Methodology and analysis for the recognition of tumour tissue based on FDG-PET

Master Thesis
in Applied Mathematics
by
G.R. Lagerweij

April 18, 2012

Chairman: Prof. Dr. H.J. Zwart
Prof. Dr. A.A. Stoorvogel
Dr. Ir. E.A. van Doorn
Dr. J.A. van Dalen

UNIVERSITEIT
TWENTE.

Medisch Spectrum △Twente

Abstract

For diagnosing the cause of a patient's illness medical images are used more and more. Each medical problem can be visualized with its own specific image. For the detection of tumour tissue the main medical image is a Positron Emission Tomography (PET) image.
A patient with suspicious tissue is injected with a radio-active substance. This active substance will decline and a positron is formed. This element will bump into an electron and two gamma photons appear due to this collision. When two opposite detectors in the scanner detect an arrival of a gamma photon within a certain time interval, the computer receives a signal. The intensity of these signals gives an indication of the density and location of the radio-active substance.
Active organs and tumour tissue attract this radioactive substance since it is combined with glucose, therefore these parts are visible on a PET scan. So by making a PET scan the location and amount of the tumour tissue can be determined.
Not all photons are detected correctly, so noise arises. This noise affects the accuracy of the image. It is important to get an indication of the effect and consequences of the noise.
Besides the noise an extra difficulty of a PET image is the difference between two various PET images, each patient has its own PET image with specific intensity values. The reason is that each patient is different by gender, age or level of metabolism. Another difficulty is the origin of a high intensity pixel, a pixel can have a high intensity for many reasons. For example, a high intensity pixel describes a part of an active organ, but it can also be caused by an infection. Therefore it is very difficult to determine whether or not a pixel is suspicious based on its intensity only.

A literature study on PET images and the effect of noise is performed. A PET image looks like a photo, so perhaps it is possible to use algorithms used in the photo-industry, unfortunately this does not seem possible. There are big differences between pictures and PET images, the main difference is the number of pixels. One slice of a PET image exists of 168×168 pixels, this slice is a vertical cross-cut of your body. Note that a PET scan is three-dimensional, therefore it consists of multiple two-dimensional slices.
A normal photo has at least twice as much pixels as a PET image, therefore a PET image is not accurate enough and that is why it is so important to distinguish tumour tissue from the noise. To illustrate this, one strange pixel can be a tumour of 0.5cm^2.
In a photo it is easy to detect a strange pixel, because of the precision. This strange pixel can be determined based on its neighbors, a white pixel stands out from a group of black pixels. In a PET image it is unknown whether a pixel should be white or black.
Other algorithms for clustering pixels into groups based on their gradation are not useful but still it is important to understand the effects of the noise. Therefore the distribution of the noise is investigated and the data from a PET scan is estimated with a distribution. This results in a difficult distribution which can not be implemented properly, therefore instead a normal distribution is used to perform the simulations.
The methodology of the simulations and the performed steps are described extensively. From the data set of a phantom study a normal distribution is estimated. With this distribution a data set with correlated random values is generated. All pixels with a high intensity, the outliers, are determined for a certain percentage (1% or 5%) and one pixel is increased to a certain value.
The hypothesis from the derivation of the noise model is that there exists a certain correlation between pixels. A pixel with a high intensity due to noise, will have neighbors with a high intensity as well. However a suspicious pixel is high because of itself. Therefore a suspicious pixel has less correlation (similarity) with its neighbors. This knowledge is used to distinguish a strange pixel from the outliers.

An outlier has more correlation with its neighbors and hence the difference in intensity between such a pixel and its neighbors is relatively small. For an increased pixel it is the opposite; the difference in intensity will be relatively large. The simulations support this hypothesis. These differences in intensities are used to determine confidence intervals for both outliers and increased pixels. The next step is to find a (variable) criterion based on these confidence intervals. The best criterion is based on a trade-off between well judged pixels and wrongly judged outliers. Wrongly judged outliers are actual outliers which are identified as suspicious pixels.

The simulations work rather well, still the question is whether it is a realistic implementation of PET images. To investigate this, a validation of the distribution for the noise is performed. The theoretical values for the increased pixel can be derived using a normal distribution. The experimental values converge within 1000 simulations to these theoretical values, this is true for both the expected value and the variance of the distribution of an increased pixel.
The derivation of the theoretical values for the outliers is a lot harder, the distribution of an outlier changes since the data set has changed (only the outliers are evaluated, not all pixels). We have to deal with a conditional probability because of the condition that the intensity of the pixels is above some value. It turns out that the experimental values of the outliers converge a lot slower to their theoretical values (within 2000 simulations), but also these values converge.

The results of the simulation are promising, however an optimal criterion can not be established because it is highly dependent on the situation and preferences of the doctor. On the other hand the derived percentages for correctly and wrongly judged pixels are also promising.
Next the methodology used in the simulations is applied on PET images. These results are less convenient because the necessary confidence intervals can not be established. Therefore the results of applying the methodology on PET images give rise for further research.

The research on noise on PET images is not finished yet, more analysis on patients is necessary. It might also be possible to determine another approach to distinguish a suspicious pixel from a data set, instead of the distance with its neighbors.

Preface

Nothing about my life has been easy,
but nothing is gonna keep me down,
cause I know a lot more today,
than I knew yesterday,
so I am ready to carry on.

It is over, I am finishing my study and I will have to say goodbye to Enschede. It is amazing how fast the years past by.

More than six years ago I moved to Enschede, a big change in my life. Still I remember the introduction, dancing on a table though that was never the plan, for the first time in my life I went to the disco, got nights with less sleep then ever but it was wonderful. I have been a member of the board of study association W.S.G. Abacus. It was an incredible year, I learned a lot, earned some study credits, had long evenings (and nights) and laughed a lot with everybody. After this year, I went on studying and now almost three years later, I am finishing my final project.
In this way my autobiography looks nice, but that is not totally true. I had to work very hard to get to this point. The costs were a lot of tears, sweat, frustrations, blood and conversations with people. Sometimes I though I had some syrup to attract difficult situations. All these "bad" things have resulted in something nice; I grew up. They have made me stronger, more secure and positive. The uncertain and litte girl has changed into a powerful woman.
After almost six years of working I am glad that it is over. I am grateful for all the things I have learned, all the people I have met and all the things I have experienced.

To conclude, I have had the time of my life.

Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description	Page
$\bar{x}$	Mean value	21
s	Sample standard deviation	22
s^2	Sample variance, can be determined by squaring the (sample) standard deviation s	30
$\rho_{x,z}$	Correlation coefficient between the points x and z	27
p_{space}	Length between two pixels, approximately 4.07283 mm	28
v_{limit}	Limit value for a certain percentage outliers	33
d	Difference between the intensity of a point and its neighbor	35
d^2	Squared distance between a point and its neighbor	35
$\bar{d}$	Sum of all distances, for all neighbors, divided by the number of neighbors	34
d_{r1}	Sum of all distances for round 1 neighbors, divided by the number of neighbors	34
d_{r2}	Sum of all distances for round 2 neighbors, divided by the number of neighbors	34
d_o	Average distance for an outlier	34
d_p	Average distance for an increased pixel	34
p_{good}	Percentage well judged suspicious pixel	45
p_{wrong}	Percentage wrongly judged suspicious pixel	45
v_{pixel}	The intensity of the increased pixel	46
CIO_L	The left limit value for the confidence interval for an outlier	45
CIO_R	The right limit for the confidence interval for an outlier	46
CIP_L	The left limit for the confidence interval for an increased pixel	45
CIP_R	The right limit for the confidence interval for an increased pixel	45
k_{in}	The difference between the right limit of the confidence interval for an outlier (CIO_R) and the left limit of the confidence interval for an increased pixel (CIP_L)	46

1 Introduction

The subject of this thesis is a specific type of medical images, this chapter presents the background of these images.

1.1 Positron Emission Tomography

We begin with a short introduction on PET images in general. The abbreviation PET stands for Positron Emission Tomography, a technique that visualizes the distribution of a radioactive substance in a human body, more precisely this substance consists of a positron emitter. A radioactive substance is characterized by its decay after some time, during the decline of the active substance used in PET studies, a positron is formed. This positron moves and collides with an already present electron. Both particles disappear (annihilation) and two gamma photons arise, these gamma photons move freely but in opposite directions.
The patient is positioned in a ring with gamma detectors. As soon as two opposite located detectors detect a trigger (an arrival of a gamma photon) within a certain time interval, a signal is send to the computer. This way, it is possible to determine the location and amount of annihilation.

The detection of two triggers at exactly the same time instant is called a coincidence event [2, 21]. Coincidence events in PET can be divided into four categories: true, scattered, random and multiple. All four coincidences are shown in Figure 1.

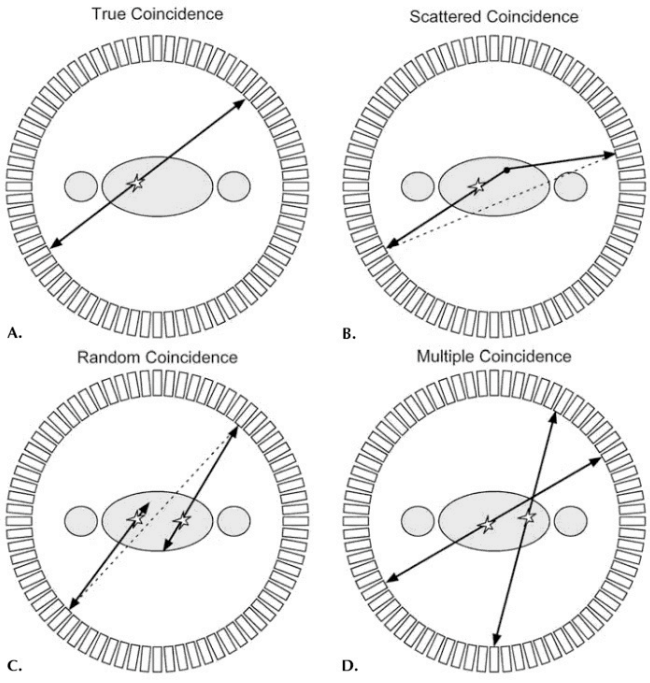

Figure 1: Coincidence events of a PET image [21]

True coincidences occur when both photons from an annihilation event are detected by a pair of detectors within the coincidence time-window. A coincidence time-window is the resolving time of a PET scanner during which two events captured by two opposed located detectors will be considered as a

coincidence event. This coincidence time-window is in the order of nanoseconds.
For a true event photons do not have any interaction with other photons before detection. Furthermore no other events are detected within the specific coincidence time-window.
For a scattered coincidence, see Figure 1 right top, one of the photons has a changed direction during the process. This type of a coincidence adds noise to the actual signal and results in an incorrect positioning of the event.
The third form of coincidences is called random events. These events occur when two photons, not coming from the same annihilation event, are detected within the coincidence time-window of the system. These random coincidences also add noise to the data.
Multiple coincidence events are the last form of coincidence events. These events include that three or more photons are detected simultaneously within the same coincidence time-interval. Again this type of coincidences causes noise on the data.

1.2 FDG-PET

PET images have their main application in determining the presence of tumour tissue and its severity rate. The amount and type of nuclear substance used differs per patient and tissue. For investigating suspicious tissue a radioactive substance combined with an amount of fluorine-atoms is used, this substance (2-fluoro-2-deoxy-D-glucose) is abbreviated with FDG. Human body parts that consume a lot of energy absorb this substance. Examples of organs that consume a lot of energy are the brains, the heart and the liver. Not only active organs will transmit radiation but the tumour tissue will too. This is because tumours are active and fast growing tissues and will therefore attract more glucose than the healty tissues. It is also likely to find FDG in the bladder and kidneys because the radioactive substances will be secreted in these organs.
Due to PET scans it is possible to visualize the location of the radioactive substance FDG and with this to detect the location of possible suspicious tissue.

1.3 Reconstruction

As described in Subsection 1.1 the number of coincidence events gives an indication of the location of the annihilation, only a limited number of the released gamma photons reach the detectors. The remaining photons stay in the human body. Of all the photon couples that reach the ring of detectors only a small part can be used to locate the annihilation (the true coincidence events). It is not true that all photons go in a straight direction, because sometimes a photon arrives with a graduated corner (scatter). Therefore the precise location of this annihilation is difficult to establish.
After the PET scanner has detected all coincidences a reconstruction method is applied [12], after this reconstruction the PET image is ready. Now the distribution of the injected radioactive fluid in the patient's body is known.
The limited number of photons causes another uncertainty and therefore additional noise arises. Since it is not possible to detect all released photons, the effect of noise is large. One possible way of improving a PET image is upgrading the reconstruction method, but the way of reconstruction differs per scanner. So this would not lead to a general solution which we desire.

A natural question that arises is how to deal with the noisy images. It is important to know what the actual signal is and what is not in order to diagnose a patient. Another question is how large the effect of the noise on the PET image is.

2 Problem formulation

Although it is not always obvious, every signal (e.g. electrical) is affected by noise. The effect of noise can be very small, but it can also be large. Roughly speaking, noise is nothing more than a perturbation of the desired signal, an example of noise on a signal is "snow" on a television. High noise levels can change the meaning of a message in (electronic) communication, which may result in making wrong decisions. Therefore it is important to know the amount of noise on a signal.

2.1 Noise on PET images

Unfortunately the effect of noise is large in medical imaging, so it is hard to distinguish the actual signal from the noise. Particularly the noise in FDG-PET images plays an important part. An FDG-PET image is used to detect suspicious tissue like a tumour, this detection depends on the contrast between the activity of the tumour and its background. The tumour detection becomes easier if the noise level is low. However, the noise is affected by many factors; gender, age, metabolism level, injected dose radioactivity, etc.

As described in Chapter 1, it is difficult to determine the exact location of the annihilation. A second difficulty is the unknown amount of annihilations, hence there are already uncertainties before the reconstruction takes place. Thus the reconstruction method will affect the data with noise.
The precise reconstruction method is unknown because of the confidentiality of the manufacturer, therefore it is not that easy to upgrade the reconstruction method. Some researchers have attempted to reproduce the reconstruction method. However for reducing the effect of noise in PET images the details from the following (literature) studies are not sufficient: [19, 24, 26].

The question remains how to distinguish the noise from the actual signal. A disadvantage of PET images is that a large pixel intensity does not mean that this pixel is a tumour, for example a pixel in the liver-area or brain-area can also have a high intensity. A high intensity pixel can even be an infection. For this reason, it turns out that it is rather impossible to create a probability distribution of tumour and of not-tumour tissue. The question there is whether it is possible to say with a certain probability that a pixel is noise or not.

2.2 Goal

The conclusion is that the amount of effect of noise is unknown. Thus it is difficult to distinguish suspicious pixels from a noisy background, the main goal in this thesis is to analyze the noise on PET images. A second purpose is to find a method to distinguish suspicious pixels from harmless outliers in a noisy image.

3 Literature review

Before it is possible to acquire more information from the FDG-PET images, it is necessary to perform some research on the available literature. In this chapter we give an overview of this research and the results on improving FDG-PET images. After this literature study the methodology we worked with is explained in Chapter 4.

It is a known fact that FDG-PET images have a low resolution. From early literature studies it is known that the effect of noise on PET images is large. Therefore the aim is to find a method that will reduce the effect of noise ion FDG-PET images.
The literature study has to be taken widely. There is not much literature of PET images in combination with noise and improving the quality of the images. A question that comes up naturally is, what the difference between a picture of a photo camera and a PET image is. Can an algorithm of a photo camera be used for improving a PET image?
The main difference between a photo and a PET image is the resolution. A photo has a larger resolution, which means more pixels in one image. The format of one pixel of a photo is smaller and therefore a picture of a photo is more precise. Another difference is the prior knowledge of the photo. If you see a photo with a white pixel surrounded by black pixels, you know it is not correct. The reason that you visualize this problem, is the comparison with the neighbors of this deviant pixel. In a PET image this is not possible because based on its neighbors only, a deviant pixel is not necessarily good, i.e. healthy. It is likely that the neighborhood has some influence, but taking the neighborhood as real is not right. The neighborhood can as well be effected by noise. Therefore it is difficult to apply algorithms or methods that are used in the photo-industry.
A second point of interest is the correlation between the input and noise. It is known that the activity concentration is related to the signal to noise ratio (SNR) and the noise equivalent counts (NEC) [15]. This should be observed by determining the right algorithm. The situation is easier when the noise of a picture is not statistically dependent on the original image. In that case it is possible to recover the actual image from a noisy image [22], but earlier research shows that the noise and the original image are dependent.

A further literature study shows another way for detecting and clustering (deviant) points [13, 16, 17]. For example, the algorithm Fuzzy C-Means (FCM) [14] clusters all points into groups with a certain degree. The number of groups is defined by the user. The neighborhood for each investigated pixel is taken into account.
The objective function for partitioning a set $\{x_k\}_{k=1}^N$ into c clusters is given by [11, 18, 28]:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^p ||x_k - v_i||^2 \tag{1}$$

with $\{v_i\}_{i=1}^c$ the prototypes of the clusters, p as weighting exponent on each fuzzy membership and $\{u_{ik}\}_{k=1}^N = U$ the partition matrix. This partition matrix is of the form

$$U = \left\{ u_{ik} \in [0,1] \,\middle|\, \sum_{i=1}^{c} u_{ik} = 1 \quad \forall k \quad \text{and} \quad 0 < \sum_{k=1}^{N} u_{ik} < N \quad \forall i \right\}. \tag{2}$$

Some extensions of the basic Fuzzy C-Means (FCM) algorithm are established in order to adapt the algorithm to the desired goal. All these extensions have the basic objective function, with only a few

changes. The change can be an addition in the form of a feature direction for the pixel intensity or a distance attraction for the spatial position of the neighbors which depends on the structure of the neighborhood. Both additions affect the attraction between the neighborhood and the investigated pixels. Taking a kernel operator instead of the absolute value of the difference [28] can also be a modification. One of the Fuzzy C-Mean [11] methods is programmed in Matlab by a former student of the University of Twente. The BCFCM (Bias Corrected Fuzzy C-Means) algorithm is tested on MRI data and the results are good. The code is likewise applied on PET data, but the results are not hopeful. After some research and a meeting with a professor of Electrical Engineering (Prof. Dr. F. van der Heijden), it is clear that there is just one possibility of using an FCM method. The PET data must satisfy the constraint that the two (or more) probability distributions have a small relatively overlap. If the overlap is large, it is hard to distinguish pixels from each other.
In our case, clustering deviant pixels from normal pixels, the goal is to find a criterion for whether a point is suspicious or not. In most cases a suspicious pixel is tumour tissue. Therefore we have to deal with the probability distributions of normal tissue and tumour tissue. A pixel with a high intensity is not automatically suspicious. For example, pixels with high intensity in the brains have another interpretation than high intensity pixels in the liver. A pixel with a high intensity in the brain can be a tumour, but in the liver it is not. The reason could be for example the difference in metabolism. The intensity gives an impression of the metabolism level in tissue. Brain tissue has a higher level of metabolism than liver tissue.
We can conclude that it is very difficult to realize a probability distribution function of tumour tissue and one of not-tumour tissue that do not overlap.

To summarize, it is not possible to use a FCM method for clustering PET data into groups, because of the overlap in their probability density distributions. Another reason for not using a clustering method is the problem with the neighbors. As stated before it is not correct to eliminate a deviant pixel because of its neighborhood.

So, it seems that changing our approach is required. It is not likely that the solution for the detection of deviant pixels on PET image can be found in the available literatue.
It is true that the neighborhood has influence on a pixel, but it is also true that a point can be high by itself. An important question is what precisely is the influence of the neighborhood in noisy images?

4 Derivation of the noise model

Since some parts of the body absorb more nuclear substance than others, the resulting PET image is heterogeneous. This means that the image exists of roughness, which are causes by the presence of organs, the level of metabolism and the type of patient. As described in Chapter 3 not much literature is known on PET images in combination with noise. It is difficult to determine the effect of noise because of the rugged character of the PET images.
First we will focus on homogeneous images which have a constant input. A homogeneous PET image can be made by filling a phantom with a radioactive fluid and scanning this phantom. A phantom is a body without contents, e.g. a sphere, a bottle or a bucket. In our case this is a cylinder filled with water and the radioactive substance FDG. These scans are normally used for the reliability of the PET camera.
Literature shows that a homogeneous PET image has approximately a Poisson or Gaussian distribution [25, 27]. Does this estimated distribution describes the noise correctly? Can this estimated distribution be used for detecting deviant pixels? These question will be answered in the next subsection.
In the next subsection a distribution of the noise is estimated. A question which we want to answer is; does it agree with the results found in [25, 27]?

4.1 Distribution

Based on the phantom-studies it is possible to determine the distribution of the noise. In the ideal situation the determined signal of the PET scan equals the injected radioactive substance. The difference between these two signals is the noise.
If the distribution of the noise is known, it is possible to neglect (a large part of) the effect of the noise on the signal. This cause the actual signal to become visual and therefore the determination of suspicious pixels is possible.
There are several phantom-studies available for determining the distribution of the noise. These studies also include different input activities, which helps to determine the precise effect of the activity on the distribution.

$$\overline{x} \;:=\; \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

Unfortunately, these studies show some curious phenomena. For instance, the difference between the mean, see (3), of the picture and the injected activity is large. For reliability reasons, the difference between these two numbers must be smaller than 10% of the input value. One reason for this difference might be caused by the rescale slope. This slope transforms the pixel values of the image into values which are meaningful for the application. This is required for the image to present the right intensity values measured in the human body.
After some more studies, we have decided to take only a recent phantom study and not to use the other studies. The difference between the mean of this study and its input value is less than 5%, which is acceptable. There is only one negative consequence of using only this (new) data set; now it is not possible to take a look at the effect of different radio activities as input.

A normal PET study of a patient has a data matrix of $168 \times 168 \times 374$. The actual data set of the phantom consists of approximately 74 slices of 168×168 pixels each. From the original data a

matrix of $25 \times 25 \times 55$ pixels is taken because the borders of the PET scan are not suitable. Therefore the innermost part of the PET scan with a size of 25×25 pixels is taken. For all calculations, only one slice is chosen which results in a data matrix of 625 pixels.
First a histogram is plotted for this particular slice. This histogram gives an indication of the probability density function. Based on the histogram our hypothesis is that the data set is distributed normally with a certain mean and standard deviation. The mean of the noise will be zero, because the average intensity of the image has to be approximately the input, i.e. the injected radio-activity. And since the input is constant, the average of the noise becomes zero. The common estimation for the standard deviation is given by (4)

$$s := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\overline{x} - x_i)^2}, \tag{4}$$

with $\bar{x}$ the mean value. We remark that the standard deviation is different for each study, therefore for each patient. Further details about the difference for the standard deviation are explained later.

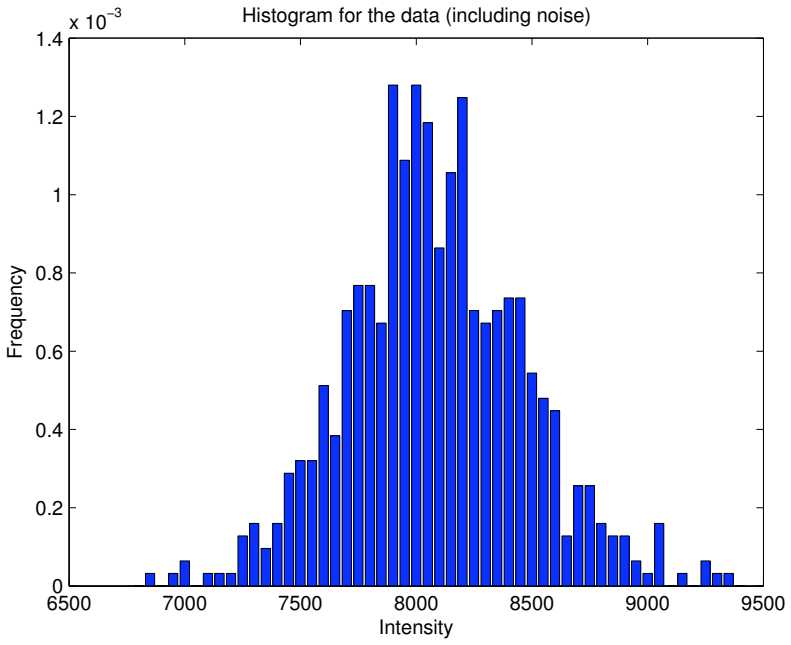

Figure 2: Histogram of the phantom data including the noise

The histogram alone, see Figure 2, is not sufficient to determine the distribution. There are several tests for fitting a distribution to a dataset, see Matlab [1] for some demos.
The null-hypothesis is that the noise is distributed normally. The alternative hypothesis H_1 is that the distribution is arbitrary. The used tests, Kolmogorov-Smirov and Chi-Squared, reject the null-hypothesis several times. The corresponding p-values are $p_{KS} = 0$ and $p_{\chi^2} = 4.3926 \cdot 10^{-20}$.
The only conclusion that can be taken is that the noise is *not* distributed normally.

Since the number of outliers is large, the normal distribution does not fit exactly with the data, see

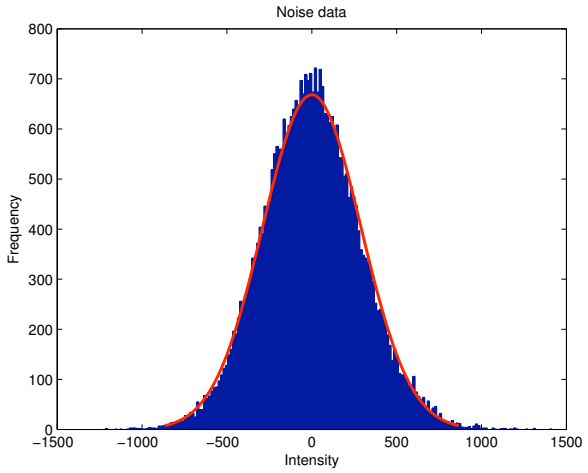

Figure 3: Histogram fit of the noise

Figure 3. Still these outliers need to be judged whether they are innocent or suspicious.
In Figure 3 the histogram of the noise together with an estimated normal density function is drawn. This is one of the many applications of Matlab for fitting a distribution through data. Another application creates a normal probability plot of the data. For our noise data this plot is shown in Figure 4. This figure shows that the outliers are not fitted by a normal distribution.

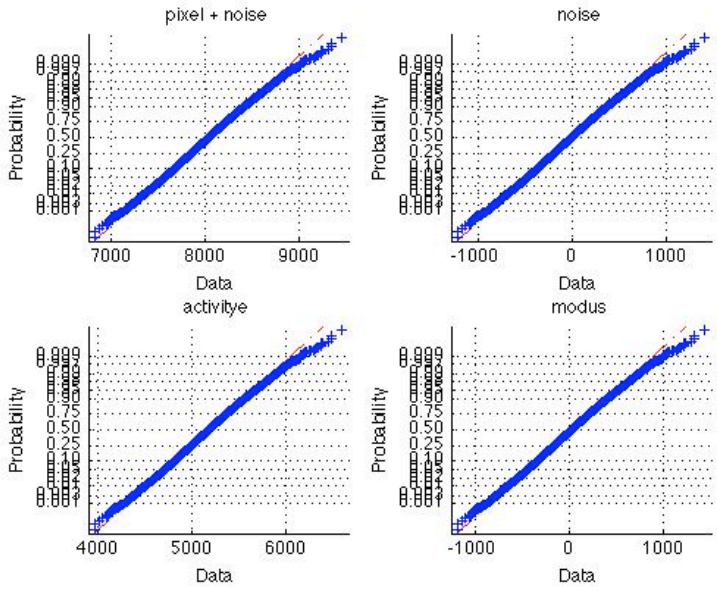

Figure 4: Plot of normal probability

A last attempt to fit a normal distribution to our data is by making a box plot [23] and to compare the results with earlier outcomes. This diagram is an effective method of presenting and summarizing the data. A box plot gives an indication of the centering of the data (median), the spread of the data, the presence of outliers and the symmetry or asymmetry of the data. Especially, the latter is a problem.

The box plot shows that de data is not distributed normally, since for a normal distribution it is allowed that 0.7% of the data is outlier whereas the PET data has 1.6% outliers.

The skewness and kurtosis of the distribution are also calculated. These numbers are a measure of skewness of the distribution and their formulas can be found in Equation (5) and (6).

$$skew := \frac{\sum_{i=1}^{N}(x_i - \overline{x})^3}{(N-1)s^3} \tag{5}$$

$$kurtosis := \frac{\sum_{i=1}^{N}(x_i - \overline{x})^4}{(N-1)s^4} - 3 \tag{6}$$

The skewness says something about the number of outliers. It is not odd that the noise shows skewness because of the large number of large outliers. The skewness is that large that the noise can not be distributed normally and hence the noise could have a skew normal distribution, see Figure 5. Investigating this possibility will cost a lot of time because of the complexity of determining its parameters. Secondly there is not much known about a skew normal distribution, it is hard to say some about the distribution and its behavior. Therefore the skew normal distribution is not chosen as the distribution for the noise.

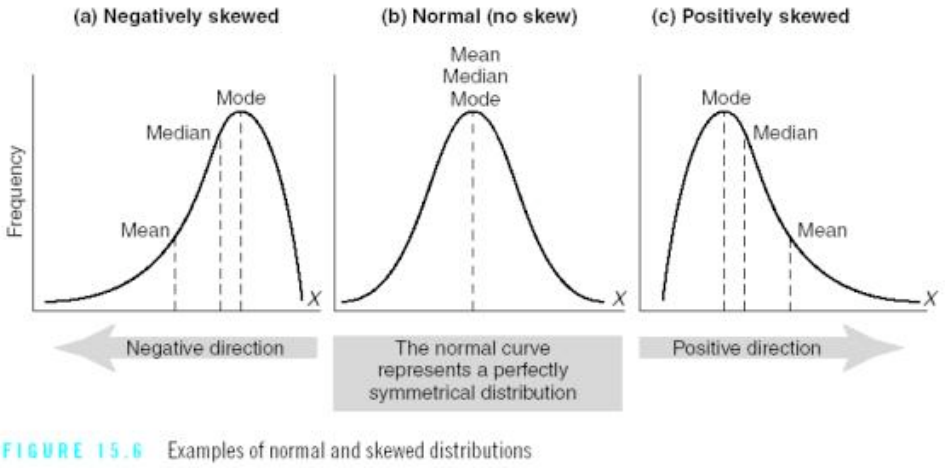

FIGURE 15.6 Examples of normal and skewed distributions

Figure 5: A plot of a (skew) normal distribution [9]

All together we can conclude that the data is not distributed normally.

Although the data is not distributed normally, it has strong similarities with a normal distribution, see Figure 3. In probability theory and combinatorics special polynomials are known, these polynomials are named after Charles Hermite [8].
The probabilists' Hermite polynomials are of the following form:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{\frac{-x^2}{2}}. \tag{7}$$

These polynomials describe the data perhaps better because of the possibility of using multiple order

polynomials. Figures 3 and 4 show that the number of outliers is the main reason why a normal distribution does not fit. To get a feeling for these polynomials, in Figure 6, the Hermite polynomials of degree zero, one, two and three are plotted.

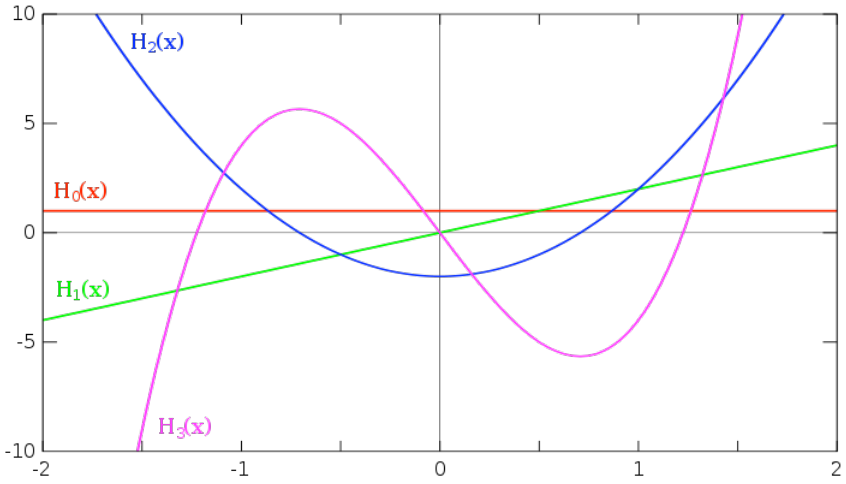

Figure 6: A plot of four hermite polynomials [7]

The aimed Hermite polynomial must be a probability density function, which means that the area under the function must equal one. Calculating the area under a probability density function means taking the integral from $-\infty$ to ∞, i.e. $\int_{-\infty}^{\infty} f(x)dx = \|f\|_{L^1}$. This is also known as the L^1-norm and this case the L^1-norm must be one.
For measuring the similarity between the Hermite polynomial and the probability density function of the data, the L^1-norm of the difference between the Hermite polynomial and the plot of the data is taken. Taking an integral is a first order problem. Therefore a second order norm (L^2-norm), the euclidean norm, is not necessary.

So for determining the right Hermite polynomial, the area under the function has to be calculated and it must equal one. To have this, the possible Hermite polynomial is normalized. This is done by calculation the area A under the Hermite polynomial. The Hermite polynomial is divided by this calculated area A such that the area under the function is one.
After the normalization, the L^1-norm of the difference between the two functions (probability density and Hermite) can be calculated. Just one small change in the Hermite polynomial leads to new calculations of the area under the function and several steps before this L^1-norm can be calculated.

From Figure 6 can be seen that a second order Hermite polynomial is the best option. Figure 7 shows the plot for the data and a second order polynomial.
As can be seen the second order Hermite polynomials fits the data the best, but still the positive outliers are not fitted precisely by the second order Hermite polynomial and these outliers are the important points. It is possible to expand the second order Hermite polynomial, with the result that the peak is not fitted exactly anymore. So it turns out that a Hermite polynomial is not the best fit of the data.
A solution for the outlier-problem is add some peaks at the location of the outliers to a second order Her-

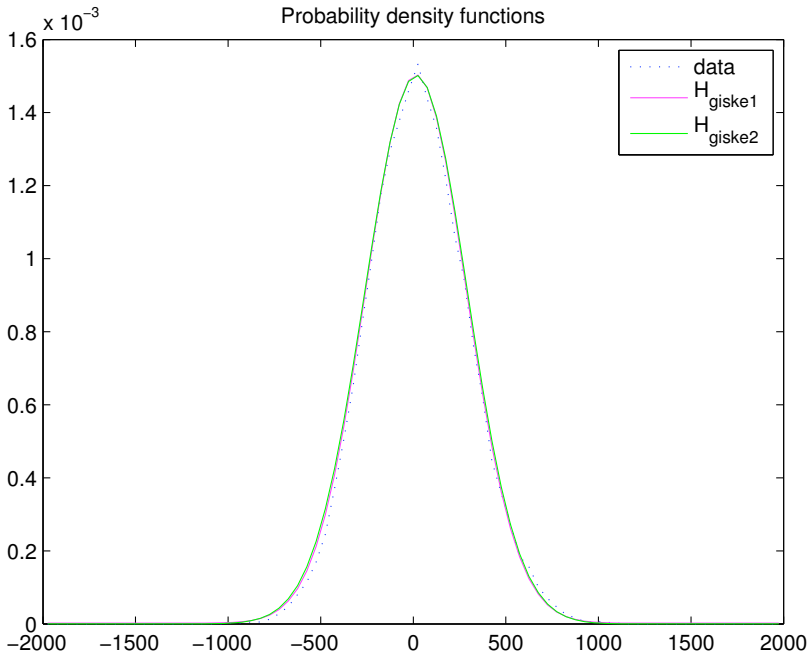

Figure 7: A plot of two determined Hermite polynomials and the data

mite polynomial. Only the positive outliers are important, so only one second order polynomial is added.

Figure 8 shows two derived density functions and the probability density function of the data set. Equation (8) and (9) give the formulas for two estimated probability density functions. These functions are a sum of two Hermite polynomials and are found by trial-and-error.

$$H_{giske1}(w) = \frac{1}{1.026197606} \cdot \left[0.001502\mathrm{e}^{\frac{-(w-15)^2}{385^2}} + 2 \cdot 10^{-6}\mathrm{e}^{\frac{-(w-750)^2}{350^2}}\right] \quad (8)$$

$$H_{giske2}(w) = \frac{1}{1.035548931} \cdot \left[0.001502\mathrm{e}^{\frac{-(w-15)^2}{375^2}} + 10^{-10}(20+9w^2) \cdot \mathrm{e}^{\frac{-w^2}{360^2}}\right] \quad (9)$$

The norms of the difference for both functions with the data are given by (10) and (11).

$$\|data - H_{giske1}\|_1 = 0.0351 \quad (10)$$

$$\|data - H_{giske2}\|_1 = 0.0366 \quad (11)$$

The best plot, the function with the smallest norm, is the density function H_{giske1}. This function estimates the probability density function of the data the best.

4.2 Correlation

The behavior of the noise is described rather well by (9). Figure 9 shows an intensity plot for the data. Here can be seen that the pixels are correlated with each other. In order to prove this presumption first

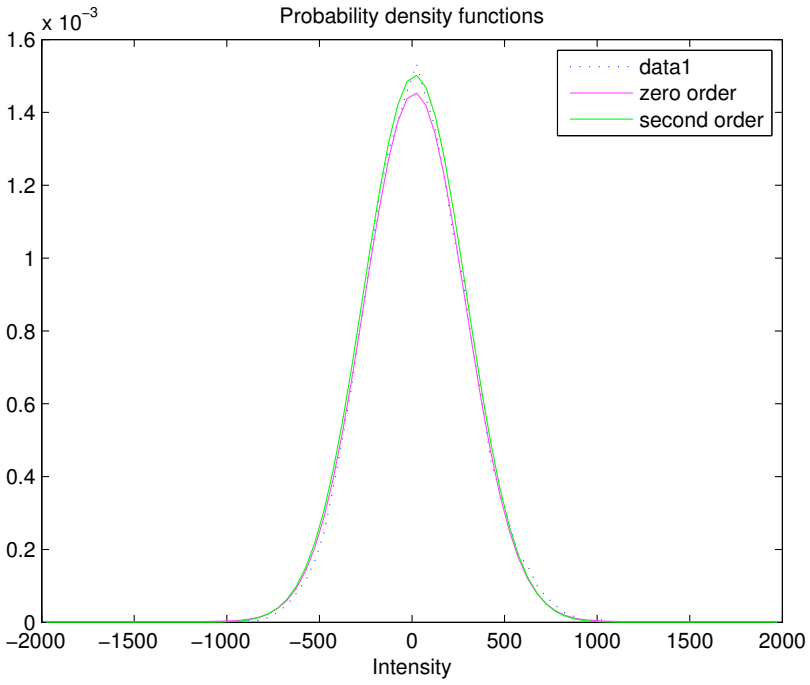

Figure 8: A plot of two determined Hermite polynomials

the correlation coefficient are calculated. The correlation coefficient between x and z is defined as

$$\rho_{x,z} := \frac{\mathbb{E}[(x-\mathbb{E}(x))(z-\mathbb{E}(z))]}{\sigma_x\sigma_z} \tag{12}$$

$$= \frac{\mathbb{E}[(x-\bar{x})(z-\bar{z})]}{\sigma_x\sigma_z} \tag{13}$$

$$\simeq \frac{\sum_{i=1}^{N}(x_i-\bar{x})(z-\bar{z})}{(N-1)s_x s_z} \tag{14}$$

$$\simeq \frac{\sum_{i=1}^{N}(x_i-\bar{x})(z-\bar{x})}{(N-1)s_x^2}. \tag{15}$$

Equation (15) is the one programmed in Matlab. We calculate the correlation coefficient between two pixels in one data set. Therefore the means for x and z are the same. For pixel z we take a neighbor of pixel x and this is repeated for the eight (direct) neighors. The location of the eight neighbors are given in Table 1, where x has the general location (g, h). The edges of the image have less neighbors therefore they are not taken along.

The correlation coefficient for the neighbor one slice up or down is calculated similar. This coefficient is the correlation for pixel x with the same pixel one slice up or down (neighbors s_{t+1} and s_{t-1}, respectively). Table 2 shows the calculated correlations. As can be seen the correlation between the right and left, and the upper and lower neighbor are the same. This is not surprising because if a pixel is the right neighbor of a certain pixel, then the latter pixel is the left neighbor of the first pixel.

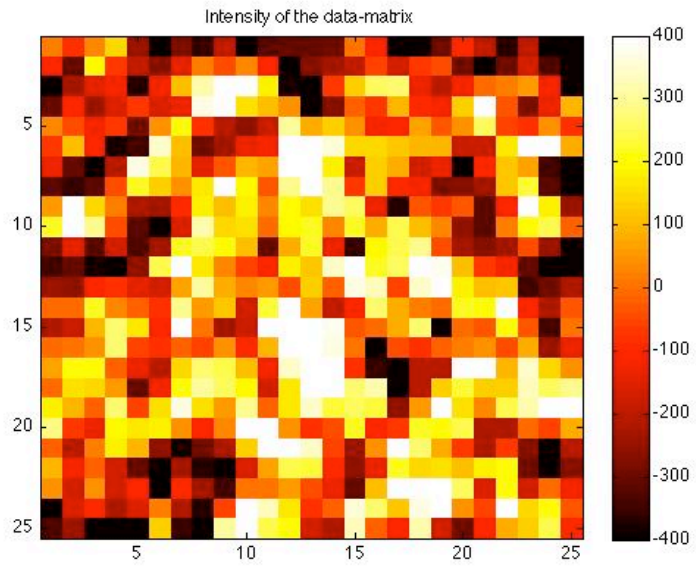

Figure 9: Plot of the matrix of the phantom study

neighbor	location	direction
z_1	(g-1,h)	north
z_2	(g-1,h+1)	north-east
z_3	(g,h+1)	east
z_4	(g+1,h+1)	south-east
z_5	(g+1,h)	south
z_6	(g+1,h-1)	south-west
z_7	(g,h-1)	west
z_8	(g-1,h-1)	north-west

Table 1: Correlation for eight neighbors

The reason for the smaller value of the diagonal neighbor is the distance between the pixels. The pixel-spacing (p_{space}) is $4,07283$mm, which results is a distance between a pixel and his diagonal neighbor of $\sqrt{2} \cdot p_{space}$. Therefore the correlation coefficient for the diagonal pixel is smaller. The distance between two slices, the slice location, is 3 mm. This explains the larger coefficient value for the upper (and lower) slice-correlation coefficient.

As can be seen in the plot of the intensity matrix of the phantom data (Figure 9) the original data matrix has correlated noise. For instance, pixels with a high intensity are clustered. From the correlation coefficients it can similarly be concluded that the data is correlated.

4.3 Covariance matrix

With the calculated correlation coefficients it is possible to determine the corresponding covariance matrix. Correlation coefficients are relevant but they do not tell everything about correlation in a data

neighbor	correlation
z_1	0,4134
z_2	0,2624
z_3	0,4443
z_4	0,2630
z_5	0,4135
z_6	0,2624
z_7	0,4443
z_8	0,2630
s_{t-1}	0,5556
s_{t+1}	0,5596

Table 2: Correlation for ten neighbors

z_8	z_1	z_2
z_7	**x**	z_3
z_6	z_5	z_4

Table 3: First round of neighbors

z_{23}	z_{24}	z_9	z_{10}	z_{11}
z_{22}				z_{12}
z_{21}		**x**		z_{13}
z_{20}				z_{14}
z_{19}	z_{18}	z_{17}	z_{16}	z_{15}

Table 4: Second round of neighbors

set. From a covariance matrix it is possible to examine which pixels are correlated, in combination with the knowledge of the position of the pixels. The equation for the covariance is

$$\begin{aligned} \text{cov}(x,z) &:= \mathbb{E}[(x-\mathbb{E}(x))(z-\mathbb{E}(z))] && (16)\\ &= \mathbb{E}[(x-\mu_x)(z-\mu_z)] && (17)\\ &= \mathbb{E}(xz)-\mathbb{E}(x)\mathbb{E}(z) && (18)\\ &= \rho_{x,z}\cdot\sigma_x^2 && (19) \end{aligned}$$

Normally the covariance between two vectors or inside a vector is calculated. In this situation the covariance of a matrix is necessary. Therefore the size of the covariance matrix becomes much larger. For each point of the data set a covariance vector must be derived. All these covariance vectors will be combined into one large covariance matrix R.
Table 2 show that there is a correlation between pixel x and its vertical neighbor z. For instance, if pixel x has position $(1,1)$ in the data matrix then there exists a correlation with pixel z in the matrix with position $(26,1)$ in the matrix. In the first covariance vector of point x with $(1,1)$ this means a value on position 1 and 26. The data set in our situation has the size 25×25, which means 625 pixels and that 625 covariance vectors have to be determined. These vectors are also correlated with each other.
To simplify the filling of the covariance matrix R, we choose to use another approach. Covariance matrix R is split up in blocks. These blocks describe a covariance matrix for each row with a certain row.
The matrices on the diagonal present the covariances inside one row. The matrices on the sub-diagonal

describes the covariance between two rows.

$$R = \sigma^2 \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,m} \\ R_{2,1} & R_{2,2} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ R_{n,1} & \cdots & \cdots & R_{n,m} \end{bmatrix}.$$

The variable n indicates the row and m indicates the column. For a 4×4 matrix, the covariance matrix is of the form

$$R_{1,1} = \sigma^2 \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{2,1} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{3,1} & \rho_{3,2} & 1 & \rho_{3,4} \\ \rho_{4,1} & \rho_{4,2} & \rho_{4,3} & 1 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho_{1,2} & 0 & 0 \\ \rho_{1,2} & 1 & \rho_{2,3} & 0 \\ 0 & \rho_{2,3} & 1 & \rho_{3,4} \\ 0 & 0 & \rho_{3,4} & 1 \end{bmatrix} = R_{1,1}^T.$$

The first observation is that the covariance of any pixel x with itself is the estimated variance (σ^2). Secondly the correlation between pixels 1 and 2 ($\rho_{1,2}$) is the same as the correlation coefficient $\rho_{2,1}$, see Table 2. So this explains the first matrix operation.
The correlation between pixel x and its n^{th} neighbors, for $n > 1$, are calculated. These values become smaller as n gets larger. Therefore we assume that that there is no correlation between a pixel and its n^{th} neighbor, for $n > 1$. The main diagonal of the big covariance matrix R has the covariance matrices $R_{1,1} \cdots R_{n,n}$, with n as the row indicator.
The upper and lower diagonal of matrix R represent the covariance matrices between rows. As shown in the correlation-table, Table 2, there is correlation between the pixels on row t and $t+1$. This matrix R_s of the same size (4×4) looks like,

$$R_s = \sigma^2 \begin{bmatrix} \rho_s & 0 & 0 & 0 \\ 0 & \rho_s & 0 & 0 \\ 0 & 0 & \rho_s & 0 \\ 0 & 0 & 0 & \rho_s \end{bmatrix} = R_s^T.$$

The resulting covariance matrix R is of the form

$$R = \sigma^2 \begin{bmatrix} R_{1,1} & R_s & 0 & 0 \\ R_s & R_{2,2} & R_s & 0 \\ 0 & R_s & R_{3,3} & R_s \\ 0 & 0 & R_s & R_{4,4} \end{bmatrix} = R^T.$$

The size of this matrix R is 16×16. If the concept data matrix has a size of $n \times m$, the covariance matrix becomes $(n \cdot m) \times (n \cdot m)$. The main diagonal matrices are of size $m \times m$ and that for n rows.

The data set of the PET data has a size of 25×25 so the covariance matrix is 625×625. A covariance matrix possesses some properties, among others it is symmetric and positive-semidefinite. The first property ($M = M^T$) is easy to check since the sub-covariance matrix (R_s) of the upper and lower diagonal are the same and symmetric. The covariance on the main diagonal is symmetric because of equal correlation coefficients for the horizontal neighbors. For this reason the covariance matrix R is symmetric.

The second property (positive semi-definiteness) is more difficult to check but important. A matrix $M \in \mathbb{C}^{n\times n}$ is positive semi-definite if $x^*Mx \geq 0$ for all $x \in \mathbb{C}^n$, or for all $x \in \mathbb{R}^n$ if M is a real matrix. Another way of checking positive semi-definiteness is showing that all the sub-determinants of the matrix are nonnegative. All 625 sub-determinants are calculated and the result is that the matrix is not positive semi-definite while a covariance matrix is positive semi-definite by definition. To conclude, our choice of covariance matrix is not correct and needs to be adjusted.

There are several explanations for the incorrectness of the proposed covariance matrix R. One possibility is that the matrix is filled incorrectly but this seems not be the case. The first solution we tried was to add more correlation coefficients because the covariance between pixel x and its direct horizontal neighbor is approximately 34506 ($\rho_{1,2} \cdot \sigma^2$), which is large.
The covariance for pixel x and its second horizontal neighbor z_{13} is said to be zero, because the calculated correlation coefficient is small. However because of the large variance the covariance ($\rho \cdot \sigma^2$) is too large to neglect, so the covariance of the second (horizontal) neighbors are added. These neighbors are pixels z_{13} and z_{21} respectively, see Table 4.

The other correlations (and covariances) are observed and they are too large to neglect as well. Thus the covariance of the diagonal neighbor (pixel z_2 or z_4) and the second vertical neighbors (pixel z_9 and z_{17}) are added. The result of these operations is that the covariance matrix is finally positive semi-definite. The final covariance matrix R, after all additions, is of the form

$$R = \sigma^2 \begin{bmatrix} R_{1,1} & R_s & R_{s2} & 0 \\ R_s & R_{2,2} & R_s & R_{s2} \\ R_{s2} & R_s & R_{3,3} & R_s \\ 0 & R_{s2} & R_s & R_{4,4} \end{bmatrix}.$$

The sub covariance matrices are of the following form

$$R_{1,1} = \sigma^2 \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & 0 \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ 0 & \rho_{2,4} & \rho_{3,4} & 1 \end{bmatrix}, R_s = \sigma^2 \begin{bmatrix} \rho_s & \rho_d & 0 & 0 \\ \rho_d & \rho_s & \rho_d & 0 \\ 0 & \rho_d & \rho_s & \rho_d \\ 0 & 0 & \rho_d & \rho_s \end{bmatrix}, R_{s2} = \sigma^2 \begin{bmatrix} \rho_{s2} & 0 & 0 & 0 \\ 0 & \rho_{s2} & 0 & 0 \\ 0 & 0 & \rho_{s2} & 0 \\ 0 & 0 & 0 & \rho_{s2} \end{bmatrix},$$

with ρ_s the degree of correlation between a pixel and his direct (horizontal or vertical) neighbor (pixel z_1 or z_3). The correlation coefficient between a pixel and his second direct neighbor is given by ρ_{s2} ($= \text{cov}(x, z_9) = \text{cov}(x, z_{17})$). The variable ρ_d gives the correlation coefficient between a pixel and a diagonal (upper or lower) neighbor (pixel z_2 or z_4).

4.4 Random values from the distribution

Now we have more information about the noise and so with the determined distribution of the noise it is possible to perform experiments. These experiments have the purpose to simulate the reality and hopefully give an idea how to distinguish suspicious pixels from normal pixels.
For these experiments random data sets are needed because it is statistically not possible to use pixels from the original data. Because of the difficulty of using the Hermite polynomial, the decision is made to use a vector valued normal distribution with correlation. After these simulations it is hopefully possible

to use a sum of two Hermite polynomials to appoint deviant tissue. Moreover the two added Hermite polynomials have to be estimated and adapted for every patient.

The covariance matrix R is required for generating random values of a correlated normal distribution. In mathematical terms such a distribution is called a multivariate normal distribution [3, 10, 20]. This is the generalisation of a one-dimensional (univariate) normal distribution to higher dimensions. Matlab can generate random values from a multivariate normal distribution. The determined covariance matrix is an input, together with a zero vector $\bar{x}$, for generating a correlated sequence. After the simulations it is (hopefully) possible to have a criterion for excluding noise from pixels with some deviation. The latter is probably suspicious tissue.

The covariance matrix has the size 625×625 and the zero vector $\bar{x}$ has 625 elements. The resulting multivariate vector has 625 elements and can be reshaped to a 25×25 simulation matrix with correlated elements. Figure 11 shows the intensity matrix of the simulation matrix whereas in Figure 10 it is shown how an uncorrelated data set looks like. It can be seen that a correlated data set should be simulated instead of a uncorrelated (and simpler) data set, see Figure 9. Uncorrelated noise shows no clustering of points, while Figure 9 (an actual PET data-image) and 11 show clustering of high intensity points.

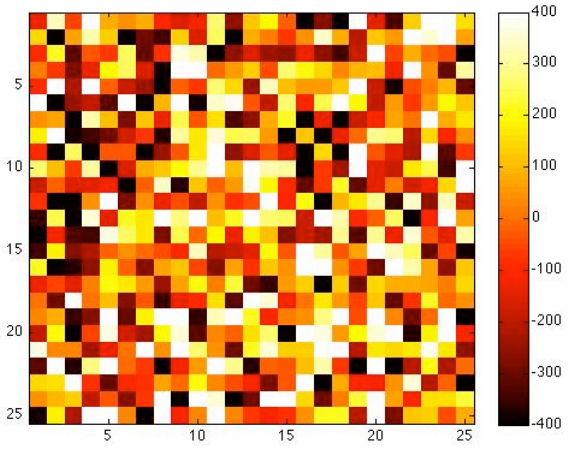

Figure 10: Uncorrelated noise

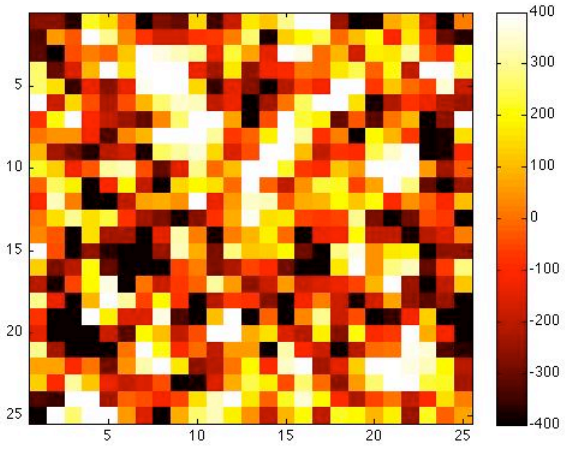

Figure 11: Correlated noise

So, a correlated data set is necessary to mimic the PET images. The data set will be generated by a Matlab command and they will be used in Chapter 5 for further research and experiments.

5 Simulations

The random values based on the multivariate distribution generated in Chapter 4 are used for simulations where the goal is to simulate a real situation. The simulation matrix will be influenced artifically by increasing one pixel. We refer to this pixel as the *increased pixel*. A to-be-designed algorithm has the task to detect the increased pixel. In our simulations we know the location of the increased pixel. For this reason it is possible to test the performance of the algorithm.

5.1 Determining an outlier

A pixel in the data set is called an *outlier* if its intensity is above a certain value. This limit value, v_{limit}, is a constant number for a (normal) distribution.
Usually the number of outliers of a data set is defined as a percentage. In Figure 12, it can be seen that for a normal distribution approximately 95% of the values lie within two standard deviations from the mean. This criterion is known as the empirical rule.

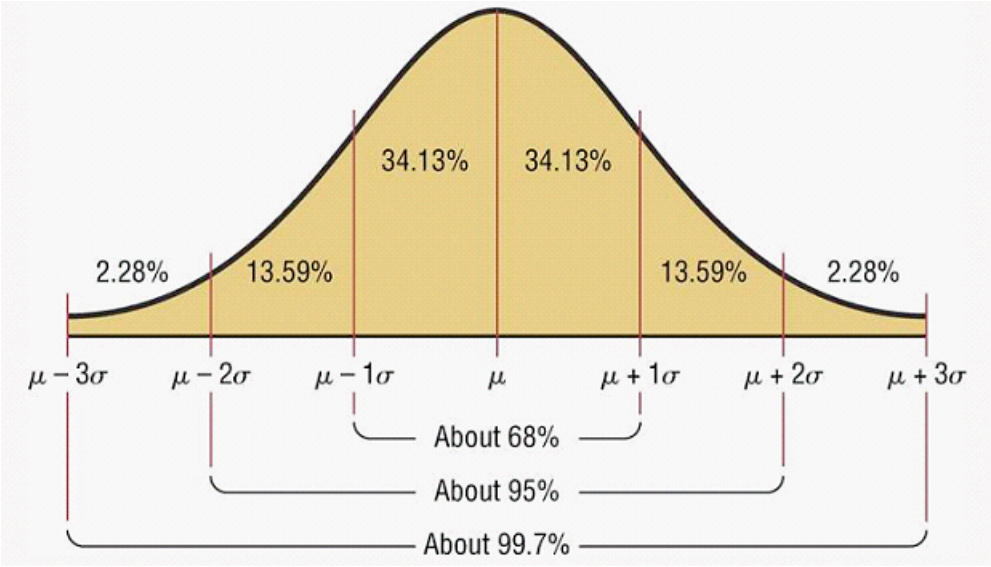

Figure 12: Empirical rule of normal distribution [5]

With the error-function [6] corresponding to the normal distribution it is possible to calculate the interval that contains any percentage of the data. For the mean and variance of the nosie, estimated from the orginal PET image, the limit value for 5% outliers is 475. (For 10% outliers the limit is 370 and for 1% it is 672.) The 95% interval contains positive outliers only and is called one sided, i.e.

$$\int_{-\infty}^{v_{limit}} f(x)dx = 0.95.$$

5.2 Distance

The data set with random values is simulated and the outliers are found. The next step is to determine how to distinguish a suspicous pixel from the outliers. One possibility is to look at the intensity. As

described before a high intensity does not automatically mean that it is suspicious. It has been shown that pixels are correlated with each other. The hypothesis is that a suspicious pixel distinguishes itself from an outlier by a smaller correlation with its neighbors although it is hard to calculate this correlation. Therefore the difference between an outlier and a suspicious pixel could be found by looking to the differences in distance.
The distance (d) is the difference between the intensity of a pixel and one of its neighbors, see (20). The average distance $\overline{d}$ is the sum of all distances, for all h neighbors, divided by the number of neighbors for pixel i.

$$d := |x - z| \tag{20}$$

$$\overline{d_i} := \frac{1}{h}\sum_{k=1}^{h} |x_i - z_k| \tag{21}$$

The neighbors of the outlier are probably also relatively high because of the correlation between these pixels. Therefore the hypothesis is that the average distance for an outlier (d_o) will be smaller than the average distance for an increased pixel (d_p).

5.3 One increased pixel

A possible way of distinguishing a suspicious pixel from an outlier is described in Subsection 5.2. For simulation of a PET image one pixel in the data set will be increased. This increased pixel can be seen as a suspicious pixel in reality. The intensity of the increased pixel is raised to a value above v_{limit}.
The methodology of the simulations is to identify all outliers, including the increased pixel, and to calculate all distances. The first round takes all eight direct neighbors. The second round consists of the sixteen neighbors $z_9 - z_{24}$.
Equations (22) and (23) give the formula for the distance for the first and second neighborhood round.

$$d_{r1,i} := \frac{1}{8}\sum_{k=1}^{8} |x_i - z_k| \tag{22}$$

$$d_{r2,i} := \frac{1}{16}\sum_{k=9}^{24} |x_i - z_k| \tag{23}$$

The results of the two rounds are shown in Table 5 for different percentage outliers. It turns out that the average distance between an outlier and its direct eight neighbors is twice as small as the average distance for the increased pixel. This ratio is smaller for the second round.

This means that the results of the first round are sufficient to distinguish an outlier from the increased pixel. The second round is not necessary anymore. As shown before these neighbors have less correlation with the outlier because of the pixel space (p_{space}). The correlation between a pixel and its n^{th} neighbor decreases rapidly for n larger than one which means that the influence of the noise is larger than the influence of the correlation between pixels. Also the diagonal neighbors may not have a significant influence because of the pixel distance p_{space}. Therefore only the direct neighbors (z_1, z_3, z_5, z_7) are taken into account. Table 6 gives the definition of the neighbors and their corresponding direct neighbors from Table 3.
In Table 7 the average distance as well as the average squared distance are presented. The reason for

1%	d_{r1}	d_{r2}
outlier	518	759
pixel	1002	1008
5%		
outlier	414	594
pixel	714	711
10%		
outlier	365	513
pixel	558	562

Table 5: Different intensity values for two rounds of neighbors and different percentage outliers

Pixel	**Location**
n_1	z_1
n_2	z_3
n_3	z_5
n_4	z_7

Table 6: Four direct neighbors and their location

the squared distance is the possibility to say something about the theoretical background. It is difficult to say something about the expected value for the absolute distance, while the expected value for the squared distance has some similarities with a (co)variance or standard deviation.

For the comparison of the outliers with the increased pixel, the distance between a pixel and its neighbors is calculated. Table 7 shows the results for these calculations. The formula for d_o and d_p is the same. The only difference is the pixel x. For d_o this is an outlier and in the case of d_p it is the increased pixel.
For the sake of simplicity Equation (25) is used to calculate the squared distance. The average squared distance is determined by adding all squared values (for all outliers) and dividing by the number of outliers. The average distance for the increased pixel with its neighbors is just the value calculated by Equation (25) since there is just one increased pixel. If there are more increased pixels, the mean has to be taken for determining the average distance.

$$\overline{d_i} = \frac{1}{4}\sum_{k=1}^{4}|x_i - n_k| \tag{24}$$

$$\overline{d_{i,2}} = \frac{1}{4}\sum_{k=1}^{4}|x_i - n_k|^2 \tag{25}$$

The increased pixel has an intensity of $1.5 \cdot v_{limit}$, with v_{limit} the limit value as explained in Subsection 5.1. As given before the limit value for 1% outlier is 370 and 475 for 5%.
This value intensity value is chosen arbitrary such that the pixel is larger than v_{limit} but smaller than the maximum pixel.

$\beta = 5\%$			$\beta = 1\%$		
	Outlier	Pixel		Outlier	Pixel
d	368	711		459	1010
d^2	189391	588245		266465	1083509

Table 7: Difference in distance for two percentage outliers

A confidence interval is used to indicate the reliability of an estimator. The interval is observed, which means that the values are calculated from observations. For a 95% confidence interval in 95% of the cases the average (squared) distance will be between the calculated endpoints. However in 5% of the cases, it will not be.
Every confidence interval is of the following form,

$$[\bar{d} - \frac{c \cdot s}{\sqrt{n}}, \bar{d} + \frac{c \cdot s}{\sqrt{n}}] \tag{26}$$

with c being a value from the cumulative standard normal distribution function.
The variable α is a fixed number and gives the change (in percentage) that the average distance will not be between the endpoints. The choice of the number z is described below.

$$\begin{aligned} P(-z \leq X \leq z) &= 1 - \alpha \qquad \Leftrightarrow \\ P(|X| > z) &= \alpha \qquad (27) \\ P(-z < X) &= P(X > z) = \frac{1}{2}\alpha. \qquad (28) \end{aligned}$$

Because of the symmetry of the distribution around 0, we get

$$P(X \leq z) = 1 - \frac{1}{2}\alpha. \tag{29}$$

This construction is necessary because of the values in the cumulative standard normal distribution table [23].

$$\phi(c) = P(Z \leq c).$$

The variable c gets the value $\phi^{-1}(z)$, with $z = 1 - \frac{1}{2}\alpha$. For $\alpha = 0.05$ we have $c = 1.96$. The value of s is the estimated standard deviation calculated by Equation (4) on page 22. $\bar{d}$ is the average of the distance and n is the number of simulations used for the confidence interval.

Table 8 shows 95% confidence intervals for the average distance and the squared distance, both for an outlier and an increased pixel.

$\beta = 5\%$	Outlier	Pixel	$\beta = 1\%$	Outlier	Pixel
d	[367 ; 370]	[704 ; 719]		[456 ; 462]	[1003 ; 1018]
d^2	[188204 ; 190578]	[576482 ; 600008]		[263347 ; 269583]	[1067289 ; 1099728]

Table 8: 95% Confidence interval for two percentage outliers

As it can be seen, the confidence intervals for the outlier and pixel do not overlap. This is nice because the confidence intervals are used for determining whether a pixel is noise or suspicious tissue.

We remark, only the squared distance is used from now on because of the similarities with (co)variances and standard deviations. Henceforth from now on by the distance we mean the average squared distance.

6 Validation of the model

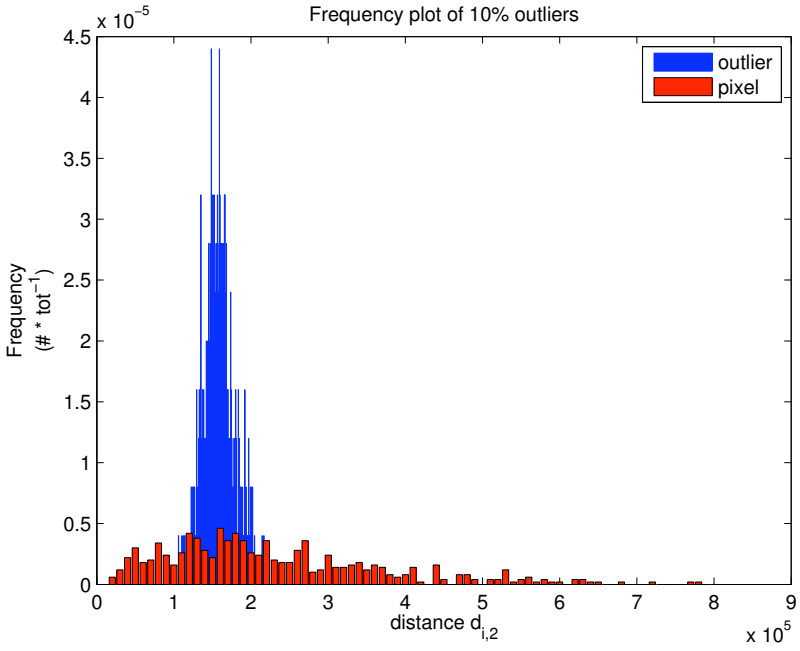

Figure 13: Histogram of the squared distance for 10% outliers

Figure 13 shows the frequency plots of the distance of the outlier and of the increased pixel where the distance is calculated by (25). This figure shows a large overlap between the support of the two plots. However the confidence intervals, as defined in (26), for an outlier and an increased pixel show no overlap and thus some questions regarding the credibility of the histogram arise. What is the expected value for the distance of an outlier and increased pixel? Will these theoretical values for the mean distance and variance of the distance for the outlier and pixel agree with the values resulting from the simulations?
In the next subsection we reproduce the formulas for the expectation and the variance of the increased pixel. These derivations are relatively simple because an increased pixel is fixed and hence it does not have a probability distribution. It is harder to derive the theoretical values for the outliers. The derivatioins shown here are rather short, the complete deduction can be found in Appendix B, on page 63.

6.1 Theoretical values of an increased pixel

The formula of the expected value for the distance between an increased pixel and its neighbors is explained here. Thus we calculate

$$\mathbb{E}\left[\frac{1}{4}\sum_{b=1}^{4}(y-x_b)^2\right] \quad =: \quad \mathbb{E}[W], \tag{30}$$

where x_b is a neighbor of y, the increased pixel.
The expectation of a sum is the sum of the expectations, therefore the formula can be written as

$$\mathbb{E}\left[\frac{1}{4}\sum_{b=1}^{4}(y-x_b)^2\right] = \frac{1}{4}\sum_{b=1}^{4}\mathbb{E}[(y-x_b)^2] \tag{31}$$

For each neighbor x_b the expectation is the same and the correlation between the increased pixel y (with constant value) and its neighbor x_b is known.

$$\begin{aligned}
\mathbb{E}[(y-x_b)^2] &= \mathbb{E}[(y-\mu+\mu-x_b)^2], \quad \text{with } \mathbb{E}(x_b)=\mu && (32)\\
&= \mathbb{E}[((y-\mu)-(x_b-\mu))^2] && (33)\\
&= \mathbb{E}[(y-\mu)^2 - 2(y-\mu)(x_b-\mu) + (x_b-\mu)^2] && (34)\\
&= (y-\mu)^2 + \sigma_x^2 && (35)\\
&= y^2 + \sigma_x^2, && (36)
\end{aligned}$$

where we used that $\mu = 0$, see page 31. The variable σ_x^2 is the variance of the noise, i.e. the standard deviation from (4) squared.
Deriving the variance of the increased pixel is a little harder. The equations become very large if all elements are described. For this reason, only the variance of one neighbor is described.

$$\begin{aligned}
\operatorname{var}(W) &= \frac{1}{16}\operatorname{var}\left(\sum_{b=1}^{4}(y-x_b)^2\right), \qquad \text{since } \operatorname{var}(ax) = a^2\operatorname{var}(x) \text{ for } a\in\mathbb{R} && (37)\\
&= \frac{1}{16}\sum_{b=1}^{4}\operatorname{var}(W_b) + 2\sum_{b<j}\operatorname{cov}(W_b, W_j), \qquad \text{with } W_b = (y-x_b)^2. && (38)
\end{aligned}$$

First the derivation of the variance for each neighbor is explained shortly. By (66), we have

$$\operatorname{var}(W_b) = \mathbb{E}[(y-x_b)^4] - (\mathbb{E}[(y-x_b)^2])^2 \tag{39}$$

The expectation of the fourth moment is not that easy. Using $\mathbb{E}(x_b) = 0$ and (98) we find

$$\mathbb{E}[(y-x_b)^4] = y^4 + 6y^2\sigma_x^2 + 3\sigma_x^4 \tag{40}$$

Now we have derived the fourth moment. The expectation of $(y-x_b)^2$ is shown previously in (36), so the square of it is also known. The formula for the variance is found by combining (66) with the results above.

$$\operatorname{var}(W_b) = 2\sigma_x^4 + 4y^2\sigma_x^2 \tag{41}$$

The equation for the covariance of an increased pixel is derived by using (71).

$$\begin{aligned}
\operatorname*{cov}_{b<j}(W_b, W_j) &:= \mathbb{E}[W_bW_j] - \mathbb{E}[W_b]\mathbb{E}[W_j],\\
\mathbb{E}[W_bW_j] &= y^4 + 2\sigma_x^2y^2 + \sigma_x^4 + \rho_{b,j}(4y^2 + 2\rho_{b,j}),\\
\operatorname*{cov}_{b<j}(x_b, x_j) &= \rho_{b,j} && (42)\\
\mathbb{E}[W_b]\mathbb{E}[W_j] &= \mathbb{E}[W_b]^2, \qquad \text{for } b = 1,2,3,4\\
\operatorname*{cov}_{b<j}(W_b, W_j) &= \sigma_x^2\rho_{b,j}(4y^2 + 2\sigma_x^2\rho_{b,j}) \quad \text{for } b<j && (43)
\end{aligned}$$

The last step is summing up all the variances and covariances, see Equation (38), and hence $\operatorname{var}(W)$ is derived.

6.2 Theoretical values of an outlier

An outlier comes from the same data set as the increased pixel, henceforth it is assumed that the set with outliers is distributed normally. The steps shown above are also used for deriving the theoretical values for an outlier. After some extensive calculations the following formula is derived for the expectation of the distribution of an outlier:

$$\mathbb{E}[(x - x_b)^2] = \sigma^2(1 - \rho_v - \rho_h), \tag{44}$$

with ρ_v and ρ_h the correlation coefficient for a vertical and horizontal neighbor, respectively. Pixel x is an outlier and x_b is a neighbor.
Only the theoretical values do not match with the results obtained from the simulations. The assumption that the outliers are distributed normally is not correct. The reason is that there is a condition for determining the outlier point x. This condition states that x is above the limit value (v_{limit}). This condition must be taken into account when performing the calculations and therefore the distribution of x changes into a conditional distribution.
Point X is an pixel and point T is a neighbor and say X and T are distributed normally. The joint probability density function of a multivariate normal distribution is given by (45)

$$f_{\mathbf{X}}(x) = \frac{1}{(2\pi)^k \left|\Sigma\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}, \tag{45}$$

with $\mathbf{X} = \begin{bmatrix} X \\ T \end{bmatrix}$. Since X and T have the same expectation and variance, we find

$$f_{X,T}(x,t) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)\sigma^2}[(x-\mu)^2+(t-\mu)^2-2\rho(x-\mu)(t-\mu)]}, \tag{46}$$

where ρ is the correlation coefficient between point X and T.

Given this joint probability density function we derive a formula for the conditional probability density function

$$f_{X,T|X>a}(x,t), \tag{47}$$

where X is an outlier and T is a neighbor.
The cumulative conditional density function for the outlier case is given by

$$\begin{aligned}
F_{X,T|X>a}(x,t) &= P(X \le x, T \le t \,|X > a\,) && (48)\\
&= \frac{1}{P(X > a)} \cdot P(a \le X \le x, T \le t) \quad \text{for } x > a && (49)\\
&= \frac{F(x,t) - F(a,t)}{1 - F(a)}, && (50)\\
\text{where} \quad F(a) &= P(X \le a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx && (51)\\
\text{with} \quad P(\mathbf{X} \le x) &= \int_{-\infty}^{x} f_{\mathbf{X}}(x) dX && (52)\\
& && (53)
\end{aligned}$$

Where Bayes Rule [23] is used to go from (48) to (49).

By using

$$F_{X,T|X>a}(x,t) = \int_{-\infty}^{x}\int_{-\infty}^{t} f_{X,T|X>a}(\varphi,\tau)\,d\varphi\,d\tau, \tag{54}$$

we see that the conditional probability density function becomes

$$\begin{aligned} f_{X,T|X>a}(x,t) &= \frac{\partial^2}{\partial x \partial t} F_{X,T|X>a}(x,t) && (55)\\ &= \frac{1}{1-F(a)} f_{X,T}(x,t) \qquad \text{for } x > a, -\infty < t < \infty, && (56) \end{aligned}$$

with $f_{X,T}$ given by (46).
For a standard normal distribution, the expectation of X^2 can be derived as

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx \qquad \text{with the probability function } f_X(x).$$

Thus the expected value of the distance of an outlier is given by (58).

$$\begin{aligned} \mathbb{E}[(X-T)^2\,|X>a] &= \int_a^{\infty}\int_{-\infty}^{\infty} (x-t)^2 f_{X,T|X>a}(x,t)\,dt\,dx && (57)\\ &= \frac{1}{1-F(a)}\int_a^{\infty}\int_{-\infty}^{\infty} (x-t)^2 f_{X,T}(x,t)\,dt\,dx && (58) \end{aligned}$$

Figure 14 shows the two dimensional plot for the conditional probability density function from (56).

6.3 Conditional variance

The theoretical value for the variance of the outlier is derived in the same way [4].

$$\begin{aligned} \operatorname{var}(X\,|\,Y) &= \mathbb{E}[(X-\mathbb{E}[X\,|\,Y])^2\,|\,Y] && (59)\\ &= \mathbb{E}[X^2\,|\,Y] - (\mathbb{E}[X\,|\,Y])^2 && (60) \end{aligned}$$

For the variance of the outlier the following equation can be used

$$\begin{aligned} \operatorname{var}((X-T)^2\,|\,X>a) &= \mathbb{E}[((X-T)^2)^2\,|\,X>a] - (\mathbb{E}[(X-T)^2\,|\,X>a])^2 && (61)\\ &= \mathbb{E}[(X-T)^4\,|\,X>a] - (\mathbb{E}[(X-T)^2\,|\,X>a])^2 && (62) \end{aligned}$$

The first part of (62) can be derived by (58) with $(x-t)^4$ instead of $(x-t)^2$. The second part of the variance is the expected value of an outlier squared, which is also known.

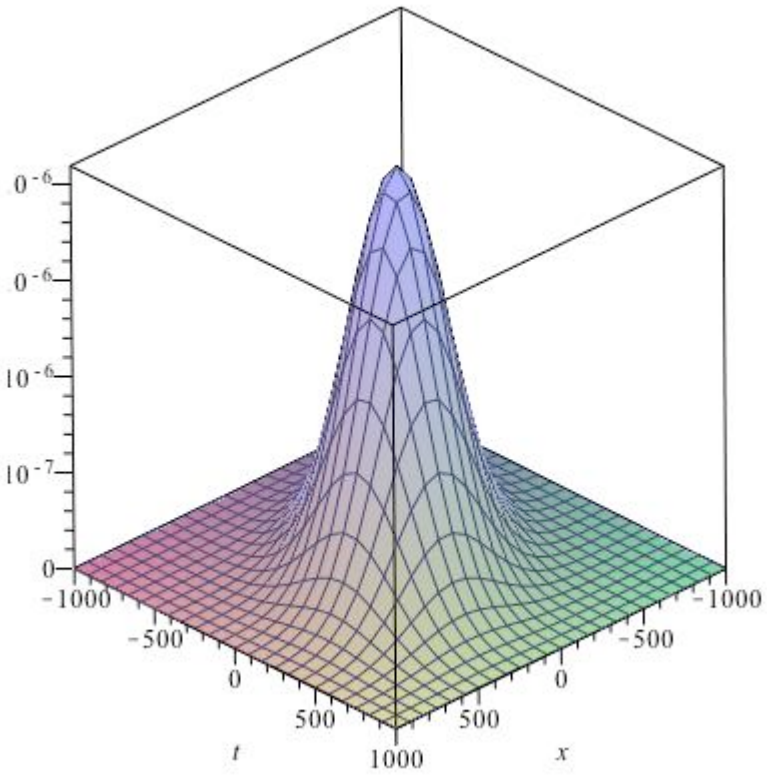

Figure 14: 2D plot for the joint probability function

	Expected value of the distance		
	Experimental	Theoretical	Discrepancy
Outlier	$1.8958 \cdot 10^5$	$1.9548 \cdot 10^5$	3.01%
Increased pixel	$3.5638 \cdot 10^5$	$3.6738 \cdot 10^5$	4.67%

Table 9: Experimental and theoretical expected values for 5% outliers performed by 2000 simulations

6.4 Validation of the values

A theoretical model is produced and a question that comes up naturally is, what is the performance of the model? Does the experimental and theoretical values match? The values for the simulations and validation are shown in Table 9.

The values obtained from the simulations match with the values from the theoretical background. For the standard deviation the determined values match less. Table 10 shows the discrepancy in percentage for the standard deviation of the increased pixel. The calculations are performed for two different numbers of simulations.

Table 10 shows that the discrepancy for the standard deviation of an increased pixel decreases if the number of simulations increases. We have to remark that for another round of simulations the values

	Discrepancy		
	$2 \cdot 10^5$ simulations	$4 \cdot 10^5$ simulations	$8 \cdot 10^5$ simulations
Outlier	86%	86%	86%
Increased pixel	15%	11%	0.28%

Table 10: Discrepancy of the standard deviation for 5% outliers and $v_{pixel} = 1.5 \cdot v_{limit}$

change. For example for three rounds of $8 \cdot 10^5$ give a absolute discrepancy of 20%, 0.28% and 13%. From Table 10 it can also be concluded that the standard deviation of an outlier converges much slower than the standard deviation of an increased pixel. For the same number of simulations, the discrepancy does not change. It will take much more simulations before the standard deviation of an outlier converges to its theoretical value. A reason which causes the slower convergence could be the difference in distribution of an increased pixel and outlier.

Chapter 7 shows different methods for qualifying the outliers. What options to distinguish a suspicious pixel from the outliers are possible? Therefore the task is to find a criterion to judge the outliers; which outlier is effected by noise and which one is not?

7 Results

7.1 Judgement of the outliers

The first goal of this thesis has been achieved. All pixels in the data set have been screened and the outliers (harmless or not) have been detected. The second question is how to distinguish the suspicious pixels from the outliers.
As can be seen in Chapter 5, it is possible to make a discrimination between an outlier and an increased pixel based on their distances. This distance is calculated via (25) on page 33.
The confidence intervals of the distances, Table 8 on page 36, are used for the classification of the outliers. Table 7 shows a significant difference between the distances of an outlier and an increased pixel. Although looking at the histogram for 1% and 5% outliers, Figures 15 and 16, indicates a (possible) problem.

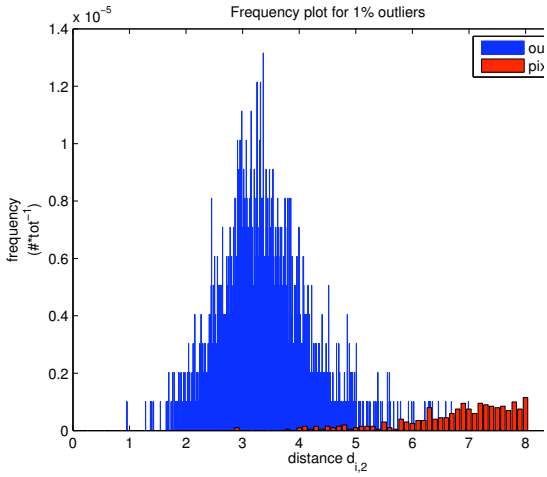

Figure 15: Histogram for a limit of 1%

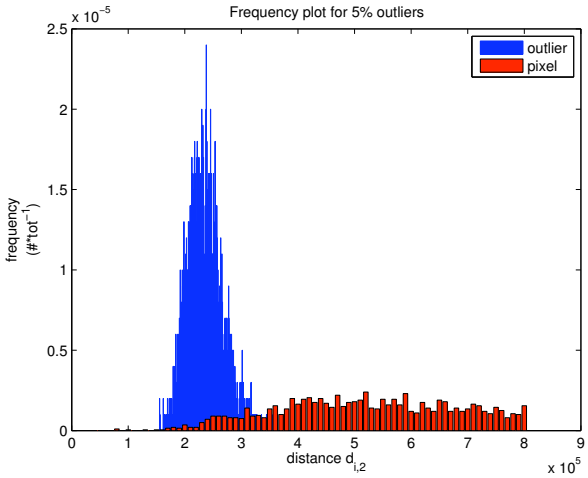

Figure 16: Histogram for a limit of 5%

The problem is the overlap in distances. By accident an increased pixel can also have a relatively small distance if the increased pixel is located in an area with high intensities. In this case the increased pixel will be marked as not suspicious. It is interesting to know how often this happens. The goal is to minimize the percentage of wrongly judged (suspicious) pixels and to have an acceptable percentage wrongly judged pixels. It is preferable to find (almost) all suspicious pixels and additionally assess some outliers as suspicious then to judge (almost) all outliers correctly but assess some suspicious pixels as outlier. The first error is a type I error, the latter is a type II error.
When the percentage of outliers (including the pixel) decreases, the histogram of "tumour" (red rods) shifts to right. The overlap between the outliers and pixels becomes smaller, as can be seen in Figures 15 and 16. For a smaller number of pixels with a large intensity, the suspicious pixels can easier be distinguish because of its relation with its neighbors.
We aim to find a criterion with a high percentage correctly judged suspicious pixels (p_{good}) and an allowable percentage wrongly judged outliers (p_{wrong}). Wrongly jugded outliers are actual outliers which are identified as suspicous.
In order to indicate whether data differs statistically or not, it is possible to use a confidence interval. A confidence interval is different for every data set. From Table 8 it can be seen that the confidence interval for an outlier $[CIO_L, CIO_R]$ and the interval of an increased pixel $[CIP_L, CIP_R]$ are different.

This information can be used to judge the data. Table 11 shows the result for a criterion for the distance with a constant value 200.000 for 5% outliers and 400.000 for 1% outliers. These constant criterions (200.000 and 400.000) are chosen arbitrary from Table 8. The upper limit of the confidence interval (CIO_R) for the squared distance of an outlier is approximately 190.000. In the case, the constant criterion that indicates suspicious pixel must be larger than 190.000 and so we choose 200.000. In the same way the value 400.000 has been taken as constant criterion in case of 1% outliers. The constant criterion for 1% outlier is taken significant higher because of the higher values for the confidence interval of the increased pixel $[CIP_L, CIP_R]$.
The variable v_{limit} is the limit value for a percentage of 1% outliers or 5% outliers, $v_{limit} = 372$ or $v_{limit} = 475$, respectively. The variable v_{pixel} is the intensity of the increased pixel.

	p_{good}	p_{wrong}
1%	98.9%	13.0%
5%	95.9%	32.2%

Table 11: Constant criterion, with $v_{pixel} = 1.5 \cdot v_{limit}$

The results for a large number of simulations and with a constant criterion are shown in Table 11. The percentage wrongly judged outliers is a little high, but that is not terrible. For example, 5% outliers of 625 pixels in this case means approximately 33 outliers including the increased pixel. Without the pixel there are 32 outliers and 32.2% of them are judged wrongly. So there are 11 wrongly judged points out of 625 pixels. Therefore, a high percentage wrongly judged pixels does not make the criterion bad.
A disadvantage of judging pixels with a constant criterion is that the influence of the patient and the radio-activity (the input) are not taking into account. Another radio-activity as input changes the whole situation, examples of this are shown in the next paragraph, Subsection 7.2.
Table 11 shows the results for an increased pixel with the value $1.5 \cdot v_{limit}$ for a constant criterion. The results for the same constant criterions (200.000 and 400.000) are worse if the increased pixel gets a lower value. Results can be seen in Appendix C, on page 67. The fraction correctly assessed increased pixels decreases while the fraction wrongly judged outliers stays the same. It turns out that the standard deviation changes if the injected radio-activity changes. It is also known that every patient is different; gender, age, level of metabolism. There are so many parameters with an effect on the intensity of the pixel. For this reason a constant criterion is not the right way for judging outliers and thus it is better to adapt the criterion to the patient and the injected dose activity.

In the example of the constant criterion $(200.000, 400.000)$, see Table 8, the criterion to judge pixels is larger than CIO_R. The question is how much larger the criterion must be in order to have a fine judgement. Is two times CIO_R suitable? Or is it more convenient to set the limit value somewhere halfway between CIO_R and CIP_L? Is one criterion enough or is it better to use several criterions, one for a every percentage of outliers? Evidently, there are a lot possibilities.

Table 12 shows the results for one variable criterion value based on the confidence intervals and k_{in}, where k_{in} is the difference between CIO_R and CIP_L;

$$k_{in} := CIP_L - CIO_R.$$

The natural assumption is made that $k_{in} > 0$.

Based on Figure 15 and 16 and Table 8, it seems wise to include the distance of the outliers as well as the distance for the increases pixel. By trial-and-error the variable criterion is chosen as $\frac{1}{4} \cdot k_{in}$ and $\frac{1}{2} \cdot k_{in}$.
In Appendix C more variable criterion values are presented, see page 67.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
1%	97.4%	6.68%	87.0%	0.836%
5%	88.9%	14.6%	76.4%	5.62%

Table 12: Variable criterion, with $v_{pixel} = 1.5 \cdot v_{limit}$

7.2 Variation in the simulations

In the previous paragraph some variations in simulations have been explained in this paragraph. Another experiment investigates the behavior of the model with one or more increased pixels already present in the model.

So far Model_1 is used for the simulations. Model_1 estimates the mean and standard deviation from the phantom data set. A large number of multivariable normal distribution data sets are simulated. A limit value v_{limit} is determined for a certain percentage of outliers. One pixel is increased to the intensity (v_{pixel}) depending on the limit value. All pixels with an intensity above the limit value are assessed as outliers. The algorithm takes a better look at all outliers and determines which outliers are influenced by noise and which are not, based on the distance of the neighboring points. See Chapter 5 for the details.

The algorithm used in the simulations determines the correct pixel in more than 90% of the cases both with a constant and a variable criterion. Does the algorithm also have such good results for PET images?
A large difference between the reality (PET scan) and these simulations is the unknown distribution of the noise. We have to determine the mean and standard deviation from the PET image while the distribution of the noise for the simulations is known. A large problem is the presence of possible suspicious pixels. These suspicious pixels affect the mean and standard deviation for sure. The question is, how large is the influence of these suspicious pixels on the mean and standard deviation?
To answer these questions the basic model, which we will call Model_1, is extended to two models, Model_2 and Model_3.

Model_2 is a generated data set with random values from Model_1 already containing one increased pixel. This generated random data set with one increased pixel is the input. For Model_1 the input is a data set without the increased pixel.
The standard deviation of the data set of Model_2 differs from the standard deviation of Model_1 (and the original PET data set) because of the presence of the increased pixel. With the new mean and standard deviation a large number of data sets is simulated, the new limit value v_{limit} is calculated based on the simulations and used for the delineation of (suspicious pixels and) outliers. From all distances of the outliers and known increased pixels the confidence intervals are calculated. These confidence interval are used for the assesment of again a large number of (newly) simulated data sets and the fraction of

correctly and wrongly judged pixels is calculated.
Model_3 is the same as Model_2 only with two increased pixels instead of one. These two increased pixels are not located abreast because the expectation is that two abreast located pixels are visible on a PET image. The difficulty in this thesis is to find suspicious pixels anywhere in the PET image.

The confidence intervals of one study for all three models are presented in Table 13. The percentage outliers is 5% and the intensity value of the increased pixel is $1.5 \cdot v_{limit}$.

	d_o^2	d_p^2
Model_1	[188204 ; 190578]	[576482 ; 600008]
Model_2	[194491 ; 197128]	[582837 ; 606290]
Model_3	[194037 ; 196630]	[581732 ; 605076]

Table 13: Confidence interval for 5% outliers and three models

The confidence intervals for Model_2 and Model_3 differs from Model_1. The differences are in the order of approximately 10%.
A constant criterion gives the results presented in Table 14. The constant value for 1% and 5% outliers is taken 400.000 and 200.000 respectively. The intensity of the increased pixel v_{pixel} is taken $1.5 \cdot v_{limit}$

	$\beta = 1\%$		$\beta = 5\%$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	98.9%	13.0%	95.9%	32.2%
Model_2	98.8%	14.5%	96.1%	33.7%
Model_3	97.9%	46.3%	91.5%	38.4%

Table 14: Three models with a constant criterion for two percentage outliers

In Table 14 for both percentage outliers (harmless or not) can be seen that the percentage correctly judged pixels stays almost the same. The difference is three correctly judged suspicious pixels 2000 simulations. This is not desirable but not very significant.
The percentage wrongly judged pixels is another story. For 5% outliers the percentage increases a little, from 32.2% to 38.4%, which means two wrongly judged pixels more but this is not significant. The percentage for 1% increase much more with the result of ten more wrongly judged pixels (of the 31 outliers).

To judge whether an outlier is suspicious or not, a variable criterion is better. An advantage of a variable criterion is the possibility to variate per patient. Each patient is different with its own standard deviation so if the standard deviation changes, the associated confidence interval changes as well, see Subsection 7.1.
The results for a variable criterion the values are given in Table 15. The simulations are performed with 5% outliers and pixel intensity $v_{pixel} = 1.5 \cdot v_{limit}$.

The results of the simulations are promising, so a natural question that comes up is, what are the results on PET images? The outcomes of these experiments are shown in the next subsection.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	88.9%	14.6%	76.4%	5.6%
Model_2	86.2%	14.5%	78.0%	5.4%
Model_3	78.3%	19.1%	55.8%	8.7%

Table 15: Three models for 5% outliers with a two variable criterions

7.3 PET images

The approach for the PET images is based on the idea of Model_3. The reason for using Model_3 is the presence of suspicious pixels. For Model_1 the distribution of the noise, i.e. the mean and standard deviation, are already known. The distribution of the noise in a PET image is unknown and have to be determined from the image. The expectation is that there are more than one suspicious pixels present. That is why the approach of Model_3 is used for the PET images.
The first step is to estimate the distribution of the actual PET image. The associated values (expected value $\bar{x}$, standard deviation s and v_{limit}) are determined and used to generate a large number of data sets. Each data sets receives one increased pixel. The outliers are determined by the value v_{limit} and the distances (d_o and d_p) are calculated. For these distances confidence intervals are determined and with these intervals the original PET image is judged. The main question is if there are points present which differ significantly from the background?
For the best similarity with Model_3, PET images of the liver are taken since liver tissue has the best homogeneous background. The used PET images are from actual patients that can be categorized as (most likely) healthy patients or patients with some disease in other parts of the body.

Figure 17 shows the PET image of a part of the liver of a (most likely) healthy patient. The four points describe the area that is taken for the simulations.

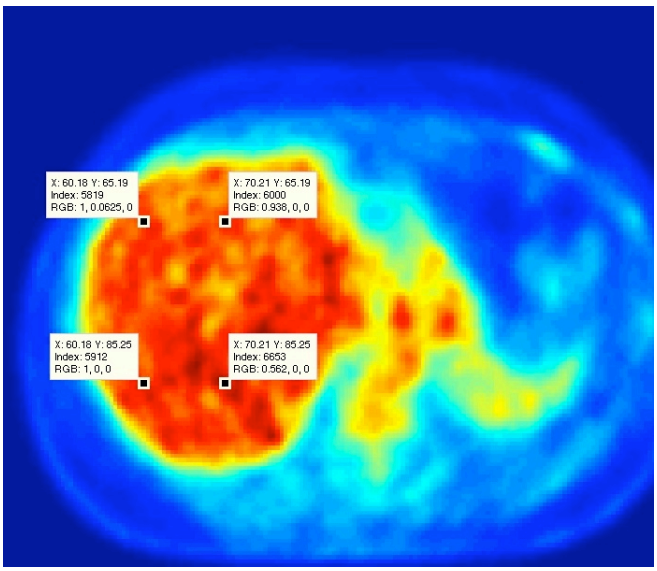

Figure 17: PET image of a liver from a patient

Figure 18 and 19 show the intensity figures for two (most likely) healthy patients. For the young and old patient the same size of data set is chosen.

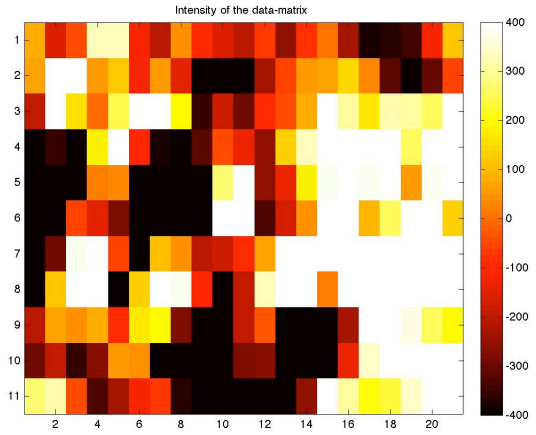

Figure 18: Young patient

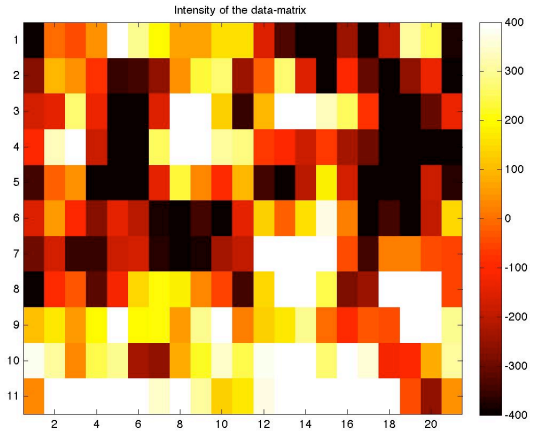

Figure 19: Old patient

These figures show the differences between two patients, both with a (most likely) healthy liver. Table 16 shows the correlation coefficients for the PET images of these (probably) healthy livers.

	Young patient	Old patient	Discrepancy
z_1	0.56	0.72	-29%
z_2	0.50	0.42	$+16\%$
z_3	0.72	0.67	$+7\%$
z_4	0.34	0.40	-18%
s_{t-1}	0.85	0.95	-12%
s_{t+1}	0.80	0.84	-5%

Table 16: Correlation coefficient for two (most likely) healthy livers

Table 16 shows the correlation coefficients for neigbors z_1-z_4. The correlation coefficients for the other half (z_5-z_8) have the same values as the values described in Table 16 because of symmetry.
As can be seen for the young patient in Table 16 the correlation coefficients for equal pixel space p_{space} differ. Table 17 show the discrepancy between values for the neighbors, compared with the values from the phantom study (Table 2).
Table 17 show that the discrepancy between neighbors with the same pixel space p_{space} is the largest for the young patient. The discrepancy from the old patient match better with the phantom study but still it is significant different from each other.

Table 18 shows the correlation coefficients with the n^{th} horizontal neighbor. The 1^e neighbor is point z_3 and point z_{13} is the 2^{nd} neighbor ($n = 2$), see page 29. The results for the correlation coefficients for the vertical n^{th} neighhor are presented in Table 19. Table 18 and 19 show the correlation coefficients for n horizontal and vertical neighbors.

	Young patient	Old patient	Phantom Study
$z_1 - z_3$	29%	7%	7%
$z_2 - z_4$	32%	5%	0.2%
$s_{t-1} - s_{t+1}$	6%	12%	0.7%

Table 17: Absolute value of the relative discrepancy between correlation coefficients for two (most likely) healthy livers

n	Young patient	Old patient
1	0.72	0.67
2	0.38	0.28
3	0.24	0.07
4	0.23	0.04
5	0.22	0.08

Table 18: Correlation with horizontal neighbors

n	Young patient	Old patient
1	0.56	0.65
2	0.17	0.33
3	0.11	0.26
4	0.09	0.25
5	0.04	0.17

Table 19: Correlation with vertical neighbors

From Table 18 can be seen that for the young patient the correlation coefficients hardly decrease for the n^{th} neighbor, with $n = 2, 3, 4, 5$ while for the same neighbors the values for the old patient are promising. Table 19 shows the correlation coefficients for the vertical neighbors and these are opposite compared with Table 18. The values for the young patient decrease for neighbor $2, 3, 4$ and 5 while for the old patient the values are stable for the same neighbors $n = 2, 3, 4, 5$.
After taking a look at the variance for the young and old patient, we see large values, $\text{var}_{jong} = 486^2$ and $\text{var}_{oud} = 342^2$. This values combined with the strange and large correlation coefficients give a negative semi-definte covariance matrix R, see page 28. The covariance matrix is necessary to simulate data sets and further steps for the delineation of suspicious pixel in a noisy image.

The next result is a sick patient with tumour tissue, but not in the liver. One of the questions of a doctor is, whether or not there are suspicious pixels in the liver.

Figure 20 shows the intensity plot of the sick patient. It can be seen that the pixels are less correlated, so perhaps now it is possible to get a positive semi-definite covariance matrix. Table 20 shows the outcomes for the first round neighbors.

	Sick patient
z_1	0.49
z_2	0.45
z_3	0.64
z_4	0.23
s_{t-1}	0.73
s_{t+1}	0.73

Table 20: Correlation coefficients for a sick patient with a (most likely) healthy liver

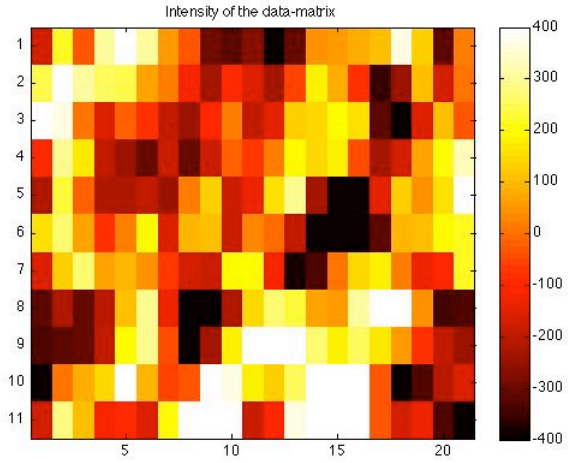

Figure 20: Intensity plot of the data of sick patient

	Sick patient	
n	Horizontal neighbors	Vertical neighbors
1	0.64	0.47
2	-0.03	0.15
3	-0.15	-0.09
4	-0.18	0.23

Table 21: Correlation coefficients for n horizontal and vertical neighbors

As it can be seen in Table 21, the correlation coefficients are smaller compared with Table 18 and 19. Still it is not possible to derive a positive semi-definite covariance matrix R. Without a covariance matrix it is not possible to distinguish suspicious pixels in a PET scan based on the distance.
In PET images we do not have the location of the suspicious pixels and therefore the distance can not be calculated. The distance for an increased pixel has to be estimated by simulations but that is possible because of a non positive semi-definite matrix R.
The experiments with the PET images end here.

8 Conclusion

The results are presented in Chapter 7. This chapter will summarize the outcomes and presents the conclusion based on these outcomes.

8.1 Simulations

Chapter 7 shows the results for different criterions for distinguishing a suspicious pixel from outliers. A constant criterion is not accurate because of the different parameters. The radio-activity as an input affects the average intensity of the PET image and therefore also the distance, see (25) on page 35. Another cause of variable pixel intensities is the patient because every patient is different, in e.g. gender, age and level of metabolism. Therefore it is necessary that the criterion depends on the actual PET image, and thus on the input and the patient specific parameters.
Table 12 shows the outcomes for a variable criterion based on the confidence intervals of an increased pixel and outlier. It is hard to find the best trade-off between a high percentage correctly judged pixels and a low percentage wrongly judged outliers. The best trade-off does not exist, i.e. it depends on the circumstances and the opinion of the doctor. Since not all patients are the same, it is the physician who has to decide which trade-off is the best in what situation.

Chapter 7 shows promising outcomes for different variable criterions on the simulations. Even when the criterions are applied on the extended models the results are still good.
However there are some differences between the theoretical and experimental values, see Chapter 6. The theoretical and experimental values for the expected value match which implies that the simulations are an accurate implementation of the reality. However the differences for the standard deviation are significant. The reason can be found in the degree of convergence and the distribution, see Chapter 6.

To conclude, the simulations are promising. With some adjustments the experimental and theoretical values will fully match. An example of a possible adjustment is performing more simulations. The results from the simulations differ per round of simulations. The results for four rounds of simulations are shown in Appendix C.1, Table 22. More simulations means less differences in outcomes and actual random values. From experiments it is known that the "random values" generated by Matlab are not always "random". This explains partially the different outcomes of the simulations. Another reason for the differences in the outcomes could be found in the written Matlab code due to human mistakes.
Another possibility is to take a look at the conditional distribution, both for the expected value and variance. Perhaps another approach of defining the formulas will give better results.
For now there is no reason not to use the simulations to find a criterion in order to judge the outliers.

8.2 PET images

The outcomes of the simulations are hopeful however the goal of this thesis is to recognize small tumour tissue in noisy FDG-PET images. A PET image will have suspicious pixels at an unknown location, therefore the simulations are extended to two other models. These models estimate the distribution of the noise based on the given data set. The outcomes for distinguishing increased pixels for these models (Model_2 and Model_3) are shown in Table 15 on page 49. Model_3 is used as a model to analyze actual PET images as well.
The results of the analysis of some PET images are shown in Subsection 7.3. These PET images present parts of (most likely) healthy livers. An important difference between the simulations and the analysis

of PET images is the data set. The simulations are performed with a data set of 625 pixels whereas a liver area has at most 150 pixels, i.e. a size of 10×15.
This appears to be a problem when calculating the correlation coefficient since the more pixels available the better the accuracy of these calculations. It is difficult and statistically not correct to say something based on a few data points. More data, in this case more pixels, means more reliable calculations.
As can be seen in Table 20, on page 51, the correlation coefficients are large because of the small number of data points. In this situation it is necessary to take more neighbors into account. The latter is a problem because the selected liver area is not large enough. For calculating the correlation with the 5^{th} neighbor, of a liver with size 10×15 pixels, approximately only 100 pixels are available.

Another consequence of the large correlation coefficients is the second condition of the covariance matrix R, i.e. it should be positive semi-definite, see page 28. It turns out that our matrix R is not positive semi-definite when there are large correlation coefficients. Without a covariance matrix it is not possible to generate data sets with correlated values. These data sets are necessary to determine confidence intervals for the distances of the outlier and pixel (d_o and d_p respectively), based on Model_3. No confidence intervals means no criterion to identify significant strange pixels in a noisy PET image.

At this stage it is hard to determine suspicious pixels in an actual PET image however we have succeeded to establish a model to simulate correlated noisy PET images and to detect suspicious pixels. Because of variations per patient further adjustment for the model are necessary such that the model is able to identify suspicious pixels from a noise actual PET image.

8.3 Goal

The first goal of this thesis, the analysis of the noise on PET images, has been achieved. Furthermore we succeeded to achieve the second goal partially: we have found a method to distinguish suspicious pixels from harmless (noisy) outliers. This method works well in the simulations, based on PET image from a phantom study. However, it does not work as we hoped in practice yet.

9 Discussion and further research

Chapter 8 presents the conclusion based on the results shown in Chapter 7. As can be read in Chapter 8 the research on PET images is not finished yet, because the outcomes on the PET images are not satisfying. In this chapter the research on PET images based on simulations will be discussed and also suggestions for further research are presented.

For the simulations some assumptions are made to simplify the situation. These assumptions could cause the bad outcomes for the PET images.
The first assumptions are made before the actual simulations took place. A PET image of a human body is modeled with simulations based on phantom studies. Therefore the background on a PET image is assumed to be homogeneous instead of heterogeneous (Chapter 4) on page 21.
For detecting suspicious pixels in an heterogeneous background it might be possible to simulate such a background. By adding extra noise specifically per organ it is possible to get a heterogeneous image that resembles a PET image of that specific organ. A necessity of this approach is the a priori knowledge of the noise which will hopefully improve the analysis. The noise is added in such a way that the heterogeneous image looks like an actual PET image. However a possible huge disadvantage of this approach is the differences between PET images. For every (kind of) patient another noise has to be added.
A second assumption is the dimension of a PET image. A PET image is three-dimensional while the data sets from the simulations are two-dimensional. From the entire two-dimensional image a limited region of interest of 25×25 is used.
These described assumptions probably cause the difference between the reality (PET images) and the approximation (simulations).

Another assumption we have made is the actual estimation of the noise model. The derived probability density function is a sum of two Hermite polynomials, see Chapter 4 on page 26. The data sets used in the simulations exist of values which are generated by a multivariate correlated normal distribution. Implementing the function given by (8), on page 26 is not possible because of the determined relations between pixels, the correlation coefficients.
This assumption also contributes to the discrepancies between the reality and the simulations.

The research for small deviations in PET images is not complete yet. The results from this thesis give an impression of the effect of noise and a possible approach to distinguish a strange pixel from a data set.
The approach used in the simulations, Chapter 5, based on the PET results is probably not optimal. The results from these simulations are acceptable, but in combination with PET image they fail. For improving the outcomes on the PET images in combination with the approach from this thesis, more data is necessary. The idea of using confidence intervals to distinguish strange pixels is probably correct. But the difficulty in determining the confidence intervals for an outlier and an increased pixel is a problem. Perhaps it is better to determine these confidence intervals by analyzing data from all kind of patients. This idea will cost a lot of time but hopefully it will give better results. A large advantage of analyzing the data is to get a general overview of the differences between patients. This information is also usable in other research.

Another cause of problems in analyzing the PET image is the small data set. The uncertainties caused by this problem can disappear if there is more knowledge available about the patient and the organs.

When comparing Tables 2 and 16, it can be seen that there are significant differences. The reason for these differences is probably due to the size of the data sets.
First of all, it has been shown that the variance differs per input concentration. Since the phantom and PET study have different radioactive concentration as input, the variance of the PET study will not be the same as the variance determined in the phantom study. This means that the correlation coefficients do not match either because the correlation coefficients depend on the variance.
Secondly, the actual values of the correlation coefficients are unknown, the values for the phantom study are an estimation of the real correlation coefficients. This is also the case for the PET study, however the estimation of the phantom study is much more accurate because it has much more pixels. To illustrate this, it is statistically incorrect to estimate a variable based on only three points.
If the phantom and PET study have the same input it might be possible that their correlation coefficients are the same as well, but it is hard to show this based on a large and a relatively small data set.
To summarize, the differences between the correlation coefficients of the two studies are caused by the different inputs and the size of the data set, i.e. the number of pixels.

Except for more patient data, the approach of the simulations and experiments on PET image could also be changed. In this thesis only two dimensional data sets are used and investigated. It is possible that a three dimensional investigation will give more and better information about the correlation between pixels. With this extra information it might be possible to end up with a positive semi-definite covariance matrix R.

As described earlier a covariance matrix R is necessary to generate data sets. The covariance matrix is positive semi-definite by definition, but because of large correlation coefficients the derived matrix in Subsection 7.3 is not positive semi-definite. Because of the large correlation between pixels, it is statistically difficult to say something relevant about the PET image.
One way to create a covariance matrix is by taking more correlation coefficients into account. Another possibility is changing the way of filling the matrix. It might be possible that filling the matrix differently, more correlation coefficients and an extra dimension will result in an actual covariance matrix.
An idea for another way of filling the matrix is by using a rule of thumb. From a theory used in Time Series Analysis [20], we know that the correlation between pixels with a larger length converges to zero with the natural assumption that the correlation coefficient $|a| < 1$. The first neighbor has a correlation of a and the second neighbor has a correlation of approximately a^2 etc. This idea of filling the matrix in a recursive way is illustrated in Appendix D on page 71.
The problem with the small data set is now almost gone if Using this way of filling the covariance matrix Without a covariance matrix it is not possible to generate data sets and without data sets no confidence intervals for identifying significant strange pixels in a noisy PET image can be derived. As described in Subsection 8.1 it is impossible to use a constant criterion, therefore the confidence intervals are of vital importance.
A possible solution is to use only the average (squared) distance of an outlier. This means we have to detect all outliers and calculate the average distance. So this idea is only possible if there are only a few suspicious pixels present in the image because then the influence of suspicious pixels is small. Substantial more suspicious pixels means much higher pixel intensities, which means a large effect on the mean and standard deviation, see Table 15. With more high intensity pixels the effect of the suspicious pixels on the mean and the standard deviation is very large and it becomes rather impossible to distinguish tumour pixels out of the harmless outliers. So we have a huge problem if there are many suspicious pixels.

A remark on the PET study is that the correlation coefficients for the PET study remain large. This is the main difference between the correlation coefficients for the phantom and PET study. In the phantom study the correlations decrease if the length between pixels gets larger. Furthermore the correlation coefficients for neighbors with a larger length are determined based on more data points. The hypothesis is that with more data points, it might be possible to calculate more reliable values for, for example the 5^{th} neighbor, and perhaps to derive a covariance matrix.

The last assumption that is made has to do with the size of the tumour tissue. In the simulations and PET study a tumour tissue with the size of one pixel is investigated. The reason for this assumptions is based on the hypothesis that tumour tissue with a size of two or more pixels (abreast) is visible with the human eye, therefore only tumour tissue of the size of one pixel is investigated.
Tumour tissue with a larger size can also be identified by the derived models with some adjustments. These adjustments should take the information of increased abreast pixels into account or an extra input with this information could help.

To conclude, the noise distribution derived in Chapter 4 is a rather good realization of PET images. The largest incorrectness between the simulations and a PET image is probably caused by the assumption of a normally distributed data set instead of the not normally distributed data set. A second cause of incorrectness is the shortage of information, i.e. the number of pixels.

Acknowledgement

The work for my final project has been performed. It is time to look back and to thank the lovely people who helped me during my study.

First of all, I would like to begin with mentioning my first supervisor, Hans Zwart. In the beginning we did not have much contact but I was always welcome, if he was present in Enschede. During the last months of my thesis, most of the time I was here at the University and the contact became more intensive. Mr. Zwart, thank you for all your help, good ideas, incredible knowledge about everything (sometimes it was found in a book), wonderful proverbs and positive energy to proceed my final project. It was pleasant working and discussing with you.

Secondly, I would like to thank my other supervisor, Jorn van Dalen. During my traineeship in Medisch Spectrum Twente he became my supervisor and he stayed it until he left to Zwolle to work there. Jorn, thank you for all your help, medical knowledge and questions. Especially answering your question and explaining what I had done, helped me attempting to solve the (sub)problems. Sometimes it was difficult to understand each other, but I learned a lot from you. I liked our time working together during my traineeship and thesis.

Furthermore I would like to say thanks to my family for all their understandings and the opportunity to run back to my parental environment. I know that I have changed during the years, but I was always welcome. It was nice to be home and to walk with you (and Sep). A last note for my mother; Maarn will always be home (thuis) for me.
Certainly not important are my friends. Thanks to you all, for listening to my strange stories and "big" problems, borrowing your coffee-card, playing cards with me and reading my enormously large e-mails. It did not matter how far away I was, you were there for me. Thanks a lot.

Last but not least, I would to thank my friend. Dear Almar, you are the best. Especially in the beginning I was a little sceptic about my thesis but look at me now; I am finishing proudly my final project. Thank you for believing in me, answering my stupid questions and helping me with everything.

Appendices

A Definitions

First a reenactment of two formulas for the variance [23] and covariance [23] which are used for the derivation of the equations in this thesis.

$$\begin{aligned} \operatorname{var}(x) &= \mathbb{E}[(x - \mathbb{E}(x))^2] && (63)\\ &= \mathbb{E}[x^2 - 2x\mathbb{E}[x] + (\mathbb{E}[x])^2] && (64)\\ &= \mathbb{E}[x^2] - 2\mathbb{E}[x]\mathbb{E}[x] + (\mathbb{E}[x])^2 && (65)\\ &= \mathbb{E}[x^2] - (\mathbb{E}[x])^2 && (66)\\ &= \sigma_x^2, && (67) \end{aligned}$$

with σ_x^2 the variance calculated by the standard deviaton from (4) squared.

$$\begin{aligned} \operatorname{cov}(x, y) &= \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] && (68)\\ &= \mathbb{E}[xy - x\mathbb{E}[y] - y\mathbb{E}[x] + \mathbb{E}[x]\mathbb{E}[y]] && (69)\\ &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[y]\mathbb{E}[x] + \mathbb{E}[x]\mathbb{E}[y] && (70)\\ &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] && (71)\\ &= \rho_{x,y} \cdot \sigma_x \sigma_y, && (72) \end{aligned}$$

with σ_x the standard deviation from (4) and $\rho_{x,y}$ the correlation coefficient between x and y, derived from (15) on page 27.

B The increased pixel

All steps made for the derivation of the expectation and variance of the distribution of an increased pixel are described below. A prior knowledge is that the increased pixel y is constant and the neighbors x_b are stochastic, is used for the derivation of these values.

B.1 Expectation

$$\mathbb{E}\left[\frac{1}{4}\sum_{b=1}^{4}(y-x_b)^2\right] = \mathbb{E}[W] \tag{73}$$

$$\frac{1}{4}\mathbb{E}\left[\sum_{b=1}^{4}(y-x_b)^2\right] = \frac{1}{4}\sum_{b=1}^{4}\mathbb{E}[(y-x_b)^2] \tag{74}$$

For each neighbor x_b the expectation is the same and the correlation between the increased pixel y and its neighbor x_b is derived below. Note that the increased pixel y has a constant value.

$$\mathbb{E}[(y-x_1)^2] = \mathbb{E}[(y-x_2^2)] = \mathbb{E}[(y-x_3^2)] = \mathbb{E}[(y-x_4^2)] \tag{75}$$

$$\mathbb{E}[(y-x_b)^2] = \mathbb{E}[(y-\mu+\mu-x_b)^2], \qquad \text{with } \mathbb{E}(x)=\mu \tag{76}$$

$$= \mathbb{E}[((y-\mu)-(x_b-\mu))^2] \tag{77}$$

$$= \mathbb{E}[(y-\mu)^2 - 2(y-\mu)(x_b-\mu) + (x_b-\mu)^2] \tag{78}$$

$$= \mathbb{E}[(y-\mu)^2 - 2(y-\mu)\mathbb{E}(x_b-\mu) + \mathbb{E}(x-b-\mu)] \tag{79}$$

since $\mathbb{E}[x_b] = \mu \Rightarrow \mathbb{E}[x_b-\mu] = 0$

$$= \mathbb{E}[(y-\mu)^2] + \mathbb{E}[(x_b-\mu)^2] \tag{80}$$

$$= (y-\mu)^2 + \sigma_x^2 \tag{81}$$

The noise is centered around zero, see Subsection 4.4 on page 31, therefore $\mu = 0$ and the expectation of the distance of the distribution of an increased pixel is derived en given by (82)

$$\mathbb{E}[(y-x_b)^2] = y^2 + \sigma^2. \tag{82}$$

B.2 Variance

$$\operatorname{var}(W) = \frac{1}{16}\operatorname{var}\left(\sum_{b=1}^{4}(y-x_b)^2\right), \qquad \text{since } \operatorname{var}(ax) = a^2\operatorname{var}(x) \text{ for } a \in \mathbb{R} \tag{83}$$

$$= \frac{1}{16}\sum_{b=1}^{4}\operatorname{var}(W_b) + 2\sum_{b<j}\operatorname{cov}(W_b, W_j), \qquad \text{with } W_b = (y-x_b)^2. \tag{84}$$

First we take a look at the variance

$$\begin{aligned}
\mathrm{var}((y-x_b)^2) &= \mathbb{E}[((y-x_b)^2 - \mathbb{E}[(y-x_b)^2])^2], \qquad \text{for } b=1,2,3,4 && (85)\\
&= \mathbb{E}[((y-x_b)^2)^2] - (\mathbb{E}[(y-x_b)^2])^2 && (86)\\
\mathbb{E}[((y-x_b)^2)^2] &= \mathbb{E}[(y-x_b)^4] && (87)\\
&= \mathbb{E}[(y-\mu+\mu-x_b)^4], \qquad \text{with } \mathbb{E}(x_b)=\mu. && (88)\\
&= \mathbb{E}[((y-\mu)-(x_b-\mu))^4] && (89)\\
&= \mathbb{E}[(y-\mu)^4 - 4(y-\mu)^3(x_b-\mu) + 6(y-\mu)^2(x_b-\mu)^2 - \\
&\quad 4(y-\mu)(x_b-\mu)^3 + (x_b-\mu)^4] && (90)
\end{aligned}$$

The expectation of a sum is the sum of the expectations

$$\begin{aligned}
\mathbb{E}[(y-x_b)^4] &= \mathbb{E}[(y-\mu)^4] - 4\mathbb{E}[(y-\mu)^3(x_b-\mu)] + \\
&\quad 6\mathbb{E}[(y-\mu)^2(x_b-\mu)^2] - 4\mathbb{E}[(y-\mu)(x_b-\mu)^3] + \mathbb{E}[(x_b-\mu)^4] && (91)\\
&= \mathbb{E}[(y-\mu)^4] - 4(y-\mu)^3\mathbb{E}[(x_b-\mu)] + 6(y-\mu)^2\mathbb{E}[(x_b-\mu)^2] - \\
&\quad 4(y-\mu)\mathbb{E}[(x_b-\mu)^3] + \mathbb{E}[(x_b-\mu)^4] && (92)\\
&\overset{Eq.(114)}{=} (y-\mu)^4 - 4(y-\mu)^3\times 0 + 6(y-\mu)^2\sigma_x^2 - 4(y-\mu)\times 0 + 3\sigma_x^4 && (93)\\
&= (y-\mu)^4 + 6(y-\mu)^2\sigma_x^2 + 3\sigma_x^4, \qquad \mu=0, \text{ see page 31} && (94)\\
&= y^4 + 6y^2\sigma_x^2 + 3\sigma_x^4 && (95)
\end{aligned}$$

Taking everything together gives the next formula for the variance of an increased pixel,

$$\begin{aligned}
\mathrm{var}((y-x_b)^2) &= y^4 + 6y^2\sigma_x^2 + 3\sigma_x^4 - (y^2+\sigma_x^2)^2 \text{ using Eq. (67)} && (96)\\
&= y^4 + 6y^2\sigma_x^2 + 3\sigma_x^4 - (y^4 + 2\sigma_x^2y^2 + \sigma_x^4) && (97)\\
&= y^4 + 6y^2\sigma_x^2 + 3\sigma_x^4 - y^4 - 2\sigma_x^2y^2 - \sigma_x^4 && (98)\\
&= 4y^2\sigma_x^2 + 2\sigma_x^4 && (99)
\end{aligned}$$

The equations for the covariance are shown below

$$\underset{b<j}{\operatorname{cov}}(W_b, W_j) \overset{Eq.(71)}{=} \mathbb{E}[W_b W_j] - \mathbb{E}[W_b]\mathbb{E}[W_j] \tag{100}$$

$$\begin{aligned}
\mathbb{E}[W_b W_j] &= \mathbb{E}[(y-\mu+\mu-x_b)^2(y-\mu+\mu-x_j)^2] \qquad \mathbb{E}[x_b] = \mu && (101)\\
&= \mathbb{E}[((y-\mu)-(x_b-\mu))^2((y-\mu)-(x_j-\mu))^2] && (102)\\
&= \mathbb{E}[((y-\mu)^2 - 2(y-\mu)(x_b-\mu) + (x_b-\mu)^2) \\
&\quad ((y-\mu)^2 - 2(y-\mu)(x_j-\mu) + (x_j-\mu)^2)] && (103)\\
&= \mathbb{E}[(y-\mu)^4 - 2(y-\mu)^3(x_j-\mu) + (y-\mu)^2(x_j-\mu)^2 - \\
&\quad 2(y-\mu)^3(x_b-\mu) + 4(y\mu)^2(x_j-\mu)(x_b-\mu) - 2(y-\mu)(x_b-\mu)(x_j-\mu)^2 + \\
&\quad (x_b-\mu)^2(y-\mu)^2 - 2(y-\mu)(x_b-\mu)^2(x_j-\mu) + (x_b-\mu)(x_j-\mu)] && (104)\\
&= \mathbb{E}[(y-\mu)^4] - 2\mathbb{E}[(y-\mu)^3(x_j-\mu)] + \mathbb{E}[(y-\mu)^2(x_j-\mu)^2] - \\
&\quad 2\mathbb{E}[(y-\mu)^3(x_b-\mu)] + 4\mathbb{E}[(y\mu)^2(x_j-\mu)(x_b-\mu)] - \\
&\quad 2\mathbb{E}[(y-\mu)(x_b-\mu)(x_j-\mu)^2] + \mathbb{E}[(x_b-\mu)^2(y-\mu)^2] - \\
&\quad 2\mathbb{E}[(y-\mu)(x_b-\mu)^2(x_j-\mu)] + \mathbb{E}[(x_b-\mu)(x_j-\mu)] && (105)\\
&= (y-\mu)^4 - 2(0) + (y-\mu)^2\sigma_x^2 - 2(y-\mu)^3\mathbb{E}(x_b-\mu) + \\
&\quad 4(y-\mu)^2\operatorname{cov}(x_b,x_j) - 2(0) + (y-\mu)^2\sigma_x^2 - 2(0) + \\
&\quad \sigma_x^4 + 2\operatorname{cov}(x_b,x_j)^2 && (106)\\
&= (y-\mu)^4 + (y-\mu)^2\sigma_x^2 + 4(y-\mu)^2\operatorname{cov}(x_b,x_j) + \\
&\quad (y-\mu)^2\sigma_x^2 - 2(0) + \sigma_x^4 + 2\operatorname{cov}(x_b,x_j)^2 && (107)\\
&= (y-\mu)^4 + 2(y-\mu)^2\sigma_x^2 + \sigma_x^4 + \operatorname{cov}(x_b,x_j)(4(y-\mu)^2 + 2\operatorname{cov}(x_b,x_j)) && (108)\\
&\overset{Ch.4.4}{=} y^4 + 2\sigma_x^2 y^2 + \sigma_x^4 + \operatorname{cov}(x_b,x_j)(4y^2 + 2\operatorname{cov}(x_b,x_j)) && (109)\\
&= y^4 + 2\sigma_x^2 y^2 + \sigma_x^4 + \rho_{b,j}(4y^2 + 2\rho_{b,j}), && (110)
\end{aligned}$$

with $\rho_{b,j}$ is the correlation between pixel x_b and x_j (=$\operatorname{cov}(x_b, x_j)$) from Equation (72), which is the correlation coefficient calculated from the orginal PET images.

The final equations become

$$\begin{aligned}
\mathbb{E}(W_b W_j) &= y^4 + 2\sigma_x^2 y^2 + \sigma_x^4 + \rho_{b,j}(4y^2 + 2\rho_{b,j}) && (111)\\
\mathbb{E}(W_b)\mathbb{E}(W_j) &= \mathbb{E}(W_b)^2 \qquad \text{for } b = 1,2,3,4 && (112)\\
\underset{b<j}{\operatorname{cov}}(W_b, W_j) &= \sigma_x^2\rho_{b,j}(4y^2 + 2\sigma_x^2\rho_{b,j}). && (113)
\end{aligned}$$

The derivations of higher moments of the multivariate normal distribution [10] are given in the next equations.

$$\begin{aligned}
\mathbb{E}[X_i^4] &= 3\sigma_i^4 && (114)\\
\mathbb{E}[X_i^3 X_j] &= 3\sigma_i^2\sigma_{ij} \qquad \sigma_{ij} = \operatorname{cov}(X_i, X_j) && (115)\\
\mathbb{E}[X_i^2 X_j^2] &= \sigma_i^2\sigma_j^2 + 2(\sigma_{ij})^2 && (116)\\
\mathbb{E}[X_i^2 X_j X_k] &= \sigma_i^2\sigma_{jk} + 2\sigma_{ij}\sigma_{ik} && (117)
\end{aligned}$$

C Tables

C.1 Validation values for expected value

	Discrepancy		
	$2 \cdot 10^5$ simulations	$4 \cdot 10^5$ simulations	$8 \cdot 10^5$ simulations
Outlier	20%	20%	20%
Increased pixel	0.48%	1.1%	1.4%

Table 22: Absolute value of the relative discrepancy of the expected value for 5% outliers and $v_{pixel} = 1.5 \cdot v_{limit}$

C.2 Constant criterion

Table 23 shows the percentage wrongly and well judged outliers. The judgement is performed with a constant criterion of 200.000 and 400.000 for 1% and 5% outliers respectively. The value of the increased pixel (v_{pixel}) varies.

	v_{pixel}			
	$1.10 \cdot v_{limit}$		$1.25 \cdot v_{limit}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
1%	79.0%	13.3%	91.4%	13.3%
5%	75.9%	32.3%	85.7%	32.4%

Table 23: Constant criterion with values for v_{pixel}

C.3 Confidence interval for three models

Table 24 has 1% outliers and the intensity of the increased pixel (v_{pixel}) is $1.5 \cdot v_{limit}$.

	$d^2_{outlier}$	d^2_{pixel}
Model_1	[263347 ; 269583]	[1067289 ; 1099728]
Model_2	[274814 ; 282119]	[1086057 ; 1119124]
Model_3	[280191 ; 287947]	[1081411 ; 1114364]

Table 24: Confidence interval for 1% outliers for the three models

C.4 Different variable criterion

Recall that k_{in} is the difference between the upper limit (CIO_R) and the lower limit (CIP_L).

$$k_{in} := CIP_L - CIO_R > 0$$

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
1%	90.5%	13.2%	80.0%	4.03%
5%	79.7%	21.0%	69.4%	12.1%

Table 25: Variable value as criterion, with $v_{pixel} = 1.25 \cdot v_{limit}$

C.5 Three different models

C.5.1 $v_{pixel} = 1.25 \cdot v_{limit}$

Table 26 gives the results for the simulations with 5% outliers.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	79.7%	21.0%	69.4%	12.1%
Model_2	78.7%	21.5%	68.3%	12.5%
Model_3	86.3%	25.9%	84.0%	25.3%

Table 26: Three models for 5% outliers with a variable criterion and $v_{pixel} = 1.25 \cdot v_{limit}$

For 1% outliers the results are presented in Table 27.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	90.5%	13.2%	80.0%	4.03%
Model_2	86.5%	8.71%	72.5%	2.85%
Model_3	72.1%	32.1%	57.2%	24.3%

Table 27: Three models for 1% outliers with a variable criterion and with $v_{pixel} = 1.25 \cdot v_{limit}$

C.5.2 $v_{pixel} = 1.5 \cdot v_{limit}$

For 1% outliers and $v_{pixel} = 1.5 \cdot v_{limit}$ the results are shown in Table 28.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	97.4%	6.68%	87, 0%	0.836%
Model_2	97.2%	6.23%	88.4%	0.803%
Model_3	94.3%	39.3%	76.6%	30.0%

Table 28: Three models for 1% outliers with a variable criterion and with $v_{pixel} = 1.5 \cdot v_{limit}$

Table 29 gives the results for an pixel intensity (v_{pixel}) of $1.5 \cdot v_{limit}$ and 5% outliers.

	$CIO_R + \frac{1}{4} \cdot k_{in}$		$CIO_R + \frac{1}{2} \cdot k_{in}$	
	p_{good}	p_{wrong}	p_{good}	p_{wrong}
Model_1	88.9%	14.6%	76.4%	5.6%
Model_2	86.2%	14.5%	78.0%	5.4%
Model_3	78.3%	19, 1%	55.8%	8.7%

Table 29: Three models for 5% outliers with a variable criterion and $v_{pixel} = 1.5 \cdot v_{limit}$

D Covariance matrix

For the recursive way of filling the matrix, explained in Subsection 9.2 the natural assumption is made that

$$\operatorname{cor}(x, z_1) = a \qquad \text{with } |a| < 1. \tag{118}$$

The covariance matrix R is of the following form

$$R = \sigma^2 \begin{bmatrix} R_{1,1} & R_s & R_{s2} & R_{s3} \\ R_s & R_{2,2} & R_s & R_{s2} \\ R_{s2} & R_s & R_{3,3} & R_s \\ R_{s3} & R_{s2} & R_s & R_{4,4} \end{bmatrix},$$

with

$$R_{1,1} = \sigma^2 \begin{bmatrix} 1 & a & a^2 & a^3 \\ a & 1 & a & a^2 \\ a^2 & a & 1 & a \\ a^3 & a^2 & a & 1 \end{bmatrix} \qquad \text{and} \qquad R_s = \sigma^2 \begin{bmatrix} a & a^{\sqrt{2}} & a^{\sqrt{5}} & a^{\sqrt{10}} \\ a^{\sqrt{2}} & a & a^{\sqrt{2}} & a^{\sqrt{5}} \\ a^{\sqrt{5}} & a^{\sqrt{2}} & a & a^{\sqrt{2}} \\ a^{\sqrt{10}} & a^{\sqrt{5}} & a^{\sqrt{2}} & a \end{bmatrix}.$$

The other matrices are filled in the same way. In this way, the correlation between the points $b = (i, j)$ and $c = (k, l)$ from the original PET image is given by

$$\operatorname{cor}(b, c) = a^{\sqrt{(i-k)^2+(j-l)^2}}.$$

References

[1] Matlab - the language of technical computing. http://www.mathworks.nl/products/matlab/, 1994.

[2] Introduction to PET Physics. http://depts.washington.edu/nucmed/IRL/pet_intro/intro_src/section2.html, 1999.

[3] The Multivariate Gaussian Distribution. http://cs229.stanford.edu/section/gaussians.pdf, 2008.

[4] Conditional Expectation and Variance. http://isfaserveur.univ-lyon1.fr/~stephane.loisel/prerequis_esp_cond.pdf, 2012.

[5] Empirical rule. http://www.shannonforbesblog.com/wp-content/uploads/2012/02/EmpiricalRule.png, 2012.

[6] Normal Distribution. http://mathworld.wolfram.com/NormalDistribution.html, 2012.

[7] Hermite polynomials. https://commons.wikimedia.org/wiki/File:Hermite_polynomial.svg, 2012.

[8] Hermite Polynomial. http://mathworld.wolfram.com/HermitePolynomial.html, 2012.

[9] Descriptive Statistics. http://www.southalabama.edu/coe/bset/johnson/lectures/lec15.html, 2012.

[10] Multivariate Normal Distribution. http://en.wikipedia.org/wiki/Multivariate_normal_distribution#Cumulative_distribution_function, 2012.

[11] M. Ahmed. A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE Transactions on Medical Imaging*, 21(3):193–199, 2002.

[12] R. Boellaard. Standards for PET Image Acquisition and Quantitative Data Analysis. *Department of Nuclear Medicine and PET Research*, 50:11–20, 2009.

[13] I. Buvat. A bootstrap approach for analyzing the statistical properties of spect and pet images. *IEEE Transactions on Nuclear Science*, 1:1–5, 2002.

[14] W. Cai. Fast and robust fuzzy c-means clustering algorithms incorporation local information for image segmentation. *Department of Computer Science and Engineering*, 1:1–27, 2007.

[15] E. Clementel. Comparison of image signal-to-noise ratio and noise equivalent counts in time of flight pet. *IEEE*, 1:3622–3625, 2010.

[16] M. Dahlbom. Estimation of image noise in pet using the bootstrap method. *IEEE Transactions on Nuclear Science*, 49(5):2062–2066, 2002.

[17] I. Despotović. Noise-robust method for image segmentation. *Department of Telecommunications and Information Processing*, 1:1–10, 2010.

[18] Y.A. Ghassabeth. Mri fuzzy segmentation of brain tissue using ifcm algorithm with genetic algorithm optimization. *IEEE/ACS International Conference on Computer Systems and Applications*, 1:665–668, 2007.

[19] H. Kitamura. Estimation of local statistical noise in pet images induced by attenuation inside the body. *Ann Nuclear Med*, 24(1):197–205, 2010.

[20] G. Meinsma and H. Kwakernaak. *Time Series Analysis and System Identification*. Department of Applied Mathematics, 2009.

[21] M.E. Phelps. *PET ; Physics, Instrumention and Scanners*. Springer, 2006.

[22] K. Rank. Estimation of image noise variance. *IEEE Image Signal Process*, 146(2):80–84, 1999.

[23] J. Rice. *Mathematical Statistics and Data Analysis*. Cengage Learning, Inc, 2006.

[24] C. Riddell. Noise reduction in oncology fdg pet images by iterative reconstruction: A quantive assesment. *The Journal of Nuclear Medicine*, 42(9):1316–1323, 2001.

[25] M. Rzeszotarski. The AAPM/RSNA Physics Tutorial for Residents. *RadioGraphics*, 19(1):765–782, 1999.

[26] G. Tarantola. Pet instrumentation and reconstruction algorithms in whole body applications. *The Journal of Nuclear Medicine*, 44(5):756–769, 2003.

[27] C.C. Watston. New, faster, image-based scatter correction for 3d pet. *IEEE Transactions on Nuclear Science*, 47(4):1587–1594, 2000.

[28] D. Zhang. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 32(1):37–50, 2004.