Autocalibration method for non-stationary CT bias correction

Gonzalo Vegas-Sánchez-Ferrero [a,b,*], Maria J. Ledesma-Carbayo [b], George R. Washko [a], Raúl San José Estépar [a]

[a] *Applied Chest Imaging Laboratory (ACIL), Brigham and Women's Hospital, Harvard Medical School, 1249 Boylston St. 02115, Boston, MA, USA*
[b] *Biomedical Image Technologies Laboratory (BIT), ETSI Telecomunicacion, Universidad Politecnica de Madrid, and CIBER-BBN, Madrid, Spain*

ABSTRACT

Computed tomography (CT) is a widely used imaging modality for screening and diagnosis. However, the deleterious effects of radiation exposure inherent in CT imaging require the development of image reconstruction methods which can reduce exposure levels. The development of iterative reconstruction techniques is now enabling the acquisition of low-dose CT images whose quality is comparable to that of CT images acquired with much higher radiation dosages. However, the characterization and calibration of the CT signal due to changes in dosage and reconstruction approaches is crucial to provide clinically relevant data. Although CT scanners are calibrated as part of the imaging workflow, the calibration is limited to select global reference values and does not consider other inherent factors of the acquisition that depend on the subject scanned (e.g. photon starvation, partial volume effect, beam hardening) and result in a non-stationary noise response. In this work, we analyze the effect of reconstruction biases caused by non-stationary noise and propose an autocalibration methodology to compensate it. Our contributions are: 1) the derivation of a functional relationship between observed bias and non-stationary noise, 2) a robust and accurate method to estimate the local variance, 3) an autocalibration methodology that does not necessarily rely on a calibration phantom, attenuates the bias caused by noise and removes the systematic bias observed in devices from different vendors. The validation of the proposed methodology was performed with a physical phantom and clinical CT scans acquired with different configurations (kernels, doses, algorithms including iterative reconstruction). The results confirmed the suitability of the proposed methods for removing the intra-device and inter-device reconstruction biases.

© 2017 Published by Elsevier B.V.

1. Introduction

Computed tomographic imaging has become almost universally available in clinical and research settings. Since its introduction, it has grown to be part of routine clinical practice, and it is estimated that over 80 million CT scans are performed each year in the United States (Hess et al., 2014). While it continues to provide new insight into the characterization and prognostication of disease, this high utilization has also raised concerns about the implications of radiation exposure to clinical populations (Brenner and Hall, 2007). Those concerns have tempered the growth in CT imaging studies (Hess et al., 2014) and propelled technological innovations for the implementation of low-dose and ultra-low-dose techniques in clinical practice (Mayo-Smith et al., 2014).

The characterization and calibration of the CT signal due to changes in dosage and reconstruction approaches is crucial for the advent of quantitative imaging as a clinically relevant tool. Quan-

titative imaging (QI) is the process of reducing functional, biological and morphological processes to a measurable quantity employing medical imaging. The uses of QI are even greater in the light of a new healthcare delivery system that becomes more personalized and tries to tailor therapies to the underlying pathophysiology. QI contributes to the radiological interpretation by assessing the degree of a given condition (Buckler et al., 2011). QI has been adopted in clinical studies and trials to obtain more sensitive and precise endpoints. The advancement in techniques to automatically interpret and quantify medical images has been recognized by regulatory agencies that have now proposed guidelines for the qualification of image-based biomarkers to be used as valid endpoints in clinical trials (e.g. the Quantitative Imaging Biomarkers Alliance, QIBA, at www.rsna.org/qiba). However, the utility of quantitative imaging is hampered by the lack of standardization among vendors due to the variations in the acquisition and reconstruction processes such as signal-to-noise ratio, spatial resolution, slice thickness, image reconstruction algorithms among others (Mulshine et al., 2015).

The characterization and calibration of the CT signal due to changes in dosage and reconstruction approaches are foundational for the transition of quantitative image analysis from biomedical research to clinical care (Sieren et al., 2012). This transition has heretofore been limited by inter- and intra-scanner variability which has inhibited efforts to perform and interpret large-scale cross-sectional and longitudinal studies (Chen-Mayer et al., 2017; Parr et al., 2004). Any observed variability in such efforts has been ascribed to an inseparable admixture of poor standardization and biology. The current approach to this challenge is to create larger cohorts of subjects for clinical investigation with the hope that sample size will allow to detect biological effects despite the noise (Regan et al., 2011).

Although CT scanners are calibrated as part of the imaging workflow, the calibration is typically based on selecting global reference values such as air and water (Millner et al., 1978). These limited calibration points are unable to account for inherent factors of the acquisition (e.g. photon starvation, partial volume effect, beam hardening) and so the resulting CT signal is more variable than desired (Hsieh, 2003). These effects are particularly important to account for when creating a quantitative metric that is consistent among vendors and free of confounding factors due to changes in patient weight and size to fulfill requirements of accuracy and precision (Uppot et al., 2007). Recent validation studies comparing different low-dose reconstruction approaches show the variability of quantitative traits, therefore, suggesting the need for calibration procedures (Choo et al., 2014).

Among all those issues, a clear effect of low-dose and ultra-low-dose CT protocols is the increase of image noise (Kim et al., 2015). CT noise is a major factor that has been carefully studied during the last decades at the detector level as part of the transmission process (Whiting, 2002). The non-monochromatic nature of the X-ray signal, the amount of total X-ray energy defined by tube current coupled with the effects of the reconstruction and the interaction between X-ray and matter within the scanning field of view make the noise characterization in the reconstructed image a complex process. One of the main effects of this complexity is the lack of stationarity. It is well understood that fan-beam tomography introduces non-stationary frequency components and non-stationary noise by the nature of the scanning geometry (Zeng, 2004).

The advent of iterative methods to deal with ultra-low-dose reconstruction provides a more complex scenario in which the underlying preprocessing affects differently the reconstructed signal. Those reconstruction methods affect the attenuation levels differently depending on their assumptions and may result in a deviation of the desired calibration as commonly seen in PET attenuation correction techniques using ultra-low dose CT protocols (Xia et al., 2012).

The aforementioned factors highlight the many challenges present in the quantitative comparison of CT images acquired under different conditions. This paper analyzes the effect of reconstruction inconsistencies depending on the dose, reconstruction algorithm, and acquisition parameters. The analysis is especially focused on the effect of the non-stationarity of noise, which is a signal-dependent source of variability that cannot be prevented due to the physics of the acquisition. The effect of bias due to the non-stationary variance has been observed in the literature due to the strong deviations perceived in the attenuation level for the air cavities such as the trachea (Parr et al., 2004). On the other hand, the kernels and iterative algorithms used for reconstruction may arbitrarily affect the average intensities. If these effects are not considered, the reconstructions become useless for multicenter studies or the analysis of disease progression.

Several efforts have been coordinated to conduct phantom studies in a variety of scanner models to establish a baseline for assessing the variations in patient studies that can be attributed to scanner calibration and measurement uncertainty. One of those is Chen-Mayer et al. (2017), in which a phantom study provides an assessment of the accuracy and precision of the density metrics across platforms due to machine calibration. This study, however, does not consider the potential effects of non-stationary noise and spatially variant biases as the ones observed in Parr et al. (2004).

In our previous work Vegas-Sánchez-Ferrero et al. (2017), we proposed a statistical framework to describe the non-stationary behavior of noise and proposed a stabilization scheme to transform it into a stationary Gaussian process, enabling the local comparison of histograms between different doses and reconstruction kernels. This methodology would suffice to standardize scanners under the same calibration assuming the reconstruction methods do not introduce any bias. However, the non-stationary and positively skewed nature of noise also introduces an intrinsic bias that depends spatially on the object scanned. Even the same imaging device exhibits different bias fields depending on the subject being imaged. This fact is further complicated in multivendor studies where different systematic biases may be observed (Sieren et al., 2012).

To define and correct the aforementioned effect, we performed a bias study through a statistical exploratory data analysis in a series of acquisitions provided with different devices, reconstruction kernels, and doses. Then, we establish a functional relationship between the observed attenuation level and the local moments. This relationship paves the way for the derivation of an autocalibration method. The local variance has to be carefully estimated to remove the spatial dependence between bias and noise. We propose a robust methodology to estimate the local variance that avoids artifacts due to tissue boundaries and non-homogeneities based on the statistical characterization and variance stabilization methods recently proposed in Vegas-Sánchez-Ferrero et al. (2017). Then, an autocalibration scheme is proposed to establish a common framework for comparison of studies. This method successfully removes the bias caused by non-stationary noise and calibrates according to certain reference attenuation levels (e.g. anatomical references or well-defined homogeneous materials). The proposed methodology does not require any phantom, and it is designed to deal with studies with heterogeneous calibrations that may differ remarkably in their reconstruction methods or doses without the need for any parameter specification such as the reconstruction method. This method makes use of anatomical regions to establish common references and successfully removes the bias due to spatially-variant noise. The systematic bias induced by reconstruction methods and devices is also successfully corrected.

The contributions of this paper are: 1) The derivation of a functional relationship between the bias and non-stationary variance, 2) a robust method to calculate non-stationary variance, and 3) the autocalibration methodology.

The evaluation of the methods is performed with phantom acquisitions for different doses, reconstruction methods, and devices. The bias due to noise is successfully removed with a reduction over 90% in most of the cases, and the systematic bias is successfully removed. The evaluation with clinical CT scans is tested considering low-dose acquisitions, different reconstruction kernels, and iterative reconstruction methods. The observed intensities in well-defined anatomical structures such as trachea show a significant discrepancy in non-calibrated images. The autocalibration methodology successfully removes the bias and provides a uniform response across the different acquisition conditions.

The paper is structured as follows: Section 2 presents the exploratory data analysis of a phantom acquired with different kernels, doses, and devices. The exploratory analysis carefully tests functional relationships between the statistics of noise and the observed attenuation level. The descriptive analysis leads to the definition of a functional relationship between observed attenuation
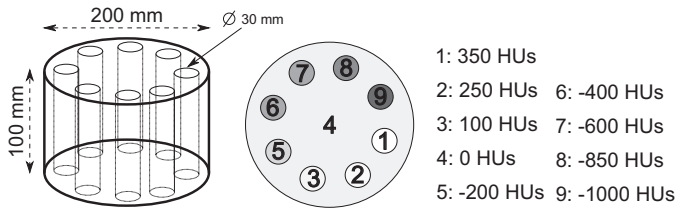
Fig. 1. Scheme of the cylindrical phantom studied. The legend specifies in descending order the different attenuation levels per material.

Fig. 2. Axial view of the acquired 3D volumes with different devices (Top: General Electric; Bottom: Siemens) for different kernels and doses. a) STD HD, b) STD LD, c) BONE HD, d) BONE LD, e) B31f HD, f) B31f LD, g) B45f HD, h) B45f LD.

and local variance of noise (Section 2.1). The methods to estimate local moments and the non-stationary variance are described in Section 3.1. Then, in Section 3.3, the autocalibration methodology is described. The validation of the proposed methodology is given in Section 4. Finally, in Section 5 we present our conclusions.

2. Exploratory data analysis of bias

The exploratory analysis is carried out in a physical phantom to see the deviations from the nominal attenuation value given by manufacturers. The deviations will be studied considering the non-stationary nature of noise, which is signal dependent.

We used the 8-step linearity module of the *Lung Cancer Screening CT Phantom.*[1] The phantom is schematically described in Fig. 1. It consists of a cylindrical structure (200 mm diameter, 100 mm height) made of a homogeneous material that contains other 8 concentric cylindrical structures with different attenuation levels (from 340 HUs to −1000 HUs).

The phantom was acquired with two different devices with the following reconstruction protocol:

- **General Electric Discovery STE.** Four volumes of size $512 \times 512 \times 313$ were acquired at Brigham and Women's Hospital with various doses (400 mA and 100 mA) and reconstruction kernels (Standard, Bone). All of them with a KVP: 120 kV, slice thickness 0.625, pixel spacing 0.7×0.7, with software 07MWDVCT36.4. We will refer to these volumes as STD HD, STD LD, BONE HD, BONE LD for the different arrangements of kernels and doses (HD: high dose; LD: low dose).
- **Siemens Definition.** Similarly, four volumes of size $512 \times 512 \times 313$ were acquired at Brigham and Women's Hospital for the same configurations of doses (400 mA and 100 mA) and reconstruction kernels B31f, B45f. All of them with a KVP: 120 kV, slice thickness 0.75, pixel spacing 0.98×0.98, with software *Syngo*.CT 2007C. Following the same convention as before, we will refer to them as B31f HD, B31f LD, B45f HD, B45f LD.

Fig. 2 shows an example of the acquired images for all the kernels, devices, and doses considered.

In what follows we will denote $X : \Omega \to \mathbb{R}$ as the CT volume defined in Ω, the location coordinates as $\mathbf{r} = (x, y, z) \in \Omega$, and $\Omega_n \subset \Omega$ the locations of the n-th homogeneous material.

2.1. Functional dependence on local variance

The study was performed in a set of samples collected from each tissue identified by the numbered regions from 1 to 9 of the CT images (see the designation of regions in Fig. 1). The samples were acquired by manually selecting a circular region in the axial view laying within each tissue type. More than 20,000 samples were obtained in each region.
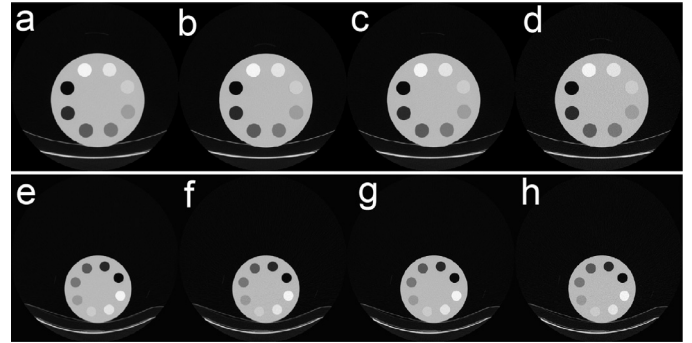
The deviations from the nominal attenuation level per region were firstly analyzed by considering the spatial mean and (unbiased) variance estimators computed in the longitudinal direction as:

$$\widehat{\mu}(x, y) = \frac{1}{N} \sum_{z=1}^{N} X(x, y, z) \tag{1}$$

$$\widehat{\sigma}^2(x, y) = \frac{1}{N-1} \sum_{z=1}^{N} \left(X(x, y, z) - \widehat{\mu}(x, y) \right)^2 \tag{2}$$

where $X(x, y, z)$ is the voxel value at location (x,y,z), and N is the number of slices considered. Note that this calculation takes advantage of the invariance in the z-axis due to the cylindrical shape of the phantom. This analysis allows us to obtain the functional relationship between the mean attenuation level and the local noise variance. Fig. 3 shows this relationship between $\widehat{\mu}$ and $\widehat{\sigma}^2$ for the different regions, doses, and devices.

At first sight, the average attenuation levels seem to be independent of the local variance. However, as lower densities are considered, both the mean and variance become more related. We have zoomed in the samples of the lowest attenuation level and superimposed the regression line for a better visual inspection. Note that the observed correlation exhibits an increasing bias as the variance grows. This bias will be more noticeable in those acquisitions with more intrinsic noise (e.g. low-dose or sharper kernels) or near high-density structures.

On the other hand, the observed correlation decreases for higher densities, meaning that local variance introduces a bias in tissues typically characterized by lower attenuation levels such those related to air (e.g. lung parenchyma, airways).

This bias is due to the lower limit imposed by the physics of the acquisition, which assigns -1000 HU to the minimum attenuation perceived –corresponding to air– and 0 HU to the attenuation level of water (Hsieh, 2003). Obviously, this constraint on the lower attenuation level enforces a positive skewness in the distribution of noise. The positive skewness was confirmed in Vegas-Sánchez-Ferrero et al. (2017), where the statistical response of noise was studied for different doses, reconstruction kernels, and devices. In Fig. 4 we show the probability density functions observed for both the GE and Siemens devices where the overall statistical behavior can be observed. Note that the density functions may take values below −1000 HU. This is because the least attenuation level is set to −1000 HU on average. Usually, vendors set a lower attenuation limit to a value below −1000 HU. We will refer to this value as δ and, in our particular case, $\delta = -1024$ HU.

It is also worth mentioning that, in some cases, the lower attenuation value set by vendors may generate a truncated distribution, which clearly has a positive skewness (the distributions depicted in
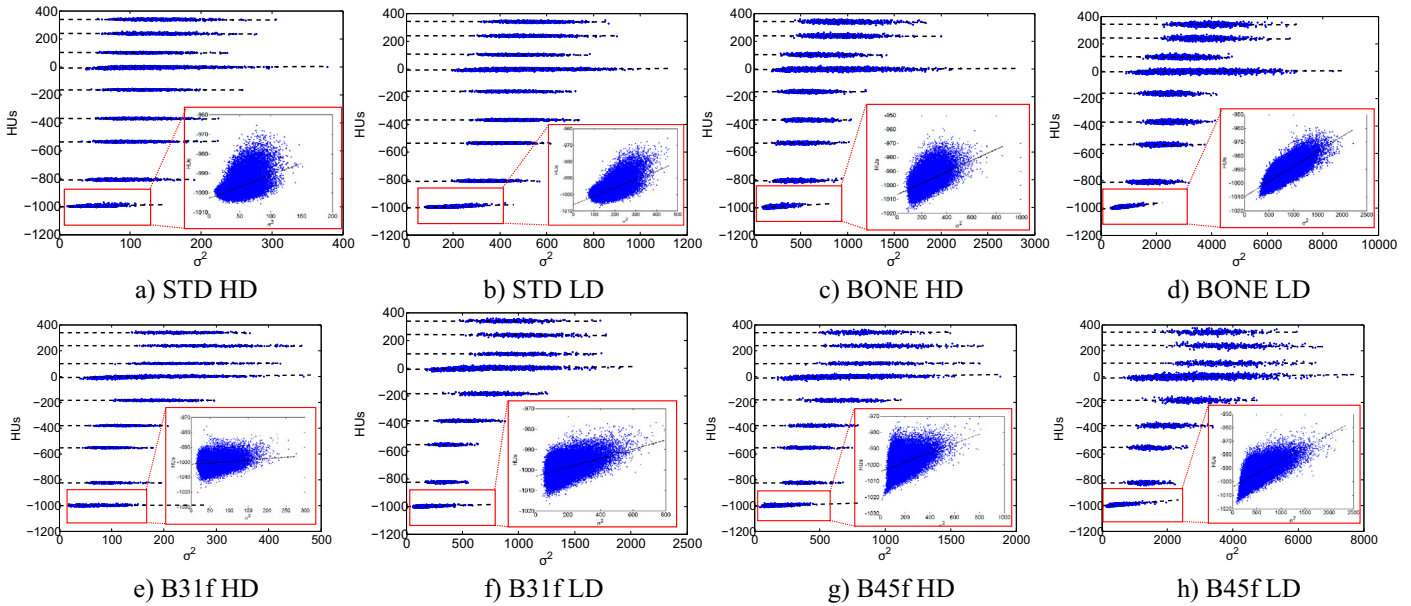
Fig. 3. Functional relationship of local variance and local mean for the different tissue regions and different arrangements of devices, kernels, and doses. Top: GE, Bottom: Siemens. The regression lines are superimposed to the samples as dashed lines. The zoomed region shows the positive bias introduced by the increase of variance.
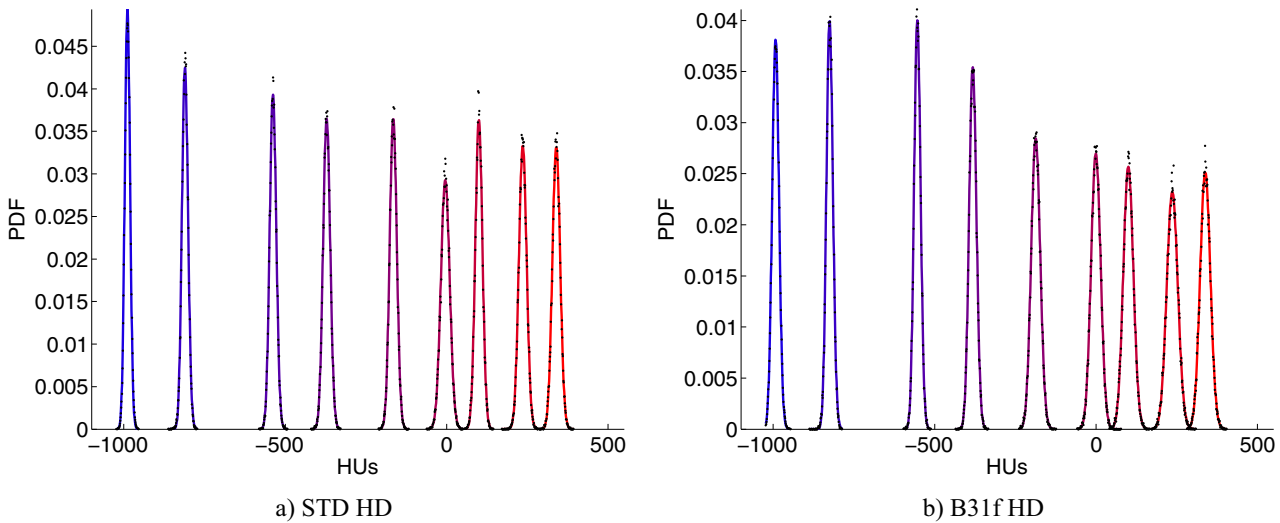
Fig. 4. Empirical probability density functions of samples for different acquisitions of the phantom (the rest of kernel-dose configurations showed similar distributions and were omitted for brevity). Note that the distributions exhibit different variance and a positive skewness especially for low CT numbers.

Fig. 4 do not show this effect, but other configurations with higher noise may exhibit truncated distributions).

In both possible scenarios, the skewness will link the CT number with the local variance of noise resulting in a non-stationary bias that strongly affects the lower attenuation levels.

The positive skewness can contribute to the bias in several functional ways. For simplicity, we test the linear relationship between the mean and the variance with a linear regression model. The results gathered for the regression analyses in air regions are shown in Table 1 and the regression lines are depicted as dashed lines in Fig. 3. We exclude the rest of tissues because the coefficient of determination rapidly decreases for higher densities, resulting in a negligible bias due to the rapid reduction of skewness for higher densities.

The increasing bias with the variance of noise is confirmed with an F-test for the null hypothesis "H_0: *The average attenuation level does not depend on the local variance.*" The coefficient of determi-

Table 1
Linear regression between variance and mean.

	Regression Line	Bias interval 95% (HU)	R^2	p-value
STD LD	$\hat{\mu} = 0.05\sigma^2 - 1006.18$	$[-0.97, 8.24]$	0.39	$< 10^{-4}$
STD HD	$\hat{\mu} = 0.12\sigma^2 - 1002.86$	$[-0.52, 6.71]$	0.26	$< 10^{-4}$
BONE LD	$\hat{\mu} = 0.02\sigma^2 - 1009.67$	$[0.34, 16.55]$	0.58	$< 10^{-4}$
BONE HD	$\hat{\mu} = 0.04\sigma^2 - 1006.62$	$[-1.55, 8.80]$	0.24	$< 10^{-4}$
B31f LD	$\hat{\mu} = 0.02\sigma^2 - 1003.18$	$[-1.36, 4.52]$	0.20	$< 10^{-4}$
B31f HD	$\hat{\mu} = 0.02\sigma^2 - 1001.44$	$[-0.97, 0.77]$	0.07	$< 10^{-4}$
B45f LD	$\hat{\mu} = 0.02\sigma^2 - 1003.79$	$[0.36, 15.27]$	0.45	$< 10^{-4}$
B45f HD	$\hat{\mu} = 0.03\sigma^2 - 1004.21$	$[-1.88, 5.47]$	0.17	$< 10^{-4}$

nation R-squared that accounts for the explained variance shows a strong linear relationship between the attenuation level and the local variance with up to 58% of the variance explained by a linear model. Finally, the p-value obtained for the F-test confirms
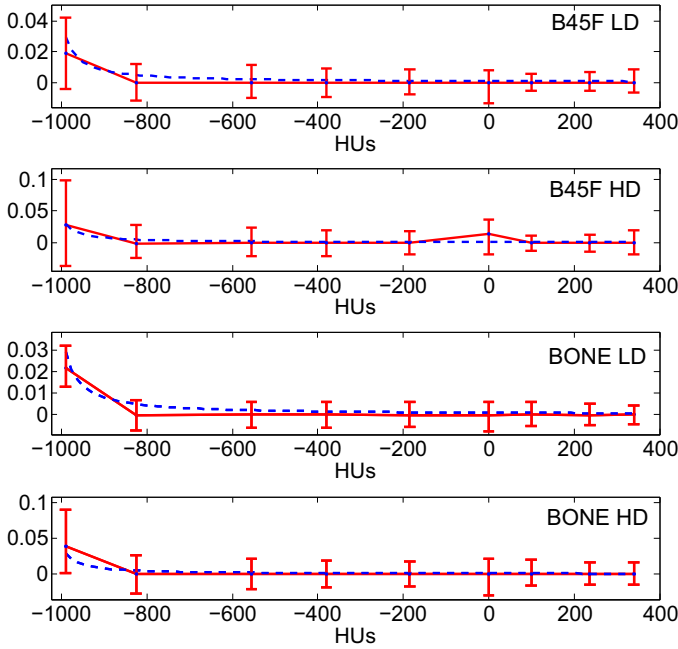
Fig. 5. Functional dependence of bias for the linear coefficient of bias (kernels B31f and STD were omitted for brevity, although the behavior is the same).

with a strong evidence (p-value $< 10^{-4}$) that both parameters – attenuation and local variance– exhibit a linear relationship and, therefore, the null hypothesis can be rejected. Note that although the linear coefficient may seem small, the variance observed in the images make this factor non-negligible. Actually, the bias interval calculated for the 95% of confidence may go up to 15 HU (see Table 1), which is a remarkably high number that may affect quantitative measures based on thresholding such as emphysema or gas trapping (Parr et al., 2004).

Assuming the linear model $\mu = a\sigma^2 + b$ confirmed by the statistical test, we now study the influence of the nominal attenuation level on the linear coefficient, a, for each tissue $n = \{1, \ldots, 9\}$. Then, the assumed linear model becomes:

$$\mu_n = a_n \sigma^2 + b_n, \tag{3}$$

The functional relationship of the linear coefficient with respect to the nominal attenuation level can be derived by writing a_n as a function of the local mean and variance according to the linear model of Eq. (3):

$$a_n = \frac{\widehat{\mu}_n - b_n}{\widehat{\sigma}_n^2}, \qquad \text{with } n = 1, \ldots, 9, \tag{4}$$

where b_n is the intercept coefficient of Table 1 and $\widehat{\mu}_n$ and $\widehat{\sigma^2}_n$ are the local mean and variance for locations of each tissue type Ω_n, i.e. $\widehat{\mu}_n(x, y) = \widehat{\mu}(x, y)$ and $\widehat{\sigma}_n^2(x, y) = \widehat{\sigma}^2(x, y)$ for $(x, y) \in \Omega_n$. The results for the different configurations of acquisitions are shown in Fig. 5 where the 95% confidence interval is shown in red. Note that the value of a_n rapidly decreases as the attenuation level grows. This behavior is well described by an inversely proportional relationship with $\mu - \delta$, which always falls within the 95% of confidence for all the different configurations (kernels and doses).

According to this analysis, we can formulate a mathematical model describing the bias induced by non-stationary noise considering the linear relationship with σ^2 previously tested and the inverse relationship with $\mu - \delta$:

$$X(\mathbf{r}) = \mu(\mathbf{r}) + \underbrace{\frac{\sigma^2(\mathbf{r})}{\mu(\mathbf{r}) - \delta}}_{\substack{\text{non-stationary} \\ \text{bias}}} + \underbrace{b}_{\substack{\text{systematic} \\ \text{bias}}} \tag{5}$$

where $\mu(\mathbf{r})$ is the actual attenuation coefficient and X the observed attenuation at location $\mathbf{r} \in \Omega$. The linear term with $\sigma^2(\mathbf{r})$ accounts for the increasing bias with higher variance and the inverse relationship with $\mu(\mathbf{r})$ shows a higher influence for lower attenuation levels, whereas the systematic bias, b, depends on the calibration of the device or the DC contribution of the reconstruction kernels.

3. Methods

3.1. CT signal model

According to Eq. (5), the bias correction will require the calculation of accurate estimates of local variance and local mean. This is achieved by considering the statistical characterization of noise proposed in Vegas-Sánchez-Ferrero et al. (2017), which adopts a non-central Gamma (nc-Γ) model as a suitable distribution that models the stochastic behavior of homogeneous tissues in CT scans with different reconstruction kernels, doses or devices. This model is a three-parameter distribution whose density function (PDF) is defined as:

$$f_X(x|\alpha, \beta, \delta) = \frac{(x - \delta)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x-\delta}{\beta}}, \quad x \geq \delta \text{ and } \alpha, \beta > 0, \tag{6}$$

where $\Gamma(x)$ is the Euler Gamma function, α is the shape parameter, β is the scale parameter, and δ is defined as the least attenuation level (typically around −1000 HU, in our case −1024 HU). In Fig. 4 we show the nc-Γ distribution (continuous lines) fitted to the data.

Its generalization for describing the heterogeneous nature of tissues is formulated as a non-central Gamma Mixture model (nc-ΓMM), whose PDF reads as:

$$p(x|\Theta) = \sum_{j=1}^{J} \pi_j f_X(x|\Theta_j), \tag{7}$$

with π_j the weights of the mixture, and Θ_j the parameters of each jth component. As in Vegas-Sánchez-Ferrero et al. (2017), the mixture model parameters are estimated by Expectation-Maximization.

The local moments of the observed intensity, X, are estimated taking advantage of the tissue characterization provided by the mixture model conditioned to each component:

$$E\{X(\mathbf{r})^k\} = \sum_{j=1}^{J} \pi_j E\{X(\mathbf{r})^k | Z(\mathbf{r}) = j\} \approx \sum_{j=1}^{J} \pi_j \langle X(\mathbf{r})^k | Z(\mathbf{r}) = j \rangle, \tag{8}$$

where the variable $Z(\mathbf{r})$ describes the membership of the random variable (RV) at location $\mathbf{r}$ to the different nature of tissues (e.g. air, muscle, fat). Besides, the approximation to the local conditioned moments is performed with the operator $\langle \cdot | Z(\mathbf{r}) = j \rangle$ in a neighborhood $\eta(\mathbf{r})$ at location $\mathbf{r}$:
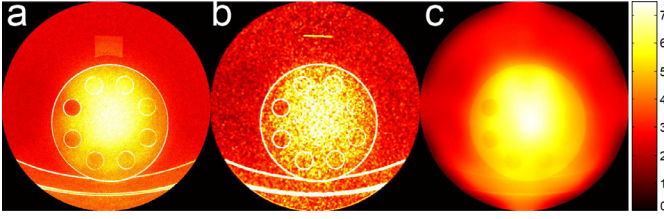
$$\langle X(\mathbf{r})^k | Z(\mathbf{r}) = j \rangle = \frac{\sum_{\mathbf{s} \in \eta(\mathbf{r})} x(\mathbf{s})^k \gamma_j(\mathbf{s})}{\sum_{\mathbf{s} \in \eta(\mathbf{r})} \gamma_j(\mathbf{s})}, \tag{9}$$

and $\gamma_j(\mathbf{r})$ is the posterior probability for the jth tissue class derived by the Bayes theorem as:

$$\gamma_j(\mathbf{r}) = p(Z(\mathbf{r}) = j | x(\mathbf{r}), \Theta) = \frac{p(x(\mathbf{r})|\Theta_j) p(Z(\mathbf{r}) = j | \Theta_j)}{p(x(\mathbf{r})|\Theta_j)}. \tag{10}$$

3.2. Homomorphic local variance estimation

The non-stationarity of variance depends on the tissue nature and also on the nearby structures that may induce an increase

Fig. 6. Local standard deviation of noise observed in the Phantom acquired with the BONE LD configuration in the GE scanner. a) sample standard deviation calculated in the long-axis direction; b) Local estimate obtained from one single slice; c) Homomorphic estimation with the proposed method.

of variance due to photon starvation and beam hardening. Unfortunately, the estimation of variance according to local neighborhoods will show a significant deviation from the actual value in the boundaries of structures. Fig. 6.a shows the standard deviation calculated in the longitudinal axis following Eqs. (1) and (2), and can be considered as a gold standard due to the cylindrical shape of the phantom. The local estimate of variance calculated with Eqs. (8)–(10) over one single slice is shown in Fig. 6.b. Note that this estimate exhibits a particular granular pattern. This pattern is due to the presence of different densities within the neighborhood and the relatively low number of samples in the neighborhood. As a result of this, it does not accurately describe the spatial variation but the edges of structures within the local neighborhood. To overcome this important limitation, we develop a homomorphic approach inspired by the methodology presented in Aja-Fernandez et al. (2015), where the authors apply a statistical model of noise to provide a homomorphic decomposition for the non-stationary variance (assumed to be smooth) and the noise fluctuations themselves. The estimation is finally achieved by taking advantage of the Gaussian convergence as the signal-to-noise ratio (SNR) increases. Additionally, some correction functions are provided for low SNRs. Pieciak et al. (2016) extended this approach by increasing the converge to a Gaussian distribution by a functional transformation.

In our extension to CT scans, we adopt the nc-ΓMM characterization as the model of noise, and the stabilization scheme for nc-ΓMM to increase the convergence range to a Gaussian distribution, both proposed in Vegas-Sánchez-Ferrero et al. (2017).

Formally, let us suppose the image $X : \Omega \to \mathbb{R}$ is distributed as a nc-ΓMM with central parameters $\delta_j = \delta$ and $Y(\mathbf{r}) = X(\mathbf{r}) - \delta$ the centered Gamma mixture model.

This mixture model can be transformed to a more treatable mixture model with the stabilizing function proposed in Vegas-Sánchez-Ferrero et al. (2017) for nc-ΓMM:

$$f_{\text{stab}}(Y(\mathbf{r})) = \sigma(\mathbf{r}) \cdot \frac{\sqrt{Y(\mathbf{r})} - E\left\{\sqrt{Y(\mathbf{r})}\right\}}{\sqrt{\text{Var}\left\{\sqrt{Y(\mathbf{r})}\right\}}} = \sigma(\mathbf{r}) \cdot N(\mathbf{r}), \qquad (11)$$

where $\sigma(\mathbf{r})$ accounts for the unknown non-stationary standard deviation and $N(\mathbf{r})$ is the estimate of the stabilized noise, i.e. zero mean and standard deviation equals 1. This transformation shows important features to obtain a good estimate of the non-stationary variance:

1. It transforms each component of the mixture model into its corresponding stabilized counterpart, i.e. each component shows a more independent behavior between mean and variance.
2. The stabilization accelerates the Gaussian convergence as a function of the signal-to-noise (SNR) ratio.

The homomorphic filter can now be applied considering that the non-stationary variance varies smoothly compared to the vari-
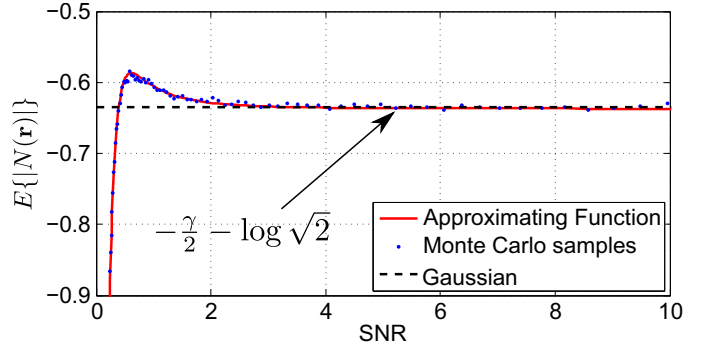
Fig. 7. Expectation of the log-stabilized signal $\log |N(\mathbf{r})|$ as a function of the signal-to-noise ratio. The fast convergence to a $-\gamma/2 - \log\sqrt{2}$ allows applying the Gaussian estimation of non-stationary variance for SNR > 2, whereas a correction factor can be easily used for SNR ≤ 2.

ability of noise itself (i.e. $\sigma(\mathbf{x})$ is a low-frequency signal compared to noise). The low-frequency contribution can be retrieved with a low-pass filter (LPF):

$$\text{LPF}\{\log|f_{\text{stab}}(Y(\mathbf{r}))|\} = \underbrace{\log\sigma(\mathbf{x})}_{\substack{\text{lower} \\ \text{frequency}}} + \underbrace{\text{LPF}\{\log|N(\mathbf{r})|\}}_{\substack{\text{higher} \\ \text{frequency}}}. \qquad (12)$$

The LPF can be seen as a local averaging that approximates the expectation of the random variable:

$$\text{LPF}\{\log|f_{\text{stab}}(Y(\mathbf{r}))|\} \approx \log\sigma(\mathbf{x}) + E\{\log|N(\mathbf{r})|\}. \qquad (13)$$

It is important to note that the calculation of $E\{\log|N(\mathbf{r})|\}$ can be done considering each of the components of the mixture model separately since the expectation is a linear operator and the PDF of the transformed mixture model remains a mixture model (i.e. the PDF is still a convex sum of the PDFs of the transformed components). This way, for this derivation we can assume $Y(\mathbf{r})$ to be a central Gamma distribution with parameters $\alpha(\mathbf{r})$ and $\beta(\mathbf{r})$, and $W = \sqrt{Y(\mathbf{r})}$ becomes a generalized Gamma distribution (Stacy, 1962; Vegas-Sanchez-Ferrero et al., 2012) with PDF:

$$f_W(w) = 2\frac{w^{2\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{w^2}{\beta}}, \quad \text{with } \alpha, \beta > 0, \qquad (14)$$

and moments:

$$E\{W^k\} = \beta^{k/2} \frac{\Gamma(\alpha) + \frac{k}{2}}{\Gamma(\alpha)}. \qquad (15)$$

Therefore, the expectation of $\log|N(\mathbf{r})|$ becomes:

$$E\{\log|N(\mathbf{r})|\} = \int_0^\infty \log\left|w - \frac{\Gamma(\alpha) + 1/2}{\Gamma(\alpha)}\right| 2\frac{w^{2\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{w^2}{\beta}} \, dw$$
$$- \frac{1}{2}\log\left(\alpha - \left(\frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)}\right)^2\right) \qquad (16)$$

Note that Eq. (16) no longer depends on β due to the normalization applied in the calculation of $N(\mathbf{r})$.

This integral was estimated by Monte Carlo trials of 10^5 samples for SNR $\in [10^{-3}, 20]$ logarithmically sampled. Fig. 7 shows the results where a fast convergence to $-\log\sqrt{2} - \gamma/2 \approx -0.6352$ is clearly observed, being γ the Euler-Mascheroni constant. This specific value comes from the expectation of a Gaussian RV with zero mean and σ^2 variance, $\mathcal{N}(0, \sigma^2)$, since $E\{\log|\mathcal{N}(0, \sigma^2)|\} = \log\sigma - \log\sqrt{2} - \gamma/2$ (see (Aja-Fernandez et al., 2015) for more de-
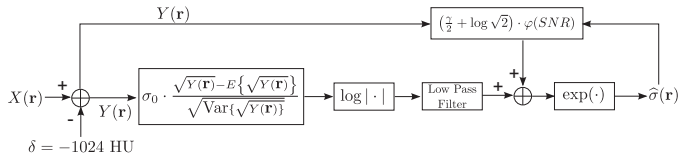
Fig. 8. Pipeline of $\sigma(\mathbf{r})$ estimation assuming noise distributed as a nc-ΓMM.

tails). This fact exhibits the rapid convergence to a Gaussian distribution of the stabilized random variable as the SNR increases.[2]

For practical purposes, we provide an approximating function of Eq. (16) also represented in Fig. 7:

$$E\{\log|N(\mathbf{r})|; \text{SNR} = x\} \approx -\left(\frac{\gamma}{2} + \log\sqrt{2}\right) \cdot \varphi(x) \qquad (17)$$

with

$$\varphi(x) = \exp\left(\frac{49x^4 + 255x^3 - 1808x^2 + 665x^1 + 2}{x^5 + 21287x^4 - 6169x^3 + 2060x^2 + 365x}\right). \qquad (18)$$

Finally, $\sigma(\mathbf{r})$ is derived by rearranging Eq. (13) with the result of Eq. (17):

$$\widehat{\sigma}(\mathbf{r}) = e^{\text{LPF}\{\log|f_{\text{stab}}(Y(\mathbf{r}))|\} + (\gamma/2 + \log\sqrt{2}) \cdot \varphi(\text{SNR})} \qquad (19)$$

Note that this estimation requires the SNR as a parameter that can be directly calculated with Eqs. (8)–(10). This formulation can be iteratively refined until convergence by introducing $\widehat{\sigma}(\mathbf{r})$ in the calculation of the SNR in each iteration.

The pipeline of this method is depicted in Figs. 8 and 6.c shows the estimate of $\sigma(\mathbf{r})$ obtained in a single slice of the phantom acquired in a GE scanner with the BONE LD configuration with the proposed iterative scheme after convergence is reached (4 iterations in this case to get a relative change below 1%).

The mean squared error (MSE) of the estimate was calculated using the z-axis estimate shown in Fig. 6.a as the reference. Results obtained with our approach (Fig. 6.c, MSE = 268.97) clearly overcomes the previously mentioned limitations (Fig. 6.b, MSE = 336.87) and provides a smooth and reasonably accurate estimate of the non-stationary variance.

3.3. Autocalibration

The proposed autocalibration method for bias correction is based on the functional relationship derived in Section 2.1. From Eq. (5) we can remove the non-stationary bias as:

$$\widetilde{X}(\mathbf{r}) = X(\mathbf{r}) - \frac{\widehat{\sigma}^2(\mathbf{r})}{\widehat{\mu}(\mathbf{r}) - \delta} \qquad (20)$$

where $\widehat{\mu}(\mathbf{r})$ and $\widehat{\sigma}^2(\mathbf{r})$ are the local mean and local variance estimates calculated with the conditioned moments of Eqs. (8)–(10).

The systematic bias is then removed by imposing the attenuation levels of certain anatomical areas with well-defined references. For chest scanners, air cavities offer a well-defined anatomic reference in which air attenuation levels should be clearly present. This attenuation is set to -1000 HU according to the definition of the Hounsfield unit scale. On the other hand, for higher attenuation levels one can find relatively homogeneous regions either in muscle, fat, or blood.

Although the selection of fairly homogeneous regions can be easily achieved by manual segmentation, the statistical characterization proposed in Vegas-Sánchez-Ferrero et al. (2017) provides a suitable way that enables the automatic detection of those regions. Note that we use the same formulation to characterize tissues and to calculate conditioned moments in local neighborhoods.
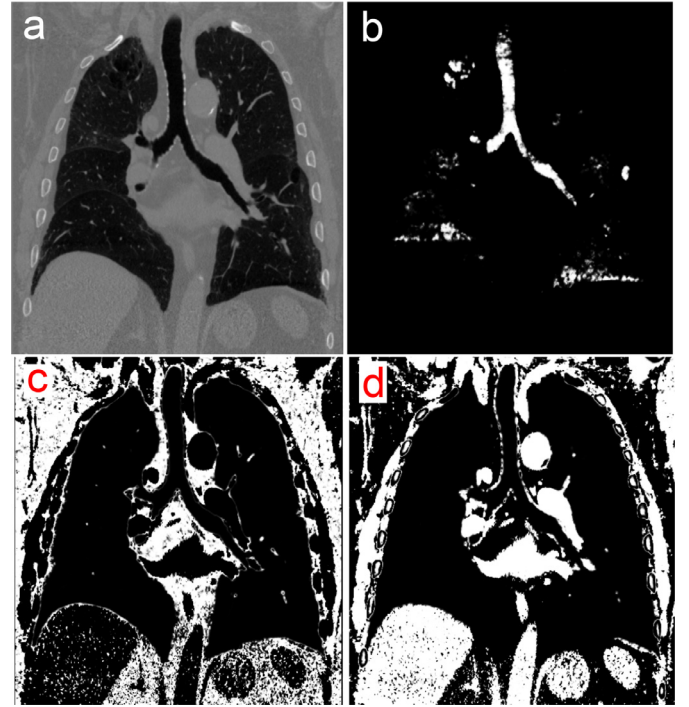
---

[2] In this case, we define the SNR as μ/σ, which in terms of a Gamma distribution of parameters α and β becomes SNR = $\sqrt{\alpha}$.

Fig. 9. Posterior probability maps for a conventional CT scan, a) coronal view; b) posterior probability for air; c) Posterior probability for fat; d) posterior probability for blood/muscle.

Thus, the classification of tissues (e.g. *air* and *blood*) can be done automatically considering Eq. (10) with no extra cost. The results can be defined as the regions Ω_i are those locations with a posterior probability higher than a certain threshold, t, for each tissue class:

$$\Omega_i = \{\mathbf{r} \in \Omega : p(Z(\mathbf{r}) = j|y(\mathbf{r}), \Theta) > t\}, \text{ with } i = 1, \dots, n \qquad (21)$$

with $Y(\mathbf{r}) = X(\mathbf{r}) - \delta$ the centered random variable at location $\mathbf{r}$ and $y(\mathbf{r})$ its observed realization.

In Fig. 9 we show the automatic selection of different regions according to the tissue characterization provided by the mixture model of Eq. (7) proposed in Vegas-Sánchez-Ferrero et al. (2017). The resulting posterior probabilities clearly define the trachea as an air cavity, while the visceral fat and muscle are also identified.

The removal of the systematic bias is extended to the whole range of attenuation levels by a piecewise linear interpolation which centers the average attenuation of each of the reference regions Ω_i to their respective nominal value, μ_i, as follows:

$$\widehat{X}(\mathbf{r}) = \widetilde{X}(\mathbf{r}) - E\{\widetilde{X}(\mathbf{r})\} + \widehat{b}(\widetilde{X}(\mathbf{r})), \qquad (22)$$

where the calibrated region is set according to the local average by the piecewise function $\widehat{b}(\cdot) : \mathbb{R} \to \mathbb{R}$ defined for $\{\Omega_i\}_{i=1}^n$ regions is:

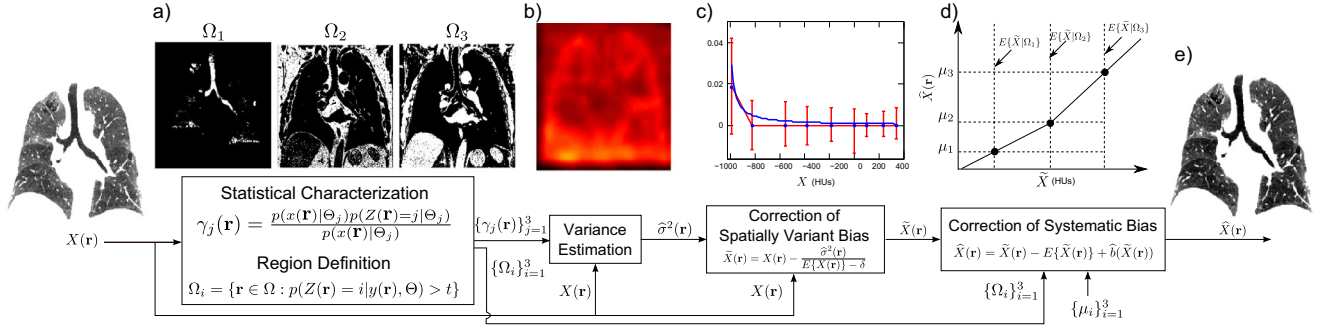$$\widehat{b}(\widetilde{X}(\mathbf{r})) = \begin{cases} M_1(\widetilde{X}(\mathbf{r})) & \text{if } \widetilde{X} < E\{\widetilde{X}|\Omega_2\} \\ \vdots \\ M_i(\widetilde{X}(\mathbf{r})) & \text{if } E\{\widetilde{X}|\Omega_i\} \le \widetilde{X} < E\{\widetilde{X}|\Omega_{i+1}\} \\ \vdots \\ M_{n-1}(\widetilde{X}(\mathbf{r})) & \text{if } E\{\widetilde{X}|\Omega_{n-1}\} \le \widetilde{X} \end{cases} \qquad (23)$$

where $E\{\widetilde{X}|\Omega_i\}$ is defined in Eq. (9), and $M_i(\cdot)$ is the linear interpolator defined as:

$$M_i(\widetilde{X}) = \mu_i + (\mu_{i+1} - \mu_i)\frac{\widetilde{X} - E\{\widetilde{X}|\Omega_i\}}{E\{\widetilde{X}|\Omega_{i+1}\} - E\{\widetilde{X}|\Omega_i\}} \qquad (24)$$

with μ_i the nominal attenuation level for region Ω_i.

Fig. 10. Autocalibration scheme considering three nominal attenuation levels $\{\mu_i\}_{i=1}^3$. a) The autocalibration is accomplished considering three different regions, $\{\Omega_i\}_{i=1}^3$, defined by the tissue characterization of Eq. (7). b) The variance estimate is computed following the scheme shown in Fig. 8. c) the correction of the spatially variant bias is accomplished by applying the functional relation of Eq. (20). d) The systematic bias is corrected by assigning the nominal values $\{\mu_i\}_{i=1}^3$ to the observed average attenuation levels after non-stationary bias correction with a piecewise linear interpolation. e) The autocalibrated image.

A comprehensive scheme of the proposed autocalibration method is described in Fig. 10, where we show the different steps of the for the calibration considering three different regions, $\{\Omega_i\}_{i=1}^3$, defined by the tissue characterization of Eq. (7) (Fig. 10.a). In Fig. 10.b, the variance estimate is computed by following the scheme shown in Fig. 8. Then, in Fig. 10.c, the correction of the spatially variant bias is accomplished by applying Eq. (20). Finally, the systematic bias is corrected by assigning the nominal values $\{\mu_i\}_{i=1}^3$ to the observed average attenuation levels after non-stationary bias correction with a piecewise linear interpolation, Fig. 10.d. The resulting image, Fig. 10.e, is now calibrated with no stationary bias introduced by noise.

4. Results

The autocalibration methodology is validated firstly evaluating the reduction of the intra-device bias due to non-stationary noise. This is performed in the phantom acquired with different configurations per device. Then, we carry out a similar evaluation with a subject acquired with different configurations (doses and reconstruction methods) by evaluating the bias within the trachea. Finally, we test the performance of the autocalibration to correct the inter-device bias.

4.1. Intra-device validation

4.1.1. Phantom study

The autocalibration methodology proposed in Section 3.3 is evaluated with the Phantom already introduced in the exploratory analysis of bias. This evaluation will demonstrate the robustness of our methodology, the importance of the bias due to the non-stationary variance for low attenuation level, and the efficiency of the suggested scheme in removing systematic biases that depend on the device. Two regions were considered for the autocalibration: water density (0 HU) and air (-1000 HU). The regions Ω_{water} and Ω_{air} were automatically selected as described in Eq. (21) for $t = 0.9$.

The analysis of bias due to noise was measured with respect to the average attenuation level of each tissue in the regions described in Fig. 1 for a 95% confidence interval. The reduction of bias is shown in Table 2 and was calculated as:

$$\text{Bias Reduction} = 100 \cdot \left(1 - \frac{\max(|B_{\text{w}}|)}{\max(|B_{\text{w/o}}|)}\right)\%, \tag{25}$$

were B_{w} and $B_{\text{w/o}}$ are the bias intervals for 95% of confidence with and without removal of non-stationary variance.

Note that the bias is reduced in most of the cases to less than a Hounsfield unit, which implies that the reduction is always over

Table 2
Correction of bias due to non-stationary noise.

Air Density	Non-calibrated Bias interval 95% (HU)	Calibrated Bias interval 95% (HU)	Bias Reduction
STD LD	$[-3.79, 11.81]$	$[-0.41, 1.88]$	84.09%
STD HD	$[1.09, 9.91]$	$[-1.79, 0.13]$	81.91%
BONE LD	$[0.98, 14.31]$	$[-0.28, 4.07]$	71.17%
BONE HD	$[1.04, 8.32]$	$[-0.23, 0.41]$	95.04%
B31f LD	$[-0.97, 3.09]$	$[-0.01, 0.04]$	98.68%
B31f HD	$[-0.59, 1.54]$	$[-0.04, 0.01]$	97.64%
B45f LD	$[2.58, 10.49]$	$[0.05, 1.67]$	84.12%
B45f HD	$[-0.08, 3.54]$	$[-0.13, 0.01]$	96.32%

Water Density	Non-calibrated Bias interval 95% (HU)	Calibrated Bias interval 95% (HU)	Bias Reduction
STD LD	$[-11.12, -8.40]$	$[-0.25, 0.50]$	95.52%
STD HD	$[-12.43, -9.65]$	$[-0.11, 0.05]$	99.08%
BONE LD	$[-9.82, -7.27]$	$[-0.62, 0.68]$	93.04%
BONE HD	$[-14.07, -7.89]$	$[-0.31, 0.32]$	97.70%
B31f LD	$[-11.03, 1.86]$	$[-0.25, 0.21]$	97.76%
B31f HD	$[-6.61, 3.62]$	$[-0.09, 0.33]$	95.00%
B45f LD	$[-12.84, 1.89]$	$[0.09, 0.27]$	97.94%
B45f HD	$[-10.23, 3.48]$	$[0.07, 0.26]$	97.48%

70%. The worst scenario is the BONE low-dose reconstruction, that shows a bias up to 4 HU, which is a fairly good result considering the initial bias was over 14 HU.

4.1.2. Clinical CT scans

In this experiment, we considered two different acquisitions of the same subject with different doses (400 mA and 100 mA) in the Siemens scanner[3] A coronal view of the subject is depicted in Fig. 9.a. The reconstruction kernels applied to higher doses were B31f and B45f, whereas the B31f was used for the LD reconstruction. Additionally, three different implementations of the Siemens iterative reconstruction methods were applied for the LD acquisition: I31f2, I44f2 and I31f5, providing a reconstruction from a softer to a sharper effect in the reconstruction respectively. Both acquisitions were registered with *Advanced Normalization Tools*[4] for easier comparison.

The autocalibration was done considering regions with blood/muscle density (30 HU (Aubrey et al., 2014)) and air (-1000 HU). The regions Ω_{air} and $\Omega_{\text{blood/muscle}}$ were automatically selected as described in Eq. (21) for $t = 0.9$ (both regions are depicted in Fig. 9.b and d). A coronal view of the reconstructions is shown in Fig. 11 jointly to the autocalibrated reconstruction

---

[3] The data was obtained at Brigham and Women's Hospital (Boston, MA, USA) with the approval of its ethics committee and informed consent of the subject.

[4] http://stnava.github.io/ANTs/.
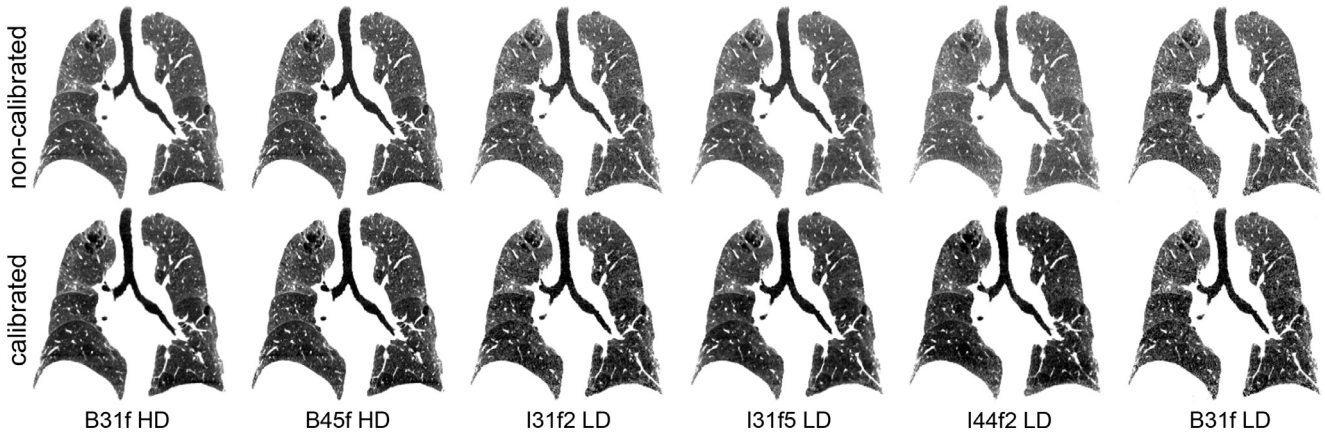
Fig. 11. Coronal view of the reconstructed images with different configurations for visualization window [−1024, −500] HU. Columns are ordered by reconstruction configurations (both with systematic and spatially-variant bias), while rows represent the non-calibrated (top) and calibrated (bottom) acquisitions (after systematic and spatially-variant bias correction).
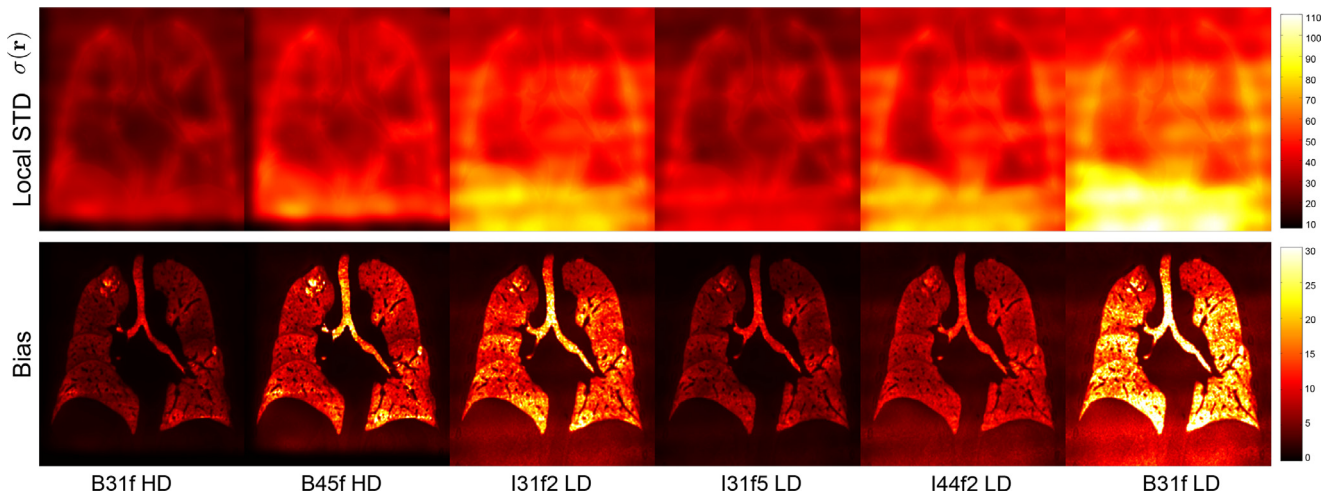
Fig. 12. Local standard deviation estimate (top row) and induced non-stationary bias (bottom row). The bias observed is especially affecting to the lung parenchyma and trachea due to the higher skewness of attenuations levels for low densities. Average biases observed in the trachea are: B31f HD 8.30 ± 1.83 HU; B45f HD 17.68 ± 4.87 HU; I31f2 LD 21.36 ± 7.07 HU; I31f5 LD 9.38 ± 2.21 HU; I44f2 LD 11.65 ± 2.06 HU; B31f LD 31.65 ± 12.78 HU.

(after the systematic and non-stationary bias correction) for an intensity window of [−1024, −500] HU to enhance the effect of bias in lower attenuation levels.

The local standard deviation, $\sigma(\mathbf{r})$, estimated according to the methodology proposed in Section 3.1 is shown in Fig. 12 (top row). The results show a remarkable increase of variance in the lower part due to the higher densities of soft tissues compared to lung parenchyma. The non-stationary bias induced by the local variance is shown in the bottom row of Fig. 12. As expected, the bias observed is especially affecting the lung parenchyma and trachea due to the higher skewness of attenuations levels for low densities. The effect of this bias is markedly higher in the B31f LD acquisition, which does not take advantage of the iterative reconstruction methodologies. Iterative methods partially attenuate the bias observed with reconstruction kernels. However, the observed bias in trachea and parenchyma are not negligible. The best case is for I31F5 showing an average bias in the trachea of 9.38 ± 2.06 HU, while the worst scenario was for I31f2 showing a bias of 21.36 ± 7.07 HU which is comparable to the observed for the reconstruction kernel B31f for low-dose (31.65 ± 12.72 HU).

A careful analysis of the local standard deviations and the reconstructed images shows the overall effect of the different reconstruction methods. The low-dose reconstructions exhibit a higher variance of noise, which induces an increase of the local attenua-

tion level in the trachea. There is also a systematic bias introduced by the iterative methods I31f2 and I44f2 that can be observed in the overall reduction of contrast in the lung parenchyma. The histograms in the trachea depicted in Fig. 13 quantitatively confirm the bias introduced by the reconstruction methods and the spatial-varying variance of noise (the sample histograms are represented as dots, and the parametric approximation with the nc-Γ distribution is shown for a better visualization). Note that the low-dose acquisitions are clearly shifted to higher values than those obtained for high-dose. This behavior shows a clear inconsistency in the mean value due to the different reconstruction methods. The box-plots evidence the strong deviation with respect the nominal air attenuation level (−1000 HU) of more than 50 HU in each reconstructed image and the notable discrepancy among them. On the other hand, the autocalibration method is able to remove the deviations and exhibits a consistent attenuation level across all the reconstructions.

4.2. Inter-device study

The analysis of the systematic bias introduced by the reconstruction method in each device is studied by calculating the deviation from the nominal attenuation level in the phantom. The results for both devices are depicted in Fig. 14, where the biases

Table 3
Bias observed for scanners without calibration (W/C) and autocalibrated (Cal) for the different acquisitions. Bias under 2 HU are marked in bold. The calibration reduces dramatically the deviations from the nominal attenuation levels, which now remains around 0 HU.

Region	STD LD		STD HD		BONE LD		BONE HD		B31f LD		B31f HD		B45f LD		B45f HD	
	W/C	Cal	W/C	Cal	W/C	Cal	W/C	Cal	W/C	Cal	W/C	Cal	W/C	Cal	W/C	Cal
1	24.81	**0.24**	23.57	−0.09	19.55	**0.08**	21.13	−0.13	12.41	−0.02	12.43	−0.08	8.21	−0.10	10.06	−0.10
2	22.22	**1.14**	26.29	**1.02**	19.11	**1.14**	23.80	**0.51**	14.00	**1.00**	13.49	**0.79**	12.15	**1.08**	12.48	**0.83**
3	3.85	−0.67	14.91	−0.07	2.50	−0.65	13.76	**0.09**	−0.09	−0.41	0.73	−0.20	−1.11	−0.49	−0.37	−0.18
4	10.58	**1.37**	11.94	**0.91**	8.92	**1.29**	11.01	**0.64**	4.28	**0.78**	3.82	**0.64**	3.91	**0.63**	3.45	**0.50**
5	−35.72	−1.61	−28.51	−1.30	−34.81	−1.67	−27.97	−1.11	−13.41	−0.75	−15.33	−0.65	−11.70	−0.64	−13.94	−0.63
6	−28.69	**0.92**	−26.26	**0.55**	−27.42	**0.82**	−24.99	**0.47**	−17.18	**0.02**	−17.53	**0.03**	−15.64	**0.09**	−16.27	−0.00
7	−69.62	−1.98	−64.82	−1.84	−64.37	−1.73	−62.87	−1.60	−47.12	−1.74	−47.17	−1.70	−45.45	−1.52	−44.96	−1.46
8	−52.78	**0.06**	−47.64	**0.12**	−50.45	−0.08	−46.26	**0.05**	−27.07	**0.36**	−26.50	**0.40**	−27.56	**0.11**	−26.90	−0.01
9	**−1.80**	**0.23**	−4.04	**0.16**	−6.24	**0.33**	−4.09	**0.21**	−0.43	**0.06**	0.03	**0.06**	−5.35	**0.07**	−0.88	**0.05**
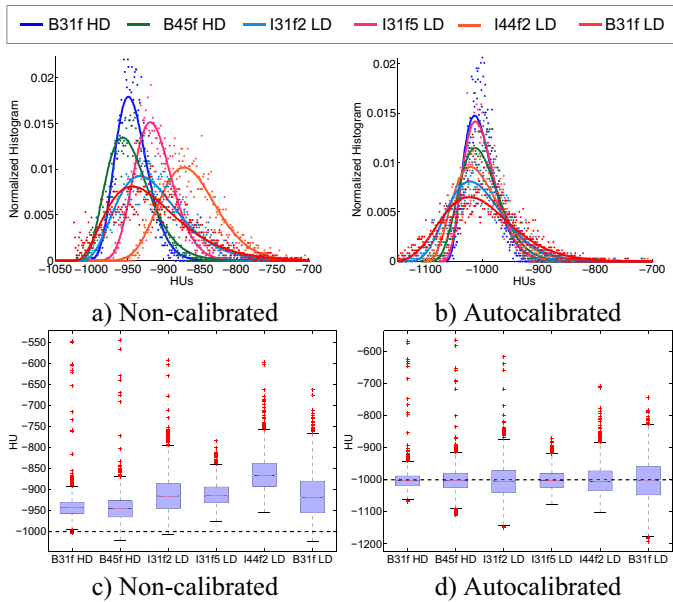
Fig. 13. Histograms of the trachea region for the different reconstructions shown in Fig. 11 and Box-plots for non-calibrated and autocalibrated data. The non-calibrated histograms show a important discrepancy with the nominal attenuation level of air (−1000 HU) and an inconsistent bias depending on the reconstruction method and doses. Conversely, autocalibrated histograms show a consistent average location among them and with the nominal attenuation value of air.
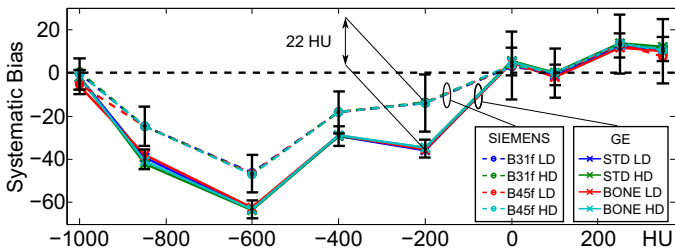
Fig. 14. Systematic bias observed in each device. This bias depends on the calibration of the device and the DC contribution of the reconstruction.

within devices are consistent across reconstruction kernels and doses. However, the bias exhibits clear differences between both devices that can be over 20 HU in some cases. This confirms the importance of a common calibration in heterogeneous cohorts in which different devices and reconstruction methods are applied.

We evaluated the performance of the autocalibration method to reduce the discrepancies between devices already shown in Fig. 14. The results are gathered in Table 3. Note that the systematic bias

is virtually removed by the autocalibration method for both the Siemens and the GE acquisitions.

It is worth noting that the observed bias for both devices shows two well-defined regions, one with a positive bias (for CT numbers $> 0 HU$) and the other for negative bias (CT numbers $< 0 HU$). This interesting behavior evidences that a non-linear response in the attenuation levels is fitted by a linear function with reference points those of the water and air attenuations. This behavior is consistent with the common calibration protocols applied in the scanners, that use the air and water references. A higher order polynomial behavior has been reported previously in the literature as a result of the interplay between the Photoelectric and Compton components of the attenuation coefficient (Jackson and Hawkes, 1983; Watanabe, 1999).

5. Conclusion

This paper analyzes the effect of reconstruction inconsistencies depending on the dose, reconstruction algorithm, and acquisition parameters. These inconsistencies show a two-folded nature. First, the variance of noise introduces a positive bias that artificially increases the intensity values. This effect has been systematically noticed in the literature due to the strong deviations observed in the air cavities such as the trachea. Second, the kernels and iterative algorithms used for reconstruction may affect the average intensities. This results in a strong intensity deviation that may differ from the nominal values in more than 50 HUs, making the reconstructions useless for multicenter studies or analysis of disease progression.

The analysis was carried out through a statistical exploratory data analysis of bias in a series of acquisitions provided with different devices, reconstruction kernels, and doses. The analysis concluded that the bias introduced by the local variance contributes linearly to the mean attenuation level, being remarkably higher for lower attenuations levels. This bias may increase the intensity level up to 30 HUs in low-dose reconstructions. The linear relationship decreases quickly with higher tissue densities, i.e. the linear coefficient decreases with the nominal attenuation coefficient of tissues.

In the light of all these results, a functional dependence of the observed attenuation with the local variance can be established as a useful model to retrieve the unbiased signal. As a first contribution, we propose a model that considers a positive bias as a function of the local moments. The model also considers the systematic bias due to the different behaviors of reconstruction kernels or algorithms.

The retrieval of the unbiased signal requires an accurate estimate of the non-stationary local variance. Our second contribution is the derivation of an accurate and robust estimate of non-stationary variance by adapting the probabilistic model of noise in CT scans proposed in Vegas-Sánchez-Ferrero et al. (2017).

The functional relation adopted to model the bias allows us to correct the two different manifestations of bias. First, the spatially variant bias induced by the non-stationary noise. Second, the systematic bias introduced by different reconstruction methodologies or doses. The proposed methodology does not require a phantom, and it is designed to establish a common framework for the analysis of CT scanners without a common calibration protocol, different reconstruction methodologies or doses.

The autocalibration methodology was tested in a phantom resulting in a bias removal of more than 90% in most of the cases for the bias due to noise variance. Additionally, the systematic bias was virtually removed. These results evidence the suitability of the bias model proposed and the methodology for bias removal.

The experiments carried out in clinical CT images considered several acquisition scenarios (two doses, different reconstruction kernels and iterative methods). The biases originated by reconstruction methods, and different doses are clear at first sight. These differences become more evident when a quantitative analysis of an anatomical reference is performed. The autocalibration was evaluated in the trachea to study the consistency across the different acquisitions. The non-calibrated signals showed a systematic bias higher than 50 HU compared to the nominal intensity expected for air and an arbitrary behavior in terms of variance and location for the histogram due to the non-stationarity of noise and undesired effects of the iterative reconstruction methods. After calibration, the histograms align around the nominal intensity for air showing an excellent consistency across acquisitions.

The overall effect of autocalibration is an increase of the contrast in low-intensity values due to the positive bias introduced by the local variance dependence. This effect makes easier to detect low attenuation lesions, such us emphysema in chronic pulmonary obstruction disease (COPD) patients, due to the consistency across different conditions, and facilitates the analysis of progression. It is worth noting that the proposed autocalibration scheme can be perfectly combined with other state-of-the-art methodologies to mitigate or stabilize the noise just after the autocalibration (Kim et al., 2016; Vegas-Sánchez-Ferrero et al., 2017).

We believe the methodologies provided in this paper will enable the preprocessing of CT images before quantitative imaging analysis tasks are performed to control the accuracy of the extracted biomarkers. Our results can be especially relevant to those studies that track the progression of a disease. The changes in transmission medium due to the reduction of dose in longitudinal studies or oncological situations make the consistency in the intensity levels critical.

Acknowledgments

References

Aja-Fernandez, S., Pieciak, T., Vegas-Sanchez-Ferrero, G., 2015. Spatially variant noise estimation in MRI: a homomorphic approach. Med. Image Anal. 20 (1), 184–197.

Aubrey, J., Esfandiari, N., Baracos, V.E., Buteau, F.A., Frenette, J., Putman, C.T., Mazurak, V.C., 2014. Measurement of skeletal muscle radiation attenuation and basis of its biological variation. Acta Physiologica 210 (3), 489–497.

Brenner, D.J., Hall, E.J., 2007. Computed tomography — an increasing source of radiation exposure. N. Engl. J. Med. 357 (22), 2277–2284.

Buckler, A.J., Bresolin, L., Dunnick, N.R., Sullivan, D.C., Group, F., 2011. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 259 (3), 875–884.

Chen-Mayer, H.H., Fuld, M.K., Hoppel, B., Judy, P.F., Sieren, J.P., Guo, J., Lynch, D.A., Possolo, A., Fain, S.B., 2017. Standardizing CT lung density measure across scanner manufacturers. Med. Phys. 44 (3), 974–985.

Choo, J.Y., Goo, J.M., Lee, C.H., Park, C.M., Park, S.J., Shim, M.-S., 2014. Quantitative analysis of emphysema and airway measurements according to iterative reconstruction algorithms: comparison of filtered back projection, adaptive statistical iterative reconstruction and model-based iterative reconstruction. Eur. Radiol. 24 (4), 799–806.

Hess, E.P., Haas, L.R., Shah, N.D., Stroebel, R.J., Denham, C.R., Swensen, S.J., 2014. Trends in computed tomography utilization rates: a longitudinal practice-based study. J. Patient Saf. 10 (1), 52–58.

Hsieh, J., 2003. Computed Tomography: Principles, Design, Artifacts, and Recent Advances. SPIE Hoboken, N. J. John Wiley & Sons, Bellingham, Wash.

Jackson, D.F., Hawkes, D.J., 1983. Energy dependence in the spectral factor approach to computed tomography. Phys. Med. Biol. 28 (3), 289–293.

Kim, J.H., Chang, Y., Ra, J.B., 2016. Denoising of polychromatic ct images based on their own noise properties. Med. Phys. 43 (5), 2251–2260.

Kim, Y., Kim, Y.K., Lee, B.E., Lee, S.J., Ryu, Y.J., Lee, J.H., Chang, J.H., 2015. Ultra-low–dose CT of the thorax using iterative reconstruction: evaluation of image quality and radiation dose reduction. AJR Am. J. Roentgenol. 204 (6), 1197–1202.

Mayo-Smith, W.W., Hara, A.K., Mahesh, M., Sahani, D.V., Pavlicek, W., 2014. How i do it: managing radiation dose in CT. Radiology 273 (3), 657–672.

Millner, M.R., Payne, W.H., Waggener, R.G., McDavid, W.D., Dennis, M.J., Sank, V.J., 1978. Determination of effective energies in ct calibration. Med. Phys. 5 (6), 543–545.

Mulshine, J.L., Gierada David S. III, S.G.A., Avila, R.S., Yankelevitz, D.F., Kazerooni, E.A., Mcnitt-gray, M.F., Buckler, A.J., Sullivan, D.C., 2015. Role of the quantitative imaging biomarker alliance in optimizing CT for the evaluation of lung cancer screen e detected nodules. J. Am. College Radiol. 12 (4), 390–395.

Parr, D.G., Stoel, B.C., Stolk, J., Nightingale, P.G., Stockley, R.A., 2004. Influence of calibration on densitometric studies of emphysema progression using computed tomography. Am. J. Respir. Crit. Care Med. 170 (8), 883–890.

Pieciak, T., Aja-Fernandez, S., Sanchez-Ferrero, G.V., 2016. Non-stationary rician noise estimation in parallel mri using a single image: a variance-stabilizing approach. IEEE Trans Pattern Anal Mach Intell (99). 1–1.

Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., Curran-Everett, D., Silverman, E.K., Crapo, J.D., 2011. Genetic epidemiology of COPD (COPDGene) study design. COPD 7 (1), 32–43.

Sieren, J.P., Newell, J.D., Judy, P.F., Lynch, D.A., Chan, K.S., Guo, J., Hoffman, E.A., 2012. Reference standard and statistical model for intersite and temporal comparisons of CT attenuation in a multicenter quantitative lung study. Med. Phys. 39 (9), 5757–5767.

Stacy, E.W., 1962. A generalization of the gamma distribution. Ann. Math. Stat. 33, 1187–1192.

Uppot, R.N., Sahani, D.V., Hahn, P.F., Gervais, D., Mueller, P.R., 2007. Impact of obesity on medical imaging and image-guided intervention. AJR Am. J. Roentgenol. 188 (2), 433–440.

Vegas-Sanchez-Ferrero, G., Aja-Fernandez, S., Palencia, C., Martin-Fernandez, M., 2012. A generalized gamma mixture model for ultrasonic tissue characterization. Comput. Math. Methods Med. 2012.

Vegas-Sanchez-Ferrero, G., Ledesma-Carbayo, M.J., Washko, G.R., San José Estépar, R., 2017. Statistical characterization of noise for spatial standardization of CT scans: enabling comparison with multiple kernels and doses. Med. Image Anal. 40, 44–59.

Watanabe, Y., 1999. Derivation of linear attenuation coefficients from CT numbers for low-energy photons. Phys. Med. Biol. 44 (9), 2201–2211.

Whiting, B.R., 2002. Signal statistics of x-ray computed tomography. In: SPIE 4682, Medical Imaging. Physics of Medical Imaging 2002, pp. 53–60.

Xia, T., Alessio, A.M., De Man, B., Manjeshwar, R., Asma, E., Kinahan, P.E., 2012. Ultra-low dose CT attenuation correction for PET/CT. Phys. Med. Biol. 57 (2), 309–328.

Zeng, G.L., 2004. Nonuniform noise propagation by using the ramp filter in fan-beam computed tomography. IEEE Trans. Med. Imaging 23 (6), 690–695.