Utrecht University

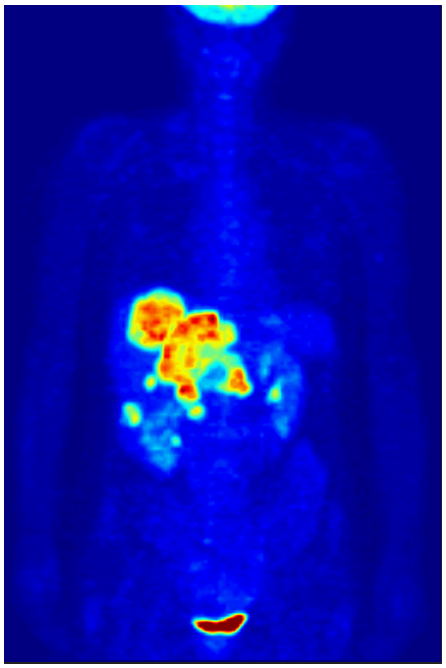
UMC Utrecht

Faculty of Science

Detection of Lesions in Positron Emission Tomography

Bachelor Thesis

Nina L.F. Bezem
Study: Physics and Astronomy



Supervised by:
Dr. Andre Mischke
Utrecht University, Institute for Subatomic Physics (SAP)

Dr. ir. Casper Beijst
UMC Utrecht, Imaging Department

January, 2017

Abstract

Observer studies are often used to compare the quality of Positron Emission Tomography (PET) images. One of the purposes of observer studies is to assess whether a lesion is detectable in an image or not (i.e., determine the detection threshold). However, observer studies are time consuming and the results may depend on the specific observer. Computer models may also be used to predict the detection threshold. The purpose of this study was to develop a novel method to quantify the detection threshold of a lesion, based on the distribution of the signal in the background. PET acquisitions of a phantom with six spheres were used. The distribution of the signal in the background was assessed by looking at the histogram of the contrast-to-noise ratio (CNR) in the background. The shape of the histogram was fitted and normalized to obtain a probability density function. This was tried with both a gamma distribution and a skew normal distribution. This probability density function was used to determine whether a fixed probability could be calculated that would predict the detection threshold. The results showed that for the two smallest spheres, a fixed threshold probability could be found when fitting a gamma distribution. The skew normal distribution fit was subject to numerical errors and therefore a slightly different approach was tried, in which the CNR corresponding to a fixed probabilty p=0.0001 was compared to the reference values. A linear relationship was found, which suggested that it may be possible to express the detection threshold in terms of a fixed probability, although this study did not succeed in quantifying this probability. In conclusion, the novel method that was investigated in this study did not result in an accurate quantification of the detection threshold. Improvements can be made in relation to numerical errors, the choice of probability density function, the method for determining the threshold scans and the measure used to describe how well the human eye can detect lesions.

Image on front page:
By Jens Langner (http://www.jens-langner.de/) [Public domain], via Wikimedia Commons. Retrieved on 13/01/17 from:
$https://commons.wikimedia.org/wiki/File\%3APET-MIPS\_anim\_frame.PNG$.

Contents

1 Introduction

A Positron Emission Tomography (PET) scan is often conducted to investigate if a patient has cancer. The images that are obtained from a PET scan are studied by radiologists, who draw a conclusion from what they see on an image about whether the patient has cancer or not. The quality of such medical images can influence how well lesions (tumors) can be distinguished from healthy tissue. The quality of images in nuclear medicine that have been obtained using different imaging techniques (e.g. scanners, reconstruction algorithms etc.) is often compared using observer studies, which involves a group of people studying and comparing medical images. However, observer studies are time consuming and the results may depend on the specific observer. The development of computers and technology in the last decades has opened up for computer modelled techniques to compare the quality of images, with hopefully more accurate and reproducible results than with humans. One example is model observers, which are simulated observers that are used to replace human observers, which benefits reproducibility and speed. One of the tasks of model observers can be to assess whether a lesion is detectable or not. Since it is not very efficient to have observers determine the detectability of lesions for every image that is produced, it is useful to define a generalized *detection threshold*, a critical point for a certain quantity that can be calculated from the image, at which the lesion goes from being detectable to being undetectable. A detection threshold allows human or computer simulated (model) observers to determine whether the area in an image that could be interpreted as a lesion really is a lesion, or whether it is a result of noise, with a small probability for errors. A detection threshold that is often applied since its development in the 1970's is the Rose criterion [1]. However, the Rose criterion relies on the assumption that the distribution of the signal in the background (the quantity that is measured by the detectors of a PET scanner in the surroundings of regions of interest) of an image follows a normal distribution, which is not always the case in PET images. This gives rise to the idea of trying to find a detection threshold that also can be applied if the signal in the background follows a different distribution. If such a model were to be developed, it would allow clinicians to draw more accurate conclusions about whether or not they observe a lesion in an image.

This motivation led to the following **research question**: Is it possible to develop a method that accurately quantifies the detection threshold for lesions in Positron Emission Tomography (PET)?

2 Theory

2.1 Phantom studies

For research purposes in medical imaging, an object called a *phantom* is often scanned instead of a human in order to avoid that humans are excessively exposed to radioactivity. A phantom is constructed in such a way that its shape is similar to a (part of a) living body. It often consists of a large container in the shape of a human torso with subelements (e.g. spheres) that can be filled separately. The subelements usually represent lesions.

Figure 1: The Siemens Biograph TruePoint PET/CT scanner, an example of a PET/CT scanner used in hospitals. Retrieved on 07/12/16 from the website of Siemens Healthcare: https://www.healthcare.siemens.nl/molecular-imaging/pet-ct/biograph-truepoint-petct.

2.2 Positron Emission Tomography (PET)

Positron Emission Tomography (PET) is a method for tomographic imaging, i.e. making "2-D representations of structures lying within a selected plane in a 3-D object," that is used in nuclear medicine [2]. PET uses detector systems that are placed or rotate around the patient or object of interest, which enables the detector system to take images from many different angles. The images of the different planes are reconstructed mathematically to form a complete 3-D image. Figure (1) shows an image of a typical PET scanner.

2.2.1 PET detector system

The detector system of a PET scanner consists of a ring of detectors, where each detector element consists of a scintillator (e.g. LSO, lutetium oxyorthosilicate), which emits ultraviolet and/or visible light when it interacts with ionizing radiation, followed by a photomultiplier, which converts the light to an electric signal [3].

2.2.2 Positron emission

For PET, positron emitting radionuclides are used [2]. A solution containing a substance that decays by positron emission is administered to the patient before a PET scan. This solution is chemically composed in such a way that it is more likely to accumulate in lesions than in healthy tissue. For example, the glucose-like compound $^{18}F - fluorodeoxyglucose$ (FDG) is sometimes used for PET scans. "FDG accumulation in tissue is proportional to the amount of glucose utilisation. Increased consumption of glucose is a characteristic of most cancers" [4]. Hence, the lesion will have a higher uptake of the radioactive compound. As a result, the lesion will appear to radiate more than the surrounding tissue as the radioactive material decays. The radiation coming from different positions is converted to photons (explained below) and measured by the scanner, which processes this information to produce an image of the region that was scanned by assigning each voxel (pixel in 3 dimensions) a *voxel value* between 0 and 1 according to its colour. The images obtained from a PET scan consist of

black, white and grey voxels. A black voxel indicates that no photons were emitted from that position and is assigned the voxel value 0. A white voxel indicates that a high number of photons was emitted from that position and is assigned the voxel value 1. Voxel values between 0 and 1 correspond to grey voxels of different shades. Consequently, a lesion will appear brighter in the image than its surroundings, which allows clinicians to distinguish between a lesion and healthy tissue.

Now, it will be explained how the radiation emitted from the positron emitting radionuclide is transformed to photons which can be detected by the detectors of the PET scanner. The emitted positrons annihilate with electrons in the body of the patient, as described by the equation:

$$e^+ + e^- \longrightarrow \gamma + \gamma$$

This ususally happens between some tenths of millimetres and a few millimetres from the emission site [2]. Since positrons and electrons are anti-particles, they annihilate upon collision and their masses are converted into energy in the form of two photons of approximately 511 keV. This stems from the relation between the energy E and rest mass m of an electron-positron pair: $E = \sqrt{(pc)^2 + (mc^2)^2}$, where p is the total momentum of the electron-positron pair and the rest mass of an electron or positron is approximately $m_e = \frac{m}{2} = 511$ keV/c^2 [5]. If we assume that the electron and positron are at rest when they annihilate, the total momentum of the electron-positron system is zero and the energy of the photon-photon system is given by $E = mc^2$. In that case, the photons are emitted in exactly opposite directions because of conservation of momentum (which must then also be zero in the photon-photon system). Thus, in theory, they will travel along a line (the *line of response*) in opposite directions away from the annihilation site. Although the positron is slowed down as it moves through matter, the positron (as well as the electron) may have some momentum when the collision takes place. Therefore, the photons will deviate slightly from the ideal line and the angle between them will not be exactly 180°. This effect is called *non-colinearity*. The order of such deviations is typically a few tenths of a degree [2]. The energy of the photons will also deviate slightly from 511 keV.

2.2.3 Detection of radiation: Annihilation coincidence detection (ACD)

Radiation is detected by annihilation coincidence detection (ACD). ACD is based on the idea that if two photons of energy 511 keV arrive at different detectors in the detector ring within a *coincidence timing window* of typically 6-12 nanoseconds, i.e. nearly simultaneously, they are considered to originate from the same position and a line is constructed between the two detectors that measured them. If this process is repeated for many photon pair emissions, the lines will intersect at the position of the lesion. The arrival times of the two photons are recorded by the electronics in the detectosr, which attach digital "time-stamps" to the photons that enter the detectors, typically with a precision of 1 to 2 nanoseconds [2].

2.2.4 Reconstruction

Different types of reconstruction methods can be used in medical imaging. In PET, the Ordered subsets expectation maximization (OS-EM) reconstruction method is usually used

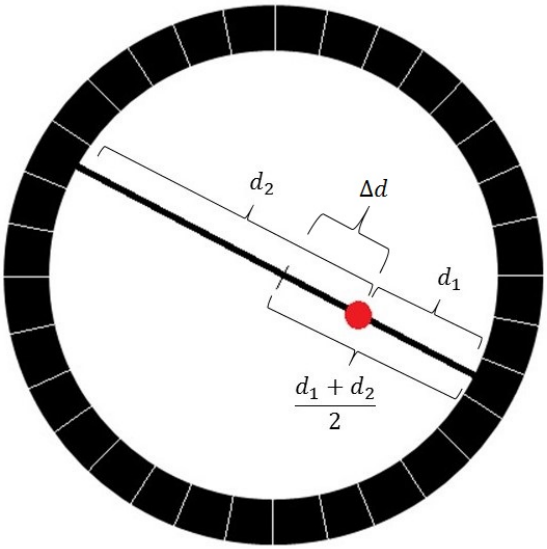


Figure 2: Sketch of a TOF PET reconstruction. The line along which the photons travel connects two detectors. The red spot indicates the annihilation point on this line, from which the photons were emitted. d_i is the distance between the point of emission and the detector for photon i with $i = 1, 2$. The distance between the red spot and the midpoint between the two detectors, Δd, can then be calculated from the arrival times of the two photons.

[6]. This reconstruction method will not be explained further in this thesis, but the interested reader can read more about OS-EM in [6]. A technique used in this reconstruction method and others is *Time-of-flight (TOF) PET*. TOF is used to determine the point of emission along the line of response and in so doing uses the measurements of the arrival times of the photons. Figure (2) illustrates how the TOF technique works. If the arrival times of the photons are given by t_1 and t_2, and the distances travelled by the photons, which can be calculated from t_1 and t_2, are d_1 and d_2, respectively, then the location of the annihilation event (and emission of photons) with respect to te midpoint between the two detectors is given by:

$$\begin{aligned} \Delta d &= \frac{d_1 + d_2}{2} - d_1 \\ &= \frac{d_2 - d_1}{2} \\ &= \frac{\Delta t \times c}{2} \end{aligned}$$

where $\Delta t = t_2 - t_1$ and c is the speed of light [2].

2.2.5 Image degrading effects

Sometimes, detection errors may occur, in which the line between two detectors is not the line on which the annihilation occurred. Apart from a true event, which is detected and interpreted correctly, scatter, random or prompt gamma events can occur [7]. Such events are illustrated in Figure (3). If a photon is scattered, it makes an angle in its path and

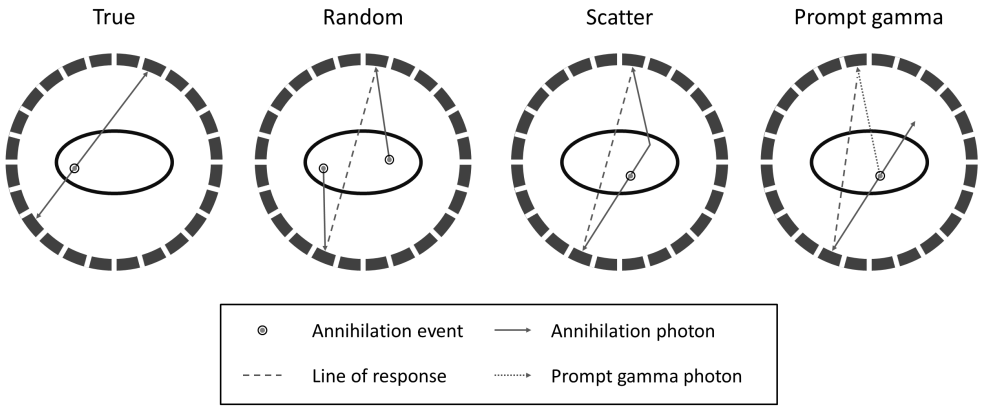


Figure 3: Possible events in PET. Courtesy of Casper Beijst [7].

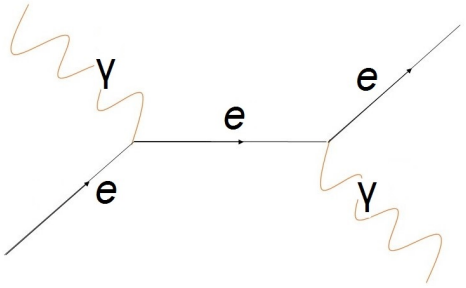


Figure 4: Feynman diagram for Compton scattering. Physical situation: A photon ejects an electron from the outer shell of an atom, thereby changing direction and leaving with a slightly lower energy.

will arrive at a different detector than if it had travelled straight forward. The line that is constructed between the two detectors will not intersect with the point of emission of the photons. Scattering of photons can occur by Compton scattering, the photoelectric effect and pair production, where "Compton scattering is the predominant interaction of x-ray and gamma-ray photons in the diagnostic energy range with soft tissue" [3]. Figure (4) shows the Feynman diagram for Compton scattering. Compton scattering occurs when a photon interacts with an electron in the outer shell of an atom. The electron is ejected and absorbs some of the energy of the incoming photon. Thus, the photon is scattered due to conservation of momentum and is left with a lower energy than before the interaction due to conservation of energy. A random event is an event where two photons from different annihilation events are detected within the coincidence timing window. A prompt gamma event is an event that may occur when an isotope is used that emits a cascade of positrons together with photons [7]. When an annihilation photon and a photon from this cascade are detected within the coincidence timing window, the event is called a prompt gamma event. In a phantom where the positions of radioactive material are known, light-coloured spots at other locations than those of the radioactive material can be a result of scatter, random or prompt gamma events. This is often what is meant by noise in an image (See section 2.3.2). Sometimes, such spots can be mistaken for lesions.

2.3 The quality of images in nuclear medicine

The quality of images in nuclear medicine can be assessed by quantitatively measuring physical characteristics such as spatial resolution, contrast and noise, or by qualitative analysis in observer performance studies [2]. This research project mainly looks into the quantitave measures of image quality, in particular contrast and noise.

2.3.1 Contrast

Contrast is defined as "the ratio of signal change of an object of interest, such as a lesion, relative to the signal level in the surrounding parts of the image" [2].

2.3.2 Noise

Noise can be described as a signal outside the object of interest that may interfere with the signal of the object of interest. In PET images, noise can be described as random or systematic variations in the voxel values of voxels in the volume surrounding a volume of interest, and can be observed as grey or white spots in the image that seem to be randomly distributed over the background (or, in the case of systematic variations, a spot that reoccurs even though it is not a lesion). If clinicians interpret these spots as lesions while they are noise in the background, the conclusion they draw is called a *false positive*.

2.4 Contrast-to-noise ratio (CNR)

The two concepts contrast and noise can be combined into a number called the contrast-to-noise ratio (CNR), although the definition CNR does not explicitly express it as the ratio between the contrast and noise. If μ_l is the mean voxel value in a spot representing a lesion, μ_b the mean voxel value in the background and σ_b the standard deviation of the voxel values in the background, then the contrast-to-noise ratio is defined [8] as:

$$CNR = \frac{\mu_l - \mu_b}{\sigma_b}. \tag{1}$$

The CNR is a measure of how well a volume of interest in the image can be distinguished from the surrounding volume. The detection threshold of lesions is often expressed in terms of the CNR: If the CNR of a volume of interest is above a certain critical value or threshold, one may conclude that the volume of interest is a lesion. Otherwise, the spot that is observed is simply a result of noise.

2.5 The Rose criterion: Detectability of lesions

A lesion can be observed in an image from a scan as a white spot. The Rose criterion quantifies the threshold value for the CNR. The criterion, as introduced by Albert Rose in 1977, states that "to reduce the number of false alarms [false positives] to below unity, we will need a signal whose amplitude is 4-5 times larger than the rms [root mean square, i.e. the square root of the mean square] noise" [1].

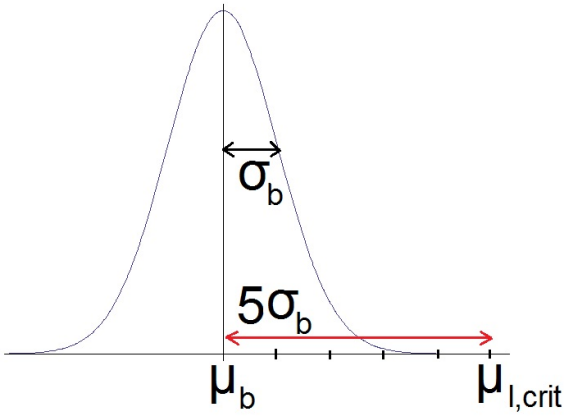


Figure 5: Visualization of the Rose criterion. The curve shows a normal distribution of the pixel values of pixels in the background. On the horizontal axis, the mean pixel value of pixels in the background μ_b and the threshold value for the mean pixel value of the lesion $\mu_{l,crit}$ are shown. Note that the threshold value $\mu_{l,crit}$ lies 5 standard deviations in the background σ_b away from μ_b. This corresponds to a CNR of 5.

In other words, the CNR of a detectable lesion is greater than 4-5. The CNR can be calculated for 2 dimensions or 3 dimensions, where the difference lies in whether the voxel values or the pixel values of an image are considered, and whether volumes or areas of interest are considered. The Rose criterion applies to a CNR calculated in 2 dimensions. In this research project the CNR was calculated in 3 dimensions since the images obtained from the scans are 3-D, and it was assumed that the Rose criterion in 3-D is of the same order of magnitude as in 2-D. Figure (5) shows a visualization of the Rose criterion where it is placed in the context of a distribution of pixel values of pixels in the background.

3 Method

The aim was to develop a method that accurately quantifies the detection threshold of lesions in PET images. In this section, the approach to reach this aim is explained. Firstly, the set-up for the data analysis and the acquisition of phantom data are described. Then, it is explained how the reference detection threshold values were determined and finally, how the novel method for determining the detection threshold was tested.

3.1 Set-up of MATLAB script and choice of data

After a preparatory study into recent research on the topic of image quality in medical imaging in order to obtain some background knowledge on the topic, this study was initiated by trying to understand and expand a script in MATLAB (one of the programmes that was used for the data analysis) that was aimed to process the data that were obtained from scan sessions on a PET scanner. This script could convert the data to a 2-D image that showed one slice obtained from the PET scan. Futhermore, the script could be used to determine the pixel values of the pixels in the image, which could then be used to calculate the contrast-to-noise ratio (CNR) of regions in the background and of the spheres. This was used for further data analysis as explained below, e.g. to plot a histogram of the CNRs in

Figure 6: The NEMA 2012/IEC 2008 phantom, similar to the NEMA 2007 phantom. The phantom is torso-shaped. It contains six spheres and a background compartment that can be separately filled. Retrieved on 07/12/16 from the website of Biodex: http://www.biodex.com/nuclear-medicine/products/pet-positron-emission-tomography/pet-phantoms/pet-phantom-nema-2012iec-200.

the background. The script was evaluated for different scans from three scan sessions which were performed using different radionuclides: ^{18}F (18-fluorine), ^{124}I (124-iodine) and ^{68}Ga (68-gallium). These scan sessions had been conducted using different ratios of the activity in the spheres to the activity in the background. In the scan session with ^{18}F there was zero activity in the background. It appeared from the data that the histograms that presented the distribution of CNRs in the background were sharply peaked at low CNRs, which made it difficult to predict what kind of function could describe the shape well. Also, zero activity in the background is not a good representation of a clinical situation, since there will always be some radioactivity absorbed by healthy tissue in patients. In the scan session with ^{68}Ga the ratio between the activities in the spheres and the background was 17.6:1. Although the ratio that is used in other research varies, it was noted that in several other studies a ratio close to or equal to 10:1 is used ([8],[9],[10],[11]). Therefore, it would be difficult to compare results with existing research using a significantly higher ratio. Also, the histograms from the scan session with ^{68}Ga were similar to ^{18}F and thus sharply peaked around low CNRs, which again made it difficult to accurately determine which function could describe the shape of the histogram. Since a ratio of 10:1 was used in the scan session with ^{124}I, it was decided that this study would focus on that scan session.

3.2 Phantom measurements

In this research project, the data from a scan session on a Siemens Biograph mCT Time-of-Flight (TOF) PET/CT scanner (Siemens Healthcare, Erlangen, Germany) were analyzed. The object that was scanned was a torso-shaped NEMA-2007 image quality phantom with 6 spheres of diameters 10, 13, 17, 22, 28 and 37 mm which are labeled sphere 1, 2, 3, 4, 5 and 6, respectively, and a background compartment. An image of such a phantom can be seen in Figure (6). The spheres and background compartment were filled with a solution containing the positron-emitting isotope of iodine ^{124}I, with an activity ratio between the spheres and the background compartment of 10:1. The initial activity concentration was 1.8×10^4 Bq/ml

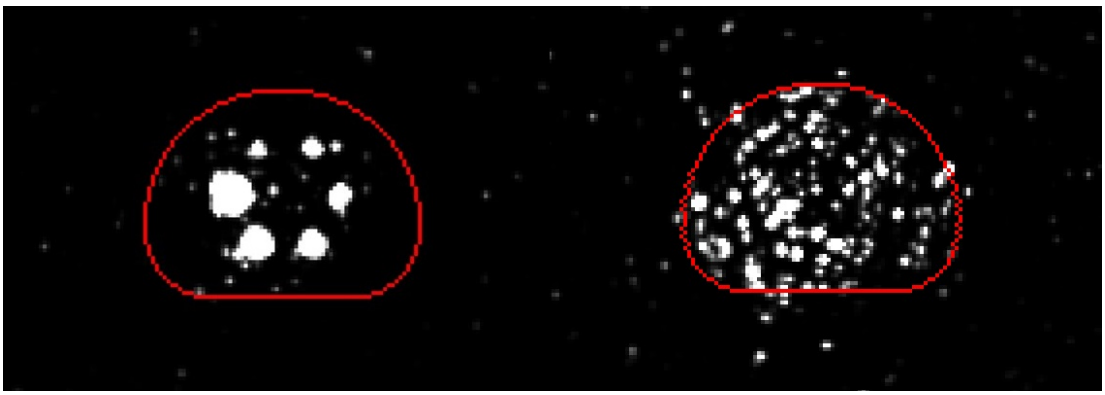


Figure 7: An example of a 2-D representation of the cross section of the phantom that goes through the centres of the spheres. This figure shows two images. The right image is a later scan than the left image and thus corresponds to a lower activity present in the spheres. The comparison of the two images illustrates the image quality degradation over time.

in the spheres and 1.8×10^3 Bq/ml in the background. The scan session consisted of a number of successive scans, each performed at a lower activity than the previous. This was achieved by letting the radioactive nuclei decay over time. The activity concentration in the spheres ranged from 1.8×10^4 Bq/ml to 4.6 Bq/ml. The two reconstruction methods that were used were both ordered subset expectation maximization (OS-EM) 3D reconstruction methods but with different corrections. The details of these reconstruction methods will not be included here since they are not directly relevant for this research project. The image data from single scans were imported in MATLAB for the data analysis.

3.3 Determination of reference values

From the image data, the 2-D slice through the centres of the spheres was presented in an image using the DipImage toolbox for MATLAB. The spheres could be recognized as clusters of white pixels in the shape of discs in this image, while the surrounding areas were mostly black with grey and white spots in seemingly random positions. An example of such an image can be seen in Figure (7). For each sphere, consecutive scans of decreasing activity from the scan session were compared and since the image quality degraded as the activity decreased it was possible to determine the scan in which a sphere could no longer be distinguished from the variation in pixel values in the background. These scans were called the *threshold scans*. Figure (7) shows a comparison of two slices from scans performed at different times, and thus with different levels of activity in the spheres. In the right image, a lower level of activity is present in the spheres than in the left image. The figure clearly shows the degradation of the quality of the images with time. The CNRs of the spheres in their respective threshold scans were calculated and were called the *reference threshold CNRs*. These values represent the CNRs of the spheres at which their 2-D representations (discs) can no longer be distinguished from the background. The Rose criterion states that a lesion is detectable if the difference between the mean pixel value of the lesion and the mean pixel value of the background is larger than 5 times the standard deviation of the background pixel values. This corresponds to a probability, which can be calculated by evaluating the cumulative density function (CDF) of a standard normal distribution at $x = 5\sigma = 5$.

3.4 Quantification of the detection threshold

Since the aim was to find a general method to quantify the detection threshold, several methods were tried, two of which are explained in detail below. A third method which was unsuccessful was one where the reference threshold CNRs were expressed as a factor times the CNR at the peak of the distribution of the background CNRs. This idea was discarded because no clear patterns were observed when comparing different spheres. It must be noted however, that trying to find a working method cost a lot of time, because running the script at this stage of the project was a slow process (up to 10 minutes), so that making adjustments repeatedly, for every sphere and up to five scans (for sufficient data), was very time consuming.

The main method that was tested was aimed to determine a fixed detection threshold that can be applied to any kind of distribution of the signal in the background. In order to make the detection threshold applicable to any kind of distribution, a measure had to be found which did not depend on the shape of the distribution. In this study it is suggested that the area under the distribution of CNRs in the background can be used as such a measure, or in the case of normalized values, the area under the probability density function (PDF) that describes the distribution of CNRs in the background, i.e., it is a measure of probability. A fixed detection threshold in terms of a probability could consequently be used to predict the critical value for the CNR of a region that may be a lesion, by calculating the value corresponding to such an area under the PDF. This critical value in terms of the CNR was expected to differ for different region sizes, activities and distributions of background CNR. However, the threshold probability was expected to be fixed for all distributions, thus making it a widely applicable detection threshold. In comparison to the normal distribution that is assumed for the Rose criterion, a probability was expected that is in the same order of magnitude as one minus the cumulative probability corresponding to a sphere CNR of 5 for the normal distribution.

In order to try to quantify the detection threshold, the aim was to compare the reference threshold CNRs to the distribution of the CNR values of regions of the same size as the sphere in the background. A histogram would show the number of regions in the background that had a certain CNR. If the distribution were to show that a number of regions had a CNR value equal to the reference threshold CNR of the sphere, there would be a probability for a human interpreting the image to mistake a spot in the background for a lesion. The next step in the analysis was to try to quantify this probability based on the reference threshold CNRs determined from the threshold scans. Firstly, a histogram was made of the CNR values in the background. An example of such a histogram is shown in Figure (8). Secondly, a probability density function (PDF) was fitted to the histogram, normalizing the values and giving the probability that a region of size equal to that of the sphere that is being considered had a certain CNR. Literature suggested [12] that gamma and normal distributions would describe noise in PET images reconstructed with filtered-backprojection and row-action maximum likelihood algorithms. Although these are different reconstruction methods than those used for the scan session analyzed in this research project, it was assumed that the gamma distribution could also be applied here. The skew normal distribution adds a third parameter to quantify the skewness, compared to the normal distribution. Therefore, it was argued that a skew normal distribution would be more likely to show good

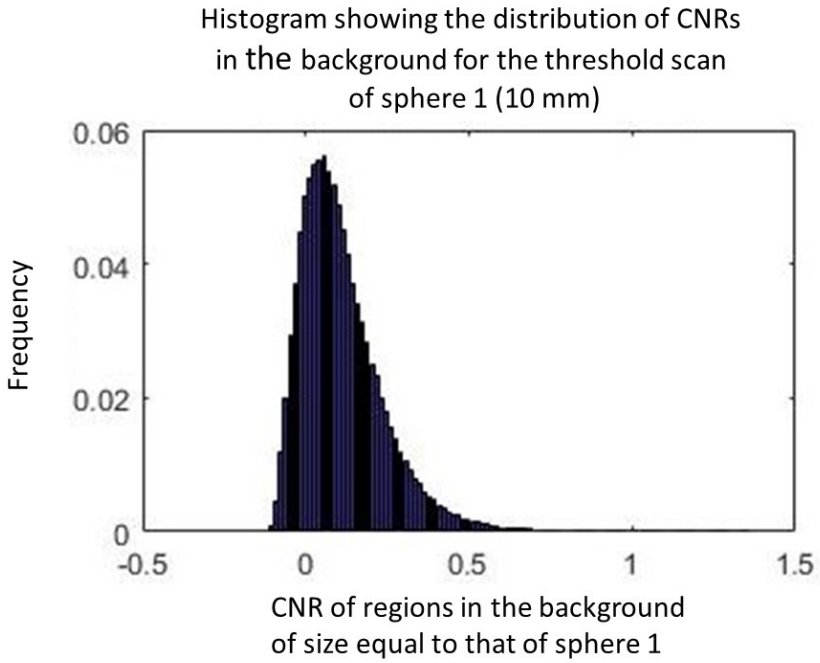


Figure 8: An example of a histogram of the background CNR values for the threshold scan of sphere 1 (diameter = 10 mm), the smallest sphere. This plot was made with the MATLAB programme. The histogram shows the number of regions that are equal in size to the sphere, that have a CNR equal to that given by the value on the horizontal axis for a particular bin.

correspondence with the noise (background CNR) distribution, especially since it could be observed that the distribution was not symmetric around the mean.

3.4.1 The gamma distribution

The first approach was to fit a gamma distribution to the histogram of the CNRs in the background, for each of the spheres in their respective threshold scans. This yielded six probability density functions and, correspondingly, six cumulative density functions. For each sphere the cumulative density function was evaluated at the reference threshold CNR of that sphere. This resulted in a probability. Finally, the probability that the background contained a region with CNR equal to that of the reference threshold CNR of the relevant sphere was given by 1 minus the probability obtained from the CDF. The reason for this was that this probability would correspond to the probability of a false positive, thus giving it a statistical meaning.

3.4.2 The skew normal distribution

For the skew normal distribution, the function for the PDF of a skew normal distribution was fitted in the MATLAB Curve Fitting Tool since it is not a built-in PDF in MATLAB. The function that was fitted was

$$\mathrm{pdf}_{Skewnormal}(x) = \frac{b}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\xi}{\omega}\right)^2} \left(1 + \mathrm{erf}\left(\frac{\alpha(x-\xi)}{\omega\sqrt{2}}\right)\right),$$

where ξ represents the location of the peak on the horizontal axis, ω represents the width of the peak and α represents the skewness. The parameter b was added to scale the function.

Contrary to the gamma distribution, which is a built-in PDF in MATLAB, the probability could not be calculated directly by evaluating the CDF, but had to be calculated by integration of the function that the fit yielded, from $-\infty$ to the reference threshold CNR, and dividing by the integral from $-\infty$ to $+\infty$ for normalization, followed by calculating 1 minus this result. However, this did not seem to work as some of the resulting probabilities were smaller than zero, which is in conflict with the definition of probability. Since very small probabilities were expected and MATLAB could only calculate the integrals numerically, it was argued that the reason was most likely numerical errors. The approach was adjusted in an attempt to work around such numerical errors. The main adjustment was that instead of calculating the probability that the CNR of a region equal to the size of the sphere would be equal to the reference threshold CNR of that sphere, a fixed probability greater than zero and in the order of the expected probability was chosen. The CNR corresponding to this probability was calculated for three scans and all six spheres. The aim was to investigate whether a relation existed between this CNR value and the reference threshold CNR, which would suggest that it might still, in theory, be possible to express the detection threshold in terms of some fixed probability if numerical errors could be corrected for. Firstly, the programme *Mathematica* was used instead of MATLAB, because *Mathematica* does have a built-in PDF for the skew normal distribution (however, applying the same procedure as with the gamma distribution was not succesful because it resulted in all probabilities equal to zero). Secondly, the data set that was used as input for the fit was changed from the histogram values to the unbinned list of background CNRs. The length of this list was of the order of 10^5, so it was argued that a sufficiently large data set would remain if a sample was created by taking every 100th element of the complete data set. This would speed up the programme's search for a fit significantly, since running the script for the complete data set was estimated to take an hour or more, whereas running the script for the sample took approximately 15 minutes. *Mathematica* needed starting values for the fit, and therefore the parameter values obtained from the fit in MATLAB were used for this purpose (although a different data set was fitted, a similar shape was expected). *Mathematica* estimated the parameter values using the "Maximum Likelihood" method, which is the default for the FindDistributionParameters command. Next, the CNR corresponding to a probability of 0.0001 was calculated for three scans and all six spheres. The choice of this value for the probability was based on the consideration of a probability close to what was expected versus one that was sufficiently large so as to be less vulnerable to numerical errors such as those experienced previously. The CNR values corresponding to a probability of 0.0001 were compared to the reference threshold CNRs and an investigation was done to evaluate whether there was a relationship between them.

3.5 Calculation of uncertainties

Uncertainties were calculated for the CNRs that were determined from the probability density functions. Equation (1) gives the relationship between the CNR and the mean signal of the spheres μ_l and the mean signal of the background μ_b. The uncertainties in μ_l

Sphere	Reference threshold CNR
1	1.777 ± 1.391
2	1.160 ± 0.833
3	1.060 ± 1.228
4	0.780 ± 0.777
5	0.337 ± 0.304
6	0.879 ± 0.485

Table 1: Reference threshold CNR values calculated for the threshold scan of each sphere

and μ_b, $\Delta\mu_l$ and $\Delta\mu_b$, respectively, were approximated by the standard deviation of the sample mean, based on the laws of large numbers from statistics [13], i.e. $\Delta\mu_l = \frac{\sigma_l}{\sqrt{N_l}}$ and $\Delta\mu_b = \frac{\sigma_b}{\sqrt{N_b}}$, where σ_i is the standard deviation of the list of signal values and N_i the number of samples/values used to calculate σ_i, for $i = l, b$. If we write Equation (1) as $CNR = f(\mu_l, \mu_b)$ with f a function, then the uncertainty in the CNR, ΔCNR, is given by:

$$\Delta CNR = \sqrt{\left(\frac{\partial f}{\partial \mu_l}\right)^2 (\Delta\mu_l)^2 + \left(\frac{\partial f}{\partial \mu_b}\right)^2 (\Delta\mu_b)^2}.$$

From Equation (1) follows that $|\frac{\partial f}{\partial \mu_l}| = |\frac{\partial f}{\partial \mu_b}| = \frac{1}{\sigma_b}$. One calculation of $\Delta\mu_b$ shows that $\frac{\Delta\mu_l}{\Delta\mu_b} \sim 10^3$ so $\Delta\mu_b$ has a very small contribution to ΔCNR. Therefore, $\Delta\mu_b$ is neglected in the calculation of uncertainties, leaving:

$$\Delta CNR = \left|\frac{\sigma_l}{\sigma_b\sqrt{N_l}}\right|.$$

4 Results

In this section, the results are presented.

4.1 Reference values

Table (1) shows the reference threshold CNRs that were determined visually to serve as a reference for subsequent analyses.

The threshold probability predicted by the Rose criterion is:

$$\begin{aligned} p_{Rose} &\approx 1 - 0.9999997133 \\ &= 2.867 \times 10^{-7} \end{aligned}$$

4.2 Results of the gamma distribution fit

The probabilities that were calculated by fitting a gamma distribution to the histogram of CNRs in the background and evaluating the CDF at the reference threshold CNRs, followed

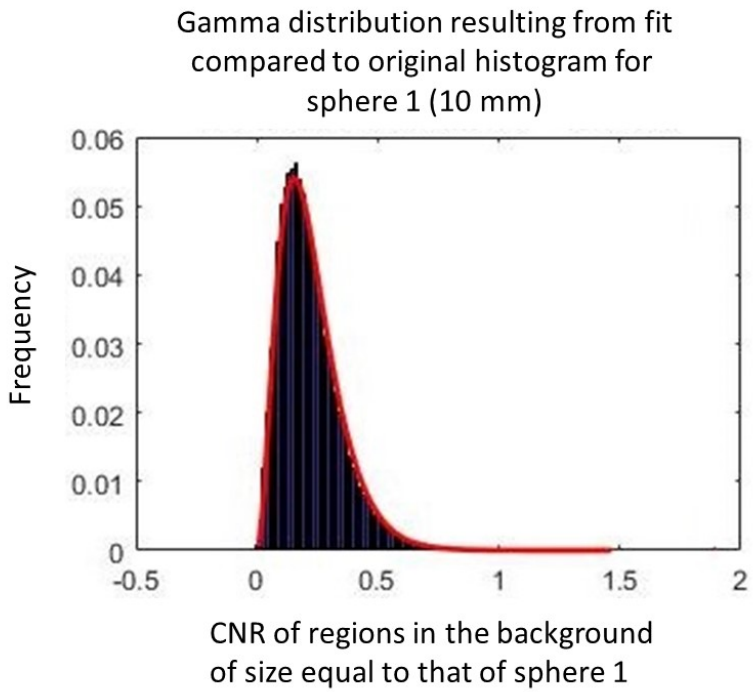


Figure 9: An example of the plot of the probability density function resulting from fitting the data, compared to the histogram of the data. This figure shows the resulting gamma distribution (red) obtained by fitting the histogram (black and blue) for sphere 1 (diameter= 10 mm).

Sphere	**Threshold probability**
1	1.500×10^{-9}
2	2.554×10^{-15}
3	0
4	0
5	0
6	0

Table 2: Probabilities from the gamma fit corresponding to the reference threshold CNRs.

by calculating 1 minus this result, are presented in Table (2). In addition, Figure (9) shows an example of the gamma distribution resulting when fitting to the histogram. Note that from Table (2) it follows that the method that was tried only worked for the two smallest spheres. Also, the values differ from the probability predicted by the Rose criterion by a factor of 10^2 (sphere 1) and 10^8 (sphere 2).

4.3 Results of the skew normal distribution fit

The results of the skew normal distribution fit are shown in Table (3), Table (4) and Figure (10). Table (3) shows the values of 1 minus the CDF of the skew normal distribution that was obtained from the fit and was evaluated at the reference threshold CNR of a sphere. The

Sphere	Threshold probability
1	1.299×10^{-14}
2	4.197×10^{-14}
3	-4.529×10^{-11}
4	-5.995×10^{-15}
5	7.805×10^{-14}
6	-

Table 3: Probabilities from the skew normal fit corresponding to the reference threshold CNRs.

Sphere	CNR corresponding to CDF=0.9999		
	Scan 1	Scan 2	Scan 3
1	0.773 ± 1.391	0.938 ± 1.391	-0.044 ± 1.391
2	0.286 ± 0.833	0.269 ± 0.833	-
3	0.007 ± 1.228	0.020 ± 1.228	0.080 ± 1.228
4	-0.067 ± 0.777	-0.057 ± 0.777	-0.015 ± 0.777
5	-0.099 ± 0.304	-0.084 ± 0.304	-0.042 ± 0.304
6	-0.039 ± 0.485	-0.039 ± 0.485	-0.052 ± 0.485

Table 4: CNR values corresponding to CDF=0.9999 for all spheres and 3 scans and the skew normal distribution.

absolute value of each probability is very small, ranging from the order of 10^{-15} to 10^{-11}. For such small magnitudes, numerical errors have a significant effect. This may explain why some of the probabilities are negative, which is not a valid result. Note that the value for Sphere 6 is missing. This value was not included because the resulting function from the fit was a horizontal line, which did not agree with the input data at all. Table (4) shows the CNRs corresponding to a fixed probability of 0.0001 (equivalent to CDF=0.9999). Note that one value in the table is missing for sphere 2. The reason for this is that the fit did not converge. The values from Table (4) are compared to the reference threshold CNRs from Table (1) in Figure (10) by plotting the CNRs corresponding to a probability p=0.0001 against the reference threshold CNRs. A weighted fit with a first order polynomial was conducted, although it must be noted that other functions had also been possible due to the large error bars. The linear relationship that was found is:

$$\mathrm{CNR}_{\mathrm{p=0.0001}} = (-0.466 \pm 0.183) + (0.558 \pm 0.147)\ \mathrm{CNR}_{\mathrm{ref,threshold}}.$$

This suggests that there is indeed a relationship between the visually determined reference threshold CNRs and the CNRs corresponding to a fixed probability. Hence, it cannot be excluded that there may exist some fixed probability that corresponds to the detection threshold. However, the method applied in this study has not been successful in quantifying that probability.

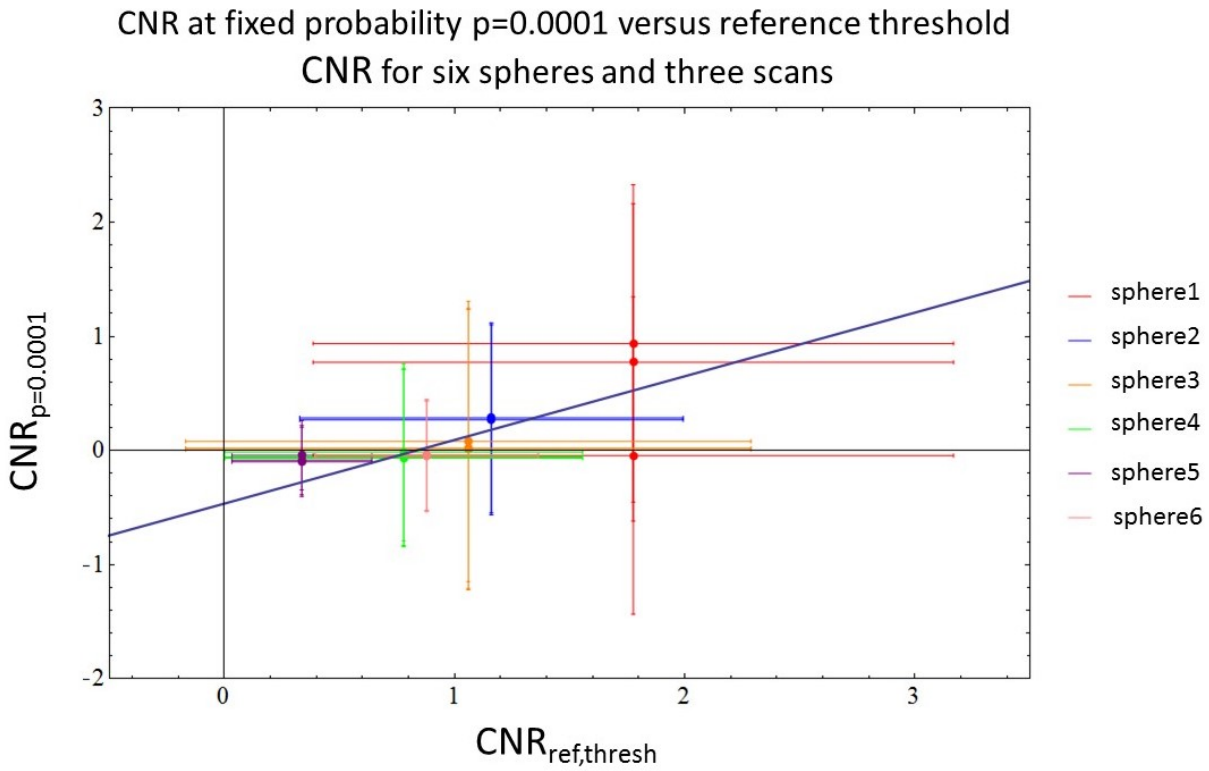


Figure 10: A plot showing the CNRs calculated for a fixed probability p=0.0001, $CNR_{p=0.0001}$, plotted against the reference threshold CNRs, $CNR_{ref,thresh}$, for all six spheres of sizes 10 mm (sphere 1, red), 13 mm (sphere 2, blue), 17 mm (sphere 3, orange), 22 mm (sphere 4, green), 28 mm (sphere 5, purple) and 37 mm (sphere 6, pink). For each sphere, $CNR_{p=0.0001}$ was calculated for three scans, which can be recognized as three y-values (which in some cases overlap) per x-value. A first order polynomial was fitted through the plotted quantities and is represented by a dark blue line.

5 Discussion

The results indicate that the method that has been tested does not quantify the detectability threshold for lesions in PET images satisfactorily, and thus does not validate the hypothesis that a probability could predict the threshold value. A possible reason for this could be numerical errors. When calculating the CNR for a fixed probability of 0.0001 in the hope of finding a constant factor between the reference threshold CNRs and the CNRs for p=0.0001, a linear relationship was found when comparing the resulting values to the reference threshold CNRs of the six spheres. This indicates that although the detection threshold in terms of probability was not quantified in this study, it may still exist. However, building further on the linear relationship that was found to derive some criterion would probably not directly lead to a result that can easily be applied in a clinical setting. Thus, the results are somewhat unclear since they do not verify the hypothesis, nor clearly falsify it.

The main problem with the method that has been tested is probably numerical precision. The expectation was that it would be possible to find a probability that could predict a threshold CNR such that a lesion could be considered detectable if its CNR is larger than the threshold. Under the assumption that the result should roughly agree with the

Rose criterion, a probability of the order 10^{-7} was expected. In principle, MATLAB and Mathematica should be able to calculate numbers to precisions higher than 7 decimals. However, if one considers that there is some uncertainty in the fit parameters and reference threshold CNRs, significantly larger than this precision level in the case of the CNRs, it seems possible that the programmes could not correctly calculate the small values for the probability to high precision. An observation that is related to numerical precision is that since a probability density function is normalized, the sum of the area under its curve tends to 1 as the CNR goes to infinity, but very often, the tail of the distribution of CNRs of the background did not seem to extend to the reference threshold CNR of the relevant sphere on the horizontal axis. This meant that its right tail decayed to zero at a value lower than the reference threshold CNR. Evaluating the CDF at the reference threshold CNR thus resulted in 1. However, if the numerical precision of the input values had been higher, the programmes would count contributions to the area all the way to infinity, which is where the right tail actually tends to zero. Hence, numerical precision may have been the reason that the gamma distribution resulted in threshold probabilities of zero for spheres 3 to 6. The reason why it did give some results for the two smallest spheres could be that their threshold scans were scans where the activity was higher, implying that there were a lot more emissions taking place which probably resulted in a larger data set for the CNR in the background. A larger data set often improves the precision of data analysis.

On the other hand, the reference threshold CNRs may also have been too high. Lower values would more clearly intersect with the tail of the probability density function fitted to the background CNRs. However, this mainly affects the accuracy of the obtained values and should not exclude the possibility that a threshold probability exists. The reference threshold CNRs were determined by first determining the threshold scans. This was done only by me personally, and more accurate results could have been obtained if an observer study had been performed with a group of different observers, including clinicians. Another possibility would have been to use model observers. However, since this research project was mainly aimed at investigating whether a method that accurately quantifies the detection threshold of lesions in PET images exists, such observer studies were intially not considered to be necessary, and would also be too time-consuming for the limited time that was available. However, this could be a point of improvement for further research on this topic.

As mentioned above in the context of numerical precision, the probability density functions that resulted from fitting the distribution of CNRs in the background seemed to decay to zero quite steeply. Although the general shape seemed to agree with the original distribution, it may be that the probability density functions that were fitted, the gamma distribution and skew normal distribution, did not describe the distribution well at the tail. A distribution with a longer extended tail might be more suitable. The results from the gamma distribution and skew normal distribution gave considerably different results, which suggests that the choice of distribution has significant influence. The assumption that the gamma and skew normal distribution would correspond well with the distribution of the CNR in the background may thus have been incorrect. A reason for this may be that a different reconstruction method was used than the reconstruction methods used in the article that suggested the gamma and (skew) normal distributions [12]. However, no other sources in literature were found that had suggestions in the case of the OS-EM reconstruction method. Therefore, there are also no additional sources that validate the

results from the article that was used as a source to decide which probability density functions would be tested and compared in this study. Improvements of the method will thus rely on the developments in research on this topic.

Another example of poor precision is the large uncertainty in the CNRs. This can be observed in Figure (10), where it can be seen that the error bars are very large compared to the values themselves. The conclusions that have been drawn about the existence of a linear relationship are therefore not very reliable.

The skew normal distribution fit could perhaps have given more accurate results if a larger sample of the list of CNRs in the background had been used, or if the entire list had been used. The data showed that when the sample was plotted, the points were a lot sparser than in a plot of the entire list, so some information was lost when taking a sample. It must be noted that using a larger sample was found to be very time-consuming with the programmes and computer used for this study (running the script for the entire data set was estimated to take an hour or longer, compared with running the script for the sample, which took approximately 15 minutes). The time it cost to run the script was also limiting for this study in another way, since it made it very time consuming to repeatedly test and adjust the method. Perhaps there exists another programme that can perform such a fit more accurately or that can process a larger sample in a smaller amount of time.

It must also be noted that although the image that was used to determine the threshold scans was 2-D, the CNRs were calculated in 3-D, as explained in Section 2.5. This may imply that the values calculated cannot be compared to the value predicted by the Rose criterion accurately, because the Rose criterion applies to 2-D images. It was expected that values of the same order of magnitude would be obtained, and this was not the case in this study, as can be seen from the threshold probabilities obtained from the gamma distribution fit.

Finally, another limitation of this study is that the investigation was based on CNR values only. However, whether the human eye can detect a lesion or not depends not only on the mean and standard deviation of pixel values, but also on many other factors such as the size and shape of the lesion and how the noise pixels are distributed. For example, if the noise is evenly distributed, a spot representing a lesion is easier to distinguish from the noise than if there are clusters of light grey or white noise pixels that look similar to lesions, although the mean and standard deviation of the noise can be the same in these situations. Hence, the CNR may not be the best measure to base the detection threshold on.

If further research is to be conducted in the future, care must be taken with numerical precision. The method can also be improved by incorporating model observers. In any case, the first step should probably be to investigate which probability density functions best describe noise in PET images for a wider range of reconstruction methods, in particular the OS-EM reconstruction method. Investigating what the best measure is to describe whether the human eye can detect a lesion is also important, since CNR may not be the best measure.

6 Conclusions

In conclusion, this study has not succeeded in accurately quantifying the detection threshold of lesions in PET images and the method that has been tested in this study does not yet seem adequately accurate to serve as a model to quantify the detection threshold. The

main reasons for the unconvincing results are probably numerical errors, a dubious choice of probability density functions that may not accurately describe the data and the CNR being an incomplete measure for quantifying how well the human eye can detect a lesion. However, since the results from the skew normal fit indicate that a relationship may exist between the reference threshold CNRs and the CNR corresponding to a fixed probability, they do not exclude the possibility that the detection threshold can, in theory, be expressed in terms of a fixed probability and that a method that accurately quantifies the detection threshold in PET images can be developed. Further research will have to be conducted, with the improvements that have been suggested in this study, to gain more certain conclusions. Accurately quantifying the detection threshold remains an important matter to aid clinicians in the diagnosis of cancer.

References

[1] A. Rose. *Vision: Human and Electronic*. Plenum Press, New York, third edition.

[2] S.R. Cherry, J.A. Sorenson, and M.E. Phelps. *Physics in Nuclear Medicine*. Saunders, Philadelphia, third edition, 2003.

[3] J.T. Bushberg, J.A. Seibert, E.M. Leidholdt Jr., and J.M. Boone. *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins, Philadelphia, second edition, 2002.

[4] R. Boellaard, M.J. O'Doherty, W.A. Weber, F. M. Mottaghy, M.N. Lonsdale, S.G. Stroobants, W.J.G. Oyen, J. Kotzerke, O.S. Hoekstra, J. Pruim, P.K. Marsden, K. Tatsch, C.J. Hoekstra, E.P. Visser, B. Arends, F.J. Verzijlbergen, J.M. Zijlstra, E.F.I. Comans, A.A. Lammertsma, A. Paans, A.T. Willemsen, T. Beyer, A. Bockisch, C. Schaefer-Prokop, D. Delbeke, R.P. Baum, A. Chiti, and B.J. Krause. Fdg pet and pet/ct: Eanm procedure guidelines for tumour pet imaging: version 1.0. *European Journal of Nuclear Medicine and Molecular Imaging*, 37:181–200, 2010.

[5] P.A. Tipler and R.A. Llewellyn. *Modern Physics*. W.H. Freeman and Company, New York, sixth edition, 2012.

[6] H. Chen, X. Lei, and Y. Dezhong. An improved ordered subsets expectation maximization positron emission computerized tomography reconstruction. *Computers in Biology and Medicine*, 37:1780–1785, 2007.

[7] C. Beijst. *Imaging Techniques for Guidance of Radionuclide Therapy*. PhD thesis, Utrecht University, 2016.

[8] C. Beijst, J.W. Kist, M. Elschot, M.A. Viergever, O.S. Hoekstra, B. de Keizer, and H.W.A.M. de Jong. Quantitative comparison of 124i pet/ct and 131i spect/ct detectability. *Journal of Nuclear Medicine*, 57(1):103–108, 2016.

[9] M. Oehmigen, S. Ziegler, B. W. Jakoby, J.-C. Georgi, D. H. Paulus, and H. H. Quick. Radiotracer dose reduction in integrated pet/mr: Implications from national electrical manufacturers association phantom studies. *Journal of Nuclear Medicine*, 55(8):1361–1367, 2014.

[10] A.A. Attarwala, F. Molina-Duran, K.-A. Büsing, S.O. Schönberg, D.L. Bailey, K. Willowson, and G. Glatting. Quantitative and qualitative assessment of yttrium-90 pet/ct imaging. *PloS one*, 9(11):e110401, 2014.

[11] K.P. Willowson, M. Tapner, D.L. Bailey, et al. A multicentre comparison of quantitative 90y pet/ct for dosimetric purposes after radioembolization with resin microspheres. *European Journal of Nuclear Medicine and Molecular Imaging*, 42(8):1202–1222, 2015.

[12] A. Teymurazyan, T. Riauka, H.-S. Jans, and D. Robinson. Properties of noise in positron emission tomography images reconstructed with filtered-backprojection and row-action maximum likelihood algorithm. *Journal of Digital Imaging*, 26(3):447–456, 2013.

[13] R. Durrett. *Elementary Probability for Applications*. Cambridge University Press, New York, 2009.

Acknowledgements

I have very much enjoyed to do this project and would like to thank the people who made it possible. Firstly, I would like to thank Andre Mischke for giving me the opportunity to do this project and making the connection with Hugo de Jong. Hugo, thank you for speaking with your colleagues when I asked if there was a project available, and giving me the opportunity to speak with Casper Beijst about a possible project. I would like to thank Casper for his excellent supervision on a weekly basis at UMC. I felt like I could always stop by your office if I had questions. I would also like to thank the rest of the research group that I was involved with at UMC Utrecht. I really enjoyed the lunches and coffee breaks we had together and it was nice to get to know all of you. I learned a great deal about your field of research from the group meetings. Wilco and Leo, I am very happy that you were willing to share your office with me. It was a great place to work!