Check for updates

Received: 23 May 2022 | Revised: 22 September 2022 | Accepted: 10 November 2022

DOI: 10.1002/mp.16133

RESEARCH ARTICLE

MEDICAL PHYSICS

Generative compressed breast shape model for digital mammography and digital breast tomosynthesis

Marta C. Pinto[1] | **Koen Michielsen**[1] | **Ramyar Biniazan**[2] | **Steffen Kappler**[2] | **Ioannis Sechopoulos**[1,3,4]

[1]Dept. of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands

[2]Siemens Healthcare GmbH, Forchheim, Germany

[3]Dutch Expert Centre for Screening (LRCB), Nijmegen, The Netherlands

[4]Technical Medicine Centre, University of Twente, Enschede, The Netherlands

Correspondence
Ioannis Sechopoulos, Radboud University Medical Centrum, Postbus 9101, 6500 HB Nijmegen (766), Geert Grooteplein 10 (route 767), The Netherlands.
Email: ioannis.sechopoulos@radboudumc.nl

Abstract

Background: Modelling of the 3D breast shape under compression is of interest when optimizing image processing and reconstruction algorithms for mammography and digital breast tomosynthesis (DBT). Since these imaging techniques require the mechanical compression of the breast to obtain appropriate image quality, many such algorithms make use of breast-like phantoms. However, if phantoms do not have a realistic breast shape, this can impact the validity of such algorithms.

Purpose: To develop a point distribution model of the breast shape obtained through principal component analysis (PCA) of structured light (SL) scans from patient compressed breasts.

Methods: SL scans were acquired at our institution during routine craniocaudal-view DBT imaging of 236 patients, creating a dataset containing DBT and SL scans with matching information. Thereafter, the SL scans were cleaned, merged, simplified, and set to a regular grid across all cases. A comparison between the initial SL scans after cleaning and the gridded SL scans was performed to determine the absolute difference between them. The scans with points in a regular grid were then used for PCA. Additionally, the correspondence between SL scans and DBT scans was assessed by comparing features such as the chest-to-nipple distance (CND), the projected breast area (PBA) and the length along the chest-wall (LCW). These features were compared using a paired t-test or the Wilcoxon signed rank sum test. Thereafter, the PCA shape prediction and SL scans were evaluated by calculating the mean absolute error to determine whether the model had adequately captured the information in the dataset. The coefficients obtained from the PCA could then parameterize a given breast shape as an offset from the sample means. We also explored correlations of the PCA breast shape model parameters with certain patient characteristics: age, glandular volume, glandular density by mass, total breast volume, compressed breast thickness, compression force, nipple location, and centre of the chest-wall.

Results: The median value across cases for the 90^{th} and 99^{th} percentiles of the interpolation error between the initial SL scans after cleaning and the gridded SL scans was 0.50 and 1.16 mm, respectively. The comparison between SL and DBT scans resulted in small, but statistically significant, mean differences of 1.6 mm, 1.6 mm, and 2.2 cm^2 for the LCW, CND, and PBA, respectively. The final model achieved a median mean absolute error of 0.68 mm compared to the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

scanned breast shapes and a perfect correlation between the first PCA coefficient and the patient breast compressed thickness, making it possible to use it to generate new model-based breast shapes with a specific breast thickness.
Conclusion: There is a good agreement between the breast shape coverage obtained with SL scans used to construct our model and the DBT projection images, and we could therefore create a generative model based on this data that is available for download on Github.

KEYWORDS
breast tomosynthesis, compressed breast shape, generative model, mammography, principal component analysis, statistical shape model, structured light scanning

1 | INTRODUCTION

Digital mammography (DM) and digital breast tomosynthesis (DBT) require mechanical compression of the breast during image acquisition. Therefore, many state-of-the-art image processing algorithms involve simulation of the breast undergoing compression. Some examples include scatter correction[1] applications, in which multiple realizations of realistic breast shapes are used to develop an x-ray scatter deep learning-based model that is then used to correct DBT projections, or image reconstruction, in which the prior knowledge of the breast shape can reduce reconstructions artifacts near the breast edge[2] and help to further optimize the design of DBT systems for clinical practice.[3]

Modelling of the human anatomy is the subject of intense research, involving anywhere from the whole body,[4,5] only the face,[6,7] or single organs, such as the heart.[8] One approach to create such models is by examining a representative sample of shapes by statistical means, also known as statistical shape modelling. This can be performed by first obtaining the mean shape representation from the dataset, together with possible variations from the mean representation. The analysis of these statistical shape models can also help to study and identify certain shape features in a patient population, which are capable of characterizing possible pathologies and of providing estimations of the body composition.[5,6] When an object shape is represented as a set of points distributed across a surface, it can be defined as a Point Distribution Model (PDM). This is a type of statistical shape model and therefore can be statistically analysed through principal component analysis (PCA). For instance, the work by Bennett et al.[5] showed how statistical shape models can be used to model the human body using PCA and to provide detailed shape features capable of predicting metabolic health risks.

Currently, another approach to model the human anatomy is to create a learning-based shape analysis framework with deep learning networks.[9] However, this type of methods requires a large image dataset to capture the finer shape changes and, in many cases, especially in applications requiring medical images, datasets of such magnitude are not yet available.

Recently, the work by Rodriguez-Ruiz et al.[10] applied the aforementioned PCA to a PDM of the compressed breast curvature. The work set out to create a full 3D model of the compressed breast by sequentially combining the information in multiple vertical profiles of the 3D breast surface with a previously developed 2D model of the breast shape projected onto mammograms.[11] The 3D-generated shape from this combination included breast shape deformations between the detector cover and the compression paddle. That 3D model was achieved by imaging the patients' breast surface with state-of-the-art structured light (SL) technology while the patient underwent a normal cranio-caudal (CC) view DBT image acquisition. The resulting model was used to characterize the breast curvature under compression. However, that initial proof-of-concept work used a limited number of vertical 2D profiles of the breast curvature instead of the full 3D breast surface information. Furthermore, it was limited in the number of patients included in the study and, due to hardware resources, in the angular coverage of the breast. This restricted the dataset used to develop the breast PDM.

In the present work, a larger study population is scanned with an optimized 3D SL setup to develop an improved PDM of the entire breast shape under compression in the CC view. The correlations of certain patient characteristics with the breast PDM parameters were also evaluated.

2 | MATERIALS AND METHODS

2.1 | Study population

In our study, we aimed to represent the breast screening population, by including breast thicknesses ranging from 30 to 90 mm. Based on previous work,[10] we estimated that the smallest (< 40 mm) and largest (> 80 mm) thickness groups represent roughly 8% of all cases. Therefore, to include approximately 20 cases in these extreme thickness groups, we aimed to collect 250 cases. We invited female patients, age 40 or older, undergoing a DBT exam due to clinical concerns or recall from screening, at Radboud University Medical Center,

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 1 SL scanning system setup: one projector with two cameras is placed on each side of the DBT system (see close-up image) and markers were added to the compression paddle to ensure good alignment in the post-processing step (top blue arrows). DBT, digital breast tomosynthesis; SL, structured light.

Nijmegen, the Netherlands, to participate in this study between March 2019 and February 2020. Exclusion criteria were confirmed pregnancy, bilateral mastectomy, breast implants, or having undergone surgery of both breasts. The study was approved by the medical ethics committee and all enrolled subjects gave written informed consent. Only one breast was scanned per patient, with being artifact-free and matching the DBT breast coverage as selection criteria for using the scans.

2.2 | Image acquisition and SL goodness of fit

During acquisition of a clinical CC-view DBT exam, patients are seated while their breast is positioned and compressed. For this study, the SL scan was performed concurrently with the DBT acquisition. An SL scanning system (HP Inc., Palo Alto, CA, USA) was positioned to each side of a clinical DBT system (MAMMOMAT Inspiration, Siemens Healthineers, Erlangen, Germany), as can be seen in Figure 1. This setup was chosen so that the SL scans did not extend the clinical workflow (and breast compression) by more than 15 s.[12] Since the DBT and SL scans are acquired at the same time, identical information from both image techniques is recorded.

With each set of SL systems consisting of two cameras and a single projector, an increased angular coverage of the breast was achieved when compared to prior work,[10] even for thicker breasts. Furthermore, markers were placed on each side of the compression paddle to ensure an easier alignment in the post-processing step between left and right scans (Figure 1).

2.2.1 | SL scans post-processing

The 3D representation of the scanned breast was generated from the information recorded by the cameras by the accompanying software (HP 3D scan V5, HP Inc., Palo Alto, CA, USA). Since the scans record spurious surfaces from the system beyond the breast surfaces themselves, appropriate cleaning by removal of these surfaces had to be carried out before merging the scans from the left and right sides using the markers placed on the compression paddle (Figure 2). These steps were performed in MeshLab (Visual Computing Lab, ISTI - CNR, Pisa, Italy).

An essential requirement for building shape models with PDMs is that all training examples have their set of points located across the surface at corresponding positions.[13] Therefore, to model the compressed breast surface between the support table and the compression paddle, we started by simplifying and rearranging the breast scan surface points of the training examples.

For this, first, we removed all the points from the breast surface scan (BSS) in which the coordinates were outside the volume delimited by the detector area or outside the compressed breast thickness (T), defined by the height between the support table and the compression paddle, as reported in the DICOM header of the corresponding DBT acquisition. Subsequently, breast shapes represented by between 200 000 and 1 000 000 points were subsampled to approximately 10 000 points using Poisson-disk sampling method.[14] The remaining points were converted to a cylindrical coordinate system and interpolated on a grid with 256 regular angular steps ranging from [−90°, 90°] and 1 mm high layers, in order to produce a regularly sampled set of points for all cases. The 1 mm-apart points represented the centres of each layer, so they were positioned from 0.5 mm to T − 0.5 mm. We used the kriging method,[15] assuming a spherical model, not only as our interpolation method to produce a gridded 2D array with the desired estimated point locations, but also as an extrapolation method to fill in missing points in small areas for that same gridded array. The regularly sampled scans were then saved as a 2D distance map, specifying only the cylindrical coordinates (Figure 3), and also converted back to a point cloud as a new fitted BSS. In this distance map, the pixel intensity matched the distance between the axis through the centre of the detector at the chest-wall side (origin of distance map) to the breast surface (i.e., the radial distance in the cylindrical coordinate system, Figure 3, left), whereas the breast height (i.e., the height in the cylindrical coordinate system) was recorded as the y-axis of the image, and the points at predefined angles were recorded as the x-axis of the image (i.e., the angular position in the cylindrical coordinate system, Figure 3, right). This origin of the distance maps was kept fixed across patient cases.

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

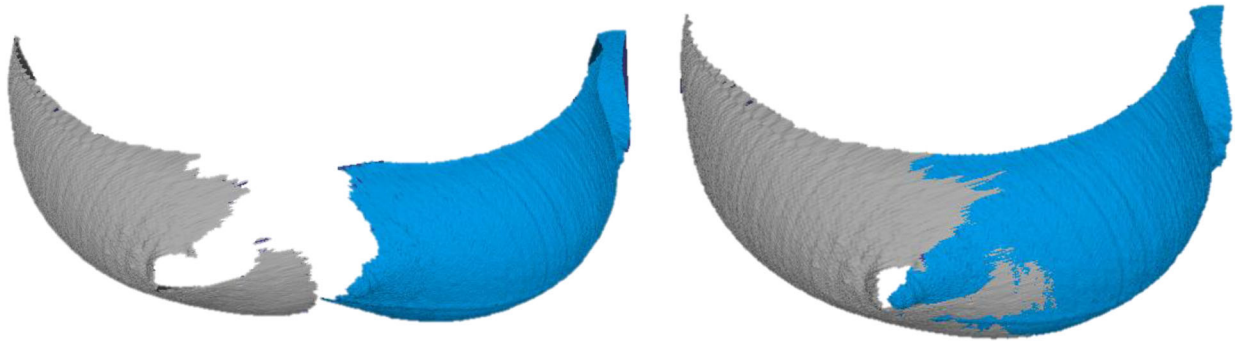

FIGURE 2 Merging of a cleaned scan in MeshLab (left) before and (right) after.

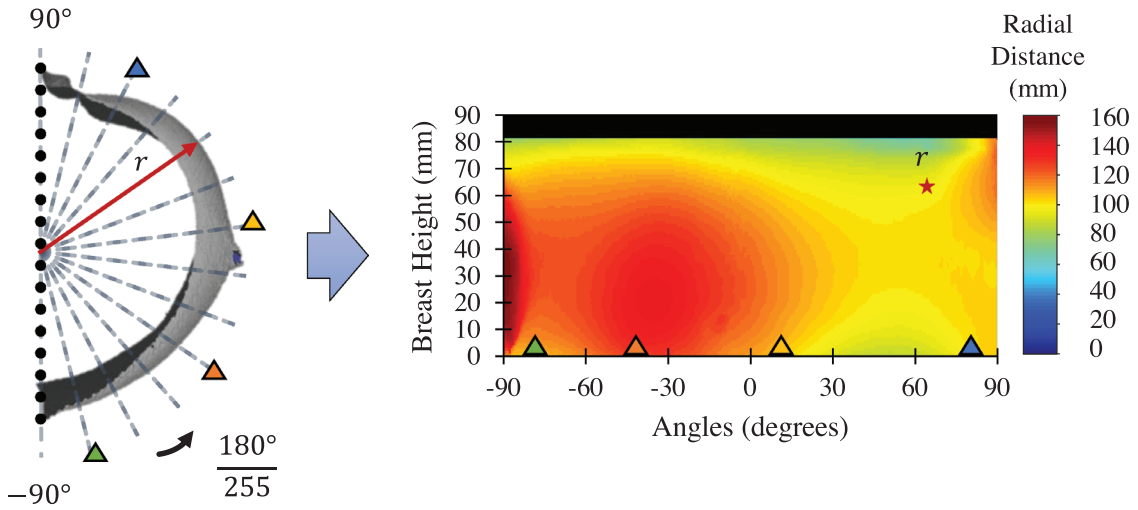

FIGURE 3 Graphical example of a distance map for an 84 mm phantom (*r* represents the radial distance vector).

2.2.2 | Goodness of fit in BSSs

To evaluate the performance of the interpolation process, a comparison between the initial BSS after cleaning and the new fitted scans was performed to determine the absolute difference between these scans. Additionally, the correspondence between SL and DBT scans was assessed by looking into features such as the chest-to-nipple distance (CND), the projected breast area (PBA) and the length along the chest-wall (LCW). These features were obtained from both the DBT exams and the BSSs and compared using a paired t-test, in normal distributed data, or its equivalent non-parametric test, the Wilcoxon signed rank sum test, when the data was not normally distributed (with a significance level of 0.05).

2.3 | Dimensionality reduction and data analysis

After simplification of the BSSs to result in distance maps, the next step was to reduce the dimensionality of the dataset through PCA. In this manner, a smaller set of independent parameters comprising most of the information in the data points[16] can be obtained. This was performed in MATLAB 2018b (MathWorks, Natick, MA, USA).

The crucial requirement for PCA to work in this case is that we have the same number of points for each realization of the data, and that each data point represents corresponding locations across each realization. Because the distance maps of the BSSs represent the breast surface points at the same angular coordinates but at varying heights across cases, each case was then resampled by linear interpolation to a set of 30 equidistant heights along the breast thickness. The start and end of the breasts surface points were kept fixed at their original positions. The cases representing a left-side breast were also mirrored to match the right-side cases before performing the PCA. This ensured that breast shape deformations that occurred during breast compression, and were described by the PCA components, were related to medial and lateral shape variations, and not averaged out by analysing opposite sides (medial vs. lateral) of contralateral breasts.

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The coefficients obtained using PCA were then plotted as individual histograms together with the corresponding Gaussian distribution fit (mean μ, standard deviation σ), and a normality test (D'Agostino and Pearson's[17,18] test with a p-value = 0.05) was performed for each component.

2.4 | Breast PDM

2.4.1 | Breast shape generation

To generate new and realistic breast shape distance maps, new coefficients were sampled randomly from these Gaussian fit functions. The sampling was done even for coefficients that did not follow the normality assumption, but it could optionally be performed using exclusively the empirical cumulative distribution function of the coefficients. The PCA coefficients obtained can parameterize a given breast shape as an offset from the sample means.

2.4.2 | Evaluation of PDM residuals

The PCA residuals (prediction errors) were evaluated by calculating the mean absolute error, in millimetres, (MAE, Eq. 1) in order to determine whether the model had adequately captured the information in the dataset.

$$MAE\ (mm) = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right|, \quad (1)$$

where n is the number of cases in our dataset, X_i is the matrix with all the fitted breast shape data points and $\hat{X}_i$ is the predicted breast shape data points through PCA.

2.4.3 | Correlation between breast PDM parameters and patient characteristics

To visually examine the influence each coefficient has in changing the mean breast shape, each coefficient value was varied independently, from $(\mu - 2\sigma)$ to $(\mu + 2\sigma)$, while the remaining coefficients were held constant at their mean values. The ability to describe each PCA coefficient using patient features was also evaluated by computing linear correlations between relevant PCA coefficients and each individual patient feature. The following patient features were either available from the exam DICOM header or determined using Volpara™ (v1.5, Volpara Health, Wellington, New Zealand) from the DBT images acquired concurrently with the BSSs: age, glandular volume (GV), glandular density by mass (GDM), total breast volume (TBV), compressed breast thickness (T), and compression force (CF). The nipple location (NL) and the centre of the chest-wall (CCW) features were measured based on the BSSs. The linear relationships found between PCA coefficients and individual patient features, were then used to develop a PDM capable of modelling and generating new examples of the entire breast shape under compression in a CC view.

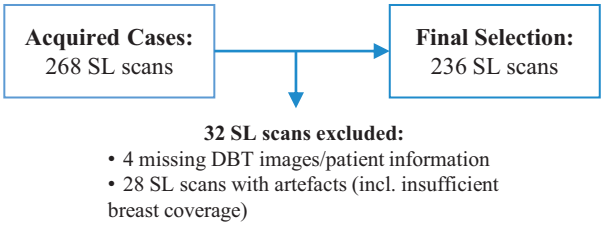

FIGURE 4 Flowchart of SL scans selection. SL, structured light.

2.4.4 | Generative model – python module

The final implementation of the PDM takes advantage of the strong correlation found between one of the PCA coefficients and the breast thickness, to generate breast shapes with specific or random thicknesses. The PDM was made available as a python module and can be found on GitHub (https://github.com/radboud-axti/abreast-generator). The module can use the fitted gaussian distribution to all the PCA coefficients or, alternatively, the empirical cumulative distribution functions for each of the coefficients. The user can decide between generating a random or average compressed breast shape, with or without setting a specific breast thickness (further information can be found with the module package).

3 | RESULTS

3.1 | Study dataset

Figure 4 illustrates how the final dataset of 236 cases was assembled for this study, with Table 1 showing some of this population characteristics after the post-processing step of the BSS. The minimum and maximum breast thickness values in our dataset were of 30 and 92 mm, respectively.

3.2 | Goodness of fit in BSSs

From the included cases, the 90th and 99th percentile of the interpolation error between the initial BSS after cleaning and the corresponding fitted surface scans, was calculated and the median value for the 90th and 99th percentile resulted in a 0.50 and 1.16 mm error

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

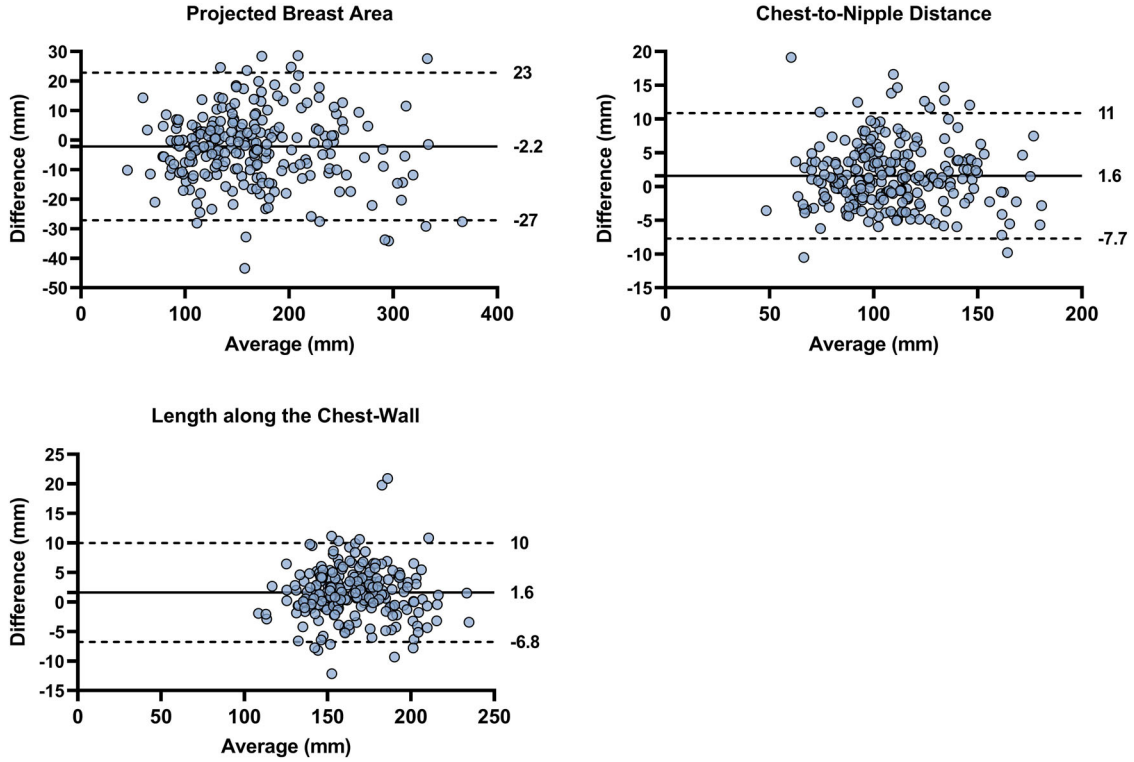

FIGURE 5 Bland-Altman plots comparing PBA (in cm^2), CND (in mm) and LCW (in mm) measurements across BSSs and DBT scans. Mean difference is shown with a solid line, together with $\pm\sigma$ shown as dashed lines. CND, chest-to-nipple distance; LCW, length along the chest-wall; PBA, projected breast area.

TABLE 1 Population characteristics with median values and interquartile ranges in brackets

Population characteristics	Median (interquartile range)
Age (years)	60 (54–67)
Imaged breast laterality	Left: 117; Right: 119
Compression force (N)	82.3 (68.9–97.3)
Breast thickness (mm)	62.0 (52.0–69.8)
Glandular density by mass[a] (%)	12.2 (8.5–19.1)
Glandular volume[a] (cm^3)	50.4 (37.4–69.8)
Total breast volume[a] (cm^3)	775.7 (555.2–1145.4)

[a]Values obtained using Volpara™.

between the surfaces, respectively. Further analysis on these fitted BSS and DBT exams was carried out and is reported in Table 2 for the PBA, CND, and LCW, and compared across the two different scans in Figure 5. To avoid mismatch due to non-breast tissue visible in the DBT that was removed in the SL scan, the LCW measurement was performed at 50% distance of the CND instead of at chest wall. Table 2 and Figure 5 show that the small mean differences of 1.6 mm (0.96%) and 1.6 mm (1.51%) for the LCW and CND, respectively, were statistically different (at a significance level of 0.05), as was the projected breast area with a mean difference of 2.2 cm^2 (1.59%).

3.3 | Principal component analysis

The PCA performed in our data captures 96.61% of the cumulative data variance for the first PCA component, while 99.76% is included in the first three PCA components, and 99.98% with 15 PCA components. Figure 6 illustrates the distribution of the estimated 15 PCA components and the corresponding Gaussian distribution fits. All components, except the 5th, 8th, 9th, 11th, and 12th, were found to be normally distributed (p-value < 0.05).

The analysis of the prediction errors from the PCA showed a median MAE of 0.68 mm (maximum of 3.53 mm and minimum of 0.44 mm). Most of the prediction errors were at the borders of the breast (see Figure 7), which may be due to the variability found in the posterior part of the scanned breast, mostly due to skin folds.

3.4 | Breast PDM characterization

Visual inspection of the influence that each component could have on the deformation of the mean breast shape seemed to indicate that the first three PCA components were related with the size of the breast, while the rest of the components were needed to capture smaller shape variations (see Figure 8 and the Figure S1).

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

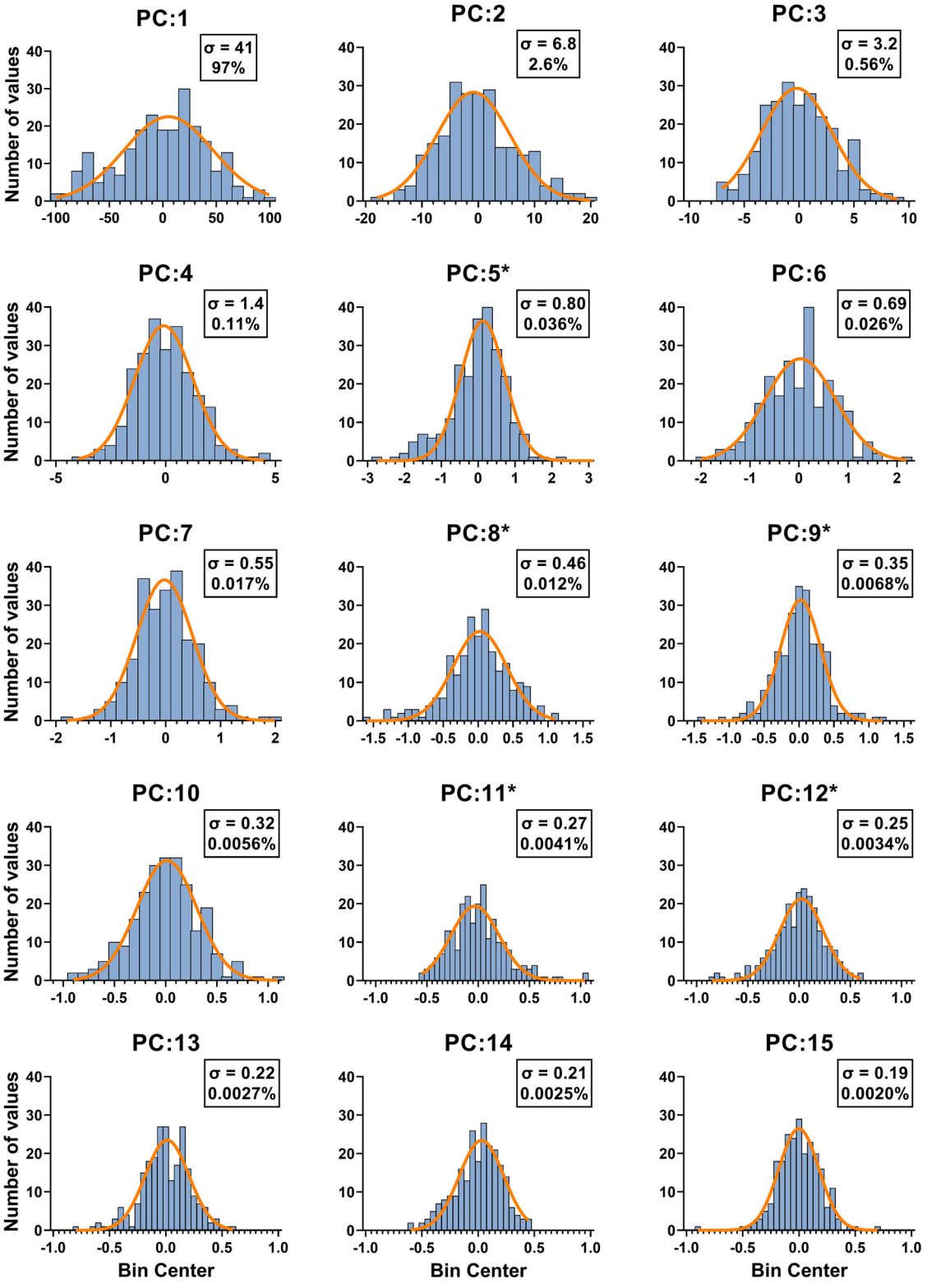

FIGURE 6 Gaussian distributions of the 15 PCA components with mean zero and the specified standard deviation and the data variance percentage represented by each component. The * indicates the components that are not normally distributed. PCA, principal component analysis.

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Comparison of breast tissue coverage in BSSs and DBT scans

	Projected breast area (cm^2)	Chest to nipple distance (mm)	Length along chest-wall[a] (mm)
DBT	157.1 (122.9–196.5)	103.9 (89.8–122.8)	162.9 (148.4–178.4)
BSS	156.2 (116.5–195.1)	105.8 (89.7–122.7)	166.4 (148.9–180.9)
Mean difference	−2.2 ($p < 10^{-4}$)	1.6 ($p < 10^{-4}$)	1.6 ($p < 10^{-4}$)
Percentage difference	−1.59%	1.51%	0.96%

Median values are reported together with interquartile range within parentheses. The *p*-values report the result of the statistical test with the null hypothesis that the two groups do not differ statistically at a significance level of 0.05.
Abbreviations: BSS, breast surface scan; CND, chest-to-nipple distance; DBT, digital breast tomosynthesis.
[a]Measured at 50% distance of the CND.

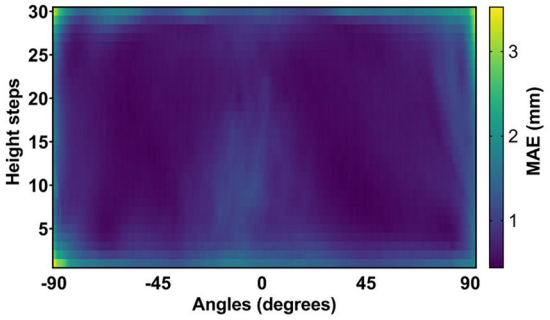

FIGURE 7 MAE (in mm) over the mean breast shape representation: the y axis shows the 30 equidistant height steps along the breast thickness, while the x axis represents points of the mean breast shape representation at predefined angular positions. The lower error in the centre of the breast (dark blue) compared to the borders of the breast, especially top and bottom corners (yellow). MAE, mean absolute error.

Figure 9 shows the relationships between the three first PCA coefficients and the recorded individual patient features with correlation larger than 0.4. The correlations plotted here match what was observed in the above Figure 8 through visual inspection.

A very strong simple linear relationship was found between the first PCA coefficient, PC1, and the breast thickness($p < 0.05$). For the second and third coefficients, PC2 and PC3, we found that the CND and the position of the breast along the chest-wall edge of the detector is correlated with these coefficients. We could observe a good correlation between PC2 and the CND ($p < 0.05$) and between PC3 and the centre of the chest-wall ($p < 0.05$) and the nipple location ($p < 0.05$). No strong correlations were found for the remaining coefficients and patient features ($R^2 < 0.4$ in all other instances).

4 | DISCUSSION

The simulation of the breast under compression is a requirement for certain image processing algorithms when working with mammography and DBT. The present work aimed to improve a previously reported breast shape model.[10] Our model was based on a larger dataset with 236 patient cases and included an optimized 3D SL scan acquisition setup, which enabled the full coverage of the breast surface under compression in the CC view. Compared to previous work,[10,12] the breast PDM development was updated in various ways. First, two 3D SL scanning systems were used, one on each side of the DBT machine, with two cameras per projector (stereo scanning mode, enhancing coverage). Second, the scanned 3D breast surface was analysed as a whole, as opposed to analysing a limited number of vertical 2D profiles. Finally, this whole-surface analysis resulted in a complete 3D model, rather than achieving one by combining an independent 2D model of the breast shape projection with the model of the 3rd dimension from the surface scans. When compared to merging information from two complementary perspectives (the breast shape projection and curvature),[10] modelling the 3D shape of the breast directly has the advantage of including shape correlations and deformations that are not accounted otherwise. This enabled us to characterize the model regarding certain patient features obtained from the concurrently acquired DBT scans. The capability to directly correlate between patient features and the model, makes this a straightforward approach when compared to a deep-learning based one, where a block-based framework makes it not so easy to implement[9] and where fine tuning of several parameters might be needed.

Our DBT dataset results in comparable values to previous studies, particularly for thickness (mean = 60 mm) and projected breast area (mean = 160 cm^2),[10,11,19] which points to our sample being representative for the breast cancer population. Additionally, our study included more cases with thicknesses at least one standard deviation from the mean (below 46 mm and above 76 mm), when compared to the previous dataset[10] (39 vs. 8 and 19 vs. 3, respectively). This has the advantage that our model generates breast shapes within a range that is based on actually acquired data when compared to the previous model, in which for some thicknesses the data was instead extrapolated. The fact that the post-processing of the scanned breasts resulted in a 99th percentile median error of 1.16 mm also shows that

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

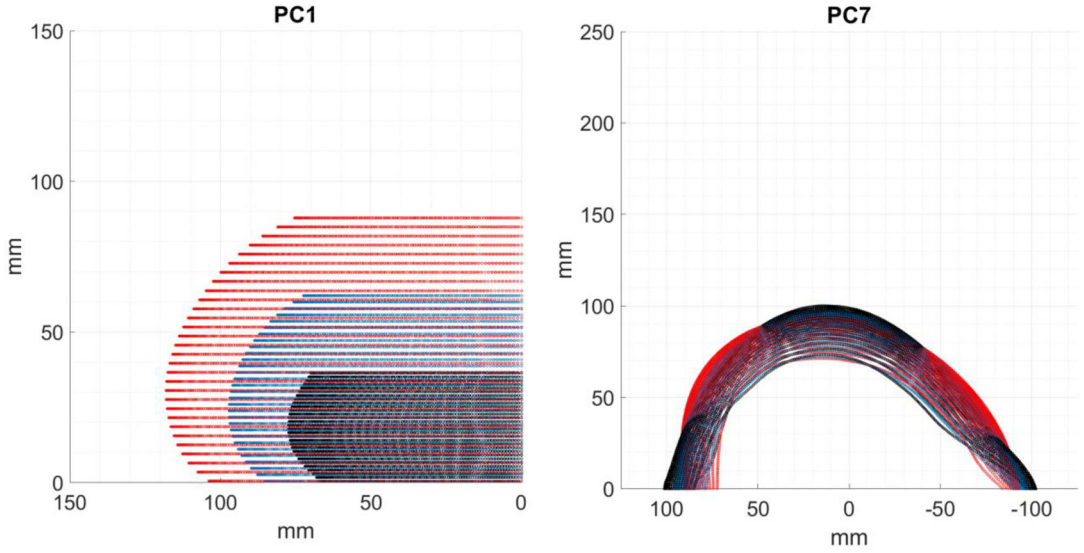

FIGURE 8 Influence of the first PCA component in the generated breast shape (left) and of the seventh component (right): Mean breast representation (blue), mean breast representation + 2 standard deviations (red) and – 2 standard deviations (black). The breast shapes generated can range in thickness between the minimum and maximum values reported in our dataset (i.e., 30–92 mm): the left side figure presents thicknesses of 35 mm (black), 62 mm (blue), and 90 mm (red). PCA, principal component analysis.

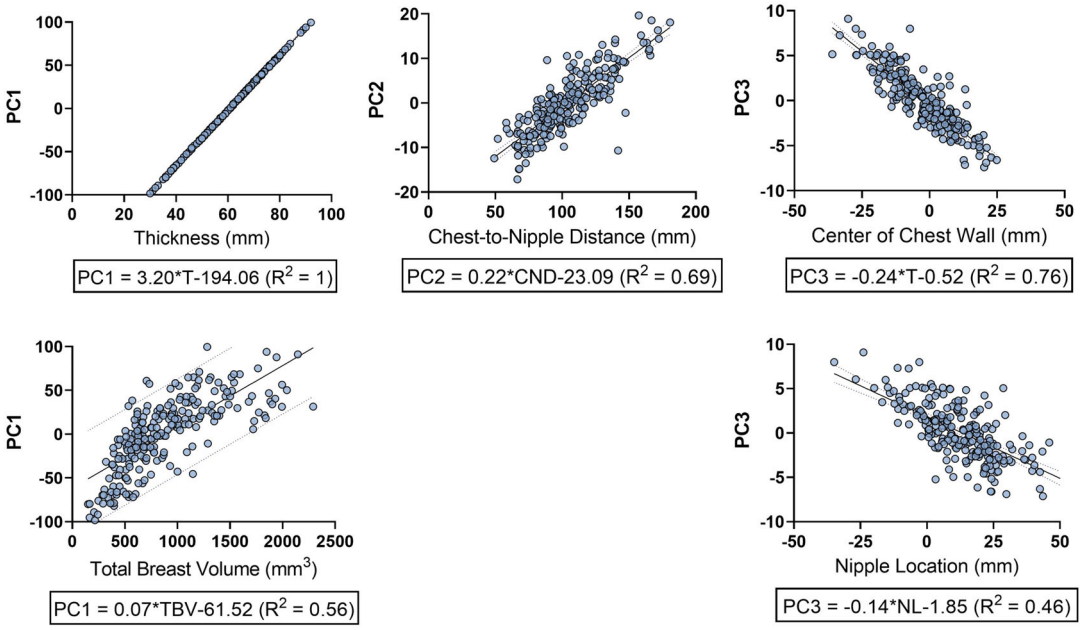

FIGURE 9 Simple linear regressions found relevant ($R^2 \geq 0.4$) between the first three PCA coefficients and some patient features. PCA, principal component analysis.

our fitted BSS, used as inputs for our model, were truthful to the original BSS and did not deform the original shape substantially.

The posterior comparison between the new fitted BSS and the DBT exams in Table 2, together with Figure 5, demonstrated that the reported projected breast area, length along the chest-wall, and chest-to-nipple distance were statistically different (at a significance level of 0.05). We believe these small yet statistically significant differences are caused by the cumulative effects present in the post-processing of the BSS scans, such as the previously-reported error of 3 mm found in our scanning system,[20] the merging of left- and right-side scans, the alignment to the DBT data, and the fitting of the BSS. However, the median values reported in Table 2 show that there is a good agreement on the breast shape coverage between SL cameras and DBT images, with the reported values showing a mean difference smaller

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

than the presented interquartile range. This confirms that, overall, our fitted BSS capture the correct dimensions of the breast shape, but that for some cases where the quality of the original scan is poorer, special attention is needed when post-processing the BSS. Additionally, we are certain that no further systematic biases were introduced in the model's dataset since our breast PDM achieved a median MAE of 0.68 mm among all the fitted breast shapes and the ones predicted through PCA, which is smaller than the error obtained during the pre-processing of the scanned breasts.

Our choice to use the DICOM reported compressed breast thickness in pre-processing the SL scans resulted in a perfect correlation between the first PCA coefficient and the patient breast compressed thickness, which made it possible to use it as a variable that can be user defined when generating new model-based breast shapes. The described model is now available as a python module on GitHub. The moderate correlations found, regarding the second and third PCA coefficients and certain patient features (chest-to-nipple distance, centre of the chest-wall and the nipple location), were not used in our model as their coefficient of determination (R^2) were insufficient to ensure that the generated breast shape could present the patient feature assigned by the user. Nonetheless, these correlations mean that our generated model breasts also include variation in the chest-to-nipple distance and shifts parallel to the detector edge, as seen in clinical practice.

Our work had some limitations, namely, the fact that for some cases artifacts arose while using the SL scanning systems and the fact that the post-processing of BSS limited the details retrieved of the breast shape (resampled done regarding the height of the breast removed the nipple, as shown in Figure 8, and small variations across breast curvature). However, we believe the data validity analysis results provide insights into the good accuracy obtained by the model and its capability of generating realistic breast compressed shapes.

As the next step, for future clinical and research studies, we would like to complete this compressed breast shape model by including the breast shape obtained during a medio-lateral oblique view image acquisition.

5 | CONCLUSION

We developed a patient-based 3D generative model of the compressed breast shape for DM and DBT in CC view and made it publicly available. The model enables the simulation of as many randomly shaped compressed breasts as needed, even for a specific defined thickness. Thus, the model can be used for future studies in which simulation of the compressed breast shape is required, such as image processing and/or reconstruction research.

ACKNOWLEDGEMENTS

We would like to thank the collaboration of Ruby Egging and Floor Schampers in the collection and processing of data. This work was supported by Siemens Healthcare.

CONFLICT OF INTEREST

The authors have the following conflicts of interest to disclose: I. Sechopoulos–research agreements: Canon Medical Systems, Siemens Healthcare, Sectra Benelux, ScreenPoint Medical, Volpara Health; R. Biniazan and S. Kappler are employees of Siemens Healthcare.

DATA AVAILABILITY STATEMENT

The described Point Distribution Model is publicly available for download as a python module on GitHub (https://github.com/radboud-axti/abreast-generator).

REFERENCES

1. Pinto MC, Michielsen K, Rodríguez-Ruiz A, Biniazan R, Kappler S, Sechopoulos I. Deep learning scatter estimation for breast tomosynthesis based on a realistic compressed breast shape model in the CC view. *Proc. SPIE 12031*. 2022;120310C:1605-7422. 10.1117/12.2610772
2. Michielsen K, Rangelova T, Sechopoulos I, Patient evaluation of breast shape-corrected tomosynthesis reconstruction. In: *Proc. SPIE 10948*. 2019; 1094813:38. 10.1117/12.2513436
3. Acciavatti RJ, Reiser I, Sechopoulos I, et al. Analysis of volume overestimation artifacts in the breast outline segmentation in tomosynthesis. 2018;1057359(March 2018):195. 10.1117/12.2293253
4. Tian IY, Ng BK, Wong MC, et al. Predicting 3D body shape and body composition from conventional 2D photography. *Med Phys*. 2020;47(12):6232-6245. 10.1002/mp.14492
5. Ng BK, Sommer MJ, Wong MC, et al. Detailed 3-dimensional body shape features predict body composition, blood metabolites, and functional strength: the Shape Up! studies. *Am J Clin Nutr*. 2019;110(6):1316-1326. 10.1093/ajcn/nqz218
6. Hennessy RJ, McLearie S, Kinsella A, Waddington JL. Facial shape and asymmetry by three-dimensional laser surface scanning covary with cognition in a sexually dimorphic manner. *J Neuropsychiatry Clin Neurosci*. 2006;18(1):73-80. 10.1176/jnp.18.1.73
7. Brunton A, Salazar A, Bolkart T, Wuhrer S. Review of statistical shape spaces for 3D data with comparative analysis for human faces. *Comput Vis Image Underst*. 2014;128:1-17. 10.1016/j.cviu.2014.05.005
8. Mansi T, Voigt I, Leonardi B, et al. A statistical model for quantification and prediction of cardiac remodelling: application to tetralogy of fallot. *IEEE Trans Med Imaging*. 2011;30(9):1605-1616. 10.1109/TMI.2011.2135375
9. Gutiérrez-Becker B, Sarasua I, Wachinger C. Discriminative and generative models for anatomical shape analysis on point clouds with deep neural networks. *Med Image Anal*. 2021;67:101852. 10.1016/j.media.2020.101852
10. Rodríguez-Ruiz A, Agasthya GA, Sechopoulos I. The compressed breast during mammography and breast tomosynthesis: in vivo shape characterization and modeling. *Phys Med Biol*. 2017;62(17):6920-6937. 10.1088/1361-6560/aa7cd0
11. Rodríguez-Ruiz A, Feng SSJ, Mann JR, D'Orsi CJ, Sechopoulos I. Improvement of an objective model of compressed breasts undergoing mammography: Generation and characterization of breast shapes. *Eur Congr Med Phys*. 2017;44(6):2161-2172
12. Pinto M, Egging R, Rodríguez-Ruiz A, Michielsen K, Sechopoulos I, Compressed breast shape characterization and modelling

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

during digital breast tomosynthesis using 3D stereoscopic surface cameras. *Proc. SPIE 11513*. 2020;1151307: 1151307-1-1151307-7. 10.1117/12.2563823
13. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. *Med Image Anal*. 2009;13(4):543-563. 10.1016/j.media.2009.05.004
14. Corsini M, Cignoni P, Scopigno R. Efficient and flexible sampling with blue noise properties of triangular meshes. *IEEE Trans Vis Comput Graph*. 2012;18(6):914-924. 10.1109/TVCG.2012.34
15. Stytz MR, Parrott RW. Using kriging for 3d medical imaging. *Comput Med Imaging Graph*. 1993;17(6):421-442. 10.1016/0895-6111(93)90059-V
16. Jolliffe I. Principal component analysis. In: Lovric M, ed. *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg; 2011:1094-1096. 10.1007/978-3-642-04898-2_455
17. D'Agostino RB. An omnibus test of normality for moderate and large size samples. *Biometrika*. 1971;58(2):341-348.
18. D'Agostino RB, Pearson ES. Tests for departure from normality . empirical results for the distributions of b2 and $\sqrt{}$ b1. *Biometrika*. 1973;60(3):613-622.
19. Boone JM, Lindfors KK, Cooper VN, Seibert JA. Scatter/primary in mammography: comprehensive results. *Med Phys*. 2000;27(10):2408-2416. 10.1118/1.1312812
20. Schampers F, Pinto MC, Michielsen K, Accuracy and precision of a structured light scanning system for acquiring 3D information of the breast under compression. In: *Proc. SPIE 11595*. 2021:115954T. 10.1117/12.2581913

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pinto MC, Michielsen K, Biniazan R, Kappler S, Sechopoulos I. Generative compressed breast shape model for digital mammography and digital breast tomosynthesis. *Med Phys*. 2022;1-11. https://doi.org/10.1002/mp.16133

24734209, 0, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16133 by University Of Twente Finance Department, Wiley Online Library on [18/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License