UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS FÍSICAS
INSTITUTO DE FÍSICA

BREAST GLANDULAR TISSUE VOLUME AND VOLUMETRIC GLANDULAR RATIO
MEASUREMENT USING DUAL-ENERGY MAMMOGRAPHY

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS (FÍSICA MÉDICA)

PRESENTA:
GUSTAVO PACHECO GUEVARA

TUTOR PRINCIPAL

DRA. MARÍA ESTER BRANDAN SIQUÉS
INSTITUTO DE FÍSICA-UNAM

MIEMBROS DEL COMITÉ TUTOR

DR. JORGE ALBERTO MÁRQUEZ FLORES
INSTITUTO DE CIENCIAS APLICADAS Y TECNOLOGÍA-UNAM

DRA. KARLA PAOLA GARCÍA PELAGIO
FACULTAD DE CIENCIAS-UNAM

M.EN.C JORGE PATRICIO CASTILLO LÓPEZ
INSTITUTO NACIONAL DE CANCEROLOGÍA

CIUDAD DE MÉXICO, ENERO 2021

Agradecimientos

Breast Glandular Tissue Volume and Volumetric Glandular Ratio measurement using Dual-Energy Mammography

Resumen. La densidad mamaria se define como la fracción de tejido glandular dentro de la mama, y está asociada a un incremento en el riesgo de padecer cáncer de mama. El objetivo de este trabajo fue desarrollar un método para medir volumen de tejido glandular y densidad mamaria a partir de mamografías digitales de energía dual. Dicho método está basado en el formalismo de descomposición en base de materiales y, a partir de una calibración inical, permite medir volumen de tejido glandular y volumen total de la mama a partir de imágenes mamográficas adquiridas con técnicas de energía dual. El método se aplicó en imágenes clínicas de 14 pacientes provenientes de un estudio multimodalidad. Se evaluó el acuerdo entre las medidas obtenidas con la técnica de energía dual y estudios de resonancia magnética de las mismas pacientes. Se encontró un buen acuerdo entre ambas modalidades para el volumen glandular (prueba de Wilcoxon, $p > 0.05$), Sin embargo, la mamografía por energía dual subestimó el volumen total mamario, lo cual resultó en una sobre-estimación de la densidad mamaria. El método propuesto es computacionalmente barato, fácil de calibrar y produce medidas comparables con las obtenidas en estudios de resonancia magnética.

Introducción

La glándula mamaria está compuesta principalmente por dos tipos de tejido: fibroglandular (denominado simplemente "glandular", en adelante) y adiposo. El término "densidad mamaria" se usa para describir la proporción de tejido glandular dentro de la mama. La evaluación de la densidad mamaria es un problema de interés clínico desde finales de los años sesenta, cuando se empezó a investigar su asociación con el incremento en riesgo de padecer cáncer de mama [1, 2, 24, 25, 5].

Tradicionalmente, la densidad mamaria es evaluada de manera visual por un radiólogo y reportada según el sistema BI-RADS [6]. Esta evaluación es subjetiva y depende del entrenamiento y experiencia de quien la realiza [7, 8]. Debido a esto, la densidad mamaria no puede ser validada como un parámetro cuantitativo ni un indicador de riesgo en sentido estricto.

Para responder a la necesidad de medidas cuantitativas, recientemente se han desarrollado y comercializado herramientas para calcular densidad mamaria a partir de mamografías digitales convencionales (es decir, de energía única). Dos ejemplos de estas herramientas son VolparaDensity (Volpara Solutions, Ltd.) [11] y Quantra (Hologic, Inc.) [12] , ambas validadas para su uso clínico en EUA. Debido a las limitaciones de las imágenes mamográficas convencionales, estas herramientas no son capaces de medir volumen directamente. Otras modalidades de imagen como tomografía computada (CT) de mama, resonancia

magnética (RM), tomosíntesis mamaria y mamografía digital de energía dual (DEDM, por sus siglas en inglés) son capaces de proveer información volumétrica directamente [13].

Medir densidad mamaria usando DEDM es, en esencia, un problema de descomposición de materiales. Para resolver este problema, se combinan dos imágenes adquiridas con diferentes espectros ("baja energía" y "alta energía") para producir mapas bidimensionales (2D) de espesor de tejido, los cuales se integran píxel a píxel para obtener medidas volumétricas. Trabajos como los de Ducote et al [18, 19], y Kappadath et al [20] han explorado la descomposición de materiales en imágenes mamográficas de energía dual. En particular, en los trabajos de Ducote y Molloi [18, 19] se midió densidad mamaria en placas de material equivalente a tejido mamario, reportando el error en esta cantidad así como la dosis impartida por cada adquisición de energía dual. Sin embargo, estos trabajos carecieron de un análisis de incertidumbre, así como de una metodología completa para procesar imágenes clínicas y, finalmente, comparaciones con otras modalidades de imagen.

El objetivo de este trabajo fue desarrollar y evaluar un método para medir densidad mamaria y volúmenes mamarios (de tejido glandular y total) usando DEDM. El método desarrollado se aplicó en imágenes clínicas de 14 pacientes de alto riesgo (con lesiones unilaterales en la mama contralateral a la estudiada), adquiridas con una técnica de energía dual. Los resultados se compararon con los obtenidos al medir estas cantidades usando imágenes de RM de las mismas pacientes.

Material y métodos

Todas las imágenes mamográficas analizadas en este trabajo fueron adquiridas en una estación Amulet Innovality (FUJIFILM Corporation). Los parámetros de adquisición de las imágenes de baja y alta energía fueron los siguientes:

- Baja energía (LE). Potencial del tubo de Rayos-X: 31 kV. Ánodo: Tungsteno. Filtro: Rodio (50µm).
- Alta energía (HE). Potencial del tubo de Rayos-X: 45 kV. Ánodo: Tungsteno. Filtro: Aluminio (700µm) + 5 mm de Al externo.

Se procesaron imágenes de 14 pacientes provenientes de un estudio clínico previo [65]. El criterio de inclusión del protocolo fue sospecha de cáncer de mama multicéntrico. El protocolo incluyó la adquisición de secuencias de RM (con y sin contraste) y un estudio de mamografía contrastada. Este último fue un estudio temporal, en el cual se adquirieron imágenes de la mama sospechosa antes de la inyección de medio de contraste, así como 1, 2, 3, y 4 minutos después de esta. Posteriormente se adquirieron mamografías de energía dual de la mama contra-lateral, las cuales fueron procesadas en este trabajo. Dado que la adquisición fue realizada entre 5 y 6 minutos despúes de la inyección de medio de contraste, no se puede descartar la presencia de trazas del mismo en la mama contralateral. Está pendiente la cuantificación de yodo en las imágenes procesadas, así como su posible efecto en los cálculos presentados. Los estudios de RM fueron adquiridos en un resonador GE de 3T (General Electric Medical Systems, y se usaron secuencias IDEAL (Fat-only) para los cálculos de volumen.

La figura 0.0.1 resume el procedimiento para calcular volumen glandular, total y su cociente, denominado fracción glandular volumétrica o VGR por sus siglas en inglés.
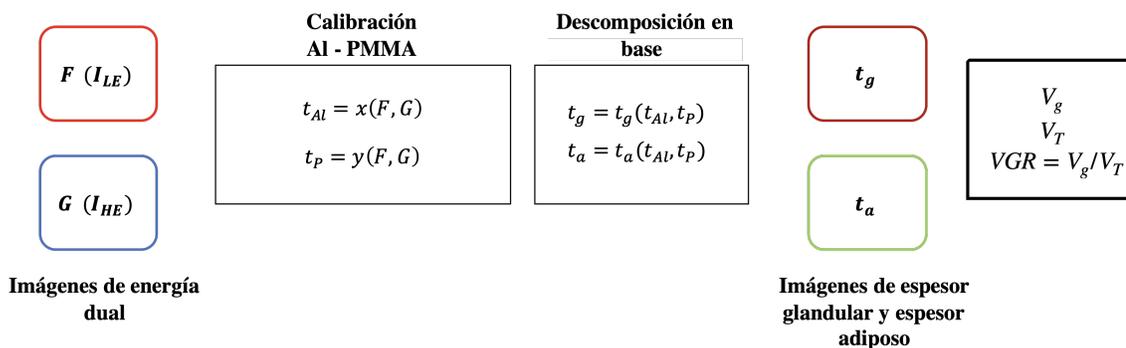
FIGURA 0.0.1. Esquema del método para calcular V_g, V_T y VGR a partir de mamografías de energía dual.

La calibración Al-PMMA consiste en obtener un par de imágenes de energía dual (F, G) de un objeto en forma de escalón, el cual contiene regiones con diferentes combinaciones de espesor de aluminio y PMMA (t_{Al}, t_P). Para un par de imágenes de energía dual y un par de materiales, se puede encontrar una función que relaciona espesor de material con valor de píxel. Usando los datos del conjunto de calibración, se ajustaron las siguientes funciones para recuperar (t_{Al}, t_P) de las imágenes:

$$t_{Al} = \frac{a + bF + cG + dF^2 + eFG + kG^2}{1 + mF + nG} \qquad t_P = aF + bG$$
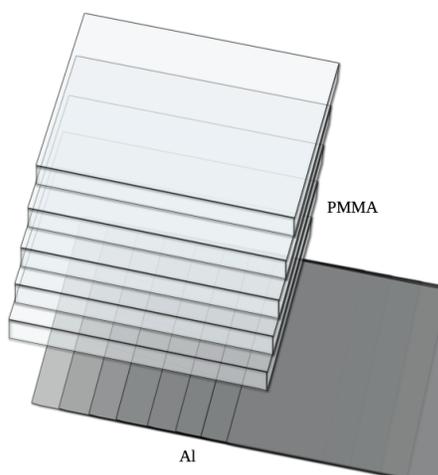
FIGURA 0.0.2. Ilustración del arreglo de calibración.

Los materiales de interés son tejido glandular y tejido adiposo. En ausencia de material tejido equivalente para la calibración, se optó por usar aluminio y PMMA en la calibración.

Para relacionar espesores de Al y PMMA (t_{Al}, t_P) con espesores de tejido glandular y adiposo (t_g, t_a), se usó el formalismo de descomposición en base de materiales [**16**, **17**, **26**]. A partir de este formalismo se obtienen expresiones lineales que relacionan los espesores de ambos pares de materiales. Después de aplicar estas expresiones, se obtienen imágenes -o mapas- de espesor de tejido glandular $t_g(i,j)$, adiposo $t_a(i,j)$ y espesor total $T(i,j) = t_g(i,j) + t_a(i,j)$. Al integrar estas imágenes sobre el área de la mama se obtienen las medidas volumétricas.

Para medir volúmenes a partir de los estudios de RM, se segmentó el tejido adiposo dentro de la mama. Esto se hizo implementando una rutina de umbralaje adaptativo basada en el método de Wellner [**70**, **71**]. El resultado son cortes binarizados, donde un valor de 1 corresponde a un píxel adiposo, y un valor de 0 corresponde a un píxel glandular. Para obtener los volúmenes, se cuenta el número de píxeles y se multiplica por las dimensiones de vóxel.

Resultados y discusión

La figura 0.0.3 muestra ejemplos de los mapas de espesor $t_g(i,j)$, $t_a(i,j)$, $T(i,j)$ obtenidos a partir de un par de imágenes clínicas de energía dual.
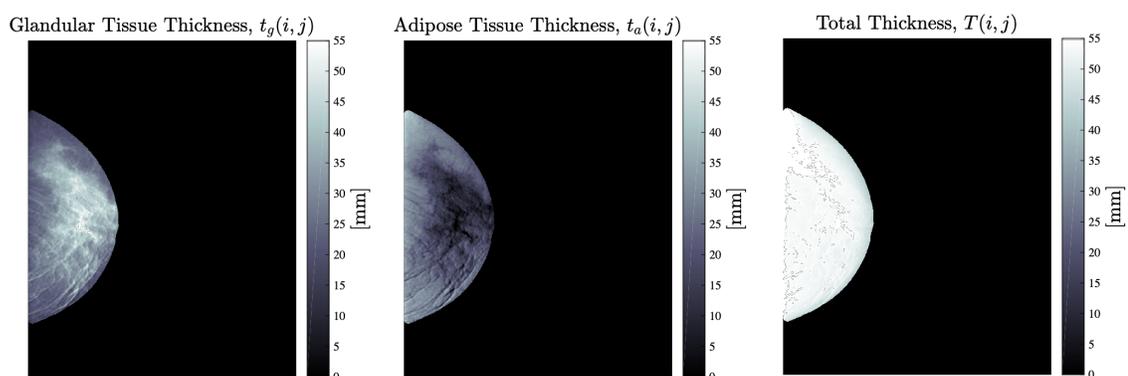
FIGURA 0.0.3. Mapas 2D generados a partir de un conjunto de imágenes de energía dual.

De acuerdo con estudios previos [**42**, **44**, **72**], la densidad mamaria decrece conforme el espesor de la mama comprimida aumenta. La densidad mamaria calculada en este trabajo, VGR_{DEDM}, muestra este mismo comportamiento. También, VGR_{DEDM} disminuye como función del volumen total de la mama, comportamiento que está de acuerdo con estudios de biopsia [**48**, **49**]. Se ha reportado [**35**, **74**] que la densidad mamaria disminuye como función de la edad, pero no se encontró dependencia respecto a esta variable en VGR_{DEDM}. Sin embargo, el tamaño de población de este estudio es demasiado pequeño como para confirmar o descartar definitivamente estas tendencias. Las gráficas en la figura 0.0.4 muestran el comportamiento de VGR_{DEDM} como función del espesor de la mama comprimida T_c y del volumen total de la mama V_T.
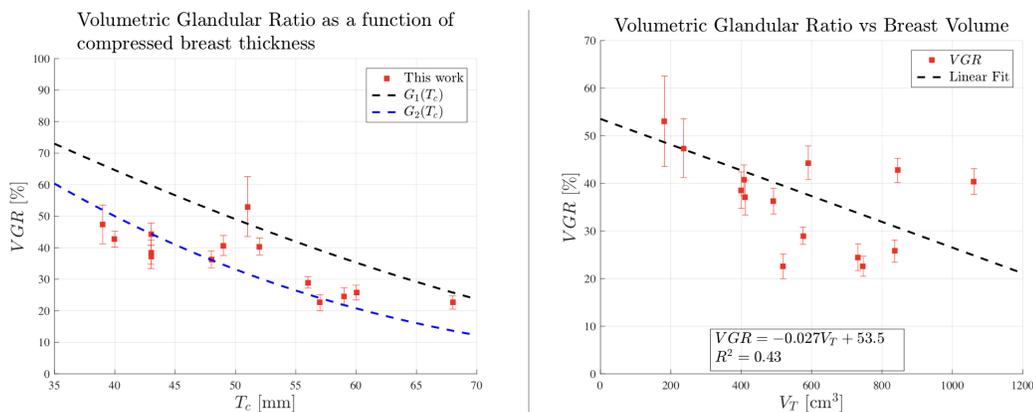
FIGURA 0.0.4. Izquierda: VGR vs espesor de mama comprimida. Las líneas punteadas representan modelos G_1, G_2 que predicen el comportamiento decreciente [**44**]. Derecha: VGR vs V_T. La línea punteada representa un ajuste lineal.

Finalmente, se evaluó tanto la correlación como el acuerdo entre las medidas obtenidas usando DEDM y las obtenidas usando RM. Las figuras 0.0.5 - 0.0.7 muestran diagramas de dispersión y gráficas de Bland-Altman para las tres cantidades calculadas (volumen glandular, volumen total, y VGR). En las gráficas de Bland-Altman se visualizan las diferencias entre ambas técnicas[1] graficadas como función de su promedio. Los límites de acuerdo (LoA) en la gráfica de Bland-Altman se calculan de la siguiente manera:

$$LoA = \bar{\Delta V} \pm 1.96\sigma_{\Delta V},$$

donde $\bar{\Delta V}$ representa el promedio de las diferencias, y $\sigma_{\Delta V}$ su desviación estándar.
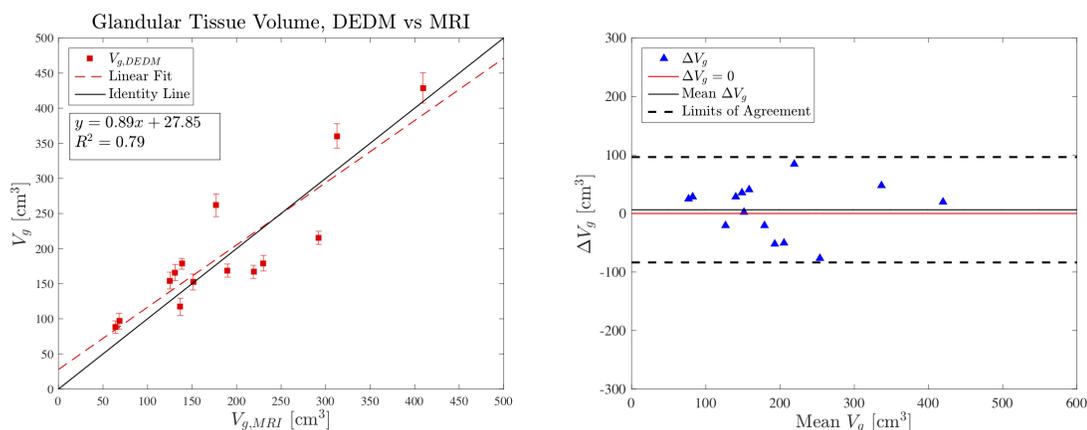
FIGURA 0.0.5. Izquierda: diagrama de dispersión de $V_{g,DEDM}$ vs $V_{g,MRI}$. Derecha: gráfica de Bland-Altman para ΔV_g.

---

[1]Estas diferencias se definieron como la cantidad calculada con RM restada de la cantidad calculada con DEDM. Como ejemplo, la diferencia en volumen glandular es $\Delta V_g = V_{g,DEDM} - V_{g,MRI}$
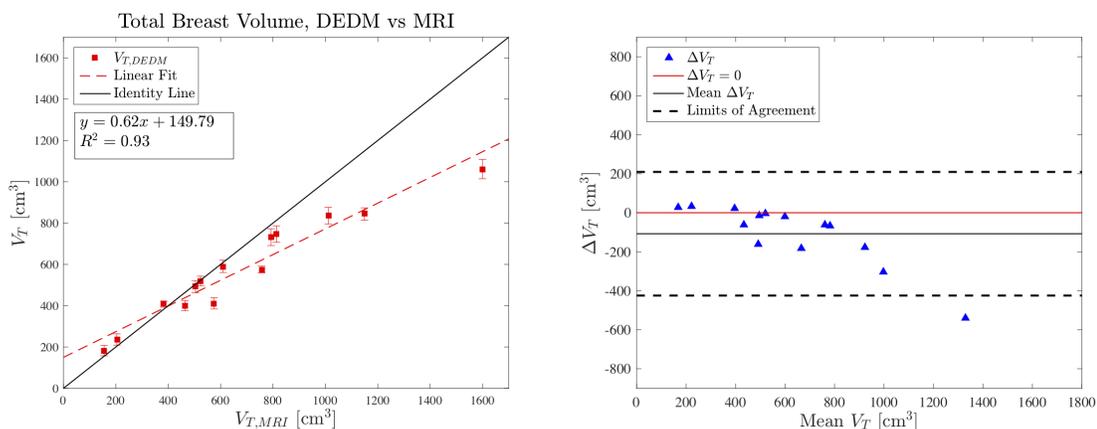
FIGURA 0.0.6. Izquierda: diagrama de dispersión de $V_{T,DEDM}$ vs $V_{T,MRI}$. Derecha: gráfica de Bland-Altman para ΔV_T.
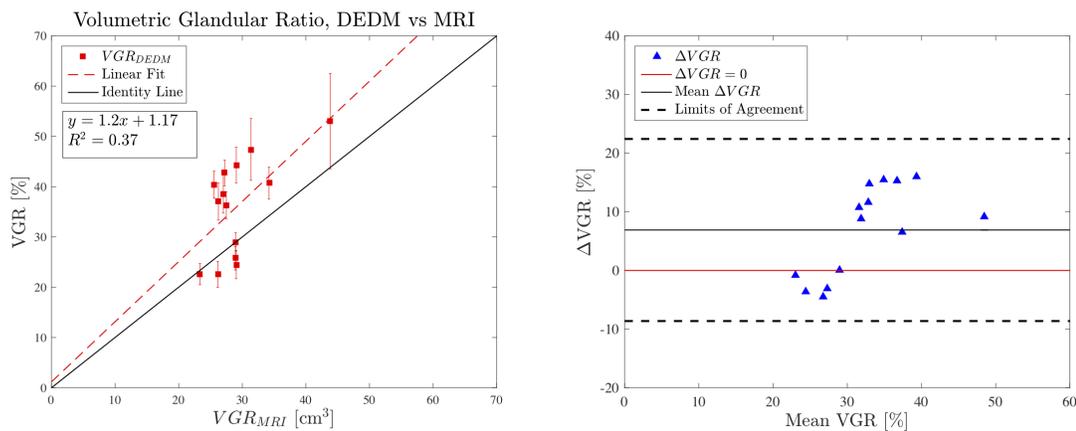
FIGURA 0.0.7. Izquierda: diagrama de dispersión de VGR_{DEDM} vs VGR_{MRI}. Derecha: gráfica de Bland-Altman para ΔVGR.

La correlación entre los resultados de ambas técnicas se evaluó calculando los coeficientes de correlación de Spearman (ρ). Los resultados se muestran en el cuadro 0.0.1. Valores de $p < 0.05$ indican correlaciones estadísticamente significativas con un nivel de confianza de 95 %.

Cantidad	ρ	p
V_g	0.87	< 0.001
V_T	0.98	< 0.001
VGR	0.58	0.03

CUADRO 0.0.1. Resumen de correlaciones entre medidas obtenidas usando DEDM y MRI.

Para V_g y V_T, se encontró una correlación fuerte ($\rho \geq 0.7$) entre las medidas obtenidas usando ambas técnicas. Por otro lado, la correlación encontrada en VGR fue moderada ($0.5 < \rho < 0.7$). En todos los casos, las correlaciones fueron estadísticamente significativas con un nivel de confianza de 95 %.

El cuadro 0.0.2 muestra el promedio, la mediana y el rango intercuartil (IQR) de las diferencias ΔV_g, ΔV_T y ΔVGR.

Cantidad	Promedio	Mediana	IQR
ΔV_g	6.4cm^3	21.9cm^3	55.9cm^3
ΔV_T	-107.6cm^3	-62.2cm^3	170.6cm^3
ΔVGR	6.9%	9.0%	15.6%

CUADRO 0.0.2. Resumen del acuerdo entre medidas obtenidas usando DEDM y MRI.

Para determinar si las diferencias fueron estadísticamente significativas, se usó la prueba de Wilcoxon con un nivel de significancia $\alpha = 0.05$. Se encontró que la diferencia en volumen glandular no fue significativa ($p = 0.58$), mientras que las diferencias en volumen total y VGR lo fueron ($p = 0.02$ y $p = 0.01$, respectivamente). Esto permite concluir que DEDM tiende a sub-estimar el volumen total de la mama, lo cual resulta en una sobre-estimación de VGR en comparación con las medidas de RM

La sobre-estimación de la densidad mamaria -respecto a otras modalidades 3D de imagen como RM- es una característica bien reportada de los métodos basados en mamografía [51, 54, 79, 52]. En particular, Tagliafico et al [54] reportan una sobre-estimación promedio de 16.2 % relativa a medidas de RM. En el caso de este trabajo, la sobre-estimación relativa promedio es de 18.5 %. Rahbar y colaboradores [80] compararon volúmenes medidos usando Quantra y Volpara con resultados de RM, encontrando un buen acuerdo en V_T con diferencias significativas en V_g. En contraste, el método descrito en este trabajo muestra un buen acuerdo en V_g , con diferencias significativas en V_T.

Debido a que no hay métodos estandarizados para calcular volúmenes y densidad mamaria usando RM o mamografía (excluyendo las alternativas comerciales), es difícil realizar comparaciones concluyentes entre estudios y modalidades. Por ejemplo, Wang et al [52] reportaron un coeficiente de correlación de 0.38 entre V_g calculada usando Quantra y usando RM. Por su parte, Kontos et al [51] reportaron un coeficiente de 0.15 comparando la misma cantidad medida con Quantra y RM. Citando un último ejemplo, Holland et al [56] reportaron coeficientes de correlación de hasta 0.86 cuando se compara V_g calculado con un método basado en mamografía con RM.

Los resultados y literatura discutidos sugieren que las medidas de volumen y densidad mamaria, provenientes de diferentes modalidades de imagen, no son equivalentes. Para interpretar estas diferencias, se deben tomar en cuenta variables como características del

grupo de pacientes y procedimientos de adquisición al intentar comparar resultados entre estudios y modalidades diferentes.

Conclusiones

Se desarrolló y evaluó un método que permite medir volúmenes de tejido glandular y adiposo en mamografías de energía dual. Como parte del trabajo, se desarrolló una rutina de pre-procesamiento de imágenes mamográficas que mejora los tiempos de cálculo y disminuye el ruido en las imágenes. El método tiene poca complejidad computacional, se puede implementar en equipos de uso personal, y su calibración requiere el uso de materiales comúnmente hallados en la clínica.

El método se aplicó en imágenes de 14 pacientes adquiridas en una unidad comercial, y los resultados se compararon con medidas de RM de las mismas pacientes. Las medidas volumétricas de RM requirieron de la programación de una rutina de segmentación basada en umbralaje adaptativo.

Las medidas de densidad mamaria volumétrica producidas por ambas modalidades se encontraron en intervalos comparables. Se evaluó la correlación y el acuerdo entre ambas modalidades. El nivel de correlación fue de alto ($\rho = 0.87$ y $\rho = 0.98$ para volumen glandular y total, respectivamente) a moderado ($\rho = 0.58$ para densidad mamaria), y hubo un buen acuerdo en ambas medidas de volumen glandular. También se encontró un buen acuerdo para volumen total, pero las diferencias entre ambos métodos aumentan para mamas más grandes. Debido a esto, el método basado en DEDM tiende a sobre-estimar la densidad mamaria en comparación con RM.

Una primer factor limitante del trabajo es la posible presencia de yodo en las mamas analizadas. El objetivo de trabajos futuros será cuantificar su presencia y determinar si afecta o no las medidas volumétricas.

Finalmente, debido a que el número de imágenes procesadas es pequeño y provienen de pacientes de alto riesgo, los resultados no se pueden comparar directamente con otros estudios. En trabajos futuros, se buscará incluir un conjunto mayor de pacientes, así como realizar un análisis de robustez del método.

Contents

CHAPTER 1

Introduction

The term "breast density" dates back to the days of analog mammography, and it was coined to describe the composition of the breast in terms of fatty tissue (dark areas on mammograms) and fibroglandular tissue (bright areas on mammograms). Breast density assessment has been a problem of clinical interest since the late sixties, when its association with an increase in breast cancer risk was first investigated [1].

Since then, many studies have found that a higher breast density -as assessed in screening mammograms- implies an increased risk of breast cancer. The reported increase in risk varies between 1.8- and 6-fold depending on the study [2, 24, 25, 5].

Traditionally, breast density is assessed by a radiologist's eye and categorized using the Breast Imaging Reporting and Database System (BI-RADS). BI-RADS was developed by the American College of Radiology (ACR), and it divides breast density in four categories: I to IV, where I represents an almost fatty breast with less than 25% density, and IV represents a breast with 75% or greater density [6]. This categorization procedure is subjective and depends strongly on the clinician's training and expertise, which introduces a degree of intra- and inter-observer variability to the results[1] [7, 8]. The qualitative nature of this parameter prevents it from being a true risk predictor. As such, there is a current need for quantitative methods for breast density determination.

One first step into a more quantitative direction is the segmentation of fibroglandular tissue from digitized or natively digital mammograms, achieved by the interactive (i.e. human-controlled) thresholding of the image. This thresholding approach is at the heart of many software utilities such as Cumulus (University of Toronto, Canada) and other in-house developed ones. The result is an area-based percentage mammographic density, sometimes referred to as 2D breast density. These methods have two main drawbacks: 1) they still depend on an operator's input to define the thresholds, and as such exhibit a degree of inter- and intra-observer variability [9] and 2) the area-based measurements cannot take into account the 3D distribution of glandular tissue, which results in breast density over-estimation. Furthermore, it is argued that a volumetric breast density measurement would be a better breast cancer risk predictor [10].

Responding to a clinical interest for automated volumetric breast density measurements, tools such as VolparaDensity (Volpara Solutions, Ltd.) [11] and Quantra (Hologic, Inc.) [12] have been developed and commercialized. Both of them use a physics-based approach

---

[1]Ciatto et al [8] found "substantial" intra-observer agreement, reporting an average kappa-statistic value of $\kappa = 0.71\ [0.33, 0.88]$. Inter-observer agreement was found to be "moderate" at best, with an average kappa-statistic value of $\kappa = 0.54\ [0.02, 77]$.

to calculate glandular volume and volumetric breast density using conventional (single-energy) digital mammograms. Given that they don't require any human input to perform calculations, these tools have increased reproducibility and robustness. On grounds of these advantages and the growing interest in providing breast density reports, both of them have been cleared for clinical use by the FDA [6, 13].

As single-energy X-ray imaging only allows for the direct quantification of a single material, the aforementioned tools do not calculate volumes directly; instead, they calculate density as a percentage and use other information such as compressed breast thickness to estimate the total breast volume. From those measurements, the total glandular volume is derived [14]. Since these methods rely upon acquisition parameters, the potential absence or mislabelling of DICOM headers limits their application. Furthermore, as compression is applied the paddle tilts from the thoracic wall to the breast edge, and so the reported value is not representative of the thickness throughout the complete breast area [15]. Because of this, "native" volume measurements are preferable. Imaging modalities such as breast CT (bCT), MRI, digital breast tomosynthesis (DBT), and dual-energy digital mammography (DEDM) are able to provide volumetric information natively.

Breast density measurement using DEDM is, in essence, a material decomposition problem. The goal is to combine two images acquired with different X-ray imaging spectra in order to produce tissue thickness maps (fibroglandular and adipose in this case). These tissue maps can then be integrated over the breast area to produce volumetric measurements. The decomposition of dual energy images into material thickness images has been previously investigated with an emphasis in bone and soft tissue separation [16, 17]; however, as shown in [18, 19, 20] and similar studies, the formalism can also be applied to breast fibroglandular tissue separation and quantification.

The aim of this work was to develop and evaluate a method for calculating breast density and breast glandular volume using dual-energy digital mammography. The presented method is based on the inverse mapping and material decomposition formalisms, and it was applied to clinical images of 14 patients that underwent a multimodality study. The availability of MR studies from the same patients enabled us to assess the agreement between these two modalities. The proposed DEDM-based method is computationally inexpensive, fully automatic, and only requires an initial calibration using readily available materials in a clinical setting.

CHAPTER 2

Background

2.1. Dual-Energy mammography

Dual-energy X-ray imaging is based on the energy dependence of each material's attenuation properties. Figure 2.1.1 shows the energy dependence of the linear attenuation coefficient, μ, for materials of mammographic interest.
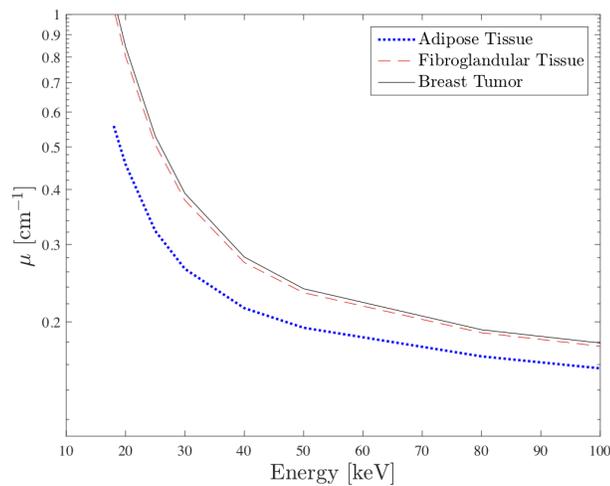
FIGURE 2.1.1. Linear attenuation coefficients for materials of mammographic interest. Data taken from [21]

Dual energy images are acquired using two different spectra, usually referred to as low-energy (LE) and high-energy (HE). The spectral features of each X-ray beam are dictated by the X-ray tube's potential (kV) and anode/filter combination, as well as any present external filtering. Figure 2.1.2 shows a pair of calculated spectra representative of those used in this work, while figure 2.1.3 shows mamograms acquired using LE and HE spectra.

Both images have appreciably different contrast characteristics, but their true utility lies in the fact that they can be combined to produce "material-enhanced" images. One such example is contrast-enhanced digital mammography (CEDM), in which a weighed subtraction is performed to generate contrast between iodine-based contrast medium and the surrounding mixture of breast tissue .The weighing factor can be altered to highlight -or eliminate- different materials. These subtraction-based methods provide contrast information, but to truly quantify material presence a calibration is needed [20].
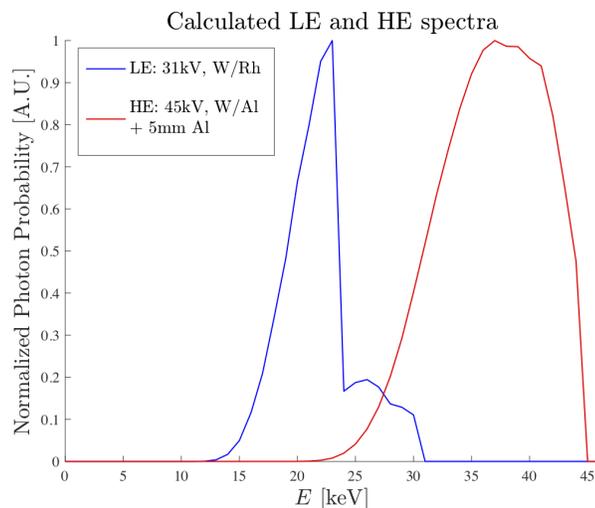
FIGURE 2.1.2. Examples of mammography LE and HE spectra. Tube potential, anode and filter listed in the figure.
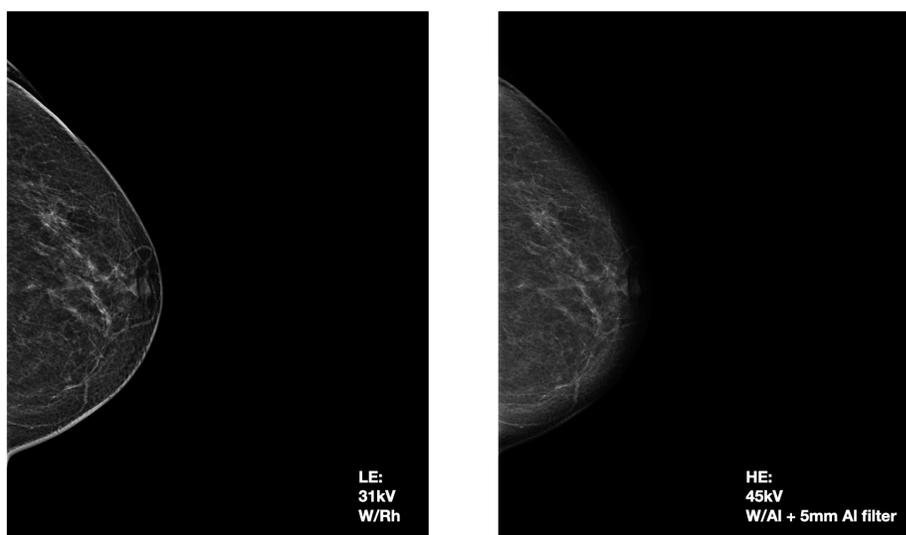
FIGURE 2.1.3. Left: Low-Energy mammogram. Right: High-Energy mammogram. Both images have logarithmic "for display" intensity values, where higher gray values (white) are indicative of greater attenuation.

2.2. Dual-energy material calibration surfaces

Consider a mono-energetic acquisition system with energy E_0 and an imaged object of thickness T, composed of materials a and b with thicknesses t_a and t_b, respectively (figure 2.2.1). Let S be the signal registered by the detector after imaging the previously described object. Assuming simple exponential attenuation, the logarithmic signal intensity registered by the detector, normalized by the background signal S_0 (no object present), can be written as [22]:

$$f = -\ln\left(\frac{S}{S_0}\right) = -\ln\left(\frac{Q(E_0)\Phi_0 e^{-\mu_a(E_0)t_a - \mu_b(E_0)t_b}}{Q(E_0)\Phi_0}\right) = \mu_a(E_0)t_a + \mu_b(E_0)t_b,$$

where $Q(E)$ represents the detector's quantum efficiency (i.e. the signal produced per inciding x-ray photon) and Φ_0, the total incident photon flux.
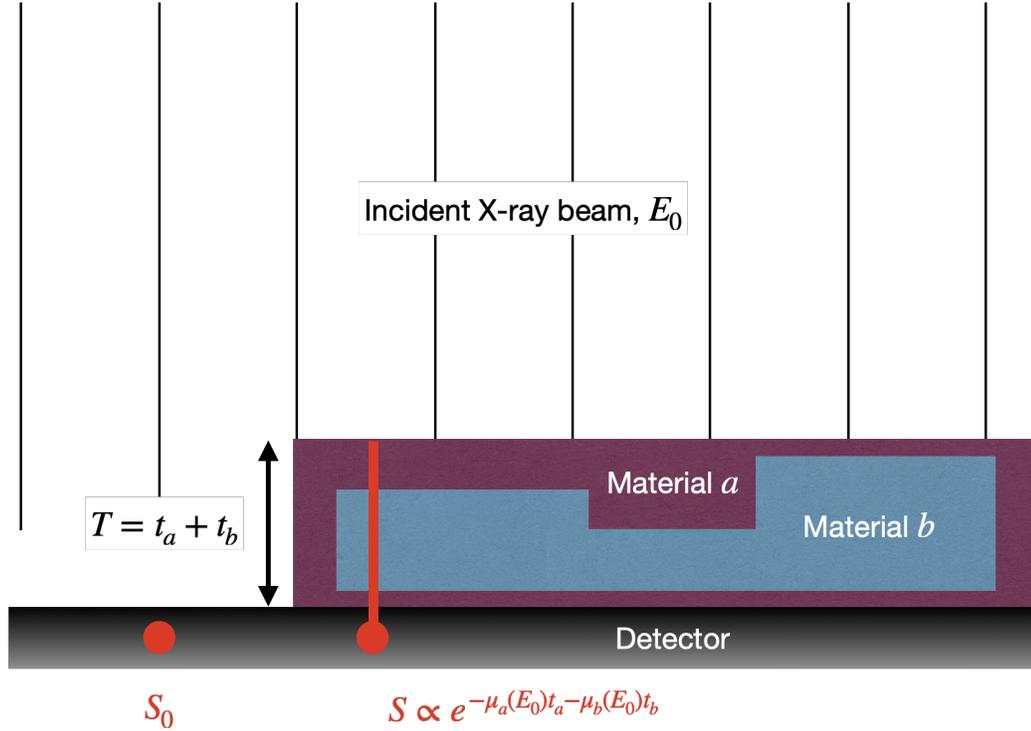
$$f = -\ln(S/S_0) = \mu_a(E_0)t_a + \mu_b(E_0)t_b \quad \boxed{\text{Log, normalized signal}}$$

FIGURE 2.2.1. Schematic representation of the two material and detector X-ray imaging system.

Now, consider independent acquisitions of the same object, using two different beams with single energies E_f (LE) and E_g (HE). The following system of equations for the low- and high-energy log signals arises:

$$f = \mu_a(E_f)t_a + \mu_b(E_f)t_b \ ,$$

$$g = \mu_a(E_g)t_a + \mu_b(E_g)t_b$$

Provided $E_f \neq E_g$, this system can be inverted to find the material thicknesses t_a, t_b. This yields the following expressions:

$$t_a = P_a(f, g) \ ,$$

$$t_b = P_b(f, g) \ ,$$

where P_a and P_b are linear polynomials in f and g; that is, first order surfaces in the (f, g, t_a) and (f, g, t_b) spaces. By imaging materials of known t_a and t_b distributions, and measuring pixel values on the dual-energy images, calibration surfaces can be found.

However, real systems are poly-energetic, and non-linear effects such as beam hardening and scattering are present. Because of this, it's expected that the surfaces P_a and P_b will be non-linear functions of f and g. Cardinal et al [22] first evaluated the performance of different higher-order surfaces, and it was found that the following eight-parameter rational function produced the best calibration results:

$$(2.2.1) \qquad t_i(f, g; a, b, c, d, e, m, n) = \frac{a + bf + cg + df^2 + efg + kg^2}{1 + mf + ng} \ ,$$

where t_i represents material thickness, f and g pixel values in log-normalized HE and LE images, and $\{a, b, c, d, e, k, m, n\}$ are the surface parameters determined by solving the non-linear least squares problem. As the surface equations proposed and evaluated by Cardinal et al make no physical assumptions, they have been used to solve the dual-energy decomposition problem across different mammography systems and techniques [19, 23, 24, 25].

2.3. Basis material decomposition formalism

Ideally, the calibration surfaces discussed in section 2.2 are found by imaging samples of the materials of interest. In many cases such samples are not available, or the calibration phantoms may prove hard to accurately manufacture. The basis material decomposition formalism provides a way to find an equivalence between two pairs of materials in dual-energy images.

This formalism is based on the following proposition: over the diagnostic X-ray energy range, where photoelectric absorption and Compton scattering are the most prevalent photon interactions, one material's mass attenuation coefficient can be written as a linear combination of other two materials' attenuation coefficients [16, 17, 26, 27]. The latter materials are called *basis materials*:

$$\frac{\mu_a}{\rho_a} = k_1 \frac{\mu_\xi}{\rho_\xi} + k_2 \frac{\mu_\eta}{\rho_\eta} \ ,$$

where the subindex a represents the material of interest; ξ and η, the chosen basis materials. k_1 and k_2 are the basis material decomposition coefficients, assumed to be constant throughout the diagnostic X-ray energy range.

The proposed method was calibrated using known aluminum (Al) and acrylic (PMMA) thicknesses. Therefore, Al and PMMA were decomposed in terms of the adipose and glandular tissue base in order to relate both pairs of materials. The Al and PMMA attenuation coefficients are expressed as the following linear combinations:

$$(2.3.1) \qquad \frac{\mu_{Al}}{\rho_{Al}} = a_1 \frac{\mu_g}{\rho_g} + a_2 \frac{\mu_a}{\rho_a} \ ,$$

$$(2.3.2) \qquad \frac{\mu_P}{\rho_P} = b_1 \frac{\mu_g}{\rho_g} + b_2 \frac{\mu_a}{\rho_a} \ ,$$

where the subindexes Al, P, a, g are used to indicate Al, PMMA, adipose tissue, and glandular tissue respectively. The method used to find $\{a_1, a_2, b_1, b_2\}$ is discussed in section 3.4. The expressions relating material thicknesses are derived next.

The following derivation assumes simple exponential attenuation conditions. If an object comprised of glandular tissue thickness t_g -with density ρ_g- and adipose tissue thickness t_a -with density ρ_a- is imaged using two mono-energetic beams with energies L and H, the log-attenuation equations can be written as follows:

$$C_H = t_g \mu_g^H + t_a \mu_a^H \ ,$$

$$C_L = t_g \mu_g^L + t_a \mu_a^L \ ,$$

where C_H and C_L represent the logarithmic attenuation produced by the high and low-energy beams, respectively. Now, let's suppose a combination of Al and PMMA thicknesses produces equivalent attenuations:

$$C_H = t_{Al} \mu_{Al}^H + t_P \mu_P^H \ ,$$

$$C_L = t_{Al} \mu_{Al}^L + t_P \mu_P^L \ .$$

The system

$$(2.3.3) \qquad \begin{cases} t_{Al} \mu_{Al}^H + t_P \mu_P^H = t_g \mu_g^H + t_a \mu_a^H \\[2mm] t_{Al} \mu_{Al}^L + t_P \mu_P^L = t_g \mu_g^L + t_a \mu_a^L \end{cases}$$

has an unique solution (t_{Al}, t_P) for each pair of energies (L, H) provided $L \neq H$. This is to say, there exists an unique combination of (t_{Al}, t_P) that produces the same attenuation as the combination of (t_g, t_a). Substituting 2.3.1 and 2.3.2 in the system 2.3.3:

(2.3.4)
$$\begin{cases} t_g\mu_g^H + t_a\mu_a^H = (A_1 + B_1)\mu_g^H + (A_2 + B_2)\mu_a^H \\ \\ t_g\mu_g^L + t_a\mu_a^L = (A_1 + B_1)\mu_g^L + (A_2 + B_2)\mu_a^L \end{cases}$$

where

$$A_1 = t_{Al}a_1\frac{\rho_{Al}}{\rho_g}; \quad A_2 = t_{Al}a_2\frac{\rho_{Al}}{\rho_a}$$

$$B_1 = t_Pb_1\frac{\rho_P}{\rho_g}; \quad B_2 = t_Pb_2\frac{\rho_P}{\rho_a}.$$

System 2.3.4 can be expressed in matrix form as

$$\mathcal{M}\overrightarrow{T} = \begin{bmatrix} (A_1 + B_1)\mu_g^H + (A_2 + B_2)\mu_a^H \\ (A_1 + B_1)\mu_g^L + (A_2 + B_2)\mu_a^L \end{bmatrix}$$

where $\mathcal{M} = \begin{bmatrix} \mu_g^H & \mu_a^H \\ \mu_g^L & \mu_g^H \end{bmatrix}$ is the matrix of tissue attenuation coefficients, and $\overrightarrow{T} = \begin{bmatrix} t_g \\ t_a \end{bmatrix}$ is the thickness vector for which we wish to solve. This system has the following unique solution, provided $\mathcal{M}$ is non-singular $(L \neq H)$:

(2.3.5)
$$t_g = A_1 + B_1 = t_{Al}a_1\frac{\rho_{Al}}{\rho_g} + t_Pb_1\frac{\rho_P}{\rho_g} ,$$

$$t_a = A_2 + B_2 = t_{Al}a_2\frac{\rho_{Al}}{\rho_a} + t_Pb_2\frac{\rho_P}{\rho_a} .$$

Equations 2.3.5 are used to map Al and PMMA thicknesses, obtained from the calibration, to equivalent glandular and adipose tissue thicknesses.

While it is true that real systems do not conform to the exponential attenuation assumption, this assumption is a compromise made in order to derive analytical, linear relationships between material thicknesses. In particular, the potential error introduced by the mono-energetic spectra assumption is addressed in the uncertainty calculations (see Appendix A).

2.4. Breast Density measurements using mammography

This section oresents a brief overview of mammography-based methods for breast density measurement. The findings of population breast density studies will also be discussed.

Mammography-based methods can be divided into four categories: visual examination of mammograms, area-based methods, single-energy absorptiometry (SXA), and dual-energy material decomposition. Visual examination provides a qualitative density measurement, while area-based methods rely on segmentation techniques (such as thresholding) to produce bidimensional breast density measurements. Finally, both SXA and dual-energy methods are able to provide volumetric measurements i.e. volumetric breast density.

Visual examination of mammograms is the oldest method, and it remains the standard clinical practice as of yet. The Breast Imaging Reporting and Data System (BI-RADS) system provides a way to classify breasts according to their mammographic density, as illustrated in figure 2.4.1.
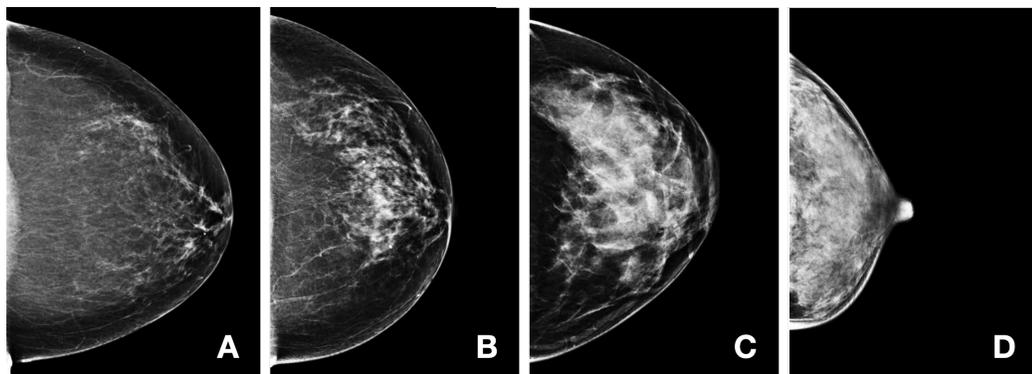
FIGURE 2.4.1. Craneo-caudal (CC) mammographic views depicting the four BI-RADS categories. Breast A is almost entirely fatty, B shows scattered areas of glandular content, C is heterogenously dense, and D is extremely dense. Figure adapted from [28].

As previously stated, the qualitative nature of visually-assessed breast density prevents it from being a true risk predictor or biomarker.

The first step towards a quantitative breast density measurement was taken well before the introduction of digital imaging, in the form of interactive thresholding methods. In [29], a method involving the digitization of radiographic films is presented. After digitizing the film, a histogram-based thresholding was performed to segment breast from background, and to determine its glandular tissue proportion. The aforementioned thresholding was manually carried out by an operator, whose task was to identify gray values corresponding to glandular tissue. Once the threshold was set, breast density was calculated as the ratio of pixels above the threshold (assumed to be glandular) to total pixels within the breast

area. This is the working principle of algorithms such as Cumulus [1] [29, 30] and Madena [31].

While showing better reproducibility than visual assessment, interactive thresholding methods still rely on trained human input to define the density threshold. Because of this, attempts to automate the measurement process were made. The resulting automatic area-based methods employed more advanced forms of segmentation such as maximum entropy [32], adaptive fuzzy c-means [33], and machine learning approaches [34]. These algorithms eliminated the human component and thus improved reproducibility. However, doubts have been cast on the accuracy with which 2D breast density reflects the glandular tissue proportion within the breast, and whether volumetric measurements would be a more adequate risk predictor [9, 10]. Figure 2.4.2 exemplifies how area-based breast density measurements fail to take into account the tridimensional distribution of glandular tissue: it represents two breasts for which the craneo-caudal projection of glandular tissue covers the same area, but the glandular tissue distribution is different along the longitudinal axis. These two breasts would have the same 2D breast density, even though one of them has a larger fraction of glandular tissue.

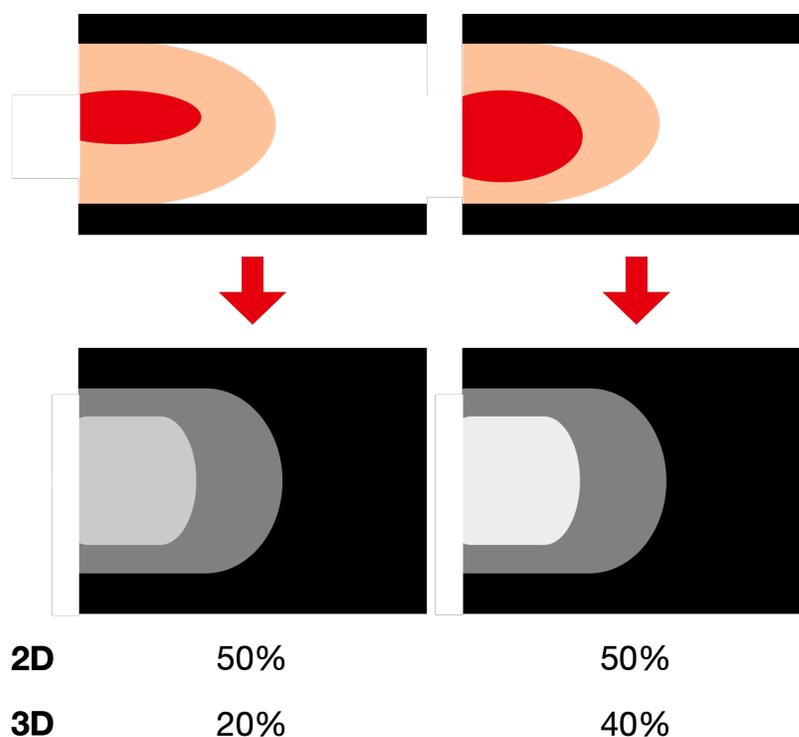|		
2D	50%	50%
3D	20%	40%

FIGURE 2.4.2. Top: lateral view of compressed breasts, in which the red regions represent glandular tissue. Bottom: X-ray illustrations demonstrating how area-based methods would fail to capture the increase in glandular tissue volume.

---

[1]Not to be confused with Cumulus V, which does provide volumetric measurements.

In their watershed paper [35], Yaffe et al used prototype breast CT scanners to measure volumetric breast density. Up to that date, the prominent belief was that the average breast had a density of 50%. What they found, instead, was an average breast density of 19% among their patient cohort, hence bringing to attention "the myth of the 50/50 breast". Thus, interest started to shift towards automated volumetric breast density measurements.

Volumetric methods measure the differences in attenuation produced by adipose and glandular tissue, and thus can be classified as SXA. Within this family of methods, there exists another sub-classification: methods that rely on calibration objects, and methods that utilize physics models.

Briefly, calibration-based methods such as the ones presented in [36, 37, 38], involved imaging phantoms (tissue equivalent, acrylic or aluminum) either alongside the breast or before the acquisition. Then, the attenuation in each pixel was referenced against the calibration array or calibration curve in order to determine the percentage of glandular tissue.

Physics modeling methods, such as Quantra and Volpara, are based on the so-called "h_{int} representation" first developed by Highnam and Brady [39, 40]. This formalism models the energy imparted to the detector at every pixel, $E^{imp}(i,j)$, by the primary X-ray beam that goes through the compression paddle, breast and detector system:

$$(2.4.1) \qquad E^{imp}(i,j) = \phi(V_t, i, j) A_p t_s \int_0^{E_{max}} N_0(V_t, \varepsilon) G(\varepsilon) D(\varepsilon) e^{-\mu_{luc}(\varepsilon) h_{plate}} e^{-h\mu(\varepsilon)} d\varepsilon,$$

where ϕ is the photon flux, V_t is the tube voltage, A_p is the area of the pixel, t_s is the time of exposure, E_{max} is the maximum photon energy, N_0 is the relative number of incident x-ray photons at the specific energy ε, D is the detector efficiency, G is the grid transmission, μ_{luc} is the linear attenuation coefficient of Lucite, h_{plate} is the thickness of the Lucite breast compression plate and $h\mu(\varepsilon)$ is the linear x-ray attenuation coefficient due to breast tissues in the column of tissue above pixel (i,j). Assuming each column of breast tissue is composed only of adipose and glandular tissues, $h\mu$ can be written as follows:

$$h\mu(\varepsilon) = h_{int}\mu_{int}(\varepsilon) + h_{fat}\mu_{fat}(\varepsilon) ,$$

where h_{int} is the thickness of the tissue of interest (glandular), h_{fat} represents adipose tissue thickness, and μ_i their respective linear attenuation coefficients. Furthermore, if the total breast thickness $H = h_{int} + h_{fat}$ is known, the previous equation can be rewritten as:

$$(2.4.2) \qquad h\mu(\varepsilon) = h_{int}(i,j)(\mu_{int}(\varepsilon) - \mu_{fat}(\varepsilon)) + H\mu_{fat}(\varepsilon).$$

If parameters such as tube voltage, current, time of exposure, breast thickness, and material attenuation coefficients are known, after substituting 2.4.2 in 2.4.1 the only unknown is $h_{int}(i,j)$. By equating the primary energy found in the practical case with the theoretical value and solving the resulting nonlinear equation, $h_{int}(i,j)$ can be found.

The Quantra algorithm follows this approach, with the addition of incorporating information from the DICOM headers to help calibrate the model to the specific imaging system being used, and to determine H [41]. This methodology is commonly reffered to as "absolute physics modeling".

Volpara, on the other hand, uses what's known as "relative physics modeling". This method involves finding a pixel signal corresponding to purely adipose tissue, and using this value as a reference [14]:

$$h_{int}(i,j) = \frac{\ln\left(P(i,j)/P_{fat}\right)}{\mu_{fat} - \mu_{int}}.$$

$P(i,j)$ represents the pixel value at coordinates (i,j), while P_{fat} represents the pixel value corresponding to the entirely adipose region. The values in the denominator are the effective x-ray linear attenuation coefficients for fat and glandular tissue at the particular acquisition conditions (target, filter, tube voltage and recorded breast thickness combination). Currently, Volpara and Quantra have risen to prominence on grounds of being fully automated, requiring no calibration, and being FDA-approved for clinical use.

As single-energy X-ray imaging can only be used to quantify one material at a time, these methods rely on breast compression models and other acquisition data to estimate total breast volume and derive volumetric measurements. DEDM allows for the simultaneous quantification of two materials, and thus is able to provide natively volumetric measurements.

The use of dual-energy decomposition to solve the breast density determination problem has been investigated in the past, particularly by Ducote and Molloi [18, 19]. The method proposed in the referenced works was based on the dual energy material calibration formalism discussed in section 2.2, and was calibrated tissue-equivalent material slabs. Percent volumetric breast density was measured on tissue-equivalent phantoms, and they reported a root mean square error of 5% in their measurements. However, these works did not report the uncertainty associated with their breast density measurements, and the proposed method was not applied to clinical images. Finally, it is important to point out that dual-energy methods are not limited to glandular and adipose tissue decomposition. Works such as [23] have experimimented with decomposing the breast in protein, water and lipid content.

This concludes the review on mammography-based breast density measurement methods. The remainder of this section will be devoted to summarizing the known characteristics of breast density as measured using mammography - hereinafter referred to as "mammographic breast density".

Mammographic breast density has been found to be dependent on three variables, namely: compressed breast thickness, patient's age, and total breast volume. The works carried out by Klein et al [42], Young et al [43] and Dance et al [44] first showed the breast thickness dependence. These three studies calculated breast density on screening populations by using measurements of the mAs (determined by the AEC) to draw a relationship between the compressed breast and a phantom material. Dance proposed the following cubic models -divided by age group- to describe this behavior:

$$(2.4.3) \qquad G_i(T_c) = aT_c^3 + bT_c^2 + cT_c + d,$$

where $i = 1$ represents the age 40-49 group and $i = 2$, the age 50-64 group. The values of the coefficients are given in table 2.4.1.

Coefficient	Age 40-49	Age 50-64
a	0.00005209	-0.0001118
b	0.00125494	0.03932
c	-1.988	-4.544
d	138.8	176.0

TABLE 2.4.1. Coefficients for Dance's polynomial model of glandularity as a function of breast thickness (in cm) [44].

By proposing different functions for each age group, Dance's work accounted for the age dependence of breast density. Klein also noted this dependence, pointing out that their calculated glandular fraction (breast density) decreased from about 65% (20 years) to about 30% (at the age of 75 years). This age dependence was also noted in semi-quantitative studies such as [45], where the breast density of 1353 women was visually assessed. Breast density was found to progressively decrease from the age cohort of 25- to 29-year-olds, in which 38% of patients had predominantly (>50%) fatty breasts, through the cohort of 75- to 79- year-olds, in which 76% had predominantly fatty breasts.

In a large-scale, quantitative study (N=15351) Alonzo-Prolux et al [46] reported a decrease in volumetric breast density from 45% to 25% as age increased from 35 to 75 years, and an increase to 30% at 80 years. This work also presented the measured breast density and glandular volume distributions, which were found to be non-normal. Though this may indicate a more complex age-dependence, they argued the increase past age 80 could be a due to the fact that women in that older age group continued to present for mammography imaging because of a higher breast cancer risk, which might imply a higher breast density.

Breast CT studies [35, 47] also confirmed this dependence. In particular, Huang et al [47] presented a thorough analysis of volumetric breast density as a function of different biological and anatomical variables, one of them being age. In this study, the breast

volumes (adipose and glandular) and volumetric breast density of 219 patients were plotted as a function of their age. They found that, as women aged, the volume of adipose tissue increased steadily while the volume of fibroglandular tissue decreased gradually. The results from bCT indicate that the age dependence is not exclusive to mammographic breast density.

Lastly, mammographic density has also been observed to be dependent on total breast volume. Alonzo-Prolux et al [46] reported a steady decline in volumetric breast density with increasing breast size, possibly explained by the association between breast size and body fat. The previously referenced bCT study [47] found an association between breast diameter and volumetric breast density, in that the latter decreases as the former increases.

The reports on the total volume dependence are limited, though biopsy studies such as [48, 49] have shown that smaller breasts have greater areas of collagen and glands. These results suggest higher volumetric glandular densities on smaller breasts.

2.5. Breast Density measurements using MRI

MRI has been recognized as a superior modality for volumetric breast density measurements, and is often used as a reference when assessing the performance of methods based on image modalities such as DBT, Breast CT and Digital Mammography [55, 57, 56]. By nature, it produces 3D images with strong contrast between fibroglandular tissue and adipose tissue. Furthermore, tissue overlap is not a problem in 3D imaging techniques.

Measuring breast tissue volumes using MRI is a segmentation problem; that is, to assign each voxel a degree of membership to the glandular and adipose tissue categories. There are several ways to perform this segmentation, such as adaptive thresholding [?], fuzzy c-means (FCM) clustering [45], and deep learning-based methods [59]. In this work we implemented an adaptive thresholding solution, and the details will be discussed on section 3.8.

Modern MR systems provide a wide breadth of acquisition sequences to choose from, each with their own contrast properties (see figure 2.5.1). Because of this, works such as [47] have assessed their performance in the breast density measurement problem. In the referenced study, Tagliafico et al. compared the performance of T1-turbo spin-echo (T1-tSE), T2-turbo spin-echo (T2-tSE), VIBRANT (Volume Imaging for Breast Assessment) and IDEAL sequences. They concluded that T1-weighed sequences remain the most clinically viable and are sufficient for breast density measurement purposes, but IDEAL sequences -if available- may provide a truer tissue separation because tissue contrast in these sequences are a product of biochemical differences between the tissues.
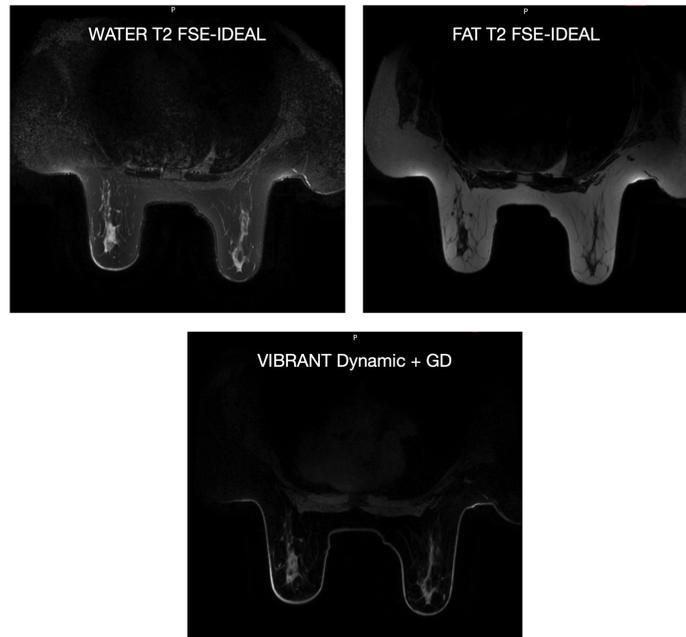
FIGURE 2.5.1. Matched axial MRI slices of the same patient, acquired using three different sequences (indicated on each image).

In this work, breast volumes were measured using IDEAL sequences. IDEAL stands for "Iterative Decomposition of water and fat with Echo Asymmetry and Least squares estimation", and is an implementation of Dixon's method [61] proprietary of GE systems (General Electric Medical Systems, Milwaukee, WI, USA) [62]. In essence, this method relies on in- and out-of-phase imaging (acquiring two images at different echo times), exploiting the difference in chemical shift between water and fat. These sequences provide "water-only" and "fat-only" images, which allows for the quantification of both materials within a voxel.

2.6. Multimodality studies

The ever-increasing breadth of modalities and algorithms that can be used to assess breast density raises an obvious question: are the resulting measurements equivalent? (And if not, which technique should be recognized as the "gold standard"?). To help answer this question, several multimodality studies have been carried out. In this section, the main findings of such studies are be summarized.

Klifa et al [50] compared MRI breast density measurements (FCM implementation on fat-supresed gradient recall echo sequences) to digitized mammography results (customized threshold-based method). Breast density was calculated for 35 women at high-risk for breast cancer. Mammographic breast density was found to be higher than the MRI counterpart, and better correlation between the two measurements was found for lower density breasts (<20%, $R^2 = 0.732$). For higher densities (>20%), $R^2 = 0.264$. The overall correlation coefficient without the density separation was $R^2 = 0.67$. Finally, the

authors observed that the range in MRI density measurements was smaller than the range of corresponding mammographic densities.

Kontos et al [51] compared two methods: Quantra, MRI (FCM implementation on T1 sequences). Glandular volume, total volume, and volumetric breast density (VBD) were measured for 32 women in a high risk population, with recently detected abnormalities (only the contralateral breast images were processed). The authors found that VBD showed good correlation ($R^2 = 0.8$), though the measured values were higher for the mammography-based method. Also, there was good agreement in total volume, as the distributions' means showed no statistical difference. However, lower agreement was found in glandular volume, particularly for lower density breasts, with MRI providing lower values. Finally, they found a better correlation between the two methods when for higher VBD values.

In [52], Wang et al compared Quantra measurements to MRI measurements. Glandular volume, total volume and volumetric breast density (VBD) were measured for 123 screening patients. The authors performed linear regressions on the Quantra measurements as a function of MRI measurements, and found the following R^2 coefficients: 0.89 for total volume, 0.38 for glandular volume, and 0.50 for VBD. To be noted, the regression equation slope parameters were significantly different than one, and intercepts significantly different than zero. The reported root mean square errors (with respect to MRI) were 110.55 cm^3 for total volume, 55.10 cm^3 for glandular volume, and 6.48% for VBD.

In a follow up work [53] the same group compared four methods: SXA (custom implementation), Quantra, Volpara and MRI. Glandular volume, total volume and volumetric breast density (VBD) were calculated for 99 screening patients. They used the MRI results as ground truth, and evaluated the correlation between them and the remaining three methods. Correlation coefficients for volumetric breast density were 0.78, 0.51 and 0.73 for SXA, Quantra and Volpara, respectively. In this work, the non-normality of the glandular volume distribution was pointed out, and the authors performed a log normalization before carrying out the statistical analysis. By calculating the kappa statistic, substantial agreement was found between all percent fibroglandular tissue measures (κ=0.72 to 0.63), but only moderate agreement for log glandular volumes. The kappa statistics for all percent density measures were highest in the comparisons of the SXA and MRI results. The largest error source between MRI and the mammography techniques was found to be differences in measures of total breast volume. The RMSE in the evaluated quantities ranged from 4.4% (Volpara) to 10% (SXA) for VBD, 0.37 (Volpara) to 0.57 (Quantra) for log glandular volume, and 108 cm^3 (SXA) to 121 cm^3 (Volpara) for TBV.

Tagliafico et al [54]compared full field digital mammography (FFDM), DBT and MRI measurements. The used algorithms were all based based on Shannon's maximum enthropy segmentation method. They measured percent breast density in a group of 48 patients that underwent diagnostic imaging (only images of unaffected breasts were processed). They found good ($R^2 = 0.9$) correlation between MRI and DBT measurements, and between DBT and Digital mammography ($R^2 = 0.94$). Correlation between MRI and Digital mammography was lower ($R^2 = 0.78$). It was also found that mammography over-estimated VBD by 16.2% in comparison to MRI, and 15.1% in comparison to

DBT. The increased VBD values in FFDM can be explained by the fact that they used a thresholding method on 2D images, which resulted in area density measurements.

In another three-modality study, Pertuz et al [55] compared FFDM (Volpara), DBT (In-house method) and MRI (FCM implementation on T1 sequences). The authors measured total volume, glandular volume, and VBD in a population of 80 women undergoing high-risk screening. The performed statistical analysis showed significant differences in VBD and glandular tissue volume, but not in total breast volume, among the studied modalities. Substantial correlation among VBD measurements was also found, with the strongest being between MRI and DBT. High correlationns were also observed for estimates of total breast volume across the three modalities.

Finally, Holland et al [56] compared a custom FFDM method (similar to the Volpara implementation of the h_{int} representation) to MRI. Total volume, glandular volume and VBD were measured in a group of 202 patients (no other information is provided). High correlations in total volume were found, with R^2 values ranging from 0.77 to 0.81, depending on the FFDM method's parameters. However, it was found that FFDM tended to over-estimate this quantity. VBD showed substantial differences among the two techniques, especially for dense breasts. Lower correlations were found in glandular volume, with R^2 ranging from 0.62 to 0.74.

Results from multimodality studies can be summarized by the following bullet points:

- There is good agreement in total volume, when comparing mammography-based methods to 3D modalities such as MRI.
- The main source of disagreement in VBD across these modalities stems from differences in measured glandular volume.
- Agreement in VBD seems to be better for less dense breasts.
- Mammography-based methods tend to over-estimate glandular tissue volume, which in turn results in over-estimation of VBD. Additionally, the variation in VBD among a given population seems to be higher when measured using mammography, with the range in MRI measurements being markedly compressed in comparison.
- Even though inter-modality correlations range from decent to excellent, the measured quantities may not be equivalent. As there is no agreed-upon ground truth, multimodality studies can only evaluate relative performance between methods.
- Many of the cited studies involve high risk patients, and therefore a selection bias is not out of the question.

Materials and methods

3.1. DEDM image acquisition

All mammography images used in this work were acquired at the Radiodiagnostic Department of Mexico's National Cancer Institute (INCAN), using a commercial Amulet Innovality digital mammography and DBT station (FUJIFILM Corporation). This station is equipped with a tungsten anode X-ray tube and selectable Rh (50μm) and Al (700μm) filters. Detector-wise, it features a logarithmic response a-Se flat panel detector, in which individual detector elements are set in a hexagonal array. This system is capable of producing digital mammograms with square pixels down to 50μm in side [63, 64].

Acquisitions were performed in "manual mode" (AEC bypassed). The dual-energy technique was optimized in a previous study [65], and the acquisition parameters are shown in table 3.1.1.

Technique	kV	mAs	Filter
LE	31	63	Rh
HE	45	63	Al + 5mm Al (external)

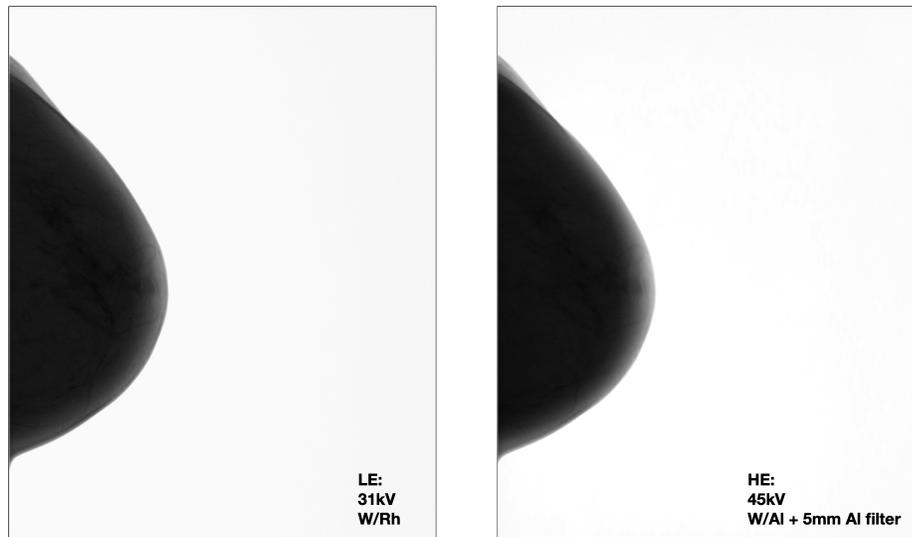TABLE 3.1.1. DEDM acquisition parameters.

FIGURE 3.1.1. Raw mammograms of the same patient as figure 2.1.3. Display window has been adjusted to cover the entire the dynamic range.

All calibration and clinical images were acquired as craneo-caudal projections with right breast orientation (RCC). They were exported and processed as "raw" images (.std files). In this format, the images have a size of 2364x2964 px, with each pixel being 100µm in side. The pixel value (PV) - signal intensity relationship is linear and non-inverted; this is, higher PV, displayed as whiter shades of gray, are indicative of lower attenuation. Given the detector's logarithmic response, raw images need no log-transformation prior to processing. Figure 3.1.1 shows an example of a pair of clinical raw images.

3.2. DEDM image pre-processing

The pre-processing routine applied to all DEDM images (both calibration and clinical) involves the following two steps:

(1) Detector non-uniformity correction.
(2) Downsampling.

The pre-processing was applied in the stated order, and each step will be discussed in the following sections.

3.2.1. Non-uniformity correction. Figure 3.2.1 shows a map of percentage differences (relative to the mean) corresponding to an uniform PMMA slab 4cm in thickness, which covered the entire detector area. These maps, and therefore the unprocessed images, exhibit intensity gradients: the LE image, a falloff near the outer edge; the HE image, an increase.
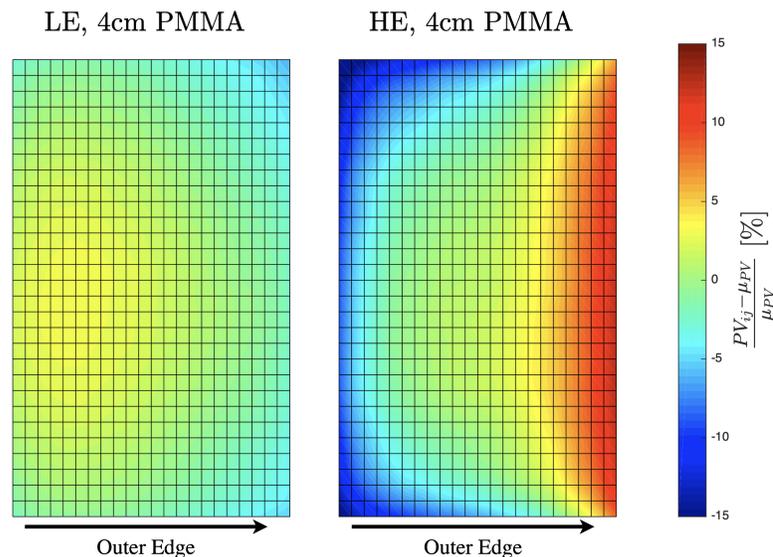

FIGURE 3.2.1. Pixel value difference maps (relative to the mean value) corresponding to an uniform PMMA slab, displayed using the same scale. Note that the relative difference can reach values of up to 15% for the HE image.

If the response were thoroughly uniform, the imaged slab would translate to uniform images, and the maps in figure 3.2.1 would be equal to zero in every region. This non-uniformity affects both the calibration and volume calculations, and must be accounted for before any further processing takes place.

To this end, the following correction was devised: First, an "uniformity image" U^0 was acquired. U^0 is obtained by imaging, in Quality Assurance (QA) mode, the aforementioned PMMA slab using the clinical DE technique. In order to minimize quantum noise effects on the correction, U^0 was convolved with a 40x40 pixel median filter H:

$$U = U^0 * H .$$

If I_{ij} is the unprocessed image, the corrected image is given by:

$$(3.2.1) \qquad\qquad I'_{ij} = I_{ij} \left(\frac{\overline{U}}{U_{ij}} \right) ,$$

where $\overline{U}$ is the mean value of the uniformity image. The subindexes i, j represent arbitrary pixel coordinates in each image, so eq. 3.2.1 indicates pixel-by-pixel operations.

3.2.2. Downsampling. The purpose of this pre-processing stage is two-fold:

- Improve the DE calibration quality.
- Reduce resulting noise and computation times.

In order to prevent aliasing, downsampling is usually preceded by a low-pass filter application. In this case, a Gaussian filter was used. Gaussian filtering and downsampling were handled using the ImageJ implementation contained in the *Downsample* function [66], using parameters $\sigma_{source} = \sigma_{target} = 1$. Image dimensions were reduced by a factor of 4, so the resulting images had a size of 591x741 px, with a pixel area of 0.16 mm^2.

Downsampling, by definition, reduces spatial resolution. For this application, the tradeoff is well worth it since the interest lies in measuring overall breast composition, rather than visualizing fine details.

3.3. Dual-energy calibration

The calibration discussed in section 2.2 required imaging an Al-PMMA phantom with different thickness combinations. The step-wedge configuration shown in figure 3.3.1 provided a total of 43 data points for use in the calibration: 42 (t_{Al}, t_P) combinations, plus the background signal region.
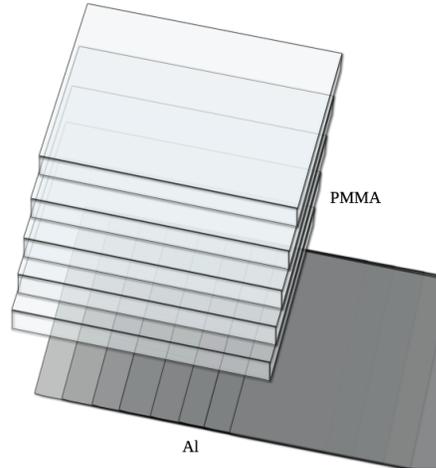

FIGURE 3.3.1. Illustration of the Al-PMMA step-wedge array used to calibrate the method.

The calibration set is described mathematically as $\mathcal{C} = P \times A$. $P=\{0,10, 20, 30, 35, 40\}$ and $A=\{0, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7\}$ were the PMMA and Al thickness steps in mm, respectively. Once the DE calibration images had been pre-processed, mean pixel values (MPV) were measured inside each thickness combination region using a square 20x20 px ROI (figure 3.3.2). ROIs placements were recorded and applied using an ImageJ macro, which resulted in MPV measurements corresponding to the same pixels in LE and HE images.
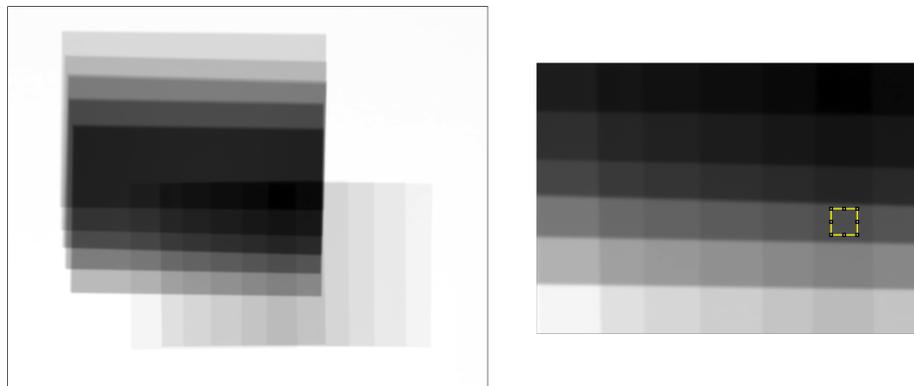

FIGURE 3.3.2. Left: LE calibration image. Right: Closeup showing ROI definitions.

The measured pixel values in each region were subtracted from the "background" signal; that is, the MPV measured in a region where $t_{Al} = t_P = 0$:

$$F = f(0,0) - f(t_{Al}, t_P) \, ,$$

$$G = g(0,0) - g(t_{Al}, t_P) \, ,$$

where F represents PV in the LE image and G represents PV in the HE image. The calibration dataset consisted of points in the (F, G, t_{Al}) and (F, G, t_P) spaces. Two different functions were then fitted to the dataset: a first order surface, and the 8-parameter rational function (eq. 2.2.1):

$$t_{i,1} = aF + bG \, ,$$

$$t_{i,2} = \frac{a + bF + cG + dF^2 + eFG + kG^2}{1 + mF + nG} \, .$$

3.4. Basis material decomposition

Going back to equations 2.3.1 and 2.3.2, for each pair of energies (L, H) a system of 2 equations can be defined. For the Al attenuation coefficient decomposition, the pair of equations becomes:

$$\frac{\mu_{Al}^L}{\rho_{Al}} = a_1 \frac{\mu_g^L}{\rho_g} + a_2 \frac{\mu_a^L}{\rho_a} \, ,$$

$$\frac{\mu_{Al}^H}{\rho_{Al}} = a_1 \frac{\mu_g^H}{\rho_g} + a_2 \frac{\mu_a^H}{\rho_a} .$$

In matrix form, the systems for the Al and PMMA attenuation coefficient decomposition become

$$(3.4.1) \qquad \mathcal{M}\vec{a} = \begin{bmatrix} \mu_g^L/\rho_g & \mu_a^L/\rho_a \\ \mu_g^H/\rho_g & \mu_a^H/\rho_a \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mu_{Al}^L/\rho_{Al} \\ \mu_{Al}^H/\rho_{Al} \end{bmatrix} \, , \text{ and}$$

$$(3.4.2) \qquad \mathcal{M}\vec{b} = \begin{bmatrix} \mu_g^L/\rho_g & \mu_a^L/\rho_a \\ \mu_g^H/\rho_g & \mu_a^H/\rho_a \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \mu_P^L/\rho_P \\ \mu_P^H/\rho_P \end{bmatrix} .$$

By inverting systems 3.4.1 and 3.4.2, the set of parameters $\{a_1, a_2, b_1, b_2\}$ is determined. Given the energy dependence of effective atomic numbers, the solutions to these systems are not the same for each pair of energies considered [67]. As the dual-energy spectra are poly-energetic, the systems were solved for each possible pair of energies within the [10, 45] keV range, in steps of 1keV.

The material attenuation coefficient data were obtained from the XCOM database (NIST). XCOM calculates the attenuation coefficients for compounds and mixtures by performing a weighted sum over the attenuation coefficients of the constituting elements. The weight factors are obtained from the material composition percentage (mixtures) or the chemical formula (compounds) [68]. Table 3.4.1 shows the compositions of the materials of interest in this work.

	H	C	O	N	S,P,K,Ca	Al	ρ [g/cm^3]
Breast glandular tissue	10.2	18.4	67.7	3.2	0.5	0	1.04
Breast adipose tissue	11.2	61.9	25.1	1.7	0.1	0	0.93
PMMA	8.05	59.98	31.96	0	0	0	1.19
Al	0	0	0	0	0	100	2.7

TABLE 3.4.1. Material compositions (percentage) and densities. Breast tissue percentage compositions were taken from [69].

3.5. Material decomposition accuracy test

In order to test the material decomposition accuracy, a step-wedge arrangement of tissue equivalent slabs was imaged (figure 3.5.1).
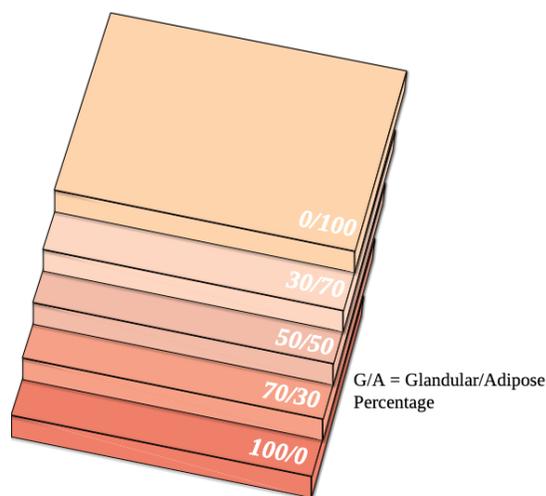

FIGURE 3.5.1. Tissue-equivalent slab configuration used to test the material decomposition relationships. Each slab is 10mm in thickness and equivalent to different proportions of glandular/adipose tissue.

A LE image of the illustrated arrangement is shown in figure 3.5.2. The "steps" of the arrangement define regions with different glandular ratio values.

Region	Glandular ratio
1	100
2	85
3	73.3
4	62.5
5	50

FIGURE 3.5.2. LE image of the slab configuration. Glandular ratio inside each region is also shown.

The equivalent nominal glandular thickness t_e of each slab was obtained by multiplying its glandular fraction by the slab thickness. For instance, $t_e = 7$ mm for the "70/30" slab. The nominal glandular ratio inside region i was calculated as $GR_i = {}^{100 \cdot \sum t_e}/_{10n}$, where n is the number of superimposed slabs in region i and $\sum t_e$ is the sum of their equivalent glandular thicknesses.

Glandular tissue and adipose tissue thickness maps $t_g(i,j)$ and $t_a(i,j)$ were generated, as well as total thickness maps $T(i,j) = t_g(i,j) + t_a(i.j)$. Finally, the glandular ratio map was calculated as $GR(i,j) = {}^{t_g(i,j)}/_{T(i,j)}$. For each map, MPV inside each region were measured and compared to the nominal quantities.

3.6. Patient images

The patient images used in this work were acquired as part of a completed clinical study [65]. The protocol included twenty-six patients with suspected multicentric breast cancer, and it involved the acquisition of various MRI sequences (both pre and post-contrast) and a CEDM study. The latter consisted of a single energy temporal (SET) acquisition of the suspicious breast at 1, 2, 3 and 4min post-contrast injection, followed by a DE acquisition of the contralateral (non-suspect) breast. Only the DE images of the contralateral breast were processed in this work.

Out of the twenty-six available image sets, twelve were discarded because they either showed artifacts, or were acquired with a different technique to that used in the DEDM calibration. This left us with a population of $N = 14$, with characteristics shown in table 3.6.1:

Patient ID	Age [years]	Examined Breast	T_c [mm]
1	48	R	59
2	44	R	68
3	43	R	60
4	61	R	40
5	61	R	43
6	58	L	43
7	55	R	52
8	39	R	43
9	36	R	51
10	61	R	56
11	58	R	39
12	45	R	49
13	52	R	57
14	56	R	48

TABLE 3.6.1. Patient group characteristics. Mean age was 51.2 years. T_c represents the compressed breast thickness, as registered by the mammography unit.

The MRI studies were acquired on a 3T GE system at INCAN's Radiodiagnostic Department. The segmentation of breast tissues was performed on pre-contrast IDEAL (FatOnly) slices. These sequences consisted of 40 axial slices with square pixels 0.7813 mm in side, and 5.5 mm inter-slice spacing (voxel size = 3.35 mm^3).

3.7. Clinical DEDM image processing and volumetric calculations

In addition to the pre-processing described in section 3.2, clinical DEDM images required an extra step of processing: masking the breast region. This masking served two purposes:

- The mask defined the breast area, necessary for calculating volumes.
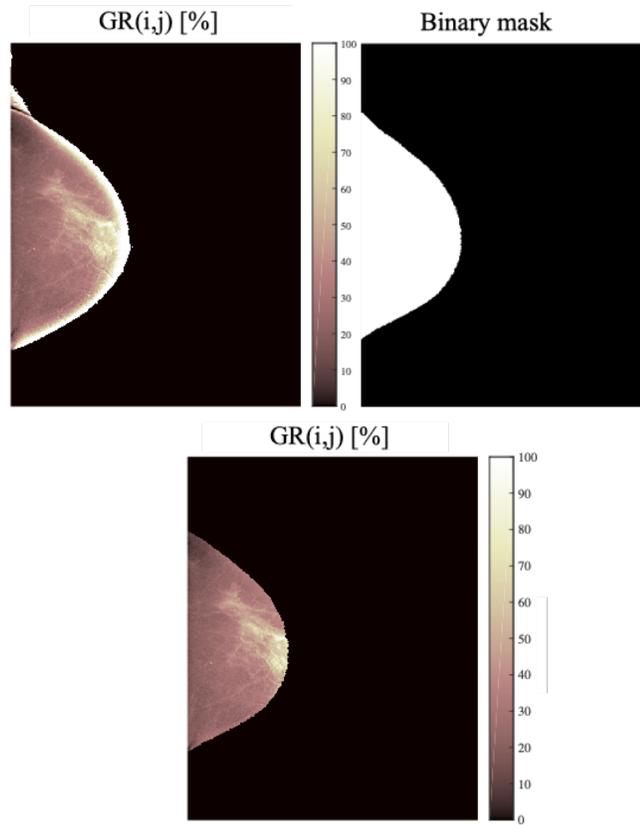- It excluded the breast border and the associated artifact (see figure 3.7.1) from the calculation.

FIGURE 3.7.1. Left to right: glandular ratio map without breast masking. Binary mask. Glandular ratio map calculated after applying the binary mask to the original images. The high GR region along the edge is considered an artifact in the calculation, since there is no glandular tissue present in that area.

A bespoke mask was generated for each pair of DE images. This was done in the following way: first, a preliminary glandular ratio map $GR(i,j)$ was calculated. From $GR(i,j)$, a gradient magnitude image $G(i,j)$ was generated. As GR increases abruptly near the edge of the breast, $G(i,j)$ has the highest values in that region - which means this parameter can be used as a criterion for elimination. The following image was then computed:

$$T^*(i,j) = \begin{cases} 0 & G(i,j) > G_0 \\ T(i,j) & \text{otherwise} \end{cases}$$

where $T(i,j)$ is the preliminary total thickness image and G_0 is a empirically determined threshold. Next, $T^*(i,j)$ is binarized in order to obtain a mask. As the resulting image is likely to have imperfections near the edge, the following morphological operations were applied: filling, spur pixel removal, and erosion along the breast edge. Finally, a gaussian filter was applied and a final thresholding was performed in order to smooth out the mask. This process is illustrated in figure 3.7.2.
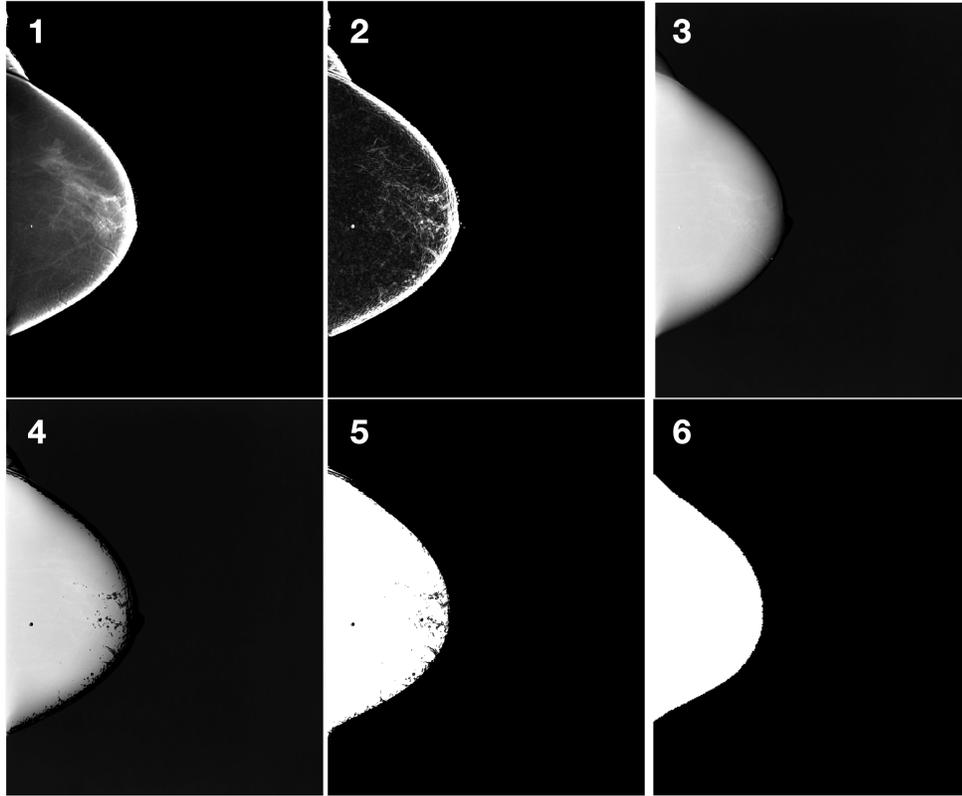
FIGURE 3.7.2. 1) Preliminary GR map. 2) Gradient magnitude image $G(i,j)$. 3) Preliminary total thickness image $T(i,j)$. 4) $T^*(i,j)$. 5) Binarized $T^*(i,j)$. 6) Resulting binary mask after applying the morphological processing.

Given a pair of clinical DE images, I_{LE} and I_{HE}, the following processing steps were taken to obtain the volumetric measurements:

1. Pre-processing (non-uniformity correction and downsampling).

2. Normalization of the DE images. As the calibration was performed in terms of background-normalized PV, the LE image is defined as $F = I_{LE,0} - I_{LE}(i,j)$; similarly the HE image is $G = I_{HE,0} - I_{HE}$. The background PV was measured in regions with no breast present.

3. Pixel-by-pixel application of the Al-PMMA calibration. Al and PMMA thickness maps were generated using the following equations:

$$t_{Al} = \frac{a + bF + cG + dF^2 + eFG + kG^2}{1 + mF + nG} \quad \text{and} \quad t_P = aF + bG \,.$$

4. Pixel-by-pixel application of the basis material decomposition relationships 2.3.5 . At this stage, preliminary tissue thickness and GR maps were generated:

$$t_g = t_{Al}a_1\frac{\rho_{Al}}{\rho_g} + t_P b_1\frac{\rho_P}{\rho_g} \ , \quad t_a = t_{Al}a_2\frac{\rho_{Al}}{\rho_a} + t_P b_2\frac{\rho_P}{\rho_a} \ , \quad T = t_g + t_a \text{ and } GR = 100 \cdot (t_g/T).$$

5. Breast masking generation and application. The binary mask was applied to the preliminary maps generated in step 4.

6. Volume calculations. The tissue maps were integrated over the entire image to produce volumetric measurements:

$$(3.7.1) \qquad\qquad V_g = \sum_{i,j} t_g(i,j)dA, \qquad V_T = \sum_{i,j} T(i,j)dA,$$

where dA is the pixel area (0.16 mm^2). As the binary masks were applied beforehand, pixels outside the breast area do not contribute to the area integrals. Finally, the **volumetric glandular ratio**, or volumetric breast density, was computed as:

$$(3.7.2) \qquad\qquad VGR[\%] = 100\left(\frac{V_g}{V_T}\right).$$

3.8. Clinical MRI processing and volume segmentation

As previously stated, each patient MRI study consisted of 40 slices. Each one of them was a 16-bit depth DICOM image, 512x512 px in size. The individual slices were exported as text matrices and processed in this format.

In terms of pre-processing, a background intensity gradient correction (sometimes referred to as bias-field correction) was performed, followed by the application of a mask to define the breast region. Let S be a MRI slice. The background-corrected slice S^* is given by:

$$S^* = \frac{S}{S * G_{256}}$$

where G_{256} is a Gaussian filter with $\sigma = 256$px. The convolution in the denominator has the effect of erasing all high-frequency content, leaving only the low-frequency background gradient. Once each slice was corrected, a ROI surrounding the breast of interest was manually defined on the center slice, as shown in figure 3.8.1. The complement to this ROI was mapped to zero for each slice, and the resulting images were binarized. Each slice was morphologically processed (filling and closing) and the erosion operation was applied along the edge of the breast in order to exclude this region. The resulting binary slices $B_M(i,j)$ constituted the breast mask, and it was applied before the tissue segmentation took place.
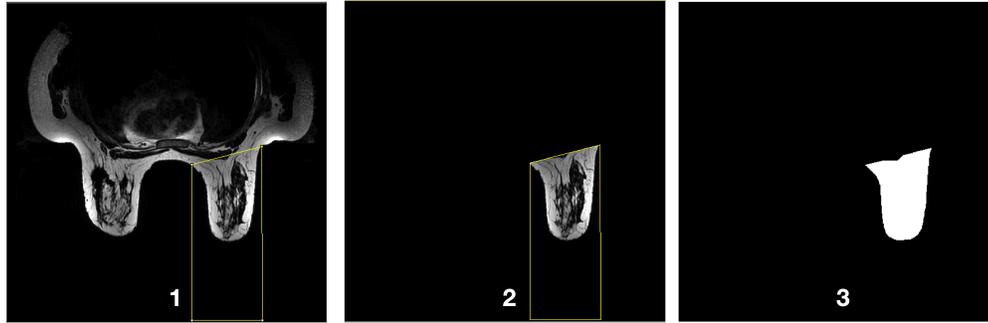
FIGURE 3.8.1. 1) Center slice of a clinical MR volume, yellow lines indicate the segmented breast region. 2) All pixels outside the selected ROI are mapped to zero. 3) Resulting binary mask for the same slice.

In order to segment adipose tissue from non-adipose tissue -assumed to be glandular-, an adaptive thresholding method was applied. This method is a customized implementation of Wellner's method [70, 71]. It consists of comparing each pixel to an average of the surrounding neighborhood, and if the value of the current pixel is t percent lower than the average then it's set to zero. Mathematically, each slice $S^*(i,j)$ was binarized thus:

$$B_a(i,j) = \begin{cases} 0 & S^*(i,j) \leq \overline{S_n}(1 - t/100) \\ 1 & \text{otherwise} \end{cases},$$

where $\overline{S_n}$ is the average PV in the $n \times n$ neighborhood surrounding pixel (i,j), and t is the threshold value. For this application, the empirically determined values $n = 10$ and $t = 5$ were used. $B_a(i,j)$ is the binary image in which a value of 1 represents a pixel that is comprised of adipose tissue.
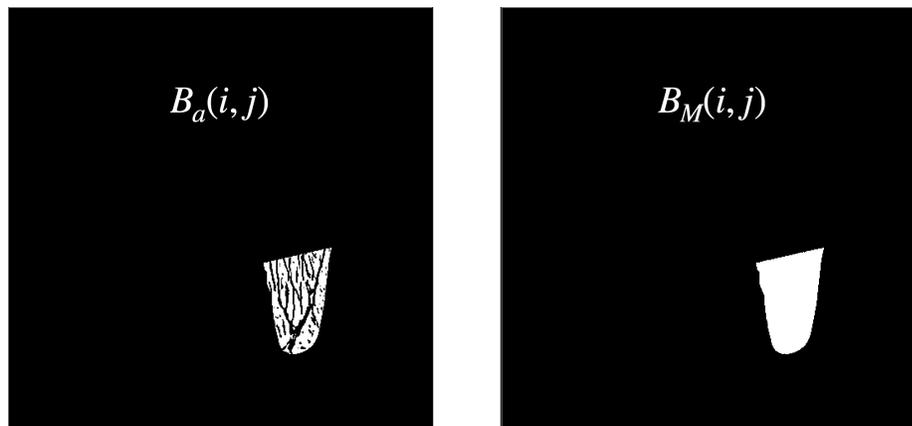

FIGURE 3.8.2. Left: Binary adipose tissue slice. Right: Binary breast mask slice.

The adipose tissue volume was calculated by counting the non-zero pixels in every $B_a(i, j)$ slice, and then multiplying that number by the voxel dimensions. Similarly, the total volume was obtained by counting the non-zero pixels in $B_M(i, j)$. The glandular volume V_g was calculated by subtracting the adipose volume from the total volume.

$$P_{a,k} = \sum_{i,j} B_a(i, j), \qquad P_{T,k} = \sum_{i,j} B_M(i, j),$$

$$V_{a,MRI} = dV \sum_k P_{a,k}, \qquad V_{T,MRI} = dV \sum_k P_{T,k},$$

$$V_{g,MRI} = V_{T,MRI} - V_{a,MRI},$$

with $P_{x,k}$ representing the number of adipose ($x = a$) or total ($x = T$) pixels inside the breast area in a given slice k, and dV the voxel dimensions. Finally, the volumetric glandular ratio was computed as the quotient of V_g and V_T:

$$(3.8.1) \qquad VGR_{MRI}[\%] = 100 \left(\frac{V_{g,MRI}}{V_{T,MRI}} \right).$$

Results and discussion

4.1. Dual-Energy calibration

Figure 4.1.1 shows the calibration data points obtained from the step-wedge array shown in figure 3.3.1, while table 4.1.1 shows the fitted parameters for the material thickness functions as well as the R^2 value and the root-mean-square error (RMSE).
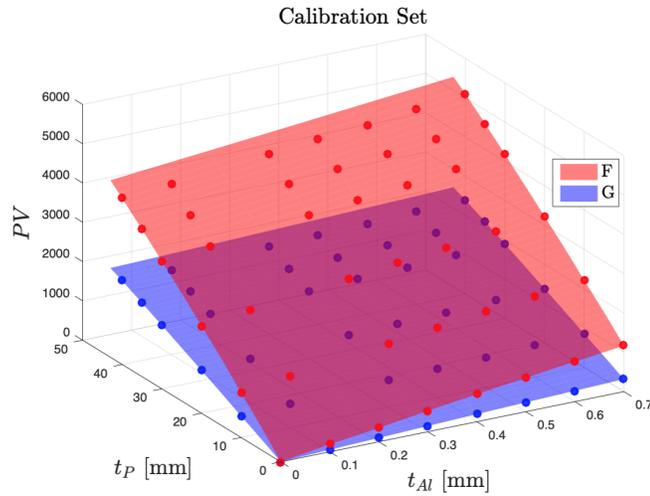

FIGURE 4.1.1. DE Calibration data set, plotted as background-normalized PV vs material thickness. LE PV in red, HE PV in blue.

Material	Function	Parameters	Value	R^2	RMSE [mm]
Al	$t_{Al} = \frac{a+bF+cG+dF^2+eFG+kG^2}{1+mF+nG}$	a	-2.11E-04	0.9971	0.014
		b	1.34E-03		
		c	-2.86E-03		
		d	-1.03E-06		
		e	4.43E-06		
		k	-4.65E-06		
		m	-7.75E-04		
		n	1.64E-03		
PMMA	$t_P = aF + bG$	a	-0.01128	0.9999	0.1263
		b	0.04113		

TABLE 4.1.1. DE calibration fit parameters and goodness of fit metrics

It was found that a linear surface provided a good fit for t_P. This function was preferred over the 8-parameter one because linear expressions propagate less uncertainty and noise. Figure 4.1.2 shows absolute and relative errors in the calibration points, as a function of known thickness.
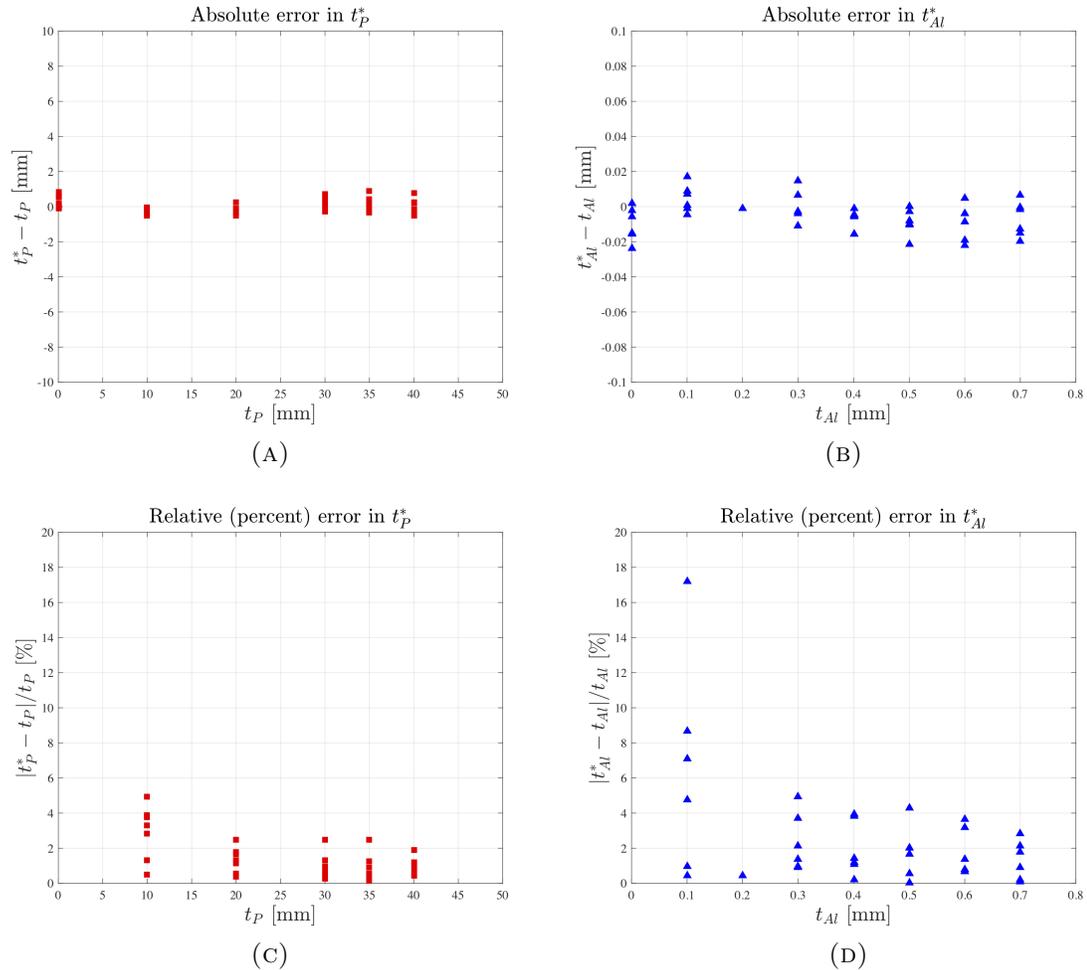

FIGURE 4.1.2. Absolute and relative errors in the calibration data. t_i^* represents the calculated thickness; t_i, the nominal values.

For both materials, the absolute error was fairly independent of material thickness, as can be seen on figure 4.1.2 A and B. This means that the relative error in the calculations decreased as the thickness increases, and its values were typically below 5% (figure 4.1.2 C and D). However, for very small aluminum thicknesses (0.1mm), this error went as high as 17%.

For a true separation of materials, the estimates of one material thickness must be independent from the second material thickness. Figure 4.1.3 shows the estimates of one material thickness when the second one varied in thickness (i.e., thickness values along each "step" region). Figure 4.1.4 shows the calculated maps for the calibration phantom in which the adequate separation can be visually confirmed.
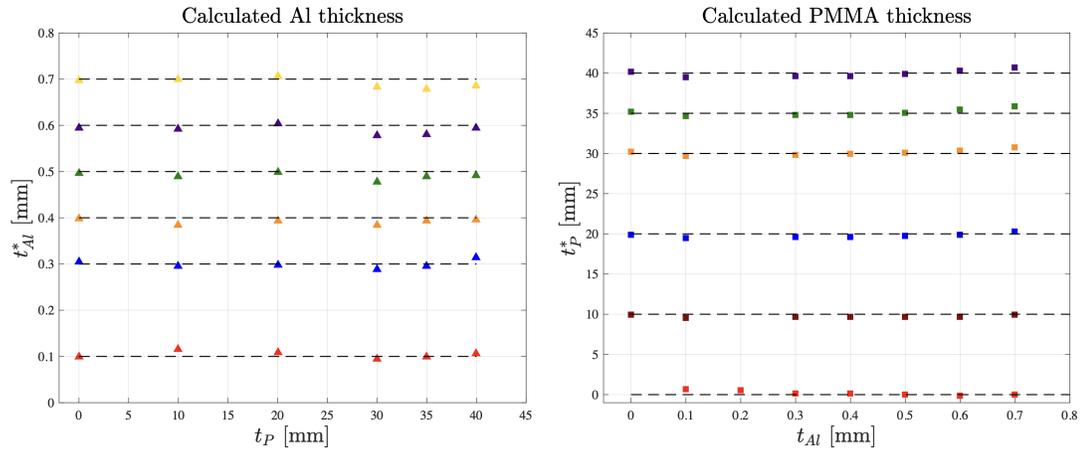
FIGURE 4.1.3. Left: variation in the Al thickness estimate when the superimposed PMMA varied from 0 to 40 mm. Right: variation in the PMMA thickness estimate when the underlying Al layers varied from 0 to 0.7 mm. Dotted lines indicate the actual material thickness.
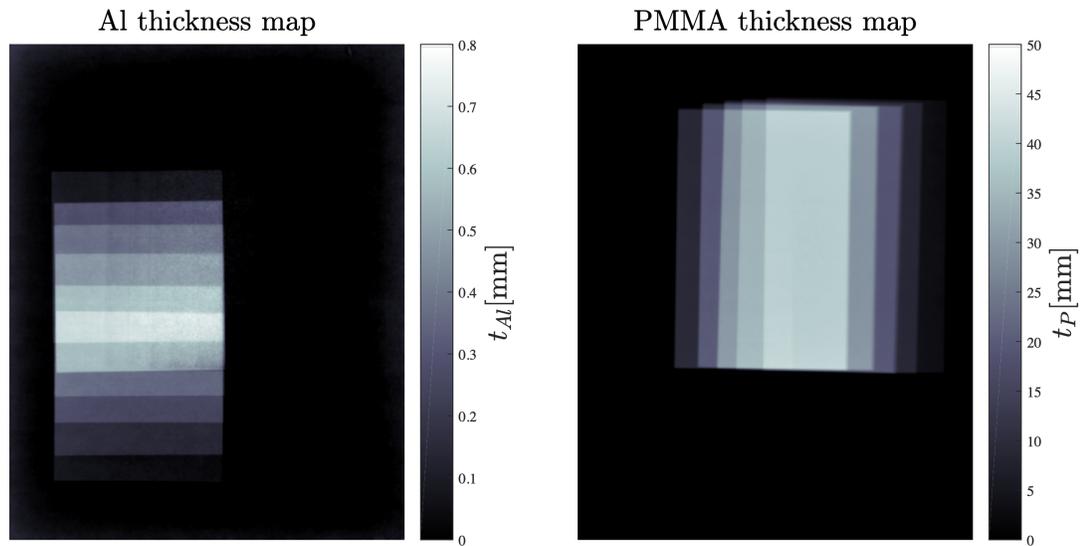

FIGURE 4.1.4. $t_{Al}(i,j)$ and $t_P(i,j)$ calculated from the calibration phantom images.

To conclude the Al-PMMA calibration evaluation, a second step-wedge phantom with a different material arrangement was imaged. This array consisted of the following thickness steps t_P={0,10, 20, 30, 40, 45, 50} mm and t_{Al}={0, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7} mm . The resulting material thickness maps are shown in figure 4.1.5. Mean thickness and its standard deviation were measured inside each material "step" region, and compared to the known material thickness (figure 4.1.6).
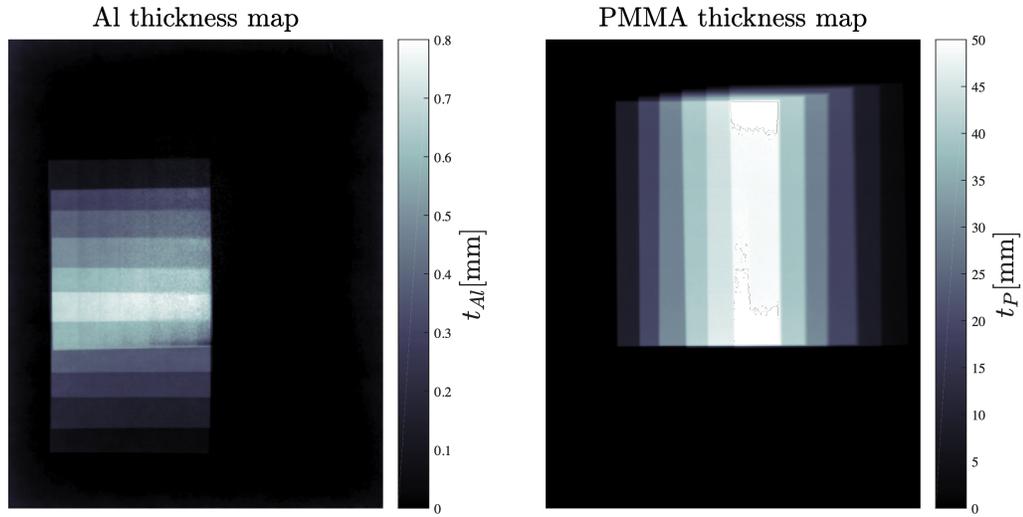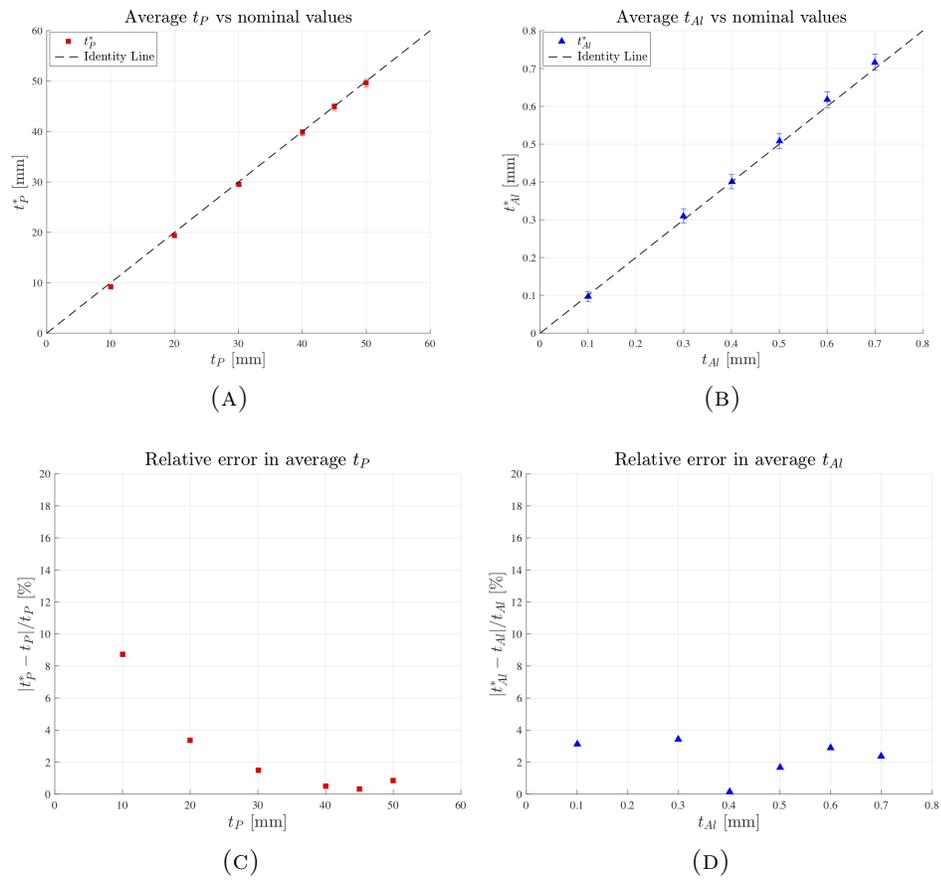
Al thickness map

PMMA thickness map

FIGURE 4.1.5. $t_{Al}(i,j)$ and $t_P(i,j)$ calculated from a second Al-PMMA array.

(A)

(B)

(C)

(D)

FIGURE 4.1.6. Errors in thickness estimation on an independent Al-PMMA array. The length of each error bar represents the standard deviation of the measured thickness values inside each "step" ROI.

Figure 4.1.5 shows that separation of the two materials was achieved, though there were some artifacts in $t_{Al}(i,j)$ introduced by edges in the PMMA slabs. The mean percent error in material thickness across all regions was 2.3% for Al and 2.5% for PMMA, and for both materials it was typically below 4%.

4.2. Basis material decomposition parameters

Figure 4.2.1 shows maps containing the solutions $\{a_1, a_2, b_1, b_2\}$ to systems 3.4.1 and 3.4.2, for each pair of energies $(L < H)$ in the [10,45] keV interval.
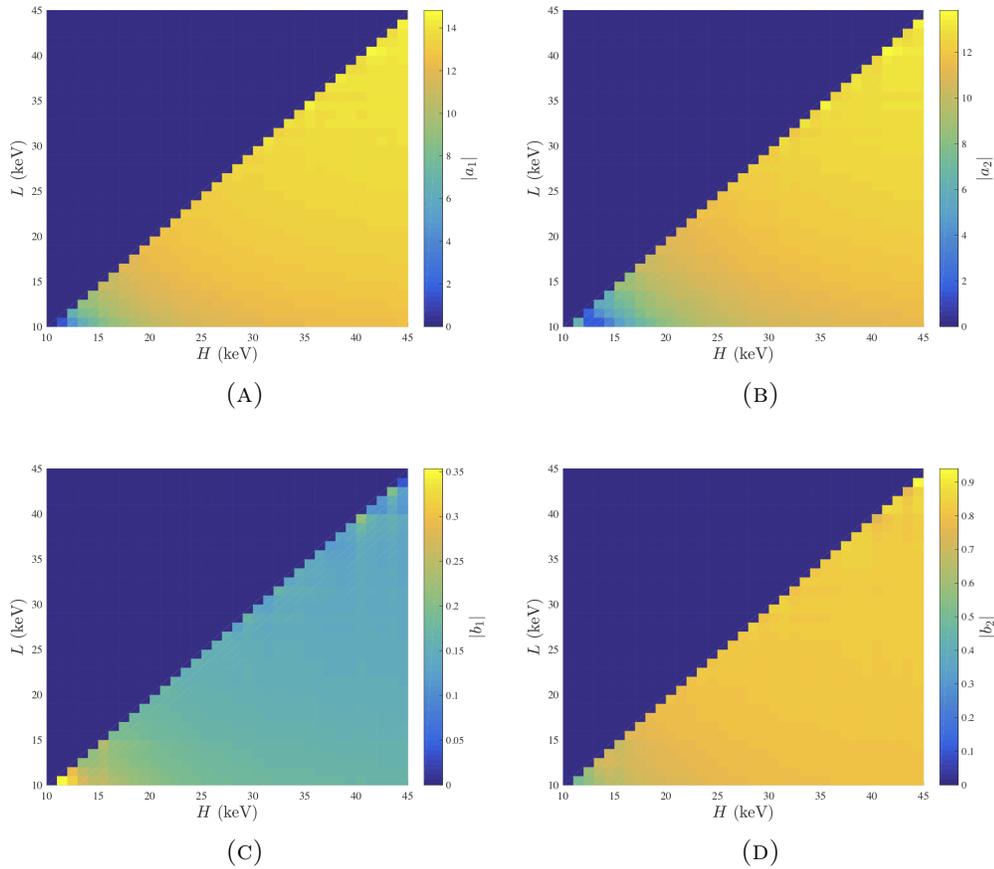

(A)

(B)

(C)

(D)

FIGURE 4.2.1. Calculated basis material decomposition parameters for each pair of energies. Absolute values are plotted for figure consistency.

The presence of gradients in each parameter map confirms the energy dependence of $\{a_1, a_2, b_1, b_2\}$ mentioned in section 2.3. In order to find the average parameters and their standard deviation, only energy pairs within the $[10, 31] \times [10, 45]$ keV region were taken into account, as 31 keV and 45 keV are the respective energy cutoffs in the LE and HE spectrum. Results are shown in table 4.2.1.

Parameter	Mean Value	Standard Deviation
a_1	12.72	0.07
a_2	-11.16	0.11
b_1	0.1709	0.0015
b_2	0.8015	0.0022

TABLE 4.2.1. Basis material decomposition parameters

4.3. Material decomposition test

Figure 4.3.1 shows the calculated thickness and GR maps corresponding to the step-wedge array described in section 3.5 and illustrated in figure 3.5.1.
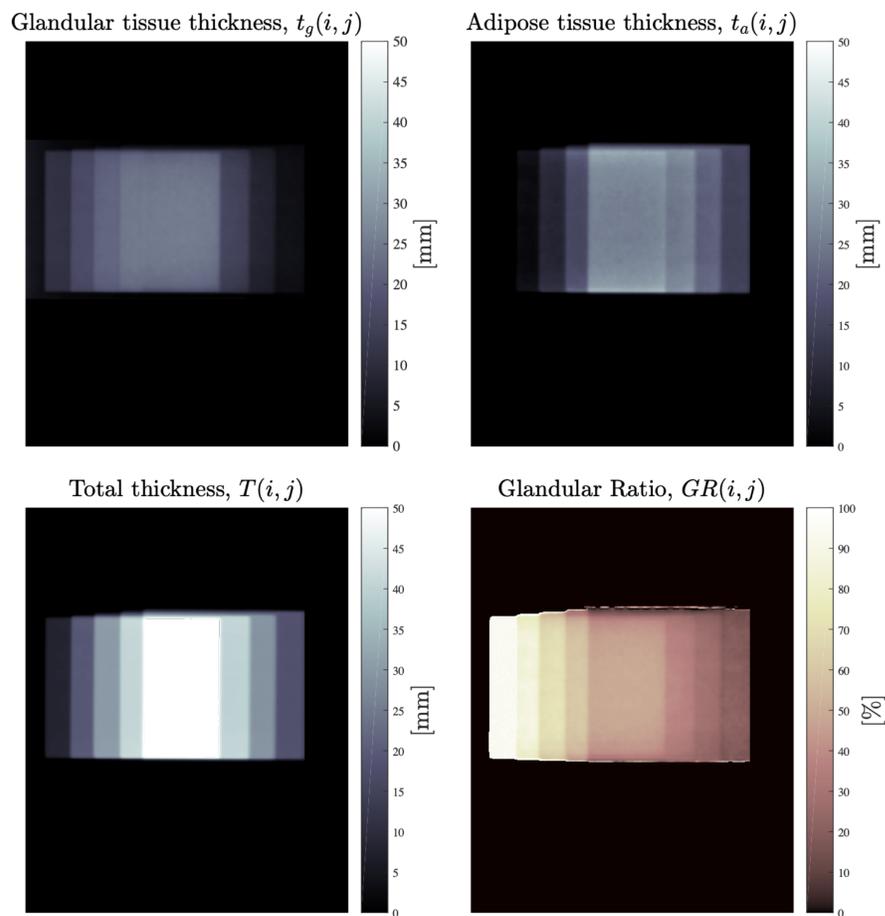

FIGURE 4.3.1. Tissue thickness and GR maps corresponding to the phantom configuration shown in figure 3.5.1

Table 4.3.1 compares the measured mean values inside each region to their nominal counterparts. Measured quantities are identified with an asterisk (*), and all thicknesses are reported in mm.

Region	t_g	t_g^*	t_a	t_a^*	T	T^*	$GR[\%]$	$GR^*[\%]$
1	10	11.26 ± 0.23	0	0.03 ± 0.60	10	11.29 ± 0.64	100	99.74 ± 2.06
2	17	16.83 ± 0.33	3	2.62 ± 0.37	20	19.45 ± 0.50	85	86.51 ± 1.68
3	22	21.50 ± 0.49	8	8.84 ± 0.58	30	30.34 ± 0.76	73.3	70.86 ± 1.61
4	25	24.32 ± 0.73	15	16.56 ± 0.84	40	40.88 ± 1.12	62.5	59.49 ± 1.79
5	25	24.53 ± 1.03	25	26.84 ± 1.25	50	51.37 ± 1.62	50	47.75 ± 2.01

TABLE 4.3.1. Tissue thickness and GR inside each one of the regions (steps) defined in figure 3.5.2.

The mean absolute errors in the calculations were as follows: for t_g, 0.61 mm; for t_a, 0.92 mm; and for GR, 1.89%. It is worth pointing out that regions 4 and 5 had practically the same calculated t_g, even though there was an additional adipose tissue slab in the latter region. This indicates that the Al-PMMA material separation discussed in section 4.1 translated to an adequate glandular and adipose tissue separation.

4.4. DEDM calculations in clinical images
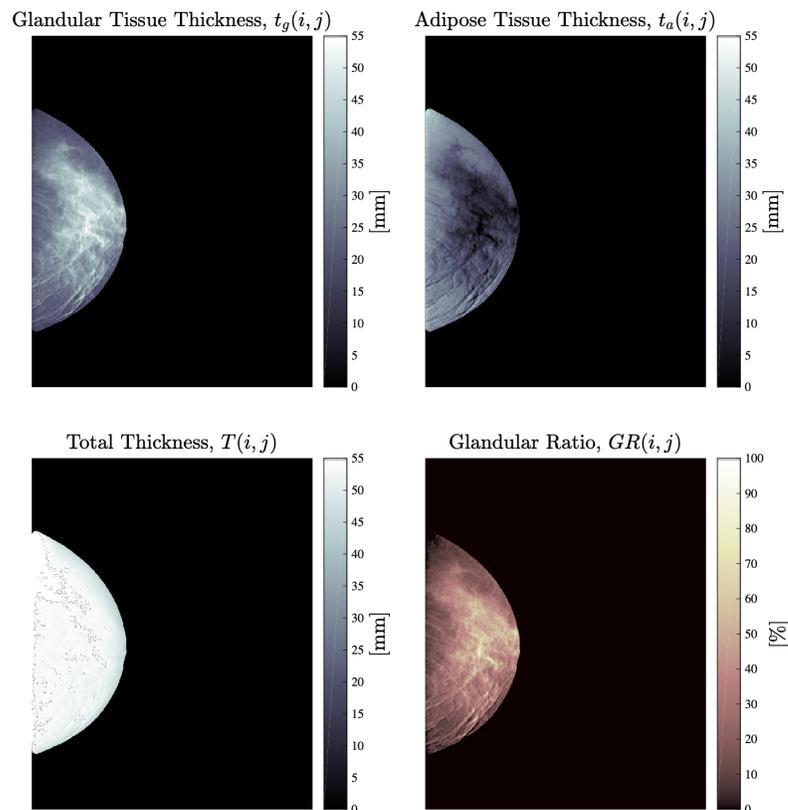

FIGURE 4.4.1. 2D maps generated from patient 10's set of DEDM images.

Figure 4.4.1 shows $t_g(i,j)$, $t_a(i,j)$, $T(i,j)$ and $GR(i,j)$ obtained from one patient's set of DEDM images. Such maps were generated for all 14 patient images, and glandular volume V_g, total breast volume, V_T and volumetric glandular ratio VGR were calculated. Table 4.4.1 shows the volumetric measurements (all reported in cm^3) obtained for each patient, using both DEDM and MRI.

ID	$V_{g,DEDM}$	$V_{T,DEDM}$	$VGR_{DEDM}(\%)$	$V_{g,MRI}$	$V_{T,MRI}$	$VGR_{MRI}(\%)$
1	179.3 ± 11.1	731.7 ± 40.6	24.5 ± 2.8	230.2 ± 23.4	792.8 ± 42.1	29.0 ± 3.3
2	168.6 ± 9.4	746.4 ± 39.9	22.6 ± 2.1	189.8 ± 21.5	813.2 ± 40.1	23.3 ± 2.9
3	215.3 ± 9.3	835.8 ± 40.2	25.8 ± 2.3	292.2 ± 35.1	1010.7 ± 49.3	28.9 ± 3.7
4	360.5 ± 17.3	844.0 ± 29.9	42.7 ± 2.5	312.7 ± 52.1	1149.1 ± 45.6	27.2 ± 4.7
5	261.6 ± 16.1	590.4 ± 30.0	44.3 ± 3.5	177.1 ± 28.9	609.8 ± 29.6	29.0 ± 4.9
6	154.4 ± 12.0	400.3 ± 24.7	38.6 ± 3.8	125.2 ± 18.0	463.8 ± 27.5	27.0 ± 4.2
7	428.7 ± 21.5	1061.3 ± 46.8	40.4 ± 2.7	409.3 ± 60.4	1600.0 ± 53.4	25.6 ± 3.9
8	152.3 ± 11.3	411.3 ± 27.0	37.0 ± 3.7	150.9 ± 22.7	575.1 ± 24.9	26.2 ± 4.1
9	96.5 ± 11.1	181.9 ± 24.9	53.0 ± 9.5	68.0 ± 3.8	155.2 ± 14.0	43.8 ± 4.7
10	166.9 ± 9.3	575.4 ± 16.3	29.0 ± 1.8	218.9 ± 24.1	756.4 ± 32.3	28.9 ± 3.4
11	88.4 ± 8.7	236.1 ± 27.0	47.4 ± 6.1	63.9 ± 6.6	203.7 ± 15.1	31.4 ± 4.0
12	165.9 ± 11.5	407.5 ± 14.4	40.7 ± 3.2	131.1 ± 12.9	383.4 ± 22.2	34.2 ± 3.9
13	117.1 ± 12.3	519.2 ± 23.2	22.6 ± 2.6	136.9 ± 22.2	523.5 ± 25.8	26.1 ± 4.4
14	178.3 ± 7.7	491.7 ± 29.7	36.3 ± 2.7	138.4 ± 19.4	503.6 ± 25.8	27.5 ± 4.1

TABLE 4.4.1. Volumetric measurements corresponding to each patient.

For a detailed description of the uncertainty calculations in both modalities, refer to appendix A.

VGR values calculated using DEDM were viewed as a function of three variables, namely: compressed breast thickness T_c, patient's age, and total breast volume. This was done because, as discussed in section 2.4, mammographic breast density has been found to exhibit particular behaviors as a function of the first two variables, and biopsy studies have provided insight between the breast density-volume relationship.

Breast density, as calculated using mammography, is known to decrease as a function of compressed breast thickness [42, 44, 72]. Figure 4.4.2 shows the VGR measurements for each of the 14 patients as a function of compressed breast thickness, as registered by the mammography station (T_c). Also included in this plot are breast density values predicted by Dance's model (as defined by equation 2.4.3).

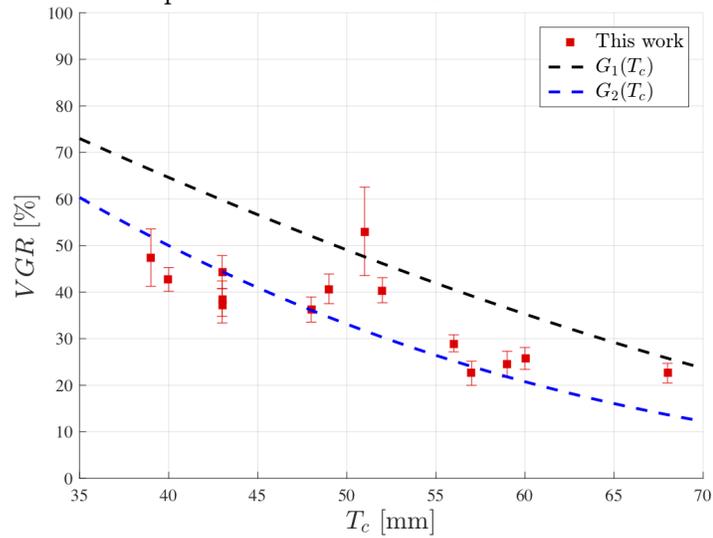Volumetric Glandular Ratio as a function of compressed breast thickness

FIGURE 4.4.2. VGR vs T_c plot. Dotted lines show Dance's model for two age groups: 40-49 years in black (G_1), and 50-64 in blue (G_2).

The VGR measured in this work tends to decrease as a function of T_c. In this sense, VGR calculated using DEDM shows the same behavior one would expect from "traditional" breast density measurements.

The age dependence of glandular fraction has also been noted in the literature [35, 46, 73, 74]. Across multiple studies, it has been observed that the glandular tissue fraction within the breast tends to decrease as age increases. Figure 4.4.3 shows this work's VGR plotted as a functon of age.
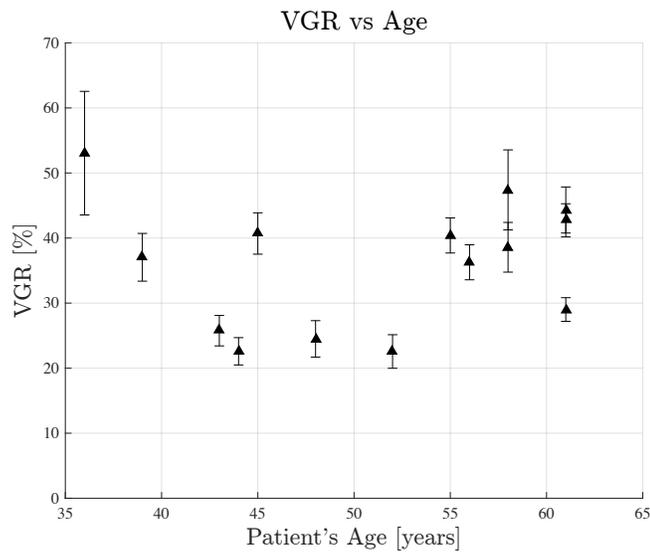

FIGURE 4.4.3. VGR vs patient's age.

The results obtained from the analyzed population show no apparent age dependence. However, the following caveat applies when comparing the presented results to those reported in other mammography-based studies:

As our sample size is considerably smaller than the population referenced in section 2.4, and given the racial differences in breast composition [**74**], we cannot conclude much about the conformance (or non-conformance) of the calculated values to their expected behaviors in terms of age and compressed breast thickness. Moreover, since our sample population consists of breast cancer suspects -who are likely to have higher breast densities, given its association with breast cancer-, an overall bias in breast glandularity is not out of the question.

Biopsy studies [**48, 49**] have shown that smaller breasts have greater areas of collagen and glands, which should result in a greater glandular density, or in our case, VGR. Figure 4.4.4 shows both V_g and V_T as a function of V_T.

Inspecting figure 4.4.4, it can be seen that there is a linear increase of V_g as a function of V_T. It can also be seen that the linear relationship has a slope lesser than 1, which means that while V_g increases for larger breasts (in terms of V_T), their ratio -VGR- gets smaller (both volumes progressively distance from one another in the plot).

This dependence can be better visualized in figure 4.4.5, where VGR is plotted as a function of V_T. A moderate negative correlation (Spearman's $\rho = -0.4$) between the two variables can be observed: smaller breast volumes tend to have higher glandular ratios. However, the low statistical significance of the correlation ($p = 0.17$) prevents us from confirming any total volume dependence. In this work, age, T_c and V_T were regarded as independent variables. A multivariate analysis would allow for the assessment of possible variable interactions and their effect on VGR. However, such analyses require larger databases to yield results with significant statistical power.
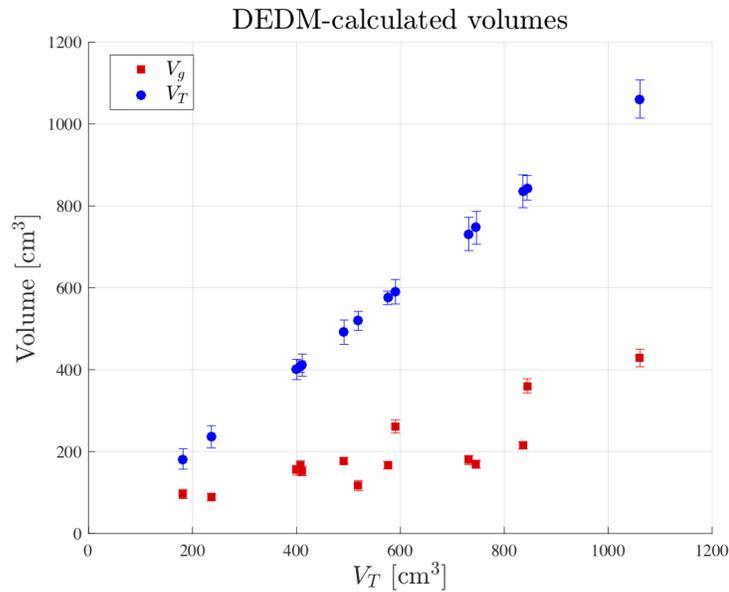

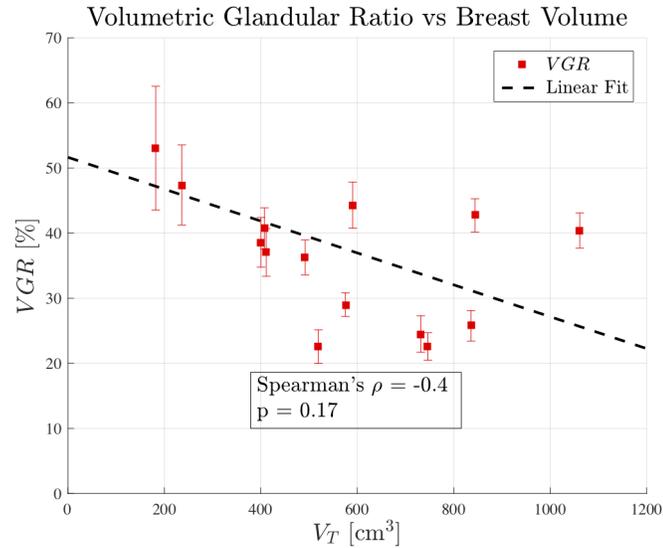FIGURE 4.4.4. V_g (red squares) and V_T (blue circles) as a function of V_T

FIGURE 4.4.5. VGR vs V_T. Dotted line shows a linear fit. A $p < 0.05$ value indicates statistical significance of the correlation.

4.5. DEDM-MRI comparison

In this last section, the correlation -and more importantly- agreement between DEDM and MRI volumetric measurements is discussed. Three variables were compared: V_g, V_T, and VGR.

Table 4.5.1 and figures 4.5.1 - 4.5.3 summarize both modalities' measurement distributions. The solid red line in the boxplots indicates the median of the distribution, while the box indicates the inter-quartile range (IQR), and the "whiskers" define the data range. Outliers are plotted as red crosses.
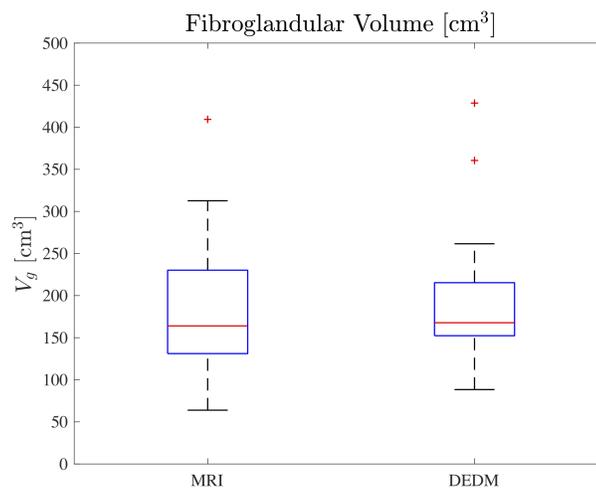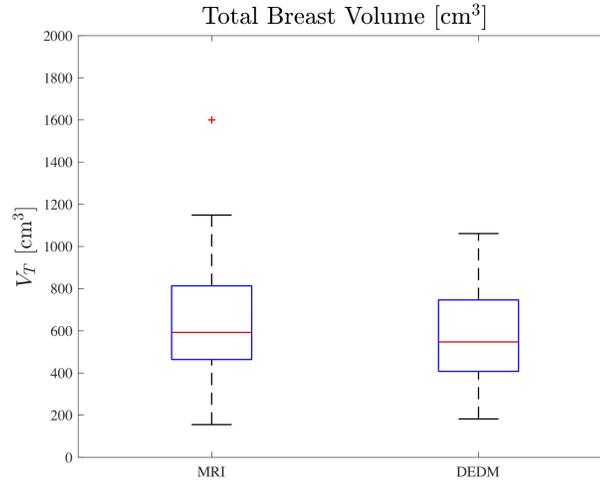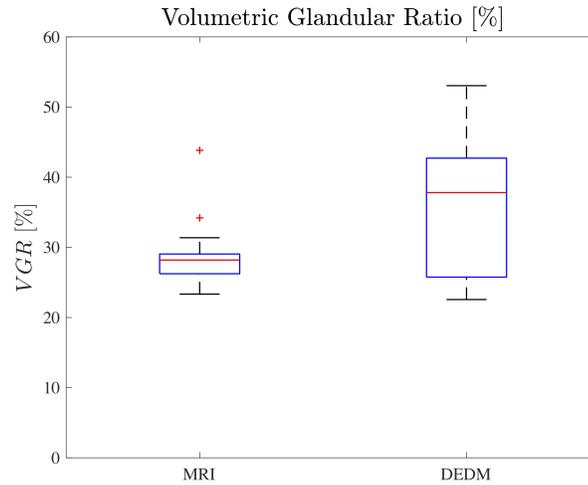

FIGURE 4.5.1. Boxplots summarizing the V_g distributions for each modality.

FIGURE 4.5.2. Boxplots summarizing the V_T distributions for each modality.

FIGURE 4.5.3. Boxplots summarizing the VGR distributions for each modality.

Quantity	Mean	Median	IQR	Min	Max
$V_{g,DEDM}$	195.29	167.78	53.46	88.38	428.74
$V_{g,MRI}$	188.92	164.02	94.83	63.90	409.32
$V_{T,DEDM}$	573.80	547.31	334.29	181.93	1061.27
$V_{T,MRI}$	681.45	592.47	334.37	155.23	1599.99
VGR_{DEDM}	36.06	37.81	15.64	22.56	53.05
VGR_{MRI}	29.17	28.20	2.61	23.34	43.83

TABLE 4.5.1. V_g, V_T and VGR distribution summary for DEDM and MRI. Values related to V_g and V_T are reported in cm^3, while those related to VGR are reported as percentages.

The distributions showed volume measurements were in similar ranges. On the other hand, the VGR_{MRI} range was markedly compressed when compared to its DEDM counterpart, as observed by Klifa et al [50].

Previous studies [53, 75, 76, 77] have noted the non-normality of the volume distributions. The Shapiro-Wilk test ($\alpha = 0.05$) was applied to the obtained volume and VGR distributions and it was found that, for some of them, the null hypothesis (the volume values do not follow a normal distribution within the sampled population) cannot be rejected. As normality cannot be assumed, there are two options: transform the data to make it normal, or run non-parametric tests. Given the small amount of samples, the latter option was deemed more adequate.

In order to assess the correlation between DEDM and MRI measurements, Spearman's correlation coefficients (ρ) were calculated. V_g and V_T showed strong correlations ($\rho \geq 0.7$) between DEDM and MRI measurements, while the correlation was only moderate ($0.5 < \rho < 0.7$) for VGR. For the three quantities, the correlations were statistically significant at a 95% confidence level ($p < 0.05$). These results are shown in table 4.5.2.

Quantity	ρ	p
V_g	0.87	< 0.001
V_T	0.98	< 0.001
VGR	0.58	0.03

TABLE 4.5.2. Correlation summary for DEDM and MRI measurements.

Figures 4.5.4 through 4.5.6 show the DEDM-calculated variables plotted against their MRI-calculated counterparts, along with linear fits.
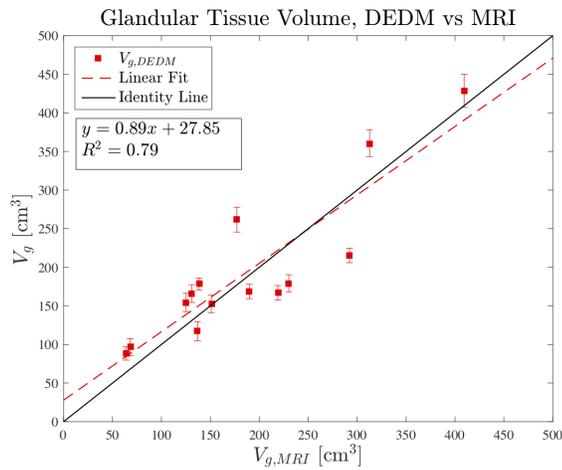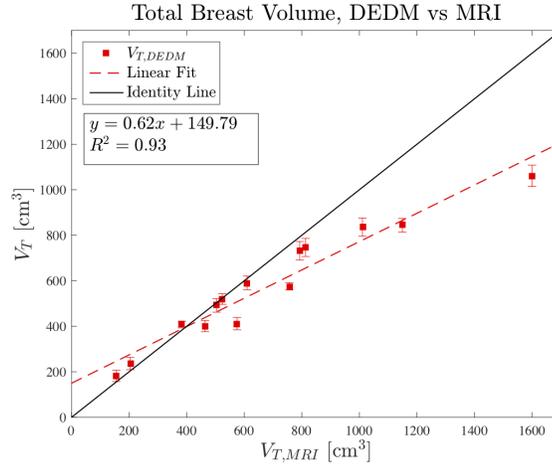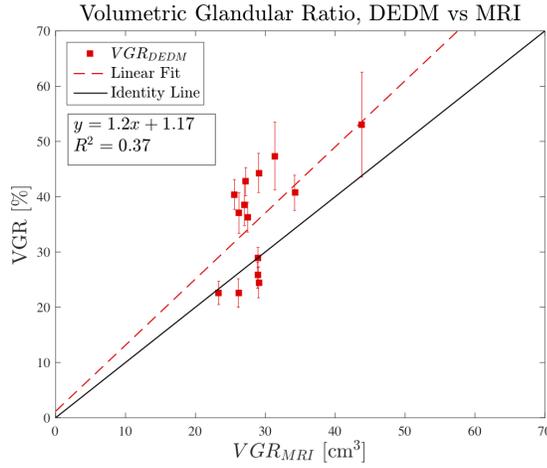

FIGURE 4.5.4. $V_{g,DEDM}$ vs $V_{g,MRI}$

FIGURE 4.5.5. $V_{T,DEDM}$ vs $V_{T,MRI}$

FIGURE 4.5.6. VGR_{DEDM} vs VGR_{MRI}

The results showed strong correlations between DEDM-calculated volumes and their MRI counterparts, and a moderate correlation between VGR measurements. However, correlation does not imply agreement. To assess the latter, the following distributions were considered:

$$\Delta V_g = V_{g,DEDM} - V_{g,MRI},$$

$$\Delta V_T = V_{T,DEDM} - V_{T,MRI,} \text{ and}$$

$$\Delta VGR = VGR_{g,DEDM} - VGR_{g,MRI}.$$

Figures 4.5.7 through 4.5.9 show boxplots corresponding to ΔV_g, ΔV_T and ΔVGR.
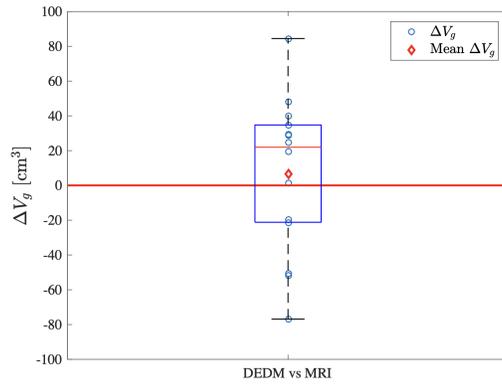
ΔV_g			
	Mean	**Median**	**p (sign)**
DEDM vs MRI	6.3750	21.9500	0.5830

FIGURE 4.5.7. ΔV_g distribution.
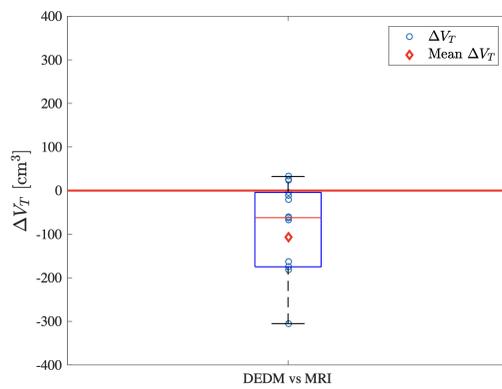
ΔV_T			
	Mean	**Median**	**p (sign)**
DEDM vs MRI	-107.65	-62.25	0.0166

FIGURE 4.5.8. ΔV_T distribution.
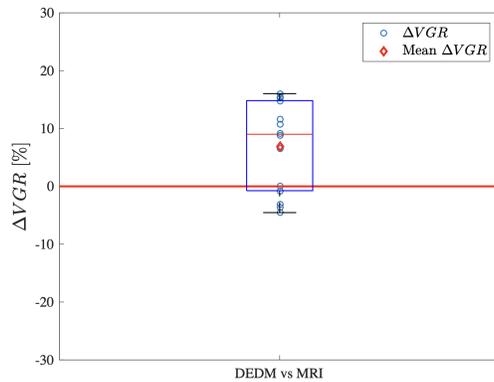
ΔVGR			
	Mean	**Median**	**p (sign)**
DEDM vs MRI	6.89	9	0.0134

FIGURE 4.5.9. ΔVGR distribution.

As these distributions inherit the previously discussed non-normality, the Wilcoxon signed rank test ($\alpha = 0.05$) was performed to test the null hypothesis that their medians are equal to zero. If the hypothesis holds ($p > 0.05$), we can conclude there is no statistically significant difference between the median difference and zero, and thus, there is agreement between the two measurements.

In order to assess the limits of agreement between the two modalities and to visualize any dependence or bias in the discrepancies, Bland-Altman plots were constructed (figures 4.5.10 through 4.5.12). These plots are constructed by plotting the differences between the two techniques as a function of their mean [78]. In the presented plots, dotted black lines indicate the lower and upper limits of agreement (LoA) i.e. the interval in which the differences are likely to be found. The red line indicates the mean of the differences. RPC stands for reproducibility coefficient, and CV stands for coefficient of variation. The Bland-Altman statistics were computed as follows :

$$LoA = \bar{\Delta V} \pm 1.96\sigma_{\Delta V},$$

$$RPC = 1.96\sigma_{\Delta V}, \quad \text{and}$$

$$CV = 100 \left(\frac{\sigma_{\Delta V}}{\bar{\Delta V}} \right),$$

where $\bar{\Delta V}$ represents the average difference and $\sigma_{\Delta V}$ its standard deviation.

Bland-Altman Measurement Agreement Summary for V_g

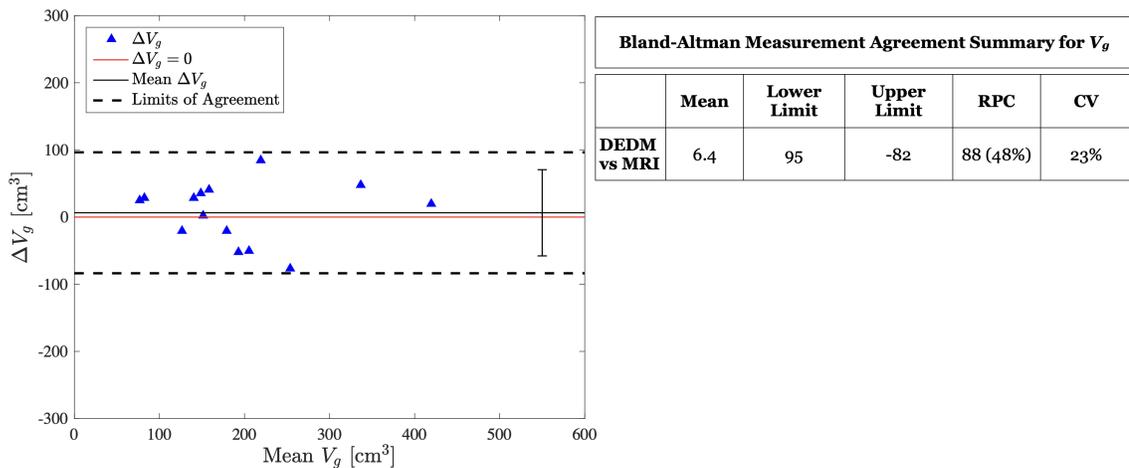|	Mean	Lower Limit	Upper Limit	RPC	CV
DEDM vs MRI	6.4	95	-82	88 (48%)	23%

FIGURE 4.5.10. Bland-Altman plot for V_g. Black error bar shows the largest uncertainty in ΔV_g
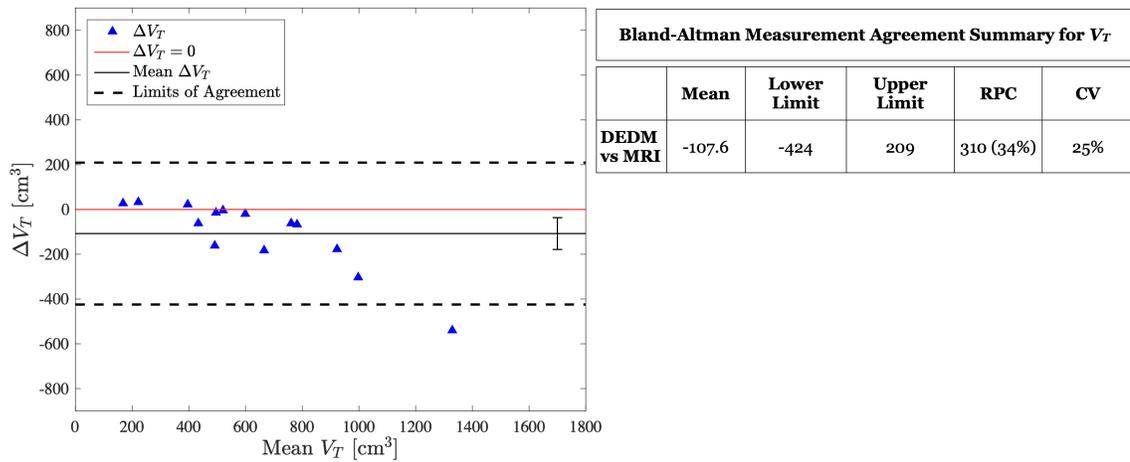
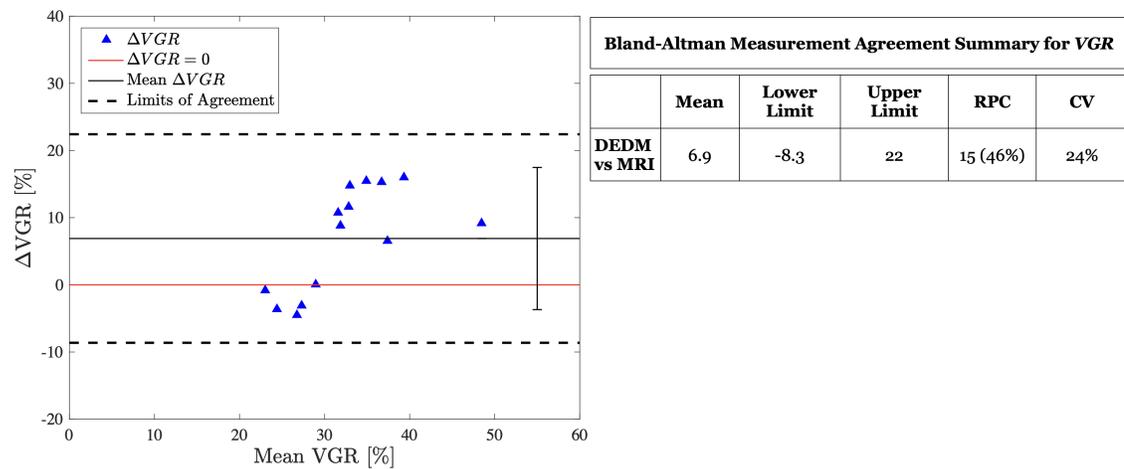FIGURE 4.5.11. Bland-Altman plot for V_T. Black error bar shows the largest uncertainty in ΔV_T

FIGURE 4.5.12. Bland-Altman plot for VGR. Black error bar shows the largest uncertainty in ΔVGR

The Wilcoxon signed rank test let us conclude that there was agreement in V_g, while statistically significant differences were found for V_T and VGR. Specifically, the proposed DEDM method showed a tendency to under-estimate V_T and in result, over-estimate VGR. Going back to the discussion in section 2.6, the over-estimation of VGR (or VBD), relative to MRI, is a common feature of mammography-based methods [51, 52, 54, 56, 79]. The mean differences in V_T (107.6 cm^3) and VGR (6.9%) are remarkably similar to those reported by Wang [52] (110.55 cm^3 and 6.48%, respectively). In our case, however, the disagreement in VGR was driven mainly by discrepancies in V_T.

The Bland-Altman plot for ΔV_T shows a trend in total volume discrepancy: it tends to be larger as the total breast volume increases. Also to be noted is that the range in VGR

is smaller in MRI, which is consistent with results reported in other MRI-mammography comparative studies [45].

As there are no standardized methods to calculate breast density using MRI nor digital mammography (besides commercial software in the latter modality), definitive comparisons between studies and modalities are hard to make. For instance, [52] reports a correlation coefficient of 0.38 in V_g when comparing Quantra (digital mammography) to MRI. Another study [51] comparing Quantra to MRI measurements reports a correlation coefficient of 0.15 in the same quantity. To cite a final example, [56] reports correlation coefficients of up to 0.86 when comparing V_g measurements made using an in-house mammography-based method to MRI.

When comparing image modalities, one must take into account that absolute volume agreement, or lack thereof, is influenced by each technique's field of view. The volumes present in the image are not exactly the same as result of factors such as MRI breast region segmentation, mammography placement, and amount of compression. Because of this, VGR is a more robust measurement and should be favored over absolute volume measurements, especially when comparing two different image modalities. This is elaborated further on appendix B.

The previously discussed results and literature suggest that breast density and volume measurements across different modalities may not be equivalent, and many variables, ranging from patient cohort characteristics to study acquisition methodology, must be considered before attempting to draw comparisons between studies and image modalities.

Since the clinical DEDM images used in this work were acquired after contrast medium injection, the presence of iodine in the analyzed breast is possible. The extent and significance of its presence are yet to be determined. Another limitation of this work is that it all images were acquired on a single mammography station, and the robustness of the calibration method remains to be assessed. Finally, a more advanced segmentation method (such as FCM) for MRI calculations could provide more accurate reference values, as the utilized thresholding method binarizes the image in glandular and adipose content, while other segmentation methods are able to provide fractional glandular/adipose values.

CHAPTER 5

Conclusions

In this work, a method that enables the measurement of breast glandular and adipose tissue in DEDM images was developed and evaluated. This method provides inherently volumetric measurements, which are comparable to those produced by image modalities such as MRI and breast CT. Unlike many commercially available mammography-based tools, the results provided by the proposed method have associated uncertainties. The method is fully automatable once the calibration images have been acquired, which represents an improvement over thresholding-based interactive software.

A DEDM image pre-processing routine was developed. This routine addresses factors such as background image inhomogeneity and noise, and it improves the material thickness calculations and computation times. As a result of this pre-processing, and given the fact that the calculations involve only arithmetic operations, the method is computationally inexpensive and it can be properly implemented in consumer-grade computers. It is also easy to calibrate since it only requires aluminum sheets and PMMA slabs, two commonly found materials in a clinical setting. However the calibration is not universal, and it must be carried out for each clinical DE protocol.

The method was applied to images from 14 patients acquired on a commercial mammography unit, and the results were compared with MRI volumetric measurements of the same patients. The MRI volumetric breast density measurements required the development of a custom segmentation routine based on adaptive local thresholding. To the author's knowledge, this work is the first to implement the dual-energy material decomposition formalism to process clinical mammography images and to compare the results with MRI measurements of the same patients.

The volumetric and breast density measurements produced by both modalities were in comparable ranges. Both correlation and agreement between the two modalities was assessed. Strong to moderate correlations were found in the three evaluated quantities (glandular volume, total volume and VGR), and good agreement was found in absolute glandular volume. There was also a good agreement in total volume for smaller breasts, but the discrepancy between both modalities increased for larger breasts. Because of this, it was found that the DEDM-based method tended to over-estimate breast density when compared to MRI. The difference in volume estimations between both techniques could be the result of the inherent field-of-view mismatch, as discussed in Appendix B.

As the number of processed images was quite small, no conclusions could be drawn about the population. Future works will seek to include a larger patient group, as well as to perform robustness assessments.

Bibliography

[1] JN Wolfe. The prominent duct pattern as an indicator of cancer risk. Oncology 1969.

[2] JA Tice, SR Cummings, E Ziv & K Kerlikowske. Mammographic Breast Density and the Gail Model for Breast Cancer Risk Prediction in a Screening Population. Breast Cancer Research and Treatment 94, 2005.

[3] Y Nagao, Y Kawaguchi, Y Sugiyama, S Saji & Y Kashiki, Relationship between Mammographic Density and the Risk of Breast Cancer in Japanese Women: A Case-Control Study. Breast Cancer 10 (3), 2003.

[4] MT Mandelson, N Oestreicher, PL Porter, D White, CA Finder et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 92, 2000.

[5] NF Boyd, GS Dite, J Stone, A Gunasekara, DR English, MR McCredie, GG Giles, D Tritchler, A Chiarelli A, & MJ Yaffe, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med 347(12), 2002.

[6] AT Wang, CM Vachon, KR Brandt & K Ghosh. Breast Density and Breast Cancer Risk. A Practical Review. Mayo Clinic Proceedings 89(4), 2014.

[7] K Kerlikowske, D Grady, J Barclay, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology breast imaging reporting and data system. JNCI 90, 1998.

[8] S Ciatto, N Houssami, A Apruzzese, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. Breast 14, 2005

[9] MJ Yaffe: Mammographic density. Measurement of mammographic density. Breast Cancer Research 10, 209, 2008

[10] DB Kopans. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. Radiology 246, 2008

[11] VolparaDensity. https://volparasolutions.com/our-products/volparadensity/

[12] Hologic Breast Analytics. https://www.hologic.com/hologic-products/breast-skeletal/image-analytics

[13] JH Chen, G Gulsen & M Su. Imaging Breast Density: Established and Emerging Modalities. Translational Oncology 8 (6), 2015.

[14] R Highnam, M Brady, MJ Yaffe, N Karssemeijer & J Harvey. Robust Breast Composition Measurement-Volpara. In: J Martí, A Olvier, J Freixenet, R Martí (eds) Digital Mammography. IWDM 2010. Lecture Notes in Computer Science, vol 6136, 2010

[15] EU Ekpo et al., Breast composition: Measurement and clinical use. Radiography, 2015.

[16] WR Brody, G Butt, A Hall & A Macovski. A method for selective tissue and bone visualization using dual energy scanned projection radiography. Med. Phys. 8(3), 1981.

[17] LA Lehmann, RE Alvarez, A Macovski & W.R. Brody. Generalized image combinations in dual KVP digital radiography. Med. Phys. 8(5), 1981.

[18] JL Ducote & S Molloi. Quantification of breast density with dual energy mammography: A simulation study. Med. Phys. 35 (12), 2008.

[19] JL Ducote & S Molloi. Quantification of breast density with dual energy mammography: An experimental feasibility study. Med. Phys. 37 (2), 2010.

[20] SC Kappadath & CC Shaw. Dual-energy digital mammography: Calibration and inverse mapping techniques to estimate calcification thickness and glandular tissue-ratio. Med. Phys. 30(6), 2003.

[21] Johns, P. C., & Yaffe, M. J. (1987). X-ray characterisation of normal and neoplastic breast tissues. Physics in Medicine and Biology, 32(6), 675–695

[22] HN Cardinal & A Fenster. An accurate method for direct dual energy calibration and decomposition. Med. Phys 17, 1990.

[23] Laidevant, A. D., Malkov, S., Flowers, C. I., Kerlikowske, K., & Shepherd, J. A. (2009). Compositional breast imaging using a dual-energy mammography protocol. Medical Physics, 37(1), 164–174.

[24] Ding, H., & Molloi, S. (2012). Quantification of breast density with spectral mammography based on a scanned multi-slit photon-counting detector: a feasibility study. Physics in Medicine and Biology, 57(15), 4719–4738.

[25] Lee, S., Choi, Y.-N., & Kim, H.-J. (2014). Quantification of breast composition by using a dual-energy technique with a photon-counting detector: Monte Carlo simulation studies. Journal of the Korean Physical Society, 64(2), 305–312.

[26] RE Alvarez and AMacovski. Energy-selective reconstructions in X-ray computerised tomography. Phys. Med. Biol. 21 (733), 1976.

[27] EL Gingold, BH Hasegawa. Dual-Energy X-ray Processing for Quantitative Projection Imaging. 1990 IEEE Nuclear Science Symposium Conference Record pp. 1147-1150, 1990.

[28] Sickles, EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology; 2013.

[29] Byng, J W; Boyd, N F; Fishell, E; Jong, R A; Yaffe, M J (1994). The quantitative analysis of mammographic densities. Physics in Medicine and Biology, 39(10), 1629–1638.

[30] Byng, J W; Yaffe, M J; Jong, R A; Shumak, R S; Lockwood, G A; Tritchler, D L; Boyd, N F (1998). Analysis of mammographic density and breast cancer risk from digitized mammograms.. RadioGraphics, 18(6), 1587–1598.

[31] Ursin, G., Astrahan, M. A., Salane, M., Parisky, Y. R., Pearce, J. G., Daniels, J. R., Pike, M. C., & Spicer, D. V. (1998). The detection of changes in mammographic densities. Cancer Epidemiology Biomarkers and Prevention, 7(1), 43-47.

[32] Tagliafico, A., Tagliafico, G., Tosto, S., Chiesa, F., Martinoli, C., Derchi, L. E., & Calabrese, M. (2009). Mammographic density estimation: Comparison among BI-RADS categories, a semi-automated software and a fully automated one. The Breast, 18(1), 35–40.

[33] Keller, B. M., Nathan, D. L., Wang, Y., Zheng, Y., Gee, J. C., Conant, E. F., & Kontos, D. (2012). Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Medical Physics, 39(8), 4903–4917.

[34] Eriksson, M., Li, J., Leifland, K., Czene, K., & Hall, P. (2018). A comprehensive tool for measuring mammographic density changes over time. Breast Cancer Research and Treatment, 169(2), 371–379.

[35] Yaffe, M. J., Boone, J. M., Packard, N., Alonzo-Proulx, O., Huang, S. Y., Peressotti, C. L., Al-Mayah, A., & Brock, K. (2009). The myth of the 50-50 breast. Medical Physics, 36(12), 5437-5443.

[36] Diffey, J., Hufton, A., & Astley, S. (2006). A New Step-Wedge for the Volumetric Measurement of Mammographic Density. In Digital Mammography (pp. 1–9). Springer Berlin Heidelberg.

[37] Shepherd, J. A.; Herve, L.; Landau, J.; Fan, B.; Kerlikowske, K.; Cummings, S. R. (2005). Novel use of Single X-Ray Absorptiometry for Measuring Breast Density. Technology in Cancer Research & Treatment, 4(2), 173–182.

[38] Kaufhold, J.; Thomas, J. A.; Eberhard, J. W.; Galbo, C. E.; Trotter, D. E. González (2002). A calibration approach to glandular tissue composition estimation in digital mammography. Medical Physics, 29(8), 1867–.

[39] Highnam, R., Brady, M., & Shepstone, B. (1996). A representation for mammographic image processing. Medical Image Analysis, 1(1), 1–18.

[40] Highnam, R; Pan, X; Warren, R; Jeffreys, M; Smith, G Davey; Brady, M (2006). Breast composition measurements using retrospective standard mammogram form (SMF). Physics in Medicine and Biology, 51(11), 2695–2713.

[41] Hartman, K., Highnam, R., Warren, R., & Jackson, V. (2008). Volumetric Assessment of Breast Tissue Composition from FFDM Images. In Digital Mammography (pp. 33–39). Springer Berlin Heidelberg.

[42] Klein, R., Aichinger, H., Dierker, J., Jansen, J. T. M., Joite-Barfuß, S., Säbel, M., Schulz-Wendtland, R., & Zoetelief, J. (1997). Determination of average glandular dose with modern mammography units for two large groups of patients. Physics in Medicine and Biology, 42(4), 651–671

[43] Young, K. C., Ramsdale, M. L., & Bignell, F. (1998). Review of Dosimetric Methods for Mammography in the UK Breast Screening Programme. Radiation Protection Dosimetry, 80(1), 183–186.

[44] Dance, D. R., Skinner, C. L., Young, K. C., Beckett, J. R., & Kotre, C. J. (2000). Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol. Physics in Medicine and Biology, 45(11), 3225–3240.

[45] Stomper, P. C., D'Souza, D. J., DiNitto, P. A., & Arredondo, M. A. (1996). Analysis of parenchymal density on mammograms in 1353 women 25-79 years old. American Journal of Roentgenology, 167(5), 1261–1265.

[46] Alonzo-Proulx, O., Jong, R. A., & Yaffe, M. J. (2012). Volumetric breast density characteristics as determined from digital mammograms. Physics in Medicine and Biology, 57(22), 7443–7457

[47] Huang, S.-Y., Boone, J. M., Yang, K., Packard, N. J., McKenney, S. E., Prionas, N. D., Lindfors, K. K., & Yaffe, M. J. (2011). The characterization of breast anatomical metrics using dedicated breast CT. Medical Physics, 38(4), 2180–2191.

[48] Li, T. (2005). The Association of Measured Breast Tissue Characteristics with Mammographic Density and Other Risk Factors for Breast Cancer. Cancer Epidemiology Biomarkers & Prevention, 14(2), 343–349.

[49] Vandeweyer, E., & Hertens, D. (2002). Quantification of glands and fat in breast tissue: An experimental determination. Annals of Anatomy - Anatomischer Anzeiger, 184(2), 181–184.

[50] C Klifa, J Carballido-Gamio, L Wilmes, A Laprie, J Shepherd, et al. Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. Magn Reson Imaging 28, 2010.

[51] Kontos, D., Xing, Y., Bakic, P. R., Conant, E. F., & Maidment, A. D. A. (2010). A comparative study of volumetric breast density estimation in digital mammography and magnetic resonance imaging: results from a high-risk population. In N. Karssemeijer & R. M. Summers (Eds.), Medical Imaging 2010: Computer-Aided Diagnosis. SPIE.

[52] Wang, J., Aziz, A., Newitt, D., Joe, B. N., Hylton, N., & Shepherd, J. A. (2012). Comparison of Hologic's Quantra Volumetric Assessment to MRI Breast Density. In Breast Imaging (pp. 619–626).

[53] Wang, J., Azziz, A., Fan, B., Malkov, S., Klifa, C., Newitt, D., Yitta, S., Hylton, N., Kerlikowske, K., & Shepherd, J. A. (2013). Agreement of Mammographic Measures of Volumetric Breast Density to MRI. PLoS ONE, 8(12), e81653.

[54] Tagliafico, A., Tagliafico, G., Astengo, D., Airaldi, S., Calabrese, M., & Houssami, N. (2013). Comparative estimation of percentage breast tissue density for digital mammography, digital breast tomosynthesis, and magnetic resonance imaging. Breast Cancer Research and Treatment, 138(1), 311–317.

[55] Pertuz, S., McDonald, E. S., Weinstein, S. P., Conant, E. F., & Kontos, D. (2016). Fully Automated Quantitative Estimation of Volumetric Breast Density from Digital Breast Tomosynthesis Images: Preliminary Results and Comparison with Digital Mammography and MR Imaging. Radiology, 279(1), 65–74.

[56] Holland, K., Gubern-Mérida, A., Mann, R. M., & Karssemeijer, N. (2017). Optimization of volumetric breast density estimation in digital mammograms. Physics in Medicine and Biology, 62(9), 3779–3797.

[57] Chen, J.-H., Chan, S., Lu, N.-H., Li, Y., Tsai, Y. C., Huang, P. Y., Chang, C.-J., & Su, M.-Y. (2016). Opportunistic Breast Density Assessment in Women Receiving Low-dose Chest Computed Tomography Screening. Academic Radiology, 23(9), 1154–1161.

[58] CH Wei, Y Li, PJ Huang, CY Gwo, & SE Harms. Estimation of breast density: An adaptive moment preserving method for segmentation of fibroglandular tissue in breast magnetic resonance images. European Journal of Radiology, 81(4), 2012

[59] BHM van der Velden, MHA Janse, MAA Ragusi, CE Loo, & KGA Gilhuijs. Volumetric breast density estimation on MRI using explainable deep learning regression. Scientific Reports, 10(1), 2020.

[60] A Tagliafico, B Bignotti, G Tagliafico, D Astengo, L Martino, et al. Breast Density Assessment Using a 3T MRI System: Comparison among Different Sequences. PLoS ONE 9(6), 2014.

[61] WT Dixon. Simple proton spectroscopic imaging. Radiology, 153(1), 1984.

[62] Reeder, S. B., Pineda, A. R., Wen, Z., Shimakawa, A., Yu, H., Brittain, J. H., Gold, G. E., Beaulieu, C. H., & Pelc, N. J. (2005). Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging. Magnetic Resonance in Medicine, 54(3)

[63] Amulet Innovality | Fujifilm Global. https://www.fujifilm.com/products/medical/products/digital_mammography/innovality/

[64] Public Health England. Technical evaluation of Fujifilm AMULET Innovality digital breast tomosynthesis system. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685174/Technical_evaluation_of_Fujifilm_AMULET_Innovality_digital_breast_tomosynthesis_system.pdf

[65] Castillo Lopez, J. P., Brandan, M. E., Berumen, F., Villaseñor-Navarro, Y., Galván, H. A., & Cruz Rodríguez, J. C. (2018). Optimization of acquisition parameters for the detection of secondary breast lesions applying temporal contrast enhanced digital mammography. In E. A. Krupinski (Ed.), 14th International Workshop on Breast Imaging (IWBI 2018). Proc. of SPIE Vol. 10718

[66] ImageJ: Downsample. https://imagej.net/Downsample

[67] Gingold, E. L., & Hasegawa, B. H. (1992). Systematic bias in basis material decomposition applied to quantitative dual-energy x-ray imaging. Medical Physics, 19(1), 25–33.

[68] NIST XCOM: Element/Compound/Mixture.Retrieved from https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html

[69] Hammerstein, G. R., Miller, D. W., White, D. R., Masterson, M. E., Woodard, H. Q., & Laughlin, J. S. (1979). Absorbed Radiation Dose in Mammography. Radiology, 130(2), 485–491.

[70] Wellner, P. D. (1993). Adaptive thresholding for the digitaldesk. Tech. Rep. EPC-93-110, EuroPARC

[71] Bradley, D., & Roth, G. (2007). Adaptive Thresholding using the Integral Image. Journal of Graphics Tools, 12(2), 13–21.

[72] Kawaguchi, A., Matsunaga, Y., Asada, Y., Suzuki, S., & Chida, K. (2016). Glandularity Estimation in Japanese Women by Using a Breast Model Made from Mammographic Findings of European Women. Japanese Journal of Health Physics, 51(4), 238–244

[73] Ziv, E., Shepherd, J., Smith-Bindman, R., & Kerlikowske, K. (2003). Mammographic Breast Density and Family History of Breast Cancer. JNCI Journal of the National Cancer Institute, 95(7), 556–558.

[74] Nie, K., Su, M.-Y., Chau, M.-K., Chan, S., Nguyen, H., Tseng, T., Huang, Y., McLaren, C. E., Nalcioglu, O., & Chen, J.-H. (2010). Age- and race-dependence of the fibroglandular breast density analyzed on 3D MRI. Medical Physics, 37(6Part1), 2770–2776.

[75] Morrish, O. W. E., Tucker, L., Black, R., Willsher, P., Duffy, S. W., & Gilbert, F. J. (2015). Mammographic Breast Density: Comparison of Methods for Quantitative Evaluation. Radiology, 275(2), 356–365.

[76] Eng, A., Gallant, Z., Shepherd, J., McCormack, V., Li, J., Dowsett, M., Vinnicombe, S., Allen, S., & dos-Santos-Silva, I. (2014). Digital mammographic density and breast cancer risk: a case–control study of six alternative density assessment methods. Breast Cancer Research, 16(5).

[77] Thompson, D. J., Leach, M. O., Kwan-Lim, G., Gayther, S. A., Ramus, S. J., Warsi, I., Lennard, F., Khazen, M., Bryant, E., Reed, S., Boggis, C. R., Evans, D. G., Eeles, R. A., Easton, D. F., & Warren, R. M. (2009). Assessing the usefulness of a novel MRI-based breast density estimation algorithm in a cohort of women at high genetic risk of breast cancer: the UK MARIBS study. Breast Cancer Research, 11(6).

[78] Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet (London, England), 1(8476), 307–310.

[79] Moon, W. K., Lo, C.-M., Goo, J. M., Bae, M. S., Chang, J. M., Huang, C.-S., Chen, J.-H., Ivanova, V., & Chang, R.-F. (2014). Quantitative Analysis for Breast Density Estimation in Low Dose Chest CT Scans. Journal of Medical Systems, 38(3).

[80] Rahbar, K., Gubern-Merida, A., Patrie, J. T., & Harvey, J. A. (2017). Automated Volumetric Mammographic Breast Density Measurements May Underestimate Percent Breast Density for High-density Breasts. Academic Radiology, 24(12), 1561–1569.

APPENDIX A

Uncertainty calculations

The uncertainty associated to the proposed volumetric DEDM measurements has three components:

- Calibration uncertainty (repeatability).
- Basis material decomposition uncertainty.
- Processing method (breast masking).

The calibration uncertainty component was obtained by imaging 3 calibration sets and finding the corresponding fit parameters for each. The average $\bar{p}_i$ and standard deviation σ_{p_i} of each parameter set was obtained, and the former was taken to be the uncertainty associated to that parameter. The parameter uncertainties were propagated in order to generate Al and PMMA thickness uncertainties δt_i:

$$\text{(A.0.1)} \qquad \delta t_j = \sqrt{\sum_i \left(\frac{\delta t_j}{\delta p_i}\right)^2 (\delta p_i)^2}$$

where t_j represents the fitted functions for either Al or PMMA thickness, shown in table 4.1.1. Equation A.0.1 was applied pixel-by-pixel to generate material thickness uncertainty maps, $\delta t_{Al}(i,j)$ and $\delta t_P(i,j)$.

Next up is the basis material decomposition uncertainty component. The standard deviations of $\{a_1, a_2, b_1, b_2\}$ (table 4.2.1) were taken as the uncertainties of these parameters. These were propagated, pixel by pixel, from the material change equations 2.3.5 and the Al-PMMA uncertainty maps A.0.1. This resulted in glandular and adipose tissue uncertainty maps $\delta t_g(i,j)$ and $\delta t_a(i,j)$.

Finally, the breast masking introduces its own independent uncertainty component. There were two controllable variables in this process: the gradient intensity G_0 used as threshold, and the breast edge erosion radius r.

G_0 defines the edge section to be removed from the image. By quantifying the changes in volume produced by varying G_0 in an interval $G_o \in [G^* - \Delta G, G^* + \Delta G]$, an uncertainty was associated to the variation in this parameter. Figure A.0.1 illustrates the effect of varying G_0.
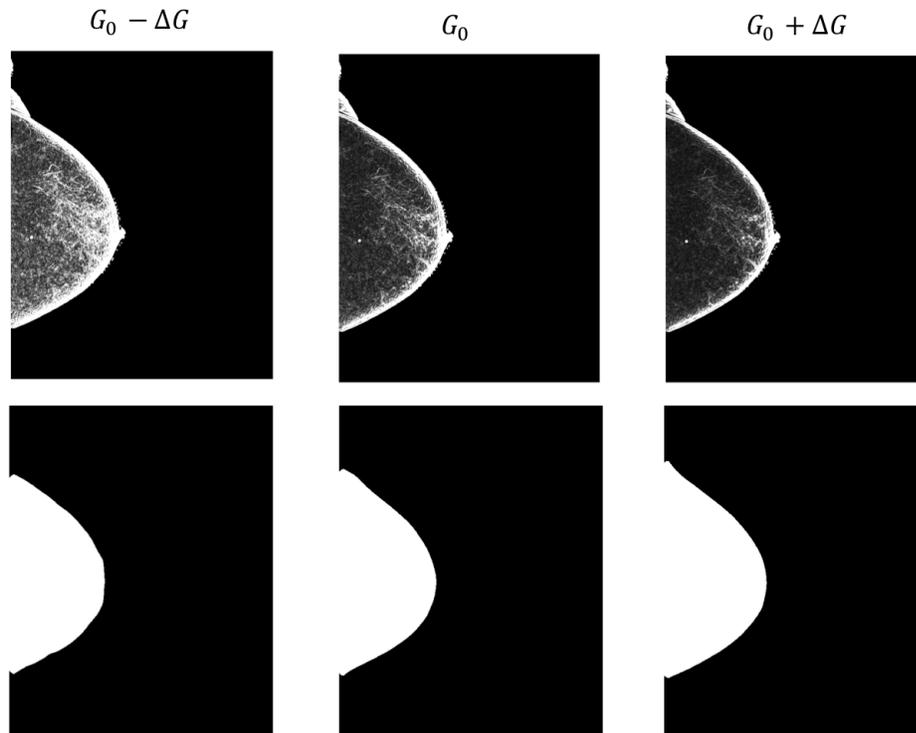
FIGURE A.0.1. Masks generated by different G_0 values. The resulting area difference, when generating the mask using $G_0 = 0.005$ (left) and $G_0 = 1.5$ (right), was 15.48 cm^2.

The second variable, r, also defines the mask dimensions, and therefore the calculated volume. Figure A.0.2 illustrates this effect.
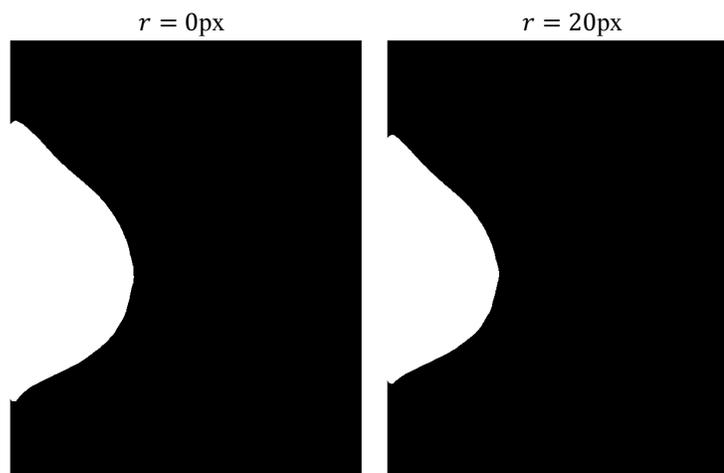

FIGURE A.0.2. Masks generated by using $r = 0$ px and $r = 20$ px as erosion radius. The area difference between both masks was 22.31 cm^2.

For each patient image, 220 masks were generated. These masks feature varying values of G_0 and r: $G_0 \in [0.005, 1.500]$ (in steps of 0.001) and $r \in [0, 20]px$ (in steps of 1px). Volume calculations were carried out using the different masks, and the average and standard deviation were calculated. Figure A.0.3 shows an example of the 220 volume points for one patient.
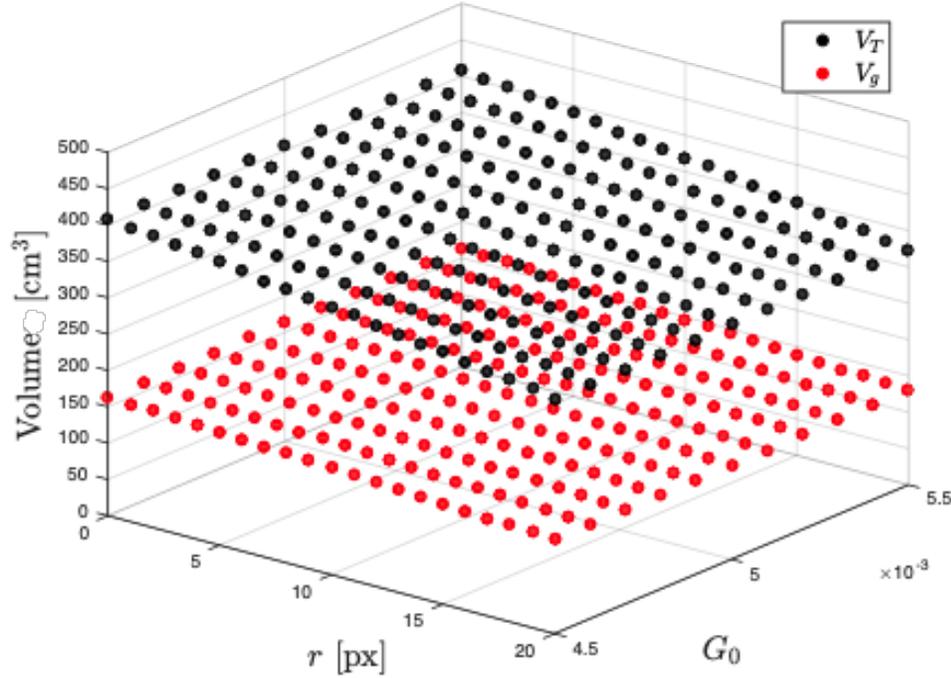

FIGURE A.0.3. Calculated total and glandular volume distributions (for a single patient) obtained by varying mask parameters G_0 and r.

The standard deviations of these volume distributions are denoted by $\sigma_{M,g}$ (glandular) and $\sigma_{M,T}$ (total), and they were the uncertainty component associated to the masking process. The average masking uncertainty components were 4.5% for glandular volume and 6% for total volume. These values represent the change in breast area -and thus calculated volume- when the mask parameters are chosen differently. The variation was, predictably, larger in total volume than in glandular volume, since the latter is mainly concentrated towards the chest wall, where masking had a reduced effect. Considering the three components, the uncertainty in glandular and total volume was calculated as:

$$\delta V_g = \sqrt{dA^2 \sum_{i,j} (\delta t_g(i,j))^2 + \sigma_{M,g}^2}, \text{ and}$$

$$\delta V_T = \sqrt{dA^2 \sum_{i,j} (\delta T(i,j))^2 + \sigma_{M,T}^2}.$$

Finally, the uncertainty in VGR was given by:

$$\delta VGR = \sqrt{\frac{V_T^2 \delta V_g^2 + V_g^2 \delta V_T^2}{V_T^4}} * 100.$$

For DEDM, average relative uncertainty in VGR was close to 10%.

Compared to the DEDM uncertainty calculation, MRI was fairly straightforward. In this case, the following image processing parameters had an effect on the calculated volumes:

- Erosion radius r used for breast mask generation.
- Adaptive thresholding neighborhood dimension N.
- Adaptive thresholding percentage threshold T.

For each patient, volume measurements were performed varying the $[r, N, T]$ parameters. Figure A.0.4 shows an example of the resulting distribution.
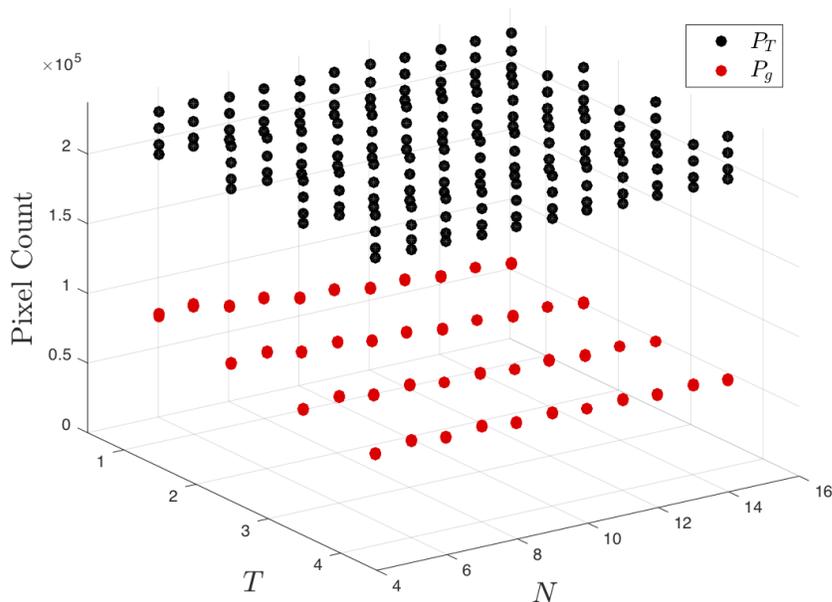
FIGURE A.0.4. Calculated volume distribution obtained by varying parameter set $[r, N, T]$. $r \in [0, 3]$ px, $N \in [5, 15]$, and $T \in [1, 4]$.

The standard deviations of these volume distributions were denoted by $\sigma_{MRI,g}$ (glandular) and $\sigma_{MRI,T}$ (total), and they were the uncertainty component associated to the MRI volume calculations. Therefore, the VGR uncertainty was given by

$$\delta VGR_{MRI} = \sqrt{\frac{V_T^2 \sigma_{MRI,g}^2 + V_g^2 \sigma_{MRI,T}^2}{V_T^4}} * 100$$

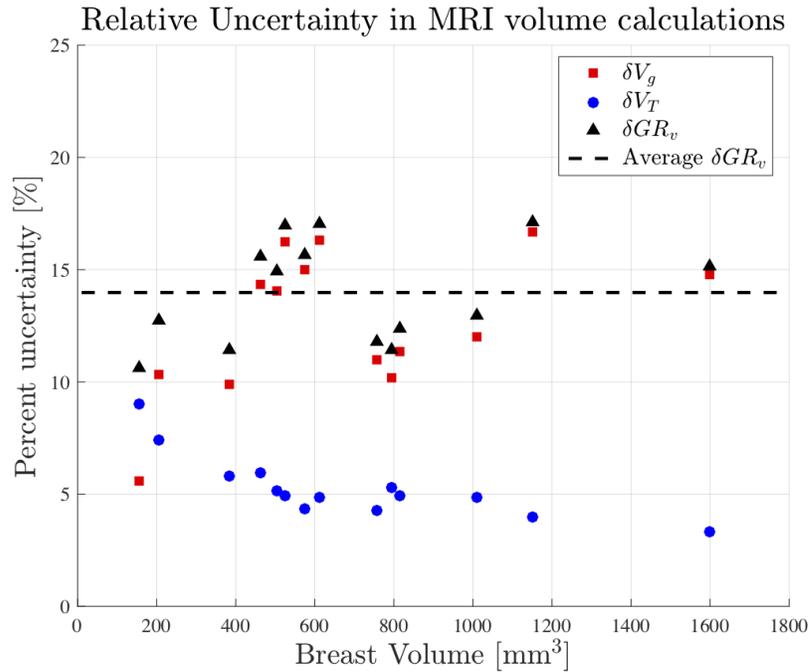Figure shows a breakdown of the uncertainty calculations for $V_{g,MRI}$, $V_{T,MRI}$ and VGR_{MRI}.

FIGURE A.0.5. Relative uncertainty in MRI calculations as a function of total breast volume.

Average relative uncertainty in VGR_{MRI} was close to 14%, and the main contributing factor was the uncertainty in glandular volume estimation i.e. the change produced by varying the adaptive thresholding parameters.

The uncertainties associated to the data shown in table 4.4.1, as well as the plots presented throughout sections 4.4 and 4.5, were calculated using the methods described in this section.

There is one additional uncertainty component, namely: that related to image acquisition. For DEDM, this encompasses patient positioning differences, compression force, and the use of different mammography units. As shown in appendix B, VGR is not susceptible to changes in the field of view resulting from positioning differences. The effect of the other two factors could be evaluated as part of a future robustness assessment. The extent to which acquisition and reconstruction parameters could affect the MRI measurements is yet to be investigated.

Field-of-view mismatch between MRI and DEDM

As mentioned in section 4.5, the breast region present in each modality is likely to be different and as such, each modality measures different organ volumes. In order to assess the difference produced by a possible "Field-of-view (FOV) mismatch", the nipple-pectoral distance (NPD) was measured. This measurement was carried out in the DEDM images and in the center slice of the MRI volume, where the breast projection is largest.

Across all DEDM images, NPD was smaller than in MRI. In fact, the pectoral muscle wasn't visible in the mammography projections. Figure B.0.1 shows ΔV_T as a function of $\Delta NPD = NPD(MRI) - NPD(DEDM)$.
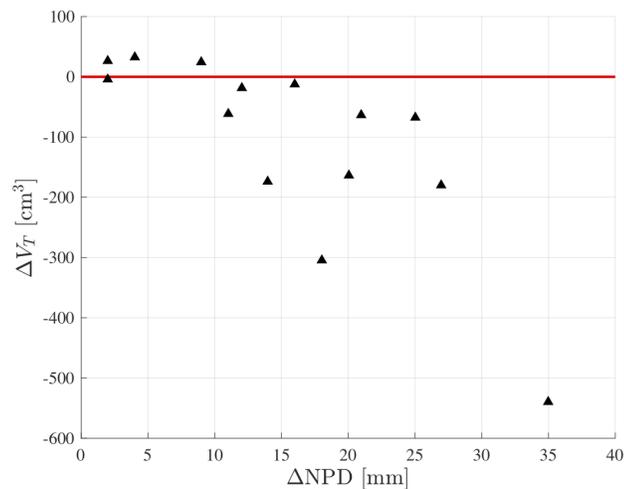
FIGURE B.0.1. ΔV_T vs ΔNPD plot for the 14 patients included in this work.

The difference in total volume increased as the difference in NPD did. Because of this, differences in V_T across both modalities -as observed in figure 4.5.11- could be attributed to the FOV mismatch.

Next, MRI breast regions were manually redefined to match the DEDM NPD measurements, and volumetric measurements were performed using the newly defined regions. Figures B.0.2, B.0.3 and B.0.4 show boxplots of the ΔV_g, ΔV_T and ΔVGR distributions obtained using NPD-corrected MRI volumes next to the ones obtained without such correction.
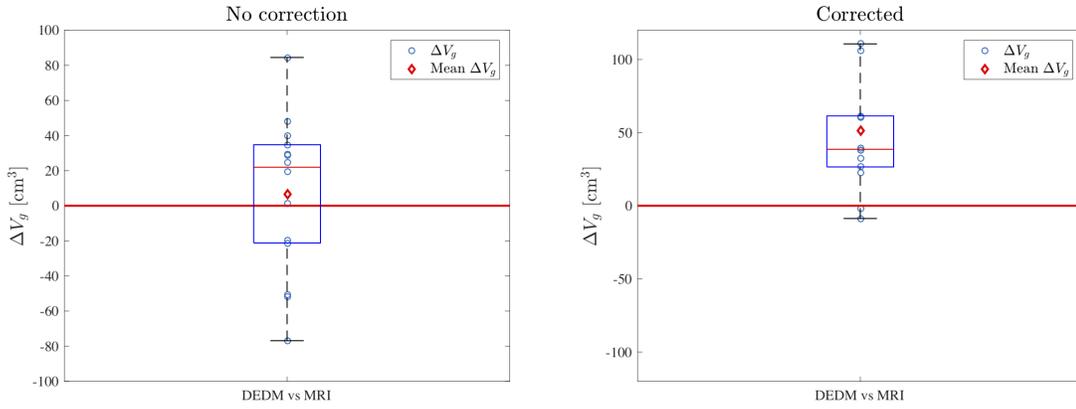
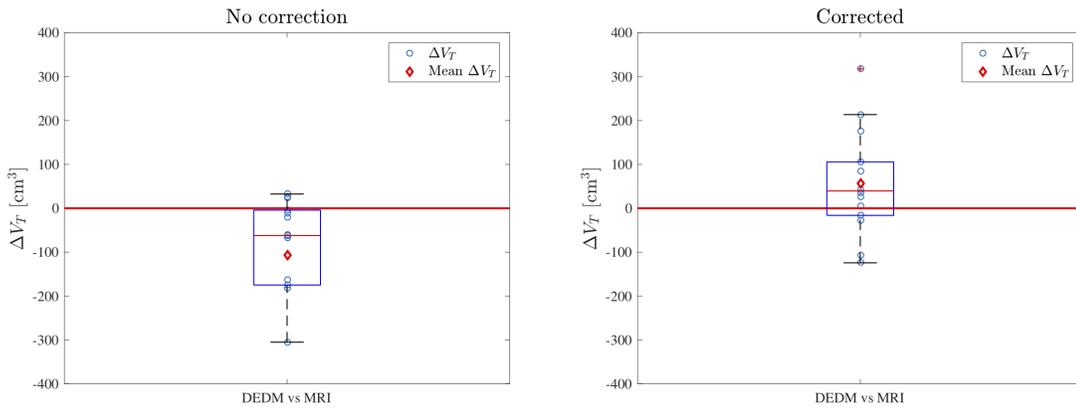FIGURE B.0.2. Effects of NPD correction on the ΔV_g distribution.

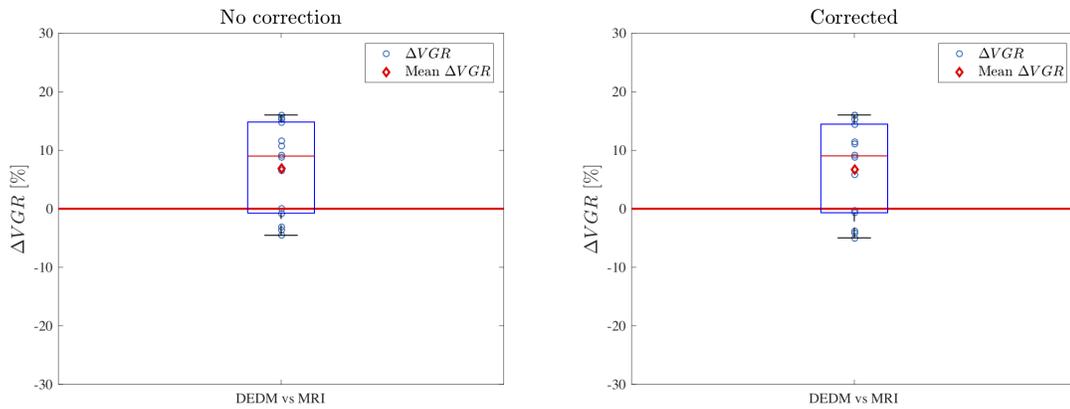FIGURE B.0.3. Effects of NPD correction on the ΔV_T distribution.

FIGURE B.0.4. Effects of NPD correction on the ΔVGR distribution.

The NPD correction brought an agreement to the total volume measurements, but increased the discrepancies in glandular volume. However, the VGR distribution remained practically unaffected. Table B.0.1 summarizes the distributions.

Quantity	Mean	Median	p (Wilcoxon test)
ΔV_g^*	6.37	21.95	0.58
ΔV_g	51.07	38.60	6.1E-04
ΔV_T^*	-107.65	-62.25	0.016
ΔV_T	56.46	39.42	0.119
ΔVGR^*	6.89	9.00	0.013
ΔVGR	6.7	9.03	0.016

TABLE B.0.1. ΔV_g, ΔV_T and ΔVGR distribution summary. Values related to V_g and V_T are reported in cm^3, while those related to VGR are reported as percentages. Asterisks indicate values without NPD correction.

To clearly see the effect the NPD correction had on V_g and V_T, the NPD-corrected MRI volumes were plotted as a function of MRI volumes without correction (figures B.0.5 and B.0.6).

The slopes of both fits were practically the same for both volumes ($a = 0.7379$ for glandular volume, $a = 0.7391$ for total volume). This means both volumes were reduced in the same proportion (approximately 26%). Due to this, their ratio remained constant (figure B.0.7).
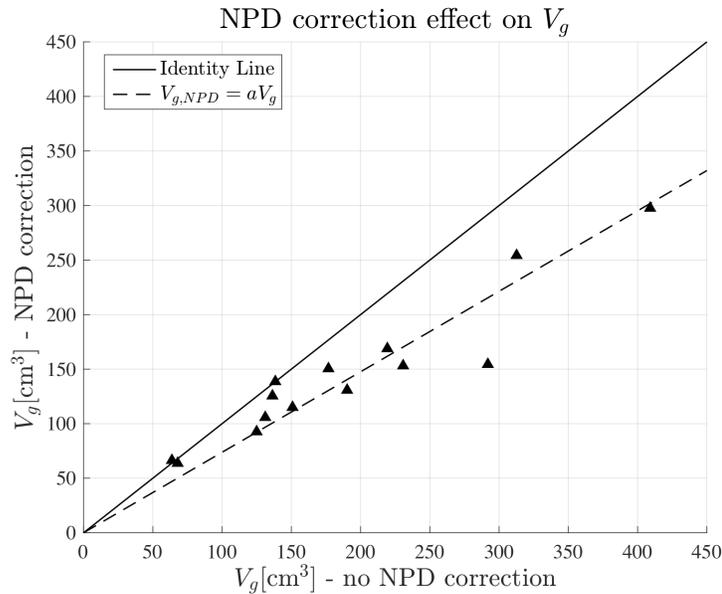

FIGURE B.0.5. NPD-corrected V_g vs non-corrected V_g. Dotted line shows linear fit.
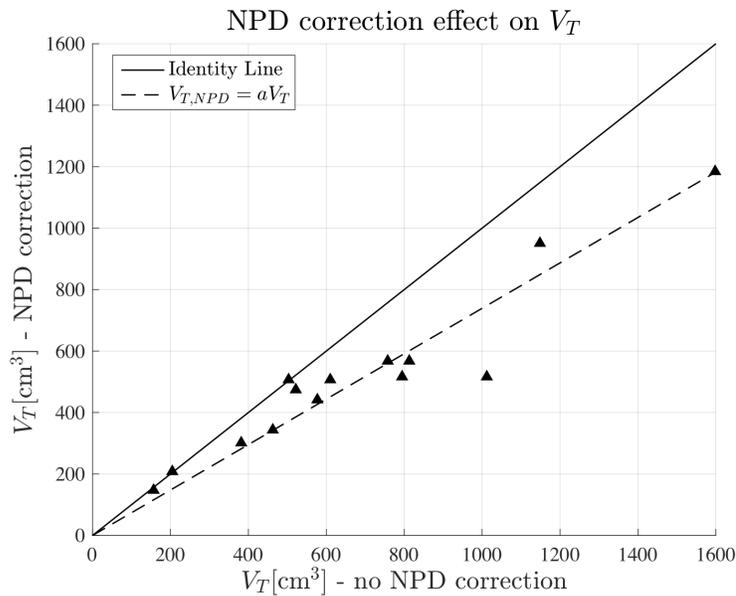
FIGURE B.0.6. NPD-corrected V_T vs non-corrected V_T. Dotted line shows linear fit.
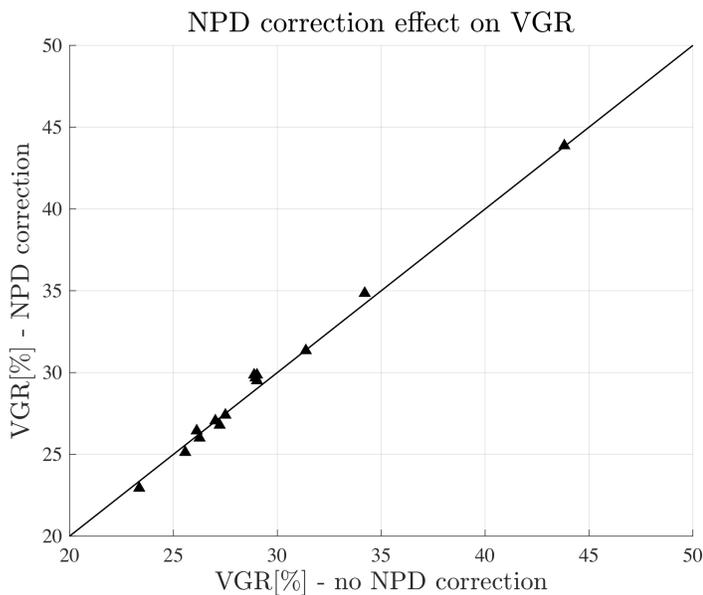

FIGURE B.0.7. NPD-corrected VGR vs non-corrected VGR. Solid black line represents the identity line.

This experiment led to the conclusion that, while glandular and total volume measurements were highly sensitive to breast region definition, VGR was fairly robust. As such, the volume ratio should be preferred when comparing different image modalities, where an exact matching of breast regions is impossible to achieve.