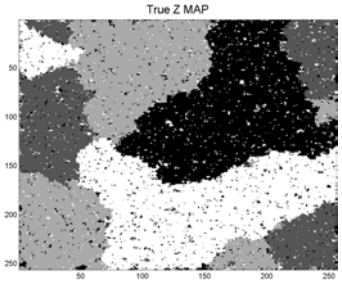
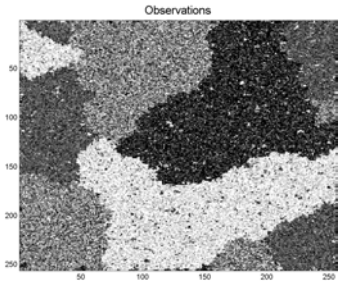


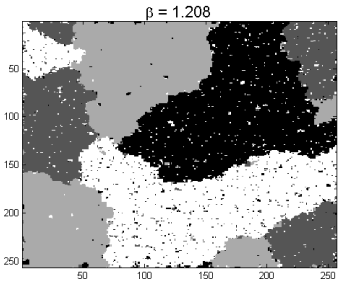
Figure 4.6: Probability density functions of Gaussian mixture components.



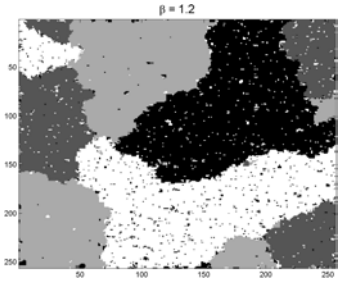
(a) True Labels ($\beta = 1.2$)



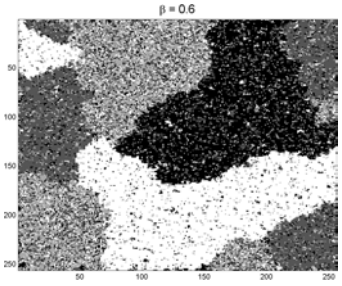
(b) Observations



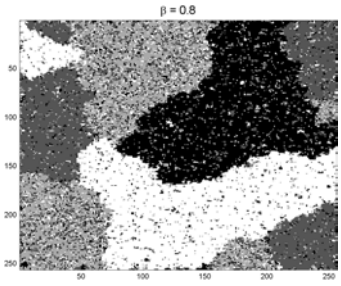
(c) Estimated β



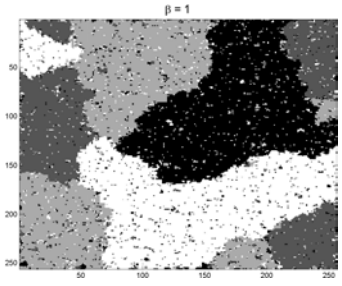
(d) True $\beta = 1.2$



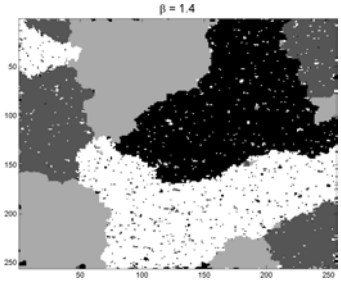
(e) $\beta = 0.6$



(f) $\beta = 0.8$

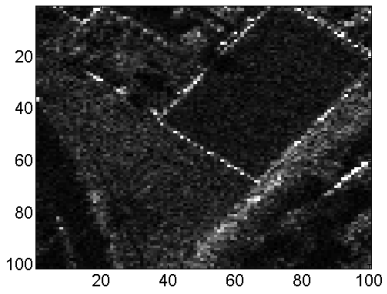


(g) $\beta = 1.0$

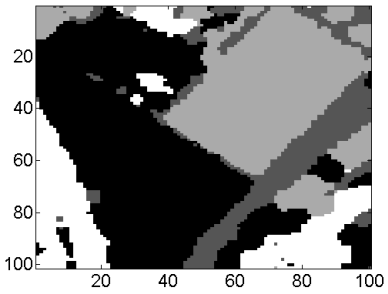


(h) $\beta = 1.4$

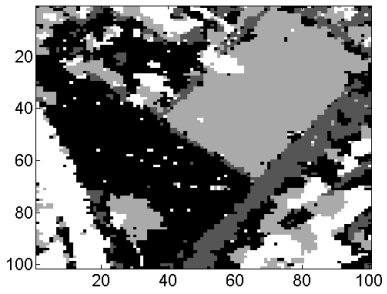
Figure 4.7: Gaussian mixture: Estimated labels using the MAP estimators. (a) Ground truth, (b) observations, (c) proposed algorithm (estimated β),(d) true $\beta = 1.2$, (e)-(h) fixed $\beta = (0.6, 0.8, 1.0, 1.2, 1.4)$.



(a) Multilook SAR Image



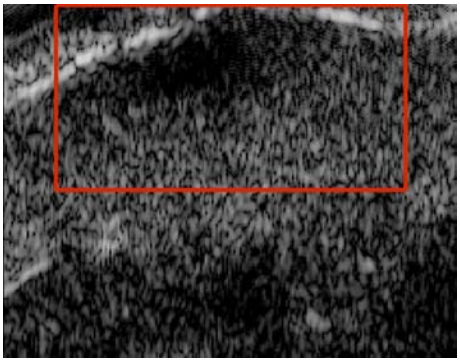
(b) Labels ($\hat{\beta}$ = 1.62)



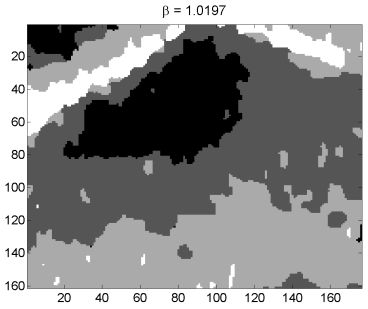
(c) Labels (β=1)

(d) Optical Image of Toulouse

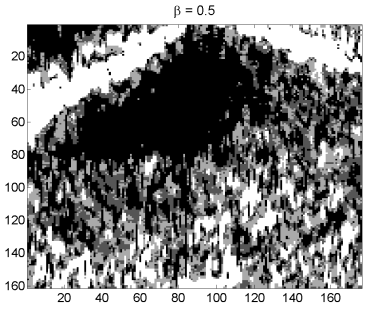
Figure 4.8: Pixel classification in a multilook SAR image (c). MAP labels when β is estimated (d) and β = 1 (e). Figs. (a)-(b) provide optical images of the same region.



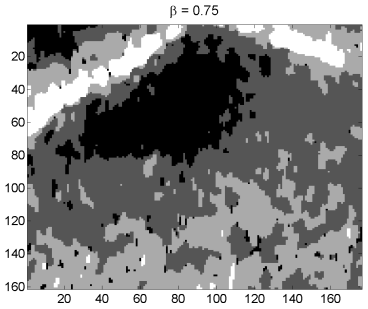
(a) Dermis view with skin lesion (ROI = $160 \times 175 \times 16$).



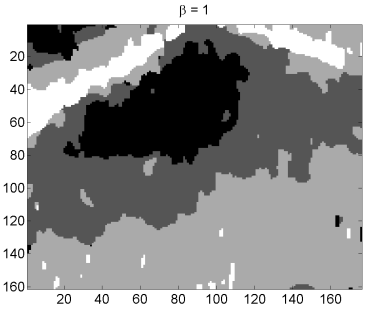
(b) (Estimated $\hat{\beta}$)



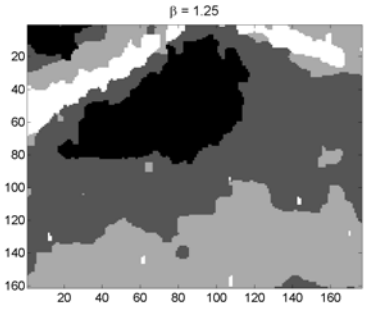
(c) ($\beta = 0.5$)



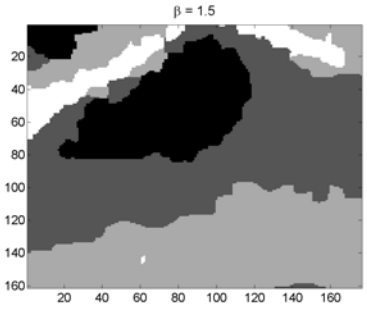
(d) ($\beta = 0.75$)



(e) ($\beta = 1.0$)



(f) ($\beta = 1.25$)



(g) ($\beta = 1.5$)

Figure 4.9: Log-compressed US images of skin lesion and the corresponding estimated class labels (*lesion* = black, *epidermis* = white, *pap. dermis* = dark gray, *ret. dermis* = light gray). MAP estimates of the class labels. Fig. (b) shows the results obtained r with the proposed method. Figs. (c)-(g) show the results obtained with the algorithm [4] for $\beta = (0.5, 0.75, 1, 1.25, 1.5)$.

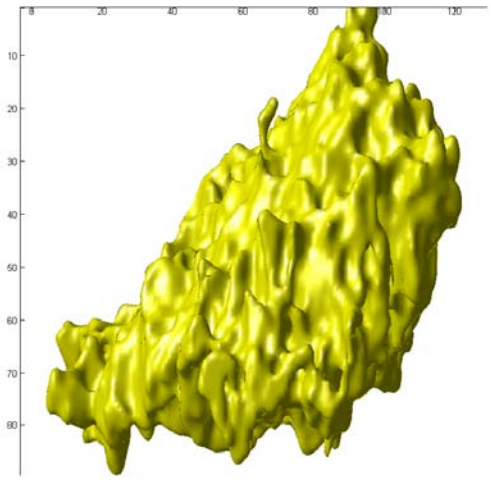


Figure 4.10: Frontal viewpoint of a 3D reconstruction of the skin lesion.

Calcul de la borne de Cramér-Rao d'un champ aléatoire d' Ising-Markov ou de Potts-Markov

L estimation de paramètres dans les modèles statistiques ayant une vraisemblance non calculable (dits intractables) est un problème difficile, ayant reçu une grande attention dans les communautés statistiques et traitement de signal [92, 124, 126]. En particulier, l estimation des paramètres d un champ aléatoire de Markov est un sujet de recherche actif en traitement d images [10, 91, 99, 152]. Plusieurs nouveaux estimateurs basés principalement sur des méthodes Monte Carlo ou des approximations variationnelles ont été proposés récemment [92, 93, 124, 126, 127]. On pourrait estimer une distribution approchée à l aide des méthodes référencées plus haut, puis calculer la borne numériquement. Cependant, cela aboutirait à des approximations faibles à cause des opérations numériques mal-conditionnées dans le calcul de la borne.

Ce chapitre est consacré à la dérivation de la borne de Cramer-Rao (CR) [153] pour les estimateurs de paramètres de MRF. La borne CR établit la limite inférieure de la quantité d information qu un ensemble de données apporterait sur des paramètres inconnus. Spécifiquement, elle définit la variance minimale de tout estimateur non biaisé de ce paramètre en termes d erreur aux moindres carrés. Malheureusement, la borne CR de la plupart des champs aléatoires de Markov est difficile à évaluer car leurs vraisemblances sont non calculables [107, Ch. 7].

On propose d exploiter une propriété intéressante de la famille des distributions exponentielles pour calculer directement la borne CR numériquement. Précisément, on montre qu il est possible d exprimer la borne en termes de moments statistiques de second ordre. Un algorithme Monte Carlo original est proposé pour estimer ces moments et calculer la borne. La méthode proposée est illustrée sur les modèles d Ising et de Potts, largement utilisés dans le traitement d images pour représenter la cohérence spatiale entre pixels dans des problèmes de segmentation [4, 79, 101].

Le chapitre est organisé comme suit : La famille de distributions considérée dans ce travail est présentée dans la section 5.2. . La section 5.3.1 montre que pour ces distributions, la borne de Cramer-Rao peut être exprimée comme une mesure de covariance. L algorithme Monte Carlo pour estimer cette covariance est présenté dans la section 5.3.2. L application de la méthode proposée aux modèles d Ising et de Potts est présentée dans la section 5.4. Des conclusions sont finalement reportées en section 5.5.

Chapter 5

Computing the Cramer-Rao bound of the granularity coefficient of an Ising or a Potts Markov random field

This chapter addresses the problem of computing the Cramer-Rao bound for the estimators of Markov random field parameters. This bound depends on the derivatives of a likelihood that are generally intractable. It is established that by exploiting a property of the exponential family, this intractable bound can be related to the statistical moments of the Gibbs potential of the Markov random field. A derivative-free Monte Carlo algorithm is then proposed to estimate the moments and compute the bound. To illustrate the interest of this method, the proposed algorithm is successfully applied to the Ising and the Potts models. The resulting bounds are used to assess the statistical efficiency of three state-of-the art estimators of the parameter of these Markov random fields.

5.1 Introduction

The estimation of the parameters involved in intractable statistical models (i.e., with intractable likelihood distributions) is a difficult problem that has received significant attention in the computational statistics and signal processing literature [92, 124, 126]. Particularly, the estimation of the parameters of a Markov random field (MRF) is an active research topic in image processing [10, 91, 99, 152]. Several new estimators have been recently derived, mainly based on efficient Monte Carlo or variational approximations [92, 93, 124, 126, 127]. On the other hand, to the best of our knowledge, little work has been conducted on estimation theory for these parameters.

This chapter focuses on the problem of deriving the Cramer-Rao bound (CRB) [153] for estimators of MRF parameters. The CRB establishes a theoretical lower limit on how much information a set of observation carries about an unknown parameter. Specifically, it defines the minimum variance for any unbiased estimator of this parameter. Moreover, the CRB is practically used as a means to characterize the performance of unbiased estimators in terms of mean square error (i.e., estimation variance). Unfortunately, the CRB for most MRF models is difficult to compute because their probability density function (pdf) is intractable [107, Ch. 7]. One could think of exploiting the estimation techniques referenced above to first compute an approximate pdf and then evaluate the CRB numerically. However, this would lead to a poor approximation because the numerical operations involved in the computation of the CRB are ill-conditioned.

Conversely, this chapter proposes to exploit an interesting property of the exponential family to compute the CRB directly. Precisely, it is shown that it is possible to express the CRB in terms of second order statistical moments. An original Monte Carlo algorithm is then proposed to estimate these moments and compute the CRB. The proposed CRB estimation method is illustrated on specific MRF models that have been widely used in the image processing community, namely the Ising and Potts models. Indeed, these models have been successfully used to capture spatial correlations between neighboring pixels

in several segmentation and/or classification problems [4, 79, 101]. The remainder of the chapter is organized as follows: The class of distributions considered in this work is introduced in Section 5.2. Section 5.3.1 shows that for these distributions the CRB can be expressed as a covariance measure. The Monte Carlo algorithm to estimate this covariance is presented in Section 5.3.2. The application of the proposed method to the Ising and the Potts models is presented in section 5.4. Conclusions are finally reported in Section 5.5.

5.2 Problem Statement

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ be an unknown parameter vector and $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$ an observation vector whose elements take their values in a set Ω. This chapter considers the case where $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are related by the following generic pdf

$$f(\boldsymbol{z}|\boldsymbol{\theta}) \triangleq \frac{1}{C(\boldsymbol{\theta})} \exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right] \tag{5.1}$$

where $\Phi(\boldsymbol{z}) : \Omega^N \rightarrow \mathcal{R}^M$ is a sufficient statistic and $C(\boldsymbol{\theta})$ is the normalizing constant given by

$$C(\boldsymbol{\theta}) = \int_{\Omega^N} \exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right] d\boldsymbol{z}. \tag{5.2}$$

Note that the model (5.1) defines a subclass of the exponential family. It comprises several standard distributions, such as Gaussian or Poisson distributions, as well as multivariate distributions frequently used in signal and image processing applications, such as Markov random fields [107]. In this latter case, the normalizing constant, also called *partition function* [89], is generally intractable.

The Cramer-Rao bound of $\boldsymbol{\theta}$ establishes a lower bound on the covariance matrix of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ [153]. Because the existence of the bound requires that $f(\boldsymbol{z}|\boldsymbol{\theta})$ verifies some weak regularity conditions, we will assume that $C(\boldsymbol{\theta})$ is twice continuously differentiable (i.e., $C(\boldsymbol{\theta}) \in C^2$). Then the CRB is equal to the inverse of the Fisher informa-

tion matrix (FIM) of $\boldsymbol{\theta}$ [153], i.e.,

$$\mathrm{cov}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \geq [\mathcal{I}(\boldsymbol{\theta})]^{-1} \tag{5.3}$$

where $\mathcal{I}(\boldsymbol{\theta})$ is an $M \times M$ positive definite matrix whose element (i, j) is given by [153]

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}) \triangleq -E_z\left[\frac{\partial^2}{\partial\theta_i\theta_j}\log[f(\boldsymbol{z}|\boldsymbol{\theta})]\right] \tag{5.4}$$

and where E_z denotes the expectation operator with respect to $\boldsymbol{z}$. By applying the definition (5.4) to (5.1) it can be shown that

$$\begin{aligned}\mathcal{I}_{i,j}(\boldsymbol{\theta}) &= E_z\left[\frac{\partial^2}{\partial\theta_i\theta_j}\log[C(\boldsymbol{\theta})] - \frac{\partial^2}{\partial\theta_i\theta_j}\boldsymbol{\theta}\Phi(\boldsymbol{z})\right] \\ &= \frac{\partial^2}{\partial\theta_i\theta_j}\log[C(\boldsymbol{\theta})]\end{aligned} \tag{5.5}$$

Unfortunately, evaluating (5.5) exactly for MRF models is rarely possible because of the intractability of $C(\boldsymbol{\theta})$. One could think of addressing this difficulty by approximating $\frac{\partial^2}{\partial\theta_i\theta_j}\log[C(\boldsymbol{\theta})]$ by: i) first using a Monte Carlo scheme to estimate $\log[C(\boldsymbol{\theta})]$ over a grid of points akin to [99], then ii) applying a crude numerical differentiation method. However, numerical differentiation is ill-conditioned (i.e., amplifies estimation errors), it may be difficult to calibrate and it introduces a discretization error that is hard to evaluate.

The next section proposes an alternative derivation of the Fisher information matrix that avoids any differentiation of $\log[C(\boldsymbol{\theta})]$. Then a Monte Carlo algorithm is proposed to estimate the FIM (5.5) directly, without estimating $C(\boldsymbol{\theta})$ or computing derivatives.

5.3 Computing the Fisher Information Matrix

This section proposes a derivative-free Monte Carlo algorithm to estimate $\mathcal{I}(\boldsymbol{\theta})$. The proposed method is based on the fact that the distributions with pdf in (5.1) belong to the exponential family.

5.3.1 Derivative-free estimation of $\mathcal{I}(\boldsymbol{\theta})$

For distributions in the exponential family, the derivatives of the log-partition function can be directly related to the statistical moments of their sufficient statistic.

Proposition 1. *The Fisher information matrix of distributions with pdf in* (5.1) *is equal to the covariance matrix of their sufficient statistic, i.e.,* $\mathcal{I}(\boldsymbol{\theta}) = cov\left[\Phi(\boldsymbol{z})\right]$.

Proof. Let $\Phi_i(\boldsymbol{z}) : \Omega^N \rightarrow \mathcal{R}$ denote the ith component of the vector $\Phi(\boldsymbol{z}) = [\Phi_1(\boldsymbol{z}), \ldots, \Phi_M(\boldsymbol{z})]^T$. Then, the first and second statistical moments of $\Phi_i(\boldsymbol{z})$ are given by

$$E_z\left[\Phi_i(\boldsymbol{z})\right] = \int_{\Omega_K^N} \Phi_i(\boldsymbol{z}) \frac{\exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right]}{C(\boldsymbol{\theta})} d\boldsymbol{z} \tag{5.6}$$

and

$$E_z\left[\Phi_i(\boldsymbol{z})\Phi_j(\boldsymbol{z})\right] = \int_{\Omega_K^N} \Phi_i(\boldsymbol{z})\Phi_j(\boldsymbol{z}) \frac{\exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right]}{C(\boldsymbol{\theta})} d\boldsymbol{z}. \tag{5.7}$$

By developing (5.5)

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}) = \frac{C^{(i,j)}(\boldsymbol{\theta})}{C(\boldsymbol{\theta})} - \frac{C^{(i)}(\boldsymbol{\theta})C^{(j)}(\boldsymbol{\theta})}{C^2(\boldsymbol{\theta})} \tag{5.8}$$

where $C(\boldsymbol{\theta})$ has been defined in (B.2) and where $C^{(i)}(\boldsymbol{\theta})$ and $C^{(i,j)}(\boldsymbol{\theta})$ are given by

$$\begin{aligned} C^{(i)}(\boldsymbol{\theta}) \triangleq \frac{\partial}{\partial \theta_i} C(\boldsymbol{\theta}) &= \int_{\Omega_K^N} \Phi_i(\boldsymbol{z}) \exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right] d\boldsymbol{z} \\ &= C(\boldsymbol{\theta}) E_z\left[\Phi_i(\boldsymbol{z})\right] \end{aligned} \tag{5.9}$$

$$
\begin{aligned}
C^{(i,j)}(\boldsymbol{\theta}) \triangleq \frac{\partial^2}{\partial\theta_i \partial\theta_j} C(\boldsymbol{\theta}) &= \int_{\Omega_K^N} \Phi_i(\boldsymbol{z})\Phi_j(\boldsymbol{z}) \exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right] d\boldsymbol{z} \\
&= C(\boldsymbol{\theta}) E_{\boldsymbol{z}}\left[\Phi_i(\boldsymbol{z})\Phi_j(\boldsymbol{z})\right].
\end{aligned} \tag{5.10}
$$

Finally, by substituting (5.9) and (5.10) into (5.8)

$$
\begin{aligned}
\mathcal{I}_{i,j}(\boldsymbol{\theta}) =& E_{\boldsymbol{z}}\left[\Phi_i(\boldsymbol{z})\Phi_j(\boldsymbol{z})\right] - E_{\boldsymbol{z}}\left[\Phi_i(\boldsymbol{z})\right] E_{\boldsymbol{z}}\left[\Phi_i(\boldsymbol{z})\right] \\
=& \operatorname{cov}\left[\Phi_i(\boldsymbol{z}), \Phi_j(\boldsymbol{z})\right].
\end{aligned} \tag{5.11}
$$

□

Note that this alternative expression for $\mathcal{I}_{i,j}$ does not involve derivatives. Unfortunately (5.11) is still generally intractable and cannot be evaluated exactly. However, unlike (5.5), the moments $\operatorname{cov}\left[\Phi_i(\boldsymbol{z}), \Phi_j(\boldsymbol{z})\right]$ in (5.11) can be efficiently estimated by Monte Carlo integration [94, Chap. 3].

5.3.2 Proposed Monte Carlo method

Based on Proposition 1, we present a Monte Carlo method to estimate $\operatorname{cov}\left[\Phi_i(\boldsymbol{z}), \Phi_j(\boldsymbol{z})\right]$ in (5.11). Precisely, we present a Monte Carlo algorithm to generate samples that are distributed according to $f(\Phi(\boldsymbol{z})|\beta)$. The generated samples are then used to approximate the statistical moments $\operatorname{cov}\left[\Phi_i(\boldsymbol{z}), \Phi_j(\boldsymbol{z})\right]$.

Algorithm 5 Proposed Monte Carlo algorithm

1: Input: $\boldsymbol{\theta}$, number of Monte Carlo samples P.
2: **for** $n = 1$ to N **do**
3: Generate $\boldsymbol{z}^{(n)} \sim f(\boldsymbol{z}|\boldsymbol{\theta}) = \frac{1}{C(\boldsymbol{\theta})} \exp\left[\boldsymbol{\theta}\Phi(\boldsymbol{z})\right]$
4: Set $\boldsymbol{\phi}^{(n)} = (\Phi_1(\boldsymbol{z}^{(n)}), \ldots, \Phi_M(\boldsymbol{z}^{(n)}))^T$
5: **end for**

Once the samples have been generated, the elements of the FIM can be approximated using the sample covariance $\hat{\mathcal{I}}_{i,j}(\boldsymbol{\theta}) = \frac{1}{N-1} \sum_{n=1}^{N} \left(\phi_i^{(n)} - \overline{\phi}_i\right)\left(\phi_j^{(n)} - \overline{\phi}_j\right)$. Finally, note that step 3 of Algo. 5 requires the simulation of variables exactly distributed according to $f(\boldsymbol{z}|\boldsymbol{\theta})$. For some distributions in (5.1) this might not be possible. In that case, one can resort to

a Markov chain Monte Carlo method [94] to generate samples approximately distributed according to $f(\boldsymbol{z}|\boldsymbol{\theta})$. For instance, most MRFs can be efficiently sampled using a Gibbs sampler [94], which does not require knowing $C(\boldsymbol{\theta})$.

5.4 Application to the Ising and Potts MRF

5.4.1 Ising and Potts models

This section presents experimental results conducted to assess the performance of the proposed method. The CRB of two important intractable models have been computed, the Ising and the Potts MRF. In these experiments the simulation of variables $\boldsymbol{z} \sim f(\boldsymbol{z}|\boldsymbol{\theta})$ has been approximated by using a Gibbs sampler [94], which does not require knowing $C(\boldsymbol{\theta})$. Therefore the proposed method has the advantage of avoiding the need of estimating $C(\boldsymbol{\theta})$ and of computing derivatives numerically. Finally, to simplify their visual interpretation, results are displayed using a logarithmic scale. For completeness the Ising and the Potts models are recalled below.

Let $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$ be a discrete random vector whose elements take their values in the finite set $\Omega_K = \{1, \ldots, K\}$. The Ising and the Potts MRF are defined as follows

$$f(\boldsymbol{z}|\beta) \triangleq \frac{1}{C(\beta)} \exp\left[\beta \Phi(\boldsymbol{z})\right] \tag{5.12}$$

with

$$\Phi(\boldsymbol{z}) = \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'}) \tag{5.13}$$

where $\mathcal{V}(n)$ is the index set of the neighbors of the nth element, $\delta(\cdot)$ is the Kronecker function and $\beta \in \mathcal{R}^+$ is the granularity coefficient or *inverse temperature* parameter. The Gibbs distribution (5.12) corresponds to the Ising MRF when $K = 2$, and to the Potts MRF for $K \geq 3$. In our experiments $\mathcal{V}(n)$ will be considered to be a bidimensional first-

order (i.e., 4-pixel) neighborhood structure. However, the proposed method is valid for any correct neighborhood structure (see [107] for more details).

5.4.2 Validation of the proposed method

To validate the proposed Monte Carlo method under controlled ground truth conditions (i.e., known CRB), the algorithm described in Section 5.3.2 was first applied to an Ising model defined on a *toroidal* graph (i.e., with cyclic boundary conditions) of size $N = 32 \times 32$. Unlinke most MRF models, the pdf of this particular MRF has a known analytical expression [154] that leads to the following FIM [152]

$$\mathcal{I}(\beta) = \frac{1}{\sinh^2(\beta)} + 12N \sum_{n=1}^{N} \frac{\nu'(\beta)^2 \Delta_n(\beta) - \nu''(\beta)\bar{\Delta}_n(\beta)}{\bar{\Delta}_n(\beta)^{\frac{3}{2}}} \tag{5.14}$$

with

$$\nu'(\beta) = \cosh(\beta) - \frac{\cosh(\beta)}{\sinh^2(\beta)}$$

$$\nu''(\beta) = \sinh(\beta) + \frac{1 + \cosh^2(\beta)}{\sinh^3(\beta)}$$

and

$$\Delta_n(\beta) = \frac{\cosh^2(\beta)}{\sinh(\beta)} - \cos\left[\frac{(2n-1)\pi}{2N}\right]$$

$$\bar{\Delta}_n(\beta) = \Delta_n(\beta)^2 - 1$$

Fig. 5.1 compares the true CRB values computed with (5.14) with the estimates obtained using the proposed method. These estimates have been computed from $P = 10\,000$ independent Monte Carlo samples, generated using 10 000 Gibbs moves per field component (these estimates are represented as red crosses). In order to illustrate the effect of the number of Gibbs moves used to approximate step 3 of Algo. 5, this experiment was repeated using 100 000 Gibbs moves (estimates are represented as black circles). To ease visual interpretation, results are displayed using a logarithmic scale. We observe that the

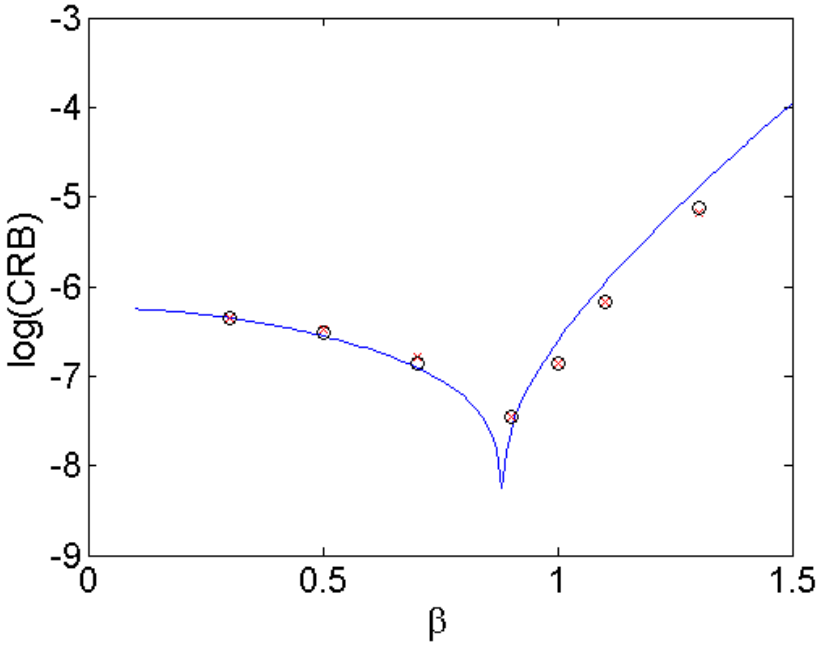


Figure 5.1: Cramer-Rao bound for an Ising ($K = 2$) defined on a toroidal graph.

estimates obtained with the proposed method are in good agreement with the true values of the CRB, and that the approximation introduced by using a Gibbs sampler within Algo. 5 has not affected the estimation results significantly. We also observe that the CRB varies significantly with the value of β and that it reaches a minimum at approximately $\beta = 0.9$. Note that the location of this minimum coincides with the phase-transition temperature of the Ising MRF ($\beta_c = \log(1 + \sqrt{2}) \approx 0.88$). This result is in agreement with the fact that the log-partition function $\log C(\beta)$ has its highest derivatives around the phase-transition temperature, as noticed in [99, 152]. Moreover, Fig. 5.2 illustrates the effect of the number of particles P on the CRB estimates for different values of β. We observe that in these experiments the estimates stabilize for $P > 1\,000$.

5.4.3 Asymptotic study of the CRB

The second set of experiments shows the evolution of the CRB with the size of the observation vector z (i.e., the number of field coordinates N). The CRB has been computed for the

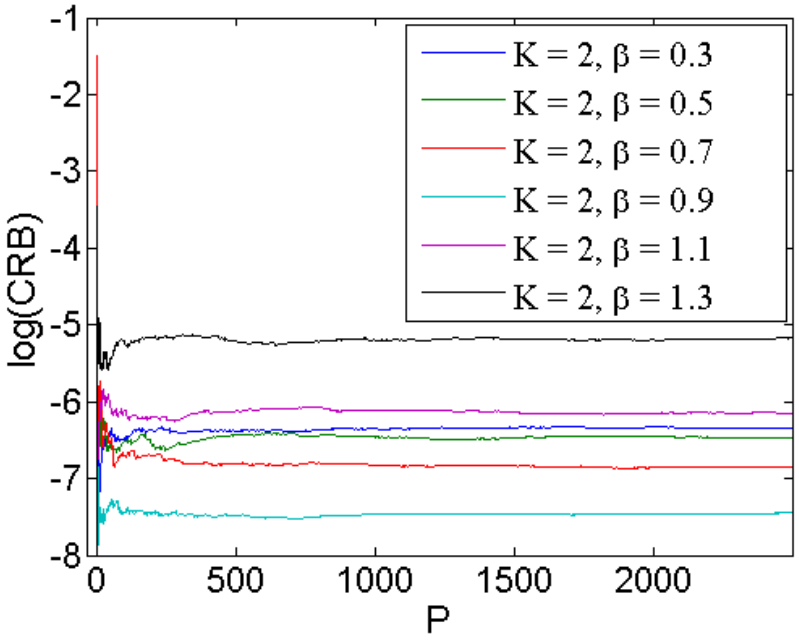


Figure 5.2: Cramer-Rao bound for an Ising ($K = 2$) defined on a toroidal graph.

following 5 field sizes $N = (2^8, 2^{10}, 2^{12}, 2^{14}, 2^{16})$, these values correspond to bidimensional MRF of size 16×16, 32×32, 64×64, 128×128 and 256×256. Experiments have been performed using an Ising MRF, a 3-state and a 4-state Potts MRF (i.e., $K = 2$, $K = 3$ and $K = 4$ respectively) defined on a plane (not a toroid). CRB estimates have been computed using 10 000 Gibbs moves per field component and $P = 2\,500$ particles. Finally, for each model, the parameter β was set to the critical phase-transition value, i.e., $\beta_c = \log(1 + \sqrt{K})$ to introduce a strong dependence between the components of the MRF. Fig. 5.3 shows the resulting CRB values in double logarithmic scale.

We observe that for all models the logarithm of the CRB increases almost linearly with the logarithm of the number of field components. This result shows that the strong dependence between the field components did not modify significantly the linear behavior that is generally observed for models with statistically independent components. We also observe that the CRB increased with the number of states K, indicating that an accurate estimation of β for the Ising model is perhaps more difficult than for a Potts MRF.

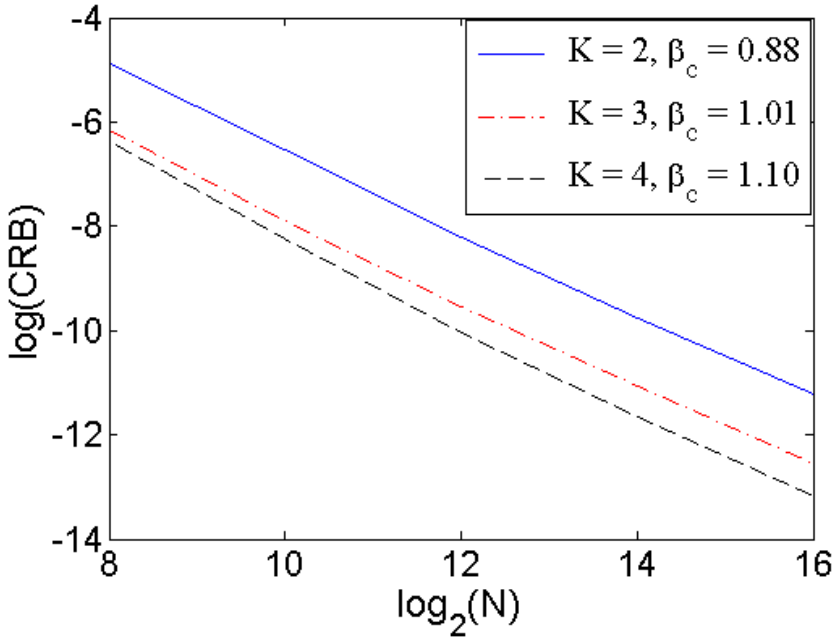


Figure 5.3: Cramer-Rao bound for an Ising ($K = 2$) and two Potts MRF ($K = 3$ and $K = 4$) at phase-transition $\beta_c = \log(1 + \sqrt{K})$ and for different field sizes N. Results are displayed in *log-log* scale.

5.4.4 Evaluation of state-of-the art estimators of β

The third set of experiments compares the CRB to the empirical variance of three state-of-the art estimation methods, the *auxiliary variable* [92], *exchange* [126] and *ABC* [10] algorithms. As explained previously, the CRB is often used as a means to measure the *efficiency* of unbiased estimators in term of mean square error. In this chapter, these three algorithms have been applied to compute an approximate maximum-likelihood (ML) estimation of β for the Ising and the 3-state Potts MRF. Experiments were conducted as follows. First a random vector $z \sim f(z|\beta)$ was generated using an appropriate Gibbs sampler. Once z had been generated, each estimation method was used to produce an approximation of the intractable likelihood $f(z|\beta)$. Precisely, each algorithm was used to generate 1 000 samples approximately distributed according to the likelihood function. Finally, for each set of samples, an ML estimate was computed by maximizing a non-parametric kernel density estimation. This procedure was repeated 2 500 times to obtain the variance of the ML estimates. All

algorithms used 250 burn-in steps and 10 Gibbs moves per auxiliary variable coordinate, which are realistic implementation conditions for signal processing applications [10, 141]. Moreover, the auxiliary variable method [92] was implemented using the true value of β as auxiliary estimate, and the tolerance of the ABC method [10] was set to 1% (please see [11] for more details about these methods and their application to the Ising and the Potts MRF).

Fig. (5.4) compares the CRB values computed for the toroidal Ising model and the variance of the ML estimates obtained with the state-of-the art algorithms and with the known analytical partition function [152, 154]. These values have been computed for $\beta < \beta_c$ which is the range of interest in signal processing applications (for $\beta > \beta_c$ all the field components are almost surely of the same color). We observe that the ML estimator based on the known analytical pdf is efficient (i.e., close to the CRB) for all the values of β. Similarly, the ML estimators based on the exchange [126] and the ABC [10] methods are also efficient for small values of β (i.e., $\beta < 0.5$). However, their efficiency decreases progressively for $\beta > 0.5$. This deterioration is due to a degradation of the approximate likelihood and could be improved by increasing the number of iterations of the Gibbs sampler, or by using perfect sampling [139]. Moreover, the ML estimator based on the auxiliary variable method [92] is considerably less efficient than the other estimators. This result is in accordance with the experiments reported in [10, 126, 127].

Furthermore, Fig. 5.5 shows the CRB values computed for a planar 3-state Potts MRF of size 32×32 and $\beta < \beta_c = \log(1 + \sqrt{3})$. Again, these values are compared to the empirical variance of the state-of-the art methods. We observe that for the Potts MRF the CRB also varies significantly with the value of β and that it reaches a minimum at approximately $\beta = 1.0$. Again, the location of this minimum coincides with the phase-transition temperature $\beta_c \approx 1.01$. Similarly to 5.4, the ML estimates based on the Exchange [126] and the ABC [10] methods are close to the CRB for small values of β, and depart progressively as β increases.

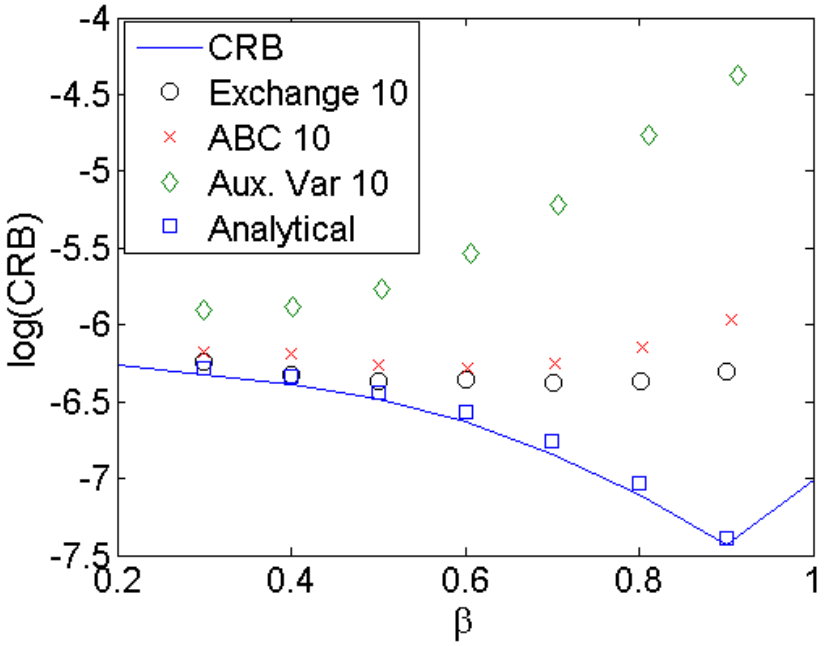


Figure 5.4: Cramer-Rao bound for a toroidal Ising model of size 32×32. Results are displayed in logarithmic scale.

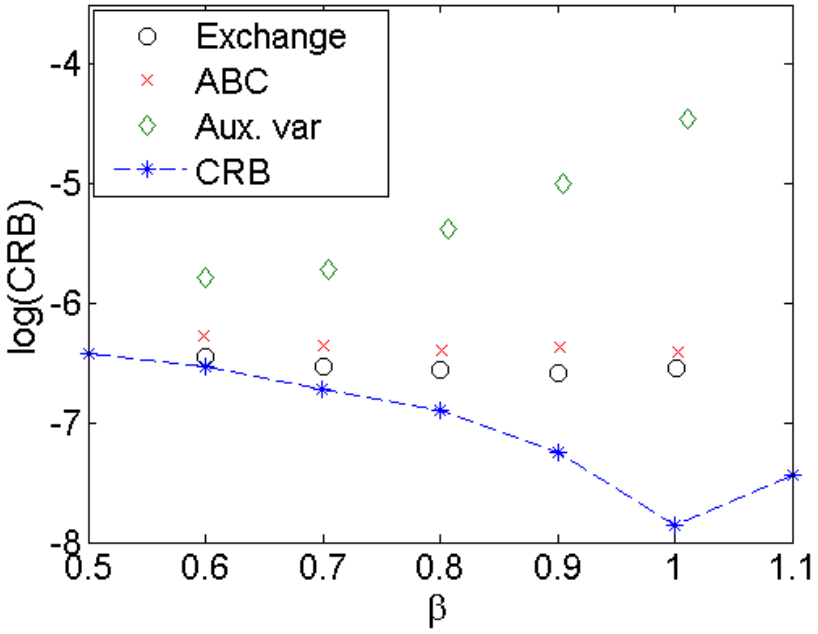


Figure 5.5: Cramer-Rao bound for an 3-state Potts model of size 32×32. Results are displayed in logarithmic scale.

5.5 Conclusion

This chapter studied the problem of computing the CRB of the parameters of Markov random fields. It has been shown that for these distributions the CRB depends on the derivatives of the normalizing constant or partition function $C(\boldsymbol{\theta})$, which is generally intractable. This problem has been addressed by proposing an original Monte Carlo algorithm based on a property of the exponential family. Precisely, that the derivatives of $C(\boldsymbol{\theta})$ can be related to the second order statistical moments of the Gibbs potential of the MRF. Based on this result, an original Monte Carlo method was proposed to compute the Fisher Information matrix of the MRF and therefore the CRB. To illustrate the interest of this method, the proposed algorithm has been successfully applied to the Ising and the Potts models, which are frequently used in signal processing applications. The resulting bounds have been used, in turn, to assess the statistical efficiency of three state-of-the art estimation methods that are interesting for image processing applications. These results revealed that for high temperatures (i.e., low values of β) the estimation methods are close to being statistically efficient. Perspectives for future work include deriving the error bounds of Bayesian estimators for intractable models, and the application of the proposed method to other MRFs for which the CRB is unknown.

Chapter 6

Concluding Remarks

This thesis presented image processing methods developed to aid dermatologists to use in vivo ultrasound imaging to explore the structure of human skin in general and lesions in particular. More precisely, Bayesian methods and MCMC algorithms have been considered to perform tissue segmentation in high-frequency envelope ultrasound images of skin. Four main methodological contributions have been proposed in Chapters II, III, IV and V.

Chapter II studied the statistical distribution of single-tissue (i.e., homogeneous) regions in dermatological ultrasound images. Starting from the widely accepted point scattering model, mathematical developments have shown that ultrasound signals from skin tissues follow a Levy Flight and have α-stable statistics. As a result, envelope signals are distributed according to a heavy-tailed Rayleigh distribution. It has also been established that α-stable distributions model a new case of non-Gaussian statistics, where both the number of scatterers and the variance of their cross-sections tend to infinity (a configuration that is believed to arise in the dermis). In addition, analytical expressions have been provided to relate the α-stable parameters to scatterer properties. Simulations and experimental results supported by excellent goodness-of-fit tests confirm the proposed analytical results, which provide new insight into non-Gaussian statistics and set the basis for new ultrasound image processing methods.

Chapter III addressed the problem of multiple-tissue segmentation in 2D and 3D B-mode ultrasound images of skin. A spatially coherent finite mixture of heavy-tailed Rayleigh distributions was proposed to represent the image statistics. Spatial correlation was introduced into the model by using a Potts Markov random field that promotes dependance between neighbor pixels. An original Bayesian algorithm combined with a Markov chain Monte Carlo (MCMC) method was then proposed to jointly estimate the mixture parameters and classify voxels into different tissues in 2D and 3D ultrasound images. Experimental results showed that the proposed technique outperforms a state of the art method in the segmentation of in-vivo lesions. A tridimensional reconstruction of a melanoma tumor suggested that the resulting segmentation can be used to assess lesion boundaries in dermatologic oncology.

The segmentation results obtained by the method presented in Chapter III depend on the hyperparameter β, which controls the amount of spatial correlation introduced by the Potts model. This important parameter had to be set heuristically by cross-validation because conventional MCMC methods cannot estimate β from the data. This difficulty has been addressed in Chapter IV by presenting an original hybrid Gibbs sampler for estimating the Potts parameter β jointly with the unknown parameters of a Bayesian model. The estimation of β has been included within an MCMC method using an ABC likelihood-free Metropolis-Hastings algorithm, in which intractable terms have been replaced by simulation-rejection schemes. The ABC distance function has been defined using the Potts potential, which is the natural sufficient statistics of the Potts model. The proposed method has been successfully applied to perform unsupervised tissue segmentation in a 3D ultrasound image.

Chapter V addressed the problem of computing the Cramer-Rao bound of the granularity coefficient β of an Ising or a Potts Markov random field. It has been shown that for both distributions the bound depends on the 2nd order derivative of the logarithm of the partition function, which is generally intractable. This problem has been addressed by proposing an

original Monte Carlo algorithm based on an identity of the exponential family. More precisely, it has been established that this derivative is equivalent to the variance of the MRF potential, which can be efficiently estimated using standard Monte Carlo techniques. The proposed algorithm has been successfully applied to several Ising and Potts models. These results have revealed that some of the state-of-the art algorithms to estimate β are close to being efficient.

The work presented in this thesis can be further developed in several ways. For instance, perspectives for future work include enhancing the accuracy of the proposed segmentation method by combining it with ultrasound image restoration techniques. More precisely, by considering Bayesian methods to perform tissue segmentation jointly with ultrasound image deconvolution (note that restoration of ultrasound images is considered a challenging problem because imaging systems generally have space-variant and partially unknown point-spread-functions). Similarly, segmentation accuracy could also be improved by studying new prior models that include topological constraints derived from skin anatomy (i.e., layer structure). For example, it would be possible to consider hybrid prior distributions that combine the Potts model with global anatomical information encoded using digital topology. Alternatively, prior distributions defined through topologically constrained active contours or appearance models.

Another important possibility for future work is the development of faster Bayesian algorithms capable of performing tissue segmentation in real-time. One perspective is to study deterministic optimization methods (i.e., graph-cut) to substitute the MCMC algorithm proposed in Chapter III. The estimation of the Potts parameter β could be integrated by using approximation methods based on off-line sampling. Also, computing time could be reduced by studying fast initialization techniques using other models (i.e., parametric active contours, clustering, etc).

Finally, the tissue segmentation methods proposed in Chapters III and IV have not been tested extensively. Future works should characterize their performance in a clinical context by applying them to a large set of ultrasound images and comparing the resulting segmentations with the annotations of several experts and the subsequent histology results.

Appendix A

Statistical properties of the scattering cross-section

Let $F_A(a_i)$, $F_P(p_i)$ and $F_Z(z_i)$ be the cumulative distribution functions of the following random variables

- $a_i \in \mathbb{R}^+$
- $p_i = p(t - \tau_i) \in [-p^*, p^*]$
- $z_i = a_i p_i \in \mathbb{R}$

where $0 < p^* < \infty$. We assume that $F_A(a_i)$, $F_P(p_i)$ and $F_Z(z_i)$ are absolutely continuous and denote by $f_A(a_i)$, $f_P(p_i)$ and $f_Z(z_i)$ the density functions of a_i, p_i and z_i.

A.1 Conditions on $f_A(a_i)$ and $f_P(p_i)$ for convergence towards $S\alpha S$ distributions with $\alpha < 2$

The addition of *i.i.d.* random variables falls in the domain of attraction of a stable distribution with exponent $\alpha < 2$ if and only if these variables satisfy the Doebling & Gnedenko

conditions [34, p.175]. Specifically, if the distribution of $X_t = z_1 + \ldots + z_M$ converges to a $S\alpha S$ distribution with $\alpha < 2$, then it is positive that $F_Z(z_i)$ verifies the following conditions:

$$\lim_{z_i \to \infty} \frac{F_Z(-z_i)}{1 - F_Z(z_i)} = \frac{C_+}{C_-} = 1 \tag{A.1}$$

$$\lim_{z_i \to \infty} \frac{1 - F_Z(z_i) + F_Z(-z_i)}{1 - F_Z(l\, z_i) + F_Z(-l z_i)} = l^\alpha, \quad \forall l > 0 \tag{A.2}$$

where $C_+ = C_-$ due to symmetry.

The first condition (A.1) establishes that the tail of $F_Z(-z_i)$ and $1 - F_Z(z_i)$ are asymptotically equivalent. In Landau notation, condition (A.1) is equivalent to

$$F_Z(-z_i) = [1 - F_Z(z_i)] + o\,(1 - F_Z(z_i))\,. \tag{A.3}$$

By replacing (A.3) in the second condition (A.2)

$$\lim_{z_i \to \infty} \frac{2[1 - F_Z(z_i)] + o\,(1 - F_Z(z_i))}{2[1 - F_Z(l z_i)] + o\,(1 - F_Z(l z_i))} = \lim_{z_i \to \infty} \frac{1 - F_Z(z_i)}{1 - F_Z(l z_i)} = l^\alpha. \tag{A.4}$$

Recalling that $z_i = a_i p_i$ and using Rohatgi s result on the distribution of products of random variables [155]

$$F_Z(z_i) = \begin{cases} \int_{-p^*}^{0} f_P(p_i)[1 - F_A(z_i/p_i)]\, dp_i & \text{if } z_i < 0 \\ F_Z(0^-) + \int_0^{p^*} f_P(p_i) F_A(z_i/p_i)\, dp_i & \text{if } z_i \geq 0 \end{cases} \tag{A.5}$$

Finally, by replacing (A.5) in (A.4)

$$\lim_{z_i \to \infty} \frac{\int_0^{p^*} f_P(p_i)[1 - F_A(z_i/p_i)]\, dp_i}{\int_0^{p^*} f_P(p_i)[1 - F_A(l z_i/p_i)]\, dp_i} = l^\alpha, \quad \forall l > 0 \tag{A.6}$$

which is equivalent to

$$\lim_{z_i \to \infty} \frac{\int_0^{p^*} f_P(p_i)[\int_{z_i/p_i}^{\infty} f_A(\rho)\, d\rho]\, dp_i}{\int_0^{p^*} f_P(p_i)[\int_{z_i/p_i}^{\infty} f_A(l\rho) l\, d\rho]\, du} = l^{\alpha}, \quad \forall l > 0. \tag{A.7}$$

Therefore the product $a_i p_i$ falls in the domain of attraction of a $S\alpha S$ distribution with $\alpha < 2$ if and only if $F_A(a_i)$ and $f_P(p_i)$ verify (A.6) or $f_A(a_i)$ and $f_P(p_i)$ verify (A.7).

Conditions (A.6) and (A.7) are for instance verified when a_i follows a Pareto distribution with tail exponent α, i.e.,

$$f_A(a_i) = \alpha \frac{a_m^{\alpha}}{a_i^{\alpha+1}}, \quad F_A(a_i) = 1 - a_m^{\alpha} a_i^{-\alpha} \tag{A.8}$$

where $a_m > 0$ is the distribution s mode. By replacing $F_A(a_i)$ (A.8) in (A.6), $\forall l > 0$

$$\begin{aligned} \lim_{z_i \to \infty} \frac{\int_0^{p^*} f_P(p_i) a_m^{\alpha} p_i^{\alpha} z_i^{-\alpha}\, dp_i}{\int_0^{p^*} f_P(p_i) a_m^{\alpha} p_i^{\alpha} z_i^{-\alpha} l^{-\alpha}\, dp_i} &= \lim_{z_i \to \infty} \frac{\int_0^{p^*} f_P(p_i) a_m^{\alpha} p_i^{\alpha} z_i^{-\alpha}\, dp_i}{l^{-\alpha} \int_0^{p^*} f_P(p_i) a_m^{\alpha} p_i^{\alpha} z_i^{-\alpha}\, dp_i} \\ &= \lim_{z_i \to \infty} l^{\alpha} \\ &= l^{\alpha}. \end{aligned} \tag{A.9}$$

In fact, a_i need not be exactly distributed according to a Pareto distribution to verify (A.7). This condition is also satisfied by all the distributions that are asymptotically equivalent to a Pareto distribution. These generalizations of the Pareto distribution, denominated power-law distributions, have the following form

$$f_A(a_i) = L(a_i) a_i^{-(\alpha+1)} \tag{A.10}$$

where $L(a_i)$ is any function such that $\lim_{a_i \to \infty} L(a_i) = \alpha a_m^{\alpha}$. Power-law distributions (A.10) are asymptotically equivalent to a Pareto distribution, i.e.,

$$L(a_i) a_i^{-(\alpha+1)} = \alpha a_m^{\alpha} a_i^{-(\alpha+1)} + o(a_i^{-(\alpha+1)}) \tag{A.11}$$

since $\lim_{a_i\to\infty} \frac{L(a_i)a_i^{-(\alpha+1)}}{\alpha a_m^{\alpha} a_i^{-(\alpha+1)}} = 1$. Similarly, the survival function of power-law distributions is asymptotically equivalent to the survival function of Pareto distributions

$$1 - F_A(a_i) = \int_{a_i}^{\infty} L(\rho)\rho^{-(\alpha+1)}d\rho = a_m^{\alpha}a_i^{-\alpha} + o(a_m^{\alpha}a_i^{-\alpha}) \tag{A.12}$$

By multiplying both sides of (A.12) by $f_P(p_i)$ and integrating over $[0, p^*]$ it can be shown that

$$\int_0^{p^*} f_P(p_i)[1 - F_A(z_i/p_i)]\,dp_i = \int_0^{p^*} f_P(p_i)a_m^{\alpha}a_i^{-\alpha}dp_i + o\left(\int_0^{p^*} f_P(p_i)a_m^{\alpha}a_i^{-\alpha}dp_i\right) \tag{A.13}$$

Finally, in view of (A.9), the power-law cumulative distribution function $F_A(a_i) = \int_0^{a_i} L(\rho)\rho^{-(\alpha+1)}d\rho$ verifies the condition (A.6)

$$\begin{aligned}
\lim_{z_i\to\infty} \frac{\int_0^{p^*} f_P(p_i)[1 - F_A(z_i/p_i)]\,dp_i}{\int_0^{p^*} f_P(p_i)[1 - F_A(lz_i/p_i)]\,dp_i} &= \lim_{z_i\to\infty} \frac{\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}z_i^{-\alpha}dp_i + o\left(\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}z_i^{-\alpha}dp_i\right)}{\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}l^{-\alpha}z_i^{-\alpha}dp_i + o\left(\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}l^{-\alpha}z_i^{-\alpha}dp_i\right)} \\
&= \lim_{z_i\to\infty} \frac{\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}z_i^{-\alpha}dp_i}{\int_0^{p^*} f_P(p_i)a_m^{\alpha}p_i^{\alpha}l^{-\alpha}z_i^{-\alpha}dp_i} \\
&= l^{\alpha}.
\end{aligned} \tag{A.14}$$

We conclude that for modeling and physical interpretation purposes the scattering cross-sections can be assumed to follow a Pareto distribution with tail exponent α, even though $f_A(a_i)$ could in fact be any power-law asymptotically equivalent to that distribution.

Appendix B

Physical interpretation of the $S\alpha S$ parameter γ

The physical interpretation of γ can be derived from the Generalized Central Limit Theorem . Given that $a_i p(t-\tau_i)$ is in the domain of attraction of non-Gaussian distributions, then as $M \to \infty$ a properly normalized point scattering model converges to a standard $S\alpha S$ law [50, p.22]

$$\sum_{i=1}^{M} \frac{a_i p(t-\tau_i)}{\sqrt[\alpha]{D(\alpha) M (C_- + C_+)}} \xrightarrow{d} S\alpha S(\alpha, 1) \tag{B.1}$$

where $D(\alpha) = \frac{\pi}{2\Gamma(\alpha)\sin(\frac{\pi\alpha}{2})}$, M is the number of scatterers and

$$\begin{aligned} C_+ &= \lim_{z_i \to \infty} z_i^{\alpha} \left[1 - F_Z(z_i)\right] \\ C_- &= \lim_{z_i \to \infty} z_i^{\alpha} \left[F_Z(-z_i)\right] \end{aligned} \tag{B.2}$$

where it is recalled that $F_Z(z_i)$ is the cumulative distribution function of z_i. From the normalization property of stable laws

$$\gamma^{\alpha} = 2D(\alpha) M \left(\frac{C_- + C_+}{2}\right). \tag{B.3}$$

We observe that the terms C_- and C_+ depend exclusively on the asymptotic behavior of $F_Z(z_i)$, which is determined by the asymptotic behavior of $F_A(a_i)$. Assuming that the distribution of a_i is asymptotically equivalent to a Pareto distribution with exponent α

$$\begin{aligned}\left(\frac{C_- + C_+}{2}\right) &= \lim_{z_i\to\infty} z_i^\alpha \int_0^\infty f_P(p_i)\left(\frac{a_m p_i}{z_i}\right)^\alpha dp_i + \lim_{z_i\to-\infty} z_i^\alpha \int_{-\infty}^0 f_P(p_i)\left(\frac{a_m p_i}{z_i}\right)^\alpha dp_i \\ &= a_m^\alpha \int_{-\infty}^\infty f_P(p_i) p_i^\alpha \, dp_i \\ &= a_m^\alpha \langle p_i^\alpha \rangle\end{aligned} \tag{B.4}$$

where $p_i = p(t-\tau_i)$, f_P and $\langle p_i^\alpha \rangle$ are respectively the density function and the α-th fractional moment of p_i, and a_m is given by

$$a_m = \lim_{a_i\to\infty} a_i^\alpha F_A(a_i). \tag{B.5}$$

Finally, by considering γ independent of t and replacing (B.4) in (B.3)

$$\gamma = D^*(\alpha)\sqrt[\alpha]{M} a_m \tag{B.6}$$

where $D^*(\alpha) = \sqrt[\alpha]{\frac{2\pi\langle p_i^\alpha\rangle}{\Gamma(\alpha)\sin\left(\frac{\pi\alpha}{2}\right)}}$.

Appendix C

Gamma mixture model

The gamma mixture model states that observations in the kth class are fully described by the gamma distribution with parameters L and m_k

$$r_n|z_n = k \sim \left(\frac{L}{m_k}\right)^L \frac{r_n^{L-1}}{\Gamma(L)} \exp\left(-\frac{Lr_n}{m_k}\right) \tag{C.1}$$

where $\Gamma(t) = \int_0^{+\infty} u^{t-1}e^{-u}du$ is the standard gamma function and L (the number of looks) is assumed to be known.

This study uses a conjugate inverse gamma distribution with hyperparameters a_0 and b_0 as prior for m_k

$$m_k \sim \mathcal{IG}(a_0, b_0), \quad k = 1, \ldots, K. \tag{C.2}$$

The generation of samples according to the conditional distributions $f(\boldsymbol{m}|L, \boldsymbol{z}, \boldsymbol{r})$ can then be easily achieved by sampling $(m_1, \ldots, m_K)$ coordinate-by-coordinate using the following Gibbs moves

$$m_k^{(t)} \sim \mathcal{IG}\left(a_0 + n_k L, b_0 + L \sum_{\{n|z_n=k\}} r_n\right) \tag{C.3}$$

where $n_k = \sharp\{n|z_n = k\}$ and where (C.3) results from the conjugacy property of the prior distribution [156, p. 265]. In this work, the hyperparameter a_0 and b_0 are fixed in order to obtain a vague prior (i.e., the hyperparameters have been set to $a_0 = 1$ and $b_0 = 1$ in our

experiments). However it would be possible to incorporate more specific prior information about m_k by setting a_0 and b_0 differently (see discussions in [157]).

Appendix D

α-Rayleigh mixture model

The α-Rayleigh mixture model states that the observations in the kth class are fully described by an α-Rayleigh distribution

$$r_n|z_n = k \sim p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k) \triangleq r_n \int_0^\infty \lambda \exp\left[-(\gamma_k\lambda)^{\alpha_k}\right] J_0(r_n\lambda)\, d\lambda \tag{D.1}$$

where α_k and γ_k are the parameters associated with the kth class and where J_0 is the zeroth order Bessel function of the first kind.

Inference on $(\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^T$ requires defining priors for these parameters. Assuming a priori independence between the parameters α_k and γ_k, the joint prior for the α-Rayleigh parameters is

$$f(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = f(\boldsymbol{\alpha}) f(\boldsymbol{\gamma}) = \prod_{k=1}^{K} f(\alpha_k) f(\gamma_k) \tag{D.2}$$

where the prior $f(\alpha_k)$ ($k = 1, \ldots, K$) is a uniform distribution on $(0, 2]$ (this interval covers all possible values of this parameter)

$$f(\alpha_k) = \mathcal{U}(0, 2) \tag{D.3}$$

and the prior for γ_k is an inverse gamma distribution with hyperparameters a_0 and b_0

$$f(\gamma_k) = \mathcal{IG}(a_0, b_0), \quad k = 1, \ldots, K \tag{D.4}$$

where the hyperparameters are fixed in order to obtain a vague prior ($a_0 = 1$ and $b_0 = 1$ will be used in our experiments).

The generation of samples according to the conditional distributions $f(\boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{r})$ and $f(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r})$ is not easy to achieve. In this work $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are sampled coordinate-by-coordinate using random walk MH moves [94, p. 245], as proposed in [4]. Accordingly, the proposal distribution associated with α_k is a truncated normal distribution centered on the previous value of the chain with variance $\sigma^2_{\alpha,k}$

$$\alpha_k^* \sim \mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}, \sigma^2_{\alpha,k}) \tag{D.5}$$

where α_k^* denotes the proposed value at iteration t and $\alpha_k^{(t-1)}$ is the previous state of the chain. The hyperparameters $\sigma^2_{\alpha,k}$ are adjusted to ensure an acceptance ratio close to $\frac{1}{3}$, as recommended in [103, p. 316]. Finally, since the prior for α_k is uniform, the MH acceptance rate of the proposed move can be expressed as follows

$$\text{ratio} = \min\left\{1, \frac{\mathcal{N}_{(0,2)}(\alpha_k^{(t-1)}|\alpha_k^*, \sigma^2_{\alpha,k})}{\mathcal{N}_{(0,2)}(\alpha_k^*|\alpha_k^{(t-1)}, \sigma^2_{\alpha,k})} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k^*, \gamma_k)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k^{(t-1)}, \gamma_k)}\right\}. \tag{D.6}$$

Moreover, the proposal distribution associated with γ_k is a truncated normal distribution centered on the previous value of the chain with variance $\sigma^2_{\gamma,k}$

$$\gamma_k^* \sim \mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}, \sigma^2_{\gamma,k}\right) \tag{D.7}$$

where γ_k^* denotes the proposed value at iteration t, $\gamma_k^{(t-1)}$ is the previous state of the chain and $\mathcal{N}_{\mathbb{R}^+}$ is the Gaussian distribution truncated on $\mathbb{R}^+$. The acceptance ratio for this move is

$$\text{ratio} = \min\left\{1, \frac{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^{(t-1)}|\gamma_k^*, \sigma^2_{\gamma,k}\right)}{\mathcal{N}_{\mathbb{R}^+}\left(\gamma_k^*|\gamma_k^{(t-1)}, \sigma^2_{\gamma,k}\right)} \times \prod_{\{n|z_n=k\}}^{N} \frac{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^*) f(\gamma_k^*|a_0, b_0)}{p_{\alpha\mathcal{R}}(r_n|\alpha_k, \gamma_k^{(t-1)}) f(\gamma_k^{(t-1)}|a_0, b_0)}\right\} \tag{D.8}$$

where the prior distribution $f(\gamma_k|a_0, b_0)$ has been defined in (D.4).

Appendix E

Gaussian mixture model

The Gaussian mixture model states that the observations in the kth class are fully described by a Gaussian distribution with mean μ_k and variance σ_k^2

$$r_n|z_n = k \sim \mathcal{N}(\mu_k, \sigma_k^2).$$

This study uses conjugate priors for μ_k and σ_k^2

$$\mu_k \sim \mathcal{N}(0, \xi)$$

$$\sigma_k^2 \sim \mathcal{IG}(\kappa, \nu)$$

where ξ, κ and ν have been chosen to yield vague priors (i.e., ξ will be set to a large value and κ and ν will be set to small values in our experiments).

The generation of samples according to $f(\boldsymbol{\mu}|\boldsymbol{\sigma^2}, z, \boldsymbol{r})$ and $f(\boldsymbol{\sigma^2}|\boldsymbol{\mu}, z, \boldsymbol{r})$ can be easily achieved by sampling $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$ and $\boldsymbol{\sigma^2} = \left(\sigma_1^2, \ldots, \sigma_K^2\right)$ coordinate-by-coordinate using the following Gibbs moves

$$\mu_k{}^{(t)} \sim \mathcal{N}\left(\frac{\sum_{(n|z_n=k)} \frac{r_n}{\sigma_k^2}}{\frac{1}{\xi} + \frac{n_k}{\sigma_k^2}}, \frac{1}{\frac{1}{\xi} + \frac{n_k}{\sigma_k^2}}\right) \tag{E.1}$$

$$\sigma_k^{2(t)} \sim \mathcal{IG}\left(\kappa + \frac{n_k}{2}, \nu + \sum_{(n|z_n=k)} \frac{(r_n - \mu_k)^2}{2}\right) \tag{E.2}$$

where $n_k = \sharp\{n|z_n = k\}$ and where (E.1) and (E.2) result from the conjugacy property of the prior distribution [156, p. 265].

Publications

Journals

M. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Estimating the Granularity Coefficient of a Potts-Markov Random field within an MCMC algorithm,”, submitted.

M. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Segmentation of skin lesions in 2D and 3D ultrasound images using a spatially coherent generalized Rayleigh mixture model,” *IEEE Trans. Med. Imag.*, to appear.

M. Pereyra and H. Batatia, “Modeling Ultrasound Echoes in Skin Tissues using Symmetric α-Stable Processes,” *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 59, no. 1, pp. 60U72, Jan. 2012.

Pipelined Journals

M. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Computing the Cramer-Rao bound of the granularity coefficient of an Ising or a Potts Markov field”.

Z. Irace, M. Pereyra, N. Dobigeon and H. Batatia, “A Riemannian Population MCMC method for Negative Binomial mixture regression and classification,”.

Conferences

Z. Irace, M. A. Pereyra, N. Dobigeon and H. Batatia, “Un modèle de mélange de lois Poisson-Gamma pour segmenter des images TEP," in Actes de la Conférence Reconnaissance des Formes et Intelligence Artificielle (RFIA), Lyon, France, Jan. 2012, in French, to appear.

M. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Un modèle de mélange de lois Rayleigh généralisées spatialement corrélées pour la segmentation d échographies cutanées,” in *(GRETSI)*, Bordeaux, France, Sep. 2011.

M. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Segmentation of ultrasound images using a spatially coherent generalized Rayleigh mixture model,” in *European Signal Proc. Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011.

Z. Irace, M. A. Pereyra, N. Dobigeon and H. Batatia, “Bayesian Segmentation of Chest Tumors in PET Scans using a Poisson-Gamma Mixture Model,” in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Nice, France, June 2011.

M. A. Pereyra, N. Dobigeon, H. Batatia and J.-Y. Tourneret, “Labeling skin tissues in ultrasound images using a generalized Rayleigh mixture model,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, Prague, Czech Republic, May 2011.

M. A. Pereyra and H. Batatia, “A Levy Flight Model for Ultrasound in Skin Tissues,” in *Proc. IEEE Eng. Med. Bio. Soc. (EMBC)*, San Diego (CA), USA, Oct. 2010.

M. A. Pereyra and H. Batatia, “An Alpha-Stable Model for Ultrasound Speckle Statistics in Skin,” in *Proc. IEEE Eng. Med. Bio. Soc. (EMBC)*, Buenos Aires, Argentina, Sep. 2010.

Other

Poster presentation at the 9th IEEE EMBS Int. Summer School on Biomed. Imaging, Berder, France, June 2010.

Guest lecturer for IEEE Argentina, Oct. 2009 and Sep. 2010.

Bibliography

[1] A. Sarti, C. Corsi, E. Mazzini, and C. Lamberti, "Maximum likelihood segmentation of ultrasound images with Rayleigh distribution," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 52, no. 6, pp. 947–960, June 2005.

[2] S. MacNeil, "Progress and opportunities for tissue-engineered skin," *Nature*, vol. 445, no. 1, pp. 874–880, Feb. 2007.

[3] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Labeling skin tissues in ultrasound images using a generalized Rayleigh mixture model," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, Prague, Czech Republic, May 2011.

[4] ——, "Segmentation of skin lesions in $2d$ and $3d$ ultrasound images using a spatially coherent generalized Rayleigh mixture model," *IEEE Trans. Med. Imag.*, 2011, to appear.

[5] M. A. Pereyra and H. Batatia, "Modeling ultrasound echoes in skin tissues using symmetric α-stable processes," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 59, no. 1, pp. 60–72, Jan. 2012.

[6] ——, "A Levy ight model for ultrasound in skin tissues," in *Proc. IEEE Ultrason. Symp.*, San Diego (CA), USA, Oct. 2010.

[7] ——, "An alpha-stable model for ultrasound speckle statistics in skin," in *Proc. IEEE Eng. Med. Bio. Soc. (EMBC)*, Buenos Aires, Argentina, Sept. 2010.

[8] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Segmentation of skin lesions in **2**d and **3**d ultrasound images using a spatially coherent generalized Rayleigh mixture model," University of Toulouse, IRIT/INP-ENSEEIHT, France, Tech. Rep., Dec. 2011. [Online]. Available: http://pereyra.perso.enseeiht.fr/pdf/Pereyra_TMI_TR_2011.pdf

[9] M. A. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Segmentation of ultrasound images using a spatially coherent generalized Rayleigh mixture model," in *European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, Sep 2011, pp. 664–668.

[10] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity parameter of a Potts-Markov random field within an MCMC algorithm," submitted to IEEE Trans. Image Process.

[11] ——, "Estimating the granularity parameter of a Potts-Markov random field within an MCMC algorithm," University of Toulouse, IRIT/INP-ENSEEIHT, France, Tech. Rep., Feb. 2012. [Online]. Available: http://pereyra.perso.enseeiht.fr/pdf/PereyraIEEETIPtr2012.pdf

[12] B. I. Raju, "High frequency ultrasonic characterization of human skin in vivo," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, June 2002.

[13] K. S. Nehal, D. Gareau, and M. Rajadhyaksha, "Skin imaging with re ectance confocal microscopy," *Seminars in Cutaneous Medicine and Surgery*, vol. 27, no. 1, pp. 37 – 43, 2008.

[14] A. Mcheik, "Modélisation statistique du Speckle en OCT. Application à la segmentation d images de la peau," Ph.D. dissertation, Université de Toulouse, Toulouse, France, Oct. 2010.

[15] T. Szabo, *Diagnostic Ultrasound Imaging*. Boston,USA: Academic Press, 2004.

[16] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, no. 2, pp. 021 314–021 325, 2010.

[17] P. Morse and K. Ingard, *Theoretical Acoustics*. Princeton (NJ): Princeton University Press, 1987.

[18] P. M. Shankar, J. M. Reid, H. Ortega, C. W. Piccoli, and B. B. Goldberg, "Use of non-Rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast," *Medical Imaging, IEEE Transactions on*, vol. 12, no. 4, pp. 687 –692, Dec. 1993.

[19] P. Shankar, "A general statistical model for ultrasonic backscattering from tissues." *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 47, no. 3, pp. 727–736, 2000.

[20] P. M. Shankar, V. A. Dumane, J. M. Reid, V. Genis, F. Forsberg, C. W. Piccoli, and B. B. Goldberg, "Classification of ultrasonic B-mode images of breast masses using Nakagami distribution," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 48, no. 2, pp. 569 –580, March 2001.

[21] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *Sonics and Ultrasonics, IEEE Transactions on*, vol. 30, no. 3, pp. 156 – 163, May 1983.

[22] R. F. Wagner, M. F. Insana, and D. G. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," *J. Opt. Soc. Am. A*, vol. 4, no. 5, pp. 910–922, May 1987.

[23] V. Dutt and J. F. Greenleaf, "Ultrasound echo envelope analysis using a homodyned K distribution signal model," *Ultrasonic Imaging*, vol. 16, no. 4, pp. 265–287, 1994.

[24] P. M. Shankar, "A compound scattering pdf for the ultrasonic echo envelope and its relationship to K and Nakagami distributions," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 50, no. 3, pp. 339 –343, March 2003.

[25] B. I. Raju and M. A. Srinivasan, "Statistics of envelope of high-frequency ultrasonic backscatter from human skin in vivo," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 49, no. 7, pp. 871 –882, July 2002.

[26] P. M. Shankar, "Ultrasonic tissue characterization using a generalized Nakagami model," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 48, no. 6, pp. 1716 –1720, Nov. 2001.

[27] T. Eltoft, "The Rician inverse Gaussian distribution: a new model for non-Rayleigh signal amplitude statistics," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1722 –1735, Nov. 2005.

[28] R. Agrawal and Karmeshu, "Ultrasonic backscattering in tissue: characterization through Nakagami-generalized inverse Gaussian distribution," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 166 – 172, 2007.

[29] R. N. Mantegna and H. E. Stanley, "Stochastic process with ultraslow convergence to a Gaussian: The truncated levy ight," *Phys. Rev. Lett.*, vol. 73, no. 22, pp. 2946–2949, Nov. 1994.

[30] F. Destrempes and G. Cloutier, "A critical review and uniformized representation of statistical distributions modeling the ultrasound echo envelope," *Ultrasound Med. Biol.*, vol. 36, no. 7, pp. 1037–1051, 2010.

[31] L. C. Gilman, "First-order statistics of pulsed-sinusoid backscatter from random media: basic elements of an exact treatment," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 44, no. 4, pp. 798 –804, July 1997.

[32] J. M. Bordier, M. Fink, A. L. Brun, and F. Cohen-Tenoudji, "The in uence of multiple scattering in incoherent ultrasonic inspection of coarse grain stainless steel," in *Proc. IEEE Ultrason. Symp.*, vol. 2, Dec. 1991, pp. 803 –808.

[33] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*. New York (NJ): Chapman & Hall/CRC, 2000.

[34] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Boston: Addison Wesley Publishing Co., 1954.

[35] P. A. Petropulu, V. T. Nasis, O. Tretiak, and C. W. Piccoli, "Benign versus malignant classification of breast tumors based on the the PLSN model for the ultrasound RF echo and homomorphic filtering," in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, vol. 1, Sept. 2004, pp. 21 –24.

[36] A. Achim, B. Buxton, G. Tzagkarakis, and P. Tsakalides, "Compressive sensing for ultrasound RF echoes using alpha-stable distributions," in *Proc. IEEE Eng. Med. Bio. Soc. (EMBC)*, Buenos Aires, Argentina, Sept. 2010.

[37] M. A. Kutay, A. P. Petropulu, and C. W. Piccoli, "On modeling biomedical ultrasound RF echoes using a power-law shot-noise model," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 48, no. 4, pp. 953 –968, July 2001.

[38] A. Petropulu, J.-C. Pesquet, and X. Yang, "Power-law shot noise and its relationship to long-memory alpha-stable processes," *Signal Processing, IEEE Transactions on*, vol. 48, no. 7, pp. 1883 –1892, July 2000.

[39] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distribution and Applications*. New York (NJ): Wiley, 1995.

[40] J. Goodman, "A random walk through the field of speckle," *Opt. Eng.*, vol. 86, p. 610U612, 1986.

[41] L. A. Goodman, "On the exact variance of products," *Journal of the American Statistical Association*, pp. 708–713, Dec. 1960.

[42] A. Achim, E. Kuruoglu, and J. Zerubia, "SAR image filtering based on the heavy-tailed Rayleigh model," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2686–2693, Sept. 2006.

[43] Z. Sun and C. Han, "Heavy-tailed Rayleigh distribution: A new tool for the modeling of SAR amplitude images," in *Proc. (IGARSS 08). IEEE Int. Geosc. and Remote Sensing Symp.*, vol. 4, July 2008, pp. 1253–1256.

[44] E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, April 2004.

[45] J. S. Bodenschatz and C. L. Nikias, "Maximum-likelihood symmetric alpha-stable parameter estimation," *Signal Processing, IEEE Transactions on*, vol. 47, no. 5, pp. 1382 –1384, May 1999.

[46] G. A. Tsihrintzis and C. L. Nikias, "Fast estimation of the parameters of alpha-stable impulsive interference," *Signal Processing, IEEE Transactions on*, vol. 44, no. 6, pp. 1492 –1503, June 1996.

[47] J. P. Nolan, "Maximum likelihood estimation and diagnostics for stable distributions," in *Lévy Processes*, O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick, Eds. Boston: Brikhäuser, 2001, pp. 379–400.

[48] J. H. McCulloch, "Simple consistent estimators of stable distribution parameters," *Communications in Statistics - Simulation and Computation*, vol. 15, pp. 1109 – 1136, 1986.

[49] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics - Stochastic Models*, vol. 13, no. 4, pp. 759–774, 1997.

[50] ——, *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhauser, 2011.

[51] T. Hergum, S. Langeland, E. Remme, and H. Torp, "Fast ultrasound imaging simulation in k-space," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 56, no. 6, pp. 1159 –1167, June 2009.

[52] J. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 39, pp. 262 – 267, 1992.

[53] J. Jensen, "Field: A program for simulating ultrasound systems," in *Proc. 10th Nordic-Baltic Conf. on Biomed. Imag.*, 1996.

[54] M. Lebertre, "Echographie quantitative haute-frequence: Proprietes du derme humain et potentiel diagnostique," Ph.D. dissertation, University of Tours, Tours, France, 2001.

[55] P. M. Shankar, C. W. Piccoli, J. M. Reid, F. Forsberg, and B. B. Goldberg, "Application of the compound probability density function for characterization of breast masses in ultrasound B scans," *Phys. Med. Biol.*, vol. 50, no. 10, pp. 2241–2248, March 2001.

[56] J. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.

[57] A. Belaid, D. Boukerroui, Y. Maingourd, and J.-F. Lerallut, "Phase-based level set segmentation of ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 138–147, Jan. 2011.

[58] R. Schneider, D. Perrin, N. Vasilyev, G. Marx, P. del Nido, and R. Howe, "Mitral annulus segmentation from 3*d* ultrasound using graph cuts," *IEEE Trans. Med. Imag.*, vol. 29, no. 9, pp. 1676–1687, Sept. 2010.

[59] K. Somkantha, N. Theera-Umpon, and S. Auephanwiriyakul, "Boundary detection in medical images using edge following algorithm based on intensity gradient and texture gradient features," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 567–573, March 2011.

[60] P. Yan, S. Xu, B. Turkbey, and J. Kruecker, "Adaptively learning local shape statistics for prostate segmentation in ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 633–641, March 2011.

[61] G. Unal, S. Bucher, S. Carlier, G. Slabaugh, T. Fang, and K. Tanaka, "Shape-driven segmentation of the arterial wall in intravascular ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 3, pp. 335–347, May 2008.

[62] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1342–1355, Sept. 2008.

[63] P. Paramanathan and R. Uthayakumar, "Tumor growth in the fractal space-time with temporal density," in *Control, Computation and Information Systems*, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2011, vol. 140, pp. 166–173.

[64] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Automatic segmentation of breast lesions on ultrasound," *Medical Physics*, vol. 28, no. 8, pp. 1652–1659, 2001.

[65] I.-S. Jung, D. Thapa, and G.-N. Wang, "Automatic segmentation and diagnosis of breast lesions using morphology method based on ultrasound," in *Fuzzy Systems and Knowledge Discovery*, ser. Lecture Notes in Computer Science, L. Wang and Y. Jin, Eds. Springer Berlin / Heidelberg, 2005, vol. 3614, pp. 491–491.

[66] A. Madabhushi and D. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 155–169, Feb. 2003.

[67] Y.-L. Huang and D.-R. Chen, "Watershed segmentation for breast tumor in $2-d$ sonography," *Ultrasound in Medicine and Biology*, vol. 30, no. 5, pp. 625–632, 2004.

[68] X. Guofang, M. Brady, J. Noble, and Z. Yongyue, "Segmentation of ultrasound B-mode images with intensity inhomogeneity correction," *IEEE Trans. Med. Imag.*, vol. 21, no. 1, pp. 48–57, Jan. 2002.

[69] D. Boukerroui, A. Baskurt, J. Noble, and O. Basset, "Segmentation of ultrasound images–multiresolution $2d$ and $3d$ algorithm based on global and local statistics," *Pattern Recog. Letters*, vol. 24, no. 4-5, pp. 779–790, 2003.

[70] J. Dias and J. Leitao, "Wall position and thickness estimation from sequences of echocardiographic images," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 25–38, Feb. 1996.

[71] E. Brusseau, C. de Korte, F. Mastik, J. Schaar, and A. van der Steen, "Fully automatic luminal contour segmentation in intracoronary ultrasound imaging-a statistical approach," *IEEE Trans. Med. Imag.*, vol. 23, no. 5, pp. 554–566, May 2004.

[72] C. Chesnaud, P. Refregier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1145–1157, Nov. 1999.

[73] M. H. R. Cardinal, J. Meunier, G. Soulez, R. L. Maurice, E. Therasse, and G. Cloutier, "Intravascular ultrasound image segmentation: a three-dimensional fast-marching method based on gray level distributions," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 590–601, May 2006.

[74] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comput. Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[75] L. Saroul, O. Bernard, D. Vray, and D. Friboulet, "Prostate segmentation in echographic images: A variational approach using deformable super-ellipse and Rayleigh distribution," in *Proc. IEEE Int. Symp. Biomed. Imag (ISBI)*, May 2008, pp. 129–132.

[76] F. Destrempes, J. M. J., M. F. Giroux, G. Soulez, and G. Cloutier, "Segmentation in ultrasonic B-mode images of carotid arteries using mixture of Nakagami distributions and stochastic optimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 215–229, Feb. 2009.

[77] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[78] F. Lecellier, J. Fadili, S. Jehan-Besson, G. Aubert, M. Revenu, and E. Saloux, "Region-based active contours with exponential family observations," *J. Math. Imaging Vis.*, vol. 36, pp. 28–45, Jan. 2010.

[79] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[80] D. Van de Sompel and M. Brady, "Simultaneous reconstruction and segmentation algorithm for positron emission tomography and transmission tomography," in *Proc. IEEE Int. Symp. Biomed. Imag (ISBI)*, May 2008, pp. 1035 –1038.

[81] B. Scherrer, F. Forbes, C. Garbay, and M. Dojat, "Distributed local MRF models for tissue and structure brain segmentation," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1278 –1295, Aug. 2009.

[82] L. Risser, J. Idier, P. Ciuciu, and T. Vincent, "Fast bilinear extrapolation of 3D Ising field partition function. Application to fMRI image analysis," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Cairo, Egypte, Nov. 2009, pp. 833–836.

[83] J. Marroquin, E. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, Nov. 2003.

[84] F. Y. Wu, "The Potts model," *Rev. Mod. Phys.*, vol. 54, no. 1, pp. 235–268, Jan. 1982.

[85] M. Woolrich, T. Behrens, C. Beckmann, and S. Smith, "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data," *IEEE Trans. Med. Imag.*, vol. 24, no. 1, pp. 1–11, Jan. 2005.

[86] D. Salas-Gonzalez, E. E. Kuruoglu, and D. P. Ruiz, "Finite mixture of α-stable distributions," *Digit. Signal Process.*, vol. 19, pp. 250–264, March 2009.

[87] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[88] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. Ser. B*, vol. 36, no. 2, pp. 192–236, 1974.

[89] R. Kindermann and J. L. Snell, *Markov random fields and their applications*. Providence: RI: Amer. Math. Soc., 1980.

[90] Z. Zhou, R. Leahy, and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs prior," *IEEE Trans. Image Process.*, vol. 6, no. 6, pp. 844–861, June 1997.

[91] X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," *IEEE Trans. Image Process.*, vol. 8, no. 7, pp. 945–963, June 1999.

[92] J. Moller, A. N. Pettitt, R. Reeves, and K. K. Berthelsen, "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, vol. 93, no. 2, pp. 451–458, June 2006.

[93] C. McGrory, D. Titterington, R. Reeves, and A. Pettitt, "Variational Bayes for estimating the parameters of a hidden Potts model," *Statistics and Computing*, vol. 19, no. 3, pp. 329–340, Sept. 2009.

[94] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[95] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[96] J. Diebolt and E. H. S. Ip., "Stochastic EM: method and application," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996.

[97] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2675–2685, May 2010.

[98] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059 –2070, Sept. 2009.

[99] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 1059 –1074, April 2010.

[100] K. Kayabol, E. Kuruoglu, and B. Sankur, "Bayesian separation of images modeled with MRFs using MCMC," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 982 –994, May 2009.

[101] M. Mignotte, "Image denoising by averaging of piecewise constant simulations of image partitions," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 523 –533, Feb. 2007.

[102] J. Gonzalez, Y. Low, A. Gretton, and C. Guestrin, "Parallel Gibbs sampling: From colored fields to thin junction trees," in *Proc. Artifical Intell. Stat. (AISTATS)*, Ft. Lauderdale, FL, May 2011.

[103] G. O. Roberts, "Markov chain concepts related to sampling algorithms," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 259–273.

[104] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Stat. Sciences*, vol. 7, no. 4, pp. 457–511, 1992.

[105] C. P. Robert and S. Richardson, "Markov Chain Monte Carlo methods," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 1–25.

[106] A. Gelman, J. B. Carlin, H. P. Robert, and D. B. Rubin, *Bayesian Data Analysis*. London: Chapman & Hall, 1995.

[107] S. Z. Li, *Markov random field modeling in image analysis*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2001.

[108] L. Cordero-Grande, G. Vegas-Sanchez-Ferrero, P. Casaseca-de-la Higuera, and C. Alberola-Lopez, "A Markov random field approach for topology-preserving registration: Application to object-based tomographic image interpolation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2047 –2061, April 2012.

[109] D. Mahapatra and Y. Sun, "Integrating segmentation information for improved MRF-based elastic image registration," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 170 –183, Jan. 2012.

[110] T. Katsuki, A. Torii, and M. Inoue, "Posterior mean super-resolution with a causal Gaussian Markov random field prior," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2187–2197, April 2012.

[111] S. Jain, M. Papadakis, S. Upadhyay, and R. Azencott, "Rigid motion invariant classification of 3D-textures," *IEEE Trans. Image Process.*, 2012, to appear.

[112] L. Risser, T. Vincent, J. Idier, F. Forbes, and P. Ciuciu, "Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions. Application to the joint detection-estimation of brain activity in fMRI," *J. Sig. Proc. Syst.*, vol. 60, no. 1, pp. 1–14, July 2010.

[113] S. Pelizzari and J. Bioucas-Dias, "Oil spill segmentation of SAR images via graph cuts," *IEEE Trans. Geosci. Remote Sens.*, 2010, submitted. [Online]. Available: http://arxiv.org/abs/1007.4969

[114] M. Picco and G. Palacio, "Unsupervised classification of SAR images using Markov random fields and $calG_I^0$ model," *IEEE Trans. Geosci. Remote Sens.*, vol. 8, no. 2, pp. 350 –353, March 2011.

[115] L. Wang, J. Liu, and S. Z. Li, "MRF parameter estimation by MCMC method," *Pattern Recognition*, vol. 33, no. 11, pp. 1919 – 1925, July 2000.

[116] Y. Yu and Q. Cheng, "MRF parameter estimation by an accelerated method," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1251 – 1259, Aug. 2003.

[117] I. Murray and Z. Ghahramani, "Bayesian learning in undirected graphical models: Approximate MCMC algorithms," in *Proc. Conf. Uncertainty in Artificial Intell. (UAI)*. Arlington, Virginia: AUAI Press, 2004, pp. 392–399.

[118] Y. Cao, H. Sun, and X. Xu, "An unsupervised segmentation method based on MPM for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 2, no. 1, pp. 55 – 58, Jan. 2005.

[119] J. Besag, "Statistical analysis of non-lattice data," *J. Roy. Stat. Soc. Ser. D*, vol. 24, no. 3, pp. 179–195, Sept. 1975.

[120] C. J. Geyer and E. A. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data (with discussions)," *J. Roy. Statist. Soc.*, vol. B, no. 54, pp. 657–699, April 1992.

[121] D. B. Rubin, A. Gelman, C. John, and H. Stern, *Bayesian Data Analysis (2nd ed.)*. Boca Raton: Chapman & Hall, 2003.

[122] R. Reeves and A. Pettitt, "Efficient recursions for general factorisable models," *Biometrika*, vol. 91, pp. 751U–757, Dec. 2004.

[123] A. Gelman and X. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling," *Statistical Science*, vol. 13, no. 2, pp. 163–185, 1998.

[124] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. Roy. Stat. Soc. Ser. B*, vol. 72, no. 3, May 2010.

[125] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. Ser. B*, vol. 68, no. 3, May 2006.

[126] I. Murray, Z. Ghahramani, and D. MacKay, "MCMC for doubly-intractable distributions," in *Proc. (UAI 06) 22nd Annual Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA, USA, July 2006, pp. 359–366.

[127] R. G. Everitt, "Bayesian parameter estimation for latent Markov random fields and social networks," *J. Comput. Graphical Stat.*, 2012, to appear.

[128] P. Marjoram, J. Molitor, V. Plagnol, and S. TavarÃl , "Markov chain Monte Carlo without likelihoods," *Proc. Nat. Academy Sci.*, vol. 100, no. 26, pp. 15 324–15 328, Dec. 2003.

[129] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, "Population growth of human y chromosomes: a study of y chromosome microsatellites." *Molecular Biology and Evolution*, vol. 16, no. 12, pp. 1791–1798, 1999.

[130] J.-M. Marin, P. Pudlo, C. P. Robert, and R. Ryder, "Approximate Bayesian Computational methods," *Stat. Comput.*, vol. 21, no. 2, pp. 289–291, Oct. 2011.

[131] J.-M. Marin and C. P. Robert, *Bayesian core: a practical approach to computational Bayesian statistics*. New-York: Springer, 2007.

[132] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum rand estimator for image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1610 –1624, June 2010.

[133] K. Kayabol, E. Kuruoglu, J. Sanz, B. Sankur, E. Salerno, and D. Herranz, "Adaptive Langevin sampler for separation of t-distribution modelled astrophysical maps," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2357–2368, Sept. 2010.

[134] X. Zhou, Y. Lu, J. Lu, and J. Zhou, "Abrupt motion tracking via intensively adaptive Markov-chain Monte Carlo sampling," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 789 –801, Feb. 2012.

[135] F. Destrempes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov random field models and its Bayesian estimation," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2920 –2935, Oct. 2006.

[136] C. Nikou, A. Likas, and N. Galatsanos, "A Bayesian framework for image segmentation with spatially varying mixtures," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2278 –2289, Sept. 2010.

[137] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, and J.-C. Olivo-Marin, "Bayesian estimation for optimized structured illumination microscopy," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 601 – 614, Feb. 2012.

[138] J. G. Propp and D. B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Rand. Struct. Algorith.*, vol. 9, pp. 223–252, Aug. 1996.

[139] A. M. Childs, R. B. Patterson, and D. J. C. MacKay, "Exact sampling from nonattractive distributions using summary states," *Phys. Rev. E*, vol. 63, no. 3, pp. 36 113–36 118, Feb. 2001.

[140] A. Grelaud, J. M. Marin, C. Robert, F. Rodolphe, and F. Tally, "Likelihood-free methods for model choice in Gibbs random fields," *Bayesian Analysis*, vol. 3, no. 2, pp. 427–442, Jan. 2009.

[141] F. Liang, "A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants," *J. Stat. Comp. Simulation*, vol. 80, no. 9, pp. 1007–1022, 2010.

[142] T. McKinley, A. Cook, and R. Deardon, "Inference in epidemic models without likelihoods," *Int. Journal of Biostatistics*, vol. 5, no. 1, pp. 1–38, Sept. 2009.

[143] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate Bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.

[144] J.-Y. Tourneret, M. Doisy, and M. Lavielle, "Bayesian retrospective detection of multiple changepoints corrupted by multiplicative noise. Application to SAR image edge detection," *Signal Processing*, vol. 83, no. 9, pp. 1871–1887, Sept. 2003.

[145] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046 – 2057, Oct. 2004.

[146] H. Deng and D. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 528 – 538, March 2005.

[147] Y. Yang, H. Sun, and C. He, "Supervised SAR image MPM segmentation based on region-based hierarchical model," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 4, pp. 517 –521, Oct. 2006.

[148] Y. Li, J. Li, and M. Chapman, "Segmentation of SAR intensity imagery with a Voronoi tessellation, Bayesian inference, and reversible jump MCMC algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, April 2010.

[149] P. J. Green, "Reversible jump Markov chain Monte Carlo methods computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[150] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *J. Roy. Stat. Soc. Ser. B*, vol. 59, no. 4, pp. 731–792, 1997.

[151] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1âAŞ4, pp. 259 – 268, 1992.

[152] J.-F. Giovannelli, "Ising field parameter estimation from incomplete and noisy data," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Sept. 2011, pp. 1853 –1856.

[153] H. L. V. Trees, *Detection, estimation, and modulation theory: Part I.* New York: Wiley, 1968.

[154] L. Onsager., "A two-dimensional model with an order-disorder transition," *Phys. Rev.*, vol. 65, no. 3, pp. 117–149, Feb. 1944.

[155] V. M. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*. New York (NJ): Wiley-Interscience, 1976.

[156] J. Bernardo and A. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[157] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Proc.*, vol. 6, no. 4, pp. 352–372, July 1998.