Simulation and assessment of realistic breast lesions using fractal growth models

A Rashidnasab, P Elangovan, M Yip, O Diaz, D R Dance, K C Young, and K Wells

1 Centre for Vision, Speech and Signal Processing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK.
2 National Coordinating Centre for the Physics of Mammography, Royal Surrey County Hospital, Guildford, Surrey, GU2 7XX, UK
3 Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK

E-mail: a.rashidnasab@surrey.ac.uk

Abstract. A new method of generating realistic three dimensional simulated breast lesions known as diffusion limited aggregation (DLA) is presented, and compared with the random walk (RW) method. Both methods of lesion simulation utilize a physics-based method for inserting these simulated lesions into 2D clinical mammogram images that takes into account the polychromatic X-ray spectrum, local glandularity and scatter. DLA and RW masses were assessed for realism via a receiver operating characteristic (ROC) study with 9 observers. The study comprised 150 images of which 50 were real pathology proven mammograms, 50 were normal mammograms with RW inserted masses and 50 were normal mammograms with DLA inserted masses. The average area under the ROC curve for the DLA method was 0.55 (95% confidence interval 0.51-0.59) compared to 0.60 (95% confidence interval 0.56-0.63) for the RW method. The observer study results suggest that the DLA method produced more realistic masses with more variability in shape compared to the RW method. DLA generated lesions can overcome the lack of complexity in structure and shape in many current methods of mass simulation.

Accepted at: Physics in Medicine and Biology
doi:10.1088/0031-9155/58/16/5613
Online at stacks.iop.org/PMB/58/5613
1. Introduction

National breast screening programmes using X-ray mammography have produced a significant decline in breast cancer mortality among women (Karim-Kos et al. 2008). In recent years the traditional film-screen image receptor has been replaced by digital receptors (Young & Dance 2010) and digital breast tomosynthesis is now available. However, the impact of digital systems on the clinical detection task still needs to be established. Clinical trials are conventionally used to study these systems but have many associated limitations such as the ethical issues of extra exposure and the excessive time scales needed for generating statistically meaningful datasets. However, simulation approaches that utilize realistic models of pathology which can be inserted into normal mammograms offer the potential to overcome these limitations. Thus, the development of lesion models with the appearance of realistic pathology can enable the assessment and optimization of 2D digital imaging and emerging methods of 3D breast imaging. Two approaches have been used in the past to simulate the appearance of masses. In the first approach real masses have been characterized from clinical data and then masses synthesised by randomly sampling from the mass appearance parameter space (Saunders et al. 2006, Highnam et al. 2000, Timberg et al. 2010, Caulkin et al. 2000, Berks et al. 2010a). Some of these have been evaluated by receiver operating characteristics (ROC) studies, with area under the ROC curve (AUC) and 95% confidence interval (CI) ranging from 0.65 (0.58-0.72) (Saunders et al. 2006) to 0.70 (0.61-0.79) (Berks et al. 2010b).

The second main approach for lesion simulation uses mathematical models to generate masses which may then be inserted into clinical images. Bliznakova et al. (2003) used simple shapes (e.g. spheres or ellipsoids) to simulate masses. A random walk method of simulating malignant masses has been used by several authors (Bliznakova et al. 2003, Ruschin et al. 2005, Gong et al. 2006, Hintsala et al. 2009). Ruschin et al. (2005) have also attempted to simulate spiculations by adding star-shaped objects to their model. Most of the mass models used for this second approach employ adhoc empirical image processing steps to enhance the realism of the mass appearance when inserted into the mammogram. More recent work by de Sisternes et al. (2010), used a stochastic Gaussian random sphere model to simulate a central tumour and a fractal branching algorithm to model spicules. However, none of these mathematical approaches appears to have been evaluated with observer studies but instead, have relied on subjective assessments. There is therefore a need for a validated model of breast masses appearance which is capable of producing a wide variety of representative pathology.

This paper presents and validates a new method of generating 3D masses using diffusion limited aggregation (DLA). For comparison purposes, masses have also been generated using a random walk (RW) method, based on a previously published method (Bliznakova et al. 2003). The generated masses are inserted into normal 2D digital mammograms using a physics-based approach thus avoiding the need for adhoc
morphological operations to produce acceptable appearance. The mass generation methods and insertion procedure are described in Section 2. Section 3 describes the validation step performed on the realism of the simulated masses using an ROC-based approach, with results and discussion reported in Section 4.

2. Methodology

Figure 1 presents the overall simulation framework. Masses were generated in 3D using either RW (Section 2.1), or DLA (Section 2.2). Section 2.3 describes the method used to insert the simulated mass into a 2D mammogram including calculating the primary transmission using local glandularity, image degradation caused by the imaging system’s acquisition process, accounting for scatter and image processing.

![Image simulation framework](image)

**Figure 1.** Image simulation framework

2.1. Generation of masses in 3D: Random walk

The random walk approach used here is based on the work of Bliznakova et al. (2003) and utilizes a recursive random walk in a simple stochastic growth model to produce malignant-like masses with irregular boundaries in 3D.

2.1.1. Method

The algorithm initiates by first incrementing the central voxel of a 3D binary matrix which represents the mass growth centre. New random walks are recursively initiated from this central voxel. At each step of the random walk, an adjacent voxel neighbour is randomly selected and its value incremented to one, i.e. assigned as part of the mass cluster. This is repeated until the random walk reaches its end point. However, each voxel value can only be incremented once in case of multiple path crossing. The two key parameters used in the RW approach are \( n \), the number of steps in each RW, and \( R \), the number of RWs used to build the mass. Each walk terminates if it reaches a predefined number of steps or collides with a predefined bounding box. Masses were simulated with the diameter ranging from 5 to 20 mm
in order to produce both subtle and obvious cases to provide a range of clinically challenging tasks. A voxel size of 17.5\(\mu\)m was chosen which was 1/4 of the imaging system detector pixel size (70\(\mu\)m for the Hologic Selenia used in this study). This facilitated the simulation of fine detail before insertion that would then be filtered and sub-sampled to the detector pixel size through the physics of the insertion process.

2.1.2. Parameter appearance dependency

Figure 2 shows how the appearance of a RW simulated mass varies when changing the parameters in the growth algorithm. The \(R\) and \(n\) parameters control the texture/density and the size of the simulated mass respectively. Figure 2(a) shows how the size of the simulated mass increases when \(n\) is increased. Figure 2(b) shows how the inner structure of the mass fills when increasing \(R\) for a constant \(n\), while the size or overall diameter remains roughly unchanged. The dotted circle defines the outer boundary of the mass. Over 200 simulations were used to study the production of RW masses with different sizes and density and bounding box. The resulting parameters were then used to produce volumes of size and density appropriate for a realistic breast mass, as summarized in Table 1.

![Figure 2](image)

**Figure 2.** Parameter appearance dependency in RW masses. Projection of the 3D RW masses for (a) constant \(R\) and increasing \(n\), (b) constant \(n\) and increasing \(R\) from left to right, the dotted circle defines the outer boundary of the mass.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(n)</th>
<th>(R)</th>
<th>bounding box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>6,000-40,000</td>
<td>6,000-100,000</td>
<td>cube, ellipsoid, super-ellipsoid</td>
</tr>
</tbody>
</table>

2.2. Generation of masses in 3D: Diffusion limited aggregation

Diffusion limited aggregation (DLA) is a type of fractal growth, proposed here to simulate the appearance of uneven lesion textures of real breast cancers. DLA was originally introduced in modelling crystal formation (Witten & Sander 1981) but this is the first occasion it has been investigated for simulating malignant masses. DLA involves the diffusion of particles modelled as a Brownian motion/random walk which then aggregate under a set of predefined rules to form a cluster.
Simulation and assessment of realistic breast lesions using fractal growth models

2.2.1. Method

DLA growth begins with assigning a central voxel of a 3D binary matrix as a mass or “sticky” cluster or growth centre, as illustrated in Figure 3, shown in 2D for ease of visualization. Launch sites are defined as a series of concentric spheres formed around the central voxel or cluster centre. The radial distance between launch sites and the particle density in each of these is user-defined. Randomly moving particles are then launched, one at a time, from the launch sites (Figure 3(a)). Each of these mobile particles moves using a random walk process with a user-defined step size. The trajectory of the random walk may eventually be located adjacent to the mass or cluster centre (Figure 3(b)), in which case the random walk particle may become aggregated to the cluster (Figure 3(c)) with a certain probability known as the sticking probability. If the sticking probability criterion is not satisfied the particle is discarded. Thus, the cluster forms by addition of random particles reaching adjacent voxels lining the growth centre, i.e. entering a single-voxel wide “sticky-zone” surrounding the growth centre. If a moving particle fails to reach the sticky cluster or mass growth centre (as is often the case) and travels further than a pre-defined distance, it is then discarded (Figure 3(d,e)). Branch-like shapes can be created by using an appropriate choice of parameters, and are formed via the higher probability of the random walk particle hitting the outer surface of the sticky cluster before reaching the inner part, thereby creating a porous volumetric appearance.

Figure 4 shows some examples of 3D DLA masses and corresponding 2D projections. Table 2 lists the DLA growth parameters, which facilitate greater flexibility in controlling the microscopic appearance of the simulated masses compared to the RW method. A voxel size of 35μm was used in growing the DLA masses, which was 1/2 of the imaging system detector pixel size used in this study (70μm for the Hologic Selenia). This increase in voxel size compared to that used for RW masses (17.5μm) was because of the excessive computational overhead associated with the DLA process, whilst still preserving a relatively high level of detail prior to insertion.

2.2.2. Parameter appearance dependency

To illustrate the wide variation in the appearance of the simulated masses that can be produced using DLA, several growth prescriptions in 2D were performed changing the parameters shown in Table 2. Figure 5 shows a selection of 2D DLA simulated masses for different launching prescriptions. Figure 5(a) demonstrates that inappropriate combinations of the distribution of points...
Simulation and assessment of realistic breast lesions using fractal growth models

Figure 4. Examples of simulated DLA masses (a-c) in 3D, (d-f) and the corresponding projections in 2D.

Table 2. DLA parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Impact on the mass appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>Number of launching particles</td>
<td>Size</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Density of particles on launching sphere</td>
<td>Texture</td>
</tr>
<tr>
<td>$\Delta C$</td>
<td>Step size in launching sphere diameter</td>
<td>Size and texture</td>
</tr>
<tr>
<td>$\Delta S$</td>
<td>Step size of the random walk</td>
<td>Density</td>
</tr>
<tr>
<td>$P_s$</td>
<td>Sticking probability</td>
<td>Density</td>
</tr>
</tbody>
</table>

and launching circles can result in unrealistic clusters. The size of the cluster increases as the number of launching points, $N$, increases (Figure 5(b)). For a constant $\rho$, an increase in the $\Delta C$ parameter results in a smaller cluster as it is less probable to hit a small sticky cluster from a more distant launching point (Figure 5(c)). Figure 5(d) shows that by increasing $\Delta S$, a more concentrated mass can be produced where the dendritic arms are in closer proximity. This was used to produce masses with a denser core by starting with a high $\Delta S$ and then decreasing $\Delta S$ to produce branch-shaped structures. A study of these parameters in 2D was initially used as a guide in selection of parameters in subsequent 3D mass generation, as this allowed easier visualisation of the internal structure.

Over 200 DLA simulations were also performed to explore the parameter space for simulating structures resembling breast masses. Table 3 summarizes the range of parameters which produced the most realistic DLA masses. It should be noted however...
that DLA parameters are not independent of each other.

![Figure 5](image)

**Figure 5.** Parameter appearance dependency of DLA masses. From left to right: (a) use of exponential, negative exponential, uniform and linear distribution for launching circles and number of points per circle, (b) increasing $N$ increases the size of the cluster, (c) larger spaced launching circles result in smaller cluster formation for a constant $\rho$, (d) increase in $\Delta S$ results in a smaller denser structure.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N$</th>
<th>$\rho$</th>
<th>$\Delta C$</th>
<th>$\Delta S$</th>
<th>$P_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>$10^6 - 5 \times 10^6$</td>
<td>constant</td>
<td>1-5</td>
<td>16-64</td>
<td>0.001-1</td>
</tr>
</tbody>
</table>

**Table 3.** Range of optimal DLA parameters

### 2.3. Lesion insertion method

A set of anonymised normal (i.e. pathology-free) mammograms was acquired from UK breast screening centres using the Hologic Selenia mammography system. The insertion site on the normal mammogram was selected manually. This was undertaken based on local morphology and visual appearance and aided through reference to a geographical distribution of real screen-detected breast cancers (Brown et al. 2001), which suggests more probable areas in the breast where malignancy may be found. Knowledge of lesion locations derived from the literature (Tabár et al. 2012) was also employed. The steps then used to insert masses into these images are described below as depicted in Figure 1.

#### 2.3.1. Calculation of primary transmission

The insertion of a mass into a mammogram is assumed to modify the primary photon flux detected in the image receptor. But given the relatively small size of the mass, it has been assumed that its insertion has negligible effect on the overall scatter distribution. We first address the relative change in primary signal in the region of interest (ROI) where the mass is to be inserted. The primary photon flux at each pixel before insertion ($p_0$) was calculated using:

$$ p_0 = \sum_E E \epsilon(E) N(E) e^{-\sum_i \mu_i(E) l_i - \mu_{br}(E) l_{br}} \Delta E $$

(1)
where $E$ is the incident photon energy, $\varepsilon(E)$ is the detector energy absorption efficiency, $N(E)$ is the relative number of X-ray photons at energy $E$ calculated from the spectral model of Boone et al. (1997) with a bin size of $\Delta E$. The spectra were matched with the associated mammogram acquisition settings in terms of target/filter combination, tube voltage, filter thickness and half-value layer derived from the associated Dicom header. The primary beam traverses several different materials (compression paddle, breast support plate, etc.) with attenuation coefficients $\mu_i$ and thickness $t_i$. $t_{br}$ is the thickness of breast traversed by the beam $\mu_{br}$ represents the attenuation coefficient of the breast tissue, assumed to be a mixture of adipose and glandular tissues. The composition of these tissues was taken from the work of Hammerstein et al. (1979) and the linear attenuation coefficients were calculated using data from NIST (Berger et al. 2010).

The attenuation of the primary photon beam after insertion ($p_1$) was calculated assuming the equivalent volume of breast tissue is replaced by the mass volume:

$$p_1 = \sum E \varepsilon(E) N(E) e^{-\sum \mu_i(E)t_i - \mu_{br}(E)t_{br} - \mu_{mass}(E)t_{mass} + \mu_{br}(E)t_{mass}} \Delta E \quad (2)$$

where $\mu_{mass}$ represents the attenuation coefficient of the breast mass assumed to have the same attenuation coefficient as glandular tissue and $t_{mass}$ is the thickness of the mass.

Parallel beam geometry was assumed in the calculation of different tissue thickness, and x-ray path length was estimated for pixels corresponding to mixtures of glandular, adipose and mass tissues according to equation 1 and 2. For the calculation of $\mu_{br}$, an estimate of the local glandularity of the breast over each image pixel was used. For this purpose, volumetric breast composition estimation was performed on the mammogram using the Volpara software tool (Highnam et al. 2010) which uses physics modelling and information derived from the image. The use of Volpara in our insertion process enabled more accurate tissue replacement by the inserted mass compared to a global estimate of the density of the breast.

2.3.2. Image degradation chain

The process of image degradation caused by the imaging system’s acquisition process (intrinsic blurring and sampling processes) needs to be modelled when inserting a simulated mass template into a clinical mammogram, thus avoiding the need for empirical image processing to achieve realistic contrast (Ruschin et al. 2005, Hintsala et al. 2009). An approach proposed by Yip et al. (2011) was employed in this work which uses a Gaussian model of the modulation transfer function (MTF) of the system.

Thus, the primary templates $p_0$ and $p_1$ (pre- and post- insertion) were filtered with a model MTF of the associated imaging system of the normal mammogram used in insertion:

$$p_i^{\text{filtered}} = \text{FFT}^{-1}\{\text{FFT}(p_i) \cdot \text{MTF}_{\text{system}}\} \quad (3)$$

The resultant images were then downsampled to the associated system detector pixel size using weighted neighbourhood averaging to produce the arrays $p_0^{\text{modified}}$ and $p_1^{\text{modified}}$. The ratio of these two quantities, named as the relative transmission factor (RTF), was then calculated.
2.3.3. Mass insertion and scatter inclusion  Having calculated the primary arrays from
the above, the scatter was taken into consideration in the insertion process. The pixel
intensity values in the raw image, acquired on the system prior to processing can be
calculated using equation 4:
\[ I_i = c \cdot (p_i + s_i) \]  (4)
where \( c \) is a constant relating the primary (\( p \)) and scattered (\( s \)) photon energy deposition
to the pixel value and \( i = 0 \) or \( 1 \) corresponding to the absence or presence of the mass,
respectively. As noted earlier, it was assumed that the scatter remains unchanged after
insertion of the mass, i.e. \( s_0 = s_1 \). Therefore, using equation 4, \( I_1 \) the intensity of the
raw image after mass insertion can be written as:
\[
I_1 = c P_1^{\text{modified}} \cdot \left(1 + \frac{s_1}{P_1^{\text{modified}}} \right)
\]  (5)
\[
= c P_0^{\text{modified}} \cdot \frac{1}{\text{RTF}} \cdot \left(1 + \frac{s_1}{P_0^{\text{modified}}} \cdot \text{RTF} \right)
\]  (6)
\[
= \frac{I_0}{1 + \text{SPR}} \cdot \frac{1}{\text{RTF}} \cdot \left(1 + \text{SPR} \cdot \text{RTF} \right)
\]  (7)
where \( I_0 \) the pixel value of the raw mammogram before insertion is known and RTF,
the relative transmission factor, is the ratio of \( P_0^{\text{modified}} \) to \( P_1^{\text{modified}} \), calculated using
the method described above. The scatter primary ratio (SPR) was calculated using the
validated Monte Carlo model of Diaz et al. (2010) which is based on the Geant4 toolkit
(Agostinelli et al. 2003). A series of uniform phantoms with thicknesses in the range of
3-7 cm and various glandularities (0%, 25%, 50%, 75%) were used. \( 10^9 \) photons emerging
from a point source at 66 cm from the detector were used for each simulation, irradiating
the entire imaging area of the receptor. The energy absorbed in a specific region of the
receptor from primary and scattered photons was scored and the SPR calculated. The
resulting values were tabulated so that the SPR for any breast thickness and glandularity
could be found by interpolation. The statistical error for each value calculated was less
than 0.23% and the compression paddle, breast support and the particular anti-scatter
grid of the Hologic Selenia system were included in the simulation. The cellular anti-
scatter grid was designed following the manufacturer’s specification. Anti-scatter septa
were aligned with the X-ray source. Further details can be found in Diaz et al. (2010).

2.4. Image processing
Digital mammography systems routinely employ post acquisition image processing prior
to the display of the image. As the final step in the mass insertion process, the raw
mammogram with the inserted mass was processed using the Hologic LORAD FFDM
Selenia V5.0 image processing package.

3. Validation
The simulation procedure was validated using human observer studies and ROC analysis.
A pre-validation pilot study was performed initially where a small limited set of
images containing inserted simulated masses were presented to a group of three expert radiologists to evaluate the realism of the masses in terms of their shape, margin and density. In this pilot study and thereafter, all the masses were inserted into the images by the first author. Radiologist feedback was used to better understand realistic compared to unrealistic insertion sites. The pre-validation pilot study was used to refine the parameters selected for subsequently generating >200 RW and >200 DLA candidate masses for use in the observer study as summarised in Table 1 and Table 3 for RW masses and DLA masses respectively.

In order to produce a similar range of sizes and shape variability as seen in the 50 real cases used in the subsequent observer study, 60 RW and 60 DLA masses were selected from the RW and DLA groups. This was based on their pre-insertion appearance and knowledge gained in the pilot study. Note, many other masses with realistic pre-insertion appearance remained in the discarded group, and were not analysed further. Each group of 60 masses (from each method) were then inserted in the individual mammograms. The images were shown to one of the expert radiologists prior to the start of the observer study wherein a small number (~10) of these cases were removed as they were deemed invisible, or not realised as a mass. Fifty such inserted masses with good radiologist feedback (i.e. rated as realistic and realised as an apparent lesion by a radiologist) were then used in the observer study. The above experts were excluded from the subsequent validation study.

3.1. Observer study

An observer study was performed to validate the realism of each of the proposed mass simulation methods against real cancers. 150 single view images, cranial-caudal or mediolateral-oblique view, were used in the study, of which 50 images contained real biopsy-proven cancers, 50 images contained simulated RW masses and 50 images contained simulated DLA masses. All the images used in the study were anonymised routine breast screening images collected from local breast screening centres. These images were acquired using Hologic Selenia systems with a detector pixel size of 70 μm and a target/filter combination of molybdenum/molybdenum or molybdenum/rhodium. The 100 images used for insertion of simulated masses were selected from the images classified as normal in breast screening. In each normal image, one simulated mass (either RW or DLA) was inserted using the methodology described above.

The regions of breast tissue into which the simulated masses were inserted presented a broad (>10 fold) range of local glandularity and background tissue variability. The locations of the inserted masses were distributed in the breast avoiding insertion close to the skin and nipple. In ~20% of the insertions for each method (DLA and RW), the mass was inserted in the close vicinity of non-malignant calcifications, in order to replicate this effect in the real cancer images used in the observer study, which also had calcifications present in close vicinity to the real mass (~30% real cases). Simulated masses with a size range of 5 to 15 mm were generated for the purpose of this study.
The total image set was displayed, in random order, one image at a time on a high resolution monitor and under the lighting conditions used for reading routine screening mammograms. Nine radiologists with 3-18 years of experience in reading mammograms, ranked the realism of each mass on a 5-point scale such that 1: definitely simulated, 2: probably simulated, 3: indeterminate, 4: probably real and 5: definitely real. It is worth noting that DLA and RW produce essentially mass appearance models, rather than biological growth models. As the intention was to generate masses of malignant appearance, radiologists were also asked for comment on the apparent level of suspicion/malignancy. The level of suspicion was ranked using a five-point scale as used in clinical assessments, i.e. 1: normal, 2: benign, 3: indeterminate, 4: suspicious and 5: malignant. The observers were given no information on the number of cases in each category, but the location of the mass in the mammogram was provided. In-house software developed as an ImageJ Java plug-in (Abramo et al. 2004), was designed for the validation study which allowed observers to zoom and search freely within the image without time constraints. A teaching dataset was used in the beginning of each study to familiarise the observer with the validation software. Receiver operating characteristic (ROC) analysis (Barrett & Myers 2004) was performed on the observers’ ranking of realism of the masses for (1) RW masses and real cancers and (2) DLA masses and real cancers, following the method previously used by other simulated lesion studies (Saunders et al. 2006, Berks et al. 2010b, Skiadopoulos et al. 2003, Shaheen et al. 2011).

4. Results and Discussion

Figures 6 and 7 show exemplar images of simulated mass insertions. The variable appearances of the masses were achieved using different simulation prescriptions. The results demonstrate a range of different lesion morphologies including embedded masses presenting with well-defined, ill-defined and mixed margins, which appear to blend with the background. The DLA model can produce masses with more irregular shapes and less symmetry compared to RW masses. Both models produce masses with a porous volumes, which is characteristic of real masses; a real mass is an inhomogeneous volume of cancer cells and malignant structure intermingled with other types of breast or necrosed tissue (Tot et al. 2002).

Figure 8 shows some examples of inserted simulated masses that were ranked as unrealistic by the observers. Figures 8(a) and 8(b) are examples of unrealistic shape and margin and show that an inappropriate choice of parameters in the mathematical models can produce unrealistic structures. Figure 8(c) shows a mass with a density that is too low so that the background structures are too easily seen through the mass. Figure 8(d) shows an example of lack of interaction of the mass with the background. The current work does not address interaction of the mass with the background in terms of tissue displacement or simulation of masses with spiculations, which requires further research. This also highlights the importance of insertion location and the background structure and its affect on the realism of the inserted lesions. A manual insertion approach was
adopted because it was not appropriate to develop a fully automatic insertion method until the overall method was fully developed and validated. Our objective was to develop a series of masses that look realistic over a range of insertion sites. To address the subjectivity implied by manual insertion, we utilised a wide range of locations through the breast ensuring variability in the local glandularity and background structure for insertion sites of both RW and DLA masses.

Figure 9 presents ROC curves plotted for each of the nine observers for RW masses (Figure 9(a)) and DLA masses (Figure 9(b)). The 50:50 chance line representing random guessing is also shown. The closer the ROC curve lies to the chance line, the more challenging was the task for the observer to correctly differentiate real and simulated masses. The data from the study were analysed further by computing the area under the ROC curve (AUC) using the trapezoidal rule along with the 95% confidence interval (CI) for each observer (Hanley & McNeil 1982). Table 4 shows the AUC and 95% CI values for each observer. For eight out of nine observers the 95% CIs of DLA masses include the 0.5 value, whereas for RW masses this was achieved for only five observers. For all the observers, except observer 4, the comparison of the individual observer results shows reduced AUC for DLA masses compared to RW masses. In addition, the average
AUCs over all readers and the corresponding 95% CIs were calculated using the standard error for nine AUCs. The average AUC and corresponding CI for RW were 0.60 (0.56, 0.63), whereas for DLA masses the corresponding values were 0.55 (0.51, 0.59). This suggests that DLA masses were more challenging to distinguish as real or simulated for observers than RW masses, and ranked as real more often than RW masses. This may be because the RW approach can only simulate a small variation in lesion appearance which is over-represented in the dataset. However, DLA lesions can be more variable in appearance and can represent a wider range of pathology.
These results may be compared with similar validations of synthetic mass pathology performed by other authors: Saunders et al. (2006) reported an AUC with 95% CIs of 0.68 (0.61,0.75) for benign masses and 0.65 (0.58,0.72) for malignant masses. In their study 3 expert mammographic readers scored a total of 200 images containing approximately equal numbers of real and simulated benign masses, malignant masses and microcalcifications. Berks et al. (2010b) reported an AUC with 95% CIs of 0.70(0.61,0.79) for simulation of malignant masses. In their study, 10 experts scored a set of 60 images, with 30 real masses and 30 simulated masses. Note that both of the above studies used digitized film-screen mammograms for insertion while in this study digital 2D mammograms were used. While Saunders et al. (2006) and Berks et al. (2010) presented observers with a limited size ROI on the mammogram including the mass, in our study the full mammogram was presented to the observers with the ROI highlighted. Such differences in the design and condition of the study and display may affect direct comparison of the statistics above. In our study, a total of 150 images were assessed by 9 expert mammographic readers. This produced an average AUC for DLA masses in the present study which lies below the confidence interval provided in previous work and closer to 0.5, implying a more realistic mass appearance.

Given that some ROC results are closer to the line of unity than others, we have also analysed individual responses with years of screening experience (Table 4). This showed that observers with >6 years experience produced an AUC statistic of 0.57-0.65 for RW masses, with somewhat more convincing performance for DLA masses (0.52-0.64).

Figure 9. ROC curves for individual observers for (a) RW and real masses,(b) DLA and real masses

Figure 10 compares the overall ranking for total real, RW and DLA cases for all the observers. This highlights the (mis-)classification of real masses. One might therefore expect that a “convincing” simulated mass would produce similar mis-classification as an authentic real mass. It can be seen that although both sets of simulated masses offer similar behaviour, particularly in the equivocal case, it is the DLA masses which most
Table 4. AUC and 95 CI for each observer and average of all observers for RW and DLA masses.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Yrs of exp.</th>
<th>RW</th>
<th>DLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>0.65 (0.54,0.76)</td>
<td>0.59 (0.48,0.71)</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>0.58 (0.47,0.69)</td>
<td>0.52 (0.40,0.63)</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0.65 (0.54,0.76)</td>
<td>0.58 (0.47,0.69)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.54 (0.43,0.66)</td>
<td>0.56 (0.45,0.68)</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>0.64 (0.53,0.75)</td>
<td>0.64 (0.53,0.75)</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.54 (0.43,0.65)</td>
<td>0.47 (0.35,0.58)</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>0.57 (0.46,0.68)</td>
<td>0.52 (0.40,0.63)</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0.62 (0.51,0.73)</td>
<td>0.54 (0.43,0.65)</td>
</tr>
<tr>
<td>9</td>
<td>5.5</td>
<td>0.58 (0.47,0.69)</td>
<td>0.51 (0.39,0.62)</td>
</tr>
<tr>
<td>Mean</td>
<td>8.16</td>
<td>0.60 (0.56,0.63)</td>
<td>0.55 (0.51,0.59)</td>
</tr>
</tbody>
</table>

Figure 10. Global histogram of all observers for the real and simulated masses realism ranking.

Table 5. Summary of pooling all observers’ results using majority voting for each mass. This shows the proportion of masses ranked as “real”, the proportion of masses ranked as “recall”, and the proportion of the masses ranked as both “real” and “recall” by the majority voting.

<table>
<thead>
<tr>
<th>Mass type</th>
<th>Real</th>
<th>RW</th>
<th>DLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of cases</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>proportion ranked as real</td>
<td>94%</td>
<td>68%</td>
<td>84%</td>
</tr>
<tr>
<td>proportion ranked as recall</td>
<td>86%</td>
<td>84%</td>
<td>78%</td>
</tr>
<tr>
<td>proportion ranked as real and recall</td>
<td>80%</td>
<td>54%</td>
<td>62%</td>
</tr>
</tbody>
</table>
closely match the classification pattern of the real cases.

In Table 5, the ranking of particular masses produced by pooling all observer results and using majority voting for each mass is summarized. Here, a mass was classified as “real” if the majority of the observers voted 3 or more on mass realism ranking (based on the 5-point realism scale as described in section 3.1.). As the table shows, DLA gave better results for producing masses of real appearance (84%) compared to RW (68%).

In the UK, a patient is recalled if the clinical assessment of a screen-detected lesion is ranked as indeterminate, suspicious or malignant. Therefore, in this study a mass was classified as a “recall” case if the majority of the observers voted 3 or more on its level of suspicion (as described in section 3.1.). This table also shows the proportion of masses ranked as both “real” and “recall”. The data for each category of genuine real, simulated RW and simulated DLA masses were analysed separately. The results show that although the RW masses were more likely to be ranked as recall, they were also more likely to be ranked as being simulated. Therefore, the DLA method is more likely to produce masses that look both real and ranked as a recall. The ranking results for real cases used in the study are also presented in this table for comparison. These results, taken together with the ROC analysis, may have been improved upon, if in retrospect, better matching of the real vs simulated pathologies had occurred.

5. Conclusion

This work presents a computational model of breast mass appearance using fractal growth which can exhibit a range of lesion appearances, and thus overcomes some of the drawbacks in previous work. Masses generated using RW and DLA models were inserted into raw digital 2D mammograms using a physical model of the imaging process and compared. Local glandularity was used for the first time to account for local breast composition, combined with polychromatic X-ray spectra to produce realistic lesion attenuation. In addition, the raw digital images with inserted simulated masses were processed with the manufacturer’s image processing software; this post-processing step is essential to preserve overall clinical appearance.

The physics-based insertion process proposed here for both approaches avoids the need for ad hoc image enhancements. An ROC study of realism gave an average AUC and corresponding 95% CIs of 0.55 (0.51,0.59) for DLA masses. This suggests that the DLA approach appears to produce a more realistic range of mass appearances compared to the RW approach, which achieved an AUC of 0.60 (0.56,0.63). Both results demonstrate improvement compared to previously published ROC studies of realism of the simulated masses. The mass simulation models may be used subsequently as part of a tool to evaluate different breast imaging technologies and their performance in the detection task. These mass models combined with realistic simulated microcalcifications (e.g. Shaheen et al. (2011)) and inserted in 2D digital mammograms can be used for such studies. Since these models are in three dimensions, they could also be used to study and compare 2D and 3D breast imaging. These topics are the subject of on-going
work.

Acknowledgements

This work is part of the OPTIMAM project and is supported by the Cancer Research UK & Engineering and Physical Sciences Research Council Cancer Imaging Programme in Surrey, in association with the Medical Research Council and Department of Health (England). We thank Ralph Highnam for granting access to the Volpara breast volume measurement tool and providing technical assistance and Hologic for granting access to the image processing software. We are grateful to colleagues at NCCPM for their help in image collection and our clinical observers at the Jarvis Breast Screening Centre, Guildford, St. George’s Hospital, London and Addenbrooke’s Hospital, Cambridge.

References

Barrett H H & Myers K J 2004 Foundation of image science John Wiley and Sons Inc.
Simulation and assessment of realistic breast lesions using fractal growth models


Tot T, Tabár L & Dean P 2002 *Practical breast pathology* Thieme.

