# Practical applicability of Model Observers in breast tomosynthesis

**Hilde Bosmans, Dimitar Petrov, Nicholas Marshall**

*Department of Radiology, University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium*

[EUSOBI, Digital Breast Tomosynthesis Course 2016, Vienna, AT]

UZ Leuven | Herestraat 49 B - 3000 Leuven | www.uzleuven.be tel. +32 16 33 22 11 | UNIVERSITY HOSPITALS LEUVEN

## Overview

- Introduction
- Formulation of the task
- Recipee
- Applications:
  - Reproducibility
  - Dose sensitivity
  - Applicability on different systems
  - Different image reconstruction methods
- Alternative in the Fourier domain
- Future outline
- Conclusion

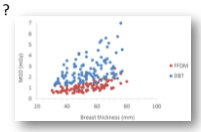
## Some abbreviations

- MO: Model Observer
- DBT: digital breast tomosynthesis, 'tomo'
- FFDM: full field digital mammography, '2D mammography'
- HO: Human observer
- CHO: Channelized Hotelling Observer
- 4-AFC: four alternative forced choice test
- Calcs: microcalcifications
- PS: power spectrum

## Challenge 1

**Clinical:** How should we be using DBT (in screening) ?

- DBT versus 2D FFDM ?
- How to use DBT ? As substitute or as add-on ?



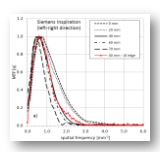
- Clinical trial ?
  - For every reconstruction algorithm ?
  - For every dose level?
  - For every monitor?
  - For every level of recall – rate (every country or screening tradition?)

## Statement: it is impossible to run a trial for every question that pops up

- but it would be good, even desirable, if we could provide at least some answers

## Challenge 2

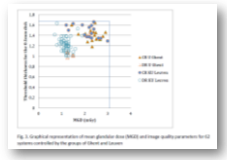
**Technical:** How to assess the technical quality of DBT?



- How to proof DBT outperforms 2D FFDM ?
- How to optimize a DBT modality ?
  - Large versus smaller angle ?
  - Dose and quality balance ?
- Quality follow-up in time ?

## Statement: the use of (non structured) phantoms of FFDM is not going to provide any answers

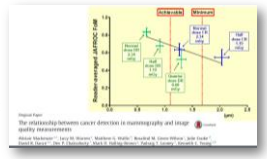
- There is a need for a new approach
- CDMAM has played a major, POSITIVE role in ensuring a minimum quality standard for mammography screening… DBT may benefit from a similar, critical performance test too.



## Challenge 3

If we had a suitable '3D structured contrast – detail' phantom, could phantom reading then be automated with a human predicting measure?

- Success story of CDMAM…
  - Detectability of calcs versus cdmam score (gold disks)
  - Cdcom versus human reading
  - d' versus cdcom and human reading



- Time to focus also on masses, example: spiculated lesions and non spiculated lesions ?

## Statement

DBT is as digital as FFDM…

- Detectability of microcalcifications may also be thé item of interest for testing in DBT
- It may be a proper moment in history to focus also on masses

UZ LEUVEN

## Research questions

- Describe the performance of DBT (versus FFDM) in terms of the task of detectability of microcalcifications, non spiculated masses and spiculated masses
- Have model observers take over the tedious observer work

UZ LEUVEN

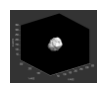
## The choice is yours ….

- Which (3D?) MO?
  - In the image (spatial) domain or in the fourier domain ?
  - Ideal model observer or antropomorphic model observer
  - Type of input images (simulated or real)
- Categorisation of mammographic background (in DBT). Which type of mammographic background to use?
- Categorisation of calcifications, spiculated and non spiculated masses (in DBT)… what is a "representative" (spiculated) mass? which signal template ?????
- Figure of merit?

UZ LEUVEN

## Basic idea to cope with unknown "representative" lesions ('signals')

1. Start with a particular model of a mass and apply the MO…
   is it detectable?



2. How well is another mass detectable?
3. …
4. How well is a particular calcification detectable ?

…

10. If all these N tested calcifications and masses are detectable, maybe all calcifications and masses are detectable

UZ LEUVEN

## A classical research question in human vision research:

How well are 'gaussian' signals detectable in a particular (homogenous or isotropic) background ?

## 'Our' research question:

How well is a specific or a representative calcification, non spiculated or spiculated mass detectable in a DBT reconstructed plane?

UZ LEUVEN

## Do you agree….

Classical, as in vision studies:

- Select or create background images
- Include signal of choice
- Study or develop an MO for the signal put in the image

'Our' application:

- Select (real looking) microcalcification or masses of interest
- Make sure it appears in 'realistic' background images in a realistic way. Example: use a phantom that includes the lesion model
- Apply the MO for the template of the object (not as it is visualized by the system)
- Example: signal template of a mass is the midslice in the 3D computer model. As for microcalc, a Gaussian blob with realistic FWHM

UZ LEUVEN

## Literature review of MO in DBT

- Stefano Young ; Subok Park ; S. Kyle Anderson ; Aldo Badano ; Kyle J. Myers, et al. "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms", Proc. SPIE 7258, Medical Imaging 2009: Physics of Medical Imaging, 72580O (March 13, 2009);
- Lynda Ikejimba ; Stephen J. Glick ; Ehsan Samei and Joseph Y. Lo "Comparison of model and human observer performance in FFDM, DBT, and synthetic mammography ", Proc. SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging, 978325 (March 22, 2016);
- …

## Recipee

1. Determine (develop) MO algorithm, test statistic & FOM
2. Images, with and without signals
   - Acquire phantom images or generate simulated images
   - Signal templates
3. Have images read by humans (HO)
4. Prepare data set for training MO and train MO
5. Acquire images for testing MO and apply MO
6. (Correlate with human observation)
7. (Improve)
8. Apply for specific application

## 1. Determine (develop) MO algorithm, test statistic & FOM

Based on literature, we started with a CHO, that was then further tuned to our models

- Anthropomorphic MO
- Gabor channels
- No internal noise added
- 2D application, applied on 3 planes around the in focus plane
- For microcalcifications, 243 times repeated (in 9 x 9 x 3 adjacent points)
- For non spiculated masses, 27 times repeated (in of a 5 x 5 x 3 region)

## Determine (develop) MO algorithm, test statistic & FOM

(1) CHO type

$$t(v) = \Delta\bar{v}K_v^{-1}v_i$$

with t(v) is discriminant function

$\Delta\bar{v}$ – signal mean channel output

$K_v^{-1}$ - covariance matrix

$v_i$ – channel output

## 2. Images, with and without signals

Our approach:
(1) 3D structured phantom with 3D printed lesion models

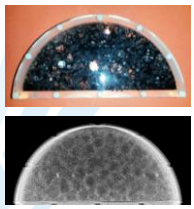
## 3D structured phantom with 3D printed lesions

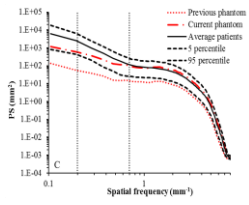
- Starting point was the work of Siewerdsen's group
  - Gang *et al* (2010) described a 'Clutter phantom' based on principles of fractal self similarity
- The power spectrum is used to characterize scene statistics (Torralba *et al* (2003))
- Realization:
  - Acrylic spheres of six different diameters (15.88 mm to 1.58 mm)
  - Acrylic semi-circular container of thickness 48 mm and diameter 200 mm
  - Space between spheres filled with water
  - Important: after shaking the phantom, the background is different, yet the characteristics remain very similar

Gang *et al*. 2010 Anatomical background and generalized detectability in tomosynthesis and cone-beam CT *Med. Phys.* **37**
Torralba and Oliva 2003 Statistics of natural image categories *Network: Comput. Neural Syst*. 14

## The phantom's 3D structure & its power spectrum

- Diameter range and material determine the power spectrum
- Power spectrum






L Cockmartin, submitted for publication

## The phantom's 3D structure & breast tissue simularity

- Height, expressed in breast equivalent thickness

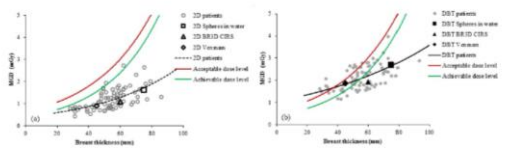
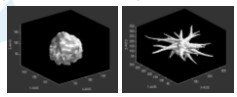


**Figure 2.10:** Mean glandular doses (MGD) of 2D planar (a)and DBT acquisitions (b) for patients and phantom exposures for the Siemens Inspiration system.

## 3D lesion models

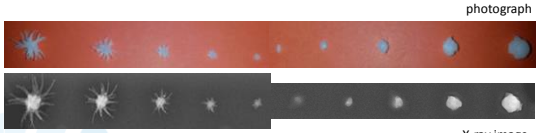
- Microcalcifications: CaCO3 particles glued on a 2 mm thick PMMA plate with liquid PMMA in groups of five single calcs
  - Five calc diameters (µm): 90 -100, 112 - 125, 140 - 160, 180 - 200 and 224 - 250 µm
  - Five calcifications each
- Masses: validated voxel models of a non-spiculated and spiculated masses (Shaheen et al, 2014 )
  - Rescaled to five different sizes (x,y,z) (approx 1.6 to 6.5 mm)
  - 3D printing of models (Materialise, Belgium)



Shaheen et al. 2014, The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis. *Med. Phys.*

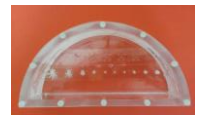
## 3D lesion models

photograph



X-ray image

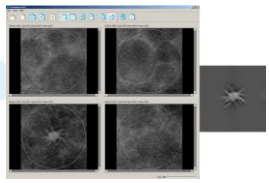
| | PMMA equivalent thickness (mm) | Al equivalent thickness (mm) | Linear attenuation coefficient at 20 keV ($cm^{-1}$) |
|---|---|---|---|
| Next | 1.09 ± 0.02 | 0.010 ± 0.001 | 0.83 ± 0.002 |
| Waterclear | 1.09 ± 0.01 | 0.012 ± 0.001 | 0.85 ± 0.001 |
| Xtreme | 1.12 ± 0.01 | 0.030 ± 0.001 | 1.04 ± 0.001 |
| TuskXC2700T | 1.08 ± 0.01 | 0.038 ± 0.001 | 1.08 ± 0.002 |
| CIRS BR Gland 2066-A-2 | 0.88 ± 0.00 | 0.021 ± 0.000 | 0.79 ± 0.001 |
| CIRS BR Fat 2641 | 0.82 ± 0.01 | -0.0034 ± 0.000 | 0.53 ± 0.001 |
| PMMA | 1.00 ± 0.01 | 0.00031 ± 0.000 | 0.68 ± 0.000 |



UZ LEUVEN

### 3. Have images read by human readers

- 4 AFC, separately for all sizes and lesion types
- Requires a reading platform
  - Thanks Guozhi Zhang (UZ leuven)
- a **team** of readers (medical physicists)



UZ LEUVEN

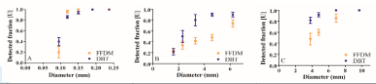
### Have images read by human readers

- Processing Inspired also by CDMAM analysis & contrast detail in general

-> Percentage correct as a function of diameter

-> Psychometric curve fit

-> standard error on the mean & condifence intervals



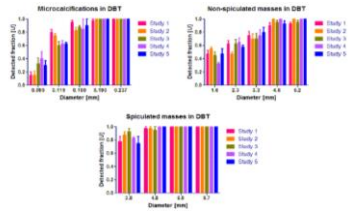
| | | Microcalcifications | | Non-spiculated masses | | Spiculated masses | |
|---|---|---|---|---|---|---|---|
| | | FFDM | DBT | FFDM | DBT | FFDM | DBT |
| System A | $d_0$ | NC | 0.106 | NC | 2.80 | 4.79 | 3.13 |
| | $C_{lin}$ | NC | [0.102-0.110] | NC | [2.31-3.29] | [4.08-5.50] | [2.59-3.66] |
| | $R^2$ | NC | 0.89 | NC | 0.69 | 0.67 | 0.52 |

L. Cockmartin, submitted for publication

UZ LEUVEN

### Have images read by human readers

- Reproducibility: humans are known to have interreader and intrareader variability, and this phantom may lead to some variability too…
  - Tested with 30 DBT series, with shaking. Analysis in groups of 10 DBT acquisitions, 4 readers.




L Van Coillie, master thesis, 2016

UZ LEUVEN

**4. Prepare data set for training MO and train MO;**
**5. Prepare data set for testing MO and run the test**

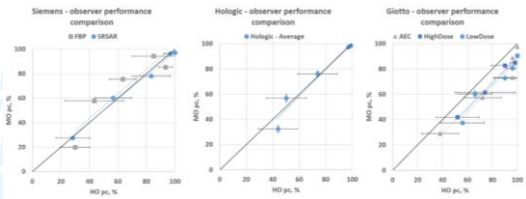
All above items ….

Training:
Example: 15 phantom acquisitions per condition;
42 ROIs without signal per acquisition
14 useful lesions in the phantom

UZ LEUVEN

**6 –> 8. First results**

1. Correlation Human observation – CHO




D Petrov, Submitted to SPIE 2017

UZ LEUVEN

**First results**

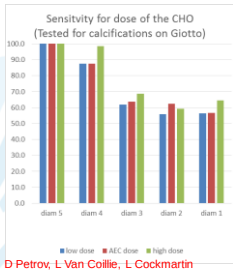
2. Reproducibility, including phantom (shaking), CHO & x-ray system

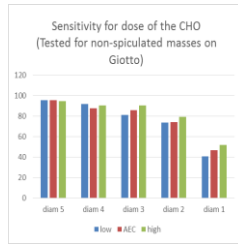

Courtesy: D Petrov, L Van Coillie, L Cockmartin

UZ LEUVEN

## First results

3. Dose sensitvity



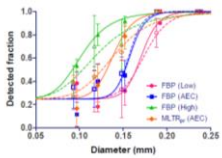



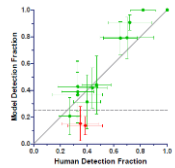

Courtesy: D Petrov, L Van Coillie, L Cockmartin

UZ LEUVEN

## First results

4. Comparison of image reconstruction algorithms
(slightly different CHO algorithm: author K. Michielsen)




K Michielsen, PhD thesis, 2015

UZ LEUVEN

## An alternative & its recipee

UZ LEUVEN

## 1. Determine (develop) MO algorithm, test statistic & FOM

*(1) CHO*

- *Anthropomorphic CHO*
- *Gabor channels*
- *2D*
- *For each image segment, CHO repeated in 27 adjacent 3D points; maximum result is used*

(2) Frequency domain, d'

- Calculated from reconstructed planes (with structure)
- Calculated from contrast, MTF, noise power of structured background
- Visual transfer function
- Signal represented by Bessel function

UZ LEUVEN

## Determine (develop) MO algorithm, test statistic & FOM

*(1) CHO*

$$t(v) = \Delta\bar{v} K_v^{-1} v_i$$

with t(v) is discriminant function

$\Delta\bar{v}$ – signal mean channel output

$K_v^{-1}$ - covariance matrix

$v_i$ – channel output

(2) Frequency domain, d'

$$d' = \frac{\sqrt{2\pi C}\int_0^\infty |S(f)MTF(f)VTF(f)|^2 df}{\sqrt{\int_0^\infty |S^2(f)MTF^2(f)VTF^4(f)NNPS(f)|f df}}$$

$$S(f) = \frac{d}{2}\frac{J_1(\pi d f)}{f}$$

$$VTF(f) = 29.5 f^2 \exp(-4f)$$

UZ LEUVEN

## 2. Images, with and without signals

Our experience with:

*(1) 3D structured phantom with 3D printed lesion models* — **Used with CHO**

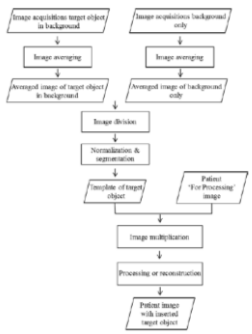
(2) DBT images of real patients with simulated lesions — **Used with d'**

UZ LEUVEN

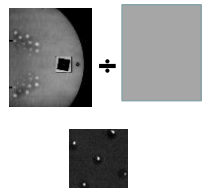
## DBT images of real patients with simulated lesions

- Collect DBT images of (normal) patients / normal breasts
- Acquire images of 'lesions of interest', and get the lesion templates
  - Spherical densities
  - Microcalcifications

UZ LEUVEN



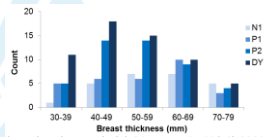

- In short:



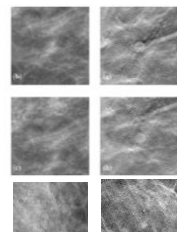
UZ LEUVEN

## DBT images of real patients with simulated lesions

- Can be used with all patient types or in selected groups
- This approach incorporates all characteristics of the system to be tested
- The same lesions can be simulated into different types of modalities (2D FFDM & DBT) or in modalities of different vendors
- 'spherical densities' represent specific cysts; fairly anthropomorphic




L. Cockmartin, *Phys Med Biol.* 2015 May 21;60(10):3939-58



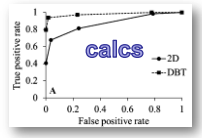
## DBT images of real patients with simulated lesions

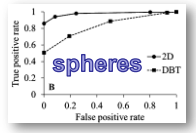
- Repeated high dose acquisitions
- A lot of images – huge data amount
- Careful: time consuming – requires careful scientists, especially for subtle lesions
- Position of the lesion cannot be adjusted after acquisition
- Has to be repeated for all thicknesses
- Requires help of manufacturer to reprocess or re-reconstruct the hybrid images

## 3. Have images read by human readers

- Detection study: ROC type (confidence)
- A team of motivated **radiologists**
- Sara² visualization tool
  - Thanks Jurgen Jacobs, Qaelum NV
- Lesions marked with a box
- 5-point confidence rating scale
- Statistical analysis: DBMMRMC 2.1

-> Area under the curve

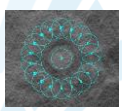


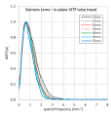



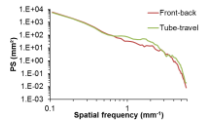

L. Cockmartin, *Phys Med Biol*. 2015 May 21;60(10):3939-58
RSNA 2013 & RSNA 2014

## 4. Collect a limited number of (technical) images

- Measure the MTF of the system: in DBT with a wire at set height
- Measure the contrast of a disk.
- Measure the power spectrum of the (3D structured) images (will also require step response)
- Run excel !



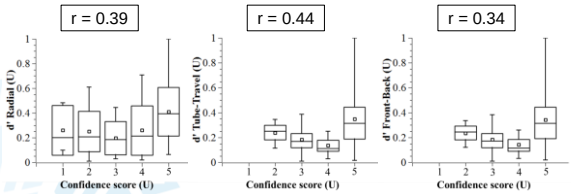





$$d' = \frac{\sqrt{2\pi}C\int_0^\infty |S(f)MTF(f)VTF(f)|^2 fdf}{\sqrt{\int_0^\infty |S^2(f)MTF^2(f)VTF^4(f)NNPS\,(f)|fdf}}$$

UZ LEUVEN

## 6 – 8. First results
## Human scores versus d' of spherical densities



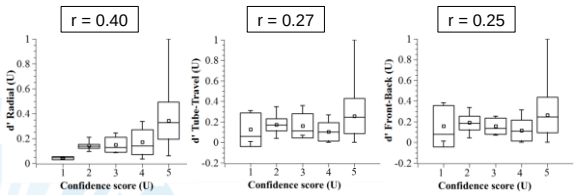

2D DBT (FB) DBT (TT)

L. Cockmartin, IWDM 2014



UZ LEUVEN

## First results: Human scores versus d' of calcs

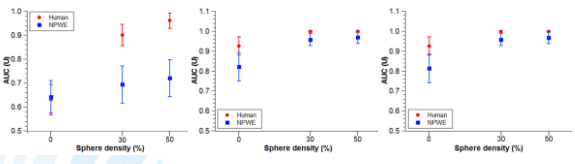



2D DBT (FB) DBT (TT)



UZ LEUVEN

## Comparison ROC results of HU and d' for densities

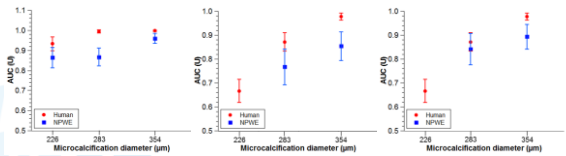



2D DBT (FB) DBT (TT)

L. Cockmartin, IWDM 2014

UZ LEUVEN

## Comparison ROC results of HU and d' for calcs



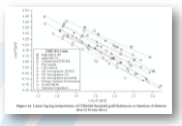

2D DBT (FB) DBT (TT)

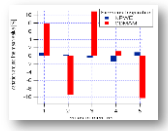
48

UZ LEUVEN

## Results with d'…

While convincing in 2D homogenous background….





It is not yet well enough understood for 3D structured DBT planes.

Dead end?

Fig on the left: Monnin P, *Phys Med Biol*. 2011 Jul 21;56(14):4221-38
Fig on the right: Courtesy N Marshall

UZ LEUVEN

## Future applications

- Understand all results
  - Image domain based approaches
  - Fourier domain approaches
- Apply current CHO approaches on more systems and under more conditions
- Fix first version recipees

## Future developments

- More efficient channels/computer code
- More realistic phantoms & signals; truly antropomorphic ?
- Other (related) tasks
  - estimation of the specific features of the lesion
  - search
- Other channels, for spiculated lesions (maybe not an issue for DBT?)
- Inclusion of the 3D character of DBT (see, between others, the work of Ljiljana Platisa)
- Other Figures of Merit
- Combine vision theory – (radiological) lesion interpretation – under time constraints
- Thinking out of the box…. Would radiologists care about a cluster if all calcs were tiny? (Courtesy K. Michielsen, J Nuyts, SPIE2013)

## Future applications

- Optimize actual DBT systems
  - Hardware & software
  - Exposure settings
  - Reconstruction & viewing settings (example: increment)
  - Viewing display
- Go beyond existing systems (with simulated data) (design – optimize)
- Combine the scores of FFDM & DBT
- Guide introduction of synthetic 2D, maybe….
- Quality assurance
  - Sensitivity of the MO for system problems
  - Predict which deviations cause noticeable effects on human performance

## Future applications

- Large scale application
  - All systems
  - All types of input images
  - All types of lesions
- Consensus & standardisation
- Make the components available to the community
  - Phantom
  - Data sets
  - Human reading results
- Use of the data (in the cloud? Radiomics ?)
- Extend to and/or learn from other imaging modalities, such as plain x-rays, CT, interventional radiology, PET-CT, …

## Conclusion

- In front of a magical experience

## Conclusion

… that can be unraveled right now. After all, DBT is a digital modality.

## Acknowledgement


- All intl. scientists who developed the basics of MOs
- All our human observers !
- The medical physics QA term in Leuven & scientific co-workers
- Our Phd students and master thesis students
- Lesley Cockmartin & Guozhi Zhang
- Our breast radiologists, and more in particular prof C Van Ongeval
- The team of radiographers
- The Lausanne team, with P Monnin, prof F Verdun & prof F Bochud
- The Optimam team, guided by prof. K Young & dr Alistair Mackenzie
- The Maxima team, and more in particular K. Bliznakova
- The scientists in the medical device companies (X-ray equipment and display systems)
- Prof R Oyen, head of the radiology department

UZ LEUVEN

Welcome to Leuven and Europe for our EUTEMPE-RX teaching courses
www.eutempe-rx.eu