PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Estimation of in vivo noise in clinical CT images: comparison and validation of three different methods against ensemble noise gold-standard

Ria, Francesco, Smith, Taylor, Abadi, Ehsan, Solomon, Justin, Samei, Ehsan

Francesco Ria, Taylor B. Smith, Ehsan Abadi, Justin B. Solomon, Ehsan Samei, "Estimation of in vivo noise in clinical CT images: comparison and validation of three different methods against ensemble noise gold-standard," Proc. SPIE 11595, Medical Imaging 2021: Physics of Medical Imaging, 115952P (15 February 2021); doi: 10.1117/12.2580954

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

Estimation of *in vivo* noise in clinical CT images: comparison and validation of three different methods against ensemble noise gold-standard

Francesco Ria[a,b], Taylor B. Smith[b], Ehsan Abadi[a,c], Justin B. Solomon[a,b], Ehsan Samei[a,b,c]

a. Carl E. Ravin Advanced Imaging Labs, Duke University Health System, Durham (NC) USA
b. Clinical Imaging Physics Group, Duke University Health System, Durham (NC) USA
c. Medical Physics Graduate Program, Departments of Radiology, Physics, Biomedical Engineering, and Electrical and Computer Engineering, Duke University, Durham (NC) USA

Abstract

Image quality estimation is crucial in modern CT with noise magnitude playing a key role. Several methods have been proposed to estimate noise surrogates *in vivo*. This study aimed to ascertain the accuracy of three different noise-magnitude estimation methods. We used ensemble noise as the ground truth. The most accurate approach to assess ensemble noise is to scan a patient repeatedly and assess the noise for each pixel across the ensemble of images. This process is ethically undoable on actual patients. In this study, we surmounted this impasse using Virtual Imaging Trials (VITs) that simulate clinical scenarios using computer-based simulations. XCAT phantoms were imaged 47 times using a scanner-specific simulator (DukeSim) and reconstructed with filtered back projection (FBP) and iterative (IR) algorithms. Noise magnitudes were calculated in lung (ROIn), soft tissues (GNI), and air surrounding the patient (AIRn), applying different HU thresholds and techniques. The results were compared with the ensemble noise magnitudes within soft tissue (En). For the FBP-reconstructed images, median En was 30.6 HU; median ROIn was 46.6 HU (+52%), median GNI was 40.1 HU (+31%), and median AIRn 25.1 HU (-18%). For the IR images, median En was 19.5 HU; median ROIn was 31.2 HU (+60%), median GNI was 25.1 HU (+29%), and median AIRn 18.8 HU (-4%). Compared to ensemble noise, GNI and ROIn overestimate the tissue noise, while AIRn underestimates it. Air noise was least representative of variations in tissue noise due to imaging condition. These differences may be applied as adjustment or calibration factors to better represent clinical results.

1. Introduction

Modern optimization and justification in radiology rely on effective estimation of risks and benefits associated with a procedure (1,2). In CT, several metrics and methods have been developed to estimate the radiological risks associated with the radiation burden. Conversely, the procedure benefit, that is the effective diagnosis, can be only indirectly assessed. The likelihood of an accurate diagnosis is affected by the quality of the CT images which can be measured using surrogates such as image noise, resolution, and contrast (3,4). In particular, noise magnitude is one of the main indicator of image quality and it can be used as a surrogate for the quality assessment of the diagnostic image properties (4).

Medical Imaging 2021: Physics of Medical Imaging, edited by Hilde Bosmans, Wei Zhao,
Lifeng Yu, Proceedings of SPIE Vol. 11595, 115952P · © 2021 SPIE
CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2580954

Several methods have been introduced over the years to measure noise both in phantoms and in patients (4–8). The most accurate approach is to assess ensemble noise by scanning a patient multiple times and sampling each pixel noise within the ensemble of images, an ethically undoable repeated imaging process (5). Other approaches rely on different strategies to segment the images, apply HU thresholds, and define regions of interest in which the noise magnitude is calculated. As a consequence, it is impossible to compare the results obtained from these methods. This scenario is further complicated by the dose reduction techniques introduced in modern CT, such as Automated Tube Current Modulation and Iterative Reconstruction algorithms (IR) that have changed the traditional statistical relationships between radiation dose and image quality (9). Inconsistency in noise magnitude estimations can negatively affect the overall evaluation of the radiological procedure, impacting the assessment of the technology, as well as justifications and optimizations. Therefore, it is crucial to compare each methods with the results obtained using a standard accurate approach.

This study measures noise magnitude values obtained in patients (*in vivo*) using three different methodologies and compares them with the measurements performed in an ensemble of virtually generated images, which is considered to be the best noise surrogate ethically obtainable. Such comparisons determine if a method can be used to represent tissue noise and can provide unique information paving the way towards the definition of adjustment or calibration factors that may be applied to different methodologies to closely represent clinical results.

2. **Materials and Methods**

This study included an XCAT-phantom (Figure 1 (10–13)) imaged virtually using a scanner-specific simulator (DukeSim (14)) modeling a commercial scanner geometry (Siemens Definition Flash). In particular, as described in other publications, XCAT phantoms are based on clinical CT images: major organs were segmented using a semi-automatic process and the remaining organs were added to the phantoms by deforming template models based on segmented organs from human body images. (10,11). DukeSim uses the computational phantom, together with scanner and protocol information, to estimate primary and scatter photons incident the virtual CT detector using ray-tracing and Monte Carlo techniques, respectively. Such projection data are finally processed and can be reconstructed using Filtered Back Projection (FBP) and Iterative Reconstruction (IR) Algorithms (14).

Following the standard definition (4), the noise magnitude for the virtually generated chest exams was calculated as the pixel value standard deviation in a circular ROI (area: 3 cm^2) manually placed in the lung in the central slice of each exam (*ROIn*). Noise magnitudes were also calculated in soft tissues (Global Noise Index – *GNI*), and air surrounding the patient (*AIRn*), applying [-300, 100] HU and HU < -900 thresholds, respectively (7) across all the 47 image datasets. Furthermore, for each pixel in the *GNI* threshold, the ensemble noise magnitudes in soft tissues (*En*) were calculated across images. Noise magnitude median values from different methods were compared in terms of percentage difference with correspondent *En* median values. The analysis was performed separately for the images reconstructed with Filtered Back Projection (FBP) and Iterative Reconstruction (IR) algorithms.

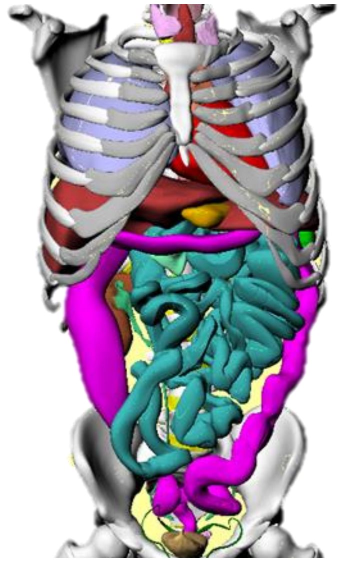

Figure 1. Example of an X-CAT phantom torso.

3. **Results**

Table 1 summarizes median *ROIn*, *GNI*, and *AIRn* and the related percentage differences with *En*. The IR images showed lower noise magnitude than the FBP images by 36% for *En*, 33% for *ROIn*, 37% for *GNI*, and 25% for *AIRn*. The noise calculated in the lung and in the soft tissues overestimate the ensemble noise with FBP and IR techniques. However, the difference between *GNI* and *En* in both FBP and IR is quite consistent. The noise calculated in the air surrounding the patient underestimates *En* and the underestimation in FBP reconstructed images is 4.5 times more than the underestimation in images reconstructed with IR.

Table 1. Median *ROIn*, *GNI*, and *AIRn* for Filtered Back Projection and Iterative Algorithm reconstructed images and percentage difference with ensemble noise magnitude in soft tissues (*En*).

	median *En*	**median *ROIn* (% difference with *En*)**	**median *GNI* (% difference with *En*)**	**median *AIRn* (% difference with *En*)**
FBP	30.6 HU	46.6 HU (+52%)	40.1 HU (+31%)	25.1 HU (-18%)
IR	19.5 HU	31.2 HU (+60%)	25.1 HU (+29%)	18.8 HU (-4%)

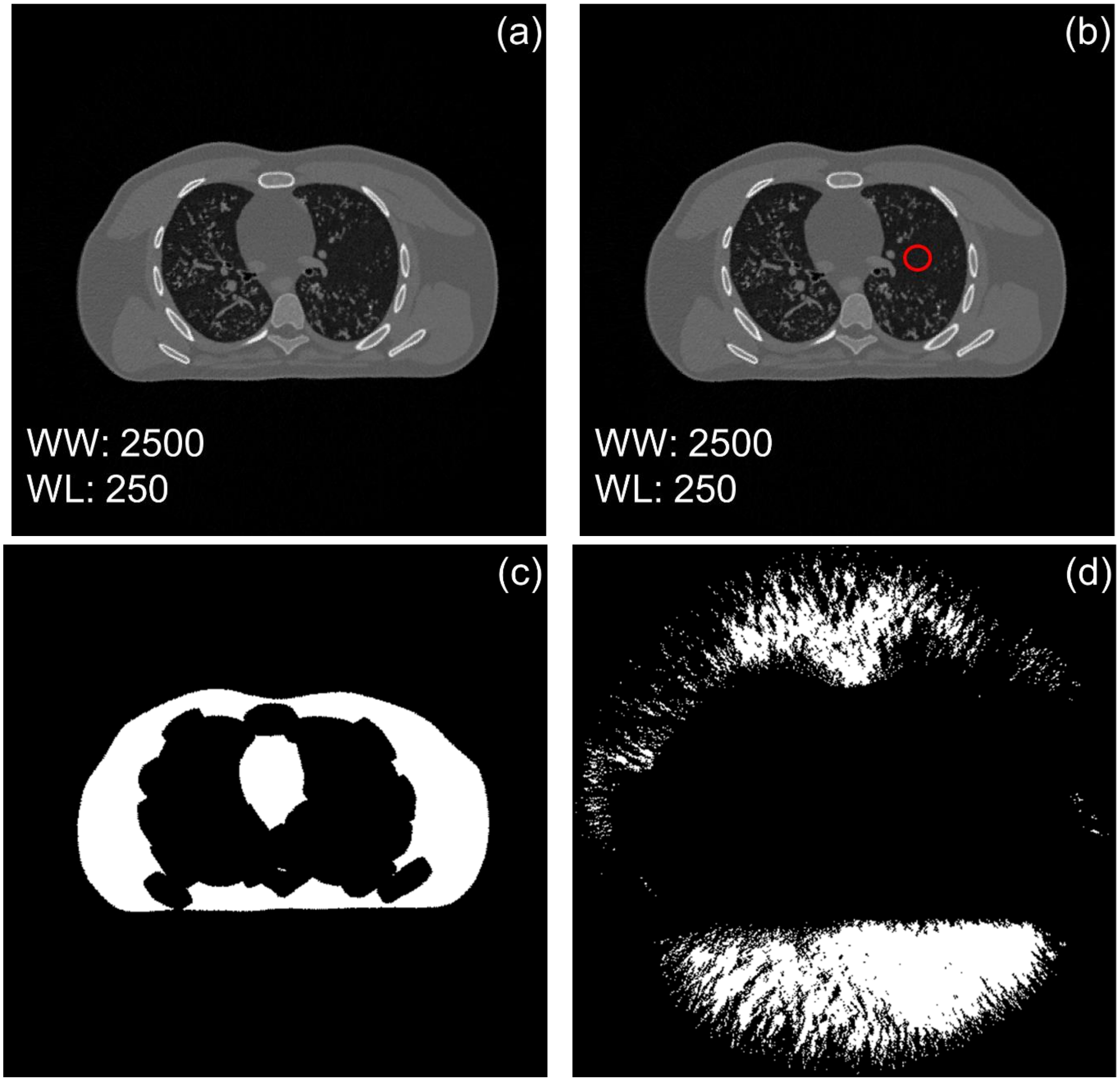

Figure 2. Example of one image considered in the study reconstructed with FBP (a); circular ROI manually placed in the lung (b); when GNI threshold is applied (c); when AIR threshold is applied (d).

4. **Discussion**

In vivo noise magnitude values obtained in patients using three different methodologies were compared with noise magnitude measured in an ensemble of virtually generated images, which is considered to be the best noise surrogate ethically obtainable. As expected, the noise magnitude in images reconstructed with Iterative Reconstruction algorithm is lower than the noise in Filtered Back Projection images. *ROIn*, and *GNI* overestimate the ensemble noise whereas *AIRn* underestimates it. Both *ROIn*, and *GNI* can represent spatial noise in tissues, however, when different reconstruction methods are considered, the overestimation is relatively constant only applying the Global Noise Index method. Such consistency enables the definition of adjustment or calibration factors that can be applied to better estimate tissue noise. Conversely, correction factors cannot be univocally defined when calculating noise magnitude in the air surrounding the patients because such method underestimate the ensemble noise in a reconstruction-dependent fashion.

Noise magnitude estimation inconsistencies can negatively affect the overall evaluation of the radiological procedure, impacting the assessment of the technology, the radiologists preference, as well as justification and optimization (3,4). Therefore, it is crucial to compare each method with the results obtained using a standard accurate approach as proposed in this study. In particular, when comparing the quality of the images obtained with different scanners, reconstruction methods, or protocols, becomes essential to provide also the method applied in the calculation of the noise magnitude. If such information is not available, it is not possible to draw any conclusion about the performance of a scanner or the actual impact of an implemented optimization action.

In radiology, the standard practical tool to promote and achieve optimization is Diagnostic Reference Level (DRL) (15). DRL quantities in CT are radiation dose related: volume Computed Tomography Dose Index ($CTDI_{vol}$), Dose Length Product (DLP), Size-Specific Dose Estimate (SSDE), etc. (16,17). However, the International Commission of Radiological Protection (ICRP) and the International Atomic Energy Agency (IAEA) recognized that radiation dose is only one of the steps in the overall process of optimization and that DRL should be expanded to include diagnostic image quality (15,18). Recently, few attempts were made to extend the concept of DRL to simultaneously include radiation dose and image quality information, particularly noise magnitude (19,20). It follows that effective optimization can only be attained by targeting accuracy in noise estimation and that inconsistencies from the ensemble noise method must be corrected applying validated correction or calibration factors.

The estimation of noise magnitude is also essential for the calculation of the so called detectability index (*d'*) that can be correlated with radiologist detection accuracy in a specific task (21). Smith et al. calculated *d'* in patients applying the Noise Power Spectrum calculated in a phantom and correcting it using noise magnitude as a scaling factor (21). In the same study, Smith et al. also showed that the *d'* is predictive of human-observer detection outcomes for cancer lesions in images also reconstructed with both FBP and IR. In this scenario, achieving consistent and accurate noise magnitude evaluation is the only way to inform detectability index and the effective assessment of the benefit connected with a radiological procedure, namely the likelihood of an accurate diagnosis.

This study was limited to only one anatomical district, one acquisition protocol, and one dose level. The presented methodology can be easily extended to include more anatomical regions, protocols, and dose levels. Such approach can provide unique information towards more consistent image quality evaluation and definition of effective calibration factors that can be applied to correct spatial noise (namely GNI) into the tissue noise represented by ensemble noise.

5. Conclusion

The gold standard in noise magnitude estimation is the ensemble noise obtained by scanning a patient multiple times and sampling each pixel noise within the ensemble of images, a repeated imaging process that can be pursue only with Virtual Imaging Trials (VITs) allowing to simulate clinical scenarios using computer-based simulations. The comparison of different noise magnitude calculation methods *in vivo* showed in this work, enables effective optimization and the improvement of consistency performance evaluation in CT imaging. To what extent spatial noise versus ensemble noise need to be the primary metric informing protocol optimization requires additional research.

References

1. Basic Safety Standards. GSR part 3. Int At Energy Agency, Vienna. 2014;3:471.
2. ICRP. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP Publication 103. Ann ICRP [Internet]. 2007 Apr 17;37(2–4):9–34. Available from: http://journals.sagepub.com/doi/10.1016/j.icrp.2007.10.003
3. Cheng Y, Abadi E, Smith TB, Ria F, Meyer M, Marin D, et al. Validation of algorithmic CT image quality metrics with preferences of radiologists. Med Phys [Internet]. 2019 Nov 20;46(11):4837–46. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13795
4. Samei E, Bakalyar D, Boedeker KL, Brady S, Fan J, Leng S, et al. Performance evaluation of computed tomography systems: Summary of AAPM Task Group 233. Med Phys [Internet]. 2019 Nov 11;46(11). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13763
5. Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys [Internet]. 2017 Jun;44(6):2173–84. Available from: http://doi.wiley.com/10.1002/mp.12240
6. Wilson JM, Christianson OI, Richard S, Samei E. A methodology for image quality evaluation of advanced CT systems. Med Phys [Internet]. 2013 Feb 20;40(3):031908. Available from: http://doi.wiley.com/10.1118/1.4791645
7. Christianson O, Winslow J, Frush DP, Samei E. Automated Technique to Measure Noise in Clinical CT Examinations. Am J Roentgenol [Internet]. 2015 Jul;205(1):W93–9. Available from: http://www.ajronline.org/doi/10.2214/AJR.14.13613
8. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, et al. Image quality in CT: From physical measurements to model observers. Phys Medica [Internet]. 2015 Dec 1;31(8):823–43. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1120179715003294
9. Ria F, Solomon JB, Wilson JM, Samei E. Technical Note: Validation of TG 233 phantom methodology to characterize noise and dose in patient CT data. Med Phys [Internet]. 2020 Apr 3;47(4):1633–9. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.14089
10. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. Med Phys [Internet]. 2010 Aug 24;37(9):4902–15. Available from: http://doi.wiley.com/10.1118/1.3480985
11. Segars WP, Bond J, Frush J, Hon S, Eckersley C, Williams CH, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. Med Phys [Internet]. 2013 Mar 15;40(4):043701. Available from: http://doi.wiley.com/10.1118/1.4794178
12. Samei E, Ria F, Tian X, Segars PW. A database of 40 patient-based computational models for benchmarking organ dose estimates in CT. Med Phys [Internet]. 2020 Jul 19;mp.14373. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.14373
13. Abadi E, Sturgeon GM, Agasthya G, Harrawood B, Hoeschen C, Kapadia A, et al. Airways, vasculature, and interstitial tissue: anatomically informed computational modeling of human lungs for virtual clinical trials. In: Medical Imaging 2017: Physics of Medical Imaging. 2017.
14. Abadi E, Harrawood B, Sharma S, Kapadia A, Segars WP, Samei E. DukeSim: A Realistic, Rapid, and Scanner-Specific Simulation Framework in Computed Tomography. IEEE Trans Med Imaging [Internet]. 2019 Jun;38(6):1457–65. Available from:

https://ieeexplore.ieee.org/document/8573883/

15. Vañó E, Miller DL, Martin CJ, Rehani MM, Kang K, Rosenstein M, et al. ICRP Publication 135: Diagnostic Reference Levels in Medical Imaging. Ann ICRP [Internet]. 2017 Oct 25;46(1):1–144. Available from: http://journals.sagepub.com/doi/10.1177/0146645317717209
16. Valentin J. Managing patient dose in multi-detector computed tomography(MDCT). ICRP Publication 102. Ann ICRP. 2007;37(1).
17. Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology [Internet]. 2017 Jul;284(1):120–33. Available from: http://pubs.rsna.org/doi/10.1148/radiol.2017161911
18. IAEA. Radiation Protection of Patients (RPOP) – Diagnostic Reference Levels (DRLs), International Atomic Energy Agency, 2017 [Internet]. Available from: https://www.iaea.org/resources/rpop/healthprofessionals/%0Aradiology/diagnostic-reference-levels
19. Ria F, Davis JT, Solomon JB, Wilson JM, Smith TB, Frush DP, et al. Expanding the concept of diagnostic reference levels to noise and dose reference levels in CT. Am J Roentgenol. 2019;213(4):889–94.
20. Hernandez AM, Shin DW, Abbey CK, Seibert JA, Akino N, Goto T, et al. Validation of synthesized normal-resolution image data generated from high-resolution acquisitions on a commercial CT scanner. Med Phys [Internet]. 2020 Oct 5;47(10):4775–85. Available from: https://onlinelibrary.wiley.com/doi/10.1002/mp.14395
21. Smith TB, Solomon J, Samei E. Estimating detectability index in vivo: development and validation of an automated methodology. J Med Imaging [Internet]. 2017 Dec 11;5(03):1. Available from: https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-5/issue-03/031403/Estimating-detectability-index-in-vivo--development-and-validation-of/10.1117/1.JMI.5.3.031403.full