

# Harmonization of PET image Reconstruction HyperParameters in Simultaneous PET/MRI

Richard Laforest (✉ [laforestr@mir.wustl.edu](mailto:laforestr@mir.wustl.edu))

Mallinckrodt Institute of Radiology, Washington University

Mehdi Khaligi

Stanford University

Yutaka Natsuaki

University of California San Francisco

Abhejit Rajagopal

University of California San Francisco

Dharshan Chandramohan

University of California San Francisco

Darrin Byrd

University of Washington

Hongyu An

Washington University in Saint Louis School of Medicine

Peder Larson

University of California San Francisco

Sara St. James

University of California in San Francisco

John J. Sunderland

University of Iowa

Paul E Kinahan

University of Washington

Thomas A. Hope

University of California San Francisco

---

## Original research

**Keywords:** PET/MR, Quantitative PET, Harmonization, Image Reconstruction

**Posted Date:** September 14th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-29815/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at EJNMMI Physics on November 5th, 2021.

See the published version at <https://doi.org/10.1186/s40658-021-00416-0>.

# Abstract

**Objective:** Simultaneous PET/MRIs vary in their quantitative PET performance due to inherent differences in the physical systems and differences in the image reconstruction implementation. This variability in quantitative accuracy confounds the ability to meaningfully combine and compare data across scanners. In this work, we define image reconstruction hyperparameters that lead to comparable contrast recovery curves across simultaneous PET/MRI systems.

**Method:** The NEMA NU-2 image quality phantom was imaged on one GE Signa and on one Siemens mMR PET/MRI scanner. The phantom was imaged at 9.8:1 contrast with standard spheres (diameter 10, 13, 17, 22, 28, 37mm) and with custom spheres (diameter: 8.5, 11.5, 15, 25, 32.5, 44 mm) using a standardized methodology. Analysis was performed on a 30 minute listmode data acquisition and on 6 realizations of 5 minutes from the listmode data. Images were reconstructed with the manufacturer provided iterative image reconstruction algorithms with and without point spread function (PSF) modeling. For both scanners, a post-reconstruction Gaussian filter of 3 to 7 mm in steps of 1 mm were applied. Attenuation correction was provided from a scaled Computed Tomography (CT) image of the phantom registered to the MR-based attenuation images and verified to align on the non-attenuated corrected PET images. For each of these image reconstruction parameter sets, contrast recovery coefficients (CRCs) were determined for the  $SUV_{mean}$ ,  $SUV_{max}$  and  $SUV_{peak}$  for each sphere. A hybrid metric combining the root mean squared discrepancy (RMSD) and the absolute CRC values were used to simultaneously optimize for best match in CRC between the two scanners while simultaneously weighting towards higher resolution reconstructions. The image reconstruction hyperparameter set were identified as the best candidate reconstruction for each vendor for harmonized PET image reconstruction.

**Results:** The range of clinically relevant image reconstruction hyperparameters demonstrated widely different quantitative performance across cameras. The best match of CRC curves were obtained at the lowest RMSD values with: for  $CRC_{mean}$ , 2 iterations -7mm filter with PSF on the GE Signa and 4 iterations -6mm filter on the Siemens mMR, for  $CRC_{max}$ , 2 iterations -7mm filter on the GE Signa, 4 iterations - 6mm filter on the Siemens mMR and for  $CRC_{peak}$ , 4 iterations-7mm filter with PSF on the GE Signa and 3 iterations-6mm filter on the Siemens mMR. Over all reconstructions, the RMSD between CRCs were 2.4%, 3.1% and 2.3% for CRC mean, max and peak, respectively. The solution of 2 iterations-3mm on the GE Signa and 4 iterations-3mm on Siemens mMR, both with PSF, led to simultaneous harmonization and with high CRC and low RMSD for CRC mean, max and peak with RMSD values of 3.4%, 5.5% and 3.0 %, respectively.

**Conclusions:** For two commercially-available PET/MRI scanners, user-selectable hyperparameters that control iterative updates, image smoothing, and PSF-modeling provide a range of contrast recovery curves that allow harmonization in harmonization strategies of optimal match in CRC or high CRC values. This work demonstrates that nearly identical CRC curves can be obtained on different commercially available scanners by selecting appropriate image reconstruction hyperparameters.

## Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

## Tables

Table 1. Compilation of representative scanner parameters.

	Siemens Biograph mMR	General Electric Signa
Crystal size	4x4x20 mm <sup>3</sup>	4x5.3x25mm <sup>3</sup>
Crystal material	LSO	LYSO
Light sensing device	Avalanche Photo-Diode	Si-PhotoMultiplier
Ring diameter	65.6 cm	66 cm
Field of View	59.4 cm x25.6 cm	60 cm x25 cm
PSF/TOF capable	Yes/No	Yes/Yes
Sensitivity	15 cps/kBq	21 cps/kBq
Peak NECR	183.5 kcps @ 23.1 kBq/mL	210 kcps @ 23.1 kBq/mL
Spatial Resolution at 1cm	4.3/4.3 mm	4.3/5.34 mm
Volumetric Spatial Resolution at 10cm	165 mm <sup>3</sup>	173 mm <sup>3</sup>

Table 2. Image reconstruction parameters leading to three optimization criteria for  $CRC_{mean}$ ,  $CRC_{max}$ , and  $CRC_{peak}$  for 30 minute data acquisition. All Siemens mMR reconstructions were done with 21 subsets, while all GE Signa reconstructions were done with 16 subsets and TOF (Time of Flight).

Metric		Optimized High CRC	Intermediate	Lowest RMSD
CRC-Mean	Siemens mMR	4it - 3mm + PSF	4it - 3mm	4it - 6mm
	GE Signa	2it - 3mm + IR	2it - 5mm + IR	2it - 7mm
RMSD		3.36	2.62	2.37
CRC- Max	Siemens mMR	4it - 3mm + PSF	4it - 5mm + PSF	4it - 6mm
	GE Signa	2it - 3mm + IR	2it - 6mm + IR	2it - 7mm
RMSD		5.46	3.81	3.13
CRC-Peak	Siemens mMR	4it - 3mm + PSF	4it - 3mm	3it - 6mm
	GE Signa	2it - 3mm + IR	4it - 5mm + IR	4it - 7mm
RMSD		3.02	2.97	2.32

Table 3. Image reconstruction parameters leading to three optimization criteria  $CRC_{mean}$ ,  $CRC_{max}$ , and  $CRC_{peak}$  for 5 minute data acquisition. All Siemens mMR reconstructions were done with 21 subsets, while all GE Signa reconstructions were done with 16 subsets and TOF (Time of Flight).

Metric		Optimized High CRC	Intermediate	Lowest RMSD
CRC-Mean	Siemens mMR	4it - 3mm + PSF	4it - 3mm	4it - 4mm
	GE Signa	2it - 4mm + IR	2it - 6mm IR	2it - 7mm + IR
RMSD		2.64	2.32	2.18
CRC- Max	Siemens mMR	4it - 4mm + PSF	4it - 5mm + PSF	4it - 6mm
	GE Signa	2it - 4mm + IR	2it - 6mm + IR	2it - 7mm
RMSD		5.92	3.77	2.66
CRC-Peak	Siemens mMR	4it - 3mm + PSF	4it - 3mm	4it - 7mm+PSF
	GE Signa	4it - 4mm + IR	4it - 6mm + IR	4it - 7mm
RMSD		2.97	2.99	2.84

## Introduction

Positron Emission Tomography (PET) allows for the measurement of absolute activity concentration of radiotracers in vivo with high sensitivity and high accuracy. Repeatable and reproducible measurements of tracer uptake in terms of tracer uptake, for example, standardized uptake value (SUV) are essential for monitoring and quantifying tumor response to therapy or progression of disease. In the context of multi-center clinical trials where data are pooled from different sites with different makes and model of scanners, vastly different quantitative performance characteristics may exist, limiting the ability to draw meaningful conclusions from these trials. This is the case whether it is a trial limited to PET/MRI scanners exclusively, or where PET/MRI data are included along with PET/CT data.

Variability of SUV measurements in clinical PET has been described before [1] [2] and has been associated to three main causes: technical (absolute scanner calibration, dose calibrator calibration, residual activity in the syringe after administration, clock synchronization), biological (patient preparation, inherent physiologic variability which is also tracer dependent, patient movement) or physical (acquisition and image reconstruction hyperparameters, method of defining regions of interest). Reproducible PET imaging involves an adequate quality control program of the scanner to ensure operation with accurate calibration to minimize bias but also with standardization in patient preparation and imaging to minimize biological variability. [3]. Physical variability can be minimized by ensuring that the acquisition parameters and image reconstruction hyperparameters be chosen to minimize the difference in SUVs across scanners from different sites or manufacturers. In particular, attention has been given to the choice of image reconstruction hyperparameters; sites cannot change the systems installed, but with careful choice of image reconstruction hyperparameters, differences between sites may be minimized. Quantitative PET is substantially affected by the choice of image reconstruction hyperparameters which may differ between institutions, and also by different scanner technology and differences in implementation of image reconstruction algorithms among vendors. Numerous efforts have been implemented to minimize the variability of clinical PET, mainly for  $^{18}\text{F}$ -FDG, such as RSNA-QIBA [4], SNMMI-CTN [5], EANM [6] [7] in the context of PET/CT. These efforts aimed at proposing specifications and requirements in the patient preparation, injection and imaging in order to provide comparability and consistency for quantitative FDG-PET across scanners in oncology.

The aim of harmonization in PET is different than standardization. Standardization implies that sites use a uniform procedure with the goal to minimize variations, while harmonization aims at achieving comparable results across manufacturers or sites even though slightly different procedures are used. Both harmonization and standardization reduce variations, but harmonization encompasses standardization which is stricter but in this work we chose to concentrate on harmonization strategies. Harmonization aims at achieving the same level of accuracy and thus minimizing variations, or determine limits in tolerable variation. Most commonly in PET/CT, this harmonization has been performed through the use of carefully tuned scanner model-specific post-reconstruction filtration [8]. Image generation for PET from PET/MRI is functionally the same as for PET/CT. The main differences are the scanner geometry which typically utilizes a smaller ring diameter and longer axial field of view as compared to PET/CT and the use of MR-based attenuation correction techniques.

Currently, in the US two manufacturers (Siemens and General Electric) offer PET/MRI scanners, although a third vendor is poised to enter the US market (United Imaging). These two currently available PET/MRI scanners differ in the choice in scintillation crystal (size and material), overall detector geometry, photomultiplication technology (avalanche photodiode vs silicon photomultiplier array) and in their image reconstruction algorithms. Although both vendors provide ordered-subset estimation-maximization algorithm (OSEM), the implementation of the algorithm varies between vendors and the performance of the algorithm is expected to differ due the use of different number of subsets, number of iterations, data compression prior to reconstruction, implementation of the system matrix, crystal size and axial plane thickness as well as from the different implementations of the point spread function modeling. The aim of this study was to determine sets of mutable and clinically relevant image reconstruction hyperparameters on both systems that yield the optimal match in quantitative performance as a function of object size for tumor like objects. The study was performed on one scanner from each vendor with controlled phantom experiment to minimize errors in phantom filling and focus on isolating effects on accuracy of measurements from the specifics of scanner hardware and image reconstruction algorithms.

## Methods

PET imaging was performed using the Siemens Biograph mMR (Siemens Healthineers, Erlangen, GE) and with the General Electric Signa PET/MRI (General Electric, Wisconsin, USA). Scanner characteristics are compared in Table 1 and their performance evaluation was previously reported [9, 10]. Phantom imaging was performed using the NEMA IEC Phantom [11] (Data Spectrum Corp., North Carolina, USA) using the standard set of spheres (diameters: 10, 13, 17, 22, 28 and 38 mm) and a custom set of spheres (diameters: 8.5, 11, 15, 25, 32.5 and 44 mm). The second set of spheres contains intermediate sphere sizes and provide more data points for measurement and optimization. A standardized filling procedure was implemented to achieve a 9.8:1 ratio between the spheres and the water background - The procedure followed EANM FDG PET PET/CT guidelines [12]. Two doses of approximately 20 MBq of  $^{18}\text{F}$  were used. The first dose was diluted in 1000 mL of water and the second dose was diluted into the approximately 9700 mL background water volume of the phantom chamber. The water volume of the phantom was determined by weight. The phantom was then centered in the PET field of view ensuring all six spheres are in the same imaging plane. On the Siemens scanner, this was ensured by placing the phantom on a foam cradle. PET data acquisitions were performed in listmode for 30 minutes. A two-point DIXON [13] (LavaFlex on GE Signa) attenuation scan was acquired but was only used to register a phantom CT attenuation template as described below.

Images were reconstructed using the vendor provided image reconstruction software with a range of hyperparameters encountered in the clinic for oncologic whole-body PET. Data from GE Signa scanner was reconstructed offline with the GE's *Duetto* PET reconstruction toolbox(v02.06) using 3D-OSEM (Ordered-Subset Expectation-Maximization) [14] algorithm with Time of Flight (TOF), at 2 and 4 iterations, 16 subsets, with and without point spread function (PSF) resolution modeling. On the Siemens mMR, images were reconstructed with *e7tools* (VE11P-SP2) using 3D-OSEM, 1 to 4 iterations, 21 subsets, with

and without PSF resolution modeling. On each system, the number of subsets are fixed by the manufacturers at 16 and 21 for clinical operations, respectively. The number of iterations on the GE Signa were selected from clinical use protocols for oncological PET. On the Siemens system, the number of iterations were selected to encompass the number of image updates (defined by the product of number of iterations times the number of subsets). As such GE Signa images were reconstructed at 32 and 64 image updates, and the images on Siemens mMR were reconstructed with a range of 21 to 84 updates.

Time of Flight (TOF) was not turned off on the GE Signa as non-TOF PET reconstruction is clinically not typically performed on systems allowing TOF. TOF benefits have been well documented to reduce image variance thereby reducing signal to noise, improving convergence rate and reducing artifacts [15] [16] and is typically used whenever available. Siemens mMR do not have the TOF option available. All reconstructions were repeated with a post-reconstruction Gaussian filter ranging from 3 to 7 mm. Images were reconstructed on a common voxel size of 2.34 mm in the transverse direction on a 256 matrix. GE Signa images were reconstructed on the native 2.78 mm slice thickness with standard axial filtering, while the Siemens mMR images were reconstructed on the native 2.027 mm slice thickness. Measured attenuation of the phantom by MRI methods such as DIXON or LavaFlex leads to inaccurate attenuation maps since the phantom material (water and plastic phantom wall) do not accurately mimic human tissue; water-like material will appear distorted, while plastics do not show at all [17]. The phantom also contained a 50 mm diameter cylindrical plastic insert filled with polystyrene and water to mimic the lungs. Consequently, standard tissue segmentation algorithm will fail when applied to the reconstructed image of the phantom. To avoid these issues, attenuation correction of the NEMA phantom is provided by the manufacturers using a template stored in the system. The CT-based template phantom attenuation map was used on the GE Signa and registered to the TOF-NAC PET images via rigid registration. For the mMR reconstructions, attenuation correction taken from a scaled CT image [18] [19] of the phantom aligned to the MR DIXON attenuation map and included attenuation of the patient bed as a 'hardware' attenuation map. The CT images of the phantom were aligned to MR using rigid registration (Euler transform) using the Elastix software [20, 21] with three level multi-resolution along with b-spline interpolation, advanced mutual information metric and the adaptive stochastic gradient descent optimizer. Accuracy of the registration of the template attenuation was visually inspected and further verified by inspection of the resulting mu-map over the non-attenuation corrected PET images.

In a first analysis, images were reconstructed using the entire 30 minute listmode data acquisition. In a second analysis, the 30 min listmode data was fragmented into 6 frames of approximately 5 minutes (277, 286, 295, 304, 314, and 324 seconds). The increasing frame duration ensured approximately equal number of collected events in each realization when accounting for radio-active decay. Contrast recovery coefficients (CRC) as defined by Liow and Strother [22] were calculated using the scanner measured sphere activity concentration, the scanner measured background water activity concentration and the sphere and background water activity concentration calculated from the assayed activities, dilution data and decay corrected to scan time.

$$CRC = \frac{(S/B)_{measured}^{-1}}{(S/B)_{theory}^{-1}} \quad (1)$$

All images were resampled to provide cubic voxels of approximately 1 mm<sup>3</sup>. Contrast recovery coefficients were computed using three methodologies: from the average sphere activity in a spherical volume of interest (VOI) drawn with diameter of the physical inner sphere diameter (CRC<sub>mean</sub>), from the maximum value in each sphere (CRC<sub>max</sub>) and from the peak value in each sphere defined as the average of a 1 cm<sup>3</sup> VOI with highest value (CRC<sub>peak</sub>) within the physical sphere. The definition of SUV<sub>peak</sub> was applied to define CRC<sub>peak</sub> [23]. As defined, the CRC<sub>peak</sub> may or may not include the hottest pixel with the sphere. Results of the two sets of spheres were combined to provide CRC curves against the 12 sphere sizes ranging from 8.5 to 44 mm. The background water activity was measured as the average over two 50 mm diameter circular regions of interest localized in five adjacent image planes of the phantom without spheres. These ROIs were used to define the image roughness which is a measure of the apparent noise [24]. The average coefficient of variation (COV) over those ten regions defines the image roughness (IR) and is calculated by

$$IR(\%) = 100 * \frac{1}{10} \sum_k \frac{STD_k}{Mean_k} \quad (2)$$

Where STD<sub>k</sub> is the standard deviation of the pixel intensity and the Mean<sub>k</sub> is the average in region k. The root mean squared discrepancy (RMSD) for all 800 image reconstruction hyperparameter combination pairs were then calculated.

$$RMSD = 100 * \sqrt{\frac{1}{12} \sum_{spheres} [CRC_{Signa}^i - CRC_{mMR}^j]^2} \quad (3)$$

Where *i,j* are the image reconstruction index from the GE Signa and Siemens mMR and the summation extends over all 12 spheres. The optimized PET imaged reconstruction hyper parameter set (#iterations, filter width, or use of PSF or not) was determined by selecting parameters that minimized the following hybrid metric:

$$F(\text{iterations}, \text{filter}, \text{PSF}/\text{IR}) = \arg \min \left\{ \text{RMSD} / a - \frac{\beta}{12} \sum \text{CRC}_{mMR}^i \text{CRC}_{\text{Signa}}^i \right\} \quad (4)$$

RMSD was determined for CRC<sub>mean</sub>, CRC<sub>max</sub> and CRC<sub>peak</sub> and the constant  $a$  is chosen to normalize RMSD so that

$\sum \text{CRC}_{mMR}^i \text{CRC}_{\text{Signa}}^i$  (referred to sum CRC product) to be on the same scale ( $a$  was set to 20 throughout) and was set to

the maximum value of RMSD observed in our analysis over the hyperparameter set. The sum of the product of the CRC coefficients will reach a maximum value for images with higher spatial resolution and thus with less smoothing. The

constant  $\beta$  was used to select solutions for either lowest RMSD ( $\beta=0$ ) or maximize CRC ( $\beta \sim 1$ ). An intermediate value of  $\beta$

was chosen to select solutions which present a compromise on high CRC values, but still with acceptably low RMSD.

Optimization was performed by systematically varying the number of iterations, level of filtration and toggling the use of point spread function.

The analysis was performed using the entire 30 minutes of data first, and in a second step, by averaging the results of the six independent realizations of approximately 5 minutes. Five minute of listmode data more closely simulates the statistics of a clinical scan oncology FDG. CRC curves agreement were generated for mean, max and peak CRC

## Results

On the mMR, the phantom preparation resulted in average activity concentrations of 1767 Bq/mL  $\pm$  5.0% and 17053 Bq/mL  $\pm$  6.0% for the background volume and the spheres, respectively, at imaging times. These activity concentrations correspond to an average ratio of 9.66  $\pm$  0.10. For the GE Signa, these values were 1622 Bq/mL  $\pm$  3.4% and 15526 Bq/mL  $\pm$  5.1%, corresponding to an average ratio of 9.57  $\pm$  0.22. The average water volume of the phantom as determined by weight was 9737  $\pm$  11 mL. Sorting the listmode data into six realizations of approximately 5 minutes resulted in an average of 47.3  $\pm$  0.18 million trues per realization for the GE Signa and 42.0  $\pm$  0.08 million trues per realization for the mMR.

The range of clinically relevant image reconstruction hyperparameters employed demonstrated widely different quantitative performance across the two manufacturers with regards to recovery of activity measurement as a function of object size. By varying both the number of iterations and post reconstruction filter level, bands of CRC curves were obtained that showed significant overlap between the two PET/MRI scanners. Contrast recovery coefficient curves for both systems using the full 30 minutes of acquisition data and four iterations and post-reconstruction filters of 3 and 7 mm are shown in Figure 1 (only these two filters are presented for clarity). CRC curves are presented for mean, max and peak hereafter referred to as  $\text{CRC}_{\text{Mean}}$ ,  $\text{CRC}_{\text{Max}}$ ,  $\text{CRC}_{\text{Peak}}$ . These plots depict the range of CRC values, and thus size-dependent SUV values, that are obtained in the clinical setting by varying the post-

reconstruction filter and using the resolution recovery algorithm option. In this figure, the effect of increasing post-reconstruction filtration (from 3 to 7mm) is illustrated; as expected, less filtration consistently led to higher CRC values. The largest effect of the resolution recovery algorithm is observed on the  $CRC_{max}$ , with smaller consequences associated with  $CRC_{mean}$  and  $CRC_{peak}$ . Of note, the  $CRC_{peak}$  for spheres with diameter less than 13 mm would typically not be defined as these spheres have a volume less than one  $cm^3$ . The  $CRC_{peak}$  for these smaller spheres are thus less than their  $CRC_{mean}$  as the  $VOI_{peak}$  includes background surrounding activity. These are nevertheless included for comparison and completeness of the present study.

Plots of the image roughness vs  $CRC_{peak}$  are presented in Figure 2 for each scanner for the 30 min and 5 min scans and for the NEMA phantom with the standard set of spheres. Only the 3mm filtration and  $CRC_{peak}$  is presented for brevity. Plots of  $CRC_{mean}$  and  $CRC_{max}$  show are similar (not shown). The filter is applied post-reconstruction and will only contribute to decrease the values of  $CRC_{peak}$  and image roughness (reduce noise). On the mMR, as the number of iteration increases from 1 to 4, the  $CRC_{peak}$  values increase and reach a maximum value while the COV (image roughness) increases. This is especially seen on the smallest spheres. For the largest spheres, the  $CRC_{peak}$  value is constant for all iteration number and is consistent with the fact that iterative reconstruction converges faster for larger objects. The PSF reconstructions, represented by the dashed lines, show higher CRC and lower image roughness (noise) at the same number of iterations. This is also consistent with PSF which improve resolution and decrease image noise. For the largest spheres, 'convergence' is approached as early as 2 iterations, but for the spheres of 17mm and smaller, the  $CRC_{peak}$  values tend to increase slowly indicating more iterations are required for convergence. The image roughness values on the GE Signa datasets are of the same magnitude as for the Siemens mMR, indicating comparable signal/noise characteristics of images from the Signa datasets as compared to the mMR. Slightly lower Image roughness are observed on the Signa possibly due to the use of TOF. The  $CRC_{peak}$  values at 2 and 4 iterations are approximately identical indicating close to convergence even for the smallest spheres. These observations are consistent with the use of TOF which is known to increase convergence rate and lower noise. The plots for the 5 minute reconstructions of image roughness vs  $CRC_{peak}$  are presented in the bottom row. The error bars are calculated from the standard deviation over the six noise realizations and correspond to the ensemble noise. The same observations can be made as for the 30min datasets and show that the signal and image noise is highly comparable between the experiments performed on each scanner and that convergence is approached similarly.

Figure 3 presents the distribution the RMSD vs sum CRC-product for all 600 image reconstruction pairs plotted for  $CRC_{mean}$ ,  $CRC_{max}$  and  $CRC_{peak}$  (top row). The reconstruction on the mMR with 1 iteration were omitted for clarity as they were shown to have not converged in Figure 2. The bottom row presents the same plots but with an expanded scale on RMSD with labels for specific reconstruction sets. On these graphs, each reconstruction pair is represented by a box whose color depicts the overall amount of filtration, ranging from dark blue (with minimal filtration – both 3mm on mMR and Signa) to red

(maximal filtration, both 7mm on mMR and Signa), with a gradual progression from blue to red for intermediate levels of filtration. In this respect, points on the right indicate reconstructions with low filtration and low agreement (high RMSD), points on the bottom left indicate solutions with high agreement (low RMSD) of CRC curves but low resolution (low sum CRC), and solutions on the upper left indicate highest resolution but still with high degree of agreement according to RMSD.

Solutions on the left are all achieved with low RMSD (good agreement of CRC curves) and varied level of resolution. This indicate multiple optimal solutions exist depending on the imaging task. Setting  $b$  to 0 leads to the solutions with lowest RMSD. The solutions with the lowest numerical value of RMSD is obtained at the highest filtration of 6 and 7mm, however these also correspond to solutions with low sum CRC product. Indeed, optimization by RMSD alone will tend to select solutions with low resolution and reflect more the influence of the post-reconstruction filter rather than the scanner performance. By setting the parameter  $b$  to 1.0, the solution that minimizes Equation 4 is represented by the green symbol. These plots show that a common optimized image reconstruction pairs exists at 4 iterations with PSF and 3mm filter for mMR and 2 iterations with IR and 3 mm filter on Signa. This solution correspond to a high CRC product and an acceptable RMSD value, and show that harmonization solution exists in which both high-resolution and excellent agreement in CRC curves can be achieved. The RMSD values using the  $CRC_{peak}$  are also, on average, smaller than the RMSD values of  $CRC_{mean}$  and  $CRC_{max}$ .

Figure 4 presents the CRC curves for the reconstruction hyperparameter set that provide the best match for  $CRC_{mean}$ ,  $CRC_{max}$ , and  $CRC_{peak}$  using the full 30 minutes of acquired data to provide either lowest RMSD (smoother images – right), (high resolution - left) and an intermediate solution (middle). The solutions with highest Sum CRC and low RMSD provides a simultaneous optimization of resolution and CRC agreement. The optimized solution for  $CRC_{max}$  show the highest level of noise and thus we provide an intermediate solution with slightly increased filtering as a compromise solution in which excellent agreement in CRC curves is achieved at less noise but still with a low RMSD. The corresponding image reconstruction hyperparameters pairs are reported in Table 2. Excellent agreement between the scanners are observed for the three scenarios, with only subtle differences. At best RMSD match, 100% of recovery ( $CRC=1.0$ ) is obtained generally for  $CRC_{max}$  or  $CRC_{peak}$  for spheres larger than 20 mm diameter and generally for lower CRC values. The RMSD values for three harmonization strategies are reported in Table 2 and Figure 4. It is interesting to note that lower RMSD values are observed for  $CRC_{peak}$ , than by  $CRC_{mean}$  and finally  $CRC_{max}$ , suggesting that closer ‘harmonization’ can be achieved using  $SUV_{peak}$  rather than by  $SUV_{max}$  or  $SUV_{mean}$ . In this figure, the green dash lines represents the EANM suggested limits in  $CRC_{max}$  values for qualification of PET/CT where the Recovery Coefficients (RC) values of [25, 26] were converted to CRC by using a known ratio of 10.

#### *Effect of scan time on harmonized hyperparameters*

Figure 5 show the harmonized CRC curves for mean, max, and peak values using only 5 minutes of listmode data. Average values were derived from the six noise realizations. Error bars on these plot

correspond to the ensemble noise on the CRC values over the 6 noise realizations. Five minutes of listmode data corresponds to a more clinically relevant scenario as it more closely mimics the data acquisition and the level of statistics encountered in oncology FDG PET/CT scans. Table 3 contains the harmonized image reconstruction hyperparameters obtained using 5 minutes of listmode data along the three optimization criteria. As in Figure 4, excellent agreement between the two scanners are found, and similarly harmonized image reconstruction hyperparameters can be determined. The RMSD values are approximately equal to the 30 minute acquisition.

CRC curves of images reconstructed using only 5 minutes of listmode data were compared, employing the image reconstruction hyperparameters that provided the best match from the 30 minute scans (data shown in Figure S1 in supplemental data). Very good agreement in  $CRC_{mean}$  was observed for 5 minutes acquisition using the 30 minutes harmonization hyperparameters as reflected in the CRC curves and in the value of RMSD. The best agreement for  $CRC_{max}$  was found when selecting the 30-minutes harmonized reconstruction hyperparameters for the intermediate and best RMSD optimization while the worst agreement is observed for best  $CRC_{max}$ . The mean and peak CRC curves show generally better agreement than the  $CRC_{max}$  curves. The largest differences were observed when looking at  $CRC_{max}$  for best 30-minutes harmonized reconstruction hyperparameters at the high CRC values which reflect the highest noise encountered in the 5min data. However, in all cases, the overall RMSD values are acceptably small.

## Discussion

In this study, we determined ‘harmonized’ image reconstruction hyperparameters for  $CRC_{mean}$ ,  $CRC_{max}$  and  $CRC_{peak}$  for the Siemens mMR and GE Signa PET/MRI systems. The experiments were performed in a controlled setting focus on variability caused by scanner hardware design and image reconstruction settings. This work excludes errors in measurements due to subjective manual regions of interest definition. CRC variability under clinically relevant range of image reconstruction hyperparameters (iterative updates, Gaussian filtration), algorithm implementation (3D OSEM, 3D OSEM plus resolution recovery) by each vendor were systematically varied. The imaging protocol was designed and executed to minimize the variability in the phantom preparation by using a rigorous phantom filling procedure, phantom alignment and imaging protocol.

The reconstructions were performed using an attenuation map template of the phantom. Since the effect of attenuation correction is decoupled from the choice of reconstruction hyperparameters, our work establishes image reconstruction hyperparameters across the two systems that will allow the study of the consequences in quantitation regarding the accuracy of the recovery coefficients of lesions due to the choice of attenuation correction strategy, as well as subtle implementation differences among the two vendors. Once the image reconstruction hyperparameters are harmonized from the PET data, effects such as choice of attenuation correction, positioning aid and others can be more accurately studied for a

given scanner and across scanner vendors. Ultimately, complete harmonization of simultaneous PET/MR scanners will need to include an attenuation correction as measured from the scanner.

This study aimed at identifying harmonized image reconstruction hyperparameters for the two most widely used simultaneous PET/MRI scanners with a multi-fill well-controlled experiment and differ from a multi-site phantom study. Measured variability of quantitative performance between sites using the same make and model scanners comes from two main sources. The first, and most major, is variability of phantom fill. We minimized this variability through performance of rigorous filling procedure, identical at each site. In this study, we minimized the variability by used long scans, used identical fill activities, all activities were measured in dose calibrators calibrated to a NIST traceable 511keV source, weights were used to access phantom fill volumes.

The second source of error is associated with fundamental intrinsic quantitative performance differences between studies performed on two physically different, but same make and model scanner. These differences, as manifested in the scanner model specific performance CRC curve on an appropriately calibrated and tuned scanner are quite small. In fact, precise CRC performance using the NEMA IQ phantom (the same as used in our studies) is used by vendors as acceptance criteria for scanner installations. This variability is small compared to other sources of error, most significantly fill accuracy and precision.

Remarkably similar quantitative performances were achieved through mutual tuning of reconstruction hyperparameters for both the 30 minute low-noise use-case, and the more clinically relevant 5 minute acquisitions. The clinical implication is that if patients are imaged under technically and biologically controlled conditions, but on different PET/MRI systems, prospectively-used harmonized reconstruction parameter sets will result in nearly identical quantitative measurements independent of the system used. This conclusion is independent of lesion size. This aspect has important consequences to multi-center clinical trials where data will be aggregated from different models of PET/MR systems.

It should be noted that the 'harmonized' image reconstruction hyperparameters are not necessarily those that would yield to the highest CRC values across all spheres. Indeed, we have identified that a harmonization approach relying solely on the lowest RMSD values leads to solutions of high-level of filtering and therefore will correspond to very smooth images. This solution will be detrimental for imaging task of lesion detection albeit providing the good agreement in CRC values. This solution emphasize more the effect of filtering rather than the performance of the scanner, which is appropriate for some clinical and clinical trial applications, but certainly not all. A solution with high CRC values (but still with acceptably low RMSD) would be typically obtained at larger number of iterations and minimal filtering (as shown) but images would be subject to higher noise levels. In particular, the  $CRC_{max}$  reaches values significantly higher than 1.0 for image reconstruction with 3mm post-reconstruction filters and using resolution recovery. We have shown that harmonization solutions exist for which, depending on the imaging task, being either lesion detection (high CRC) or quantitation accuracy across sites minimizing RMSD, excellent agreement in CRC values can be achieved between these two scanners.

The phantom was prepared under conditions mimicking conditions typically encountered in clinical practice with  $^{18}\text{F}$ -FDG in PET/CT. Imaging protocols suggests imaging at 60 minutes post-injection of 370-740 MBq (10-20 mCi)  $^{18}\text{F}$ -FDG from head to mid-thigh in a series of slightly overlapping bed positions, each with axial field of view of 20-25 cm. Although, substantial variability exists in clinical PET/CT, typical acquisition times are of the order of 2 – 4 minutes per bed position. Therefore, assuming uniform distribution in an average sized human, a typical injection yields to approximately 5000 Bq/mL (e.g. for 555 MBq (15mCi) injection administer and imaged 60 minute post-injection in a 75 kg patient). In this work, the phantom was prepared with a nominal background activity concentration of  $\sim$ 1600-1800 Bq/mL, and thus the 5min scan would yield similar count statistics to a clinical  $^{18}\text{F}$ -FDG acquisition of 2 min per bed position with the 30-min study resulting in 6 times the counts as a typical clinical study. The 30 minutes acquisition data are used to determine the optimal harmonized hyperparameters in images with minimal noise and thus be able to determine image reconstruction hyperparameters that yield most comparable CRC coefficients free from limitations due to statistical noise.

The average activity concentration at imaging time was less in the experiments performed on the GE Signa by approximately 8%. However, this scanner benefits from a higher sensitivity (21 vs 15 cps/kBq) relative to the Siemens mMR and thus when accounting for the relative scanner sensitivity, more counts were acquired on GE Signa ( $\sim$ 12% more). In addition, the GE Signa employs Time of Flight (TOF), while the Siemens mMR does not. The main advantage of TOF is faster convergence and higher signal to noise. This may explain, at least in part, why the best matching CRC curves are obtained with 2 iterations/16 subsets on the GE Signa scanner as opposed to 4 iterations/21 subsets on the Siemens mMR.

$\text{CRC}_{\text{mean}}$  and  $\text{CRC}_{\text{peak}}$  appear to be more robust metrics used as the basis for harmonization when comparing quantitative results from PET/MRI scanners than  $\text{CRC}_{\text{max}}$  (and thus  $\text{SUV}_{\text{max}}$ ), as is expected. This is likely an effect of statistical noise even for 30 minutes data sets, and this effect is greater for 5 minute acquisition times. The image noise (image roughness) depends on a variety of factors in the image reconstruction chain (number of iterations and post-reconstruction filter), and includes the choice of algorithm, use of TOF, and especially the use of resolution recovery. As such, image noise cannot be rigorously compared. However, the reconstructed noise was determined by the image roughness in the reconstructed images and for these matched experiments, (identical fill and imaging in similar conditions), comparable signal and noise were achieved between the two cameras. In phantom studies,  $\text{SUV}_{\text{mean}}$  is highly robust since the lesion volume is known and the activity distribution within the lesion is uniform. This is not the case in patient studies and extreme variability is observed in segmentation volume making it of little clinical use, currently. So lesion  $\text{SUV}_{\text{mean}}$  is not recommended within the context of clinical trial response assessment.  $\text{SUV}_{\text{max}}$  is most typically used. Inter-reader measurement variability of  $\text{SUV}_{\text{max}}$  is small, and it is a robust measurement, although impacted significantly by image noise.  $\text{SUV}_{\text{peak}}$  has slowly been gaining acceptance as a more robust (less sensitive to noise) metric of response, although literature support for its use is less prevalent, currently. Similarly, our data indicate that  $\text{SUV}_{\text{peak}}$  is likely the most repeatable measure among the three and that  $\text{SUV}_{\text{max}}$  being more affected

by noise.  $SUV_{peak}$  will generate higher SUV values than  $SUV_{mean}$ , however it can only be defined for lesion larger than 1cm. In studies where quantitative harmonization is a critical aspect to the trial's response assessment, then tighter harmonization appears to be achievable when using the  $SUV_{peak}$  metric. The  $SUV_{peak}$  metric seems more independent on the choice of region of interest and the level of smoothing thus making this metric more amenable to harmonization.

A limitation of this study is that the experiments were performed in phantoms at a set count density in the spheres and background activity. Extrapolation to human imaging is not directly translatable as OSEM algorithm are not linear and performance will be dependent on the specific patient activity distribution and count density. This limitation is common to all studies in phantoms. An alternative approach could be to insert synthetic lesions of varied activity (SUV), size and shapes in clinical patient datasets. This is currently an active area of research that we are investigating.

**Our data indicates that the harmonized images reconstruction hyperparameters proposed here (both in the ideal long scanning acquisition and in clinical conditions) for PET/MR scanners can be achieved and that comparable size-dependent recovery coefficients, or size-dependent tumor SUV values can be obtained and well within** the limits proposed by EARL-EANM. PET data acquired on PET/MR scanners would thus be acceptable to be included in multi-center clinical trials, at least as defined by the EARL-EANM criteria. However, this study goes beyond EARL-EANM as it determines image reconstruction hyperparameters that provide practically identical CRC curves between these two scanners and thereby show that variability of in small lesions quantitation can be largely eliminated by controlling the image reconstruction hyperparameters. This conclusion is important as it will allow to further study other factors affecting quantitative PET in PET/MRI such as the specific choice of attenuation correction technique (including the level at which the bones are included), patient positioning aids and others.

## Conclusion

Quantitative PET is influenced by a variety of technical, biological and physical factors. This work demonstrates that harmonization of reconstruction hyperparameters in PET in simultaneous PET/MR is possible and can yield images with nearly identical quantitative performance in terms of CRC measurements over a range of lesion sizes. For the two commercially-available PET/MRI scanners evaluated, user-selectable hyperparameters that control iterative updates, image smoothing, and PSF-modeling provide a range of contrast recovery curves that allow harmonization. This work demonstrates that essentially identical CRC curves can be obtained on two commercially available scanners by a proper choice of image reconstruction hyperparameters. This work will form the basis of further study on the quantitative performance related to the choice of attenuation correction strategy.

## Declarations

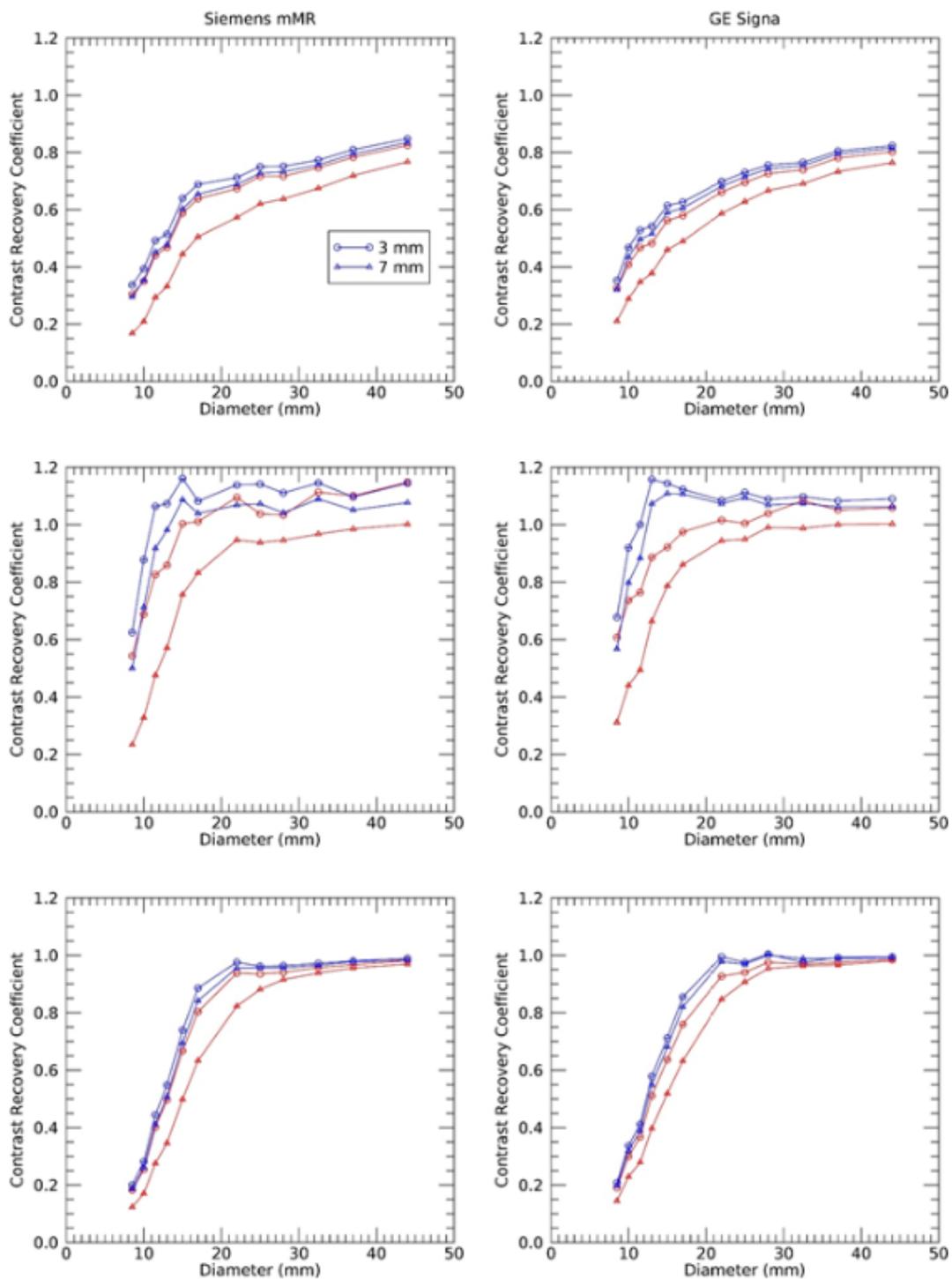
## Acknowledgments

This work was performed under the NIH grant QUALIFICATION AND HARMONIZATION OF PET/MRI FOR CANCER CLINICAL TRIALS, R01 CA202148, PIs: T.Hope/R.Laforest. Siemens mMR experiments were performed at the Center for Clinical Imaging Research at Washington University and F-18 activity was produced by Mallinckrodt Institute of Radiology Cyclotron Facility. GE SIGNA experiments were performed at the University of California in San Francisco. We wish to thank the personnel of those facilities for their support.

## **Disclosures**

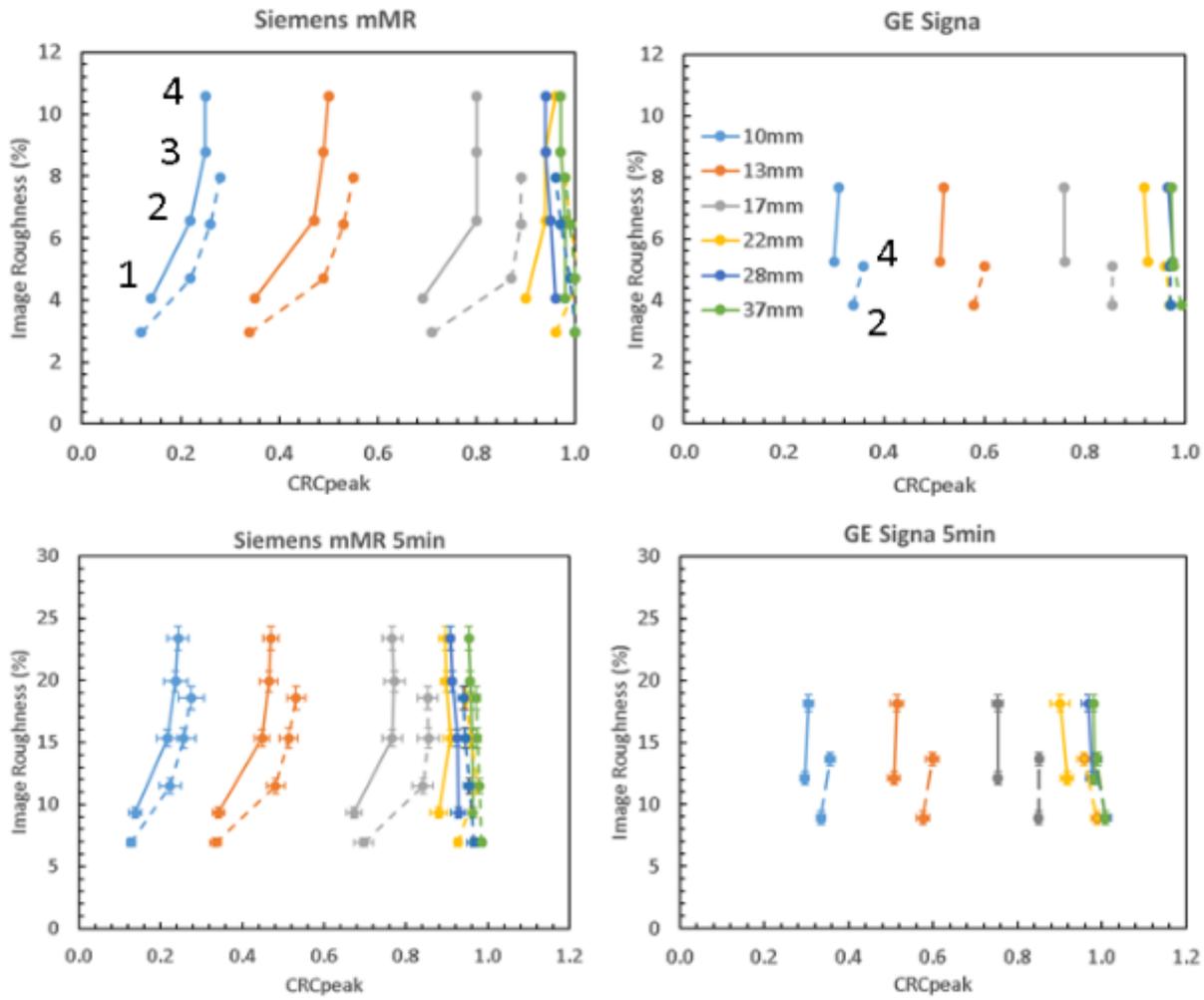
The authors have nothing to disclose.

## **Figures**



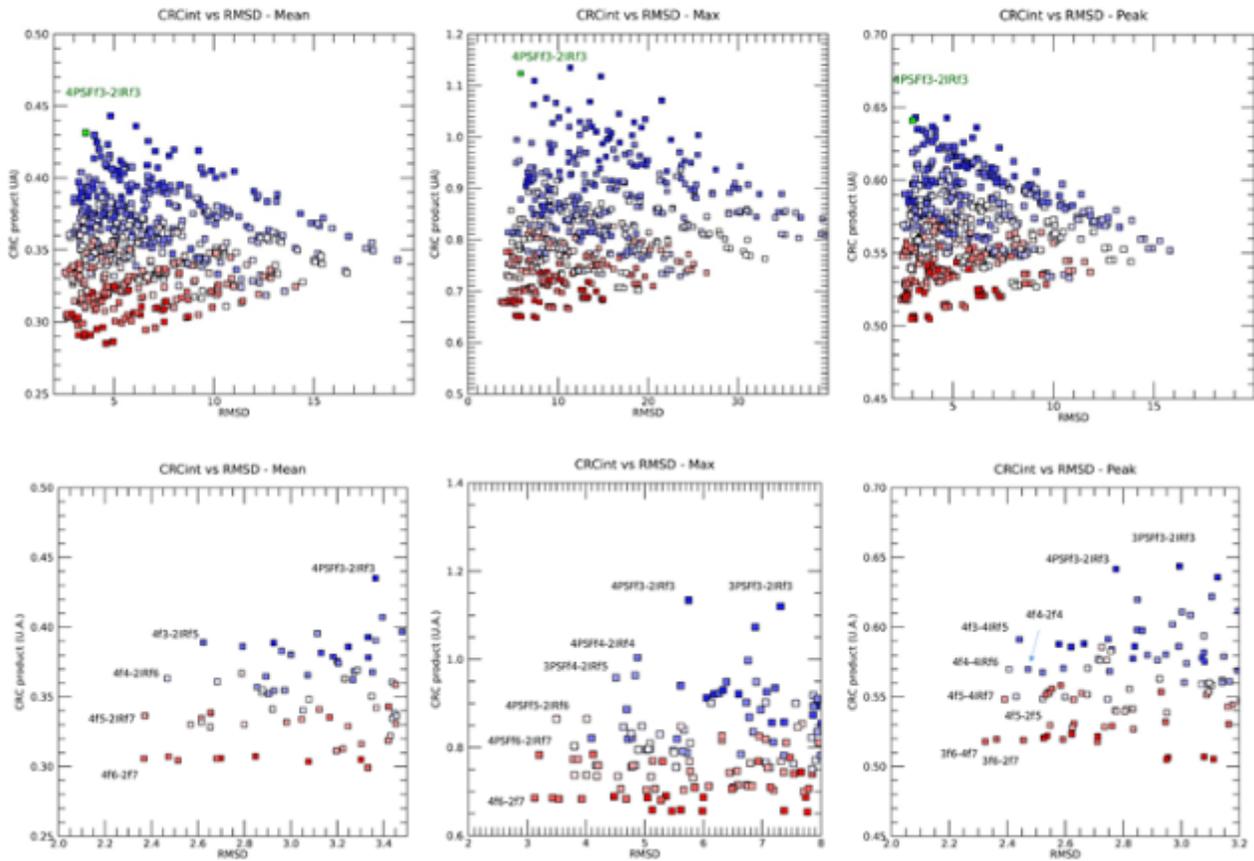
**Figure 1**

Contrast Recovery Coefficient (CRC) curves for mean (top), maximum (middle), and peak (bottom), without resolution recovery (red), and with resolution recovery (blue) for the GE Signa (right) and Siemens mMR (left) scanners. CRC curves are presented for two levels of filtration where the least filtration (3mm) and highest filtration (7mm) are shown.



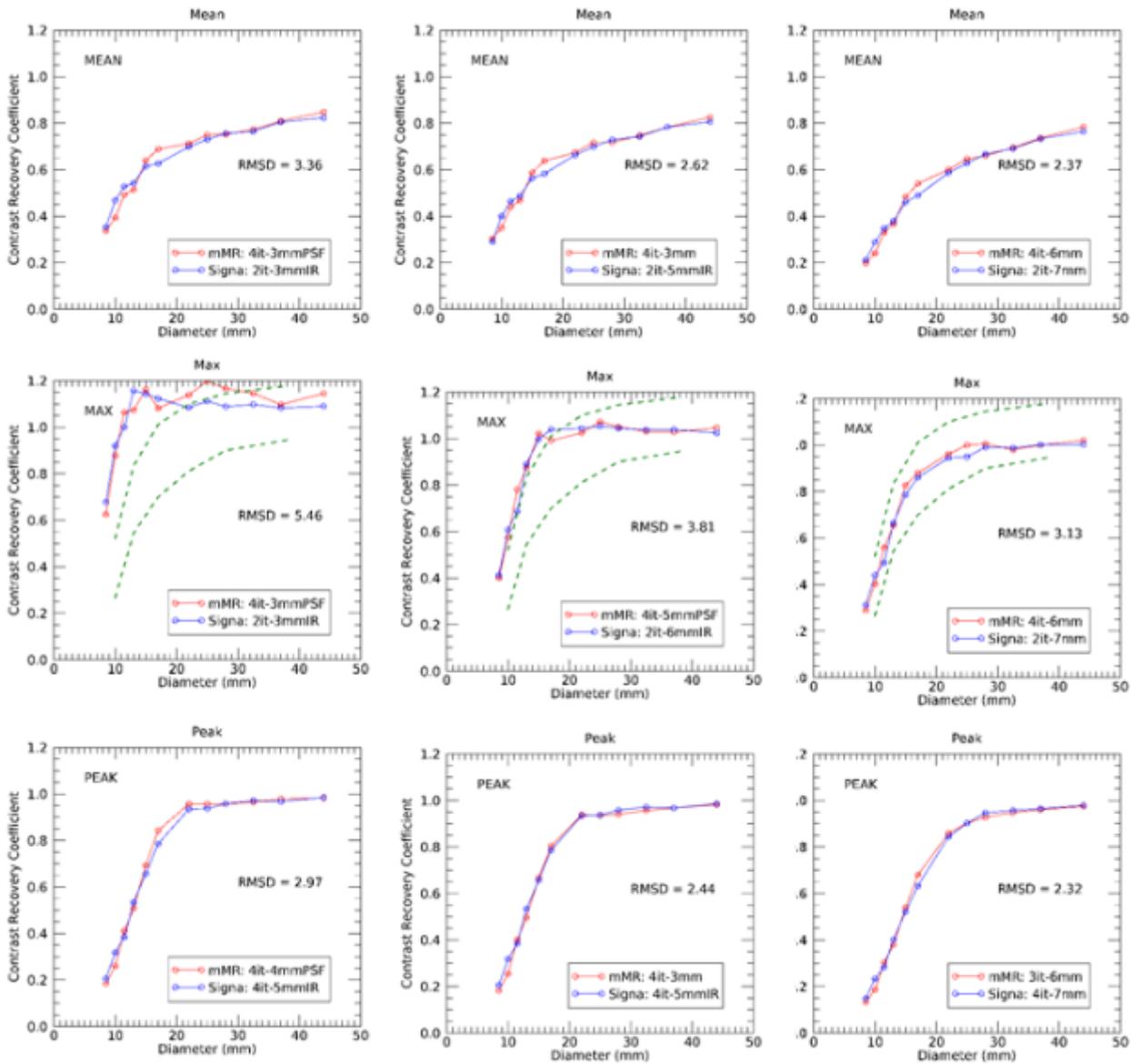
**Figure 2**

Image roughness vs CRCpeak plots for the NEMA standard set of spheres for the Siemens mMR (left) and GE Signa (right). The top row shows plots for the whole 30min of listmode data and the bottom row, the average value over six frames of 5 minutes duration. Error bars in the bottom plot show the ensemble noise over the six realizations.



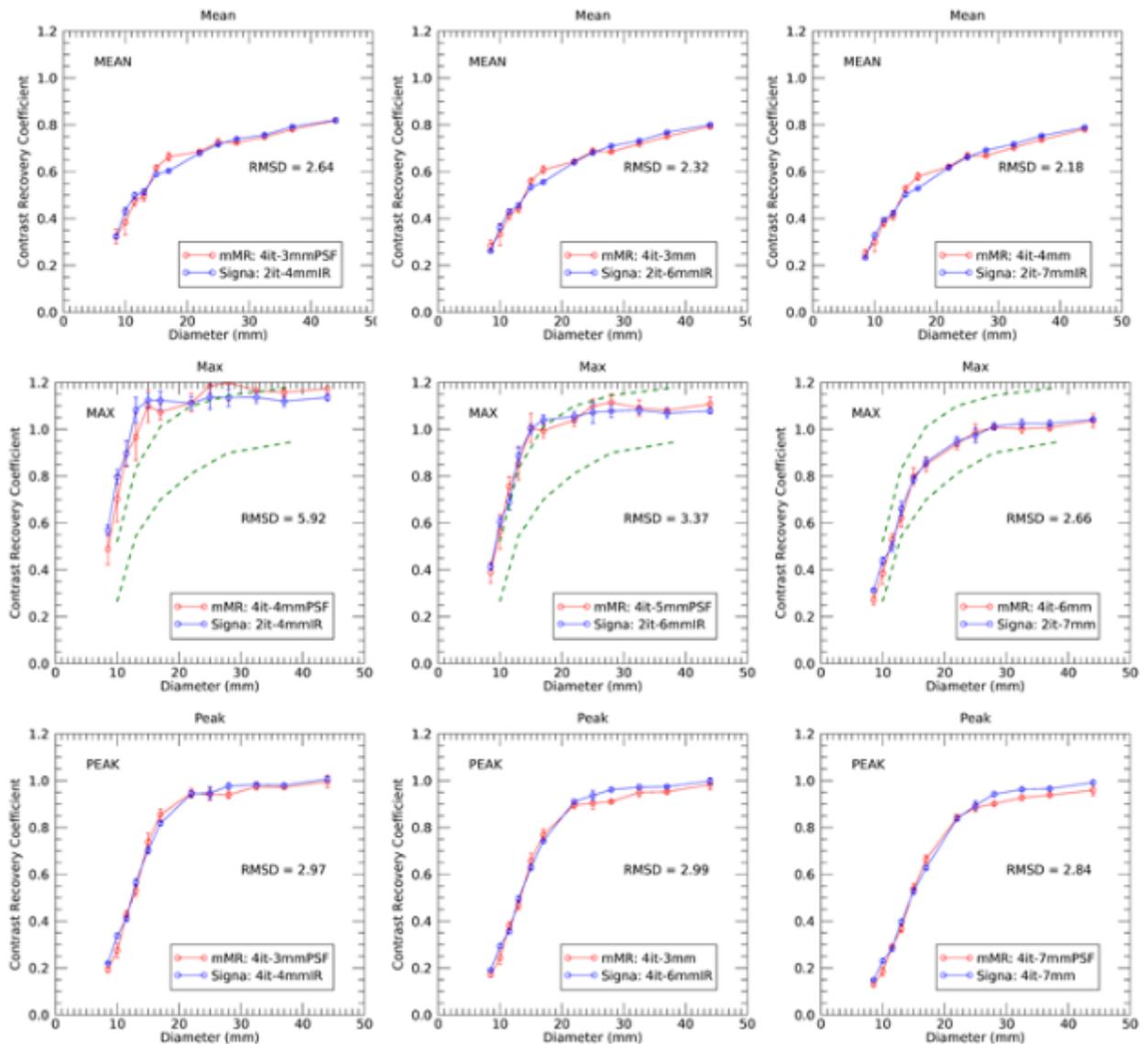
**Figure 3**

CRCmean,max,peak–product vs RMSD box plot. Each of the 600 pairs of image reconstructions hyperparameters combinations is represented by a box (top row). The color scale from blue to red represents the different level of overall filtration. Dark blue boxes represents combinations with least overall filtration (i.e. 3mm for both mMR and Signa images) and dark red correspond to most filtration (7mm on mMR and Signa images). Bottom row, expanded view at low RMSD to identify candidate image reconstruction parameters for optimization. The solution for best RMSD and CRC is indicated in green.



**Figure 4**

Matching CRC curves for mean (top), max (middle), and peak values (bottom) for 30 minutes of data corresponding to three optimization scenario: right) optimized for best resolution or highest CRC values, middle) optimized for both resolution and RMSD, and left) for best RMSD . The corresponding RMSD values are indicated in each panel. Results from the Siemens Biograph mMR are in red, and from the GE Signa are in blue. The dash lines represent the EANM limits on CRC max.



**Figure 5**

Matching CRC curves for mean (top), max (middle), and peak values (bottom) for 5 minutes of data corresponding to three optimization scenario: right) optimized for best resolution or highest CRC values, middle) optimized for both resolution and RMSD, and left) for best RMSD . The corresponding RMSD values are indicated in each panel. Results from the Siemens Biograph mMR are in red, and from the GE Signa are in blue. The dash lines represent the EANM limits on CRC max.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig1.png](#)

- [PETMRHamonizationsupplementaldataAugust20.docx](#)