

THÈSE DE DOCTORAT

de
L'UNIVERSITÉ PARIS-SACLAY

École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : Ecole normale supérieure Paris-Saclay

Établissement d'accueil : Ecole normale supérieure Paris-Saclay

Laboratoire d'accueil : Centre de mathématiques et de leurs applications, UMR
8536 CNRS

Spécialité de doctorat : Mathématiques appliquées

Zhijin LI

Task-based optimization of 3D breast x-ray imaging using
mathematical observers

Date de soutenance : 06 octobre 2017

Après avis des rapporteurs :

XAVIER DESCOMBES (INRIA Sophia Antipolis)

STEPHEN GLICK (U.S. Food and Drug Administration)

Jury de soutenance :

ANN-KATHERINE CARTON (GE Healthcare, PhD)

Co-encadrante de thèse

XAVIER DESCOMBES (INRIA Sophia Antipolis)

Rapporteur

AGNÈS DESOLNEUX (CNRS et ENS Paris-Saclay)

Directrice de thèse

ANNE ESTRADE (MAP5, Université Paris Descartes)

Examnatrice

STEPHEN GLICK (U.S. Food and Drug Administration)

Rapporteur

DOMINIQUE JEULIN (Mines ParisTech)

Examinateur

SERGE MULLER (GE Healthcare, PhD, HDR)

Examinateur

ALAIN TROUVE (ENS Paris-Saclay)

Examinateur

Acknowledgment

This PhD thesis research is a collaboration between GE Healthcare France and laboratory Centre de mathématiques et de leurs applications (CMLA), Ecole Normale Supérieure de Paris-Saclay. It is partially funded by the Association Nationale de la Recherche Technique (ANRT) under CIFRE grant n°2013/1052.

I am sincerely thankful to everyone who has helped, encouraged and inspired me through these three and half years. I would like to express my gratitude to my two thesis advisors Prof. Agnès Desolneux and Dr. Ann-Katherine Carton. Agnès, your passion for mathematics, as well as your rigorous and scientific thinking keep inspiring until today. I still remember those precious and intriguing discussions we had, around the whiteboard in your office, about probability and stochastic geometry. Ann-Katherine, thank you for teaching me how to be concise on my writing and for all the effort and trust that you put in me. I learned so much from you from the day I started working at GE; and I'm sure that these three years of learning will be one of the most valuable things in my whole life. One would feel extremely lucky if he has met a great mentor in his life; and me, I have met two.

I would like to thank Dr. Xavier Descombes and Prof. Stephen Glick for accepting being reviewers for my PhD thesis manuscript. I would like to express my profound appreciation for all your constructive comments and questions that helped improve the quality of this research. I would also like to thank Prof. Dominique Jeulin, Prof. Anne Estrade and Prof. Alain Trouvé for accepting being jury members of this thesis.

I would also like to thank Dr. Serge Muller for welcoming me to his team and for giving me the opportunity to embark on this amazing research project. To all the current and former WHARé team mates (or “les WHARéens”): Giovanni, Pablo, Razvan, Sylvie, Laurence, Viviane, Maéva, thanking you all for making me feel like working at home! You are not only great colleagues from whom I can learn great things, but also great friends to me. To all-star GE PhD students: Anna, Ruben, Maissa, Aymeric, Emanuelle, Ketan, Lucas, thank you for your constant support and for all the fascinating moments that we spent together. I am most certainly that you will all achieve huge success with your PhD program! Also, a big thank you to all the GE colleagues that I have directly or indirectly worked with over these years: Rémy, Charlotte, Thomas Almecija, Cyrill, Régis, Romain, Maxime, Alyson, Thomas Benseghir, Antoine, Luc, Camille, Fabio, Hiba and so many others whose name I forgot to mention.

I would like to thank Prof. Nicolas Vayatis for welcoming me to the CMLA team. Thank all CMLA’s staff members, Véronique, Micheline, Sandra, Virginie and Alina for all your kindness and help over the years. Thank my fellow researchers and friends at CMLA: Cédric, Thibaut, Miguel, Lara, Carlo, Charles, Thomas, Samy and many others. You have made CMLA such a cool place to work and learn. I will always cherish the memories and the fun that we’ve had!

Finally, thank you my dearest friends Tuo, Chunyu and Kunyan for your help and encouragement, and thank you for the great friendship since 2011. To my mom, dad and my little sister: thank you for your love and warmth from across the ocean that helped me

get through all the questions and doubts during all the difficult moments. To Yan, my love, you have been by my side, supporting me with all you heart over these three and half years. I could not have done it without you.

Contents

Acknowledgement	2
Table of Contents	7
Abbreviations	8
Abstract	9
Résumé	13
1 Introduction and motivation	26
1.1 Breast cancer and x-ray breast imaging	26
1.1.1 Full field digital mammography	27
1.1.2 Digital breast tomosynthesis	28
1.2 Performance assessment and virtual clinical trials	32
1.3 Simulation of x-ray breast images	34
1.3.1 Random field breast texture models	34
1.3.2 Anthropomorphic breast phantoms	36
1.4 Assessment of clinical task performance using model observers	37
1.4.1 Mathematical model observers	39
1.5 Conclusion and manuscript organization	42
2 Three-dimensional random field breast texture models	45
2.1 General mathematical formulation	46
2.2 Power-law Gaussian random field	47
2.2.1 Mathematical formulation	47
2.2.2 Simulation example	49
2.2.3 Properties of 2D slice and projection	49
2.3 Shot-noise random field	53
2.3.1 Mathematical formulation	54
2.3.2 Simulation examples	55
2.3.3 Properties of 2D slice and projection	55
2.4 Conclusion and discussion	59

3 A novel 3D stochastic solid breast texture model for x-ray breast imaging	60
3.1 Input materials: clinical dedicated breast computerized tomography image database	61
3.1.1 Segmentation of fibroglandular tissue from bCT images	61
3.1.2 Analysis and observations of segmented bCT images	62
3.2 The new 3D stochastic geometric solid breast texture model	64
3.2.1 Sampling from the 3D solid breast texture model	66
3.2.2 Impact of irregular adipose compartment boundaries	69
3.2.3 Empirical determination of model parameters	70
3.3 Preliminary simulations using empirical model parameters	70
3.4 Psycho-physical realism assessment through two-alternative forced choice experiment	72
3.4.1 Experimental set-up	72
3.4.2 Experiment results and analysis	73
3.5 Removal of the small scale texture artifact	75
3.6 Conclusion and discussion	77
4 Statistical inference of medium scale 3D solid breast texture model parameters	80
4.1 Problem statement	81
4.1.1 Formulation of the medium scale texture model	81
4.1.2 The ground truth	81
4.2 Existing parameter inference methods for marked point processes	82
4.3 The method of inference from reconstruction	83
4.3.1 The reconstruction step: approximating segmented bCT data using ellipsoids	84
4.3.2 Inference step: fitting a marked point process model to reconstructed ellipsoids	88
4.3.3 Input segmented bCT volumes of interest	92
4.4 Results	92
4.4.1 The reconstruction step	93
4.4.2 The inference step	95
4.4.3 Simulations using fitted medium scale texture model parameters .	101
4.5 Conclusion and discussion	101
5 Statistical characteristics of x-ray breast images	104

5.1	Breast density	105
5.1.1	Background	105
5.1.2	Assessment of breast density	105
5.1.3	A new stereology-based algorithm for volumetric breast density assessment from DBT projections	107
5.1.4	Performance evaluation of the stereology-based volumetric breast density estimator	110
5.2	The β metric	112
5.2.1	State-of-the-art algorithm for β assessment from breast images: the Welch method	113
5.2.2	Analytical derivation of the bias and variance of the β estimator based on the Welch method	116
5.2.3	Numerical evaluation of β estimation accuracy	123
5.2.4	Discussion	127
5.3	Conclusion and discussion	128
6	Mathematical model observer based on the <i>a contrario</i> theory for microcalcification detection in 2D and 3D breast images	130
6.1	Background	131
6.2	Related work	132
6.3	Design of the <i>a contrario</i> observer for microcalcification detection in 2D and 3D breast imaging	133
6.3.1	Construction of multi-scale neighborhoods	134
6.3.2	Computation of the NFA image	135
6.3.3	Extension for μ_{calc} detection in 3D DBT slices	135
6.3.4	The <i>a contrario</i> observer for location-known-exactly tasks	135
6.4	Validation of the proposed <i>a contrario</i> observer	136
6.4.1	Theoretical proof of global false positive control for the proposed <i>a contrario</i> observer	137
6.4.2	Simulation experiment	140
6.5	Conclusion and discussion	141
7	Application to microcalcification detectability in full field digital mammography, digital breast tomosynthesis and synthetic two-dimensional images	144
7.1	Materials and methods	145
7.1.1	Simulation of test objects	145
7.1.2	Simulation of FFDM and DBT image acquisitions	146

7.1.3	Image processing	149
7.1.4	Microcalcification detection experiment	151
7.2	Results	156
7.2.1	Evaluation of the optimal channel parameters	156
7.2.2	Comparison of the performance of the ACO, the CHO and the human observer	156
7.2.3	Comparison of μ_{calc} detection performance in FFDM, DBT and S2D images	157
7.3	Conclusion and discussion	161
8	Conclusion and directions for future research	163
Appendix A	State-of-the-art anthropomorphic breast phantoms	169
Appendix B	Channelized Hotelling observer	171
Appendix C	Additonal results	175
Bibliography		180

Abbreviations

ACO	<i>A contrario</i> observer
AUC	Area under the receiver operating characteristic curve
bCT	Dedicated breast computerized tomography
BI-RADS	Breast imaging reporting and data system
CHO	Channelized Hotelling observer
CHO2	Two-dimensional channelized Hotelling observer
CHO3msa	Multi-slice channelized Hotelling observer of type A
CHO3msb	Multi-slice channelized Hotelling observer of type B
CHO3msc	Multi-slice channelized Hotelling observer of type C
CHO3vol	Volumetric channelized Hotelling observer
CLB	Clustered lumpy background
CNR	Contrast-to-noise ratio
CSF	Contrast-sensitivity function
DBT	Digital breast tomosynthesis
FFDM	Full field digital mammography
MCE	Minimum contrast estimation
MBDS	Multiple births, deaths and shifts algorithm
MPP	Marked point process
NPS	Noise power spectrum
NPW	Non-prewhitening matched filter
NPWE	Non-prewhitening matched filter with Eye filter
OLS	Ordinary least square
PCF	Pair correlation function
RJMCMC	Reversible jump Markov chain Monte Carlo
RMSE	Root mean square error
ROC	Receiver operating characteristic
ROI	Region of interest
S2D	Synthetic two-dimensional mammography
SID	Source-to-image distance
SIsoD	Source-to-rotation isocenter distance
SNR	Signal-to-noise ratio
stCSF	Spatio-temporal contrast-sensitivity function
μ_{calc}	Microcalcification
VOI	Volume of interest
VBD	Volumetric breast density
VCT	Virtual clinical trial

Abstract

Background

Breast cancer is one of the most common cancer among women in North America and Europe; one over eight women will be diagnosed with breast cancer during their lifetime. Today, full field digital mammography (FFDM), a 2D x-ray imaging technique, is the standard modality for screening and diagnosis of breast cancer. Due to its projection nature, the primary limitation of FFDM is the superimposition of tissue that can decrease visibility or even obscure lesions, and may mimic the appearance of malignancy. To partially alleviate this limitation and thus increase diagnostic accuracy, digital breast tomosynthesis (DBT) was recently introduced as an advancement of FFDM. From multiple projection images acquired over a limited angular range, a “quasi” 3D representation of the breast can be reconstructed with high spatial resolution parallel to the detector and limited spatial resolution perpendicular to the detector.

For optimal image quality assessment and time-efficient image review of DBT, continuous research has been performed on innovative image acquisition modes, dedicated 3D reconstruction algorithms, image processing and image review modes. DBT projections are typically reconstructed into a large number of closely spaced cross-sectional slices. Projection of the reconstructed slices into a synthetic 2D mammogram (S2D) allows for a 2D representation of the breast.

Traditionally, to assess the performance of new image acquisition modes, 3D reconstruction algorithms, or image processing and image review modes, clinical imaging trials are required. Key components of clinical trials include: the image acquisition system, the imaged test object, the clinical task, human observers and the task-based performance figure-of-merit(s) used to assess the clinical task. However, clinical trials are burdensome, expensive and may impose some risk with no direct benefit to the patient due to additional radiation exposure. Virtual clinical trials (VCT) aim to offer a more cost-effective and risk-free alternative. In VCT, clinical trials components are simulated to replace, or used in combination with, real components. VCT need to be designed to predict radiologist’s decision performance for a given clinical task.

To study the impact of different image acquisition modes, 3D reconstruction algorithms, or image processing and image review modes, active research is ongoing for the development of VCT components dedicated to the unique aspects of DBT. For these purposes, 3D breast texture models allowing to simulate realistic 3D and 2D images, and 3D mathematical observers dedicated to DBT are needed.

Thesis objectives

The goal of this PhD thesis was to extend the development of objective task-based VCT components to optimize 3D imaging systems and to assess its performance versus 2D FFDM. More specifically, we proposed:

- To develop and validate a 3D breast texture model allowing to simulate realistic 3D and 2D breast x-ray images representative for the population of women seen in clinical practice;
- To develop and validate a new 3D mathematical model observer based on the *a contrario* theory allowing to model microcalcification detection in DBT;
- To apply the developed VCT components to compare microcalcification detection performance in FFDM, DBT slices and synthetic 2D images processed with state-of-the-art 3D reconstruction and processing algorithms.

Methods and results

First, to understand the statistical realism of existing x-ray breast texture models, random field breast texture models were formulated in 3D under a unified discrete mathematical setting. Analytical results were obtained allowing to characterize the 3D models using their 2D projections or 2D slices. Our results suggest that, while their visual realism is limited, existing random field breast texture models have analytical advantages in terms of mathematical tractability. Breast volumetric density, a first-order statistical metric and the power spectral index (a.k.a. the β metric), a second order breast texture metric and their estimation methods were investigated. A new breast density estimation algorithm was developed using a model-based stereology method on DBT projection images. Preliminary scoring on synthetic breast texture images with known ground truth has shown that our method allows for accurate breast density estimation with absolute bias and standard deviation less than 3% of the ground truth value for a large range of breast densities. Regarding the β metric, we investigated the accuracy of its statistical estimation using the most commonly applied algorithm based on the Welch method. Analytical β estimation bias and variance were obtained in function of different algorithm parameterizations. The analytical results were also validated by numerical experiments using synthetic FFDM images, DBT reconstructed slices and dedicated breast CT (bCT) reconstructed slices with known ground truth. Our results show that the estimation of β based on the Welch method has limitations, mainly due to its non-negligible systematical estimation error, which is smaller for 3D breast images than for 2D images.

Based on an analysis of clinical bCT images, a new 3D solid breast texture model was developed for three breast densities types, varying from almost entirely adipose to extremely dense. The proposed model uses elements from stochastic geometry to simulate the morphology and distribution of the small and medium scale breast fibroglandular and adipose tissue. The proposed model was formulated as a marked point process (MPP). In first instance, prototype model parameters were empirically determined for all breast density types. Simulated 2D and 3D breast x-ray images using the prototype models appear to have high visual realism compared with clinical images. Further, a psycho-physical experiment involving human observers was performed to formally assess the realism of simulated DBT reconstructed slices using the prototype models for three types of breast density. Experiment results demonstrate that the prototype model with empirical parameters can simulate fairly realistic DBT reconstructed slices for less dense breast types. Finally, to obtain more objective model parameters, a statistical MPP parameter inference algorithm based on dynamic Monte Carlo methods was applied, using clinical bCT

reconstructed volumes as the ground truth. Objective model parameters were obtained for all three density types. Preliminary evaluation of the 2D and 3D breast images simulated from 3D texture volumes generated using new parameters shows fairly high visual realism. The breast tissue variability in new simulated images is larger than in simulated images using the prototype models with empirical parameters.

Next, a new *a contrario* mathematical observer (ACO) was developed for microcalcification (μ_{calc}) detection for 2D and 3D breast imaging. The proposed ACO is an extension of an existing ACO for mass lesion detection in FFDM developed as part of the PhD thesis of Bénédicte Grosjean in 2007. The here proposed ACO is based on pixel-wise statistical tests with a global control of the average number of false detections by a pre-defined upper bound. For each pixel, surrounding rectangular regions with various sizes were used to construct multiple null-hypotheses. Based on the null-hypotheses, the ACO then computes the number of false alarms (NFA) associated to the pixel using its pixel intensity value. The NFA was finally compared to a pre-defined upper bound, to determine whether the corresponding pixel intensity value is detected as abnormal. Theoretical justification was obtained proving that the average number of false detections of the proposed ACO was indeed bounded by an analytically tractable term having the same order of magnitude as the pre-defined upper bound. A simulation experiment was performed that demonstrated the global control of false detections for the proposed ACO.

Finally, the 3D breast solid texture model and the ACO were applied in a complete VCT experiment to compare μ_{calc} detection performance in FFDM, DBT and S2D images. Performance of the ACO was compared to 2D and 3D state-of-the-art channelized Hotelling observer (CHO) models. A preliminary analysis was also performed to compare the performance of the ACO and the CHO with a human observer. Different μ_{calc} sizes and attenuations as well as different background types were considered. Our study found that for all experimental conditions, the CHO models outperformed the ACO but no clear negative or positive correlation was found between the two types of model observers. The preliminary comparison to the human observer performance showed that both the ACO and the CHO tend to be positively correlated with human observer performance. For both ACO and CHO, the detection performance increases with μ_{calc} size at a given μ_{calc} attenuation characteristic. For both ACO and CHO, the detection performance increases with μ_{calc} attenuation characteristics at a given μ_{calc} size. These results are consistent for all background types. Under an ideal system topology, DBT tends to outperform FFDM and S2D for the CHO models. The ACO results do not allow for a clear ranking of the different modalities.

Conclusion

This thesis contributed two new components to the development of VCT. First, a new 3D solid breast texture model using stochastic geometric elements was developed. The new breast texture model is characterized by a high visual realism. Second, a new mathematical model observer based on the *a contrario* theory was developed. The *A contrario* theory represents an alternative holistic modeling of human vision in detection tasks. Contrary to existing ACOs for 2D FFDM, the new ACO uses localized image statistics, instead of the global image model to compute the decision variable.

These two developed components were then applied in a VCT on comparing μ_{calc} de-

tectability in FFDM, DBT and synthetic 2D images, demonstrating the possibility to design and perform complete VCT experiments using our developed new components.

We believe that stochastic geometry, introduced here for breast texture modeling, could be extended to develop more realistic models of the various components of breast anatomy. Our application of ACO also suggests future extensions of our proposed VCT components to other clinical tasks.

Publications

- Li, Z., Desolneux, A., Muller, S., de Carvalho, P. M., Carton, A. K. Comparison of microcalcification detectability in FFDM and DBT using a virtual clinical trial. Accepted for publication in *SPIE Medical Imaging* 2018 (paper number 10577-12). International Society for Optics and Photonics.
- Li, Z., Carton, A. K., Muller, S., Iordache, R., Desolneux, A. (2015, March). The effect of NPS calculation method on power-law coefficient estimation accuracy in breast texture modeling. In *SPIE Medical Imaging* (pp. 94161I). International Society for Optics and Photonics.
- Carton, A. K., de Carvalho, P. M., Li, Z., Dromain, C., Muller, S. (2015, March). Assessment of mass detection performance in contrast enhanced digital mammography. In *SPIE Medical Imaging* (pp. 941608). International Society for Optics and Photonics.
- Desolneux, A., Li, Z., Carton, A.K., Muller, S. (2015). Coupes et projections de modèles 3D de textures gaussiennes ou granulaires, *actes du GRETSI*.
- Li, Z., Desolneux, A., Muller, S., Carton, A. K. (2016, June). A novel 3D stochastic solid breast texture model for x-ray breast imaging. In *International Workshop on Digital Mammography* (pp. 660-667). Springer International Publishing.

Résumé

Le cancer du sein est l'un des cancers les plus fréquents chez la femme. Une femme sur huit va être affectée par le cancer du sein au cours de sa vie. La détection précoce de ce cancer à l'aide de l'imagerie par rayons X permet de réduire la mortalité dans la population dépistée. Dans les images du sein par rayons X, les signes radiologiques typiques associés aux tissus cancéreux sont des microcalcifications, des masses et des distorsions architecturales (Figure 1).

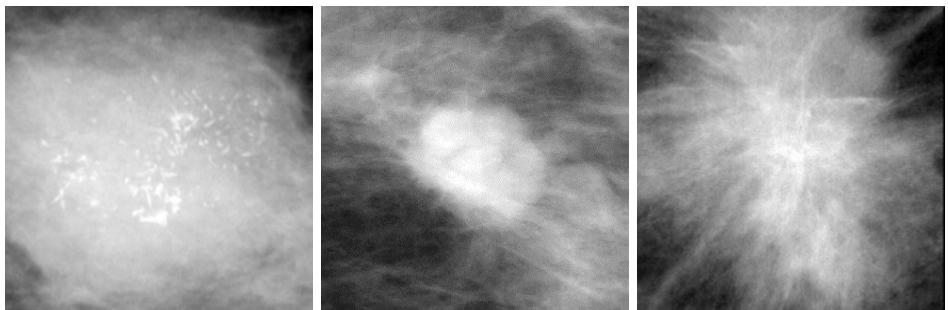


Figure 1: Exemples de microcalcifications (à gauche), d'une masse (au milieu) et d'une distortion architecturale (à droite) dans les images de mammographie.

La mammographie a montré son efficacité pour le dépistage et le diagnostic du cancer du sein. Une image de mammographie représente la projection bidimensionnelle des structures du sein. Le contraste de ces structures sur l'image dépend de l'atténuation aux rayons X des matériaux traversés et caractérisés par leurs coefficients d'atténuation linéaire. La limite principale de la mammographie est due à la superposition des tissus qui pourra potentiellement diminuer la visibilité des lésions, voire les masquer complètement. La superposition des tissus pourra aussi imiter des signes radiologiques dans les images, et par conséquent aboutir à un plus grand nombre de diagnostics erronés en faveur de la malignité (faux positifs). Au cours de ces dernières années les modalités d'imagerie tridimensionnelles, notamment la tomosynthèse numérique du sein, ont été développées pour réduire l'effet délétère de la superposition des tissus en mammographie. En tomosynthèse, une série d'images de projection est alors acquise sous différents angles sur une ouverture angulaire limitée (typiquement entre 15° et 50°). Un algorithme de reconstruction tomographique permet ensuite de fournir un ensemble de coupes parallèles au détecteur. L'ouverture angulaire limitée de la tomosynthèse ne permet cependant pas une reconstruction 3D parfaite. Néanmoins cette modalité offre beaucoup d'informations complémentaires comparée à la mammographie.

Durant ces dernières années, plusieurs études cliniques ont démontré que la tomosynthèse, utilisée seule ou en combinaison avec la mammographie permet d'augmenter le taux de la détection du cancer du sein, et aussi de diminuer significativement le taux de diagnostics faussement positifs. Néanmoins, des problématiques liées au développement de la tomosynthèse se présentent encore. Aujourd'hui il n'y a pas de consensus sur la performance de la détection des microcalcifications en tomosynthèse par rapport à la mammographie. Des études cliniques ont obtenu des résultats variés. Cette vari-

abilité peut être attribuée aux différences dans l'acquisition des images, à l'algorithme de reconstruction 3D et au mode de revue des images entre ces études. Une étude plus approfondie pour comprendre les raisons de ces différences est alors nécessaire. L'image de mammographie synthétique, créée à partir d'une acquisition en tomosynthèse, a été introduite pour potentiellement remplacer l'acquisition mammographique accompagnant un examen de tomosynthèse. Des études préliminaires ont démontré que la performance diagnostique de la mammographie synthétique était inférieure à celle de la mammographie, surtout en ce qui concerne la restitution des microcalcifications et des masses spiculées. Le développement des algorithmes pour générer des images de mammographie synthétique est en cours, et des études sont attendues pour comparer la performance de la détection et de la caractérisation des lésions à partir d'images de mammographie synthétique en comparaison avec les images de mammographie conventionnelle.

Aujourd'hui, les études cliniques sont nécessaires, afin de pouvoir comparer la performance clinique des différentes modalités d'imagerie, des différents algorithmes de reconstruction 3D et du traitement des images et des différentes façons de scruter des images en tomosynthèse. Les principaux composants d'une étude clinique incluent un système d'acquisition d'images, des objets test, une tâche clinique, des observateurs et enfin un facteur de mérite permettant de quantifier objectivement le succès de la tâche clinique. Bien que les études cliniques soient importantes, elles sont coûteuses et elles peuvent accroître les risques pour les patient(e)s, à cause de l'exposition aux rayons X. Les études cliniques virtuelles ont pour objectif d'offrir une approche alternative moins coûteuse, plus rapide à mettre en œuvre, et n'impliquant pas de sujets humains pour lesquels la nouvelle modalité d'imagerie peut ne pas présenter de bénéfice direct.

Dans une étude clinique virtuelle, certains des composants d'une étude clinique physique pourront être simulées pour les remplacer en partie (Figure 2). Le système d'acquisition d'images peut être simulé à l'aide d'un simulateur de rayons X qui permet de modéliser la formation des images de manière réaliste. Un patient ou une patiente peut être remplacé(e) par un fantôme réaliste de la partie d'anatomie concernée. Aussi, les images produites peuvent être lues par un observateur mathématique qui est un modèle mathématique ayant pour objectif de prédire la performance des humains pour une même tâche clinique. Concernant la tâche clinique, une tâche visuelle simplifiée peut être utilisée à condition que le facteur de mérite évalué dans la tâche simplifiée soit corrélé avec celui évalué dans la vraie tâche clinique.

Aujourd'hui, afin de pouvoir investiguer de manière efficace des problématiques liées au développement de la tomosynthèse, une recherche active est en cours pour développer des études cliniques virtuelles en prenant en compte les aspects spécifiques de la tomosynthèse. En particulier, un modèle de texture du sein qui permet de simuler des images réalistes du sein en 2D et en 3D, et un observateur mathématique dédié à la tomosynthèse sont attendus.

Cette thèse s'est réalisée en trois parties principales dans lesquelles nous avons développé deux nouveaux composants pour les études cliniques virtuelles, que nous avons appliqués pour implémenter une étude clinique virtuelle complète.

La première partie de cette thèse est consacrée au développement d'un modèle de texture du sein en 3D pour simuler des images du sein par rayons X avec un réalisme comparable à celui des images cliniques.

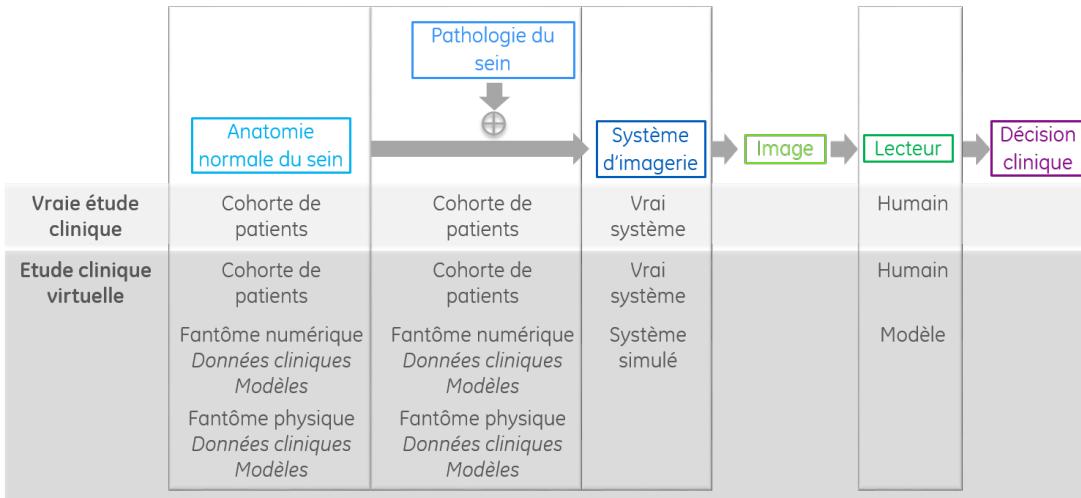


Figure 2: Différentes approches possibles pour réaliser une étude clinique virtuelle.

Les fantômes numériques anthropomorphiques du sein et les modèles de champs aléatoires représentent les deux approches principales pour la simulation des images par rayons X du sein en 2D et en 3D. Les fantômes numériques anthropomorphiques du sein sont des modèles 3D qui permettent de simuler les différentes structures anatomiques du sein à petite échelle, à échelle intermédiaire et à grande échelle. Les réalisations des fantômes numériques se présentent sous forme de volumes discrets ou de maillages. Un simulateur de rayons X est utilisé pour produire des images. Le réalisme et la variabilité anatomique des fantômes numériques varient selon les différentes approches de modélisation. La caractérisation analytique des fantômes numériques est en général difficile et peu de recherches ont été faites dans ce domaine. Les modèles de champs aléatoires, originellement proposés pour générer des images de texture du sein en 2D sont définis par des modèles mathématiques. De ce fait, ils admettent une caractérisation analytiquement accessible. Certaines études ont démontré la possibilité de déduire des métriques analytiques de performance dans des tâches de détection simplifiées avec des champs aléatoires grâce à leur caractérisation analytique. Toutefois, comparés avec des fantômes numériques, les champs aléatoires ont une capacité limitée pour simuler des images réalistes, surtout aux grandes échelles.

Des extensions des champs aléatoires en 3D ont aussi été proposées pour générer des textures anatomiques du sein sous forme de volumes. Ces volumes sont alors utilisés pour simuler des images par rayons X à l'aide d'un simulateur de rayons X. Le champ gaussien 3D en loi de puissance et le champ aléatoire granulaire 3D représentent les deux types de modèles les plus utilisés en imagerie du sein. Un champ gaussien en loi de puissance est stationnaire et isotrope, caractérisé par une densité spectrale Γ sous forme d'une loi de puissance $\Gamma(\xi) \propto \frac{1}{|\xi|^\beta}$. Ici β est appelé l'indice spectral et une valeur proche de 3 est typiquement utilisée dans la littérature pour simuler la texture volumique du sein. Un champ aléatoire granulaire consiste en une superposition de fonctions de noyau $g(\cdot, m)$ finies, centrées sur une réalisation d'un processus ponctuel simple $\Phi_C = \{y_i\}_{i \in \mathcal{I}}$ défini sur un domaine continu. Une fonction noyau $g(\cdot, m)$ est caractérisée par une marque m qui suit une loi notée \mathcal{M} . Un champ granulaire est entièrement déterminé par la donnée du processus ponctuel sous-jacent Φ_C , la fonction de noyau $g(\cdot, m)$ et la loi \mathcal{M} sur les marques. Un champ granulaire plus particulièrement utilisé dans le contexte de l'imagerie

du sein est celui appelé *clustered lumpy background*. Avec un *clustered lumpy background*, le processus Φ_C devient un processus de clustering de Thomas modifié, et les fonctions de noyau deviennent des fonctions elliptiques orientées aléatoirement dans l'espace 3D.

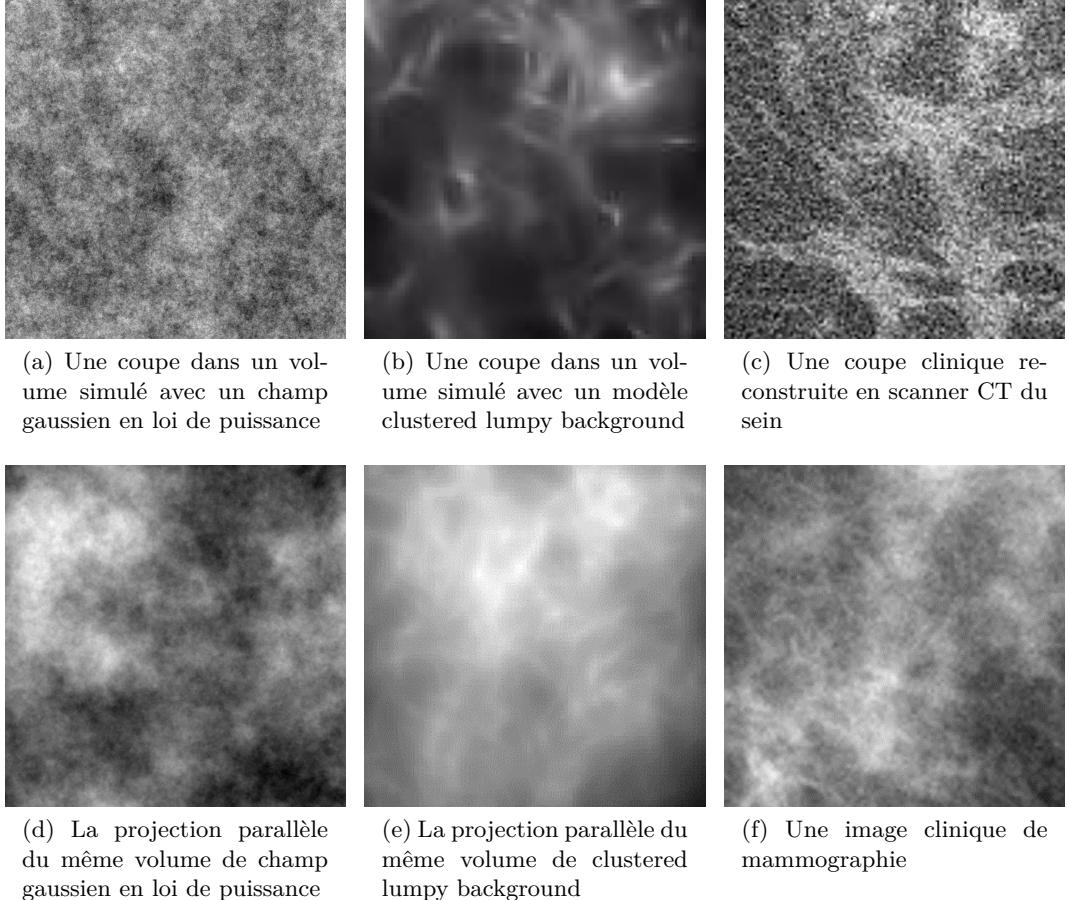


Figure 3: Exemples de simulations d'un champ gaussien en loi de puissance (indice spectral $\beta = 2.83$, à gauche) et d'un modèle clustered lumpy background (au milieu). Les volumes en 3D de taille $3.5 \text{ cm} \times 3.5 \text{ cm} \times 3.5 \text{ cm}$ ont été simulés, et leur coupe et projection parallèle sont présentées. Une coupe clinique reconstruite en scanner CT du sein (avec la permission de Prof. John M. Boone, UC Santa Davis) et une image clinique de mammographie sont aussi montrées pour références (à droite). Les images cliniques sont des régions d'intérêt de même taille que celle des images simulées.

La Figure 3 montre des exemples de simulation d'un champ gaussien en loi de puissance (indice spectral $\beta = 2.83$) et un modèle de clustered lumpy background. Une coupe clinique obtenue par reconstruction tomographique de données acquises à l'aide d'un tomodensitomètre (CT scanner) dédié à l'imagerie du sein et une image clinique de mammographie sont aussi montrées pour références (à droite). On remarque que le réalisme des images simulées est limité par rapport aux images cliniques, surtout lorsque l'on compare les coupes des volumes 3D simulés avec la coupe clinique reconstruite en scanner CT du sein.

Malgré une limite en réalisme, certaines propriétés statistiques des images créées avec le champ gaussien en loi de puissance et avec le clustered lumpy background correspondent

bien à celles des images cliniques. Cet aspect a conduit à de nombreuses applications des champs aléatoires pour investiguer la performance clinique de différentes modalités d'imagerie du sein.

Pour mieux comprendre les propriétés statistiques du champ gaussien en loi de puissance et du clustered lumpy background et pour caractériser analytiquement la projection parallèle et une coupe reconstruite (supposée “parfaite”) de ces deux modèles, nous avons effectué une étude théorique.

Les résultats obtenus montrent une stabilité analytique de la loi de la projection parallèle 2D et de la loi d'une coupe 2D d'un champ aléatoire discret f 3D, par rapport la loi originale de f . Plus précisément, la projection parallèle 2D et une coupe 2D de f appartiennent à la même classe de modèle que f , pour le champ gaussien en loi de puissance et pour le champ aléatoire granulaire. Quand f est un champ gaussien en loi de puissance, la projection parallèle de f est un champ gaussien en loi de puissance défini en 2D avec le même indice spectral β . Une coupe 2D de f est approximativement un champ gaussien en loi de puissance en 2D, avec un indice spectral égal à $\beta - 1$. Quand f est un clustered lumpy background, la projection 2D de f et une coupe 2D de f sont tous deux des champs aléatoires granulaires 2D, caractérisés par de nouveaux processus ponctuels sous-jacents et de nouvelles fonctions de noyau qui sont analytiquement accessibles.

Cette étude théorique démontre les avantages des modèles de champs aléatoire et avance notre compréhension de la simulation des images du sein par rayons X. Nous identifions avec cette étude le besoin de développer un modèle qui admet une caractérisation analytique et en même temps serait capable de simuler des images du sein réalistes comparées à celles rencontrées en clinique.

Pour répondre à ce besoin identifié, nous proposons ensuite un nouveau modèle de texture du sein en 3D basé sur la géométrie stochastique. Ce modèle est inspiré par une analyse de la morphologie et de la distribution des tissus fibroglandulaires (régions contrastées sur la Figure 3) et des tissus adipeux inter-glandulaires (régions moins contrastées sur la Figure 3). Nous avons observé qu'en utilisant une cube de taille $5\text{ cm} \times 5\text{ cm} \times 5\text{ cm}$ d'un volume reconstruit en scanner CT du sein avec les tissus fibroglandulaires segmentés, on arrivait à simuler des images mammographiques et des coupes reconstruites de tomosynthèse de très haut réalisme. Un exemple d'un cube segmenté est illustré en Figure 4.

Le nouveau modèle de texture utilise des éléments 3D de la géométrie stochastique pour modéliser la morphologie et la distribution des tissus fibroglandulaires ainsi que des compartiments de tissus adipeux inter-glandulaires à petite échelle et à échelle intermédiaire dans les volumes segmentés de scanner CT du sein. Nous avons proposé de modéliser les compartiments adipeux à échelle intermédiaire par des ellipsoïdes aléatoires avec chevauchement (Figure 4a). Nous avons aussi proposé de modéliser des irrégularités à petite échelle, observées sur les bords des compartiments adipeux dans des coupes de scanner CT segmentées (Figure 4b), et également dans des images histologiques du sein. Cette modélisation a été réalisée en remplaçant les bords lisses par des cellules de Voronoi de taille inférieure à 1 mm^3 . Le modèle de texture permet d'échantillonner des volumes binaires discrets. L'algorithme d'échantillonnage est illustré en Figure 5. Le modèle proposé peut être formulé comme un processus ponctuel marqué, admettant ainsi une caractérisation explicite.

Nous avons proposé tout d'abord trois jeux de paramètres du modèle pour simuler trois

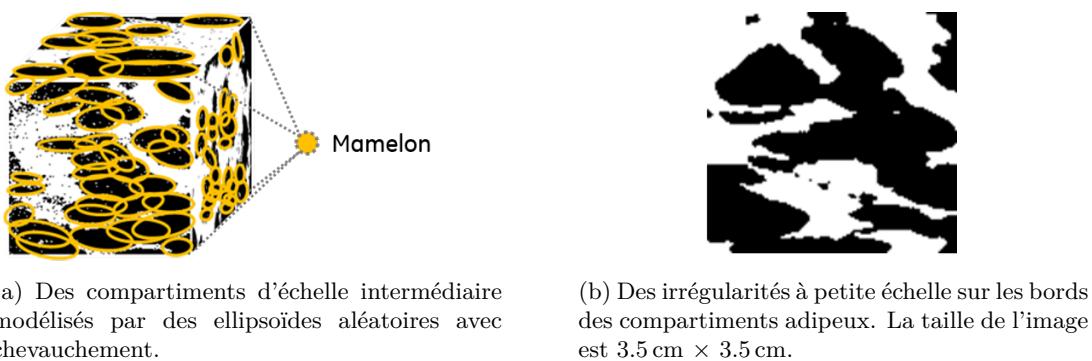


Figure 4: La morphologie et la distribution des tissus fibroglandulaires et des compartiments des tissus adipeux inter-glandulaires à petite échelle et à une échelle intermédiaire dans les volumes segmentés de scanner CT du sein.

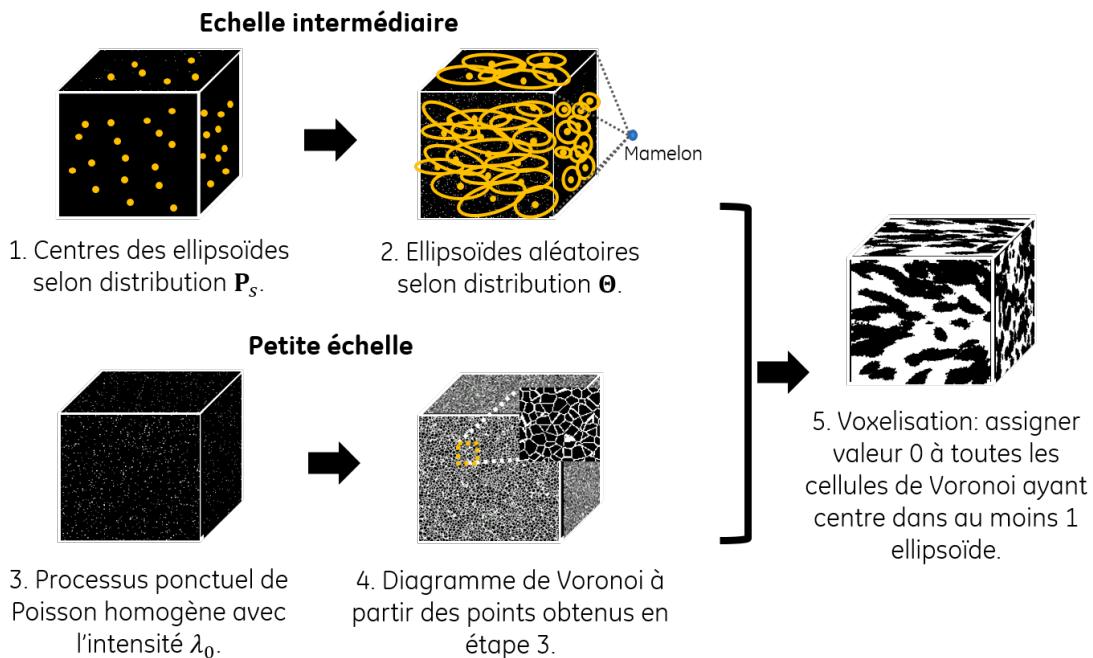


Figure 5: Différentes étapes pour échantillonner un volume binaire à partir du modèle 3D de texture proposé.

types différents de texture du sein, allant de l'extrêmement adipeux au dense hétérogène. Ces jeux de paramètres ont été empiriquement déterminés en comparant des images simulées en mammographie et en tomosynthèse avec des images cliniques. A partir de simulations préliminaires (Figure 6), on observe que le réalisme visuel des images de mammographie et des coupes reconstruites de tomosynthèse simulées est bonne en comparaison avec des images simulées obtenues à partir des volumes CT segmentés. Une validation formelle sous forme d'expériences psychophysiques a été effectuée en utilisant des coupes reconstruites en tomosynthèse simulées à partir des trois jeux de paramètres. Le résultat de cette expérience démontre que les images simulées à partir de notre modèle sont réalistes.

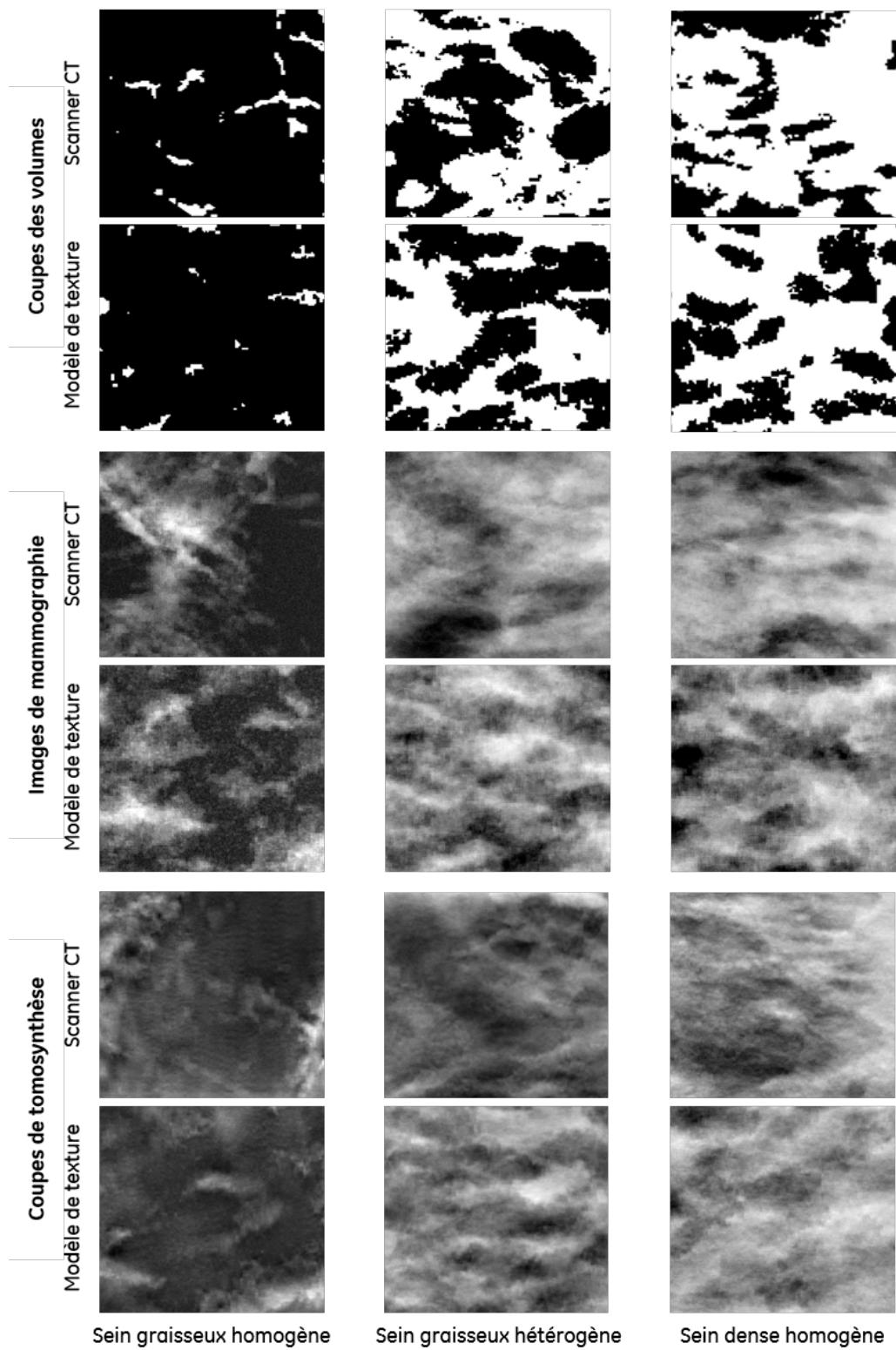


Figure 6: Exemples de régions d'intérêt de taille $3.5 \text{ cm} \times 3.5 \text{ cm}$ de coupes de volumes 3D, de projections mammographiques et de coupes reconstruites en tomosynthèse, simulées à partir de notre modèle 3D de texture avec des paramètres empiriques, et à partir des volumes CT segmentés.

Néanmoins, nous avons remarqué qu'avec les jeux de paramètres empiriques, la variabilité anatomique des images du sein simulées dans chaque catégorie était limitée. Ceci est partiellement lié à la modélisation des centres des ellipsoïdes par un processus de Poisson homogène qui n'introduit pas d'interaction entre les ellipsoïdes. Pour améliorer la variabilité dans chaque catégorie de simulation, nous nous sommes intéressés ensuite à inférer des paramètres du modèle de texture proposé par une méthode automatique basée sur les vraies données cliniques. Pour cela, nous avons proposé une méthode d'inférence statistique en utilisant des volumes CT cliniques du sein comme vérité terrain.

Tout d'abord, nous avons formulé le problème d'inférence comme une inférence d'un processus ponctuel marqué paramétrique. La difficulté ici provient de la non-observabilité des centres des ellipsoïdes, c'est-à-dire qu'à partir des volumes CT segmentés (Figure 4), on ne peut pas discerner les ellipsoïdes individuellement. Cette difficulté rend complexe l'application des approches classiques en inférence paramétrique. Pour résoudre les problèmes causés par cette non-observabilité, nous avons appliqué une méthode appelée «inférence par reconstruction». Cette méthode consiste en deux étapes. La première étape a pour objectif de reconstruire un ensemble d'ellipsoïdes à partir d'un volume CT segmenté. Nous avons formulé cette reconstruction comme un problème d'optimisation et un algorithme de type Monte Carlo dynamique a été appliqué pour trouver la solution optimale. La deuxième étape consiste à proposer un modèle de processus ponctuel marqué paramétrique et à l'estimer à partir de l'ensemble d'ellipsoïdes reconstruits. Pour cette étape, nous avons d'abord analysé la fonction de paire-corrélation empirique des centres des ellipsoïdes reconstruits ainsi que des statistiques empiriques des marques, c'est-à-dire les longueurs des demi-axes et les orientations des ellipsoïdes. Enfin, la méthode de minimum-contraste a été appliquée pour inférer un processus ponctuel sur les centres des ellipsoïdes. Des densités indépendantes ont été proposées pour modéliser les distributions des marques.

Une série de 12 volumes segmentés en scanner CT du sein a été utilisée comme données d'entrée de l'algorithme d'inférence proposé. Ceci nous a permis d'obtenir 12 jeux de paramètres objectifs. Des simulations préliminaires ont été réalisées en utilisant les nouveaux jeux de paramètres pour générer des images mammographiques et des coupes de tomosynthèse, partiellement illustrées sur la Figure 7. Visuellement, on observe que les images simulées contiennent plus de variabilité anatomique que celles simulées avec des jeux de paramètres empiriques. Ceci répond à l'objectif de cette inférence statistique basée sur des données cliniques.

Les images cliniques du sein par rayons X sont caractérisées par les caractéristiques statistiques liées à la variabilité anatomique du sein et aux bruits présents dans le système d'imagerie. Par conséquent, les caractéristiques statistiques ont un impact important sur la performance de la détection des lésions dans les images du sein. Une étude approfondie des caractéristiques statistiques des images du sein nous a permis de mieux comprendre, de comparer, et de quantifier les différences de performance de détection de lésions entre différentes modalités d'imagerie du sein. De plus, une telle étude nous a permis d'améliorer notre compréhension du réalisme statistique des images simulées.

Nous avons étudié tout d'abord la densité du sein, une caractéristique du premier ordre des images du sein. Elle s'exprime comme le ratio entre le volume de tissu fibroglandulaire et le volume du sein entier. Nous avons proposé une approche stéréologique de type "model-based" pour estimer la densité du sein à partir de projections en tomosynthèse. Nous avons commencé tout d'abord par modéliser la structure 3D du sein, composée de

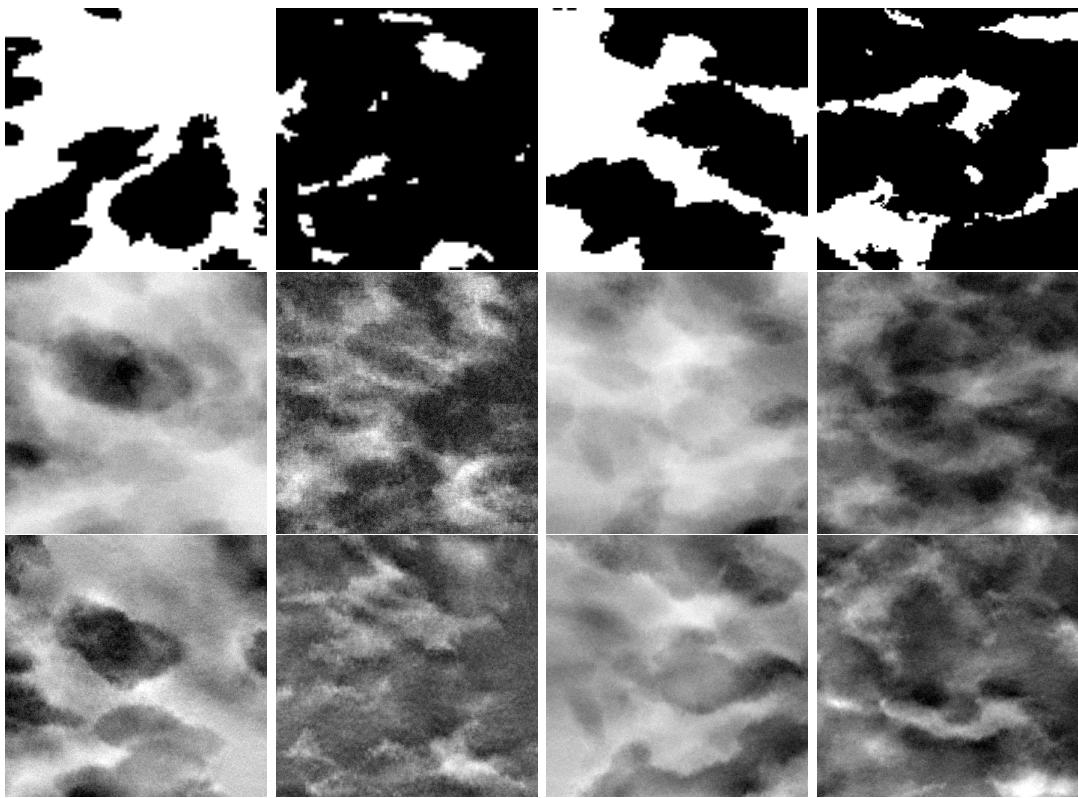


Figure 7: Exemples de régions d'intérêt de taille $3.5 \text{ cm} \times 3.5 \text{ cm}$ de coupes de volumes 3D (en haut), de projections mammographiques (au milieu) et de coupes reconstruites de tomosynthèse (en bas), simulées à partir de notre modèle 3D de texture avec des paramètres inférés.

tissus fibroglandulaires et de tissus adipeux, par un champ Booléen 3D. Le champ 3D a été supposé stationnaire et la distribution de sa projection a été supposé invariante aux différents angles de projection en tomosynthèse. Ensuite, nous avons modélisé le processus simplifié de la formation des images de projections en tomosynthèse pour en déduire la relation entre la fraction de volume glandulaire dans le volume original et l'intensité des pixels dans les projections. Grâce à cette relation, nous en avons déduit un estimateur analytique en utilisant l'ensemble des projections d'une acquisition en tomosynthèse. Une validation préliminaire a été réalisée pour tester notre méthode sur des données synthétiques où la vraie densité est connue. Le biais et l'écart-type de l'estimateur proposé sont inférieurs à 3% de la vérité terrain. Cette étude préliminaire démontre le potentiel d'utilisation de la méthode stéréologique pour mesurer la densité du sein, ainsi que les futures améliorations possibles pour ce premier estimateur stéréologique proposé.

Ensuite, nous avons étudié l'indice spectral β , une caractéristique du second ordre des images du sein. Il est défini comme l'indice spectral d'un champ gaussien en loi de puissance quand l'image du sein est modélisée par un tel champ aléatoire. Des études précédentes ont démontré que la valeur de β est liée à la performance de la détection des lésions dans les images du sein. Nous nous sommes intéressés au biais et à la variance d'un algorithme d'estimation de β couramment utilisé dans la littérature, basé sur la méthode de Welch. L'objectif est d'étudier l'impact des différents paramètres utilisés dans chaque étape de la

méthode de Welch sur le biais et la variance de l'estimateur de β . Nous avons commencé par donner une formulation mathématique précise pour chaque étape de la méthode de Welch. Puis, à l'aide de cette formulation, le biais et la variance analytique ont été dérivés en fonction des différents paramètres dans chaque étape. Les résultats analytiques ont été aussi validés par des expériences numériques en utilisant des images synthétisées simulant des images mammographiques, des coupes reconstruites de scanner CT du sein et des coupes reconstruites de tomosynthèse. Nos résultats montrent que le biais et la variance de l'estimateur de β basé sur la méthode de Welch ne sont pas négligeables et que leurs valeurs dépendent des paramètres dans chaque étape de l'estimation, ainsi que de la vraie valeur de β . Cette étude sur β souligne l'importance de bien choisir les paramètres de la méthode d'estimation de β dans les images cliniques du sein, surtout lorsque β est utilisé comme métrique pour valider le réalisme statistique des textures du sein simulées ou pour différencier la performance clinique de différentes modalités d'imagerie, ou de différents algorithmes de traitement d'images et de reconstruction 3D.

La deuxième partie de cette thèse s'intéresse au développement d'un observateur mathématique basé sur la théorie *a contrario* pour la détection des microcalcifications dans les images du sein en 2D et en 3D.

Dans une étude clinique virtuelle, un observateur mathématique peut être utilisé à la place d'un observateur humain pour évaluer la performance d'une tâche clinique (Figure 8).



Figure 8: Le processus typique pour évaluer la performance d'une tâche clinique.

En imagerie médicale, les tâches cliniques typiques sont: la tâche de détection des lésions (pour le dépistage) et la tâche de caractérisation des lésions (pour le diagnostic). Les deux types de tâche peuvent être formulés comme une classification binaire basée sur des tests statistiques. Dans une tâche de détection, l'hypothèse nulle H_0 est l'absence de lésion et l'hypothèse alternative H_1 est alors la présence de lésion. Mathématiquement, on a,

$$\begin{aligned} H_0 &: \mathbf{g} = \mathbf{b} + \mathbf{n}, \\ H_1 &: \mathbf{g} = \mathbf{b} + \mathbf{n} + \mathbf{s}, \end{aligned} \tag{1}$$

où \mathbf{g} représente une image, \mathbf{b} représente l'image dite «de fond», \mathbf{n} représente le bruit dans l'image et \mathbf{s} est la lésion (appelée aussi le signal). On utilise la convention que toutes les images ont la même taille et sont traitées comme des vecteurs colonnes.

Etant donnée une image g , un observateur mathématique spécifie une fonction de test, avec laquelle on en déduit une variable de décision λ , un scalaire qui donne le niveau de confiance sur la présence d'une lésion dans l'image g . Aujourd'hui les observateurs linéaires représentent l'état de l'art des observateurs mathématiques en imagerie médicale. Un observateur linéaire utilise un terme \mathbf{w} appelé le *template* pour calculer sa variable de décision λ . Ceci s'exprime mathématiquement comme

$$\lambda = \mathbf{w}^T g. \tag{2}$$

Classiquement, le template utilisé par les observateurs linéaires contient des informations sur la taille, le contraste et la forme du signal à détecter. Selon les différents modèles, ces informations sont soit exactement ou statistiquement connues, soit estimées à partir d'un

ensemble d'images d'apprentissage. L'observateur linéaire le plus étudié est le *channelized Hotelling observer* (CHO). Le CHO utilise des images "canalisées" en entrée. Une image canalisée est obtenu en faisant le produit scalaire de l'image avec des images de "canaux", qui sont souvent choisies pour représenter des caractéristiques spatiales ou fréquentielles des images en entrée. Typiquement, le CHO fonctionne en deux phases. La première phase, appelée la phase d'apprentissage, consiste à estimer un template à partir des images avec et sans le signal à détecter. Le template du CHO contient l'information qui permet de décorréliser le fond texturé des images en entrée. Ensuite, la deuxième phase, appelée la phase de test, consiste à appliquer l'équation (2) à une nouvelle image pour en déduire une variable de décision. Le CHO a été proposé pour les données en 2D et en 3D.

L'observateur *a contrario* est basé sur la théorie *a contrario*, une approche statistique pour la détection des objets dans des images. La théorie *a contrario* est inspirée par la psychophysiologie *Gestaliste*, plus particulièrement le principe de *Helmholtz*. L'idée générale du principe de *Helmholtz* est que certaines configurations d'objets peuvent être facilement perçues par le système visuel humain dans une réalisation aléatoire des objets. L'approche *a contrario* est une formulation mathématique pour quantifier comment une configuration observée dévie d'une réalisation aléatoire. Dans l'approche *a contrario*, une réalisation aléatoire est spécifiée par une hypothèse nulle H_0 appelée le *modèle naïf*, et le niveau de déviation est quantifié par une fonction appelée le nombre de fausses alarmes (NFA). Etant donnée un ensemble de variables aléatoires réelles $\{X_i\}_{i=1,\dots,N}$ ayant une distribution jointe spécifiée par le modèle naïf H_0 , le NFA est une fonction réelle qui satisfait,

$$\mathbb{E}(|\{i|\text{NFA}(i, X_i) \leq \epsilon\}|) \leq \epsilon, \quad (3)$$

pour $\epsilon \in \mathbb{R}^+$. L'ensemble $\{X_i\}_{i=1,\dots,N}$ est associé aux *mesures* réalisées sur les positions aléatoires dans l'image. L'équation (3) signifie que la famille de tests $\{\text{NFA}(i, X_i) \leq \epsilon\}_{i=1,\dots,N}$ sous l'hypothèse nulle H_0 admet un nombre moyen de faux positifs borné par ϵ .

Précédemment, l'approche *a contrario* a été appliquée pour modéliser la détection des lésions en mammographie. Les modèles *a contrario* existants utilisent des statistiques globales de l'image pour en déduire le NFA. Néanmoins, il a été montré dans plusieurs études psychophysiques que l'humain utilise plutôt des statistiques locales pour détecter les microcalcifications dans les images du sein. Pour mieux s'adapter à la détection des microcalcifications dans les images du sein, nous avons donc proposé un observateur *a contrario* avec un NFA calculé sur des statistiques locales de l'image. L'observateur *a contrario* que l'on propose calcule le NFA pour tous les pixels d'une image \mathbf{u} donnée en 2D. Au pixel x , nous avons commencé par construire un ensemble de voisinages multi-échelles \mathcal{V}_x (Figure 9). Un voisinage $V_x \in \mathcal{V}_x$ est défini par deux rectangles R_x^i et R_x^o centrés en x , avec la taille de R_x^o strictement supérieure que celle de R_x^i . Le voisinage V_x est alors obtenu comme $V_x = R_x^o \setminus R_x^i$. Nous avons fait varier la taille de R_x^i et de R_x^o . Ceci nous a permis de construire l'ensemble \mathcal{V}_x . Finalement, la formulation suivante de «NFA» a été proposée:

$$\text{NFA}(x) = |\Omega| |\mathcal{V}_x| \min_{V_x \in \mathcal{V}_x} \Phi \left(\frac{\mathbf{u}(x) - \mu_{V_x}}{\sigma_{V_x}} \right). \quad (4)$$

Pour déterminer si le pixel x appartient à une microcalcification, il suffit de comparer le $\text{NFA}(x)$ calculé avec une valeur de ϵ pré-définie.

Pour traiter le cas de la détection des microcalcifications sur des coupes reconstruites en tomosynthèse, une extension de l'observateur *a contrario* en 3D a été proposée. Puisque

la taille typique d'une microcalcification est plus petite que la distance entre deux coupes reconstruites en tomosynthèse, une microcalcification reconstruite est souvent focalisée sur une seule coupe reconstruite. Basé sur ce fait, nous avons proposé de calculer le NFA séparément pour chaque coupe reconstruite en tomosynthèse.

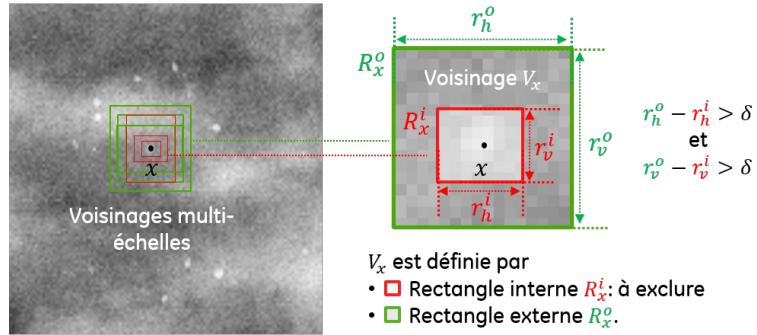


Figure 9: La construction d'un ensemble de voisnages multi-échelle \mathcal{V}_x autour d'un pixel x . Un voisinage $V_x \in \mathcal{V}_x$ est défini par deux rectangles R_x^i et R_x^o centrés en x , avec la taille de R_x^o strictement supérieure à celle de R_x^i .

Pour montrer que la formulation de «NFA» proposée dans l'équation (4) satisfait la définition donnée par l'équation (3), nous avons tout d'abord apporté une preuve théorique. On démontre que le «NFA» proposé permet de majorer le nombre de faux positifs moyen par une fonction analytique $\mathcal{C}(\epsilon)$ de ϵ , et que $\mathcal{C}(\epsilon)$ a le même ordre de grandeur que ϵ . Ce résultat théorique a été ensuite validé par une expérience numérique avec des images et des microcalcifications simulées. Le résultat de cette expérience est illustré sur la Figure 10.

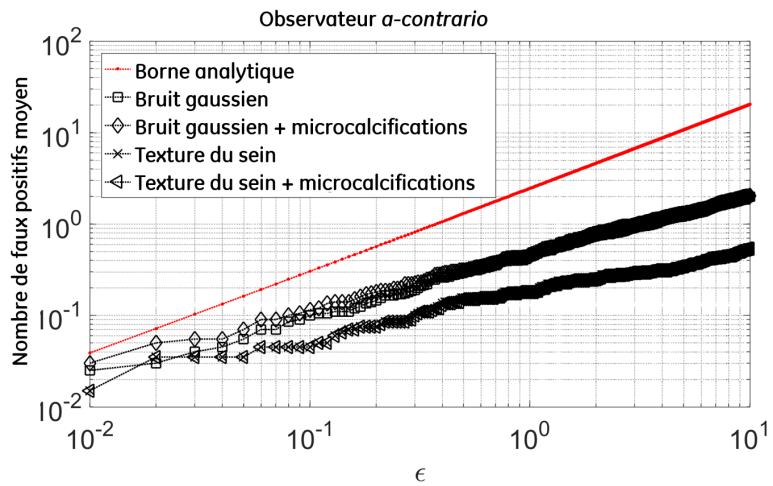


Figure 10: Résultat d'une expérience numérique avec des images et des microcalcifications simulées pour valider le fait que le nombre moyen de faux positifs de l'observateur *a contrario* proposé est borné par une fonction analytique de ϵ ayant le même ordre de grandeur que ϵ (en rouge).

Dans la dernière partie de cette thèse, nous avons implémenté une étude clinique virtuelle en utilisant les composants numériques développés précédemment, en particulier le modèle de texture 3D et l'observateur *a contrario*. Cette étude a trois objectifs. Premièrement, on

voudrait comparer la performance de l'observateur *a contrario* proposé et du CHO pour une tâche de détection. Deuxièmement, on souhaiterait comparer les performances de l'observateur *a contrario* et du CHO avec celle de l'humain. Troisièmement, on souhaiterait différencier la performance de la détection des microcalcifications en mammographie, en tomosynthèse et en mammographie synthétique. Une telle étude est importante pour pouvoir comprendre les avantages et les limitations de l'observateur *a contrario* développé, par rapport au CHO traditionnellement utilisé. Une telle étude a aussi une importance clinique car la performance de la détection des microcalcifications en mammographie, en tomosynthèse et à partir d'images de mammographie synthétiques reste aujourd'hui encore une problématique clinique clé.

Trois types de volumes du sein avec des densités différentes ont été simulés pour cette étude. Deux types de volumes texturés ont été simulés en utilisant notre nouveau modèle de texture en 3D. Le troisième type consiste en des volumes uniformes à valeur constante. Des microcalcifications de forme ronde, et de taille et atténuation variées ont été simulées. Puis, les acquisitions d'images en mammographie et en tomosynthèse ont été calculées à l'aide un simulateur d'imagerie par rayons X. La topologie et les caractéristiques physiques des systèmes d'imagerie ont été modélisées en utilisant un système commercial comme référence. Une méthode dédiée a été développée pour insérer une microcalcification dans le centre de chaque volume simulé. Les images de projection ont été traitées par les algorithmes implémentés dans le système commercial pour générer finalement des images mammographiques, des coupes reconstruites en tomosynthèse et des images de mammographie synthétiques. Les images générées ont été évaluées ensuite par l'observateur *a contrario* et le CHO. Les variables de décision de ces deux observateurs sous chaque condition expérimentale ont été utilisées pour tracer leur courbe ROC (receiver operating characteristic), et la surface sous la courbe ROC a été mesurée comme critère de performance. Les résultats finaux montrent que la performance de détection évaluée par l'observateur *a contrario* est inférieure à celle évaluée par le CHO pour toute condition expérimentale. Les performances de l'observateur *a contrario* et du CHO sont positivement corrélées avec celle de l'humain. Aussi, le résultat du CHO montre que la performance de la détection des microcalcifications à partir d'images de mammographie synthétiques est inférieure à celle que l'on obtient à partir d'images de mammographie et de tomosynthèse.

Les trois parties de cette thèse ont contribué au développement de composants pertinents pour la réalisation d'études cliniques virtuelles entièrement numériques. Le modèle de texture 3D proposé nous permet de simuler des images du sein par rayons X réalistes et avec une variabilité anatomique améliorée par rapport à l'état de l'art. L'observateur *a contrario* proposé introduit une approche alternative pour modéliser la perception humaine dans la détection des microcalcifications à partir des statistiques locales dans les images. Aussi, nous avons introduit l'utilisation de la géométrie stochastique et de la stéréologie pour des problématiques en imagerie du sein par rayons X. Les résultats encourageants démontrent le potentiel de la géométrie stochastique pour développer des modèles anatomiques du sein avancés.

Chapter 1

Introduction and motivation

1.1 Breast cancer and x-ray breast imaging

Breast cancer is one of the most common cancers among women in North America and Europe. For instance, in France, breast cancer was the main cause of cancer deaths among women in 2016 [35]. Breast cancer alone accounts for 25% of all cancer cases and 15% of all cancer deaths among women worldwide in 2012 [190].

A number of factors are associated with an increased risk of breast cancer [125]. They include being a woman, getting older, having genetic mutations (BRCA1 & BRCA2), having a family or personal history of breast cancer, having dense breasts, delivering a first pregnancy at term after the age of 30, first menarche occurring at an early age, menopause starting late, smoking and having an increased alcohol consumption, undergoing hormone replacement therapy and having a passive life style.

Studies have shown that early detection is key to improve breast cancer survival rate. When breast cancer is diagnosed at an early stage, the patient can be cured in nine over ten cases [180]. Imaging techniques play an important role in breast cancer detection, diagnosis and therapy follow-up. For screening and diagnosis of the breast cancer, x-ray breast imaging techniques, including *full field digital mammography* and *digital breast tomosynthesis* are the most clinically used modalities.

In x-ray breast images, radiologists typically look for radiological signs that indicate breast cancers. Typical radiological signs are *microcalcifications*, *masses* and *architectural distortions*.

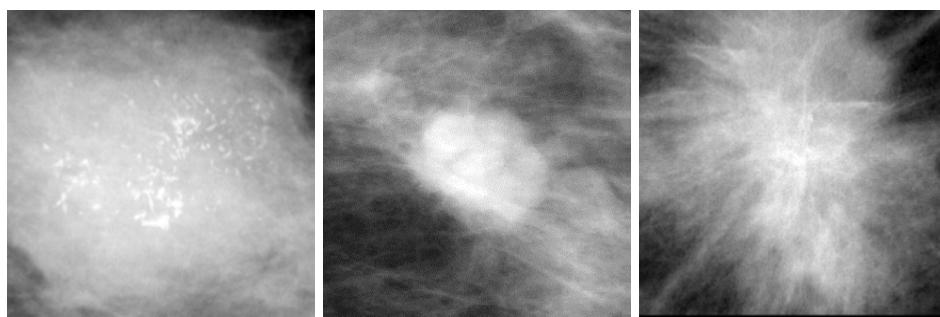


Figure 1.1: Examples of microcalcifications (left), mass (middle) and architectural distortion (right) in x-ray breast images. Courtesy of [78].

Microcalcifications are composed of calcium-containing minerals embedded in a protein matrix. They have x-ray attenuation coefficients substantially higher than breast tissue and therefore appear as bright spots in the x-ray images. The diameter of a microcalcification is typically inferior to 1 mm [86]. While microcalcifications are not always associated with malignant lesions, they might be an early sign of breast cancer when grouped in a

cluster and presenting some specific morphological and densitometric characteristics. A mass is a radiological finding demonstrating an increased density versus the surrounding tissue. Its size varies typically between few millimeters and few centimeters. Its shape, margin and radiographic density are analyzed by the radiologist to differentiate malignant from benign lesions. An architectural distortion describes the distortion of the normal breast tissue (often in radial pattern) with no definitely visible mass [72]. Figure 1.1 shows examples of microcalcifications, mass and architectural distortion in x-ray breast images.

1.1.1 Full field digital mammography

Full field digital mammography (FFDM) is a two-dimensional x-ray imaging technique dedicated to the imaging of the breast.

Figure 1.2a demonstrates an FFDM imaging system and a schematic illustrating the image acquisition. During the FFDM image acquisition, the breast is positioned on a support table. Then, the breast is mechanically deformed so as to reduce the breast thickness to a near uniform thickness. Breast compression is important to improve image quality and decrease radiation dose. By immobilizing the breast, the probability of patient motion image artifacts is reduced. By decreasing the breast thickness, scattered radiation is also reduced resulting in improved image contrast. By compressing the breast, breast tissue is spread out over a larger area, reducing the impact of confounding or superimposed breast tissue. X rays are emitted from an x-ray tube. A fraction of them are transmitted through or absorbed by the breast tissue. The x rays reaching the digital detector are converted in electronic charges that are read to deliver a digital image. The pixel intensity of this projection image of the breast reflects the breast tissue attenuation along the x-ray trajectory.

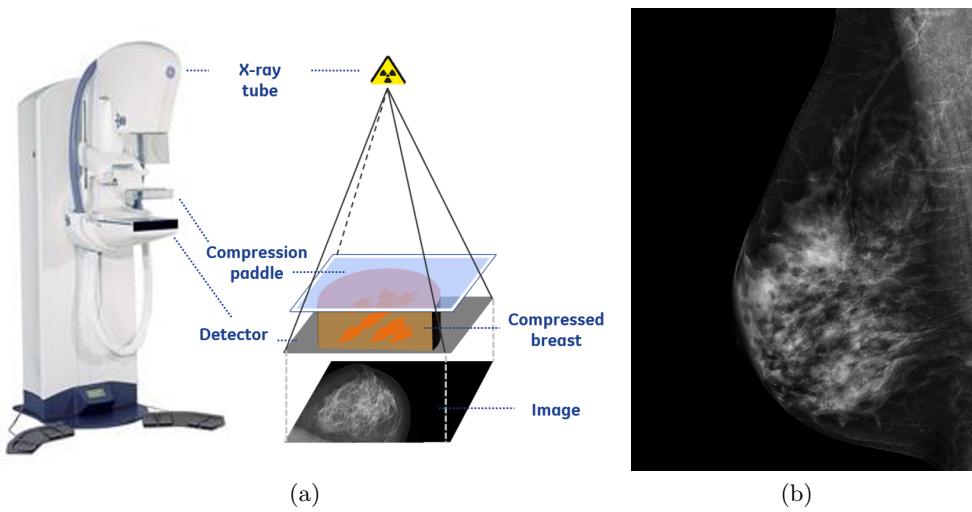


Figure 1.2: The GE Senographe Essential FFDM imaging system and a schematic of an FFDM image acquisition (left). Example of an FFDM image from the same imaging system (right).

The formation of an FFDM image on the detector is theoretically expressed by the *Beer-Lambert law*:

Lambert law, which explicitly formulates the attenuation of x-ray photons as a function of the thickness and the linear attenuation coefficient of the breast tissue.

Assuming a monochromatic x-ray spectrum, the *Beer-Lambert* law is mathematically expressed as:

$$I(x, y) = I_0 \exp \left(- \int_{r \in L(x, y)} \mu(r) dr \right), \quad (1.1)$$

where $L(x, y)$ represents the x-ray beam path between the x-ray source and the position (x, y) on the detector. Here I_0 is the x-ray photon intensity incident upon the breast, $I(x, y)$ is the x-ray photon intensity at the position (x, y) on the detector and μ represents the linear attenuation coefficient of the breast tissue material. Breast tissue is mainly composed of *fibroglandular tissue* and *adipose tissue*. They have different x-ray attenuation properties. In consequence, a contrasted image is obtained at the detector reflecting pixel-wised tissue composition of the imaged breast. Figure 1.2b demonstrates an example of an FFDM image (often called a mammogram) acquired from the system in Figure 1.2a.

Today, FFDM is widely acknowledged to be the most effective way for breast cancer screening. Systematic population screening using FFDM has been proved to help reducing breast cancer mortality over an asymptomatic population, when performed in the context of high-standard programs on the target population, typically women aged between 40 and 74 [119]. In France, since 2003 women aged between 50 and 74 were systematically invited to get a screening FFDM exam every two years [154].

Although the cost-effectiveness of FFDM is widely recognized, its sensitivity varies from 47.8% for the densest breasts to 98% for the least dense breasts (77.6% in average) [102]. The primary limitation of FFDM sensitivity is due to the overlap of breast tissue. Overlapping tissue can decrease the visibility of (malignant) lesions or even completely obscure them. This effect is higher in women with denser breasts and results in reduced cancer detection rate in FFDM [102]. Overlapping breast tissue may also mimic radiological findings (“ghost lesions”), resulting in false positive diagnosis. This may lead to unnecessary recall of women for further diagnostic tests, leading to an increase in patient anxiety and cost of healthcare [128].

FFDM is also used for diagnostic purpose. A woman typically gets a diagnostic mammogram if she has previously been identified with an abnormal tissue area detected during a breast palpation or in a screening mammogram, or if she was previously treated for breast cancer [49]. The purpose of a diagnostic mammogram is not to determine whether the identified potential abnormality actually indicates a cancer, but it is to help determining if the symptoms found in a FFDM exam are indicative of the presence of a cancer. In diagnostic FFDM, additional x-ray images are taken, providing views of the breast from multiple incidences. Magnification views may be acquired in a specific area of the breast where there is a suspicion of abnormality.

1.1.2 Digital breast tomosynthesis

Digital breast tomosynthesis (DBT) was recently introduced to partially alleviate the issue of breast tissue overlap in FFDM. It is a technique for breast imaging based on the FFDM platform. Figure 1.2a demonstrates a DBT imaging system and a schematic illustrating the image acquisition. During the DBT acquisition, the breast is positioned in a similar

way as during an FFDM acquisition. Then, multiple projection images of the breast are acquired whereby the x-ray tube moves along an arc around the breast within a limited scan angle (typically from 15° to 50°) while the detector remains stationary [174].

DBT projection images are typically processed by a reconstruction algorithm to generate a large number of closely spaced cross-sectional images (or slices), allowing for a 3D representation of the imaged breast. As an alternative, thicker slabs may be computed by combining adjacent slices. This provides the radiologists with a more manageable number of images to read. Projection of the reconstructed slices into a synthetic 2D mammogram allows for a 2D representation of the breast, with no need for an additional x-ray exposure to get a 2D mammogram.

Since DBT projections are acquired over a limited angular range, the spatial resolution of DBT slices in the depth direction, *i.e.* the direction perpendicular to the plane of the projection images, is limited. The in-plane resolution of the DBT slices, which is limited by the detector resolution, is similar to the resolution of the projection images. Figure 1.3b shows a typical reconstructed slice produced by the same system.

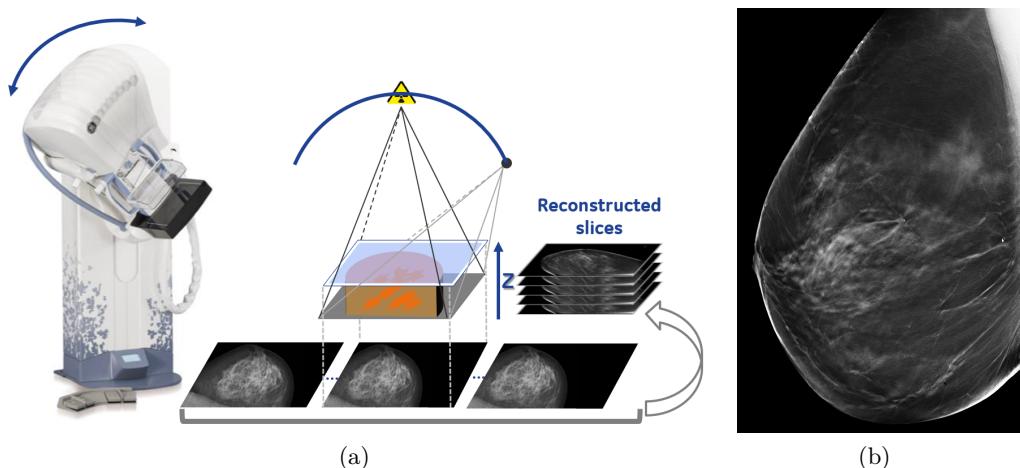


Figure 1.3: The GE SenoClaire DBT imaging system and a schematic of a DBT acquisition (left). An example of tomosynthesis reconstructed slice from the same imaging system (right).

Since the introduction of DBT was to partially alleviate the limitations of FFDM primarily due to breast tissue overlap, early expectations of DBT were to improve lesion conspicuity, to improve differentiation between benign and malignant signs and to reduce false positives. Also due to the 3D nature of DBT reconstructed data set, an improved lesion localization was expected. Thus overall expectation of DBT was to improve clinical performance, especially in dense breasts.

During the last decade, several large population clinical trials have been performed to assess the diagnostic value of DBT in terms of cancer detection. Either DBT was used in stand-alone and its performance was compared to FFDM, or DBT was combined with FFDM and the performance was compared to FFDM alone. Four examples of such clinical trials are:

- The OSLO trial: OSLO Tomosynthesis Screening Trial [185] [184]

- The STORM trial: Screening with Tomosynthesis OR Mammography trial [47]
- The MBTST trial: Malmö Breast Tomosynthesis Screening Trial [106]
- The Reggio Emilia randomized clinical trials [92]

Set-ups and main findings of the four clinical trials are described in Table 1.1. These trials have demonstrated that the use of DBT in stand-alone or DBT combined with FFDM in a screening environment resulted in a significantly higher cancer detection rate and a lower false positive recall rate compared to FFDM in stand-alone. Higher cancer detection rate can be attributed to a better visibility of breast tissue distortions (which are signs often related to invasive cancers), masses and architectural distortions as well as to the reduced tissue overlap.

Trial	Trial type	Comparison	Number of patients	Cancer detection rate	False positive recall rate
OSLO	Prospective screening trial	2-view DBT + 2-view FFDM versus 2-view FFDM	18000	27% increase	13% decrease
STORM	Prospective screening trial	2-view DBT versus 2-view FFDM	7982	34% increase	17% decrease
MBTST	Prospective screening trial	1-view DBT versus 2-view FFDM	15000 (so far only results of first half 7500 women)	40% increase	46% decrease
Reggio Emilia	Randomized clinical trial	2-view DBT + 2-view FFDM versus 2-view FFDM	9782 (FFDM) and 9623 (FFDM + DBT)	80% increase	Not reported

Table 1.1: Set-ups and findings of the four trials to assess the diagnostic value of DBT for cancer detection. The findings of the four clinical trials are compared to FFDM in stand-alone.

Some other studies have also shown that DBT improves the characterization of masses and architectural distortions. This holds particularly for dense breasts [121] [34]. Another study has shown that DBT allows for a more accurate lesion localization [2].

Today, several issues are remaining, thereby providing opportunities for further research.

- *Detection performance of microcalcifications in DBT.*

Today there is no consensus on microcalcification detection performance with DBT when used in stand-alone compared with FFDM. One of the primary reasons for performing FFDM in combination with DBT is the concern that microcalcification clusters may not be as readily and easily detected in DBT images. Some studies focusing on

microcalcification detection performance found that the detection was better with DBT than with FFDM [104] [56], while some other studies found that the detection performance was equal between DBT and FFDM [129] [159] or even was worse with DBT compared with FFDM [187]. Differences might be attributed to several aspects. The DBT image acquisition techniques in these studies were different. DBT image acquisition can be designed with a step-and-shoot or continuous tube motion, which may impact the level of blur in projections. The acquisition time might also be different, with a potential impact on patient motion induced blur in the projected views. The image reconstruction algorithms and the image review modes were also different in these studies. A more profound understanding of the contribution of the image acquisition system design, image reconstruction algorithms and image review modes on microcalcification detection and characterization performance is needed.

- *Efficiency of the image review process.*

Combining DBT and FFDM in a single exam increases image review time of the radiologist due to the large number of images to be reviewed. Previous studies reported doubled review time for DBT compared with FFDM (90 s versus 45 s in [183] and 125 s versus 67 s in [198]). This increase in review time may negatively impact the clinical workflow and increase the healthcare cost. The longer reading time needs further assessment so that efficient workflow tools can be designed to help the radiologists find the suspicious signs much more quickly.

- *Synthetic 2D mammograms from DBT acquisitions.*

One method that has been proposed to make the DBT clinical workflow more efficient and reduce radiation dose of the exam was the introduction of synthetic 2D mammograms as a replacement for the FFDM images in a DBT exam composed of both DBT and FFDM images (sometimes called "combo mode"). Synthetic 2D mammograms are created from DBT reconstructed 3D data sets. The idea is that a radiologist first reviews the synthetic 2D mammograms. In case the radiologist finds suspicious signs of lesions, she/he then reviews the DBT reconstructed slices or slabs. Synthetic 2D mammograms might also replace the acquisition of standard mammograms in a DBT exam, and thus avoid double x-ray exposure. For this to become a clinical reality, the clinical performance of the synthetic 2D mammograms complementing DBT in place of a separate 2D FFDM acquisition needs to be proven. In a preliminary study by Gur *et al.* [81], testing an early method for generating synthetic 2D mammograms, radiologist's performance has shown to be lower when using synthetic 2D mammograms combined with DBT rather than FFDM in combination with DBT. More recently introduced algorithms for generating synthetic mammograms have shown improvements. In the study of Skaane *et al.* [183], the overall cancer detection performance when using a newer version of synthetic 2D mammograms combined with DBT was found comparable to when using the standard FFDM images combined with DBT. Clinical investigations have not evaluated the diagnostic performance of synthetic mammography and FFDM in stand-alone. However, carefully conducted phantoms study have shown the inferiority of synthetic 2D compared to FFDM for the detection of small microcalcifications [142] [90]. To further improve the diagnostic value of synthetic 2D mammograms, the development of more efficient algorithms is under active research. More comprehensive studies to assess the performance of synthetic 2D mammograms versus standard mammograms are needed.

In this dissertation, we focus on comparing the microcalcification detection performance in FFDM images, DBT reconstructed slices and synthetic 2D mammograms. The impact of the image review time on microcalcification detectability is out-of-scope.

1.2 Performance assessment and virtual clinical trials

Clinical trials are carefully conducted human research studies aiming to assess the safety and effectiveness of new medical devices (including imaging systems and image processing algorithms). Clinical trials are often required for new breast x-ray imaging devices to be approved by regulatory agencies for a particular intended use before they can be marketed. Clinical trials are requested when laboratory measurements are not sufficient to assess the efficiency and safety of the device.

The set-up of clinical trials is often very burdensome and expensive. First, an appropriate clinical research site needs to be found to conduct the study. A clinical protocol, a data management plan and a statistical data analysis methodology are required to define how the study will be executed. Before starting the clinical trials, the clinical site may need specific device training to execute the research protocol. During the clinical trials, patients are closely monitored. An informed consent process needs to be consolidated aiming to protect the ethical standards; people considering a clinical trial should be given all the facts about a clinical trial before they decide to take part. A legal contract between the medical device provider (company) and the research investigation site is required. An independent ethical committee review and approval is required to insure that the patients' rights are protected throughout the study. The clinical trial needs to be registered so patients become aware of the research study and learn about how to participate. Regulatory agency notifications are required to conduct the study.

In addition, clinical trials may also impose additional x-ray exposures for the patients. This can lead to higher risk for the patients due to the extra exposures to ionizing radiation, while without the guarantee of any additional benefits. This occurs in trials whereby the patients are imaged with the new device only, or with the new device in combination with the device representing the standard of care.

There are also constraints in the budget and the timing of clinical trials. Due to these constraints, the trials are often dimensioned to demonstrate one or a selected set of intended uses. The timing of a clinical trial may also become excessive when follow-up is required to assess the patient state (benign/malignant). In addition, the ground-truth cancer information is only available in histological images.

Virtual clinical trials (VCT) may offer a cost-effective and fast alternative to clinical trials [118] [100]. Figure 1.4 shows different components for real clinical trials and for VCT. In VCT, the key components of clinical trials, including the healthy and disease patient cohort, the imaging system and the radiologist may be entirely or partially replaced by models. The imaging system can be modeled using a software x-ray simulator that realistically models the image formation process. The human test patients can be interchanged by physical or digital phantoms that realistically model the relevant human anatomy. Images can be scored by model observers which are mathematical models that aim to predict human performance for a particular visual task. Simplified visual tasks can also be modeled aiming to be correlated to real clinical tasks. Finally, the figures-of-merit for real

clinical trials can be used to assess the observer performance in VCT.

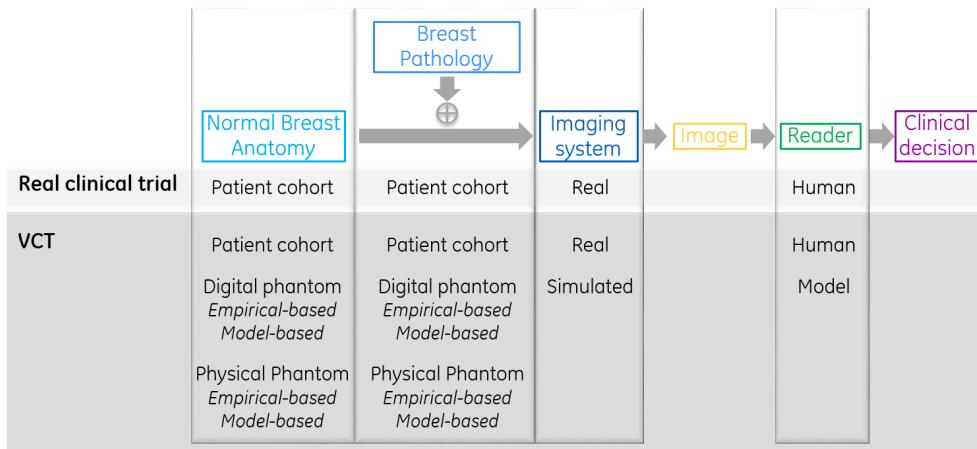


Figure 1.4: Components for real clinical trials and for VCT. Real and simulated components can be combined in different ways, representing different design options for VCT implementations.

To set up a clinical trial, a clinical protocol needs to be defined. Similarly, to set up a VCT study, a virtual clinical protocol needs to be defined. The constraints of VCT tools and real clinical trial components are very different. To design a VCT that will answer clinical questions, a profound understanding of the VCT tool constraints is key. Several aspects need to be considered. First, the level of realism of the VCT tools required to design an effective VCT needs to be well understood. For instance, today most computational imaging systems can model the key physical processes occurring during image formation. However, modeling of some elements in the acquisition chain with high precision (f.e. an anti-scatter grid) is still challenging. Therefore, depending on the VCT specifications, a design with a real or a computational imaging system may be preferred. Second, the impact of the dimension of the VCT on the choice of available VCT tools needs to be considered. For instance, today, a large number of physical anthropomorphic phantoms with different configurations is typically not available; while creating a large number of digital breast phantoms with different configurations is more straightforward, provided that a digital phantom model is available. Therefore, to demonstrate clinical feasibility, the use of physical anthropomorphic phantoms might be sufficient; while for the design of a large VCT, digital phantoms are preferred. Finally, the impact of the available time frame for a VCT on the choice of VCT tools needs to be considered. For instance, x-ray image simulation, including Monte Carlo scatter simulation, may be very time-consuming. The trade-off between the impact of the effectiveness of the VCT when including such time-consuming components and the available time frame need to be well understood.

In this thesis we will investigate the microcalcification detection performance in FFDM images, DBT reconstructed slices and synthetic 2D mammograms using a VCT approach. To allow for such a comparison, a large number of random realizations of different textured background types is required. Therefore, we decided to consider digital models of breast anatomy, pathology and the x-ray imaging system. To review the images, we will propose a new model observer for DBT and 2D breast imaging. Its performance will be tested with state-of-the-art model observers and a human observer as reference.

In the following sections, we will give a comprehensive overview of state-of-the-art anthropomorphic breast phantoms, texture models and mathematical model observers that have been used in VCT.

1.3 Simulation of x-ray breast images

Over the last two decades, various models have been proposed for 2D and 3D breast image simulation. Based on different construction methodologies, we can generally categorize existing models into two categories: *random field breast texture models* and *anthropomorphic breast phantom models*.

1.3.1 Random field breast texture models

Random field breast texture models are based on the mathematical random field theory [196]. They are characterized by their mathematical parametric formulations. Originally 2D random field breast texture models were proposed to model and synthesize FFDM textures [31] [24] [27] [39]. Two-dimensional models are used to model or simulate directly the signal intensities of the perceived FFDM images. With the extensive development of 3D breast breast imaging, 3D extensions of the 2D models have also been proposed [26] [111] [162]. Three-dimensional random field models have been used to generate 3D volumes aiming to mimic the background anatomical breast texture. The generated 3D volumes can further be projected using a realistic x-ray imaging simulator to create x-ray breast images.

Commonly applied random field breast texture models are:

- *The power-law Gaussian random field* model. It is a stationary and isotropic Gaussian random field model defined by a inverse power-law shaped spectral density. The 2D power-law Gaussian random field was first introduced to model stationary texture regions of FFDM images [31]. The 3D power-law Gaussian random field has also been adapted to simulate both 2D and 3D breast images, either as a stand-alone model [162] or in conjunction with anthropomorphic breast phantoms [109]. The power-law Gaussian random field is uniquely determined by a single parameter referred to as the power spectral index, often denoted as β .
- The *clustered lumpy background* model. It belongs to the family of *shot-noise random field* models, which are mathematically defined as the superposition of kernel functions with centers following some statistical distribution. The first generation of 2D clustered lumpy background was first introduced to synthesize 2D FFDM image textures [27]. Later-on a second generation of 2D clustered lumpy background was proposed with optimization of the model realism using a superposition of two clustered lumpy background realizations with parameters optimized with respect to clinical FFDM images [39]. Three-dimensional adaptation of the clustered lumpy background has also been investigated [111].

Figure 1.5 shows $3.5\text{ cm} \times 3.5\text{ cm}$ regions of interest (ROI) of an image simulated from a 2D power-law Gaussian random field, an image simulated from a 2D clustered lumpy background and a clinical mammogram. Figure 1.6 shows $3.5\text{ cm} \times 3.5\text{ cm}$ ROIs of a slice from a volume simulated using a 3D power-law Gaussian random field, a volume simulated

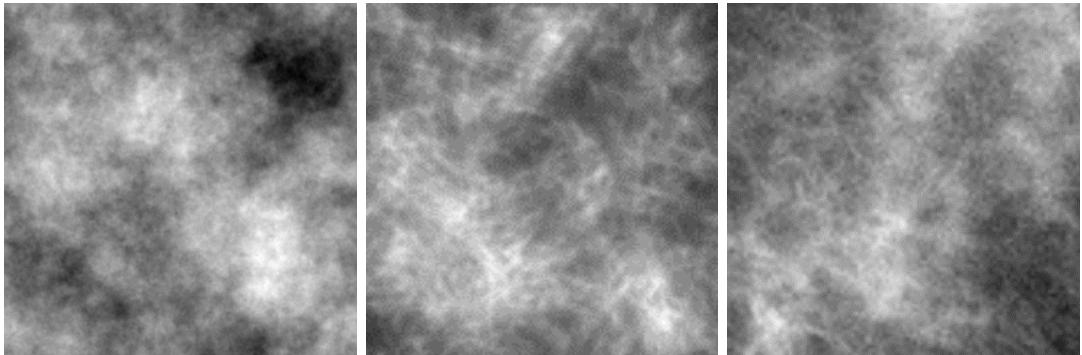


Figure 1.5: Left: image simulated from a 2D power-law Gaussian random field. Middle: image simulated from 2D a clustered lumpy background. Right: an ROI from a clinical mammogram. All images have size $3.5\text{ cm} \times 3.5\text{ cm}$.

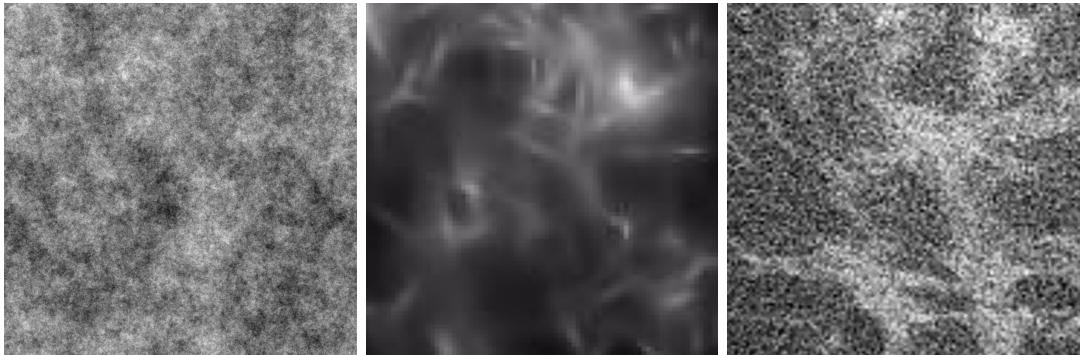


Figure 1.6: Left: an ROI of a slice from a volume simulated using a 3D power-law Gaussian random field. Middle: an ROI of a slice from a volume simulated using a 3D clustered lumpy background. Right: an ROI of a slice from a clinical bCT reconstructed data set. All images have size $3.5\text{ cm} \times 3.5\text{ cm}$.

using a 3D clustered lumpy background and a clinical dedicated breast computerized tomography (bCT) reconstructed data set.

Random field breast texture models have several advantages. Random field textures can be generated straightforwardly using well-established sampling methods. This enables a straightforward and flexible way to generate enough data for experiments demanding huge amount of images. Since they are parametric models and are completely characterizable, analytical derivation of their statistical properties is feasible. This can enable theoretical studies of performance metrics in simplified clinical tasks. For instance, it was previously demonstrated that, when stationary FFDM image regions are modeled as a power-law Gaussian random field, the β index is related to the detection performance of human observers in an analytical equation [32]. Random field breast texture models also have limitations. The primary limitation is the lack of visual realism and morphological variability when comparing breast images simulated using random field models with clinical breast images.

1.3.2 Anthropomorphic breast phantoms

Anthropomorphic breast phantoms are three-dimensional models that aim at modeling different breast anatomical structures in small, medium and large scales. A compression model can be used to further mechanically deform the breast as in clinical imaging practices. In order to produce x-ray projection images, an anthropomorphic phantom is usually subjected to either a virtual x-ray image acquisition system, or a real one if the phantom is fabricated as a physical object.

Several 3D software anthropomorphic breast phantoms have been proposed by simulating the arrangement of large, medium and small scale anatomical structures:

- *Large scale* anatomical structures: a region composed predominantly of adipose tissue and a region composed predominantly of fibroglandular tissue.
- *Medium scale* anatomical structures: adipose compartments, Cooper's ligaments, blood vessels.
- *Small scale* anatomical structures: fibroglandular tissue.

Depending on the methodologies used to generate different breast anatomical structures, existing anthropomorphic breast phantoms can be classified into three categories:

- *Model-based phantoms*. These anthropomorphic breast phantoms use mathematically defined geometric primitives to simulate different breast anatomical structures. Commonly used primitives include stochastic tessellations such as the Voronoi diagram, spheres, ellipsoids, cylinders and Bézier tubes. Model-based phantoms can be sampled as multi-resolution surface-meshes [37] [117] or discrete volumes at specific voxel size [14] [15] [16] [20] [158][19] [43] [157] [76] [115].
- *Empirical-based phantoms*. These anthropomorphic breast phantoms use clinical data or imaged surgical mastectomy specimens as input where several breast anatomical structures are extracted through a segmentation algorithm [110] [89] [127]. The segmentation aims to assign each spatial position of the input clinical data to an anatomical property. However, these models are not an exact representation of the breast anatomical components due to limited gray scale spatial resolution of the input empirical data and the segmentation or feature extraction algorithm.
- *Hybrid phantoms*. These anthropomorphic breast phantoms are modeled using different combinations of geometric primitives, clinical data and 3D random fields. Phantoms combining model-based phantoms and 3D random fields have been proposed. These phantoms are discrete volumes that use 3D random fields to simulated the background breast fibroglandular and adipose tissue. Other breast anatomical structures are simulated using geometric primitives. Three-dimensional random fields can be gray-scale 3D volumes [25] [26] [63] or be thresholded to binary volumes [109] [42] with a threshold value depending on the desired ratio of fibroglandular tissue. A phantom model combining empirical-based phantoms with 3D random fields and model-based phantom components has recently been proposed by Chen *et al.* [44]. Their phantom model adds higher-frequency components to breast structures segmented from low spatial resolution clinical breast CT data. A 3D power-law Gaussian random field is added to ensure connectivity between broken segmented structures and to create a micro-texture on the edge of the fibroglandular tissue. Geometric primitives, modeling Cooper's ligaments, ductal network, and blood vessels are also fused into the model.

Voxel-based and mesh-based formats have been proposed. In voxel-based phantoms, the choice of the voxel size is often a trade-off between the smallest relevant anatomical component and the manageability of the digital file size. In mesh-based phantoms, the flexibility to adapt mesh resolution to the complexity of the anatomical structures may allow to optimize phantom and image generation time as well as the digital file size.

An overview of the design characteristics of the state-of-the-art anthropomorphic breast phantoms can be found in Appendix A. Existing anthropomorphic breast phantoms have different mathematical tractability, variability of simulated breast images compared with clinical images and visual realism in large, medium and small scale anatomical structures. The mathematical tractability of existing model-based phantoms [14] [158] [117] [37] [76] is limited and little research has been done regarding this aspect. Since hybrid phantoms [25] [26] [109] [63] include 3D random field models that are analytically defined, they are in part mathematically tractable. The mathematical tractability of existing empirical-based phantoms [110] [89] is poor.

Anthropomorphic breast phantoms allow to simulate 2D and 3D images with large variations in visual realism. For example, the visual realism from the large scale anatomical components from empirical-based phantoms is much higher than in today's hybrid phantoms. Model-based phantoms allow to model large scale anatomical components with limited to fair visual realism. The visual realism from the medium scale anatomical components in model-based, empirical-based and hybrid phantoms can be considered all fair. The visual realism from the small scale anatomical components in empirical-based phantoms is limited. This can in part be attributed to the poor spatial resolution of the input clinical images used to design the empirical-based phantoms. The visual realism of the small scale anatomical components in model-based phantoms is poor to high and hybrid phantoms allow for a high visual realism.

Compared to the morphological variation seen in clinical images, the morphological variation in images from anthropomorphic breast phantoms is still limited; the variation is smallest for hybrid phantoms and larger for model-based and empirical-based phantoms.

In this dissertation, we aim to develop a 3D breast texture model that shares the advantages of the random field texture models and anthropomorphic breast phantoms. Specifically, we aim to develop a mathematically tractable 3D breast texture model capable of simulating realistic 2D and 3D breast images that are representative for the population of women seen in clinical practice.

1.4 Assessment of clinical task performance using model observers

The typical workflow for observer performance assessment in clinical tasks is illustrated in Figure 1.7. In the following sections, we give a comprehensive overview of its key elements.



Figure 1.7: Typical workflow for observer performance assessment in clinical tasks

In medical imaging, typical clinical tasks are lesion detection (screening) and lesion char-

acterization (diagnostics) tasks. The two tasks can be formulated as binary classification tasks using hypothesis testing. In a lesion detection task, two hypotheses are associated to an input image: lesion-absent (H_0) and lesion-present (H_1). The two hypotheses are mathematically represented as:

$$\begin{aligned} H_0 : \mathbf{g} &= \mathbf{b} + \mathbf{n}, \\ H_1 : \mathbf{g} &= \mathbf{b} + \mathbf{n} + \mathbf{s}. \end{aligned} \tag{1.2}$$

Here \mathbf{g} represents the input image, \mathbf{b} represents the image of background texture, \mathbf{n} denotes the noise in the image and \mathbf{s} denotes the lesion (or signal). We use the vectorized notation (bold letters in this document) to denote all images, treated as column vectors.

Lesion characterization tasks can be formulated in an analogous way and it is not further discussed here.

A decision variable λ is a scalar response variable reflecting the visual strategy used by the observer. During the scoring process, a decision variable is computed by the observer for each input image. The scoring process is represented by a *decision function* f that often depends on many variables.

The way a decision outcome D is obtained from a decision variable depends on the type of detection task. Typical detection tasks are the *binary* task, the *rating-scale* task and the *M-alternative forced-choice* task. Each detection task type is associated with a different decision-making scheme. In a binary task, also named yes/no task, the decision outcome is obtain using one threshold. In a N -point rating-scale task, N confidence levels of the observer are obtained using $N - 1$ possible threshold values. The rating-scales can also be continuous. In this case it is obtained as a transformation of the decision variable. In an M -alternative forced-choice task there are M decision variables. The decision outcome is assigned the index of the maximum decision variable. Analogous reasoning can be made for lesion characterization tasks. For a more detailed description of decision outcome we refer for example to the *Handbook of Medical Imaging* [194] (volume 1, chapter 11).

Depending on the scheme of decision-making, different figures-of-merit have been used to assess the performance of the observers. Common figures-of-merit are the probability of a correct answer (P_c) [194], the detectability index (d') [116] and the area under the receiver operating characteristic (ROC) curve (AUC) [30].

P_c describes the percentage of correct decisions over the total number of decisions. It is often used as the figure-of-merit in binary tasks and M -alternative forced-choice tasks. Alternatively, one can also compute the true positive fraction, the true negative fraction, the false positive fraction and the false negative fraction from all decisions outcomes.

In rating-scale tasks, d' and AUC are often used. The detectability d' is the standardized difference between the decision variables of the signal-present class and signal-absent class. An analytical link between d' and AUC exists when the decision variables of the signal-present class and signal-absent class are Gaussian distributed [21]. For a more detailed description of figures-of-merit we also refer to the *Handbook of Medical Imaging* [194] (volume 1, chapter 10).

1.4.1 Mathematical model observers

In the design of VCT, mathematical model observers have the objective to model the decision function f for the specific clinical task. Figure 1.8 shows commonly used model observers and their knowledge of the image background and the lesion. Detailed descriptions of different model observers can be found in [194].

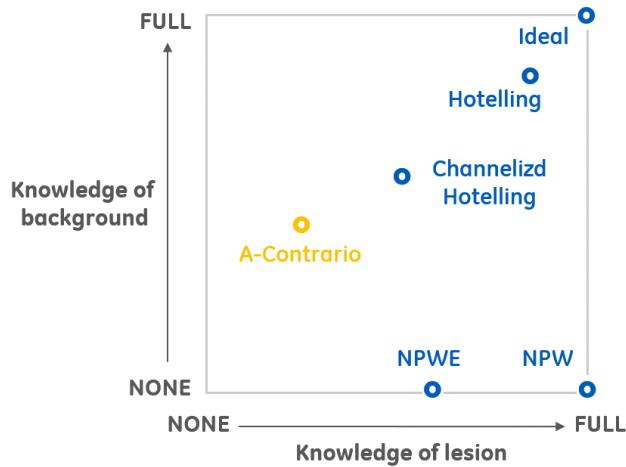


Figure 1.8: Different model observers and their knowledge of the image background and the lesion.

State-of-the-art model observers can be classified into the following categories.

Bayesian or ideal linear observer

The Bayesian ideal observer or ideal linear observer describes the performance of the optimum decision maker on a given decision task. They are aware and make use of the full knowledge in the image formation process. In clinical practice, Bayesian ideal observer and ideal linear observer are often hard to compute since the required multidimensional distribution of the images are not accessible for tasks involving complex backgrounds [148].

Non-ideal observer

The designs of non-ideal observers differ in the amount of information they include about the lesion and the background (Figure 1.8) [194]. Some non-ideal observers are equivalent to the Bayesian ideal observer when images are Gaussian [21]. Non-ideal observers can be further divided into non-anthropomorphic model observers and anthropomorphic model observers. The non-anthropomorphic model observers set an upper bound for the performance of the non-ideal observers. The anthropomorphic model observers on the other hand aim to predict the performance of human observers for the same tasks. This is achieved by including information-processing constraints intended to reflect the human visual or brain systems. Non-ideal observers sometimes outperform human observers [194] [84]. The internal noise [194] has been investigated as a mean to degrade the performance

of non-ideal observers to match human. Typically, the internal noise is modeled as a Gaussian random variable or vector and is added into the decision function f of non-ideal observers.

Non-ideal observers can be linear or non-linear. The decision variable λ of a linear model observer is written as:

$$\lambda = \mathbf{w}^T \mathbf{g}, \quad (1.3)$$

where \mathbf{w} is referred to as the template of the observer [194]. The template \mathbf{w}^T has the same number of elements as the image \mathbf{g} . Different linear non-ideal observers differ in the derivation of the observer template. Non-linear observers typically do not necessarily use a template to compute the decision variable. We describe hereunder common linear and non-linear non-ideal observers in lesion detection tasks. Analogous reasoning can be made for lesion characterization tasks.

Between the late 1990s and early 2000s, various model observers have been studied for lesion detection performance in 2D FFDM projection images. In 1999, Bochud *et al.* applied the non-prewhitening matched filter with Eye filter (NPWE) to study the detection performance of masses and microcalcifications in FFDM images [28]. As shown in Figure 1.8, the non-prewhitening matched filter (NPW) uses only information of the signal and does not use the background information. NPWE takes into account the differential human visual sensitivity to different spatial frequencies by filtering the NPW with a contrast sensitivity function (CSF) [194] [22]. Bochud *et al.* demonstrated that the performance of NPWE and human were highly correlated for the detection of microcalcifications; whereas the correlation was poor for the detection of masses.

In 2006, Grosjean *et al.* introduced an alternative non-linear non-ideal observer referred to as the *a contrario* observer (*a contrario* observer) for mass detection in FFDM-like background textures [80]. The *a contrario* observer is based on pixel-wise statistical testing with a global control of the number of false alarms and it differs from classical non-ideal observers which use a template to derive the decision variable. The decision variable of *a contrario* observer is derived from a global null-hypothesis on the background statistics and a measurement of pixel-wise features that necessitates few knowledge of the signal. The *a contrario* observer will be explained in detail in Chapter 6. Grosjean *et al.* showed that the contrast detail curves obtained by the *a contrario* observer were very similar to the contrast detail curves assessed from validations by humans [80].

Later-on in 2009, Castella *et al.* [38] used the NPW, the NPWE and the channelized Hotelling observer (CHO) to assess the detection performance of masses in FFDM-like background textures. Signal-known-exactly and signal-known statistically cases were studied. CHO uses channelized images as input. A channelized image is obtained by computing the inner product of the input image with multiple channel images of the same size, representing different spatial or frequency features of input images. CHO typically has two phases: the training and the test phases [140]. The training phase aims at estimating the CHO template from input signal-present and signal-absent images using linear discriminant analysis [124] [21]. The test phase applies the estimated template to new images to compute the decision variables. Further details of CHO are described in Appendix B. Castella *et al.* found that NPWE and CHO with internal noise correlated well with humans; while the correlation was worse for NPW with internal noise.

Further on, non-ideal observers have been applied to study several practical aspects in FFDM. These studies had the objective to assess the image quality of FFDM [137] [200]

[172] [195] [161] [203], to assess the automatic exposure control in FFDM system [171], or to compare various image processing algorithms applied in FFDM [200] [199]. These studies considered uniform background texture [137] [200] [172] [195] and FFDM (or FFDM-like) background texture [200] [199] [161] [171] [203]. The detection performance of masses [161] [171] [203] [203] and microcalcifications [137] [200] [172] [199] [195] [171] [203] were both considered. Images were obtained either using physical [137] [200] [172] [195] [203] or digital phantoms or from clinical acquisitions [200] [199] [161] [171].

Compared to 2D FFDM, the introduction of 3D DBT brings several new aspects in image acquisition, image processing, reconstruction and image review. To encounter these new aspects, dedicated models observers for 3D DBT need to be carefully designed. Very often DBT reconstructed slices are reviewed in stack mode. A lesion may span across several slices; as a consequence, there is a spatial correlation between DBT slices. In stack review of DBT slices, the radiologist may use different image browsing speeds. Also, the display monitor may have temporal variation in the luminance level and lag between image displays. These typical aspects of DBT need to be taken into account in the model observers.

Early and even some more recent studies applied 2D dedicated model observers to a single DBT reconstructed slice with the lesion in-focus for mass [40] [41] [108] and microcalcification [164] detection tasks [149] [130] [153]. Although the detection performance of single-slice non-ideal observers have shown to be correlated to human observers, these studies have limitations since none of the above described aspects inherent to DBT are taken into account.

Over recent years, new developments have been gradually introduced to existing 2D model observers to encounter the aspects dedicated to 3D DBT.

To introduce the spatial correlation between slices, in 2011 Platiša *et al.* [155] extended the CHO to 3D datasets. Two extensions of the CHO have been introduced for the detection of spherical signals in 3D datasets: the volumetric CHO model and the multi-slice CHO models. The volumetric CHO treats each set of input slices as a rendered volume and performs a single channelization on each input set using 3D channels. This channelization does not realistically model the stack review mode of volume slices in DBT. The multi-slice CHO models encounter the stack review mode of volume slices by first performing a slice-by-slice channelization, then summarizing all channelized slices according to different designs. These 3D extensions of CHO are described in detail in Appendix B. Later-on in 2011, Platiša *et al.* also incorporated the temporal display luminance of the monitor in the multi-slice CHO for the detection of spherical signals in 3D datasets [156]. The correlation to human was not investigated in the studies of Platiša *et al.*

To enhance the CHO model with the capability to search for various locations in the input images for lesions, different visual search models have been introduced. In 2013 Michielsen *et al.* [131] included a local visual search model in a multi-slice CHO model for the detection and characterization of microcalcifications in DBT reconstructed slices. The authors demonstrated that the results of the model observers were linearly correlated with the human observer results for both microcalcification detection and characterization tasks. However, a local search might not be a realistic modeling of how human searches in DBT slices for lesions. In the same year, Lau *et al.* [108] included a global visual search model on the entire image in single-slice CHO for the detection of masses in DBT reconstructed slices. Lau *et al.* showed that the introduction of global visual search

in single-slice CHO allowed to produce area under the localization ROC curve values similar to those from human observers. However, more research is needed to increase the robustness of the visual search algorithm with respect to human performance.

To incorporate the modeling of the browsing speed of DBT slices to a multi-slice CHO model introduced by Platiša *et al.*, in 2013 Avanaki *et al.* [4] combined the the multi-slice CHO with a spatio-temporal contrast sensitivity function (stCSF) [22] for the detection of masses and microcalcifications in DBT reconstructed slices. The stCSF is an extension to the CSF that incorporates extra aspects in the modeling of human visual system such as the browsing speed for a stack of images. Avanaki *et al.* demonstrated that the modeling of browsing speed using stCSF allowed to obtain model observer results that conformed to those of an earlier human observer study by Diaz *et al.* [57]. Later-on in 2014, Avanaki *et al.* proposed improvements of the previous multi-slice CHO with stCSF using non-linear methods to better match human observer performance as a function of viewing distance, effective contrast, maximum luminance, and browsing speed [5]. Their experiments indicated that the improved model yielded better match to human observer performance.

Recently in 2016, the performance of lesion detection using only DBT projection images has been investigated by Park *et al.* [150]. To introduce the modeling of the spatial correlations between DBT projection in CHO, Park *et al.* incorporated this spatial correlations in different CHO channels for the detection of spherical and elliptical 3D signals in DBT projection images. The correlation to humans was not studied. This study might has limited clinical relevance since the DBT projection images are typically not reviewed during clinical exams.

To date, very few studies have applied these 3D-adapted model observers to investigate practical aspects in DBT. A limited number of studies have been published investigating the impact of different parameters of the DBT reconstruction algorithm on the detection performance of microcalcifications in DBT [130] [168], or comparing the detection performance of microcalcifications in FFDM, DBT and synthetic 2D mammography [90]. More elaborated studies are welcome to further test and validate these 3D-adapted model observers.

In this dissertation, we aim to develop a new *a contrario* model observer to assess microcalcification in DBT slices. Our design is triggered by the satisfactory result Grosjean *et al.* obtained for mass detection in FFDM images. We aim to extend the 2D *a contrario* observer to 3D. The new *a contrario* observer will be tested against state-of-the-art CHO models and then used to compare the microcalcification detection performance in FFDM images, DBT reconstructed slices and synthetic mammograms.

1.5 Conclusion and manuscript organization

For nearly two decades now, full field digital mammography (FFDM) has been the standard for breast cancer screening. FFDM has been proven to reduce breast cancer mortality in a cost-effective way when used in organized screening programs. Recently, as a technical advancement of FFDM, 3D digital breast tomosynthesis (DBT) was introduced. DBT, allowing for the display of thin sections through the breast, reduces the masking effect of overlapping breast tissue. Prospective studies have shown improved sensitivity and

specificity for cancer detection.

Compared to FFDM, DBT faces several challenges. There is no consensus on the performance of microcalcification detectability in DBT compared with FFDM; some studies have shown improved detection performance while others have shown reduced detection performance. The increased number of images doubles image interpretation time and this negatively impacts clinical workflow. There is no convincing evidence that the diagnostic value of synthetically reconstructed mammograms is equivalent to 2D mammograms. Comprehensive studies assessing the impact of the image acquisition sequence design and DBT reconstruction algorithms on microcalcification detection performance are needed. Objective task-based methodologies allowing to analyze image review efficiency and allowing to compare lesion detectability in synthetic mammograms and real mammograms is very welcome.

Performance assessment of breast x-ray imaging systems through clinical trials is burdensome, expensive and may result in radiation dose to the patient. As an alternative, several research groups have been investigating the potential of virtual clinical trials (VCT). There is a long history in the development of random field breast texture models and anthropomorphic breast phantoms as inputs for VCT. While 2D mathematical random field breast texture models have been shown to be a good facsimile to simulate mammographic textured background and have shown to be locally statistically equivalent, their 3D extensions show very limited visual realism. Despite this limitation, some of their statistical properties may still match those of real breast textures and may still make them a reasonable candidate to better understand the value of various imaging modalities. Anthropomorphic breast phantoms on the other hand are characterized by an improved visual realism, but they are mathematically not tractable. Nowadays there is still room for further and better understanding of mathematical breast texture models. The development of a novel 3D mathematical breast texture model that offers both the advantages of random field breast texture models and anthropomorphic breast phantoms should be further investigated. A mathematically tractable 3D breast texture model capable of simulating 2D and 3D breast images with realistic morphological variations as seen in clinical images and satisfactory large, medium and small scale visual realism is very welcome. Such a 3D model may allow for more advanced VCT studies considering sub-populations of breasts. Such a model will also facilitate the performance evaluation and optimization of 3D DBT imaging acquisition topologies, reconstruction algorithms and image review methods.

Between the late 1990s and early 2000s, various model observers have been studied and were found to be more or less successful for the assessment of lesion detection performance in 2D FFDM projection images. The introduction of DBT brought several new aspects to image review; the most prominent being image review in stack mode. To encounter these unique aspects of DBT image review, new model observers have been recently proposed. Model observers have been proposed that take into account the correlation between slices in stack review, the spatio-temporal contrast sensitivity of the human eye when browsing through a stack of slices at various speeds, the temporal lag of medical displays, and the distance between the reviewer and the display. Despite the advances of these new DBT dedicated model observers, there is still plenty of room for further investigation before the acceptance of a preferred design for anthropomorphic model observers dedicated to DBT.

This manuscript presents our contributions to the development of a novel 3D breast texture model and a new mathematical model observer based on the *a contrario* theory for the

implementation of VCT requiring a high level of realism. It is organized as follows.

In Chapter 2, we provide a theoretical-mathematical framework for the 3D power-law Gaussian random field and 3D clustered lumpy background models. An analytical derivation of the statistical properties of the 3D texture representations and their relation with 2D projection images and reconstructed DBT slices simulated from these input textures is presented.

In Chapter 3, we introduce a novel 3D stochastic solid breast texture model capable of simulating a large variability of 2D and 3D breast images in terms of small and medium scale fibroglandular and inter-glandular adipose tissue in breast. The proposed 3D texture model is formulated as a marked point process with mathematically tractable characterization.

In Chapter 4, we propose a methodology to objectively infer the medium scale parameters of the 3D solid breast texture model proposed in Chapter 3. The proposed inference method is based on an automatic analysis of the clinical dedicated breast computerized tomography datasets. The inference aims to allow the 3D solid breast texture model to simulate a larger variability of breast textures, thus making it more robust.

In Chapter 5, we investigate the common statistical characteristics of typical 2D and 3D x-ray breast images encountered in clinical practices. The first order statistical characteristic of images, breast density, is investigated first. A new volumetric breast density assessment algorithm using DBT projections, based on the model-based stereology is proposed. Numerical evaluations are conducted to study the estimation bias and variance of the proposed algorithm. The power spectral index (β), a second order statistical characteristic is investigated next. We investigate a state-of-the-art β computation method focusing on its bias and variance under different parameterizations. Analytical derivation of the results is demonstrated. Numerical evaluations are performed to validate the analytical results.

In Chapter 6, we propose a new mathematical model observer based on the *a contrario* theory for microcalcification detection task in 2D and 3D x-ray breast imaging. The new *a contrario* observer is based on pixel-wise statistical tests and aims to globally control the average number of false positives. Theoretical proof and experimental analysis will be conducted to demonstrate the global false positive control for the proposed *a contrario* observer.

In Chapter 7, we implement a complete virtual clinical trial using the 3D stochastic solid breast texture model and the *a contrario* observer introduced in Chapter 3 and 6. Our objective is three-fold. First, we aim to compare the performance of the *a contrario* observer with state-of-the-art channelized Hotelling observer. Second, we aim to preliminarily compare the performance of the *a contrario* observer and the channelized Hotelling observer with the performance of human observer. Third, we aim to compare the performance of microcalcification detection in 2D FFDM, 3D DBT and synthetic 2D images.

Chapter 2

Three-dimensional random field breast texture models

In Chapter 1, the advantages and limitations of current mathematical random field texture models were discussed. Two-dimensional power-law Gaussian random field and clustered lumpy background have been shown to be a good facsimile to simulate mammographic textured background. Their first and second order statistical features are similar as those observed in clinical mammograms. The 2D power-law Gaussian random field and the clustered lumpy background have been extended to 3D as a mean to simulate breast anatomical textures [42] [162] [111]. Although these 3D random field texture models are limited in terms of visual realism, the fact that some of their statistical properties match those of real breast textures may still make them reasonable candidates to better understand the performance of various imaging modalities for particular visual tasks [111] [126].

The goal of this chapter is to provide a theoretical mathematical framework for the 3D power-law Gaussian random field and the 3D clustered lumpy background. We will also formulate the statistical properties of their 3D texture models and derive their relation with 2D projection images and reconstructed DBT slices simulated from these 3D texture models.

Our work in this chapter serves as a comprehensive study of state-of-the-art breast texture models and aims to advance our understanding of x-ray breast texture simulation. We believe that a description of the statistical properties of 3D random field texture models and the propagations of these properties in their simulated 2D projection images and reconstructed slices may help to progress designs of more advanced 3D mathematical breast texture models. Also, studying the advantages and limitations of different existing models will give us a better insight into the key issues to be addressed for the development of a future model.

2.1 General mathematical formulation

As described in Section 1.3.1, random field breast texture models are mathematically tractable models based on the random field theory [196]. In x-ray breast imaging, mainly two categories of models have been investigated to simulate breast textures:

1. *The power-law Gaussian random field* model, a stationary Gaussian random field model completely characterized by its power-spectral density, also called noise power spectrum in the medical imaging context. The name power-law originates from the fact that the noise power spectrum of the power-law Gaussian random field has a power-law function shape.
2. *The shot-noise random field* model, a general class of random field models defined as the superposition of function kernels centered at a simple spatial point process. The most commonly studied shot-noise random field model in breast imaging is the clustered lumpy background [27] [39].

We provide hereunder a generalized mathematical formulation of the two models under a unified discrete 3D setting. Consider a 3D discrete grid $D := \{1, 2, \dots, N_1\} \times \{1, 2, \dots, N_2\} \times \{1, 2, \dots, N_3\}$ where $N_1, N_2, N_3 \in \mathbb{N}$ are the numbers of discrete elements along the canonical X , Y and Z axis respectively.

Definition 2.1. (real-valued discrete random field)

A real-value discrete random field f defined on D is a measurable map

$$f : (\Omega, \mathcal{F}, P) \times D \rightarrow \mathbb{R},$$

where (Ω, \mathcal{F}, P) is a probability space.

Definition 2.2. (2D slice and parallel projection)

Let f be a real-valued discrete random field defined on D . A 2D slice f_z of f , orthogonal to the $X - Y$ plane, at position $z \in \{1, 2, \dots, N_3\}$ of the Z -axis and a 2D parallel projection f_P of f onto the $X - Y$ plane are defined as:

$$f_z(x) = f(x, z), \quad (2.1)$$

$$f_P(x) = \sum_{x_3=1}^{N_3} f(x, x_3), \quad (2.2)$$

for $x \in \{1, 2, \dots, N_1\} \times \{1, 2, \dots, N_2\}$.

This definition is illustrated in Figure 2.1. Compared to x-ray breast imaging, the parallel projection f_P is considered to represent the projection image in digital mammography (FFDM) or digital breast tomosynthesis (DBT), assuming a parallel x-ray beam. An arbitrary slice f_z can be considered a slice of the reconstructed 3D volume of the breast, assuming that an infinite number of projections of f are acquired at an infinite number of angles and a perfect 3D reconstruction algorithm is applied. Definition 2.2 will be further used to derive the analytical links between the models in 3D and the statistical properties of their 2D parallel projection f_P and 2D slice f_z .

The following sections provide detailed analyses of the 3D power-law Gaussian random field and the 3D shot-noise random field.

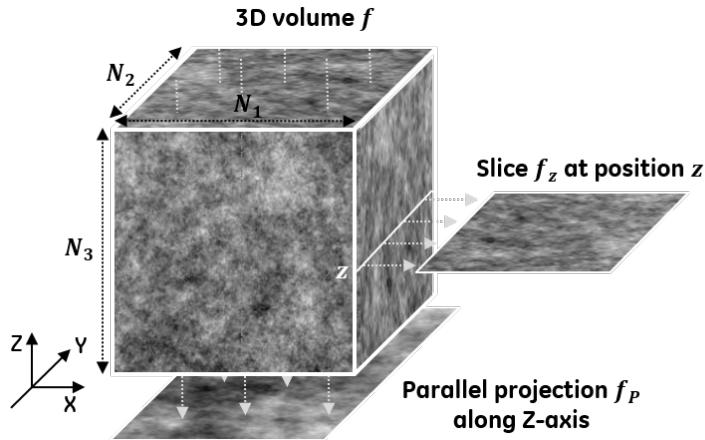


Figure 2.1: An arbitrary slice and the parallel projection of a 3D random field

2.2 Power-law Gaussian random field

The power-law Gaussian random field is the most commonly studied random field breast texture model. In 1999, Burgess [31] first applied the 2D power-law Gaussian random field to model stationary breast morphology seen in clinical 2D FFDM images. Later-on, 3D extensions of the 2D power-law Gaussian random field have also been introduced to model 3D breast anatomical texture, either as a stand-alone model [162] or embedded in anthropomorphic breast phantoms [42] [109].

2.2.1 Mathematical formulation

The power-law Gaussian random field is a particular case of Gaussian random fields.

Definition 2.3. (real-valued 3D discrete Gaussian random field)

Consider the previously defined 3D discrete grid D . A real-valued random field f defined on D is a real-valued discrete Gaussian random field if and only if the random vector $\mathbf{v} = \{f(x_1), f(x_2), \dots, f(x_d)\}$ follows a multivariate normal distribution for any collection $\{x_1, x_2, \dots, x_d\}$ with $d \leq |D|$. That is, the probability density function of \mathbf{v} is:

$$p(\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{v} - \bar{\mathbf{v}})^T \Sigma^{-1} (\mathbf{v} - \bar{\mathbf{v}})\right). \quad (2.3)$$

Here \det denotes the matrix determinant and $(\cdot)^T$ denotes the matrix transpose. The vector $\bar{\mathbf{v}}$ is the mean of \mathbf{v} and Σ is the covariance matrix.

Moreover, for any $x \in D$, we have

$$\bar{\mathbf{v}}(x) = \mu(x) = \mathbb{E}(f(x)), \quad (2.4)$$

where μ is the mean of f . Also, for any $i, j \in \{1, \dots, d\}$, the covariance matrix Σ satisfies:

$$\Sigma_{ij} = c(x_i, x_j) = \mathbb{E}((f(x_i) - \mu(x_i))(f(x_j) - \mu(x_j))), \quad (2.5)$$

where c , a real-valued non-negative definite function, is the covariance function of f .

The distribution of a Gaussian random field f is uniquely determined by its mean and covariance function. A Gaussian random field f is stationary if its distribution is translation-invariant. This implies that its mean μ is a constant real scalar and its covariance function c is translation-invariant. This means that for any $x, y \in D$, $c(x, y) = c(\mathbf{s})$, where $\mathbf{s} = x - y$. The Gaussian random field f is isotropic if its distribution is rotation-invariant. This further implies that $c(x, y) = c(Rx, Ry)$ for an arbitrary 3D rotation matrix R . A stationary and isotropic Gaussian random field f has a constant mean μ and a covariance function c that depends only on the distance between two positions, that is, $c(x, y) = c(s)$, with $s = \|y - x\|$.

The following proposition provides an analytical characterization of a real-valued discrete periodic stationary Gaussian random field.

Proposition 2.1. *Let f be a real-valued discrete periodic stationary Gaussian random field in \mathbb{Z}^3 . Without loss of generality, assume also that the mean of f is zero. Further, let w be a white Gaussian random field. That is, for any finite set $\{x_1, x_2, \dots, x_d\} \subset \mathbb{Z}^3$, $w(x_1), w(x_2), \dots, w(x_d)$ are identically independently distributed following the standard normal distribution $\mathcal{N}(0, 1)$. Under these assumptions, f can be expressed as the convolution of w and a square-integrable kernel function $g \in L^2$:*

$$f = w * g. \quad (2.6)$$

Here $*$ denotes the periodic convolution operator. In this representation, the covariance function c of f can be expressed using the kernel function g :

$$c(t) = g * g^-(t), \quad (2.7)$$

with $g^-(t) = g(-t)$, $\forall t \in \mathbb{Z}^3$.

A proof of the proposition can be found in [138, p111]. Using the Wiener-Khintchine theorem [60], Equation 2.7 can be rewritten as:

$$\Gamma_f(\xi) := \mathcal{F}_c(\xi) = \|\mathcal{F}_g(\xi)\|^2, \quad (2.8)$$

where \mathcal{F} denotes the discrete Fourier transform and ξ denotes the discrete Fourier frequency. Here Γ_f , the Fourier transform of c , is referred to as the power spectral density of f . Proposition 2.1 and Equation (2.8) indicate that a discrete stationary Gaussian random field is characterized by its power spectral density.

Now we can formally define the power-law Gaussian random field.

Definition 2.4. (real-valued 3D discrete power-law Gaussian random field)

Consider a real-valued discrete stationary Gaussian random field f defined on D . Without loss of generality, assume that f has zero mean. Then f is a power-law Gaussian random field if and only if its spectral density Γ_f exists and has the following form:

$$\Gamma_f(\xi) = \begin{cases} 0 & \text{if } \|\xi\| = 0, \\ \frac{k}{\|\xi\|^\beta} & \text{if } \|\xi\| \neq 0. \end{cases} \quad (2.9)$$

Here $\xi \in \{0, \dots, N_1 - 1\} \times \{0, \dots, N_2 - 1\} \times \{0, \dots, N_3 - 1\}$ denotes the 3D discrete frequency. The parameter $\beta \in \mathbb{R}^+$ is called the power spectral index of f . The parameter $k \in \mathbb{R}$ is used to control the magnitude of f .

From Equation (2.9) we see that the spectral density of a power-law Gaussian random field is rotation-invariant. Therefore a power-law Gaussian random field is also isotropic. A power-law Gaussian random field is characterized by the power spectral index parameter β .

Numerical simulation of such a random field is straightforward using Proposition 2.1. A realization \hat{f} of a power-law Gaussian random field can be obtained from the convolution of a white Gaussian random field w with a kernel function g satisfying

$$\mathcal{F}_g(\xi) = \frac{k}{\|\xi\|^{\beta/2}}, \quad (2.10)$$

where $k \in \mathbb{R}$ is used to control the magnitude of \hat{f} . The convolution can be done in the frequency domain by computing the point-wise product of \mathcal{F}_w and \mathcal{F}_g . Then the inverse discrete Fourier transform \mathcal{F}^{-1} is applied to the product to obtain the realization \hat{f} . That is,

$$\hat{f} = \mathcal{F}^{-1}(\mathcal{F}_w \odot \mathcal{F}_g). \quad (2.11)$$

Here \odot denotes the point-wise multiplication.

2.2.2 Simulation example

Figure 2.2 shows a slice and the parallel projection of a discrete 3D cube simulated using a 3D power-law Gaussian random field with $\beta = 2.83$. The side length of the cube is 3.5 cm and the size of the isotropic discrete voxels is 0.1 mm. The parameter $\beta = 2.83$ represents the average β value used in a previously study by Chen *et al.* in an attempt to simulate 3D breast background textures [42]. A region of interest (ROI) from a clinical breast computerized tomography (bCT) reconstructed coronal slice and from a clinical mammogram are also shown as references. The isotropic pixel sizes of the bCT slice and the mammogram are respectively 0.35 mm and 0.1 mm.

2.2.3 Properties of 2D slice and projection

Consider now a 2D slice f_z given by (2.1), and the 2D parallel projection f_P given by (2.2) of a 3D discrete power-law Gaussian random field f . The statistical properties of f_z and f_P can be analytically characterized by the following proposition.

Proposition 2.2. *Let f be a real-valued 3D discrete power-law Gaussian random field defined on D with spectral density Γ_f defined in (2.9) and with spectral index β . Then we have:*

- A 2D slice f_z of f is a 2D stationary isotropic Gaussian random field with spectral density given by:

$$\Gamma_{f_z}(\xi_2) = \sum_{x_3=0}^{N_3-1} \Gamma_f(\xi_2, x_3), \quad (2.12)$$

for any 2D discrete frequency $\xi_2 \in \{0, 2, \dots, N_1 - 1\} \times \{0, 2, \dots, N_2 - 1\}$.

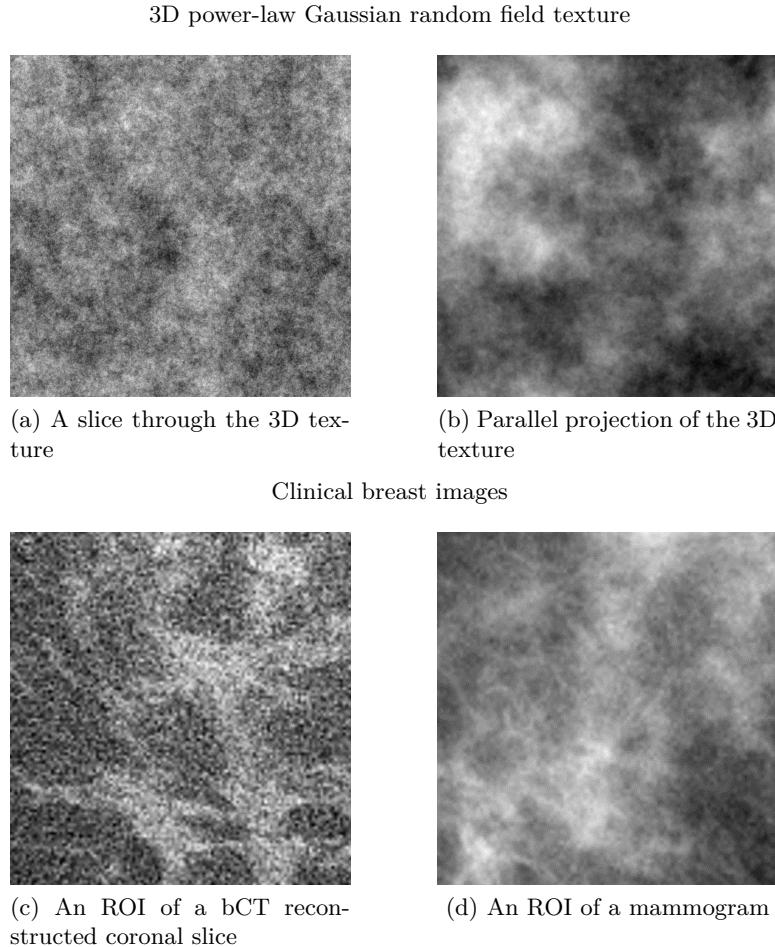


Figure 2.2: A slice (upper left) and the parallel projection (upper right) of a discrete 3D cube simulated using a 3D power-law Gaussian random field with $\beta = 2.83$. The side length of the cube is 3.5 cm and the size of the isotropic discrete voxels is 0.1 mm. An ROI from a clinical bCT reconstructed coronal slice (lower left, courtesy of Prof. John M. Boone, UC Santa Davis) and from a clinical mammogram (lower right) are also shown as references. The isotropic pixel sizes of the bCT slice and the mammogram are respectively 0.35 mm and 0.1 mm.

- The 2D parallel projection f_P of f is a 2D power-law Gaussian random field with the same spectral index β . That is, its spectral density Γ_{f_P} satisfies

$$\Gamma_{f_P}(\xi_2) = \begin{cases} 0 & \text{if } \|\xi_2\| = 0, \\ \frac{k}{\|\xi_2\|^\beta} & \text{if } \|\xi_2\| \neq 0, \end{cases} \quad (2.13)$$

for any 2D discrete frequency ξ_2 defined in the same way above.

Proof. The proof of Proposition 2.2 is an application of the Fourier projection-slice theorem stated as follows [143].

Theorem 2.1. Let f be an L^2 function defined on D . Consider its 2D projection f_P and

a slice f_z defined in Definition 2.2. Then the following equation holds:

$$\mathcal{F}_{f_P}(\xi_2) = \mathcal{F}_f(\xi_2, 0), \quad (2.14)$$

for any 2D discrete frequency $\xi_2 \in \{0, \dots, N_1 - 1\} \times \{0, \dots, N_2 - 1\}$.

The formula for the power spectral density Γ_{f_P} in (2.13) can be obtained by directly applying Theorem 2.1 to the spectral density Γ_f defined in (2.9).

Notice that Theorem 2.1 still holds if we change the Fourier transform operator \mathcal{F} to the inverse Fourier transform operator \mathcal{F}^{-1} and replace f by its 3D Fourier transform \mathcal{F}_f . This means that the Fourier transform of f_z is equal to the projection of \mathcal{F}_f onto the $X - Y$ plane in the Fourier domain. That is,

$$\Gamma_{f_z}(\xi_2) = \sum_{x_3=0}^{N_3-1} \Gamma_f(\xi_2, x_3). \quad (2.15)$$

□

Notice that f_z is a stationary isotropic Gaussian random field; But it is no longer a power-law Gaussian random field. This can be seen by further developing the spectral density of f_z :

$$\Gamma_{f_z}(\xi_2) = \sum_{x_3=0}^{N_3-1} \frac{k}{\sqrt{\|\xi_2\|^2 + x_3^2}^\beta}. \quad (2.16)$$

Let $k = 1$ without loss of generality, the discrete sum in (2.16) can be treated as a Riemann approximation to the integral:

$$\int_0^{N_3-1} \frac{1}{\sqrt{\|\xi_2\|^2 + \tau^2}^\beta} d\tau. \quad (2.17)$$

This integral evaluates to:

$$\begin{aligned} \int_0^{N_3-1} \frac{1}{\sqrt{\|\xi_2\|^2 + \tau^2}^\beta} d\tau &= \int_{\|\xi_2\|}^{\sqrt{\|\xi_2\|^2 + (N_3-1)^2}} \frac{1}{u^{\beta-1}} \frac{1}{\sqrt{u^2 - \|\xi_2\|^2}} du \\ &= \frac{C}{\beta-1} \frac{1}{\|\xi_2\|^{\beta-1}} {}_2F_1 \left(1, \frac{1}{2} - \frac{\beta}{2}; \frac{3}{2} - \frac{\beta}{2}, \frac{(N_3-1)^2}{\|\xi_2\|^2} \right). \end{aligned} \quad (2.18)$$

Here we applied a change of variable $u = \sqrt{\|\xi_2\|^2 + \tau^2}$. We use ${}_2F_1(\cdot, \cdot; \cdot, \cdot)$ to denote the hyper-geometric function and $C \in \mathbb{R}$ is a constant. Equation (2.18) suggests that the spectral density of f_z is related to a power-law function with exponent $\beta - 1$.

Numerical evaluations were performed to visualize the relation between the spectral density of f_z in (2.16) and a power-law function with exponent $\beta - 1$. The evaluations were performed for $\beta = 1.5, 2, 3$ and 4 to cover the range of textures encountered in 2D and 3D breast x-ray imaging, and N_3 in (2.16) was set to 256 . Each function was evaluated with $\|\xi_2\|$ varying from 1 to 255 with a step size of 1 . The result of the evaluation is shown in Figure 2.3. We can see that for $\beta = 3$ and $\beta = 4$, the spectral density of f_z in (2.16) is

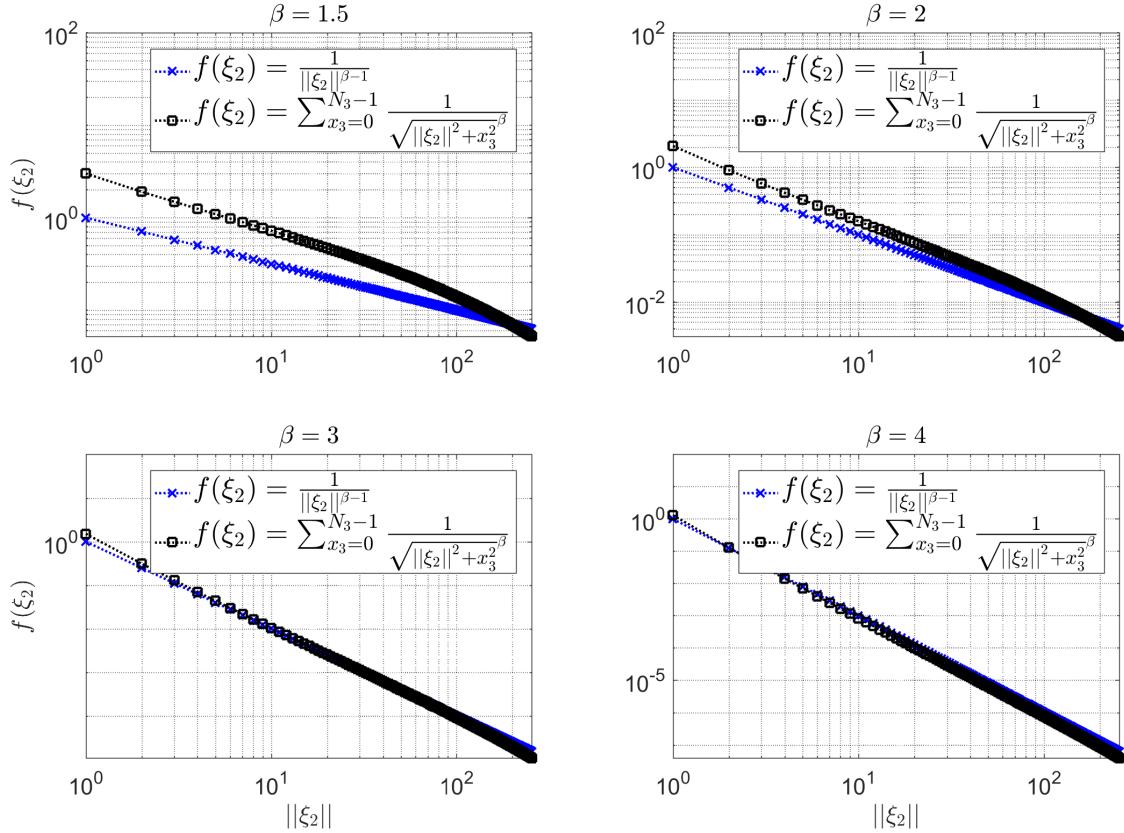


Figure 2.3: Comparison of the spectral density of f_z in (2.16) and a power-law function with exponent $\beta - 1$ for $\beta = 1.5, 2, 3$ and 4 and $N_3 = 256$. The two functions were numerically evaluated for a range of $\|\xi_2\|$ values varying from 1 to 255 with a step size of 1 .

well approximated by the power-law function with spectral index $\beta - 1$, while this is not the case for $\beta = 1.5$ and $\beta = 2$.

To further illustrate Proposition 2.2, a simulation experiment was performed. Four sets of 3D volumes of discrete size $256 \times 256 \times 256$ were simulated respectively from the power-law Gaussian random fields with spectral index $\beta = 1.5, 2, 3$ and 4 . The value of k in (2.9) was set to 1 . Each set contains 100 realizations, from which 100 central slices perpendicular to the Z -axis and 100 parallel projection images along the Z -axis were created. First, the spectral indices of the slices β_z and the projections β_P were empirically estimated for all realizations using the method described in [87]. Let $\hat{\beta}_z$ denote the estimate of β_z and let $\hat{\beta}_P$ denote the estimate of β_P . Then, to check if Proposition 2.2 was satisfied, two Student's t -tests were performed for each investigated β value. The first test was performed using all estimated $\hat{\beta}_z$ values with null-hypothesis that $\mathbb{E}(\hat{\beta}_z)$ is equal to $\beta - 1$. The second test was performed using all estimated $\hat{\beta}_P$ values with null-hypothesis that $\mathbb{E}(\hat{\beta}_P)$ is equal to β . Table 2.1 demonstrates the average of $\hat{\beta}_z$ and $\hat{\beta}_P$ for all investigated β values, as well as the p -values for the t -tests at 5% significance level. We can see that the simulation results are in accordance with Proposition 2.2. For all cases, $\mathbb{E}(\hat{\beta}_P)$ does not statistically deviate from the β value of the original 3D volume at 5% significance level. This is however not the case when we compare $\mathbb{E}(\hat{\beta}_z)$ with $\beta - 1$. The result of this simulation experiment is

consistent with Proposition 2.2.

Previous studies have found similar results regarding the statistical properties of the 2D projection and slice of the 3D power-law Gaussian random field. Metheany *et al.* showed that the β value of a slice of a 3D power-law Gaussian random field is equal to the β of the projection images minus one [126]. The authors also demonstrated the consistency of the analytical result with the β values empirically estimated from a set of clinical mammograms (average: 3.01, standard deviation 0.32) and bCT slices (average 1.86, standard deviation: 0.38). The difference between the two estimated average β values was very close to one.

3D volume	$\beta = 1.5$	$\beta = 2$	$\beta = 3$	$\beta = 4$
Slice				
	$\bar{\beta}_z = 0.876,$ $p = 2e-16$	$\bar{\beta}_z = 1.267,$ $p = 2e-16$	$\bar{\beta}_z = 2.128,$ $p = 2e-16$	$\bar{\beta}_z = 3.065,$ $p = 2e-16$
Projection				
	$\bar{\beta}_P = 1.498,$ $p = 0.548$	$\bar{\beta}_P = 1.994,$ $p = 0.084$	$\bar{\beta}_P = 3.006,$ $p = 0.141$	$\bar{\beta}_P = 4.012,$ $p = 0.548$

Table 2.1: Result from a numerical experiment demonstrating the relationship between the β of a 3D power-law Gaussian random field and its 2D slice and projection. Four sets of 3D volumes, each containing 100 3D power-law Gaussian random field realizations with $\beta = 1.5, 2, 3$ and 4 were simulated. Empirical β values of the central slice ($\hat{\beta}_z$) and the projections ($\hat{\beta}_P$) were estimated for each realization, using the method described in [87]. To compare the difference between $\mathbb{E}(\hat{\beta}_z)$ and $\beta - 1$ and the difference between $\mathbb{E}(\hat{\beta}_P)$ and β , Student's t -tests were performed. Statistically significant differences at 5% significance level ($p < 0.05$) are indicated in bold.

2.3 Shot-noise random field

Shot-noise random field, also referred to as the spot-noise random field, consists of the superposition of function kernels centered at points from a Poisson point process [71]. Rolland and Barrett first applied the shot-noise random field to simulate image textures encountered in positron emission tomography [167]. Their model was named the *lumpy background* and it consists of Gaussian kernels centered at positions from a homogeneous Poisson point process. In mammography, the shot-noise random field was modified by

Bochud *et al.* under the name *clustered lumpy background* for mammographic texture synthesis [27]. The modified model changed the underlying point process from a homogeneous Poisson point process to a cluster point process. Later on, an optimization of the clustered lumpy background was proposed by Castella *et al.*, where the parameters were objectively inferred from a database of clinical mammograms using a genetic algorithm [39].

2.3.1 Mathematical formulation

We hereby refer to the a model that consists of the superposition of function kernels centered at points from an arbitrary simple point process as a general shot-noise random field.

Definition 2.5. (real-valued 3D discrete general shot-noise random field)

A real-valued 3D discrete general shot-noise random field f , defined on D is mathematically expressed as:

$$f(x) = \sum_{i \in \mathcal{I}} g(x - y_i; m_i), \quad (2.19)$$

for $\forall x \in D$. Here $\Phi_C = \{y_i\}_{i \in \mathcal{I}}$ is the center point process: a simple point process defined in \mathbb{R}^3 . Moreover, $g(\cdot; m) : \mathbb{Z}^3 \times \mathbb{R}^d \rightarrow \mathbb{R}$, referred to as the kernel function, is a real-valued function parametrized by m , referred to as the marks. The marks are independent from Φ_C and are distributed according to the mark distribution \mathcal{M} defined on \mathbb{R}^d . The kernel function $g(\cdot; \cdot)$ satisfies:

$$\int_{\mathbb{R}^d} \sum_{y \in D} |g(y; m)| \mathcal{M}(dm) < +\infty. \quad (2.20)$$

A general shot-noise random field is characterized by its center point process Φ_C and its kernel function $g(\cdot; m)$ determined by the marks m with mark distribution \mathcal{M} .

Under similar mathematical formulation, we can formally define the 3D clustered lumpy background as a special case of the general shot-noise random field.

Definition 2.6. (3D clustered lumpy background)

A 3D clustered lumpy background f defined on D is a 3D discrete general shot-noise random field where,

- The center point process Φ_C is a modified Thomas cluster point process[136] expressed as:

$$\Phi_C = \bigcup_{j \in \mathcal{J}} \bigcup_{i=1}^{N_j} \{y_{i,c_j}\}. \quad (2.21)$$

- Here $\Phi_P = \{c_j\}_{j \in \mathcal{J}}$ is a “parent” point process: a homogeneous Poisson point process defined on \mathbb{R}^3 with intensity λ_0 .
- For any $j \in \mathcal{J}$, N_j is sampled from a Poisson random variable with mean N_c . Then a collection of independent identically distributed “children” points $\{y_{i,c_j}\}_{i=1, \dots, N_j}$ are sampled according to the multivariate Gaussian distribution $\mathcal{N}(c_j, \sigma^2 I_3)$, where I_3 is the 3×3 identify matrix.

Φ_C is constructed by collecting of all the “children” points.

- The kernel function g is expressed as

$$g(x; m) = \exp\left(-a\mathcal{E}(R_{\phi_x, \phi_y, \phi_z}x)^b\right), \quad (2.22)$$

where

$$\mathcal{E}(s) = \sqrt{\left(\frac{s_x}{L_x}\right)^2 + \left(\frac{s_y}{L_y}\right)^2 + \left(\frac{s_z}{L_z}\right)^2}, \quad (2.23)$$

for $\forall s = (s_x, s_y, s_z) \in D$, is the analytical function associated to a canonical ellipsoid having (L_x, L_y, L_z) as the half lengths of its principle axes. The random rotation matrix $R_{\phi_x, \phi_y, \phi_z}$ is determined by three uniformly distributed Euler angles (ϕ_x, ϕ_y, ϕ_z) . The parameters a, b are deterministic values defined on \mathbb{R}^+ . Notice that the marks are $m = (a, b, L_x, L_y, L_z, \phi_x, \phi_y, \phi_z)$, in which only ϕ_x, ϕ_y and ϕ_z are random.

2.3.2 Simulation examples

Figure 2.4 shows a slice and the parallel projection of a discrete 3D cube simulated using a 3D clustered lumpy background with parameters listed in Table 2.2. The side length of the cube is 3.5 cm and the size of the isotropic discrete voxels is 0.1 mm. The parameters were empirically chosen by visually inspecting the appearance of the simulated volume. To our knowledge, no previous studies has been done investigating the optimization of 3D clustered lumpy background parameters for clinical realism and this optimization was also considered out-of-scope in this study. A region of interest (ROI) from a clinical breast computerized tomography (bCT) reconstructed coronal slice and from a clinical mammogram are also shown as references. The isotropic pixel sizes of the bCT slice and the mammogram are respectively 0.35 mm and 0.1 mm.

Parameter	Value
λ_0	9.6 mm^{-3}
N_c	10
σ	11
a	5.0
b	0.59
L_x, L_y, L_z	$L_x = 4.19 \text{ mm}, L_y = 1.63 \text{ mm}, L_z = 1.63 \text{ mm}$
ϕ_x, ϕ_y, ϕ_z	Uniformly distributed on $(0, \pi)$

Table 2.2: Parameters of the 3D clustered lumpy background used for the simulation in Figure 2.4.

2.3.3 Properties of 2D slice and projection

Consider now a 2D slice f_z given by (2.1) and the 2D parallel projection f_P given by (2.2) of a real-valued 3D discrete general shot-noise random field f . The statistical properties of f_z and f_P can be analytically characterized by the following proposition.

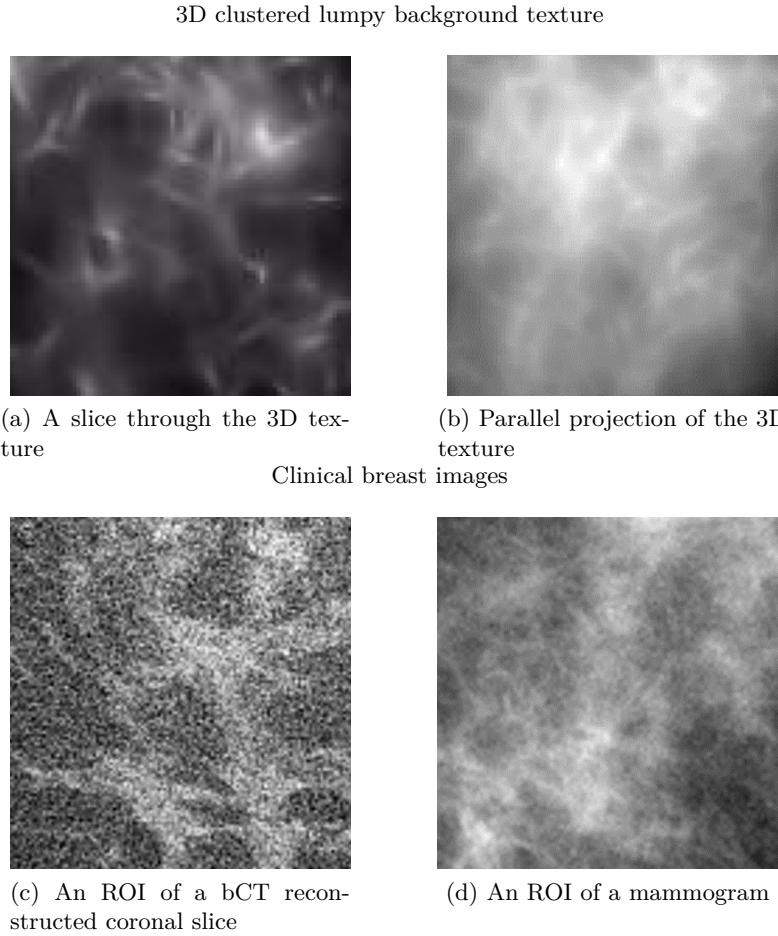


Figure 2.4: A slice (upper left) and the parallel projection (upper right) of a discrete 3D cube simulated using a 3D clustered lumpy background with empirical parameters listed in Table 2.2. The side length of the cube is 3.5 cm and the size of the isotropic discrete voxels is 0.1 mm. An ROI from a clinical bCT reconstructed coronal slice (lower left, courtesy of Prof. John M. Boone, UC Santa Davis) and from a clinical mammogram (lower right) are also shown as references. The isotropic pixel sizes of the bCT slice and the mammogram are respectively 0.35 mm and 0.1 mm.

Proposition 2.3. Let f be a real-valued 3D discrete general shot-noise random field defined in D , with center point process Φ_C and kernel function $g(\cdot; m)$, determined by marks m with distribution \mathcal{M} . A 2D slice f_z and parallel projection f_P of f are characterized as follows.

- A 2D slice f_z is a 2D general shot-noise random field characterized by the following center point process and kernel function.
 - The center point process $\tilde{\Phi}_C$ is a simple point process defined on \mathbb{R}^2 , obtained by projecting Φ_C along the Z -axis.
 - For any $x \in \mathbb{Z}^2$, the kernel function g_z is expressed as

$$g_z(x; \tilde{m}) = g_z(x; (z, m)) := g((x, z); m). \quad (2.24)$$

It is defined on $\mathbb{Z}^2 \times \mathbb{R}^{d+1}$. Let p_z be the marginal distribution of the Z -axis component of the center points from Φ_C . Then the new marks $\tilde{m} = (z, m)$ follow the distribution $p_z \mathcal{M}$.

- The parallel projection f_P is a 2D general shot-noise random field characterized by the following center point process and kernel function.
 - The center point process Φ_P is the same point process $\tilde{\Phi}_C$ as for the 2D slice f_z .
 - For any $x \in \mathbb{Z}^2$, the kernel function g_P is expressed as

$$g_P(x; m) = \sum_{x_3=1}^{N_3} g((x, x_3); m). \quad (2.25)$$

It is defined on $\mathbb{Z}^2 \times \mathbb{R}^d$ and is determined by the same marks m with the same distribution \mathcal{M} .

Let us illustrate this proposition with two examples.

Example 2.1. Consider a shot-noise random field f where, the center point process Φ_C is a homogeneous Poisson point process with intensity parameter λ . The kernel function is the indicator function of a sphere with radius m . That is

$$g(x; m) = \mathbb{1}_{\|x\| \leq m}, \quad (2.26)$$

where the distribution of m is the Dirac at value r .

- An arbitrary slice f_z of f , as shown in Figure 2.5a is a discrete 2D shot-noise random field.
 - The center point process $\tilde{\Phi}_C$ of f_z is a homogeneous Poisson point process with intensity λN_3 , obtain by projecting the original homogeneous Poisson point process Φ_C along the Z -axis.
 - The new kernel function g_z can be formulated as

$$g_z(x; \tilde{m}) = \mathbb{1}_{\|x\| \leq \tilde{m}}. \quad (2.27)$$

It is the indicator function of a disc with radius \tilde{m} , a discrete random variable uniformly distributed in $\{\sqrt{r^2 - \min(r^2, k^2)}\}_{k=-r, -r+1, \dots, 0, \dots, N_3-r-1}$.

- The projection f_P of f , as shown in Figure 2.5b is a discrete 2D shot-noise random field.
 - The center point process is the same homogeneous Poisson point process $\tilde{\Phi}_C$ with intensity λN_3 .
 - The new kernel function g_P can be formulated as

$$g_P(x; m) = \sum_{x_3=1}^{N_3} \mathbb{1}_{x_3^2 \leq m^2 - \|x\|^2} = 2 \lfloor \sqrt{m^2 - \|x\|^2} \rfloor + 1. \quad (2.28)$$

The distribution of m is the same Dirac at r .

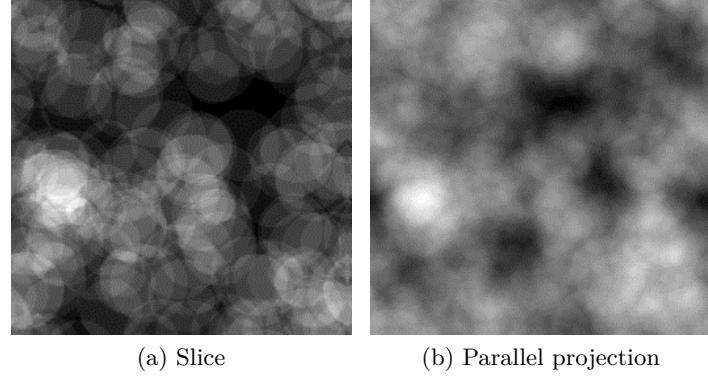


Figure 2.5: A slice and the parallel projection of a 3D cube simulated using a discrete shot-noise random field f . The side length of the cube is 2.56 cm and the size of the isotropic discrete voxels is 0.1 mm. The center point process of f is a homogeneous Poisson point process with intensity $\lambda = 40 \text{ mm}^{-3}$. The kernel functions of f is specified by the sphere with fixed radius $r = 0.3 \text{ mm}$. The slice and the parallel projection are both 2D shot-noise random fields characterized by new center point processes and new kernel functions.

Example 2.2. Consider now that f is the 3D clustered lumpy background defined in 2.6. From Proposition 2.3 we have:

- An arbitrary slice f_z of f , as shown in Figure 2.4a is a discrete 2D general shot-noise random field.
 - The center point process $\tilde{\Phi}_C$ is still a modified Thomas cluster point process. We can formulate $\tilde{\Phi}_C$ as:

$$\tilde{\Phi}_C = \bigcup_{j \in \mathcal{J}} \bigcup_{i=1}^{N_j} \{\tilde{y}_{i,c_j}\}. \quad (2.29)$$

Here $\tilde{\Phi}_P = \{c_j\}_{j \in \mathcal{J}}$ is a homogeneous Poisson point process with intensity $\lambda_0 N_3$, obtain by projecting the original homogeneous Poisson point process Φ_P with intensity λ_0 along the Z-axis. For each c_j with $j \in \mathcal{J}$, N_j follows the same Poisson distribution as in Definition 2.6. The “children” points $\{\tilde{y}_{i,c_j}\}_{i=1, \dots, N_j}$ are independent identically distributed according to the bivariate Gaussian distribution $\mathcal{N}(c_j, \sigma^2 I_2)$, where I_2 is the 2×2 identify matrix.

- The new kernel function g_z is formulated as

$$g_z(x; \tilde{m}) = g_z(x; (z, m)) := g((x, z); m). \quad (2.30)$$

The new marks $\tilde{m} = (z, m)$ follows the distribution $p_z \mathcal{M}$ where p_z is the marginal distribution of the Z-axis component of the points from Φ_C .

- The projection f_P of f , as shown in Figure 2.4b is a discrete 2D general shot-noise random field.
 - The center point process is the same modified Thomas cluster point process $\tilde{\Phi}_C$ as for the slice f_z .

- The kernel function g_P is analytically formulated as

$$g_P(x; m) = \sum_{x_3=1}^{N_3} g((x, x_3); m). \quad (2.31)$$

The marks m follow the same distribution \mathcal{M} of the original marks.

2.4 Conclusion and discussion

Three-dimensional random field breast texture models have limited visual realism. However, some of their statistical characteristics match those found in clinical breast images. A handful of studies have shown that they may be used to assess the performance of certain 3D breast imaging applications.

In this chapter, we investigated the analytical characterization of the 3D power-law Gaussian random field and the 3D clustered lumpy background under a general 3D discrete mathematical formulation. These two 3D models have been used in breast imaging. Theoretical results showed that 2D parallel projections and 2D slices of both 3D random field models belong to the same model class and can be analytically characterized. The 2D parallel projection of a 3D power-law Gaussian random field is a 2D power-law Gaussian random field characterized by the same spectral index β . An arbitrary 2D slice of a 3D power-law Gaussian random field can be approximated as a 2D power-law Gaussian random field with spectral index $\beta - 1$. The 2D parallel projection and an arbitrary 2D slice of a 3D clustered lumpy background are also general shot-noise random field models, characterized by new center point processes and new kernel functions.

Chapter 3

A novel 3D stochastic solid breast texture model for x-ray breast imaging

Chapter 1 and Chapter 2 described anthropomorphic breast phantoms and 3D random field breast texture models for use in virtual clinical trials (VCT). While 3D random field breast texture models are mathematically tractable, the restricted visual realism of their simulated 3D breast x-ray images and their limited capability to mimic the morphological variability seen in clinics are key drawbacks. Today's anthropomorphic breast phantoms on the other hand are not mathematically tractable. The visual realism of anthropomorphic breast phantoms in simulated 3D breast x-ray images and their capability to mimic the morphological variability are fair to satisfactory, depending on the phantom model.

In this chapter, we propose a 3D solid breast texture model that is capable of simulating a large variability of realistic 2D and 3D breast images in terms of small and medium scale fibroglandular and inter-glandular adipose tissue in breast. Analytical characterization of the proposed model is tractable thanks to its mathematical formulation using marked point process theory.

The proposed model is inspired by the morphology and distribution of medium and small scale fibroglandular and inter-glandular adipose tissue observed in clinical breast computerized tomography (bCT) images. For medium scale breast tissue, we used a system of stochastically distributed overlapping ellipsoids to depict the adipose compartments, and the complement of the ellipsoid system to depict the fibroglandular tissue. A small scale texture aspect was introduced by replacing the smooth ellipsoid boundaries by Voronoi cells with average volume less than 1 mm^3 .

The texture model was voxelized. X-ray images were simulated using an x-ray simulator. As a first prototype implementation, we empirically determined the model parameters for three different breast density categories by visually inspecting the simulated FFDM and DBT images using clinical images as reference. A formal realism assessment was performed for simulated DBT reconstructed slices using a two-alternative forced choice experiment involving human observers.

3.1 Input materials: clinical dedicated breast computerized tomography image database

The construction of our 3D solid breast texture model is based on the observations from clinical dedicated breast computerized tomography (bCT) image datasets.

The bCT datasets are reconstructed volumes from bCT projection images acquired using the bCT system designed at the University of California Davis Medical Center [112]. During the bCT image acquisition, the patient lay on a table in prone position with the breast to be imaged hanging through a hole on the table. An x-ray tube and a flat panel detector rotated around the breast in the horizontal plane underneath the table [112] [205]. The physical size of the detector is 40 cm in width and 30 cm in height. The size of the detector pixel matrix is 1024×768 pixels. This results in $0.388\text{ mm} \times 0.388\text{ mm}$ detector element size. A 2×2 detector binning was performed to achieve a 30 frames per second acquisition rate. Five hundred projection images were acquired at 80 kVp with an angular range slightly over 360° [205]. The mean glandular dose ranges from 2.5 mGy to 10.3 mGy depending on the patient breast size [112]. The projection images were used to reconstruct 300 to 500 slices with size 512×512 pixels, parallel to the coronal plane using the filtered back-projection algorithm with a Shepp-Logan filter [126]. The pixel size of the reconstructed slices ranges from 0.23 mm to 0.41 mm and the number of the reconstructed slices varies from patient to patient depending on the size of the breast [126].

Figure 3.1 demonstrates four examples of bCT datasets representing four fibroglandular breast tissue densities varying from almost entirely fat (BI-RADS breast density 1) to extremely dense (BI-RADS breast density 4).

3.1.1 Segmentation of fibroglandular tissue from bCT images

In a first step, the distribution and morphology of the fibroglandular and inter-glandular adipose tissue in the bCT reconstructed volumes were carefully analyzed. For this purpose, these types of tissue were segmented from the bCT reconstructed volumes. The bCT reconstructed volumes are characterized by low-frequency artifacts (Figure 3.2) due to beam hardening, x-ray scatter and due to image reconstruction. To remove the low-frequency artifacts, an automatic fibroglandular tissue segmentation algorithm was developed. The segmentation algorithm was developed as part of an internship performed in 2015 in our team [3].

The developed algorithm performs the segmentation in 3D. In summary, the algorithm consists of the following steps. First, a k -means clustering [139, p352] was performed on the histogram of the input 3D bCT volume to separate the 3D breast mask from the background. Then the breast skin was removed by masking the breast tissue with the 3D breast mask processed by morphological erosions. The low-frequency artifacts were estimated by fitting a polynomial in 3D with degree 4 to the skin-removed breast tissue using least square. The fitted polynomial was subtracted from the breast tissue to remove the low-frequency artifacts. Next, a segmentation using hysteresis thresholding was performed on the breast tissue to extract the fibroglandular tissue. Finally, a morphological closing followed by a 3D median filter with a $2 \times 2 \times 2$ kernel were performed on the extracted fibroglandular tissue to reduce noise.

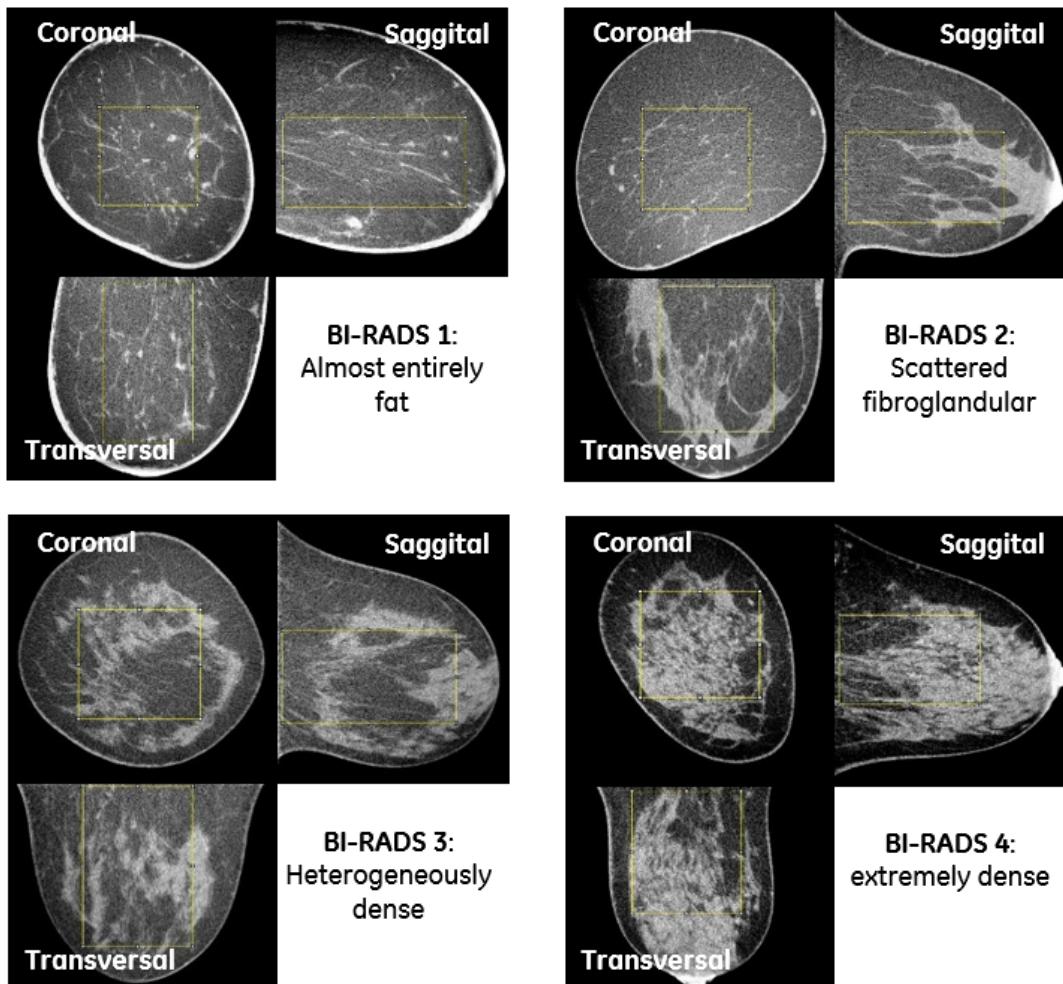


Figure 3.1: Four examples bCT reconstructed volumes representing four breast density types varying from almost entirely fat to extremely dense.

The segmented bCT datasets were then used as input to a series of analyses and observations to determine our new 3D breast texture model.

3.1.2 Analysis and observations of segmented bCT images

Previously, Li *et al.* [110] has shown that by projecting segmented clinical bCT datasets, mammographic textures depicting realistic morphological variations perceived in clinical mammograms can be obtained. This observation offers a reasonable justification for using segmented bCT datasets as the ground truth for the construction of a new breast texture model.

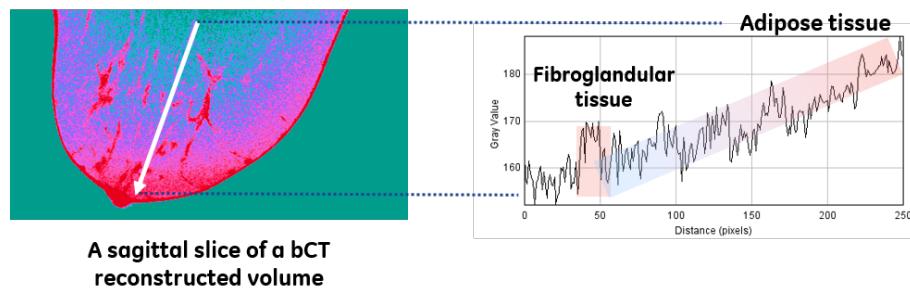


Figure 3.2: Demonstration of the low-frequency artifacts in bCT reconstructed volumes. A profile of the signal intensity is shown for a saggital slice of a bCT reconstructed volume. The profile demonstrates that, due to the low-frequency artifacts the signal intensity in a adipose tissue region can be higher than the signal intensity in a more distant fibroglandular tissue region. This prevents a correct segmentation of the original volume into fibroglandular and adipose tissue.

Simulation of DBT projections and reconstructed slices using segmented bCT images

To assess the visual realism of simulated DBT breast images using the segmented clinical bCT data, we conducted a simulation experiment in analogy to the experiment of Li *et al.* DBT projections and reconstructed slices were recreated from the bCT datasets.

Only the medium and small scale fibroglandular and inter-glandular adipose tissue types were considered for the experiment. For this, cubic volumes of interest (VOI) with size $5\text{ cm} \times 5\text{ cm} \times 5\text{ cm}$ of the original segmented bCT volumes were extracted from a subset of the bCT datasets. The VOIs were extracted from the center of the segmented breasts, leaving at least 2 cm distance to the breast skin and the chest-wall, depending on the breast size. The size of the VOIs was chosen small enough so that the fibroglandular tissue distribution does not vary drastically. For each bCT VOI, the voxels classified as fibroglandular tissue were assigned the linear attenuation coefficient corresponding to that of breast fibroglandular tissue at 20 keV. Similarly, the voxels classified as adipose tissue were assigned the linear attenuation coefficient corresponding to that of breast adipose tissue at 20 keV.

The VOIs were then projected using a previously described x-ray simulator [132] to simulate DBT projection images. The x-ray simulator was adjusted to model the topology of the GE SenoClaire DBT imaging system. Each VOI was placed at the center-left of the detector, such that the left edge of the VOI cube was aligned with the detector's left edge. A mono-energetic x-ray source at 20 keV was used to simulate nine DBT projection images with an angular range of $(-12.5^\circ, 12.5^\circ)$. Only Poisson x-ray noise was modeled and the electronic noise was not modeled. X-ray scattering and blurring on the detector were not modeled. The size of the isotropic detector pixels is $100\text{ }\mu\text{m}$. Mechanical breast texture deformation to mimic breast compression during the DBT image acquisition was not modeled. Projection images were processed by the ASIR-DBT 3D reconstruction algorithm (version 1.3.4, GE Healthcare, Buc, France) to obtain DBT reconstructed slices with 1 mm slice thickness.

Figure 3.3 shows regions of interest (ROI) of DBT central projections and reconstructed

slices simulated from segmented clinical bCT VOIs and from clinical DBT scans. The ROIs represent four different breast densities varying from BI-RADS breast density 1 to BI-RADS breast density 4. The ROIs sizes are $3.5\text{ cm} \times 3.5\text{ cm}$. The clinical DBT reconstructed slices were obtained using the same reconstruction algorithm. We observe that DBT central projections and reconstructed slices simulated using bCT VOIs visually appear to be very realistic compared with clinical images with similar breast density.

Observations of medium and small scale fibroglandular and adipose tissue morphology and distribution in segmented bCT images

In order to have some preliminary guidance on how to realistically model the distribution and morphology of medium and small scale fibroglandular and inter-glandular adipose tissue in segmented bCT VOIs, we conducted visual observations of the segmented bCT VOIs. The following remarks were made from our observations:

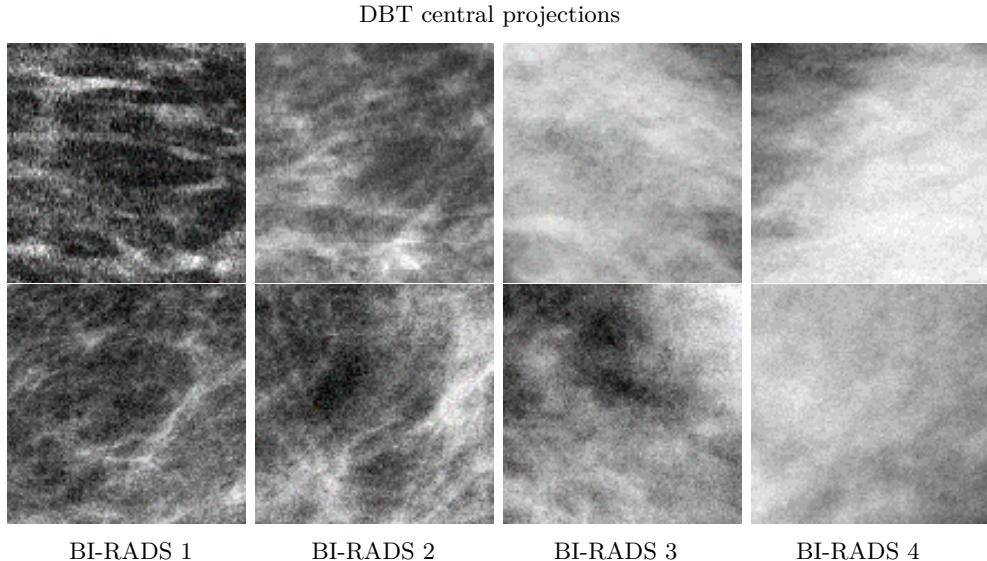
- It appears that each medium scale adipose compartment can be well approximated by a union of overlapping ellipsoids. The ellipsoids seem to point towards the nipple position with randomly distributed tilt angles. This is illustrated in Figure 3.4a. Notice that this observation is consistent with what has been found in previous studies [14] [117].
- There seems to be small scale irregularities at the adipose compartment boundaries. These irregularities seem to contribute to the micro-textures in simulated breast images. Figure 3.4b illustrates this irregularity in a $3.5\text{ cm} \times 3.5\text{ cm}$ ROI of a slice of segmented clinical bCT volume.

Both the limited spatial resolution of the Boone breast CT data and the fact that imaging was performed using continuous tube motion might limit the ability to accurately model very small-scale breast tissue using these data. The finite spatial resolution of the pixels of the imaging system will introduce a bias in the ability to accurately depict the fibroglandular tissue borders in case the contours of the voxels are larger than the actual contours of the fibroglandular tissue. In addition, a continuous exposure leads to motion blur image artifacts; indeed, short exposures are needed to obtain sharp images. In our study, we suppose that the low spatial resolution of binary bCT images is not the primary reason for the adipose compartment boundary irregularity. In fact, histological images demonstrate that the adipose compartment boundary irregularity is real. Bakic *et al.* [19] [127] have previously illustrated irregular variations around adipose compartments in microscopic breast histological images.

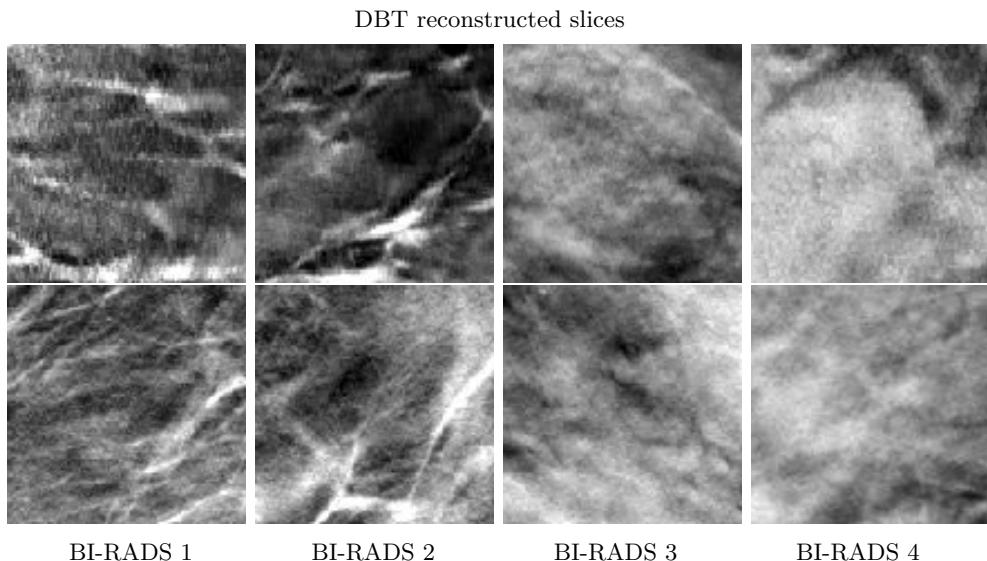
3.2 The new 3D stochastic geometric solid breast texture model

3D solid texture is a concept introduced in [50]. Contrary to 2D surface textures and shape descriptors which are defined on the surfaces of 3D objects, 3D solid textures also encompass information of the internal structure of 3D objects [50]. Our 3D breast texture model falls into the category of the 3D solid textures. Indeed, the proposed model simulates 3D voxelized volumes, describing both the surface and the interior structure of the 3D fibroglandular tissue and inter-glandular adipose tissue.

Based on the analysis and observations of segmented bCT VOIs, we decided to use the fol-



(a) DBT central projections simulated from segmented clinical bCT VOIs (top-row) and from clinical DBT scans (bottom-row). The ROIs represent four different breast densities varying from BI-RADS breast density 1 to BI-RADS breast density 4. The ROIs sizes are $3.5\text{ cm} \times 3.5\text{ cm}$.



(b) DBT reconstructed slices simulated from segmented clinical bCT VOIs (top-row) and from clinical DBT scans (bottom-row). The ROIs represent four different breast densities varying from BI-RADS breast density 1 to BI-RADS breast density 4. The ROIs sizes are $3.5\text{ cm} \times 3.5\text{ cm}$.

Figure 3.3: Illustration of DBT central projections and DBT reconstructed slices from segmented clinical bCT VOIs and from clinical DBT scans. DBT central projections and reconstructed slices simulated using bCT VOIs visually appear to be very realistic compared with clinical images with similar breast density.

lowing stochastic geometric models to simulate the medium and small scale fibroglandular and inter-glandular adipose tissue:

- To depict the medium scale fibroglandular and inter-glandular adipose tissue, we created a system of random ellipsoids whose union accurately exhibits the distribution

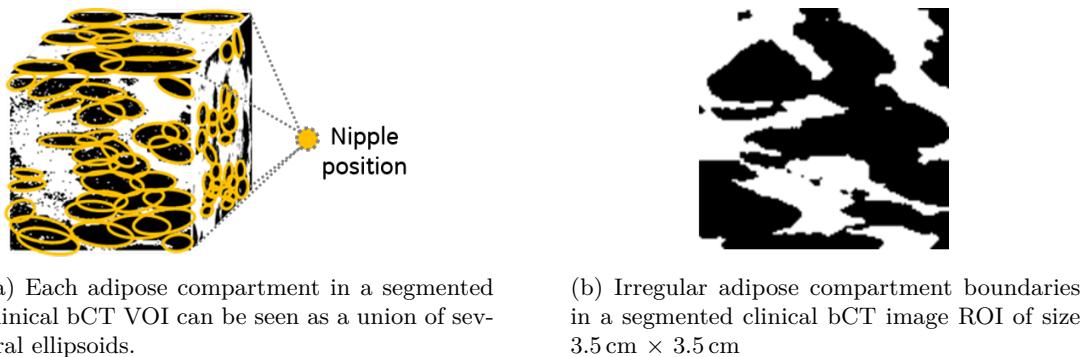


Figure 3.4: Illustration of the morphology and distribution of medium and small scale fibroglandular and adipose tissue in a segmented bCT VOI.

and morphological characteristics of the medium scale adipose compartments observed from segmented bCT datasets. The use of random ellipsoids has been investigated in some previously proposed phantom models, such as the model of Bakic *et al.* [14] and the model of Mahr *et al.* [117].

- To simulate the small scale texture aspect, we introduced small Voronoi cells at the surface boundaries of the ellipsoids, simulating irregular adipose compartment boundaries as observed in segmented clinical bCT datasets.

Since our objective is to simulate medium scale breast tissue, we restrict the simulation to only a cubic VOI with size $5\text{ cm} \times 5\text{ cm} \times 5\text{ cm}$ at the center of the breast. The global geometric configuration for our simulations is illustrated in Figure 3.5. In our configuration, the origin of the global coordinate system is placed at the center of the texture cube W . The y - z plane is parallel to the coronal plane of the breast. The nipple is placed at a position q in the half-space that contains the positive half of the x -axis.

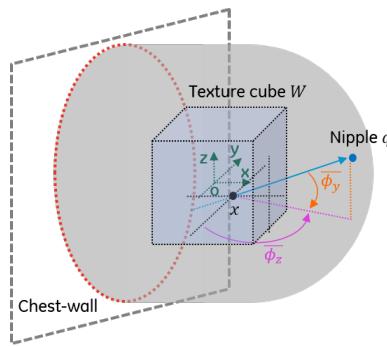


Figure 3.5: Global positioning of the simulated texture cube W and determination of $\bar{\phi}_z$ and $\bar{\phi}_y$ from an ellipsoid center position x and the nipple position q .

3.2.1 Sampling from the 3D solid breast texture model

The algorithm to sample a realization from the 3D solid breast texture model is described in Algorithm 3.1. The different algorithm steps are illustrated in Figure 3.6. From the figure we can distinguish the modeling of the medium scale (steps 1 and 2 in Algorithm 3.1) and

the small scale (steps 3 to 5 in Algorithm 3.1) fibroglandular and inter-glandular adipose tissue.

Algorithm: sampling from the 3D solid breast texture model

- **In continuous domain**

- **Step 1 (medium scale): sample seed points.**

Given a cube W , create a set of points \mathcal{S} by drawing a realization from a spatial point process Φ_s in W with distribution \mathbf{P}_s . This set of points \mathcal{S} is named the seed points.

- **Step 2 (medium scale): sample random ellipsoids.**

For each seed point $s \in \mathcal{S}$, assign a random ellipsoid $\mathcal{E}_s(\boldsymbol{\theta})$ centered at s , with parameter vector $\boldsymbol{\theta}$. The parameter vector $\boldsymbol{\theta}$ is sampled according to a pre-defined distribution \mathbf{P}^θ that depends on the nipple position q .

- **Step 3 (small scale): sample initial points.**

Independently from step 1 and step 2, uniformly generate N points in W , with N following a Poisson distribution of parameter $\lambda_0|W|$, where $|\cdot|$ denotes the volume measure and $\lambda_0 \in \mathbb{R}^+$ is a positive real number. This generates a realization of a homogeneous Poisson point process Φ_0 [45].

- **Step 4 (small scale): generate Voronoi diagram.**

Generate a Voronoi diagram [45] using points obtained from step 3 as cell centers.

- **In discrete domain**

- **Step 5: discretization.**

Discretize the whole cube W with voxel size ν . Check each point x of step 1. If for any $s \in \mathcal{S}$, x is inside the ellipsoid \mathcal{E}_s within a precision parameter ϵ , then assign all voxels of its corresponding Voronoi cell a value 0. Assign all other voxels value 1.

Algorithm 3.1

Parametrization of the random ellipsoids

The parameters of a 3D ellipsoid (step 4 in Algorithm 3.1) involve its center $x \in \mathbb{R}^3$, the half lengths of its three principle axis L_a, L_b and L_c and the orientation of its principle axes determined by three elemental rotations with angles ϕ_x, ϕ_y and ϕ_z around the x, y and z axes of the coordinate system respectively. The rotated ellipsoid is obtained by combining the three elemental rotations.

In our study we used the extrinsic rotation convention [69], meaning that the x, y and z axes of the global coordinate system remain stationary when the ellipsoid rotates. We denote ϕ_x, ϕ_y and ϕ_z as the elemental rotation angles around the x, y and z axes of the

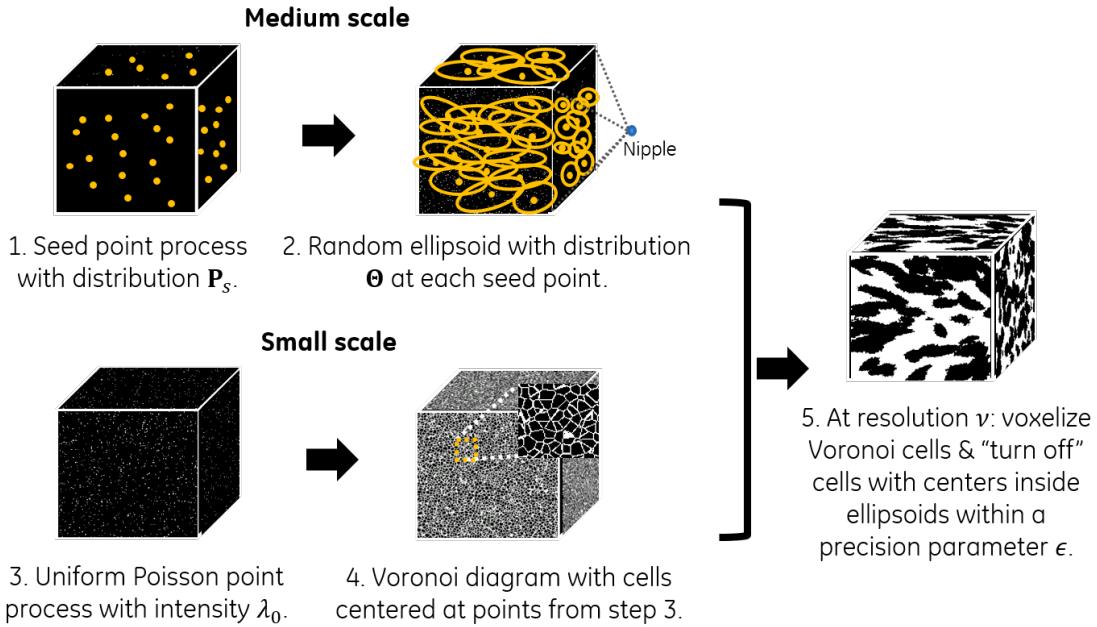


Figure 3.6: Different steps to sample from the 3D solid breast texture model. The modeling of the medium scale (steps 1 and 2 in Algorithm 3.1) and the small scale (steps 3 to 5 in Algorithm 3.1) fibroglandular and inter-glandular adipose tissue types are distinguished.

global coordinate system respectively. Directions of the elemental rotations follow the right-hand rule. Using this convention, a rotation of an ellipsoid with angles ϕ_x, ϕ_y, ϕ_z is the chaining of the extrinsic elemental rotations around the fixed global x , y and z axes in $x \rightarrow y \rightarrow z$ order and in counterclockwise direction. Detailed steps of such a rotation are illustrated in Figure 3.7.

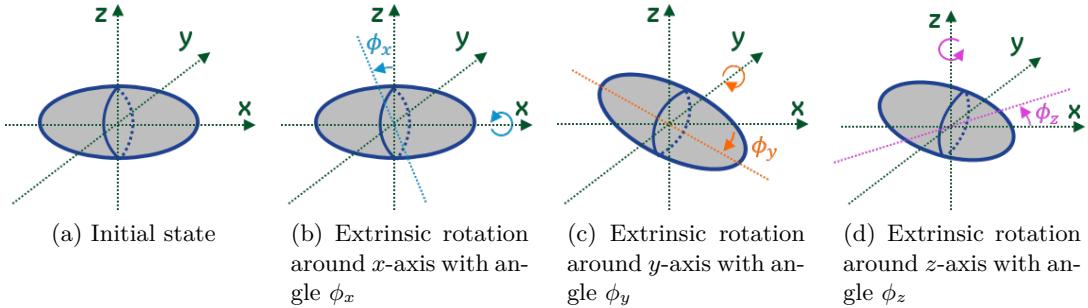


Figure 3.7: Rotation of an ellipsoid by chaining three elemental extrinsic rotations in $x \rightarrow y \rightarrow z$ order and in counterclockwise direction.

Let $\boldsymbol{\theta} = (l_a, l_b, l_c, \delta_{\phi_x}, \delta_{\phi_y}, \delta_{\phi_z})$ be a random vector with $l_a, l_b, l_c \in \mathbb{R}^+$ and $\delta_{\phi_x}, \delta_{\phi_y}, \delta_{\phi_z} \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Let $\mathbf{P}^{\boldsymbol{\theta}}$ be the joint distribution of $\boldsymbol{\theta}$.

For an ellipsoid \mathcal{E}_x centered at $x \in \mathbb{R}^3$, we assigned random values to the half lengths L_a, L_b and L_c of its principle axes and its rotation angles ϕ_x, ϕ_y and ϕ_z according to $\mathbf{P}^{\boldsymbol{\theta}}$ using the following method.

We first randomly sampled a vector $(l_a, l_b, l_c, \delta_{\phi_x}, \delta_{\phi_y}, \delta_{\phi_z})$ from \mathbf{P}^θ . The half lengths L_a, L_b and L_c were assigned directly with the value l_a, l_b, l_c respectively. The rotation angles ϕ_x, ϕ_y and ϕ_z were generated as follows. Using the nipple position q and the ellipsoid center x , we first computed the unit vector $\vec{u} = \frac{\vec{xq}}{\|\vec{xq}\|}$. Then from \vec{u} , we determined two polar angles $\overline{\phi_z}$ and $\overline{\phi_y}$. Here $\overline{\phi_z}$ is the polar angle in the x - y plane and $\overline{\phi_y}$ is the polar angle in the z - x plane, as illustrated in Figure 3.5. Finally, we set $\phi_x = \delta_{\phi_x}$, $\phi_y = \overline{\phi_y} + \delta_{\phi_y}$ and $\phi_z = \overline{\phi_z} + \delta_{\phi_z}$. Notice that for any given location x , $\overline{\phi_z}$ and $\overline{\phi_y}$ are deterministic and the only randomness is attributed to the random tilt angles $\delta_{\phi_x}, \delta_{\phi_y}$ and δ_{ϕ_z} sampled from \mathbf{P}^θ .

Discretization of Voronoi cells

Figure 3.8a illustrates the idea of how we implemented the discretization of Voronoi cells (step 5 in Algorithm 3.1). Consider a Voronoi cell \mathcal{C} and a voxel at an arbitrary position x of the discrete grid. The discretization process consists in determining whether x is part of the cell \mathcal{C} . For this, we first looped through all the faces of \mathcal{C} . For each face \mathcal{F} , we then computed its normal vector pointing inward \vec{n} , as well as the vector \vec{p} defined by the position x and the centroid of \mathcal{F} . Let $D(x)$ be the decision variable determining whether the voxel x belongs to the Voronoi cell \mathcal{C} (true) or not (false). Then $D(x)$ is formulated as:

$$D(x) = \begin{cases} \text{true} & \text{if } \vec{n} \cdot \vec{p} \geq -\epsilon, \text{ for all faces of } \mathcal{C}, \\ \text{false} & \text{otherwise,} \end{cases} \quad (3.1)$$

where \cdot is the inner product and $\epsilon \in \mathbb{R}^+$ is a precision parameter.

The reason that we needed to introduce a precision parameter ϵ was that, with the software for computing the Voronoi diagram (`Voro++`, version 0.4.6, [169]), we were not able to perfectly align the inward normals of the separating face of two adjacent Voronoi cells. As a consequence, in some cases voxels close to the separating face were assigned to neither of the two Voronoi cells. Such a situation is illustrated in Figure 3.8a. When such a situation occurs and when the value of ϵ is set to 0, we observed discretization residuals around the separating faces of adjacent Voronoi cells due to the missed assignments. The residuals are illustrated in Figure 3.8b. Therefore a positive number was set to ϵ in our simulations to ensure that the discretization residuals did not appear.

3.2.2 Impact of irregular adipose compartment boundaries

To demonstrate the impact of the irregular adipose compartment boundaries using small Voronoi cells on the improved visual realism of simulated images, we compared the images simulated with textures having smooth and irregular ellipsoid boundaries. Smooth boundaries were simulated by starting directly from the step 3 in Algorithm 3.1 and by modifying the step 5 to assign each time the value 0 to a voxel if it falls inside at least one ellipsoid.

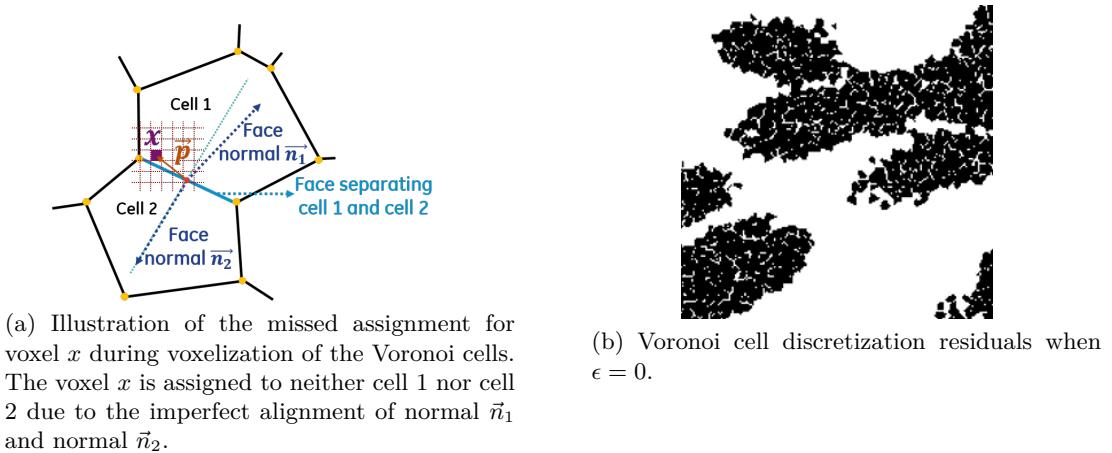


Figure 3.8: Illustration of the discretization of Voronoi cells and the effect of the precision parameter ϵ on Voronoi cell discretization residuals.

3.2.3 Empirical determination of model parameters

In the first instance, the seed point process Φ_s is set to be a homogeneous Poisson point process (*i.e.* all ellipsoid centers are independently uniformly distributed). Regarding the distribution \mathbf{P}^θ , we assume that $\mathbf{P}^\theta = (p_{L_a}, p_{L_b}, p_{L_c}, p_{\delta_{\phi_x}}, p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}})$ where p_{L_a} , p_{L_b} , p_{L_c} , $p_{\delta_{\phi_x}}$, $p_{\delta_{\phi_y}}$ and $p_{\delta_{\phi_z}}$ are independent distributions. Parameters were empirically determined for almost entirely adipose (BI-RADS breast density 1) breasts, scattered fibroglandular dense (BI-RADS breast density 2) breasts and heterogeneously dense (BI-RADS breast density 3) breasts. They are listed in Table 3.1.

3.3 Preliminary simulations using empirical model parameters

Figure 3.9 shows examples of $3.5 \text{ cm} \times 3.5 \text{ cm}$ slices through volumes simulated from the 3D breast texture model and segmented clinical bCT reconstructed volumes, as well as mammographic projections and DBT reconstructed slices simulated from these volumes. The segmented bCT VOIs have isotropic voxels with typical size between 0.2 mm and 0.4 mm. The voxel size ν for each simulated texture volume was adjusted to match bCT data resolution.

Mammograms and DBT projection images were simulated by virtually projecting the texture volume using the previously described breast x-ray imaging simulator [132]. The x-ray simulator was adjusted in the same way as described in Section 3.1.2. Mechanical breast texture deformation to mimic breast compression was not modeled for neither mammography and DBT simulation acquisitions. Projection images were processed by the ASIR-DBT 3D reconstruction algorithm (version 1.3.4, GE Healthcare, Buc, France) to obtain DBT reconstructed slices with 1 mm slice thickness.

Preliminary evaluations on ROIs of simulated mammograms and DBT reconstructed slices indicate a high visual realism compared to images simulated using bCT VOIs.

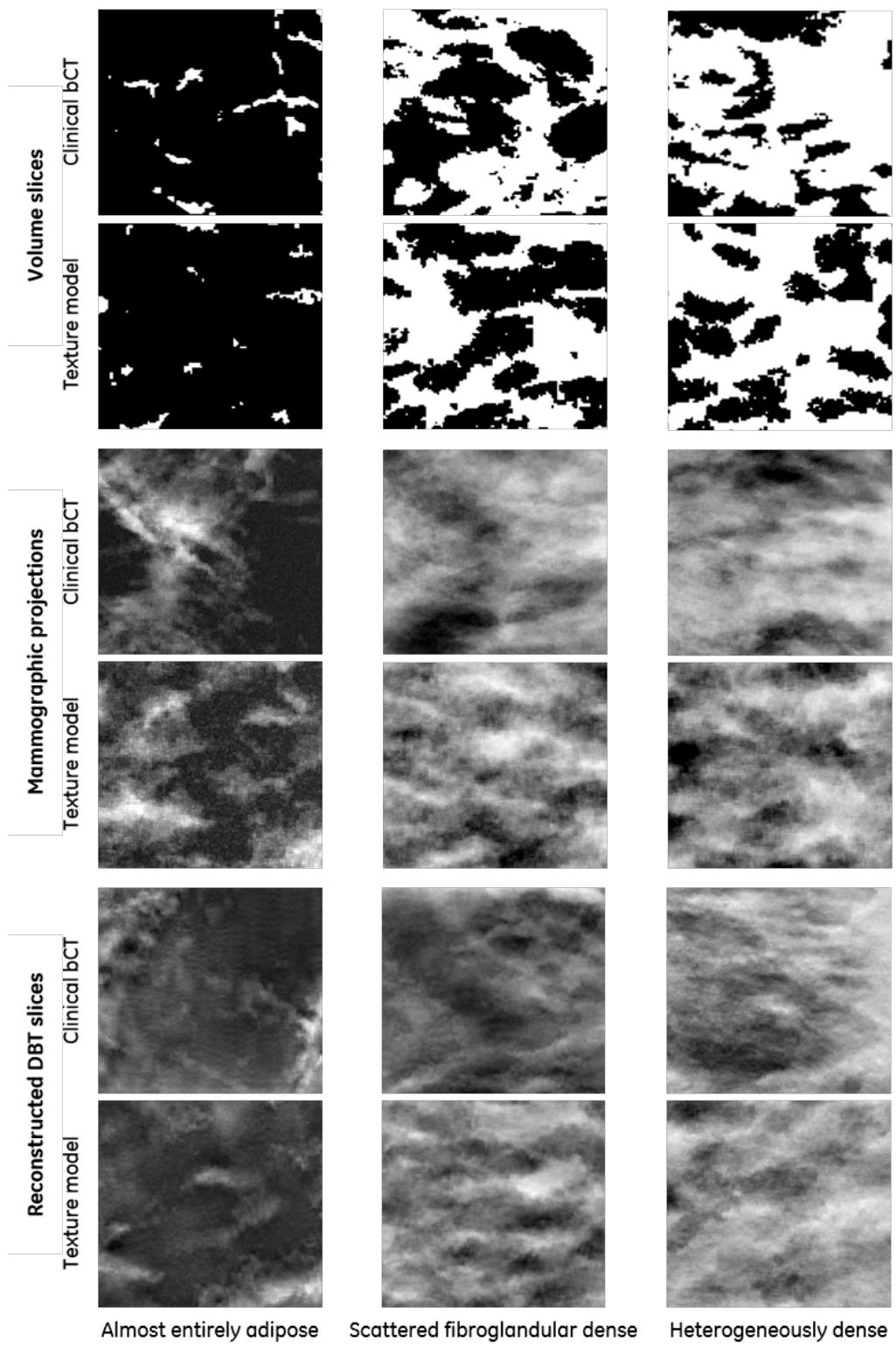


Figure 3.9: The top-rows show slices through volumes simulated from the 3D breast texture model and segmented clinical bCT reconstructed volumes. The middle-rows show mammographic projections simulated from these volumes. The bottom-rows show DBT reconstructed slices simulated from these volumes. The sizes of the images are $3.5\text{ cm} \times 3.5\text{ cm}$.

Parameter	BI-RADS density	BI-RADS density	BI-RADS density
	1	2	3
W	5 cm \times 5 cm \times 5 cm voxelized cube centered at the origin $o(0, 0, 0)$		
Voxel size ν	User defined		
Precision ϵ	0.3		
Nipple q	(0, 0, 40 mm)		
λ_0	2 mm^{-3} (<i>i.e.</i> the average volume of a Voronoi cell $\approx 0.5 \text{ mm}^3$)		
Φ_s intensity	$3.5 \times 10^{-3} \text{ mm}^{-3}$	$5 \times 10^{-3} \text{ mm}^{-3}$	$2.3 \times 10^{-3} \text{ mm}^{-3}$
$p_{L_a}, p_{L_b}, p_{L_c}$ (in mm)	$p_{L_a} = \mathcal{N}(7.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(4.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$
$p_{\delta_{\phi_x}}, p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}}$ (in radian)	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2}), p_{\delta_{\phi_y}}, p_{\delta_{\phi_y}} = \mathcal{N}(0, \frac{\pi}{12})$		

Table 3.1: Empirically determined 3D solid breast texture model parameters. Here $\mathcal{N}(\mu, \sigma)$ specifies a Gaussian distribution with mean μ , standard deviation σ , and $\mathcal{U}(l, u)$ specifies a uniform distribution with lower bound l and upper bound u .

Figure 3.10 shows examples of 2 cm \times 2 cm regions of interest of simulated mammograms and DBT reconstructed slices from a breast texture volume according to a heterogeneously dense breast. Ellipsoids with smooth boundaries and ellipsoids with irregular boundaries introduced by small Voronoi were considered. These examples illustrate that irregular adipose compartment boundaries may improve the visual realism of simulated images.

3.4 Psycho-physical realism assessment through two-alternative forced choice experiment

A two-alternative forced choice experiment was performed to assess the visual realism of simulated DBT reconstructed images.

3.4.1 Experimental set-up

Pairs of 3.5 cm \times 3.5 cm ROIs, extracted from images simulated from our texture model and clinical bCT data were displayed side-by-side. Images on the left always came from the clinical bCT dataset, while images on the right had 50% chance to be from the texture model and 50% chance to be from clinical bCT data. Similar level of glandular density was maintained for each image pair. In total, 144 image pairs were presented; 52 pairs from almost entirely adipose breasts, 56 pairs from scattered fibroglandular dense breasts and 36 pairs from heterogeneously dense breasts. Images were displayed on 5M pixels grayscale portrait monitors (SMD 21500 G, Siemens AG; Munchen, Germany) at 100% resolution.

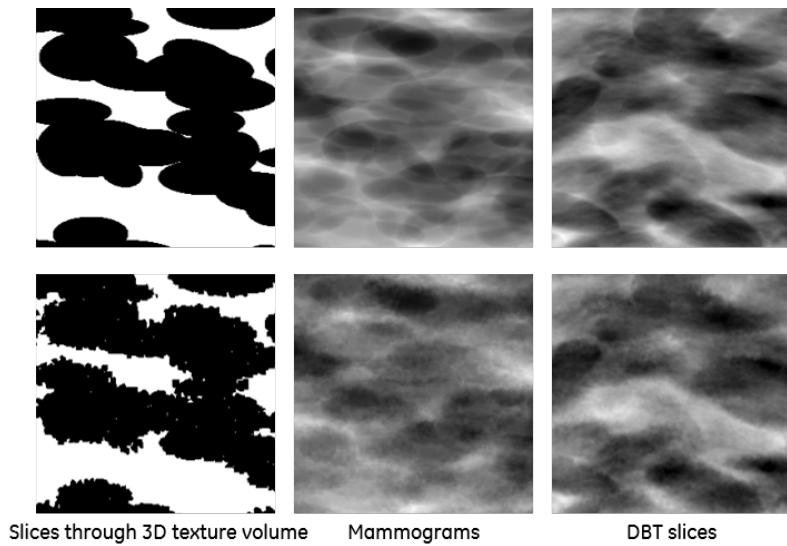


Figure 3.10: Illustration of the impact of the adipose compartment irregularities on the realism of simulated images. Slices of 3D breast texture volumes and ROIs of corresponding mammograms and DBT reconstructed slices are shown. Images simulated with irregular ellipsoid boundaries introduced by small Voronoi cells (bottom row) have improved small scale texture realism compared with images simulated with smooth ellipsoid boundaries (top row). All images have size 2 cm × 2 cm.

Four readers participated in the experiment, all GE Healthcare engineers. Reader 1, 2 and 3 had no prior knowledge of our texture model while reader 4 had been involved in the construction of the texture model sampling algorithm. The reader had to tell whether the image on the right was from the clinical bCT dataset or from the texture model. The reading distance was set to be one meter. Each image pair was displayed during five seconds, followed by the display of a uniform gray-level image for another five seconds; the readers were thus imposed to make a decision within ten seconds. A short training session with 10 image pairs of known ground truth was performed before the real experiment with no time constraint.

The percentage of correct answers, P_c , was calculated as an indication for the realism of the simulated images from the texture model. Under the null hypothesis that images simulated from the model and from clinical bCT data cannot be distinguished, $P_c = 0.5$.

3.4.2 Experiment results and analysis

Figure 3.12 shows the P_c value of the four readers for the two-alternative forced choice experiment. Readers 1 and 2 had overall P_c values of 0.65 and 0.66, while reader 3 had a P_c value of 0.73 indicating that he could more easily distinguish texture images from clinical bCT images. Reader 4 had the highest P_c value of 0.82. He reported that he was able to infer simulated images from his prior knowledge of the intermediate steps of the model construction.

Readers 1, 3 and 4 reported that they had observed a quantification effect, appearing like *small scale square blocks artifacts* in some simulated images from the texture model, as

illustrated in Figure 3.11. The artifacts appeared in particular around the area of the glandular and adipose tissue transitions. When this artifact was present, the readers used it as criterion to distinguish simulated images using the texture model from those using clinical bCT data. Reader 2 did not report the same primary criteria.

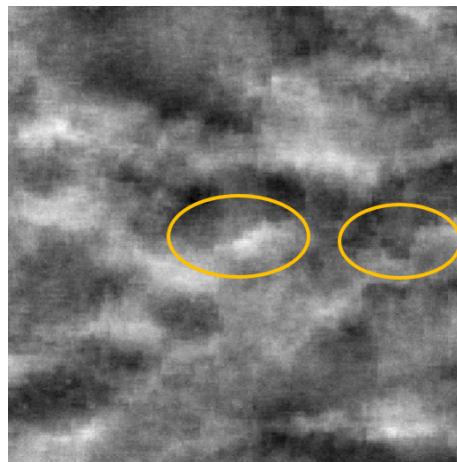


Figure 3.11: Small scale texture artifacts in a $3.5\text{ cm} \times 3.5\text{ cm}$ ROI of a simulated DBT slice from the 3D solid breast texture model.

Table 3.2 reports P_c values for each reader by type of breast density. The p -value indicates whether the reader was or was not able to differentiate images simulated from the texture model and images simulated from bCT data. Despite the reported small scale texture artifact, the P_c values for almost entirely adipose breasts were relatively low for all readers compared with other density types. This can be explained by the relatively lower number of glandular and adipose tissue transitions, thus fewer small block artifacts in this type of breasts.

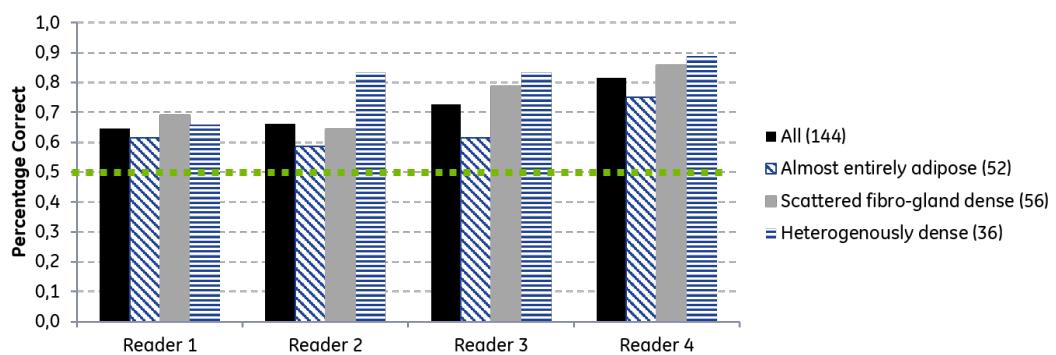


Figure 3.12: The P_c values of four readers for the two-alternative forced choice experiment, calculated together for images of all density types and separately for images of different density types.

	BI-RADS density 1	BI-RADS density 2	BI-RADS density 3	All
# of images	52	56	36	144
Reader 1	$P_c = 0.60,$ $p > 0.05$	$P_c = 0.69,$ $p < 0.05$	$P_c = 0.66,$ $p > 0.05$	$P_c = 0.65,$ $p < 0.05$
Reader 2	$P_c = 0.58,$ $p > 0.05$	$P_c = 0.64,$ $p < 0.05$	$P_c = 0.83,$ $p < 0.05$	$P_c = 0.66,$ $p < 0.05$
Reader 3	$P_c = 0.62,$ $p > 0.05$	$P_c = 0.79,$ $p < 0.05$	$P_c = 0.83,$ $p < 0.05$	$P_c = 0.73,$ $p < 0.05$
Reader 4	$P_c = 0.75,$ $p < 0.05$	$P_c = 0.86,$ $p < 0.05$	$P_c = 0.89,$ $p < 0.05$	$P_c = 0.82,$ $p < 0.05$

Table 3.2: The P_c values and p -values for each reader by type of breast density. At significance level 5%, $p < 0.05$ means the reader was able to differentiate simulated data from real data.

3.5 Removal of the small scale texture artifact

In order to improve the small scale texture realism, an investigation was conducted regarding the *small scale texture artifact* identified during the two-alternative forced choice experiment. Our investigation revealed that, the artifact was actually attributed to the following 2 aspects:

1. The precision factor ϵ (explained in Section 3.2.1) used to control the discretization in the voxelization (Step 5 in Algorithm 3.1): It was found that when ϵ was set to 0.3, some small Voronoi cells were deformed.
2. The size of the Voronoi cells: The average size of the Voronoi cells is inverse-proportional to the parameter λ_0 , *i.e.* the intensity of the initial Poisson point process. It was found that the value of $\lambda_0 = 2 \text{ mm}^{-3}$ used in our experiment was too small; thus leading to the introduction of unwanted artifacts.

The magnitude of ϵ parameter represents a trade-off between the quality of Voronoi cell voxelization and the deformation of Voronoi cells. As explained in Section 3.2.1, when the magnitude of ϵ is high, the discretization residuals are less present. However the Voronoi cells are more deformed, since we are less precise on the assignment of voxels around the faces of the Voronoi cells. Inversely, when the magnitude of ϵ is small, the Voronoi cells are less deformed, but the discretization residuals are more present. To demonstrate this trade-off, Figure 3.13 illustrates simulation results with three different magnitude values of ϵ : 0.05, 0.1 and 0.4. ROIs with size $1.5 \text{ cm} \times 1.5 \text{ cm}$ of slices through three simulated texture volumes are shown. The three volumes contain the same medium scale ellipsoids and the only difference was the value of ϵ used during the discretization of Voronoi cells. Isotropic voxels with size 0.1 mm were used for all texture volumes. Other texture volume parameters were the same as described in Table 3.1 for the BI-RADS 2 density. ROIs

of the same size of DBT reconstructed central slices, simulated from the three texture volumes are also illustrated under the corresponding volume slice. They were simulated using the same method described in Section 3.3.

Figure 3.13 demonstrates that the discretization residuals are less present when the magnitude of ϵ increases. The presence of too many discretization residuals seem to cause the appearance of grain-type noise in the DBT reconstructed slices, as seen in the lower left image. The Voronoi cell deformation are more deformed when the magnitude of ϵ increases. For a relative large ϵ value such as 0.4, the small scale block texture artifact appears very visible. We found that setting ϵ to 0.1 provides a satisfactory simulation result that minimizes both the grain-type noise caused by discretization residuals and small scale texture artifacts.

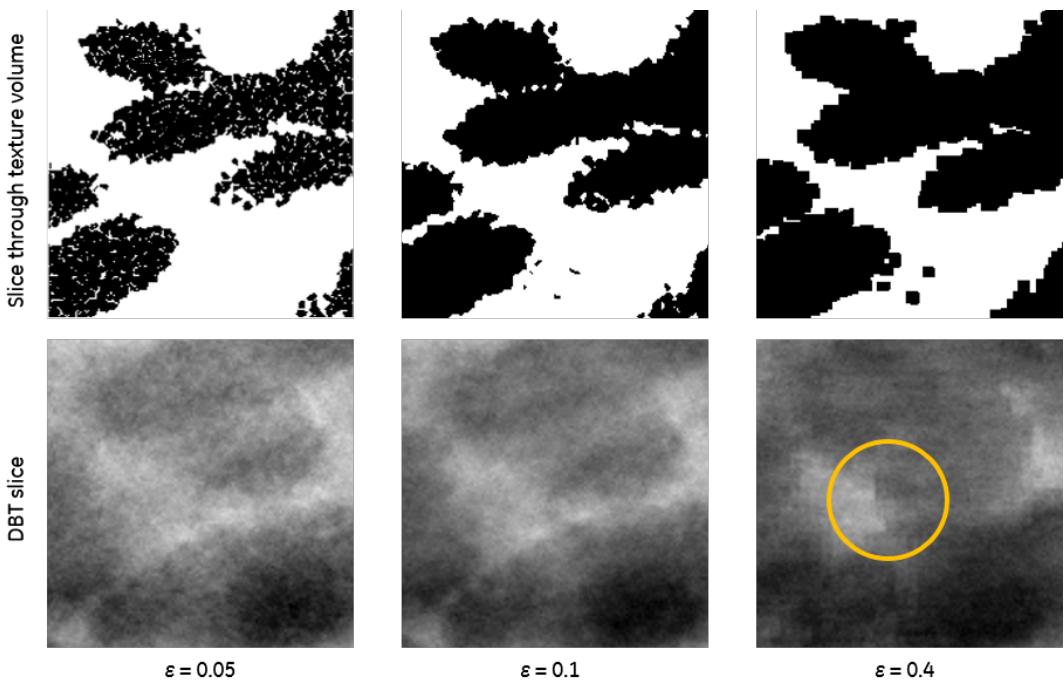


Figure 3.13: Effect of the precision parameter ϵ on Voronoi cell discretization residuals (top-row) and on the small block artifact in DBT reconstructed central slice ROIs (bottom-row). The small block artifact is emphasized with a yellow circle. Images were issued from simulated 3D texture volumes representing the BI-RADS 2 breast density. All images have size $1.5 \text{ cm} \times 1.5 \text{ cm}$.

Figure 3.14 illustrates the effect of λ_0 on the visual appearance of simulated DBT reconstructed slices. ROIs with size $3.5 \text{ cm} \times 3.5 \text{ cm}$ of the simulated DBT reconstructed slices from four simulated texture volumes are shown. The four volumes contain the same medium scale ellipsoids and the only difference was the value of λ_0 for the initial Poisson point process. Four λ_0 values: 10 mm^{-3} , 5 mm^{-3} , 3.3 mm^{-3} and 2 mm^{-3} were used. The ϵ value was set to 0.1 for all texture volumes to have optimal Voronoi cell voxelizations. Other texture volume parameters were the same as described in Table 3.1 for the BI-RADS 2 density.

To obtain a reasonable value for λ_0 , an informal observer experiment was conducted without rigorous control of the experimental conditions. Three readers were asked to

observe and compare the four different simulated DBT reconstructed ROIs using the DBT reconstructed slices simulated from segmented clinical bCT volumes shown in Figure 3.9 as the reference. Their opinions were asked regarding the level of small scale texture artifact in the four ROIs simulated from texture volumes compared with the reference images. All readers agreed that for $\lambda_0 \geq 5 \text{ mm}^{-3}$, there was no significantly visible small scale texture artifact.

Based on our investigation, the empirical parameters for our 3D solid breast texture models were adjusted. The final adjusted parameters are shown in Table 3.3.

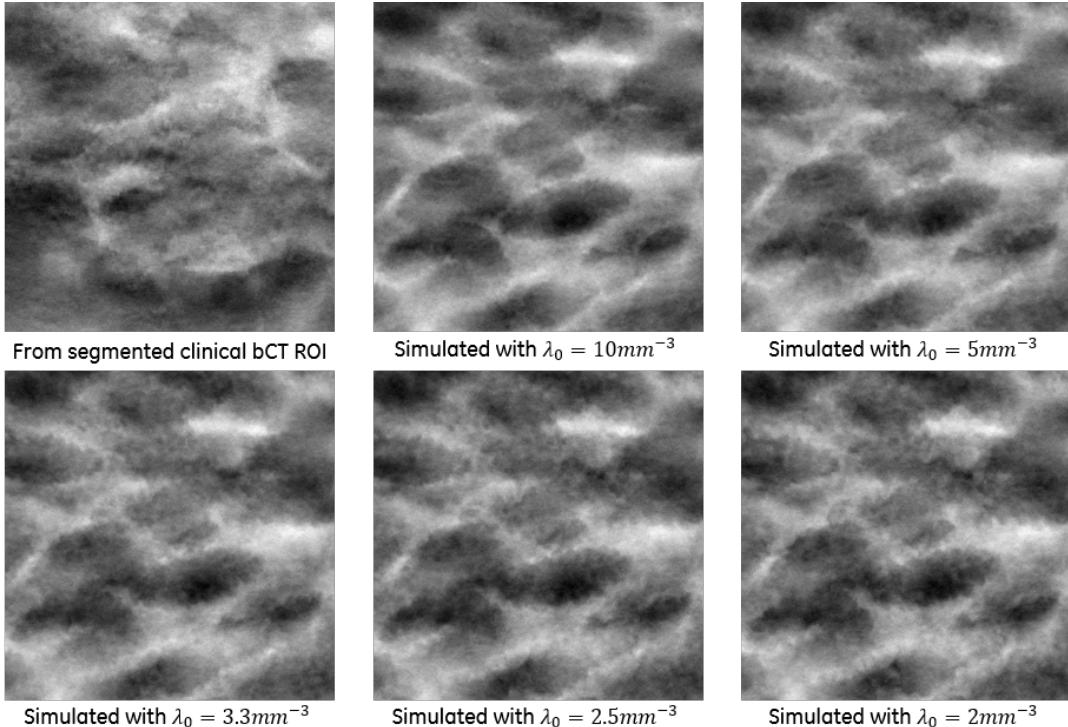


Figure 3.14: Effect of λ_0 on the small scale texture artifact in DBT reconstructed slices simulated from a 3D texture volume. All images have size $3.5 \text{ cm} \times 3.5 \text{ cm}$. Compared with the DBT reconstructed slices simulated from segmented clinical bCT volumes shown in Figure 3.9, there was no significantly visible small scale texture artifact in simulated DBT reconstructed slices for $\lambda_0 \geq 5 \text{ mm}^{-3}$.

3.6 Conclusion and discussion

In this chapter, an alternative 3D solid breast texture model based on segmented clinical breast computerized tomography (bCT) images was proposed for the small and medium scale fibroglandular and inter-glandular breast tissue. The model is constructed using stochastic geometric elements and can be formulated as a marked point process. Our model shares the advantage of 3D random field models in terms that it is analytically defined. Our model has the additional benefit that the visual realism of simulated DBT reconstructed slices is fair to high and a large variety of breast density types can be simulated.

Parameter	BI-RADS density		
	1	2	3
W	$5 \text{ cm} \times 5 \text{ cm} \times 5 \text{ cm}$ voxelized cube centered at the origin $o(0, 0, 0)$		
Nipple q	$(0, 0, 40 \text{ mm})$		
Voxel size ν	User-defined		
Precision ϵ	0.1		
λ_0	10 mm^{-3} (<i>i.e.</i> the average volume of a Voronoi cell $\approx 0.1 \text{ mm}^3$)		
Φ_s intensity	$3.5 \times 10^{-3} \text{ mm}^{-3}$	$5 \times 10^{-3} \text{ mm}^{-3}$	$2.3 \times 10^{-3} \text{ mm}^{-3}$
$p_{L_a}, p_{L_b}, p_{L_c}$ (in mms)	$p_{L_a} = \mathcal{N}(7.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(4.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$
$p_{\delta_{\phi_x}}, p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}}$ (in radians)	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2}), p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}} = \mathcal{N}(0, \frac{\pi}{12})$		

Table 3.3: Adjusted empirical model parameters to remove small scale texture artifacts. Here $\mathcal{N}(\mu, \sigma)$ specifies a Gaussian distribution with mean μ , standard deviation σ and $\mathcal{U}(l, u)$ specifies a uniform distribution with lower bound l and upper bound u .

The fibroglandular tissue model proposed by Graff [76] is in many ways similar to our model. The fibroglandular tissue of Graff's model is encapsulated in a large scale parenchymal cone. Voronoi tessellation is used to create 20 medium-scale fibroglandular compartments. Similarly, as in our model, Graff creates intra-glandular adipose tissue by embedding randomly shaped prolate spheroids with major axis oriented towards the nipple in the medium-scale glandular compartments. Graff does not detail the sample distribution of the ellipsoid centers. The system of random ellipsoids in our texture model is explicitly characterized as a marked point process. This characterization allows for a flexible assignment of different distributions to the ellipsoid centers and shape parameters. In addition, compared with the prolate spheroids used in Graff's phantom, ellipsoids may be asymmetric around the major axis, allowing for a larger morphological variability. To create a small-scale intra-glandular irregular boundary, Graff alters the spheroid surface shape using a Perlin noise function. We also perturbed the boundaries of the ellipsoids, however a different method, based on stochastic geometry, was applied: the ellipsoid surfaces are perturbed by small Voronoi cells. Another difference between Graff and our model may be the dimension of the surface perturbation. In our model, the Voronoi perturbations have volumes less than 0.1 mm^3 , while Graff does not specify the dimensions of the Perlin noise.

The statistical properties of simulated 2D and 3D breast images using the prototype models with empirical parameters were not investigated. This should be performed as a next step to further validate the prototype models. A limitation of our approach involves the model for the seed points, which is a homogeneous Poisson point process. The Poisson point process does not introduce interactions between the seed points. Due to this inherent characteristic, when comparing the simulated breast images with clinical breast images,

there is a limited morphological variability within each breast density category. Hence it is likely that the use of the Poisson point process for the seed points was inappropriate. To mitigate this limitation, the spatial interactions between the seed points should be further analyzed using clinical bCT data. Also, more advanced spatial point processes capable of modeling the spatial interactions between seed points should be investigated. In particular, a point process with clustering interactions might be worth investigating. Finally, it would also be interesting to compare simulations using the proposed 3D breast texture model with simulations using binary Gibbs models, such as the Ising model.

In this prototype implementation, the model parameters were empirically obtained by visually inspecting the segmented bCT volumes and the simulated images. A more objective method to directly infer model parameters from bCT volumes would be an approach of interest. This will be addressed in the next chapter.

Chapter 4

Statistical inference of medium scale 3D solid breast texture model parameters

Chapter 3 described a prototype implementation of a new 3D solid breast texture model allowing for the simulations of realistic DBT reconstructed slices representing a large variety of breast density types. However, further observations show that the simulated DBT slices have a limited morphological variability within each breast density type. To address this limitation, we propose in this chapter a methodology to objectively infer the medium scale parameters of the 3D breast texture model. The proposed inference method is based on an automatic analysis of the clinical dedicated breast computerized tomography (bCT) datasets. The inference aims to allow the 3D solid breast texture model to simulate a larger variability of breast textures. A larger texture variability is of interest in virtual clinical trials, where the impact of various breast tissue types on clinical task performance needs to be understood.

First, the parameter inference problem was mathematically formulated as a classical parametric marked point process inference problem. The challenges of the inference problem brought by the data unobservability was emphasized in particular.

We applied a two-step inference approach referred to as the *inference from reconstruction* method. The first step aims at reconstructing a set of random ellipsoids from an input binary bCT segmented dataset. We formulated the the reconstruction step as an optimization problem, where a dynamic Monte Carlo algorithm was proposed to find the optimal solution. The second step consists of proposing a parametric marked point process model based on statistical analysis of the centers, the half lengths and the orientation angles of the reconstructed ellipsoids. Based on the statistical analysis, we fitted a new cluster point process model to the reconstructed ellipsoid centers using the minimum contrast method. Independent distributions were also fitted to the half lengths and the orientation angles of the reconstructed ellipsoids based on the analysis of their empirical distributions.

The inference from reconstruction method was applied to the volumes of interests of a subset of clinical bCT datasets covering BI-RADS density from one to three. Finally, mammographic images and DBT reconstructed slices were generated from texture volumes simulated using inferred model parameters. New simulated images were visually compared to images generated in Chapter 3 to have a preliminary indication of the realism of the fitted models.

4.1 Problem statement

The parameters of our previously proposed 3D solid breast texture model can be further classified into two categories (Figure 3.6):

1. The parameters which are related to the *small scale* breast fibroglandular and inter-glandular adipose tissue. These parameters include the intensity λ_0 of the Poisson point process for the Voronoi cell centers and the precision parameter ϵ of the voxelization process.
2. The parameters which are related to the distribution and morphology of the *medium scale* breast fibroglandular and inter-glandular adipose tissue. These parameters include the distribution \mathbf{P}^s of the ellipsoid center point process Φ_s and the joint distribution \mathbf{P}^θ of θ , the random vector related to the shape and orientation of the random ellipsoids.

The effects of the small scale parameters λ_0 and ϵ were studied in the previous chapter. In Section 3.5, optimal values for the two small scale parameters were proposed and validated by an informal observer experiment.

In this chapter, we focus on obtaining objective medium scale parameters \mathbf{P}^s and \mathbf{P}^θ , based on a statistical inference from ground truth data. We start by mathematically formulating the inference problem.

4.1.1 Formulation of the medium scale texture model

To formally define the inference problem, we mathematically formulate the medium scale model as a *marked point process* (MPP) $\mathbf{Y} = \{\Phi_s, \theta\}$ defined on a product space $Y = \mathbb{R}^3 \times \mathbb{R}^6$ [45].

- The point process $\Phi_s \subset \mathbb{R}^3$ is referred to as the *ground point process*. It is a simple point process with distribution \mathbf{P}^s .
- The random vector $\theta \subset \mathbb{R}^6$ is referred to as the *marks*. In our case the marks specify random ellipsoids, whose union depicts the morphology and distribution of the medium scale adipose tissue in the 3D solid breast texture model. As discussed in Section 3.2.1, $\theta = (L_a, L_b, L_c, \delta_{\phi_a}, \delta_{\phi_b}, \delta_{\phi_c})$, where L_a, L_b, L_c are the half lengths of the principle axes of the ellipsoids and $\delta_{\phi_x}, \delta_{\phi_y}, \delta_{\phi_z}$ are random tilt angles of the ellipsoid. The distribution \mathbf{P}^θ is referred to as the *mark distribution*.

4.1.2 The ground truth

To statistically infer the medium scale model parameters, we used a subset of clinical dedicated breast computerized tomography (bCT) reconstructed volumes described in Section 3.1 as the input data. The bCT reconstructed volumes were processed by the 3D segmentation algorithm described in Section 3.1.1 to create the ground truth for our parameter inference.

To limit the inference to the medium scale breast tissue, only volumes of interest (VOI) with size $3.5\text{ cm} \times 3.5\text{ cm} \times 3.5\text{ cm}$ of the original segmented bCT volumes were considered. The VOIs were extracted following the same principles described in Section 3.1.2.

Mathematically, a segmented bCT VOI can be expressed as a binary volume \mathcal{D} , where

$$\mathcal{D}(x) = \begin{cases} 1 & \text{if the voxel } x \text{ is classified as fibroglandular tissue,} \\ 0 & \text{if the voxel } x \text{ is classified as adipose tissue,} \end{cases} \quad (4.1)$$

for all $x \in \Omega \subset \mathbb{Z}^3$. Here Ω represents the 3D discrete spatial domain of \mathcal{D} .

The inference problem therefore consists of first proposing parametric models for \mathbf{Y} , then optimizing the model parameters with respect to the ground truth \mathcal{D} .

As pointed out in Section 3.1.2, the segmented bCT VOIs depict adipose compartments from which it is impossible to discern the individual ellipsoids. This fact brings several challenges to the statistical inference. First, from the ground truth data \mathcal{D} , the positions of the ellipsoid centers are *unobservable*. This unobservability makes it difficult to propose an appropriate parametric model for the ellipsoid center point process. Also, depending on the type of breast, the boundaries of ellipsoids might also be unobservable from the adipose compartments in \mathcal{D} . This aspect suggests that multiple configurations of ellipsoids may be possible to approximate the adipose tissue in \mathcal{D} , making the inference problem ill-posed.

4.2 Existing parameter inference methods for marked point processes

Due to the challenges of the inference problem described in the previous section, direct application of commonly applied inference approaches such as the *likelihood*-based approaches [136] [11] [23] [94] is no longer possible, since they require the observability of the ellipsoid centers [51]. Several classical inference method used in stochastic geometry can potentially mitigate the issue of data unobservability.

The *minimum contrast estimation* (MCE) method can be configured such that the knowledge of point positions is no longer necessary.

Let Θ denote the set of all parameters for \mathbf{Y} . Let $\mathcal{C}(\mathcal{D}, \Theta)$ denote a contrast function depending on the ground truth \mathcal{D} and the parameters Θ . The minimum contrast estimator is defined as [45]:

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \mathcal{C}(\mathcal{D}, \Theta). \quad (4.2)$$

Often,

$$\mathcal{C}(\mathcal{D}, \Theta) = \iint_{Y \times Y} \left(S(y_1, y_2; \Theta) - \hat{S}(y_1, y_2; \mathcal{D}) \right)^2 dy_1 dy_2, \quad (4.3)$$

where $S(\cdot, \cdot; \Theta)$ is the analytical formula of one, or a weighted sum of several second-order summary statistics of \mathbf{Y} ; and $\hat{S}(\cdot, \cdot; \mathcal{D})$ denotes the empirical counterpart of $S(\cdot, \cdot; \Theta)$ measured from the ground truth \mathcal{D} . Popular choices of summary statistics are the two-point function [58] and the contact distribution function [85] etc.

The MCE method has limitations that make its application in our case difficult. The major limitation is the difficulty in the derivation of analytical summary statistics when

the underlying model \mathbf{Y} deviates from the Poisson MPP model. Previous applications of the MCE almost always assumed \mathbf{Y} to be a Poisson MPP with disc or spherical marks [58] [85]. These assumptions seem to be too strong in our inference problem. When the parametric form of \mathbf{Y} becomes complicated, the derivation of summary statistics may involve some numerical integration technique and the minimization can also be non-trivial.

Another statistical estimation approach based on the *Takacs-Fiksel estimation* method [51] has recently been proposed for a particular class of MPP named the *Quermass model* [98]. The Takacs-Fiksel estimation method consists in choosing a number of test functions that are analytically or computationally accessible and can be empirically measured from ground truth data even with unobservable point positions. The parameter estimations is then carried out by minimizing an objective function defined by the analytical test functions and their empirical measurements. The method has been applied to Quermass model of 2D discs for binary images [51] and was proven to give more satisfactory result than previously proposed methods [134].

The Takacs-Fiksel estimation method also has limitations. The authors in [51] pointed out that, in order to avoid the identifiability problem, the number of test functions should be much greater than the number of model parameters. Considering the number of parameters in our texture model, it might be challenging to find enough test functions to apply the TFE method. Also, when the number of parameters is too large, the TFE method might yield non-satisfactory results due to strong statistical variability [51].

4.3 The method of inference from reconstruction

Recently, Thiedmann *et al.* proposed a novel inference method for parametric MPP models using a two-step approach that can effectively mitigate the challenges in our inference problem caused by data unobservability [189]. Inspired by their work, we hereby formally present this method and refer to it as the *inference from reconstruction* method.

The inference from reconstruction method consists of two steps:

1. *The reconstruction step* aims at recovering the unobserved points and marks from the observed data through stochastic random sampling.
2. *The inference step* is a parametric inference step where a marked point process models is proposed and fitted to the reconstructed points and marks.

This two-step mechanism provides two advantages that classical MCE method and the TFE method do not share. First, the recovered points and marks provide intuitions to determine which model should be used for the inference step. For example, once the points and marks are available, we can visualize and analyze their summary statistics to study statistical properties of the reconstructed points and marks. Second, the inference step using reconstructed points and marks becomes more straightforward since all point positions are observable. For instance, the reconstructed points can be fitted separately from the marks, using methods described in Section 4.2, and the marks can be fitted later by separately analyzing their empirical statistics.

In the literature, limited studies have investigated the inference from reconstruction method for parameter estimation of stochastic geometric models. In [189], the authors first recon-

structed a set of random Boolean discs from the binary Li-ion battery images using a stochastic disc sampling algorithm. Then a planar Matérn elliptical cluster point process was fitted to the recovered disc centers using the minimum contrast method with pair correlation function. The distribution of disc radii was analyzed separately. In [51], Dereudre *et al.* has applied the reconstruction method described in [189] in order to facilitate the empirical measurements of test functions from the input images during the TFE procedure for the Quermass disc model. To our knowledge, the method of inference from reconstruction has not yet been investigated for more complicated MPP models such as our case.

Due to the previously described advantages, we decided to apply the inference from reconstruction method to fit a parametric MPP model to the ground truth segmented bCT data. In the following section we describe in detail the methodology used for the reconstruction step and the inference step in our study.

4.3.1 The reconstruction step: approximating segmented bCT data using ellipsoids

Consider the case where the density function $f_{\mathbf{Y}}$ of the underlying MPP model \mathbf{Y} has a Gibbsian form [135]. Let $\mathbf{u} \in Y$ be a set of ellipsoids, referred to as a configuration, then we have:

$$f_{\mathbf{Y}}(\mathbf{u}) = \frac{1}{Z} \exp\left(-\frac{1}{T}U(\mathbf{u})\right). \quad (4.4)$$

Here the term $U(\cdot)$ is called the *energy*. The parameter $T \in \mathbb{R}^+$ is often referred to as the temperature in physics. The term $Z = \int \exp(-U(\mathbf{u})) d\mathbf{u}$ is a normalization constant. The energy term is often decomposed into two parts [105] [53]:

$$U(\mathbf{u}) = \mathcal{L}(\mathbf{u}, \mathcal{D}) + \mathcal{P}(\mathbf{u}), \quad (4.5)$$

where $\mathcal{L}(\mathbf{u}, \mathcal{D})$ is a *data term* representing how well a configuration \mathbf{u} matches the observed dataset \mathcal{D} ; and $\mathcal{P}(\cdot)$ is a *prior term* containing the a-priori information of the underlying MPP model \mathbf{Y} . Finding the best set of ellipsoids \mathbf{u}^* is equivalent to solving the following optimization problem:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}} f_{\mathbf{Y}}(\mathbf{u}) = \arg \min_{\mathbf{u}} U(\mathbf{u}, \mathcal{D}) = \arg \min_{\mathbf{u}} (\mathcal{L}(\mathbf{u}, \mathcal{D}) + \mathcal{P}(\mathbf{u})). \quad (4.6)$$

We formulated $\mathcal{L}(\mathbf{u}, \mathcal{D})$ and $\mathcal{P}(\mathbf{u})$ as follows:

- The *data term* $\mathcal{L}(\mathbf{u}, \mathcal{D})$ quantifies the approximation quality of the current configuration \mathbf{u} with respect to the input \mathcal{D} . It consists of two terms \mathcal{L}_1 and \mathcal{L}_2 .

The first term \mathcal{L}_1 measures the total approximation error of the current configuration \mathbf{u} given input \mathcal{D} . For a given ellipsoid $\mathcal{E} \in \mathbf{u}$, its approximation error term $d_{\mathcal{E}}(\mathcal{D})$ can be expressed as:

$$d_{\mathcal{E}}(\mathcal{D}) = \frac{|\{x \in \Omega | x \in \mathcal{E} \text{ and } \mathcal{D}(x) = 1\}|}{|\{x \in \Omega | x \in \mathcal{E}\}|}. \quad (4.7)$$

The term \mathcal{L}_1 is then defined by summing the error terms of all ellipsoids. That is,

$$\mathcal{L}_1(\mathbf{u}, \mathcal{D}) = \sum_{\mathcal{E} \in \mathbf{u}} d_{\mathcal{E}}(\mathcal{D}). \quad (4.8)$$

The second term \mathcal{L}_2 measures the proportion of the input \mathcal{D} that is not covered by the current configuration \mathbf{u} . That is,

$$\mathcal{L}_2(\mathbf{u}, \mathcal{D}) = 1 - \frac{|\{x \in \Omega | x \in \mathbf{u} \text{ and } \mathcal{D}(x) = 0\}|}{|\{x \in \Omega | \mathcal{D}(x) = 0\}|}. \quad (4.9)$$

The sum of \mathcal{L}_1 and \mathcal{L}_2 defines the final data term \mathcal{L} . That is,

$$\mathcal{L}(\mathbf{u}, \mathcal{D}) = \mathcal{L}_1(\mathbf{u}, \mathcal{D}) + \mathcal{L}_2(\mathbf{u}, \mathcal{D}). \quad (4.10)$$

- Regarding the *prior term* $\mathcal{P}(\mathbf{u})$, we prefer to obtain a model estimate without imposing too much a-priori information on the distribution of the ellipsoids. Hence, only a weak constraint on the overlap ratio between ellipsoids is used to formulate $\mathcal{P}(\mathbf{u})$. That is:

$$\mathcal{P}(\mathbf{u}) = \sum_{\mathcal{E} \in \mathbf{u}} q(\mathcal{E}, \mathbf{u} \setminus \mathcal{E}), \quad (4.11)$$

where

$$q(\mathcal{E}, \mathbf{u} \setminus \mathcal{E}) = \begin{cases} 0 & \text{if } \frac{|\{x \in \Omega | x \in \mathcal{E} \text{ and } x \in \mathbf{u} \setminus \mathcal{E}\}|}{|\{x \in \Omega | x \in \mathcal{E}\}|} \leq 0.95, \\ +\infty & \text{otherwise.} \end{cases} \quad (4.12)$$

This means that the only a-priori information of the model \mathbf{Y} is that there are no ellipsoids that can have more than 95% percent of its volume covered by other ellipsoids in the same configuration.

The optimization given by (4.6) finally becomes:

$$\begin{aligned} \mathbf{u}^* &= \arg \min_{\mathbf{u}} U(\mathbf{u}, \mathcal{D}) \\ &= \arg \min_{\mathbf{u}} \mathcal{L}_2(\mathbf{u}, \mathcal{D}) + \sum_{\mathcal{E} \in \mathbf{u}} (d_{\mathcal{E}}(\mathcal{D}) + q(\mathcal{E}, \mathbf{u} \setminus \mathcal{E})). \end{aligned} \quad (4.13)$$

The analytical solution of (4.13) is difficult to obtain and the Markov chain Monte Carlo sampling is often applied to find the optimum [105]. In the literature, the *reversible jump Markov chain Monte Carlo* (RJMCMC) sampling is the most commonly applied method for MPP models [75] [77] [191] [53] [74] [105] [193] [182].

A typical RJMCMC procedure consists of iteratively simulating a Markov chain of configurations $\{\mathbf{u}_t\}_{t \in \mathbb{N}}$ that converges to the target density $f_{\mathbf{Y}}(\mathbf{u})$. At each iteration t , a modification of the current configuration \mathbf{u}_t is proposed to create the next configuration \mathbf{u}_{t+1} . The term “jump” refers to the fact that the cardinality, *i.e.* the number of marked points of the current configuration might change during the modification. A modification in RJMCMC is performed according to a density function $Q(\mathbf{u}_t, \mathbf{u}_{t+1})$, referred to as a proposition kernel. The modifications are local, in the sense that for each iteration only one or two marked points in the current configuration are modified. Typically, $Q(\mathbf{u}, \cdot)$ is a combination of several sub-proposition kernels:

$$Q(\mathbf{u}, \cdot) = \sum_n p_n Q_n(\mathbf{u}, \cdot), \quad (4.14)$$

where p_n is the probability of the occurrence of the sub-proposition kernels $Q_n(\cdot, \cdot)$, such that $\sum_n p_n \leq 1$ [52] [197]. Frequently investigated sub-proposition kernels include:

- The *Birth proposal*, in which a marked point is added to current configuration \mathbf{u}_t , according to a birth kernel denoted as $Q_b(\mathbf{u}, \cdot)$. That is, $\mathbf{u}_{t+1} = \mathbf{u}_t \cup \{u\}$, where u is drawn according to $Q_b(\mathbf{u}, \cdot)$.
- The *Death proposal*. This is the reverse process of the birth proposal, in which a marked point is chosen and deleted from current configuration \mathbf{u} according to a death kernel $Q_d(\mathbf{u}, \cdot)$. That is, $\mathbf{u}_{t+1} = \mathbf{u}_t \setminus \{u\}$, with $u \in \mathbf{u}_t$ chosen according to $Q_d(\mathbf{u}, \cdot)$.

The combination of birth and death proposals ensures that the Markov chain is able to switch between configurations with different cardinality.

- *Perturbation proposal*. This proposal consists of changing the parametrization of a marked point u in the current configuration \mathbf{u} according to a perturbation kernel $Q_p(\mathbf{u}, \cdot)$. There are various choices for the type of perturbation. Commonly applied choices include affine transformations of a mark [52], diffusion based on the Langevin equations [105], perturbation driven by observed data [191], or even switching to a different type of mark [105] [77].

In practice, the RJMCMC procedure might suffer from prohibitive rate of convergence, since each iteration brings only one or two marked points into play [52]. To mitigate this issue, a *multiple births and deaths* algorithm has recently been proposed by Descombes *et al.* [53], allowing for the births of multiple marked points at each iteration, offering a more effective sampling procedure. In our study, we adapted the multiple births and deaths algorithm design proposed in [53]. For each iteration, the original multiple births and deaths algorithm consisted of a birth step of multiple marked points and a death step that examines all marked points in current configuration. Additionally, to achieve a more effective exploration of each configuration, we proposed to add an extra perturbation step to the multiple births and deaths algorithm, named as the *shift* step.

To describe the shift step used in our adapted multiple births, deaths and shifts (MBDS) algorithm, we adopt the notion of *Legendre ellipsoid* of a convex body in classical mechanics [114]. Given a volume $K \subset \mathbb{Z}^3$, the Legendre ellipsoid $\mathcal{L}(K)$ is the unique ellipsoid defined as [113]:

$$\mathcal{L}(K) = \{x \in K | x^T \Sigma^{-1} x \leq 1\}, \quad (4.15)$$

where

$$\Sigma = \frac{\sum_{x \in K} (x - \mu)(x - \mu)^T}{|K|}, \quad (4.16)$$

with $\mu = \frac{1}{|K|} \sum_{x \in K} x$. Notice that $\mathcal{L}(K) = K$ if K is itself an ellipsoid.

In our proposed MBDS algorithm, the shift of an ellipsoid \mathcal{E} consists in replacing it by the Legendre ellipsoid computed from the part of the adipose tissue of the observed data \mathcal{D} inside \mathcal{E} . That is,

$$\mathcal{E} \rightarrow \mathcal{L}(K_{\mathcal{E}}(\mathcal{D})) \text{ with } K_{\mathcal{E}}(\mathcal{D}) = \{x \in \Omega | x \in \mathcal{E} \text{ and } \mathcal{D}(x) = 0\}. \quad (4.17)$$

The complete description of the proposed MBDS algorithm is given by Algorithm 4.1.

The decrease of temperature by a factor α in each iteration performs the simulated annealing [101] with geometric cooling schedule. Starting from a high temperature, the death probability of a marked point is relatively small. This means that the births and shifts

Algorithm: Multiple births, deaths and shifts

- **Initialization.**

Set the initial configuration $\mathbf{u}_0 = \emptyset$, the initial temperature $T = T_0$. Let Φ be a Poisson point process with initial intensity parameter $\lambda = \lambda_0$. Let f_θ be a multivariate proposal distribution used to sample $(L_a, L_b, L_c, \delta\phi_1, \delta\phi_2, \delta\phi_3)$, i.e. the marks.

- **Iterations**

Iterate the following steps in order, until convergence is reached.

- **Multiple births**

Generate a random configuration u_b with ellipsoid centers drawn from Φ and ellipsoid parameters drawn from f_θ . Update current configuration \mathbf{u} :

$$\mathbf{u} \rightarrow \mathbf{u} \cup u_b. \quad (4.18)$$

- **Computation of death probabilities**

For each ellipsoid $\mathcal{E} \in \mathbf{u}$, the death probability p_d is given by [53]:

$$p_d = \frac{r\lambda}{1 + r\lambda}, \quad (4.19)$$

with $r = \exp\left(\frac{U(\mathbf{u}) - U(\mathbf{u} \setminus \mathcal{E})}{T}\right)$, where $U(\cdot)$ is given in (4.13).

- **Deaths and shifts**

For each ellipsoid $\mathcal{E} \in \mathbf{u}$, draw a random variable $v \sim \text{uniform}(0, 1)$.

if $v < p_d$, perform a death operation: $\mathbf{u} \rightarrow \mathbf{u} \setminus \mathcal{E}$

else, shift \mathcal{E} to its Legendre ellipsoid: $\mathcal{E} \rightarrow \mathcal{L}(\mathcal{K}_{\mathcal{E}(\mathcal{D})})$, as described in (4.17).

- **Update parameters**

Decrease the Poisson intensity λ and the temperature T :

$$\lambda \rightarrow \lambda \cdot \alpha \text{ and } T \rightarrow T \cdot \alpha, \quad (4.20)$$

where $\alpha \in (0, 1)$.

- **Test of convergence**

For each iteration t , record its energy value and the energy values of its next nine successive iterations. This yields a set of ten energy values $E_t = \{U_t, U_{t+1}, \dots, U_{t+9}\}$. Then compute $\max E_t$ and $\min E_t$. If for a predefined small real number ϵ , $\max E_t - \min E_t \leq \epsilon$, then the convergence is considered reached.

Algorithm 4.1

are more favored than deaths in the beginning. This allows the Markov chain to broadly explore the solution space, and prevents it from getting stuck in local optima. As the temperature decreases, the death probability keeps increasing. The algorithm favors more deaths than births and shifts, since it assumes that the Markov chain is more concentrated on the global optimum as iterations advance.

The parameters for the MBDS algorithm are summarized in Table 4.1. Values for T_0 , λ_0 and α were empirically determined such that the convergence of the MBDS algorithm was obtained in a reasonable number of iterations (Section 4.4.1). We assumed that the proposal distribution $f_\theta = \{f_{L_a}, f_{L_b}, f_{L_c}, f_{\delta\phi_x}, f_{\delta\phi_y}, f_{\delta\phi_z}\}$, where f_{L_a} , f_{L_b} , f_{L_c} , $f_{\delta\phi_1}$, $f_{\delta\phi_2}$ and $f_{\delta\phi_3}$ are independent densities to sample the half lengths L_a , L_b , L_c and the tilt angles $\delta\phi_x$, $\delta\phi_y$, $\delta\phi_z$. We set these densities to the ones used in the empirical texture model for the BI-RADS breast density 2 type breasts (Section 3.2.3) because these previously validated values might provide a good starting-point for the MBDS algorithm and might accelerate the convergence of the algorithm.

Parameter	Value
T_0	100
λ_0	0.005
α	0.99
ϵ	0.001
f_{L_a}	$\mathcal{N}(6.4, 1.5)$
f_{L_b}, f_{L_c}	$\mathcal{N}(2.5, 0.7)$
$f_{\delta\phi_x}$	$\mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$
$f_{\delta\phi_y}, f_{\delta\phi_z}$	$\mathcal{N}(0, \pi/12)$

Table 4.1: Parameters of the MBDS algorithm to reconstruct a set of ellipsoids from each bCT VOI.

4.3.2 Inference step: fitting a marked point process model to reconstructed ellipsoids

Once a segmented bCT VOI is represented as a system of ellipsoids with known spatial positions and shape parameters, we can fit a parametric MPP model to the reconstructed ellipsoids using methods described in Section 4.2. To reduce the complexity of the inference step, we assume in the first instance that the marks of the MPP model is independent from the ground point process. This allows us to fit the center point process of the ellipsoids and their shape parameters separately.

The point process for reconstructed ellipsoid centers

To gain some intuition on the type of point process model Φ_s to fit to the reconstructed ellipsoid centers, we first conducted statistical analysis to extract distributional information of the ellipsoid centers. In stochastic geometry, the analysis of summary statistics is the most commonly performed technique for this purpose [45]. In this study, we chose to analyze the *pair correlation function* (PCF), a second-order statistical characteristic that can help reveal comprehensive structural information of a point process.

Definition 4.1. (Second order moment density, pair correlation function) [8, p34]

The pair correlation function $g(\cdot, \cdot)$ of a simple point process Φ with intensity function

$\lambda(\cdot)$ is defined as:

$$g(x, y) = \frac{\rho^{(2)}(x, y)}{\lambda(x)\lambda(y)}. \quad (4.21)$$

Here $\rho^{(2)}(\cdot, \cdot)$ is the second order moment density of Φ , satisfying

$$\int_{B_1 \times B_2} \rho^{(2)}(x, y) dx dy = \sum_{x, y \in \Phi}^{x \neq y} \mathbb{E}(\mathbf{1}(\{x, y\} \in B_1 \times B_2)), \quad (4.22)$$

for arbitrary bounded Borel sets B_1 and B_2 .

The PCF can be used to interpret important structural information of a point process. In particular,

- For all Poisson processes, $g(x, y) = 1$.
- If $g(x, y) > 1$, an *attraction* between points at locations x and y exists.
- If $g(x, y) < 1$, a *repulsion* between points at locations x and y exists.

For a stationary and isotropic point process, the PCF depends only on the relative distance between two spatial positions. That is, $g(x, y) = g(r)$, where $r = \|x - y\|$ is referred to as the *interpoint distance*. In our study, we assumed that the centers of the reconstructed ellipsoids come from a stationary and isotropic point process. Following this assumption, the PCF of reconstructed ellipsoid centers from each segmented bCT VOI was estimated, using the PCF estimator described in [91, p232]. Analytically, the PCF estimator is expressed as:

$$\hat{g}(r; \Phi_s) = \sum_{x, y \in \Phi_s \cap W}^{x \neq y} \frac{\mathbf{k}(\|x - y\| - r)}{4\pi r^2 \nu(W_x \cap W_y) \hat{\lambda}^2}, \quad (4.23)$$

where Φ_s is the collection of all ellipsoid centers reconstructed from a dataset and $\hat{\lambda}$ is an estimate of its intensity parameter, expressed as

$$\hat{\lambda} = \frac{|\Phi_s|}{\nu(W)}. \quad (4.24)$$

The function $\mathbf{k}(\cdot)$ is a smoothing kernel. We use $\nu(\cdot)$ to denote the volume measure and W is the observation window; *i.e.* a $3.5 \text{ cm} \times 3.5 \text{ cm} \times 3.5 \text{ cm}$ cube in our case. Finally, W_x denotes the translation of W by x . The division by $\nu(W_x \cap W_y)$ instead of by $\nu(W)$ acts as an edge correction for points falling outside of the observation window W [145].

The estimation was performed using the `pcf3est` function implemented in the R software package `spatstat` [12] with its default setting. In this setting the *Epanechnikov* smoothing kernel [45] was used. Mathematically, the *Epanechnikov* kernel is defined as:

$$\mathbf{k}(s) = \begin{cases} \frac{3}{4\delta} \left(1 - \frac{s^2}{\delta^2}\right) & \text{if } -\delta \leq s \leq \delta, \\ 0 & \text{otherwise.} \end{cases} \quad (4.25)$$

It has a bandwidth parameter δ to tune. In the default setting of `pcf3est`, δ is set according to the rule-of-thumb: $\delta = \sqrt[3]{\hat{\lambda}}$ [12].

Once the empirical estimates of PCFs were obtained, we first checked if the Poisson property can or can not be rejected based on the PCF estimates. This check was performed using the envelope test described in [9].

In spatial point process theory, the envelope test is a commonly used statistical test procedure to determine whether empirically estimated summary statistics deviate from the same theoretical summary statistics of a null hypothesis point process model [166] [9]. The test procedure consists in comparing the empirical summary statistics with the envelope of the same summary statistics obtained from several simulations of the null hypothesis model.

Let Φ be the null hypothesis model. Let S denote some second-order summary statistics. We assume that Φ is stationary and isotropic, hence S is a function of the interpoint distance r . To construct the envelope of S using Φ as the null hypothesis model, we first simulate N realizations of Φ , denoted as the set $\{\Phi_n\}_{n=1,\dots,N}$. For a given realization Φ_n , an empirical estimate of S is obtained, denoted as \hat{S}_n . Then we construct a point-wise envelope using the set of all empirical estimates $\{\hat{S}_n\}_{n=1,\dots,N}$ as follows. At each interpoint distance r , we rank the values in $\{\hat{S}_n(r)\}_{n=1,\dots,N}$ in ascending order. Then given a non-negative integer k , the envelope upper bound is assigned to the k^{th} highest value in $\{\hat{S}_n(r)\}_{n=1,\dots,N}$. Similarly, the envelope lower bound is assigned to the k^{th} lowest value in $\{\hat{S}_n(r)\}_{n=1,\dots,N}$.

Let Φ_b be a spatial point pattern and let the null-hypothesis be that Φ_b is drawn from the null hypothesis model Φ . Let \hat{S}_b be the empirical estimate of S using Φ_b . Then the envelope test is performed as follows. For a fixed interpoint distance value r , we reject the null-hypothesis at r , if $\hat{S}_b(r)$ falls outside of the envelope at r . It has been proved that this procedure consists in a two-sided point-wise Monte Carlo test with significance level $\frac{2k}{1+N}$ [88] [9].

We conducted the envelope test for each set of reconstructed ellipsoid centers. For a given set of reconstructed ellipsoid centers Φ_s , the null hypothesis model was the homogeneous Poisson point process with intensity parameter equal to $\hat{\lambda}$ estimated from Φ_s using (4.24). We chose $k = 5$ and $N = 199$, which yields a test significance level of 5%.

We will show in Section 4.4.2 that the analysis of the PCFs revealed a clustering interaction between reconstructed ellipsoid centers. To model the clustering interaction, we proposed to fit a three-dimensional *Matérn cluster process* [8] to the reconstructed ellipsoid centers. The 3D Matérn cluster process is a two-step process described as follows.

1. First, a set of “parent points” $\{y_i\}_{i \in \mathcal{I}} \subset \mathbb{R}^3$ are sampled from a homogeneous Poisson point process with intensity parameter κ . For each “parent point” y_i with $i \in \mathcal{I}$, a sphere with radius R centered at y_i is generated.
2. Then, inside each obtained sphere, a set of “children points” are sampled from another homogeneous Poisson point process with intensity parameter λ_0 . A realization of the Matérn cluster process is obtained as the collection of all “children points”.

We can see that a Matérn cluster process model Φ_M is a stationary and isotropic point process completely determined by the three parameters: κ , λ_0 , and R . Moreover, theoretical formula for the PCF of a 3D Matérn cluster process is analytically accessible and is given by the following proposition [91, p376].

Proposition 4.1. (Intensity and PCF of the Matérn cluster process)

Let Φ_M be a Matérn cluster process with parameters κ, λ, R , then its intensity parameter λ is:

$$\lambda = \frac{4}{3}\pi R^3 \kappa \lambda_0. \quad (4.26)$$

Its pair correlation function g is:

$$g(r; \kappa, \lambda_0, R) = \begin{cases} 1 + \frac{3}{8\pi\kappa R^6}(R - \frac{r}{2})^2(2R + \frac{r}{2}) & \text{if } 0 < r \leq 2R, \\ 1 & \text{if } r > 2R. \end{cases} \quad (4.27)$$

To fit a Matérn cluster process with parameters κ, λ, R to a set of reconstructed ellipsoid centers Φ_s , we applied the minimum contrast estimation (MCE) method described in Section 4.2. For a given Φ_s , the contrast function \mathcal{C} was computed based on the analytical and empirical PCF. That is,

$$\mathcal{C}(r; \kappa, \lambda_0, R, \Phi_s) = \sum_{r \in \mathcal{R}} (g(r; \kappa, \lambda_0, R) - \hat{g}(r; \Phi_s))^2, \quad (4.28)$$

where $\hat{g}(r; \Phi_s)$ is given in (4.23) and $g(r; \kappa, \lambda_0, R)$ is given in (4.27). Here \mathcal{R} denotes the set of interpoint distances considered by the MCE estimator. Additionally, (4.26) was used as an equality constraint. Let $\Theta = (\kappa, \lambda_0, R)$, for a given set of reconstructed ellipsoids Φ_s , the MCE estimator $\hat{\Theta}$ is formally expressed as:

$$\begin{aligned} \hat{\Theta} &= \arg \min_{\Theta} \sum_{r \in \mathcal{R}} (g(r; \kappa, \lambda_0, R) - \hat{g}(r; \Phi_s))^2, \\ &\text{subjected to: } \hat{\lambda} = \frac{4}{3}\pi R^3 \kappa \lambda_0. \end{aligned} \quad (4.29)$$

The optimization (4.29) was numerically solved using the function `fmincon` implemented in the **Matlab** software (version 2016b, The MathWorks Inc., Natick, Massachusetts, United States). The default setting of `fmincon` was used. In this setting, the *interior point* optimization method is applied, with the Hessian of the contrast function $\mathcal{C}(\cdot)$ approximated using the *Broyden–Fletcher–Goldfarb–Shanno* algorithm. We set \mathcal{R} to be a set of values increasing from 0.2 mm to 30 mm with a step size of 0.2 mm.

All optimizations were run with an initial condition $\kappa = 0.1$, $\lambda_0 = 0.1$ and $R = 1$. The convergence was considered reached when the contrast function was non-decreasing in all feasible directions, within the a tolerance value of $1e-6$. Since we were able to obtain fairly good fit for all input datasets (Section 4.4.2), the impact of different configurations of the optimization function was considered out-of-scope in our study and was not investigated.

The mark distribution

As discussed earlier, the mark distribution \mathbf{P}^θ was assumed independent from the ellipsoid center point process Φ_s . Based on this assumption, empirical statistics of each individual mark were examined separately from the ellipsoid centers.

Histograms of the half lengths L_a , L_b and L_c of the reconstructed ellipsoids were obtained to visualize the empirical distributions of L_a , L_b and L_c . For each half length of the reconstructed ellipsoids, the empirical mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ of were estimated. Next, two one-sample Kolmogorov-Smirnov tests [122] were performed for each axis.

- The first test has the objective to check if the distributions of the half lengths deviate from Gaussian distributions. It has the null-hypothesis that L_a , L_b and L_c follow Gaussian distributions with mean and standard deviation equal to the corresponding empirical estimates. The p -value of each test was computed and denoted as p_1 .
- The second test has the objective to check if the distributions of the half lengths deviate from the proposal distributions f_{L_a} , f_{L_b} , f_{L_c} specified in Table 4.1 respectively. This test has the null-hypothesis that L_a , L_b and L_c follow the proposal distributions f_{L_a} , f_{L_b} and f_{L_c} . The p -value of each test was computed and denoted as p_2 .

In a similar fashion, the histograms of the tilts angles $\delta\phi_x$, $\delta\phi_y$ and $\delta\phi_z$ of the reconstructed ellipsoids were also obtained. For each set of reconstructed ellipsoids, the empirical minimum and maximum of $\delta\phi_x$, as well as the empirical mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ of $\delta\phi_y$ and $\delta\phi_z$ were estimated. Similar to the axis lengths, two Kolmogorov-Smirnov tests were performed for each tilt angle.

- For $\delta\phi_x$, the first test has the objective to check if its distribution deviates from a uniform distribution. It has the null-hypothesis that $\delta\phi_x$ follows a uniform distribution with minimum and maximum equal to the empirical estimates. For $\delta\phi_y$ and $\delta\phi_z$, the first test has the objective to check if the two distributions deviate from Gaussian distributions. It has the null-hypothesis that $\delta\phi_y$ and $\delta\phi_z$ follow Gaussian distributions with mean and standard deviation equal to the empirical estimates. The p -value of each test was computed and denoted as p_1 .
- The second test has the objective to check if the distributions of the tilt angles deviate from the proposal distributions $f_{\delta\phi_x}$, $f_{\delta\phi_y}$ and $f_{\delta\phi_z}$ specified in Table 4.1 respectively. It has the null-hypothesis that $\delta\phi_x$, $\delta\phi_y$ and $\delta\phi_z$ follow the proposal distributions $f_{\delta\phi_x}$, $f_{\delta\phi_y}$ and $f_{\delta\phi_z}$ respectively. The p -value of each test was computed and denoted as p_2 .

Based on the empirical statistical analysis, independent distributions were proposed for each mark.

4.3.3 Input segmented bCT volumes of interest

A collection of 16 sets of segmented clinical bCT VOIs were selected for this study. The sizes of the selected bCT VOIs are $3.5 \text{ cm} \times 3.5 \text{ cm} \times 3.5 \text{ cm}$. The selected VOIs have BI-RADS density (the fourth edition, [160]) from category one to three and represent a subset of breasts with considerable variability in medium scale fibroglandular and inter-glandular adipose tissue. None of the VOIs contained a lesion. Table 4.2 summarizes the density and the size (in voxels) of the selected bCT VOIs, as well as and the size of the isotropic voxels of each VOI.

4.4 Results

In this section, we demonstrate the results of the inference from reconstruction method focusing on examples of four bCT VOIs (VOI #1, #3 #7, #11). We chose these four VOIs

VOI	Glandular density	Size (in voxels)	Voxel size
#1	BI-RADS 3 (57.2%)	99 × 99 × 99	0.357 mm
#2	BI-RADS 2 (40.6%)	99 × 99 × 99	0.354 mm
#3	BI-RADS 2 (25.1%)	94 × 94 × 94	0.376 mm
#4	BI-RADS 1 (23.5%)	103 × 103 × 103	0.341 mm
#5	BI-RADS 1 (17.4%)	103 × 103 × 103	0.343 mm
#6	BI-RADS 2 (37.2%)	100 × 100 × 100	0.351 mm
#7	BI-RADS 2 (32.9%)	112 × 112 × 112	0.314 mm
#8	BI-RADS 2 (29.0%)	96 × 96 × 96	0.366 mm
#9	BI-RADS 2 (25.7%)	97 × 97 × 97	0.361 mm
#10	BI-RADS 1 (13.9%)	101 × 101 × 101	0.349 mm
#11	BI-RADS 1 (14.5%)	128 × 128 × 128	0.275 mm
#12	BI-RADS 1 (09.8%)	87 × 87 × 87	0.405 mm
#13	BI-RADS 2 (27.1%)	107 × 107 × 107	0.329 mm
#14	BI-RADS 2 (25.0%)	87 × 87 × 87	0.405 mm
#15	BI-RADS 1 (20.5%)	124 × 124 × 124	0.283 mm
#16	BI-RADS 1 (17.2%)	126 × 126 × 126	0.279 mm

Table 4.2: Summary of the glandular density, the size (in voxels) and the size of the isotropic voxels of the selected 16 sets of segmented clinical bCT VOIs.

since their glandular densities cover BI-RADS breast density 1, BI-RADS breast density 2 and BI-RADS breast density 3 categories and they represent a fairly large variability of medium scale fibroglandular and inter-glandular adipose breast tissue. Further analyses were performed for all datasets listed in Table 4.2 and similar results were obtained.

4.4.1 The reconstruction step

The result of the reconstruction step is demonstrated using the reconstructed VOIs shown in Figure 4.1. Reconstructed VOIs are binary volumes having the same spatial and voxel size as their corresponding bCT VOIs. They were created by voxelizing the reconstructed ellipsoids from the MBDS algorithm and assigning value 0 to the ellipsoid interior. The coronal and sagittal slices of the original VOIs and the reconstructed VOIs at the same positions are shown in Figure 4.1. Moreover, projections images of the original VOIs and the reconstructed VOIs are also demonstrated. The projections were obtained by averaging the VOIs in the direction perpendicular to the transverse plane.

From Figure 4.1 we can see that the reconstructions by ellipsoids are not perfect. Despite this, the distribution and morphology of the medium scale fibroglandular and inter-glandular adipose tissue in reconstructed VOIs agree fairly well with the original segmented clinical bCT VOIs. Also, the medium scale texture variations in the projections of the original VOIs are preserved in the projections of the reconstructed VOIs. The result of the reconstruction step provides sufficiently good input for follow-up inference step which focuses only on the medium scale fibroglandular and inter-glandular tissue.

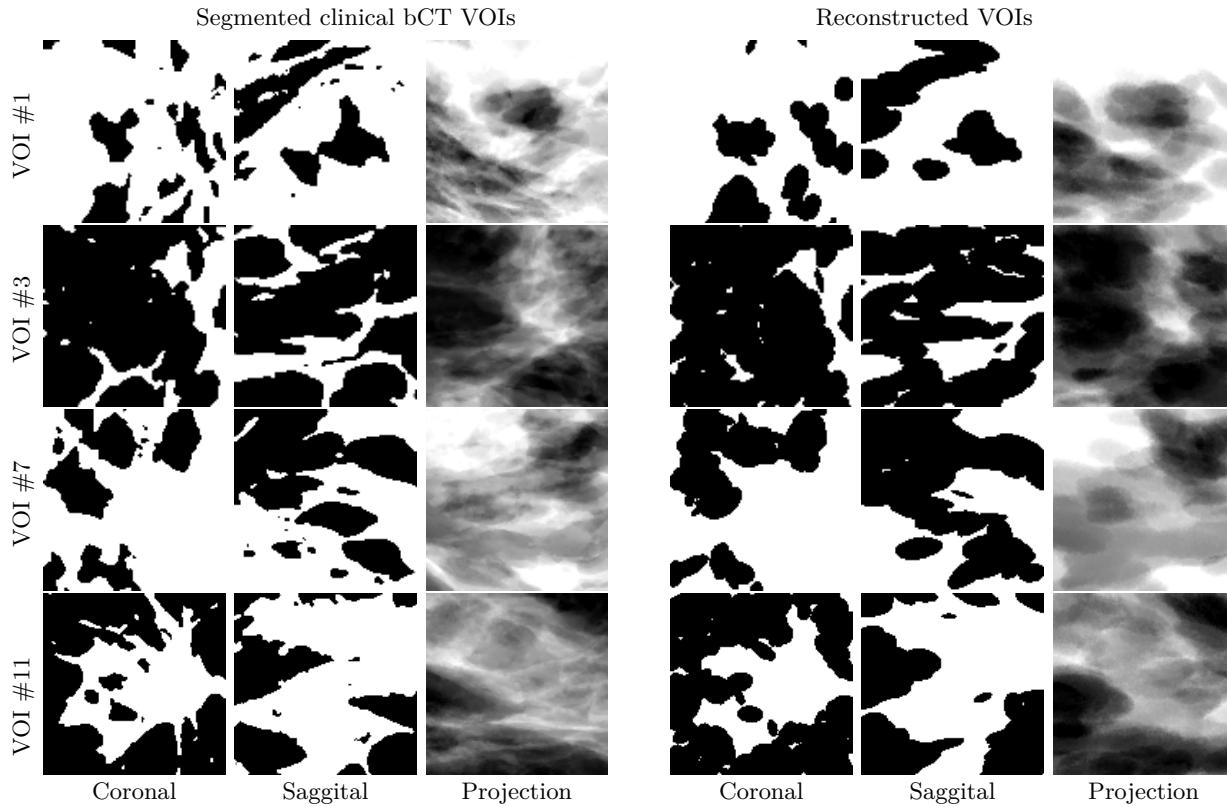


Figure 4.1: From left to right, the three columns on the left shows: coronal slices through four segmented clinical bCT VOIs, sagittal slices through the same bCT VOIs and projections of the same bCT VOIs. From left to right, the three columns on the right shows: coronal slices through the reconstructed VOIs, sagittal slices through the reconstructed VOIs and projections of the reconstructed VOIs. The sizes of the VOIs are $3.5\text{ cm} \times 3.5\text{ cm} \times 3.5\text{ cm}$. Reconstructed VOIs are binary volumes with the same size and resolution as their corresponding segmented clinical bCT VOIs. They were created by voxelizing the reconstructed ellipsoids from the MBDS algorithm and assigning value 0 to the ellipsoid interior. Projections images were obtained by averaging the VOIs in the direction perpendicular to the transverse plane. The distribution and morphology of the medium scale fibroglandular and inter-glandular adipose tissue in reconstructed VOIs agree fairly well with the original segmented clinical bCT VOIs

To demonstrate the convergence of the MBDS algorithm, Figure 4.2 shows the decrease of the energy with the increasing number of iterations for the four examples of bCT VOIs. From this figure we can see that the convergence was reached after about 7500 iterations for all the cases. Similar results were obtained for other investigated bCT VOIs.

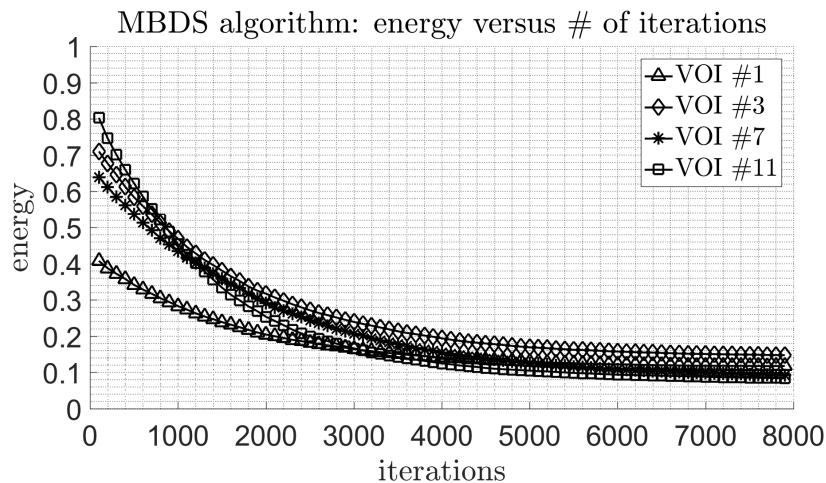


Figure 4.2: Illustration of the energy defined in (4.5) as a function of the number of iterations in the MBDS algorithm. The illustration considers the four examples of segmented clinical bCT VOIs shown in Figure 4.1. The convergence of the MBDS algorithm was reached after about 7500 iterations for all the cases.

4.4.2 The inference step

We first demonstrate the analysis of empirical PCFs estimated from the reconstructed ellipsoid centers for all selected bCT VOIs using methodologies described in Section 4.3.2.

Our analysis shows that, among the 16 selected bCT VOIs, there are four cases where the Poisson null-hypothesis can not be rejected. These four case were VOI #5, #12 #15 and #16. Figure 4.3 compares the empirical PCFs estimated from the reconstructed ellipsoid centers for these four bCT VOIs and the theoretical Poisson PCFs. The envelopes of the Poisson PCFs generated using the method described in Section 4.3.2 are also shown. Based on the envelope test described in Section 4.3.2, the Poisson null-hypothesis can not be rejected at any interpoint distance. Due to this finding, the VOI #5, #12 #15 and #16 were left out for the follow-up study since they represented a much smaller population among all bCT VOIs.

Figure 4.4 compares the empirical PCFs estimated from the reconstructed ellipsoid centers for the four bCT VOIs shown in Figure 4.1 and the theoretical Poisson PCFs. The envelopes of the Poisson PCFs generated using the method described in Section 4.3.2 are also shown. Figure 4.4 indicates that the empirical PCFs estimated from the reconstructed ellipsoid centers of the four VOIs exhibit a clustering effect. This can be seen if we look at the interpoint distance $r = 3\text{ mm}$ emphasized in Figure 4.4. The envelope test rejects the null-hypothesis for all demonstrated VOIs at $r = 3\text{ mm}$. The same clustering effect was found for all other investigated bCT VOIs.

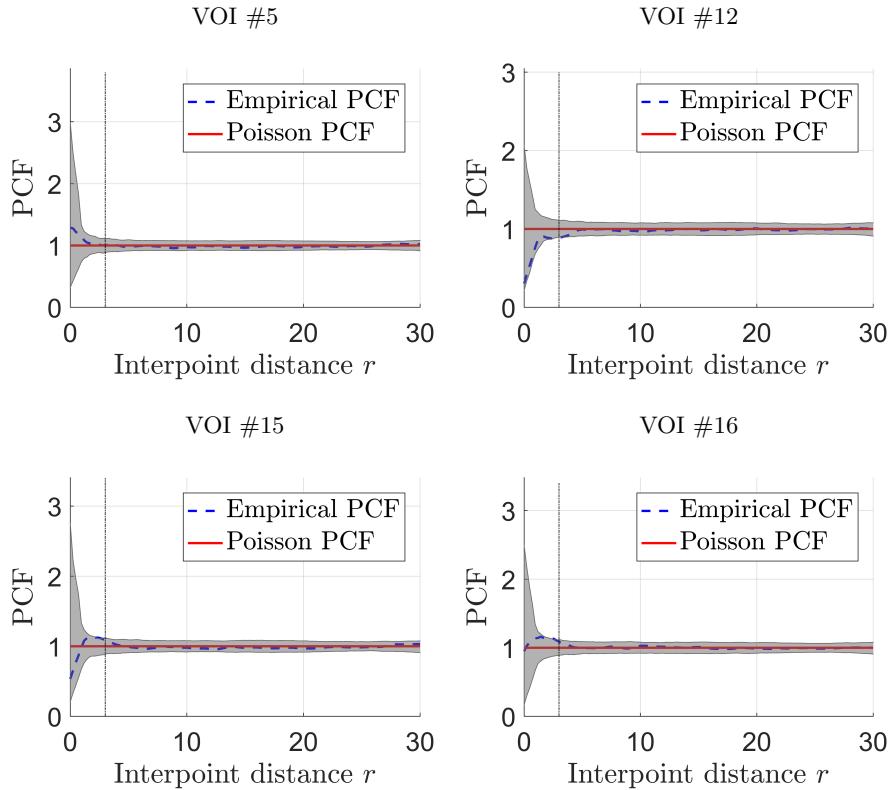


Figure 4.3: Comparison of the empirical PCFs estimated from the reconstructed ellipsoid centers for four bCT VOIs (#5, #12, #15 and #16) and the theoretical Poisson PCFs. Dashed lines are the empirical PCFs estimated from the reconstructed ellipsoid centers. Solid lines are theoretical Poisson PCFs. Gray surfaces are the envelopes of the Poisson PCFs generated using the method described in Section 4.3.2. Based on the envelope test described in Section 4.3.2, the Poisson null-hypothesis can not be rejected at any interpoint distance.

Parameters of the 3D Matérn cluster process

As discussed in Section 4.3.2, a 3D Matérn cluster process was fitted to each set of reconstructed ellipsoid centers using the MCE method. Table 4.3 lists the fitted Matérn cluster process parameters $\hat{\kappa}$, $\hat{\lambda}_0$ and \hat{R} for ellipsoid centers reconstructed from all bCT VOIs (excluding VOI #5, #12, #15 and #16).

Figure 4.5 compares the empirical PCFs estimated from the reconstructed ellipsoid centers for the four bCT VOIs shown in Figure 4.1 and the theoretical PCFs of the Matérn cluster processes fitted to the same VOIs. The PCF envelopes of the fitted Matérn cluster processes were also generated using the method described in Section 4.3.2, but with the fitted Matérn cluster process as the null hypothesis model. It can be seen that for all VOIs, the empirically estimated PCFs fall inside the envelopes at all interpoint distances. Based on this, we conclude that the fits are fairly good.

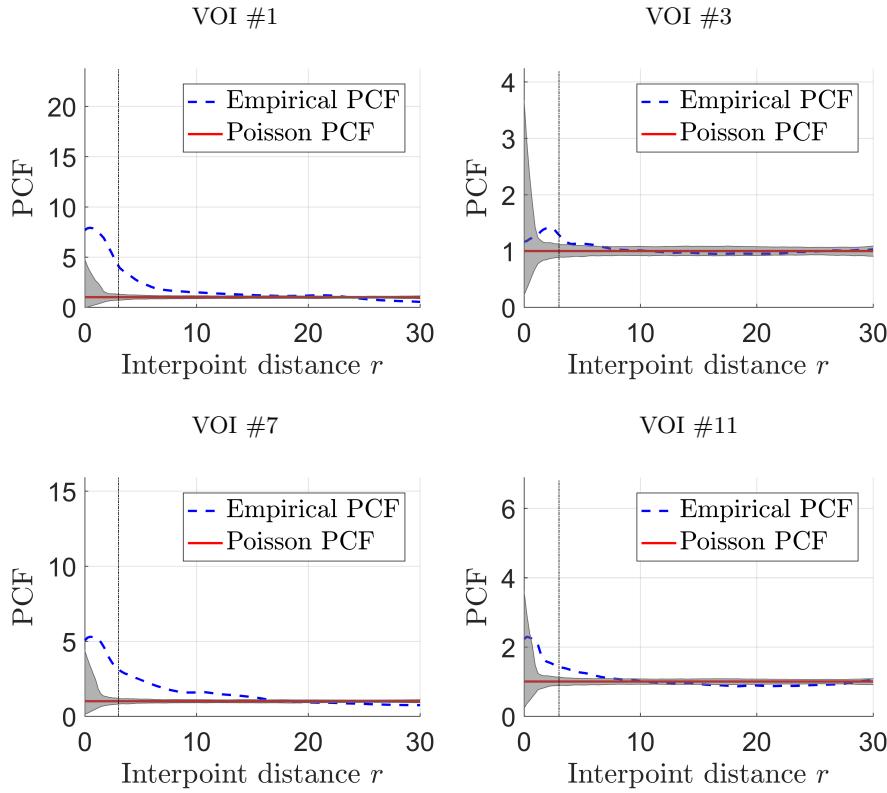


Figure 4.4: Comparison of the empirical PCFs estimated from the reconstructed ellipsoid centers for the four bCT VOIs shown in Figure 4.1 and the theoretical Poisson PCFs. Dashed lines are the empirical PCFs estimated from the reconstructed ellipsoid centers. Solid lines are theoretical PCFs of the Poisson point processes. Gray surfaces are the envelopes of the Poisson PCFs generated using the method described in Section 4.3.2. The envelope test rejects the Poisson null-hypothesis for all four VOIs at interpoint distance $r = 3$ mm.

VOI	$\hat{\kappa}$	$\hat{\lambda}_0$	\hat{R}
# 1	4.237e-04	2.812e-02	4.221
# 3	3.245e-03	5.985e-03	5.988
# 7	2.876e-04	3.092e-02	5.827
# 11	3.405e-03	1.923e-02	3.853

Table 4.3: Fitted Matérn cluster process parameters for the four bCT VOIs shown in Figure 4.1.

The mark distribution

Figure 4.6 shows the histograms of the half lengths L_a , L_b and L_c of the reconstructed ellipsoids for the four bCT VOIs shown in Figure 4.1. The empirical mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$, as well as the p_1 and p_2 values of the Kolmogorov-Smirnov tests (described in Section 4.3.2) for each half length are shown under the histograms.

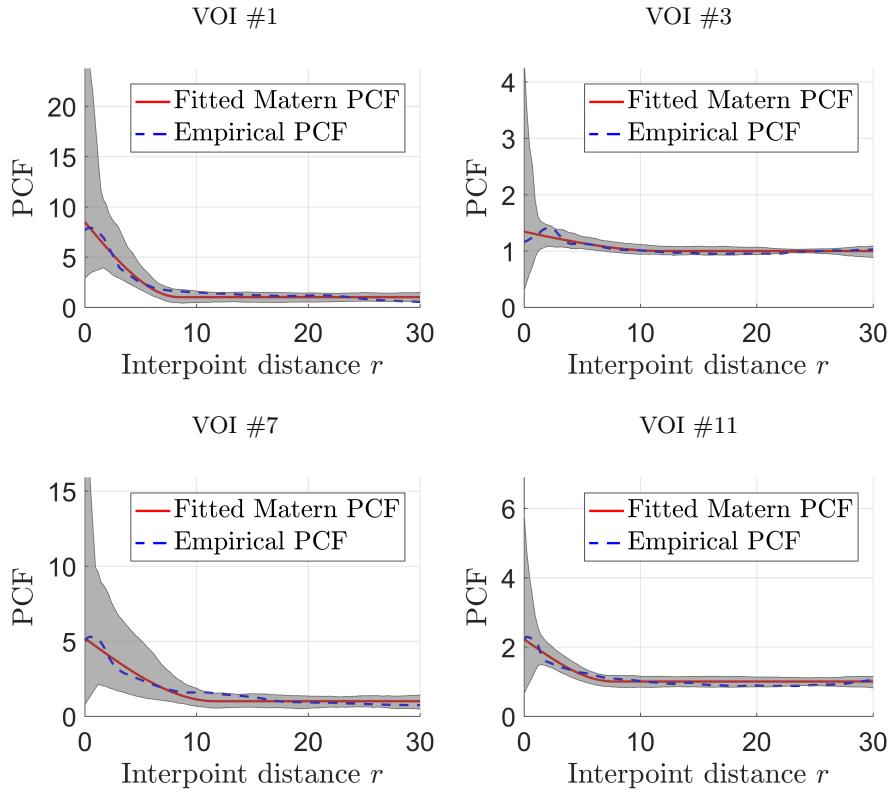


Figure 4.5: Comparison of the empirical PCFs estimated from the reconstructed ellipsoid centers for the four bCT VOIs shown in Figure 4.1 and the theoretical PCFs of the Matérn cluster processes fitted to the same VOIs. Dashed lines are the empirical PCFs estimated from the reconstructed ellipsoid centers of the four VOIs. Solid lines are theoretical PCFs of the Matérn cluster processes fitted to the four VOIs. Gray surfaces are the envelopes of the PCFs of the fitted Matérn cluster processes generated using the method described in Section 4.3.2. The empirical PCFs fall inside the envelopes at all interpoint distances.

According to the p_1 values, L_a , L_b and L_c remain to be Gaussian distributed for all cases at significance level 5%. However, according to the p_2 values the distributions of L_a , L_b and L_c deviate significantly from the proposal distributions f_{L_a} , f_{L_b} and f_{L_c} respectively, for all cases at the same significance level.

Figure 4.7 shows the histograms of the tilts angles $\delta\phi_x$, $\delta\phi_y$ and $\delta\phi_z$ of the reconstructed ellipsoids for the four bCT VOIs shown in Figure 4.1. The empirical minimum and maximum for $\delta\phi_x$, the empirical mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ for $\delta\phi_y$ and $\delta\phi_z$, as well as the p_1 and p_2 values of the Kolmogorov-Smirnov tests (described in Section 4.3.2) for each tilt angle are shown under the histograms.

According to the p_1 and p_2 values for $\delta\phi_x$, the distribution of $\delta\phi_x$ does not deviate significantly from the uniform distribution $\mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ for all cases at significance level 5%. According to the p_1 values for $\delta\phi_y$, $\delta\phi_y$ remains to be Gaussian distributed in three of the four cases at significance level 5%. However, according to p_2 values for $\delta\phi_y$, the distribution of $\delta\phi_y$ deviates significantly from the proposal distribution $f_{\delta\phi_x}$ in three of the four cases at significance level 5%. According to the p_1 values for $\delta\phi_z$, $\delta\phi_z$ remains to be

Gaussian distributed in two of the four cases at significance level 5%. However, according to the p_2 values for $\delta\phi_z$, the distribution of $\delta\phi_z$ deviates significantly from the proposal distribution $f_{\delta\phi_x}$ for all cases at significance level 5%. Similar observations were obtained for all other investigated bCT VOIs.

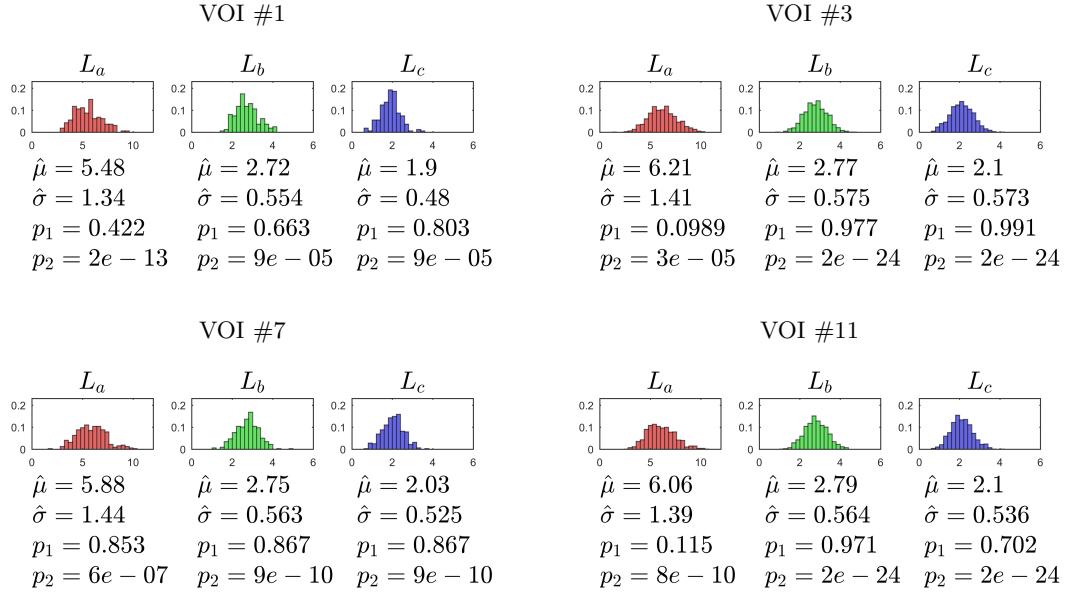


Figure 4.6: Histograms of the half lengths L_a , L_b and L_c of reconstructed ellipsoids for the four bCT VOIs shown in Figure 4.1. For each half length, $\hat{\mu}$ and $\hat{\sigma}$ refer respectively to their empirical mean and standard deviation. For each half length, p_1 and p_2 refer to the p -values of the two Kolmogorov-Smirnov tests described in Section 4.3.2.

Based on these analyses, we proposed to model the distribution \mathbf{P}^θ of the marks $\theta = (L_a, L_b, L_c, \delta\phi_x, \delta\phi_y, \delta\phi_z)$ as independent densities p_{L_a} , p_{L_b} , p_{L_c} , $p_{\delta\phi_x}$, $p_{\delta\phi_y}$ and $p_{\delta\phi_z}$ respectively. These densities are specified as follows.

The distribution of the half lengths of the ellipsoids, namely p_{L_a} , p_{L_b} and p_{L_c} were all modeled as Gaussian distributions. The mean and standard deviation of p_{L_a} , p_{L_b} and p_{L_c} were determined by the values empirically measured from the ellipsoids reconstructed from the clinical bCT VOIs. The mean and standard deviation values were specified in Figure 4.6.

The distribution of the tilt angles $\delta\phi_x$, $p_{\delta\phi_x}$ was model as the uniform distribution $\mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, as suggested by its empirical distribution demonstrated in the previous section.

The distributions $p_{\delta\phi_y}$ and $p_{\delta\phi_z}$ were modeled as Gaussian distributions. Despite the fact that in some of the cases, the Gaussian hypothesis was rejected by the Kolmogorov-Smirnov tests, as mentioned in the previous section, in other majority of the cases the Gaussian hypothesis could not be rejected. Visual inspection of the histograms of $p_{\delta\phi_y}$ and $p_{\delta\phi_z}$ also suggests that the Gaussian distribution might still be a fairly reasonable choice. Similar to the half lengths, the mean and standard deviation of $p_{\delta\phi_y}$ and $p_{\delta\phi_z}$ are determined by the empirical estimates.

We finally obtained twelve new sets of medium scale parameters for the 3D solid breast texture model from the twelve input segmented bCT VOIs. Table 4.4 lists the fitted

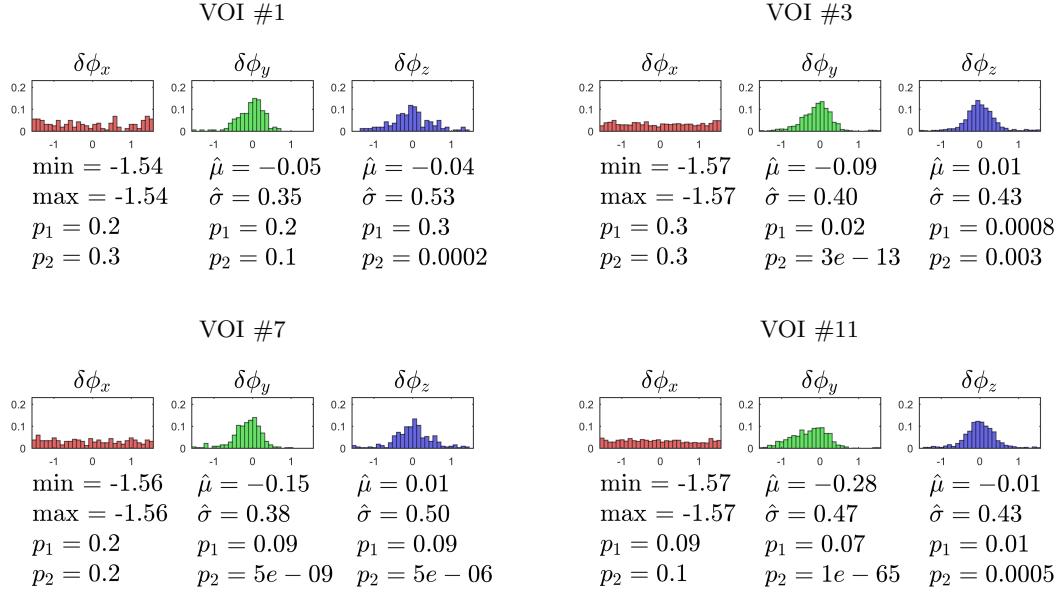


Figure 4.7: Histograms of the tilt angles $\delta\phi_x$, $\delta\phi_y$ and $\delta\phi_z$ of reconstructed ellipsoids from the four bCT VOIs shown in Figure 4.1. For $\delta\phi_x$, min and max refer respectively to its empirical minimum and maximum. For $\delta\phi_y$ and $\delta\phi_z$, $\hat{\mu}$ and $\hat{\sigma}$ refer respectively to their empirical mean and standard deviation. For each tilt angle, p_1 and p_2 refer to the p -values of the two Kolmogorov-Smirnov tests described in Section 4.3.2.

medium scale parameters for the four bCT VOIs shown in Figure 4.1. Parameters fitted using the other eight bCT VOIs can be found in Appendix C.

VOI	Φ_s (Matérn cluster process)	L_a, L_b, L_c (in mm)	$\delta\phi_x, \delta\phi_y, \delta\phi_z$ (in radian)
#1	$\kappa = 4.237e - 04$, $\lambda_0 = 2.812e - 02$, $R = 4.221$	$f_{L_a} = \mathcal{N}(5.48, 1.34)$, $f_{L_b} = \mathcal{N}(2.72, 0.55)$, $f_{L_c} = \mathcal{N}(1.90, 0.48)$	$p_{\delta\phi_x} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta\phi_y} = \mathcal{N}(-0.05, 0.35)$, $p_{\delta\phi_z} = \mathcal{N}(-0.04, 0.53)$
#3	$\kappa = 3.245e - 03$, $\lambda_0 = 5.985e - 03$, $R = 5.988$	$f_{L_a} = \mathcal{N}(6.21, 1.41)$, $f_{L_b} = \mathcal{N}(2.77, 0.58)$, $f_{L_c} = \mathcal{N}(2.10, 0.57)$	$p_{\delta\phi_x} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta\phi_y} = \mathcal{N}(-0.09, 0.4)$, $p_{\delta\phi_z} = \mathcal{N}(0, 0.26)$
#7	$\kappa = 2.876e - 04$, $\lambda_0 = 3.092e - 02$, $R = 5.827$	$f_{L_a} = \mathcal{N}(5.88, 1.44)$, $f_{L_b} = \mathcal{N}(2.75, 0.56)$, $f_{L_c} = \mathcal{N}(2.03, 0.52)$	$p_{\delta\phi_x} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta\phi_y} = \mathcal{N}(-0.15, 0.38)$, $p_{\delta\phi_z} = \mathcal{N}(0.01, 0.5)$
#11	$\kappa = 3.405e - 03$, $\lambda_0 = 1.923e - 02$, $R = 3.853$	$f_{L_a} = \mathcal{N}(6.06, 1.39)$, $f_{L_b} = \mathcal{N}(2.79, 0.56)$, $f_{L_c} = \mathcal{N}(2.10, 0.54)$	$p_{\delta\phi_x} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta\phi_y} = \mathcal{N}(-0.28, 0.47)$, $p_{\delta\phi_z} = \mathcal{N}(-0.01, 0.43)$

Table 4.4: Medium scale parameters for the 3D stochastic breast texture model fitted from the four bCT VOIs shown in Figure 4.1.

4.4.3 Simulations using fitted medium scale texture model parameters

As a preliminary validation of the fitted medium scale texture model parameters, a simulation experiment in analogy to Section 3.3 was conducted.

Four texture volumes were simulated using the four sets of medium scale parameters listed in Table 4.4. The small scale textures in these volumes were also simulated using parameters listed in Table 3.3. The nipple position and the voxel size for each volume were set to be the same as in the corresponding ground truth bCT VOI. Mammograms and DBT projection images were simulated by virtually projecting the texture volumes using the previously described breast x-ray imaging simulator [132], with the same adjustment as described in Section 3.1.2. Mechanical breast texture deformation to mimic breast compression was not modeled for neither mammography and DBT simulation acquisitions. Projection images were processed by the ASIR-DBT 3D reconstruction algorithm (version 1.3.4, GE Healthcare, Buc, France) to obtain DBT reconstructed slices with 1 mm slice thickness.

Figure 4.8 shows examples of 3.5 cm × 3.5 cm slices through volumes simulated from the 3D breast texture model with the four sets of medium scale parameters listed in Table 4.4, as well as mammographic projections and DBT reconstructed slices simulated from these volumes. Images simulated with the other eight sets of parameters can be found in Appendix C.

Visual inspection of the simulated images indicates high visual realism compared with the images simulated from segmented clinical bCT VOIs, shown in Figure 3.3. Also, compared with images simulated using the prototype implementation shown in Figure 3.9, the new model parameters are capable of simulating mammographic projections and DBT reconstructed slices with a larger morphological variability. These observations indicate an improvement of the model's realism and morphological variability compared with the prototype implementation with empirical parameters proposed in the previous chapter.

4.5 Conclusion and discussion

In this chapter, we applied a novel methodology based on inference from reconstruction to automatically and objectively infer the medium scale 3D breast texture model parameters from 3.5 cm × 3.5 cm × 3.5 cm volumes of interest (VOI) of segmented clinical breast computerized tomography (bCT) reconstructed datasets.

A multiple births, deaths and shifts (MBDS) algorithm was employed to first reconstruct a set of random ellipsoids from each ground truth bCT VOI. Visual inspection of the volumes recreated by voxelizing the reconstructed ellipsoids shows a fairly good approximation of the medium scale breast tissue in the original bCT VOIs. A Matérn cluster process was then fitted to the reconstructed ellipsoid centers using the minimum contrast method based on the pair correlation functions (PCF). This introduces a clustering interaction between ellipsoids to the previous prototype model. Statistical diagnostic analysis using the PCF suggested a fairly good fit of the reconstructed ellipsoid centers to the proposed Matérn cluster process. Distributions of the ellipsoid half lengths and orientations were finally estimated from their empirical distributions. Twelve sets of new medium scale model parameters were obtained. Preliminary evaluation of the 2D and 3D breast

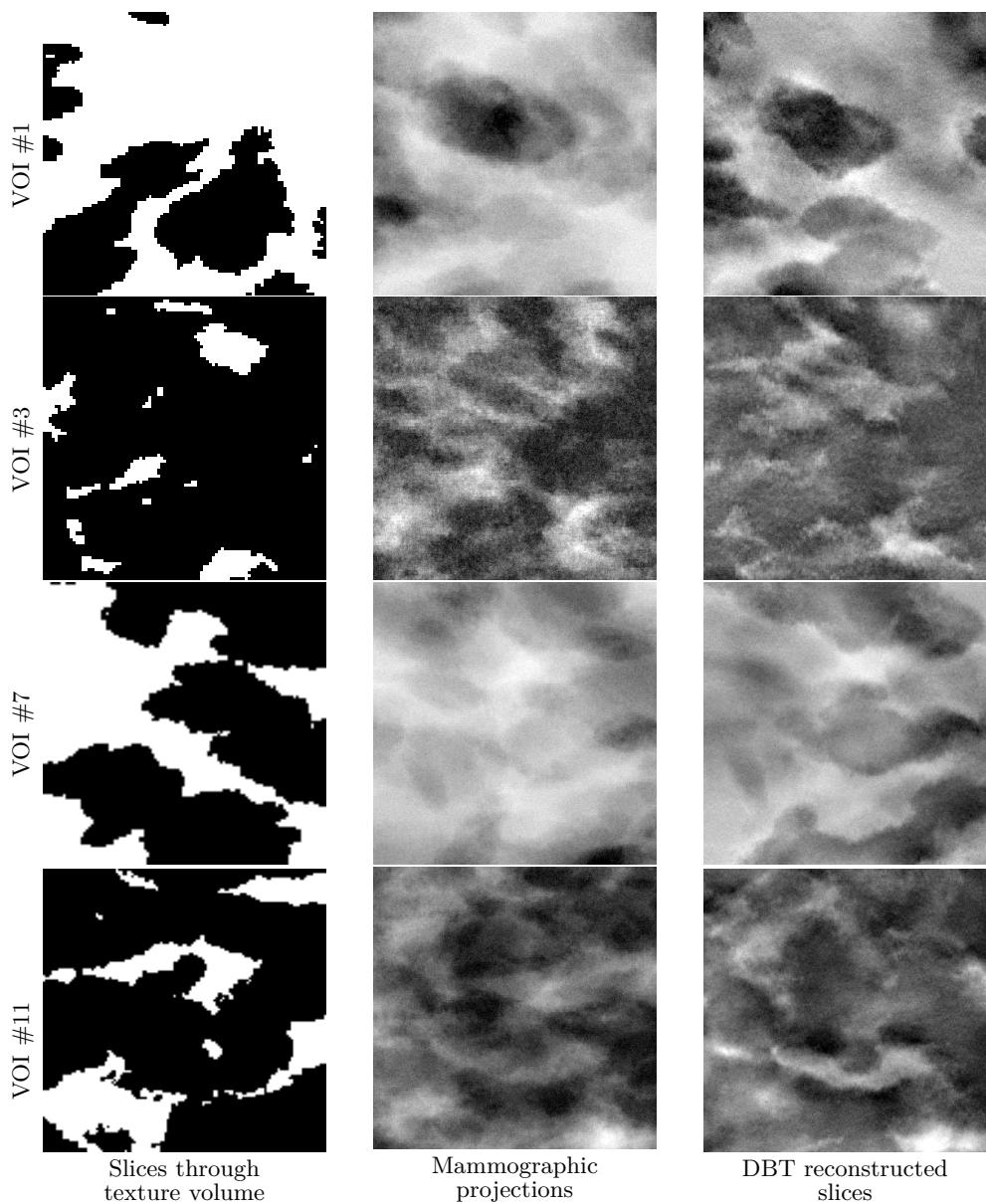


Figure 4.8: The first column shows slices through volumes simulated from the 3D breast texture model with the four sets of parameters listed in Table 4.4. The second column shows mammographic projections simulated from the simulated texture volumes. The third column shows DBT reconstructed slices simulated from the simulated texture volumes. The sizes of the images are $3.5\text{ cm} \times 3.5\text{ cm}$.

images simulated from 3D texture volumes generated using new parameters shows fairly high visual realism. The breast tissue variability in new simulated images is larger than the images simulated using previously proposed prototype implementation with empirical parameters.

The proposed method has several limitations. First, the effect of the initialization step on the result of the reconstruction should be further analyzed. A possible next step is to run the reconstruction step using different initialization schemes and to study their impact

on the reconstruction result. Secondly, the impact of the reconstruction step on the estimated parameters in the inference step was not investigated. To address this, one possible approach is to study how the inference result deviates from a ground truth model when the latter is a known Poisson marked point process. Thirdly, to reduce the optimization complexity in the inference step, the distribution of ellipsoid half lengths were estimated independently from the ellipsoid centers. This may be a simplification compared with the distribution of inter-glandular adipose compartments in real breasts. The correlation between the the half lengths and the centers of the ellipsoids could be further investigated by studying the mark correlation function of the reconstructed ellipsoids. The statistical properties of simulated 2D and 3D breast images using inferred medium scale model parameters were not studied. This should be addressed next as a further validation of the new models. Also, a formal observer psycho-physical experiment should be performed to validate the visual realism of images simulated using the fitted medium scale model parameters.

Chapter 5

Statistical characteristics of x-ray breast images

In this chapter, we aim to understand the common statistical characteristics of typical 2D and 3D x-ray breast images encountered in clinical practices.

Breast images are characterized by their statistical characteristics related to the breast anatomical parenchyma and the noise from the acquisition system. Both aspects have crucial impact on lesion detectability in breast images. These statistical characteristics have been shown to provide a way to understand, compare, or even quantify the difference of lesion detection performance in different types of breast images. In addition, knowledge of the statistical properties of x-ray breast images acquired with various imaging modalities is key to characterize the statistical realism of simulated breast texture images and may help to define more advanced breast texture models.

Breast density, the first order statistical characteristic was investigated first. An overview of the state-of-the-art breast density assessment algorithms developed for 2D and 3D breast imaging modalities was presented. A new volumetric breast density assessment algorithm using DBT projections, based on the model-based stereology was proposed. Numerical evaluations were conducted to study the estimation accuracy of the proposed algorithm.

The power spectral index (β), a second order statistical characteristic was investigated next. The most commonly applied algorithm for β assessment has many variants in literature. Different variants use similar methodology, but with different parameterizations. We investigated the impact of these different algorithm parameterizations on the bias and variance of β estimation. Analytical derivation of the results was demonstrated. Numerical evaluations were performed to validate the analytical results.

5.1 Breast density

5.1.1 Background

As mentioned in Section 1.1, breast density is an important breast cancer risk factor. Higher breast density means that the breast has a greater amount of fibroglandular breast tissue than adipose tissue. It has been shown that women with high mammographic breast density are about four to six times more likely to receive a breast cancer diagnosis compared to women with less dense breast [120] [29]. In addition, an important effect of increased breast density is the decrease in FFDM sensitivity due to possible masking of lesions by fibroglandular tissue [70]. Consequently, women with dense breast tissue may benefit from a tailored screening regimen, with additional exams using ultrasound or magnetic resonance imaging. In this respect, as of June 2017, 31 states in the United States of America have introduced laws on breast density notification after FFDM exams, intended to give women the necessary information so they can decide on further action in case they have dense breasts [62].

5.1.2 Assessment of breast density

Breast density assessment can be performed using FFDM or DBT images. Since DBT has gained acceptance as an adjunct to FFDM in screening, it is increasingly important that the methods for breast density assessment are consistent between FFDM and DBT. The methods for breast density assessment can be area-based or volumetric. We present hereunder a brief overview of area-based and volumetric breast density assessment methods for FFDM and DBT. For a more detailed description of the state-of-the-art breast density assessment methods for FFDM and DBT, we refer to the review paper by He *et al.* [83] and the references therein.

Area-based breast density assessment

The area-based breast density is defined as the ratio of the fibroglandular tissue region over the total area of the breast in a 2D breast image. Area-based breast density does not take into account the thickness of the breast.

Traditionally, in FFDM, area-based breast density is estimated based on a visual assessment of the proportion of fibroglandular tissue in the mammogram. The most commonly used visual assessment method is the four-category classification defined in the Breast imaging reporting and data system (BI-RADS, fourth edition, [160]) (Table 5.1):

- BI-RADS breast density 1: almost entirely adipose
- BI-RADS breast density 2: scattered fibroglandular
- BI-RADS breast density 3: heterogeneously dense
- BI-RADS breast density 4: extremely dense

The BI-RADS breast density classification is an assessment with limited intra- and inter-reader agreement [144]. In the newest edition of the BI-RADS atlas, the density categories 1, 2, 3 and 4 are replaced by letters *a*, *b*, *c* and *d* where *a* refers to almost entirely adipose and *d* refers to extremely dense [59]. The new BI-RADS density notation takes into

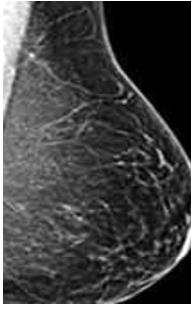
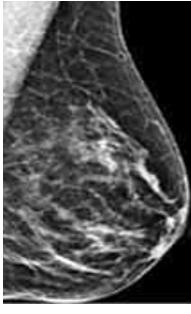
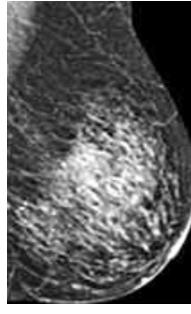
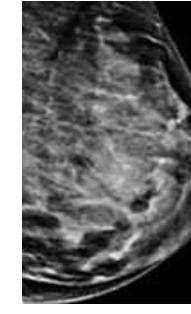
BI-RADS breast density 1	BI-RADS breast density 2	BI-RADS breast density 3	BI-RADS breast density 4
			
Almost entirely adipose	Scattered fibroglandular	Heterogeneously dense	Extremely dense

Table 5.1: BI-RADS classification of breast density in mammographic images.

account the risk of lesion masking in local regions of a mammogram. If local dense regions are identified in a mammogram, the breast could be classified as high density even if the overall quantitative breast density is low [59].

To minimize the inter- and intra-reader variability, as well as to make breast density assessment easier and more accurate, objective and automated breast density assessment methods are preferred. Several algorithms have been developed to objectively assess area based breast density from 2D FFDM and 3D DBT images. These algorithms rely on fully- or semi-automatic segmentation of the fibroglandular tissue in 2D breast images [83].

Assessment of area-based breast density in 2D FFDM images consists of computing the ratio of the segmented fibroglandular tissue region over the whole breast region. Several fibroglandular tissue segmentation algorithms have been developed [83], using methods such as thresholding, clustering, statistical texture analysis, integration of different approaches and analysis of parenchymal patterns etc. [133] [7] [67] [208] [146] [96] [82]. In DBT, same area-based breast density assessment methods have been used on DBT central projection image [17].

Volumetric breast density assessment

The volumetric breast density (VBD) refers to the true fibroglandular density of a breast, defined as the ratio of fibroglandular tissue volume over the breast volume. Contrary to area-based breast density, the thickness of the breast is taken into account.

VBD assessment in FFDM has been achieved by prior calibration [97], in-image phantom calibration [178] and modeling the physical process of FFDM image acquisition [64]. Existing algorithms analytically formulate the pixel intensity in FFDM images using parameters of FFDM image acquisition, including in particular the breast thickness [83]. Several commercially available and U.S. Food and Drug Administration approved softwares have been developed [93] [46].

VBD assessment in DBT, first investigated by Bakic *et al.* [18] and Kontos *et al.* [103]

in 2008, is relatively recent. Several methods are under investigation. Some studies segmented all or a subset of DBT reconstructed slices separately without taking the breast tissue correlation between DBT slices into account, and computed the ratio of the total fibroglandular area in all slices over the total breast area in all slices [18] [188]. Other approaches based on automatic 3D fibroglandular tissue segmentation from DBT reconstructed images have also been investigated [152] [175].

5.1.3 A new stereology-based algorithm for volumetric breast density assessment from DBT projections

We propose a new VBD assessment algorithm from DBT projection images inspired from the theory of stereology. Stereology is a branch of mathematics where the objective is to draw inference about the information of structures in 3D using only information available in lower dimensions [45]. In stereology, mainly two approaches can be distinguished:

1. The *design-based stereology*, where the inference of the 3D information is achieved by carefully designing random sampling techniques on the available lower dimension observations [173].
2. The *model-based stereology*, where the 3D structure is first formulated as a mathematical model; then the inference of the 3D information is drawn from lower dimensional observations by studying their statistical link to the 3D model [95].

For a more detailed mathematical description of the stereology theory, we refer to publications such as [123], [192] and [10].

VBD assessment from DBT projections can be approached as a stereological problem, where we aim at inferring the first order statistical characteristic of the 3D breast using its 2D projections. The proposed algorithm adopted the approach of model-based stereology. We formulated the breast as a 3D random field model. The DBT image acquisition was modeled as a linear system with simplified formulation of the physics of image acquisition. A statistical estimator of VBD was obtained by deriving the analytical formula for the pixel intensity using parameters of the simplified image acquisition system and the 3D breast model.

Three-dimensional breast model

The 3D solid breast texture model introduced in Chapter 3 was used to model local fibroglandular and inter-glandular adipose breast tissue. Let f be a 3D discrete binary volume f sampled from the 3D breast texture model (Figure 5.1). Assume that f is defined on a discrete 3D cube $\Omega \subset \mathbb{Z}^3$. Mathematically, for any $x \in \Omega$,

$$f(x) = \begin{cases} \mu_g, & \text{if } x \in \text{fibroglandular region}, \\ \mu_a, & \text{otherwise.} \end{cases} \quad (5.1)$$

Here μ_g and μ_a are the linear attenuation coefficients of the breast fibroglandular and inter-glandular adipose tissue, respectively.

We assume that f is stationary. That is, the distribution of f is translation invariant. We also assume that the distribution of f is invariant to rotations around the X -axis

(Figure 5.1). This means that the distribution of the DBT projection of f is invariant to different projection angles. The simplified assumptions are approximately realistic when considering only small ($\approx 2.5 \text{ cm} \times 2.5 \text{ cm} \times 2.5 \text{ cm}$) regions of the same tissue type. Previous research has shown that breast tissue is locally stationary [151].

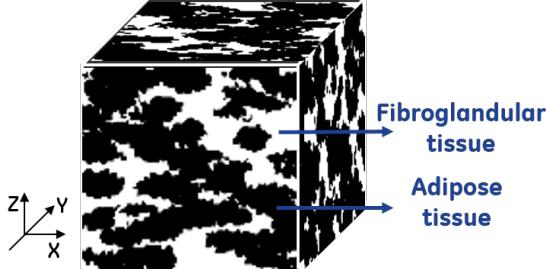


Figure 5.1: Illustration of a binary volume f that is used to model a rectangular region of fibroglandular and inter-glandular adipose tissue inside a breast. The volume is sampled from the 3D solid breast texture model introduced in Chapter 3. We assume that f is stationary and the distribution of f is invariant to rotations around the X -axis.

The VBD of f , denoted as V , is mathematically written as

$$V = \frac{|\{x \in \Omega | f(x) = \mu_g\}|}{|\Omega|} \times 100\%. \quad (5.2)$$

Notice that V is a first order statistical characteristic of f since it is proportional to the average of f .

DBT image acquisition model

Next, we proposed a simplified linear modeling of the DBT image acquisition system. We considered a GE DBT imaging system where nine projection images are acquired within an angular range of $(12.5^\circ, -12.5^\circ)$. A schematic of this simplified modeling is shown in Figure 5.2.

For a given projection angle θ , the DBT projection image of the breast volume f can be modeled using the Beer-Lambert law:

$$\frac{I^\theta(m, n)}{I_0} = \exp \left(- \int_{\mathcal{L}_\theta(m, n)} f(\mathbf{r}) d\mathbf{r} \right). \quad (5.3)$$

Here $I^\theta(m, n)$ is the photon intensity of the projection image at pixel location (m, n) and I_0 is the photon intensity at the x-ray source. The term $\mathcal{L}_\theta(m, n)$ refers to the line of the x ray defined by the x-ray source position and the pixel location (m, n) .

Equation (5.3) can be reformulated as:

$$\ln \frac{I^\theta(m, n)}{I_0} = -D_\theta(m, n)(\mu_a + p_{\mathcal{L}_\theta(m, n)} (\mu_g - \mu_a)), \quad (5.4)$$

where $D_\theta(m, n)$ is the length of the segment of $\mathcal{L}_\theta(m, n)$ that traverses the compressed breast, as indicated in Figure 5.2. The term $p_{\mathcal{L}_\theta(m, n)}$ refers to the fraction of $D_\theta(m, n)$ that

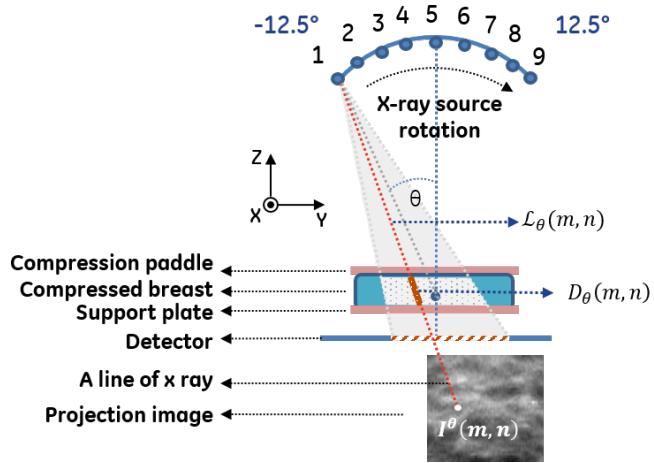


Figure 5.2: Schematic of the DBT acquisition geometry. X-ray source rotates around a stationary detector within an angular range of $(-12.5^\circ, 12.5^\circ)$. Nine projection images are acquired. At projection angle θ , $\mathcal{L}_\theta(m, n)$ is the line of x ray defined by the x-ray source position and the pixel location (m, n) in the projection image. The length of the segment of $\mathcal{L}_\theta(m, n)$ inside the compressed breast is $D_\theta(m, n)$.

corresponds to the fibroglandular tissue. Notice that $D_\theta(m, n)$ is related to the thickness of the compressed breast. If we assume that the thickness of the compressed breast is a known constant, $D_\theta(m, n)$ can be precisely computed for any arbitrary pixel location (m, n) .

From (5.4), the fraction of fibroglandular tissue along each line $\mathcal{L}_\theta(m, n)$ can be expressed as:

$$p_{\mathcal{L}_\theta(m,n)} = \frac{-\ln I^\theta(m,n) + \ln I_0 - D_\theta(m,n)\mu_a}{D_\theta(m,n)(\mu_g - \mu_a)}. \quad (5.5)$$

Analytical derivation of the volumetric breast density estimator

Using the stationarity assumption of f , for each projection image at angle θ , an estimate of the VBD, denoted as \hat{v}_θ , can be obtained by averaging $p_{\mathcal{L}_\theta(m,n)}$ across all pixel locations. That is:

$$\hat{v}_\theta = \frac{1}{N} \sum_{m,n} p_{\mathcal{L}_\theta(m,n)}. \quad (5.6)$$

Here N is the total number of pixels in a DBT projection image. Since the distribution of the projection of f is assumed to be invariant to rotations around the X -axis, \hat{v}_θ is independent of the projection angle θ . This allows us to write \hat{v}_θ as \hat{v} . Plugging (5.6) into (5.5), a linear equation can be obtained:

$$y = \hat{v}x + \alpha, \quad (5.7)$$

with

$$y = \frac{\sum_{m,n} \frac{\ln I^\theta(m,n)}{D_\theta(m,n)(\mu_g - \mu_a)}}{\sum_{m,n} \frac{1}{D_\theta(m,n)(\mu_g - \mu_a)}}, \quad (5.8)$$

$$x = \left(\frac{1}{N} \sum_{m,n} \frac{-1}{D_\theta(m,n) (\mu_g - \mu_a)} \right)^{-1}, \quad (5.9)$$

and

$$\alpha = \ln I_0 - \frac{\mu_g}{\mu_g - \mu_a}. \quad (5.10)$$

Notice that α is constant for a given image acquisition.

Let \mathbf{X} , \mathbf{Y} and $\boldsymbol{\alpha}$ be column vectors containing respectively all x , y and α obtained from all nine DBT projection images, then the final VBD estimator, denoted as \hat{V} can be obtained by solving:

$$\hat{V} = \arg \min_{\hat{v}} \|\mathbf{Y} - \hat{v}\mathbf{X} - \boldsymbol{\alpha}\|^2. \quad (5.11)$$

The optimization can be solved using linear regression. This finally yields:

$$\begin{aligned} \hat{V} &= \frac{\text{Cov}(\mathbf{Y}, \mathbf{X})}{\text{Var}(\mathbf{X})} \\ &= \frac{N\mathbf{Y}^T \mathbf{X} - \mathbf{X}^T \mathbf{1} \mathbf{Y}^T \mathbf{1}}{N\mathbf{X}^T \mathbf{X} - (\mathbf{X}^T \mathbf{1})^2}, \end{aligned} \quad (5.12)$$

where $\mathbf{1}$ is the column vector having the same size as \mathbf{X} and \mathbf{Y} and with all elements equal to 1.

5.1.4 Performance evaluation of the stereology-based volumetric breast density estimator

The accuracy of the new stereology based VBD estimator was evaluated using simulated DBT projections from the 3D breast texture model introduced in Chapter 3.

Simulation of synthetic breast volumes

Breast volumes representing BI-RADS 1, BI-RADS 2, BI-RADS 3 and BI-RADS 4 density categories were simulated with parameters described in Table 5.2. Notice that the nipple position was set to $(0, 0, \infty)$ for all simulations, making the orientation of the ellipsoids independent from the ellipsoid centers. This insures that the simulated volumes are stationary. For each density category, 20 volumes were generated. The ground truth VBD value for each generated volume was computed as the percentage ratio of the number of fibroglandular tissue voxels over the total number of voxels in the volume. Simulated breast volumes and the distribution of the ground truth VBD in each BI-RADS breast density category are summarized in Table 5.3.

Simulation of DBT image acquisitions

DBT image acquisitions were simulated using a previously described x-ray simulator [132]. The same DBT system topology and acquisition parameters as described in Section 3.1.2 were used.

To exclude the geometric non-uniformity at the physical edge of the texture cube due to the cone-beam geometry, a central region of interest (ROI) of each projection image with

Parameter	BI-RADS density 1	BI-RADS density 2	BI-RADS density 3	BI-RADS density 4
W	5 cm \times 5 cm \times 5 cm voxelized cube centered at the origin $o(0, 0, 0)$			
Nipple q		(0, 0, ∞)		
Voxel size ν		0.1 mm \times 0.1 mm \times 0.1 mm		
Precision ϵ			0.1	
λ_0	10 mm $^{-3}$ (<i>i.e.</i> the average volume of a Voronoi cell \approx 0.1 mm 3)			
Φ_s intensity	3.5×10^{-3} mm $^{-3}$	5×10^{-3} mm $^{-3}$	2.3×10^{-3} mm $^{-3}$	1.3×10^{-3} mm $^{-3}$
$p_{L_a}, p_{L_b}, p_{L_c}$ (in mms)	$p_{L_a} = \mathcal{N}(7.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(4.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$	$p_{L_a} = \mathcal{N}(6.4, 1.5),$ $p_{L_b}, p_{L_c} = \mathcal{N}(2.5, 0.7)$
$p_{\delta_{\phi_x}}, p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}}$ (in radians)		$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2}),$ $p_{\delta_{\phi_y}}, p_{\delta_{\phi_z}} = \mathcal{N}(0, \frac{\pi}{12})$		

Table 5.2: Parameters of the 3D solid breast texture model used to simulation of BI-RADS breast density 1, BI-RADS breast density 2, BI-RADS breast density 3 and BI-RADS breast density 4 type breast volumes for the performance assessment of the proposed VBD estimator.

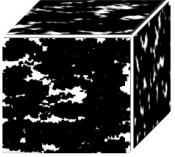
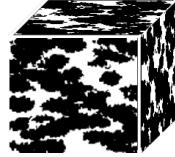
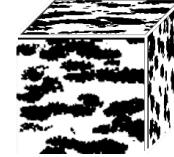
Density category	BI-RADS breast density 1	BI-RADS breast density 2	BI-RADS breast density 3	BI-RADS breast density 4
Breast volume				
Ground truth VBD	$18\% \pm 2\%$	$33\% \pm 2\%$	$60\% \pm 2\%$	$83\% \pm 2\%$
Quantity	20 volumes	20 volumes	20 volumes	20 volumes

Table 5.3: Examples of simulated volumes representing four different BI-RADS density categories. The distribution of the ground truth VBD (mean \pm standard deviation) in each category is also shown.

size 2.5 cm \times 2.5 cm were taken for the VBD estimation. Further, to account for the non-uniformity of photon intensity on the detector, each projection image was gain-calibrated using the corresponding airscan projection image. Figure 5.3 shows the airscan projections at different angles and a set of projection images from a BI-RADS breast density 3 volume

with the same angular order. Only ROIs of size 2.5 cm × 2.5 cm are shown.

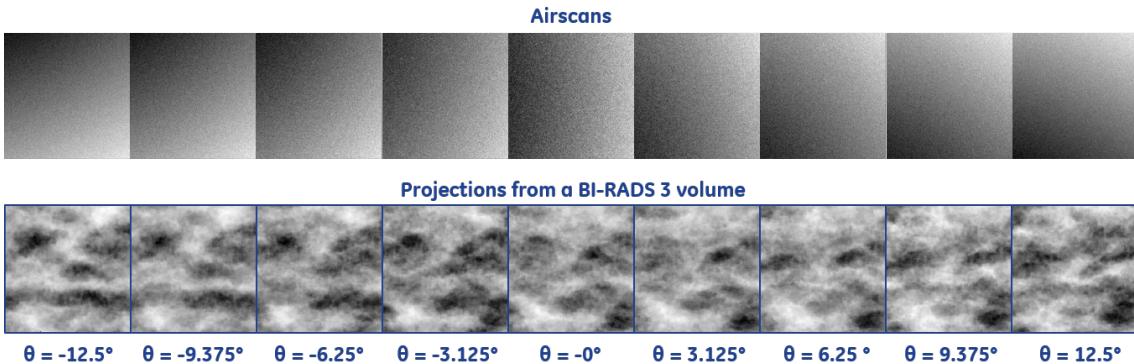


Figure 5.3: ROIs of size 2.5 cm × 2.5 cm showing DBT airscan projections and projections from a BI-RADS breast density 3 volume at the nine DBT projection angles.

Result of evaluation

The VBD estimator expressed in (5.12) was used to estimate the VBD for all simulated breast volumes. The estimated VBD values were compared to the known ground truth to measure the estimator biases, defined as the average difference between estimated VBD and the ground truth. For each density category, the standard deviation of the estimations was also computed. The result is demonstrated in Table 5.4. Based on the evaluation, the absolute value of the bias of the VBD estimator was below 3% of the ground truth value for all BI-RADS density categories. The standard deviation of the VBD estimator was below 3% of the ground truth value for all BI-RADS density categories.

	BI-RADS density 1	BI-RADS density 2	BI-RADS density 3	BI-RADS density 4
Bias	-1.52%	-2.36%	-0.07%	1.48%
Standard deviation	1.34%	1.81%	2.36%	1.24%

Table 5.4: Performance of the VBD estimator (5.12). Bias and standard deviation are percentage values with respect to the ground truth VBD.

5.2 The β metric

Since Burgess made the observation that the noise power spectrum (NPS) Γ of clinical mammograms obeys a power law relationship such as $\Gamma(\nu) \propto \frac{1}{|\nu|^\beta}$ and demonstrated that the magnitude of the power spectral index β is inversely proportional to mass detectability [31], the β metric has been extensively used as a surrogate for lesion detectability in x-ray breast images.

Several researchers have assessed the magnitude of β in clinical breast x-ray images acquired with different breast imaging modalities including full-field digital mammography (FFDM), digital breast tomosynthesis (DBT) and dedicated breast computerized tomography (bCT).

5.2.1 State-of-the-art algorithm for β assessment from breast images: the Welch method

In breast x-ray imaging, the commonly used technique to estimate β is partly based on the NPS estimation method originally proposed by Welch [201].

General description of the Welch method

As shown in Figure 5.4, the Welch method consists of five main steps:

1. First, the mean-subtracted input image is divided into partly overlapping regions of interest (ROI).
2. Then each ROI is multiplied by a 2D window function.
3. Next, the 2D NPS is estimated by averaging the squared modulus of the Fourier transform of all ROIs.
4. The 1D radial NPS is then computed, by radially averaging the 2D NPS.
5. Finally a linear regression is applied over a selected frequency range of the logarithm of the 1D radial NPS as a function of the logarithm of the radial frequencies, and the inverse of the slope of the fitted line is measured, yielding the estimated β .

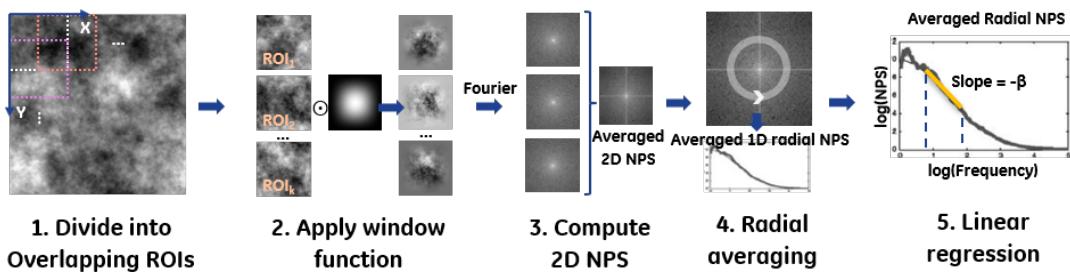


Figure 5.4: Overview of the five steps to the spectral β metric using the Welch method

Several researchers have investigated the impact of alternative parameterizations of various steps of the Welch method on the estimation bias and variance of the NPS [87] [65] [204] [181] [162]. However, to our knowledge, the propagation of these impacts to the estimation of β has not been investigated yet.

In this section, we focus on the effects of the overlapping ROIs division, the window function and radial frequency averaging on the β estimation bias and variance. The impact of the frequency range used for the linear regression was not studied and we adopted the range proposed by Wu *et al.* [204]. We first developed analytical models to evaluate the impact of the three factors. Then the analytical models were validated by simulation experiments using synthetic breast texture images with known β varying from 1 to 4 to cover the range of textures encountered in 2D and 3D breast x-ray imaging.

Mathematical formulation of the Welch method

Let u be a two-dimensional breast texture image defined on $\Omega \in \mathbb{Z}^2$, an $N \times N$ two-dimensional square discrete grid. Assume that u is a realization from a discrete power-law Gaussian random field f defined in Section 2.2.1 with spectral index β . Let Γ denote the NPS of f , we then have:

$$\Gamma_f(\nu) = \begin{cases} \frac{k}{\|\nu\|^\beta} & \text{if } \|\nu\| \neq 0, \\ 0 & \text{if } \|\nu\| = 0. \end{cases} \quad (5.13)$$

Here ν denotes the 2D discrete frequencies. We normalize the 2D discrete frequencies, such that for any ν we have $\nu \in \left[-\frac{1}{2}, \frac{1}{2}\right)$. Without loss of generality, we assume that $k = 1$ and that f has zero mean.

The five steps to empirically assess β using the Welch method can be mathematically formulated as follows.

- *Step 1.* The image u is divided into square ROIs of size $n \times n$, with α proportion of overlaps where $\alpha \in (0, 1)$. Let \mathcal{V} denote the set of all overlapping ROIs from image u .
- *Step 2.* Let w be a window function defined on the same discrete domain Ω . Given an ROI v , let v denote the windowed ROI, *i.e.* the result of pixel-wise multiplying v with w , then we have,

$$v = v \odot w. \quad (5.14)$$

Here \odot represents the pixel-wise multiplication.

- *Step 3.* The squared modulus of the Fourier transform of an image is referred to as the periodogram of the image, denoted as S [201]. Given an ROI v , its periodogram S_v is empirically measured by computing the squared modulus of the Fourier transform of the windowed ROI v . The estimate S_v can be written as:

$$S_v(\nu) = \|\mathcal{F}_v(\nu)\|^2 = \|\mathcal{F}_{v \odot w}(\nu)\|^2, \quad (5.15)$$

where \mathcal{F} refers to the 2D discrete Fourier transform and

$$\nu = (\nu_x, \nu_y) \text{ with } \nu_x, \nu_y \in \left\{-\frac{\lfloor \frac{n}{2} \rfloor}{n}, \dots, -\frac{1}{n}, 0, \frac{1}{n}, \dots, \frac{\lfloor \frac{n-1}{2} \rfloor}{n}\right\} \quad (5.16)$$

being the normalized discrete frequencies in the horizontal and vertical directions.

The final estimate of the 2D NPS, denoted as $\hat{\Gamma}$ is obtained by computing the average of the 2D NPS of all ROIs. That is,

$$\hat{\Gamma}(\nu) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} S_v(\nu). \quad (5.17)$$

Such averaging is typically applied to reduce the variance of the 2D NPS estimation [48].

- *Step 4.* The radially averaged 1D NPS $\hat{\Gamma}_r$ is obtained as follows:

1. First, divide the continuous frequency interval $\left[-\frac{1}{2}, \frac{1}{2}\right)$ into M sub-intervals with equal width $\epsilon = \frac{1}{M}$, where M is a positive integer. The sub-intervals are referred to as the frequency bins.

2. For the m^{th} frequency bin where $1 \leq m \leq M$, let τ_m denote the central value of the bin. We refer to τ_m as the radial frequency of the m^{th} bin. Then $\widehat{\Gamma}_r(\tau_m)$, the radially averaged NPS at τ_m , is obtained by averaging all estimated 2D NPS values $\widehat{\Gamma}(\nu)$ satisfying $(m-1)\epsilon \leq \|\nu\| < m\epsilon$. That is, for any $1 \leq m \leq M$,

$$\widehat{\Gamma}_r(\tau_m) = \frac{1}{|\zeta_m|} \sum_{\nu \in \zeta_m} \widehat{\Gamma}(\nu), \text{ with } \zeta_m = \{\nu | (m-1)\epsilon \leq \|\nu\| < m\epsilon\}. \quad (5.18)$$

The computation of $\widehat{\Gamma}_r$ is illustrated in Figure 5.5. The computation is performed for the radial frequencies of all the frequency bins. Studies in literature have used different values for the number of frequency bins M [87] [65] [204] [181] [162].

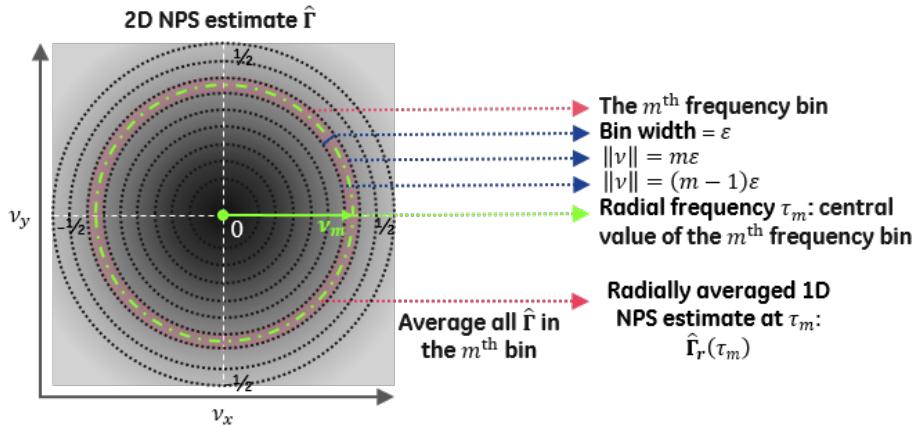


Figure 5.5: Illustration of the computation of the radially averaged empirical 1D NPS $\widehat{\Gamma}_r$ at τ_m from a 2D NPS estimate $\widehat{\Gamma}$. Here τ_m denotes the central value of the m^{th} frequency bin and is referred to as the radial frequency of the m^{th} frequency bin. Frequency bins are obtained by dividing the continuous frequency interval $[-\frac{1}{2}, \frac{1}{2}]$ into M sub-intervals with equal width ϵ . The estimate $\widehat{\Gamma}(\tau_m)$ is obtained by averaging all $\widehat{\Gamma}(\nu)$ with $(m-1)\epsilon \leq \|\nu\| < m\epsilon$.

- *Step 5.* Finally, given a lower-bound τ_l and an upper-bound τ_h of the radial frequency, the estimation of β is performed using the range of logarithm transformed radial frequencies $\mathcal{X} = \{\log(\tau_m)\}_{\tau_l \leq \tau_m \leq \tau_h}$ and their corresponding radially averaged logarithm transformed NPS $\mathcal{Y} = \{\log(\widehat{\Gamma}_r(\tau_m))\}_{\tau_l \leq \tau_m \leq \tau_h}$.

The estimate is computed by linearly fitting \mathcal{Y} to \mathcal{X} using ordinary least square (OLS) and taking the inverse of the fitted slope. Let $\widehat{\beta}_{\text{OLS}}$ denote the estimator of β , it is analytically expressed as:

$$\widehat{\beta}_{\text{OLS}} = -\frac{\text{Cov}(\mathcal{X}, \mathcal{Y})}{\text{Var}(\mathcal{X})}, \quad (5.19)$$

where Cov and Var denote respectively the covariance and the variance.

ROI division and window function application (steps 1 and 2) are typically combined since they yield a trade-off between the bias and variance of the estimated 2D NPS [48]. ROI division produces naturally a rectangular windowing effect, which causes spectral leakage in the Fourier domain and increases the bias in the estimated NPS [48]. Typically, to reduce

the spectral leakage, each ROI is multiplied by an appropriate 2D window function. A window function is designed to taper off sudden discontinuities caused by the rectangular windowing, and aims at decreasing the estimation bias by canceling the spectral leakage as much as possible. However a window function also discards part of the input image, thus increasing the NPS estimation variance. Averaging the estimated NPS from all ROIs helps reduce the variance.

5.2.2 Analytical derivation of the bias and variance of the β estimator based on the Welch method

Since the image u is a realization from a power-law Gaussian random field with spectral density described in (5.13), its theoretical discrete Fourier transform of \mathcal{F}_u can be written as:

$$\mathcal{F}_u(\nu) = \begin{cases} \frac{\mathcal{R}_u(\nu) + i\mathcal{I}_u(\nu)}{\|\nu\|^{\frac{\beta}{2}}} & \text{if } \|\nu\| \neq 0, \\ 0 & \text{if } \|\nu\| = 0. \end{cases} \quad (5.20)$$

Here $i = \sqrt{-1}$. The term \mathcal{R}_u and \mathcal{I}_u denote respectively the real and imaginary parts of the discrete Fourier transform of a zero mean white Gaussian image having the same size as u . They are independent and identically Gaussian distributed random variables with zero mean. Let σ^2 be their variance. To ease the notation, when we refer to the discrete frequency ν in the following sections, we assume that $\|\nu\|$ is always non-zero .

Let v denote an arbitrary ROI with size $n \times n$ of the image u , its theoretical discrete Fourier transform is:

$$\mathcal{F}_v(\nu) = \frac{\mathcal{R}_v(\nu) + i\mathcal{I}_v(\nu)}{\|\nu\|^{\frac{\beta}{2}}}. \quad (5.21)$$

where \mathcal{R}_v and \mathcal{I}_v denote respectively the real and imaginary parts of the discrete Fourier transform of a zero mean white Gaussian image having the same size as v . Therefore the unbiased periodogram of v is:

$$S_v(\nu) = \frac{r_v(\nu)}{\|\nu\|^{\frac{\beta}{2}}}. \quad (5.22)$$

with

$$r_v(\nu) = \sqrt{\mathcal{R}_v(\nu)^2 + \mathcal{I}_v(\nu)^2}. \quad (5.23)$$

Here for any ν , $r_v(\nu)$ is a random variable following the Rayleigh distribution with scale parameter σ [179].

In the following sections, we first conduct an analytical derivation of the β estimation bias and variance when using the Welch method. Our derivation will be in function of two aspects of the Welch method:

1. The division of the image into overlapping ROIs and the application of window functions;
2. The radial frequency averaging of the 2D NPS.

First, the impact of each aspect will be investigated separately, with other aspects ignored. To achieve this, we consider that no radial frequency averaging will be performed when investigating the impact of overlapping ROIs division and window functions. Similarly, when investigating the impact of radial frequency averaging, we consider that the impact

of overlapping ROI division is compensated by the application of a perfect window function. In the end, a few remarks will be given regarding the combining impact of the two investigated aspects on the β estimation accuracy.

Effect of overlapping ROI division and window functions

For an arbitrary ROI v and a window function w , we first develop the analytical expression of its periodogram S_v defined in (5.15). According to equation (5.14), (5.15) and (5.21), we have

$$S_v(\nu) = |\mathcal{F}_v * \mathcal{F}_w|^2 = \left| \frac{\mathcal{R}_v(\nu) + i\mathcal{I}_v(\nu)}{\|\nu\|^{\frac{\beta}{2}}} * \mathcal{F}_w(\nu) \right|^2. \quad (5.24)$$

In typical cases, the window function w is real-valued and symmetric. Moreover, to achieve the optimal reduction of the spectral leakage, w is often chosen such that \mathcal{F}_w has very small side lobes compared with the Fourier transform of a rectangular window of the same size. A commonly applied 2D window function is the Hann window [204], mathematically expressed as:

$$w(x_1, x_2) = \begin{cases} \frac{1}{4} \left(1 - \cos\left(\frac{2\pi x_1}{n-1}\right)\right) \left(1 - \cos\left(\frac{2\pi x_2}{n-1}\right)\right) & \text{if } x_1, x_2 \in \{0, \dots, n-1\}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.25)$$

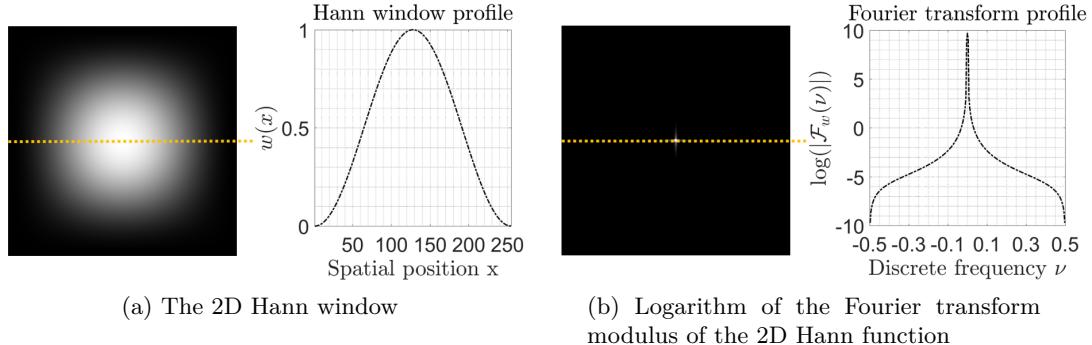


Figure 5.6: Illustration of the 2D Hann window and logarithm of its Fourier transform modulus. The discrete window size is 255×255 .

The 2D Hann window function and logarithm of its Fourier transform modulus is illustrated in Figure 5.6. We can see that the Fourier transform of the 2D Hann window is very concentrated around the zero frequency. For such as window function, \mathcal{F}_w can be numerically approximated by a finite symmetric convolution matrix D of size $n_w \times n_w$, with $n_w \ll n$. This yields,

$$S_v(\nu) \approx \left| \sum_{p,q} D(p, q) \left(\frac{\mathcal{R}_v(\nu_{p,q})}{\|\nu_{p,q}\|^{\frac{\beta}{2}}} + i \frac{\mathcal{I}_v(\nu_{p,q})}{\|\nu_{p,q}\|^{\frac{\beta}{2}}} \right) \right|^2. \quad (5.26)$$

Here $p, q \in \mathcal{L}$ where $\mathcal{L} = \{-\lfloor \frac{n_w}{2} \rfloor, -\lfloor \frac{n_w}{2} \rfloor + 1, \dots, 0, \dots, \lfloor \frac{n_w}{2} \rfloor\}$. We use $\nu_{p,q}$ to denote the translation of ν by the vector $(\frac{p}{n}, \frac{q}{n})$.

Now let $\delta_{\nu_{p,q}} = \|\nu\| - \|\nu_{p,q}\|$. Let $\{d_s\}_{s \in \{1, \dots, n_w^2\}}$ be the set enumerating all the elements in D . Let $\{R_s^v\}_{s \in \{1, \dots, n_w^2\}}$, $\{I_s^v\}_{s \in \{1, \dots, n_w^2\}}$ and $\{\delta_s\}_{s \in \{1, \dots, n_w^2\}}$ be the sets enumerating respectively all the elements in $\{\mathcal{R}_v(\nu_{p,q})\}_{p,q \in \mathcal{L}}$, $\{\mathcal{I}_v(\nu_{p,q})\}_{p,q \in \mathcal{L}}$ and $\{\delta_{\nu_{p,q}}\}_{p,q \in \mathcal{L}}$, using the same enumeration order as $\{d_s\}_{s \in \{1, \dots, n_w^2\}}$. Then equation (5.26) can be further written as

$$S_v(\nu) \approx \left| \sum_{s=1}^{n_w^2} \frac{d_s R_s^v}{(\|\nu\| - \delta_s)^{\frac{\beta}{2}}} + i \sum_{s=1}^{n_w^2} \frac{d_s I_s^v}{(\|\nu\| - \delta_s)^{\frac{\beta}{2}}} \right|^2. \quad (5.27)$$

Notice that $\max_{s \in \{1, \dots, n_w^2\}} |\delta_s|$ quantifies the maximum width of the frequency lobes of the window function in the Fourier domain. The following proposition allows us to further approximate Equation (5.27).

Proposition 5.1. *Let k be a positive integer. Let $\nu \in \mathbb{R}^2$ and $\beta \in \mathbb{R}^+$. Let $\{c_j\}_{j \in \{1, \dots, k\}}$ and $\{\delta_j\}_{j \in \{1, \dots, k\}}$ be two sets of constant real values. Let $\{Y_j\}_{j \in \{1, \dots, k\}}$ and $\{Z_j\}_{j \in \{1, \dots, k\}}$ be two sets of independent random variables following the Gaussian distribution with mean zero and variance σ^2 .*

Assume that $\max_{j \in \{1, \dots, k\}} |\delta_j|^2 \ll \frac{1}{k}$, we have,

$$\left| \sum_{j=1}^k \frac{c_j Y_j}{(\|\nu\| - \delta_j)^{\frac{\beta}{2}}} + i \sum_{j=1}^k \frac{c_j Z_j}{(\|\nu\| - \delta_j)^{\frac{\beta}{2}}} \right|^2 \approx \|\nu\|^{-\beta} r(\nu, \beta), \quad (5.28)$$

where $r(\nu, \beta)$ is a random variable following the Rayleigh distribution with scale parameter

$$\sqrt{\sum_{s=1}^k c_j^2 \left(1 + \frac{\delta_j \beta}{\|\nu\|}\right) \sigma^2}. \quad (5.29)$$

Proof. For an arbitrary $j \in \{1, \dots, k\}$, compute the Taylor series expansion of $f(t) = t^{-\frac{\beta}{2}}$ at $t = \nu - \delta_j$, we get:

$$f(\nu - \delta_j) = f(\nu) - \delta_j f'(\nu) + \mathcal{O}(|\delta_j|^2). \quad (5.30)$$

Since $\max_{j \in \{1, \dots, k\}} |\delta_j|^2 \ll \frac{1}{k}$, we can omit the quadratic term and write

$$f(\nu - \delta_j) \approx f(\nu) - \delta_j f'(\nu) = \frac{1}{\|\nu\|^{\frac{\beta}{2}}} + \delta_j \frac{\beta}{2 \|\nu\|^{\frac{\beta}{2}+1}}. \quad (5.31)$$

Now Let

$$g(\nu) = \sum_{j=1}^k c_j f(\nu - \delta_j), \quad (5.32)$$

we have:

$$g(\nu) = \sum_{j=1}^k \frac{c_j}{(\|\nu\| - \delta_j)^{\frac{\beta}{2}}} \approx \frac{1}{\|\nu\|^{\frac{\beta}{2}}} \sum_{j=1}^k c_j \left(1 + \frac{\delta_j \beta}{2 \|\nu\|}\right). \quad (5.33)$$

By plugging (5.33) into the left hand side of (5.28), we get:

$$\left| \sum_{j=1}^k \frac{c_j Y_j}{(\|\nu\| - \delta_j)^{\frac{\beta}{2}}} + i \sum_{j=1}^k \frac{c_j Z_j}{(\|\nu\| - \delta_j)^{\frac{\beta}{2}}} \right|^2 \approx \|\nu\|^{-\beta} r(\nu, \beta), \quad (5.34)$$

with

$$r(\nu, \beta) = \sqrt{\left| \sum_{j=1}^k c_j \left(1 + \frac{\delta_j \beta}{2\|\nu\|}\right) Y_j \right|^2 + \left| \sum_{j=1}^k c_j \left(1 + \frac{\delta_j \beta}{2\|\nu\|}\right) Z_j \right|^2}. \quad (5.35)$$

Since $\forall j$, Y_j and Z_j are independent and Gaussian distributed with mean zero and variance σ^2 , both the term $\sum_{j=1}^k c_j \left(1 + \frac{\delta_j \beta}{2\|\nu\|}\right) Y_j$ and the term $\sum_{j=1}^k c_j \left(1 + \frac{\delta_j \beta}{2\|\nu\|}\right) Z_j$ are also independent Gaussian random variables with zero mean and equal variance expressed as:

$$\sum_{j=1}^k c_j^2 \left(1 + \frac{\delta_j \beta}{2\|\nu\|}\right)^2 \sigma^2 \approx \sum_{j=1}^k c_j^2 \left(1 + \frac{\delta_j \beta}{\|\nu\|}\right) \sigma^2. \quad (5.36)$$

This concludes that $r(\nu, \beta)$ is indeed a Rayleigh distributed random variable with scale parameter

$$\sqrt{\sum_{j=1}^k c_j^2 \left(1 + \frac{\delta_j \beta}{\|\nu\|}\right) \sigma^2}. \quad (5.37)$$

□

Now assume that $\max_{s \in \{1, \dots, n_w^2\}} \|\delta_s\|^2 \ll \frac{1}{n_w^2}$, the direct application of the Proposition 5.1 to (5.27) yields:

$$S_v(\nu) \approx \|\nu\|^{-\beta} r_v(\nu, \beta), \quad (5.38)$$

where for any ROI $v \in \mathcal{V}$, the random variable $r_v(\nu, \beta)$ follows the Rayleigh distribution with scale parameter

$$\sqrt{\sum_{s=1}^{n_w^2} d_s^2 \left(1 + \frac{\delta_s \beta}{\|\nu\|}\right) \sigma^2}. \quad (5.39)$$

According to (5.17), the final 2D NPS estimate is the average of the 2D NPS estimate of all ROIs. Hence,

$$\hat{\Gamma}(\nu) = \sum_{v \in \mathcal{V}} S_v(\nu) \approx \|\nu\|^{-\beta} \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} r_v(\nu, \beta). \quad (5.40)$$

Taking the logarithm to both sides of (5.40) yields:

$$\log(\hat{\Gamma}(\nu)) \approx -\beta \log(\|\nu\|) + \log\left(\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} r_v(\nu, \beta)\right). \quad (5.41)$$

When a set of frequencies $\{v_k\}_{k=1, \dots, K}$ is considered, with K being a positive integer, the Equation (5.41) can be represented in matrix form. Let \mathbf{X} , \mathbf{Y} and $\epsilon(\beta, \mathcal{V})$ be $K \times 1$ column vectors, where their k^{th} element equals respectively to $\log(v_k)$, $\log(\hat{\Gamma}(v_k))$ and $\log\left(\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} r_v(\nu, \beta)\right)$. Then the following model can be established:

$$\mathbf{Y} \approx -\beta \mathbf{X} + \epsilon(\beta, \mathcal{V}). \quad (5.42)$$

Notice that \mathbf{Y} is non-linear on β since the error terms $\epsilon(\beta, \mathcal{V})$ depends on the true β . Moreover elements of $\epsilon(\beta, \mathcal{V})$ are not identically distributed, since according to (5.39), the

distribution of the k^{th} element in $\epsilon(\beta, \mathcal{V})$ depends on the frequency ν_k . In this case, the linear OLS estimator $\hat{\beta}_{\text{OLS}}$ given by (5.19) will yield biased estimation of the true β value [1]. When using still the OLS estimator $\hat{\beta}_{\text{OLS}}$, as the Welch method indicates, the mean and variance of the estimation are characterized by the following proposition.

Proposition 5.2. *For the estimation of β in Equation (5.42), the bias and variance of the OLS estimator $\hat{\beta}_{\text{OLS}}$ given by (5.19) can be expressed as:*

$$\text{Bias}(\hat{\beta}_{\text{OLS}}) \approx \frac{\sum_{k=1}^K \mathbb{E}(\epsilon_k(\beta, \mathcal{V})) (\bar{x} - x_k)}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K}, \quad (5.43)$$

$$\text{Var}(\hat{\beta}_{\text{OLS}}) \approx \frac{(\mathbf{X} - \bar{x}\mathbf{1})^T \text{Cov}(\epsilon(\beta, \mathcal{V})) (\mathbf{X} - \bar{x}\mathbf{1})}{((\mathbf{X} - \bar{x}\mathbf{1})^T (\mathbf{X} - \bar{x}\mathbf{1}))^2} - \text{Bias}^2(\hat{\beta}_{\text{OLS}}). \quad (5.44)$$

Here T denotes the matrix transposition and Cov denotes the covariance matrix. The terms x_k and $\epsilon_k(\beta, \mathcal{V})$ are the k^{th} element of \mathbf{X} and $\epsilon(\beta, \mathcal{V})$ respectively. Lastly, \bar{x} is the average of \mathbf{X} and $\mathbf{1}$ is a $K \times 1$ column vector with all elements equal to 1.

Proof. First multiply both side of (5.42) by $\frac{1}{M}\mathbf{1}^T$. We obtain:

$$\bar{y} \approx -\beta\bar{x} + \bar{\epsilon}(\beta, \mathcal{V}), \quad (5.45)$$

where \bar{x} , \bar{y} , and $\bar{\epsilon}(\beta, \mathcal{V})$ are the average of \mathbf{X} , \mathbf{Y} and $\epsilon(\beta, \mathcal{V})$ respectively. To ease the notation, for the following proof we omit the dependence on β and \mathcal{V} for $\epsilon(\beta, \mathcal{V})$ and $\epsilon(\beta, \mathcal{V})$ and write them as ϵ and $\mathbf{\epsilon}$ instead.

Using (5.45), the analytical expression of $\hat{\beta}_{\text{OLS}}$ in (5.19) can be written as:

$$\begin{aligned} \hat{\beta}_{\text{OLS}} &= -\frac{(\mathbf{X} - \bar{x}\mathbf{1})^T (\mathbf{Y} - \bar{y} - \mathbf{1})}{(\mathbf{X} - \bar{x}\mathbf{1})^T (\mathbf{X} - \bar{x}\mathbf{1})} \\ &\approx -\frac{(\mathbf{X} - \bar{x}\mathbf{1})^T (-\beta\mathbf{X} + \mathbf{C} + \epsilon + (\beta\bar{x} - \bar{c} - \bar{\epsilon})\mathbf{1})}{(\mathbf{X} - \bar{x}\mathbf{1})^T (\mathbf{X} - \bar{x}\mathbf{1})} \\ &= \beta + \frac{\mathbf{X}^T \epsilon - \bar{x}\bar{\epsilon}K}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K}. \end{aligned} \quad (5.46)$$

Taking the expectation of both sides of (5.46) gives,

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{\text{OLS}}) &\approx \beta + \frac{\mathbf{X}^T \mathbb{E}\epsilon - \bar{x} \sum_{k=1}^K K \mathbb{E}\epsilon_k}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K} \\ &= \beta + \frac{\sum_{k=1}^K \mathbb{E}(\epsilon_k)(\bar{x} - x_k)}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K}. \end{aligned} \quad (5.47)$$

Hence the bias of the estimator can be expressed as,

$$\text{Bias}(\hat{\beta}_{\text{OLS}}) \approx \frac{\sum_{k=1}^K \mathbb{E}(\epsilon_k)(\bar{x} - x_k)}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K}. \quad (5.48)$$

Finally, the variance is derived as:

$$\begin{aligned}\text{Var}(\hat{\beta}_{\text{OLS}}) &= \mathbb{E}(\hat{\beta}_{\text{OLS}}^2) - \mathbb{E}(\hat{\beta}_{\text{OLS}})^2 \\ &\approx \mathbb{E}\left(\frac{(\mathbf{X} - \bar{\mathbf{x}})^T(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\mathbf{1})(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\mathbf{1})^T(\mathbf{X} - \bar{\mathbf{x}})}{\left((\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})^T(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})\right)^2}\right) - \left(\text{Bias}(\hat{\beta}_{\text{OLS}})\right)^2 \\ &= \frac{(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})^T \text{Cov}(\boldsymbol{\epsilon}) (\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})}{\left((\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})^T(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})\right)^2} - \text{Bias}^2(\hat{\beta}_{\text{OLS}}).\end{aligned}\quad (5.49)$$

□

The following observations can be made from 5.43 and 5.44:

- The estimator $\hat{\beta}_{\text{OLS}}$ based on the Welch method is biased and the magnitude of the bias increases as the true β increases. This can be seen by looking at the term $\mathbb{E}(\epsilon_k(\beta, \mathcal{V}))$ in (5.43), for any $1 \leq k \leq K$. Recall that,

$$\mathbb{E}(\epsilon_k(\beta, \mathcal{V})) = \mathbb{E}\left(\log\left(\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} r_v(\nu_k, \beta)\right)\right), \quad (5.50)$$

where $r_v(\nu_k, \beta)$ follows a Rayleigh distribution with scale parameter equal to

$$\sqrt{\sum_{s=1}^{n_w^2} d_s^2 \left(1 + \frac{\delta_s \beta}{\nu_k}\right) \sigma^2}. \quad (5.51)$$

When the true β increases, the scale parameter of $r_v(\nu_k, \beta)$ and thus $\mathbb{E}(r_v(\nu_k, \beta))$ also increase. Therefore $\mathbb{E}(\epsilon_k(\beta, \mathcal{V}))$ becomes larger since $\log(\cdot)$ is a monotone increasing function. This yields a larger bias according to (5.43).

- The ROI size n produces a trade-off between the estimation bias and variance. This can be seen by analyzing the set $\{\delta_s\}_{s \in \{1, \dots, n_w^2\}}$ in the Rayleigh scale parameter. When the ROI size n increases, smaller and fewer side lobes occur in the frequency domain of the window function. This results in smaller δ_s values for any $s \in \{1, \dots, n_w^2\}$, yielding a smaller β estimation bias. However larger n will cause the number of ROIs to drop, yielding a larger estimation variance. Inversely, if n decreases, the estimation variance will be smaller since the estimate will be averaged over more ROIs. However the bias will increase when n increases.
- The window function also produces a bias-variance trade-off in β estimation. Compared to the rectangular window function, a smooth window function masks part of the image and thus reduces the effective image size used for the estimation, resulting in an increased estimation variance. A less smooth window results in a decreased variance, but since such a window approaches a rectangular window, it will also cause larger δ_s values, resulting in an increased bias. In the most extreme scenario where no window is used, the assumption

$$\max_{j \in \{1, \dots, k\}} |\delta_j|^2 \ll \frac{1}{k} \quad (5.52)$$

in Proposition 5.1 is no longer valid. This can result in a much more severe dependence of the estimation bias and variance on the true β value than in 5.43 and 5.44.

- Consider the case where $\delta_s = 0$ for all $s \in \{1, \dots, n_w^2\}$. This means the Fourier transform \mathcal{F}_w of the window function w is a Dirac at the zero frequency value. In this case the set $\{\delta_s\}_{s \in \{1, \dots, n_w^2\}}$ contains only the value 1. Hence the NPS estimate becomes

$$\hat{\Gamma}(\nu) = \|\nu\|^{-\beta} R(\nu), \quad (5.53)$$

where

$$R(\nu) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} r_v(\nu). \quad (5.54)$$

Here for any $v \in \mathcal{V}$, $r_v(\nu)$ follows the same the Rayleigh distribution defined in (5.23) with scale parameter σ . This indicates that the ϵ term in (5.42) becomes independent of β and the elements in ϵ becomes independent and identically distributed. In such a situation, the least square estimator $\hat{\beta}_{OLS}$ becomes unbiased and has a variance depending on the frequency range [1], which is constant in our configuration. However, a window function w with a Dirac as Fourier transform indicates that w must have infinite spatial size. This can never occur in practice. Therefore the estimator $\hat{\beta}_{OLS}$ is always biased.

Effect of radial frequency averaging

To assess the effect of radial frequency averaging of the 2D NPS estimate on the β estimation bias and variance, we assume that a perfect windowing is applied and we started from the unbiased 2D NPS estimate defined in (5.53).

Consider now that the radial frequency averaging described in the step 4 of the Welch method is applied to the unbiased 2D NPS estimate. Let $\{\tau_k\}_{k=1, \dots, K}$ be the set of radial frequencies used for the computation of the 1D radial NPS $\hat{\Gamma}_r$. Then according to (5.18) and (5.53), we have,

$$\hat{\Gamma}_r(\tau_k) = \frac{1}{N_k} \sum_{n_k=1}^{N_k} \hat{\Gamma}_r(\nu_{n_k}) = \frac{1}{N_k} \sum_{n_k=1}^{N_k} \|\nu_{n_k}\|^{-\beta} R(\nu_{n_k}). \quad (5.55)$$

Here for any $k \in \{1, \dots, K\}$, N_k denotes the number of 2D discrete frequency values in the k^{th} bin and ν_{n_k} denotes the n^{th} ordered 2D discrete frequency in the k^{th} bin. For any 2D discrete frequency ν , $R(\nu)$ is defined in (5.54). For any k and n_k , let $\delta_{n_k} = \tau_k - \|\nu_{n_k}\|$. Assume that for any k , $\max_{n_k \in \{1, \dots, N_k\}} |\delta_{n_k}|^2 \ll \frac{1}{N_k}$, by using the same approximation method described in the proof of Proposition 5.1, it can be shown that:

$$\log(\hat{\Gamma}_r(\tau_k)) \approx -\beta \log(\tau_k) + \log \left(\sum_{n_k=1}^{N_k} \frac{1 + \frac{\beta \delta_{n_k}}{\tau_k}}{N_k} R(\nu_{n_k})^2 \right). \quad (5.56)$$

Further, under the condition that for any k , $\max_{n_k \in \{1, \dots, N_k\}} \frac{\beta |\delta_{n_k}|}{\tau_k} \ll 1$, equation (5.56) can be approximated by:

$$\log(\hat{\Gamma}_r(\tau_k)) \approx -\beta \log(\tau_k) + \log \left(\sum_{n_k=1}^{N_k} \frac{R(\nu_{n_k})^2}{N_k} \right). \quad (5.57)$$

Let \mathbf{X} , \mathbf{Y} and $\boldsymbol{\epsilon}$ be $K \times 1$ column vectors, where their k^{th} element equals respectively to $\log(\widehat{\Gamma}_r(\tau_k))$, $\log(\tau_k)$ and $\log\left(\sum_{n_k=1}^{N_k} \frac{R(\nu_{n_k})^2}{N_k}\right)$. The following linear model can be immediately obtained:

$$\mathbf{Y} \approx -\beta\mathbf{X} + \boldsymbol{\epsilon}. \quad (5.58)$$

It can be noticed that the elements in $\boldsymbol{\epsilon}$ are not identically distributed since N_k varies with the bin index k . When using the OLS method on \mathbf{X} and \mathbf{Y} , the bias and variance of the β estimator $\widehat{\beta}_{\text{OLS}}$ can be directly obtained from Proposition 5.2. That is,

$$\text{Bias}(\widehat{\beta}_{\text{OLS}}) \approx \frac{\sum_{k=1}^K \mathbb{E}(\epsilon_k)(\bar{x} - x_k)}{\mathbf{X}^T \mathbf{X} - \bar{x}^2 K} \quad (5.59)$$

$$\text{Var}(\widehat{\beta}_{\text{OLS}}) \approx \frac{\sum_{k=1}^K \sigma_k^2(x_k - \bar{x})}{((\mathbf{X} - \bar{x}\mathbf{1})^T(\mathbf{X} - \bar{x}\mathbf{1}))^2} - \text{Bias}^2(\widehat{\beta}_{\text{OLS}}) \quad (5.60)$$

where σ_k^2 is the variance of ϵ_k . The following observations can be made:

- Under the condition that for all k , $\max_{n_k \in \{1, \dots, N_k\}} \frac{\beta|\delta_{n_k}|}{\tau_k} \ll 1$, both the bias and the variance in the estimated β depend only on the frequency range and the number of bins used. Once the frequency range and the number of bins are fixed in an estimation, the bias and variance are constant regardless of the magnitude of β .
- The approximation in (5.57) fails if the condition $\max_{n_k \in \{1, \dots, N_k\}} \frac{\beta|\delta_{n_k}|}{\tau_k} \ll 1$ fails for any k . This can happen when the bin width chosen is too large, or when τ_k is not the central value of the k^{th} bin. In this case the bias of $\widehat{\beta}_{\text{OLS}}$ will depend on true β value.

Combined effect of overlapping ROI division, window function and radial frequency averaging

The combined effect of ROI size, window function and radial averaging on the bias and variance in $\widehat{\beta}_{\text{OLS}}$ is hard to evaluate analytically due to the complexity of formulas (5.43) (5.44) (5.59) and (5.60). In the following section, we will evaluate the combined effect of ROI size, window function and radial averaging via numerical simulations with known ground truth.

As a first approximation, Section 5.2.2 indicates that when radial frequency averaging with small bin width is applied, its impact on the bias of $\widehat{\beta}_{\text{OLS}}$ can be neglected and the bias in $\widehat{\beta}_{\text{OLS}}$ using the complete Welch method will be approximately equal to (5.43). However, when radial frequency averaging with large frequency bins width is applied, the combined effect of overlapping ROI division, window function and will produce a bias that increases with the true β . The combining bias will be greater than the values in (5.43) and (5.59).

5.2.3 Numerical evaluation of β estimation accuracy

To represent breast texture observed in mammograms, DBT and bCT images, synthetic breast texture images with different ground truth β values were generated. A synthetic breast texture image $u \in \Omega$ was simulated using the simulation formula described in (2.11):

$$u = \mathcal{F}^{-1}(\mathcal{F}_w \odot \mathcal{F}_g). \quad (5.61)$$

where \odot denotes the pixel-wise multiplication. Here w is a white Gaussian image whose pixel intensities follow the standard normal distribution. The image g is defined on the same discrete space Ω . It represents a radially symmetric power-law kernel whose discrete Fourier transform is:

$$\mathcal{F}_g(\nu) = \begin{cases} \frac{1}{\|\nu\|^{\frac{\beta}{2}}} & \text{if } \|\nu\| \neq 0, \\ 0 & \text{if } \|\nu\| = 0. \end{cases} \quad (5.62)$$

The synthetic breast texture images were $10.24 \text{ cm} \times 10.24 \text{ cm}$, in floating-point format, with $100 \mu\text{m} \times 100 \mu\text{m}$ isotropic pixels. The value of β was varied from 1 to 4 in 0.2 steps to cover the range of textures encountered in 2D and 3D breast x-ray imaging. For each β value, 1000 synthetic breast texture images were generated. Figure 5.7 illustrates $2.56 \text{ cm} \times 2.56 \text{ cm}$ ROIs of synthetic breast texture images with β equal to 1, 2, 3 and 4.

To evaluate the impacts of overlapping ROI division, the window function and radial frequency averaging on the estimation accuracy of β , four experiments were designed according to the specifications listed in Table 5.5.

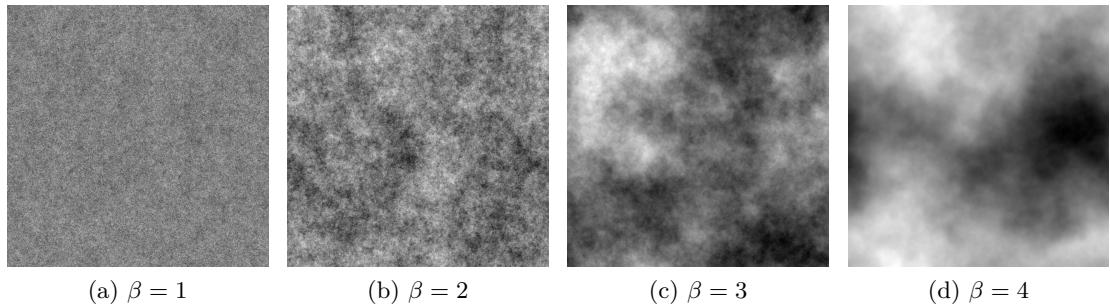


Figure 5.7: Sample ROIs of synthetic breast texture images used to assess the performance of β estimation using the Welch method. Synthetic breast texture images of size $10.24 \text{ cm} \times 10.24 \text{ cm}$ were simulated from 2D power-law Gaussian random field models with β varying from 1 to 4 with a step of 0.2. Here ROIs of images simulated with β equal to 1, 2, 3 and 4 are shown. The ROIs have size $2.56 \text{ cm} \times 2.56 \text{ cm}$.

Figure 5.8 compares the bias, standard deviation (square root of the variance) and root mean square error $\text{RMSE} = \sqrt{\text{Bias}^2 + \text{Var}}$ in the estimated β as a function of the true β of synthetic texture images for the four studied experiments. The RMSE combines both bias and variance into a single measure and was used here to assess the overall estimation performance.

Figure 5.8a shows the results of experiment 1 evaluating the effect of ROI division. Only ROIs with sizes larger than $2.5 \text{ cm} \times 2.5 \text{ cm}$ were considered, because sizes smaller than this value might not be sufficient to entirely capture the anatomical structure variations of breast textures [107]. For the two investigated ROI sizes with 50% overlap, the bias increases progressively as the true β increases when true β value is smaller than 3.6; then the bias decreases a little. When true β equals to one, the bias is almost zero. When true β equals to four, the bias equals to 0.1 for $5.12 \text{ cm} \times 5.12 \text{ cm}$ ROI division; and the bias equals to 0.2 for $2.56 \text{ cm} \times 2.56 \text{ cm}$ ROI division. The standard deviation demonstrates a similar dependence on true β value. However for $5.12 \text{ cm} \times 5.12 \text{ cm}$ ROI division, smaller standard deviations are obtained than for $2.56 \text{ cm} \times 2.56 \text{ cm}$ ROI division. There is clearly

Experiment	ROI division	Window function	Radial frequency averaging
1	2.56 cm \times 2.56 cm and 5.12 cm \times 5.12 cm ROI divisions, all with 50% overlap	Not applied	Not applied
2	2.56 cm \times 2.56 cm and 5.12 cm \times 5.12 cm ROI divisions, all with 50% overlap	2D Hann window	Not applied
3	Original 10.24 cm \times 10.24 cm without ROI division	Not applied	With 7 bins, 25 bins and 50 bins
4	2.56 cm \times 2.56 cm division with 50% overlap	2D Hann window	With 7 bins, 50 bins and without radial averaging

Table 5.5: Design of the four experiments to assess the performance of β estimation using the Welch method.

a bias-variance trade-off regarding the ROI size. The evaluation of RMSE demonstrates that, in this configuration, using 5.12 cm \times 5.12 cm ROIs is preferred.

Figure 5.8b shows the results of experiment 2 evaluating the effect of the window function. For the two investigated ROI sizes with 50% overlap, when a 2D Hann window is applied, the biases are much smaller. When true β value increases from one to four, the bias monotonically increases from 0 to approximately 0.02 for 5.12 cm \times 5.12 cm ROI division; and to 0.1 for 2.56 cm \times 2.56 cm ROI division. Larger standard deviations are obtained compared with experiment 1 when true β value is smaller than 2.5. This demonstrates a bias-variance trade-off regarding the window function. However when a 2D Hann window is applied, the standard deviations are almost independent of true β value: they are around 0.042 for 5.12 cm \times 5.12 cm ROI division; and around 0.035 for 2.56 cm \times 2.56 cm ROI division. The evaluation of RMSE demonstrates that, in this configuration, an application of a 2D Hann window is always preferred. The preference between 5.12 cm \times 5.12 cm and 2.56 cm \times 2.56 cm ROIs depends on the true β value: when true β is smaller than 2.3, using 2.56 cm \times 2.56 cm ROIs is preferred; however when β is greater than 2.3, using 5.12 cm \times 5.12 cm ROIs is preferred.

Figure 5.8c shows the results of experiment 3 evaluating the effect of radial frequency averaging. When the 2D NPS is computed from the full size input image (no ROI division) without application of an appropriate window function and 50 frequency bins (0.011 mm^{-1} frequency bin width) are used to compute the radial NPS, the bias is independent of true β and is almost zero. Note that this result is consistent with the analytical result derived in equation (5.59). When using 25 frequency bins (0.022 mm^{-1} frequency bin width), the bias increases slightly as the true β value increases. The bias equals -0.001 when the true β equal to one and increases to 0.001 when the true beta equals to four. However when only seven frequency bins (0.079 mm^{-1} frequency bin width) are used, a strong dependence of

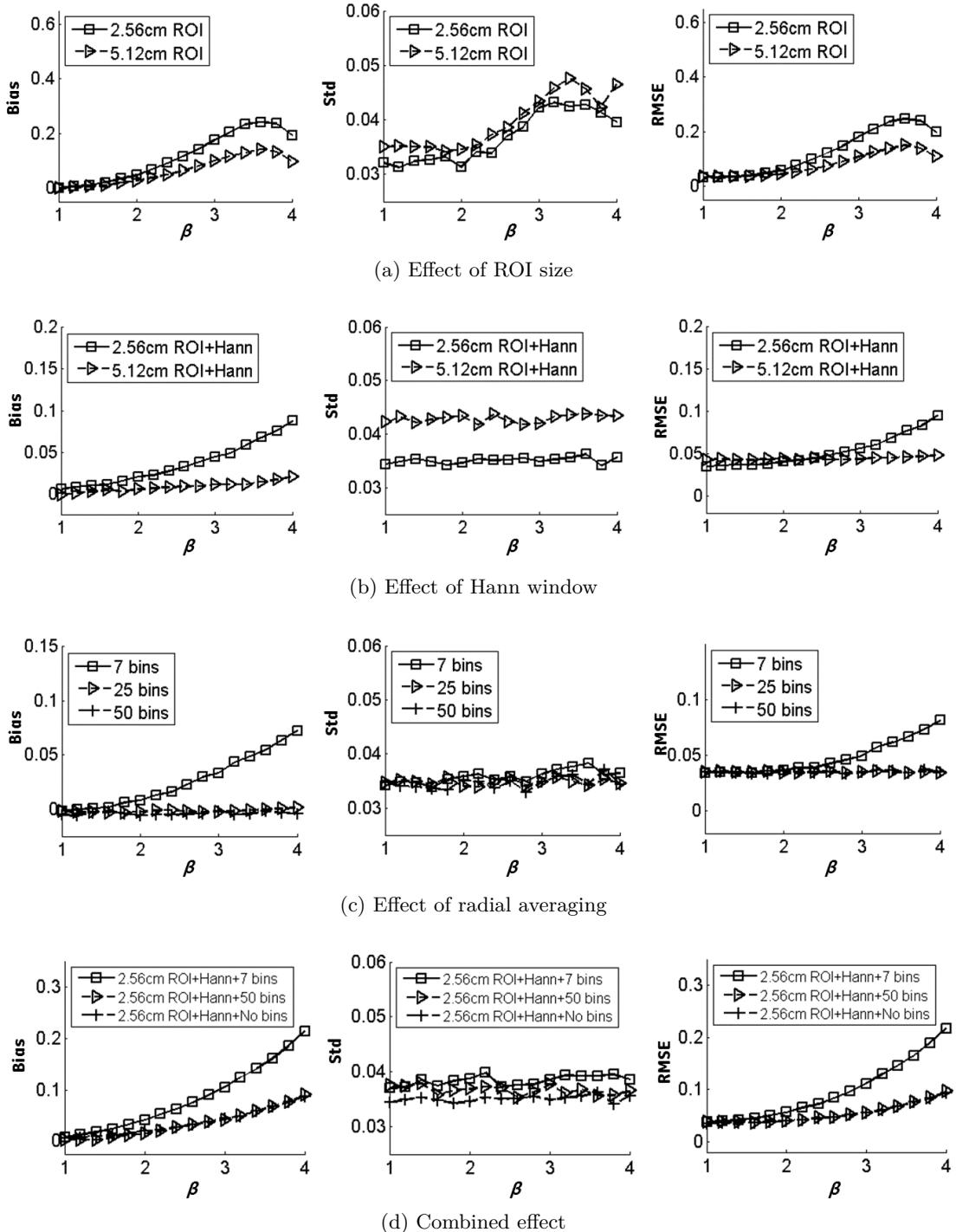


Figure 5.8: Results of the four experiments specified in Table 5.5 to assess the performance of β estimation using the Welch method. The effect of the ROI size, the Hann window function and the radial frequency averaging were first analyzed separately, then their combining effect was analyzed. For each analysis the β estimation bias, standard deviation and RMSE are plotted as functions of the true β value.

the bias on the true β value is observed; the bias equals 0 when true β equals to one and increases to 0.09 when true beta equals to four. These results are also consistent with our observations in Section 5.2.2. The standard deviation is almost independent of the true β value and stabilizes around 0.035 when 25 bins or 50 frequency bins are used. When seven frequency bins are used, the standard deviation slightly fluctuates when true β increases and is slightly larger but does not exceed 0.04. The evaluation of RMSE demonstrates that, using 25 or 50 bins for radial averaging introduces very slight estimation error independent of the true β value; while using seven bins, the estimation error depends strongly on the true β value and can be drastic.

Figure 5.8d shows the results of experiment 4 evaluating the combined effect of ROI division, window function and radial frequency averaging. When the original image is divided in 50% overlapping $2.56\text{ cm} \times 2.56\text{ cm}$ ROIs, each multiplied by a 2D Hann window function and when radial frequency averaging is applied using 50 bins, the bias increases as the true β increases. The bias is almost zero when true β equals to one and increases up to 0.09 when true β equals to four. When the experiment is repeated using the same parametrization except that radial averaging is performed using only seven bins, the bias increases faster when the true β increases. In this case, the bias is almost zero when true β equals to one and increases up to 0.22 when true β equals to four. Regardless of whether radial frequency averaging is applied or not, the standard deviation is independent of the true β value. The standard deviation does not exceed 0.037 when 50 frequency bins are used, and does not exceed 0.035 when no radial frequency averaging is applied. For all experimental conditions, the RMSE increases when the true β value increases. The RMSE values are similar when applying radial frequency averaging with 50 bins and when no radial frequency averaging is applied. In both cases, the RMSE increases from 0.04 when the true β equals to one, up to 0.1 when the true β equals to four. The RMSE is larger when only seven bins are used, and increases from 0.04 when true β equals to one, up to 0.23 when true beta equal to four.

All our experimental results are consistent with the analytically derived results.

5.2.4 Discussion

Our analytical and numerical results indicate that the RMSE in the estimated β is smallest when each input ROI is multiplied with an appropriate window and when radial averaging with small frequency bins or no radial averaging is performed to compute the (averaged) radial noise power spectrum. The ROI size yielding the smallest RMSE depends on various factors. The ROI size should be chosen considering the image resolution and the whole image size since on one hand larger ROIs have the potential to capture entirely the anatomical structure variations in breast textures and on the other hand larger ROIs result in smaller ROI numbers, which can cause increased variance. Additionally, as demonstrated in experiment 2, the choice of ROI size is also related to the true β value when each input ROI is multiplied with a 2D Hann window.

All parameters for β should be carefully chosen to reduce the estimation RMSE. As an example, when using $2.56\text{ cm} \times 2.56\text{ cm}$ ROIs, coupled with a 2D Hann window and with radial frequency averaging using 0.011 mm^{-1} frequency bins, the RMSE varies from 0.04 to 0.1 when the true β value is between one and four. However, when using $2.56\text{ cm} \times 2.56\text{ cm}$ ROIs with window function omitted and with 0.079 mm^{-1} radial frequency bins,

the RMSE varies from 0.04 to 0.23 when the true β value is between one and four.

In clinical practice, there might be a need to compromise in the choice of the ROI size to balance between the β estimation RMSE and the need to encompass enough breast texture to obtain an accurate NPS estimate. Previous studies have mentioned that ROIs with size smaller than $2.5\text{ cm} \times 2.5\text{ cm}$ might not be able to entirely capture the anatomical structure variations of breast textures [107]. However this has never been investigated in detail and is out of scope in our study.

5.3 Conclusion and discussion

In this chapter, we first proposed a new estimator to assess the volumetric breast density (VBD) from DBT projections using model-based stereology. A preliminary evaluation has shown that the absolute bias and the standard deviation of the proposed VBD estimator differ by less than 3% of the ground truth. Next, we presented a detailed evaluation of the bias and variance in the statistical estimation of the β metric inherent to different parameterizations of the commonly used estimation algorithm based on the Welch method. Our results demonstrated the limitation of the β metric, mainly due to its non-negligible systematical estimation error, which is smaller for images from 3D breast imaging systems than for 2D mammographic images.

Despite the low estimation bias and standard deviation, the proposed new VBD estimator, as well as its validation methodology have several limitations. First, the stereological estimator assumes that the thickness of the compressed breast is known a-priori. In reality, accurate estimation of breast thickness is limited and depends on the accuracy with which the compression paddle height is indicated. Also, the breast thickness was assumed to be a constant in our study. In reality, the compressed breast thickness decreases towards the nipple side. Secondly, the VBD estimator assumes that the imaged breast texture is stationary and the distribution of the projection images is invariant to different DBT projection angles. These hypotheses can be considered approximately realistic when considering only small regions of interest with the same tissue type inside the breast. However, when considering the entire breast, breast tissue components are very diverse leading to non-stationary projection images with distributions that depend on different projection angles. Design-based stereology with carefully designed sampling strategy might be considered to mitigate this limitation, since it makes less assumptions on the underlying distribution of the breast texture. Thirdly, the VBD estimator assumes a theoretical and simplified image acquisition model. Several aspects of a real image acquisition were ignored, including the cone-beam geometry, poly-energetic x-ray beam quality, x-ray scatter, image gain correction etc. Also, the preliminary performance assessment of the VBD estimator was conducted using breast texture images simulated with a simplified model of DBT image acquisition. Further study is needed to assess the impact of the simplified image acquisition model on the performance of the new VBD algorithm and to test it on a real system. Finally, the analysis and simulations performed in this study were according to the fourth edition of BI-RADS density classification. The recently published fifth edition of BI-RADS replaced the percentage density notation by a classification using a, b, c and d categories, to better encounter the masking effect of mammography. It would be interesting to test the proposed stereological estimator using the fifth edition of BI-RADS.

Our study of the β metric also has limitations. We adopted the underlying hypothesis that breast texture exhibits a radial symmetric power-law noise power spectrum. However as shown by Reiser *et al.* [163] breast texture is not radial symmetric but exhibits a rather elliptical symmetry caused by the “directionality” of fibroglandular tissue which tends to radiate from the nipple towards the chest wall. Hence for future studies, it might be of interest to investigate the estimation of β in different radial directions.

Despite these limitations, the result of the proposed VBD estimator encourage further investigation and optimization of this first model-based stereological method for breast density quantification. Our study of the β metric emphasizes the importance to carefully design the parametrization for β estimation, especially when β is used as a metric to validate the statistical realism of simulated textures or to compare the performance of different breast x-ray imaging modalities or different image processing and reconstruction algorithms.

Chapter 6

Mathematical model observer based on the *a contrario* theory for microcalcification detection in 2D and 3D breast images

In this chapter, we propose a new mathematical model observer based on the *a contrario* theory for microcalcification detection task in 2D and 3D x-ray breast images. This research is encouraged by the fact that previously developed 2D *a contrario* observer for mass detection in FFDM images has been demonstrated to produce results fairly correlated to human. However, the previously developed *a contrario* observer uses global image statistics to compute its decision variable. This might not reflect the true strategy used by human observers for the detection of microcalcifications. An *a contrario* observer that allows for a more accurate modeling of how human observers detect microcalcifications in breast images is therefore preferable.

The new *a contrario* observer is based on pixel-wise statistical tests with a global control of the average number of false detections. For each pixel, surrounding local rectangular regions with various sizes were used to construct multiple null-hypotheses. Based on the null-hypotheses, the *a contrario* observer then computes the number of false alarms (NFA) associated to the pixel using its pixel intensity value. The NFA was finally compared to a pre-defined upper bound, to determine whether the corresponding pixel intensity value is detected as abnormal. The proposed *a contrario* observer was first described for 2D breast images; then a 3D extension to DBT reconstructed slices was proposed. A theoretical proof was given to demonstrate the global control of false detections of the proposed *a contrario* observer. A simulation experiment was performed to further validate the theoretical proof.

6.1 Background

A contrario detection has been introduced by Desolneux, Moisan and Morel as a statistical detection approach inspired by the Gestalt school of human vision [55] [54]. The central idea is to define a set of “grouping laws” (alignment, closeness, parallelism *etc.*) governing the low-level human perception. The grouping laws aim to represent characteristics immediately perceived by the human in an image.

The Latin term *a contrario*, which means “by contradiction”, is used to express the grouping process by the *Helmholtz principle* [202]. This principle states that certain configurations of objects according to Gestalt grouping laws rarely happen by chance in an independent random realization of objects. Inspired by this principle, the *a contrario* detection consists of quantifying the amount of “surprise” in a “purely random” image by measuring features according to the Gestalt grouping laws. In a classical *a contrario* framework, the “purely random” image is the null-hypothesis H_0 for the detection, where the object to be detected is absent. The null-hypothesis H_0 is referred to as the *naïve model*. The amount of “surprise” of measured features is quantified by the *number of false alarms* (NFA) defined as follows [78].

Definition 6.1. (Number of false alarms)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{X_i\}_{i=1,\dots,N}$ be a set of random variables from Ω to \mathbb{R} . The number of false alarms is a real-valued function $f(i, \cdot)$ satisfying:

$$\forall \epsilon > 0, \quad \mathbb{E}(|\{i|f(i, X_i) \leq \epsilon\}|) \leq \epsilon. \quad (6.1)$$

The definition of NFA can be interpreted by analogy to the multiple statistical tests. The set of random variables $\{X_i\}_{i=1,\dots,N}$ are associated to a set of *measurements* performed at random image locations. The joint distribution of $\{X_i\}_{i=1,\dots,N}$ is specified by the naïve model H_0 . For any function that satisfies (6.1), the expected number of false alarms is less than ϵ for the family of tests $\{f(i, X_i) \leq \epsilon\}_{i=1,\dots,N}$ with H_0 as null-hypothesis. That is, the definition of NFA allows for a global control of the average number of false positives.

The canonical NFA formulation of the *a contrario* detection is given by the following proposition [78]:

Proposition 6.1. Consider $\{X_i\}_{i=1,\dots,N}$ a set of random variables. Then the function $f(i, \cdot)$, defined as:

$$f(i, x_i) = N \mathbb{P}(X_i \geq x_i) \quad (6.2)$$

is NFA defined in 6.1.

The NFA associated to the i^{th} measurement is the product of the the *number of tests* and the *tail probability* of the measurement outcome x_i conditioned on the naïve model H_0 . The detection and decision phase of the *a contrario* approach are summarized as follows:

- During the detection phase, multiple measurements are performed across the image to obtain a set of measurement values $\{x_i\}_{i=1,\dots,N}$ associated to the set of random variables $\{X_i\}_{i=1,\dots,N}$.

- The decision phase relies on the naïve model as the null-hypothesis for the joint distribution of $\{X_i\}_{i=1,\dots,N}$. Given $\epsilon > 0$, the outcome x_i of the i^{th} measurement is said to be ϵ -meaningful if $\text{NFA}(i, x_i) \leq \epsilon$ is satisfied.

The detection outcomes are obtained by fixing a value for ϵ , and collecting the indices of all ϵ -meaningful measurements.

Classical Bayesian maximum a posteriori method for object detection necessitates an *a priori* model and the detected objects are close to the *a priori* model. The definition of “close” is associated to an optimization involving decision-makings and parameters that need to be carefully tuned. The choice of a good *a priori* model and the tuning of parameters are two general difficulties in the maximum a posteriori method [139] [54]. A *contrario* detection provides an alternative statistical approach that may mitigate these two difficulties in the maximum a posteriori method. Instead of searching for objects close to an *a priori* model, the *a contrario* approach searches for large deviations, or “counter-examples” from a naïve model. The *Helmholtz principle* does not define the naïve model as a model for the images, but as a model from which our perception searches for counter-examples [54]. The naïve model may still be treated as an *a priori* model, but it does not need to be precise and can be simplified to a uniform or a Gaussian random field in most cases [54] [78]. With this advantage, the naïve model in the *a contrario* approach is usually easy to define. A *contrario* detection also does not necessitate complicated parameter tuning. The NFA bound ϵ seems to be the only parameter to tune. However the detection NFA has a weak dependency on ϵ and in practice ϵ is usually set to 1, meaning that less than one false detection is expected [54].

6.2 Related work

In breast imaging, Rick *et al.* [165] and Grosjean *et al.* [80] investigated the *a contrario* framework to model lesion detection in 2D FFDM.

In [165], Rick *et al.* applied the *a contrario* approach to model the microcalcification (μcalc) detection in 2D FFDM images. The naïve model was defined as a Gaussian noise image with *a priori* known distribution. The measurement at each pixel location was set to be the contrast between the pixel value and the mean of the Gaussian noise image at the same pixel location. Based on this set-up, NFA values were computed encountering various μcalc sizes at each pixel location of the input image. The obtained NFA values were used to study the theoretical μcalc detection false positive rate in function of a threshold level applied to the image.

In [80], Grosjean *et al.* developed an *a contrario* observer to assess the detection performance of masses in 2D FFDM images. In their approach the 2D power-law Gaussian random field was used as the naïve model for the *a contrario* observer. The measurement of the *a contrario* observer was defined as a measurement of local contrast at a pixel location. It consists of measuring the difference in the average signal intensity between two circular regions of different sizes centered at the same pixel location. The theoretical NFA formulation based on the defined naïve model and measurement was derived [79]. The authors demonstrated that the contrast-detail curve obtained by the *a contrario* observer was consistent with previous finding by Burgess *et al.* using a prewhitening matched filter [33].

Both *a contrario* models in [165] and [80] used the *a priori* known image statistics to derive the NFA. In the model of Rick *et al.*, the images statistics were the mean and the standard deviation of an *a priori* known Gaussian noise image. The limitation of the model of Rick *et al.* is that the Gaussian noise image considers only the local x-ray noise and electronic noise and does not encounter the local breast texture variation. In the model of Grosjean *et al.*, the statistics were derived from a global 2D power-law Gaussian random field. The use of global statistics has limitations for modeling how human observers detect μ calcs in breast images. Previous psycho-physical studies have pointed out that human observers use limited local image content for the detection of μ calcs in breast images [28] [206]. Also human observers tend to use a detection strategy adapted to each local image content for μ calc detection in complex non-stationary backgrounds [206].

To our knowledge, the application of the *a contrario* framework for 3D DBT has not yet been investigated. In the following sections, we propose a new *a contrario* observer to assess the μ calc detection performance in 2D breast images and 3D DBT reconstructed slices.

6.3 Design of the *a contrario* observer for microcalcification detection in 2D and 3D breast imaging

In this section we describe in detail the proposed new *a contrario* observer for μ calc detection task. We first consider μ calc detection in 2D images, then we propose a 3D extension of the *a contrario* observer to encounter μ calc detection in DBT reconstructed slice.

Let $\Omega \subset \mathbb{Z}^2$ be a 2D discrete domain where an input image \mathbf{u} is defined. The naïve model where no counter-examples should be detected was assumed to be a white Gaussian image \mathbf{u}_0 defined on Ω . Without loss of generality, we assumed that $\forall x \in \Omega$, $\mathbf{u}_0(x) \sim \mathcal{N}(0, 1)$. To address the limitations of previous related models [165] [80] described in Section 6.2, the design of our proposed *a contrario* observer encompasses the following characteristics.

- For each image pixel, the computation of NFA was limited to rectangular regions of interest with maximum side set to 5 mm \times 5 mm. This aims to encounter the local detection characteristic of human observers for μ calcs [28] [206].
- The computation of the NFA at each image location was based on local statistics empirically estimated from the corresponding local region of the input image. The empirical statistics encounter the local image noise and the local breast texture variation and aim to reflect human observers' strategy to adapt to the local image content for the detection of μ calcs [206].
- Different spatial sizes were considered for each NFA computation. This aims to encounter local detections at multiple spatial scales.

We describe hereunder the steps of the proposed *a contrario* observer in detail.

6.3.1 Construction of multi-scale neighborhoods

For a given pixel location $x \in \Omega$, the *a contrario* observer considers local regions surrounding a potential μ_{calc} centered at x . These regions are referred to as the neighborhoods. A neighborhood V_x was obtained in the following way. First, we defined two rectangles centered at x : an inner rectangle R_x^i and an outer rectangle R_x^o , with the size of R_x^o bigger than the size of R_x^i (Figure 6.1). Then, the neighborhood V_x was obtained as $V_x = \Omega \cap R_x^o \setminus R_x^i$. This indicates that, for each neighborhood V_x , the *a contrario* observer considers only pixels inside V_x . The inner rectangle R_x^i was excluded to eliminate the pixels around x that might contain a μ_{calc} . If the extent of R_x^o or R_x^i bypasses the image domain Ω , then only the pixels inside Ω are considered.

To account for multiple sizes for the μ_{calcs} , we varied the size of R_x^i at multiple scales. Similarly the size of R_x^o was also varied at multiple scales to consider different spatial extents for the neighborhoods. The variations of R_x^i and R_x^o are described as follows. Let r_h^i and r_v^i denote the horizontal and vertical side lengths of R_x^i respectively. Let r_h^o and r_v^o denote the horizontal and vertical side lengths of R_x^o respectively. The magnitudes (in pixels) of r_h^i and r_v^i were separately incremented from a minimum value r_{\min}^i to a maximum value r_{\max}^i with a step of 1 pixel. Similarly, r_h^o and r_v^o were also separately incremented from a minimum value r_{\min}^o to a maximum value r_{\max}^o with a step of 1 pixel. All combinations of R_x^o and R_x^i were collected. To insure that R_x^o is bigger than R_x^i , we imposed the following conditions for each combination:

$$r_h^o - r_h^i \geq \delta \text{ and } r_v^o - r_v^i \geq \delta. \quad (6.3)$$

Combinations that did not satisfy the condition were left out. The retained combinations were selected as the final collection of neighborhoods at pixel location x . The collection is denoted as \mathcal{V}_x . The neighborhoods constructed at all other pixel locations used the same values for $r_{\min}^i, r_{\max}^i, r_{\min}^o, r_{\max}^o$ and δ .

The construction of multi-scale neighborhoods is illustrated in Figure 6.1.

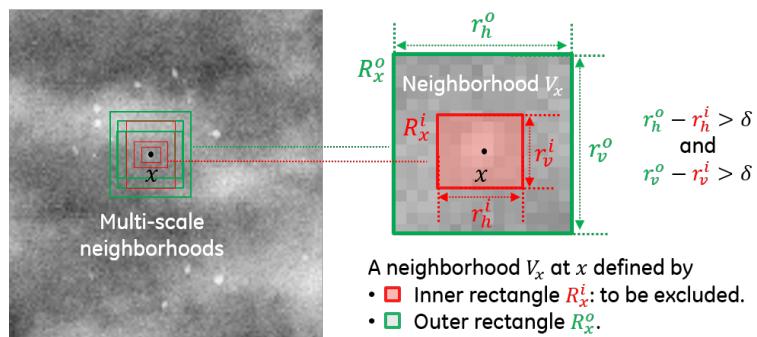


Figure 6.1: Construction of multi-scale neighborhoods. A neighborhood V_x was obtained by subtracting an inner rectangle R_x^i from an outer rectangle R_x^o to eliminate the pixels around x that might contain a μ_{calc} . The size of both R_x^i and R_x^o were varied at multiple scales to account for multiple sizes for the μ_{calcs} . To ensure that the outer rectangle is bigger than the inner rectangle, a minimum value was imposed for the difference between r_h^o and r_h^i , and the difference between r_v^o and r_v^i .

6.3.2 Computation of the NFA image

For a given neighborhood V_x at pixel position $x \in \Omega$, the mean μ_{V_x} and the standard deviation σ_{V_x} of the values of pixels in V_x were empirically estimated as:

$$\mu_{V_x} = \frac{1}{|V_x|} \sum_{y \in V_x} \mathbf{u}(y), \quad (6.4)$$

$$\sigma_{V_x} = \sqrt{\frac{1}{|V_x|} \sum_{y \in V_x} (\mathbf{u}^2(y) - \mu_{V_x}^2)}. \quad (6.5)$$

Since multiple neighborhoods exist at x , a collection of empirical statistics $\{\mu_{V_x}, \sigma_{V_x}\}_{V_x \in \mathcal{V}_x}$ were obtained at x .

The measurement $m(x)$ at x was defined as the pixel intensity value at x . That is $m(x) = \mathbf{u}(x)$. We propose to formulate the “NFA” at pixel x as:

$$\text{NFA}(x) = |\Omega| |\mathcal{V}_x| \min_{V_x \in \mathcal{V}_x} \Phi \left(\frac{\mathbf{u}(x) - \mu_{V_x}}{\sigma_{V_x}} \right). \quad (6.6)$$

Here $\Phi(\cdot)$ is the tail probability function of the standard normal distribution $\mathcal{N}(0, 1)$. The proposed “NFA” can be interpreted in the following way. At each pixel location x , the *a contrario* observer scans through all neighborhoods. For each neighborhood, it quantifies how the measurement value $m(x)$ deviates to the naïve model according to the empirically measured μ_{V_x} and σ_{V_x} . The deviation is inverse proportional to the tail probability of the Gaussian distribution $\mathcal{N}(\mu_{V_x}, \sigma_{V_x})$. Finally the smallest tail probability value was retained and was assigned to $\text{NFA}(x)$, since it represents the largest deviation possible among all searched neighborhoods.

We will demonstrate in the next section that the proposed “NFA” does not strictly satisfy the Definition 6.1, but can still be used to control the global number of false alarms. Applying (6.6) to all pixel positions in Ω , we obtained the final NFA image.

6.3.3 Extension for μ_{calc} detection in 3D DBT slices

Microcalcifications are typically smaller (200 mm to 800 mm) than the distance between DBT reconstructed slices (typically 1 mm). Therefore, a μ_{calc} appears often only in focus in maximum one slice. Due to this fact, the spatial correlation between adjacent DBT reconstructed slices is relatively weak. Based on this rationale, we propose to extend the *a contrario* observer to DBT slices by individually processing each slice, to obtain separately the NFA image of each slice. The complete algorithm for the proposed *a contrario* observer is described in Algorithm 6.1.

6.3.4 The *a contrario* observer for location-known-exactly tasks

To perform a location-known-exactly detection task at location x_l , we restricted the computation of the NFA to a small neighborhood v_l centered at x_l . In this case, the output of the *a contrario* observer is a scalar decision variable λ defined as:

$$\lambda = \min_{x \in v_l} \text{NFA}(x). \quad (6.7)$$

Algorithm: The *a contrario* observer for μ_{calc} detection

Input: An image \mathbf{u} defined on $\Omega \in \mathbb{Z}^2$, $\delta \in \mathbb{R}^+$, $r_{\min}^i, r_{\max}^i, r_{\min}^o, r_{\max}^o \in \mathbb{N}^*$
with $r_{\min}^i < r_{\max}^i$ and $r_{\min}^o < r_{\max}^o$.

Output: The NFA image.

```

1   $n = 0;$ 
2  for  $x \in \Omega$  do
3    for  $r_h^i \in \{r_{\min}^i, \dots, r_{\max}^i\}$  do
4      for  $r_v^i \in \{r_{\min}^i, \dots, r_{\max}^i\}$  do
5        for  $r_h^o \in \{r_{\min}^o, \dots, r_{\max}^o\}$  do
6          for  $r_v^o \in \{r_{\min}^o, \dots, r_{\max}^o\}$  do
7            if  $r_h^o - r_h^i \geq \delta$  and  $r_v^o - r_v^i \geq \delta$  then
8               $n \leftarrow n + 1;$ 
9               $R_x^i \leftarrow R(x, r_h^i, r_v^i), R_x^o \leftarrow R(x, r_h^o, r_v^o);$ 
10              $V_x \leftarrow \Omega \cap R_x^o \setminus R_x^i;$ 
11              $\mu_{V_x} \leftarrow \frac{1}{|V_x|} \sum_{y \in V_x} \mathbf{u}(y), \sigma_{V_x}$ 
12                $\leftarrow \sqrt{\frac{1}{|V_x|} \sum_{y \in V_x} (\mathbf{u}^2(y) - \mu_{V_x}^2)};$ 
13              $p_{V_x} \leftarrow \Phi\left(\frac{\mathbf{u}(x) - \mu_{V_x}}{\sigma_{V_x}}\right).$ 
14    $\text{NFA}(x) \leftarrow n|\Omega| \min_{V_x \in \mathcal{V}} p_{V_x};$ 
15   if DBT reconstructed slices as input then
16     repeat line 1 to 13 for all slices and divide each NFA value by the
       number of slices.

```

Algorithm 6.1

For DBT, we propose to compute the decision variable λ in the following way. First, a subset of adjacent slices encompassing the μ_{calc} was selected. Next, for the i^{th} slice in the subset, we restricted the computation of its NFA image to a small local neighborhood v_l^i centered at x_l^i , where x_l^i is the projection of the lesion location x_l onto the plane of the i^{th} slice. Then, using (6.7), a decision variable was computed for each selected slice. The minimum of this set of decision variables for all slices was selected as the final decision variable λ .

The computation of the decision variable of the *a contrario* observer for lesion known exactly tasks in DBT reconstructed slices is illustrated in Figure 6.2.

6.4 Validation of the proposed *a contrario* observer

As described in Section 6.1, the main property of the *a contrario* detection is the global control of the average number of false positives. In the following sections, we demonstrate this property of the proposed *a contrario* observer by a theoretical proof, as well as a simulation experiment.

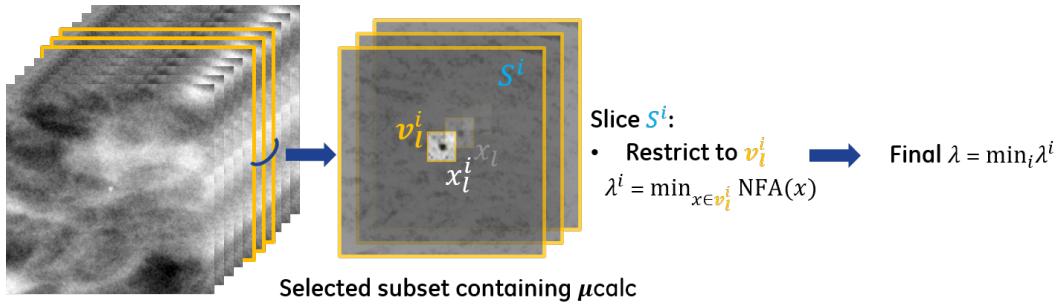


Figure 6.2: Computation of the scalar decision variable λ of the *a contrario* observer for lesion known exactly tasks in DBT reconstructed slices, whereby the location of the microcalcification is at x_l . A subset of slices that encompasses the μ_{calc} is first selected. For the i^{th} slice in the subset, we restricted the computation of its NFA image to a small local neighborhood v_l^i centered at x_l^i , where x_l^i is the projection of the lesion location x_l onto the plane of the i^{th} slice. Then a decision variable is obtained for each slice by taking the minimum of all NFA values in v_l . The final decision variable λ is obtained by taking the minimum of all the decision variables for all slices.

6.4.1 Theoretical proof of global false positive control for the proposed *a contrario* observer

The following proposition demonstrates that the proposed NFA in (6.6) allows for a global control of the average number of false positives, even though the formal definition of NFA given in Definition 6.1 is not strictly satisfied.

Proposition 6.2. *Let \mathbf{u} be an image defined on $\Omega \subset \mathbb{Z}^2$. Let $\mathcal{N}(0, 1)$ be the naïve model of the *a contrario* observer described in Algorithm 6.1 with the NFA formulation defined in (6.6). Then for any $x \in \Omega$ and $\epsilon > 0$, the NFA satisfies:*

$$\mathbb{E}(\#\{x | \text{NFA}(x) \leq \epsilon\}) < \mathcal{C}(\epsilon). \quad (6.8)$$

Let \mathcal{V} denotes all neighborhoods at any arbitrary pixel location. If for any neighborhood $V \in \mathcal{V}$, we have $|V| \gg (\Phi^{-1}(\frac{\epsilon}{t}))^2$, where $t = |\Omega||\mathcal{V}|$, then the upper bound $\mathcal{C}(\epsilon)$ has the same order of magnitude as ϵ .

Proof. $\forall x \in \Omega$ and $\epsilon > 0$, we have

$$\begin{aligned} \mathbb{E}(\#\{x | \text{NFA}(x) \leq \epsilon\}) &= \mathbb{E}\left(\sum_{x \in \Omega} \mathbb{1}_{\text{NFA}(x) \leq \epsilon}\right) \\ &= \sum_{x \in \Omega} \mathbb{P}(\text{NFA}(x) \leq \epsilon) \\ &= \sum_{x \in \Omega} \mathbb{P}\left(\min_{V_x \in \mathcal{V}_x} \Phi\left(\frac{\mathbf{u}(x) - \mu_{V_x}}{\sigma_{V_x}}\right) \leq \frac{\epsilon}{t}\right) \\ &\leq \sum_{x \in \Omega} \sum_{V_x \in \mathcal{V}_x} \mathbb{P}\left(\mathbf{u}(x) - \mu_{V_x} \geq \sigma_{V_x} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right). \end{aligned} \quad (6.9)$$

Since the multi-scale neighborhoods for all pixel locations are constructed in the same way, $t = |\Omega||\mathcal{V}|$ is a constant independent of the pixel location.

For an arbitrary neighborhood $V_x \in \mathcal{V}_x$ centered at $x \in \Omega$, we assume that x and pixels in V_x are independent. Since x is separated from V_x by the inner rectangle R_x^i , this is a reasonable assumption when R_x^i is large enough. The sample mean μ_{V_x} and standard deviation σ_{V_x} of the values of the pixels inside V_x are estimated according to (6.4) and (6.5). Based on the naïve model, μ_{V_x} is a Gaussian random variable with mean 0, and standard deviation $\frac{1}{\sqrt{v}}$. Here for the ease of notation we denote $v = |V_x|$. Therefore,

$$\mathbf{u}(x) - \mu_{V_x} \sim \mathcal{N}\left(0, \sqrt{1 + \frac{1}{v}}\right). \quad (6.10)$$

This gives

$$\begin{aligned} & \mathbb{P}\left(\mathbf{u}(x) - \mu_{V_x} \geq \sigma_{V_x} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right) \\ &= \mathbb{P}\left(\frac{\mathbf{u}(x) - \mu_{V_x}}{\sqrt{1 + \frac{1}{v}}} \geq \frac{\sigma_{V_x}}{\sqrt{1 + \frac{1}{v}}} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right) \\ &= \int_0^\infty \Phi\left(\frac{s}{\sqrt{1 + \frac{1}{v}}} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right) \times p_{\sigma_{V_x}}(s) ds. \end{aligned} \quad (6.11)$$

Here $p_{\sigma_{V_x}}$ is the distribution of σ_{V_x} . The last equality stems from the fact that μ_{V_x} and σ_{V_x} are independent random variables (a proof can be found in [68]). The distribution $p_{\sigma_{V_x}}$ can be explicitly written as [99]:

$$p_{\sigma_{V_x}}(s) = ks^{v-2} \exp\left(-\frac{vs^2}{2}\right), \quad (6.12)$$

with

$$k = 2 \frac{\left(\frac{v}{2}\right)^{(v-1)/2}}{\Gamma\left(\frac{v-1}{2}\right)}. \quad (6.13)$$

It follows that,

$$\mathbb{P}\left(\mathbf{u}(x) - \mu_{V_x} \geq \sigma_{V_x} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right) = \int_0^\infty k \Phi\left(\frac{s}{\sqrt{1 + \frac{1}{v}}} \Phi^{-1}\left(\frac{\epsilon}{t}\right)\right) s^{v-2} \exp\left(-\frac{vs^2}{2}\right) ds. \quad (6.14)$$

Furthermore we have the following lemma [66] [61]:

Lemma 6.1. *Let $\Phi(\cdot)$ denote the tail probability of the standard normal distribution $\mathcal{N}(0, 1)$, then for any $x > 0$, the following inequalities hold:*

$$\frac{x}{(x^2 + 1)\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < \Phi(x) < \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (6.15)$$

Let

$$r_\epsilon = \frac{\Phi^{-1}\left(\frac{\epsilon}{t}\right)}{\sqrt{1 + \frac{1}{v}}}, \quad (6.16)$$

we have,

$$\begin{aligned}
& \mathbb{P} \left(\mathbf{u}(x) - \mu_{V_x} \geq \sigma_{V_x} \Phi^{-1} \left(\frac{\epsilon}{t} \right) \right) \\
& < \int_0^\infty k \frac{1}{\sqrt{2\pi} r_\epsilon s} \exp \left(-\frac{r_\epsilon^2 s^2}{2} \right) s^{v-2} \exp \left(-\frac{vs^2}{2} \right) ds \\
& = \frac{k}{2\sqrt{2\pi} r_\epsilon} \left(\frac{v+r_\epsilon^2}{2} \right)^{-\frac{v-2}{2}} \Gamma \left(\frac{v-2}{2} \right) \\
& = \left(\frac{v}{2} \right)^{\frac{v-1}{2}} \frac{\Gamma \left(\frac{v-2}{2} \right)}{\Gamma \left(\frac{v-1}{2} \right)} \frac{\sqrt{1+\frac{1}{v}}}{\sqrt{2\pi} \Phi^{-1} \left(\frac{\epsilon}{t} \right)} \left(\frac{v+1+(\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2}{2 \left(1+\frac{1}{v} \right)} \right)^{-\frac{v-2}{2}} \\
& = \frac{\Gamma \left(\frac{v-2}{2} \right)}{\Gamma \left(\frac{v-1}{2} \right)} \frac{\sqrt{v+1}}{2\sqrt{\pi} \Phi^{-1} \left(\frac{\epsilon}{t} \right)} \left(\frac{v+1}{v+1+(\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2} \right)^{\frac{v-2}{2}}.
\end{aligned} \tag{6.17}$$

Denote

$$\gamma = \frac{\Gamma \left(\frac{v-2}{2} \right)}{\Gamma \left(\frac{v-1}{2} \right)}, \tag{6.18}$$

then the following lemma can be used to obtain an upper bound for γ [73]:

Lemma 6.2. *Let $a \in \mathbb{R}^+$ and $0 \leq b \leq 1$. The following inequality holds for the Gamma function Γ :*

$$\frac{\Gamma(a+b)}{\Gamma(a+1)} \leq a^{b-1}. \tag{6.19}$$

Now assume $\frac{v-3}{2} > 0$. Let $a = \frac{v-3}{2}, b = \frac{1}{2}$, we immediately have,

$$\gamma = \frac{\Gamma \left(\frac{v-2}{2} \right)}{\Gamma \left(\frac{v-1}{2} \right)} \leq \sqrt{\frac{2}{v-3}}. \tag{6.20}$$

Now plug this in 6.17 and 6.9, we finally have:

$$\mathbb{E}(\#\{x | \text{NFA}(x) \leq \epsilon\}) < \mathcal{C}(\epsilon), \tag{6.21}$$

where,

$$\mathcal{C}(\epsilon) = \frac{|\Omega|}{2\sqrt{\pi} \Phi^{-1} \left(\frac{\epsilon}{t} \right)} \sum_{V \in \mathcal{V}} \sqrt{\frac{2v+2}{v-3}} \left(\frac{v+1}{v+1+(\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2} \right)^{\frac{v-2}{2}}. \tag{6.22}$$

Furthermore, if we assume that for any $V \in \mathcal{V}$, $v = |V| \gg (\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2$, then $\mathcal{C}(\epsilon)$ can be approximate as:

$$\begin{aligned}
\mathcal{C}(\epsilon) & \approx \frac{|\Omega|}{\sqrt{2\pi} \Phi^{-1} \left(\frac{\epsilon}{t} \right)} \sum_{V \in \mathcal{V}} \left(1 - \frac{(\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2}{v+1} \right)^{\frac{v-2}{2}} \\
& \approx \frac{t}{\sqrt{2\pi} \Phi^{-1} \left(\frac{\epsilon}{t} \right)} \exp \left(-\frac{(\Phi^{-1} \left(\frac{\epsilon}{t} \right))^2}{2} \right).
\end{aligned} \tag{6.23}$$

Using again Lemma 6.1 on the last term in Equation (6.23), we can see that $\mathcal{C}(\epsilon)$ has the same of order of magnitude as

$$t\Phi\left(\Phi^{-1}\left(\left(\frac{\epsilon}{t}\right)\right)\right) = \epsilon. \quad (6.24)$$

This concludes the proof. \square

6.4.2 Simulation experiment

To further illustrate the global control of false positives for the proposed *a contrario* observer, an experiment using simulated 2D images was conducted.

Two types of input images were considered: white Gaussian noise images with pixels drawn from $\mathcal{N}(0, 1)$ and the BI-RADS breast density 2 type mammographic projection images simulated using the 3D breast texture model according to Section 3.3. All input images were defined on a square domain Ω with size $2\text{ cm} \times 2\text{ cm}$. Image pixels were isotropic with size equal to $0.1\text{ mm} \times 0.1\text{ mm}$. For each type of input images, 200 random realizations were generated. Each input image was also inserted with ten random exponential kernels to mimic the presence of microcalcifications. The i^{th} inserted exponential kernel $f_i(\cdot)$ is mathematically expressed as:

$$f_i(x) = k \exp\left(\frac{-\|x - c_i\|^2}{0.08\sigma_i^2}\right), \quad (6.25)$$

for a pixel $x \in \Omega$. Its center c_i was uniformly randomly chosen in Ω . The parameter σ_i controls the width of the kernel and was chosen according to a uniform distribution $\mathcal{U}(200\text{ }\mu\text{m}, 600\text{ }\mu\text{m})$ to represent typical microcalcification diameter. The parameter k controls the magnitude of the exponential kernel and was set to 15.0 for white noise images and to 500.0 for BI-RADS breast density 2 images. Figure 6.3 (first row) illustrates a white Gaussian noise image and a BI-RADS breast density 2 mammographic projection image with and without the insertions of exponential kernels.

The proposed *a contrario* observer was used to score both images with and without inserted exponential kernels. The parameters δ , r_{\min}^i , r_{\max}^i , r_{\min}^o and r_{\max}^o of the *a contrario* observer were set according to Table 6.1. A range of ϵ values were considered, varying from 10^{-2} to 10, with a step of 10^{-2} . For each ϵ value, we empirically measured the average number of false positive detections per input image. The false positive detections were counted using the following rules. For an input image without exponential kernels, the number of false positive detections is

$$|\{x | \text{NFA}(x) \leq \epsilon\}_{x \in \Omega}|. \quad (6.26)$$

For an input image with exponential kernels, the number of false positive detections is

$$|\{x | \text{NFA}(x) \leq \epsilon \text{ and } \forall i \in \{1, \dots, 10\}, \|x - c_i\| > 1.5\sigma_i\}_{x \in \Omega}|. \quad (6.27)$$

Figure 6.3 illustrates the NFA images (second column) obtained by the *a contrario* observer for the four input images shown in the first column of this figure. The detection results obtained by thresholding the NFA images at $\epsilon = 0.01$ (third column) and $\epsilon = 1$ (fourth column) are also shown. The false positive detections are marked with red circles in the figure.

Parameter	Value
δ	8 pixels
r_{\min}^i	3 pixels
r_{\max}^i	11 pixels
r_{\min}^o	10 pixels
r_{\max}^o	35 pixels

Table 6.1: Configuration of the parameters of the *a contrario* observer for the numerical experiment to illustrate Proposition 6.2.

Figure 6.4 compares the analytical average false positive upper-bound given by Equation (6.22) and the empirically measured average false positive detections in white Gaussian and BI-RADS breast density 2 images with and without exponential inserts. This figure shows that the empirically measured average number of false positive detections is below the analytical NFA bound for all investigated ϵ values and for all types of input images.

The result of this experiment validates the Proposition 6.2 and demonstrates that the proposed *a contrario* observer allows indeed for a global control of the average number of false positives.

6.5 Conclusion and discussion

In this chapter, we proposed a new mathematical observer for microcalcification (μ_{calc}) detection task in 2D and 3D DBT breast imaging. The proposed observer is based on the *a contrario* detection theory, a mathematical formulation of the visual perception proposed by the “Gestalt psychology”. The central idea is that perturbations of the statistical image background properties can be perceived by humans. The proposed *a contrario* observer scans the image pixel-by-pixel. At each pixel location, it uses empirically estimated local image statistics to quantify the deviation of the pixel intensity value from a null hypothesis background image model where large deviations should happen with small probability. The key analytical characteristic of the proposed *a contrario* observer is its global control of the average false positives by a pre-defined parameter ϵ . Theoretical justification was obtained proving that the average false positives of the proposed *a contrario* observer was indeed bounded by a mathematically tractable function having the same order of magnitude as ϵ . A simulation experiment was performed that demonstrated the global control of false positives for the proposed *a contrario* observer.

This model differs from classical linear model observers which use a template to derive the decision variable. The template typically contains knowledge of the size, image contrast and shape of the lesion, which depending on the model type, are either exactly or statistically known, or are estimated from a set of training images. The template may also contain the correlation characteristics of the image background texture for models such as the channelized Hotelling observer. The *a contrario* observer assumes minimum knowledge of the signal and relies on a simplified approximate background model to compute its decision variable. Except for the μ_{calc} size, no other a priori assumptions of the μ_{calc}

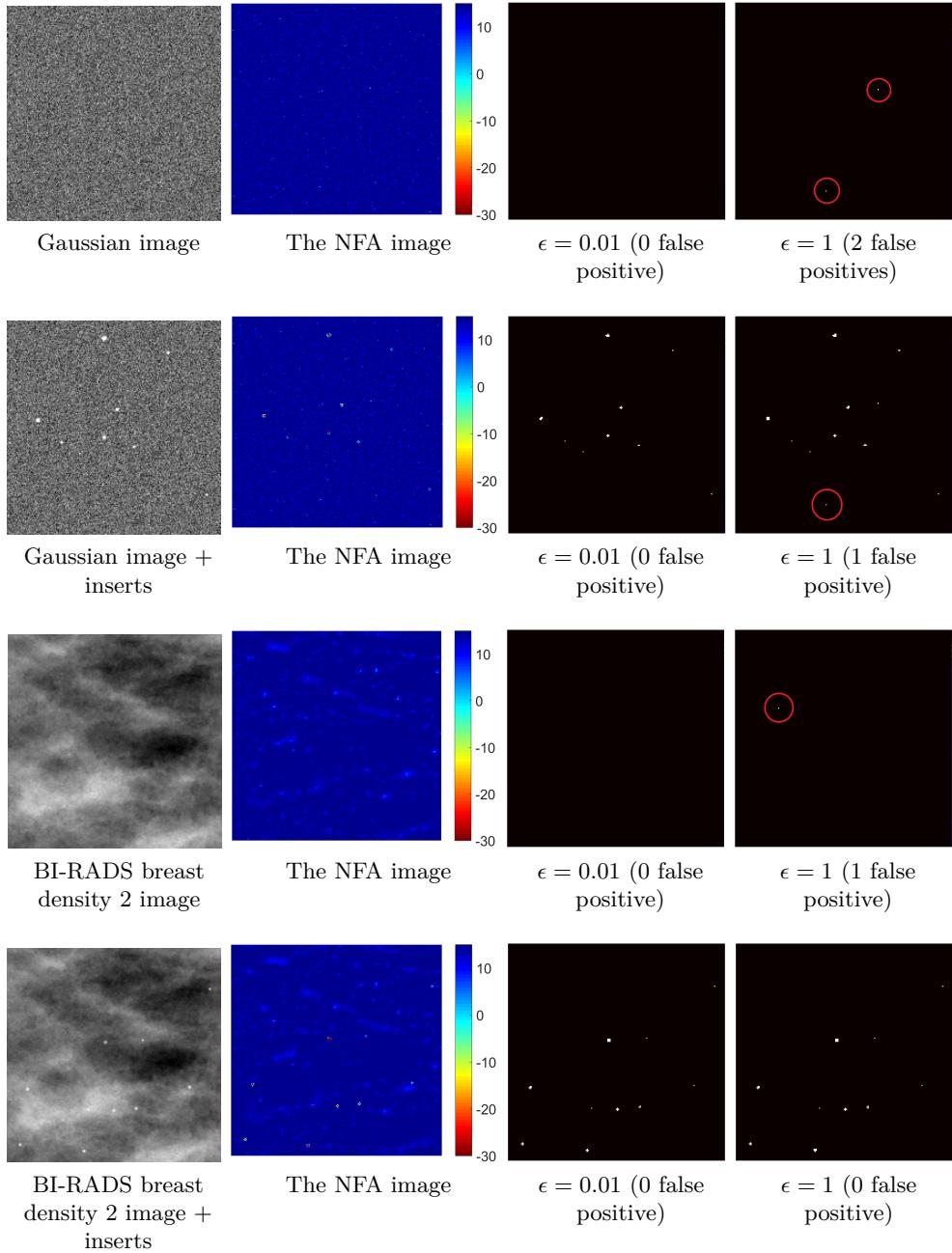


Figure 6.3: Four examples illustrating the simulation experiment used to validate the proposed *a contrario* observer. This figure shows input images to the *a contrario* observer (first column), the NFA images obtained by the *a contrario* observer (second column) and images obtained by thresholding the NFA images at $\epsilon = 0.01$ (third column) and at $\epsilon = 1$ (fourth column). The first and the third row shows the result from an input white Gaussian noise image and an input BI-RADS breast density 2 mammographic projection image respectively. The second and the fourth row demonstrate the result from the same images in the first and third row with randomly inserted exponential kernels. All images have size 2 cm \times 2 cm.

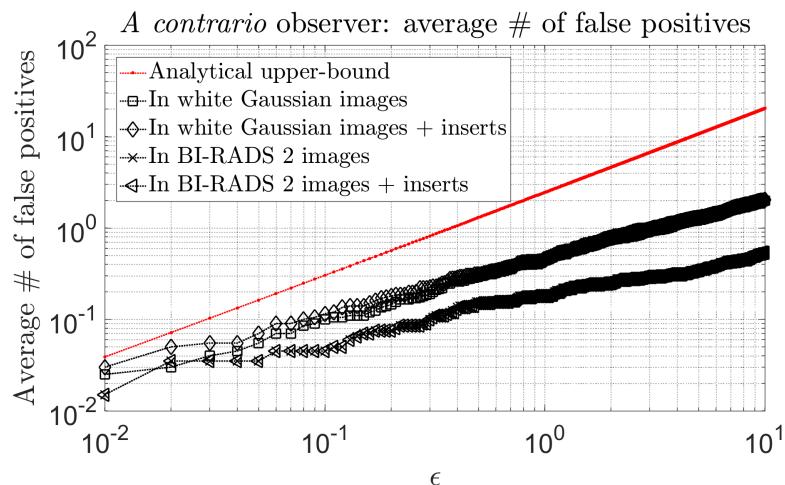


Figure 6.4: Comparison of the analytical average false positive upper-bound given by Equation (6.22) with empirically measured average false positive detections from Gaussian white noise and BI-RADS breast density 2 images with and without inserted exponential kernels.

were encountered by the *a contrario* observer.

A limitation of the current *a contrario* observer is that the correlation between adjacent DBT slices is not considered. For the detection task of μ calcs in reconstructed images corrected for out-of-focus μ calc artifacts, this model may be sufficient. However, consideration of inter-slice correlations may be key when out-of-focus μ calc artifacts are not addressed in DBT image reconstruction. The design of the experiment to validate the proposed *a contrario* observer also has limitations. The goal of the experiment was to validate the analytically demonstrated false positive control of the *a contrario* observer. It would also be interesting to test the true positive rate of the *a contrario* observer.

In the next chapter, the proposed *a contrario* observer will be compared to the performance of 2D and 3D state-of-the-art CHO models for μ calc detection.

Chapter 7

Application to microcalcification detectability in full field digital mammography, digital breast tomosynthesis and synthetic two-dimensional images

In this chapter, a virtual clinical trial (VCT) experiment was designed to assess microcalcification detection performance using VCT tools developed as part of this thesis. Our objectives were three-fold. The first goal was to compare the performance of the *a contrario* observer (ACO) developed in Chapter 6 with state-of-the-art channelized Hotelling observers (CHO). The second goal was to preliminarily compare the performance of the ACO and the CHO with the performance of human observer. The third goal was to compare the performance of microcalcification detection in 2D FFDM, 3D DBT and synthetic 2D images in a VCT set-up.

Computational models were used for all VCT components including the breasts, the microcalcifications, the imaging systems and the reviewers of the images. Breast texture phantoms of various densities were simulated using the 3D solid breast texture model introduced in Chapter 3. Round microcalcifications with various sizes and attenuation characteristics were used. Each phantom was duplicated to provide phantoms with and without microcalcification. Phantom images were simulated using a previously developed image acquisition simulator. The acquisition geometry corresponds to the nominal geometry of a commercial imaging system. Projection images were processed with commercial image processing and reconstruction algorithms. A lesion-known-statistically and location-known-exactly rating-scale detection task was designed. All images were scored by the ACO and three CHO models. A limited set of the images were scored by a human observer. The Laguerre-Gauss channel was used by the CHO models and a range of channel parameters was tested. The decision variables of ACO and CHO were analyzed using the receiver operating characteristic (ROC) methodology and the area under the ROC curve (AUC) was used as the figure-of-merit. Only the CHO with channel parameters giving the highest AUC were retained.

Experimental result was analyzed to first compare the performance of ACO and CHO. A preliminary comparison was also performed regarding the correlation of the performance of ACO and CHO to human. Then the microcalcification detection performance in FFDM, DBT and synthetic 2D images was compared for different microcalcification sizes and background types.

7.1 Materials and methods

In this section, we describe in detail the simulation of the test objects, the imaging acquisitions and the microcalcification detection task.

7.1.1 Simulation of test objects

Breast phantoms

Three types of breast phantoms with different anatomical backgrounds were simulated: phantoms with uniform background, phantoms with textured background representing the BI-RADS 2 density category and phantoms with textured background representing the BI-RADS 3 density category.

The 3D solid breast texture model described in Chapter 3 was used to generate the textured phantoms. Textured phantoms were generated as voxelized cubes with spatial size of 5 cm \times 5 cm \times 5 cm. The 5 cm thickness represents the average thickness of compressed breasts in clinical FFDM and DBT exams. Isotropic voxels with size 100 μm \times 100 μm \times 100 μm were used. The BI-RADS breast density 2 texture phantoms have average glandular density of 27%, with a standard deviation of 2%. BI-RADS breast density 3 texture phantoms have average glandular density of 55%, with a standard deviation of 2%. For each BI-RADS breast density category, 200 random realizations were generated using the corresponding set of parameters described in Table 3.1.

A uniform phantom of the same size as the textured phantoms and with 37% glandular density was employed. A thirty-seven percent density corresponds to the average density of the two textured phantoms. The phantom format was a surface mesh object. Figure 7.1 (top row) illustrates the simulated three types of breast phantoms.

Microcalcifications

Microcalcifications (μcalc) were simulated as surface mesh objects. Round-shaped microcalcifications were considered. Spheres were first simulated. Then, to add irregularities and randomness to the simulated μcalcs , the sphere surface of each μcalc was randomly deformed. The deformation consisted of two steps. First, the sphere surface was slightly deformed to a random ellipsoid along three randomly chosen axes, with maximum deformation magnitude set to five percent of the μcalcs diameter. Then the surface meshes were modified according to a stochastic Perlin noise to create irregularities. The Perlin noise deformation was done by locally displacing the vertices of a surface mesh in directions perpendicular to the mesh face. The displacement magnitude was set to 20 μm .

Microcalcifications with four diameters, $D = 100 \mu\text{m}$, $200 \mu\text{m}$, $400 \mu\text{m}$ and $600 \mu\text{m}$ were considered to account for a realistic range of μcalc sizes as seen in clinical practice. A bank of 50 μcalc realizations were generated for each diameter. Figure 7.1 (bottom row) illustrates one simulated μcalc per diameter.

Microcalcifications with x-ray attenuation properties corresponding to 20%, 40%, and 60% of the attenuation of aluminum (Al) at 22 keV and with the same volumetric mass density as Al, *i.e.* 2.72 kg m^{-3} , were designed. We denote the attenuation coefficients of

these three types of μ calcs as $\mu_{\text{Al}20}$, $\mu_{\text{Al}40}$ and $\mu_{\text{Al}60}$. The choice of 22 keV corresponded to the photon energy of the x-ray source used in our study. Al is less attenuating than the minerals composing real μ calcs which contain calcium carbonate, calcium oxalate or apatite. This is realistic since in real μ calcs, the minerals are embedded in a protein matrix [13]; therefore the attenuation of a μ calc is lower than when considering only the minerals.

To insure that the μ calc detection performance for different experimental conditions could be differentiated, the attenuation coefficients of the μ calcs were determined in an off-line preliminary detection experiment with a human observer. FFDM was used as the reference modality for the experiment. The experiment is described as follows. A BI-RADS breast density 2 background FFDM image was first simulated and was embedded with a 200 μ m μ calc with attenuation equal to 100% of the Al at 22 keV. The same simulation was then repeated, where each time the percentage value with respect to Al decreased for 10%. The observer was asked to rate the visibility of the μ calc in each simulated image using four scales: definitely visible, almost visible, almost not visible and definitely not visible. The percentage value 60% was chosen since at this value the observer first rated the μ calc as almost not visible. The percentage value 20% was chosen since at this value the observer first rated the μ calc as definitely not visible. The third percentage value 40% corresponds to the average of the first two values.

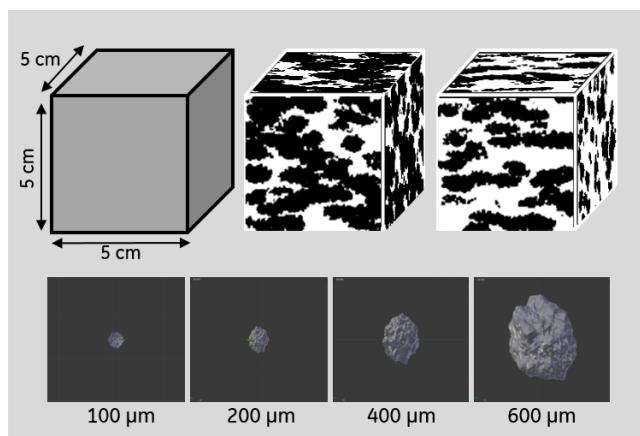


Figure 7.1: Top row: illustration of three types of phantoms used for this study. From left to right: a uniform background phantom with glandular density at 37 %; a textured phantom from the BI-RADS 2 density category and a textured phantom from the BI-RADS 3 density category. Bottom row: examples of simulated μ calcs with average diameter 100 μ m, 200 μ m, 400 μ m and 600 μ m.

7.1.2 Simulation of FFDM and DBT image acquisitions

FFDM and DBT image acquisitions were simulated using a previously described x-ray breast imaging simulator [132]. The topological and physical characteristics of the simulator were adjusted using the FFDM and DBT modes of the GE Senographe Pristina commercial 3D breast imaging system as reference.

System topology

Figure 7.2 illustrates the topological configuration of the simulated image acquisition system. Topology of the FFDM image acquisition was modeled as follows. The source-to-image distance (SID) was set to 660 mm. The detector size was set to 239.4 mm in the x -axis direction and 286 mm in the y -axis direction. Isotropic detector pixels with size of 0.1 mm were simulated. Position of the x-ray source was modeled such that its projection along the z axis onto the detector plane falls exactly at the midpoint of the detector's left edge. The bucky and the compression paddle were modeled as two planes parallel to the detector.

For the DBT image acquisition mode, the detector, the bucky and the compression paddle were modeled in the same way as for the FFDM mode. The source-to-rotation isocenter distance (SIsoD) was 620 mm. X-ray tube positions were modeled in a step-and-shoot fashion within an angular range of $(-12.5^\circ, 12.5^\circ)$, with a step size of 3.125° .

Breast phantoms were placed at the centered-left position of the detector panel, as illustrated in Figure 7.2. To reduce the processing time for projections, only a 100 mm \times 100 mm region of interest (ROI) at the centered-left position of the detector was considered. This resulted in images of size 1000 pixels \times 1000 pixels. We insured that the size of the detector ROI was set large enough to contain the entire phantom at all DBT projection angles.

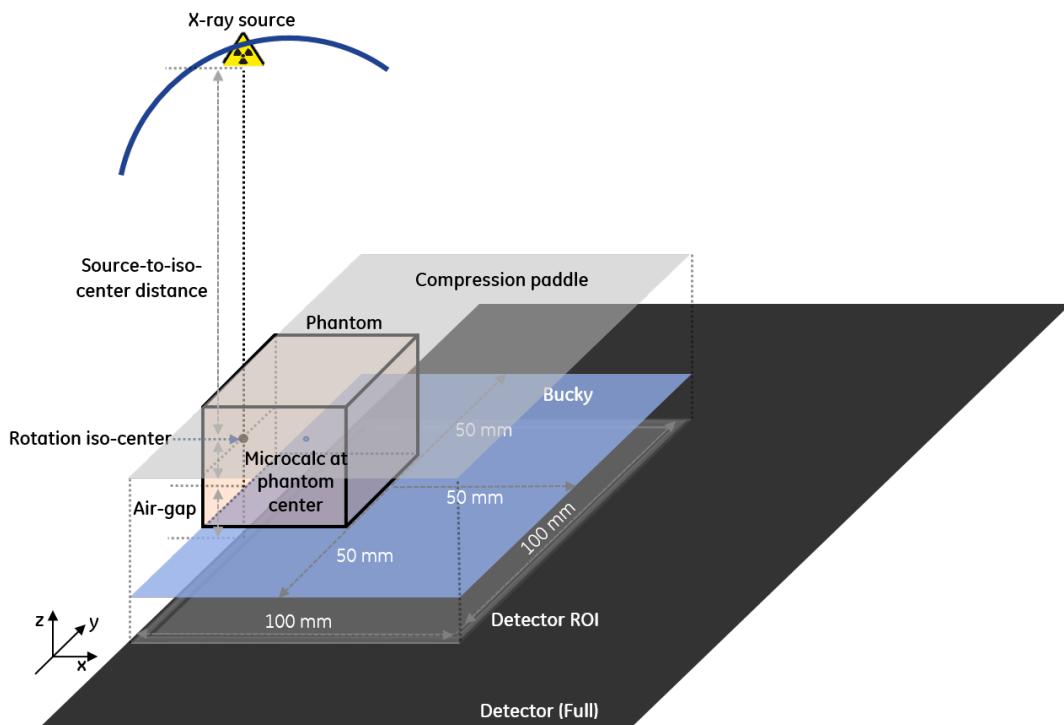


Figure 7.2: A schematic illustrating the simulated GE Senographe Pristina image acquisition topology.

Physical characteristics

Physical characteristics of the imaging system were modeled as follows. A simplified mono-energetic x-ray source was used. The photon energy level was set to 22 keV, this is equivalent to the effective x-ray energy of a 34 kVp Rhodium (Rh)/Silver (Ag) target/filter spectrum used for imaging a 50 mm compressed breast. An ideal point source focal-spot was used. In order to simulate the detector unsharpness, a modulation transfer function empirically measured from the reference system was used by the simulator. X-ray scatter from the test object was not considered. Only Poisson x-ray noise was added and the electronic noise was not modeled.

A calibration was performed to match the signal-to-noise ratio (SNR) in simulated images with the SNR obtained in experimentally acquired images using the GE Senographe Pristina system with automatic optimization of parameters (AOP). The calibration was repeated for FFDM and DBT acquisition modes. The adjustment method is described as follows. On the GE Senographe Pristina system, FFDM and DBT central raw projection images were acquired from two 50 mm thick phantoms (CIRS, Virginia, USA) with 100% adipose and 100% fibroglandular equivalent compositions. The phantoms were positioned side-by-side on the breast positioning table. Images were acquired at 34 kVp using a Rh/Ag target/filter combination and using AOP mode. SNR values were measured in the raw projection images (processed only with the manufacturer's gain, offset, and defective pixel correction), in 2 cm × 2 cm square ROIs at 4 cm from the chest wall. SNR was defined as $\frac{\langle SI \rangle}{\sigma_{SI}}$, where $\langle SI \rangle$ is the average detected signal intensity per pixel and σ_{SI} is the standard deviation in the signal intensity. The experiment was then simulated. The mAs value in the simulation was adjusted so as SNR values in the simulated and experimentally obtained images were found similar. Due to restrictions of the x-ray simulator, it was not possible to exactly match the SNR for both 100% adipose and 100% fibroglandular regions. The mAs in the simulations was therefore adjusted till SNR differed by maximum 10% in both breast-tissue equivalent.

Embedding of μ calcs in FFDM and DBT raw images

The voxel-based phantoms and surface mesh-based μ calcs use different projector implementations in our x-ray simulator. One way to obtain images with both background and μ calc in a single acquisition is to voxelize the surface mesh μ calc, and insert it directly into the voxel-based phantom. This will create a single voxel-based phantom containing both the background and the μ calc. However in doing so, there might be a loss of morphological accuracy of the μ calc during the voxelization process. The accuracy loss depends on the resolution of the voxelization. At very high resolution the loss might be insignificant; but the voxel-based phantoms must also be over-sampled at the same resolution. This can greatly increase the simulation complexity in terms of processing time and phantom storage space requirement.

Therefore, a hybrid simulation method was applied to embed a separately projected mesh-based μ calc into projection images of a voxel-based phantom. The method is similar to previously described methods [36] [176]. Detail of the method is described as follows:

1. Before imaging the test objects, an adjustment of the linear attenuation coefficient of the μ calc was performed. In fact, when the background phantom and the μ calc are

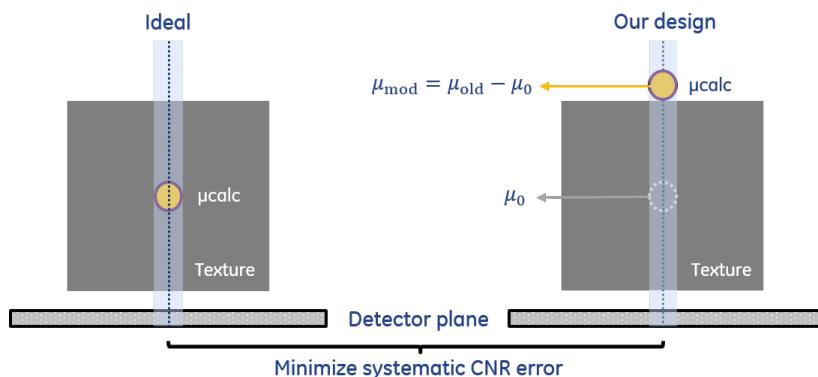
projected separately, an extra portion of breast tissue is imaged. This extra tissue consists of the voxels in the background phantom that fall inside the μ_{calc} 's surface mesh boundary, when the μ_{calc} is placed at the phantom's center. The addition of the extra tissue introduces a systematic bias to the local contrast-to-noise ratio (CNR) at the location of the μ_{calc} in the projection images. To minimize this bias, the extra tissue should be compensated. In our design, this compensation was done by adjusting the attenuation coefficient of the μ_{calc} . First, the extra voxels of the background phantom that fall inside the μ_{calc} 's surface mesh boundary were identified. Then the average attenuation coefficient of the identified voxels was computed. Finally the attenuation coefficient of the μ_{calc} was adjusted by subtracting the computed average. This step is illustrated in Figure 7.3 (top).

2. Next, using the previously adjusted image acquisition system, the voxelized phantom and the surface mesh μ_{calc} were projected separately, with their corresponding projector implementation. All projections were done without x-ray noise and with a 5×5 detector pixel over-sampling, to ensure that small μ_{calcs} are properly imaged. This step produces background-only and the μ_{calc} -only raw projection images.
3. A log transformation was then applied to all pixels of the background-only and the μ_{calc} -only raw projection images. The μ_{calc} -only images were first subtracted by a log-transformed airscan simulated using the same image acquisition system. This step has the objective to compensate the non-uniform intensity background of the μ_{calc} -only image, in order to only combine the pixel value related to the μ_{calc} to the background-only image. Then, the airscan subtracted μ_{calc} -only images and the background-only images were added pixel by pixel, to produce noiseless projection images with both background and μ_{calc} , as if they were imaged altogether, without x-ray noise. Lastly, an exponential operator was applied to all the pixels of the combined image, to transform the pixel values back to the same unit as the normal raw projections of the x-ray simulator.
4. In the final step, x-ray noise was added to the combined images, producing the final raw projection images with inserted μ_{calc} .

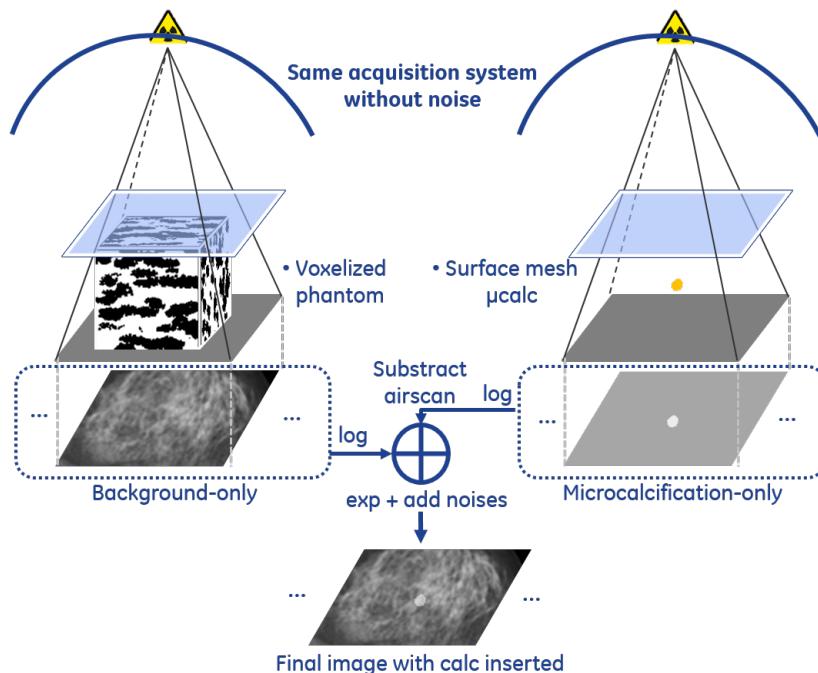
Figure 7.3 shows a schematic to illustrate the described μ_{calc} insertion method for the DBT mode. For each acquisition with μ_{calc} insert, one random μ_{calc} was chosen from the bank of 50 realizations in the same diameter category, and was embedded in the geometric center of the background texture phantom. For each background texture phantom, 12 different insert types were defined by combining the 4 different values for the μ_{calc} diameter and the 3 different values for the attenuation; Together with the phantom without μ_{calc} inserts, there were in total 13 different image acquisition conditions for each phantom in both FFDM and DBT modes. Since there were 600 different phantom realizations (200 BI-RADS breast density 2 phantoms plus 200 BI-RADS breast density 3 phantoms, and the uniform phantom was reused 200 times), there were in total 7800 image acquisitions.

7.1.3 Image processing

To produce the final mammograms, DBT reconstructed slices and S2D images, FFDM and DBT raw projections images were processed with commercially used image post-processing and 3D reconstruction algorithms on the GE Senographe Pristina reference



(a) Adjustment of μ_{calc} attenuation coefficient to minimize the systematic bias of the local CNR at the location of the μ_{calc} in the projection images. This bias is due to the extra voxels from the texture volume that are inside the μ_{calc} 's boundary when combining the separately imaged texture volume and μ_{calc} . To obtain the adjusted μ_{calc} attenuation coefficient μ_{mod} , the original μ_{calc} attenuation coefficient μ_{old} was subtracted by a value μ_0 which corresponds to the average attenuation coefficient of the extra voxels of the texture volume that are inside the μ_{calc} 's boundary when the μ_{calc} is placed at the texture volume's center.



(b) Combination of separate simulations of background and μ_{calc} .

Figure 7.3: Schematic illustrating the method to simulate raw projection images with embedded μ_{calc} .

system. Parameters of the algorithms were adjusted to represent what were typically used by the radiologists in clinical workflows.

FFDM raw projection images were processed by the PremiumView software (version 1.8.13, GE Healthcare, Buc, France) to produce the final FFDM images. The software contains

the **eContrast** and the **FineView** processing algorithms. The **eContrast** algorithm aims to increase the visibility of breast structures. By means of a frequential analysis, the image contrast in the fibroglandular tissue is enhanced while the visibility of the whole breast is preserved. The **FineView** algorithm aims to restore the blurring impact of the optical diffusion occurring in the CsI layer of the detector, thus providing an image closer to the detector input image. This is obtained by modeling this detector degradation and applying the inverse effect on the detected image by means of Wiener filter [186]. To obtain DBT reconstructed slices, the **ASIR-DBT** 3D reconstruction algorithm (version 1.3.4, GE Healthcare, Buc, France) was used. **ASIR-DBT** is an iterative reconstruction algorithm. It takes 9 DBT raw projection images as input, and produces 50 closely-spaced slices of 1 mm thickness as the output. The algorithm reduces image noise by iteratively comparing the acquired image with a modeled projection. The DBT reconstructed slices have the same image and pixel size as the input projection images. To produce the synthetic 2D (S2D) images, the **VPreview** software (version 3, GE Healthcare, Buc, France) was used. The software takes 9 DBT raw projection images as input, and produces a synthesized 2D FFDM-like image. Figure 7.4 shows examples of $2.5\text{ cm} \times 2.5\text{ cm}$ ROIs of simulated FFDM images, DBT central reconstructed slices and S2D images with μ_{calc} inserts, for the uniform, BI-RADS breast density 2 and BI-RADS breast density 3 background types. To simplify the illustration, only μ_{calcs} with diameter $D = 400\text{ }\mu\text{m}$ and attenuation $\mu_{\text{Al}60}$ are shown.

Some aspects during clinical image reviews on the review workstation were not considered in this study. When reviewing x-ray breast images, human observers typically apply zooms on the displayed images and an interpolation algorithm is usually applied which changes the visual appearance of the μ_{calcs} . This image interpolation was not modeled in this study. Also, for better image representation, the review station software usually applies a pixel intensity look-up table to the processed breast images, before displaying them on the screen. The look-up table is typically determined by a non-linear transfer function and was not modeled in this study.

7.1.4 Microcalcification detection experiment

Task

The μ_{calc} detection task was modeled as a *signal-known-statistically* and *location-known-exactly* rating-scale task. In each trial of the task, the observer knew the size/attenuation group of the μ_{calc} . The location of the μ_{calc} was exactly known to the observer.

Observers

Images were first score by the *a contrario* observer (ACO) introduced in Chapter 6 and the channelized Hotelling observer (CHO, Appendix B).

The ACO was configured using parameters described in Section 6.4.2. For DBT reconstructed slices, we applied the 3D DBT extension for ACO, described in Section 6.3.3. Since the μ_{calc} detection task is location-known-exactly, the methodology described in Section 6.3.4 was used to restrict the computation of ACO to a square region with side length $D \times 3$, centered at the known μ_{calc} center for each input image.

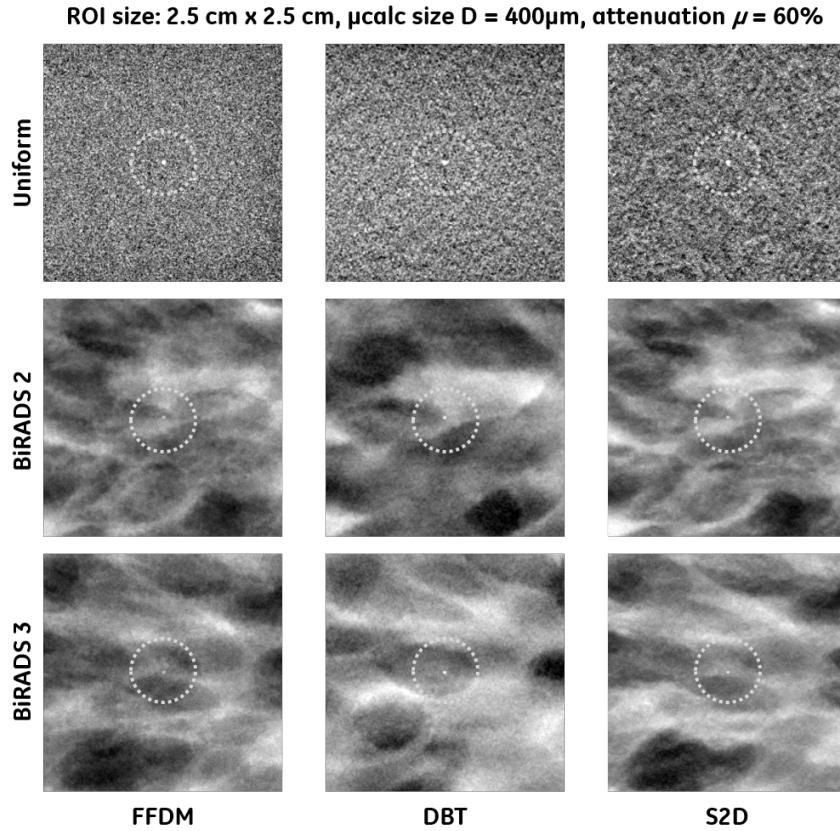


Figure 7.4: Examples of 2.5 cm × 2.5 cm ROIs of simulated processed FFDM images (first column), DBT central reconstructed slices (second column) and S2D images (third column) with μ_{calc} inserts, for uniform, BI-RADS breast density 2 and BI-RADS breast density 3 background. To simplify the illustration, only images containing μ_{calc} s with diameter $D = 400 \mu\text{m}$ and attenuation $\mu_{\text{Al}60}$ are shown.

The 2D CHO model (CHO2) was used to score 2D FFDM and S2D images. Two 3D extensions of the CHO model were considered for DBT reconstructed slices: the volumetric CHO (CHO3vol) model and the multi-slice CHO type A (CHO3msa) model (Appendix B). The Laguerre-Gauss channel was used by all three CHO models. A d -dimensional Laguerre-Gauss channel U is mathematically defined as [155]

$$U(r; p, a_u) = \frac{\sqrt{2}}{a_u} \exp\left(-\frac{2\pi r^2}{a_u^2}\right) L_p\left(\frac{2\pi r^2}{a_u^2}\right), \quad (7.1)$$

for $\forall r \in \mathbb{R}^d$. Here $a_u \in \mathbb{R}^+$ represents the spread of the Gaussian function and L_p denotes the Laguerre polynomials at degree $p \in \mathbb{N}$, defined as

$$L_p(x) = \sum_{k=0}^p (-1)^k \binom{p}{k} \frac{x^k}{k!}. \quad (7.2)$$

A range of values for a_u and p were considered for CHO2, CHO3msa and CHO3vol. The value for a_u was varied from 100 μm to 800 μm , with a step size of 100 μm . The value for p

was varied from 0 to 29 with a step size of 1. Figure 7.5 shows $5 \text{ mm} \times 5 \text{ mm}$ images (pixel size 0.1 mm) of Laguerre-Gauss channels with $a_u = 400 \mu\text{m}$, $600 \mu\text{m}$ and $p = 0, 1, 2, 3$.

Images were also score by a human observer who is very familiar with the appearance of clinical FFDM, DBT and S2D images.

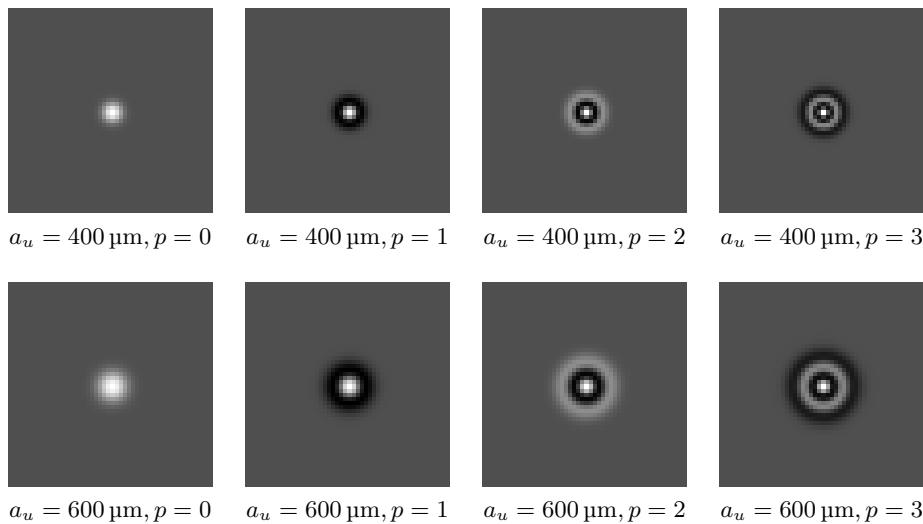


Figure 7.5: Laguerre-Gauss channels with $a_u = 400 \mu\text{m}$, $600 \mu\text{m}$ and $p = 0, 1, 2, 3$. The images have size $5 \text{ mm} \times 5 \text{ mm}$ with 0.1 mm isotropic pixels.

Image review

All images were cropped to $2.5 \text{ cm} \times 2.5 \text{ cm}$ ROIs such that the μ_{calc} was positioned in the center of the image when it was present. For each experimental condition there were 200 trials and for half of them the input images contained a μ_{calc} . The order of the trials was randomized for each experimental condition.

The model observers reviewed all experimental conditions. For each trial involving DBT reconstructed slices, the model observers reviewed only a subset of five consecutive DBT slices where the central slice passed through the μ_{calc} center. The five consecutive slices represent a breast thickness of 5 mm , which is thick enough to contain the largest μ_{calc} ($\approx 600 \mu\text{m}$). For each experimental condition, the CHO first reviewed all 200 trials for the training phase; Then it reviewed the same 200 trials in randomized order for the test phase. This choice is based on a trade-off between the fact that the image generation process is very slow (\approx three weeks to generate all experimental conditions on a machine with the Intel Xeon 5160 3.00GHz CPU and 16GB memory) and the fact that the statistical characteristics of a given background type are found fairly similar such as that computation of the template on a separate data set would not significantly impact the template. The model observers computed a continuous scalar decision variable λ after each trial during the test phase. The decision variable was considered to represent an infinite number of rating scales for the model observers.

The human observer only reviewed the experimental conditions for μ_{calcs} with diameter $D = 200 \mu\text{m}$ and attenuation μ_{Al40} . All three modalities were reviewed by the human

observer. Figure 7.6 shows an example of an FFDM image displayed during the human observer image review. For DBT reconstructed slices, the human observer reviewed only the central slice, where the μ_{calc} was in-focus when it was present. The image review by the human observer is described as follows. The review was carried out in a darkened room. Images were displayed on a 5M pixels gray-scale portrait monitor (SMD 21500 G, Siemens AG; Munchen, Germany) at 100% resolution. Zooming of the images was not allowed. The human observer maintained approximately a 40 cm distance to the display monitor. No time limitation was imposed for the human observer to make a decision. To reinforce the human observer's knowledge of the μ_{calc} location, a 5 mm diameter gray circle, centered on the possible microcalcification position (Figure 7.6) was added to each displayed image as a visual clue. For each experimental condition, the human observer had a short training session with 20 trials, where he/she was given the ground truth after each trial. Between two consecutive trials, a uniform gray-level image was displayed to minimize the visual influence of the previous image for the observer. During each trial, the human observer reported the confidence level of the μ_{calc} presence in five rating scales. The indications of the five scales are as follows. Scale 5 indicates that the image is very likely to be μ_{calc} -present; Scale 4 indicates that the image is likely to be μ_{calc} -present; Scale 3 indicates that the image is possibly μ_{calc} -present; Scale 2 indicates that the image is probably μ_{calc} -absent and scale 1 indicates that the image is very likely to contain no μ_{calc} .

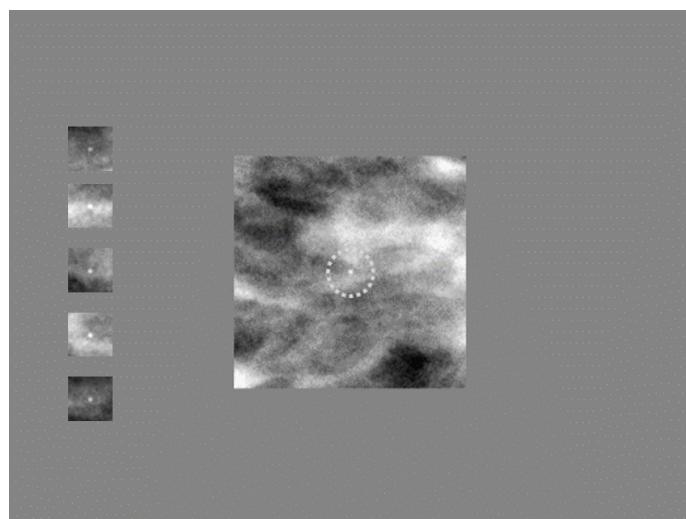


Figure 7.6: Example of an FFDM image displayed in the human observer experiment. The FFDM ROI is 2.5 cm x 2.5 cm. The ROI is surrounded by a uniform gray background with mean-luminance equal to the mean signal intensity value of the FFDM ROI. A 5 mm diameter circle is positioned at the center of the FFDM ROI (which contained in 50% of the trials a μ_{calc}) and aims to direct the visual attention of the observer to where a potential μ_{calc} may be. At the left, five μ_{calc} s of the same attenuation and size category were displayed aiming to indicate the μ_{calc} attenuation and size properties that had to be searched for. Here the ROI contains a μ_{calc} with diameter 600um and attenuation uAL60. FFDM, DBT and S2D images were presented in a similar way.

Figure-of-merit

The rating scales of the model and human observers were used as input to a receiver-operating characteristics (ROC) analysis, where the area under the ROC curve (AUC) was computed as the figure-of-merit.

The AUC computation for the model observers is described as follows. For CHO models, the AUC was estimated using the analytical equation

$$\text{AUC} = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \frac{d'}{2}, \quad (7.3)$$

where

$$d' = \frac{|\mathbb{E}(\lambda_1) - \mathbb{E}(\lambda_0)|}{\sqrt{\frac{1}{2} (\operatorname{Var}(\lambda_1) + \operatorname{Var}(\lambda_0))}}. \quad (7.4)$$

Here λ_1 and λ_0 denote respectively the decision variables from the μ_{calc} -present and μ_{calc} -absent images. Since the validity of (7.3) is based on the hypothesis that the both λ_1 and λ_0 are Gaussian distributed [194], we performed statistical normality tests for λ_0 and λ_1 for each experimental condition to check if the Gaussian hypothesis was met. The test results showed that in more than 90% of the cases the Gaussian hypothesis was not rejected. This justifies the use of equation (7.3) to compute the AUC for CHO.

Since the decision variable of the ACO represents the number of false alarms and is not Gaussian distributed, the analytical equation (7.3) cannot be used to compute the AUC for the ACO. We computed the AUC for the ACO using the empirical method described as follows. For a given experimental condition, we first thresholded the decision variables λ using 500 equally spaced threshold values varying from λ_{\min} to λ_{\max} . For each threshold, a trial was classified as μ_{calc} -absent if the λ value was less than or equal to the threshold and was classified as μ_{calc} -present otherwise. The classification outcomes were then compared to the ground truth to compute the truth positive fraction and false positive fraction. This allows us to obtain an empirical ROC curve with 501 operating points. Finally, the AUC was measured from the empirical ROC curve using the trapezoidal rule [194]. The trapezoidal rule consists of first approximating the region between each adjacent operating points under the empirical ROC curve by a trapezoid. Then the AUC is computed as the integration of the area of all trapezoids.

The same empirical method was used to compute the AUC computation for the human observer. For a given experimental condition, the human observer's rating scales were thresholded using threshold values 0, 1, 2, 3, 4 and 5. This allows us to obtain an empirical ROC curve with six operating points. The AUC was measured from the empirical ROC curve using the trapezoidal rule.

For both the model observer and the human observer, the mean and the standard deviation of the AUC were numerically computed using the bootstrap re-sampling method [177]. The bootstrap method randomly samples from the decision variables with replacement to obtain re-sampled sets of "new" decision variables. The population of each re-sampled set equals the original population. We repeated the bootstrap procedure 1000 times for each experimental condition. This way 1000 AUC values from the re-sampled decision variables were obtained, from which the mean and the standard deviation was computed.

7.2 Results

In this section, we present the result of the VCT experiment to compare the μ_{calc} detection performance of the ACO, the CHO and the human observer, as well as the μ_{calc} detection performance in FFDM, S2D and DBT images for different experimental conditions.

7.2.1 Evaluation of the optimal channel parameters

As mentioned earlier, the CHO models were evaluated with multiple Gaussian spread a_u and degree p parameters for the Laguerre-Gauss channel. We first demonstrate the impact of these different channel parameters on the performance of the CHO. Figure 7.7 shows AUC as a function of the a_u and p parameters of the Laguerre-Gauss channel for μ_{calc} with diameter $D = 200 \mu\text{m}$ and attenuation $\mu_{\text{Al}60}$. The average AUC values and errorbars, representing twice the standard deviation, are shown for all CHO models (CHO2, CHO3vol and CHO3msa) and modalities (FFDM, S2D and DBT). Figure 7.7 shows that the average AUC value initially increase rapidly as the channel degree p increases. Then the speed of the increase slows down when the degree p becomes greater than eight; it was found that from $p = 8$ to $p = 29$, the increase in the average AUC value was less than 5%. The increase might be explained by the intrinsic channel mechanism of the CHO. When p increases, more channels are added to the CHO. This means that CHO has more information to discriminate μ_{calc} -present images from μ_{calc} -absent images. This results in an estimated CHO template with more discriminative power between μ_{calc} -present and μ_{calc} -absent images, hence a higher AUC value. Figure 7.7 also shows that the AUC value of the CHO increases faster when the value of a_u is small.

For all experimental conditions, the a_u and p values of the Laguerre-Gauss channels resulting in the maximum AUC values can be found in Appendix C. For all experimental conditions, we observed similar trends showing that the AUC value increases when the degree p of the Laguerre-Gauss increases. Regarding the a_u parameter, we observed that the correlation between the magnitude of a_u and the speed of the increase of the AUC was different for different background types, modalities, μ_{calc} sizes and μ_{calc} attenuations. Further exploration is needed to understand this observation. In the remaining Section 7.2.2 and Section 7.2.3, only the maximum AUC obtained from all combinations of p and a_u for each experimental condition were retained. These maximum AUC values are hereafter denoted as AUC_{\max} .

7.2.2 Comparison of the performance of the ACO, the CHO and the human observer

Figure 7.8 compares the μ_{calc} detection performance of the ACO and the CHO. The performance of the ACO was compared separately with different CHO models (CHO2, CHO3vol and CHO3msa) for different modalities (FFDM, S2D and DBT). Figure 7.8 shows that the CHO models outperform the ACO for all experimental conditions. No clear negative or positive correlation was found between the ACO and the CHO models. For DBT, CHO3msa outperforms CHO3vol for all experimental conditions.

Figure 7.9 compares the performance of the ACO and the CHO with the performance of the human observer. All background types and modalities were considered. Only μ_{calcs}

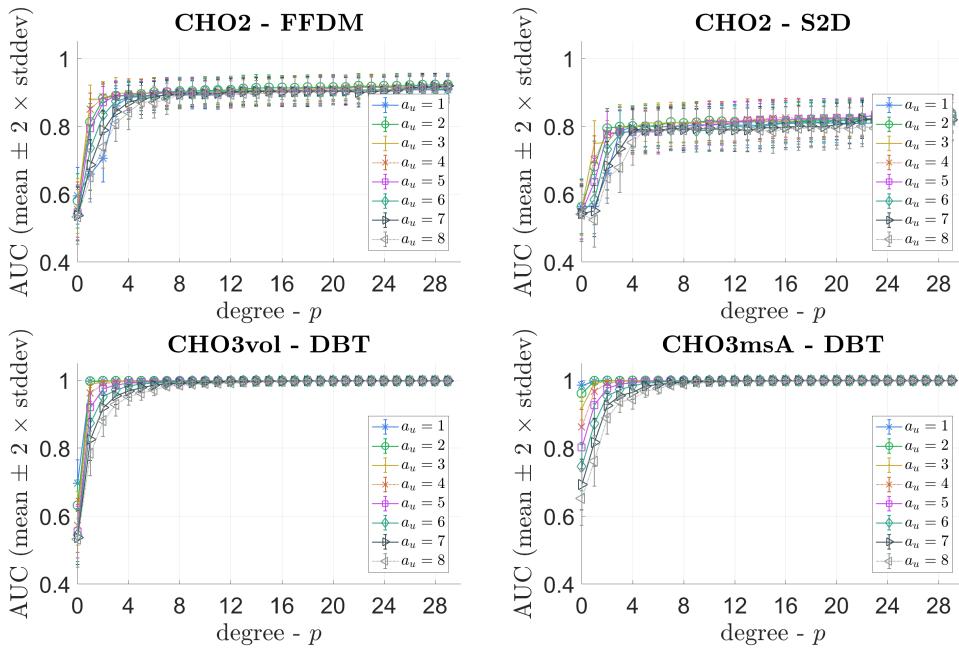


Figure 7.7: Evaluation of the AUC (mean \pm two standard deviations) for different values of the Gaussian spread a_u and the degree p of Laguerre-Gauss channels. Results are shown for all CHO models (CHO2, CHO3vol and CHO3msa) and modalities (FFDM, S2D and DBT). Only the case of μ_{calc} with diameter $D = 200 \mu\text{m}$ and attenuation $\mu_{\text{Al}60}$ is shown. For all a_u values, the average AUC first increases rapidly as the channel degree p increases; Then the speed of the increase slows down when the degree p is greater than eight.

with diameter $D = 200 \mu\text{m}$ and attenuation $\mu_{\text{Al}40}$ were considered since these were the only experimental conditions reviewed by the human observer. We observe that the CHO models outperform the human observer, except for one condition with FFDM images and BI-RADS breast density 3 background texture. On the contrary, the ACO underperforms the human observer. For uniform and BI-RADS breast density 2 backgrounds, both the ACO and the CHO are positively correlated with human observer performance. However, it is hard to make solid conclusions on the correlation between the performances of the human observer and the CHO due to the magnitude of the AUC_{\max} values. The AUC_{\max} for CHO in DBT and FFDM images are close to one. As a result, there is little room to demonstrate a performance difference. The AUC values for the ACO and the human observer on the other hand are smaller and thus there is more room to show a performance difference. For the BI-RADS breast density 3 background, the ACO is positively correlated with the human observer, but this is not the case for the CHO.

7.2.3 Comparison of μ_{calc} detection performance in FFDM, DBT and S2D images

Figure 7.10 and Figure 7.11 show the comparison of the μ_{calc} detection performance in FFDM, S2D and DBT images for the ACO and the CHO respectively. The comparisons were performed for all μ_{calc} diameters, attenuations and background types. For a given background type and μ_{calc} size, the detection performance increases for both the ACO

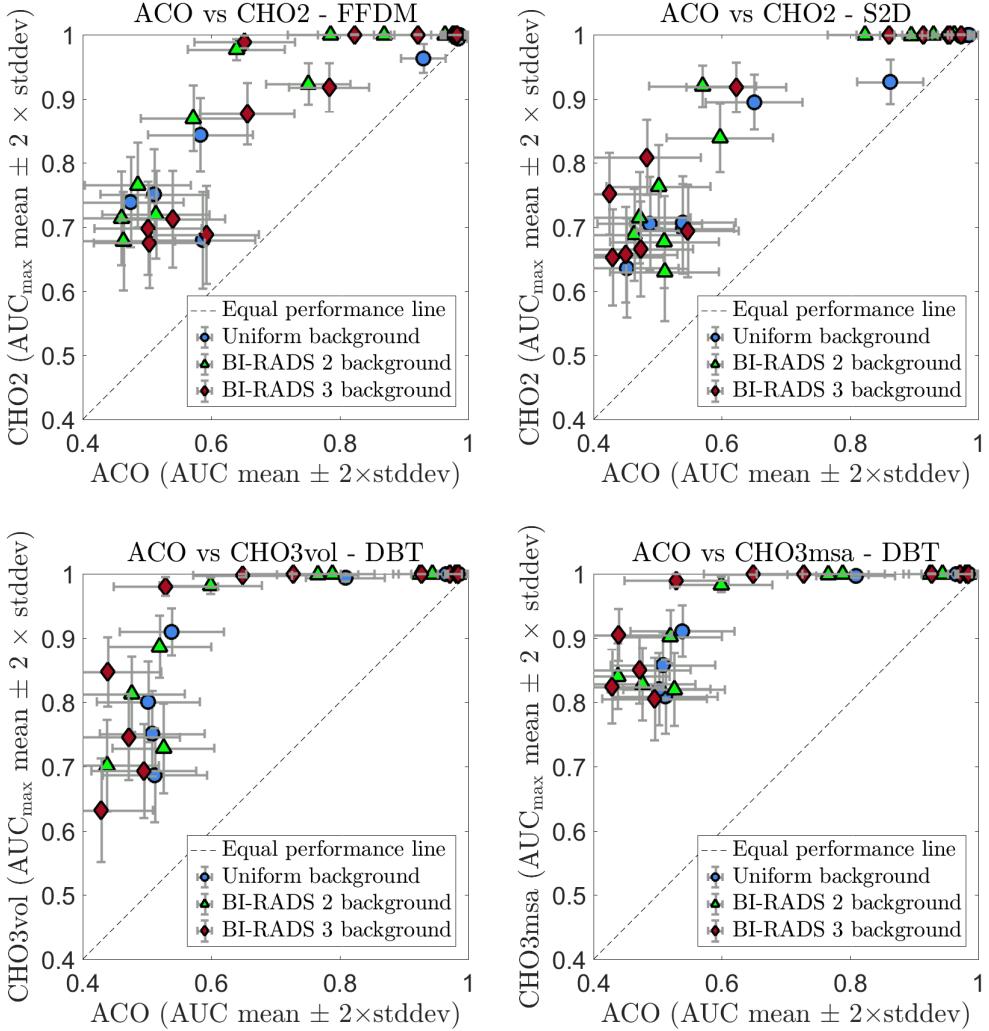


Figure 7.8: Comparison of μ_{calc} detection performance of the ACO and the CHO2 in FFDM images (upper-left), the CHO2 in S2D images (upper-right), the CHO3vol in DBT slices (lower-left) and the CHO3msa in DBT slices (lower-right). All μ_{calc} diameters, attenuations and background texture types were considered. Different background types are shown with different point markers. The CHO models outperform the ACO in all cases.

and the CHO when the μ_{calc} is more attenuating. For a given background type and attenuation, the detection performance increases for both the ACO and the CHO when the μ_{calc} size increases.

Regarding the ACO, for all three background types, the AUC of the ACO is approximately 0.5 for 100 μm diameter μ_{calcs} for all three modalities, meaning that there is no discriminative power between the three modalities for 100 μm diameter μ_{calcs} . For uniform background, the AUC of the ACO is approximately equal to 1 for 400 μm and 600 μm diameter μ_{calcs} and for all three modalities, meaning that there is no discriminative power between the three modalities for 400 μm and 600 μm diameter μ_{calcs} . The ACO does not

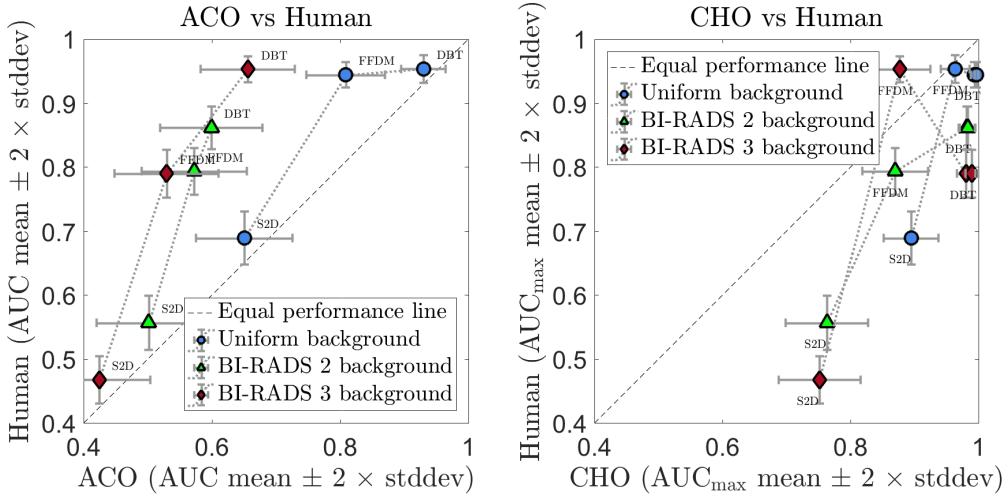


Figure 7.9: Comparison of the performance of the ACO (left) and the CHO (right) with the performance of the human observer. All background types and modalities were considered. Only μ calcs with diameter $D = 200 \mu\text{m}$ and attenuation $\mu_{\text{Al}40}$ were considered.

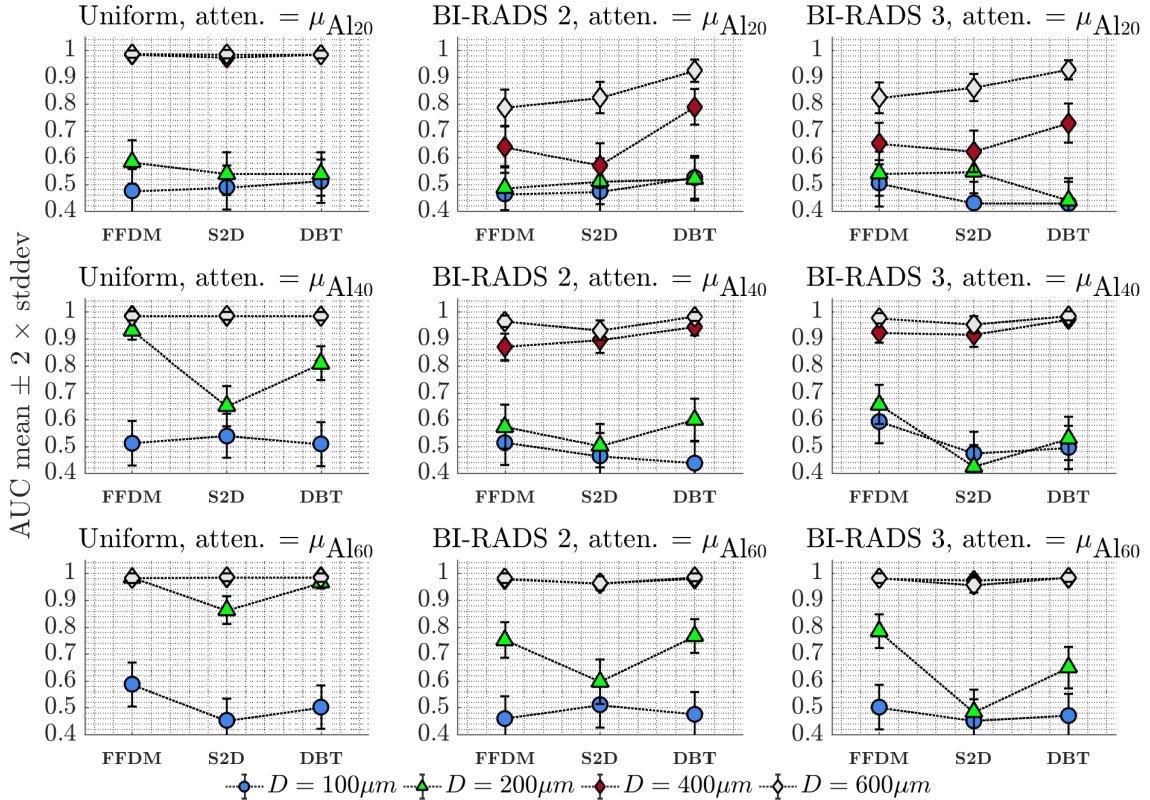


Figure 7.10: Comparison of μ calc detection performance in FFDM, S2D and DBT images for the ACO. All μ calc diameters, attenuations and background types were considered.

report a consistent AUC ranking between FFDM and DBT. Figure 7.10 indicates that for μ calcs with diameter equal to $200 \mu\text{m}$, there is a trend that the AUC in FFDM is either

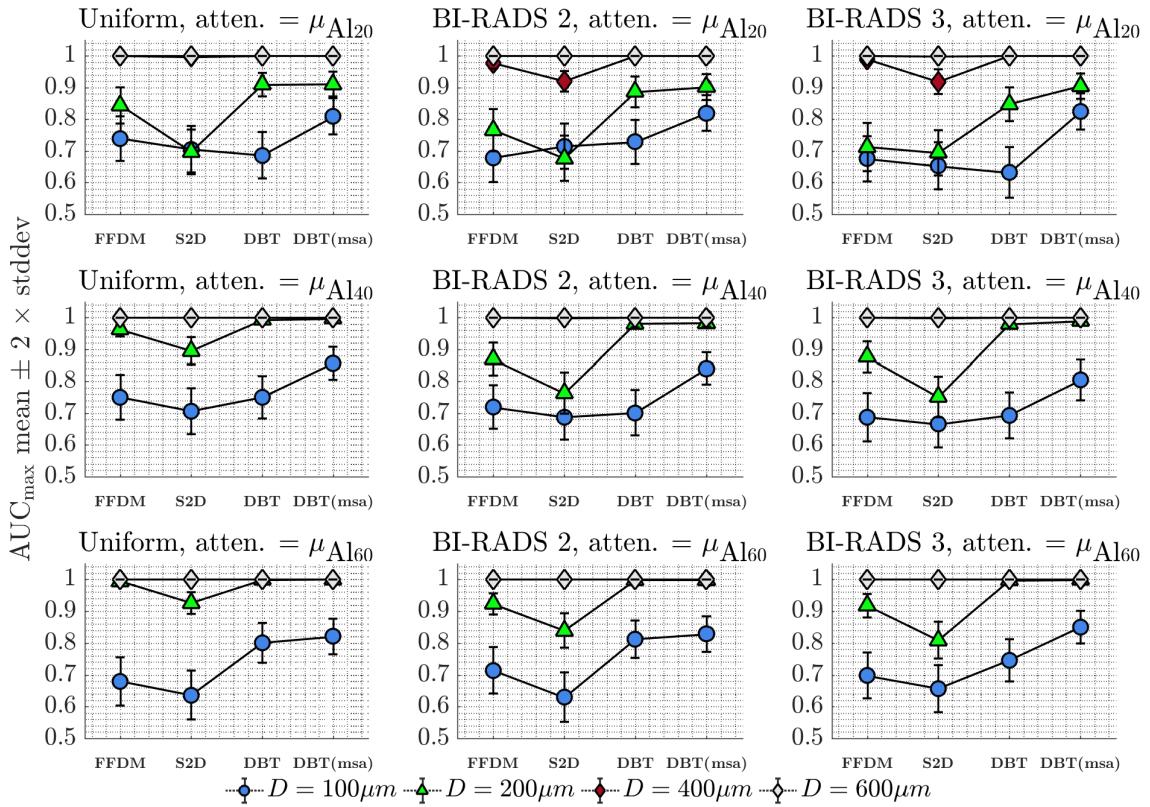


Figure 7.11: Comparison of μ_{calc} detection performance in FFDM, S2D and DBT images for the CHO. In each sub-figure, DBT on the abscissa refers to the result from the CHO3vol model and DBT(msa) refers to the result from the CHO3msa model. All μ_{calc} diameters, attenuations and background types were considered.

higher than or similar to the AUC in DBT. For μ_{calcs} with diameter equal to 400 μm , there is a trend that the AUC in FFDM is smaller than the AUC in DBT for BI-RADS breast density 2 and BI-RADS breast density 3 background types. For 200 μm and 400 μm diameter μ_{calcs} , there is a trend that the AUC in S2D is smaller than or equal to the the AUC in FFDM and DBT. For 200 μm diameter μ_{calcs} and for all μ_{calc} attenuations, there is a trend that the μ_{calc} detection performance is higher in uniform background compared to textured backgrounds. This observation does not hold for the other μ_{calc} sizes.

Regarding the CHO, for uniform background, the AUC_{max} of the CHO is approximately equal to 1 for 400 μm and 600 μm diameter μ_{calcs} , meaning that there is no discriminative power between the three modalities for 400 μm and 600 μm diameter μ_{calcs} . For BI-RADS breast density 2 and BI-RADS breast density 3 backgrounds and for 200 μm diameter μ_{calcs} , the CHO reports a rather consistent AUC_{max} ranking between the three modalities. For all μ_{calc} attenuations, the ranking is DBT (scored by CHO3msa) \geq DBT (scored by CHO3vol) $>$ FFDM $>$ S2D. For uniform background, the same ranking can not be confirmed since the AUC_{max} is close to one for FFDM and DBT. For 100 μm diameter μ_{calcs} , the CHO does not report a consistent performance ranking for FFDM and DBT. In all cases, the AUC_{max} in DBT scored by the CHO3msa is higher than the AUC_{max} in FFDM. For 200 μm diameter μ_{calcs} and for all μ_{calc} attenuations, there is a trend that the μ_{calc} detection performance is higher in uniform background compared to textured

backgrounds. This observation does not hold for the other μ_{calc} sizes.

7.3 Conclusion and discussion

In this chapter we implemented a virtual clinical trial (VCT) experiment using previously described VCT components to assess microcalcification (μ_{calc}) detection performance in FFDM, DBT and S2D images. The performance of the *a contrario* observer (ACO) proposed in Chapter 6 was compared with state-of-the-art channelized Hotelling observers (CHO). A preliminary comparison between the performances of the ACO and the CHO models and the performance of a human observer was also conducted. Different μ_{calc} sizes and attenuations as well as different background types were considered.

Our study found that for all experimental conditions, the CHO models outperformed the ACO but no clear negative or positive correlation was found between the two types of model observers. The preliminary comparison to the human observer performance showed that both the ACO and the CHO tend to be positively correlated with human observer performance. For both ACO and CHO, the detection performance increases with μ_{calc} size at a given μ_{calc} attenuation characteristic. For both ACO and CHO, the detection performance increases with μ_{calc} attenuation characteristics at a given μ_{calc} size. These results are consistent for all background types. Under this ideal topology, DBT tends to outperform FFDM and S2D for the CHO models. The ACO results do not allow for a clear ranking of the different modalities.

Several of our observations are however not yet interpreted in a fully satisfactory manner. Further exploration of the CHO and the ACO models and their relation to human observer performance will be needed to get a more profound insight and draw more solid conclusions. For that, we believe that we will need to consider μ_{calcs} with sizes and attenuations that allow to obtain CHO and ACO performances with more discriminative power. Microcalcifications with diameters between 150 μm and 350 μm and with similar or even less attenuating characteristics as the μ_{calcs} considered in this study might be a good choice. We noticed that the CHO3msa outperformed the CHO3vol for both uniform and textured backgrounds. The efficiency with which the CHO3msa and the CHO3vol exploit information about the μ_{calc} contrast and size, as well as about the spatial correlation of the background in the depth direction needs to be further investigated. A more profound analysis is required to understand the correlation between human and respectively the ACO and the CHO models. Such an in-depth investigation will be needed to allow further tuning of the design of the respective model observers to better predict human performance. We believe that a more refined analysis will also allow for a better ranking of the various imaging modalities regarding μ_{calc} detection performance.

Our study also has design limitations that need further exploration. First, for the CHO models, we used the same image data sets for training and test phases. This choice is based on a trade-off between the fact that the image generation process is very slow (\approx three weeks to generate all experimental conditions on a machine with the Intel Xeon 5160 3.00GHz CPU and 16GB memory) and the fact that the statistical characteristics of a given background type are found fairly similar such as that computation of the template on a separate data set would not significantly impact the template. However, in a next step, the latter hypothesis needs to be tested. Second, for the CHO models, the covariance

matrix was estimated using 400 images (200 with μ_{calc} and 200 without μ_{calc}). Further analysis is required to understand whether this finite sample size adequately represents the true population of each background type and allows for a stable estimation of the covariance matrix. Finally, our VCT experiment was designed using an ideal imaging system topology. We also only considered μ_{calcs} with round shape and encountered the ideal case where the μ_{calcs} appeared in-focus in a DBT reconstructed slice. However in clinical reality, the morphological variability of the μ_{calcs} is much larger and the μ_{calcs} might not be in-focus in any slice of the reconstructed data set. Further investigation is required to understand the impact of these conditions on μ_{calc} detection performance. To simulate x-ray images, the x-ray tube focus was considered to be a point source and only primary x rays and quantum noise were modeled. The impact of x-ray scatter and electronic noise were not studied. Further evaluation and validation on a real imaging system are worthwhile to explore.

Despite the limitations of this study, this VCT experiment demonstrates the potential of using our proposed 3D solid breast texture model and *a contrario* observer for the implementation of more elaborated VCT experiments in the future.

Chapter 8

Conclusion and directions for future research

Full field digital mammography (FFDM), a 2D x-ray breast imaging modality has been proved to reduce the breast cancer mortality. Today, digital breast tomosynthesis (DBT), a 3D x-ray breast imaging modality, is being integrated in clinical practice and is believed to replace standard FFDM in the near future. Although DBT is based on an FFDM imaging platform, the image acquisition architecture, image processing and image reviewing conditions of DBT and FFDM are very different. To assess the clinical performance of various aspects unique to DBT, clinical trials are needed. However, clinical trials are burdensome, expensive and may increase risk to the patient, with no direct benefit, due to additional radiation exposure. Virtual clinical trials (VCT) aim to offer a more cost-effective and faster alternative by using computational components. Today, active research is ongoing to develop VCT components dedicated to the specific aspects of DBT so as to be able to investigate the impact of hardware, image processing and review on overall image quality. In particular, a mathematically tractable 3D breast texture model allowing to simulate realistic 2D and 3D images, and a mathematical observer well suited to DBT are needed.

The purpose of this dissertation has been to extend the development of objective task-based VCT components to optimize 3D imaging systems and to assess its performance versus 2D FFDM. Our first main objective has been to develop and validate a mathematically tractable 3D breast texture model allowing the simulation of realistic 3D and 2D breast x-ray images representative of the women population seen in clinical practice. Our second main objective has been to develop and validate a new 3D mathematical model observer based on the *a contrario* theory allowing to model microcalcification detection in DBT. Finally, our third main objective has been to apply the developed VCT components to compare microcalcification detection performance in FFDM, DBT and synthetic 2D mammography (S2D) images processed with state-of-the-art 3D reconstruction and processing algorithms. A side objective has been to study the analytical characterization of state-of-the-art random field breast texture models and to analyze common first and second order statistical estimators for breast x-ray image characterization.

Main results and contributions

We first investigated the analytical characterization of the 3D power-law Gaussian random field and the 3D clustered lumpy background. These two models are state-of-the-art random field breast texture models for the simulation of 2D and 3D x-ray breast images. Although these random field breast texture models result in images with limited visual realism, the fact that some of their statistical properties match those of real breast textures may still make them reasonable candidates to better understand the performance of various imaging modalities for particular visual tasks. Understanding the theoretical properties of 3D power-law Gaussian random field and the 3D clustered lumpy background, as well

as the analytical characterization of the 2D and 3D breast images simulated from these models may help to progress designs of more advanced 3D mathematical breast texture models. While the analytical characterization of the 3D power-law Gaussian random field has been studied in literature [126], little research has been done to characterize the 3D clustered lumpy background [111]. Using a 3D discrete random field formulation, we theoretically proved that 2D projections and slices of both the 3D power-law Gaussian random field and the 3D clustered lumpy background belong to the same class as the 3D models and can be analytically characterized.

To develop a model that shares the advantages of random field texture models, characterized by their mathematical tractability, and anthropomorphic breast phantoms, characterized by improved visual realism, we proposed a novel 3D solid breast texture model that is both mathematically tractable and capable of simulating realistic 2D and 3D breast images with fairly large morphological variability. A large texture variability is of interest in VCT studies, where the impact of various breast tissue types on clinical task performance often needs to be understood. The proposed 3D breast texture model is based on an analysis of segmented clinical breast computerized tomography (bCT) images and depicts the small and medium scale fibroglandular and inter-glandular breast tissue. The model uses elements from the stochastic geometry theory. In a first phase, model parameters were based on empirical observations. Visual realism of simulated DBT reconstructed slices of this first 3D breast texture model was found to be fair to high, based on the results of a psycho-physical experiment. In a second phase, to add more morphological variability to simulated breast textures, we implemented a two-step statistical inference approach referred to as the simulation-based inference to automatically infer the medium scale 3D texture model parameters from clinical bCT reconstructed volumes. Using statistical analysis, we demonstrated that the automatically estimated model parameters reflect the underlying distribution and morphology of the medium scale fibroglandular and inter-glandular breast tissue in the clinical bCT volumes. The 3D breast texture model resulting from this effort gave simulated DBT reconstructed slices with fair to high visual realism. Moreover, this model allows for a larger morphological variability of breast textures. We believe we introduced for the first time principles of stochastic geometry and simulation-based inference for parametric estimation of a complex marked point process to develop a breast texture model. The results obtained with these techniques are very satisfactory and we believe that they have great potential for further refining our 3D breast texture model.

Next, we investigated common statistical characteristics of clinical 2D and 3D breast images. These statistical characteristics have been shown to provide a way to understand, compare, or even quantify the difference of lesion detection performance in different types of breast images. Breast density, a first order statistical characteristic of breast images was investigated first. An alternative statistical estimator for volumetric breast density using DBT projections was proposed. It is based on the approach of model-based stereology and it assumes a simplified modeling of the breast volume and the DBT image acquisition. A preliminary evaluation has shown that the absolute bias and the standard deviation of the proposed estimator differ by less than 3% of the ground truth. To the best of our knowledge, this is the first work that shows the application of the model-based stereology to the estimation of the volumetric breast density. The result of our proposed density estimator encourages further investigation and optimization of this first stereology-based approach for breast density quantification. The power spectral index (β), a second order

statistical characteristic was investigated next. We investigated the accuracy of its statistical estimation using the most commonly applied algorithm based on the Welch method. Analytical β estimation bias and variance were obtained in function of different algorithm parameterizations. The analytical results were also validated by numerical experiments using synthetic FFDM images, DBT reconstructed slices and bCT reconstructed slices with known β values. Our results showed that the estimation of β based on the Welch method has limitations, mainly due to its non-negligible systematical estimation error, which is smaller for 3D breast images than for 2D images. We believe that our study of the β metric is the first theoretical derivation of the β estimation bias and variance inherent to the most commonly used Welch method. Our result emphasizes the importance to carefully design the parametrization for β estimation, especially when β is used as a metric to validate the statistical realism of simulated textures or to compare the performance of different breast x-ray imaging modalities or different image processing and reconstruction algorithms.

To develop a new mathematical observer dedicated to DBT, we proposed a new model observer based on the *a contrario* theory for the detection of microcalcifications in 2D and 3D breast images. Our study is encouraged by the fact that a 2D *a contrario* observer, introduced by Grosjean *et al.* in 2007 for mass detection in FFDM images, has been demonstrated to produce results fairly correlated to human. Our proposed *a contrario* observer is an extension to the 2D *a contrario* observer. The *a contrario* observer developed by Grosjean *et al.* uses global image statistics to compute the decision variable. This might not reflect the true strategy used by human observers for the detection of microcalcifications since several previous studies demonstrated that human observers use local image content for the detection of microcalcifications. To encounter this aspect, the proposed *a contrario* observer uses localized image statistics that are empirically estimated from the input images. The proposed *a contrario* observer allows for a global control of the average false positive detections by a pre-defined parameter. This property was demonstrated by a theoretical justification and also by simulation experiments. The proposed *a contrario* observer differs from state-of-the-art linear observers which necessitate exact or statistical knowledge of the signal and/or the image background. The proposed *a contrario* observer assumes minimum knowledge of the signal and relies on a simplified approximate background model to compute its decision variable. Except for the microcalcification size, no other a priori assumptions of the microcalcification were encountered. Our contribution of a new *a contrario* observer extended the *a contrario* framework to 3D breast imaging and further demonstrated the potential of using this novel holistic approach of human vision to model various types of clinical detection tasks.

Finally, we implemented a VCT experiment using the proposed 3D breast texture model and the *a contrario* observer to assess microcalcification detection performance in FFDM, DBT and S2D images. Such an assessment has an important clinical value since the microcalcification detection performance in DBT and S2D compared with FFDM represents crucial aspects in the development of DBT that have not been fully understood yet. The microcalcification detection performance of the proposed *a contrario* observer, state-of-the art channelized Hotelling observers and human observers were compared. Different microcalcification sizes and attenuations as well as different background types were considered. Our study found that for all experimental conditions, the channelized Hotelling observers outperformed the *a contrario* observer but no clear negative or positive correlation was found. The preliminary comparison to the human observer performance showed that both

the *a contrario* observer and the channelized Hotelling observers tend to be positively correlated with human observer performance. Some intuitively obvious results were found; with both model observers, microcalcification detection performance increases when microcalcification size and attenuation characteristics increase. These results are consistent for all background types. Under an ideal system topology, DBT tends to outperform FFDM and S2D for the channelized Hotelling observers. However, the *a contrario* observer does not allow for a clear ranking of the different imaging modalities. Our result for the channelized Hotelling observer performance is consistent with a previously published VCT study [90] and indicates that efforts are needed for the development of more efficient algorithms for the generation of S2D images to further improve its clinical diagnostic value. This application demonstrated the possibility to implement more elaborated future VCT studies using our developed new components.

Directions for future research

In this dissertation, we demonstrated the use of our proposed 3D breast texture model and *a contrario* model observer for the design of a complete VCT study. Our VCT study considered the detection of a single round-shape microcalcification with known location in 2D and 3D breast images under an ideal imaging system topology. Medium and small scale breast fibroglandular and inter-glandular adipose tissue representing different breast density types were simulated. Small images with size $2.5\text{ cm} \times 2.5\text{ cm}$ were considered. This VCT visual task is still far away from the clinical tasks that radiologists are confronted with in breast cancer screening.

To design VCT studies that can answer clinically relevant questions, the set-up of our VCT study needs to be further optimized and extended. Our 3D breast texture model can be further refined to allow for a more realistic modeling of the morphology and distribution of the fibroglandular and inter-glandular adipose breast tissue seen in clinical practices. To include local or global microcalcification search in the detection task, there is a need to extend our local 3D breast texture model to a model of the entire breast.

To optimize our 3D breast texture model for more refined modeling of the morphology and distribution of the fibroglandular and inter-glandular adipose breast tissue at various scales, the following aspects need to be further investigated. First, the current model does not encounter the correlation between the ellipsoid shape parameters and the ellipsoid centers. However, from observations of clinical bCT images, we believe that it might be more realistic to include this correlation in the 3D texture model. Such a correlation needs to be further characterized using the ellipsoids reconstructed from clinical bCT volumes. Secondly, by applying the inference from reconstruction method, we obtained twelve different sets of parameters for the 3D breast texture model. It might be worthwhile to express the variation between these twelve sets of model parameters as a function of the glandular density of the ground truth bCT volumes. This might help to create a single model that can continuously change between the twelve sets of model parameters. Finally, the current model can be extended to model breast fibroglandular and inter-glandular adipose tissue at a larger scale to encounter the morphology and distribution of breast tissue at various positions inside the breast. All these optimizations and extensions of the proposed 3D breast texture model might enable the simulation of 2D and 3D breast images with further improved visual realism and with statistical properties closer

to clinical images. First, the current model does not encounter the correlation between the ellipsoid shape parameters and the ellipsoid center position inside the breast. However, from observations of clinical bCT images, we believe that it might be more realistic to include this correlation in the 3D texture model. Such a correlation need to be further characterized and quantified using the ellipsoids reconstructed from clinical bCT volumes. Second, the inference from reconstruction method yielded twelve separate sets of texture model parameters from twelve different ground truth bCT volumes. The relationship between these twelve sets of parameters was not studied. It might be worthwhile to express the variation between these twelve sets of model parameters as a function of the glandular density of the ground truth bCT volume. This might help to create a single model that can continuously change between the twelve sets of model parameters when the density changes.

From a theoretical perspective, although we demonstrated that the method of inference from reconstruction can allow for the inference of complex marked point process models, this method needs further analysis and optimizations. The statistical properties such as the bias and the variance of this inference method were not studied. As a future step, we should consider running the simulation from reconstruction on synthetic data sets from a Poisson marked point process with known parameters and quantify the bias and variance of the estimated parameters. In the inference step we applied the minimum contrast estimation method using the pair correlation function to fit the parameter of the Matérn cluster process to the reconstructed ellipsoid centers. A possible optimization is to include more summary statistics such as the Ripley's K function and the spherical contact distribution function into the minimum contrast estimator. This might help to encompass more statistical characteristics of the reconstructed ellipsoid centers and make the minimum contrast estimator more accurate.

The proposed *a contrario* observer could be optimized in various ways to better model the way human observers review DBT images. First, the current *a contrario* observer does not model the correlations between DBT reconstructed slices. The correlation between DBT slices may become important in case there are microcalcification reconstruction artifacts in adjacent slices (*i.e.* when a reconstruction algorithm without artifact management is used). This could be achieved by considering several neighboring slices when computing the local image statistics used to derive the number of false alarms. Secondly, the current *a contrario* observer does not encounter the viewing distance and the scrolling speed of the human observer in a stack of images. These aspects could be incorporated by first pre-processing the stack of images with a spatio-temporal contrast-sensitivity function, then running the *a contrario* observer on the pre-processed images. Avanaki *et al.* have previously demonstrated that the use of the spatio-temporal contrast-sensitivity function allowed to obtain a model observer detection performance that better matches human observer performance when testing different viewing distances and browsing speeds [5]. Finally, the current task of the *a contrario* observer could be further extended to the detection of a microcalcification cluster instead of detecting an isolated microcalcification. In a first step, it might be worthwhile to explore the capacity of the *a contrario* observer to independently compute decision variables for various image locations. This can be considered as a search mechanism. However, the assumption of independent microcalcification detection is not realistic since the ability to detect a cluster heavily depends on its characteristics (size, shape and number of microcalcifications). Michielsen *et al.* [131] have shown that when microcalcifications appear in a cluster, the visibility of a very small

individual microcalcification is reinforced by surrounding microcalcifications. By incorporating a strategy of microcalcification interaction in a multi-slice channelized Hotelling observer designed for microcalcification detection, Michielsen et al. [131] has shown that correlation between model observer and human observer increases. It might be a good choice to test and extend the design of Michielsen et al. for our proposed *a contrario* observer.

The theoretical false positive control of the proposed *a contrario* observer needs further exploration. We demonstrated that the proposed *a contrario* observer based on a white Gaussian noise naïve model allows to globally control the number of false positive detections. Our simulation experiment showed that the proposed *a contrario* observer using a white Gaussian noise as the naïve model also allows to control the number of false positive detections in FFDM-like images. This result suggests that an extension of the white Gaussian noise naïve model to more complex models might be possible. Therefore, a rigorous theoretical proof should be conducted to further investigate the extension and to interpret the experimental result. One possible starting point is to derive the theoretical average false positives of the proposed *a contrario* observer using the power-law Gaussian random field as the naïve model.

AppendixA

State-of-the-art anthropomorphic breast phantoms

Model	Type	Format	Large scale		Medium scale				Small scale
			Breast shape	Subcutaneous adipose region	Inter-glandular adipose compartments	Cooper's ligaments	Lactiferous ducts	Blood vessels	Fibroglandular tissue
UPenn phantom (2003 - 2017)	M (H later)	Voxel	Ellipsoids	Voronoi cells inside a geometrically defined region	Voronoi diagram with Mahalanobis distance	Space btw adjacent Voronoi cells	Topology: ramification matrix; spatial arrangement: peripheral & central ducts; with cylinders	Not present	Assign dense tissue to random Voronoi cells (later versions added 3D random fields as microtextures)
Bliznokova <i>et al.</i> (2003, 2010)	H	Voxel	Ellipsoid + hyperboloid	Thin ellipsoid shells	Gray-level 3D random field using random walk	Thin ellipsoid shells	Recursive trees with up to 5 nodes for each branch; with cylinders	Not explicitly modeled	Gray-level 3D random field using random walk
Li <i>et al.</i> (2009, 2013)	E	Mesh	Breast mask segmented from bCT data	Segmented from bCT data (some structures such as Cooper's ligament, lactiferous ducts and blood vessels are not fully visible / discernable due to bCT resolution and segmentation algorithm)					
Mahr <i>et al.</i> (2012)	M	Mesh	Ellipsoid + 2nd degree polynomial	Sphere inserts inside a geometrically defined region	Random ellipsoids	Thin ellipsoid shells	Same as the UPenn phantom	Not present	Spheres at duct terminal
Carton <i>et al.</i> (2014)	M	Mesh	Truncated right circular cone + ellipsoid section	Deformed Voronoi cells	Voronoi diagram with Mahalanobis distance	Space btw adjacent deformed Voronoi cells	Same as the UPenn phantom but with 3D Bézier tubes	Similar to ducts	Random growth of 3D Bézier tubes around ducts

Graff (2016)	M	Voxel	Analytical geometric functions in \neq breast quadrants	Voronoi cells	Deformed spheroids	Voxels within a specific distance to the inter-glandular adipose lobules	One recursive random tree per glandular compartment; with toroidal segments	Random trees	Assignment of dense tissue to random Voronoi cells
-----------------	---	-------	---	---------------	--------------------	--	---	--------------	--

Table A.1: Summary of the methodologies used in state-of-the-art anthropomorphic breast phantoms to model the large, medium and small anatomical structures. M stands for model-based phantom; E stands for empirical-based phantom; H stands for hybrid phantom.

AppendixB

Channelized Hotelling observer

The channelized Hotelling observer (CHO) is a variant of the Hotelling observer (HO) [21]. The HO is the linear model observer that maximizes the detectability index d' written as,

$$d' = \frac{|\mathbb{E}(\lambda_1) - \mathbb{E}(\lambda_0)|}{\sqrt{\frac{1}{2} (\text{Var}(\lambda_1) + \text{Var}(\lambda_0))}}, \quad (\text{B.1})$$

with,

$$\lambda_1 = \mathbf{w}_{\text{HO}}^T \mathbf{g}_1 \text{ and } \lambda_0 = \mathbf{w}_{\text{HO}}^T \mathbf{g}_0. \quad (\text{B.2})$$

Here \mathbf{w}_{HO} is referred to as the template of the HO. We use \mathbf{g}_1 and \mathbf{g}_0 to denote the signal-present and signal-absent images respectively. The optimal HO template that maximizes the d' can be derived using *linear discriminant analysis* [68]. This yields,

$$\mathbf{w}_{\text{HO}} \propto \mathbf{S}_{\text{HO}}^{-1} \bar{\mathbf{s}}, \quad (\text{B.3})$$

with

$$\begin{aligned} \bar{\mathbf{s}} &= \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0, \\ \mathbf{S}_{\text{HO}} &= \frac{1}{2} (\mathbf{S}_0 + \mathbf{S}_1). \end{aligned} \quad (\text{B.4})$$

Here $\bar{\mathbf{g}}_0$ and $\bar{\mathbf{g}}_1$ are the sample means of the images conditioned on signal-absent and signal-present respectively; \mathbf{S}_0 and \mathbf{S}_1 are the sample covariance matrices of the images conditioned on signal-absent and signal-present respectively. Under (1.2), $\bar{\mathbf{s}}$ acts as an estimate of the average signal.

The CHO is an extension of the HO that uses channelized images as input. A channelized image is obtained by decomposing the original image \mathbf{g} into a set of features using a series of channels with the number of channels N_c depending on the clinical task. The channels are represented by a matrix \mathbf{T} where each column $\mathbf{t}_{i, 1 \leq i \leq N_c}$ denotes a single channel image using the vectorized notation. All channel images have the same size as the original image. Channels are defined in either the spatial domain or the frequency domain. For the later case, the images are first subjected to a discrete Fourier transform before being given as input to the CHO. For an input image \mathbf{g} , the response to channel \mathbf{t}_i is derived by computing the inner-product of \mathbf{g} and \mathbf{t}_i . Hence, the channel response vector \mathbf{v} can then be represented using matrix multiplication. That is,

$$\mathbf{v} = \mathbf{T}^T \mathbf{g} \quad (\text{B.5})$$

where \mathbf{v} is referred to as the channelized image. The template \mathbf{w}_{CHO} of the CHO can be obtained using the same linear discriminant analysis as the HO. This gives,

$$\mathbf{w}_{\text{CHO}} \propto \mathbf{T} \mathbf{K}_{\text{CHO}}^{-1} \bar{\mathbf{v}}, \quad (\text{B.6})$$

with

$$\begin{aligned}\bar{\mathbf{v}}_s &= \bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_0, \\ \mathbf{K}_{\text{CHO}} &= \frac{1}{2} (\mathbf{K}_0 + \mathbf{K}_1).\end{aligned}\tag{B.7}$$

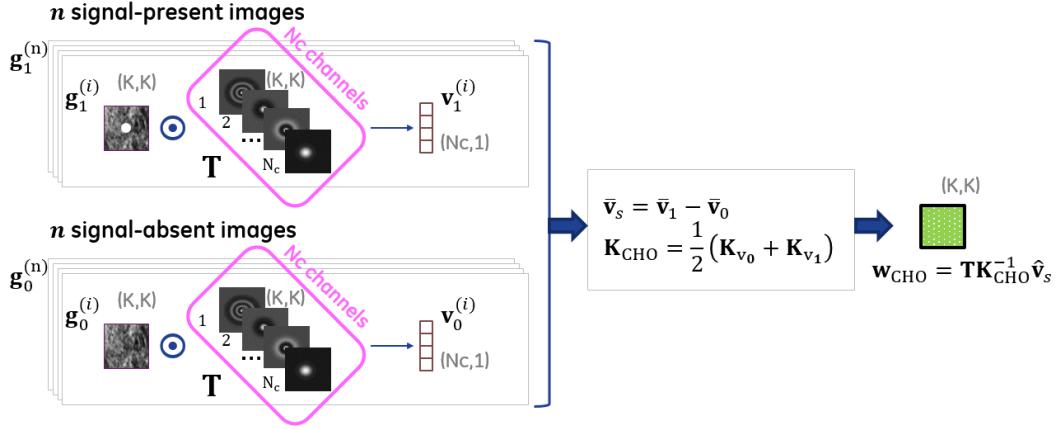
Here $\bar{\mathbf{v}}_0$ and $\bar{\mathbf{v}}_1$ are the sample means of the channelized images conditioned on signal-absent and signal-present respectively; \mathbf{K}_0 and \mathbf{K}_1 are the sample covariance matrices of the channelized images conditioned on signal-absent and signal-present respectively. Under (1.2), $\bar{\mathbf{v}}_s$ acts as an estimate of the average channelized signal. The computational steps in (B.6) and (B.7) to derive the CHO template \mathbf{w}_{CHO} are referred to as the training phase (Figure B.1a). Once the training phase is complete, the CHO can compute the decision variable λ for new images according to (1.3). This is referred to as the test phase (Figure B.1b).

The channel mechanism was initially introduced to simplify the computation of the inverse covariance matrix in HO by greatly reducing the dimension of the input images [141]. Typically, the number of channels N_c is much smaller than the number of elements in input images. As a result, there are much fewer number of elements in vector \mathbf{v} than in the original image \mathbf{g} [84]. This implies a practical advantage of CHO over HO that is, to be able to properly invert the covariance matrix, the CHO requires much fewer number of training images than HO. Later-on it was also shown that the channel mechanism may be used to model characteristics of the human visual system in certain conditions. Different spatial and frequency channel models have been investigated in the literature, depending on the application [194] [84]. Frequently applied channel models are the square channels [170] [141], the Gabor channels [207] and the Laguerre-Gauss channels [147].

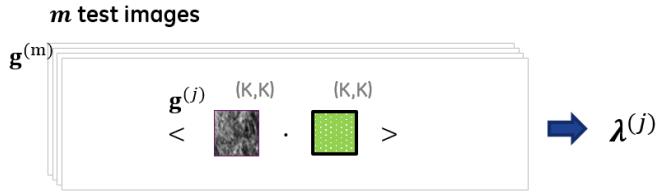
Since the definition of CHO is not limited by the dimension of the images, extensions of the CHO to 3D volumetric datasets have also been investigated in literature. We refer to the different designs of 3D CHO models by Platiša *et al.* described in [155], which has been applied in various VCT studies [6] [118]. In their original paper, Platiša *et al.* described two types of designs of CHO for 3D datasets: the *volumetric* CHO models and the *multi-slice* CHO models.

The volumetric CHO (CHO3vol) shares the same design as the 2D CHO and acts as a straight-forward extension of the 2D CHO. With CHO3vol, the channels are extended to 3D. Each input 3D dataset is channelized as a whole, by a set of 3D channels. Like in the 2D case, the response to a channel is the inner-product of the input 3D data and the 3D channel, both viewed as column vectors. Once the channelized images are obtained, the training and test phase of CHO3vol proceed in the same way as the 2D CHO.

The multi-slice CHO (CHO3ms) models use a different design than the CHO3vol. With CHO3ms models, the input 3D datasets are treated as collections of 2D slices. Instead of channelizing an input dataset as a whole by a set of 3D channels, the CHO3ms models channelize an input dataset slice-by-slice using a set of 2D channels [155]. This channelization is referred to as the multi-slice channelization. The multi-slice channelization produces a collection of channelized slices, with the number of channelized slices equal to the number of slices in the input 3D datasets. All collections of channelized slices are then used to derive the final CHO3ms template. There are three variants of CHO3ms and they differ in the schemes to compute the final CHO3ms template. These three variants are referred to as the multi-slice CHO of type A (CHO3msa), the multi-slice CHO of type B (CHO3msb) and the multi-slice CHO of type C (CHO3msc) respectively. Hereunder we



(a) The CHO training phase to compute the CHO template \mathbf{w}_{CHO} . The i^{th} input signal-present image $\mathbf{g}_1^{(i)}$ and signal-absent image $\mathbf{g}_0^{(i)}$ are first channelized by N_c channels (grouped into a channel matrix \mathbf{T}) to obtain the channelized images $\mathbf{v}_1^{(i)}$ and $\mathbf{v}_0^{(i)}$ respectively. From the channelized images, we then compute the average channelized signal $\bar{\mathbf{v}}_s$ and the covariance matrix \mathbf{K}_{CHO} . Finally the CHO template is derived as the product of \mathbf{T} , the inverse of \mathbf{K}_{CHO} and $\bar{\mathbf{v}}_s$.

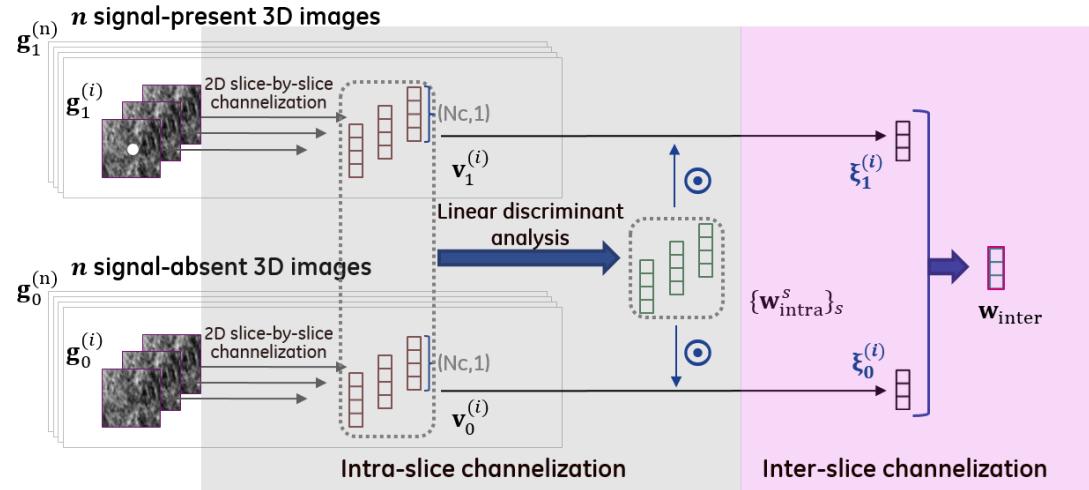


(b) The CHO test phase to compute the decision variable. For the j^{th} test image $g^{(j)}$, its associated decision variable $\lambda^{(j)}$ is computed as the inner product of $g^{(j)}$ and the CHO template \mathbf{w}_{CHO} obtained during the training phase.

Figure B.1: The training and test phases of the CHO.

briefly describe the training and the test phases of the CHO3msa model, which is the most general model of the three variants. For a more detailed description of all three CHO3ms variants, we refer to the paper by Platiša *et al.* [155]. The training phase of the CHO3msa model can be divided into two steps: the *intra-slice decorrelation* step and the *inter-slice decorrelation* step.

1. The intra-slice decorrelation step compute one template per slice using all channelized slices at the same position from all input 3D datasets. The template derived for each channelized slice is referred to as the intra-slice template. The intra-slice template for each channelized slice is obtained by applying the linear discriminant analysis (Equation (B.6) and Equation (B.7)). Then the inner-product of each channelized slice and its corresponding intra-slice template is computed, to obtain a scalar decision variable, referred to as the intra-slice decision variable. The intra-slice decorrelation step transforms each input 3D dataset into a vector of intra-slice decision variables, with the number of elements equal to number of slices in the original input 3D dataset.
2. The inter-slice decorrelation step takes the output vectors of intra-slice decision vari-



(a) The CHO3msa training phase to compute the intra-slice templates $\{w_{s,\text{intra}}^s\}_s$ and the inter-slice template w_{inter} . The i^{th} input signal-present 3D image $g_1^{(i)}$ and signal-absent 3D image $g_0^{(i)}$ are first channelized slice by slice using a set of 2D channels to obtain the channelized slices $v_1^{(i)}$ and $v_0^{(i)}$ respectively. From the channelized slices, we then apply the linear discriminant analysis (Equation (B.6) and Equation (B.7)) to obtain the set of intra-slice templates $\{w_{s,\text{intra}}^s\}_s$. Next, we compute the inner-product of the intra-slice templates and respectively the i^{th} set of signal-present channelized slices $v_1^{(i)}$ and the i^{th} set of signal-absent channelized slices $v_0^{(i)}$. This results in two vectors of intra-slice decision variables $\xi_1^{(i)}$ and $\xi_0^{(i)}$. Finally, applying the Hotelling observer template defined in (B.3) on these intra-slice decision variable vectors yields the inter-slice template w_{inter} .

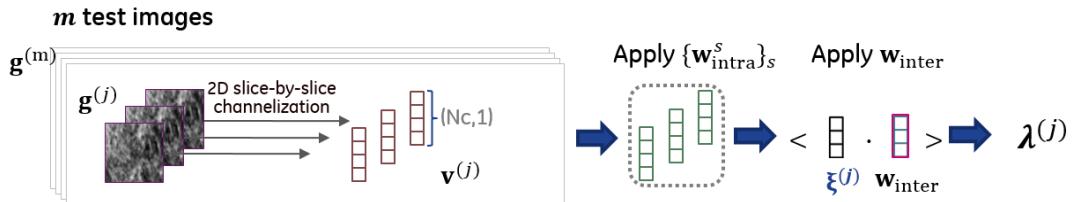


Figure B.2: The training and test phases of the CHO3msa model.

ables from the intra-slice decorrelation step as the inputs, and compute an inter-slice template. The computation is performed by directly applying the formula of the Hotelling observer template defined in (B.3).

The training phase of CHO3msa is illustrated in Figure B.2a.

In the CHO3msa test phase, a new input 3D dataset first goes through the same multi-slice channelization as in the training phase. This results in a set of channelized slices. Then the intra-slice templates and the inter-slice template are applied to the channelized slices in order. This results in the final decision variable. The test phase of CHO3msa is illustrated in Figure B.2b.

AppendixC

Additonal results

Table C.1 lists the additional eight sets of medium scale parameters for the 3D solid breast texture model inferred from segmented clinical breast computerized tomography (bCT) reconstructed volumes of interest (VOI). These VOIs are numbered as 2, 4, 6, 8, 9, 13 and 14 in Table 4.2. Figure C.1 and Figure C.2 show examples of 3.5 cm × 3.5 cm slices through volumes simulated from the 3D breast texture model with these eight sets of medium scale parameters listed in Table 4.4, as well as mammographic projections and DBT reconstructed slices simulated from these volumes. The simulation condition was detailed in Section 4.4.3.

Set	Φ_s (Matern Cluster)	L_a, L_b, L_c (in mm)	$\delta_{\phi_x}, \delta_{\phi_y}, \delta_{\phi_z}$ (in radian)
From VOI #2	$\kappa = 4.719e - 02$, $\lambda_0 = 4.331e - 02$, $R = 1.218$	$p_{L_a} = \mathcal{N}(6.06, 1.53)$, $p_{L_b} = \mathcal{N}(2.79, 0.59)$, $p_{L_c} = \mathcal{N}(2.04, 0.52)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(0, 0.26)$, $p_{\delta_{\phi_z}} = \mathcal{N}(0.01, 0.39)$
From VOI #4	$\kappa = 1.014e - 04$, $\lambda_0 = 1.515e - 02$, $R = 10.408$	$p_{L_a} = \mathcal{N}(5.98, 1.42)$, $p_{L_b} = \mathcal{N}(2.82, 0.56)$, $p_{L_c} = \mathcal{N}(2.06, 0.53)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.23, 0.43)$, $p_{\delta_{\phi_z}} = \mathcal{N}(0.04, 0.51)$
From VOI #6	$\kappa = 5.652e - 04$, $\lambda_0 = 1.185e - 02$, $R = 6.980$	$p_{L_a} = \mathcal{N}(5.93, 1.47)$, $p_{L_b} = \mathcal{N}(2.81, 0.58)$, $p_{L_c} = \mathcal{N}(2.04, 0.52)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.38, 0.53)$, $p_{\delta_{\phi_z}} = \mathcal{N}(-0.01, 0.47)$
From VOI #8	$\kappa = 1.102e - 03$, $\lambda_0 = 1.787e - 02$, $R = 5.150$	$p_{L_a} = \mathcal{N}(5.87, 1.45)$, $p_{L_b} = \mathcal{N}(2.74, 0.59)$, $p_{L_c} = \mathcal{N}(2.04, 0.53)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.19, 0.51)$, $p_{\delta_{\phi_z}} = \mathcal{N}(0, 0.47)$
From VOI #9	$\kappa = 1.752e - 03$, $\lambda_0 = 1.836e - 02$, $R = 4.467$	$p_{L_a} = \mathcal{N}(6.11, 1.49)$, $p_{L_b} = \mathcal{N}(2.80, 0.57)$, $p_{L_c} = \mathcal{N}(2.10, 0.53)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.15, 0.43)$, $p_{\delta_{\phi_z}} = \mathcal{N}(0.02, 0.49)$
From VOI #10	$\kappa = 1.374e - 02$, $\lambda_0 = 6.429e - 03$, $R = 3.643$	$p_{L_a} = \mathcal{N}(6.17, 1.44)$, $p_{L_b} = \mathcal{N}(2.85, 0.58)$, $p_{L_c} = \mathcal{N}(2.12, 0.54)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.18, 0.47)$, $p_{\delta_{\phi_z}} = \mathcal{N}(-0.02, 0.45)$
From VOI #13	$\kappa = 8.206e - 04$, $\lambda_0 = 1.380e - 02$, $R = 6.608$	$p_{L_a} = \mathcal{N}(5.97, 1.36)$, $p_{L_b} = \mathcal{N}(2.78, 0.58)$, $p_{L_c} = \mathcal{N}(2.04, 0.53)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(-0.15, 0.43)$, $p_{\delta_{\phi_z}} = \mathcal{N}(0.03, 0.48)$
From VOI #14	$\kappa = 6.643e - 04$, $\lambda_0 = 5.720e - 03$, $R = 9.993$	$p_{L_a} = \mathcal{N}(6.19, 1.47)$, $p_{L_b} = \mathcal{N}(2.79, 0.61)$, $p_{L_c} = \mathcal{N}(2.18, 0.59)$	$p_{\delta_{\phi_x}} = \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $p_{\delta_{\phi_y}} = \mathcal{N}(0, 0.26)$, $p_{\delta_{\phi_z}} = \mathcal{N}(-0.01, 0.38)$

Table C.1: Additional eight sets of fitted medium scale parameters for the 3D solid breast texture model using bCT VOIs #2, #4, #6, #8, #9, #13 and #14 listed in Table 4.2.

Table C.2, Table C.3, Table C.4 and Table C.5 list the retained Laguerre-Gauss parameters

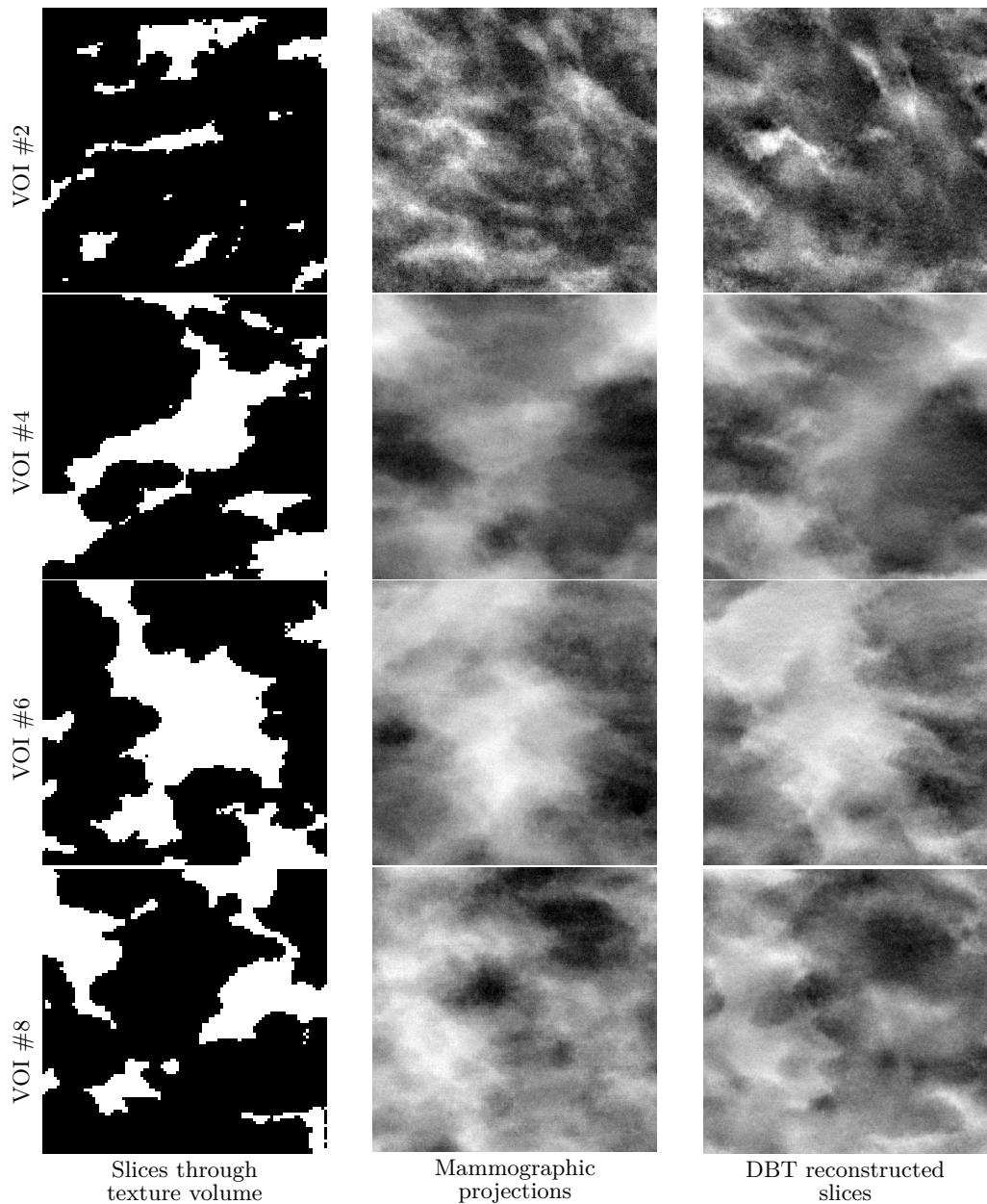


Figure C.1: The first column shows slices through volumes simulated from the 3D breast texture model with the four sets of parameters listed in Table C.1. The second column shows mammographic projections simulated from the simulated texture volumes. The third column shows DBT reconstructed slices simulated from the simulated texture volumes. The sizes of the images are $3.5\text{ cm} \times 3.5\text{ cm}$.

a_u (the Gaussian spread) and p (the degree of the Laguerre polynomial) that yielded the maximum area under the ROC curve for the CHO2, CHO3vol and CHO3msa models in all experimental conditions described in Chapter 7.

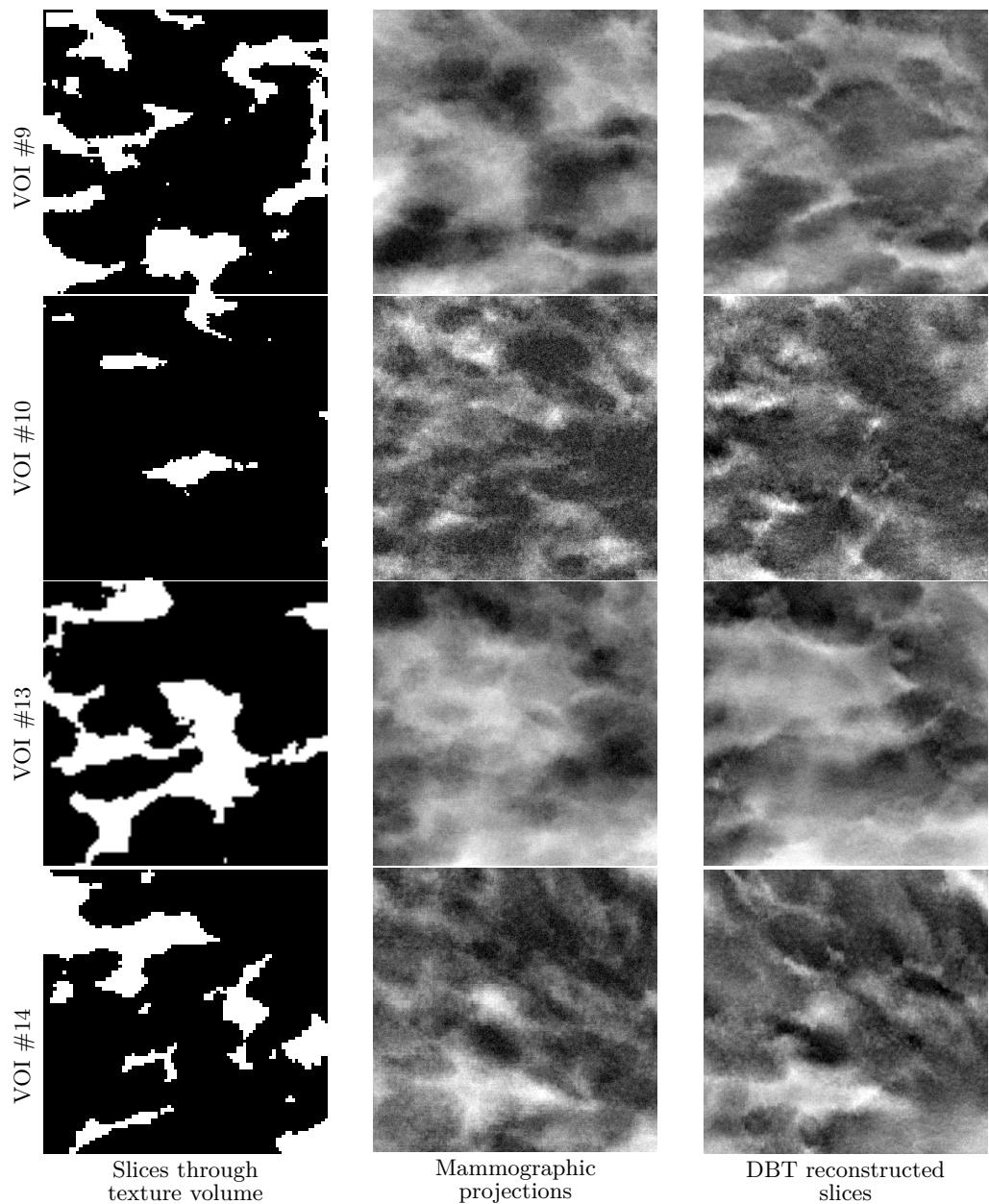


Figure C.2: Figure C.1 continued.

CHO2 in FFDM images				
d (μm)	μ	Uniform	BI-RADS breast density 2	BI-RADS breast density 3
≈ 100	20%	$a_u = 8, p = 23$	$a_u = 3, p = 29$	$a_u = 2, p = 28$
	40%	$a_u = 3, p = 28$	$a_u = 5, p = 29$	$a_u = 5, p = 29$
	60%	$a_u = 5, p = 29$	$a_u = 6, p = 27$	$a_u = 7, p = 27$
≈ 200	20%	$a_u = 4, p = 29$	$a_u = 2, p = 28$	$a_u = 3, p = 25$
	40%	$a_u = 3, p = 28$	$a_u = 4, p = 29$	$a_u = 5, p = 26$
	60%	$a_u = 1, p = 16$	$a_u = 2, p = 29$	$a_u = 8, p = 24$
≈ 400	20%	$a_u = 5, p = 29$	$a_u = 2, p = 27$	$a_u = 3, p = 27$
	40%	$a_u = 1, p = 10$	$a_u = 6, p = 28$	$a_u = 3, p = 29$
	60%	$a_u = 1, p = 0$	$a_u = 2, p = 28$	$a_u = 3, p = 29$
≈ 600	20%	$a_u = 1, p = 1$	$a_u = 2, p = 27$	$a_u = 3, p = 26$
	40%	$a_u = 1, p = 0$	$a_u = 2, p = 24$	$a_u = 3, p = 29$
	60%	$a_u = 1, p = 0$	$a_u = 1, p = 7$	$a_u = 1, p = 7$

Table C.2: Retained Laguerre-Gauss channel parameters (a_u, p) that gave maximum average AUC for μ_{calc} detection in FFDM images using CHO2. Results are shown for all combinations of μ_{calc} diameter, attenuation and background texture type.

CHO2 in synthetic 2D images				
d (μm)	μ	Uniform	BI-RADS breast density 2	BI-RADS breast density 3
≈ 100	20%	$a_u = 7, p = 29$	$a_u = 6, p = 29$	$a_u = 3, p = 26$
	40%	$a_u = 5, p = 28$	$a_u = 8, p = 29$	$a_u = 3, p = 27$
	60%	$a_u = 8, p = 22$	$a_u = 5, p = 29$	$a_u = 3, p = 29$
≈ 200	20%	$a_u = 7, p = 29$	$a_u = 3, p = 25$	$a_u = 5, p = 29$
	40%	$a_u = 3, p = 29$	$a_u = 3, p = 29$	$a_u = 2, p = 27$
	60%	$a_u = 4, p = 29$	$a_u = 6, p = 29$	$a_u = 4, p = 28$
≈ 400	20%	$a_u = 2, p = 29$	$a_u = 4, p = 28$	$a_u = 2, p = 29$
	40%	$a_u = 3, p = 26$	$a_u = 2, p = 28$	$a_u = 6, p = 22$
	60%	$a_u = 1, p = 0$	$a_u = 7, p = 29$	$a_u = 2, p = 23$
≈ 600	20%	$a_u = 1, p = 18$	$a_u = 3, p = 28$	$a_u = 6, p = 28$
	40%	$a_u = 1, p = 0$	$a_u = 3, p = 29$	$a_u = 2, p = 28$
	60%	$a_u = 1, p = 0$	$a_u = 3, p = 26$	$a_u = 4, p = 28$

Table C.3: Retained Laguerre-Gauss channel parameters (a_u, p) that gave maximum average AUC for μ_{calc} detection in S2D images using CHO2. Results are shown for all combinations of μ_{calc} diameter D , attenuation μ and background texture type.

CHO3vol in DBT reconstructed slices				
d (μm)	μ	Uniform	BI-RADS breast density 2	BI-RADS breast density 3
≈ 100	20%	$a_u = 7, p = 29$	$a_u = 3, p = 29$	$a_u = 2, p = 29$
	40%	$a_u = 2, p = 18$	$a_u = 3, p = 16$	$a_u = 5, p = 29$
	60%	$a_u = 3, p = 29$	$a_u = 7, p = 29$	$a_u = 3, p = 29$
≈ 200	20%	$a_u = 2, p = 29$	$a_u = 3, p = 29$	$a_u = 8, p = 25$
	40%	$a_u = 8, p = 29$	$a_u = 6, p = 29$	$a_u = 3, p = 23$
	60%	$a_u = 8, p = 29$	$a_u = 4, p = 29$	$a_u = 6, p = 29$
≈ 400	20%	$a_u = 3, p = 29$	$a_u = 4, p = 28$	$a_u = 3, p = 29$
	40%	$a_u = 1, p = 0$	$a_u = 6, p = 28$	$a_u = 8, p = 29$
	60%	$a_u = 1, p = 0$	$a_u = 2, p = 22$	$a_u = 6, p = 29$
≈ 600	20%	$a_u = 1, p = 0$	$a_u = 3, p = 29$	$a_u = 8, p = 29$
	40%	$a_u = 1, p = 0$	$a_u = 1, p = 7$	$a_u = 2, p = 22$
	60%	$a_u = 1, p = 0$	$a_u = 1, p = 3$	$a_u = 1, p = 3$

Table C.4: Retained Laguerre-Gauss channel parameters (a_u, p) that gave maximum average AUC for μ_{calc} detection in DBT reconstructed slices using CHO3vol. Results are shown for all combinations of μ_{calc} diameter D , attenuation μ and background texture type.

CHO3msa in DBT reconstructed slices				
d (μm)	μ	Uniform	BI-RADS breast density 2	BI-RADS breast density 3
≈ 100	20%	$a_u = 3, p = 29$	$a_u = 2, p = 29$	$a_u = 5, p = 28$
	40%	$a_u = 3, p = 29$	$a_u = 4, p = 29$	$a_u = 5, p = 28$
	60%	$a_u = 6, p = 29$	$a_u = 2, p = 27$	$a_u = 2, p = 27$
≈ 200	20%	$a_u = 3, p = 29$	$a_u = 5, p = 29$	$a_u = 2, p = 28$
	40%	$a_u = 1, p = 29$	$a_u = 3, p = 29$	$a_u = 4, p = 28$
	60%	$a_u = 2, p = 29$	$a_u = 3, p = 29$	$a_u = 7, p = 20$
≈ 400	20%	$a_u = 8, p = 28$	$a_u = 4, p = 29$	$a_u = 3, p = 25$
	40%	$a_u = 2, p = 2$	$a_u = 8, p = 9$	$a_u = 4, p = 2$
	60%	$a_u = 1, p = 0$	$a_u = 8, p = 8$	$a_u = 3, p = 1$
≈ 600	20%	$a_u = 2, p = 10$	$a_u = 8, p = 28$	$a_u = 5, p = 3$
	40%	$a_u = 1, p = 0$	$a_u = 8, p = 3$	$a_u = 5, p = 1$
	60%	$a_u = 1, p = 0$	$a_u = 8, p = 7$	$a_u = 4, p = 1$

Table C.5: Retained Laguerre-Gauss channel parameters (a_u, p) that gave maximum average AUC for μ_{calc} detection in DBT reconstructed slices using CHO3msa. Results are shown for all combinations of μ_{calc} diameter D , attenuation μ and background texture type.

Bibliography

- [1] Aitkin, A. C. “On least squares and linear combination of observations”. In: *Proc. Roy. Soc. Edin. A.* Vol. 55. 1935, pp. 42–48.
 - [2] Alakhras, M., Mello-Thoms, C., Rickard, M., Bourne, R., and Brennan, P. “Efficacy of digital breast tomosynthesis for breast cancer diagnosis”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2014, pp. 90370V–90370V.
 - [3] Almecija, T. *Development of an anthropomorphic software model of the female breast*. Master thesis. Ecole Supérieure d’Electricité, 2015.
 - [4] Avanaki, A. N., Espig, K. S., Marchessoux, C., Krupinski, E. A., Bakic, P. R., Kimpe, T. R., and Maidment, A. D. “Integration of spatio-temporal contrast sensitivity with a multi-slice channelized Hotelling observer”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2013, 86730H–86730H.
 - [5] Avanaki, A. R., Espig, K. S., Maidment, A. D., Marchessoux, C., Bakic, P. R., and Kimpe, T. R. “Development and evaluation of a 3D model observer with nonlinear spatiotemporal contrast sensitivity”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2014, pp. 90370X–90370X.
 - [6] Avanaki, A. R., Espig, K. S., Xthona, A., Kimpe, T. R., Bakic, P. R., and Maidment, A. D. “It is hard to see a needle in a haystack: modeling contrast masking effect in a numerical observer”. In: *International Workshop on Digital Mammography*. Springer. 2014, pp. 723–730.
 - [7] Aylward, S. R., Hemminger, B. M., and Pisano, E. D. “Mixture modeling for digital mammogram display and analysis”. In: *Digital Mammography*. Springer, 1998, pp. 305–312.
 - [8] Baddeley, A., Bárány, I., and Schneider, R. “Spatial point processes and their applications”. In: *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004* (2007), pp. 1–75.
 - [9] Baddeley, A., Diggle, P. J., Hardegen, A., Lawrence, T., Milne, R. K., and Nair, G. “On tests of spatial pattern based on simulation envelopes”. In: *Ecological Monographs* 84.3 (2014), pp. 477–489.
 - [10] Baddeley, A. and Jensen, E. B. V. *Stereology for statisticians*. CRC Press, 2004.
 - [11] Baddeley, A. and Turner, R. “Practical maximum pseudolikelihood for spatial point patterns”. In: *Australian & New Zealand Journal of Statistics* 42.3 (2000), pp. 283–322.
 - [12] Baddeley, A. and Turner, R. “Spatstat: an R package for analyzing spatial point patterns”. In: *Journal of statistical software* 12.6 (2005), pp. 1–42.
 - [13] Baker, R., Rogers, K., Shepherd, N., and Stone, N. “New relationships between breast microcalcifications and cancer”. In: *British journal of cancer* 103.7 (2010), pp. 1034–1039.
-

- [14] Bakic, P. R., Albert, M., Brzakovic, D., and Maidment, A. D. "Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation". In: *Medical physics* 29.9 (2002), pp. 2131–2139.
- [15] Bakic, P. R., Albert, M., Brzakovic, D., and Maidment, A. D. "Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic mammogram texture". In: *Medical physics* 29.9 (2002), pp. 2140–2151.
- [16] Bakic, P. R., Albert, M., Brzakovic, D., and Maidment, A. D. "Mammogram synthesis using a three-dimensional simulation. III. Modeling and evaluation of the breast ductal network". In: *Medical Physics* 30.7 (2003), pp. 1914–1925.
- [17] Bakic, P. R., Carton, A.-K., Kontos, D., Zhang, C., Troxel, A. B., and Maidment, A. D. "Breast Percent Density: Estimation on Digital Mammograms and Central Tomosynthesis Projections 1". In: *Radiology* 252.1 (2009), pp. 40–49.
- [18] Bakic, P. R., Kontos, D., Carton, A.-K., and Maidment, A. D. "Breast percent density estimation from 3D reconstructed digital breast tomosynthesis images". In: *Medical Imaging*. International Society for Optics and Photonics. 2008, pp. 691318–691318.
- [19] Bakic, P. R., Pokrajac, D. D., De Caro, R., and Maidment, A. D. "Realistic simulation of breast tissue microstructure in software anthropomorphic phantoms". In: *International Workshop on Digital Mammography*. Springer. 2014, pp. 348–355.
- [20] Bakic, P. R., Zhang, C., and Maidment, A. D. "Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm". In: *Medical Physics* 38.6 (2011), pp. 3165–3176.
- [21] Barrett, H. H., Yao, J., Rolland, J. P., and Myers, K. J. "Model observers for assessment of image quality". In: *Proceedings of the National Academy of Sciences* 90.21 (1993), pp. 9758–9765.
- [22] Barten, P. G. *Contrast sensitivity of the human eye and its effects on image quality*. Vol. 72. SPIE press, 1999.
- [23] Besag, J. "Statistical analysis of non-lattice data". In: *The statistician* (1975), pp. 179–195.
- [24] Biermé, H., Richard, F., Rachidi, M., and Benhamou, C.-L. "Anisotropic texture modeling and applications to medical image analysis". In: *ESAIM: Proceedings*. Vol. 26. EDP Sciences. 2009, pp. 100–122.
- [25] Bliznakova, K., Bliznakov, Z., Bravou, V., Kolitsi, Z., and Pallikarakis, N. "A three-dimensional breast software phantom for mammography simulation". In: *Physics in medicine and biology* 48.22 (2003), p. 3699.
- [26] Bliznakova, K., Suryanarayanan, S., Karella, A., and Pallikarakis, N. "Evaluation of an improved algorithm for producing realistic 3D breast software phantoms: Application for mammography". In: *Medical physics* 37.11 (2010), pp. 5604–5617.
- [27] Bochud, F. O., Abbey, C. K., and Eckstein, M. P. "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds". In: *Optics express* 4.1 (1999), pp. 33–43.
- [28] Bochud, F. O., Valley, J.-F., Verdun, F. R., Hessler, C., and Schnyder, P. "Estimation of the noisy component of anatomical backgrounds". In: *Medical physics* 26.7 (1999), pp. 1365–1370.

- [29] Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., Jong, R. A., Hislop, G., Chiarelli, A., Minkin, S., et al. "Mammographic density and the risk and detection of breast cancer". In: *New England Journal of Medicine* 356.3 (2007), pp. 227–236.
- [30] Brown, C. D. and Davis, H. T. "Receiver operating characteristics curves and related decision measures: A tutorial". In: *Chemometrics and Intelligent Laboratory Systems* 80.1 (2006), pp. 24–38.
- [31] Burgess, A. E. "Mammographic structure: Data preparation and spatial statistics analysis". In: *Medical Imaging'99*. International Society for Optics and Photonics. 1999, pp. 642–653.
- [32] Burgess, A. E., Jacobson, F. L., and Judy, P. F. "Human observer detection experiments with mammograms and power-law noise". In: *Medical physics* 28.4 (2001), pp. 419–437.
- [33] Burgess, A. E., Jacobson, F. L., and Judy, P. F. "Lesion detection in digital mammograms". In: *Medical Imaging 2001*. International Society for Optics and Photonics. 2001, pp. 555–560.
- [34] Cai, S.-Q., Yan, J.-X., Chen, Q.-S., Huang, M.-L., and Cai, D.-L. "Significance and application of digital breast tomosynthesis for the BI-RADS classification of breast cancer". In: *Asian Pac J Cancer Prev* 16 (2015), pp. 4109–4114.
- [35] cancer, I. national du. *Les cancers en France en 2016: L'essentiel des faits et chiffres*. Tech. rep. Feb. 2017.
- [36] Carton, A.-K., Bosmans, H., Van Ongeval, C., Souverijns, G., Rogge, F., Van Steen, A., and Marchal, G. "Development and validation of a simulation procedure to study the visibility of micro calcifications in digital mammograms". In: *Medical physics* 30.8 (2003), pp. 2234–2240.
- [37] Carton, A.-K., Grisey, A., Carvalho, P. M. de, Dromain, C., and Muller, S. "A virtual human breast phantom using surface meshes and geometric internal structures". In: *International Workshop on Digital Mammography*. Springer. 2014, pp. 356–363.
- [38] Castella, C., Eckstein, M., Abbey, C., Kinkel, K., Verdun, F., Saunders, R., Samei, E., and Bochud, F. "Mass detection on mammograms: influence of signal shape uncertainty on human and model observers". In: *JOSA A* 26.2 (2009), pp. 425–436.
- [39] Castella, C., Kinkel, K., Descombes, F., Eckstein, M. P., Sottas, P.-E., Verdun, F. R., and Bochud, F. O. "Mammographic texture synthesis: second-generation clustered lumpy backgrounds using a genetic algorithm". In: *Optics express* 16.11 (2008), pp. 7595–7607.
- [40] Castella, C., Ruschin, M., Eckstein, M. P., Abbey, C., Kinkel, K., Verdun, F., Tingberg, A., and Bochud, F. "Mass detection in breast tomosynthesis and digital mammography: a model observer study". In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2009, 72630O–72630O.
- [41] Chawla, A. S., Lo, J. Y., Baker, J. A., and Samei, E. "Optimized image acquisition for breast tomosynthesis in projection and reconstruction space". In: *Medical physics* 36.11 (2009), pp. 4859–4869.

- [42] Chen, B., Shorey, J., Saunders, R. S., Richard, S., Thompson, J., Nolte, L. W., and Samei, E. “An anthropomorphic breast model for breast imaging simulation and optimization”. In: *Academic radiology* 18.5 (2011), pp. 536–546.
- [43] Chen, F., Bakic, P. R., Maidment, A. D., Jensen, S. T., Shi, X., and Pokrajac, D. D. “Description and characterization of a novel method for partial volume simulation in software breast phantoms”. In: *IEEE transactions on medical imaging* 34.10 (2015), pp. 2146–2161.
- [44] Chen, X., Gong, X., Graff, C. G., Santana, M., Sturgeon, G. M., Sauer, T. J., Zeng, R., Glick, S. J., and Lo, J. Y. “High-resolution, anthropomorphic, computational breast phantom: fusion of rule-based structures with patient-based anatomy”. In: vol. 10132. 2017, pp. 10132–10132. DOI: [10.1117/12.2255913](https://doi.org/10.1117/12.2255913). URL: <http://dx.doi.org/10.1117/12.2255913>.
- [45] Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- [46] Ciatto, S., Bernardi, D., Calabrese, M., Durando, M., Gentilini, M. A., Mariscotti, G., Monetti, F., Moriconi, E., Pesce, B., Roselli, A., et al. “A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification”. In: *The Breast* 21.4 (2012), pp. 503–506.
- [47] Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., Tuttobene, P., Bricolo, P., Fantò, C., Valentini, M., et al. “Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study”. In: *The lancet oncology* 14.7 (2013), pp. 583–589.
- [48] Cox, D. D. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. 1996.
- [49] Dee, K. E. and Sickles, E. A. “Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently”. In: *American Journal of Roentgenology* 176.3 (2001), pp. 729–733.
- [50] Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D., and Müller, H. “Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities”. In: *Medical image analysis* 18.1 (2014), pp. 176–196.
- [51] Dereudre, D., Lavancier, F., and Helisová, K. S. “Estimation of the Intensity Parameter of the Germ-Grain Quermass-Interaction Model when the Number of Germs is not Observed”. In: *Scandinavian Journal of Statistics* 41.3 (2014), pp. 809–829.
- [52] Descombes, X. *Stochastic geometry for image analysis*. John Wiley & Sons, 2013.
- [53] Descombes, X., Minlos, R., and Zhizhina, E. “Object extraction using a stochastic birth-and-death dynamics in continuum”. In: *Journal of Mathematical Imaging and Vision* 33.3 (2009), pp. 347–359.
- [54] Desolneux, A., Moisan, L., and Morel, J.-M. *From gestalt theory to image analysis: a probabilistic approach*. Vol. 34. Springer Science & Business Media, 2007.
- [55] Desolneux, A., Moisan, L., and Morel, J.-M. “Maximal meaningful events and applications to image analysis”. In: *Annals of Statistics* (2003), pp. 1822–1851.

- [56] Destounis, S. V., Arieno, A. L., and Morgan, R. C. "Preliminary clinical experience with digital breast tomosynthesis in the visualization of breast microcalcifications". In: *Journal of clinical imaging science* 3 (2013), p. 65.
- [57] Diaz, I., Timberg, P., Zhang, S., Abbey, C., Verdun, F., and Bochud, F. O. "Development of model observers applied to 3D breast tomosynthesis microcalcifications and masses". In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2011, 79660F–79660F.
- [58] Diggle, P. J. "Binary mosaics and the spatial pattern of heather". In: *Biometrics* (1981), pp. 531–539.
- [59] D'Orsi, C. J. *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology, 2013.
- [60] Dudgeon, D. E. and Mersereau, R. M. "Multidimensional digital signal processing". In: *Prentice-Hall Signal Processing Series, Englewood Cliffs: Prentice-Hall*, 1984 (1984).
- [61] Duembgen, L. "Bounding standard gaussian tail probabilities". In: *arXiv preprint arXiv:1012.2063* (2010).
- [62] Durning, M. V. *Breast Density Notification Laws by State: Interactive Map*. June 2017. URL: <http://www.diagnosticimaging.com/breast-imaging/breast-density-notification-laws-state-interactive-map>.
- [63] Dustler, M., Bakic, P., Petersson, H., Timberg, P., Tingberg, A., and Zackrisson, S. "Application of the fractal Perlin noise algorithm for the generation of simulated breast tissue". In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2015, 94123E–94123E.
- [64] Engeland, S. van, Snoeren, P. R., Huisman, H., Boetes, C., and Karssemeijer, N. "Volumetric breast density estimation from full-field digital mammograms". In: *IEEE transactions on medical imaging* 25.3 (2006), pp. 273–282.
- [65] Engstrom, E., Reiser, I., and Nishikawa, R. "Comparison of power spectra for tomosynthesis projections and reconstructed images". In: *Medical physics* 36.5 (2009), pp. 1753–1758.
- [66] Feller, W. *An introduction to probability theory and its applications*. Vol. 2. John Wiley & Sons, 2008.
- [67] Ferrari, R., Rangayyan, R., Desautels, J., and Frere, A. "Segmentation of mammograms: identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc". In: *Proceedings of the 5th international workshop on digital Mammography*. 2000, pp. 573–579.
- [68] Fisher, R. A. *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.
- [69] Foley, J. D. and Van Dam, A. *Fundamentals of interactive computer graphics*. Vol. 2. Addison-Wesley Reading, MA, 1982.
- [70] Freer, P. E. "Mammographic breast density: impact on breast cancer risk and implications for screening". In: *Radiographics* 35.2 (2015), pp. 302–315.
- [71] Galerne, B. "Stochastic image models and texture synthesis". PhD thesis. École normale supérieure de Cachan-ENS Cachan, 2010.

- [72] Gaur, S., Dialani, V., Slanetz, P. J., and Eisenberg, R. L. “Architectural distortion of the breast”. In: *American Journal of Roentgenology* 201.5 (2013), W662–W670.
- [73] Gautschi, W. “Some elementary inequalities relating to the gamma and incomplete gamma function”. In: *Studies in Applied Mathematics* 38.1-4 (1959), pp. 77–81.
- [74] Ge, W. and Collins, R. T. “Marked point processes for crowd counting”. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE. 2009, pp. 2913–2920.
- [75] Geyer, C. J. and Møller, J. “Simulation procedures and likelihood inference for spatial point processes”. In: *Scandinavian journal of statistics* (1994), pp. 359–373.
- [76] Graff, C. G. “A new open-source multi-modality digital breast phantom”. In: *SPIE Medical Imaging.* International Society for Optics and Photonics. 2016, pp. 978309–978309.
- [77] Green, P. J. “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination”. In: *Biometrika* (1995), pp. 711–732.
- [78] Grosjean, B. “Lesion Detectability in Digital Mammography: Impact of Texture”. In: *Paris: Ecole Centrale Paris, PhD in Mathématiques Appliquées aux Systèmes* (2007), p. 275.
- [79] Grosjean, B. and Moisan, L. “A-contrario Detectability of Spots in Textured Backgrounds”. In: *Journal of Mathematical Imaging and Vision* 33.3 (2009), pp. 313–337.
- [80] Grosjean, B., Muller, S., and Souchay, H. “Lesion detection using an a-contrario detector in simulated digital mammograms”. In: *Medical Imaging.* International Society for Optics and Photonics. 2006, 61460S–61460S.
- [81] Gur, D., Zuley, M. L., Anello, M. I., Rathfon, G. Y., Chough, D. M., Ganott, M. A., Hakim, C. M., Wallace, L., Lu, A., and Bandos, A. I. “Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study”. In: *Academic radiology* 19.2 (2012), pp. 166–171.
- [82] He, W., Denton, E. R., Stafford, K., and Zwiggelaar, R. “Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments”. In: *Biomedical Signal Processing and Control* 6.3 (2011), pp. 321–329.
- [83] He, W., Juette, A., Denton, E. R., Oliver, A., Martí, R., and Zwiggelaar, R. “A review on automatic mammographic density and parenchymal segmentation”. In: *International journal of breast cancer* 2015 (2015).
- [84] He, X. and Park, S. “Model observers in medical imaging research”. In: *Theranostics* 3.10 (2013), pp. 774–786.
- [85] Heinrich, L. “Asymptotic properties of minimum contrast estimators for parameters of Boolean models”. In: *Metrika* 40.1 (1993), pp. 67–94.
- [86] Henrot, P., Leroux, A., Barlier, C., and Génin, P. “Breast microcalcifications: The lesions in anatomical pathology”. In: *Diagnostic and interventional imaging* 95.2 (2014), pp. 141–152.

- [87] Hill, M. L., Mainprize, J. G., Carton, A.-K., Saab-Puong, S., Iordache, R., Muller, S., Jong, R. A., Dromain, C., and Yaffe, M. J. “Anatomical noise in contrast-enhanced digital mammography. Part II. Dual-energy imaging”. In: *Medical physics* 40.8 (2013).
- [88] Hope, A. C. “A simplified Monte Carlo significance test procedure”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1968), pp. 582–598.
- [89] Hsu, C. M., Palmeri, M. L., Segars, W. P., Veress, A. I., and Dobbins, J. T. “Generation of a suite of 3D computer-generated breast phantoms from a limited set of human subject data”. In: *Medical physics* 40.4 (2013).
- [90] Ikejimba, L., Glick, S. J., Samei, E., and Lo, J. Y. “Comparison of model and human observer performance in FFDM, DBT, and synthetic mammography”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2016, pp. 978325–978325.
- [91] Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. *Statistical analysis and modelling of spatial point patterns*. Vol. 70. John Wiley & Sons, 2008.
- [92] Iotti, V., Nitrosi, A., Coriani1, C., Caffarri, S., Campari, C., Ginocchi, V., Vacondio, R., Giorgi, P., and Rossi Pattacini1, P. “A randomised controlled trial to evaluate tomosynthesis vs digital mammography screening: preliminary results on baseline detection rate”. European Congress of Radiology. 2017.
- [93] Jeffreys, M., Harvey, J., and Highnam, R. “Comparing a new volumetric breast density method (Volpara TM) to cumulus”. In: *Digital mammography* (2010), pp. 408–413.
- [94] Jensen, J. L. and Møller, J. “Pseudolikelihood for exponential family models of spatial point processes”. In: *The Annals of Applied Probability* (1991), pp. 445–461.
- [95] Jeulin, D. “Random texture models for material structures”. In: *Statistics and Computing* 10.2 (2000), pp. 121–132.
- [96] Kallenberg, M. G., Lokate, M., Van Gils, C. H., and Karssemeijer, N. “Automatic breast density segmentation: an integration of different approaches”. In: *Physics in Medicine and Biology* 56.9 (2011), p. 2715.
- [97] Kaufhold, J., Thomas, J., Eberhard, J., Galbo, C., and Trotter, D. “A calibration approach to glandular tissue composition estimation in digital mammography”. In: *Medical physics* 29.8 (2002), pp. 1867–1880.
- [98] Kendall, W. S. “On some weighted Boolean models”. In: *Advances in Theory and Applications of Random Sets* (1997), pp. 105–120.
- [99] Kenney John, F. “Mathematics Of Ststistics”. In: (1939).
- [100] Kiarashi, N. “Towards Realizing Virtual Clinical Trials for Optimization and Evaluation of Breast”. PhD thesis. Duke University, 2014.
- [101] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. “Optimization by simulated annealing”. In: *Science* 220.4598 (1983), pp. 671–680.
- [102] Kolb, T. M., Lichy, J., and Newhouse, J. H. “Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations”. In: *Radiology* 225.1 (2002), pp. 165–175.
-

- [103] Kontos, D., Bakic, P., Troxel, A., Conant, E., and Maidment, A. “Digital breast tomosynthesis parenchymal texture analysis for breast cancer risk estimation: a preliminary study”. In: *Digital Mammography* (2008), pp. 681–688.
- [104] Kopans, D., Gavronis, S., Halpern, E., and Moore, R. “Calcifications in the breast and digital breast tomosynthesis”. In: *The breast journal* 17.6 (2011), pp. 638–644.
- [105] Lafarge, F., Gimel farb, G., and Descombes, X. “Geometric feature extraction by a multimarked point process”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (2010), pp. 1597–1609.
- [106] Lång, K., Andersson, I., Rosso, A., Tingberg, A., Timberg, P., and Zackrisson, S. “Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study”. In: *European radiology* 26.1 (2016), pp. 184–190.
- [107] Lau, B. *Optimization of Breast Tomosynthesis Imaging Systems for Computer-Aided Detection*. Tech. rep. DTIC Document, 2011.
- [108] Lau, B. A., Das, M., and Gifford, H. C. “Towards visual-search model observers for mass detection in breast tomosynthesis”. In: *Proc. SPIE*. Vol. 8668. 2013, p. 86680X.
- [109] Lau, B. A., Reiser, I., Nishikawa, R. M., and Bakic, P. R. “A statistically defined anthropomorphic software breast phantom”. In: *Medical physics* 39.6 (2012), pp. 3375–3385.
- [110] Li, C. M., Segars, W. P., Tourassi, G. D., Boone, J. M., and Dobbins, J. T. “Methodology for generating a 3D computerized breast phantom from empirical data”. In: *Medical physics* 36.7 (2009), pp. 3122–3131.
- [111] Liang, H., Park, S., Gallas, B. D., Myers, K. J., and Badano, A. “Image browsing in slow medical liquid crystal displays”. In: *Academic radiology* 15.3 (2008), pp. 370–382.
- [112] Lindfors, K. K., Boone, J. M., Nelson, T. R., Yang, K., Kwan, A. L., and Miller, D. F. “Dedicated breast CT: Initial clinical experience 1”. In: *Radiology* 246.3 (2008), pp. 725–733.
- [113] Ludwig, M. “Ellipsoids and matrix-valued valuations”. In: *Duke Mathematical Journal* 119.1 (2003), pp. 159–188.
- [114] Lutwak, E., Yang, D., and Zhang, G. “A new ellipsoid associated with convex bodies.” In: *Duke Mathematical Journal* 104.3 (2000), pp. 375–390.
- [115] Ma, A. K., Gunn, S., and Darambara, D. G. “Introducing DeBRA: A detailed breast model for radiological studies”. In: *Physics in medicine and biology* 54.14 (2009), p. 4533.
- [116] Macmillan, N. A. “Signal detection theory”. In: *Stevens' handbook of experimental psychology* (2002).
- [117] Mahr, D. M., Bhargava, R., and Insana, M. F. “Three-dimensional in silico breast phantoms for multimodal image simulations”. In: *IEEE transactions on medical imaging* 31.3 (2012), pp. 689–697.
- [118] Maidment, A. D. “Virtual clinical trials for the assessment of novel breast screening modalities”. In: *International Workshop on Digital Mammography*. Springer. 2014, pp. 1–8.

- [119] Mandelblatt, J. S., Cronin, K. A., Bailey, S., Berry, D. A., De Koning, H. J., Draisma, G., Huang, H., Lee, S. J., Munsell, M., Plevritis, S. K., et al. "Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms". In: *Annals of internal medicine* 151.10 (2009), pp. 738–747.
- [120] Mandelson, M. T., Oestreicher, N., Porter, P. L., White, D., Finder, C. A., Taplin, S. H., and White, E. "Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers". In: *Journal of the National Cancer Institute* 92.13 (2000), pp. 1081–1087.
- [121] Mansour, S., Adel, L., Mokhtar, O., and Omar, O. S. "Comparative study between breast tomosynthesis and classic digital mammography in the evaluation of different breast lesions". In: *The Egyptian Journal of Radiology and Nuclear Medicine* 45.3 (2014), pp. 1053–1061.
- [122] Massey Jr, F. J. "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American statistical Association* 46.253 (1951), pp. 68–78.
- [123] Matheron, G. "Random sets theory and its applications to stereology". In: *Journal of Microscopy* 95.1 (1972), pp. 15–23.
- [124] McLachlan, G. *Discriminant analysis and statistical pattern recognition*. Vol. 544. John Wiley & Sons, 2004.
- [125] McPherson, K., Steel, C., and Dixon, J. "ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics". In: *BMJ: British Medical Journal* 321.7261 (2000), p. 624.
- [126] Metheny, K. G., Abbey, C. K., Packard, N., and Boone, J. M. "Characterizing anatomical variability in breast CT images". In: *Medical physics* 35.10 (2008), pp. 4685–4694.
- [127] Michael O'Connor, J., Das, M., Dider, C. S., Mahd, M., and Glick, S. J. "Generation of voxelized breast phantoms from surgical mastectomy specimens". In: *Medical physics* 40.4 (2013).
- [128] Michell, M. "Breast screening review—a radiologist's perspective". In: *The British journal of radiology* (2014).
- [129] Michell, M., Iqbal, A., Wasan, R., Evans, D., Peacock, C., Lawinski, C., Douiri, A., Wilson, R., and Whelehan, P. "A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis". In: *Clinical radiology* 67.10 (2012), pp. 976–981.
- [130] Michielsen, K., Nuysts, J., Cockmartin, L., Marshall, N., and Bosmans, H. "Design of a model observer to evaluate calcification detectability in breast tomosynthesis and application to smoothing prior optimization". In: *Medical Physics* 43.12 (2016), pp. 6577–6587.
- [131] Michielsen, K., Zanca, F., Marshall, N., Bosmans, H., and Nuysts, J. "Two complementary model observers to evaluate reconstructions of simulated micro-calcifications in digital breast tomosynthesis". In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2013, 86730G–86730G.
- [132] Milioni de Carvalho, P. "Low-dose 3D quantitative vascular X-ray imaging of the breast". PhD thesis. University Paris 11, 2014.

- [133] Miller, P. and Astley, S. “Classification of breast tissue by texture analysis”. In: *Image and Vision Computing* 10.5 (1992), pp. 277–282.
- [134] Møller, J. and Helisová, K. “Likelihood inference for unions of interacting discs”. In: *Scandinavian Journal of Statistics* 37.3 (2010), pp. 365–381.
- [135] Møller, J. and Waagepetersen, R. P. “Modern statistics for spatial point processes”. In: *Scandinavian Journal of Statistics* 34.4 (2007), pp. 643–684.
- [136] Moller, J. and Waagepetersen, R. P. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.
- [137] Monnin, P., Marshall, N., Bosmans, H., Bochud, F., and Verdun, F. “Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis”. In: *Physics in medicine and biology* 56.14 (2011), p. 4221.
- [138] Mumford, D. and Desolneux, A. *Pattern theory: the stochastic analysis of real-world signals*. CRC Press, 2010.
- [139] Murphy, K. P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [140] Myers, K., Barrett, H. H., Borgstrom, M., Patton, D., and Seeley, G. “Effect of noise correlation on detectability of disk signals in medical imaging”. In: *JOSA A* 2.10 (1985), pp. 1752–1759.
- [141] Myers, K. J. and Barrett, H. H. “Addition of a channel mechanism to the ideal-observer model”. In: *JOSA A* 4.12 (1987), pp. 2447–2457.
- [142] Nelson, J. S., Wells, J. R., Baker, J. A., and Samei, E. “How does c-view image quality compare with conventional 2D FFDM ?” In: *Medical physics* 43.5 (2016), pp. 2538–2547.
- [143] Ng, R. “Fourier slice photography”. In: *ACM Transactions on Graphics (TOG)*. Vol. 24. 3. ACM. 2005, pp. 735–744.
- [144] Nicholson, B. T., LoRusso, A. P., Smolkin, M., Bovbjerg, V. E., Petroni, G. R., and Harvey, J. A. “Accuracy of assigned BI-RADS breast density category definitions”. In: *Academic radiology* 13.9 (2006), pp. 1143–1149.
- [145] Ohser, J. “On estimators for the reduced second moment measure of point processes”. In: *Statistics: A Journal of Theoretical and Applied Statistics* 14.1 (1983), pp. 63–71.
- [146] Oliver, A., Freixenet, J., Bosch, A., Raba, D., and Zwigelaar, R. “Automatic classification of breast tissue”. In: *Pattern Recognition and Image Analysis* (2005), pp. 171–175.
- [147] Park, S., Barrett, H. H., Clarkson, E., Kupinski, M. A., and Myers, K. J. “Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal”. In: *JOSA A* 24.12 (2007), B136–B150.
- [148] Park, S., Witten, J. M., and Myers, K. J. “Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer”. In: *IEEE transactions on medical imaging* 28.5 (2009), pp. 657–668.
- [149] Park, S., Zhang, G. Z., Zeng, R., and Myers, K. J. “Comparing observer models and feature selection methods for a task-based statistical assessment of digital breast tomosynthesis in reconstruction space”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2014, pp. 90370M–90370M.

- [150] Park, S., Zhang, G., and Myers, K. J. “Comparison of channel methods and observer models for the task-based assessment of multi-projection imaging in the presence of structured anatomical noise”. In: *IEEE transactions on medical imaging* 35.6 (2016), pp. 1431–1442.
- [151] Pereira, S. M. P., McCormack, V. A., Moss, S. M., and Santos Silva, I. dos. “The spatial distribution of radiodense breast tissue: a longitudinal study”. In: *Breast cancer research* 11.3 (2009), R33.
- [152] Pertuz, S., McDonald, E. S., Weinstein, S. P., Conant, E. F., and Kontos, D. “Fully automated quantitative estimation of volumetric breast density from digital breast tomosynthesis images: preliminary results and comparison with digital mammography and mr imaging”. In: *Radiology* 279.1 (2015), pp. 65–74.
- [153] Petrov, D., Cockmartin, L., Marshall, N., Vancoillie, L., Young, K., and Bosmans, H. “Real space channelization for generic DBT system image quality evaluation with channelized Hotelling observer”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2017, 101360N–101360N.
- [154] Pivot, X., Rixe, O., Morere, J., Coscas, Y., Cals, L., Namer, M., Serin, D., Dolbeault, S., Eisinger, F., Roussel, C., et al. “Breast cancer screening in France: results of the EDIFICE survey”. In: *International journal of medical sciences* 5.3 (2008), p. 106.
- [155] Platiša, L., Goossens, B., Vansteenkiste, E., Park, S., Gallas, B. D., Badano, A., and Philips, W. “Channelized Hotelling observers for the assessment of volumetric imaging data sets”. In: *JOSA A* 28.6 (2011), pp. 1145–1163.
- [156] Platiša, L., Marchessoux, C., Kimpe, T., Vansteenkiste, E., Badano, A., and Philips, W. “Channelized Hotelling observers for signal detection in stack-mode reading of volumetric images on medical displays with slow response time”. In: *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*. IEEE. 2011, pp. 2697–2702.
- [157] Pokrajac, D. D., Kuperavage, A., Maidment, A. D., and Bakic, P. R. “Reduction of artifacts in computer simulation of breast Cooper’s ligaments”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2016, 97832Q–97832Q.
- [158] Pokrajac, D. D., Maidment, A. D., and Bakic, P. R. “Optimized generation of high resolution breast anthropomorphic software phantoms”. In: *Medical physics* 39.4 (2012), pp. 2290–2302.
- [159] Poplack, S. P., Tosteson, T. D., Kogel, C. A., and Nagy, H. M. “Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography”. In: *American Journal of Roentgenology* 189.3 (2007), pp. 616–623.
- [160] Radiology. BI-RADS Committee, A. C. of. *Breast imaging reporting and data system*. American College of Radiology, 1998.
- [161] Rashidnasab, A., Bemelmans, F., Salvagnini, E., Marshall, N. W., and Bosmans, H. “Detectability of malignant mass models in 2D mammograms, differentiated towards size and background glandularity”. In: *Medical Image Perception Society (MIPS) Conference XVI*. 2015, p. 58.
- [162] Reiser, I., Edwards, A., and Nishikawa, R. “Validation of a power-law noise model for simulating small-scale breast tissue”. In: *Physics in medicine and biology* 58.17 (2013), p. 6011.

- [163] Reiser, I., Lee, S., and Nishikawa, R. “On the orientation of mammographic structure”. In: *Medical physics* 38.10 (2011), pp. 5303–5306.
- [164] Reiser, I. and Nishikawa, R. “Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise”. In: *Medical physics* 37.4 (2010), pp. 1591–1600.
- [165] Rick, A., Muller, S., Bothorel, S., and Grimaud, M. “Quantitative modelling of microcalcification detection in digital mammography”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1999, pp. 32–41.
- [166] Ripley, B. D. “Modelling spatial patterns”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1977), pp. 172–212.
- [167] Rolland, J. and Barrett, H. H. “Effect of random background inhomogeneity on observer detection performance”. In: *JOSA A* 9.5 (1992), pp. 649–658.
- [168] Rose, S. D., Sanchez, A. A., Sidky, E. Y., and Pan, X. “Investigating Simulation-Based Metrics for Characterizing Linear Iterative Reconstruction in Digital Breast Tomosynthesis”. In: *arXiv preprint arXiv:1706.00405* (2017).
- [169] Rycroft, C. “Voro++: A three-dimensional Voronoi cell library in C++”. In: *Lawrence Berkeley National Laboratory* (2009).
- [170] Sachs, M. B., Nachmias, J., and Robson, J. G. “Spatial-frequency channels in human vision”. In: *JOSA* 61.9 (1971), pp. 1176–1186.
- [171] Salvagnini, E., Bosmans, H., Van Ongeval, C., Van Steen, A., Michielsen, K., Cockmartin, L., Struelens, L., and Marshall, N. W. “Impact of compressed breast thickness and dose on lesion detectability in digital mammography: FROC study with simulated lesions in real mammograms”. In: *Medical Physics* 43.9 (2016), pp. 5104–5116.
- [172] Salvagnini, E., Lemmens, K., Bosmans, H., Struelens, L., and Marshall, N. W. “Model observer detectability as a substitute for contrast detail analysis in routine digital mammography quality control”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2013, pp. 866821–866821.
- [173] Schmitz, C. and Hof, P. “Design-based stereology in neuroscience”. In: *Neuroscience* 130.4 (2005), pp. 813–831.
- [174] Sechopoulos, I. “A review of breast tomosynthesis. Part I. The image acquisition process”. In: *Medical physics* 40.1 (2013).
- [175] Shafer, C. M., Seewaldt, V. L., and Lo, J. Y. “Validation of a 3D hidden-Markov model for breast tissue segmentation and density estimation from MR and tomosynthesis images”. In: *Biomedical Sciences and Engineering Conference (BSEC), 2011*. IEEE. 2011, pp. 1–4.
- [176] Shaheen, E., Van Ongeval, C., Zanca, F., Cockmartin, L., Marshall, N., Jacobs, J., Young, K. C., R Dance, D., and Bosmans, H. “The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis”. In: *Medical physics* 38.12 (2011), pp. 6659–6671.
- [177] Shao, J. and Tu, D. *The jackknife and bootstrap*. Springer Science & Business Media, 2012.

- [178] Shepherd, J. A., Herve, L., Landau, J., Fan, B., Kerlikowske, K., and Cummings, S. R. "Novel use of single X-ray absorptiometry for measuring breast density". In: *Technology in Cancer Research & Treatment* 4.2 (2005), pp. 173–182.
- [179] Siddiqui, M. "Some problems connected with Rayleigh distributions". In: *J. Res. Nat. Bur. Stand D* 60 (1962), pp. 167–174.
- [180] Siegel, R. L., Miller, K. D., and Jemal, A. "Cancer statistics, 2016". In: *CA: a cancer journal for clinicians* 66.1 (2016), pp. 7–30.
- [181] Siewerdsen, J., Cunningham, I., and Jaffray, D. "A framework for noise-power spectrum analysis of multidimensional images". In: *Medical physics* 29.11 (2002), pp. 2655–2671.
- [182] Sirinukunwattana, K., Snead, D. R., and Rajpoot, N. M. "A stochastic polygons model for glandular structures in colon histology images". In: *IEEE transactions on medical imaging* 34.11 (2015), pp. 2366–2378.
- [183] Skaane, P., Bandos, A. I., Eben, E. B., Jebsen, I. N., Krager, M., Haakenaasen, U., Ekseth, U., Izadi, M., Hofvind, S., and Gullien, R. "Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images". In: *Radiology* 271.3 (2014), pp. 655–663.
- [184] Skaane, P., Bandos, A. I., Gullien, R., Eben, E. B., Ekseth, U., Haakenaasen, U., Izadi, M., Jebsen, I. N., Jahr, G., Krager, M., et al. "Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program". In: *Radiology* 267.1 (2013), pp. 47–56.
- [185] Skaane, P., Bandos, A. I., Gullien, R., Eben, E. B., Ekseth, U., Haakenaasen, U., Izadi, M., Jebsen, I. N., Jahr, G., Krager, M., et al. "Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration". In: *European radiology* 23.8 (2013), pp. 2061–2071.
- [186] Souchay, H., Durand, J., Colin, C., and Lefin, P. "Digital compensation of x-ray detector MTF for mammography". In: *Proc. Int. Workshop on Digital Mammography*. 2004.
- [187] Spangler, M. L., Zuley, M. L., Sumkin, J. H., Abrams, G., Ganott, M. A., Hakim, C., Perrin, R., Chough, D. M., Shah, R., and Gur, D. "Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: a comparison". In: *American Journal of Roentgenology* 196.2 (2011), pp. 320–324.
- [188] Tagliafico, A., Tagliafico, G., Astengo, D., Airaldi, S., Calabrese, M., and Housami, N. "Comparative estimation of percentage breast tissue density for digital mammography, digital breast tomosynthesis, and magnetic resonance imaging". In: *Breast cancer research and treatment* 138.1 (2013), pp. 311–317.
- [189] Thiedmann, R., Stenzel, O., Spettl, A., Shearing, P. R., Harris, S. J., Brandon, N. P., and Schmidt, V. "Stochastic simulation model for the 3D morphology of composite materials in Li-ion batteries". In: *Computational Materials Science* 50.12 (2011), pp. 3365–3376.
- [190] Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. "Global cancer statistics, 2012". In: *CA: a cancer journal for clinicians* 65.2 (2015), pp. 87–108.

- [191] Tu, Z. and Zhu, S.-C. “Image segmentation by data-driven Markov chain Monte Carlo”. In: *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002), pp. 657–673.
- [192] Underwood, E. E. “The mathematical foundations of quantitative stereology”. In: *Stereology and Quantitative Metallography*. ASTM International, 1972.
- [193] Utasi, A. and Benedek, C. “A 3-D marked point process model for multi-view people detection”. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE. 2011, pp. 3385–3392.
- [194] Van Metter, R. L., Beutel, J., and Kundel, H. L. “Handbook of Medical Imaging, Volume 1: Physics and Psychophysics”. In: SPIE. 2000.
- [195] Van Peteghem, N., Salvagnini, E., Bosmans, H., Cockmartin, L., and Marshall, N. W. “A comparison of mammographic systems for different breast thicknesses using model observer detectability”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2015, pp. 941235–941235.
- [196] Vanmarcke, E. *Random fields: analysis and synthesis*. World Scientific Publishing Co Inc, 2010.
- [197] Verdié, Y. and Lafarge, F. “Detecting parametric objects in large scenes by Monte Carlo sampling”. In: *International Journal of Computer Vision* 106.1 (2014), pp. 57–75.
- [198] Wallis, M. G., Moa, E., Zanca, F., Leifland, K., and Danielsson, M. “Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study”. In: *Radiology* 262.3 (2012), pp. 788–796.
- [199] Warren, L. M., Given-Wilson, R. M., Wallis, M. G., Cooke, J., Halling-Brown, M. D., Mackenzie, A., Chakraborty, D. P., Bosmans, H., Dance, D. R., and Young, K. C. “The effect of image processing on the detection of cancers in digital mammography”. In: *American Journal of Roentgenology* 203.2 (2014), pp. 387–393.
- [200] Warren, L. M., Mackenzie, A., Cooke, J., Given-Wilson, R. M., Wallis, M. G., Chakraborty, D. P., Dance, D. R., Bosmans, H., and Young, K. C. “Effect of image quality on calcification detection in digital mammography”. In: *Medical physics* 39.6 (2012), pp. 3202–3213.
- [201] Welch, P. “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms”. In: *IEEE Transactions on audio and electroacoustics* 15.2 (1967), pp. 70–73.
- [202] Wertheimer, M. “Untersuchungen zur Lehre von der Gestalt”. In: *Psychological Research* 1.1 (1922), pp. 47–58.
- [203] Wigati, K. T., Cockmartin, L., Marshall, N. W., Soejoko, D. S., and Bosmans, H. “Comparison of two psychometric functions in analyzing 4-AFC detection results using a task-based structured phantom for digital mammography”. In: *Physica Medica: European Journal of Medical Physics* 32 (2016), pp. 215–216.
- [204] Wu, G., Mainprize, J. G., and Yaffe, M. J. “Spectral analysis of mammographic images using a multitaper method”. In: *Medical physics* 39.2 (2012), pp. 801–810.
- [205] Xia, J. Q., Lo, J. Y., Yang, K., Floyd, C. E., and Boone, J. M. “Dedicated breast computed tomography: Volume image denoising via a partial-diffusion equation based technique”. In: *Medical physics* 35.5 (2008), pp. 1950–1958.

- [206] Zhang, Y., Abbey, C. K., and Eckstein, M. P. “Adaptive detection mechanisms in globally statistically nonstationary-oriented noise”. In: *JOSA A* 23.7 (2006), pp. 1549–1558.
- [207] Zhang, Y., Pham, B. T., and Eckstein, M. P. “The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds”. In: *IEEE Transactions on Medical Imaging* 25.10 (2006), pp. 1348–1362.
- [208] Zwigelaar, R., Planiol, P., Marti, J., Marti, R., Blot, L., Denton, E. R., and Rubin, C. M. “EM texture segmentation of mammographic images”. In: *Digital Mammography*. Springer, 2003, pp. 223–227.

Titre : Optimisation de l'imagerie tridimensionnelle du sein, basée sur les tâches du radiologue, par l'utilisation d'observateurs mathématiques.

Mots Clefs : imagerie du sein par rayons X, modèle de texture 3D du sein, observateurs mathématiques, études cliniques virtuelles, géométrie stochastique, théorie *a contrario*

Résumé :

La mammographie, une modalité 2D d'imagerie du sein par rayons X, a montré son efficacité pour réduire la mortalité par cancer du sein. Aujourd'hui, la tomosynthèse numérique du sein, une modalité 3D d'imagerie du sein par rayons X, prend une place de plus en plus importante dans la pratique clinique, et est reconnue de plus en plus souvent comme ayant le potentiel de remplacer la mammographie dans un proche avenir. Pour évaluer plusieurs aspects de la tomosynthèse, des études cliniques sont nécessaires. Mais les études cliniques sont coûteuses et présentent des risques supplémentaires pour les patientes dus à l'utilisation de radiations ionisantes. Les études cliniques virtuelles ont pour objectif d'offrir une approche alternative en utilisant des simulations numériques. Dans cette thèse, nous nous intéressons à plusieurs éléments intervenants dans un telle étude clinique virtuelle. Dans un premier temps, nous analysons l'état de l'art sur la caractérisation analytique des champs aléatoires 3D pour la simulation de la texture du sein par rayons X. Nous nous intéressons aussi à l'estimation de caractéristiques statistiques des images du sein par rayons X (densité, indice spectral). Puis nous développons un nouveau modèle de texture 3D du sein basé sur la géométrie stochastique, et qui permet de simuler des images 2D et 3D réalistes du sein. Nous considérons le problème de l'inférence d'un tel modèle à partir d'une base d'images cliniques 3D. Ensuite, nous développons un observateur mathématique basé sur la théorie *a contrario* de la perception visuelle, pour modéliser la détection des microcalcifications par des radiologues dans des images 2D et 3D du sein. Tous ces composants sont utilisés pour implémenter une étude clinique entièrement numérique. La pertinence des résultats obtenus montre l'utilité de ces études cliniques virtuelles et nous incite à en développer de plus élaborées dans le futur.

Title : Task-based optimization of 3D breast x-ray imaging using mathematical observers

Keys words : breast x-ray imaging, 3D breast texture model, mathematical observers, virtual clinical trials, stochastic geometry, a contrario theory

Abstract : Full field digital mammography, a 2D x-ray breast imaging modality has been proved to reduce the breast cancer mortality. Today, digital breast tomosynthesis, a 3D x-ray breast imaging modality, is being integrated in clinical practice and is believed to replace standard mammography in the near future. To assess the clinical performance of various aspects of tomosynthesis, clinical trials are needed. Clinical trials are burdensome, expensive and may impose increased risk to the patient due to additional radiation exposure. Virtual Clinical Trials aim to offer a more efficient alternative by using computational components. Today, active research is ongoing to develop computational components dedicated to 2D and 3D breast imaging, especially to 3D tomosynthesis. This thesis aims to advance several aspects in the development of Virtual Clinical Trials. First, we focused on analytical characterization of state-of-the-art 3D random field breast texture models. The estimation of statistical characteristics (breast density, spectral index) from clinical x-ray breast images was also studied. Next, we proposed a mathematically tractable 3D breast texture model based on stochastic geometry, that allows to simulate realistic 2D and 3D images. The statistical inference of the texture model parameters from a database of clinical 3D breast images was also tackled. We then developed a mathematical observer based on the *a contrario* theory, that allows to model the microcalcification detection process by radiologists in 2D and 3D breast images. Finally, these two proposed components were applied to implement a virtual clinical trial experiment, demonstrating their potential in the conduct of more advanced virtual clinical studies in the future.

