Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea Magistrale in Fisica

ClinQC: quality control of an X-ray imaging system using clinical images

Relatore:
Prof.ssa Maria Pia Morigi

Presentata da:
Lisa Bravaglieri

Correlatori:
Dott. Wouter J. H. Veldkamp
Dott.ssa Pieternel van der Tol
Dott. J. Chiel den Harder
Prof. Berend C. Stoel

Anno Accademico 2015/2016

Abstract

The work presented in this thesis is part of a research project of Leiden University Medical Center (LUMC) in The Netherlands. It belongs to the field of Diagnostic Radiology analysed from a Medical Physics point of view. After a short overview of the weekly quality controls of an X-ray imaging device, performed using simple phantoms, the thesis focuses on a novel approach called ClinQC (Clinical images-based Quality Control): it has the purpose to monitor the stability of imaging devices, aiming at the early detection of changes in image quality or radiation dose, by deriving quality parameters from chest images of routine patient examinations. The ClinQC algorithm extracts the noise from clinical images and derives the main dose quantities. The noise study presented in this thesis comprehends a validation of the algorithm, performed in several ways: image deteriorations, simulations, phantom studies and real clinical examples. For dose and homogeneity studies only some preliminary results are presented. The thesis collects also some ideas of improvement that can be considered for the future versions of the algorithm and to extend the ClinQC project to other X-ray anatomies and imaging modalities.
The obtained similar results for the two compared methods prove that ClinQC is able to give immediate feedbacks of the quality of the imaging devices using patient images. It provides reliable, on-the-fly and sensitive parameters of the quality of the X-ray imaging system, that have the same physical meaning and similar relative variation as the quality indicators of the gold standard QClight method. It can be concluded that the ClinQC algorithm could be already applied in clinical practice, with the initial support of the QClight weekly quality control. In this way, a comparison between the two methods in a real test period will be a guide to find the necessary adjustments of the algorithm until the final version is being installed and stably used in clinical practice.

Sommario

Il lavoro presentato in questa tesi é parte di un progetto di ricerca dell'ospedale universitario Leiden University Medical Center (LUMC) dei Paesi Bassi. La tesi si inserisce nel campo della Radiologia Diagnostica dal punto di vista della Fisica Medica. Dopo una breve panoramica sui controlli settimanali di qualitá di un sistema di diagnostica per immagini a raggi X, eseguiti con semplici fantocci, la tesi si concentra su un nuovo approccio chiamato ClinQC (Clinical images-based Quality Control). Esso ha lo scopo di monitorare la stabilitá del dispositivo, mirando alla rilevazione precoce delle variazioni nella qualitá delle immagini o nella dose di radiazioni, derivando parametri di qualitá a partire da immagini di esami di routine al torace dei pazienti. L'algoritmo ClinQC estrae il rumore dalle immagini cliniche e ricava le principali grandezze dosimetriche. Lo studio sul rumore presentato in questa tesi comprende la validazione dell'algoritmo, eseguita in diversi modi: deterioramento delle immagini, simulazioni, studi su fantocci ed esempi clinici reali. Per quanto riguarda gli studi di dose ed omogeneitá, sono illustrati solo alcuni risultati preliminari. La tesi raccoglie anche alcune idee di miglioramento per le future versioni dell'algoritmo e per estendere il progetto ClinQC ad altre anatomie e modalitá di imaging.

I risultati simili ottenuti per i due metodi a confronto dimostrano che ClinQC é in grado di dare un'immediata valutazione della qualitá del dispositivo utilizzando le immagini dei pazienti e, in particolare, fornisce affidabili e sensibili parametri di qualitá dei sistemi di diagnostica per immagini a raggi X, che hanno lo stesso significato fisico e simili fluttuazioni relative degli indicatori di qualitá utilizzati nel metodo standard di riferimento QClight. Si puó concludere che l'algoritmo ClinQC potrebbe giá essere applicato nella pratica clinica, con il supporto iniziale del controllo di qualitá settimanale QClight. Un vero periodo di prova e confronto tra i due metodi, servirá anche come guida per effettuare gli aggiustamenti necessari all'algoritmo finché la versione finale non venga installata e stabilmente utilizzata nella pratica clinica.

Contents

List of Figures

List of Acronyms

AEC Automatic Exposure Control

CDF Cumulative Distribution Function

ClinQC Clinical images-based Quality Control

CoV Coefficient of Variation

CT Computed Tomography

DAP Dose Area Product

DICOM Digital Imaging and Communications in Medicine

ECMP European Congress of Medical Physics

FMM Fast Marching Method

FOV Field Of View

JOP Young Researcher Prize

K-S Kolmogorov-Smirnov

kVp peak kilovoltage

LAT Lateral chest X-ray projection

LUMC Leiden University Medical Center

LUT Look-Up Table

NPS Noise Power Spectrum

NVKF Dutch Society for Medical Physics

PA Postero-Anterior chest X-ray projection

PACS Picture Archiving Communication System

QC quality control

QClight light (phantom-based) Quality Control

REX Reached Exposure

ROC Receiver Operating Characteristic

ROI Region Of Interest

SDD Source-DAP-meter Distance

SID Source-Image Distance

SNR Signal to Noise Ratio

Chapter 1

Introduction

This Master's thesis is the result of my six months internship within the group of Medical Physics at the Radiology Department of Leiden University Medical Center (LUMC) in The Netherlands [1]. LUMC is a modern university medical center for research, education and patient care with a high quality profile and a strong scientific orientation, that employs around 7000 people (Fig. 1.1). I had the chance

Figure 1.1: The main building entrance of LUMC.

to work at LUMC from March until August 2016 as one of the students that won the *Erasmus+ : Mobility for Traineeship* competition [2] with the ClinQC project

(Clinical images-based Quality Control).
This project started at LUMC in 2015 after some initial ideas by the two medical physicists in training J. Chiel den Harder and Pieternel van der Tol, resulting in an algorithm that was further developed under the lead of the medical physicist Wouter J. H. Veldkamp, in the group of Medical Physics of LUMC headed by Koos Geleijns. Their aim was to perform a quality control (QC) of an X-ray imaging system without the use of routine phantoms, instead performing measurements of dose and image quality on the large number of clinical images of patient studies acquired with the system everyday. The use of clinical image quality with the purpose to early detect any quality regression of diagnostic X-ray machines is an innovative approach.
I entered the project under the lead of the previously mentioned medical physicists and of Berend C. Stoel (associate professor at the Division of Image Processing, Radiology Department, LUMC), because there was the need of a student to do additional analysis to show the benefit of implementing this QC algorithm in the daily clinical practice of the imaging systems at Radiology Department. In fact, the goal of my internship and the aim of this Thesis was to validate the existing algorithm for the QCs of the diagnostic X-ray imaging device and to invent or develop new measures of the quality of the system based on the planar chest radiograph of the patients. More details about the starting point of the ClinQC project and my role in it are explained at the end of Chapter 2.
The Thesis is structured in three main subjects: Dose study (Chapter 4), Noise study (Chapter 3) and Homogeneity study (Chapter 5). Each Chapter is divided in three main Sections: Methods, Results and Discussion. The final Chapter, Discussion and Conclusions (Chapter 6), will give an overview of the ClinQC project achievement and an outlook for future steps. The focus will mostly be on the Noise study (Chapter 3), the ClinQC algorithm validation and its improvements, since I mainly worked on this topic during my internship. The Homogeneity Chapter (Chapter 5) contains a pilot study that still needs some improvements and a concrete validation before it can be applied in clinical practice.
Before going through these main topics, the specific properties of the analysed chest X-ray imaging system will be discussed in Section 2.1 and the currently existing procedure to measure its quality and stability using phantom images in Section 2.2.

Chapter 2

Background

2.1 A typical chest X-ray imaging system

The chest X-ray is one of the most commonly performed diagnostic X-ray examinations; it evaluates the lungs, heart and chest wall. The basic equipment typically used for chest X-rays consists of an X-ray tube producing a divergent beam and a wall-mounted bucky system (Fig. 2.1), containing the digital flat panel detector, the anti-scatter grid and the ionization chambers of the AEC [3]. In this kind of examination the X-ray tube is positioned at 2.00 m Source-Image Distance (SID) from the detector, this can vary for other examined anatomies. Two views of the chest are taken, one from the back where the patient stands with hands on hips and chest pressed against the image plate, Postero-Anterior chest X-ray projection (PA), and the other from the side of the body, Lateral chest X-ray projection (LAT), where the patient stands against the image recording plate holding the bucky. In this thesis only chest PA projections will be analysed.
In the next paragraphs some of the components of the imaging system matter of interest for this thesis will be illustrated, along with the patient acquisition protocol and the output image format and processing.

Figure 2.1: A chest X-ray imaging system at LUMC.

Components

Anti-scatter grid The anti-scatter grid is the mobile device that is placed between the patient and the detector and that reduces the arrival of scattered radiation to the detector, decreasing the noise and enhancing the contrast in the X-ray image. The grid is designed with a series of alternating strips of lead and air that are angled to match the divergence of the X-ray beam. The primary beam radiation passes through the space between the lead strips only if it travels parallel to them, but scattered radiation which deviates from the divergent beam encounters the lead strips at a different angle, defined by the grid focal distance, and is attenuated from the beam. This device is useful in examinations where a large quantity of scattered radiation is created, with large tissue thicknesses and high peak kilovoltage (kVp). The employment of a grid requires a greater exposure to the patient since the primary beam is also attenuated by the lead strips [4].

AEC The Automatic Exposure Control (AEC) has the purpose to deliver consistent, reproducible exposures across a wide range of anatomical thicknesses and kVp [4]. On the detector there are three ionization chambers that count the number of photons that pass through the patient in three positions of interest: the spine in the center, and the two lung regions left and right (Fig. 2.2). The manufacturer

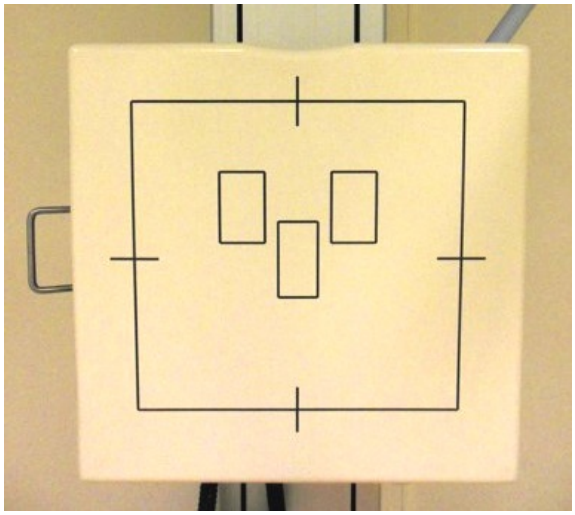

Figure 2.2: The three ionization chambers of the AEC on an LUMC bucky system.

does not share its complete knowledge about the functioning of this device, but what is known is that the results of the three chambers are combined in a unique output that allows to stop the X-ray exposure when the detector receives the desired dose in at least one of those three areas. For example, if a patient has an implant like a pacemaker, the sensor behind it will receive less photons than the other two and the AEC, according only to that output, would increase the exposure time, but the contribution of the other two ionization chambers will stop the source after a shorter exposure time.

Anode heel effect The anode heel effect appears in X-ray images as a gradient in the pixel values, due to the reduction of the intensity of the X-ray beam towards the anode side of the inner geometry of the tube [4]. This effect could change in appearance when the anode deteriorates over time, and this might produce in the images a more evident gradient in the grey values. This is an information about the

quality of the imaging system that can be retrieved from phantom images performing a QC (Section 2.2).

Acquisition protocol

The acquisition protocol is set up in order to obtain the optimal image quality for a diagnosis with the lowest patient dose. For an average weighted adult patient the typical acquisition protocol for a PA chest X-ray examination is:

- SID = 2.00 m
- kVp = 133 kV
  (gives the optimal contrast between the different body tissues and organs)
- AEC = ON
  (automatically corrects for different body thicknesses)
- image post-processing = ON

DICOM images and digital processing

The Digital Imaging and Communications in Medicine (DICOM) is the standard for handling, storing, printing, and transmitting information in medical imaging. The DICOM images are characterised by a header of DICOM tags that contain all the available technical data regarding the imaging system and the acquisition together with patient information [5].
The X-ray raw image that is captured by the detector sensors is always processed. This first image processing is the collection of operations and filtering that are applied on the digital images just after their acquisition, it is usually called pre-processing and compensates for different gain of different detector areas. This is an automated algorithm implemented by the manufacturer and is not accessible by the user: it can be said that the pre-processing is a black-box to the user.
Other operations and filtering, such as noise reduction in the background, are chosen by the user when selecting a chest X-ray protocol and can be classified as post-processing. Then there is additional post-processing, that can be adapted by the user by applying the appropriate parameter settings in the acquisition protocol and

that controls the Look-Up Table (LUT) and filtering like edge-enhancement. For patient examinations this post-processing is always used, but for some analysis with phantoms it can be useful to set it OFF switching the acquisition protocol. The post-processing is defined based on the clinicians experience with the aim to improve the diagnostic quality of the image.

2.2 QClight: phantom-based quality control

The quality control of a diagnostic X-ray imaging system comprises monitoring and evaluation of all characteristics of performance that can be measured and controlled [6]. The QC should not only measure the technical stability of the system, but also the dose given to the patients and the resulting image quality, that are relevant parameters for radiation protection and for a correct diagnosis made by the radiologist. Nowadays the most common way to measure these features is the employment of phantoms that are particularly designed to produce the same global attenuation of the X-ray beam intensity as a patient body [7].
QClight is the name of the approach and the phantom that are currently used at LUMC for weekly QCs in planar X-ray systems. QClight is a 2 mm thick copper plate phantom that has to be placed in front of the beam exit passage from the tube (Fig. 2.3). Each monday morning, before the first patient, the radiographer

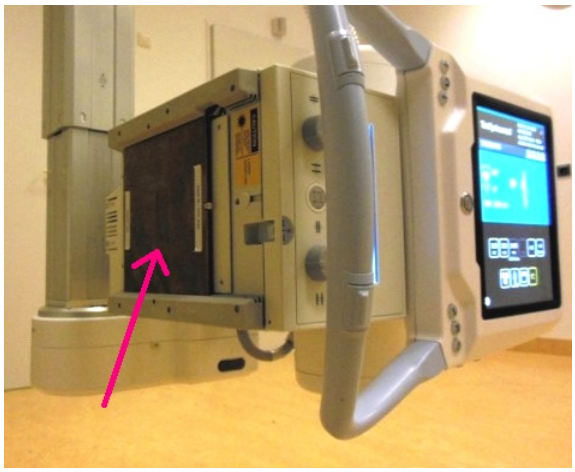

Figure 2.3: The QClight phantom is positioned for image acquisition.

places the QClight uniform phantom in this position and acquires with a specific protocol three images, activating the three ionization chambers of the AEC one at a time: only with three images it is possible to see if the three devices are working in the same consistent way to produce the desired combined AEC result (Section 2.1). The acquisition protocol of QClight phantom images is defined by some rules. The most important for our aim are:

- SID = 1.50 m
  (this distance is standardized for all the QClight acquisition protocols of LUMC: it is the same for all the X-ray rooms where different body anatomies are imaged)
- kVp = 133 kV
  (this is the same kVp used for adult patients who come for a chest X-ray examination)
- AEC = ON (in only one ionization chamber at a time)
- image post-processing = OFF

From the acquired images it is possible to retrieve all the necessary information that allow to assess a QC of the X-ray imaging system, including some dose parameters and image quality figures that are directly related to the system performances.

Dose The main dose quantities [8] are available in the DICOM tags of the QClight images: exposure, peak kilovoltage (kVp), Reached Exposure (REX) and Dose Area Product (DAP) for example. It is also possible to derive some other dose parameters to gain more complex information, combining different existing quantities given from the manufacturer of the X-ray imaging system (Chapter 4).
The constancy of all these dose parameters within their conventionally accepted ranges is investigated each monday with the QClight analysis; then, if the QC registers some differences from the usual behaviour of the system, an alarm is produced and the medical physicist job is to find out what are the causes and a possible solution.

Noise The noise in an image is the stochastic variation in the signal that compromises the image quality. The presence of the noise in X-ray images is due to three major causes. First, there is a statistical component, namely quantum noise, inherent to the fluctuations of the electromagnetic field that make the noise in the image depend on the intensity of the X-ray source: if the exposure (mAs) is increased the noise becomes higher, but the signal increases faster with increasing mAs, so the noise-to-signal ratio decreases, and in medical imaging this translates in giving higher dose to the patients. Second, all the electronic components of the system, mainly the digital detector, generate noise and interferences in their outputs that appear in the image as electronic noise. Then, the noise is due for large part to the scattered radiation that results from the interaction between the X-ray primary beam and patient tissue. A QC that includes the monitoring of some image quality figures, like the noise, gives the chance to check both the detector performances, together with all the electronic parts of the system, and the patients exposure [9].
The QClight images are homogeneous, so their pixel grey values fluctuations directly represent the noise. Selecting a Region Of Interest (ROI) of 30 % of the Field Of View (FOV) in the center, for example, makes it possible to quantify the amount of noise as the standard deviation of the pixel values that belong to this ROI. Since the system deals with an AEC that may change the exposure time for each QClight image that is acquired every week, the average grey value of the QClight image could be varying among all the acquisitions, giving to the noise measure a dependence to the image intensity that can easily be corrected. So the normalized QClight noise indicator is defined as

$$Noise_{\text{ QClight}} = \frac{\sigma_{ROI}}{\mu_{ROI}} = \frac{\sqrt{\frac{1}{NM-1}\sum_{i=1}^{N}\sum_{j=1}^{M}(p(i,j)-\mu)^2}}{\mu}, \tag{2.1}$$

where NM is the number of pixels $p(i,j)$ inside the selected ROI of the image, σ is the standard deviation of their grey values and μ is the average

$$\mu_{ROI} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}p(i,j). \tag{2.2}$$

Homogeneity Another image quality figure that is linked to the system performances is the image homogeneity. It reflects in particular the uniformity of the response of the detector. The QClight image is expected to be homogeneous as the phantom attenuation is, except for the inhomogeneities due to the detector malfunctioning or non-uniform response and inhomogeneities in the X-ray beam. There are many formulas to measure this quantity on an image of a uniform phantom, the one used at LUMC takes into account five square ROIs (R_1, R_2, ... , R_5) of width approximately 20 % of the FOV, one in the center of the image and the others close to the four corners. The average pixel value μ in each one of these five ROIs is computed using Eq. 2.2. Then the inhomogeneity is defined as:

$$Inhomogeneity\ (\%) = \frac{max\ (\mu_{R_1}, \mu_{R_2}, ..., \mu_{R_5}) - min\ (\mu_{R_1}, \mu_{R_2}, ..., \mu_{R_5})}{mean\ (\mu_{R_1}, \mu_{R_2}, ..., \mu_{R_5})}\%. \quad (2.3)$$

2.3 ClinQC: clinical images-based quality control

The long term goal of the ClinQC project will be the replacement of phantom-based QCs with measurements based directly on clinical images that will give a complete overview of the efficiency of the X-ray imaging devices. With the current imaging system for the chest, there is a need of a weekly standardized procedure to monitor its constancy using phantoms, i.e. the QClight analysis, but with other X-ray imaging modalities the QClight is not performed. This method required the training of the radiographers for the phantom acquisition procedure and 10 minutes per week to be performed. That is why it seems simpler and quicker to use a clinical images-based approach measuring the stability of the X-ray imaging device based on clinical images. The ClinQC method enables more frequent QCs, even for X-ray anatomies and imaging modalities where the QC measurements are not available. In addition, with this innovative approach, it is also possible to enter the universe of big data, using the large statistics of hospitals made by thousands of patient images: at LUMC approximately 50 patients/day come for a chest X-ray examination, which makes a total amount of 250 patients/week, 1000 patients/month and 12 000 patients/year.
In 2015 at LUMC a new clinical images-based QC was developed: the ClinQC algorithm now monitors different dose parameters and the noise of the imaging system from patient images, while the homogeneity measure is still in progress. With these

measurements on each new clinical image, the ClinQC algorithm now provides a significant knowledge about the stability of the system, and it can work side by side with the QClight method to validate the ClinQC tool.
The aim of this Thesis is to show all the steps of the ClinQC algorithm validation process that I implemented during my internship at LUMC and an important example of how the Noise and Exposure measurements would work if applied in clinical practice.
After the first steps of the ClinQC project, it has been presented in two important occasions. In 2016 J. Chiel den Harder gave a talk about the ClinQC project and received the Young Researcher Prize (JOP) for the best scientific presentation of the Dutch Society for Medical Physics (NVKF). In the same year Pieternel van der Tol presented the ClinQC project at the 1^{st} European Congress of Medical Physics (ECMP) in Athens, Greece [10]. Moreover the work shown in this Thesis will hopefully be published as a scientific paper in 2017.

Chapter 3

Noise study

3.1 Methods

In this Section all the methods adopted and developed to implement and validate the ClinQC noise extraction algorithm are illustrated.

3.1.1 ClinQC algorithm: noise extraction from clinical images

In 2015 at LUMC the ClinQC algorithm was invented [11], which extracts the noise directly from chest patient images in two main steps[1].

1. **Subtraction** from the chest *original image* (Fig. 3.1) of a *smoothed* version of the same image (Fig. 3.2). The smoothing is obtained with a Gaussian blurring filter with a kernel of 7×7 pixel size (pixelsize = 0.12×0.12 mm), shaped as a Gaussian distribution defined with $\sigma = 0.6$ pixel. The smoothed image is a version of the original image where the noise is partially removed: so it contains mainly the low spatial frequencies, which represent the anatomical signal. The output of this subtraction is the ClinQC *high spatial frequencies map* (Fig. 3.3): this image contains mainly noise but also some body signal, in the example the lung region can clearly be distinguished from the abdomen

[1] A similar method for noise extraction is applied to Computed Tomography (CT) [12].

region. To measure the noise directly on this first image would result in a significant patient dependent noise parameter, and the aim of the ClinQC project is to find patient-independent measurements of the quality and stability of the system that are not affected by patient variability.

2. **Normalization** of the ClinQC *high spatial frequencies map* by the *smoothed low frequencies image* is then required for two different reasons. First, to remove the dependence to the different image intensities that vary among all the patients examinations due to the presence of the AEC device: for the same purpose, in the QClight noise measure (Section 2.2) the normalization is represented in Eq. 2.1 by the ratio of each image to its average pixel value μ. Second, the image normalization is implemented also to try to remove the evident remaining body signal from the noise map. From the example in Fig. 3.4 it is clear that the new obtained ClinQC *normalized noise map* contains less patient anatomical signal than the output image after step 1 (Fig. 3.3), in fact only the main patient external edges can be seen.

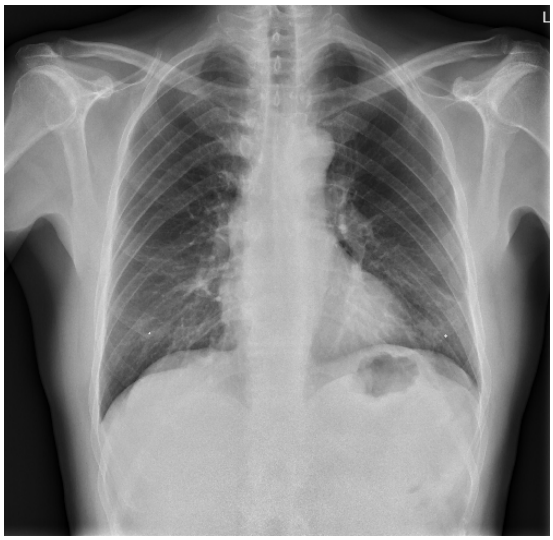

Figure 3.1: Original image.

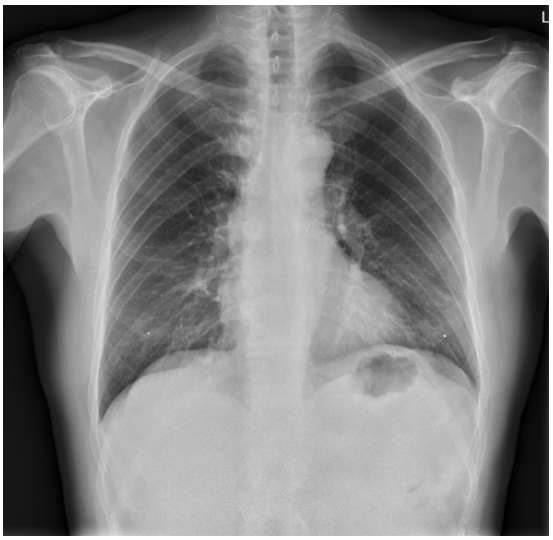

Figure 3.2: Smoothed image.

The final step to extract the noise from clinical images using the ClinQC algorithm consists in drawing a big ROI (Fig. 3.5) in the middle of the ClinQC *normalized noise image* of size 30% of the FOV, the same used in the QClight method. In this

Figure 3.3: High spatial frequencies image.

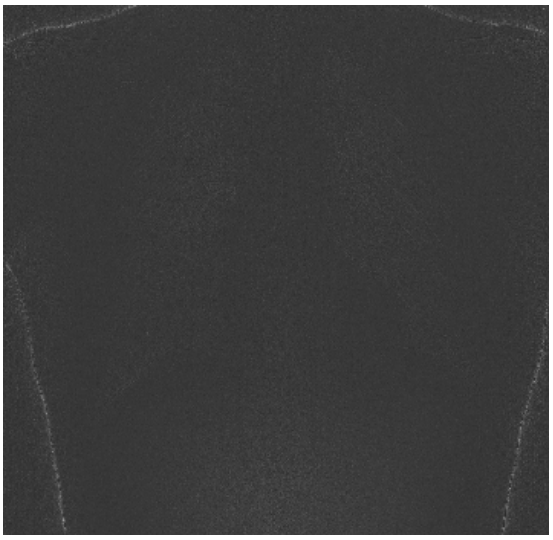

Figure 3.4: Normalized high spatial frequencies image.

way, the ClinQC *normalized noise value* can be computed as the standard deviation of the pixels $p(i, j)$ inside the defined ROI:

$$Noise_{\text{ ClinQC}} = \sigma_{ROI} = \sqrt{\frac{1}{NM-1}\sum_{i=1}^{N}\sum_{j=1}^{M}(p(i,j)-\mu)^2}, \tag{3.1}$$

where NM is the number of pixels inside the selected ROI of the ClinQC normalized noise image, σ and μ are, respectively, the standard deviation and the average of its grey values (Eq. 2.2).

Please note that Eq. 3.1 is different from the one used in the QClight method (Eq. 2.1), but the aim and the meaning are the same: the normalization here is computed pixel-by-pixel on the images at step 2, while in the QClight method there are no arithmetic operations in the phantom image and the normalization is a simple ratio of σ and μ of the pixels in the ROI of the QClight image. In this way the phantom based QC and the clinical images-based QC can be compared: the normalized QClight noise indicator has the same meaning as the ClinQC normalized noise value.

In order to perform a periodical QC of the system that can investigate the stability of the noise levels, all the ClinQC normalized noise values computed for the

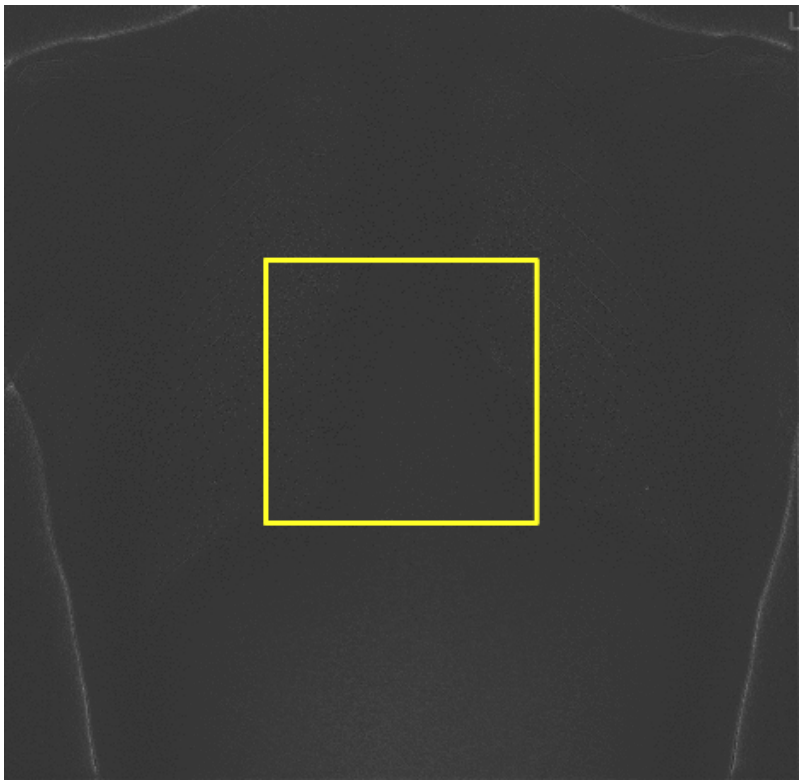

Figure 3.5: The ROI selected for the noise measurement on the ClinQC normalized noise image.

desired number of images extracted from the Picture Archiving Communication System (PACS) can be represented in a time line. Then, it is necessary to define the acceptance levels for the mean and the spread of the noise values (Section 3.2.1.2) and implement an automated algorithm that will check regularly whether the obtained values are acceptable or not. In this Thesis, the acceptance level for the mean is chosen as the average ClinQC noise value of all the images extracted from two weeks after the last calibration or maintenance of the imaging system. The fluctuations of the noise values that appear in the stability diagram usually fall within the limits of two times the standard deviation. However, a more accurate statistical analysis is required to understand how many patient images are needed, before realizing that something in the system has changed or is still changing (Section 3.1.3.2). A validation is needed to know if the ClinQC noise values are sensitive to a change in the quality of the imaging system (Section 3.1.3.2).

3.1.1.1 Properties of the ClinQC extracted noise images

The study of the effect of the ClinQC algorithm on patient images requires to look at the properties of the extracted noise images at each step of the algorithm (Section 3.1.1), before and after the normalization. Two examples of images that the ClinQC algorithm extracts are shown in Fig. 3.3 and Fig. 3.4. The method which has been chosen to compare the properties of these two images is to draw a simple but computationally slow diagram. The so called *high-low frequencies plot* (Fig. 3.11). As the name implies, this plot represents the relationship between the pixel values in the noise map (high spatial frequencies, *y-axis*) as a function of the pixel values that belong to the smoothed image (low spatial frequencies, *x-axis*). From the relationship that appears in the plot it is possible to understand whether or not the noise in each pixel has a dependence on the anatomical signal.
Since in each chest X-ray image there are around 10 thousand pixels, to plot this relationship for each pair of twin pixels, belonging respectively to the noise map and the smoothed image, requires a lot of computation time. If a correlation appears and more analyses are required to study the plot (like fitting or clustering procedures), to perform calculations on this diagram becomes a real challenge. It is easier to work directly with the *histogram* of the noise images, the distribution of the grey values in the high frequencies image or the normalized high frequencies image (the ClinQC noise maps), which is exactly the vertical distribution of the *high-low frequencies plot* along the *y-axis*.
In Section 3.2.1.1 an example of *high-low frequencies plot* and the correspondent noise image *histogram* will be shown for the ClinQC normalized noise map. An image segmentation based on the *histogram* [13] will be implemented to identify particular regions in the ClinQC *normalized noise map* where the anatomy of the patient is still visible after the normalization.
Lastly, the same *high-low frequencies plot* and correspondent *histogram* will be computed for the central ROI cropped from the ClinQC normalized noise map, to prove that the remaining patient signal it contains is not significant.

3.1.2 The ClinQC algorithm - alternative versions

The algorithm that has been used and validated in this Thesis work is the one introduced in Section 3.1.1.
In this Section new alternative versions of the ClinQC algorithm will be presented giving an outlook for the future possible improvements and extensions of the project.

3.1.2.1 Grid sampling approach

The first improvement that can be included in the ClinQC noise extraction algorithm updates a more accurate statistical analysis, that has been used also for low contrast detectability studies [14].
In the original ClinQC algorithm the final output is the *normalized noise map*, and the ClinQC noise value is the standard deviation of pixels in the selected ROI of this image. The proposed approach to modify the algorithm enters at this point as a third step:

Step 3) perform a grid sampling inside the original ROI of the ClinQC *normalized noise map*, selecting a grid of sub-ROIs (an example is shown in Fig. 3.6). For each new sub-ROI measure the ClinQC noise value as the standard deviation of its pixel values. Then compute the distribution of these many obtained ClinQC noise values. The final ClinQC-alternative noise value will be the median value of this distribution.

This is a better representation of the true value of the noise that the ClinQC algorithm should derive. In the original unmodified algorithm this final value corresponds to the *mean* value of the noise distribution, while in this alternative version it is proposed to use the *median* value. The mean value would be a good estimate of the noise distribution only in case the original ROI is uniform and does not contain any different structures, and if the noise distribution is symmetrical. Looking at the chest X-ray examinations, the selected ROI over the chest and abdomen of the patient often contains implants, pacemakers, catheters and internal or external electrodes of different medical instruments. So the ClinQC original noise extraction algorithm might misclassify some of these structures (Fig. 3.24 and Fig. 3.25), identifying them as noise because they have sharp edges (high spatial frequencies).

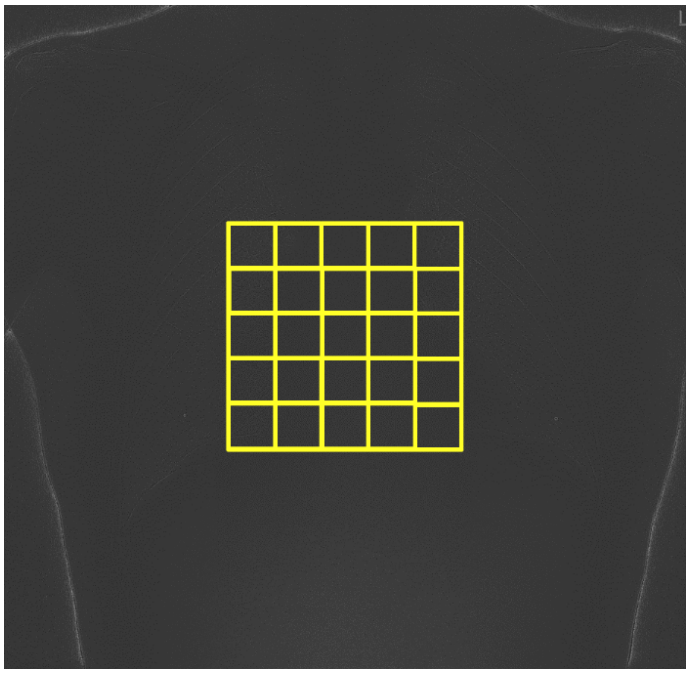

Figure 3.6: An example of the sampling grid that can be used to improve the ClinQC noise extraction algorithm.

While this sampling grid approach may help to reduce the influence of those outliers of the noise distribution, that can be due also to patients who have nodules with well defined edges.

3.1.2.2 ClinQC applied to mammography

To extend the use of this noise extraction algorithm to different X-ray anatomies and imaging modalities will be a challenge in the future of the ClinQC project. A first trial has been performed on a small dataset of *planar digital mammograms* and on a *stack of breast tomosynthesis pictures that are projected onto a single reconstructed view*.

The algorithm has been adapted only in the final step of selection of the ROI where to perform the noise measure, in order to recognize in which side of the image the breast was positioned (laterality), and to take as ROI a rectangular region smaller than the FOV. The background pixel values inside this ROI are equal to 0 from both the digital imaging systems, so they can easily be ignored in the measurement of the

standard deviation. This was a fast way to implement the new ClinQC alternative version for a preliminary test, but the ROI includes also parts of the breast edges and skin where the noise might have different dependence on the body signal than inside the breast (the same happening for chest X-ray images), so the most correct way to implement this alternative version would be to select automatically a region that contains only the inner breast tissue and that has the same amount of pixels for all the analysed patient images.
Mammograms and tomosynthesis reconstructed projections can be generated from different exposures, but also different kVp, depending on the patients breast thicknesses. This can introduce another source of variability on the noise measurement, that is not dependent on the imaging system quality and that for chest X-ray images was irrelevant since only patient images acquired with 133 kV were used in this study. In this preliminary test all the images generated from both the imaging modalities mentioned before at different exposure levels and kVp were analysed together. The preliminary results are displayed in Section 3.2.2.2.

3.1.3 Validation of the ClinQC algorithm

In this Section, all the Methods developed for the validation of the ClinQC noise extraction algorithm are presented. A validation is the process that tests the algorithm under different conditions to see if its results are reproducible and dependent on the quality of the imaging system.

3.1.3.1 Image deterioration study

The first Method that has been invented in order to simulate a decrease or an increase in the quality of the imaging system is the image deterioration study. As explained in the Introduction (Section 2.3) the chest X-ray imaging systems can be really stable over long time periods, in fact there are no examples in the last years of a slow or fast quality degradation that we can use to verify if the ClinQC algorithm can detect it. So the only way to test if the algorithm is able to detect a deterioration of the imaging system is to deteriorate the quality of the images that it produces.
Two symmetrical approaches are presented.

3.1.3.1.1 Blur To blur the images is a way to reduce their amount of noise, so the expected result of the test after blurring the images is to observe a decrease in their ClinQC noise values. If so, the decrease has to be significant to prove that the algorithm is sensitive to a change in the quality of the imaging system and in order to be recognized automatically with a detection algorithm. A reduction in the noise values could simulate an increase in the quality of the imaging system, that can occur if a new proper anti-scatter grid is installed for example (as can be observed in the examples in Section 3.2.1.3), or it can represent an increase in the exposure (mAs) given to the patients during the image acquisition. To verify if blurring the images is actually a way to simulate an increase of the exposure levels, a phantom study has been performed (Section 3.1.3.6).
The blurring study consists of smoothing at different levels different groups of images, to run the ClinQC algorithm and to plot all the results in a time line. The increasing amount of blur was obtained linearly changing the σ of the Gaussian blurring kernel used to smooth the images, from $\sigma = 0.5\ pixel$ to $\sigma = 2.05\ pixel$

(pixelsize = 0.12×0.12 mm). The first group consists of original unmodified images (instead of $\sigma = 0\ pixel$). The sizes of the Gaussian blurring kernels varied automatically with the values of σ, according to the law: $kernel\ size = 2 \cdot ceil(2\sigma) + 1\ (pixel)^2$. To study how the *average* noise value of the group of images changes, it is better not to use for every deterioration level the same group of images, but the approach used here was to randomly sub-sample different images every time. Instead, to study if and how the *spread* of the noise values changes, it is more meaningful to use the same group of images for each blurring level.

3.1.3.1.2 Gaussian noise The addition of noise to the images is done to observe an increasing trend in the ClinQC noise values, that can occur when the detector or the electronics of the system reach their end of life, or if the anti-scatter grid is not present into its slot during the image acquisition (as can be observed in the examples in Section 3.2.1.3). An increase in the noise levels can simulate also a decrease in the dose (mAs) given to the patients. Different shapes of noise distributions can be added to the images to achieve different results: in the simulation presented in this thesis the symmetric distribution of Gaussian noise has been chosen to approximate the combined effect of all these deterioration phenomena in the most simple case.
A different approach must be adopted in this case, since with this simulation we are changing something in the images that is exactly the quantity that we want to measure. So it would be a circular reasoning to add the same amount of noise to all the images and to measure if and how the ClinQC noise values change, since each image has a different initial level of noise. For example, two images A and B have different $\sigma_{noise\ A}$ and $\sigma_{noise\ B}$, if Gaussian noise with $\sigma_{noise\ ADD}$ is added to both images, the ClinQC algorithm would register only a meaningless shift in the two noise values, while their difference $|\sigma_{noise\ B} - \sigma_{noise\ A}|$ would stay the same, since the added noise is the same for both images. This inserts only a perturbation in the simulation result and does not give any new information on the extraction algorithm behaviour.
Instead, we don't want to interfere in the measurement, but we want to add noise

[2]ceil = function that rounds to the nearest integer greater than or equal to the number in parenthesis

only to simulate a controlled decrease in the quality of the imaging system. So the approach proposed here has the purpose to add different amounts of noise to the images according to their original estimated quality, and this is made possible by following three steps:

1) Estimate the original quality of each patient image by measuring its Signal to Noise Ratio (SNR). This can be made in two manners, the first one measures the SNR in decibel (dB) using the information from the entire image [15], the second one is more used in case the SNR of a single object or ROI in the image is measured:

$$SNR_{original\ image} = 10log_{10}\frac{\sigma^2_{original\ image}}{\sigma^2_{noise\ image}} \quad [dB] \tag{3.2}$$

$$SNR_{original\ image} = \frac{\mu_{original\ image\ (or\ ROI)}}{\sigma_{noise\ image\ (or\ ROI)}} \tag{3.3}$$

where the *noise image* is the ClinQC high frequencies map, so the noise image extracted at the first step of the algorithm.

2) Choose the desired levels of deteriorations of the estimated SNR that have to be achieved by adding noise to the clinical images: linearly decreased percentages of the SNR, for example from 2% to 30%.

3) Use the inverse formula of Eq. 3.2, or Eq. 3.3, to compute the variance of the Gaussian noise with 0 mean that has to be added to the clinical image, with an image processing tool (in this case MATLAB[3]), in order to achieve the desired deterioration level of the estimated SNR.

$$\sigma^2_{noise\ achieved} = \frac{\sigma^2_{original\ image}}{10^{\frac{SNR_{deteriorated}}{10}}} \quad [dB] \tag{3.4}$$

But the additional step that avoids interfering in the measurement, as said before, is to consider that each patient image has a different initial amount of noise. So the noise that has to be added to the image, in order to achieve the chosen level of deterioration of the SNR, must be reduced by the original amount of noise. Since these two amounts of noise are described by two distributions with respective standard deviations $\sigma_{noise\ added}$ and $\sigma_{original\ image}$, their addition

[3]The MathWorks Inc., Novi, MI, USA.

will result in a distribution with a standard deviation $\sigma_{noise\ achieved}$ that satisfies $\sigma^2_{noise\ achieved} = \sigma^2_{original\ image} + \sigma^2_{noise\ added}$ [16]. So the variance of the Gaussian noise with 0 mean that has to be added to the image is:

$$\sigma^2_{noise\ added} = \sigma^2_{noise\ achieved} - \sigma^2_{original\ image} \tag{3.5}$$

Also in this case, to study how the *average* noise value of the group of images changes, the approach is to random sub-sample different images for each deterioration level. Instead, to study if and how the *spread* of the noise values changes, the same group of images has to be used.

3.1.3.2 Statistical analysis with simulations

This Section presents a statistical analysis regarding the stability of the ClinQC noise values performed by implementing two different types of simulations. These simulations are designed in order to produce output curves that answer the question: *how many new clinical images have to be collected before an automated algorithm can detect a step or a trend in the noise values with a certain significance level?* So we want to assess how fast is the response of the ClinQC algorithm to the simulated deterioration of the system, and how soon will the algorithm be able to provide an alarm, based on the noise level of each new clinical image acquired.
The purpose of the Step simulation (Section 3.1.3.2.1) is to detect a sharp breaking in the system that translates in a sudden change in the noise values from one patient to another. This can happen, for example, when the anti-scatter grid is replaced or removed (as can be observed in the examples in Section 3.2.1.3). While the Trend simulation provides different examples on the detection of the long term increase or decrease in the noise values.

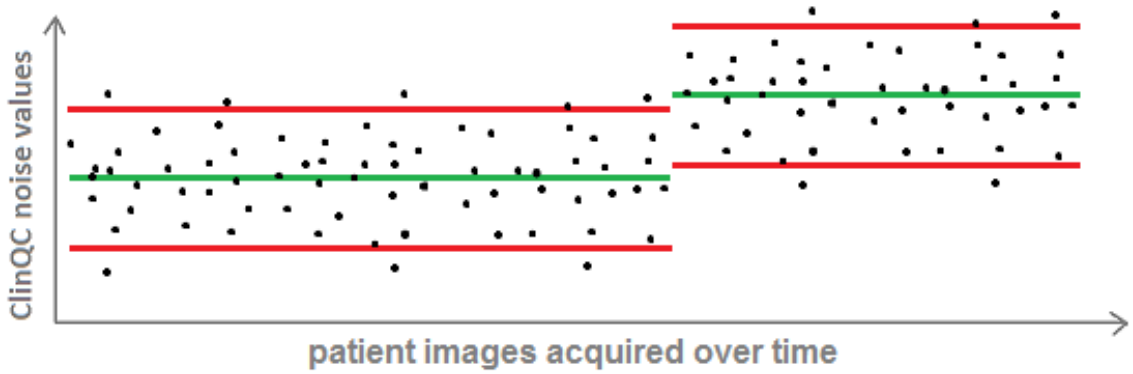

Figure 3.7: Graphic reproduction of how a step in the noise values would appear.

3.1.3.2.1 Step simulation The general idea is to simulate a step in the noise values (Fig. 3.7) with respect to the baseline (Section 3.2.1.2), and to understand how many clinical images are needed to detect it with a certain significance level. The following recipe explains the step simulation algorithm:
1) Find the parameters that describe the distribution of the ClinQC noise values

baseline dataset: μ, σ, skewness and kurtosis (Section 3.2.1.2).

2) Increase the baseline mean ($\mu_{baseline}$) with a different increasing factor at every loop.

(example: $\mu_{simulated} = \mu_{baseline} \cdot increasing\ factor$, where the increasing factors are numbers like 1.03, 1.15, ..., that indicate an increase of 3%, 15%, ...)

3) Use each new increased mean $\mu_{simulated}$, with the original other parameters that characterize the distribution of the ClinQC noise values baseline dataset σ, skewness and kurtosis, to simulate from the same designed distribution new groups of noise values of increasing sizes: for each loop (so for each new simulated mean) generate M (very high) new datasets of each size.

(example: $\mu_{simulated\ 1} = 0.0060$ (baseline mean increased by 3%)

1^{st} new dataset (size = 2 noise values) $\rightarrow$ [0.00596 , 0.00633] $\rightarrow$ $\times$M iterations

2^{nd} new dataset (size = 3 noise values) $\rightarrow$ [0.00591 , 0.00594 , 0.00621] $\rightarrow$ $\times$M iterations

...

N^{th} new dataset (size = N+1 noise values) $\rightarrow$ [0.00628 , 0.00605 , ..., 0.00594] $\rightarrow$ $\times$M iterations.)

4) For each one of the new N·M datasets of simulated groups of noise values of different sizes, perform an unpaired T-test to test the null hypothesis that the simulated datasets come from normal distributions with the same mean and equal but unknown variance as the baseline of the ClinQC noise values.

In some cases, especially when small sizes of the noise values groups are simulated, it is possible that the T-test cannot reject the null hypothesis based on so small datasets. That is why M iterations are required: in this way the rejection ratio of T-test for each one of the new dataset sizes is

$$Rejection\ Ratio = \frac{number\ of\ times\ the\ null\ hypothesis\ is\ rejected}{M\ total\ T-tests\ (iterations)}. \quad (3.6)$$

A plot of how the Rejection Ratio varies with the different sizes of the simulated datasets indicates the scoring of the T-test in detecting a change in the noise values, when a certain number of new images produces noise values distributed with a mean higher than the baseline. This results in a diagram with similar meaning and shape as a ROC [17].

5) Now if a threshold is chosen in the *y-axis* of the Rejection Ratio representation, for example at 99.5%, the first integer size of the simulated datasets that exceeds

the threshold in the rejection ratio curve is the minimum number of images that we need to say, with a high level of certainty, that the mean of the simulated noise values is actually higher than the baseline, and that the two distributions can be considered unequal.

6) If this analysis is done for each one of the increasing factors for the baseline mean, so for each different $\mu_{simulated}$, it is possible to find the minimum number of images to reject the null hypothesis in function of the baseline increase in percentage: and this is the curve that answers the original question (*how many new clinical images have to be collected before an automated algorithm can detect a step or a trend in the noise values with a certain significance level?*).

If the question is less general, in particular if the user wants to know how many images are needed to detect a step exactly like the one produced by a known effect like the anti-scatter grid removal or replacement (history examples shown in Section 3.2.1.3), the step simulation can be adapted for this purpose. In this specific case, there is no need to simulate many baseline increases, since the difference between $\mu_{baseline}$ and $\mu_{NO\ grid}$ or $\mu_{OLD\ grid}$ in percentage is already known (Section 3.2.1.3). So the Rejection Ratio with the proper threshold already tells the user how many images are needed to detect the expected step in a specific occasion.

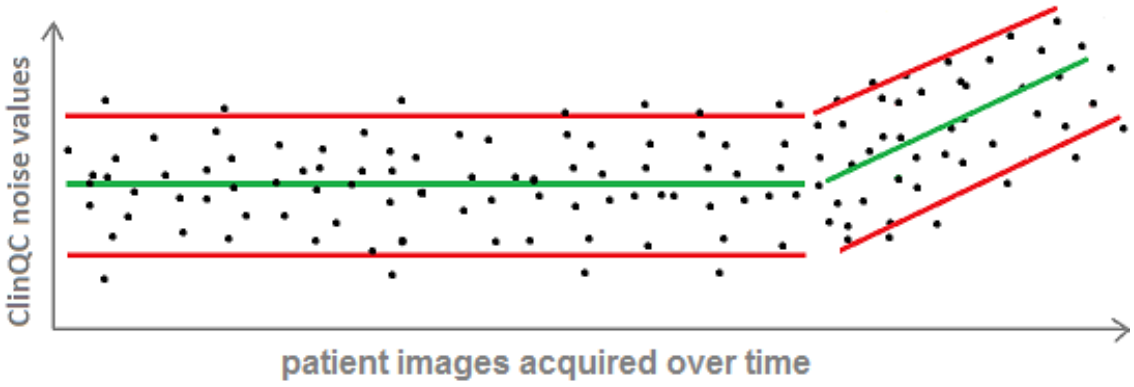

Figure 3.8: Graphic reproduction of how a trend in the noise values would appear.

3.1.3.2.2 Trend simulation The trend simulation, instead, generates a slope of noise values and uses a moving window to establish how soon the increasing trend

(Fig. 3.8) is detected.
1) Find the parameters that describe the distribution of the ClinQC noise values baseline dataset: μ, σ, skewness and kurtosis (Section 3.2.1.2).
2) Generate new baseline values from the same distribution described with the parameters found at step 1.
3) Generate new noise values from the same distribution arranged in a slope and align them to the baseline data points simulated at step 2. The number of simulated points represents a specific time line (like one week or one month of patients exams), while the slope reflects the strength of the simulated deterioration occurring in the imaging system.
4) Choose different sizes of a moving window, that can be dragged point-by-point on the simulated slope of noise values in order to perform series of unpaired T-test, between the average value of the simulated noise values that fall inside the moving window and the baseline distribution of the ClinQC noise values.
5) Each time the moving window of a chosen size is shifted, the T-test scoring is recorded and normalized by the total number of iterations. This results in the Rejection Ratio (Eq. 3.6) that can be plotted against the position in the slope when the alarm is produced.
6) Then a similar approach used in the step simulation (point 5) is adopted: a threshold of 99.5% in the Rejection Ratio plot specifies the percentage of increase of the baseline on the slope where the T-test has been able to reveal a change in the noise values in at least the 99.5% of the iterations.
7) If points 4, 5 and 6 are repeated for different sizes of the moving window, the final curve of the trend simulation can be obtained by plotting the percentage of increase of the baseline simulated by the slope, against the size of the window that was able to detect it. The most important difference between the two simulations presented in this Section is that the Trend simulation curve is dependent on the chosen time line and strength of the slope, but both simulations give curves that indicate the detectability of different conditions by using the ClinQC algorithm.

3.1.3.3 The ClinQC performance in clinical practice

How would the ClinQC noise algorithm work in clinical practice?
The stable history of the imaging system, monitored with the QClight phantom analysis, shows the device is in good health. During routine clinical use the anti-scatter grid was unintendedly inserted facing the wrong direction. This was detected during the QClight measurement the next Monday morning in July 2016. The QClight images acquired during the weekly QC were extremely inhomogeneous but symmetrical along the vertical axis. The noise measurement was lower than the average QClight baseline. The medical physicists hypothesized that the anti-scatter grid had been flipped, this was confirmed after a manual check on the imaging system.
But when exactly did it happen during the previous week? This very specific answer can't be given by a weekly QC as the QClight. So here is where the performance of the ClinQC algorithm has been tested on an actual problem of the imaging system. It has been possible to observe and study the effects of this human mistake using the ClinQC algorithm and this became an important part of its validation.
All the clinical images acquired with the X-ray imaging system during the two weeks when the problem was identified with the QClight analysis were extracted from the PACS. The ClinQC noise extraction algorithm was run on all the images producing values that fell within the range of acceptance defined by the baseline limits, while a clearly visible step down in the noise values and up in the exposure levels was observed in almost three complete days. Since the noise values after the step differ from the baseline of a certain percentage that has been measured in Section 3.2.5, according to the Step simulation output curve (Fig. 3.2.4.1), the minimum size of the moving window that can be used to detect the step in the noise values is defined. All these results and procedures are illustrated in Section 3.2.5.

3.1.3.4 Outlier analysis

The analysis of the outlier noise values produced by the ClinQC algorithm is useful to understand if there are specific categories of patients that are more susceptible to produce outliers. All the images acquired after the update of the anti-scatter grid by the end of 2014, extracted from the PACS during my thesis internship, were

the input of the ClinQC algorithm and their noise values were compared to the ClinQC baseline. First of all, a correlation between the ClinQC noise outliers and the exposure outliers was investigated. Second, the sex of the outliers was compared to the percentage of males and females undergoing a chest X-ray examination. It was not possible to compare the histogram of ages of the outliers to the distribution of ages of all the patients of one year, since it would have meant to extract more than 300 GB of additional images from the PACS archive to retrieve patient ages from their DICOM tags and draw the annual age distribution of patients in the examined imaging system. For the same reason it was not possible to study the presence in the noise outliers of patients with nodules or sharp objects in the chest area, since an automated search algorithm for these structures would have needed to run not only in the already available images but also in the annual dataset of images that was not possible to extract, in order to establish a statistical correlation.

3.1.3.5 Image Pyramids noise extraction algorithm comparison

The image Pyramids are a multi-scale representation of an image which is subject to repeated smoothing, sub-sampling and/or other manipulations. They are now widely used for image compression, object recognition and multi-scale analysis, but also for noise extraction from clinical images [18]. This last approach uses a Laplacian pyramid to decompose the chest images into many digressive spatial frequency sub-bands, in order to separate the high frequencies (mainly noise) from the low frequencies (mainly anatomical signal). A Gaussian pyramid is formed by an image G_0 that is subsequently filtered with average blurring kernels and downscaled many times (first line of the example Fig. 3.9). All the images of the G_i series (with $i = 0 \ldots N_{levels}$ of the Pyramid) can then be up-sampled back to the original size, observing an increasing smoothed appearance (second line, Fig. 3.9). The Laplacian Pyramid [19] stores the differences pixel-by-pixel of the images at each adjacent stage of the Gaussian Pyramid $L_i = G_i - G_{i-1}$, obtaining many residuals images (third line, Fig. 3.9). The first stage of the Laplacian Pyramid $L_1 = G_1 - G_0$ $_{(original\ image)}$ will show mainly the noise information, while at higher levels the noise will be gradually replaced by the low frequency coarse structures.

The ClinQC noise extraction algorithm and the Pyramids method are based on the same ideas, while the implementation is different. In fact to achieve the extraction of a noise map from a clinical image both methods combine a smoothing and a subtraction. The additional step that only the ClinQC algorithm includes is a normalization by the low frequencies image.
A comparison between these two methods has been performed using the clinical images forming the baseline dataset Section 3.2.1.2. The Laplacian Pyramid at the first stage was used as the ClinQC noise map taking a square ROI in the center and measuring a not normalized noise value but, to better compare the noise values produced by the two methods, the Laplacian Pyramid at the first stage L_1 has been also normalized by the corresponding low frequencies stage G_1. The results are illustrated in Section 3.2.7.

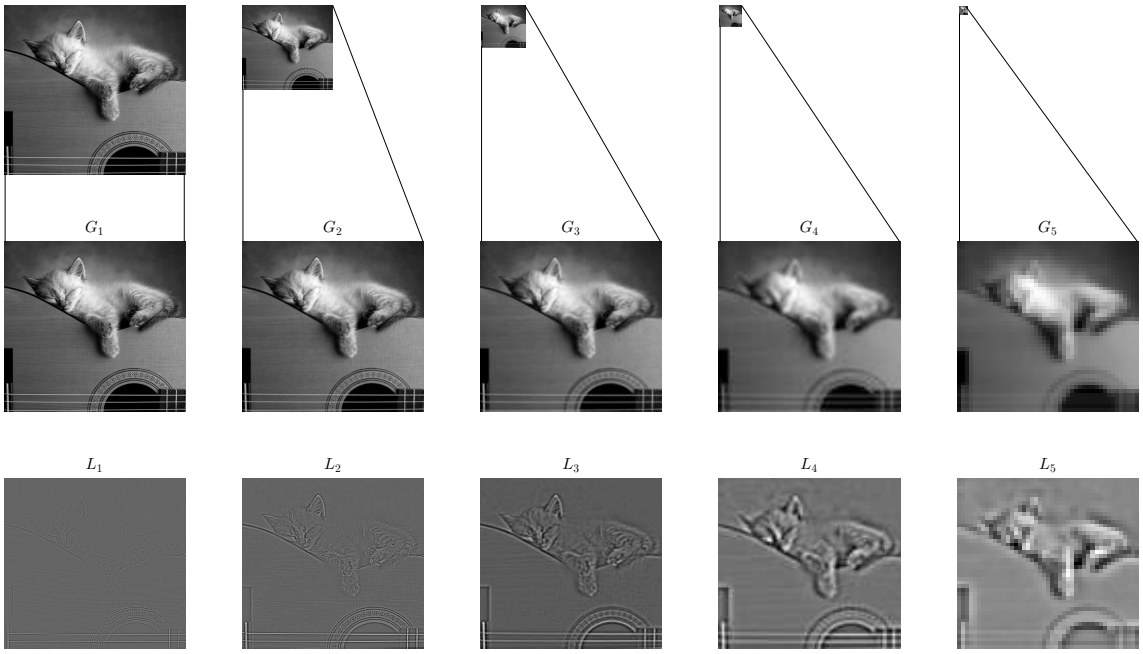

Figure 3.9: Example of Gaussian and Laplacian Pyramids.

3.1.3.6 Phantom comparisons

The ClinQC algorithm for noise extraction has been validated performing also a phantom comparison study. The two phantoms available at LUMC for this purpose were the QClight copper plate (Section 2.2) and RANDO[4] anthropomorphic chest phantom, representing the skull and trunk of a 175 cm tall male weighing 73.5 kg (Fig. 3.10).

In order to establish if the ClinQC is a reliable algorithm to extract and measure noise from clinical images, a correlation between this new patient-based approach and the QClight phantom-based method has been investigated: the patient-based technique was simulated by applying the ClinQC algorithm to RANDO anthropomorphic phantom images, acquired with the same protocol used for patients at different exposures (since images at different exposure levels of the same patient were not available). The QClight images were acquired within the same range of

[4]The Phantom Laboratory, Salem, NY, USA

Figure 3.10: RANDO anthropomorphic phantom.

mAs, using both the phantom and the patient acquisition protocols.

The QClight noise data, saved each Monday during the history of the imaging system with the actual setting, have been compared with the ClinQC noise value of one clinical image per week in the same time line. In this case, only the CoV of the noise values obtained with the two methods have been compared, while the correlation was not expected to be observed, since both the clinical and the phantom images are acquired at approximately constant exposures.

The final test that was performed using phantoms was to confirm the findings of the image deterioration study using blurring deterioration. In fact, the test was implemented in order to investigate if blurring the images is the correct way to simulate an increase of the mAs, in addition to a noise decrease that can be observed in Section 3.2.3.1. The same images of RANDO and QClight, previously acquired using, respectively, patient and phantom acquisition protocols at different exposures,

were used. The images acquired at 1.6 mAs were blurred applying filters with different increasing sigma of Gaussian blurring kernels. The resulting noise values were computed using the ClinQC algorithm for RANDO images, and with the original method for the QClight images. The dependency of the noise on the exposure and blur setting were compared by plotting the resulting noise values together in one figure.
The 2D Noise Power Spectrum (NPS) of the images obtained at different exposures and after each blurring deterioration are also displayed to compare the effect of mAs and blurring on the noise frequencies.

3.2 Results

In this Section all the results of the ClinQC noise study will be presented.

3.2.1 The ClinQC algorithm: noise extraction from clinical images

3.2.1.1 Properties of the ClinQC extracted noise images

In the *high-low frequencies plot* of the example shown in Fig. 3.11 we see three major structures: the central main cluster distributed around 0 and two other secondary clusters with different slopes.

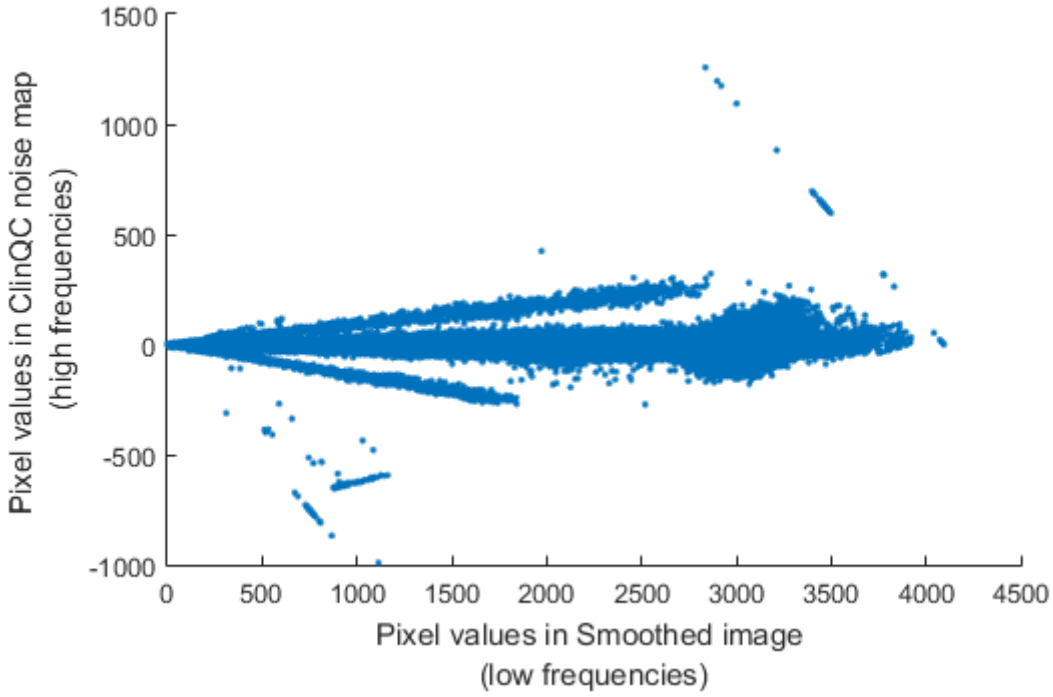

Figure 3.11: An example of *high-low frequencies plot* of the ClinQC high frequencies map, the output of the algorithm at the first step, before the normalization.

After the normalization the *high-low frequencies plot* changes significantly (Fig. 3.12): we can still see different clusters, but they are more constant and even

the secondary clusters don't correlate with the low spatial frequencies any more. The remaining offset of the secondary clusters in the *y-axis* is a consequence of the normalization of the slopes, and represents that there are regions of pixels in the normalized noise map with some higher or lower intensity than the average noise signal, which is distributed around 0. In those regions the dependence of the noise from the body signal, so the slopes in Fig. 3.11, is successfully removed by the normalization.

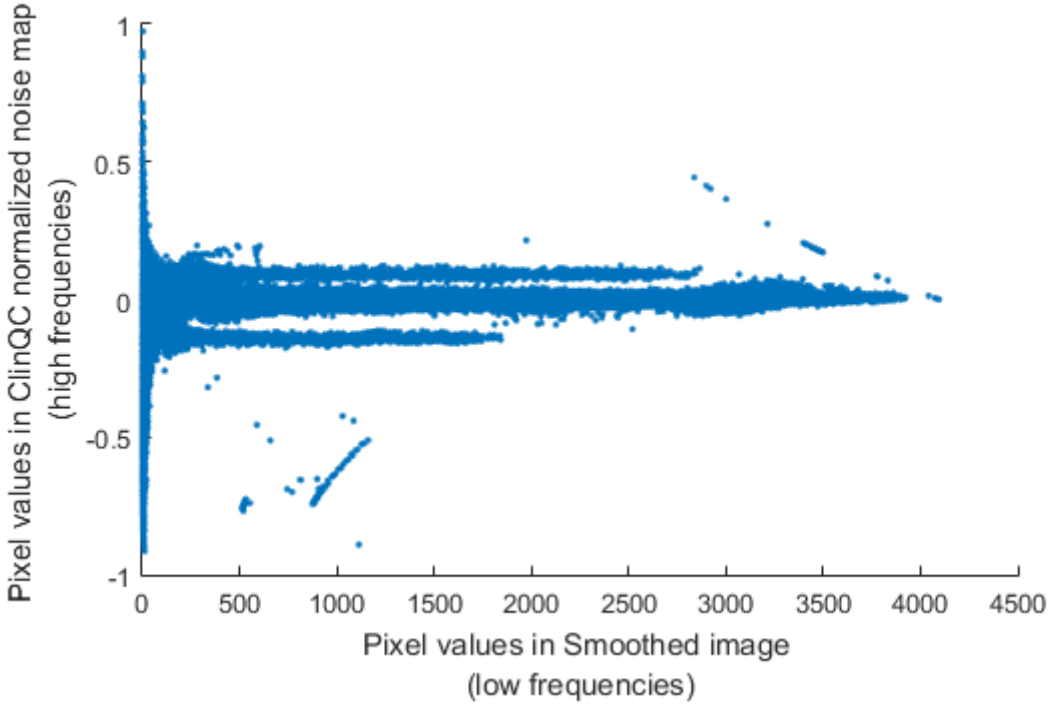

Figure 3.12: An example of *high-low frequencies plot* of the ClinQC normalized noise map, the final output of the algorithm after the second step of the normalization.

As described in the Methods (Section 3.1.1.1), the *histogram* of the ClinQC noise and normalized noise maps is a computationally faster tool that provides the same information as the *high-low frequencies plot*. Then, with a simple image segmentation based on the *histogram* of the noise maps, it is possible to identify the regions in the image where the noise was dependent on the anatomical signal and that now have a offset in their pixel intensities, so the regions of pixels that still belong to the

lateral clusters.

In Fig. 3.13 the histogram of the ClinQC normalized noise map can be compared

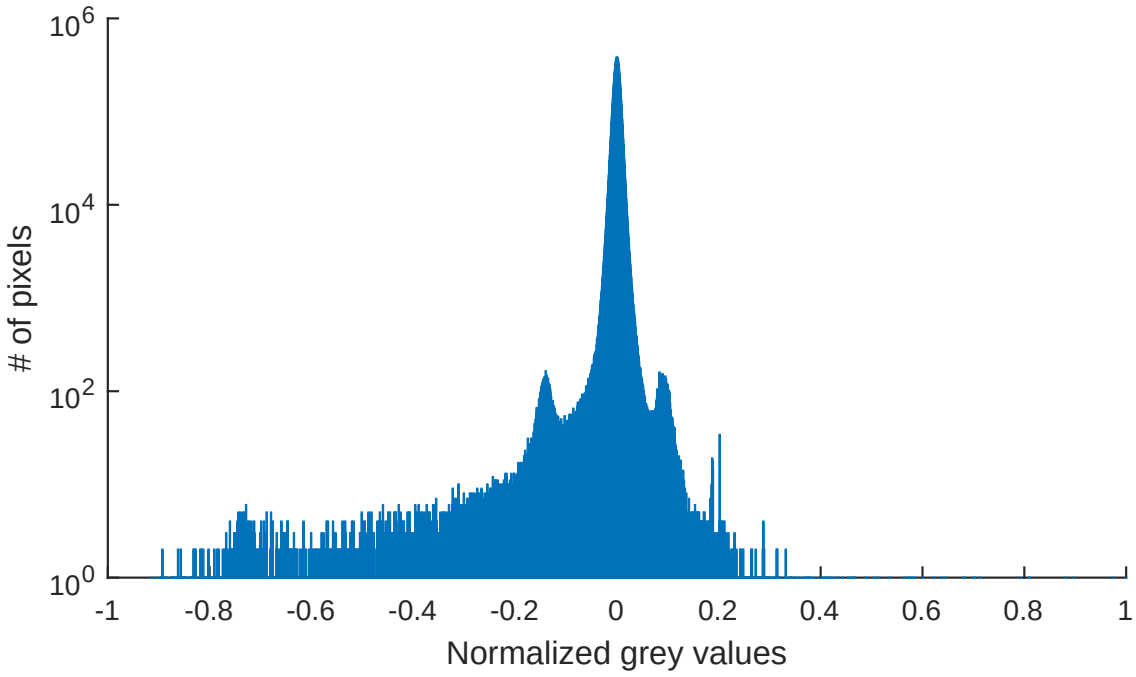

Figure 3.13: The *histogram* of the ClinQC normalized noise map. This corresponds to the *high-low frequencies plot* shown in Fig. 3.12.

to the *high-low frequencies plot* shown in Fig. 3.12. The histogram distribution reflects the vertical shape of the correspondent plot. The main additional information that the histogram can give, despite the diagram, is the number of pixels that show a certain correlation, the *y-scale* is logarithmic in Fig. 3.13: almost all the pixels belong to the central peak of the histogram, which is exactly the central cluster of the correspondent *high-low frequencies plot*, while the pixels in the lateral cluster are only a few compared to more than 10 million pixels in the whole image.

To select anatomy dependent noise and define where the lateral clusters of pixels are located in the ClinQC normalized noise map, a segmentation based on the histogram central peak has been performed. The result of the segmentation-based pixel selection is shown in Fig. 3.14, using two thresholds next to the central peak of the histogram: -0.1090 and 0.0625 normalized grey values. The binary image shows in white the few pixels belonging to the lateral peaks of the histogram, and in black

the majority of pixels coming from the central peak. The white pixels represent the remaining fingerprint of the patient external body edges and the annotations automatically burned on the image pixel values (upper right corner). The *high-low*

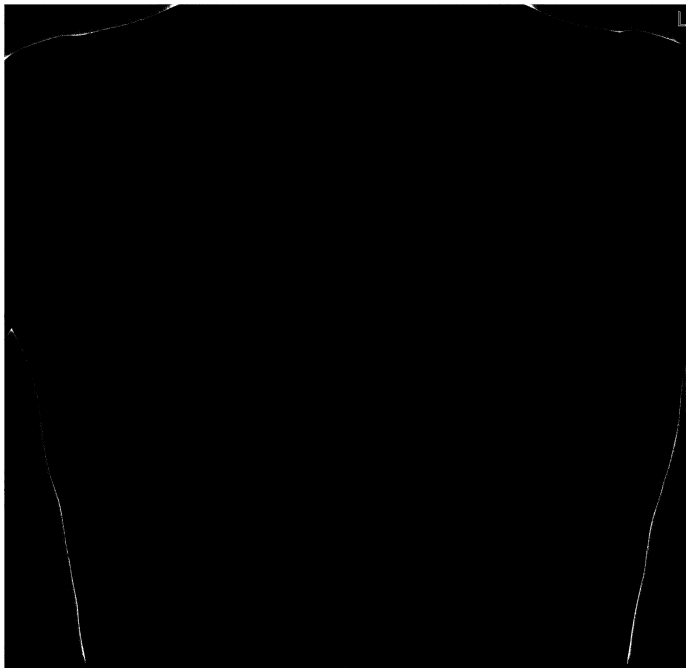

Figure 3.14: The result of the histogram-based segmentation performed on the ClinQC normalized noise image.

frequencies plot and the *histogram* of the central ROI cropped from the ClinQC normalized noise map are shown respectively in Fig. 3.15 and Fig. 3.16 in yellow colour, overlaid onto to the light blue data series belonging to the whole image, previously shown in Fig. 3.12 and Fig. 3.13. In both the equivalent examples it can be observed that the central ROI pixel values are distributed around 0 and belong to the central peak of the *histogram* without additional peaks. Taking an ROI of 30% of the FOV limits the number of pixels involved in the measurement, in fact the yellow data series in Fig. 3.15 and Fig. 3.16 do not cover respectively the central cluster and the central peak of the light blue data series just for this reason. In Section 6 an idea to better use all the pixels in the image is illustrated, for future reference.

These properties have been observed for around fifty clinical chest images, even if

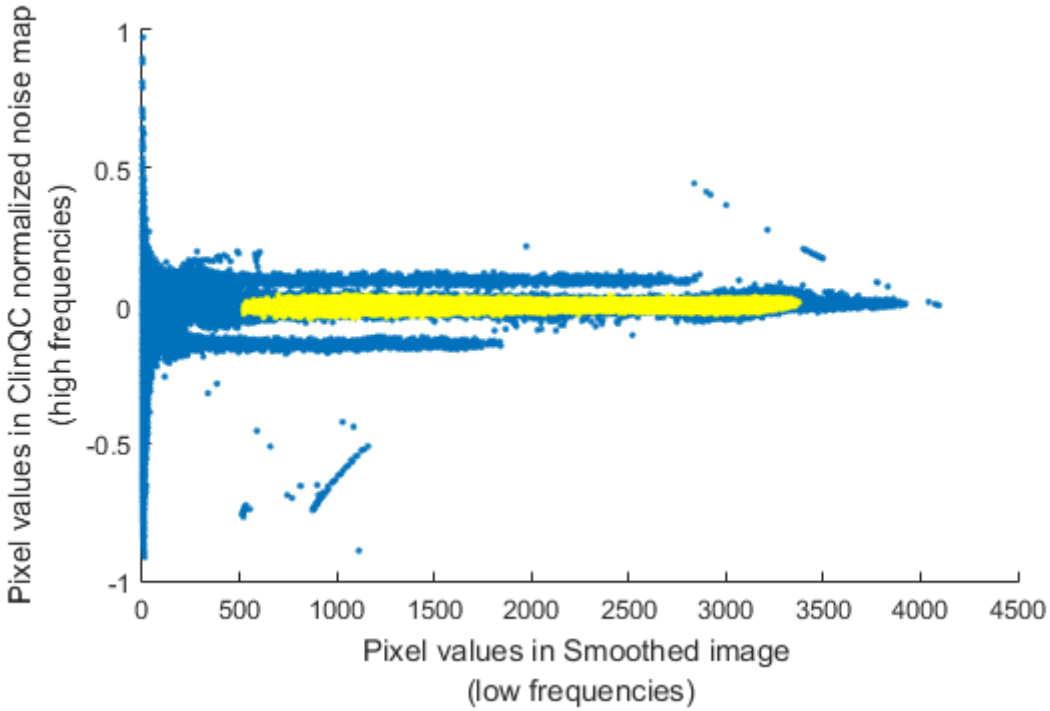

Figure 3.15: *High-low frequencies plot* for the central ROI (yellow) cropped from the ClinQC normalized noise map (light blue).

in this Section only one example was shown. The *high-low frequencies plots* and the *histograms* demonstrated to have the same characteristics for all the extracted noise images, except for different slopes, sizes and shapes of the secondary clusters, but this does not affect the general meaning of the results.
In Section 3.3.1.1 these results will be interpreted to better understand the properties of the ClinQC extracted noise images.

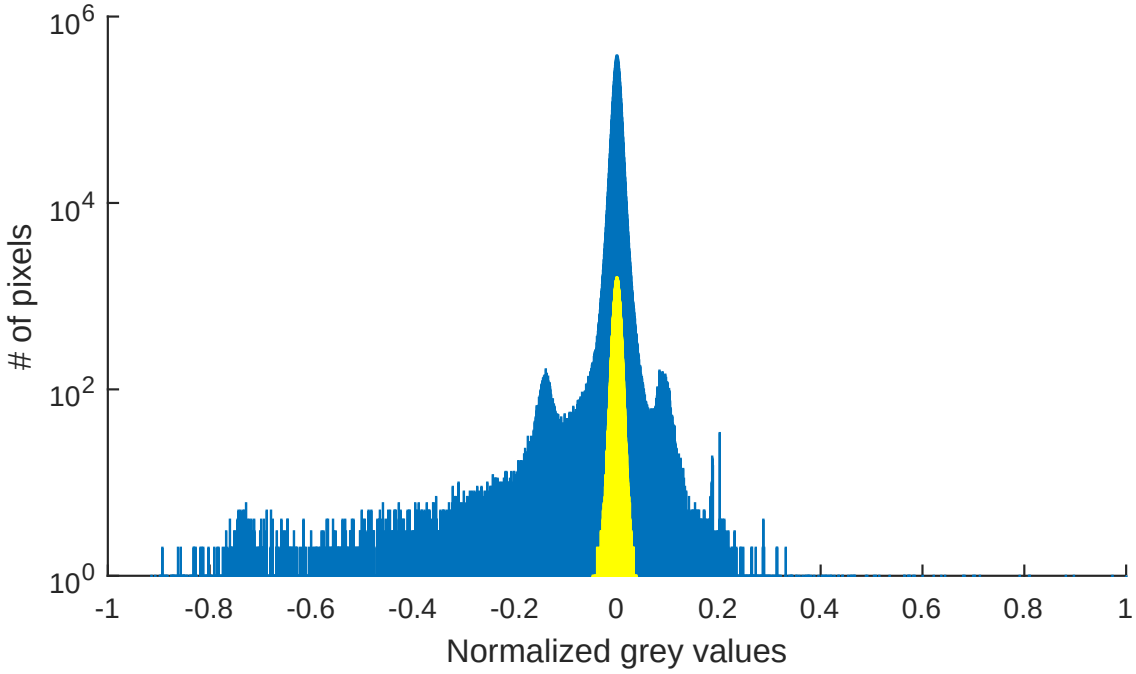

Figure 3.16: *Histogram* of the central ROI (yellow) cropped from the ClinQC normalized noise map (light blue).

3.2.1.2 The ClinQC noise values: baseline

The baseline of acceptance for the ClinQC noise values is chosen from all the clinical images of two weeks after the last maintenance of the system in April 2015. In Fig. 3.17 the baseline for the noise study is displayed, after an outliers removal of the eleven points that exceeded two times the standard deviation limit around the mean. The reason to remove the outliers is to have a more stable baseline, that could more appropriately represent future datasets. The new baseline mean of the ClinQC noise values after the outliers removal is $\mu_{baseline} = 0.0058\,\mathrm{normalizedgreyvalues}$[5], this is the reference level for all the future QCs until the next maintenance of the imaging system. The standard deviation of the noise values forming the baseline is $\sigma_{baseline} = 0.0003$. The red limits in Fig. 3.17 are placed in $\mu_{baseline} \pm 2\sigma_{baseline}$, so the

[5]The unit measure of the ClinQC normalized noise values is *normalized grey values* and will be omitted in all the future Sections and Figures for simplicity.

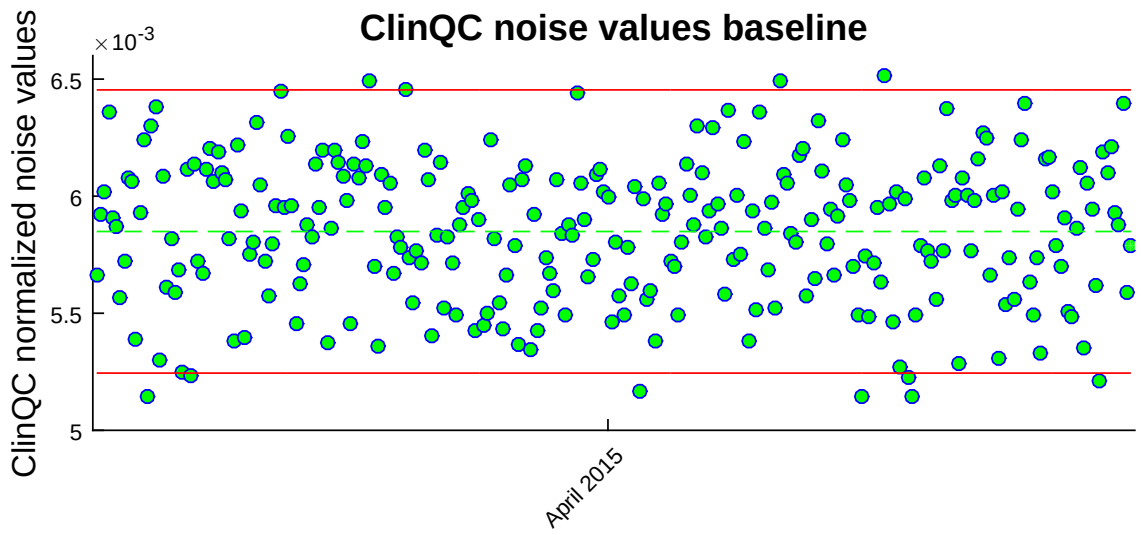

Figure 3.17: The ClinQC noise values baseline.

relative spread of acceptance around the mean is:

$$Relative\ spread\ _{baseline} = \frac{4\sigma\ _{baseline}}{\mu\ _{baseline}} \cdot 100 = 21\%, \tag{3.7}$$

that is exactly four times the commonly used Coefficient of Variation (CoV), defined as $CoV = {}^{\sigma\ _{baseline}}/_{\mu\ _{baseline}} = 5\%$. The normality of the distribution of the ClinQC noise values in the baseline dataset (Fig. 3.18) has been tested using the Kolmogorov-Smirnov (K-S) test [20]. The K-S test, with a p-value of 0.43, failed to reject the null hypothesis at 1% significance level. So the distribution of the ClinQC noise values in the baseline dataset can't be considered significantly different than a standard normal distribution. The Cumulative Distribution Function (CDF) estimated from the ClinQC baseline dataset was compared to the standard normal CDF by the K-S test, the visual similarity of the two CDFs is clear in Fig. 3.19. Since the distribution of the ClinQC noise values in the baseline dataset has a slightly skewed appearance, the additional distribution parameters skewness and kurtosis were computed for completeness.

$$Skewness\ _{baseline} = -0.17$$

$$Kurtosis\ _{baseline} = 2.44$$

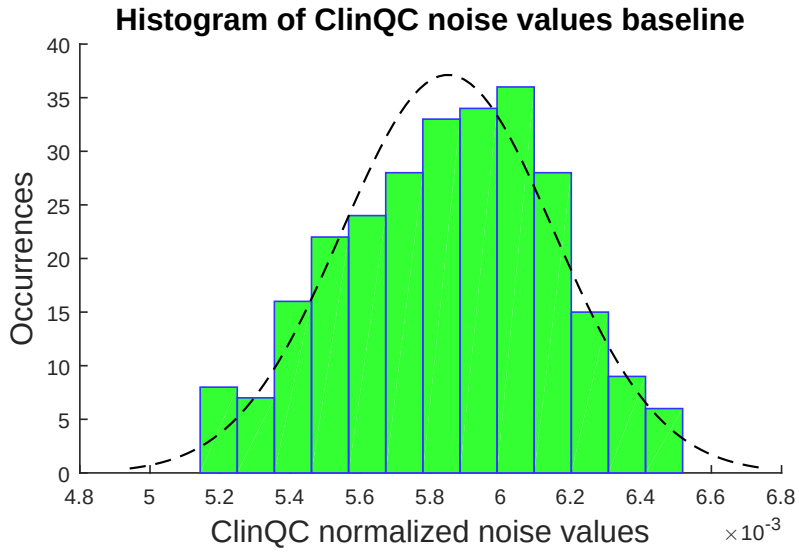

Figure 3.18: Distribution of the ClinQC noise values baseline, where the black dashed line is the normal fit of the histogram.

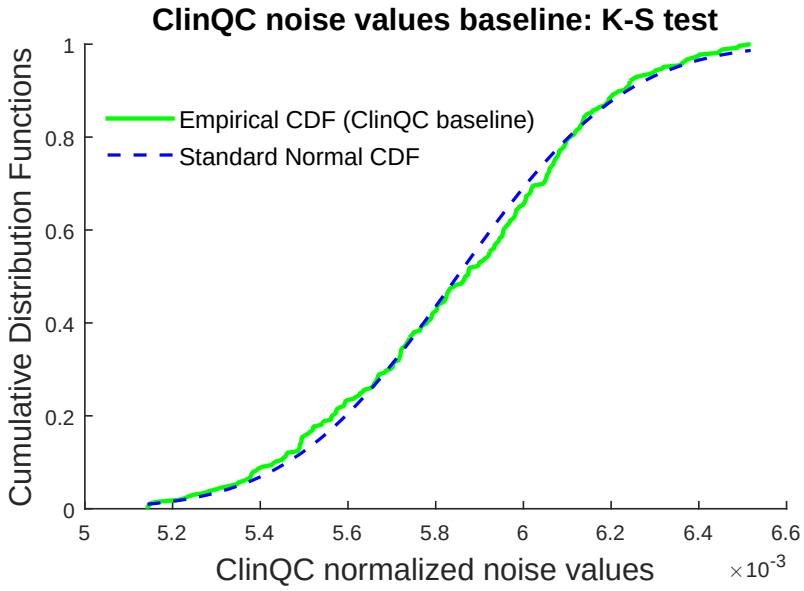

Figure 3.19: Estimated CDF of the ClinQC noise values baseline dataset, and standard normal CDF compared by the K-S test.

The skewness is negative, this means that the distribution has a tail towards low noise values, while a higher occurrence of high noise values is observed. The value is really small, and usually so small values are not considered to describe a real skewed

distribution. The same for kurtosis, which is a small value. These computed values were useful in the simulations (Section 3.1.3.2) in order to generate new random noise values from the same distribution as the ClinQC noise values baseline dataset, described by its proper mean, standard deviation, skewness and kurtosis.
An interpretation of the shape of the ClinQC noise values baseline distribution is given in Section 3.3.1.2.

3.2.1.3 The ClinQC noise values: Relevant clinical examples

In Fig. 3.20 we see some good examples of what happens to the ClinQC noise values if we change something in the imaging system or if we change the entire imaging system itself.

For this purpose, many images from different periods in the past were extracted

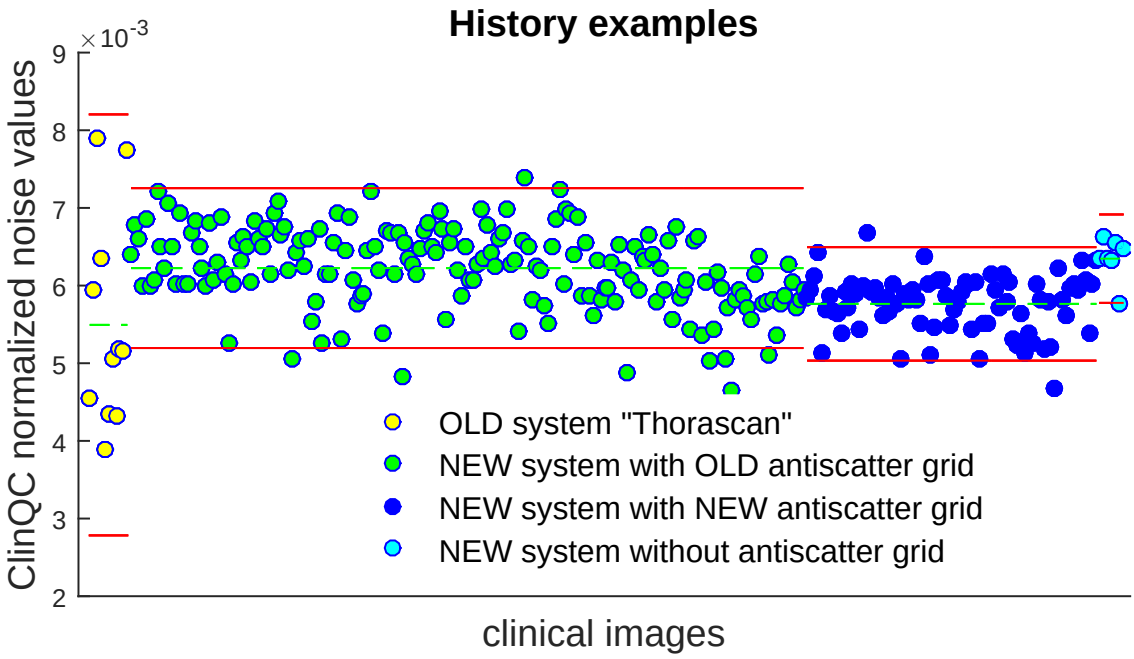

Figure 3.20: The ClinQC noise values historical examples.

from the PACS. An old imaging system called Thorascan was installed until 2013 in the same X-ray room used now for the ClinQC project. In yellow there are only few images from that system, but we can clearly see that both the mean and the standard deviation of the data series were higher if compared with the data series

of the new system. When the system was replaced, an old anti-scatter grid with a focal distance $f_{old} = 1.50\,\text{m}$ was installed, until December 2014. The average of the green noise values is 6.4% higher than the blue one, that belongs to images acquired with the system in the actual conditions, as it is used now with an anti-scatter grid that has a focal distance $f_{new} = 1.80\,\text{m}$. In light blue we see the results obtained with some patient images unfortunately acquired when there was no anti-scatter grid into the new imaging system. Even though there are only few points, it is clear that the average noise is higher than the baseline, in this case 8.5%.

3.2.2 The ClinQC algorithm - alternative versions: validation

In this Section all the preliminary results of the tests of the ClinQC alternative versions will be presented as an overview of what can be done in the future of the ClinQC project.

3.2.2.1 Grid sampling approach

Two typical distributions of all the ClinQC noise values measured on the grid of 50×50 sub-ROIs from the ClinQC normalized noise map are shown in Fig. 3.21 and Fig. 3.22 for two different patients, where the mean value of the distribution is represented in red and the median value in green.

The distribution for Patient # 1 (Fig. 3.21) is unimodal with a long tail towards

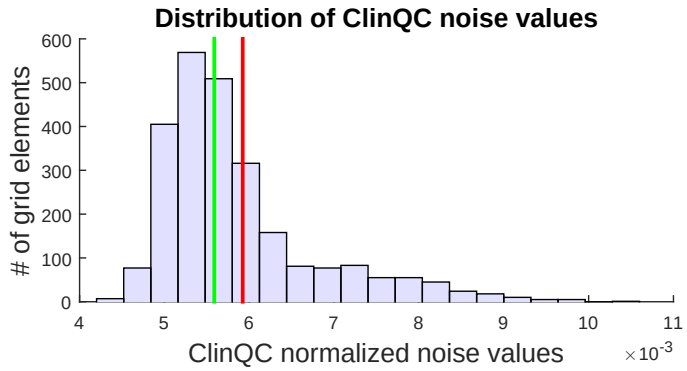

Figure 3.21: Distribution of the ClinQC noise values obtained using the grid sampling approach for Patient # 1. In red: mean value of the distribution. In green: median value of the distribution.

higher noise values. The peak is formed by almost all the sub-ROIs, while only few ROIs register higher noise levels. The median value better falls inside the peak of the noise distribution and so it represents a different estimation of the image noise than the mean value, that is shifted to higher noise levels for the presence of this low counts tail in the distribution of the ClinQC noise values. Using the mean allows to remove the sub-ROIs that give a higher standard deviation, so to remove the influence of the regions in the normalized noise map that contain more than noise, like the signal generated by nodules or implants edges for example. It can

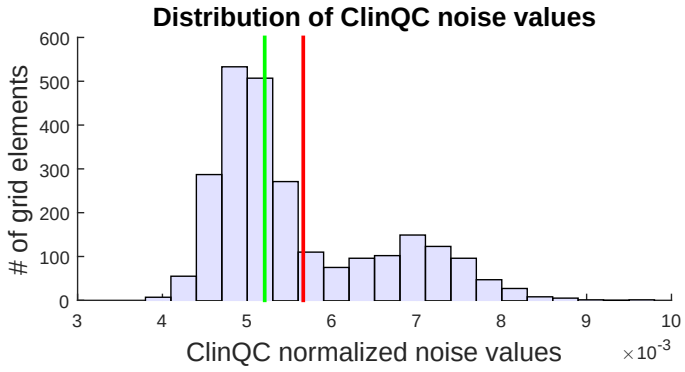

Figure 3.22: Distribution of the ClinQC noise values obtained using the grid sampling approach for Patient # 2. In red: mean value of the distribution. In green: median value of the distribution.

also happen that the distribution of the noise values is bimodal when the intruder object in the ROI (pacemaker, implant or medical device) is more prominent, like for Patient # 2 (Fig. 3.22).

Fig. 3.23 shows the comparison between the ClinQC original algorithm and the

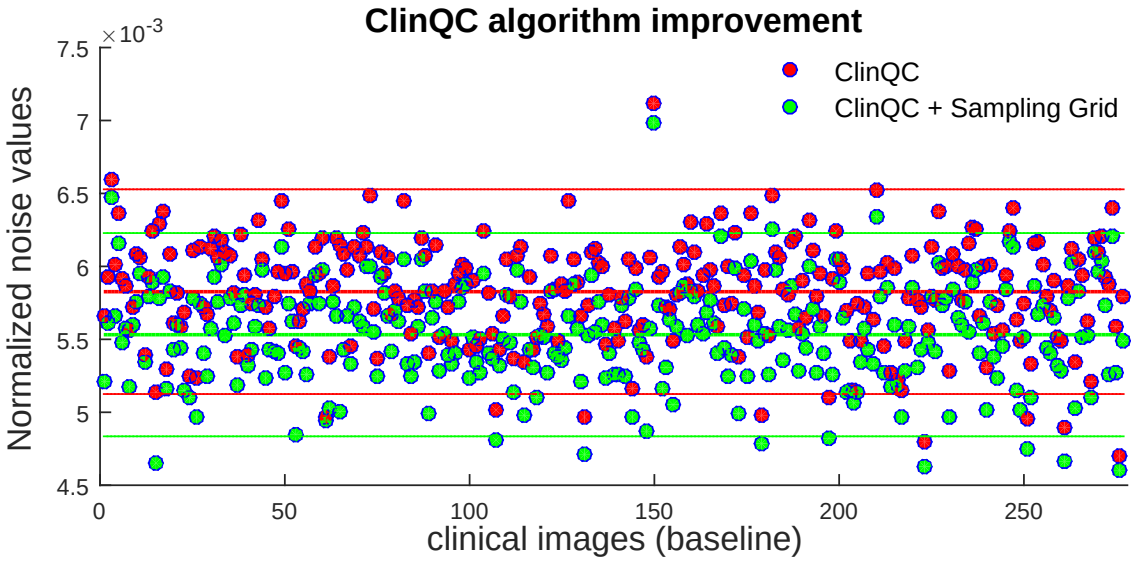

Figure 3.23: Comparison of the ClinQC original algorithm and the ClinQC modified method using the sampling grid approach for the images of the baseline (Section 3.2.1.2).

ClinQC modified algorithm using the sampling grid approach for the same group of images: the noise is globally lower using the ClinQC-alternative algorithm; this

means that taking only one big ROI, instead of doing this analysis, can make us measure something more than the noise itself for patients with implants, that origin asymmetries in the noise distribution and affect the noise measurement. The two images where we saw the biggest difference in percentage using one method or the other (the histogram of the differences between the two methods is shown in Fig. 3.26) were images of patients with exactly a pacemaker and other medical devices, suggesting that the large deviation in these cases was due to those objects. The two images are shown in Fig. 3.25 and Fig. 3.24.

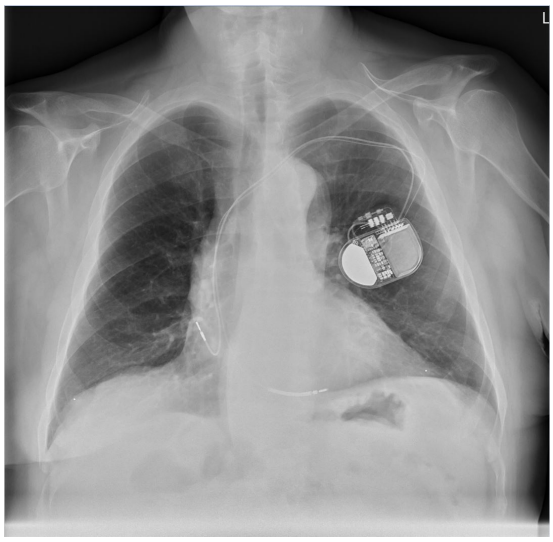

Figure 3.24: Patient with pacemaker.

Figure 3.25: Patient with medical device.

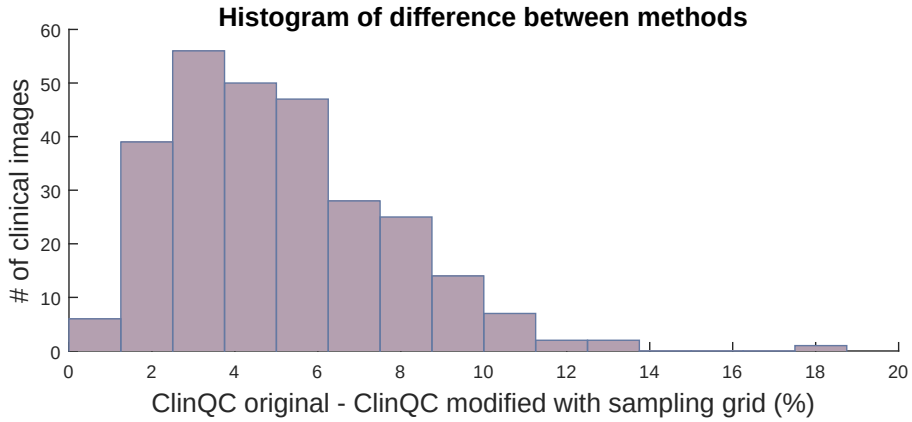

Figure 3.26: Histogram of the biggest differences in percentage in the ClinQC noise values obtained using the ClinQC original algorithm and the ClinQC modified algorithm using the sampling grid approach, for the images in the baseline period (Section 3.2.1.2).

3.2.2.2 ClinQC applied to mammography

An example of the ClinQC mammography normalized noise map and the selected ROI is displayed in Fig. 3.28 and the original image in Fig. 3.27.

The results obtained running the ClinQC-alternative algorithm on the entire dataset of mammograms and tomosynthesis reconstructed projections are shown in Fig. 3.29. The CoV of the ClinQC noise values is 14%. This might be due to the use of different kVp filtration for each patient, that could have played a more important role in the variability of the noise values, than in the ClinQC noise values baseline dataset for chest images that are acquired with a fixed kVp and present a CoV of 5%.

Four noise values are considered outliers since they exceed the red acceptance limit of $+2\sigma$ and these four images contained some surgery staples, a medical implant used during surgeries, that made significantly increase the noise values over the upper acceptance limit. Similar considerations as in Section 3.2.2.1 can be made since these artefacts can leave a clear sign in the noise map producing a shift towards higher noise values.

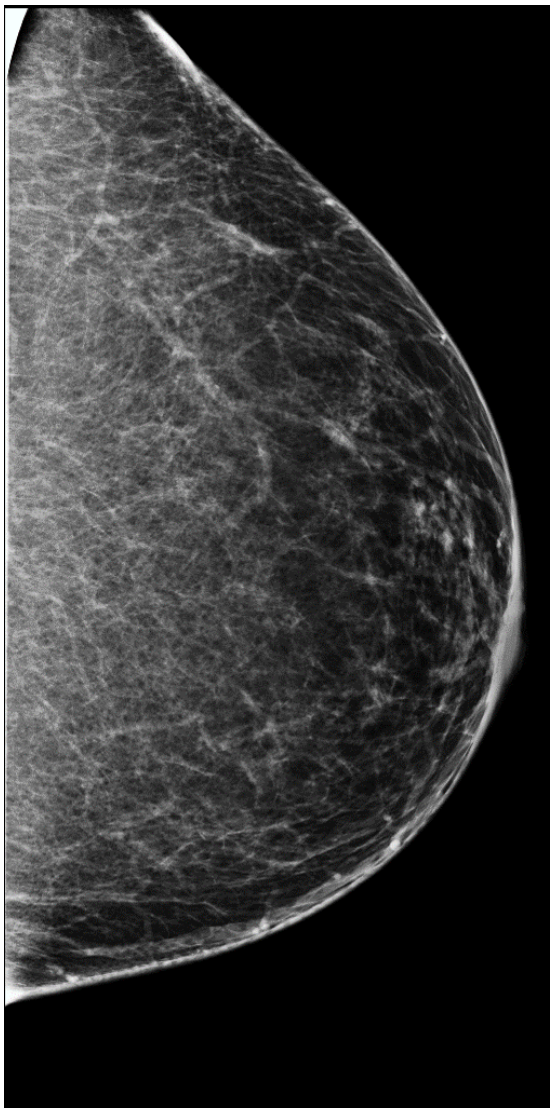

Figure 3.27: A typical mammogram.

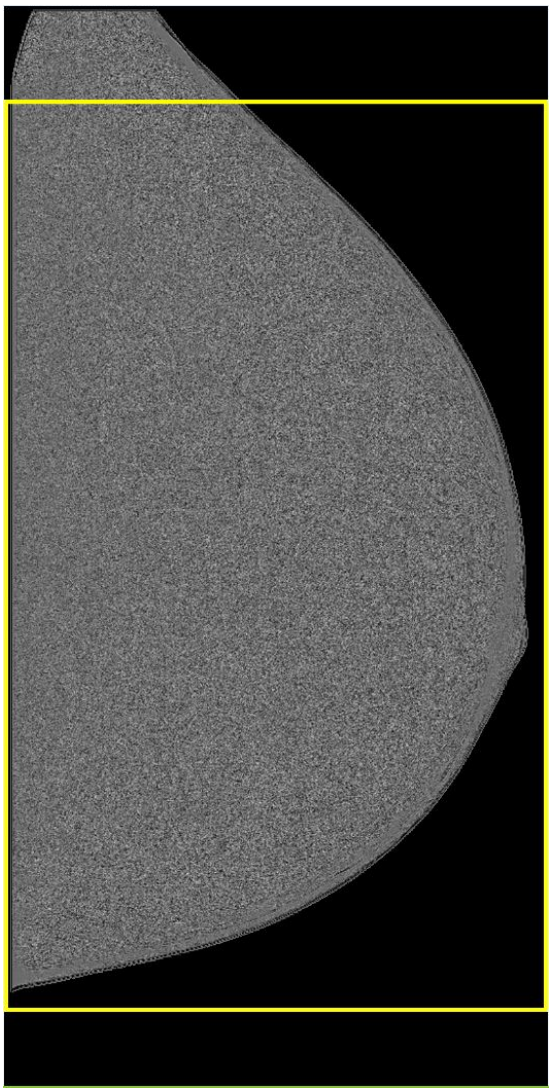

Figure 3.28: The ClinQC noise map of a mammogram with the ROI marked in yellow.

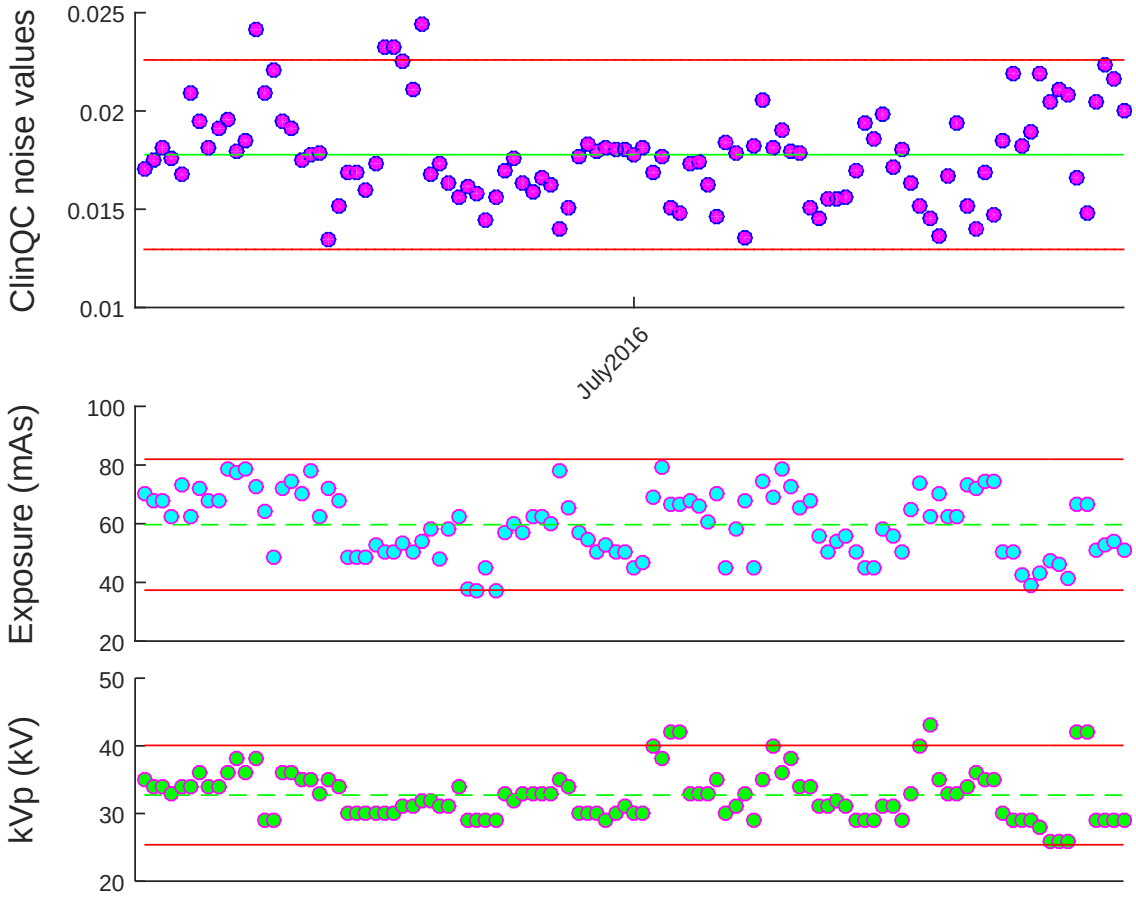

Figure 3.29: The ClinQC noise values, exposure and kVp computed on the dataset of mammograms and tomosynthesis reconstructed projections.

3.2.3 Image deterioration study

In this Section all the results of the image deterioration study (Section 3.1.3.1) will be displayed.

3.2.3.1 Blur

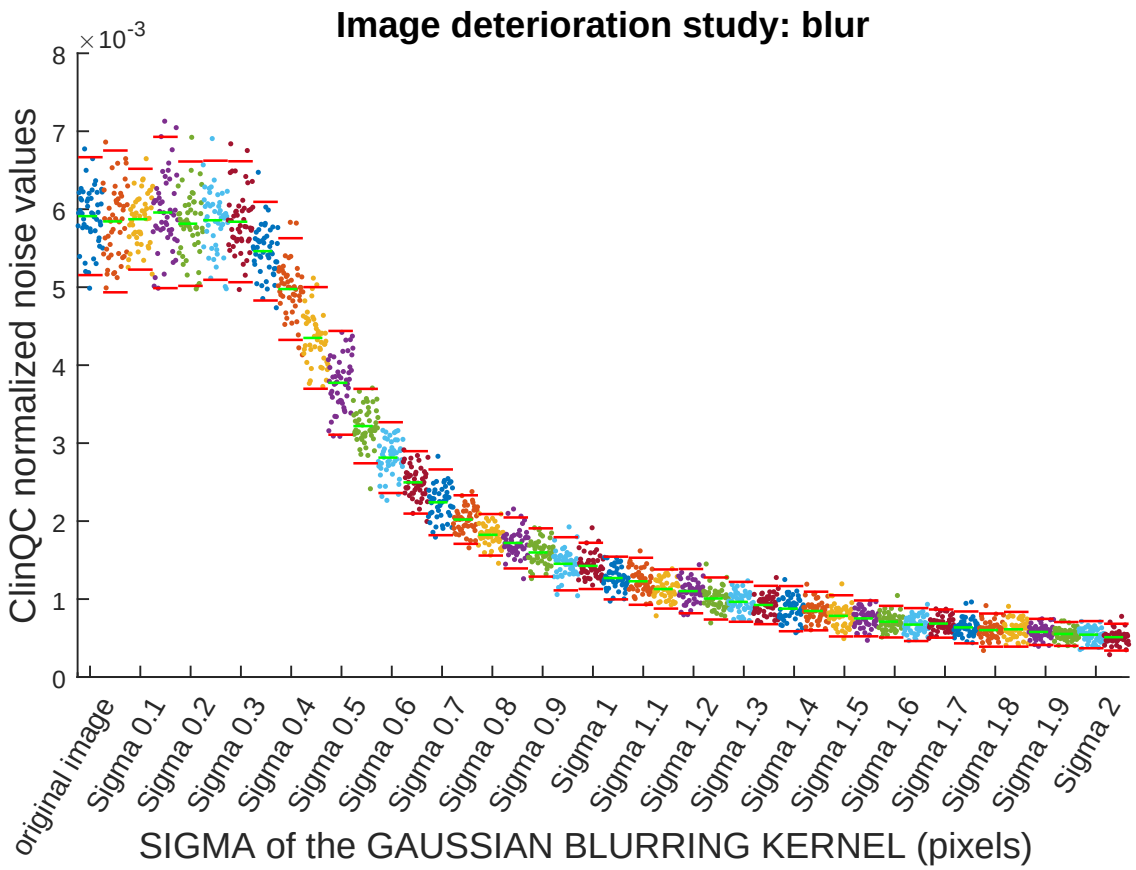

Figure 3.30: Output of the image deterioration study that shows the decrease of the *mean* of the noise values measured after blurring, with different increasing amounts of blur, 41 different groups of 50 images, together with the first blue group made of 50 original X-ray images of the chest.

In Fig. 3.30 is displayed the decrease of the *mean* of the noise values measured after blurring, with different increasing amounts of blur, 41 different groups of 50 images, together with the first blue group made of 50 original X-ray images of the chest. The ClinQC noise values start to be clearly reduced when the σ of the Gaussian

blurring kernel used to smooth the images is high enough to actually produce a noise reduction on the images, so when σ=0.4 pixel and the kernel size is 3×3pixel (pixelsize = 0.12×0.12 mm). Then a rapid fell of the ClinQC noise values is observed, until a saturation to the lowest values around $0.5 \cdot 10^{-3}$. At low noise values, after σ=1 pixel all the noise has been erased from the image by the blurring.
In Fig. 3.31 is shown the percentage of variation of the average value of each group of the ClinQC noise values represented in the previous Fig. 3.30, and the error bar is the relative spread of each group of values. The average values are normalized by the mean of the ClinQC noise values of the first group of original chest images. The decrease of the ClinQC noise indicator in percentage is extremely fast with a high slope, even with small amounts of blur.

The appearances of the same detail in the lung of a patient chest X-ray image

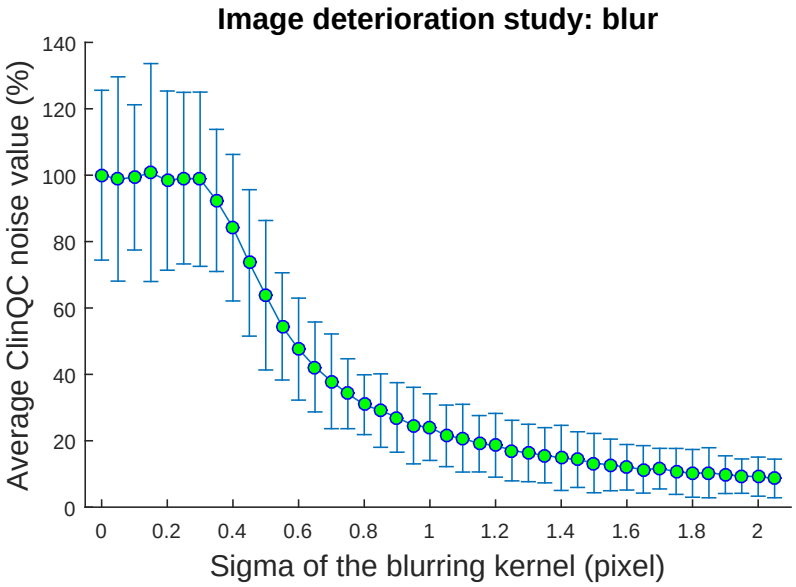

Figure 3.31: The decrease in percentage of the *mean* of the noise values measured during the image deterioration study with blurring.

used for this analysis, before and after blurring with $\sigma = 1$ *pixel*, are represented respectively in Fig. 3.32 and Fig. 3.33. This slightly smoothed effect is barely detectable by visual inspection, but the ClinQC translates it in 70% lower noise values.

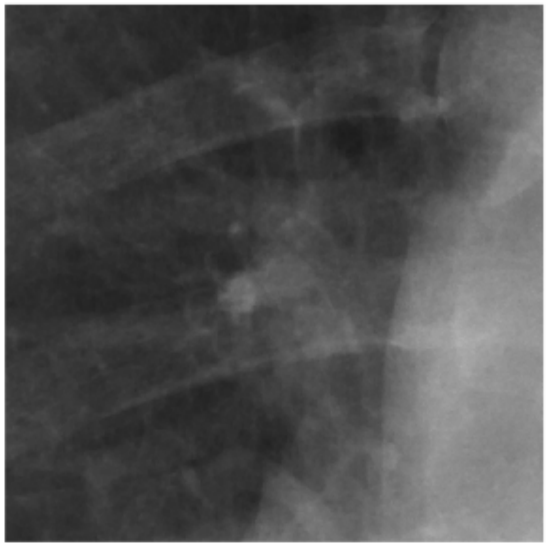

Figure 3.32: A detail in the lung of a patient chest X-ray image.

Figure 3.33: A detail in the lung of a patient chest X-ray image after blurring with $\sigma = 1$ *pixel*.

Also the variation of the relative *spread* (Eq. 3.7) of the ClinQC noise values has been investigated, using for each blurring level the same group of 25 original images (Fig. 3.34). The percentage of variation of the relative spread of each group of noise values of the deteriorated images is displayed in Fig. 3.35. When the mean of the noise values decreases by 70% (Fig. 3.31), their relative spread increases three times, following a logarithmic trend (Fig. 3.35).

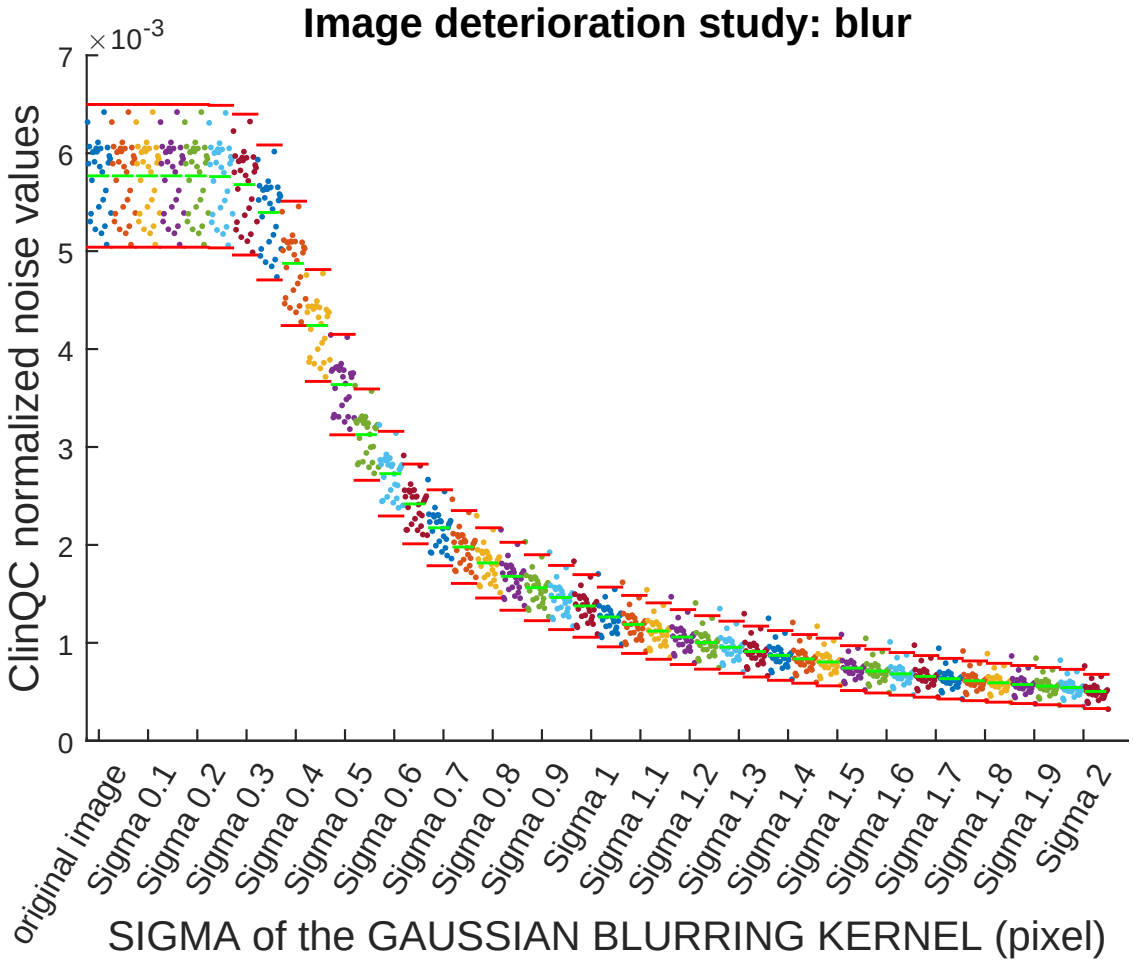

Figure 3.34: Output of the image deterioration study that shows the decrease of the *spread* of the noise values measured after blurring, with different increasing amounts of blur, the same group of 25 original X-ray images of the chest.

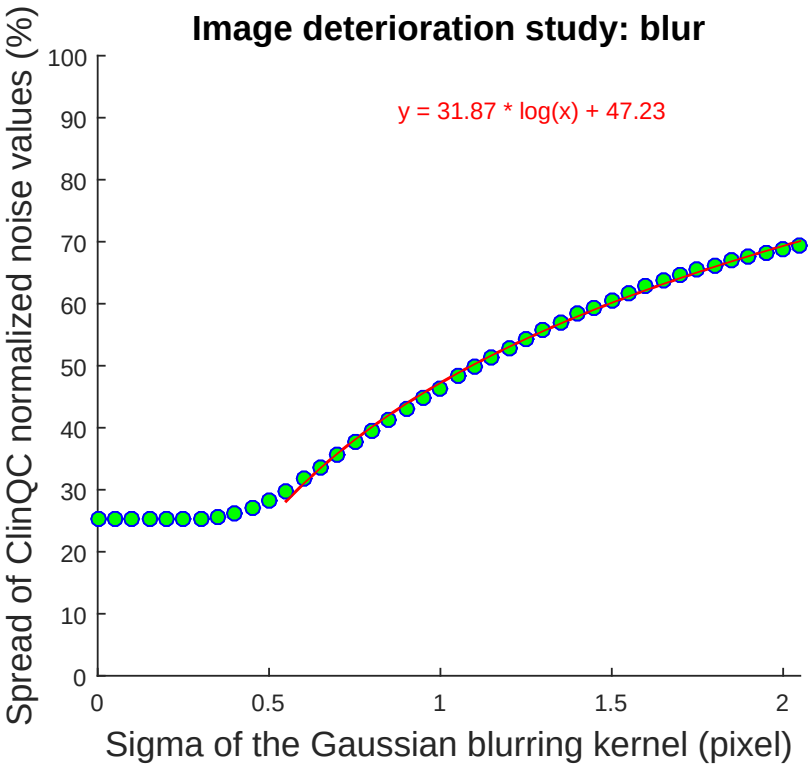

Figure 3.35: The increase in percentage of the relative *spread* of each group of noise values of the deteriorated images measured during the image deterioration study with blurring.

3.2.3.2 Gaussian noise

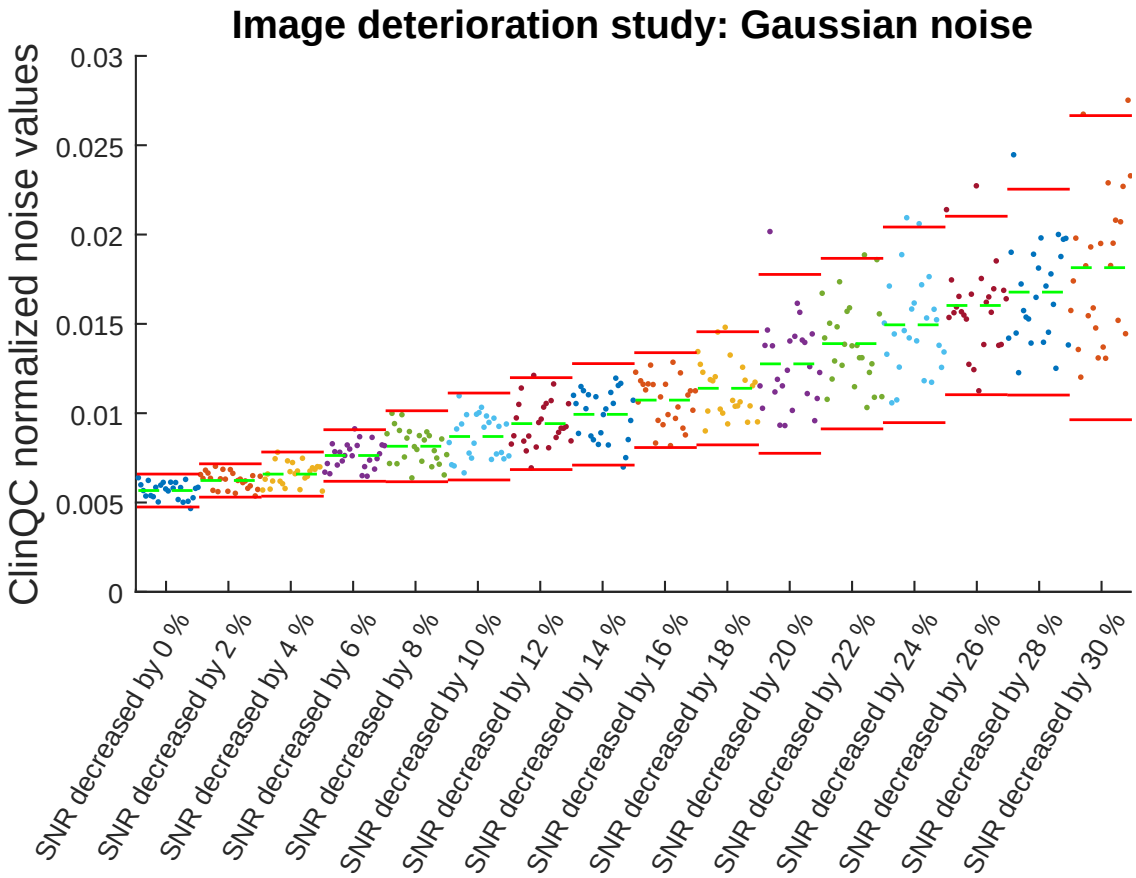

Figure 3.36: Output of the image deterioration study that shows the increase of the *mean* of the noise values, measured after the addition of Gaussian noise with 0 mean and variances computed in order to achieve prefixed levels of deterioration of the estimated SNR of different random sub-sampled groups of 25 original chest X-ray images.

Adding to different groups of 25 chest images different amounts of Gaussian noise, with 0 mean and variances computed in order to achieve prefixed levels of deterioration of the estimated SNR, such as linearly decreasing the SNR by 2% to 30%, leads to a huge and rapid increase in the ClinQC noise values.
In Fig. 3.37 the percentage of variation of the ClinQC noise values after addition of Gaussian noise is displayed setting the baseline at 100%.

The SNR was computed using the Eq. 3.2 with a unit measure in dB, since it allows to use the entire image signal and so the entire noise map information. While

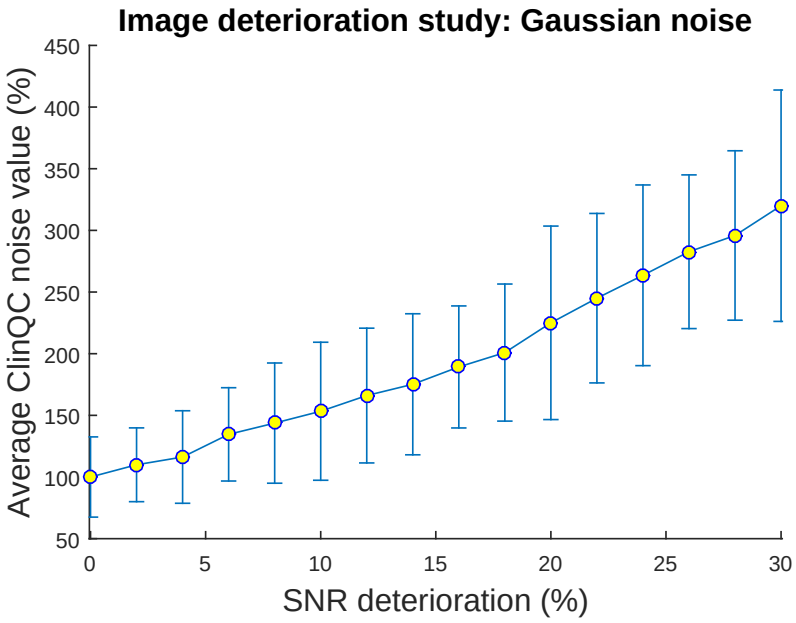

Figure 3.37: The increase in percentage of the *mean* of the noise values measured during the image deterioration study with Gaussian noise.

the traditional Eq. 3.3 required the automated selection of specific ROIs in the image, for example a rectangular region on the spine which is nearly always placed in the same central position for different patients. The traditional approach, from a first analysis, seemed to be less robust and less appropriate for this study than the method that uses the entire image signal and measures the SNR in dB.
An example of the appearance of one detail in the lung of a patient is represented, respectively before and after addition of Gaussian noise, in Fig. 3.38 and Fig. 3.39. In this example, reducing the estimated SNR by 30%, the ClinQC noise values become three times higher (300%).

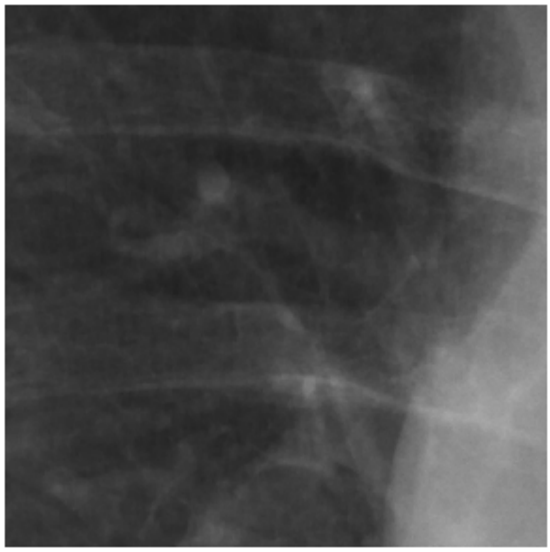

Figure 3.38: A detail in the lung of a patient chest X-ray image.

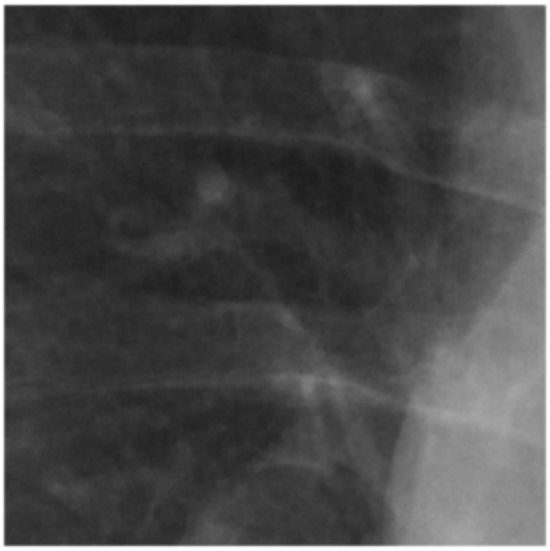

Figure 3.39: A detail in the lung of a patient chest X-ray image after addition of Gaussian noise, with 0 mean and variance computed in order to achieve a decrease of the estimated SNR by 30%.

Also in this case, the variation of the relative spread (Eq. 3.7) of the ClinQC noise values has been investigated, using for each different addition of Gaussian noise the same group of 28 original images (Fig. 3.40).

The relative spread of the ClinQC noise values of each group of deteriorated images is displayed in Fig. 3.41, and it follows a quadratic law.

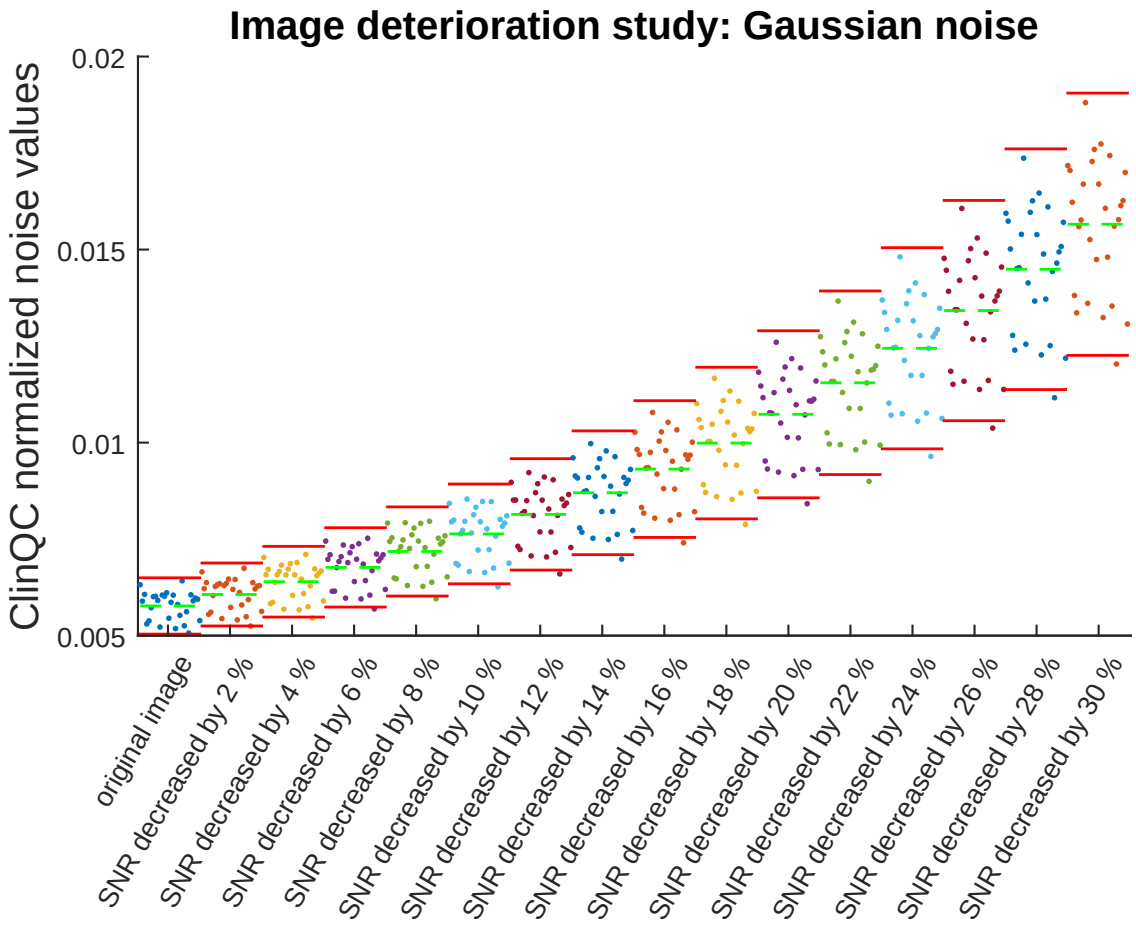

Figure 3.40: Output of the image deterioration study that shows the increase of the *spread* of the noise values, measured after the addition of Gaussian noise with 0 mean and variances computed in order to achieve prefixed levels of deterioration of the estimated SNR of different random sub-sampled groups of 28 original chest X-ray images.

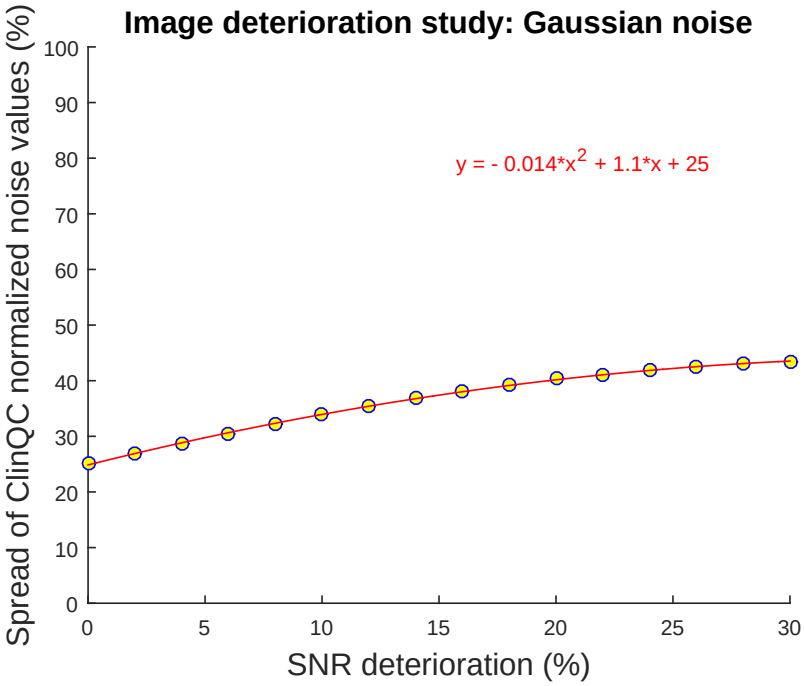

Figure 3.41: The increase in percentage of the relative *spread* of the noise values measured during the image deterioration study with Gaussian noise.

3.2.4 Statistical analysis with simulations

3.2.4.1 Step simulation

An example of the rejection ratio plot that is obtained at point 4 of the step simulation is displayed in Fig. 3.42: the mean of each simulated group of noise values was 3% higher than the baseline, and the sizes of the groups started from 2 to 150 images at steps of 2 with M = 1000 iterations. This plot represents the score of the T-test in a large number of iterations for each simulated dataset size, so the percentage of iterations when the T-test has been able to detect that the simulated noise values of a certain size where actually generated from a distribution with a mean higher than the baseline. Considering the 99.5% threshold in the rejection ratio graph,

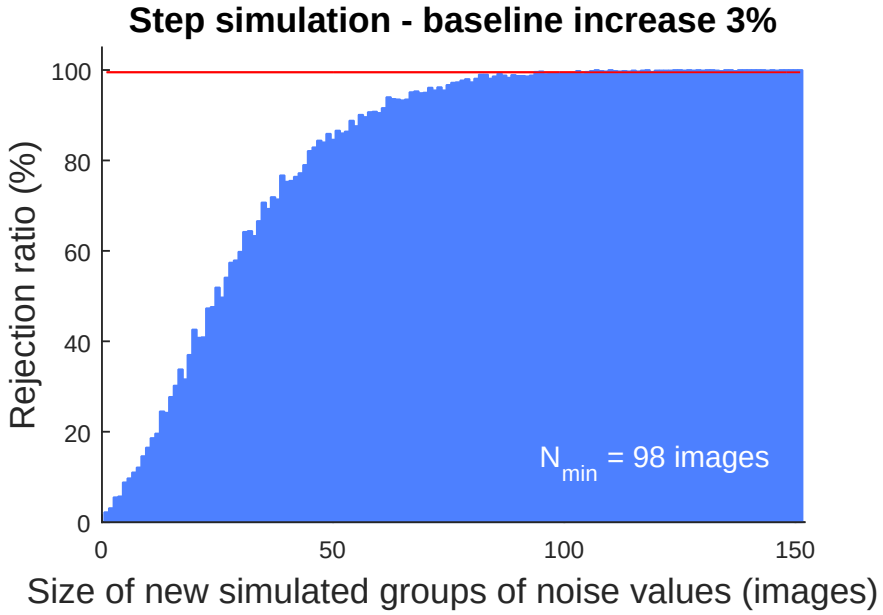

Figure 3.42: Rejection ratio (%) resulting from the step simulation when the baseline was increased by 3%.

with a dataset of at least 98 new images we are nearly always able to reject the null hypothesis that the simulated dataset come from a normal distribution with the same mean and equal but unknown variance as the baseline of the ClinQC noise values (significance level set at p¿0.005): so it is nearly always possible to detect that the simulated points are generated from a distribution with equal variance as the baseline dataset but a 3% higher mean. While with new datasets of only 25 images

the T-test is able only in 30%-40% of cases to detect this change. So if something happens to the system that produce a noise increase of 3%, an automated algorithm cannot always detect for sure a deviation from the baseline mean on the following 25 clinical images. This is why a high threshold of 99.5% is chosen for the rejection ratio graph, and this allows us to state that with at least 98 random patient images acquired after the onset of the problem, a detection algorithm based on T-test would be able to reveal a change in the noise levels.

Following the next points of the step simulation algorithm introduced in Section 3.1.3.2.1, the same analysis just shown is repeated for a set of baseline increases in percentage from 3% to 15% higher simulated mean. The final output curve of the step simulation is displayed in Fig. 3.43. This final curve shows that, simulating a

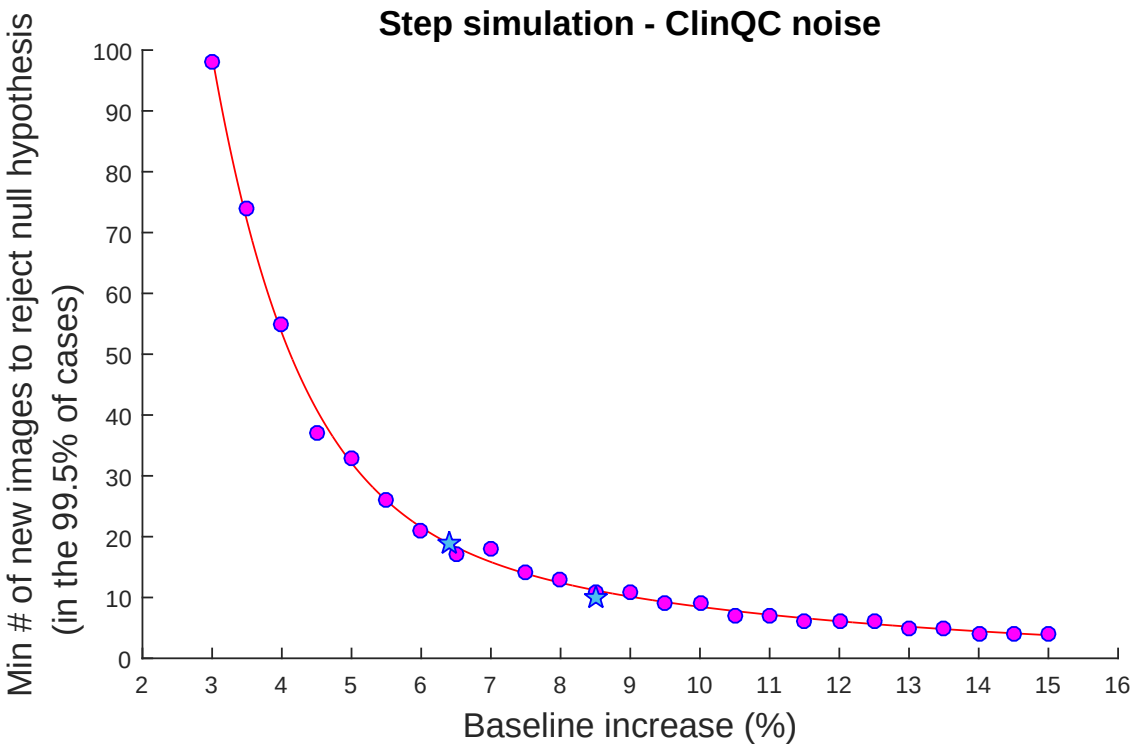

Figure 3.43: Final output curve of the step simulation (pink). The light blue stars represent the output values of the two simulations of the anti-scatter grid removal and replacement.

set of noise increases that corresponds to different possible changes in the imaging

system quality, the algorithm needs the acquisition of different numbers of new clinical images before the detection of the step in the noise levels is possible in the 99.5% of cases. This follows the exponential trend

$$y = ae^{bx} + ce^{dx}$$

with parameter values $a = 895.70$, $b = -0.83$, $c = 38.76$ and $d = -0.15$.
This curve means that in only one day, with 50 new clinical chest X-ray images, it would be possible to observe an increase in the noise levels of 4%, that can be linked to a decrease in the exposure (mAs) of 8%.

The second study that used a slightly modified version of the step simulation code is more specific to the imaging system in use. The minimum number of clinical images needed to detect an accidental anti-scatter grid removal is derived from the Rejection ratio diagram and is shown in Fig. 3.44. The simulation was run with 2000 iterations and the mean of the new groups of simulated noise values was $\mu_{NO\ grid} =$ 8.5% higher than the baseline. The minimum number of images to check with an automated algorithm based on T-test if the anti-scatter grid has been removed is 10 images.
In Fig. 3.45 the same analysis is done in case the anti-scatter grid is replaced with the old one installed in the system until 2014, or with an equivalent one with the same focal distance as the old one. Since the average noise values registered with the old anti-scatter grid were 6.4% higher than the actual baseline, the automated QC algorithm would need 19 images to detect the hypothetical downgrade of the anti-scatter grid, of course more than with the previous example, since the two distributions of the noise values of the baseline and the new simulated points are more overlapped. The two output values of these simulations correspond to the fit of the more general step simulation output curve (light blue stars in Fig. 3.43).

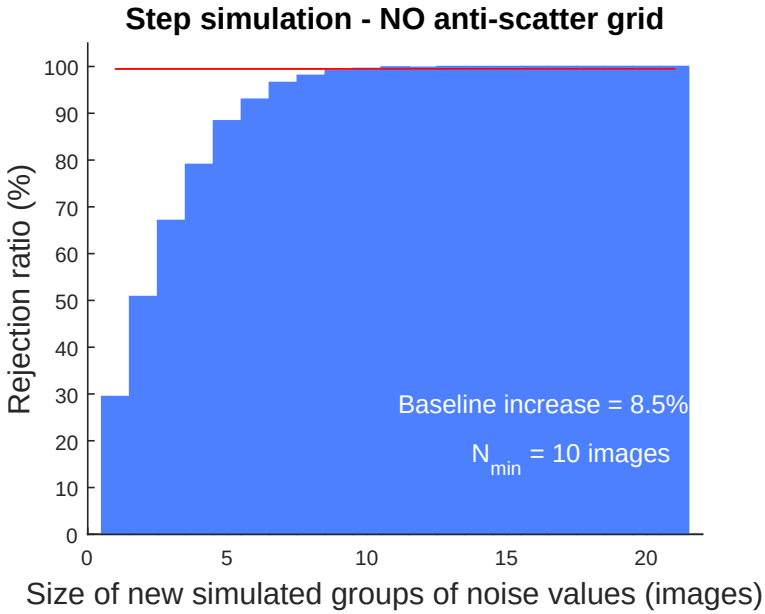

Figure 3.44: Rejection ratio (%) obtained with the step simulation of the anti-scatter grid removal.

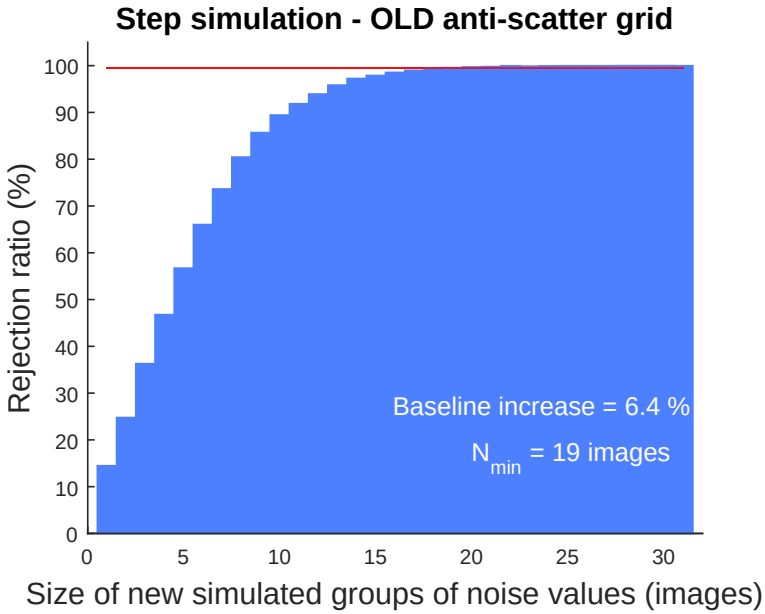

Figure 3.45: Rejection ratio (%) obtained with the step simulation of the anti-scatter grid replacement.

3.2.4.2 Trend simulation

The trend simulation was implemented to simulate two different linear time dependent deteriorations of the imaging system. The first example used a time line of one week (250 images) and the second one a time line of one month (800 images). In both the examples the increasing slope started at the baseline level and reached 30% higher noise values. Both the simulated baseline points and the simulated slope points were randomly generated at every iteration, the example of the one month time line is shown in Fig. 3.46.

The sizes of the moving windows moved along the month time line were 50, 120,

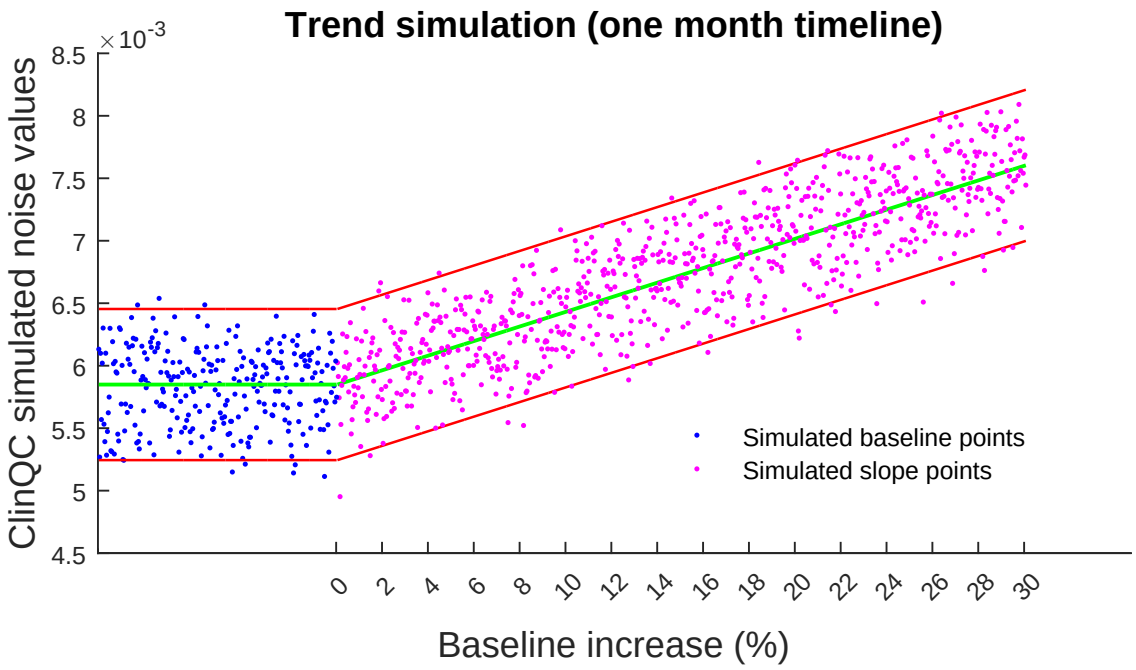

Figure 3.46: Example of the one month time line of the simulated ClinQC noise values generated by the trend simulation.

190, 260, 330 and 400 images. The Rejection ratio in Fig. 3.47 shows the scoring results of the T-test using a moving window of size 50 images on the simulated month time line. With a threshold of 99.5%, only when the moving window of 50 noise values reaches an average 5% higher than the baseline, the T-test is able to reveal that something in the noise values has changed, that in this case is not a step but a simulated trend: since the 50 points falling each time inside the moving window are

arranged on a slope, this simulation becomes a specific case of the step simulation by considering the average noise value inside the window.

Repeating this analysis for all the sizes of the moving windows listed before, the

Figure 3.47: Example of the rejection ratio (%) obtained performing the trend simulation on the month time line.

green curve in Fig. 3.48 is obtained. The data points were fitted with an exponential curve $y = ae^{bx}$ with parameter values $a = 1.13 \cdot 10^3$ and $b = -0.64$.

The same analysis was performed on the week time line, with window sizes of 10, 28, 46, 64, 82 and 100 images. The final curve is represented in the same diagram, the pink data points have been fitted using an exponential curve $y = ae^{bx}$ with parameter values $a = 7.03 \cdot 10^2$ and $b = -0.69$.

From the two final output curves it is easy to observe that a faster slope (week time line) can be detected easier than a slower slope (month time line), where more images have to be included using a larger window to detect the same baseline increase in percentage.

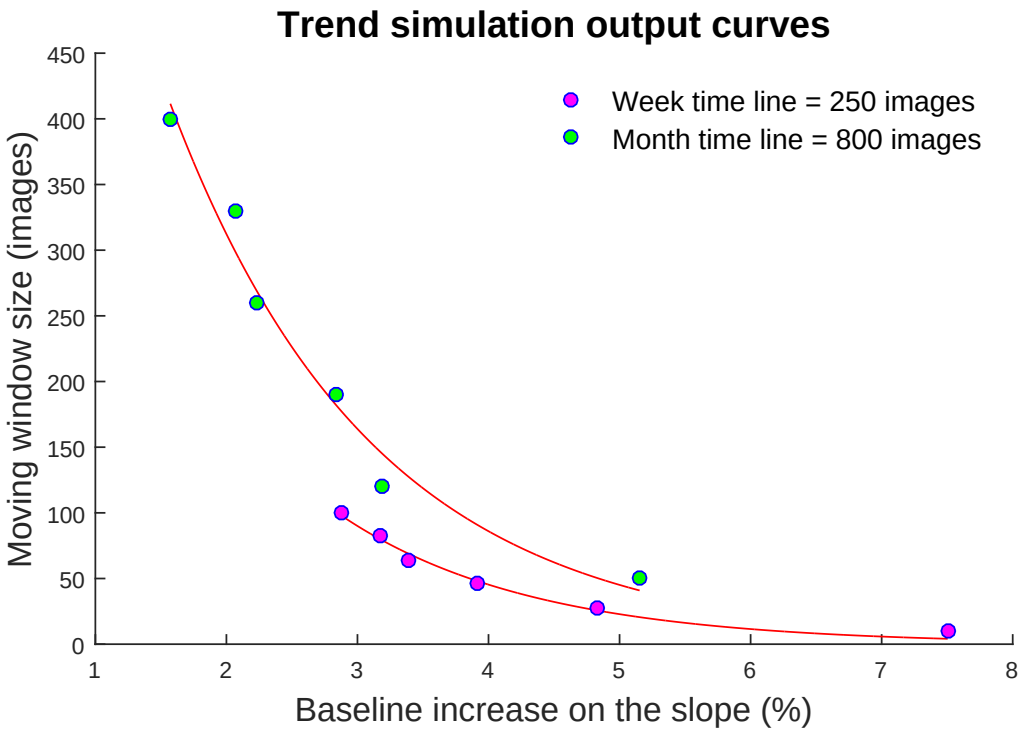

Figure 3.48: Output curves of the trend simulations performed on the one week time line and the one month time line.

3.2.5 The ClinQC performance in clinical practice

In this Section two interesting examples of the performance of the ClinQC noise extraction algorithm in clinical practice are shown.

3.2.5.1 Detection of flipped anti-scatter grid

The appearance of the QClight images acquired with the anti-scatter grid in its correct positioning and with a flipped anti-scatter grid are displayed in Fig. 3.49 and Fig. 3.50. The correct positioning makes the image appear homogeneous, while the flipped grid blocks the lateral photons that make the image appear whiter in the lateral vertical sides and darker in the center, where the majority of photons arrived.

Also the clinical image quality is strongly reduced when the anti-scatter grid is not

Figure 3.49: The appearance of a typical QClight phantom image acquired during the weekly QC.

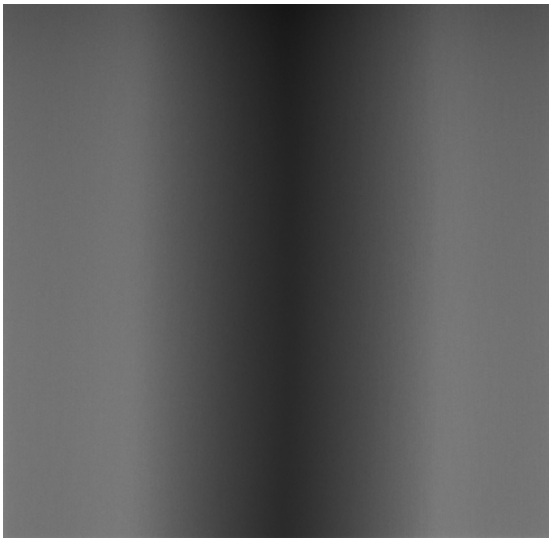

Figure 3.50: The appearance of QClight phantom image acquired with a flipped anti-scatter grid.

placed in its correct positioning. The contrast between the lung region and the surrounding chest structures becomes weaker, degrading the diagnostic quality of the image. The appearance of two images of the same patient who came for multiple examinations in the past and during the days where the anti-scatter grid was flipped are displayed respectively in Fig. 3.51 and Fig. 3.52.

Fig. 3.53 represents the noise values and the exposures measured on all the clinical images of the week when the anti-scatter grid was flipped. Monday 11 July 2016 the QClight image was homogeneous (Fig. 3.49), while on Monday 18 July 2016 the grey levels were not uniform at all (Fig. 3.50). By visual inspection, it is possible to identify the days where the anti-scatter grid was flipped since it produced higher exposures and lower noise values. The AEC uses three ionization chambers, one in the center of the acquisition field and two on the lateral sides where less photons reached them due to a flipped grid. So the AEC stopped the flux of photons at a later acquisition time and the exposures of the patients became generally higher than for the baseline. The noise in the QClight images is measured on a central ROI so where more photons reached the detector, which generally produced lower noise values than usual.

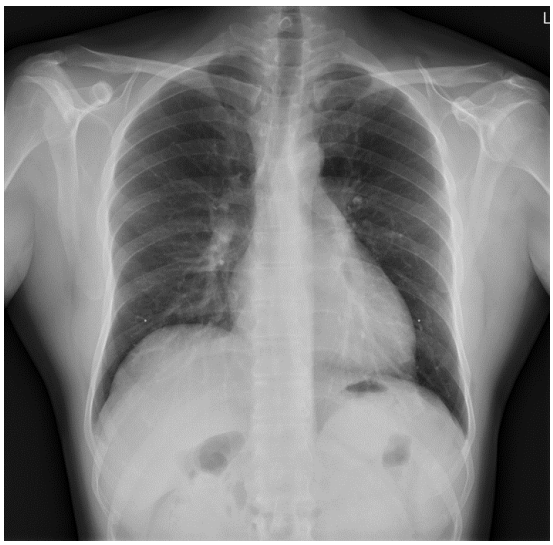

Figure 3.51: The appearance of one patient image acquired in 2015 with the same imaging system now in use.

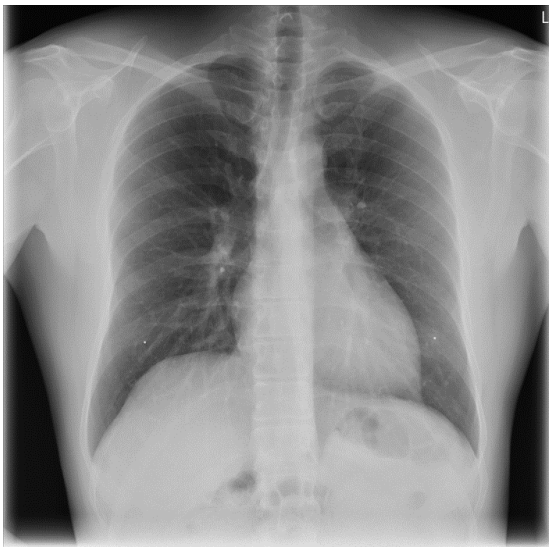

Figure 3.52: The appearance of the same patient image acquired in 2016 with the anti-scatter grid flipped.

The points where the human eye could observe lower noise values had actually an average noise level 9.3% lower than the baseline. So, according to the step simulation output curve (Fig. 3.43), the minimum number of images, coming from a distribution with a mean 9.3% lower than the baseline, that are needed to be able to detect the step using a T-test within a moving window are at least 10. The result of the alerts produced by a moving window of size 12 images moved along all the images of the complete week are shown as example in Fig. 3.54. The window was chosen slightly larger, since the step simulation does not simulate outliers in the noise distribution, that can appear instead in regular clinical practice: so the minimum number of images suggested by the step simulation output curves does not keep in account that in the real clinical practice there could be more patients giving outliers than the ones belonging to the tails of the noise values distribution. So the window is kept larger as an approximation to correct for this lack of simulated outliers patients. A red vertical line is an alert that the algorithm produces at the ending position of the moving windows when the T-test is able to detect a change. There are false positive alerts the 12^{th} of July, but the radiologists did not notice something abnormal in the image quality on the images of that day, so these false alerts are

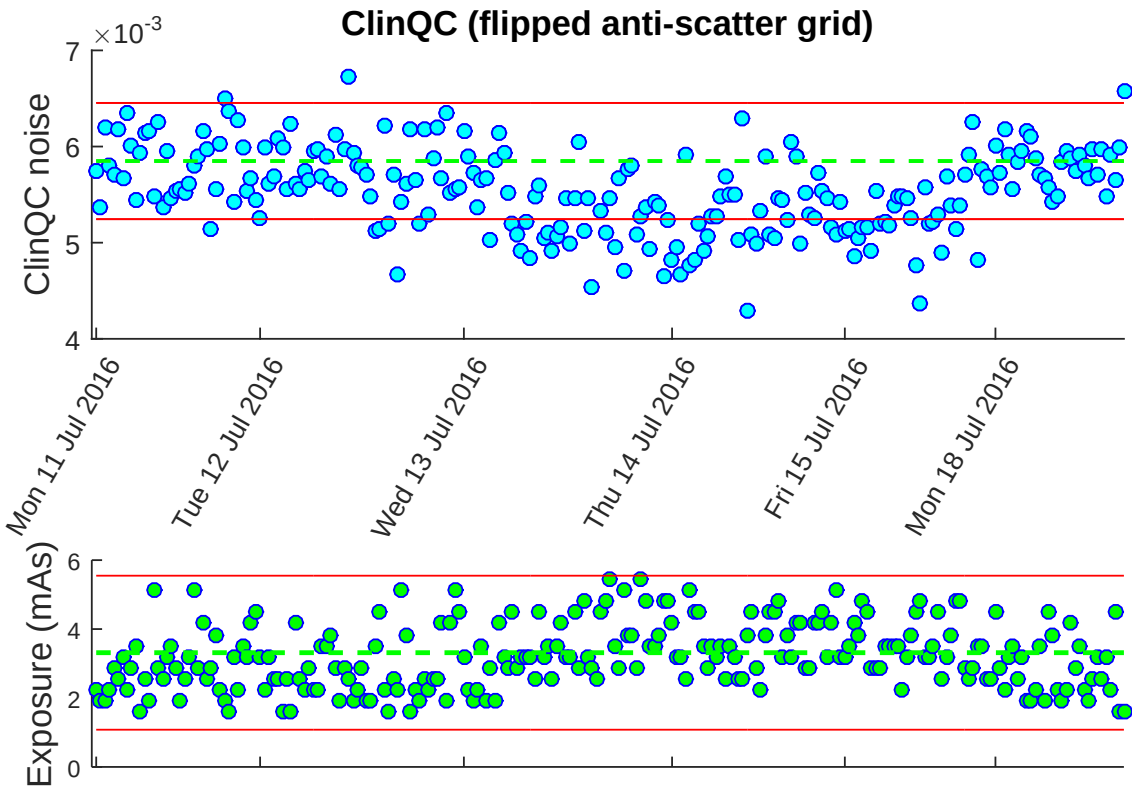

Figure 3.53: The noise values and the exposures measured on all the clinical images of the week when the anti-scatter grid was flipped in July 2016.

probably due to the presence of outliers. From this dataset, the patient images that were outliers both in terms of noise and dose were removed from the time line, since a change in the exposure is linked to a change in the noise but not necessarily to a change in the quality of the imaging system. Despite these first false alerts, the interesting thing is that in the three days after there is a continuum of alerts that corresponds perfectly to the period where also one of the radiologists working at LUMC realized a decrease in the image quality of the chest X-ray examinations. So the ClinQC detection algorithm was successfully able to select the days where the anti-scatter grid was flipped.

An improvement of the detection of specific problems in the imaging system may be achieved by using a machine learning tool [21]. The idea would be to perform a statistical classification using the noise and dose parameters derived from clinical

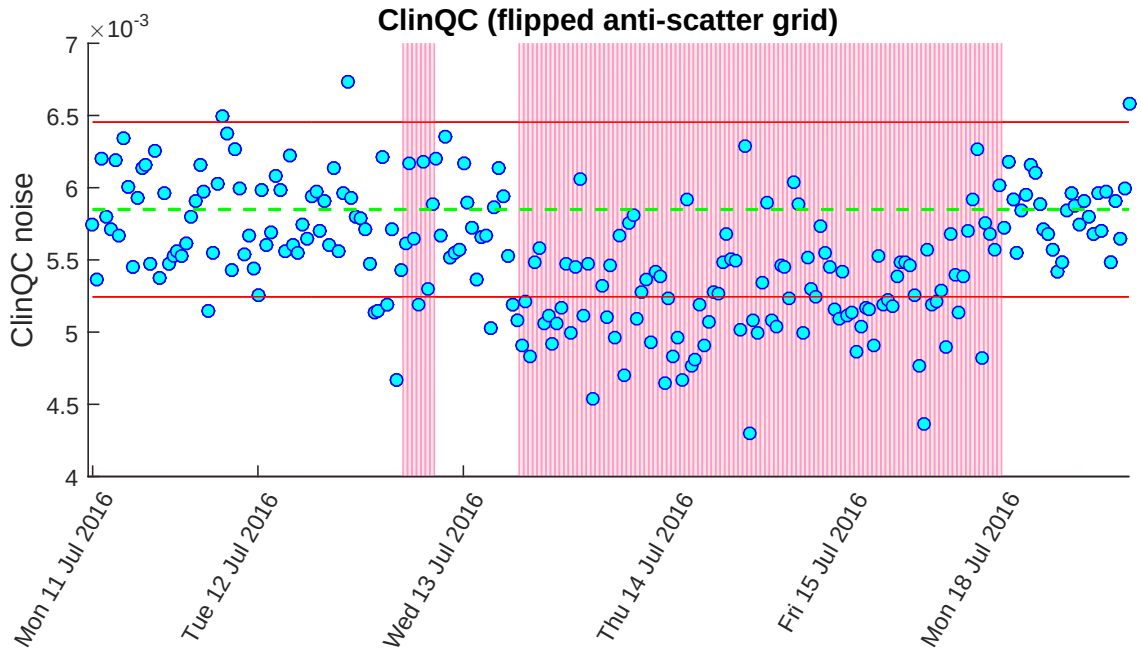

Figure 3.54: The result of the alerts produced by a moving window of size 12 images moved along the ClinQC noise values of all the images of the complete week when the anti-scatter grid was flipped in July 2016.

images as features. The example of the flipped anti-scatter grid could be partially used as training set, to train the classifier to recognize future occurrences of the same situation from the knowledge about the combination of noise and dose values. This hypothesis is supported by one first simple representation (Fig. 3.55) of the distribution of regular clinical data (in blue) and clinical data from images acquired with the flipped anti-scatter grid (in red). This diagram shows that the data of the two classes can be separated which suggests that these two features could be the correct input for a classifier used with the purpose to recognize if the anti-scatter grid has been flipped. More thorough analyses and the implementation of the classifier need to be performed in order to investigate if the use of pattern recognition tools could be a way to find a better detection parameter for the QC, and if this can be applied for the detection of other problems in the imaging system.

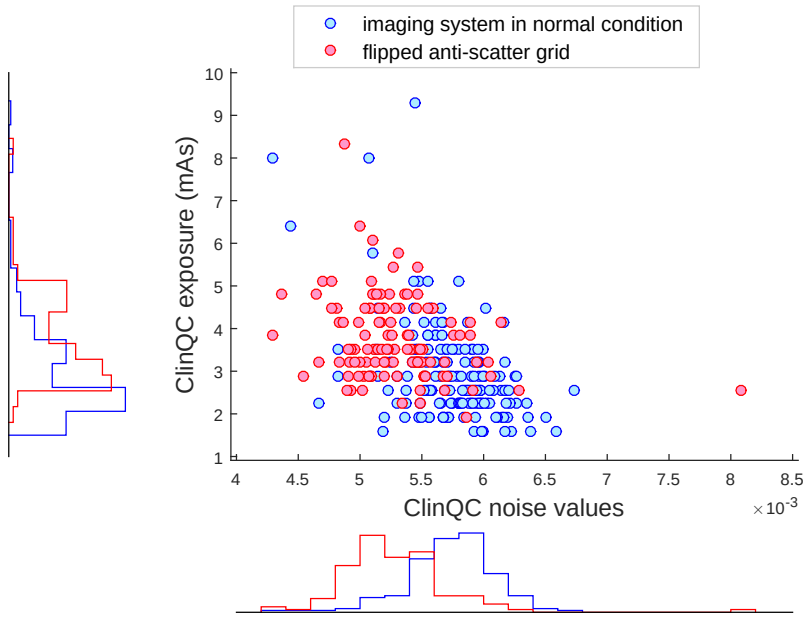

Figure 3.55: Representation of ClinQC noise values and exposures recorded from regular clinical data (in blue) and clinical data from images acquired with the flipped anti-scatter grid (in red), with lateral distributions.

3.2.5.2 Detection of anti-scatter grid replacement

The second example of the performances of the ClinQC detection algorithm in clinical practice regarded the anti-scatter grid update that happened at the end of 2014. The noise values and the exposure of a selection of images acquired from September 2014 until February 2016 are displayed in a time line in Fig. 3.56. This example required a "backward" implementation: the moving window of 19 images was moved backward to simulate the detection of a downgrade of the imaging system, so a passage from the actual anti-scatter grid to the old one. The result of the detection is successful also in this case and without any false alerts, since the period where the old anti-scatter grid was detected with a continuum of alerts between September 2014 and November 2014 where the old anti-scatter grid was still being mounted on the actual imaging system.

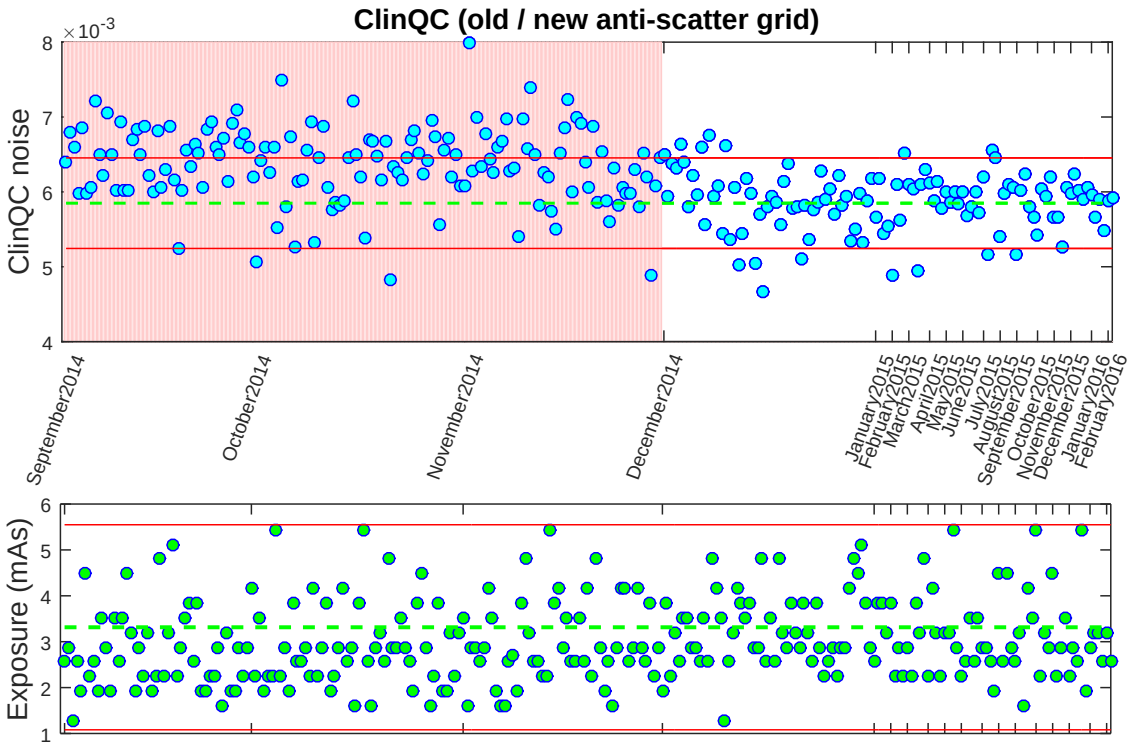

Figure 3.56: The result of the alerts produced by a moving window of size 19 images moved along the ClinQC noise values of a dataset of clinical chest X-ray images at the turn of the anti-scatter grid update occurred at the end of 2014.

3.2.6 Outlier analysis

In Fig. 3.57 the ClinQC noise values of all the clinical images acquired with the actual system setup are displayed in a time line, and the outliers are marked in red according to the acceptance limits derived from the baseline (red lines). The exposure levels of the same images are represented in Fig. 3.58, with the mAs of the ClinQC noise outliers marked in red. Only 10% of the images with noise outlier values were made with outlier exposure values, so it is not possible to imply that there is a causality between an outlier exposure value and an outlier noise value.
Fig. 3.59 shows that 70% of the ClinQC noise outliers are females, while only 30% are males. It has been possible to retrieve the sex of patients undergoing a chest X-ray examination from the PACS, without extracting all the images of one year: among approximately 15 thousands patients per year, only 47% are females. So in this case is possible to establish that there is an influence of the sex of the patients

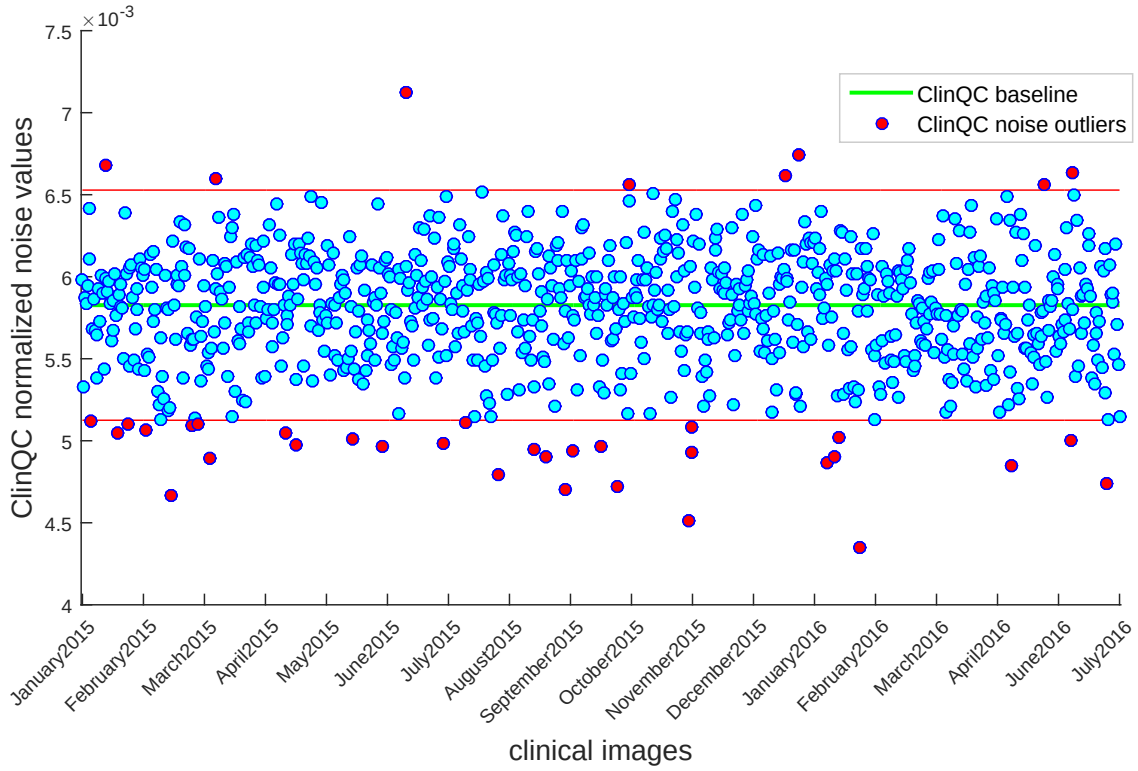

Figure 3.57: The ClinQC noise values of all the clinical images acquired with the actual system setup displayed in a time line; in red are marked the outliers according to the baseline red acceptance limits.

on the chance of finding the ClinQC noise outliers: in particular, the probability that a female patient produce a noise outlier $P(o \mid f)$ is almost 3 times higher than for a male patient (respectively 7% and 2.6%). This can be obtained from a simple probabilistic analysis applying

$$N_{female\ outliers} = N_{females} \cdot P(o \mid f).$$

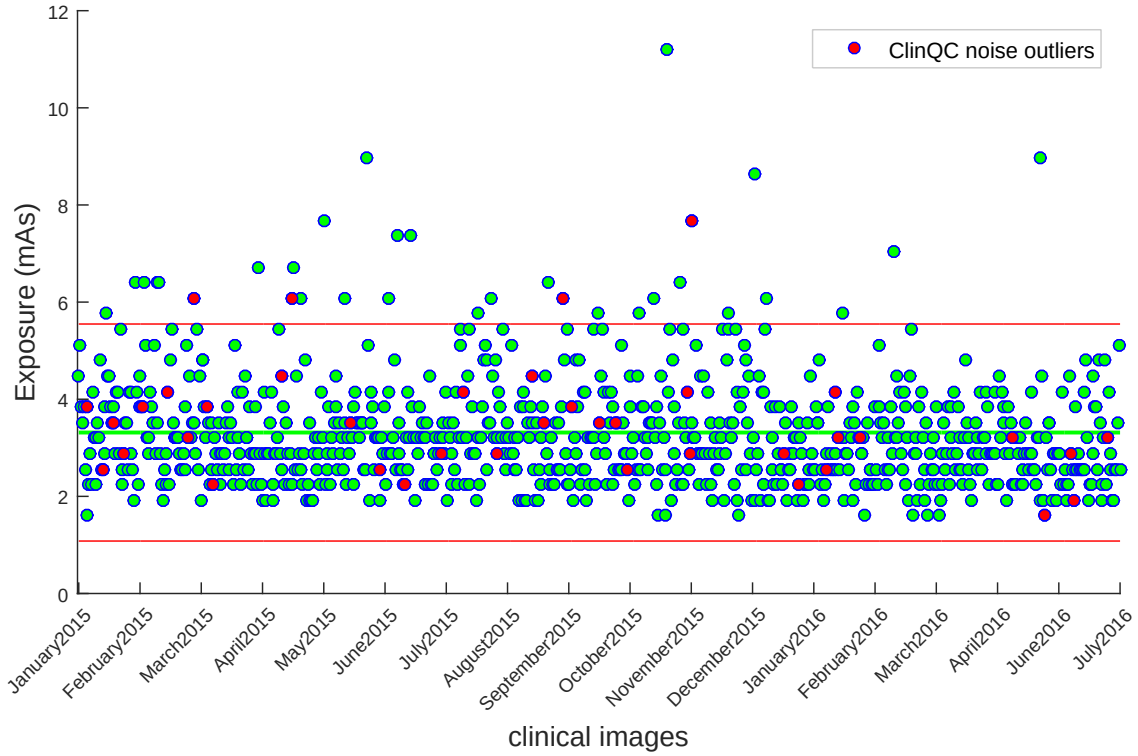

Figure 3.58: The exposures of all the clinical images acquired with the actual system setup displayed in a time line; in red are marked the exposures of the outlier images in terms of the noise (Fig. 3.57).

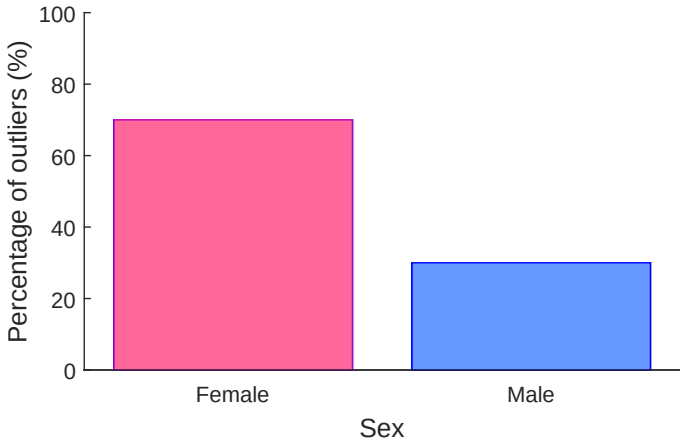

Figure 3.59: Percentage of outliers in terms of the noise that are images of female or male patients.

3.2.7 Image Pyramids noise extraction algorithm comparison

An example of the chest noise maps extracted with the Pyramids method is represented in Fig. 3.60 as first level of the Laplacian Pyramid, and Fig. 3.61 shows the same image normalized by the first level of the Gaussian Pyramid. It can be observed that in the left noise map the anatomy of the patient is still prominent, like in the intermediate image that the ClinQC algorithm produces; while, after the normalization, only the main external edges of the patient body can be distinguished.

Figure 3.60: L_1 Pyramid image example.

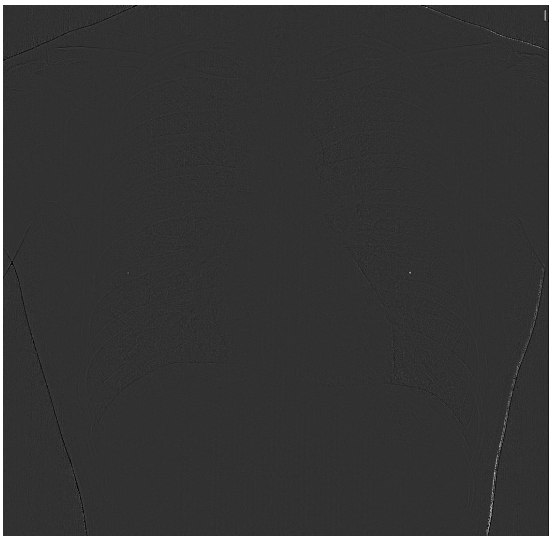

Figure 3.61: $L_{1\ norm}$ Pyramid image example.

Also the *high-low frequencies plots* (Fig. 3.62 and Fig. 3.63) and the *histograms* (Fig. 3.64 and Fig. 3.65) of the extracted Pyramid images show the same characteristics observed for the ClinQC intermediate images and normalized noise maps. Also in this case, taking a central ROI in the noise images is a way to reduce in the noise measurement the anatomical information that is present in these maps.

The comparison between the ClinQC normalized noise values and the Pyramid normalized and not normalized noise values measured on all the images forming the baseline dataset is shown in Fig. 3.66.

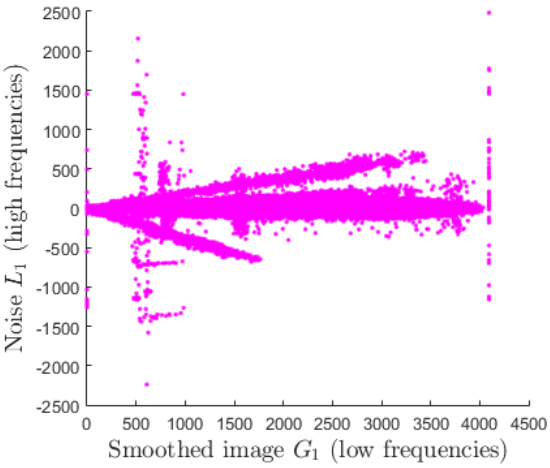

Figure 3.62: *High-low frequencies plot* of L_1 Pyramid image.

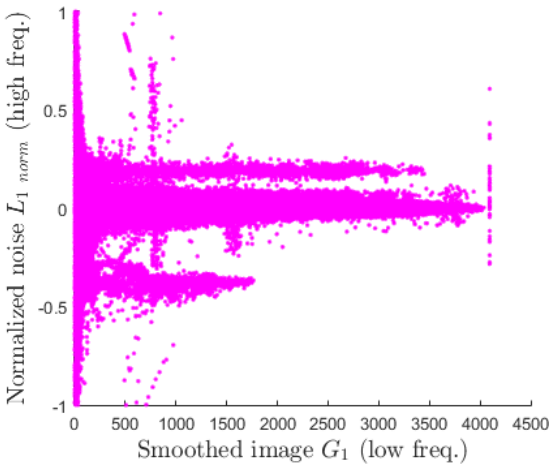

Figure 3.63: *High-low frequencies plot* of $L_{1\ norm}$ Pyramid image.

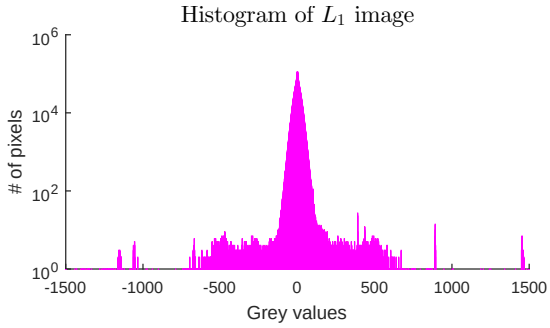

Figure 3.64: Histogram of L_1 Pyramid image.

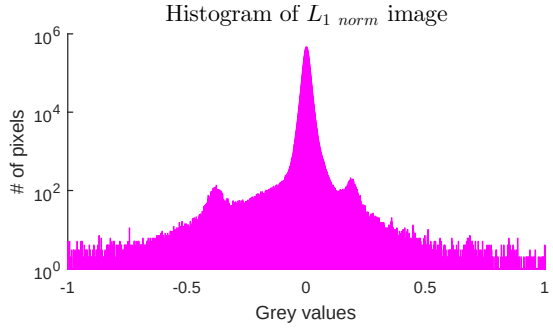

Figure 3.65: Histogram of $L_{1\ norm}$ Pyramid image.

The ClinQC algorithm and Pyramid (normalized) algorithm have 55% of their outliers in common, while the ClinQC algorithm and Pyramid algorithm have only 9% of outliers in common. The CoV of the ClinQC noise values is 5%. Instead, the CoV of Pyramid normalized noise values is 8.5%. Pearson correlation coefficient computed for the noise values obtained with the ClinQC algorithm and the Pyramid (normalized) algorithm is 0.83, the clear correlation is visually represented also in Fig. 3.67.

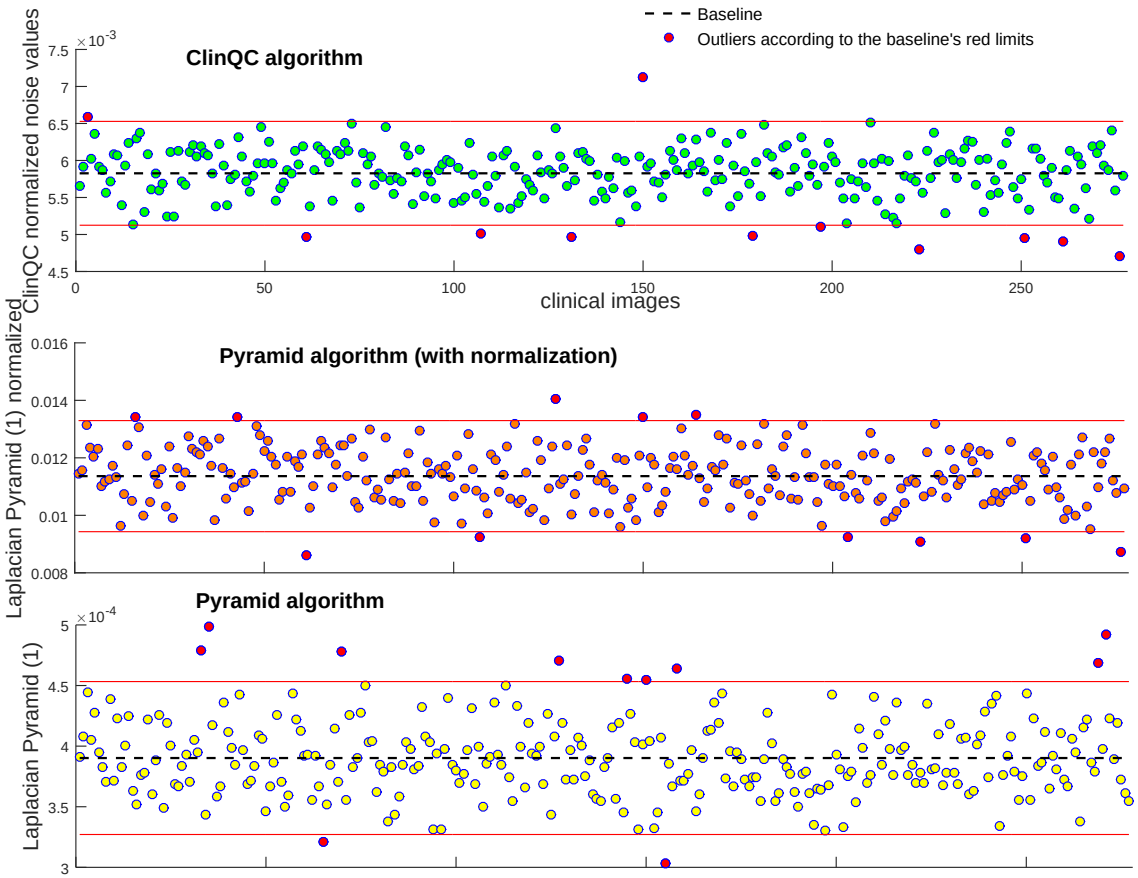

Figure 3.66: The comparison between the ClinQC normalized noise values and the Pyramid normalized and not normalized noise values measured on all the images forming the baseline dataset.

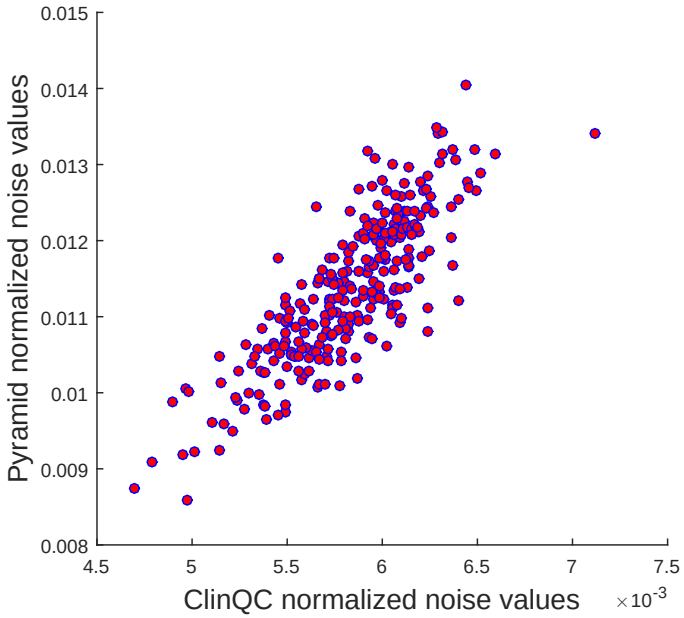

Figure 3.67: The correlation between the ClinQC algorithm and the Pyramid (normalized) algorithm.

3.2.8 Phantom comparisons

In Fig. 3.68 the red dataset represents the noise values decrease observed by acquiring 4 RANDO and 4 QClight images at increasing mAs. A clear correlation between the two methods can be observed with a high Pearson correlation coefficient of 0.98. Fig. 3.68 also displays the simulated exposure decrease achieved by blurring the images acquired at 1.6 mAs with increasing σ of Gaussian blurring kernels (in blue). The simulated and the real mAs increase lead to noise values that show very similar trends, which however do not perfectly overlap.

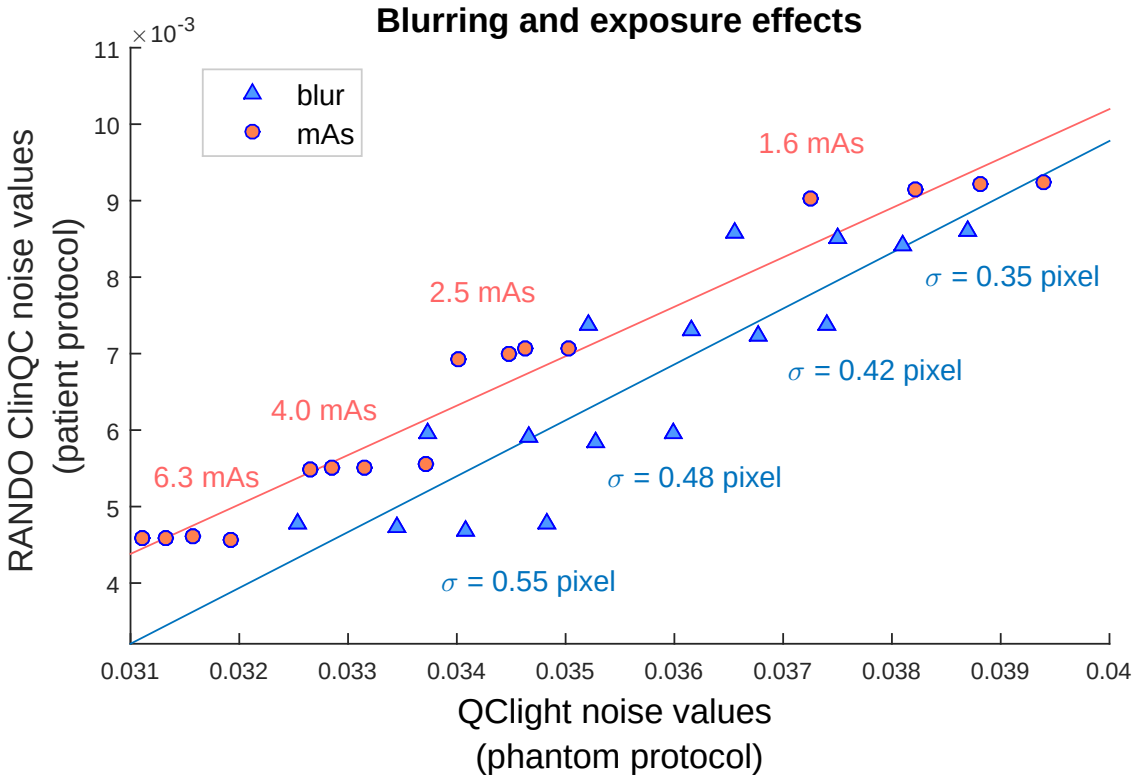

Figure 3.68: The blurring-induced decreasing effect (in blue) and the exposure-induced increase (in red) of the ClinQC noise values, measured on RANDO images, and the QClight noise values.

To better understand these two effects the 2D NPS of the images of both phantoms was computed: Fig. 3.69, Fig. 3.70 for the QClight images and Fig. 3.71, Fig. 3.72 for RANDO images. As can be observed, the increasing exposure during the image

acquisition does not significantly change the frequency distribution of the noise for both phantoms, while the blurring effect suppresses the high frequencies of the noise more efficiently, being a low-pass filter. This slightly different behaviour can explain why with an image deterioration study with blurring is not able to achieve exactly the same decrease of the noise values observed in Fig. 3.68 acquiring the images at different exposure levels. This is a limitation to be considered in the interpretation of the results of the image deterioration study with blurring.

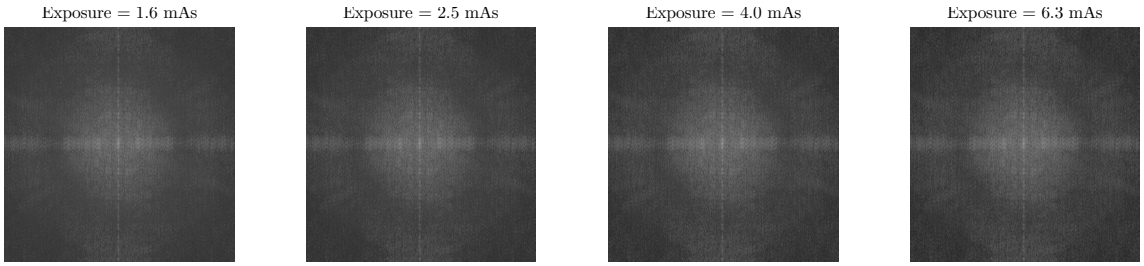

Figure 3.69: The 2D NPS of the QClight images acquired at different exposure levels.

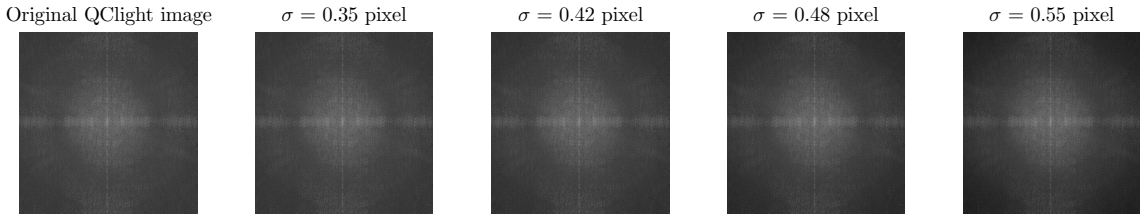

Figure 3.70: The 2D NPS of the QClight images blurred with different σ of Gaussian blurring kernels.

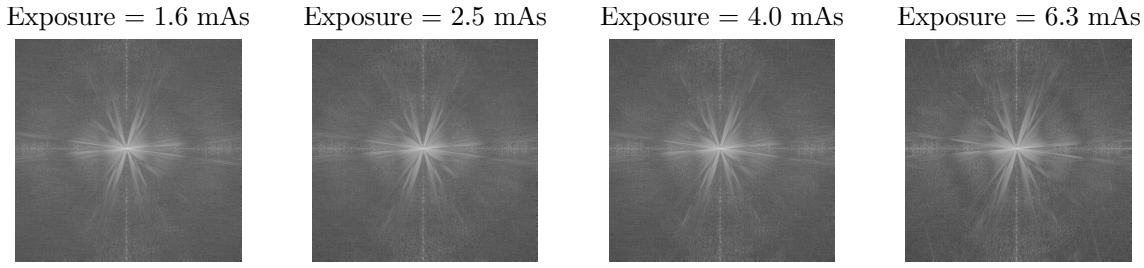

Figure 3.71: The 2D NPS of RANDO images acquired at different exposure levels.

Fig. 3.73 shows a comparison between the ClinQC and the QClight methods: the graph shows the noise values obtained applying the ClinQC algorithm on one clinical

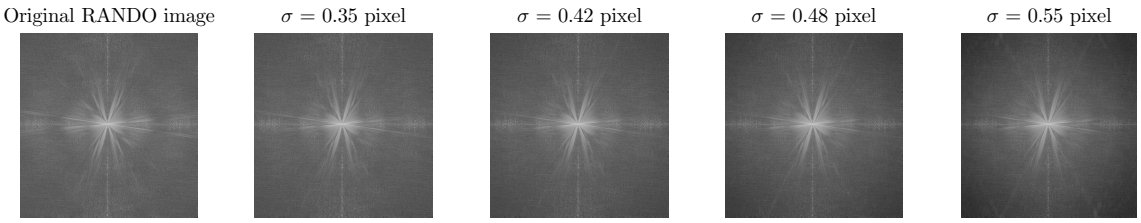

Figure 3.72: The 2D NPS of RANDO images blurred with different σ of Gaussian blurring kernels.

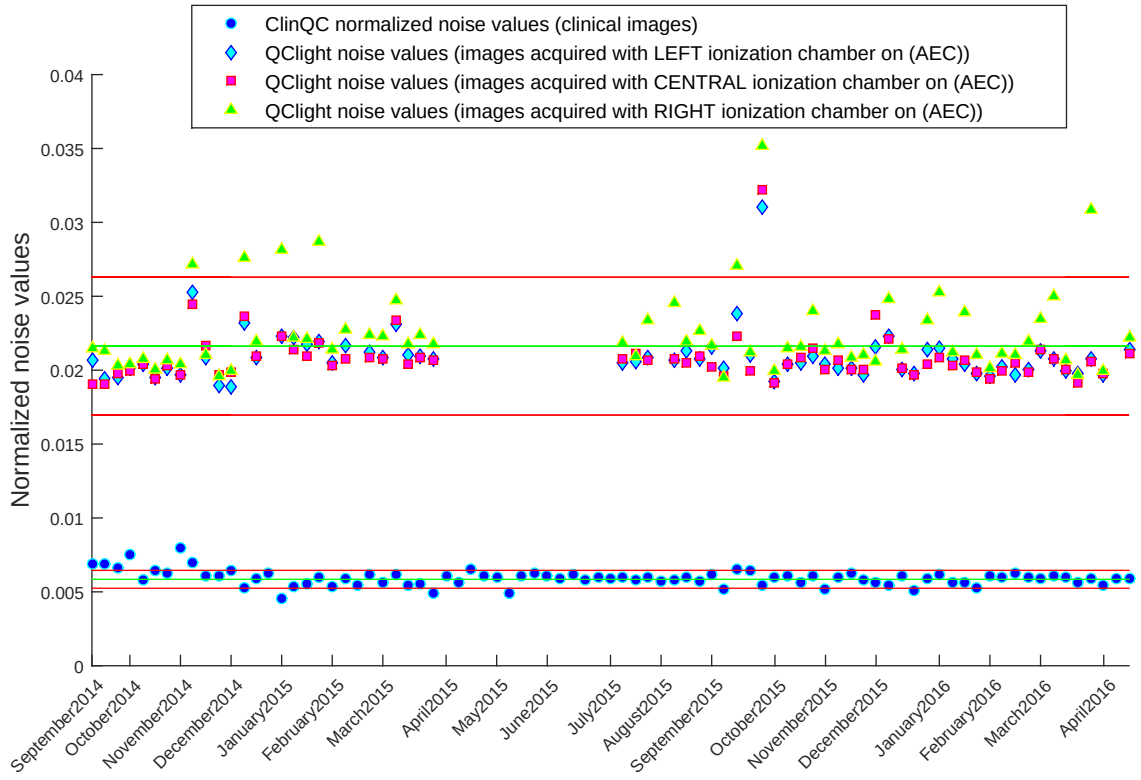

Figure 3.73: A comparison between the ClinQC and the QClight methods applied on the history dataset of, respectively, one clinical chest X-ray image per week and one phantom image per week.

image per week (in blue), acquired each Monday morning in the same time line of the QClight dataset. Only with ClinQC it is possible to observe the initial globally higher noise levels, corresponding to the period when the old anti-scatter grid was installed. This is not observable using the QClight method since the focal distance of the old grid was comparable to the SID setted for the QClight images, so the old grid was optimal for the phantom-based QC acquisition protocol, and the measured

noise was generally lower than the one registered after 2014 with the actual imaging system setup. After 2014, in fact, the new anti-scatter grid was optimal for the patients acquisition protocol and the noise in patient images was reduced by 6.4%, while with the QClight method the noise increased by 2.7%.
The CoV of the QClight history data after the anti-scatter grid update is 10%, against the 5% of the ClinQC baseline dataset.

3.3 Discussions

In this Section, all the interpretations of the results of the ClinQC noise study and the assumptions will be presented.
Since patient images are always double-processed, the implemented unknown filters and corrections might change a lot the characteristics of the raw patient images affecting the QC measurements. In the end, the information that we can get about the imaging system studying its clinical images is altered by all these operations. The assumptions that we can make after analysing the results are slightly weaker than if we had the chance to work directly on the unprocessed raw images. This is one of the main limitations of the ClinQC project and is one of the reasons why working on clinical image quality is a big challenge, especially with the purpose of deriving secondary information about the system quality directly from clinical images. For example, part of the noise is always lost after the image processing, and some filters or the flat field correction might flatten all the inhomogeneities. Even with these limitations, there is still a lot of interesting work that can be done towards this direction.

3.3.1 The ClinQC algorithm: noise extraction from clinical images

3.3.1.1 Properties of the ClinQC extracted noise images

The results in Section 3.2.1.1 show why the ClinQC noise extraction algorithm needs a normalization step. If there is a slope in the *high-low frequencies plot*, a correlation between high and low spatial frequencies, it means that the noise map extracted at the first step of the ClinQC algorithm is still patient-dependent. Because there is still the influence of the anatomy of the patient, represented by the clusters of pixels in the noise map that correlates with the body signal, and this fingerprint is also clearly visible (Fig. 3.3).
That is why in the ClinQC algorithm a normalization is included, dividing the high frequencies map (representing mainly noise) by the smoothed image a more patient-independent noise map can be observed (Fig. 3.4). The normalized *high-low frequencies plot*, computed on the final noise image after the normalization step

(Fig. 3.12), demonstrate that the correlation between the noise and the anatomical signal is reduced by the normalization.
But where do the pixels in the lateral clusters come from? The answer to this questions is the first step towards the comprehension of the ClinQC algorithm functioning. As described in the Results (Section 3.2.1.1), the lateral clusters belong to some regions in the image where the presence of the noise increased with the intensity of the body signal and where, after the normalization, the noise intensity has an offset with respect to the average noise level distributed around 0. Those pixels represent the presence of the patient in the scene and not the noise intrinsically generated during the image acquisition: that signal mustn't be included in the QC measure, otherwise the ClinQC noise value will be a patient-dependent metric, not relevant for the system stability investigation that have to be performed. Instead, the pixels in the noise map that contain only the noise generated by the imaging system and the patient during the image acquisition belong to the central cluster of the *high-low frequencies plot* and they should be the only ones involved in the ClinQC algorithm noise measurement. From the study of the properties of the ClinQC extracted noise images using the *high-low frequencies plot* and the *histogram* (Section 3.1.1.1), and after a simple image segmentation, it has been possible to understand that the ClinQC normalized noise map has only a small remaining trace of patient-dependent noise which is visible and measurable only over the external edges of the patient body and that can be easily cut off taking a square ROI in the center of the image, since the positioning of the patients is standardized and that region does not contain the external body edges. The properties of the ROI pixel values (*high-low frequencies plot* in Fig. 3.15 and *histogram* in Fig. 3.16) enlightened this conclusion. In this way, selecting an ROI in the center of the chest radiography, a less patient-dependent noise measure is obtained from the ClinQC normalized noise map; and the ClinQC noise values represent the noise generated by the imaging system during the clinical image acquisition without a strong influence of the presence of a patient.

3.3.1.2 The ClinQC noise values: baseline

As shown in Section 3.2.1.2 the distribution of the ClinQC noise values in the baseline dataset can be considered normal with small values of skewness and kurtosis. The skewness value is negative and small, this means that some patient images contains higher noise values than the average. Since the relationship between the noise and exposure in radiographic images is known, the ClinQC normalized noise map lead to measure a noise-to-signal ratio that is proportional to $1/\sqrt{mAs}$, and Section 4.2.1 shows that also the distribution of the exposure in the baseline dataset has a tail, and the peak is shifted to smaller mAs than the average. This is due to the AEC device that allows different mAs to different patients according to their body thicknesses and attenuations. Two main reasons can explain these shapes. First, the average body weight in the world is not symmetrically distributed but skewed, the peak is shifted towards smaller weights [22], so the smaller thicknesses of people could lead to non-symmetrically distributed exposures, that are then inversely reflected in non-symmetrically distributed noise values. For a complete analysis it should be better to study the distribution of patient weights only on patients that come for a chest examination in the selected X-ray room. Second reason, the optimization of the acquisition protocol could have been set to give always the lowest possible dose to patients in order to have a uniformly distributed quality of the images, so the exposure could be set non-linearly to the patients body weight. This could introduce a worse image quality in some clinical images, that could be corrected in the post-processing, in order to have a uniformly distributed image quality for all the patients even if the exposures are not linearly distributed. This has not been verified in the scope of this Thesis.

3.3.1.3 The ClinQC noise values: Relevant clinical examples

The interesting historical examples shown in Fig. 3.20 are useful to realize that the ClinQC noise values really change significantly if the imaging system is different or if something in the imaging system changes, like the focal distance of the anti-scatter grid, its presence or its correct positioning. It is easy to understand why the noise is higher without the anti-scatter grid since it is placed there to remove the noise and enhance the sharpness, and the reason why the noise is lower with the new

installed anti-scatter grid is because it has an optimal focal distance for this specific X-ray examination, since it is comparable to the source-patient distance of 2.00 m. These were interesting examples of the detection of the quality and the settings of the system but they were not strong enough to be used as the only source of validation for the project, this is the reason why other sources of validation have been implemented in this Thesis work, such as the image deterioration study, the statistical analysis with simulations and the performance of the ClinQC algorithm in clinical practice.

3.3.2 The ClinQC algorithm - alternative versions

3.3.2.1 Grid sampling approach

The grid sampling approach demonstrated to be the first possible improvement to the ClinQC noise extraction algorithm. The importance of this addition to the algorithm is due to the fact that many patients that come for a chest X-ray examination have some artificial objects placed inside or outside their body. These sharp structures leave clear signs in the noise maps extracted with the ClinQC algorithm, and this additional signal is included in the noise measurement, giving a higher estimation of the noise level than the true value. With the grid sampling approach, it is possible to remove the influence of these outliers in the noise distribution on the final ClinQC noise value. The median value of the distribution is less sensitive to implant induced deviations and hence better represents the true value of the noise generated during the imaging examination that the ClinQC algorithm wants to find. For this reason the ClinQC-alternative algorithm results may be more robust and efficient even in presence of patients with implants, catheters, pacemakers or other medical devices covering the chest area.

3.3.2.2 ClinQC applied to mammography

From the preliminary results presented in Section 3.2.2.2 we understood that more accurate analyses are required to investigate if the kVp could have played a key role in giving a higher spread in the noise values than in the chest baseline dataset. Probably, in addition to this adaptation of the acquisition protocol to different

patients, the higher spread is due to the fact that two different obtained types of images are considered together in this preliminary study.
However, the algorithm with some small updates seems to be applicable also to other imaging anatomies, giving normalized noise maps with the same visual appearance as for the chest. More analyses are needed in order to study the properties of these images and to tune the algorithm parameters, before applying this QC algorithm in clinical practice.

3.3.3 Image deterioration study

With the image deterioration study we proved that the ClinQC noise extraction algorithm provides a highly sensitive parameter for the noise measurement in an X-ray imaging system QC.

Blur With a blurring deterioration it has been possible to observe a fast decrease in the ClinQC noise values. In particular, its response curve to the introduced smoothing presents a high slope, which means that even small changes in the noise of the imaging system are translated in high variations of the ClinQC noise values. The simulation of a decrease in the noise values represents, as explained in Section 3.1.3.1, an increase in the quality of the imaging system, that can occur if a new proper anti-scatter grid is installed for example, or it can represent an increase in the exposure (mAs) given to the patients during the image acquisition. This is why from the ClinQC noise extraction algorithm it is possible to double check the AEC device that controls the flux of X-ray photons that are released from the tube, pass through the patient and reach the detector. If there is a malfunctioning of this complex system the ClinQC algorithm will detect it both with the ClinQC noise values and with the exposure registered in the DICOM tags of each image.
The increasing behaviour of the relative spread, that we observe the more we smooth the images, means that the relative distribution of the noise values becomes wider the more the noise is artificially reduced: so we might erroneously imply that low noise values correlate with wider noise distributions, and this assumption might lead to the premature consideration that if the noise in an imaging system is lower than the nominal value (baseline), so for example with future or more optimal imaging

systems, and the relative noise distribution computed from clinical images is wide, it becomes harder to detect if something in the system is changing. But these are simulated images. To make assumptions like these, the distributions of many different imaging systems should be compared, while the validation presented in this Thesis is made on one imaging system only.

Gaussian noise Also in this case we proved that the ClinQC algorithm produces a highly sensitive parameter to detect an increase of the Gaussian noise, which is a way to simulate the deterioration of the imaging system, that can occur when the detector or the electronics of the system reach their end of life, or if the anti-scatter grid is not present into its slot during the image acquisition, or if there is an increase in the exposure (mAs) given to the patients. The Gaussian noise symmetric distribution was chosen to simulate in the simplest way the effect of all the possible combined deterioration phenomena occurring in an imaging system. More specific simulations can be performed using different noise distributions.

3.3.4 Statistical analysis with simulations

3.3.4.1 Step simulation

The step simulation output curve (Fig. 3.43) mainly proves that in only one day, with 50 new clinical chest X-ray images, it would be possible, by applying the ClinQC detection algorithm, to observe an increase in the noise levels of 4% with respect to the baseline, that can be linked to a decrease in the exposure (mAs) given to the patients of 8%. This is important not only to check the noise levels of the imaging system potentially on-the-fly, depending on when the extraction of the clinical images from the PACS is planned, but also to have an automatic report about the functioning of the AEC device.

3.3.4.2 Trend simulation

The trend simulation was implemented to simulate two different linear time dependent deteriorations of the imaging system: a strong deterioration that produced 30% increase of the noise values happening in one week, and a weaker one reaching

the same noise levels in a slower time line of one month. The output curves of the trend simulation (Fig. 3.48) can be used to understand the optimal size of a moving window that is moved along the ClinQC noise data in order to detect a certain percentage of increase of the noise in one week or one month. These simulations are dependent on the simulated slopes and are a particular case of the step simulation, they give specific examples of application of how the ClinQC detection algorithm would work if applied in clinical practice.

3.3.5 The ClinQC performance in clinical practice

The performance of the ClinQC noise extraction algorithm has been tested with two important cases for clinical practice, apart from any kind of simulation. These tests proved that the ClinQC noise extraction algorithm can be used to implement a powerful detection tool, that extracts every night the images of the patients acquired during the day from the PACS and perform a statistical analysis on their ClinQC noise values in order to investigate the stability of the noise of the imaging system. In addition, the step simulation output curve demonstrated to be accurate enough in its approximation to be used as a guideline for the detection of specific steps of noise levels increase or decrease in percentage.
If with the QClight method there is 1 $^{\text{alert}}/_{\text{week}}$ with the Monday morning measurement, with the ClinQC method it could be possible to extract all the images of the day from the PACS during the night and have a detection parameter that can give 1 $^{\text{feedback}}/_{\text{day}}$ defined as the number of alerts per day normalized by the number of possible step-by-step movements of the moving window. This will estimate each day if there are serious problems in the imaging system and if it is necessary to perform an additional QClight phantom analysis. If this tool would have already been available for clinical practice in that period, the hospital might have avoided to acquire 72 new patient images with a lower diagnostic image quality due to the flipped anti-scatter grid.
Since in these first examples the few image outliers in terms of the noise that were also outliers in terms of the exposure were left out from the analysis to help the detection algorithm to better observe a contrast in the noise levels before and after a change in the system setting, it would be interesting to try to find a new clinical

detection parameter that combine these two measures of noise and exposure to optimize the detection instead of excluding them from the analysis. This search could include also some machine learning techniques.
The detection of the flipped anti-scatter grid was completely successful, but produced also some false positive alerts that are not relevant compared to the number of displacements that the moving window can do every day on the total number of images. In this specific case the number of false positive alerts is only the 10% of all the possible alerts that the algorithm could have produced in one day.
Also the detection of the anti-scatter grid replacement, that required a "backward" implementation of the moving window algorithm, was successful without any false positive alerts. This proved that the step simulation predicted well the optimal size of the moving window that has to be used for a correct check in case the anti-scatter grid now in use is replaced with another one with smaller focal distance.

3.3.6 Outlier analysis

The ClinQC noise outlier analysis proved that there is a really low probability (10%) that an image with outlier noise value was made with an outlier exposure value. This can be useful knowledge to guide the invention of a new ClinQC detection parameter that can be used in clinical practice (see previous discussion in Section 3.3.5), but more studies need to be performed in order to establish what the other factors are that induce the presence of noise outliers even when the mAs falls within the accepted range.
Instead, there is a higher chance that the ClinQC algorithm results in outlier noise values if the patient is female than male. This can be related to the fact that the major volume of female breast tissue causes more scattering of the radiation and hence increases the noise. It would be interesting to study if outlier noise values can be correlated also to patient age, using hospital data.

3.3.7 Image Pyramids noise extraction algorithm comparison

The noise maps extracted with the Pyramid method on the baseline dataset show the same properties as the ClinQC noise maps. A clear correlation between the ClinQC and the Pyramid (normalized) algorithms is representative of the fact that the two methods are based on the same idea while their implementation is different. The higher relative spread of the normalized Pyramid algorithm might suggests that it would be more difficult to detect some changes in the noise values if this methods is applied in clinical practice instead of the ClinQC. However more analyses need to be performed to establish whether or not Pyramid noise parameter would be less sensitive than the ClinQC noise value.

3.3.8 Phantom comparisons

A high correlation coefficient between the ClinQC noise values, measured from RANDO images acquired with patient protocol, and the noise values of the QClight images, acquired with phantom protocol, indicates that both methods are sensitive to a noise reduction produced by an exposure increase and either of them can be used to measure the noise of the imaging system.
The results in Section 3.2.8 demonstrate a limitation of using the blurring method in the image deterioration study to simulate an exposure increase in clinical images, however since the difference between the real and simulated phenomena is small, and can be explained by analysing the effect of blurring to the noise frequencies of both phantoms images, the image deterioration study with blurring was adopted in order to simulate an exposure increase in clinical images.
Finally, the visual comparison between the phantom-based QC and the ClinQC method showed that the ClinQC algorithm is more representative of what happens to clinical images with the actual imaging system setup. The CoV of the ClinQC baseline noise values is half the CoV of the QClight method, and since these two methods correlates, so they reacts in similar ways to the same changes in the exposure levels, it can be concluded that the ClinQC method detects similar changes in the device quality, with a sensitive and reliable noise parameter.

Chapter 4

Dose study

The dose study was conducted in parallel with the noise study and it contains preliminary results and examples.

4.1 Methods

The dose results can be extracted by applying the ClinQC algorithm in order to have an overview of the performances of the imaging system. In fact, the ClinQC algorithm for noise extraction was designed to retrieve dose related quantities from the DICOM tags of clinical images. In this way it is possible to monitor both the noise and the dose parameters from clinical images.

4.1.1 Exposure and tube output

The **Exposure** of patients to X-rays during a radiological examination is defined by the current-time product (mAs), which combines the time of radiation exposure (in seconds, s) and the tube current (mA). This is the current of electrons being accelerated from the X-ray tube cathode towards the anode where they produce X-rays (bremsstrahlung) by interaction with the anode. An increasing tube current or exposure time results in more X-ray photons, and thus a higher patient dose. The dose is proportional to mAs but also depends on other parameters, such as the kVp and filtration which shape the X-ray beam spectrum and the SID that influences the dose according to the inverse square law. The Jarque-Bera (J-B) test [23] was

used to test the normality of the baseline exposure values distribution. The use of this parametric test based on the sample skewness was preferred to the K-S test in this specific case, since the mAs distribution (Fig. 4.1) was skewed with a longer tail compared to the ClinQC noise values baseline distribution (Fig. 3.18) and the J-B test proved to have better performances in these cases [24]. The exposure and the ClinQC noise-to-signal ratio, obtained after the noise map normalization, vary together following an inverse relationship, that is expected to be

$$\frac{N}{S} \propto \frac{1}{\sqrt{mAs}},$$

considering that the noise in digital radiographies depends on the number of discrete X-ray photons reaching the detector, and so it increases with the square root of mAs ($N \propto \sqrt{mAs}$), while the signal increases linearly with mAs ($N \propto mAs$).

The **Dose Area Product** (DAP) is defined as the absorbed dose in air multiplied by the area being irradiated (mGy cm^2). This depends on the current-time product and so a measure of the DAP mainly verifies whether the AEC device works properly. Due to the divergence of the beam emitted from the X-ray source, the irradiated area A increases with the square of distance from the source (A $\propto$ SID^2), while radiation intensity I decreases according to the inverse square of distance (I $\propto$ $^1/_{SID^2}$). Therefore the product of intensity and area is independent to the distance from the source. The diverging beam is collimated for each patient to image only the relevant area. Consequently, not only the dose, but also the area, and thus the DAP are different for each patient. With the phantom-based QC this is not the case, since the geometry of the phantom being irradiated is the same every week. With a clinical images-based QC as ClinQC, the dose parameter needs to be independent to the different patient thicknesses in order to have a measure of the quality of the system that is not affected by patient variability. This is why the tube output (mGy/mAs) is introduced, defined as the dose the X-ray tube delivers per unit of tube load (i.e. charge)

$$Tube\ output = \frac{Dose\ (mGy)}{Exposure\ (mAs)} = \frac{DAP\ (mGy\ cm^2)}{Exposure\ (mAs) \cdot A\ (cm^2)} \cdot \frac{SID^2}{SDD^2},$$

where A is the image area, measured from clinical images as $A = rows \cdot columns \cdot resolution^2$, and SDD is the Source-DAP-meter Distance while the SID is the Source-Image Distance. The main reason to include the monitoring of the tube output in

ClinQC algorithm is the control of the X-ray spectrum that is related to the state of the anode of the X-ray tube and the filters.

4.2 Results

In this Section the preliminary results and examples of the dose study (Section 4.1.1) are presented.

4.2.1 Exposure

The exposure (mAs) of the ClinQC baseline dataset is displayed in Fig. 4.1 with the lateral distribution obtained using a kernel density estimation based on the histogram of data. The distribution is positively skewed with a tail to high mAs, the skewness value is 1.7, and the kurtosis of the distribution is 7.4. The Jarque-Bera test [23] rejected the null hypothesis that the data come from a normal distribution with an unknown mean and variance at 5% significance level, with a p-value of $3 \cdot 10^{-3}$. In Fig. 4.1 the mAs are discretised due to the AEC device design.

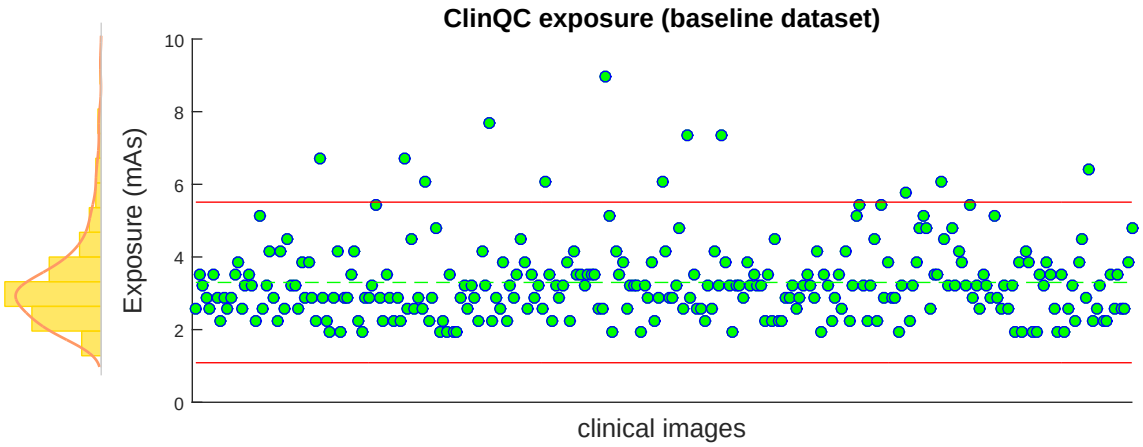

Figure 4.1: Exposure (mAs) of the ClinQC baseline dataset with lateral distribution, with red acceptance limits of $\pm 2\sigma$.

Other exposure results have already been presented alongside the results of the noise study, for example in Section 3.2.5 (Fig. 3.53 and Fig. 3.56) with the performance

of the ClinQC noise extraction algorithm in clinical practice. Here it is clear that the exposure and the ClinQC noise-to-signal ratio, obtained after the noise map normalization, vary together following the inverse relationship presented in Section 4.1.1. This relationship explains the presence of some patient images being outliers in terms of the noise while also being outliers in terms of the dose (Fig. 3.58).
In this case is not useful to perform a comparison between the clinical mAs data and the QClight phantom exposure data, since the geometry of the phantom is the same for each acquisition and also the kVp is fixed by the acquisition protocol, so if the AEC device is working properly the QClight current-time product is always the same value. If there are issues with the imaging system this value will change. This is only marginally indicative of the functioning of the AEC device, that is designed to adapt the current-time product to different patients. The measurements of exposure with different attenuations (e.g. patients) is much more telling of the performance of this device over a range of exposures.

4.2.2 Tube output

In Fig. 4.2 a representation of the stability of the tube output measured from the baseline dataset of clinical images of April 2015 is shown. The tube output combines the dose, measured by the DAP-meter, with the exposure current-time product, set by the AEC device.In a phantom-based QC all parameters in the tube output calculation would be constant, with the exception of the dose measured by the DAP-meter. Since the tube output computed with the QClight method from phantom images would be proportional to the DAP, the ClinQC tube output retrieved from clinical images is then compared to the QClight DAP in Fig. 4.2.
The CoV of tube output obtained from the clinical images of the baseline dataset is 7.7% while the CoV of the DAP values registered for the QClight phantom images is 3.2%.

The step simulation, presented as a validation tool for the ClinQC noise extraction algorithm (Section 3.1.3.2.1), has been applied to assess how many images are needed to detect a change in the clinical tube output values, producing the curve in Fig. 4.3. The result shows that in one day, with approximately 50 clinical images, it is

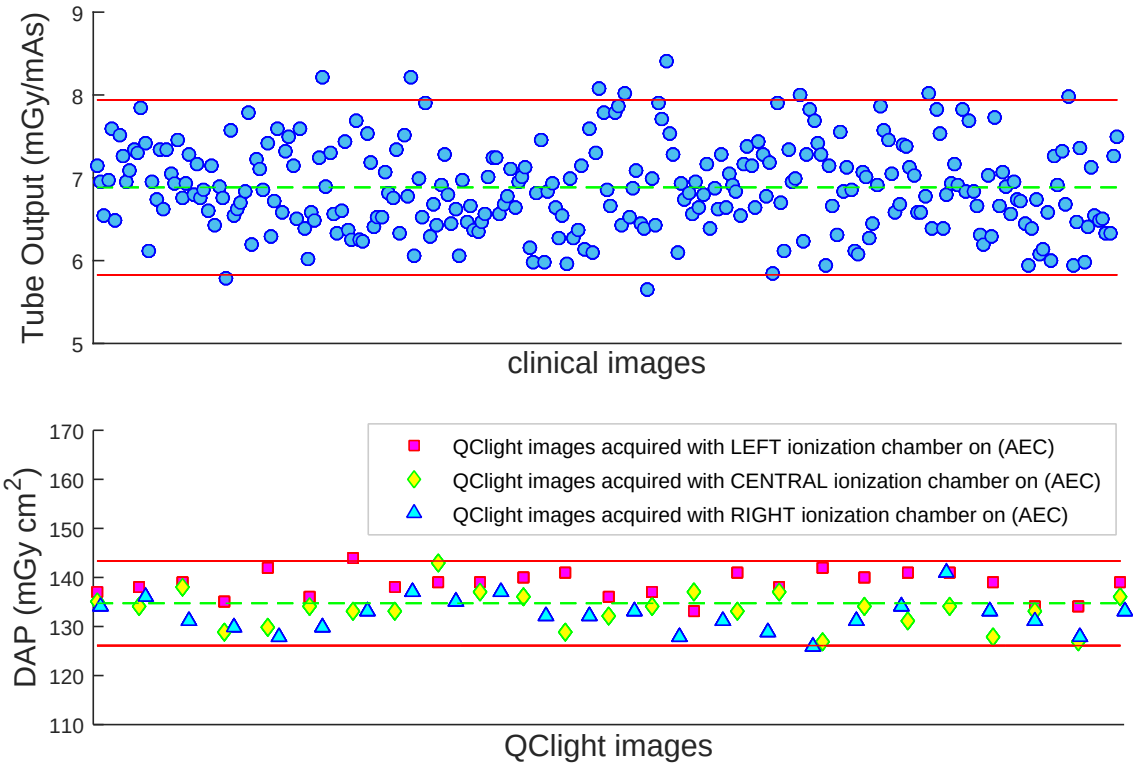

Figure 4.2: Tube output (mGy/mAs) of the ClinQC baseline dataset and DAP ($mGycm^2$) of the QClight history data of 2015/2016.

possible to detect a change in the tube output levels of at least 6%, attributable to deterioration of the anode of the X-ray tube or the filtration system.

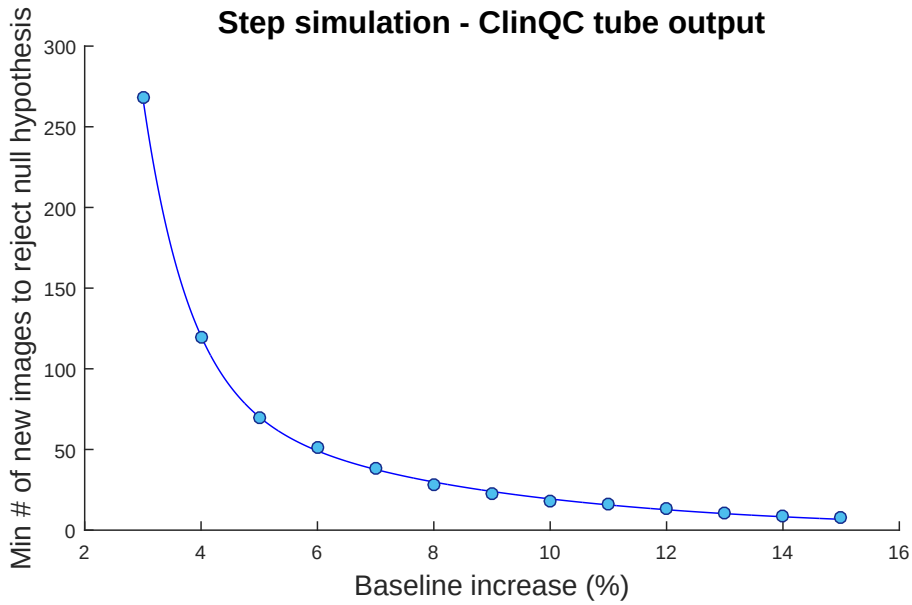

Figure 4.3: Output curve of the step simulation applied to clinical tube output values.

4.3 Discussion

With the dose study it has been possible to obtain some preliminary results, regarding the exposure (mAs) and the tube output (mGy/mAs), retrieved and computed from the DICOM tags of clinical and the QClight phantom images.
In particular, the exposure results obtained with the clinical images-based QC were analysed together with the ClinQC noise results and are more representative of the functioning of the AEC device as an entire range of attenuation is covered, rather than just one single phantom using QClight method.
The tube output that was introduced in the clinical-images based QC, was compared to the DAP of the phantom-based QC. Being independent of patient size and having a relative spread of the same order of magnitude as the QClight DAP values, the tube output can be considered a new dose indicator to be used for QCs of X-ray imaging systems from clinical images.
These dose parameters can be used not only to monitor the levels of dose absorbed by the patients, but also to monitor the imaging system functioning, from the X-ray tube to the AEC device, including the patient acquisition protocol. A change in the exposure levels can be also indicative of a change in anti-scatter grid positioning or

focal distance. To understand if this and other simulated effects can be detected also by the tube output parameter, a proper validation is needed.

Chapter 5

Homogeneity study

In this Chapter the preliminary results on a homogeneity study in clinical images are presented.

5.1 Methods

As introduced in Section 2.2 the inhomogeneity of the QClight method is computed after having extracted five square ROIs in the phantom image, one in the center and the others close to the four corners. The easiest way to produce a measure of the inhomogeneity of the imaging system from patient images, that is comparable to the QClight parameter, is to segment four ROIs close to the corners of the image that contains the background, and use the same QClight formula (Eq. 2.3) to compute the inhomogeneity value. To segment the background areas close to the corners of patient images, several thresholding algorithms have been tested (Section 5.1.1). Furthermore, in order to try to bypass some of the problems linked to the thresholding algorithms, discussed in Section 5.3, an alternative approach based on image profiles is presented (Section 5.1.2).

5.1.1 Thresholding algorithms

All the thresholding algorithms presented in this Section are intensity-dependent, so they are based directly on the image pixel values or on the image histogram of grey levels. In order to segment the information that is available in the background, all

these methods are separately applied on four rectangular ROIs respectively placed near the corners of each image: the upper left, upper right, lower left and lower right corners, where most of the patients don't cover the entire detector area with their body.
The first segmentation algorithm being tested was **Otsu method** [25] with multi-threshold. This method is widely used in image processing; the original algorithm is based on the assumption that the image histogram has a bimodal distribution, so the two histogram classes, representing background and foreground, can be separated computing the Otsu threshold that maximizes the variance between the classes. The multi-threshold version of the algorithm is more appropriate in this case, since the histogram of chest X-ray images is not bimodal but shows a more complex shape (Fig. 5.3).
Two other segmentation methods being tested are also based on the knowledge of the histogram of chest X-ray images and implement, respectively, a **fixed threshold** at the 350^{th} gray level of the image and a threshold based on the **minimum point** of the first class of the image histogram: these values have been classified as the limit of the background pixel class, as an evidence from many observations of chest X-ray image histograms Fig. 5.3.
The final algorithm being tested is the **Fast Marching Method (FMM)** [26] that is available in MATLAB[1]. This is based on the computed normalized geodesic map of differences in gray scale intensity from the image pixels to a seed. In this case, the seed is chosen as the background average gray value, which can be estimated from the knowledge of the image histogram. Then a threshold level T is chosen, and all the pixels that have normalized geodesic distance values less than T are considered background pixels; changing T, different segmentation results can be obtained.

5.1.2 Normalized profiles

The normalized profiles method was invented and developed as a trial. This algorithm divides one chest patient image in two vertical sides and draws a vertical profile of the pixels on the left rows and one on the right rows (example in Fig. 5.1).

[1]The MathWorks Inc., Novi, MI, USA.

This is done for a large number of patient images of one week when the QClight method recorded a low optimal level of inhomogeneities (approximately 200 clinical images). This way, it is possible to calculate two average profiles (left and right) that can be used to normalize all profiles of new patients. Then in each left and right normalized patient profile it is possible to select two intervals, one up and one down, and compute their average intensity values. These average intensity values are then used as input for Eq. 2.3. In this way, segmentation of the background regions is not required, giving an advantage in applying the algorithm to images where there is no visible background or it is difficult to extract automatically.

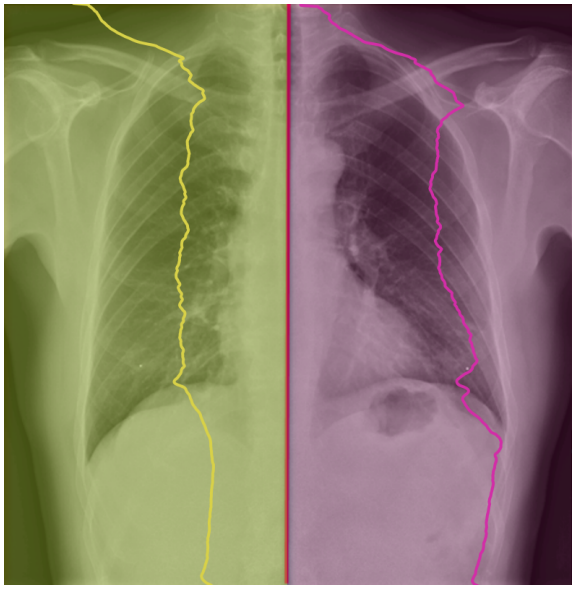

Figure 5.1: Example of how the normalized profiles algorithm divides one chest patient image in two vertical sides and draws a vertical profile of the pixels on the left rows (yellow) and one on the right rows (pink).

5.1.3 Validation method

When the appropriate robust method to measure the inhomogeneity from patient images is found, it can be validated by performing an image deterioration study, similar to what has been implemented for the validation of the ClinQC noise study (Section 3.1.3.1). For the homogeneity study, different strengths of the anode heel effect can be simulated. This way it can be verified if the homogeneity algorithm

can detect a rise in the inhomogeneity values due to the change in pixel values after the simulation. This will indicate if the chosen method is actually able to measure changes in the homogeneity values of the imaging system.

5.2 Results

In this Section all the preliminary results regarding the homogeneity study are presented.

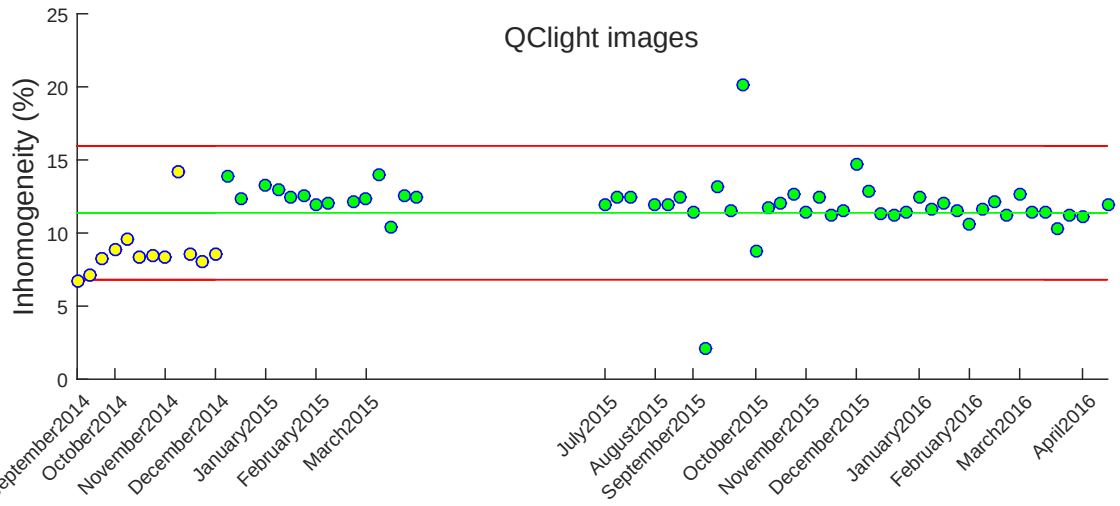

Figure 5.2: The history of the recorded QClight homogeneity parameter.

In Fig. 5.2 the history of the QClight homogeneity values is represented. In 2014, with the anti-scatter grid with focal distance of 1.40 m and the QClight SID of 1.50 m, the inhomogeneity computed from the phantom images was globally lower than with the actual anti-scatter grid; this was also the case for the QClight noise values. The reason for this is that the new anti-scatter grid with focal distance 1.80 m, instead, is optimal for the patients SID of 2.00 m, while is not well suited for the QClight phantom acquisition, with a SID of 1.50 m. For this reason the inhomogeneity measured with QClight is higher with the new anti-scatter grid.

It has been observed, after looking at the histogram of several chest X-ray images

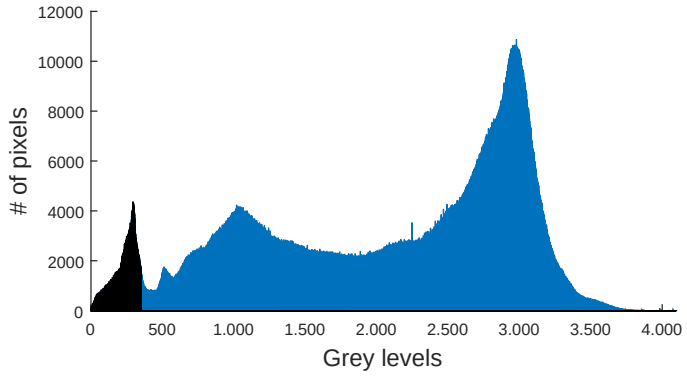

Figure 5.3: Example of the histogram of a typical chest X-ray patient image, where the highlighted black class represents mainly the background and ends at the 350^{th} grey level.

obtained with the actual imaging system, that the background pixel values constitute a well separated class in the global image histogram, that has a limit around the 350^{th} grey level (example in Fig. 5.3). In fact, one method that has been used to segment the background pixels is be to implement a fixed threshold on the 350^{th} grey level of the image. The histogram minima method is a similar approach but it looks for the real position of the minima of the first class of the histogram that can be slightly different from 350 for each patient image.

All the thresholding algorithm listed in Section 5.1.1 have been tested on several clinical chest X-ray images. After a first visual analysis, none of them seemed to be robust enough to correctly crop the background ROIs in the corners of all the patient images being analysed, since the segmented regions often contained also part of the patient skin. The best result was obtained using the FMM, the example is shown in Fig. 5.4.

All the clinical homogeneity extraction algorithms, both the segmentation tools and the normalized profiles method, have been tested on the same dataset of clinical images, acquired during one week in March 2016 when an average acceptable QClight homogeneity value was recorded. This is considered as the baseline dataset of clinical images for homogeneity.

The distributions of the inhomogeneity values obtained with all the clinical homogeneity algorithms are represented by the boxes in the left side of Fig. 5.5. On the

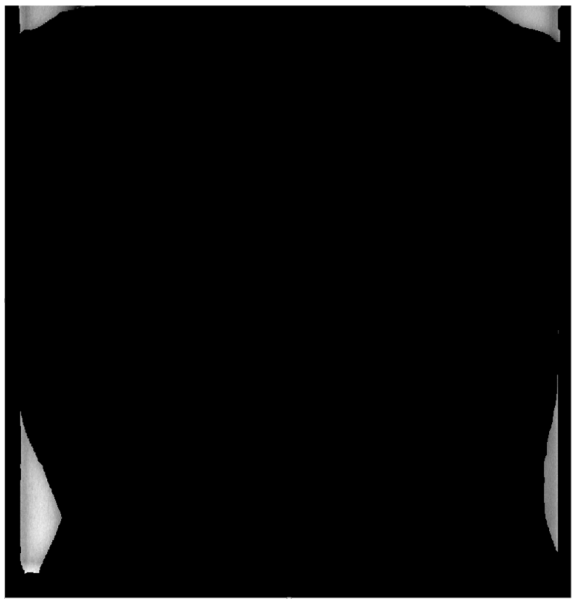

Figure 5.4: Example of the segmentation of the background pixel values in a chest X-ray image using the Fast Marching Method (FMM).

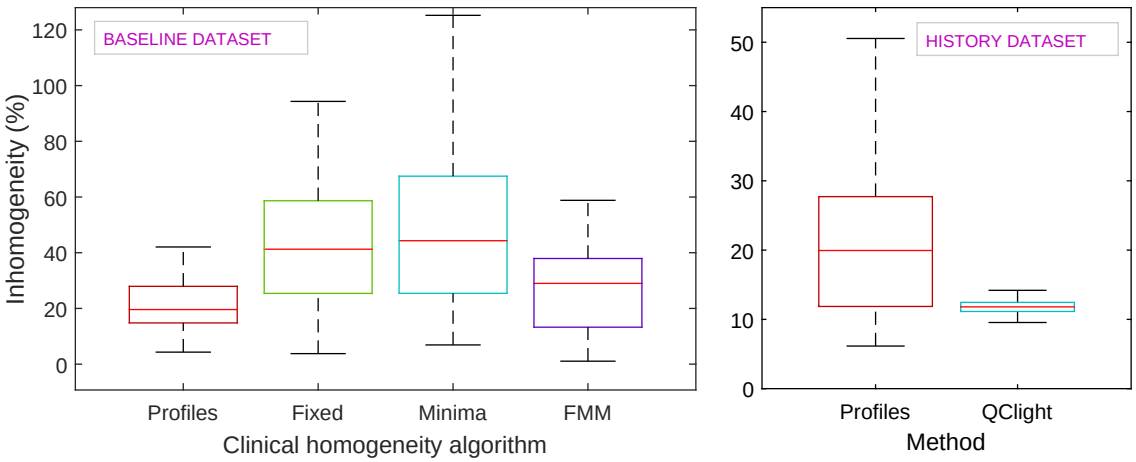

Figure 5.5: Left side: each box represents the distributions of the inhomogeneity values obtained applying the specified clinical homogeneity algorithm on the *baseline dataset*. Right side: comparison between the normalized profiles algorithm and the QClight method, both applied on the *history dataset*. The *y-scale* is zoomed.

right side a comparison between the normalized profiles algorithm and the QClight method is displayed. Both are applied on the history dataset made of respectively

one clinical and one phantom image per week in the history of the imaging system in the current setup (these comparison values are also shown in a time line in Fig. 5.6). The *y-scale* in Fig. 5.5 is zoomed so that it can be observed that the main differences between the inhomogeneity values obtained with clinical images-based algorithms and with the QClight method are the mean and the spread. The histogram minima method and the fixed threshold method lead to similar results for the baseline dataset, but the inhomogeneity values were much higher compared to the values measured with the QClight method on the history dataset, with the highest relative spread. Both the FMM and the profiles method recorded lower inhomogeneity values with even a lower variation, but were not comparable to the QClight values distribution.

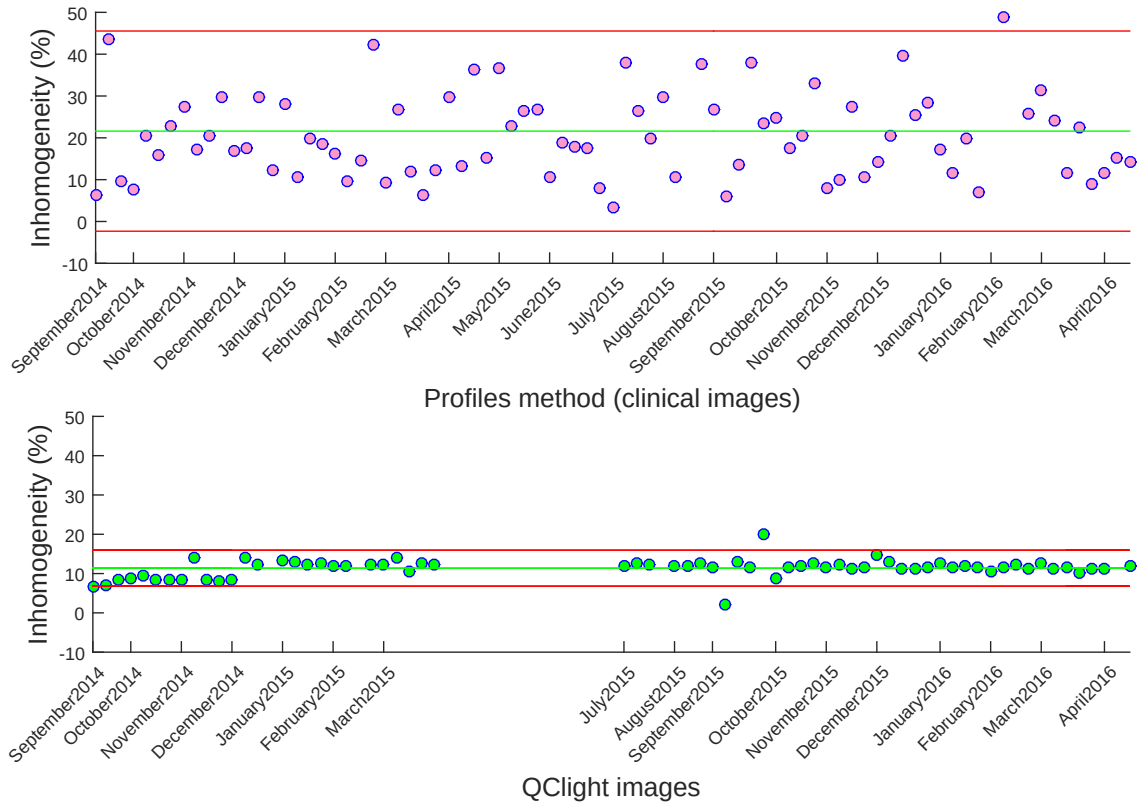

Figure 5.6: Top: the inhomogeneity values computed using the normalized profiles algorithm on the history dataset of clinical images. Bottom: the inhomogeneity values computed using the QClight method on the history dataset of QClight phantom images.

The proposed validation method (Section 5.1.3) is based on the image deterioration study with a simulation of the anode heel effect in the clinical images (example in Fig. 5.7).

When the proper clinical homogeneity algorithm is found, the validation can be

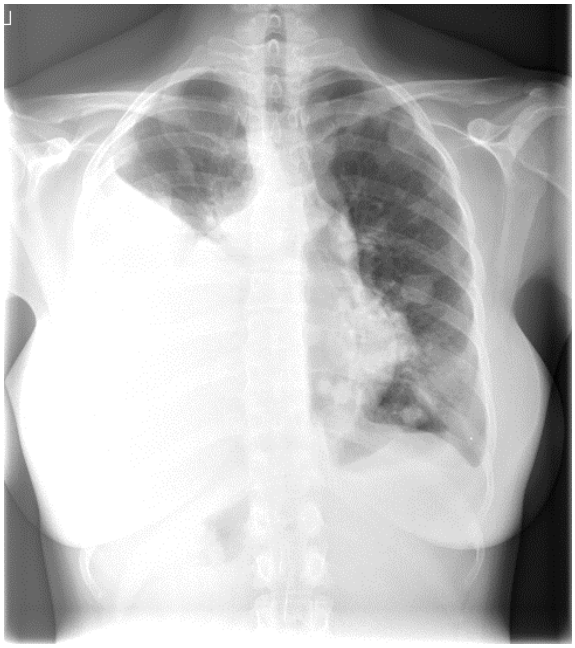

Figure 5.7: Example of a strong simulated anode heel effect on a clinical chest X-ray image.

made by simulating an increasing strength of the anode heel effect in the clinical images and recording the response of the homogeneity algorithm. An example of how this would work for the Fast Marching Method is shown in Fig. 5.8: the same clinical image is segmented before and after applying the simulated anode heel effect in two different strengths (with gradients of, respectively, 1 and 5 grey values for 100 pixel).

The result is that the segmented lower right ROI becomes smaller when the simulated anode heel effect is increased towards its direction: the reduction of the lower right ROI area of the original image is respectively 3% and 40% for the two simulated images in Fig. 5.8. This happens because all the thresholding methods are intensity-dependent. It is often segmented with the background information also the primary X-ray beam shadow, that appears like an image frame on the background

Figure 5.8: Example of the segmentation using the FMM on the same chest X-ray image, respectively in its original appearance (left) and after simulating two different strength of the anode heel effect with gradients of 1 (center) and 5 (right) grey values for 100 pixel.

areas around the patient (example in Fig. 5.9).

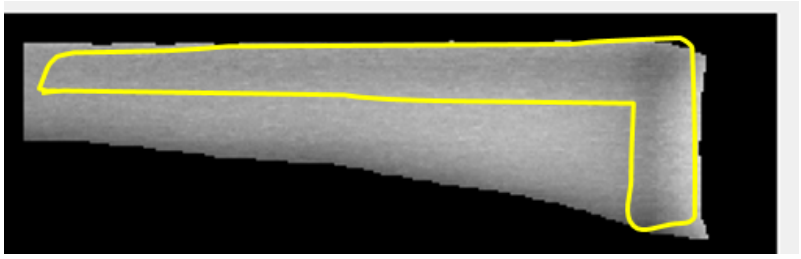

Figure 5.9: Example of the primary X-ray beam shadow, that appears as an image frame on the background areas around the patient.

5.3 Discussion

The large variations in the homogeneity values obtained applying clinical images-based algorithms demonstrate that these algorithms cannot be used to measure the inhomogeneity of an X-ray imaging system from clinical images. The high variations suggest that the tested algorithms are not robust enough to be used in clinical practice, but more tests are required in order to investigate what are the main causes and find alternative algorithms. The major limitation of the thresholding-based algorithms is that they are intensity-dependent, and therefore are affected by the patient anatomical signal nearby the background regions, by the dose that was used

to achieve a pre-fixed image quality and by the post-processing. In particular, the post-processing might equalize the image histogram, changing the average grey value of the background and/or enhancing the patient edges, darkening the regions of pixel surrounding the patient. Another reason that caused a variability in the homogeneity measures is that in the segmented ROIs is often clearly visible the primary X-ray beam shadow, that appears like an image frame on the background areas around the patient and is also emphasized by the post-processing filters. So if these large variations in pixel values are segmented with the background, they might lead to high percentages of inhomogeneity, not truly representative of the inhomogeneity of the imaging system. If a real homogeneity-related problem occurs in the imaging system, due to a damage of the detector or to a significant change in the X-ray beam uniformity or a problem with the detector calibration, a thresholding-based algorithm will hardly be able to segment the background as in a normal situation, since the background could be brighter than usual and even with adaptive filters it could be confused with the patient signal.
To find a clinical images-based algorithm that measure the inhomogeneity from clinical images avoiding the background segmentation would make the ClinQC algorithm applicable also to modalities where there is no available background or it is difficult to extract. This could have been the case of the normalized profiles method, which is more robust than the thresholding-based algorithms since it avoids the background segmentation, but it produced much highly fluctuating inhomogeneity measures compared to the gold standard QClight homogeneity method.
After this first trial, it can be concluded that the ClinQC homogeneity measure is not ready to be applied in clinical practice until the appropriate clinical images-based homogeneity extraction algorithm is found, tested and validated. A number of limitations of segmentation-based algorithms has been illustrated and can be useful input in the development of a new ClinQC homogeneity algorithm.

Chapter 6

Discussions and conclusions

The aim of the ClinQC project was to perform a QC of an X-ray imaging system using clinical images. Regarding both the dose and the noise studies, the ClinQC algorithm introduced new reliable and sensitive parameters of the quality of the imaging system, that have the same physical meaning and similar relative variation as the QClight method gold standard quality indicators.
Specifically concerning the *noise study*, the ClinQC noise extraction algorithm from clinical images has been validated in several ways with both statistical simulations, phantom studies and real clinical examples proving that it is able to give an on-the-fly and highly sensitive feedback on the performances of the imaging system with a relatively low number of false alarms. The results were also comparable with a good correlation coefficient to the published Pyramid noise extraction algorithm, and some positive preliminary results showed that it can be easily improved by implementing a grid sampling approach that will help in reducing the influence of outliers such as pacemakers, implants and other medical devices in the noise measurement.
The *dose study* suggested that monitoring the stability of the exposure of patients from clinical images, instead of using the QClight phantom images, gives a better overview of the functioning of the AEC device, and truly helps to monitor if the AEC is delivering consistent and reproducible exposures across a wide range of anatomical thicknesses. In addition, the introduced tube output metric can be considered a new dose indicator employed in QCs of X-ray imaging systems from clinical images,

comparable to the DAP used in phantom-based QCs.
Since this project is still at the beginning, there are some *limitations* to consider. First, the studied images are only chest X-ray examinations. Up to now only a preliminary test has been performed with mammographies. Second, the ClinQC algorithm contains a limited set of dose and image quality parameters and the algorithms are simple. Future updates may result in more quality parameters and even better performances. Lastly, although some first initial interesting results were obtained, for both dose and homogeneity studies more analyses and validation need to be performed before the new tools can be applied in clinical practice side-by-side with the QClight.
By contrast, the ClinQC project proved to have also many *advantages*: using the ClinQC algorithm is a fast and reliable way to monitor both the health and correct installation of the hardware of an X-ray imaging system, to supervise the patient acquisition protocols and the image post-processing. This algorithm also allows to perform a QC without additional phantom acquisitions. Lastly, the world of Big Data is accessible by using this method since the numbers of patients per day, week, month and year are surprisingly high in big hospitals, so a large amount of information can be gained from the study of patient data and their clinical images, allowing the users to perform more extended statistical analyses.
After the studies of the ClinQC noise and dose parameter performances and all the kinds of validation designed in this thesis work, it can be concluded that the ClinQC algorithm for noise and dose monitoring of an X-ray imaging system could be already applied in clinical practice, with the initial support of the QClight weekly quality control. In this way a real test period, with a comparison between the two methods, will indicate the necessary final adjustments on the algorithm until the final version is being installed. In the meantime, all the ideas presented in the next Section can be considered to improve the future versions of the algorithm and to extend the ClinQC project to other anatomies and/or imaging modalities.

Outlook for the future

In this Section some new ideas that can improve the ClinQC algorithm during the project continuation are presented.

- Taking an ROI of 30% of the FOV in the last step of the ClinQC noise extraction algorithm limits the number of pixels involved in the measurement. An idea to use all the pixels in the final noise map, in order to reach better statistics and improve the measurement accuracy, could be to use the *histogram* of the image or the *high-low frequencies plot* knowledge to crop out the patient edges, burned-in annotations and background and use all the original image area instead of the ROI to measure the ClinQC noise value. A similar method is used for CT images [27].

- The development of a new ClinQC detection algorithm that combines noise and dose information could be useful to minimize the production of false positive alerts as the output of the clinical images-based QC and could be achieved by implementing also some pattern recognition tools.

- The validation of the ClinQC tube output parameter is needed before it can be used in clinical practice, side-by-side with the DAP information retrieved from the QClight phantom images.

- With the implementation and validation of the ClinQC-alternative algorithm, updated with the sampling grid approach (Section 3.1.2.1), the influence of the structures in the ClinQC normalized noise image that change the noise distribution are reduced.

- It would be interesting to study if the ClinQC outlier noise values can be correlated to patient age, using hospital data.

- New alternative versions of the ClinQC algorithm can be adapted to different anatomies and/or imaging modalities, both introducing new noise and dose indicators or using the same noise and dose parameters, but considering that many factors can change and affect the measurements, such as the acquisition

protocols, as has been shown for mammography. The feasibility of each new method has to be investigated with a proper validation.

Bibliography

[1] Leiden University Medical Center. www.lumc.nl.

[2] Erasmus+: mobility for traineeship. http://ec.europa.eu/programmes/erasmus-plus/node_en.

[3] Radiology Info. www.radiologyinfo.org.

[4] Jerrold T Bushberg and John M Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.

[5] DICOM NEMA. http://dicom.nema.org.

[6] NCRP. Report no. 099 - quality assurance for diagnostic imaging. Technical report, NCRP, 1988.

[7] Harrell G Chotas, Carey E Floyd Jr, G Allan Johnson, and Carl E Ravin. Quality control phantom for digital chest radiography. *Radiology*, 202(1):111–116, 1997.

[8] Cornelia Schaefer-Prokop, Ulrich Neitzel, Henk W Venema, Martin Uffmann, and Mathias Prokop. Digital chest radiography: an update on modern technology, dose containment and control of image quality. *European radiology*, 18(9):1818–1830, 2008.

[9] Perry Sprawls. *Physical principles of medical imaging*. Aspen Publishers, 1987.

[10]

[11]

[12] Xiaoyu Tian and Ehsan Samei. Accurate assessment and prediction of noise in clinical ct images. *Medical Physics*, 43(1), 2016.

[13] Rafael C Gonzalez, Richard E Woods, et al. Digital image processing, 2002.

[14] EH Chao, TL Toth, EC Williams, SH Fox, DA Carleton, and NB Bromberg. A statistical method of defining low contrast detectability. In *Radiology*, volume

217, pages 162–162. RSNA, 20th and Northampton sts, Easton, PA 18042 USA, 2000.

[15] Anna Fabijanska and Dominik Sankowski. Image noise removal-the new approach. In *2007 9th International Conference-The Experience of Designing and Applications of CAD Systems in Microelectronics*, 2007.

[16] Wouter JH Veldkamp, Lucia JM Kroft, Jan Pieter A van Delft, and Jacob Geleijns. A technique for simulating the effect of dose reduction on image quality in digital chest radiography. *Journal of digital imaging*, 22(2):114–125, 2009.

[17] Seong Ho Park, Jin Mo Goo, and Chan-Hee Jo. Receiver operating characteristic (roc) curve: practical review for radiologists. *Korean Journal of Radiology*, 5(1):11–18, 2004.

[18] Yuan Lin, Hui Luo, James T Dobbins III, H Page McAdams, Xiaohui Wang, William J Sehnert, Lori Barski, David H Foos, and Ehsan Samei. An image-based technique to assess the perceptual quality of clinical chest radiographs. *Medical physics*, 39(11):7019–7031, 2012.

[19] Peter Burt and Edward Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[20] Gerhard Bohm and Günter Zech. *Introduction to statistics and data analysis for physicists*. DESY, 2010.

[21] Trevor J.. Hastie, Robert John Tibshirani, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2011.

[22] M Hermanussen, H Danker-Hopfe, and GW Weber. Body weight and the shape of the natural distribution of weight, in very large samples of german, austrian and norwegian conscripts. *International Journal of Obesity & Related Metabolic Disorders*, 25(10), 2001.

[23] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987.

[24] Thorsten Thadewald and Herbert Büning. Jarque–bera test and its competitors for testing normality–a power comparison. *Journal of Applied Statistics*, 34(1):87–105, 2007.

[25] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[26] James Albert Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999.

[27] Xiaoyu Tian and Ehsan Samei. Accurate assessment and prediction of noise in clinical ct images. *Medical physics*, 43(1):475–482, 2016.

Acknowledgements

Desidero ringraziare tutta la mia famiglia per avermi resa quella che sono e, soprattutto, per avermi insegnato a stare al mondo.

In particolare i miei genitori Margherita e Silvio, che mi hanno permesso di fare quel che volevo nella vita senza viziarmi troppo, di studiare e fare questa esperienza unica all'estero, aiutandomi a superare tutte le difficoltá del mio percorso di studi e di vita. Senza di voi non sarei certo arrivata fin qui.

Enrico, per essere il mio eroe.

Tutti gli amici, i compagni di studio, di sport, di varie follie e della triste condizione di essere viaggiatori pendolari, per il divertimento, le risate e la spensieratezza che mi hanno saputo trasmettere anche nei momenti peggiori.

Gli amici e colleghi musicisti dell'Orchestra Mutinae Plectri, che credevano di essersi liberati di me vedendomi partire per l'Olanda.

Il mio Maestro di chitarra classica Roberto per l'enorme pazienza dimostrata in questi anni di musica passati insieme.

Un particolare ringraziamento anche alla mia relatrice e Professoressa Maria Pia Morigi per tutto il supporto, anche a distanza, durante il periodo di studi, di tirocinio e di tesi.

En nu ik overschakelen naar Nederlands, well... it is better for all of us if I switch to English.

I'd like to say *dank je wel* to all the new LUMC friends and colleagues, for teaching me a wide Dutch vocabulary and not to live without pannenkoeken any more. I will always feel close to you and your wonderful country.

In particular thanks to Wouter, Pieternel, Chiel and Berend for giving me complete blind trust and support in every occasion, and giving me the chance to work on a challenging and interesting project as the ClinQC, in a familiar and really funny environment.
Special thanks to Pieternel that helped me in breaking the Physics everyday, but fixing it soon before Chiel and his oranges would have realized.

Irene, que en Leiden me hizo sentir como en casa.