Responses to PhD thesis review by Professor Xavier Descombes and Professor Stephen Glick

Questions and comments from Professor Xavier Descombes

- *Chapter 3. Il aurait également été intéressant de comparer les simulations obtenues avec un modèle de Gibbs binaire tel que le modèle d'Ising. (It would also be interesting to compared simulations of the proposed texture model with binary Gibbs models, such as the Ising model.)*

We agree with the comment. We also think that the Isling might be a good approach to model breast textures. One of our intention of developing a stochastic geometric texture model is to apply the stochastic geometry, in particular marked point process theory for the modeling of breast textures. That was the reason why we chose such an approach to develop our model. The following sentence was added to the conclusion section of chapter 3 (page 78, paragraph 2).

> "[...]. 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."

- *Chapter 4. Une validation sur des données de synthèse aurait permis de mieux évaluer l'erreur commise. Par ailleurs, on pourrait s'interroger sur l'influence de l'initialisation. (A validation of the inference from reconstruction algorithm using synthetic input data may allow for a better evaluation of the reconstruction error. We could also challenge the effect of the initialization step of the algorithm.)*

We totally agree with the remarks. As stated in the future research directions section, we discussed the fact that some further validation of the inference from reconstruction approach using synthetic input should be performed. The following sentence was added to the conclusion section of chapter 4 (page 101, paragraph 2).

> "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, [...]"

Questions and comments from Professor Stephen Glick

- *Page 36. There are some additional references that should be recognized here. In particular, Christian Graff at the FDA is developing a somewhat similar model-based phantom approach [SPIE Medical Imaging 2016]. Some discussion of how the proposed breast texture phantom differs from the Graff model should be included. Another similar model, referred as the DeBra phantom was published by (Ma, A., S. Gunn, and D.G. Darambara, Introducing DeBRa: a detailed breast model for radiological studies. Phys Med Biol, 2009. 54: p. 4533-4545.)*

The manuscript version you reviewed references Graff's phantom on page 36 (Reference [75]). We now also reference the Ma *et al.* paper describing the DeBRa phantom on page 35, paragraph "Model-based phantoms". The following paragraphs were added after paragraph 1 of the conclusion of chapter 3 (page 77).

> "The fibroglandular tissue model proposed by Graff 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 fibro-glandular 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-fibroglandular 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."

- *Another approach that should be discussed under empirical-based phantoms is using tissue specimens to make phantoms. This approach eliminates a number of limitations typically present when using patient data. Please see O'Connor, J.M., et al., Generation of voxelized breast phantoms from surgical mastectomy specimens. Med Phys, 2013. 40(4).*

We added this reference in the "Empirical-based phantoms" section (page 35).

- *Another phantom generation approach that should be discussed is to combine elements of the empirical and model-based approaches. Please see Chen, X., et al., Hybrid computational breast phantom combining high-resolution, procedurally-generated structures with patient-based anatomy. SPIE Medical Imaging 2017.*

The paragraph on "Model-based phantoms combined with 3D random fields" (page 35) was modified as follows. This paragraph was moved after the paragraph "empirical based phantom".

"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 or be thresholded to binary volumes 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. 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."

- *Fig. 2.3 is an interesting result. Metheany et al (Med Phys, 35(10), 2008) have also studied this problem and come up with a similar result. However, they do not observe the deviation from the theory at low beta. More discussion on how this work differs from Metheany et al should be provided.*

We believe that the difference in results between our study and that of Metheany *et al*. is due to the difference in the definition of the 3D power-law Gaussian random field. Metheany *et al*. defined the 3D power-law Gaussian random field in the continuous domain, with infinite support. In our study, the 3D power-law Gaussian random field was defined in a bounded discrete domain. The difference in definition modifies the limits of integration when deriving the relationship between the spectral density of the 3D random field and its 2D slice. We believe that both definitions are valid. However, our definition allows for a more straightforward correlation between analytical results and simulation experiments with 3D power-law textures which have a finite boundary.

- *Page 60. It is stated "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 (bCT) images". However, it is well-known (i.e., published), that the Boone bCT data used here had limited spatial*

resolution compared to other breast CT data. This is because the bCT patient data was collected with continuous x-ray exposure (not x-ray pulsed). Thus, the small-scale morphology might be diminished in this dataset. This should be discussed, especially since the bCT data here is being used as ground truth. This is the motivation for Chen's work on hybrid computational phantoms discussed above.

Thank you for pointing this out. We totally agree with your remarks. We added an additional paragraph on page 64, paragraph 4.

"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 fibroglanular tissue. In addition, a continuous exposure leads to motion blur image artifacts; indeed, short exposures are needed to obtain sharp images. "

This is the reason why we did not use the Boone breast CT data as the full ground truth to quantify the small-scale aspect in our texture model. Despite that, although our small-scale Voronoi model, aiming to mimic the high-frequency aspect on the fibroglandular tissue boundaries, might not be entirely accurate, the fibroglandular tissue boundaries in simulated FFDM and DBT images have a reasonable high visual realism compared to real clinical images. This demonstrates that our model is more than reasonably acceptable for the spatial resolution characteristics encountered in FFDM/DBT.

- *Page 63. In the simulation of DBT, it is not clear how the mu values were assigned.*

We added on page 62, paragraph 3 describing how mu values were assigned to the adipose and fibroglandular components.

"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."

- *Fig. 3.3. There appears to be some visual artifacts in the simulated data (especially seen in BI-RADS 1 images). These should be discussed.*

The images in Fig 3.3 are simulated using breast CT volumes segmented with an early version of our fibroglandular/adipose tissue segmentation algorithm. This early version does not properly handle noise reduction, leaving some small residuals in the segmented breast CT volume. The image artifacts are due to these residuals. For the remainder of the studies described in the manuscript, the breast CT volumes were segmented using an updated version of the segmentation algorithm, handling noise residuals in a more proper fashion.

- *Page 65, In the paragraph discussing histological images, you might refer to the O'Connor reference given above.*

We added this reference on page 65.

- *Section 3.3. The bCT VOIs have isotropic voxels of 0.2 – 0.4 mm. This is considerably higher than the detector element size (0.1 mm?). In order to minimize the null space, phantom voxels usually should be smaller than the detector element size. Fig. 3.9. Again, I am wondering about loss of fine details in the clinical bCT slices due to limited spatial resolution of the bCT clinical data.*

We agree with the remark on the loss of fine details in clinical breast CT data due to limited spatial resolution. We considered the idea to upsample the voxels so as to improve the fine detail of the breast CT data. However, we rejected this idea since upsampling might introduce a bias due to the modification of the original breast CT data.

- *The 3D stochastic solid breast texture model does not include the ductal tree that some model-based phantoms include. This omission should be discussed. Does the author think it is necessary to model the ductal tree?*

The goal of this study was restricted to design a model allowing to realistically simulate the morphology and distribution of small and medium scale fibroglandular and intra-glandular adipose tissue. Modeling of other anatomical structures occurring in the fibroglandular zone of the breast, such as the ductal trees and blood vessels, was out of the scope of our study. We realize that by including these details, the model's realism will increase and give images with more and different types of detail.

- *Section 4.3.3. When BI-RADS density categories are discussed, it is not clear what edition is being referred to. Edition 4 and 5 are quite different. In fact, Edition 5 eliminates any reference to percentages.*

We refer to the 4th edition. On page 90, we added a reference to the 4th edition.

- *Chapter 5 of the Thesis focuses on two issues of importance in statistically characterizing breast images; 1) breast density, and 2) estimation of the parameter β. I believe this is the weakest chapter in the Thesis. Although estimating breast density is certainly an important topic, there are some questionable assumptions that are made in the proposed model-based stereology methods and analysis. In particular, the approach requires a priori information of the thickness of the breast. Although, this information might be reported in the DICOM header, it is not always accurate. Furthermore, some parts of the breast near the nipple have less thickness. The analysis uses BI-RADS density categories to simulate density (see comment above on new edition of BI-RADS), however, simulated volumes are stationary. The masking effect of mammography has been getting more important of late, and because of this, I would have liked to see this method analyzed for estimating BI-RADS 5th categories.*

We agree with these limitations. Paragraph 2 of the conclusion section of chapter 4 (page 126) was modified as follows.

> "[...] 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. [....] Finally, the analysis and simulations performed in this study were according to the 4th edition of BI-RADS density classification. The recently published 5th 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 5th edition of BI-RADS."

- *Page 134. In computing the NFA image, mean and standard deviation of the values of pixels in Vx are computed. However, unlike the CHO, the covariance of the noise is not considered. The implications of this should be discussed. Is the assumption that human observers do not use noise correlations in their decision?*

We agree with the comment on the noise correlation. This is our first attempt to develop an *a contrario* model observer for microcalcification detection. As a first step, we considered a straightforward white Gaussian noise formulation of the naïve model for the *a contrario* detector, where the noise correlation was not considered. As mentioned in the future research directions section (page 165, second paragraph), in a

next step, we will investigate more complex naïve model such as the power-law Gaussian random field, which takes into account the noise correlation.

- *As mentioned in the Conclusions, calcifications are usually present in clusters. It's not clear what the implications are if multiple calcs are present in the Vx neighborhood.*

We agree with the remark. The currently proposed *a contrario* model observer assumes that the microcalcifications are isolated. It does not encounter the detection of clusters of microcalcifications. As we pointed out in the directions of future research section (page 164, third paragraph), we aim to address the cluster detection aspect in a future study.

- *Page 139. It's not clear how the k parameter was selected. From images in Fig. 6.3, it appears as if detectability performance would be high (i.e., calcs are of high contrast). Also, in Fig. 6.3 it appears as if calcs are quite separated (i.e., some by more than 1 cm).*

We agree with the comment. The experimental parameters were chosen empirically so that the generated inserts were visible. The goal of the experiment was to validate the analytically derived false positive rate of the *a contrario* observer. Therefore, we did not try to vary the k parameter to the level of minimum detection threshold, since this will mainly affect the true positive rate. However, we acknowledge that it is also interesting to decrease the k parameter value so as to test the true positive rate of the *a contrario* observer. The following sentences are added to the conclusion of chapter 6 (page 141).

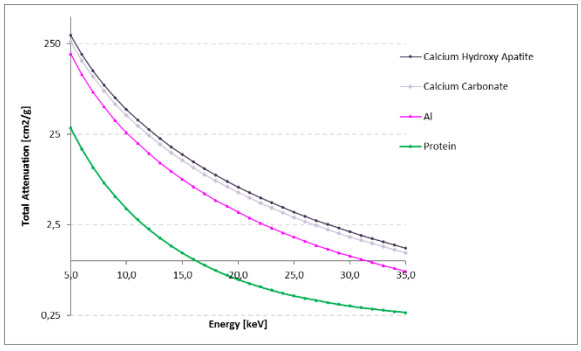
> "[…] The design of the experiment to validate the proposed a contrario observer also has limitations. The goal of the experiment was to validate the analytically demonstrated false positive control of the a contrario observer. It would also be interesting to test the true positive rate of the a contrario observer. "

The simulated microcalcification are randomly distributed across the image. As mentioned in the conclusion chapter (page 164, at the bottom), the proposed a contrario observer aims to detect isolated microcalcifications, we did not aim to simulate realistic distribution of microcalcification clusters.

- *Page 144. Seems like phantoms have unrealistic high-density. For example, the BI-RADS 3 texture phantoms have average glandular density of 55% (e.g., see Yaffe et al, "The myth of the 50-50 breast", Med Phys 2009).*

We agree with the remark. The volumetric density of our texture phantoms is higher than encountered in clinical breast images. This can be attributed to the fact that we only modeled fibroglandular and intra-glandular adipose tissue, and we did not model the subcutaneous adipose tissue layers surrounding the fibroglandular and intra-glandular adipose tissue. When deforming the breast during breast compression, subcutaneous adipose tissue appears as two layers, one on the upper and one on the lower breast bound. These layers may be up to approximately 2 cm thickness. The subcutaneous adipose tissue reduces the volumetric breast density. In future research, we will address this shortcoming.

- *Based on Warren et al "Comparison of the x-ray attenuation properties of breast calcifications, aluminum, hydroxyapatite and calcium oxalate", Med Phys 2013, hydroxyapatite (the primary component of microcalcification) is much more attenuating than Aluminum. Thus, it is questionable why the author used less attenuation than Al. On the other hand, the density of Al (2.72 kg/m) is probably too high. I understand the need for making the task challenging to allow differentiation between modalities. My guess is that the background known exactly task (i.e, no search) required a lower-contrast signal. Some discussion on this should be provided.*



As mentioned on page 144, microcalcifications contain a mineral component (typically calcium hydroxyapatite or calcium carbonate) that is more or less compactly embedded in a protein matrix *[Baker et al. ("New relationships between breast microcalcifications and cancer". British journal of cancer 2010)]*. The ensemble has a lower mass attenuation coefficient than the mass attenuation coefficient of the mineral part only. In preliminary experiments, we found that modeling µcalcs as pure calcium hydroxyapatite or calcium carbonate resulted in unrealistically high attenuation specs. Therefore, we choose to model microcalcifications with lower attenuation properties in part to obtain spec images with image contrast resembling real microcalcifications. But, also in part – as you mention - to produce a lower image contrast so to obtain a detection performance that allows to better discriminate between the imaging modalities. This means that detection performance is somewhere in the middle between 0% and 100% (perfect) detection performance.

- *Page 148, "All projections were done without x-ray noises ….". Should be x-ray noise.*

Thank you for pointing this out. This error was corrected on page 148.

- *Page 148. "This is step illustrate …".*

Thank you for pointing this out. This error was corrected on page 148.

- *Page 150. Since this is an academic thesis, more details on what the GE software is doing should be provided (i.e., regarding eContrast, FineView, and ASiR-DBT).*

We now provide additional details of the eContrast, FineView and ASIR-DBT algorithms (page 150 – paragraph 1):

> "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. (Souchay et al. "Digital Compensation of x-ray detector MTF for mammography". IWDM 2004.). ASIR-DBT is an iterative reconstruction algorithm. It takes 9 DBT raw projection images as input, and produces 50 closely-spaced slices of 1mm thickness as the output. The algorithm reduces image noise by iteratively comparing the acquired image with a modeled projection."

- *Middle page 150. "In each trial of the task, the observer knew the size/attenuation characteristics of the calc, but did not know the exactly presented calc" This is unclear, please provide more details on what the observer was looking at – maybe a screenshot of the reader software.*

The sentence “In each trial of the task, the observer knew the size/attenuation characteristics of the μcalc, but did not know the exactly presented μcalc” was modified to “In each trial of the task, the observer knew the size/attenuation group of the μcalc.”. The following figure and caption were added in page to show what the observer saw on the screen during a trial.

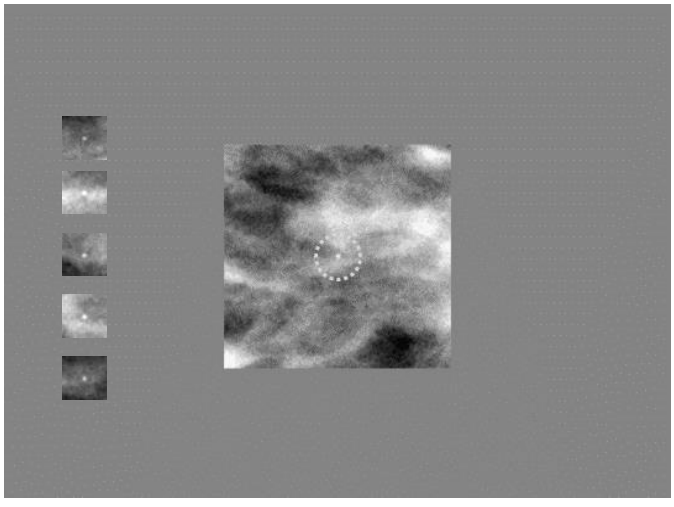


Figure 7.4: Example of an FFDM image displayed in the human observer experiment. The FFDM ROI is 2.5 cm x 2.5 cm. The ROI is surrounded by a uniform gray background with mean-luminance equal to the mean signal intensity value of the FFDM ROI. A 5-mm diameter circle is positioned at the center of the FFDM ROI (which contained in 50% of the trials a μcalc) and aims to direct the visual attention of the observer to where a potential μcalc may be. At the left, five μcalcs of the same attenuation and size category were displayed aiming to indicate the μcalc attenuation and size properties that had to be searched for. Here the ROI contains a μcalc with diameter 600um and attenuation uAL60. FFDM, DBT and S2D images were presented in a similar way.