

**TOWARDS FURTHER OPTIMIZATION OF
RECONSTRUCTION METHODS FOR DUAL-
RADIONUCLIDE MYOCARDIAL PERFUSION SPECT**

by

Xin Li

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the
degree of Doctor of Philosophy

Baltimore, Maryland

October, 2016

ABSTARCT

Coronary artery disease (CAD) is the most prevalent type of heart disease and a leading cause of death both in the United States and worldwide. Myocardial perfusion SPECT (MPS) is a well-established and widely-used non-invasive imaging technique to diagnose CAD. MPS images the distribution of radioactive perfusion agent in the myocardium to assess the myocardial perfusion status at rest and stress state and allow diagnosis of CAD and allow differentiation of CAD and previous myocardial infarctions.

The overall goal of this dissertation was to optimize the image reconstruction methods for MPS by patient-specific optimization of two advanced iterative reconstruction methods based on simulations of realistic patients population modeling existing hardware and previously optimized dual-isotope simultaneous-acquisition imaging protocols. After optimization, the two algorithms were compared to determine the optimal reconstruction methods for MPS.

First, we developed a model observer strategy to evaluate image quality and allow optimization of the reconstruction methods using a population of phantoms modeling the variability seen in human populations. The Hotelling Observer (HO) is widely used to evaluate image quality, often in conjunction with anthropomorphic channels to model human observer performance. However, applying the HO to non- multivariate-normally (MVN) distributed, such as the output from a channel model applied to images with variable signals and background, is not optimal. In this work, we proposed a novel model observer strategy to evaluate the image quality of such data. First, the entire data ensemble is divided into sub-ensembles that are exactly or approximately MVN and homoscedastic. Next, the Linear Discriminant (LD) is applied to estimate test statistics for each sub-ensemble, and a single area under the receiver operating characteristics curve (AUC) is calculated using the pooled test statistics from all the sub-ensembles. The AUC serves as the figure of merit for performance on the defect detection task.

The proposed multi-template LD was compared to other model observer strategies and was shown to be a practical, theoretically justified, and produced higher AUC values for non-MVN data such as that arising from the clinically-realistic SKS task used in the remainder of this work.

We then optimized two regularized statistical reconstruction algorithms. One is the widely used post-filtered ordered subsets-expectation maximization (OS-EM) algorithm. The other is a maximum a *posteriori* (MAP) algorithm with dual-tracer prior (DTMAP) that was proposed for dual-isotope MPS study and was expected to outperform the post-filtered OS-EM algorithm. Of importance, we proposed to investigate patient-specific optimization of the reconstruction parameters. To accomplish this, the phantom population was divided into three anatomy groups based on metrics that expected to affect image noise and resolution and thus the optimal reconstruction parameters. In particular, these metrics were the distance from the center of the heart to the face of the collimator, which is directly related to image resolution, heart size, and counts from the myocardium, which is expected to determine image noise. Reconstruction parameters were optimized for each of these groups using the proposed model observer strategy. Parameters for the rest and stress images were optimized separately, and the parameters that achieve the highest AUC were deemed optimal. The results showed that the proposed group-wise optimization method offered slightly better task performance than using a single set of parameters for all the phantoms. For DTMAP, we also applied the group-wise optimization approach. The extra challenges for DTMAP optimization are that it has three parameters to be optimized simultaneously, and it is substantially more computationally expensive than OS-EM. Thus, we adopted optimization strategies to reduce the size of the parameter search space. In particular, we searched in two parameter ranges expected to give result in good image quality. We also reduced the computation burden by exploiting limiting behavior of the penalty function to reduce the number of parameters that need to be optimized. Despite this effort, the optimized DTMAP had poorer task performance compared to the optimized OS-EM algorithm. As a result, we studied the

limitations of the DTMAP algorithm and suggest reasons of its worse performance for the task investigated.

The results of this study indicate that there is benefit from patient-specific optimization. The methods and optimal patient-specific parameters may be applicable to clinical MPS studies. In addition, the model observer strategy and the group-wise optimization approach may also be applicable both to future work in MPS and to other relevant fields.

Advisor: Eric C. Frey, Ph.D.

Reader: Yong Du, Ph.D.

ACKNOWLEDGEMENTS

I owe thanks to a lot of people that helped me in many ways during my graduate studies at Johns Hopkins University. Without them, I would not have completed this important goal in my life.

First and foremost, I would like to thank my advisor Dr. Eric C. Frey for his guidance both in research and life all through the years. He is a wonderful mentor, always supportive, inspiring, patient, and thoughtful. I am also thankful for his support, encouragement and prayer for me during my difficult times. Without him, completing my degree would have been impossible.

I would also like to thank Dr. Jonathan M. Links for his insightful and inspiring comments through emails and during meetings. I really appreciate his effort to schedule lengthy meetings with me in spite of his busy schedules. I learned a lot from his unique ways to looking at problems.

I would also like to thank Dr. Benjamin M. W. Tsui for his mentoring during my first two and half years and his effort for making our division a wonderful place to learn and thrive.

I am lucky to have Dr. Abhinav K. Jha in my research project. He spent a lot of time with me and offered a lot of important and helpful suggestions for my work. Despite his busy schedules, he was always patient to answer my naive questions. I learned a lot from him and will always appreciate his guidance and support.

I would also like to thank Dr. Michael Ghaly for his generous help during my research. I learned a lot from him about this myocardial perfusion SPECT project. He is also a great friend in life: warm-hearted, thoughtful, supportive, and informative. I will always cherish his friendship.

I would also fortunate to have met Dr. Xiaolan Wang as my lab-mate and friend. She never hesitates to give me candid suggestions and she is very thoughtful and warm-hearted.

I would also thank Dr. Yong Du, Dr. Jingyan Xu, Dr. Xing Rong, Dr. George S. K. Fung,

Dr. Si Chen, Dr. Katsuyuki Taguchi, Dr. Andrew Rittenbach, Dr. Qiulin Tang, Dr. Fatma E. Elshahaby, and Ye Li for their support and helpful discussions. I learned a lot from each of them.

I would also thank Martin Stumpf for his hard work maintaining our cluster environment and help me with programming questions.

I am also grateful to LaVahn Tunstall and Michele Ingram for her hard administrative work.

I give many thanks to all other members of the Division of Medical Imaging Physics for their friendship and support.

My sincere appreciation also goes to my committee members, Dr. Eric C. Frey, Dr. Yong Du, Dr. Arman Rahmim, Dr. Jin U. Kang, and Dr. Gerard G.L. Meyer for their time and suggestions for my dissertation.

Last but not least, I would like to thank my family, my parents and my in-laws, my husband Xinkai, and our children Jonathan and Peter. They are the source of my strength. My love to them is ineffable.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vii
LIST of TABLES	x
LIST of FIGURES	xi
1. Introduction	1
1.1 Problem domain and contributions	1
1.2 Organization	2
2. Background	4
2.1 Coronary artery disease and myocardial perfusion	4
2.2 Myocardial perfusion SPECT	5
2.2.1 SPECT system hardware	6
2.2.2 Radiopharmaceuticals for myocardial perfusion SPECT	7
2.2.3 MPS acquisition protocols	7
2.3 Image reconstruction	8
2.3.1 Image degrading factors	8
2.3.2 Reconstruction methods	11
2.4 Task-based image quality evaluation	14
2.4.1 Tasks	15
2.4.2 Observers	17
2.4.3 Figures-of-merit	22
2.5 Patient-specific optimization	25
3. Common Methods: phantom population and projection data simulation	27
3.1 Introduction	27
3.2 Methods	28
3.2.1 Population of realistic digital phantoms for MPS study	28
3.2.2 Realistic projection simulation with optimized imaging protocols	30
3.3 Summary	30
4. A sub-ensemble-based approach to evaluate detection task performance for non-multivariate-normally-distributed data using multi-template linear observer strategies	32
4.1 Introduction	32

4.2	<i>Theory</i>	35
4.2.1	Sub-ensemble-based approach	35
4.2.2	Multi-template linear observer strategy 1	36
4.2.3	Multi-template linear observer strategy 2	37
4.3	<i>Method</i>	42
4.3.1	Image reconstruction and post-processing	42
4.3.2	Implementation and evaluation of observer strategies	42
4.4	<i>Results</i>	47
4.4.1	Effect of using the LD observer instead of the HO observer for multiple sub-ensembles obtained using the partitioning method in 4.3.2	47
4.4.2	Comparison of the AUCs using the three observer strategies	49
4.4.3	Comparison of the optimal parameters obtained by the three observer strategies	51
4.5	<i>Discussion</i>	55
4.6	<i>Conclusion</i>	59
4.7	<i>Appendix</i>	60
4.7.1	Appendix A	60
4.7.2	Appendix B	61
4.7.3	Appendix C	64
4.7.4	Appendix D	65
4.7.5	Appendix E	67
5.	Toward patient-specific optimization of post-filtered OS-EM reconstruction for simultaneous-acquisition dual-isotope myocardial perfusion SPECT	69
5.1	<i>Introduction</i>	69
5.2	<i>Methods</i>	70
5.2.1	Reconstruction and post-processing	70
5.2.2	Image quality evaluation	70
5.2.3	Patient-specific optimization of reconstruction parameters	72
5.3	<i>Results</i>	75
5.3.1	Optimal parameters	75
5.3.2	Two-class task performance using parameters obtained from the four optimization approaches	78
5.3.3	Three-class task performance using parameters obtained from the four optimization approaches	82
5.4	<i>Conclusion</i>	83
6.	Toward patient-specific optimization of a maximum <i>a posteriori</i> reconstruction algorithm for simultaneous-acquisition dual-isotope myocardial perfusion SPECT	84
6.1	<i>Introduction</i>	84
6.2	<i>Methods</i>	85
6.2.1	Motivation and assumptions of the hyperparameter optimization methods	86
6.2.2	Hyperparameter optimization in the noise-smoothing-edge-preserving range	88
6.2.3	Hyperparameter optimization in the edge-preserving range	93
6.2.4	Image quality evaluation	101

6.3	<i>Results</i>	102
6.3.1	Optimal parameters	102
6.3.2	Image quality evaluation based on task performance	107
6.3.3	Image quality evaluation based on physical properties	107
6.3.4	Resolution Properties	113
6.4	<i>Discussion</i>	117
6.5	<i>Conclusions</i>	118
7.	Conclusions	120
7.1	<i>Summary of results</i>	120
7.1.1	A sub-ensemble-based approach to evaluate detection task performance for non-MVN distributed data using multi-template linear observer strategies	121
7.1.2	Toward patient-specific optimization of post-filtered OS-EM reconstruction for simultaneous-acquisition dual-isotope myocardial perfusion SPECT	122
7.1.3	Toward patient-specific optimization of a maximum <i>a posteriori</i> reconstruction algorithm with a dual-tracer prior for simultaneous-acquisition dual-isotope myocardial perfusion SPECT	124
7.2	<i>Summary of technical and methodological contributions</i>	125
7.3	<i>Future work</i>	126
7.3.1	Multi-template LD with pooled test statistic strategy evaluation	127
7.3.2	Research on the advantages and disadvantages of the DTMAP algorithm	127
7.3.3	Hyperparameter optimization for DTMAP	128
7.4	<i>Conclusions</i>	129
	REFERENCES	130
	CURRICULUM VITAE	136

LIST OF TABLES

Table 2-1. Defect status in the rest and stress images for different diagnoses.	5
Table 3-1. Defect parameters.....	29
Table 4-1. CMD for all sub-ensembles and all reconstruction parameters.....	45
Table 4-2. Determination ratio for all sub-ensembles and all reconstruction parameters.....	46
Table 4-3. Difference between AUC values for the multi-template LD with pooled test statistics and single-template HO strategies.	50
Table 4-4. Difference between AUC values for the multi-template LD with pooled test statistics and the multi-template HO with averaged AUCs strategies.	50
Table 4-5. Comparison of AUCs for Tc for individual sub-ensemble using the optimal parameters obtained using the three observer strategies.	55
Table 5-1. Four optimization approaches.....	75
Table 5-2. Optimal parameters for Tc.....	76
Table 5-3. Optimal parameters for Tl.	76
Table 5-4. Specificity at 0.8 and 0.9 sensitivity for the four optimization approaches for Tc.....	82
Table 5-5. Specificity at 0.8 and 0.9 sensitivity for the four optimization approaches for Tl.	82
Table 6-1. Ratio of image noise to the smallest edges of interest (with 10% severity) for Tc and Tl images.	90
Table 6-2. Ratio of image noise to the largest edges of interest (with 50% severity) for Tc and Tl images.	90
Table 6-3. Parameters of phantoms used to verify the effect of edge-preserving assumption.....	94
Table 6-4. Optimal values of k_s for different anatomy groups obtained in the noise-smoothing-edge-preserving range.	103
Table 6-5. The mean, minimum, and maximum values of k_r over all the population and uptake realizations for each anatomy groups obtained using the optimal k_s values.	103
Table 6-6. Optimal parameters for different anatomy groups obtained in the edge-preserving range.	104
Table 6-7. Three-class task performance obtained using the optimal hyperparameters in the two search ranges for DTMAP.....	104
Table 6-8. The mean, minimum, and maximum values of k_s/k_r for different anatomy groups obtained using the matched resolution method.....	105
Table 6-9. Task performance of optimized post-filtered OS-EM and DTMAP.....	107
Table 6-10. Parameters of phantoms used to study the image quality through physical properties..	108
Table 6-11. The ratio of the average COV in the myocardium for DTMAP divided by that of OS-EM.	113
Table 6-12. The ratio of the average fluctuations in the myocardium for DTMAP divided by that of OS-EM.....	113
Table 6-13. 2D FWHM of PSFs for 4 reconstruction algorithms and different objects.....	115
Table 6-14. 3D FWHM of PSFs for 4 reconstruction algorithms and different objects.....	115

LIST OF FIGURES

Figure 2-1. Illustration of atherosclerosis in the coronary artery. https://en.wikipedia.org/wiki/Coronary_artery_disease .	4
Figure 2-2. Schematic diagram showing the cross section of a common gamma camera detector (a copy of Figure 11.1 from [13] used by permission of Eric Frey with all rights reserved by him).	7
Figure 2-3. Six difference-of-mesa channels (top row) and their spatial-domain (bottom row).	21
Figure 2-4. An example of test statistic distribution, decision making, and ROC analysis.	24
Figure 2-5. (a) Decision variable and the decision structure for the practical 3-class ROC analysis method. (b) Sample 3-class ROC surface.	25
Figure 3-1. Sample images of six defect types 1-6 (from left to right) from the short axis view of the myocardium; the red arrows indicate defect locations.	29
Figure 4-1. Illustration of linear observer test statistic distributions for all sub-ensembles before shifting.	41
Figure 4-2. Illustration of linear observer test statistic distributions for all sub-ensembles after shifting.	41
Figure 4-3. Thirty bin histograms of the first channel feature value distribution for the entire ensemble (left) and for the SKS sub-ensemble corresponding to defect type 1 and contains the mixture of anatomies in whole patient population (right) for Tc at iteration number 5 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} .	44
Figure 4-4. Covariance matrices under the two hypotheses (top row: defect-absent, bottom row: defect-present) for the six sub-ensembles (the 1st to the 6th columns are corresponding to defect type 1 to 6, respectively) for Tc at iteration number 5 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} .	44
Figure 4-5. Diagram illustrating the data flow for the three observer strategies investigated. S_j stands for the j th sub-ensemble ($1 \leq j \leq N$), where N is the total number of sub-ensembles. Δg_j , Kg_j , and η_j denote the mean data vector, the covariance matrix of the data, and the term defined in Equation (4-4) for the j th sub-ensemble.	47
Figure 4-6. Sixty-four bin histograms of HO test statistics for different defect types for Tc at iteration number 1 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} . The graphs, from top to bottom, are for relative defect types 1-6, respectively.	48
Figure 4-7. Sixty-four bin histograms of LD test statistics for different defect types for Tc at iteration number 1 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} . The graphs, from top to bottom, are for relative defect types 1-6, respectively.	49
Figure 4-8. AUC values with the optimal parameters obtained by three observer strategies.	51
Figure 4-9. Optimal parameters for all defect types for Tc (left) and Tl (right) using (top to bottom): single-template HO, multi-template HO with averaged AUCs strategy, and multi-template LD with pooled test statistics strategies. The crosses represent the set of parameters that achieved the maximum AUC. The filled circles represent the parameter points where the difference in AUC with respect to the maximum AUC was not statistically significant. The contour line indicates the region where AUC values that differed from the optimal one by no more than 0.01, a difference considered clinically important. We used 4 subsets/iteration during the OS-EM reconstructions.	53
Figure 4-10. Thirty bin feature vector histograms of Channel 4 outputs for Tc (left) and Tl (right) for relative defect types 1 (top, anterior defect with 5% extent and 50% severity) and 2 (bottom, inferior defect with 5% extent and 50% severity) using 5 iterations (4 subsets/iteration) and a cutoff frequency of 0.1 pixel^{-1} .	54

Figure 4-11. AUC plot for Tc of three observer strategies for different cutoff frequencies (pixel ⁻¹) at iterations 2(upper) and 60 (lower) (4 subsets/iteration).	57
Figure 4-12. AUC plot for Tc of three observer strategies for different iteration numbers at cutoff frequencies 0.08 (upper) and 0.24 (lower) pixel ⁻¹	58
Figure 5-1. Illustration of the separation of phantoms into anatomy groups based on noise index (COV for 10 mCi Tc-99m) and defect to resolution ratio index (Ratio of cube root of 10% myocardium volume to average resolution): LRLN (low defect to resolution ratio low noise), LRHN (low defect to resolution ratio high noise), and HRLN (high defect to resolution ratio low noise).	74
Figure 5-2. Optimal parameter ranges for Tc (left) and TI (right) using by anatomy group approach (1 st row: for group HRLN; 2 nd row: for group LRHN; 3 rd row: for group LRLN) and using whole population approach (4 th row). The crosses represent the set of parameters that achieved the maximum AUC. The dots indicate the parameter points where the difference in AUC with respect to the maximum AUC was not statistically significant. The contour line indicates the regions where AUC values that differed from the optimal ones by no more than 0.01. We used 4 subsets/iteration during the OS-EM reconstruction.	78
Figure 5-3. Bar graph of AUC values for Tc using optimal parameters obtained with each of the four optimization approaches for the three anatomy groups and the whole population. The error bars represent the standard deviation of the AUCs.	80
Figure 5-4. Bar graph of AUC values for TI using optimal parameters obtained with each the four optimization approaches for the three anatomy groups and the whole population. The error bars represent the standard deviation of the AUCs.	80
Figure 5-5. Partial ROC curves for Tc for the whole population using optimal parameters obtained by the four optimization approaches.	81
Figure 5-6. Partial ROC curves for TI for the whole population using optimal parameters obtained by the four optimization approaches.	81
Figure 5-7. Bar graph of VUS values using optimal parameters obtained by the four optimization approaches. The error bars represent the standard deviations of the VUS values.	83
Figure 6-1. Sample images of QMAP (left) and post-filtered OS-EM (right) with matched resolution. The line indicates the position of the profile plotted in Figure 6-2.	92
Figure 6-2. Profile along the vertical line passing through the center of the left ventricular myocardium of QMAP and post-filtered OS-EM images with matched resolution shown in Figure 6-1.	93
Figure 6-3. Stress images of the phantom M.BS.HS.FS (Male, small body core size, small heart size, and small subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	95
Figure 6-4. Rest images of the phantom M.BS.HS.FS (Male, small body core size, small heart size, and small subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	96
Figure 6-5. Stress images of the phantom M.BM.HS.FS (Male, medium body core size, small heart size, and small Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	97
Figure 6-6. Rest images of the phantom M.BM.HS.FS (Male, medium body core size, small heart size, and small Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	98
Figure 6-7. Stress images of the phantom F.BL.HL.FL (Female, large body core size, large heart size, and large Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	99

Figure 6-8. Rest images of the phantom F.BL.HL.FL (Female, large body core size, large heart size, and large Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.	100
Figure 6-9. Plot of VUS versus δ for different values of β for the HRLN group and the anterior defect with 5% extent and 50% severity.....	106
Figure 6-10. Plot of VUS versus β for different values of δ for the HRLN group defect and the anterior location with 5% extent and 50% severity.	107
Figure 6-11. Sample Tc (top) and TI (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the parameters obtained in the edge-preserving range) for phantom M.BS.HS.FS with defect type 1 with 5% extent and 50% severity at the anterior location.....	108
Figure 6-12. Sample Tc (top) and TI (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the optimal parameters obtained in the edge-preserving range) for the phantom M.BL.HL.FL with defect type 1 with 5% extent and 50% severity at the anterior location.....	109
Figure 6-13. Sample Tc (top) and TI (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the parameters obtained in the edge-preserving range) for the phantom F.BL.HS.FM with defect type 1 with 5% extent and 50% severity at the anterior location.....	109
Figure 6-15. Mean, standard deviation, and COV images (generated using 100 noise realizations) of Tc images for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with a 5% extent and 50% severity at the anterior location. Note that the images in each column have the same display range [min, max], with min set to the smallest voxel value of the two images and max set to the maximum voxel value of the two images in a region that is defined by the myocardium dilated by 5 voxels.	111
Figure 6-16. Mean, Standard deviation, COV images (generated using 100 noise realizations) of TI images for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with 5% extent and 50% severity at anterior location. Note that the images in each column have the same display range [min, max], with min set to the smallest voxel value of the two images and max set to the maximum voxel value of the two images in a region that is defined by the myocardium dilated by 5 voxels.	112
Figure 6-17. Profile of a vertical line passing through the center of the defect for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with a 5% extent and 50% severity at the anterior location. The four different line styles represent four different noise realizations.	112
Figure 6-18. The local PSFs at the center of the defect obtained with (from left to right): unfiltered OS-EM (2 iterations with 4 subsets/iteration), post-filtered OS-EM (2 iterations and cutoff frequency of 0.2 pixels^{-1}), and DTMAP (DTMAP with hyperparameters obtained in the edge-preserving range) for 1) the male phantom with the large core body size, heart size and subcutaneous adipose tissue thickness with defect at an anterior position, 5% extent and 50% severity and a single uptake realization (first row); 2) same as 1) except the defect was at an inferior position (second row); 3) same as 1) except at for a different uptake realization (third row); 4) same as 1) except that the phantom had a medium body core size, small heart size and subcutaneous adipose tissue thickness (same anatomy group as 1), fourth row).....	114
Figure 6-19. Sample images of the phantom (just the myocardium is shown) and from optimal post-filtered OS-EM and DTMAP with parameters obtained in the edge-preserving range (from left to right). The phantom did not have a perfusion defect.	116

Figure 6-20. Profile from a vertical line passing through the center of the images reconstructed with optimal post-filtered OS-EM, DTMAP with parameters obtained in the edge-preserving range . The phantom did not have a perfusion defect.117

1. Introduction

1.1 Problem domain and contributions

Coronary artery disease (CAD) is the most prevalent type of heart disease and a leading cause of death both in the United States and worldwide. Myocardial perfusion SPECT (MPS) is a well-established, widely-used and cost-effective non-invasive technique to diagnose CAD. A lot of previous work has been devoted to optimizing various parameters and methods related to MPS imaging. However, any further optimization would still be highly significant and worthwhile if it can result in a better diagnosis of the disease. The overall goal of this work is to further optimize the reconstruction methods for MPS by exploring more patient-specific optimization of advanced iterative reconstruction methods based on simulations of realistic patients population and previously optimized imaging protocols. The contribution of this work can be summarized as follows.

First, we investigated two multi-template linear observer strategies, one adapted from previous work and a novel strategy proposed here referred to as the multi-template linear discriminant (LD), based on sub-ensemble-based analysis of data, which is non-multivariate-normally (MVN) distributed, from a clinically realistic signal-known-statistically (SKS) task. We compared these strategies and to a commonly-used single-template Hotelling Observer strategy in the context of optimizing reconstruction parameters for a realistic dual-isotope MPS simulated dataset. The proposed multi-template LD with pooled test statistic strategy is shown to be a general, theoretically justified, and practical strategies for handling non-MVN data such as that arising from clinically-realistic SKS tasks.

Second, this work validated the hypothesis that patient-specific reconstruction parameters would result in better task performance. Not only is this significant for MPS and the diagnosis of CAD, the patient-specific optimization strategy could be applied to other imaging modalities and

may result in significant improvement in the diagnosis and treatment of other diseases.

Third, we propose a practical group-wise patient-specific reconstruction parameter optimization strategy. The results and methods for patient-specific OS-EM optimization could be applied immediately in clinical MPS studies. Similar methods could also be developed and used for other studies.

Finally, we proposed a practical patient-specific hyperparameter optimization method for a dual-tracer MAP algorithm (DTMAP) [1] specifically developed for dual-isotope MPS studies. The optimized DTMAP had poorer task performance compared to the optimized OS-EM algorithm. We also studied the limitations of the DTMAP algorithm and suggest reasons for the lack of improvement of task performance. While this is a negative result in the sense that it fails to demonstrate the superiority of the more theoretically sophisticated DTMAP algorithm, the results provide insight into methods for improving the selection of hyperparameters for the DTMAP algorithm and design of optimal reconstruction methods for MPS.

1.2 Organization

This dissertation is organized as follows.

Chapter 2 provides background by giving an overview of CAD, MPS imaging, image degrading factors of simultaneous dual-tracer SPECT imaging, reconstruction methods, the basic concepts of task-based image quality assessment, and the motivation of patient-specific optimization.

Chapter 3 describes common methods used in this work by describing the details of the realistic digital phantom population and the generation of the simultaneous dual-isotope MPS projection images.

In Chapter 4, we propose a sub-ensemble-based approach to handle the non-MVN data that can result from clinically realistic tasks: the entire data ensemble is divided into sub-ensembles that are MVN distributed and homoscedastic. We investigated two multi-template model observer

strategies to handle the sub-ensemble-based data. Then, we compare these strategies with that of a conventional single template HO by applying them to optimize reconstruction parameters for the non-MVN distributed data from a realistic dual-isotope MPS simulation dataset. The methods proposed here were applied to the work in Chapters 5 and 6.

Chapter 5 proposes patient-specific optimization methods for OS-EM. This work was presented in 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference.

In Chapter 6, we proposed a practical method to obtain patient-specific hyperparameters for DTMAP reconstruction algorithm. We compared the performance of optimized DTMAP to optimized OS-EM. We also studied the limitation of DTMAP. Part of this work was presented in 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference.

Chapter 7 summarizes the results of this dissertation, discusses the contributions of this work, and suggests some possible areas for future work.

2. Background

2.1 Coronary artery disease and myocardial perfusion

Cardiovascular disease is the leading cause of the death worldwide and the most common manifestation is coronary artery disease (CAD) [2]. CAD is caused by atherosclerosis due to the accumulation of plaques made up of fat, cholesterol, etc., as shown in Figure 2-1. CAD can decrease the blood flow supply to the myocardium (the muscle of the heart). Insufficient myocardial perfusion (blood flow to the myocardium) can affect the ability of the heart to pump blood to the body. Due to autoregulation, myocardial perfusion may not be affected when a patient is at rest until the stenosis exceeds 90% of the artery diameter [3]. However, subcritical stenosis may still result in myocardial perfusion reduction when the patient is under pharmacological or exercise stress [4]. Myocardial perfusion imaging, which assesses the extent and severity of reduced perfusion in the myocardium when a patient is at rest and stress, can help diagnose and evaluate the prognosis of patients with CAD. There is a large literature supporting this application, e.g., [5-8].

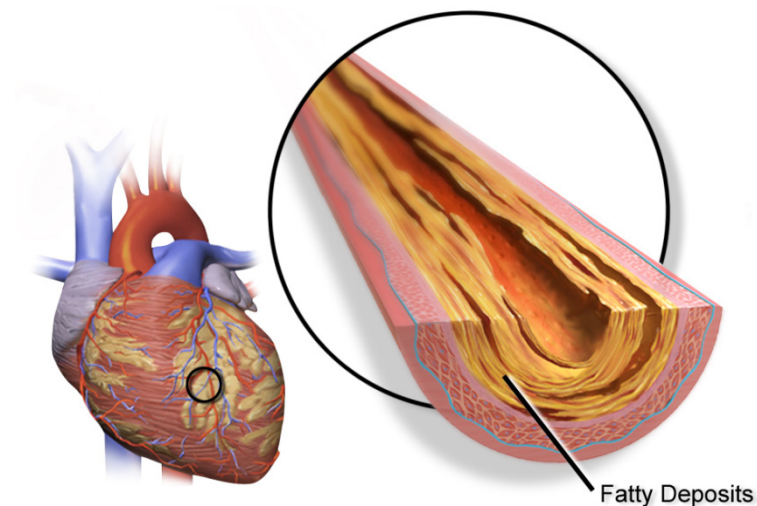


Figure 2-1. Illustration of atherosclerosis in the coronary artery.
https://en.wikipedia.org/wiki/Coronary_artery_disease.

If there is no defect (i.e., no region of reduced myocardial perfusion) for both rest and

stress image, the patient has normal perfusion status; if there is a defect only in the stress image, the patient has a reversible defect; if there are defects in both images, the patient has a fixed (non-reversible) perfusion defect. Non-reversible defects are typically an indicator of infarcted myocardial tissue as the result of a previous heart attack. The diagnostic task requires classifying the patients into one of the three categories, as shown in Table 2-1 (from Table I in [9]), using the defect status in the rest and stress images.

Table 2-1. Defect status in the rest and stress images for different diagnoses.

	Normal	Fixed defect	Reversible defect
Rest	No ^a	Yes ^b	No
Stress	No	Yes	Yes

^a No denotes that defect is absent

^b Yes denotes that defect is present

2.2 Myocardial perfusion SPECT

There are multiple medical imaging modalities that can be used to image myocardial perfusion non-invasively, including SPECT, PET, MRI, CT, and ultrasound. Each of these has their own advantages and disadvantages [10, 11]. Among them, SPECT is currently the best validated, well established, most cost-effective, and most widely used non-invasive CAD diagnostic imaging technique [3].

SPECT is a 3D non-invasive imaging technique to image the distribution of a radionuclide in a patient. First, the patient is administrated a radiopharmaceutical. The radiopharmaceutical distributes in the body and can be taken up by different organs depending on factors including the properties of the radiopharmaceuticals and the organs. The gamma photons emitted when the radionuclide portion of the radiopharmaceutical decays and can travel through the body and be detected by a gamma camera. A typical SPECT system consists of one or more gamma cameras mounted on a gantry. The gamma cameras acquire images that represent 2D projections of the 3D distribution of radioactivity inside the patient [12]. If we have enough projection views, we can estimate the activity distribution in the patient, such process is called reconstruction.

2.2.1 SPECT system hardware

A SPECT system consists of one or more gamma cameras, a gantry and a bed. The most commonly used gamma cameras in SPECT systems are usually comprised of a parallel-hole collimator, a scintillation crystal coupled to an array of photo-multiplier tubes (PMTs) and associated detection electronics (including preamplifiers, shaping and positioning, and corrections and framing), as shown in Figure 2-2 (a copy of Figure 11.1 from [13] used by permission of Eric Frey with all rights reserved by him). The collimator is analogous in function to a lens in optical imaging, and establishes a relationship between positions in the image and directions in space; ideally, only photons passing through the holes will be detected. The scintillation crystal converts the gamma photon's energy into visible light, and the PMTs convert the light into a charge signal. The positioning electronics has as its input the outputs of the PMT array and estimates the energy and position of interaction of an incident photon. The patient is positioned on the bed and moved into the gantry so the region-of-interest can be imaged by the camera. The gantry provides the ability to move and rotate the cameras to acquire projection images at multiple views to reconstruct the 3D distribution of the radioactivity inside the patient. Many modern SPECT systems are coupled to an x-ray CT scanner that provides attenuation information which is used to compensate for attenuation of photons in the patient body during image reconstruction, as will be described below. In this research, our work is based on GE Infinia dual-head SPECT system.

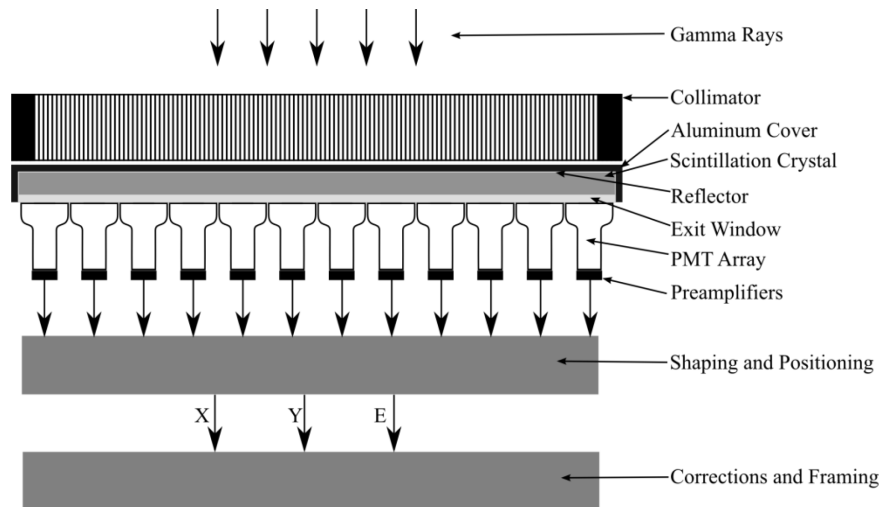


Figure 2-2. Schematic diagram showing the cross section of a common gamma camera detector (a copy of Figure 11.1 from [13] used by permission of Eric Frey with all rights reserved by him).

2.2.2 Radiopharmaceuticals for myocardial perfusion SPECT

An ideal radiopharmaceutical for MPS should have (1) a radionuclide that emits photons with energies suitable for gamma camera imaging, decays without emission of charged particles, and has a half-life that is long enough to allow for imaging and short enough so that a sufficient number of decays occur during imaging, (2) a high first-pass extraction fraction as the initial myocardial uptake after intravenous administration is determined by blood flow and first-pass extraction fraction, (3) a linear relationship between myocardial uptake and blood flow for relevant ranges of blood flow, (4) low uptake in surrounding organs such as the liver and lungs, and (5) sufficiently long retention (i.e., biological half-life in the myocardium) in the myocardium to allow for SPECT imaging [10, 14]. The commonly used radiopharmaceuticals for MPS are: Tl-201, Tc-99m labeled sestamibi and tetrofosmin. In this project, we used Tl-201 for rest study and Tc-99m sestamibi for stress study.

2.2.3 MPS acquisition protocols

Single isotope or dual isotopes can be used to obtain images of the patient at rest and stress states. The rest and stress images can be obtained separately or simultaneously when dual isotopes are used. In this project, we studied a simultaneous dual-isotope MPS protocol [15-18].

It has the advantages compared to separate acquisition protocols of saving time and improving patient throughput, which reduces cost, increasing patient comfort, and providing perfect registration of stress and rest images in time and space. In our study, we used a previously-optimized dual-isotope simultaneous-acquisition protocol [18] with administered activities of 10 mCi of Tc-99m and 2 mCi of Tl-201 (total dose of 13.5 mSv). The protocol details can be found in [18].

2.3 Image reconstruction

In order to reconstruct images from measured projection data, we first need to understand the detection process. Ideally, a SPECT system measure the simple line integral through the activity distribution process. However, in a real SPECT system, there are a number of factors that are non-ideal and needed to be modeled during image reconstruction.

2.3.1 Image degrading factors

Here, we review the image degrading factors for our dual-radionuclide MPS study.

2.3.1.1 Attenuation and scatter in patient

There are two physical effects caused by the interaction of photons with the patient that can deteriorate the image quality: attenuation and scatter.

When gamma photons travel through the body, they can interact with the body mainly by two physical interactions: photoelectric absorption and Compton scatter. In photoelectric absorption, the gamma photons interact with inner-shell electrons and result in the ejection of these electrons, hence referred to as photoelectrons, as the energy of the gamma photons is completely transferred to the photoelectrons. This interaction results in a significant reduction in the number of photons exiting the body. The probability of this interaction is dependent on the gamma photon energy and the composition of the tissues [12]. In Compton scattering, the gamma photon interacts with an outer-shell electron by elastic collision, transfers part of its energy to the

electron, and is deflected from its original path. A Compton scattering event results in a decrease in energy and a change in direction of the photon. A scattered photon can exit the body or subsequently undergo a photoelectric or one or more additional scatter events [12].

One effect of these interactions is a reduction in the number of primary photons (photons that have not interacted) exiting the body, referred to as the attenuation effect. This effect can be described mathematically by the following exponential expression [12]:

$$N = N_0 e^{-\mu \Delta x} \quad (2-1)$$

where N_0 is the number of incident photons incident on a slab of material with a thickness of Δx , N is the number of exiting photons, and μ is a constant describing the loss of photons referred to as the attenuation coefficient of the material. Thus, the attenuation effect results in a depth-dependent reduction of the number of gamma photons detected. A discretized version of the attenuation coefficient distribution, referred to the attenuation map, is typically estimated from x-ray CT images of the patient and can be used to compensate for attenuation effects.

Since a Compton scattered photon does not carry correct information about the original photon emission position, detection of it would degrade the image quality. It is desirable to eliminate scattered photons from the data or compensate for their effects in the reconstructed images. First, since Compton scatter reduces the energy of interacting gamma rays and emitted gamma rays have well-defined energies, the energy of detected photons can be used to distinguish between scattered and unscattered photons. This process is referred to as energy discrimination and involves counting photons that are detected with an energy that falls in a specified range (an energy window) around the emitted photon energies. However, due to the finite energy resolution of gamma cameras, scattered photons can pass through the collimator and be detected within the detection energy window. In this work, the distribution of scattered photons in the measured projection data was modeled using the effective source scatter estimation (ESSE) method [19] to compensate for the effects of scatter.

2.3.1.2 Collimator-detector blurring

An ideal parallel-hole collimator would only pass photons incident perpendicular to the detector. However, this would result in a vanishingly small number of detected photons, and thus, since the number of detected photons is a Poisson random variable, very high image noise. As a result, real collimators accept photons with a finite range of incident angles (typically a few degrees). This results in a loss of precision in the direction information associated with the detected photons, producing distant-dependent blurring in the projection images. In addition, the detector itself estimates the position of photon interaction in the crystal with limited precision. Together, these effects are characterized by the collimator-detector response function (CDRF), which is an image of a point source at a particular position in the field-of-view. The CDRF can be approximated analytically or estimated by Monte Carlo (MC) simulation or experimental measurements and be compensated for during image reconstruction [20].

2.3.1.3 Crosstalk between different isotopes

Crosstalk contamination between the projection images of the two radionuclides is a special problem in simultaneous-acquisition dual-isotope imaging. It happens because the photons from one isotope can be detected within the energy window of the other isotope due to scatter in the body or collimator detector system or the finite energy resolution of the detector. Crosstalk contamination can cause artifacts that result in misinterpretation of the reconstructed MPS images. Previously, model-based methods have been developed to compensate for crosstalk [21, 22]. These methods involve modeling the physical effects that result in this contamination during the image reconstruction process.

2.3.1.4 Statistical noise

Due to the randomness of radioactive decay and interactions of the photons with the body and collimator-detector system, the number of detected photons is a random variable. The number of detected photons per projection bin can be described by a Poisson distribution. A figure of

merit related to image quality that characterizes the level of image noise is the coefficient of variation (COV) of a projection bin, defined as the standard deviation of the counts in a bin divided by the mean value and is given by [23]

$$\text{COV} = \frac{\sqrt{\overline{N}}}{\overline{N}} = \frac{1}{\sqrt{\overline{N}}}, \quad (2-2)$$

where \overline{N} is the mean total number of acquired photons.

The reciprocal of the COV is the signal-to-noise (SNR) ratio and is given by [12, 23]:

$$\text{SNR} = \frac{1}{\text{COV}} = \sqrt{\overline{N}}, \quad (2-3)$$

The more counts acquired, the less noisy the SPECT projection data.

2.3.2 Reconstruction methods

In SPECT imaging, the goal is to estimate the 3D distribution of the radionuclide in the patient, denoted by $f(\vec{r})$, where \vec{r} is the spatial coordinates and $f(\vec{r})$ is a continuous function of \vec{r} . The function $f(\vec{r})$ can be approximated by equally spaced samples in the spatial domain, i.e., a discretized image. If the spatial domain is discretized into N voxels, we can use a vector $\mathbf{f} = [f_1, f_2, \dots, f_N]$ to represent the object. If we approximate the SPECT imaging system as a linear system, the relation between the projection data \mathbf{g} and \mathbf{f} can be expressed below using vector notation:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (2-4)$$

where \mathbf{H} and \mathbf{n} denote the imaging system (projection) operator and noise, respectively. If \mathbf{f} is an $N \times 1$ vector and \mathbf{g} is an $M \times 1$ vector, \mathbf{H} will be an $M \times N$ matrix and \mathbf{n} will be an $M \times 1$ vector. The imaging system operator, \mathbf{H} , characterizes all the factors in the image formation process, including all the factors mentioned in Section 2.3, with H_{ij} being the probability that a photon emitted in image voxel j is detected in projection bin i .

We can obtain an estimate $\hat{\mathbf{f}}$ of \mathbf{f} from the projection data \mathbf{g} using a reconstruction algorithm, symbolized by the operator \mathbf{O} , as follows:

$$\hat{\mathbf{f}} = \mathbf{O}\mathbf{g}. \quad (2-5)$$

There are two categories of reconstruction methods: analytical and iterative. There are Iterative methods are subdivided into statistically- and non-statistically based ones. Analytical methods, including classical filtered backprojection (FBP), are fast. However, analytical methods to invert the reconstruction problem that account for image degrading factors (e.g., attenuation, scatter, collimator-detector blurring, and crosstalk,) are presently not available. Using analytic methods that do not take into account these effects leads to inaccurate (non-quantitative), blurred and noisy reconstructed images and results in the presence of various image artifacts. These images often give rise to suboptimal diagnostic performance [24].

Iterative reconstruction methods have the ability to model all image degrading factors, and the use of statistically-based method should, at least theoretically, provide better control of image noise. Among the statistically-based iterative methods, the Maximum Likelihood (ML) [25] algorithm is the most rigorously correct approach. It finds the $\hat{\mathbf{f}}$ that maximizes the statistical likelihood of the projection data given the image, i.e., $P(\mathbf{g}|\hat{\mathbf{f}})$.

However, unregularized ML reconstruction results in very noisy images, due, in some sense, to overfitting the data. Two kinds of regularization methods have been used to control reconstructed image noise: post-smoothing the image stopped after a user-specified number of iterations (post-filtered ML) and adding a roughness penalty to the maximum-likelihood objective function to control noise, e.g. using maximum *a posteriori* (MAP) methods.

For the former method and in the case of Poisson noise, an expectation-maximum (EM) algorithm is typically usually used to find the ML estimates (ML-EM). The solution can be obtained using the following iterative scheme:

$$\hat{\mathbf{f}}^{(n+1)} = \frac{\hat{\mathbf{f}}^{(n)}}{\mathbf{H}^T \times \mathbf{e}} \left(\mathbf{H}^T \times \frac{\mathbf{g}}{\mathbf{H} \times \hat{\mathbf{f}}^{(n)}} \right), \quad (2-6)$$

where $\hat{\mathbf{f}}^{(n)}$ and $\hat{\mathbf{f}}^{(n+1)}$ are the n th and $(n+1)$ th estimate, \mathbf{e} is an $M \times 1$ vector with all elements equal to 1, and \times denotes the conventional matrix-vector multiplication. Note that the vector-vector multiplication and division operation in (2-6) is element-by-element. In ML-EM, each new update is modified through a correction term that is based on the previous update. If the initial estimate (usually an image with all ones) is non-negative, the final estimate will be non-negative.

Since ML-EM tends to converge slowly, ordered subset acceleration (OS-EM [26]) is often used to accelerate the reconstruction. In OS-EM, the projection data from all angles are split into several mutually exclusive subsets. The estimate is updated for each subset using

$$\hat{\mathbf{f}}^{(l+1, S_q)} = \frac{\hat{\mathbf{f}}^{(l, S_q)}}{\mathbf{H}_{S_q}^T \times \mathbf{e}} \left(\mathbf{H}_{S_q}^T \times \frac{\mathbf{g}^{(S_q)}}{\mathbf{H}_{S_q} \times \hat{\mathbf{f}}^{(l, S_q)}} \right), \quad (2-7)$$

where S_q is the q th subset, \mathbf{H}_{S_q} is the system operator for the q th subset, $\mathbf{g}^{(S_q)}$ is the projection data in the q th subset, and l is the local subset update number. Suppose there are S subsets. Each full iteration results in S updates using a subset of the projection data. As a result, OS-EM converges nearly S times faster than ML-EM. Again, post-reconstruction filtering is typically applied to the OS-EM reconstructed images to control noise.

In this dual-isotope MPS study, the rest and stress images can be reconstructed separately using post-filtered OS-EM. There are four reconstruction parameters that can be adjusted to achieve optimal image quality: the iteration numbers and post-filtering cutoff frequencies for the rest and stress images.

In this dual-isotope MPS study, we also studied a state-of-the-art MAP reconstruction algorithm that incorporates a dual-tracer prior [9]. This method is referred to as DTMAP in this dissertation. In this method, the estimate is obtained by minimizing the following objective

function:

$$\Phi(\mathbf{f}_r, \mathbf{f}_s) = -L(\mathbf{f}_r) - L(\mathbf{f}_s) + \beta \Lambda(\mathbf{f}_r, \mathbf{f}_s), \quad (2-8)$$

where \mathbf{f}_r and \mathbf{f}_s are the image vectors of the rest and stress images, $L(\mathbf{f}_r)$ and $L(\mathbf{f}_s)$ are the log likelihood of the rest and stress images given the projection data, and β is the weight of the dual-tracer prior $\Lambda(\mathbf{f}_r, \mathbf{f}_s)$. The prior is defined as

$$\Lambda(\mathbf{f}_r, \mathbf{f}_s) = \sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \left(\sqrt{1 + \left(\frac{\mathbf{f}_r[j] - \mathbf{f}_r[k]}{\delta} \right)^2 + \left(\frac{\mathbf{f}_s[j] - \mathbf{f}_s[k]}{\eta} \right)^2} - 1 \right), \quad (2-9)$$

where J is the number of voxels in the image, N_j refers to the neighborhood of voxel j , ω_{jk} are the fixed non-negative weights between neighboring pixels j , and k , and δ and η are two parameters that define the properties of the prior. A convergent ordered subsets algorithm has been derived to iteratively find the solution [1].

Using this DTMAP algorithm, the rest and stress images are reconstructed simultaneously. Since the algorithm is convergent, it is iterated to effective convergence (i.e., where the value of the objective function changes very slowly with iteration). Three adjustable parameters, β , δ , and η , are present in the objective function and need to be adjusted to obtain optimal image quality.

In Chapter 5 and 6, we propose methods to optimize the reconstruction parameters of both post-filtered OS-EM and DTMAP when applied the dual-isotope MPS data and a myocardial perfusion defect detection task. We then compared the quality of the images generated by each of the reconstruction methods in terms of performance on this task.

2.4 Task-based image quality evaluation

Objective assessment of image quality in a clinically-relevant way is essential in the evaluation and optimization of medical imaging systems. Traditionally, image quality has been defined in terms of physical quantities such as the image resolution, sensitivity, contrast, and noise. While such metrics are still widely used, and provide insights about factors that affect

various components of image quality, they are not directly related to the performance of an observer (e.g., a radiologist) on a specific task. A more sophisticated way to evaluate image quality is to define it based on the performance of a certain observer on a specific task of clinical interests [27]. In order to quantify the task performance, a figure of merit (FOM) that objectively quantifies the performance of the observer on the task is needed. How image quality is evaluated will be dependent on these key factors: the task, the observer, and the FOM. In the following subsections, we will discuss these key elements in detail.

2.4.1 Tasks

One of the main philosophies underlying task-based image quality evaluation is that imaging systems or images are evaluated with respect to a task. In medical imaging, the task should be clinically relevant. The tasks of interest can be either a parameter estimation task or classification tasks. An estimation task involves estimating the value of certain quantities from images. For example, tumor volume can be estimated from a CT image. In cardiac imaging, the cardiac ejection fraction of the heart, which is the fraction of blood pumped out from the left ventricle in each heart cycle, can be estimated from an echocardiogram or dynamic MPS images. Another example is that the myocardial perfusion defect extent, which is the fraction of the myocardium that has reduced blood flow and can be estimated from myocardial SPECT images. In a classification task, an observer sorts images into two or more classes. For example, in medical imaging, this could be deciding whether an image indicates the presence or absence of a certain disease. In MPS the task is typically a two-class or three-class classification task, as described below.

2.4.1.1 Binary classification task

In a binary classification task, also known as a detection task, we classify the objects, \mathbf{f} , into two categories: signal present and signal absent. Mathematically, the two hypotheses are defined as:

$$\begin{aligned}
H_0 : \mathbf{f} &= \mathbf{f}_b, \\
H_1 : \mathbf{f} &= \mathbf{f}_b + \mathbf{f}_s,
\end{aligned} \tag{2-10}$$

where \mathbf{f}_b and \mathbf{f}_s are the background and signal and H_0 and H_1 are the hypotheses for the signal-absent and -present, respectively. When the exact mean signal or background is known, the tasks are referred to as signal known exactly (SKE) and background known exactly (BKE). If the background or signal are random and described by a known statistical model then they are referred to as signal known statistically (SKS) and background known statistically (BKS) tasks.

2.4.1.2 Three-class classification task

Objects can also be classified into more than two classes. For example, in rest-stress MPS imaging, the goal is to classify patients into three classes as described in Section 2.1: normal, having a fixed perfusion defect, or having a reversible perfusion defect. Mathematically, the 3-class hypotheses can be defined by:

$$\begin{aligned}
H_N : \mathbf{f} &= \mathbf{f}_b \\
H_F : \mathbf{f} &= \mathbf{f}_b + \mathbf{f}_F, \\
H_R : \mathbf{f} &= \mathbf{f}_b + \mathbf{f}_R,
\end{aligned} \tag{2-11}$$

where \mathbf{f}_b , \mathbf{f}_F , \mathbf{f}_R are the background, fixed signal, and reversible signal and H_N , H_F and H_R are the hypotheses for signal absent, fixed and reversible defects classes, respectively.

2.4.1.3 More advanced tasks

There has also been research directed towards more advanced tasks that are, in some cases, more relevant than detection and 3-class classification tasks. These include multi-signal detection [28] and search (localization and detection) tasks [29]. These tasks are more closely related to some clinical tasks. For example, in MPS it is often desirable to know where a perfusion defect is in the heart in order to direct revascularization therapy. However, FOMs for evaluating these tasks are much less mature than those available for detection tasks. In this study, we only consider

the two- and three-class classification SKE/BKS and SKS/BKS tasks with signal location known exactly (LKE).

2.4.2 Observers

Images should be interpreted or evaluated by relevant observers. In general, observers can be divided into two categories: human and mathematical (model) observers. In medical imaging, human observers such as radiologists perform the relevant task and so, strictly speaking, should be used in evaluation and optimization of imaging methods. However, human-observer studies are time-consuming, expensive and there is variability within and between observers [30]. Further, when the end objective is finding the optimal parameter for, e.g., a reconstruction method, hundreds of images for each of several possible parameter values must be evaluated to judge image quality. It is prohibitively expensive and slow, and thus essentially impossible, to use human observers for such parameter optimization. Thus, model observers have been developed and applied for such applications. Model observers are mathematical algorithms that extract features from images and calculate test statistics indicating the presence or absence of a signal. There are a number of categories of model observers for classification tasks. One such observer is the Bayesian Ideal Observer (IO) [31], which uses all statistical information and has the best task performance of any observer in the sense of minimizing Bayes risks. A second important category is anthropomorphic observers, which are model observers that mimic the performance of human observers. In this dissertation, we are interested in developing observers for SKE/BKS and SKS/BKS with LKE tasks. In this section, we will review the IO and the most widely used anthropomorphic observer in medical imaging as they are used in the work presented in this dissertation.

2.4.2.1 The Ideal Observer (IO)

The ideal observer (IO) minimizes Bayes' risk or, equivalently, maximizes expected utility [32, 33]. The IO uses all statistical information available in an optimal way and the performance

is better than or equivalent to any other observer. For a binary classification task, the IO uses the ratio of the conditional probability distribution of the data under the signal present and absent hypotheses given the measured image. This ratio is known as the likelihood ratio (LR). The observer can also use any monotonic transformation of the LR, such as the log of the likelihood ratio, as the test statistic or decision variable [34]. The log likelihood ratio is often used because further simplifications can be made and it is defined as:

$$\lambda_{H_0, H_1}(\mathbf{g}) = \log \frac{pr(\mathbf{g} | H_1)}{pr(\mathbf{g} | H_0)}. \quad (2-12)$$

For the three-class classification task, the IO uses a pair of log-likelihood ratios as decision variables [35], defined as,

$$\lambda_{N,R}(\mathbf{g}) = \log \frac{pr(\mathbf{g} | H_R)}{pr(\mathbf{g} | H_N)}, \quad (2-13)$$

$$\lambda_{R,F}(\mathbf{g}) = \log \frac{pr(\mathbf{g} | H_F)}{pr(\mathbf{g} | H_R)}. \quad (2-14)$$

Due to its optimality, the IO can extract the most information from images for a given task. It is often applied to evaluate and optimize imaging system hardware, as this ensures that the maximum information about the task is present in the acquired data. In such tasks, the optimization is usually performed on measured (raw) data (e.g., projection data). The IO is generally not useful for optimizing image reconstruction and processing methods that are ultimately interpreted by human observers as the IO has better performance than human observers and does not suffer from some of the same limitations as the human visual system. However, the IO, as will be discussed below, is often applied to data following a data reduction process designed to model the human perception process. As a result, a discussion of the IO is relevant to the work in this dissertation.

2.4.2.2 The Hotelling Observer (HO) and Channelized Hotelling Observer (CHO)

The IO uses the LR or any monotonic transformation of the LR as the test statistic, which

is usually a non-linear function of the data \mathbf{g} . As a result, it is, in general, mathematically intractable and difficult to implement. Thus linear observers, which use a linear function of the data as the test statistic, are widely used. For a general linear observer, the test statistic can be written as:

$$\lambda(\mathbf{g}) = \mathbf{w}'\mathbf{g} \quad (2-15)$$

where \mathbf{w} is a linear template, \mathbf{g} is the image data, and t is the transpose operation.

For a binary detection task, the Hotelling Observer (HO), which is derived from linear discriminant analysis [36], is widely used. It is the linear observer that maximizes the signal-to-noise ratio (SNR), a measure of the ability of an observer to separate data from two classes, defined as [31]:

$$SNR = \frac{\langle \lambda(\mathbf{g}) \rangle_{H_1} - \langle \lambda(\mathbf{g}) \rangle_{H_0}}{\sqrt{\frac{1}{2}(\sigma_{H_1}^2 + \sigma_{H_0}^2)}} \quad (2-16)$$

where $\langle \lambda(\mathbf{g}) \rangle_{H_i}$ and $\sigma_{H_i}^2$ are the mean and variance of the test statistics under hypothesis H_i ($i=0, 1$).

For the two-class classification task, the Hotelling observer test statistic is calculated by taking the dot product of a vector, known as the Hotelling template, and the vector of measured data. The Hotelling template is generated using the first and second order statistics of the data and it is given by [37]:

$$\mathbf{w}_{N,R} = \mathbf{K}_{N,R}^{-1} \Delta \mathbf{g}_{N,R}, \quad (2-17)$$

$$\mathbf{K}_{\mathbf{g}} = \left[\frac{1}{2}(\mathbf{K}_0 + \mathbf{K}_1) \right], \quad (2-18)$$

$$\Delta \mathbf{g} = \mathbf{g}_1 - \mathbf{g}_0, \quad (2-19)$$

where \mathbf{K}_i and \mathbf{g}_i are the covariance matrix and mean vector of the data \mathbf{g} under the hypothesis H_i ($i=0, 1$).

For the three-class classification problem, the Hotelling template is generated similarly as

described in [38]:

$$w_{N,R} = \mathbf{K}_{N,R}^{-1} \Delta \mathbf{g}_{N,R}, \quad (2-20)$$

$$w_{R,F} = \mathbf{K}_{R,F}^{-1} \Delta \mathbf{g}_{R,F}, \quad (2-21)$$

$$\Delta \mathbf{g}_{N,R} = \mathbf{g}_R - \mathbf{g}_N, \quad (2-22)$$

$$\Delta \mathbf{g}_{R,F} = \mathbf{g}_F - \mathbf{g}_R, \quad (2-23)$$

$$\mathbf{K}_{N,R} = \left[\frac{1}{2} (\mathbf{K}_N + \mathbf{K}_R) \right], \quad (2-24)$$

$$\mathbf{K}_{R,F} = \left[\frac{1}{2} (\mathbf{K}_R + \mathbf{K}_F) \right], \quad (2-25)$$

where the data \mathbf{g}_i ($i=N,R,F$) is the concatenation of the stress and rest image vectors, $\mathbf{g}_i = \begin{bmatrix} \mathbf{g}_{s,i} \\ \mathbf{g}_{r,i} \end{bmatrix}$,

and \mathbf{K}_i is the covariance matrix of the data \mathbf{g}_i ($i=N,R,F$).

Calculating the HO test statistic or the SNR requires knowledge of the first and second order statistics of the data: $\bar{\Delta \mathbf{g}}$ and \mathbf{K}_g . The mean difference in the class vectors, $\bar{\Delta \mathbf{g}}$, is easy to compute. However, \mathbf{K}_g may be computationally intensive to estimate and invert because of its size.

Adding a channel mechanism to the HO can reduce the computational complexity. Channelizing the input involves applying a series of channel templates to the image data, resulting in one scalar value for each channel template, i.e.

$$v_i = \mathbf{c}_i^T \mathbf{g}, \quad (2-26)$$

where \mathbf{c}_i is the i th channel, T is the vector transpose operator, and v_i is the scalar output for the i th channel. Suppose \mathbf{g} is an $M \times 1$ vector image, \mathbf{c}_i will also be an $M \times 1$ vector. Suppose there are a total of L channels, then the channel output is an $L \times 1$ vector obtained by stacking the scalar output from all the channels together:

$$\mathbf{v} = [v_1 \ v_2 \ \cdots \ v_L]^T. \quad (2-27)$$

Usually, $L \ll M$, so the dimension of the data is greatly reduced.

Two important types of channels are anthropomorphic and efficient channels. Anthropomorphic channels are intended to model the human visual system. A common kind of anthropomorphic channels is a series of frequency selective channels that models some features of the human system. Among the proposed anthropomorphic channel models are: difference-of-mesa channels [39] (an example is shown in Figure 2-3), difference-of-Gaussian channels [40] and Gabor channels [41]. The combination of a channel operator and the HO applied to the vector of channel outputs is referred to as the channelized Hotelling Observer (CHO), and, with an appropriate channel model, this observer has been shown to agree well with human performance for SKE/BKE tasks [42, 43] and SKE-lumpy backgrounds tasks [44]. For more clinically relevant SKE/BKS and SKS/BKS tasks, there is evidence of correlation between rankings obtained from the CHO and human observers for different noise levels [45, 46], reconstruction methods and phantom groups [47], imaging systems [48], and compensation methods and post-filter cutoff frequencies [49].

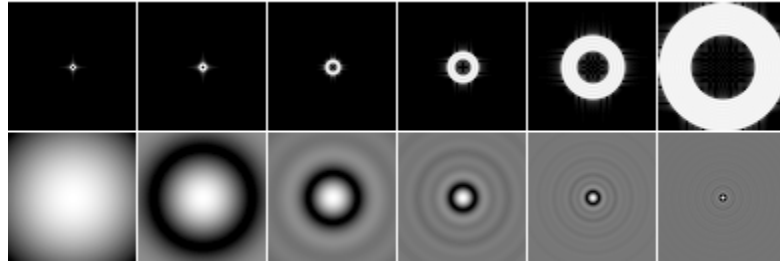


Figure 2-3. Six difference-of-mesa channels (top row) and their spatial-domain (bottom row).

Efficient channels [32] are channels that can reduce the data dimensionality (i.e., the size of the input vector), without unacceptable degradation in the observer's performance. With proper efficient channels, the CHO has also been investigated as a surrogate for the HO or IO [50]. The use of efficient channels is important for these applications because of the size of the covariance matrix that needs to be estimated and inverted with the HO and the fact that IO is nonlinear and its performance is difficult to estimate in cases where the data statistics are not known

analytically.

The HO uses the first- and second-order statistics of the feature vectors to compute the template [32]. When the feature vectors are MVN distributed and the vectors from the two classes are homoscedastic, the first- and second-order statistics fully describe the statistical distribution of the channel outputs, the performance of the HO is equivalent to the IO. The MVN and homoscedasticity conditions are often assumed to hold for SKE/BKE tasks due to the central limit theorem provided that the channels are carefully chosen. Homoscedasticity is usually satisfied for SKE/BKS tasks. However, for more clinically realistic SKS/BKS tasks, the signal size, contrast, location, and background can vary. As a result, the feature vector distributions can be very non-MVN and even multi-modal [51]. The conventional HO may not be optimal for non-MVN distributed feature vectors, and especially multi-modal feature vector distributions. Eckstein et al. [52-54] proposed a multi-template strategy to handle SKS and signal known exactly but variable (SKEV) tasks. However, these strategies are computationally expensive to implement when the number of possible signal types is large. Further, these strategies are not rigorously theoretically justified.

In Chapter 4, we propose two multi-template linear observer strategies to handle the non-MVN data in clinically realistic SKS tasks using a sub-ensemble-based analysis of the data.

2.4.3 Figures-of-merit

A proper figure-of-merit (FOM) is a scalar that can be used to rank the performance on a task. Having a general and relevant FOM is thus essential for application of task-based image quality evaluation. In the following, we describe figures of merit relevant for the binary and three-class classification tasks used in this work.

2.4.3.1 Binary classification task

In a binary classification task, the FOM is modeled as follows. First, the observer generates a test statistic, a scalar that indicates the likelihood (or confidence) that the defect is present. Next,

the test statistic is compared to a decision threshold: test statistic values below the decision threshold indicate that the signal is absent; values above the threshold indicate that the signal is present. This case is decisions to the left of the dashed vertical line, representing the threshold, in Figure 2-4. The value of the threshold needed to produce a decision that minimizes the Bayes' Risk (or maximizes the utility of the decision) depends on the utility of each of the decision outcomes and the prevalence of the classes. In this context, utility refers to the benefit or cost of a correct or incorrect decision; a decision outcome can be a false positive (signal is actually absent; decision is signal present), false negative (signal is actually present; decision is signal absent), true positive (signal is actually present; decision is signal present) or true negative (signal is actually absent; decision is signal present). Since it is very difficult to assign utilities to decision outcomes, and the prevalence is not known, it is highly desirable to use FOMs that are independent of prevalence and outcome utilities. Note that the distribution of test statistics under the two hypotheses will typically overlap. The fractions (with respect to the total number of cases) of true positives and false negatives are the true position fraction (TPF) and the false negative fraction (FNF), respectively. Similarly, the fractions of true negatives and false positives are the true negative (TNF) and the false positive fractions (FPF), respectively.

A common way of deriving a prevalence and utility independent FOM is to consider the tradeoff between the TPF and FPF. The TPF is typically referred to as the sensitivity and $1 - \text{FPF}$ is known as the specificity. Ideally, the TPF would be unity and the FPF would be zero (sensitivity equal to 1). However, when the distributions of test statistics overlap there is no decision threshold where this is possible. As shown on the left in Figure 2-4, if we move the threshold to the left the TPF and FPF increase. Thus, the sensitivity will increase and the specificity will decrease. The opposite scenario happens when we move the threshold to the right. In other words, sensitivity and specificity depend on the threshold. However, the choice of threshold depends on the utility of each of the decision outcomes and the prevalence of the classes [55].

A way to quantify task performance for all prevalence and utility combinations is to use receiver operating characteristic (ROC) analysis [55]. The ROC curve is generated by plotting the TPF relative to FPF values. Starting at (TPF, FPF) = (0, 0), the curve is traced as the decision threshold moves from $-\infty$ to ∞ , as shown on the right in Figure 2-4. This curve can be obtained by sweeping the decision threshold from the left to the right.

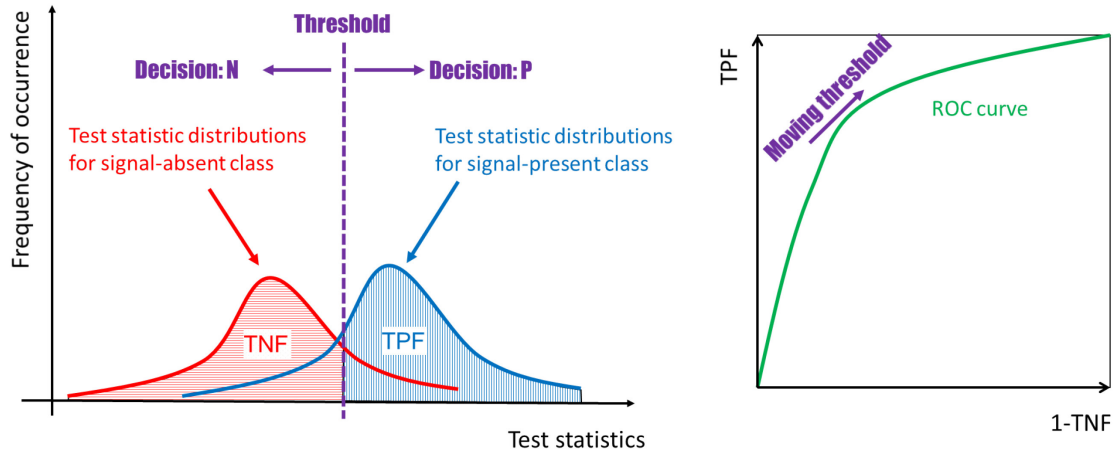


Figure 2-4. An example of test statistic distribution, decision making, and ROC analysis.

ROC analysis has been widely used to characterize task performance because it is independent of prior information (i.e., prevalence and utilities). An imaging system provides better task performance for all prevalence and utility combinations when its ROC curve is higher than the ROC curve for another system for all specificities. The area under the ROC curve (AUC) is thus often used as a FOM for binary classification task performance: in the case that the two ROC curves do not cross, a method with a higher AUC is better than one with a lower AUC in terms of binary classification performance independent of utilities and prevalence.

2.4.3.2 Three-class classification task

Three-class ROC analysis is an extension of binary classification ROC analysis. In the 3-class task, the observer sorts images in one of three categories. This is modeled using a pair of log likelihood ratios, defined by equations (2-13) and (2-14). The 3-class ROC surface is generated by sweeping the decision structure across the decision plane as shown in Figure 2-5 (a) (from Fig.

1 in [38]) and calculating the three true class fractions (T1F, T2F, and T3F, where TnF indicates the fraction of cases that were actually from class n and the decision was class n). The decision structure is defined by 3 rays with fixed slopes (one parallel to the horizontal axis, one parallel to the vertical axis, and one with slope 1) and the critical point, i.e., the intersection of the 3 rays in the decision structure. A sample 3-class ROC surface is shown in Figure 2-5 (b). The volume under the 3-class ROC surface serves as a FOM for 3-class task performance. It should be noted that this method has not been shown to be completely general and is rigorously valid only in cases where the utilities of the errors (e.g., deciding an image belongs to class m when it actually belongs to class n and $n \neq m$) are equal. More details on 3-class ROC analysis can be found in [35].

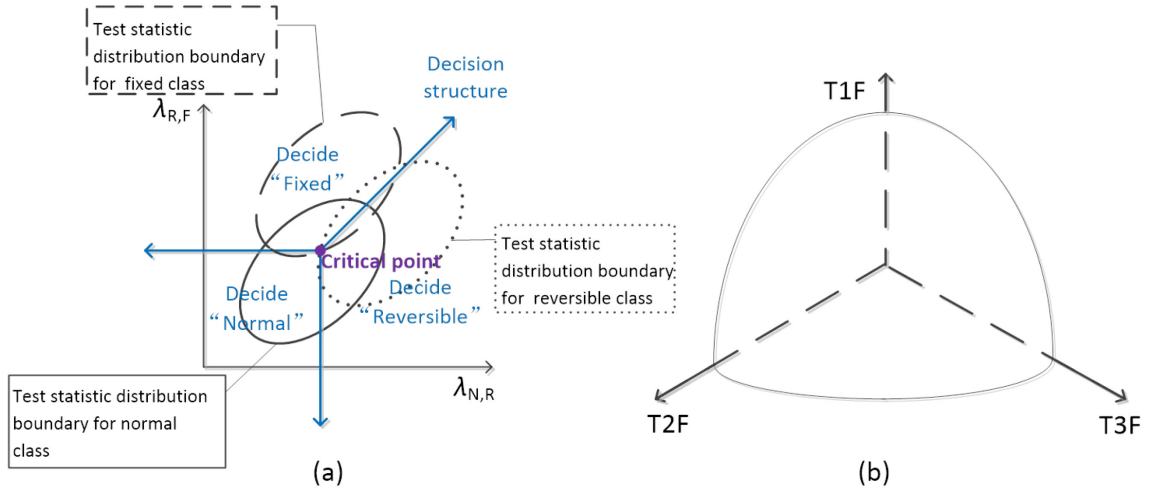


Figure 2-5. (a) Decision variable and the decision structure for the practical 3-class ROC analysis method. (b) Sample 3-class ROC surface.

2.5 Patient-specific optimization

In the era of “precision medicine”, we are interested in tailoring medical procedures (e.g., diagnostic techniques, treatment plans, etc.) to different patient or patient groups to improve patient outcomes. In this dissertation, not only are we interested in finding more optimal reconstruction methods and parameters, we are also interested in making the overall results more accurate by tailoring the parameters to the patient. Our hypothesis is that patient-specific reconstruction parameters will result in better overall diagnostic task performance. This

hypothesis is based on previous studies indicating that the optimal reconstruction method depends on patient specific parameters such as the spatial frequency power spectrum of the object and the count density of the image [56]. In this dissertation, we proposed patient-specific optimization methods for the two advanced iterative image reconstruction methods mentioned in 2.4. These methods use information about the patient that can be estimated from the measured data.

3. Common Methods: phantom population and projection data simulation

3.1 Introduction

In order to evaluate and optimize the reconstruction algorithm for dual-isotope MPS, it is essential to use data as realistic as possible. In such study, a population of patients that includes a wide range of patient, heart, and defect sizes and organ uptake variations is desirable. However, it is impractical to use real patient data because of the difficulty of obtaining the true perfusion status. It is equally impractical to use physical phantoms because of the difficulty of creating realistically phantom anatomies in phantoms with fillable compartments. For example, the thickness of the material surrounding each fillable compartment results in a portion of the phantom that contains no radioactivity. In addition, there is substantial cost involved with creating a large number of physical phantoms, filling them with activity, and imaging them on real imaging systems.

Mathematical phantoms offer a practical solution to the above problem. In order to use digital phantoms as a substitute for real patients, it is vital for them to be as realistic as possible. Much work has been done toward this goal. The three-dimensional (3D) eXtended CArdiac Torso (XCAT) phantom provides a model of human anatomy and physiology and has been widely used for medical imaging simulation [57-61]. The organ shapes are based on images from the Visible Human project [62] thus realistically modeled human anatomy. The XCAT phantom includes an enhanced heart model and models cardiac and respiratory motions, though the motion features were not used in this work. The 3D XCAT phantom has the important advantage of being flexible: it is possible to generate variations in anatomy and motion by adjusting the values of parameters that define the various organ shapes or motions.

In this work, we used a previously developed adult digital phantom population based on

the 3D XCAT phantom [63]. We generated clinically realistic projection data based on optimized MSP acquisition protocol [64]. The following provides details about the phantom population and projection data generation.

3.2 Methods

3.2.1 Population of realistic digital phantoms for MPS study

We used a digital phantom population that was specifically designed for dual-isotope MPS imaging research [65]. The population is comprised of 54 adult anatomies: all combinations of 2 genders, 3 body core sizes, 3 thickness of subcutaneous adipose tissue and 3 heart sizes. The phantoms were based on the 3D XCAT phantom, which realistically models human anatomy. For each anatomy, we modeled 6 defect types with several sizes, contrasts, and locations, as described below.

The population modeled 3 body core sizes: the standard torso and organ size of the XCAT phantom, one larger and smaller than the standard size. The standard XCAT parameters were created using a body core size that represents a 50th percentile (height and weight) male and female based on the PEOPLESIZE program [66]. The bigger and smaller body sizes were modeled by scale the torso and the standard organ sizes of the 3D XCAT phantoms to match patients with sizes 2 standard deviations above and below the average, based on the data in the Emory Cardiac Database [67].

The population also modeled 3 thickness of subcutaneous adipose tissue: the standard thickness of the XCAT phantom and thicknesses scaled by the factors 0.5 and 1.5 of the standard one. The ratio of the average body lateral to anterior-posterior sizes was set to 1.36 and 1.47 for male and female populations, respectively, matching the observed measurements in patient populations. Variations in subcutaneous adipose tissue thickness were studied as these result in different levels of attenuation of photons emitted from the myocardium, and were thus expected to have an effect on myocardial defect detectability.

Three heart sizes, small, medium, and large, were modeled for each body size by changing the XCAT heart parameters. The medium heart size was the standard XCAT phantom heart size for each body size. The large and small heart sizes were generated by scaling the heart size parameters to sizes that were one standard deviation above and below the standard size based on sizes reported in the Emory Cardiac Database. The ratios of the length to radius for the left ventricle (LV) were set to 3.2 and 3.17 for the male and female population, respectively, matching the observed measurements in the patient populations. The heart sizes were varied as they were expected to have an effect on defect detectability.

We modeled 6 defect types including 2 defect locations in the myocardium, anterior and inferior, with 3 severities (defined as the defect to normal myocardium activity ratio) and extents (defined as the volume percentage of the defect with respect to the total volume of myocardial tissue). The defect parameters are summarized in Table 3-1 (from table II in [68]). Note that in all cases the product of extent and severity was constant, representing a constant total reduction in myocardial uptake. An extent of 5% and a severity of 10% represent values that are difficult extent and severity of clinical interest. Sample images of the six defect types are shown in Figure 3-1.

Table 3-1. Defect parameters.

Defect type	Location	Extent (%)	Severity (%)
1	Anterior	5	50
2	Inferior	5	50
3	Anterior	10	25
4	Inferior	10	25
5	Anterior	25	10
6	Inferior	25	10



Figure 3-1. Sample images of six defect types 1-6 (from left to right) from the short axis view of the myocardium; the red arrows indicate defect locations.

The phantoms and defects were digitized into 0.22-cm voxels in a 256x256x228 matrix before projection simulation.

3.2.2 Realistic projection simulation with optimized imaging protocols

Low-noise projections were generated using the SimSET MC code [69] in conjunction with the angular response function (ARF) method [20]. This combination offers fast and accurate simulation both of interactions of photons with the collimator-detector system and the patient.

In the simulation, a GE Infinia dual-head SPECT system was modeled. The detector system has two large field-of-view (LFOV) NaI(Tl) scintillation crystal detectors. The thickness of the crystal was 9.5 mm. Each of the detectors was coupled to a GE low-energy high-resolution (LEHR) collimator. The projections were generated at 60 views over 180° from left posterior oblique to right anterior oblique modeling a body-contouring orbit. The projection bin size was 0.442 cm. The acquisition energy window was 20% centered at 140.5 keV for Tc-99m and 28% centered at 72 keV for Tl-201. A large number of photons were used during the simulation to generate low noise data. Attenuation, scatter, collimator-detector response and crosstalk between the two isotopes were all modeled. The low-noise projections were scaled to a count level corresponding to injected activities of 10 mCi of Tc-99m and 2 mCi of Tl-201, which are among the optimal injected activity values suggested in [18].

The low-noise projections were generated separately for the heart, liver, lung, blood pool, gall bladder, kidney and background (the rest of the body) in order to allow modeling of various uptakes in the organs by scaling and summing the organ projections. A total of 20 random uptake realizations each for Tc-99m and Tl-201, based on organ uptake distributions obtained from patient data, were generated for each anatomy by appropriately scaling the organ projections. Poisson noise was then added to the scaled and summed projection images.

3.3 Summary

In this work, we used a previously designed digital phantom population specifically designed for MPS research. The phantom population was based on realistic XCAT phantom and included variations in gender, body core size, heart size, and subcutaneous adipose tissue thickness. We modeled 6 defect types including various defect locations, severities, and extent with clinical difficult cases modeled. Projection data with realistic models and optimal injected activities for both isotopes were generated. We used this data to evaluate and optimize the reconstruction methods for dual-isotope MPS study.

4. A sub-ensemble-based approach to evaluate detection task performance for non-multivariate-normally-distributed data using multi-template linear observer strategies

4.1 Introduction

In medical imaging, image quality is objectively assessed in terms of the performance of an observer on a task of interest [70]. Relevant tasks include classification and estimation tasks. For binary classification (i.e., detection) tasks with known signal locations, performance can be characterized by the receiver operating characteristics (ROC) curve. The area under the ROC curve (AUC) is an often-used figure of merit for detection tasks.

Several model observers have been formulated for objectively evaluating image quality on detection tasks. The ideal observer (IO) uses all the statistical information about the data and yields the maximum AUC of all possible observers [31, 71]. However, the IO is often complicated and difficult to compute, and may not predict human observer performance. Therefore, linear observers, in particular, the Hotelling observer (HO), have been widely used [36]. The HO uses the first- and second-order statistics of the image data to compute an observer template. When the data follow a multivariate normal (MVN) distribution with equal covariances under the signal-absent and signal-present hypotheses, the HO has equivalent performance to the IO.

The channelized version of the HO referred to as the CHO, is also a commonly used model observer. By using appropriate anthropomorphic channels that model the human visual system, the CHO has been shown to agree well with the human performance for signal known exactly (SKE)/background known exactly (BKE) tasks [42, 43] and SKE-lumpy backgrounds tasks [44].

For SKE tasks where the channel outputs, often referred to as feature vectors, are MVN distributed and homoscedastic (i.e., when the two classes have equal covariance matrices), the CHO is theoretically optimal [34]. The CHO has been widely used to predict human performance for SKE tasks [43, 72, 73].

Clinical tasks have variability in both signal and background. The signal and the background may be known only statistically, thus resulting in a signal known statistically (SKS)/background known statistically (BKS) task. In the context of nuclear medicine imaging, signals vary in tracer uptake, size, and position. Similarly, variations in normal organ and tissue size, shape and uptake result in background variability. In recent work, we have shown that these kinds of variation can result in very non-MVN and even multi-modal [51] distributions of channel outputs.

It is thus desirable to have observer strategies that can handle both SKS and BKS data. The IO can, in principle, optimally treat non-MVN data. Since the non-MVN data (e.g., feature vectors) from an SKS task can be viewed as a combination of the data from all possible SKE tasks, the feature vectors in an SKS task can be statistically described using, for example, a Gaussian mixture model. However, in that scenario, the IO is still mathematically difficult to evaluate. As a result, observer strategies based on linear observers have been proposed to handle such data based on different approaches.

The HO has frequently been applied to data (feature vectors) with signal and background variability [74-77]. Since one HO template is applied to all the images in the ensemble, we refer to this as a single-template HO strategy. However, for non-MVN distributed data, the application of the HO to these tasks can be problematic [78].

Previous authors have proposed and applied a multi-template observer strategy to the data in SKS task [52, 53, 79-81]. In this strategy, a template (i.e., an instance of the linear observer whose dot product with an input data item yields the test statistic for that item) is generated for

each possible realization of the signal. The test statistics for each image are obtained by applying all templates to the image data (projection data or feature vector) and are then combined using the optimal sum of likelihood rule [82]. Note that this observer strategy is not equivalent to the IO, and it is applicable only to the case where there are a finite number of signal types. In these studies, empirical evidence was provided for task performance correlations between human observers and this model observer strategy for the SKS task. This strategy is computationally expensive compared to a single-template observer strategy, especially when the number of possible signal types is very large.

Eckstein et al. [52, 54] have demonstrated that human performance on an SKS task can be approximated by the performance of a simplified signal known exactly but variable (SKEV) task for the range of signal type variations in their study. An SKEV task is one in which the signals vary from image to image but the observers know the exact signal type present in the image. They also proposed a multi-template strategy for SKEV tasks. Again, this strategy can still be computationally expensive when the number of possible signal types is very large.

In this Chapter, we describe two multi-template strategies to perform detection tasks when the data are non-MVN distributed. It should be noted that the focus is not on providing a general solution to SKS or SKEV tasks, but on handling the problem of non-MVN data, which can arise due to signal or background variability. The strategies we provide are based on a sub-ensemble-based approach: the non-MVN data are divided into sub-ensembles with MVN-distributed and homoscedastic data. A different observer template is estimated and applied for each such sub-ensemble. First, we adapt a multi-template HO strategy initially proposed by Eckstein et al. for SKEV tasks [52, 54]. We also propose a novel multi-template linear discriminant strategy and discuss the theoretical motivation and the optimality of the classification performance in terms of the AUC for this strategy. We compared these strategies with the commonly used computationally-inexpensive single-template HO strategy. We applied these strategies to

optimize the reconstruction parameters for a defect detection task performed on a realistic dual isotope myocardial perfusion SPECT (MPS) simulated dataset [63]. The channel outputs from the images in the study are non-MVN. The three strategies were compared in terms of AUC and range of optimal parameters obtained from the three strategies.

4.2 Theory

In this section, we provide two multi-template linear observer strategies to handle non-MVN distributed data using a sub-ensemble-based approach. We will first discuss the sub-ensemble-based approach and then describe the two strategies.

4.2.1 Sub-ensemble-based approach

The proposed method for handling non-MVN distributed data is to divide the data into sub-ensembles that are exactly or approximately MVN distributed and homoscedastic. We can then apply optimal linear observers to each sub-ensemble and use the test statistics to compute a figure of merit. In this Chapter, we provide two such observer strategies, as described in 4.2.2 and 4.2.3.

The sub-ensemble partitioning method that provides subsets of data that meet the MVN and homoscedasticity conditions may not be unique, and we do not propose a general way to achieve this condition. Here, we provide two feasible partitioning methods that were applicable to the data used in this work.

The MVN and homoscedasticity conditions are often, though not always, satisfied when each sub-ensemble represents an SKE task. However, this SKE sub-ensemble approach is not always practical due to the large number of possible SKE tasks. For example, in our clinically-realistic dataset described later, the possible number of SKE tasks was 861840. Estimating observer templates for such a large number of SKE tasks requires generating millions of images and is thus highly impractical.

In a previous study [78], it was observed that an SKS task might also have approximately

MVN distributed data (feature vectors) when the signal and background variations are sampled from a continuous distribution. Thus, each sub-ensemble can also be an SKS task that is approximately MVN. The homoscedasticity condition can be approximately satisfied if the signal variations are relatively small compared to the background variations and the signal and background are uncorrelated, as proved in 4.7.1. Based on these observations, a feasible partition method is to divide the data into groups with signal variations smaller in comparison to background variations and with variations sampled from a continuous distribution. Note that in such sub-ensembles, the signal is known only statistically. Compared to partitioning data into SKE sub-ensembles, this method has the advantage of reducing the number of sub-ensembles and thus reducing the time and number of images required to estimate all the observer templates. An example of this partitioning method is presented in Section 4.3.2.

When each sub-ensemble is an SKE task, the entire task is SKEV. If, on the other hand, when the sub-ensembles are SKS, as in the partitioning method described above, then the overall task is SKS. It should be noted, however, that applying a linear observer to an SKS may not model ideal observer performance. Thus, in the case where the sub-ensembles are SKS, as is the case with the data presented here, the resulting observer is only optimal in a restricted sense.

4.2.2 Multi-template linear observer strategy 1

We adapted the previous multi-template observer strategy for SKEV tasks proposed by Eckstein et al. [52, 54] and applied it to the sub-ensemble based data as described below. In this strategy, a different HO observer template was estimated for each sub-ensemble, the AUC was computed for each sub-ensemble, and the AUC for the entire dataset was the weighted sum of AUCs for all sub-ensembles. The weight applied to the AUC for each sub-ensemble was the fraction of cases in each sub-ensemble to the total cases in the entire ensemble. This strategy is referred to as the multi-template HO with averaged AUCs strategy. Note the AUC used here is equivalent to the percent correct used in the strategy proposed by Eckstein et al. because they are

equivalent for any two-alternative forced choice (2AFC) experiment [32]. In addition, when the each sub-ensemble is from an SKE task, this strategy is the same as the observer strategy for SKEV task proposed by Eckstein et al.

It should be noted that averaging the AUCs from different tasks is not theoretically rigorously justified. Averaging the AUCs is equivalent to averaging the ROC curves, which is equivalent to averaging the true positive fraction (TPF) for each false positive fraction (FPF). The exact meaning of averaging the TPFs for a given FPF is difficult to define.

4.2.3 Multi-template linear observer strategy 2

In this section, we propose a novel multi-template linear observer strategy. We first present the motivation for this strategy and then discuss the properties and optimality.

4.2.3.1 Motivation: The Relation between Hotelling Observer and likelihood ratios

This section examines the relationship between the HO test statistics and likelihood ratios. This relationship is the justification for the proposed strategy, as will become clear below.

Consider the task of classifying an object into signal-absent or signal-present classes based on some measurement, denoted by a data vector \mathbf{g} . The data vectors can be projection data or feature vectors, such as vectors of channel outputs obtained from a reconstructed image. We denote the signal-absent and signal-present hypotheses by H_0 and H_1 , respectively. If \mathbf{g} follows an MVN distribution under both hypotheses, we can describe its probability distribution under both hypotheses using:

$$\text{pr}(\mathbf{g} | H_i) = \frac{1}{(2\pi)^{M/2} \sqrt{|\mathbf{K}_i|}} \exp \left[-\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}_i)^T \mathbf{K}_i^{-1} (\mathbf{g} - \bar{\mathbf{g}}_i) \right], \quad (4-1)$$

where $\bar{\mathbf{g}}_i$ and \mathbf{K}_i denote, respectively, the mean data vector and covariance matrix of the data under hypothesis H_i ($i = 0, 1$), $|\mathbf{K}_i|$ is the determinant of matrix \mathbf{K}_i , M denotes the dimension of the data vector \mathbf{g} , and T denotes vector transpose operation.

If $\mathbf{K}_1 \approx \mathbf{K}_2$, and we write the covariance matrix under both hypothesis as \mathbf{K}_g , then the log of the ratio of $\text{pr}(\mathbf{g} | H_1)$ and $\text{pr}(\mathbf{g} | H_0)$, referred to as the log of the likelihood ratio, can be written as

$$\lambda(\mathbf{g}) \approx \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \mathbf{g} + \frac{1}{2} \left(\bar{\mathbf{g}}_0^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_0 - \bar{\mathbf{g}}_1^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_1 \right), \quad (4-2)$$

where $\Delta \bar{\mathbf{g}} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0$. Note that using any monotonic transformation of the likelihood ratio (e.g. the log of the likelihood ratio) as the decision variable maximizes the classification performance in the sense that Bayes risk is minimized [33, 83, 84]. The resulting observer is referred to as the IO. Ignoring terms that are independent of \mathbf{g} yields the following expression, which gives the Hotelling test statistic, $\lambda_{\text{HO}}(\mathbf{g})$:

$$\lambda_{\text{HO}}(\mathbf{g}) = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \mathbf{g}. \quad (4-3)$$

The term $\Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1}$ is the Hotelling template. The term in (4-2) that was ignored is:

$$\eta = \frac{1}{2} \left(\bar{\mathbf{g}}_0^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_0 - \bar{\mathbf{g}}_1^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_1 \right), \quad (4-4)$$

and, as noted, is independent of the input data, \mathbf{g} .

Note that addition of the term defined in (4-4) to an observer template does not affect the value of the AUC obtained using that observer. Thus, in the scenario where the same template is applied to the entire dataset, the template as defined in (4-3) is used instead of that defined in (4-2). However, if a multi-template observer strategy is used, i.e., each sub-ensemble of data are treated using a different observer template, then (4-4) is different for different sub-ensembles since the terms $\bar{\mathbf{g}}_0$, $\bar{\mathbf{g}}_1$ and \mathbf{K}_g depend on the signal and background statistics for each different sub-ensemble.

4.2.3.2 Proposed strategy: multi-template linear discriminant (LD) with pooled test statistics

Based on the fact that the term defined in (4-4) is, in general, different for different sub-ensembles, we propose a multi-template observer strategy: For each sub-ensemble, we include the term defined in (4-4) in the test statistics, i.e., we use (4-2). The resulting test statistics for each sub-ensemble are pooled to calculate the AUCs for the entire dataset. Since equation (4-2) is referred to as the Linear Discriminant [85-88] (LD), we refer to the proposed strategy as the multi-template LD with pooled test statistics strategy. We now provide the theoretical justification for using this strategy.

4.2.3.3 Properties of the multi-template LD observer

In this section, we prove that adding the term (4-4) to the HO maximizes classification performance for SKEV tasks in terms of the AUC when shifting the HO test-statistics by a different constant for each sub-ensemble is allowed. This can be proved using the following two theorems.

Theorem 1. Consider a dataset for an SKEV task that can be grouped into multiple sub-ensembles, where each sub-ensemble is from a sub-SKE task in the SKEV task and the data are MVN distributed and homoscedastic. Consider applying different linear observer templates to each sub-ensemble, each of which yields a test statistic. If only shifting of these test statistics by an input-data-independent term is allowed, the AUC of the pooled test statistics is maximized when the distribution functions of the test statistics under the two hypotheses cross (i.e., have the same probability density) at the same test statistic value for all sub-ensembles.

The proof is as follows. Any linear observer can be defined by a template \mathbf{w} , such that, when applied to the input data \mathbf{g} , it yields a test statistic $\lambda(\mathbf{g})$, given by

$$\lambda(\mathbf{g}) = \mathbf{w}^T \mathbf{g}. \quad (4-5)$$

The test statistics of a linear observer are a linear combination of the values in the data vector. Thus, if \mathbf{g} is MVN, the test statistics will be normally distributed [89, 90]. Also, from (4-5), if the distributions of the input data vectors are homoscedastic under the two hypotheses, then the distribution of test statistics of a linear observer under the two hypotheses will have the same variance. Thus, if the input vectors under both hypotheses are MVN and homoscedastic, the test statistics will be normally distributed and homoscedastic. The MVN condition is usually satisfied in an SKE task and the homoscedastic condition can be approximately satisfied when the signal size and contrast is relatively small.

Now suppose that an SKEV task is composed of N sub-SKE tasks ($N \geq 2$) and the sub-ensemble of data from each sub-SKE task does have these properties. In this case, the test statistics obtained from each sub-ensemble will be normal and homoscedastic. To describe this mathematically, for the j th sub-ensemble, denote the standard deviation for the linear observer test statistics under the two classes by σ_j , the crossing point of the distributions under the two classes by d_j , and the means by $d_j - \alpha_j$ and $d_j + \beta_j$ for signal absent and present classes, respectively. Without loss of generality, assume $\alpha_j, \beta_j > 0$. Since the test statistics for the sub-ensemble are normally distributed and homoscedastic, $\alpha_j = \beta_j$. Define $\mu_j = \alpha_j = \beta_j$ ($\mu_j > 0$). These symbols are illustrated in Figure 4-1 for the case of N sub-ensembles.

Suppose the test statistic distributions for the j th sub-ensemble are shifted by Δd_j , where Δd_j is independent of the input data, \mathbf{g} . Denote $c_j \equiv d_j + \Delta d_j$. Then, as proven in Section 4.7.2, when $c_j = c_1 (\forall j \in [2, N])$, as shown in Figure 4-2, the pooled test statistics from the N sub-ensembles achieve the highest AUC.

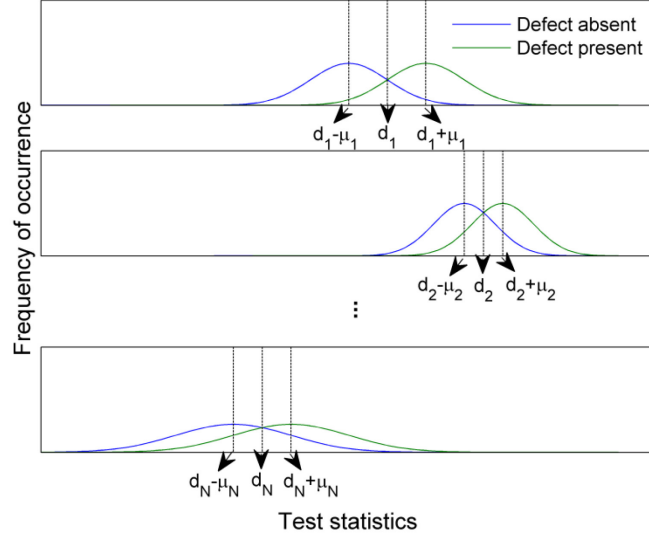


Figure 4-1. Illustration of linear observer test statistic distributions for all sub-ensembles before shifting.

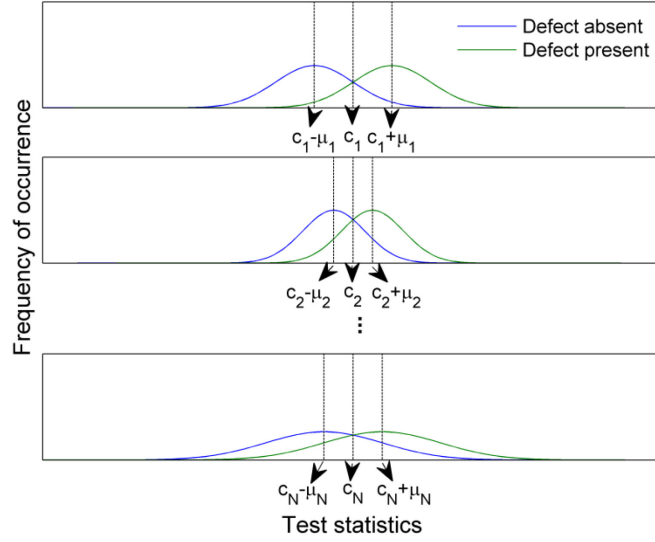


Figure 4-2. Illustration of linear observer test statistic distributions for all sub-ensembles after shifting.

Theorem 2. When the sub-ensembles of data are MVN distributed and homoscedastic, if the LD is used to generate the test statistics for each sub-ensemble, the distributions of the test statistics under the two hypotheses for all sub-ensembles cross at test statistic $\lambda=0$.

Rephrasing theorem 2 mathematically, for the j th sub-ensemble, when $\Delta d_j = \frac{1}{2} \left(\bar{\mathbf{g}}_{j0}^T \mathbf{K}_{\mathbf{g}j}^{-1} \bar{\mathbf{g}}_{j0} - \bar{\mathbf{g}}_{j1}^T \mathbf{K}_{\mathbf{g}j}^{-1} \bar{\mathbf{g}}_{j1} \right)$ (as defined by (4-4)), where $\bar{\mathbf{g}}_{ji}$ and $\mathbf{K}_{\mathbf{g}j}$ are, respectively, the mean data vector and covariance matrix of the data under hypothesis H_i ($i = 0, 1$), then $c_j = 0$. In other words, LD accomplishes the alignment of the test statistic distributions of different sub-

ensembles, as shown in Figure 4-2. This theorem is proved in Section 4.7.3.

From Theorem 1 and 2, we can conclude that when the input ensemble of data vectors can be separated into sub-ensembles that are MVN and homoscedastic, using the proposed multi-template LD strategy maximizes the AUC if only shifting of the HO test statistics by a different constant for each sub-ensemble is allowed prior to pooling them.

4.3 Method

4.3.1 Image reconstruction and post-processing

Images were reconstructed using the ordered subsets-expectation maximization (OS-EM) [26] algorithm with compensation for attenuation, collimator-detector response, scatter, and crosstalk contamination between projection data from the two radionuclides. Scatter compensation was based on the effective source scatter estimation (ESSE) method [19]. For each isotope, we used the true noise-free crosstalk projection data from the other in the crosstalk compensation, modeling an ideal crosstalk compensation method. We used four subsets per iteration for both Tc-99m and Tl-201 and evaluated images obtained after iterations 1, 2, 3, 5, 7, 10, 15, 20, 30, 45 and 60 for Tc-99m and 1, 2, 3, 5, 7, 10, 15 and 20 for Tl-201. After reconstruction, images were filtered using a Butterworth filter of order 8 and cutoff frequencies 0.08, 0.1, 0.12, 0.14, 0.16, 0.2 and 0.24 pixels⁻¹. The filtered images were then reoriented to short axis slices. A 64×64 image having the centroid of the defect at the center of the image was extracted, windowed so that the range [0, maximum in the heart] was mapped to the range [0,255], truncated to integers, and used in the observer studies.

4.3.2 Implementation and evaluation of observer strategies

We evaluated three observer strategies by applying them to optimize the reconstruction parameters for dual-isotope MPS study.

The first strategy was a conventional computationally-inexpensive HO strategy that used a

single HO template for the entire ensemble. This strategy is referred to as the single-template HO strategy. The implementation in this work was similar to that described in [77, 91]. First, feature vectors were calculated by applying six rotationally symmetric frequency channels. The first channel had a starting frequency and channel width of 1/128 cycles per pixel. Subsequent channels abutted the previous one and had double the previous width. We calculated the HO test statistics using a leave-one-out technique [77, 92]. In that technique, the HO was trained on an ensemble including all data except a single image. The resulting HO template was then applied to the remaining image to calculate a single test statistic. For each combination of iteration number and cutoff frequency, a single test statistic was calculated with 6,480 pairs of defect-present and -absent images (12,959 training images to estimate the template and 1 testing image as the input image data). This process was repeated with each image in the ensemble left out in turn, resulting in a number of test statistics equal to the number of images. ROC analysis [93] using the LABROC program [94] was then applied to this set of test statistics to estimate the AUC.

The other two strategies were the suggested multi-template strategies in 4.2.2 and 4.2.3. We first divided the dataset into multiple sub-ensembles, where each sub-ensemble had the same defect type. As explained above, the absolute value of the defect volumes and uptakes were different in a sub-ensemble, leading to an SKS dataset. However, since the absolute activity in each organ was sampled from a continuous distribution, and the organ volume was relatively continuous, the signal and background variations in each sub-ensemble were sampled from a relatively continuous distribution. Thus, based on the observations in [78], we expected each SKS sub-ensemble to have an approximate MVN distribution. This agreed with empirical observations in this study, as shown in Figure 4-3. We note that the entire ensemble had a multi-modal distribution. However, the sub-ensemble following the above partitioning strategy was approximately normally distributed.

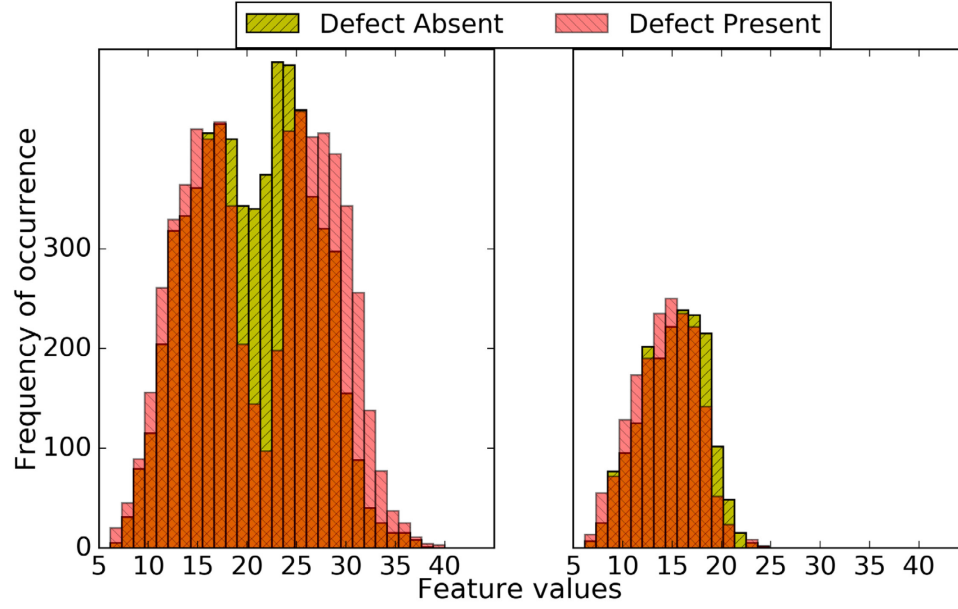


Figure 4-3. Thirty bin histograms of the first channel feature value distribution for the entire ensemble (left) and for the SKS sub-ensemble corresponding to defect type 1 and contains the mixture of anatomies in whole patient population (right) for Tc at iteration number 5 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1}

We also tested the homoscedasticity, i.e. the equality of the covariance matrix under the defect absent and defect present cases, for each of the sub-ensembles. Sample images of covariance matrices under the two hypotheses for the six sub-ensembles are shown in Figure 4-4.

Visually, the homoscedasticity condition was approximately satisfied.

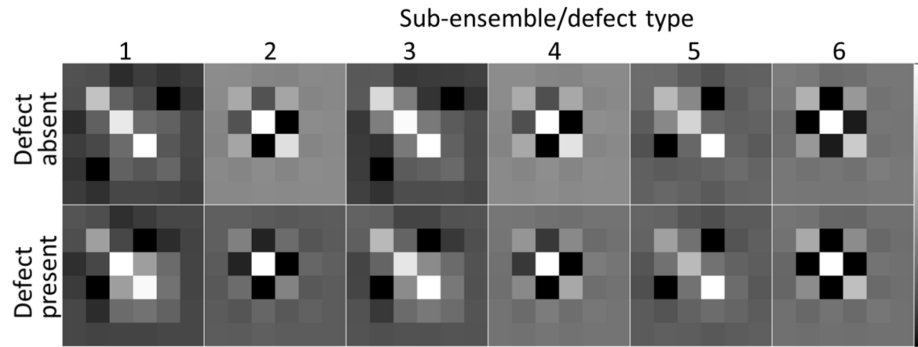


Figure 4-4. Covariance matrices under the two hypotheses (top row: defect-absent, bottom row: defect-present) for the six sub-ensembles (the 1st to the 6th columns are corresponding to defect type 1 to 6, respectively) for Tc at iteration number 5 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} .

We also tested the homoscedasticity condition based on two metrics. The first metric used

to measure the homoscedasticity is defined as [95]

$$CMD(\mathbf{K}_0, \mathbf{K}_1) = 1 - \frac{\text{tr}\{\mathbf{K}_0 \mathbf{K}_1\}}{\|\mathbf{K}_0\|_f \|\mathbf{K}_1\|_f}, \quad (4-6)$$

where $\|\cdot\|_f$ and $\text{tr}\{\cdot\}$ denote matrix Frobenius norm and trace. This metric is referred as Correlation Matrix Distance (CMD) in [95] and it measures the similarities between two positive definite matrices up to a scale. This value is between 0 and 1. It becomes zero when the two matrices are equal up to a scaling factor and increases as their extent of difference increases. The CMD values for all sub-ensembles and all reconstruction parameters for Tc and Tl are summarized in Table 4-1. The mean CMD for all sub-ensembles and all reconstruction parameters is close to 0, which indicate good similarities between the two covariance matrices up to a scale.

The second metric, defined as

$$\frac{\left| \frac{1}{2}(\mathbf{K}_0 + \mathbf{K}_1) \right|}{\sqrt{|\mathbf{K}_0| |\mathbf{K}_1|}}, \quad (4-7)$$

where $|\cdot|$ denotes matrix determinant. This metric is always greater than 1 if \mathbf{K}_0 and \mathbf{K}_1 are positive definite [96], which is true because covariance matrices are always positive definite. This metric measures the average scale difference of the two covariance matrices. We refer this metric as the determinant ratio. Suppose the two matrices are the same up to a scale m , i.e., $\mathbf{K}_0 = m\mathbf{K}_1$ ($m > 0$), the closer m is to 1, the closer this determinant ratio is to 1. The determinant ratio values for all sub-ensembles and all reconstruction parameters for Tc and Tl are summarized in Table 4-2. Since the mean of this determinant ratio is very close to 1, the average scale difference between the two covariance matrices for each sub-ensemble is small.

Table 4-1. CMD for all sub-ensembles and all reconstruction parameters.

	Mean	Minimum	Maximum
Tc	0.0293	0.0004	0.1488
Tl	0.0144	0.0005	0.0602

Table 4-2. Determination ratio for all sub-ensembles and all reconstruction parameters.

	Mean	Minimum	Maximum
Tc	1.25	1.01	2.19
Tl	1.08	1.01	1.25

Based on the above, the sub-ensembles are approximately homoscedastic. Thus, using this partitioning strategy by defect types yielded sub-ensembles that approximately satisfy the MVN and homoscedasticity conditions. The results in 4.4.1 also empirically indicate this.

For the multi-template LD with pooled test statistics strategy, we estimated the LD test statistics using the above leave-one-out strategy for each sub-ensemble. The resulting set of test statistics from all the sub-ensembles were pooled and used to compute the overall AUC. Thus, the ROC analysis was performed only once.

For the multi-template HO with averaged AUCs strategy, HO test statistics were obtained using the above leave-one-out strategy, the AUC was computed for each sub-ensemble, and the weighted sum of AUC for all sub-ensembles was calculated, where the weight was the fraction of cases in each sub-ensemble.

For each sub-ensemble and each combination of iteration number and cutoff frequency, a single test statistic was calculated with 1080 pairs of defect-present and -absent images (2159 training images and 1 testing image) for the two multi-template strategies.

A diagram illustrating the three observer strategies is shown in Figure 4-5.

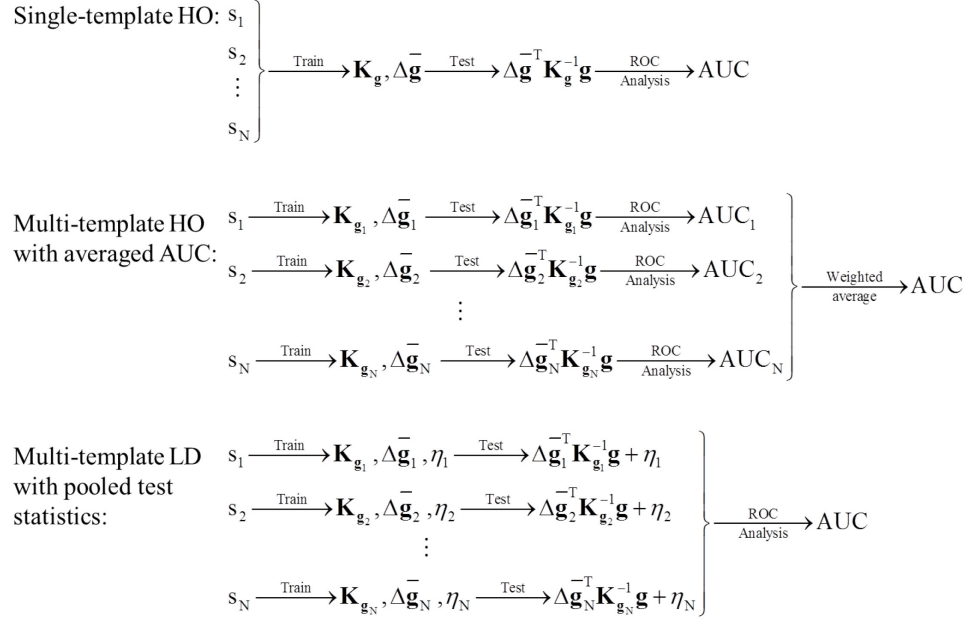


Figure 4-5. Diagram illustrating the data flow for the three observer strategies investigated. S_j stands for the j th sub-ensemble ($1 \leq j \leq N$), where N is the total number of sub-ensembles. $\Delta \bar{\mathbf{g}}_j$, \mathbf{K}_{g_j} , and η_j denote the mean data vector, the covariance matrix of the data, and the term defined in Equation (4-4) for the j th sub-ensemble.

The overall goal was to optimize the number of iterations of the OS-EM algorithm and the cutoff frequency of the post-reconstruction-low-pass-filter for Tc and Tl images. For all three strategies, the reconstruction parameters that achieved the highest AUC values for the entire ensemble were deemed optimal.

4.4 Results

4.4.1 Effect of using the LD observer instead of the HO observer for multiple sub-ensembles obtained using the partitioning method in 4.3.2

The effect of using the LD instead of the HO for the six sub-ensembles obtained using the partitioning method in 4.3.2 can be seen in Figure 4-6 and Figure 4-7. Figure 4-6 shows the test statistic distributions for the six different sub-ensembles using the HO. We observed that the ranges of the test statistic values for different sub-ensembles were different. The distributions for the LD are shown in Figure 4-7. Note that the histograms of test statistic of the two classes for all

sub-ensembles cross when the value of the test statistic is approximately zero (not strictly zero because the feature vectors in each sub-ensemble were not strictly MVN distributed and homoscedastic, and the histograms were generated from a finite number of samples). This indirectly indicates that sub-ensembles obtained using the partitioning method by defect type in this study approximately satisfies the MVN and homoscedasticity condition.

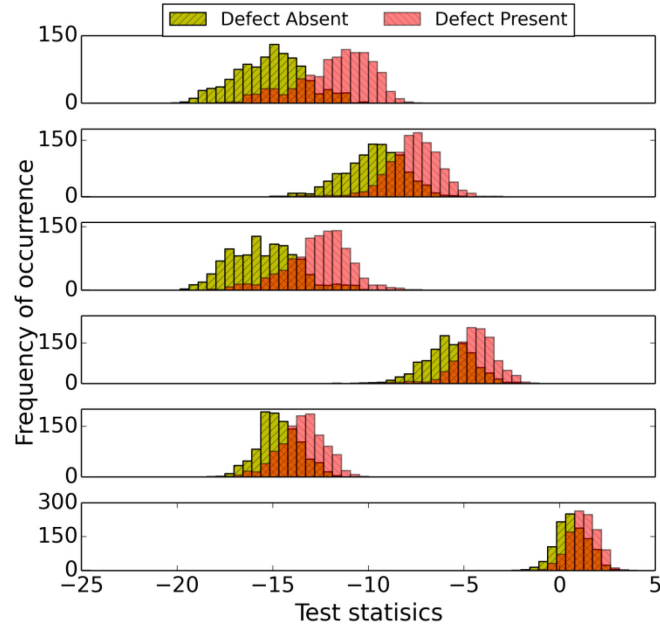


Figure 4-6. Sixty-four bin histograms of HO test statistics for different defect types for T_c at iteration number 1 (4 subsets/iteration) and cutoff frequency 0.1 pixel^{-1} . The graphs, from top to bottom, are for relative defect types 1-6, respectively.

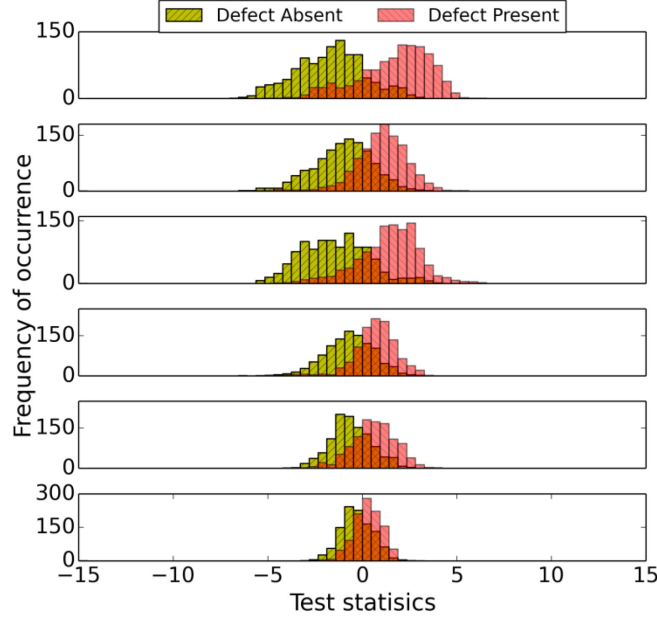


Figure 4-7. Sixty-four bin histograms of LD test statistics for different defect types for Tc at iteration number 1 (4 subsets/iteration) and cutoff frequency 0.1 pixel⁻¹. The graphs, from top to bottom, are for relative defect types 1-6, respectively.

4.4.2 Comparison of the AUCs using the three observer strategies

For each set of reconstruction parameters, we computed AUC values using the three observer strategies. The calculated AUC was highest for the multi-template LD with pooled test statistics strategy and lowest for the conventional single-template HO strategy for all the sets of reconstruction parameters investigated. The differences in the AUC values obtained using the multi-template LD with pooled test statistics strategy and single and multi-template HO strategy for Tc and Tl images are summarized in Table 4-3 and Table 4-4, respectively. The difference between multi-template strategies and single-template HO strategy is not surprising because the signal uncertainty in the entire dataset is larger than in the sub-ensembles.

The AUC values obtained using the optimal parameters for each observer strategy are shown in Figure 4-8. Note that the multi-template LD with pooled test statistics strategy gave the highest AUC value among the three strategies. The p -values for a two-tailed t -test of the differences between the three strategies estimated using bootstrapping were smaller than 0.05 for

both Tc and Tl, indicating that the differences were statistically significant. The differences in the AUC values between the strategies were all greater than 0.01. We considered a difference of 0.01 in the AUC to be clinically important since the AUC differences for images reconstructed with and without the scatter or detector response compensation were 0.01 in similar studies [97], and these combinations of compensations have been adopted and recognized as clinically significant. The differences between the three strategies were larger for Tc than Tl because the optimal parameters obtained by the three observer strategies were more different for Tc and more similar for Tl, as shown in section 4.4.3.

Table 4-3. Difference between AUC values for the multi-template LD with pooled test statistics and single-template HO strategies.

	Mean	Minimum	Maximum
Tc	0.054 ± 0.005	0.033 ± 0.004	0.082 ± 0.005
Tl	0.029 ± 0.006	0.024 ± 0.006	0.042 ± 0.006

Table 4-4. Difference between AUC values for the multi-template LD with pooled test statistics and the multi-template HO with averaged AUCs strategies.

	Mean	Minimum	Maximum
Tc	0.025 ± 0.004	0.005 ± 0.006	0.038 ± 0.005
Tl	0.015 ± 0.006	0.003 ± 0.006	0.023 ± 0.006

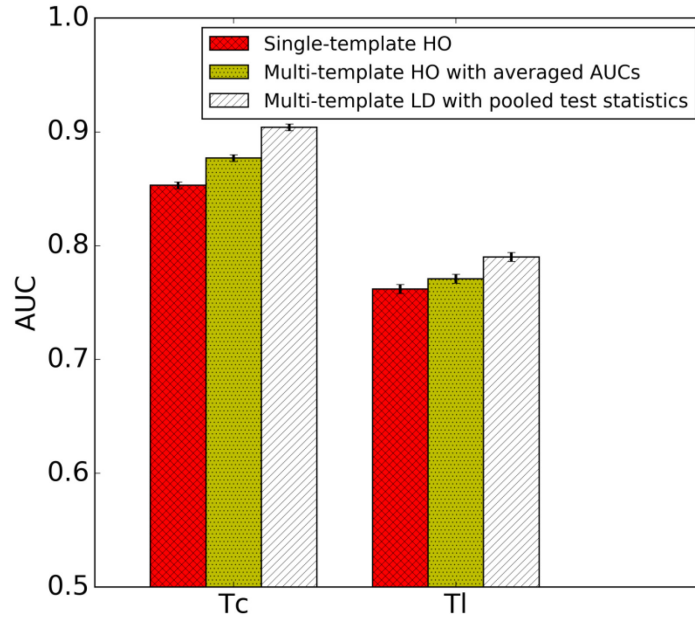


Figure 4-8. AUC values with the optimal parameters obtained by three observer strategies.

4.4.3 Comparison of the optimal parameters obtained by the three observer strategies

The optimal parameters for the entire dataset obtained by the three different observer strategies are shown in Figure 4-9 for Tc (left) and Tl (right). In these plots, the cross shows the parameter set that gave the highest AUC values; the filled circles shows sets of parameters where the difference in AUC values with respect to the optimal one was not statistically significant (p -value > 0.05); the contour line surrounds combinations of parameters for which the AUC values differed by 0.01 or less. Parameter combinations inside this curve were considered to be near-optimal.

For Tc, the two multi-template observer strategies were optimal for lower iteration numbers and higher cutoff frequencies than with the conventional HO strategy. One explanation is that a higher iteration number, which improves image resolution, is preferred by the single-template HO strategy when there is more background variability since higher resolution tends to reduce the effects on myocardial intensity of variations in anatomy and uptake in neighboring organs; a lower cutoff frequency was needed to reduce the noise when the iteration number was

high.

For Tl, the near-optimal parameter ranges using the three observer strategies were similar. This was different than for Tc and can be explained as follows.

The observer templates for the three strategies would be the same when the distributions of feature vector values were the same for different sub-ensembles. The positions of the feature vector distributions were largely determined by the pixel values near the defect position. For the anterior defect, this was largely the myocardium, as activity in neighboring structures was small. For the inferior defect, the liver also made a significant contribution because of its high uptake and proximity to the defect. Assuming that the contribution to the feature vector values from the myocardium in the two locations was the same, the position of the feature vector distributions for an anterior compared to an inferior defect was determined largely by the liver contribution. The activity in the liver relative to the myocardium for Tc was greater than for Tl by a factor of 1.29. Thus, there was a greater absolute shift in the positions of the distributions for Tc for the anterior versus inferior defect locations, as seen in Figure 4-10. The conventional HO observer is more sensitive to differences in the positions of the distributions test statistics from the two locations, for reasons described above. Thus the conventional HO strategy would be more different from the other two observers for Tc than for Tl, resulting in the possibility of differences in the optimal parameter combinations.

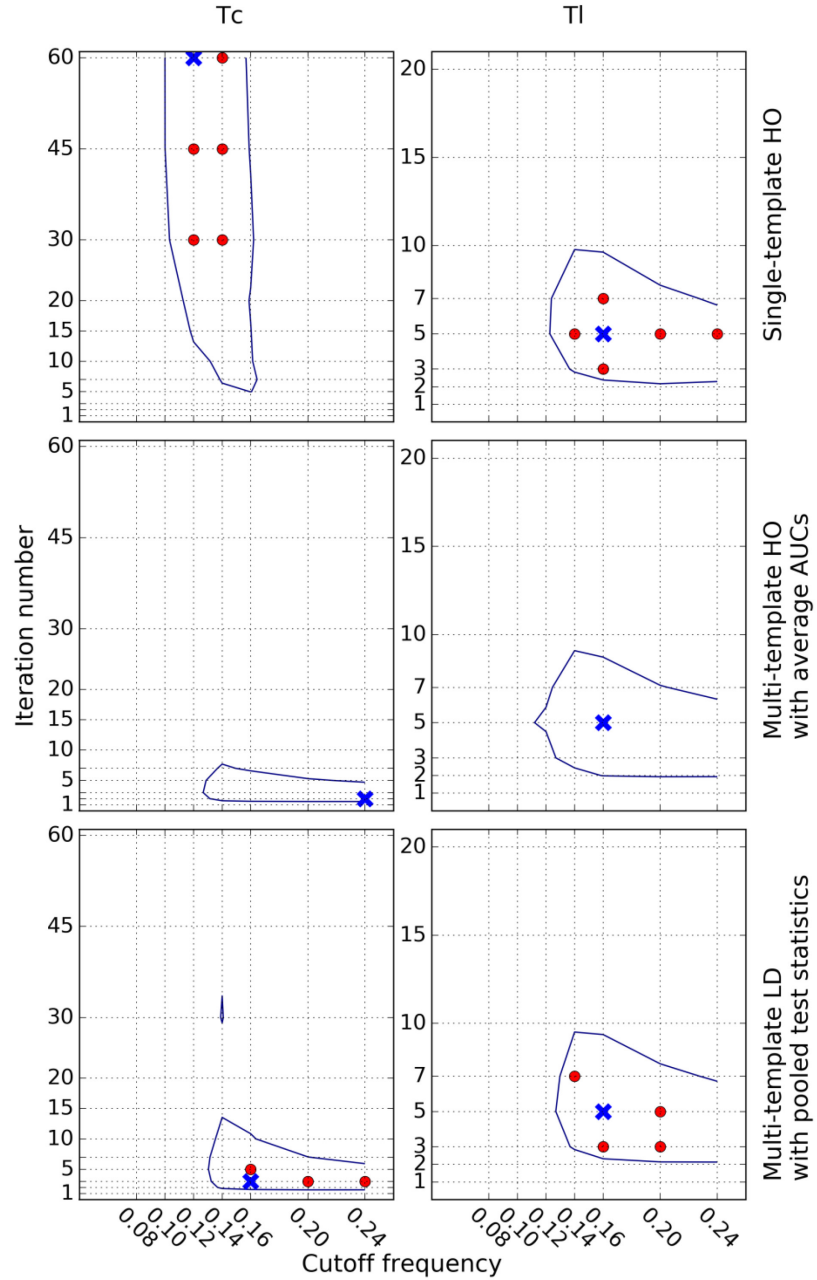


Figure 4-9. Optimal parameters for all defect types for Tc (left) and Tl (right) using (top to bottom): single-template HO, multi-template HO with averaged AUCs strategy, and multi-template LD with pooled test statistics strategies. The crosses represent the set of parameters that achieved the maximum AUC. The filled circles represent the parameter points where the difference in AUC with respect to the maximum AUC was not statistically significant. The contour line indicates the region where AUC values that differed from the optimal one by no more than 0.01, a difference considered clinically important. We used 4 subsets/iteration during the OS-EM reconstructions.

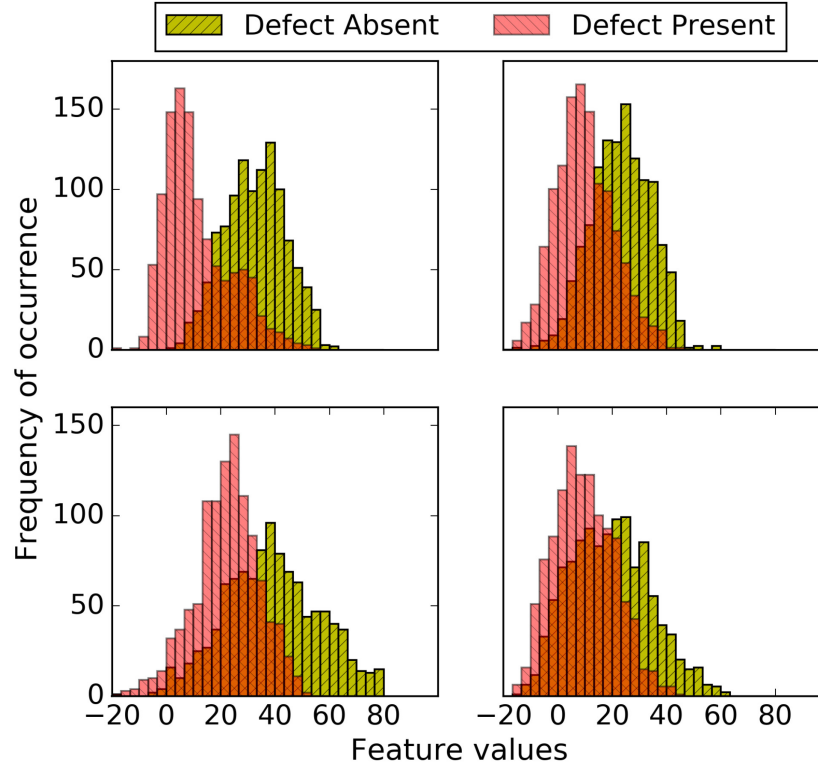


Figure 4-10. Thirty bin feature vector histograms of Channel 4 outputs for Tc (left) and Tl (right) for relative defect types 1 (top, anterior defect with 5% extent and 50% severity) and 2 (bottom, inferior defect with 5% extent and 50% severity) using 5 iterations (4 subsets/iteration) and a cutoff frequency of 0.1 pixel^{-1} .

Note that a similar argument applies to the defect types at the same location. In this case, the differences in the positions of the distributions were a function of the uptake in the myocardium. Since the myocardium had a Tc activity that was greater by a factor of 1.97 than Tl, the absolute difference in the positions of the feature vector distributions would be greater for Tc than for Tl. The conventional HO strategy would have greater difficulty dealing with this larger difference, and thus there would be a greater difference for the single-template strategy compared to the other two observers for Tc than for Tl.

To determine which observer strategy gave the truly optimal parameters for the combination of all defect types, we compared the performance for each relative defect type using the optimal parameters obtained by the three observer strategies. The results are shown in Table 4-5. Note that for each sub-ensemble (corresponding to the each relative defect type in this study),

the AUCs in Table 4-5 are calculated using HO observer method. For each relative defect type, the AUC values obtained with the optimal parameters from the multi-template LD with pooled test statistics strategy were always greater than or equal to the AUCs obtained using parameters optimal from the conventional HO strategy. The largest differences in AUC values between the multi-template and the single-template strategies were for relative defect types 5 and 6, the defects with the largest extent and lowest severity. This is likely because detecting relative defect types 1-4 which have higher contrast was easier (as shown in Figure 3-1), and the AUC was less sensitive to changes in the parameters. The AUC values for the two multi-template strategies were comparable: for some defect types our proposed strategy gives higher AUC but not for all defect types; for all mixed defect types, the two multi-templates strategies were comparable (as shown in Figure 4-9), the optimal parameters obtained by the two multi-template strategies were within the optimal parameter range of each other.

Table 4-5. Comparison of AUCs for Tc for individual sub-ensemble using the optimal parameters obtained using the three observer strategies.

Sub-ensembles	Single-template HO	Multi-template HO with averaged AUCs	Multi-template LD with pooled test statistics
1	0.968±0.003	0.980±0.002	0.977±0.003
2	0.949±0.004	0.937±0.005	0.949±0.004 ^a
3	0.954±0.004	0.950±0.004 ^a	0.957±0.004 ^a
4	0.899±0.006	0.877±0.007	0.900±0.006 ^a
5	0.709±0.011	0.793±0.009	0.774±0.010
6	0.674±0.011	0.724±0.011	0.702±0.011

The values after \pm are the standard deviations of the AUCs.

^a The AUC difference compared to the single-template strategy was either not statistically significant ($p > 0.05$) or not clinically important (difference < 0.01).

4.5 Discussion

We have developed and evaluated two multi-template strategies to classify non-MVN distributed data using a sub-ensemble-based approach. In order to use either of them to rank different systems in place of human observers, correlation of task performance calculated by these model observers and human performance on the corresponding task must be studied.

An important observation in this study was that good correlation of different model observers in a one-dimensional parameter space did not imply good correlation in a multi-dimensional parameter space. More specifically, in this study, we were optimizing two reconstruction parameters: the iteration number of the OS-EM algorithm and the cutoff frequency of a post reconstruction low-pass filter. We observed that the three observer strategies had very good correlations when comparing only the ranking of different values for one parameter with the other parameter fixed, as shown in Figure 4-11 and Figure 4-12. However, the three strategies achieved different optimal parameters when we simultaneously optimized both parameters, as shown in Figure 4-9. This happens when the ranking of different values for one parameter changes with the other parameter(s). The result is significant since it implies that the correlation of different observers in a one-dimensional parameter space, as studied previously [74, 79, 80], may not indicate the correlation in a multi-dimensional parameter space.

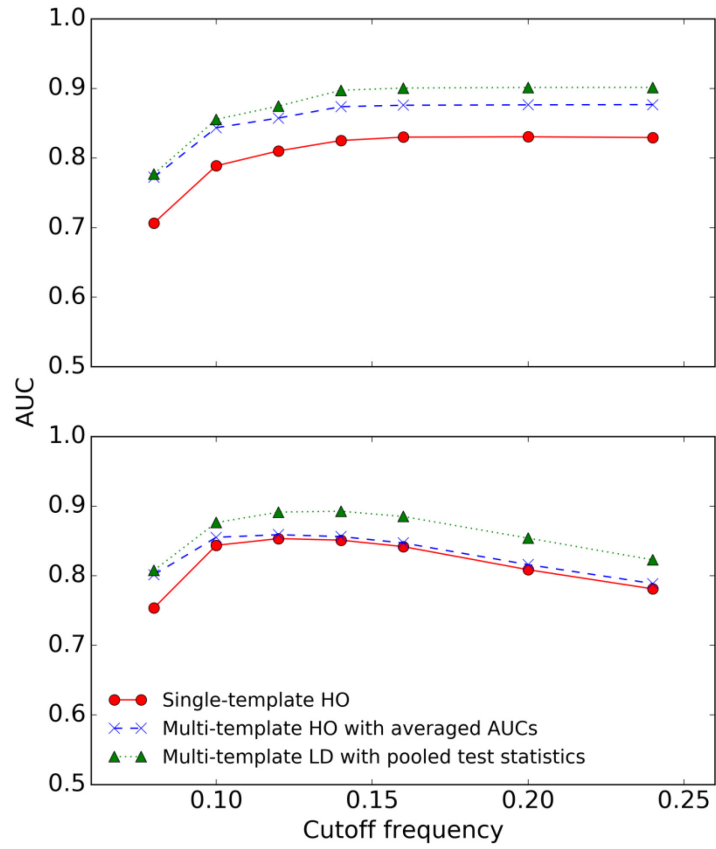


Figure 4-11. AUC plot for Tc of three observer strategies for different cutoff frequencies (pixel^{-1}) at iterations 2(upper) and 60 (lower) (4 subsets/iteration).

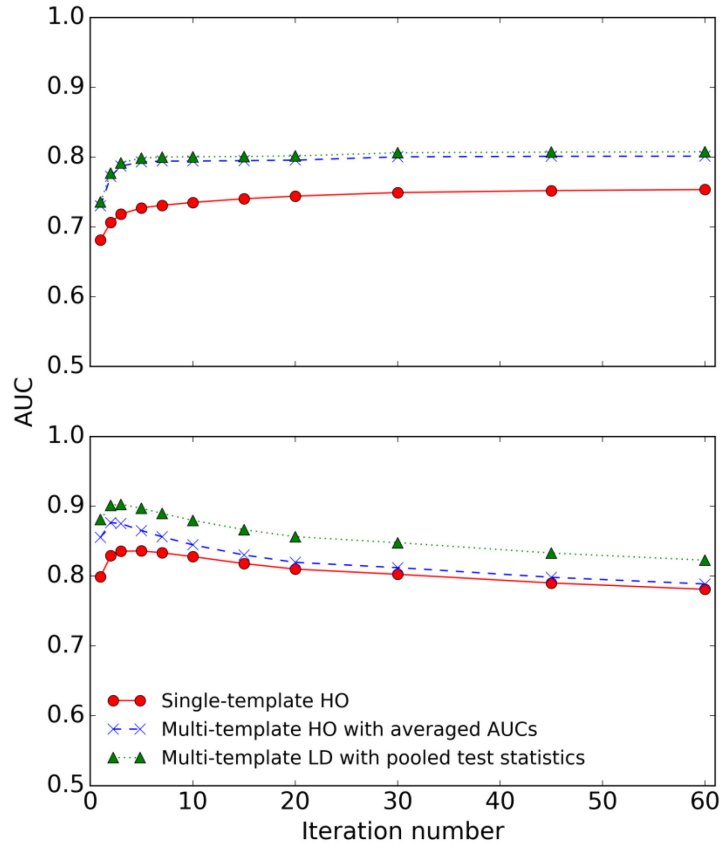


Figure 4-12. AUC plot for Tc of three observer strategies for different iteration numbers at cutoff frequencies 0.08 (upper) and 0.24 (lower) pixel⁻¹.

An unresolved question is the relative merit of the two multi-template strategies. In terms of AUCs, the multi-template LD with pooled test statistics strategy provided higher values. In terms of optimal parameter range, the two strategies obtained similar optimal parameter ranges for the data used in this study. However, it is possible that differences in ranking of different image systems could be observed for other tasks, and this remains a topic where a future investigation is required. One advantage of the multi-template LD with pooled test statistics strategy is that it requires fewer ROC analysis operations compared to the multi-template HO with averaged AUCs strategy. This could be important when there are a small number of cases available for certain sub-ensembles since AUC estimation methods are often not reliable for a small number of cases. The effect of having only a small number of cases on the performance of the two multi-template strategies is another topic of investigation. In addition, the multi-template

LD with pooled test statistics strategy is also more theoretically sound for the following reason. For the multi-template HO with averaged AUCs strategy, the AUC values for each defect type are averaged to give the overall AUC value, which is equivalent to averaging the ROC curves to get the overall ROC curve. This is equivalent to averaging true positive fraction (TPF) values at the same false positive fraction (FPF) value. However, the exact meaning of averaging the TPFs for a given FPF is difficult to define. Thus, the meaning of averaged ROC curves is not clear. For the multi-template LD with pooled test statistics strategy, the pooling of the test statistics is justified based on pooling of the likelihood ratios and the fact that the shifting provided by the LD gives the maximum overall AUC; a single ROC curve is estimated for the entire ensemble of test statistics, thus avoiding the questionable averaging of ROC curves used by the multi-template HO strategy.

It must also be emphasized that the investigated multi-template strategies are general strategies to handle non-MVN data, and are not limited to feature vectors generated using the anthropomorphic channels evaluated in this Chapter. The advantages of the strategies apply to other potentially non-MVN data, such as feature vectors from other anthropomorphic or efficient channels or even projection data. These multi-template strategies are also less computationally expensive when the number of sub-ensembles is much smaller than the possible number of signal types compared to the previous multi-template strategies for SKS and SKEV tasks.

4.6 Conclusion

We have proposed a novel multi-template linear observer strategy for analyzing detection performance in datasets that are not MVN distributed. The strategy consists of dividing the data into sub-ensembles that are MVN and homoscedastic, applying different linear discriminant (LD) templates on the different sub-ensembles, and finally pooling the test statistics. We also adapted another multi-template strategy, initially proposed in the context of SKEV tasks, for non-MVN distributed data based on the sub-ensemble approach. Both of these multi-template strategies

were compared to the conventional single-template HO strategy. The strategies were compared by applying them to optimize reconstruction parameters for the non-MVN data from a realistic simulated myocardial perfusion SPECT dataset. The two multi-template strategies yielded more optimal reconstruction parameters compared to the single-template HO strategy in terms of higher AUC for each sub-ensemble. The novel multi-template LD with pooled test statistics strategy is more theoretically justified and provided a higher AUC for the entire ensemble than the adapted multi-template strategy. The theory and results we presented provide strong evidence in favor of using the proposed multi-template LD strategy to classify the non-MVN data from clinically realistic tasks such as that used in this study.

4.7 Appendix

4.7.1 Appendix A

Here we prove the homoscedasticity condition will be approximately satisfied if the signal variations are relatively small compared to the background variations and the signal and background are uncorrelated.

Suppose \mathbf{B} and \mathbf{S} are the vectors of background and signal data. The covariance matrices of \mathbf{B} and \mathbf{S} are expressed in the following two equations:

$$\mathbf{K}_{\mathbf{B}} = \mathbb{E} \left[(\mathbf{B} - \mathbb{E}[\mathbf{B}])(\mathbf{B} - \mathbb{E}[\mathbf{B}])^T \right], \quad (\text{A-1})$$

$$\mathbf{K}_{\mathbf{S}} = \mathbb{E} \left[(\mathbf{S} - \mathbb{E}[\mathbf{S}])(\mathbf{S} - \mathbb{E}[\mathbf{S}])^T \right], \quad (\text{A-2})$$

where $\mathbb{E}[\cdot]$ denotes the expectation of the corresponding vector or matrix.

The covariance matrix of $\mathbf{B} + \mathbf{S}$ is expressed as

$$\begin{aligned} \mathbf{K}_{\mathbf{B}+\mathbf{S}} &= \mathbb{E} \left[(\mathbf{B} + \mathbf{S} - \mathbb{E}[\mathbf{B} + \mathbf{S}])(\mathbf{B} + \mathbf{S} - \mathbb{E}[\mathbf{B} + \mathbf{S}])^T \right] \\ &= \mathbf{K}_{\mathbf{B}} + \mathbf{K}_{\mathbf{S}} + \mathbb{E} \left[(\mathbf{B} - \mathbb{E}[\mathbf{B}])(\mathbf{S} - \mathbb{E}[\mathbf{S}])^T \right] + \mathbb{E} \left[(\mathbf{S} - \mathbb{E}[\mathbf{S}])(\mathbf{B} - \mathbb{E}[\mathbf{B}])^T \right]. \end{aligned} \quad (\text{A-3})$$

When B and S are not correlated, we have

$$\mathbb{E}\left[(\mathbf{B}-\mathbb{E}[\mathbf{B}])(\mathbf{S}-\mathbb{E}[\mathbf{S}])^T\right]=\mathbb{E}\left[(\mathbf{S}-\mathbb{E}[\mathbf{S}])(\mathbf{B}-\mathbb{E}[\mathbf{B}])^T\right]=0. \quad (\text{A-4})$$

Based on (A3) and (A4),

$$\mathbf{K}_{\mathbf{B}+\mathbf{S}} = \mathbf{K}_{\mathbf{B}} + \mathbf{K}_{\mathbf{S}}. \quad (\text{A-5})$$

When $\mathbf{K}_{\mathbf{S}} \ll \mathbf{K}_{\mathbf{B}}$, $\mathbf{K}_{\mathbf{B}+\mathbf{S}} \approx \mathbf{K}_{\mathbf{B}}$, which means the covariance matrices of the background with and without the signal are approximately equal if the signal variations are relatively small compared to the background variations and if the signal and background are uncorrelated.

4.7.2 Appendix B

Here we prove Theorem 1 in 4.2.3.3.

Let t be the decision threshold. For the j th sub-ensemble, the true-positive fraction (TPF) and false-positive fraction (FPF) can be expressed as

$$TPF_j(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(\lambda - c_j - \mu_j)^2}{2\sigma_j^2}\right] d\lambda \quad (\text{B-1})$$

and

$$FPF_j(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(\lambda - c_j + \mu_j)^2}{2\sigma_j^2}\right] d\lambda, \quad (\text{B-2})$$

where λ is the test statistic value.

Let P_j be the prevalence of the j th sub-ensemble. If we pool the test statistics from all sub-ensembles we obtain

$$TPF(t) = \sum_{j=1}^N P_j \times TPF_j(t) \quad (\text{B-3})$$

and

$$FPF(t) = \sum_{j=1}^N P_j \times FPF_j(t). \quad (B-4)$$

Note that

$$\sum_{j=1}^N P_j = 1. \quad (B-5)$$

Using the definition of the error function [98]

$$\frac{2}{\sqrt{\pi}} \int_z^{\infty} \exp(-t^2) dt = 1 - \text{erf}(z), \quad (B-6)$$

and equations (B-1)-(B-4), we have

$$TPF(t) = \sum_{j=1}^N \frac{P_j}{2} \left[1 - \text{erf} \left(\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j} \right) \right] \quad (B-7)$$

and

$$FPF(t) = \sum_{j=1}^N \frac{P_j}{2} \left[1 - \text{erf} \left(\frac{t - c_j + \mu_j}{\sqrt{2}\sigma_j} \right) \right]. \quad (B-8)$$

According to the definition of the AUC [32]

$$AUC = - \int_{-\infty}^{\infty} TPF(t) \frac{dFPF(t)}{dt} dt. \quad (B-9)$$

Using equations (B-7)-(B-9) and the following property of the error function [98]

$$\frac{d}{dz} \text{erf}(z) = \frac{2}{\sqrt{\pi}} \exp(-z^2), \quad (B-10)$$

the AUC can be written as

$$AUC(c_2, c_3, \dots, c_N) = \int_{-\infty}^{\infty} \left\{ \sum_{j=1}^N \frac{P_j}{2} \left[1 - \text{erf} \left(\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j} \right) \right] \right\} \left\{ \sum_{i=1}^N \frac{P_i}{\sqrt{2\pi}\sigma_i} \exp \left[- \left(\frac{t - c_i + \mu_i}{\sqrt{2}\sigma_i} \right)^2 \right] \right\} dt, \quad (B-11)$$

where we have written AUC as a function of c_2, c_3, \dots, c_N , since we are interested in finding the

values of c_2, c_3, \dots, c_N that maximize the AUC.

Using (B-5) and expanding the terms of the above yields

$$\begin{aligned}
AUC(c_2, c_3, \dots, c_N) = & \frac{1}{2} \int_{-\infty}^{\infty} \sum_{i=1}^N \frac{P_i}{\sqrt{2\pi}\sigma_i} \exp \left[- \left(\frac{t - c_i + \mu_i}{\sqrt{2}\sigma_i} \right)^2 \right] dt \\
& - \frac{1}{2\sqrt{2\pi}} \sum_{j=1}^N \frac{P_j^2}{\sigma_j} \int_{-\infty}^{\infty} \operatorname{erf} \left(\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j} \right) \exp \left[- \left(\frac{t - c_j + \mu_j}{\sqrt{2}\sigma_j} \right)^2 \right] dt \\
& - \frac{1}{2\sqrt{2\pi}} \sum_{j=1}^N \sum_{i=1, i \neq j}^N \frac{P_j P_i}{\sigma_i} \int_{-\infty}^{\infty} \operatorname{erf} \left(\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j} \right) \exp \left[- \left(\frac{t - c_i + \mu_i}{\sqrt{2}\sigma_i} \right)^2 \right] dt.
\end{aligned} \tag{B-12}$$

Note that the first and second terms in the above can be made independent of the constants c_2, c_3, \dots, c_N by change of variables. For example, in the first term, we could simply replace $t - c_i$ by t' . Thus, these terms can be ignored when maximizing the AUC. Denoting the last term by $k(c_2, c_3, \dots, c_N)$, we have:

$$k(c_2, c_3, \dots, c_N) = - \frac{1}{2\sqrt{2\pi}} \sum_{j=1}^N \sum_{i=1, i \neq j}^N \frac{P_j P_i}{\sigma_i} \int_{-\infty}^{\infty} \operatorname{erf} \left(\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j} \right) \exp \left[- \left(\frac{t - c_i + \mu_i}{\sqrt{2}\sigma_i} \right)^2 \right] dt. \tag{B-13}$$

Replacing $\frac{t - c_j - \mu_j}{\sqrt{2}\sigma_j}$ by t' , and using the following property of the error function [98]

$$\int_{-\infty}^{\infty} \operatorname{erf}(x) \exp[-(ax + b)^2] dx = - \frac{\sqrt{\pi}}{a} \operatorname{erf} \left(\frac{b}{\sqrt{a^2 + 1}} \right), \tag{B-14}$$

equation (B-13) can be written as

$$k(c_2, c_3, \dots, c_N) = \frac{1}{2} \sum_{j=1}^N \sum_{i=1, i \neq j}^N P_i P_j \operatorname{erf} \left(\frac{\mu_j + c_j - c_i + \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right). \tag{B-15}$$

Differentiating the above expression with respect to c_2, c_3, \dots, c_N and equating the result to zero yields the critical points at which $k(c_2, c_3, \dots, c_N)$, or alternatively, the AUC is maximized or minimized. Thus, differentiating the above expression with respect to any general c_l , where

$I = 2, \dots, N$, yields

$$\begin{aligned} & \frac{\partial k(c_2, c_3, \dots, c_N)}{\partial c_I} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{i=1, i \neq I}^N \frac{P_I P_i}{\sqrt{\sigma_I^2 + \sigma_i^2}} \left\{ \exp \left[- \left(\frac{c_I + \mu_I - c_i + \mu_i}{\sqrt{2(\sigma_I^2 + \sigma_i^2)}} \right)^2 \right] - \exp \left[- \left(\frac{c_i + \mu_i - c_I + \mu_I}{\sqrt{2(\sigma_I^2 + \sigma_i^2)}} \right)^2 \right] \right\}. \\ & (I = 2, \dots, N) \end{aligned} \tag{B-16}$$

The above expression is equal to zero when the exponential terms are equal, which occurs when $c_I = c_1 (I = 2, \dots, N)$. In 4.7.4, Hessian matrix of $k(c_2, c_3, \dots, c_N)$ is proved to be strictly negative definite at any (c_2, c_3, \dots, c_N) . Thus, when $c_i = c_j (\forall i, j \in [1, N])$, the term $k(c_2, c_3, \dots, c_N)$, and hence the AUC, is maximized.

4.7.3 Appendix C

Here we prove theorem 2 in 4.2.3.3.

For the j th sub-ensemble, assuming the input data are MVN distributed and homoscedastic, the probability distribution of the data vectors is

$$\text{pr}(\mathbf{g} | H_i) = \frac{1}{(2\pi)^{M/2} \sqrt{|\mathbf{K}_{\mathbf{g}j}|}} \exp \left[-\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}_{ji})^T \mathbf{K}_{\mathbf{g}j}^{-1} (\mathbf{g} - \bar{\mathbf{g}}_{ji}) \right]. (i = 0, 1) \tag{C-1}$$

Then, from equation (C-1), if $\mathbf{g} = \frac{\bar{\mathbf{g}}_{j0} + \bar{\mathbf{g}}_{j1}}{2}$, we can derive that $\text{pr}(\mathbf{g} | H_0) = \text{pr}(\mathbf{g} | H_1)$, and

$\text{pr}(\lambda(\mathbf{g}) | H_0) = \text{pr}(\lambda(\mathbf{g}) | H_1)$. This means that the distributions of the test statistics of the two

hypotheses are equal (i.e., the test statistic distributions cross) when $\mathbf{g} = \frac{\bar{\mathbf{g}}_{j0} + \bar{\mathbf{g}}_{j1}}{2}$.

The LD test statistic defined by equation (4-2) at the point $\mathbf{g} = \frac{\bar{\mathbf{g}}_{j0} + \bar{\mathbf{g}}_{j1}}{2}$ is given by

$$c_j = \lambda(\mathbf{g}) = \Delta \bar{\mathbf{g}}_j^T \mathbf{K}_{\mathbf{g}j}^{-1} \mathbf{g} + \frac{1}{2} \left(\bar{\mathbf{g}}_{j0}^T \mathbf{K}_{\mathbf{g}j}^{-1} \bar{\mathbf{g}}_{j0} - \bar{\mathbf{g}}_{j1}^T \mathbf{K}_{\mathbf{g}j}^{-1} \bar{\mathbf{g}}_{j1} \right) = 0, \tag{C-2}$$

where $\Delta \bar{\mathbf{g}}_j = \bar{\mathbf{g}}_{j1} - \bar{\mathbf{g}}_{j0}$.

Now, we have shown that shifting the HO test statistics by the term

$$\Delta d_j = \frac{1}{2} \left(\bar{\mathbf{g}}_{j0}^T \mathbf{K}_{g_j}^{-1} \bar{\mathbf{g}}_{j0} - \bar{\mathbf{g}}_{j1}^T \mathbf{K}_{g_j}^{-1} \bar{\mathbf{g}}_{j1} \right)$$

leads to the distributions of the test statistics being equal at the point where the test statistic is zero.

4.7.4 Appendix D

Here we prove that the Hessian matrix of $k(c_2, c_3, \dots, c_N)$ is negative definite.

The Hessian matrix of $k(c_2, c_3, \dots, c_N)$ is

$$H = \begin{bmatrix} \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_2^2} & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_2 c_3} & \dots & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_2 c_N} \\ \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_3 c_2} & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_3^2} & \dots & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_3 c_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_N c_2} & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_N c_3} & \dots & \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_N^2} \end{bmatrix}_{(N-1) \times (N-1)}, \quad (\text{D-1})$$

where

$$\begin{aligned} \frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_I^2} &= \frac{-2}{\sqrt{2\pi}} \sum_{j=1, j \neq I}^N \frac{P_I P_j}{\sqrt{\sigma_I^2 + \sigma_j^2}} \left\{ \exp \left[- \left(\frac{c_I + \mu_I - c_j + \mu_j}{\sqrt{2(\sigma_I^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_I + \mu_I - c_j + \mu_j}{2(\sigma_I^2 + \sigma_j^2)} \right. \\ &\quad \left. + \exp \left[- \left(\frac{c_j + \mu_j - c_I + \mu_I}{\sqrt{2(\sigma_I^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_j + \mu_j - c_I + \mu_I}{2(\sigma_I^2 + \sigma_j^2)} \right\}, \quad (I = 2, \dots, N) \end{aligned} \quad (\text{D-2})$$

and

$$\begin{aligned}
\frac{\partial^2 k(c_2, c_3, \dots, c_N)}{\partial c_I c_J} &= \frac{2}{\sqrt{2\pi}} \frac{P_I P_J}{\sqrt{\sigma_I^2 + \sigma_J^2}} \left\{ \exp \left[- \left(\frac{c_I + \mu_I - c_J + \mu_J}{\sqrt{2(\sigma_I^2 + \sigma_J^2)}} \right)^2 \right] \frac{c_I + \mu_I - c_J + \mu_J}{2(\sigma_I^2 + \sigma_J^2)} \right. \\
&\quad \left. + \exp \left[- \left(\frac{c_J + \mu_J - c_I + \mu_I}{\sqrt{2(\sigma_I^2 + \sigma_J^2)}} \right)^2 \right] \frac{c_J + \mu_J - c_I + \mu_I}{2(\sigma_I^2 + \sigma_J^2)} \right\} \cdot (I, J = 2, \dots, N, I \neq J)
\end{aligned} \tag{D-3}$$

In order to simplify the equations, we write

$$\begin{aligned}
X_{ij} &\equiv \frac{2}{\sqrt{2\pi}} \frac{P_i P_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \left\{ \exp \left[- \left(\frac{c_i + \mu_i - c_j + \mu_j}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_i + \mu_i - c_j + \mu_j}{2(\sigma_i^2 + \sigma_j^2)} \right. \\
&\quad \left. + \exp \left[- \left(\frac{c_j + \mu_j - c_i + \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_j + \mu_j - c_i + \mu_i}{2(\sigma_i^2 + \sigma_j^2)} \right\}.
\end{aligned} \tag{D-4}$$

As proved in 4.7.5, $X_{ij} > 0$.

The Hessian matrix can be written as

$$H = \begin{bmatrix} -\sum_{j=1, j \neq 2}^N X_{2j} & X_{23} & \cdots & X_{2N} \\ X_{32} & -\sum_{j=1, j \neq 3}^N X_{3j} & \cdots & X_{3N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N2} & X_{N3} & \cdots & -\sum_{j=1, j \neq N}^N X_{Nj} \end{bmatrix}_{(N-1) \times (N-1)}. \tag{D-5}$$

From (D-4) we see that

$$X_{ij} = X_{ji}. \tag{D-6}$$

Let $Y = \begin{bmatrix} y_2 & y_3 & \cdots & y_N \end{bmatrix}^T_{(N-1) \times 1}$; using (D-5) and (D-6), we have

$$Y^T H Y = - \left(\sum_{j=2}^N X_{j1} y_j^2 \right) - \sum_{i=2}^N \sum_{j=2, i \neq j}^N X_{ij} (y_i - y_j)^2. \tag{D-7}$$

When $Y \neq 0$, we have $-\left(\sum_{j=2}^N X_{j1} y_j^2\right) < 0$ and $-\sum_{i=2}^N \sum_{j=2, i \neq j}^N X_{ij} (y_i - y_j)^2 \leq 0$ since $X_{ij} > 0$.

So $Y^T H Y < 0, \forall Y \neq 0$, which means the Hessian matrix of $k(c_2, \dots, c_N)$ is negative definite.

4.7.5 Appendix E

Here we prove that X_{ij} defined in (D-4) is positive.

At the critical point, $c_i = c_j = c_1, \forall i, j \geq 2$, and since $\mu_j > 0$, from (D4), we have $X_{ij} > 0$.

At non-critical point, suppose $c_i - c_j > 0, \forall i, j \geq 2$.

When $c_i - c_j < \mu_i + \mu_j$, we have

$$\frac{c_j + \mu_j - c_i + \mu_i}{2(\sigma_i^2 + \sigma_j^2)} > 0, \quad (\text{E-1})$$

and

$$\frac{c_i + \mu_i - c_j + \mu_j}{2(\sigma_i^2 + \sigma_j^2)} > 0. \quad (\text{E-2})$$

From (E-1), (E-2) and (D-4), we have $X_{ij} > 0$.

When $c_i - c_j \geq \mu_i + \mu_j$, we have

$$0 \leq \frac{c_i - \mu_j - c_j - \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} < \frac{c_i + \mu_j - c_j + \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}}, \quad (\text{E-3})$$

$$\exp \left[- \left(\frac{c_j + \mu_j - c_i + \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] = \exp \left[- \left(\frac{c_i - \mu_j - c_j - \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] > \exp \left[- \left(\frac{c_i + \mu_j - c_j + \mu_j}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right]. \quad (\text{E-4})$$

From (E-3), (E-4), and (D-4), we have

$$\begin{aligned}
X_{ij} &> \\
&\frac{2}{\sqrt{2\pi}} \frac{P_i P_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \left\{ \exp \left[- \left(\frac{c_i + \mu_i - c_j + \mu_j}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_i + \mu_i - c_j + \mu_j}{2(\sigma_i^2 + \sigma_j^2)} \right. \\
&\quad \left. + \exp \left[- \left(\frac{c_i + \mu_j - c_j + \mu_i}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] \frac{c_j + \mu_j - c_i + \mu_i}{2(\sigma_i^2 + \sigma_j^2)} \right\} \\
&= \frac{2}{\sqrt{2\pi}} \frac{P_i P_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \exp \left[- \left(\frac{c_i + \mu_i - c_j + \mu_j}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}} \right)^2 \right] \frac{\mu_j + \mu_i}{\sigma_i^2 + \sigma_j^2} > 0.
\end{aligned} \tag{E-5}$$

So for $c_i - c_j > 0, \forall i, j \geq 2, X_{ij} > 0$.

Similarly, we can prove for $c_i - c_j < 0, \forall i, j \geq 2, X_{ij} > 0$.

So, at any points, $X_{ij} > 0$.

5. Toward patient-specific optimization of post-filtered OS-EM reconstruction for simultaneous-acquisition dual-isotope myocardial perfusion SPECT

5.1 Introduction

In this study, we used a reconstruction-based crosstalk compensation method for simultaneously acquired dual isotope data. The method is based on the ordered subset-expectation maximization (OS-EM) algorithm and post-reconstruction low-pass filtering was used to control noise. The method provides crosstalk-compensated Tc-99m stress and Tl-201 rest images. Four adjustable parameters control the image quality: the number of updates (product of number of iterations and number of subsets) and the cutoff frequencies of a post-reconstruction low-pass filter for the stress and rest images. The goal is to optimize the image quality by finding the values of parameters that maximize the image quality in terms of performance on a 3-class classification task. Previous authors have used an anthropomorphic channelized Hotelling observer (CHO) to optimize the parameters for similar studies [72, 91, 97]. In those studies a single set of parameters was found for the entire phantom population. However, different patients or different groups of patients may have different optimal parameters. There have been previous studies showing that task performance or the reconstruction or post-smoothing parameters depend on the characteristics of the imaging system and the object imaged, such as the system modulation transfer function or the count level [99, 100].

In this Chapter, we test the hypothesis that more patient-specific parameters can offer better task performance. We explored approaches to optimizing reconstruction parameters based on features of the patient including count level in the acquired projection data, a proxy for image resolution, and defect size, contrast, and location. To handle the problem of non-MVN data, we

used the Linear Discriminant (LD) with pooled test statistics observer strategy developed in Chapter 4.

5.2 Methods

5.2.1 Reconstruction and post-processing

In this study, a post-smoothing OS-EM reconstruction method was implemented as described in Section 4.3.1. Image degrading factors such as attenuation, detector response, scatter and crosstalk between the two isotopes were compensated for. Scatter compensation was based on the effective source scatter estimation (ESSE) method [19] and crosstalk compensation was based on true noise-free crosstalk between the two isotopes. In the OS-EM reconstruction we used 4 subsets/iteration and reconstructed images using 1, 2, 3, 5, 7, 10, 15, 20, 30, 45, and 60 iterations for Tc-99m and 1, 2, 3, 5, 7, 10, 15, and 20 iterations for Tl-201. We used a Butterworth filter with an order of 8 as post-reconstruction low-pass filter; cutoff frequencies used were 0.08, 0.1, 0.12, 0.14, 0.16, 0.2, 0.24 pixels⁻¹. The filtered images were reoriented to short axis slices and a 64×64 image containing the centroid of the defect at the center of the image was extracted. The resulting floating-point-format images were rescaled and rounded such that pixel values between [0, maximum in the heart] were remapped to integer images with pixel values in the range [0,255]. These 2D, rescaled, integer images were used in the observer studies.

5.2.2 Image quality evaluation

Both two-class and three-class performances were calculated using the multi-template LD with pooled test statistics observer strategy proposed in Chapter 4, as described in the following.

First, we calculated feature vectors by apply six rotationally symmetric frequency channels to each reconstructed images. The lowest frequency channel had a starting frequency and channel width of 1/128 cycles per pixel. Subsequent higher frequency channels adjoined the previous ones and had double the previous channel's width.

The feature vectors in the full ensemble were separated into sub-ensembles that were MVN distributed and homoscedastic. In this study, we partitioned the data based on defect types, as discussed in Chapter 4. Then the LD was used to calculate test statistics for each feature vector in the sub-ensemble. The test statistics were computed using a leave-one-out technique [77, 92]. In that technique, the LD template is trained (i.e., the LD template is calculated as described in subsequent paragraphs) on all the feature vectors in each sub-ensemble except one feature vector. Then the resulting template is applied to the remaining feature vector to calculate the LD test statistic for that feature vector. Each of the feature vectors in the sub-ensemble is left out and the process above repeated.

For the two-classification task (i.e., where the classes were normal and abnormal), we estimated the LD template for all the feature vectors in the same sub-ensemble. The LD template is given by:

$$\lambda(\mathbf{g}) \approx \Delta \bar{\mathbf{g}}^T \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{g} + \frac{1}{2} \left(\bar{\mathbf{g}}_0^T \mathbf{K}_{\mathbf{g}}^{-1} \bar{\mathbf{g}}_0 - \bar{\mathbf{g}}_1^T \mathbf{K}_{\mathbf{g}}^{-1} \bar{\mathbf{g}}_1 \right), \quad (5-1)$$

where \mathbf{g} is the feature vector, $\bar{\mathbf{g}}_i$ and \mathbf{K}_i denote, respectively, the mean data vector and covariance matrix of the data under the defect-absent and defect-present hypotheses, $\Delta \bar{\mathbf{g}} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0$, and $\mathbf{K}_{\mathbf{g}} = \frac{1}{2}(\mathbf{K}_0 + \mathbf{K}_1)$ suppose $\mathbf{K}_0 = \mathbf{K}_1$.

For the three-class classification problem (where the classes were normal, fixed, and reversible), we used a similar methodology as in [101]. A pair of test statistics was calculated for each feature vector, one each for the binary decisions normal versus reversible and reversible versus fixed. These test statistics were calculated using the leave-one-out technique described for the two-class case. For this, two LD templates were required and were calculated using:

$$\lambda_{N,R}(\mathbf{g}) \approx (\bar{\mathbf{g}}_R - \bar{\mathbf{g}}_N)^T \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{g} + \frac{1}{2} \left(\bar{\mathbf{g}}_N^T \mathbf{K}_{N,R}^{-1} \bar{\mathbf{g}}_N - \bar{\mathbf{g}}_R^T \mathbf{K}_{N,R}^{-1} \bar{\mathbf{g}}_R \right) \quad (5-2)$$

$$\lambda_{R,F}(\mathbf{g}) \approx (\bar{\mathbf{g}}_F - \bar{\mathbf{g}}_R)^T \mathbf{K}_g^{-1} \mathbf{g} + \frac{1}{2} \left(\bar{\mathbf{g}}_R^T \mathbf{K}_{R,F}^{-1} \bar{\mathbf{g}}_R - \bar{\mathbf{g}}_F^T \mathbf{K}_{N,R}^{-1} \bar{\mathbf{g}}_F \right) \quad (5-3)$$

where the feature vector \mathbf{g}_i ($i=N,R,F$) is the result of stacking the feature vectors for the stress

and rest images $\mathbf{g}_{s,i}$ and $\mathbf{g}_{r,i}$, i.e., $\mathbf{g}_i = \begin{bmatrix} \mathbf{g}_{s,i} \\ \mathbf{g}_{r,i} \end{bmatrix}$, \mathbf{K}_i is the covariance matrix of the stacked feature

vectors \mathbf{g}_i ($i=N,R,F$), $\mathbf{K}_{N,R} = \left[\frac{1}{2}(\mathbf{K}_N + \mathbf{K}_R) \right]$, and $\mathbf{K}_{R,F} = \left[\frac{1}{2}(\mathbf{K}_R + \mathbf{K}_F) \right]$.

For the two-class task, the FOM was the area under the ROC curve (AUC). To calculate this, the test statistics from all the sub-ensembles were pooled and served as inputs to the LABROC program[94]. This is a maximum-likelihood algorithm for estimating the parameters of a binormal ROC curve from a set of input test statistics.

For the 3-class task the FOM was the volume under the 3-class ROC surface (VUS). The test statistics were calculated using the leave-one-out technique and were pooled together for all the sub-ensembles. In this case, the VUS was calculated using a resampling procedure described in [102].

5.2.3 Patient-specific optimization of reconstruction parameters

The underlying hypothesis of this work is that patient-specific optimization of reconstruction and filtering parameters provides better task performance. At present, optimizing these parameters for each patient must be done empirically and is not practical. As a step toward full patient-specific optimization, we investigated the optimization of groups of patients that have similar properties that have an effect on image quality.

Based on previous studies, the noise level, defect size and the image resolution are key factors that affect the optimal parameters. For example, one expects that a lower-cutoff filter and a smaller number of updates would be optimal for an image with lower counts than for one with higher counts. Thus, we propose to divide the population according to image noise in the

myocardium, the minimum defect size of interest, and image resolution in the heart. The image reconstruction and filtering parameters were then optimized for each phantom group.

We have developed proxies for these three parameters that can be estimated from the measured image data. We used the average number of detected counts in the myocardium to estimate the coefficient of variation as a noise index. The coefficient of variation (COV) was estimated using:

$$COV = \frac{\sigma_N}{\langle N \rangle} = \frac{\sqrt{N_0}}{N_0} = \frac{1}{\sqrt{N_0}}, \quad (5-4)$$

where $\langle N \rangle$ and σ_N are the mean and standard deviation of the measured number of detected counts N_0 .

We obtained the average number of detected counts by dividing the sum of voxel values in the myocardium from an image reconstructed using filtered back projection by the number of voxels in the VOI. We estimate the minimum defect size of interest based on the assumption that defects with an extent $< 10\%$ are not of clinical importance. The minimum defect size of interest for a given patient can thus be estimated from the myocardial volume, which could be estimated from CT images or by appropriate segmentation of a SPECT image. We used the cube root of 10% of the myocardial volume as the lower bound of the size of defect that would be considered clinically relevant. As a proxy for image resolution in the myocardium, we used the average distance from the centroid of the heart to the collimator surface over all projection angles for that patient combined with the collimator and detector parameters to calculate the full-width-at-half-maximum (FWHM) of the image resolution in the heart. The ratio of the minimum defect size index and the resolution is an index for the effect of resolution on the ability to visualize perfusion defects: partial volume effects will have a smaller effect on reducing defect contrast when the ratio is large than when it is small.

The phantoms were divided into three anatomy groups according to two factors: the

myocardial noise index and the ratio of the defect size and image resolution indices. The data rather naturally divided themselves into 3 groups: low defect to resolution ratio and low noise (LRLN), low defect to resolution ratio and high noise (LRHN), and high defect to resolution ratio and low noise (HRLN). The thresholds and the resulting division of data are shown in Figure 5-1. Note that the absolute noise level did not affect the grouping since the thresholds were chosen based on relative noise levels within the data. As a result, The Tc and Tl images corresponding to different injected activity, although the absolute noise levels were different by some scale factor, the anatomy groups were the same. Resolution depends both on the system and patient. For example, if a circular orbit was used the resolution would be poorer and the population group that patient could, in principle, be different than with a body-contouring orbit.

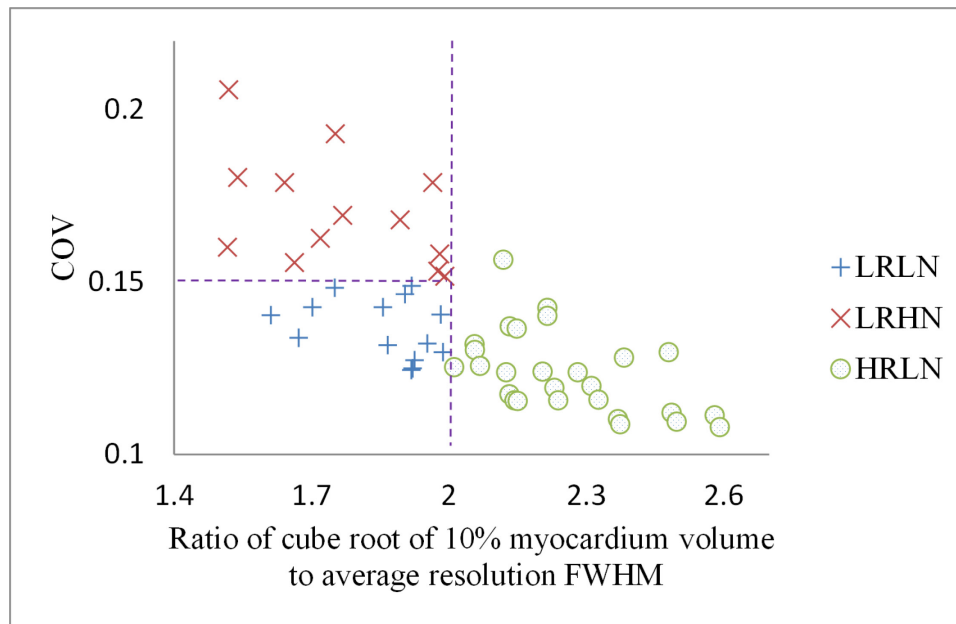


Figure 5-1. Illustration of the separation of phantoms into anatomy groups based on noise index (COV for 10 mCi Tc-99m) and defect to resolution ratio index (Ratio of cube root of 10% myocardium volume to average resolution): LRLN (low defect to resolution ratio low noise), LRHN (low defect to resolution ratio high noise), and HRLN (high defect to resolution ratio low noise).

In addition to optimizing based on parameters measurable from the images, we also investigated optimizing as a function of defect type. This might not be clinically possible, but provides information on how much improvement in task performance information about defect

type could provide.

In this Chapter, we studied and compared four optimization approaches. We determined the optimal reconstruction parameters for rest (Tl) and stress (Tc) images under the following conditions: (1) separately for each defect type and each of the three anatomy groups; (2) separately for each defect type; (3) separately for each anatomy group; (4) for the whole population. In all cases, the optimal parameters were determined in terms of 2-D class performance using multi-template LD with pooled test statistics observer strategy. The four optimization approaches are summarized in Table 5-1. Reconstruction parameters that achieved the highest AUC for each group were considered as optimal for that group.

Table 5-1. Four optimization approaches.

Approach	Optimized for each defect type	Optimized for each anatomy groups
(1) By defect type and anatomy group	Yes	Yes
(2) By defect type	Yes	No
(3) By anatomy group	No	Yes
(4) Whole population	No	No

5.3 Results

5.3.1 Optimal parameters

The optimal parameters for each of the four optimization approaches are summarized in Table 5-2 for Tc and Table 5-3 for Tl.

From these data, it is evident that variations in defect severity, extent and location affect the optimal parameters. For a fixed defect location, the optimal cutoffs tend to decrease as the defect extent increases (from defect 1 to 6). When all the defect types were combined, the optimal cutoff was a compromise between the cutoffs for every single defect type.

Different anatomy groups had different optimal parameters: the LRLN group tends to have higher iteration number than the HRLN group due to smaller defect size to resolution ratio or

higher cutoff frequencies than the LRHN group due to lower noise.

Table 5-2. Optimal parameters for Tc.

Defect type	Anatomy groups			
	HRLN	LRHN	LRLN	Whole population
1	(20, 0.14)	(5, 0.12)	(30, 0.16)	(3, 0.24)
2	(3, 0.24)	(45, 0.14)	(20, 0.14)	(7, 0.16)
3	(7, 0.12)	(3, 0.12)	(2, 0.16)	(3, 0.14)
4	(2, 0.24)	(2, 0.16)	(30, 0.14)	(7, 0.14)
5	(2, 0.1)	(3, 0.1)	(3, 0.14)	(2, 0.1)
6	(2, 0.24)	(3, 0.1)	(2, 0.1)	(5, 0.08)
1-6	(2, 0.2)	(2, 0.16)	(7, 0.14)	(3, 0.16)

^a The values in the parenthesis are (iteration number, cutoff frequency)

^b We used 4 subsets/iteration in the OS-EM reconstruction

^c The cells in red, blue, yellow, and green show the results optimized using approaches (1)-(4) in Table 5-1, respectively.

Table 5-3. Optimal parameters for Tl.

Defect type	Anatomy groups			
	HRLN	LRHN	LRLN	Whole population
1	(5, 0.16)	(3, 0.16)	(5, 0.2)	(5, 0.24)
2	(3, 0.2)	(3, 0.1)	(5, 0.24)	(3, 0.2)
3	(7, 0.12)	(3, 0.16)	(5, 0.14)	(5, 0.14)
4	(3, 0.16)	(3, 0.1)	(10, 0.16)	(3, 0.16)
5	(5, 0.12)	(2, 0.1)	(7, 0.14)	(5, 0.1)
6	(5, 0.08)	(2, 0.08)	(5, 0.1)	(5, 0.08)
1-6	(5, 0.12)	(3, 0.1)	(5, 0.16)	(5, 0.16)

^a The values in the parenthesis are (iteration number, cutoff frequency)

^b We used 4 subsets/ iteration in the OS-EM reconstruction

^c The cells in red, blue, yellow, and green show the results optimized using approaches (1)-(4) in Table 5-1, respectively.

Figure 5-2 shows the ranges of optimal parameters that were not statistically different and where the difference in AUC values was less than 0.01, a threshold deemed as indicating clinical importance, with respect to the set of parameters having the highest AUCs for the anatomy group and the whole population approaches. In these plots, a cross indicates the parameter set that achieved the highest AUC (i.e., the optimal parameter set); the dot shows sets of parameters where the difference in AUC values with respect to the optimal ones were not statistically significant (p-value for a two-tailed t-test is not greater than 0.05). The contour lines indicate parameters for which the difference in AUC with respect to the case of the optimal parameters

was less than 0.01. This was deemed to be a clinically important difference as it is equal to the AUC difference for images reconstructed with and without the scatter or detector response compensation [97]. Again we can see the ranges of optimal parameters were different for different anatomy groups. The LRLN group tended to require a larger number of iterations than the HRLN group and/or higher cutoff frequencies than the LRHN group. The differences in the optimal parameter ranges between three anatomy groups and between patient-specific and non-patient-specific approaches were smaller for Tl than for Tc, likely due to the higher noise level in the Tl images.

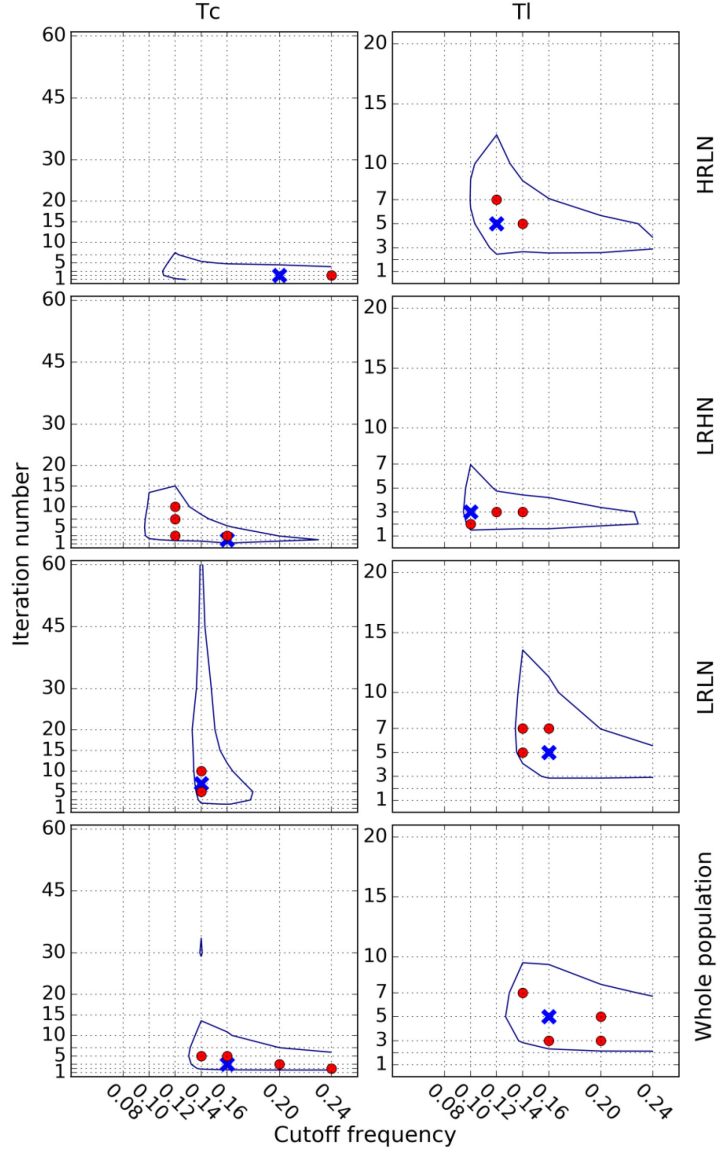


Figure 5-2. Optimal parameter ranges for Tc (left) and Tl (right) using by anatomy group approach (1st row: for group HRLN; 2nd row: for group LRHN; 3rd row: for group LRLN) and using whole population approach (4th row). The crosses represent the set of parameters that achieved the maximum AUC. The dots indicate the parameter points where the difference in AUC with respect to the maximum AUC was not statistically significant. The contour line indicates the regions where AUC values that differed from the optimal ones by no more than 0.01. We used 4 subsets/iteration during the OS-EM reconstruction.

5.3.2 Two-class task performance using parameters obtained from the four optimization approaches

The AUC values and their standard deviation for each anatomy groups and the whole

population are shown in Figure 5-3 for Tc and Figure 5-4 for Tl.

For each anatomy group with mixed defect types and the whole population, the by defect type and anatomy group approach always gave the highest performance in terms of AUCs and the whole population approach always gave the lowest AUCs; the rankings of the other two optimization approaches varied. For the whole population, the AUC differences between the different approaches were statistically significant ($p\text{-value} \leq 0.05$). However, the only optimization approaches with differences > 0.01 , and hence deemed clinically important, were between the defect type and anatomy group approach and each of the other approaches. For Tl, the AUC differences for different approaches were statistically significant (except for the by defect type and by anatomy group approaches, which were essentially equal) but only clinically important between the by defect type and anatomy group approach and the whole population approach.

The ROC curves for the whole population using parameters obtained by the four optimization approaches did not cross except for the by defect type and by anatomy group approaches for Tc. The partial ROC curves with TPF ranges from 0.8 to 0.9, a range that is typical for clinical MPS interpretation [103-105], are shown in Figure 5-5 for Tc and Figure 5-6 for Tl.

We also computed the specificities for sensitivities of 0.80 and 0.90 for Tc and Tl for the whole population using parameters obtained by the four optimization approaches; the results are shown in Table 5-4 and Table 5-5. The specificity differences for the four optimization approaches were larger for Tc than for Tl at the sensitivities of interest. The specificity differences between the practical by anatomy group and whole population approach at sensitivities of interest were about 0.01 for Tc and slightly less than 0.01 for Tl, and thus were on the borderline of clinical importance.

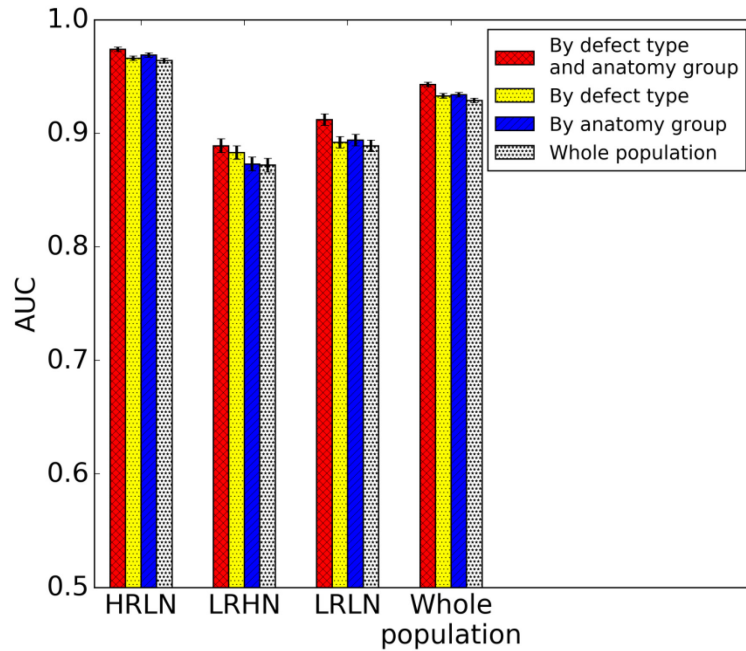


Figure 5-3. Bar graph of AUC values for Tc using optimal parameters obtained with each of the four optimization approaches for the three anatomy groups and the whole population. The error bars represent the standard deviation of the AUCs.

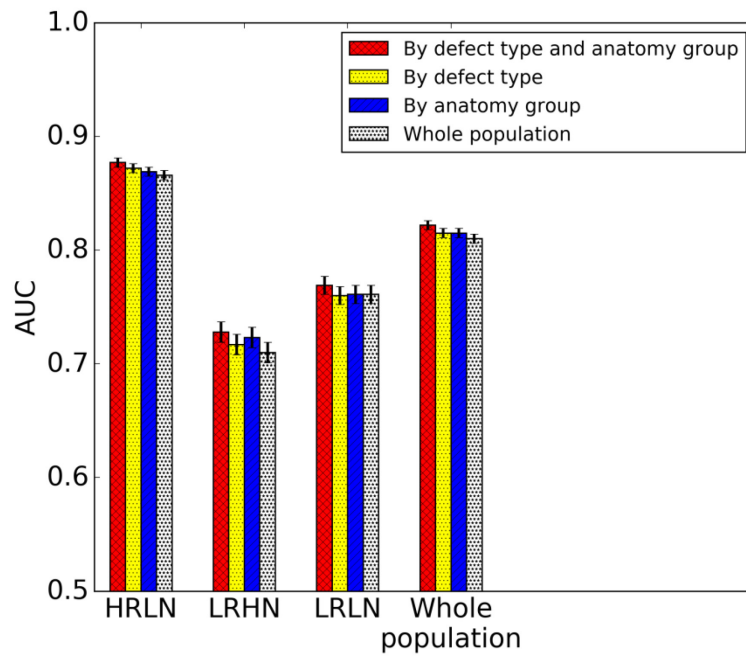


Figure 5-4. Bar graph of AUC values for Tl using optimal parameters obtained with each the four optimization approaches for the three anatomy groups and the whole population. The error bars represent the standard deviation of the AUCs.

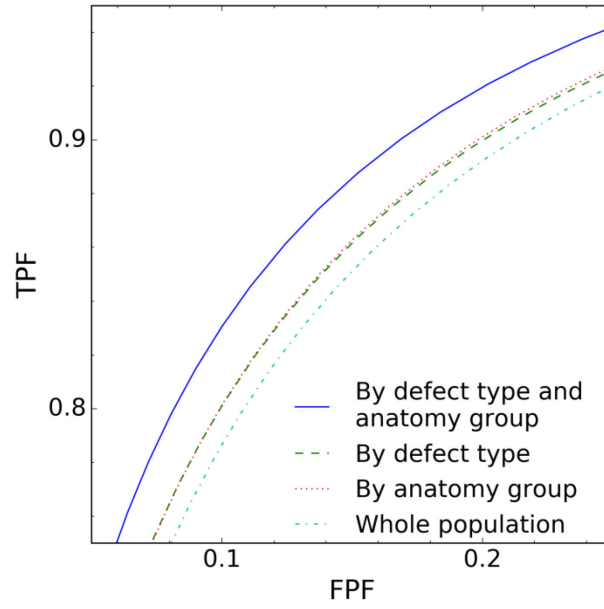


Figure 5-5. Partial ROC curves for Tc for the whole population using optimal parameters obtained by the four optimization approaches.

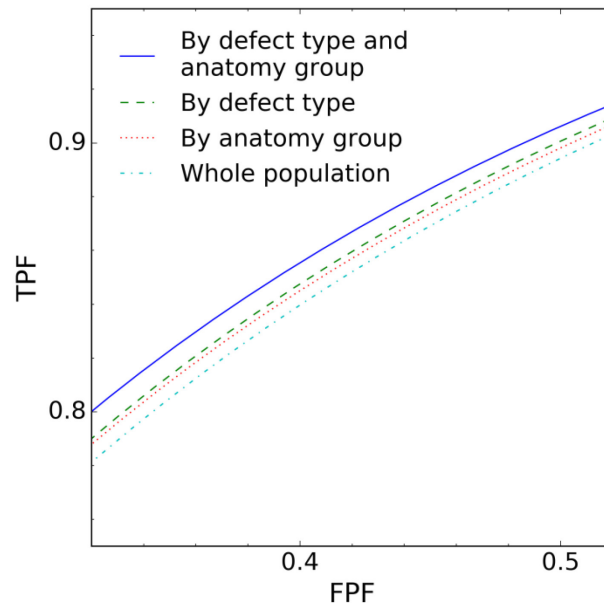


Figure 5-6. Partial ROC curves for Tl for the whole population using optimal parameters obtained by the four optimization approaches.

Table 5-4. Specificity at 0.8 and 0.9 sensitivity for the four optimization approaches for Tc.

Optimization approaches	Sensitivity	
	0.8	0.9
(1) By defect type and anatomy group	0.918	0.832
(2) By defect type	0.901	0.800
(3) By anatomy group	0.901	0.802
(4) Whole population	0.892	0.788

Table 5-5. Specificity at 0.8 and 0.9 sensitivity for the four optimization approaches for Tl.

Optimization approaches	Sensitivity	
	0.8	0.9
(1) By defect type and anatomy group	0.680	0.514
(2) By defect type	0.668	0.502
(3) By anatomy group	0.665	0.496
(4) Whole population	0.657	0.487

5.3.3 Three-class task performance using parameters obtained from the four optimization approaches

The three-class task performances of the four optimization approaches were compared using the VUS values, as shown in Figure 5-7. For each anatomy group and the whole population, the four optimization approaches ranked from high to low were: by defect type and anatomy group, by defect type, by anatomy group and whole population (except that the by defect type and by anatomy group approaches ranked the same for some patient anatomy groups and the by anatomy group and the whole population approaches ranked the same for LRLN group). The differences in the VUS values were statistically significant for the 4 approaches (except those that had same rankings). For the whole population, the by anatomy group approach had a VUS more than 0.01 larger than the non-patient-specific whole population approach.

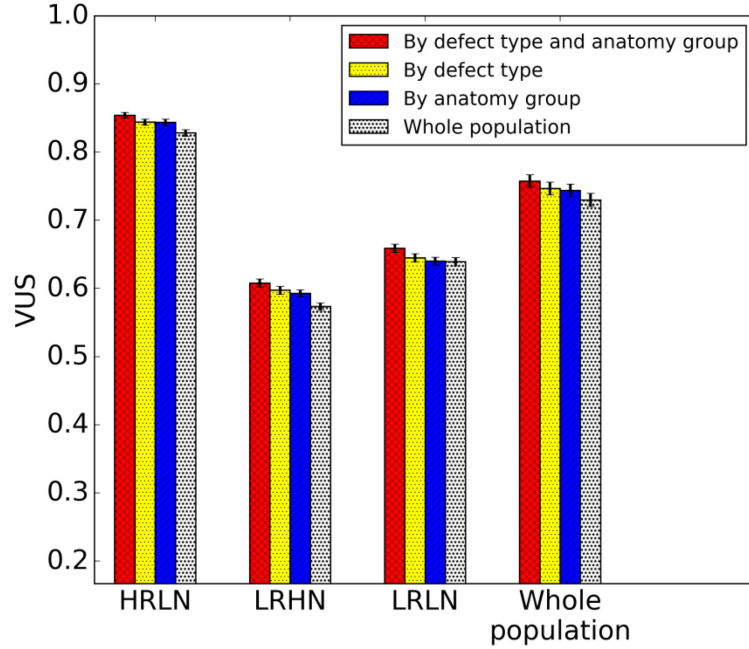


Figure 5-7. Bar graph of VUS values using optimal parameters obtained by the four optimization approaches. The error bars represent the standard deviations of the VUS values.

5.4 Conclusion

The work in this chapter verified the hypothesis that patient-specific optimization of reconstruction parameters provided better task performance. The more patient-specific information used during the optimization process, the better the overall performance: optimizing by patient anatomy groups, based on indices including the image resolution, heart size and (relative) image noise, and by defect type resulted in parameters that gave equal or higher performance for all the groups and for the population as a whole. In practice, only optimization by patient anatomy groups is feasible since defect types are not easily known beforehand. The proposed practical optimization method that optimizes parameters by patient anatomy group using the multi-template LD with pooled test statistics observer strategy resulted in better two- and three-class classification task performance than the non-patient-specific optimization method for the simulated MPS data and task investigated here. Although our results came from the simultaneous dual isotope studies, it can also be applied to MPS using separate acquisition or single radionuclides.

6. Toward patient-specific optimization of a maximum a *posteriori* reconstruction algorithm for simultaneous-acquisition dual-isotope myocardial perfusion SPECT

6.1 Introduction

Maximum a *posteriori* (MAP) algorithms have been widely used to regularize the ill-posed emission computed tomography reconstruction problem. Various priors, each with its specific properties, have been proposed. He *et al.* have previously developed a MAP algorithm based on a dual-tracer prior (DTMAP) for simultaneous-acquisition dual-isotope SPECT imaging [1]. The dual-tracer prior in [1] can couple the information from the images of the two isotopes since the two images are perfectly registered in space. The hypothesis is that this allows improvement of the quality of the resulting images. This DTMAP algorithm is reviewed in section 2.4. Although this algorithm has the advantages of preserving edges and reducing noise in uniform regions, appropriate values of three hyperparameters must be specified in order to achieve good image quality. Previously, this DTMAP algorithm was shown to produce images with better image quality compared to optimized OS-EM in terms of three-class classification task performance for a simulated simultaneous-acquisition dual-isotope MPS study based on NCAT [106] phantom population [1]. However, this previous study had some limitations. First, the NCAT phantom is somewhat less realistic in its modeling of human anatomy than the state-of-the-art XCAT phantom. Secondly, the defect size and contrast used in this previous study were not in a range where detection was not particularly challenging. Thirdly, that study did not simulate crosstalk between the two isotopes. Finally, only one hyperparameter was optimized with the other two fixed. As a result of these limitations, further evidence is needed that DTMAP is better than OS-EM for more clinically realistic data before embarking on expensive clinical evaluations of the

method. In this Chapter, we aim to provide more complete and patient-specific hyperparameter optimization methods for DTMAP with crosstalk compensation. The methods were evaluated using simulated phantoms based on more realistic simulations of a phantom population based on the XCAT phantom with clinically relevant and challenging defect types. The goal was to study the performance of DTMAP in this context with respect to post-filtered OS-EM.

Optimizing the hyperparameters for MAP in general and DTMAP in particular is difficult because there is no straightforward and intuitive way to select the parameters. One possible method is to optimize the parameters through an exhaustive grid-search. However, for DTMAP it is much more computationally expensive to produce images for a single hyperparameter setting than for OS-EM, where saving intermediate iterations and post-reconstruction filtering makes searching for the optimal number of iterations and filter cutoff frequency relatively computationally inexpensive. As an example, for DTMAP, the reconstruction of one set of rest and stress images using one set of optimal hyperparameters can take 2-5 hours of computer time; for OS-EM, generating images for all iterations and post-reconstruction filtering take about 5 minutes for one phantom. Thus, for DTMAP, a brute force grid search method for the optimization of three hyperparameters using observer studies is impractical.

In this chapter, we propose a practical method to obtain patient-specific hyperparameters for DTMAP by searching the hyperparameters in two ranges that were thought to provide good performance in terms of defect detection. In the method, patient-specific optimization was achieved by using a group-wise optimization approach that optimized the parameters for each anatomy group separately, as discussed in Chapter 5. The results provide insight into some of the properties of the hyperparameters. We compared the task performance of optimized DTMAP to optimized post-filtered OS-EM. Also, we studied some of the characteristics and limitations of the DTMAP algorithm and the images obtained from it.

6.2 Methods

6.2.1 Motivation and assumptions of the hyperparameter optimization

methods

In this section, we propose to optimize the hyperparameters, β , δ , and η , for DTMAP by searching among two possible hyperparameter ranges, as discussed below. By restricting the hyperparameter ranges, the number of hyperparameters that need to be optimized through the computationally-expensive observer studies is also reduced, and thus the optimization process is accelerated. Note that the optimal hyperparameters obtained should be considered optimal in a restricted sense since they are restricted to these specific ranges, and thus cannot be guaranteed to be global optima.

6.2.1.1 Motivation

When δ and η are greater than the differences in neighboring voxels, i.e., $\delta > \mathbf{f}_r[j] - \mathbf{f}_r[k] \equiv \Delta \mathbf{f}_r$ and $\eta > \mathbf{f}_s[j] - \mathbf{f}_s[k] \equiv \Delta \mathbf{f}_s$, resulting $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^4 \rightarrow 0$ and $\left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^4 \rightarrow 0$. Then it is reasonable to use a Taylor Series expansion and ignore the higher order terms $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^i$ and $\left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^i$ ($i \geq 4$). In this case, the prior term defined by equation (2-9) becomes

$$\beta \Lambda(\mathbf{f}_r, \mathbf{f}_s) \approx \sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \left(\frac{\beta}{2\delta^2} \Delta \mathbf{f}_r^2 + \frac{\beta}{2\eta^2} \Delta \mathbf{f}_s^2 \right), \quad (6-1)$$

and the prior acts like a quadratic prior, which tends to smooth large differences more than small differences, and thus tends to blur edges. In this dissertation, we will refer to MAP reconstruction

with a quadratic prior $\sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \left(\frac{\beta}{2\delta^2} \Delta \mathbf{f}_r^2 + \frac{\beta}{2\eta^2} \Delta \mathbf{f}_s^2 \right)$ as QMAP.

When δ and η are smaller than the differences in neighboring voxels, i.e., $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^2 \gg 1$,

the prior becomes

$$\beta\Lambda(\mathbf{f}_r, \mathbf{f}_s) \approx \sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \sqrt{\left(\frac{\beta\Delta\mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\beta\Delta\mathbf{f}_s}{\eta}\right)^2}, \quad (6-2)$$

and the prior is similar to a total variation prior[107] and thus tends to be more edge-preserving than the quadratic prior.

6.2.1.2 Hyperparameter searching ranges

A previous study demonstrated QMAP had a worse SNR at the location of the signal than post-filtered ML-EM for matched spatial resolutions [108]. Note that in that study, the SNR for voxel j in that study is defined as

$$SNR[j] = \frac{\langle \hat{f}_{H_1}[j] \rangle - \langle \hat{f}_{H_0}[j] \rangle}{\sqrt{\frac{1}{2}(\sigma_{H_1}^2[j] + \sigma_{H_0}^2[j])}} \quad (6-3)$$

where $\langle \hat{f}_{H_i}[j] \rangle$ and $\sigma_{H_i}^2[j]$ are the mean and variance of the voxel j from multiple noise realizations under the hypothesis H_i ($i=0, 1$).

Thus, we would expect the QMAP to have worse performance in terms of AUC compared to the optimized post-filtered OS-EM when their resolutions are matched. Thus, we restrict the search for optimal hyperparameters to two ranges, discussed below, which are expected to result in better task performance for DTMAP than post-filtered OS-EM. Limiting the range of hyperparameters explored in this way reduces the size of the space that must be searched to find optimal values of these hyperparameters.

The first hyperparameter range results in a near-quadratic prior in uniform regions and is more edge-preserving at edges (note that the prior at edges is still defined by Equation (2-9)). Since we are trying to detect the defects in the myocardium, δ and η should be greater than the noise inside the myocardium so that $\left(\frac{\Delta\mathbf{f}_r}{\delta}\right)^4 \rightarrow 0$ and $\left(\frac{\Delta\mathbf{f}_s}{\eta}\right)^4 \rightarrow 0$ inside the myocardium. The

same terms should not be much greater than the differences in neighboring voxels around the defect edges so that $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^4$ and $\left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^4$ do not approach 0, and the prior is not quadratic at the edges. Note that this is possible only for images where the amplitude of noise fluctuations is smaller than the amplitude of the edges. In such a hyperparameter range, DTMAP is expected to smooth noise fluctuations in uniform regions and preserve the edges. The balance of smoothing and preservation of edges is determined by the choices of δ and η . In this work, we refer to this hyperparameter range as the noise-smoothing-edge-preserving range.

The second hyperparameter range investigated is where the prior is more edge-preserving even at uniform regions. If δ and η are small compared to the noise in the myocardium, resulting $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^2 \gg 1$, the prior can be approximated by Equation (6-2) and the total number of hyperparameters is reduced to two instead of three. The justification for this range is that the choice $\delta = \eta = 1$ in a previous study resulted in better performance for DTMAP than post-filtered OS-EM [1]. In this work, this hyperparameter range is referred to as the edge-preserving range.

In the following, we propose methods to optimize the hyperparameters in these two ranges. The set of hyperparameters in the two ranges that achieve the better 3-class task performance was deemed optimal.

6.2.2 Hyperparameter optimization in the noise-smoothing-edge-preserving range

We proposed the following method to optimize the hyperparameters for DTMAP in the noise-smoothing-edge-preserving range discussed in 6.2.1.2.

First, we need to verify that the amplitude of noise in the myocardium is less than the amplitude of defect edges. We used the average number of detected counts to obtain the amplitude of the noise. The average number of detected counts per voxel was obtained by

dividing the sum of voxel values in the myocardium from an image reconstructed using filtered back projection without applying post-smoothing filters and the number of voxels in the myocardium volume. The myocardial volume could, in practice, be estimated from contrast CT images. In this study, we used the true myocardial volume calculated from the voxelized phantom image. Suppose the average number of detected counts is N_0 ; then the corresponding noise is $\sqrt{N_0}$. The differences in neighborhood voxels around the defect edges was calculated as

$$\Delta f_{s,edges} = Severity \times C_s, \quad (6-4)$$

$$\Delta f_{r,edges} = Severity \times C_r, \quad (6-5)$$

where *Severity* is the defect severity defined as the ratio of the activity concentrations in defect and normal myocardium, $\Delta f_{s,edges}$ and $\Delta f_{r,edges}$ are the differences in reconstructed voxels inside the defect compared to normal myocardium in the stress and rest images, respectively, and C_s and C_r are the average counts per voxel in the myocardium in the reconstructed images for stress and rest images, respectively. The values of defect severity used in this work are given in Table 3-1. The values of C_s and C_r can be calculated using the total number of counts in the myocardium in the stress and rest images and the volume of the myocardium. The total number of counts in the myocardium was estimated from images reconstructed by OS-EM using a large number of iterations (4 subsets/iteration and 20 iterations) with attenuation correction and no post-filtering. The use of a large number of iterations and no post-filtering in OS-EM reduces partial volume effects and is expected to produce an accurate estimate of the total counts in the myocardium. The values of the ratio of the noise $\sqrt{N_0}$ to the smallest (Δf_{edges} with the 10% severity) and largest (Δf_{edges} with the 50% severity) edges of interest are summarized in the following table. The ratios were less than 1, indicating the amplitude of noise was less than the amplitude of the edges of interest.

Table 6-1. Ratio of image noise to the smallest edges of interest (with 10% severity) for Tc and Tl images.

	Tc (stress image)	Tl (rest image)
min	0.19	0.27
max	0.56	0.66
average	0.33	0.44

Table 6-2. Ratio of image noise to the largest edges of interest (with 50% severity) for Tc and Tl images.

	Tc (stress image)	Tl (rest image)
min	0.04	0.05
max	0.11	0.13
average	0.07	0.09

The optimization method is discussed below.

The first step of the optimization method is to find the quadratic smoothing parameters

$\frac{\beta}{2\delta^2}$ and $\frac{\beta}{2\eta^2}$ in the myocardium. We proposed to do this by choosing $\frac{\beta}{2\delta^2}$ and $\frac{\beta}{2\eta^2}$ that match the resolution of QMAP to the optimized OS-EM in the myocardium for rest and stress images, respectively. We then determined the value of η that optimally balances the smoothing of noise and preservation of edges. Mathematically,

$$\eta = k_s \times C_s, \quad (6-6)$$

where C_s is the average counts per voxel in the myocardium for the stress image mentioned in the previous paragraph and k_s is to be determined by observer studies. Note that η is count-level dependent (patient-specific). For an image with a given value of C_s , the value of k_s controls the values of differences in neighboring voxels where the prior is more edge preserving versus more quadratic (smoothing). Note that a smaller value of k_s means smaller differences in neighborhood voxels will be treated as edges to be preserved. For each value of k_s , the DTMAP images can be reconstructed using η calculated from Equation 6-6 and β and δ calculated subsequently using $\frac{\beta}{2\delta^2}$ and $\frac{\beta}{2\eta^2}$ obtained by matching the resolution of DTMAP to OS-EM.

We selected the value of k_s that gave the best 3-class task performance through observer studies. Note that by this method the number of the parameters that must be optimized through observer studies is reduced to one, i.e., k_s .

Since we assume the prior is quadratic inside the myocardium, we need $\eta \gg$ the noise amplitude in the myocardium. In order to preserve defect edges, k_s should not be much greater than the edge amplitude. Thus we searched for values of k_s around 0.1. In particular, we investigated k_s with values equal to 0.03, 0.05, 0.1, and 0.2. We generated reconstructed images for 10 uptake realizations for each anatomy and defect type. We found the k_s value among those tested that gave the highest 3-class task performance for each anatomy group (HRLN, LRHN, LRLN), as discussed in Chapter 5. The values of β , δ , and η obtained using this method with the optimal value of k_s were deemed to be optimal.

Since this method does not ensure that the value of δ is in the noise-smoothing-edge-preserving range, it is necessary to verify that. Suppose

$$\delta = k_r \times C_r, \quad (6-7)$$

where C_r is the average counts per voxel in the myocardium for rest images, as mentioned above, and k_r is the parameter that defines the relationship between δ and C_r . With the data in this study, the value of k_r should be approximately 0.1 in order for δ to be in the noise-smoothing-edge-preserving range.

The above method requires matching of the resolution of QMAP to that of OS-EM. This requires a criterion for having equal resolutions, which is difficult to specify for the following reasons. First, the shapes of the point spread functions (PSFs) for QMAP and post-filtered OS-EM are different [109]. The PSF for OS-EM is near-Gaussian, while for QMAP it is very non-Gaussian. For this reason, using the full width at half maximum (FWHM) of the PSFs as a standard to match the resolution between the two algorithms is not reasonable. Second, for both algorithms, the image resolution is position dependent. The position dependency of the PSF is

especially hard to characterize for QMAP because its PSF is very dependent on the background [109]. The position dependency of the PSF for post-filtered OS-EM is arises largely from the position-dependent collimator-detector response; post-filtering of OS-EM, tends to reduce the position-dependence. Since the position dependences for QMAP and OS-EM are different, even if we can match the PSF of the two algorithms at one position, the resolution at other positions likely will not be matched. To overcome these difficulties, we matched the resolution of the two algorithms by minimizing the mean square error (MSE) of reconstructed images from the two algorithms in a region surrounding the myocardium. This region was defined as the volume just surrounding the left ventricle and chamber dilated by 3 voxels in all directions, but excluding the volume containing the liver after dilating it by 4 voxels. Sample images of QMAP and OS-EM with matched resolution are shown in Figure 6-1; the profile along the vertical line passing through the center of left-ventricle myocardium (as shown in the left image of Figure 6-1) is plotted in Figure 6-2. As can be seen in these figures, the contrast of the myocardium with respect to the background (which is affected by resolution near the myocardium) for DTMAP is comparable to that for OS-EM.

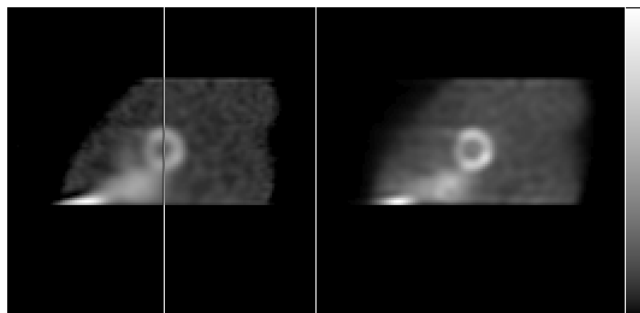


Figure 6-1. Sample images of QMAP (left) and post-filtered OS-EM (right) with matched resolution. The line indicates the position of the profile plotted in Figure 6-2.

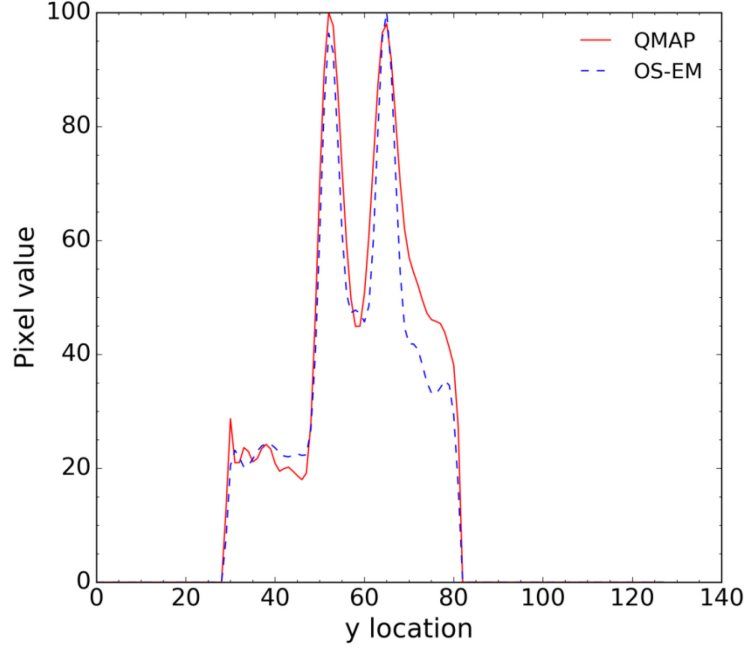


Figure 6-2. Profile along the vertical line passing through the center of the left ventricular myocardium of QMAP and post-filtered OS-EM images with matched resolution shown in Figure 6-1.

6.2.3 Hyperparameter optimization in the edge-preserving range

We proposed the following method to optimize hyperparameters for DTMAP in the edge-preserving range discussed in 6.2.1.2. From the theory in 6.2.1.1, when δ and η are much smaller than the noise amplitude in the myocardium, the prior becomes

$$\beta\Lambda(\mathbf{f}_r, \mathbf{f}_s) \approx \sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \sqrt{\left(\frac{\beta\Delta\mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\beta\Delta\mathbf{f}_s}{\eta}\right)^2} \quad (6-8)$$

in the myocardial region. Since the amplitude of the defect edges is greater than the amplitude of the noise in the myocardium, the prior can still be approximated by (6-8) in the region near the myocardium. This means the value of the prior depends on the values of these two ratios $\frac{\beta}{\delta}$ and $\frac{\beta}{\eta}$ rather than on all 3 parameters.

Since the noise amplitudes defined in 6.2.2 by the square root the average number of detected counts in the myocardium were 5.3 and 3.9 for stress and rest images, respectively, we

need $\left(\frac{3.9}{\delta}\right)^2 + \left(\frac{5.3}{\eta}\right)^2 \gg 1$, so, the prior becomes $\beta\Lambda(\mathbf{f}_r, \mathbf{f}_s) \approx \sum_{j=1}^J \sum_{k \in N_j} \omega_{jk} \sqrt{\left(\frac{\beta\Delta\mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\beta\Delta\mathbf{f}_s}{\eta}\right)^2}$.

We empirically verified of the above assertion as follows. Reconstructed images for 3 phantoms are shown in Table 6-3 obtained with the ratios $\beta/\eta=\beta/\delta$ equal to 0.001, 0.1, 0.5, 1, and 5, for values of $\eta=\delta$ equal to 100, 10, 1, and 0.1 (in each case the value of β was chosen to give the specified ratio). The reconstructed images are shown in Figure 6-3 to Figure 6-8. From the images, we can see that, when η and δ are sufficiently small (less than 10), the images with the same ratios of β/η and β/δ , i.e., the images in the columns of these figures, are visually similar.

Table 6-3. Parameters of phantoms used to verify the effect of edge-preserving assumption.

Phantom notation	Gender	Body core size	Heart size	Subcutaneous adipose tissue level	Anatomy group
M.BS.HS.FS	Male	Small	Small	Small	LRLN
M.BM.HS.FS	Male	Medium	Small	Small	HRLN
F.BL.HL.FL	Female	Large	Large	Large	LRHN

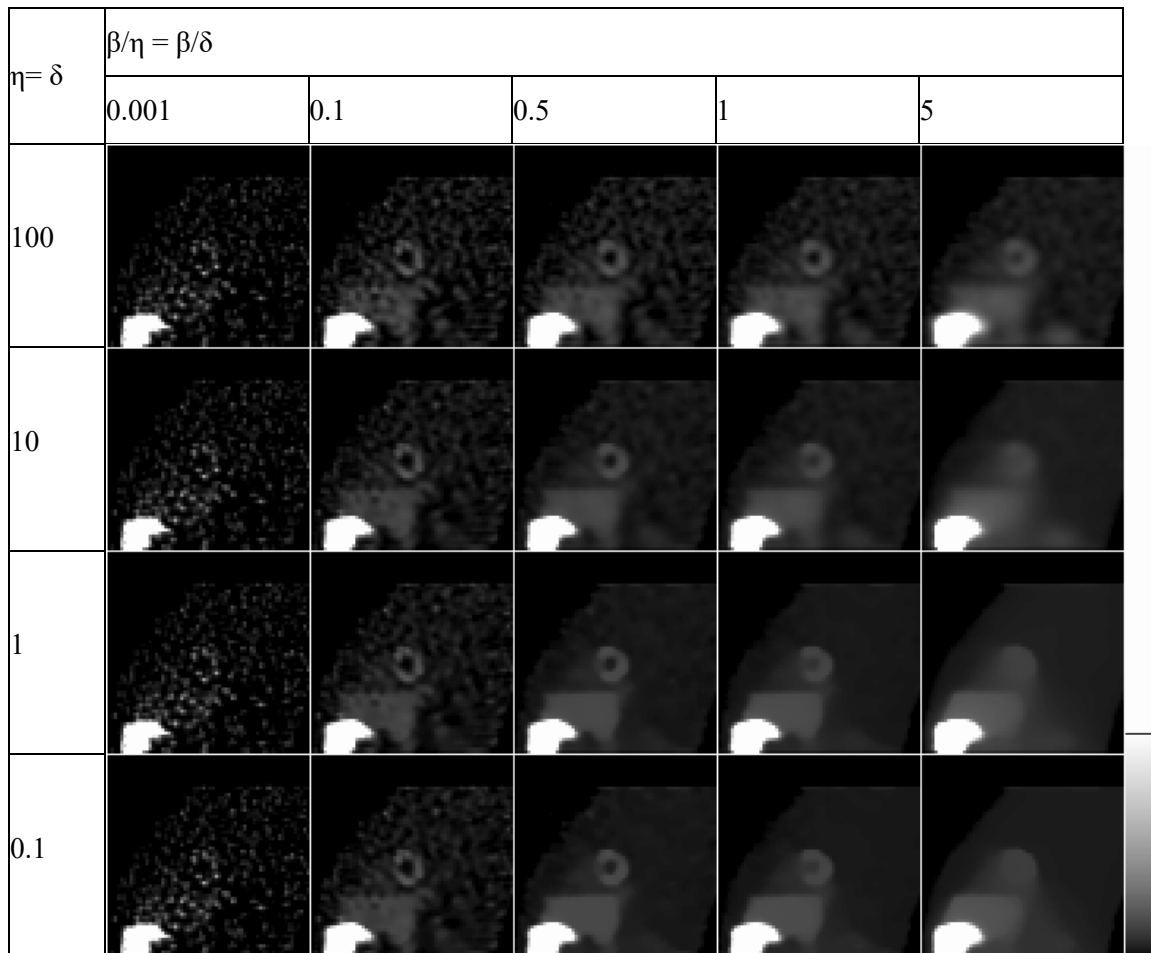


Figure 6-3. Stress images of the phantom M.BS.HS.FS (Male, small body core size, small heart size, and small subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

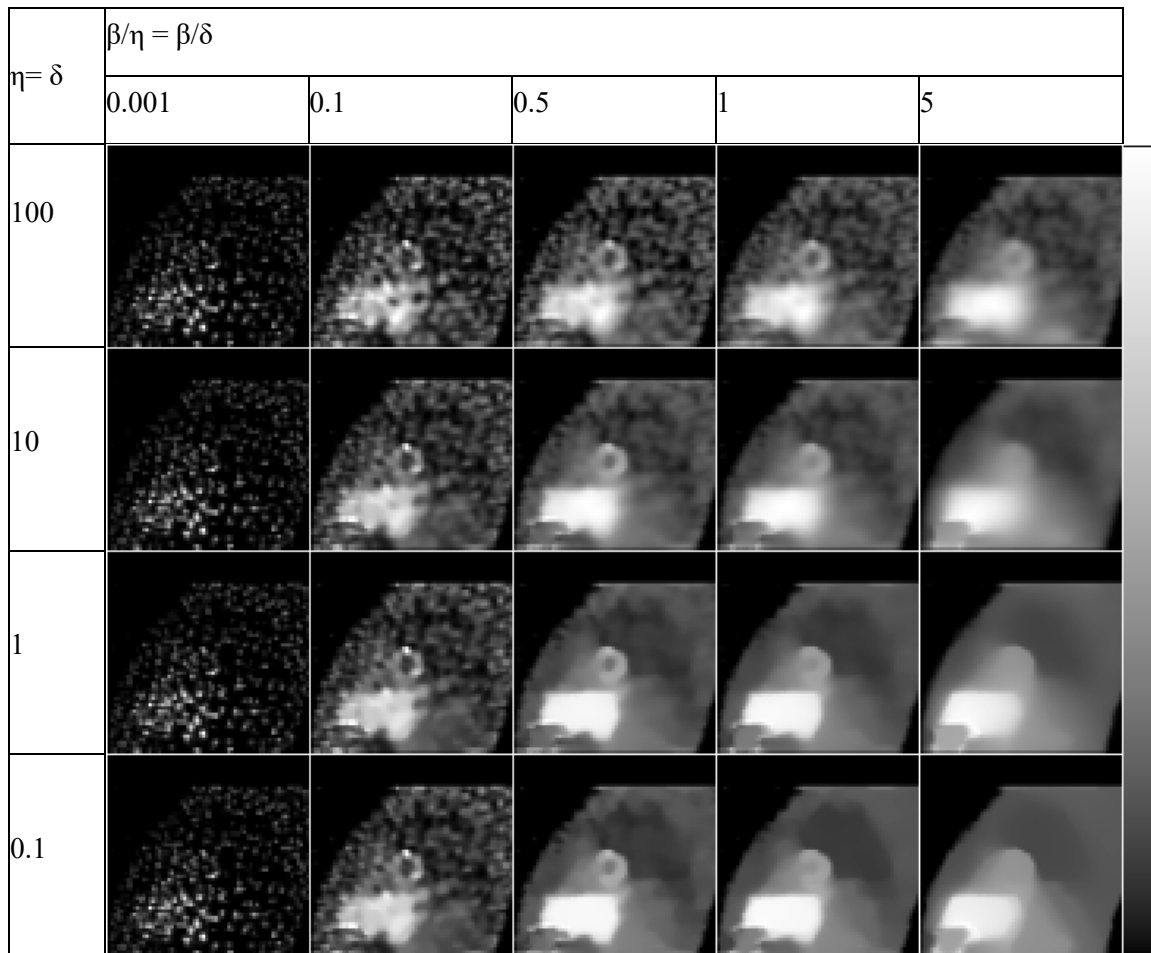


Figure 6-4. Rest images of the phantom M.BS.HS.FS (Male, small body core size, small heart size, and small subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

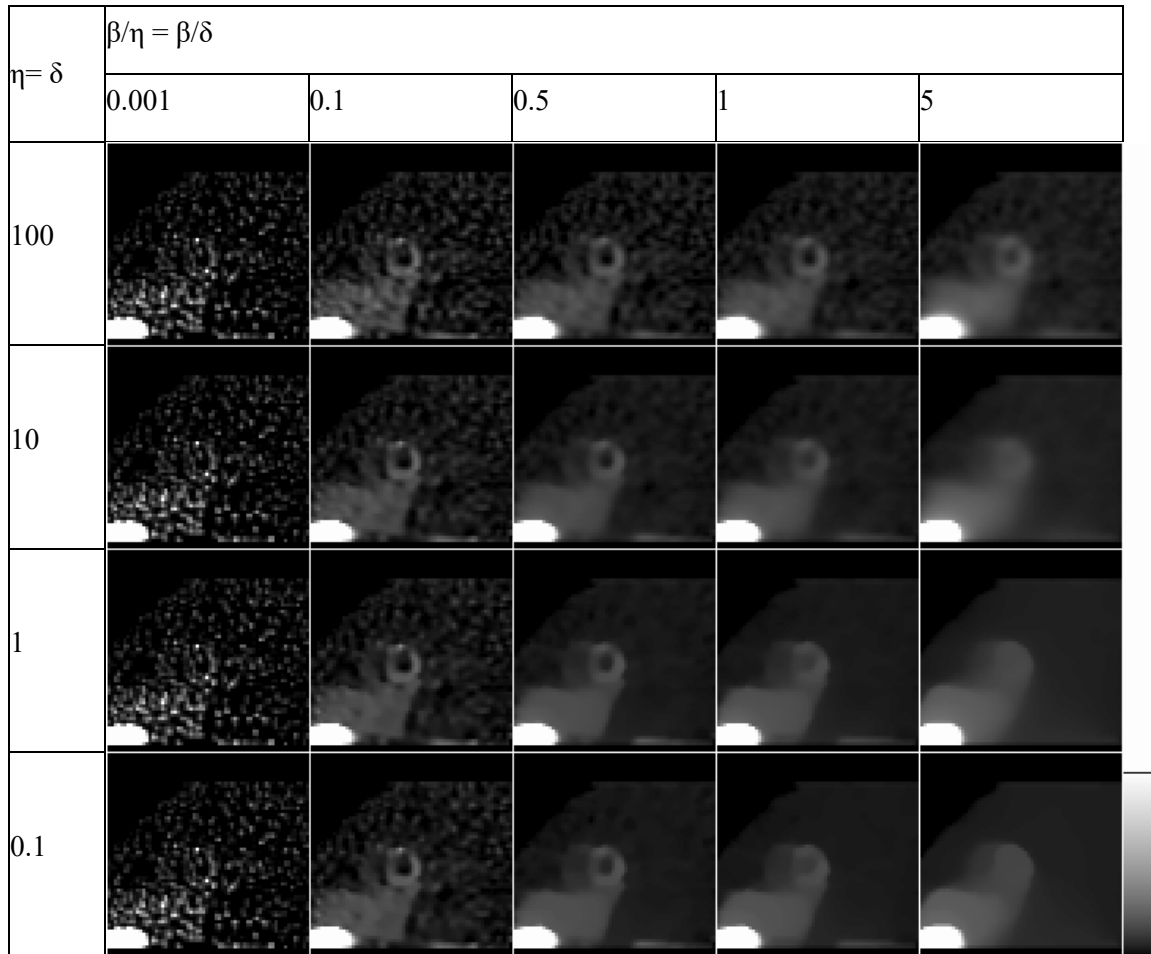


Figure 6-5. Stress images of the phantom M.BM.HS.FS (Male, medium body core size, small heart size, and small Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

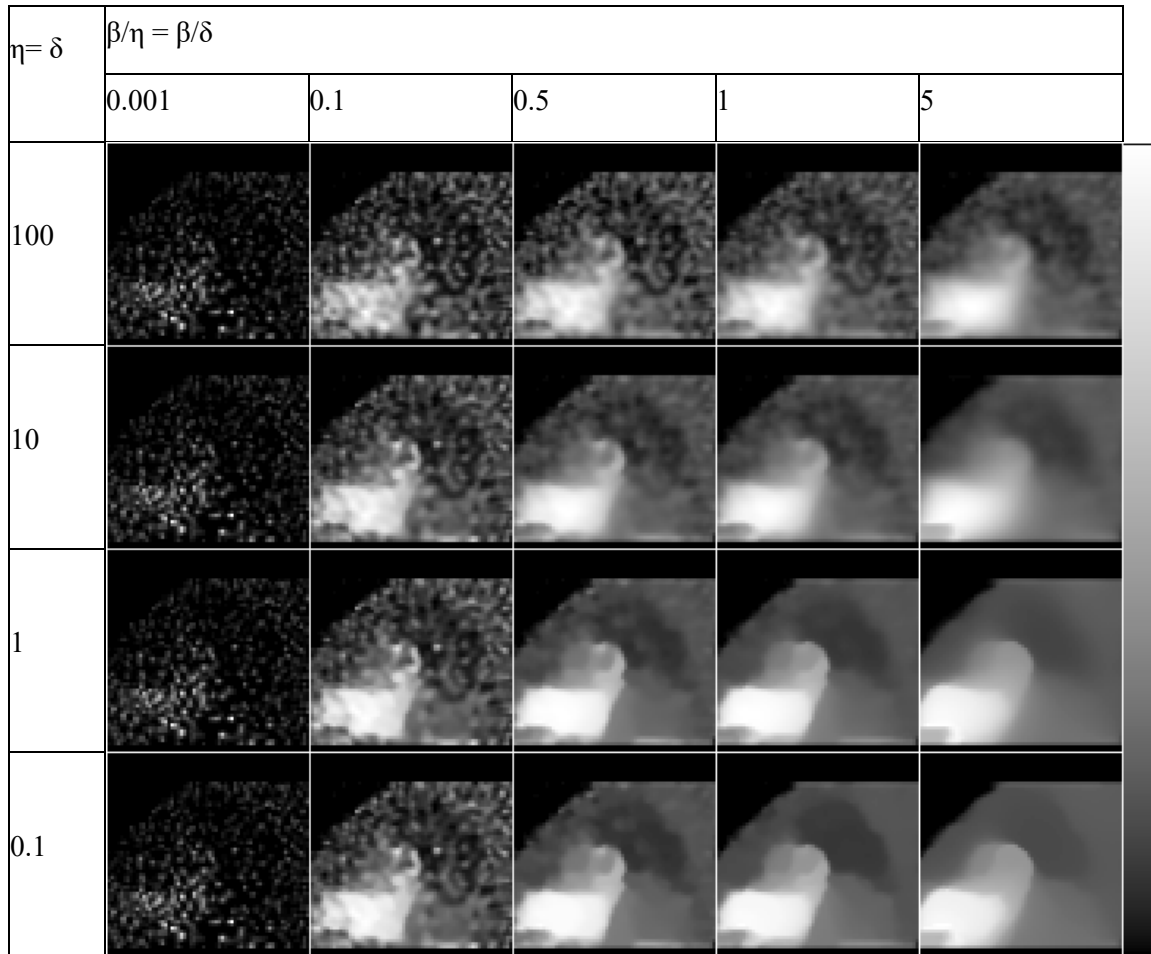


Figure 6-6. Rest images of the phantom M.BM.HS.FS (Male, medium body core size, small heart size, and small Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

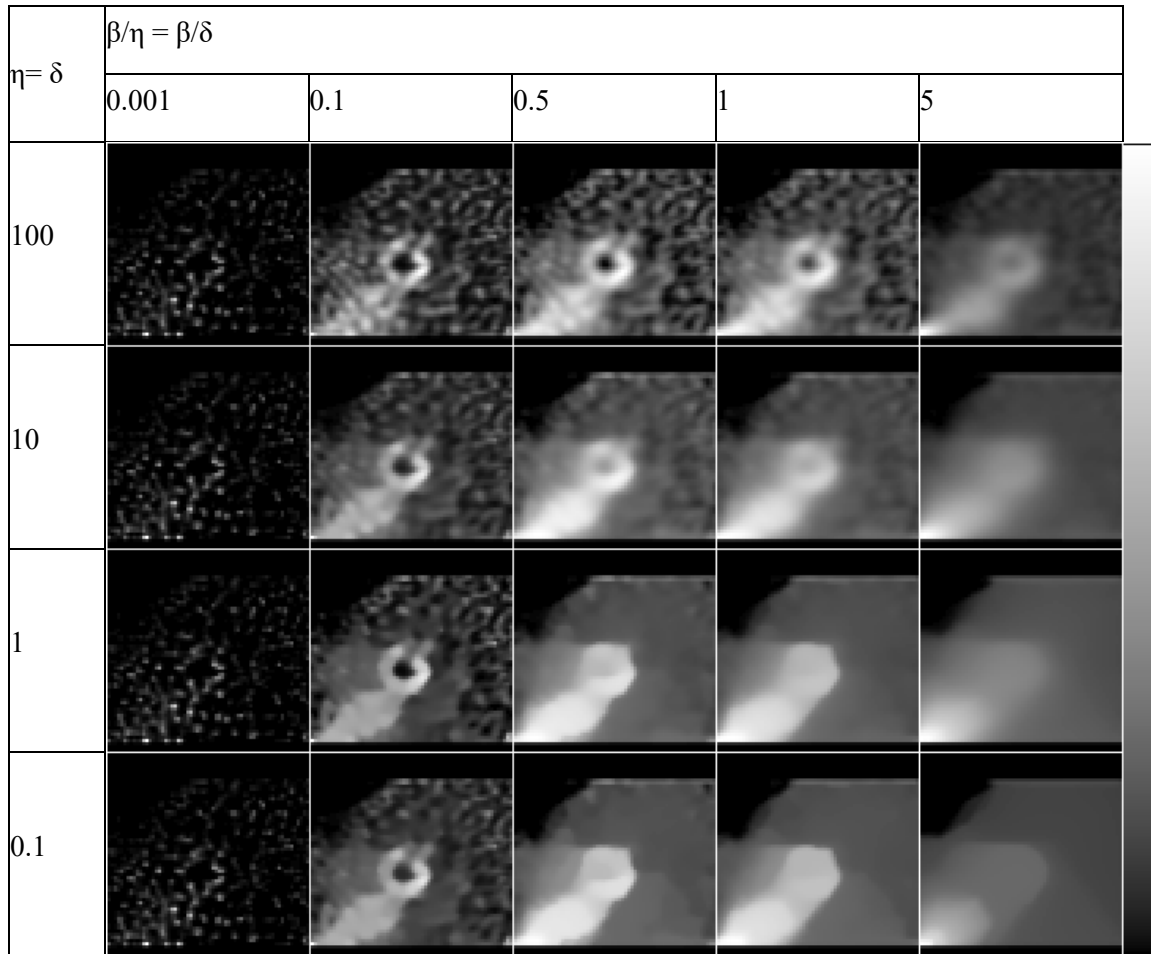


Figure 6-7. Stress images of the phantom F.BL.HL.FL (Female, large body core size, large heart size, and large Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

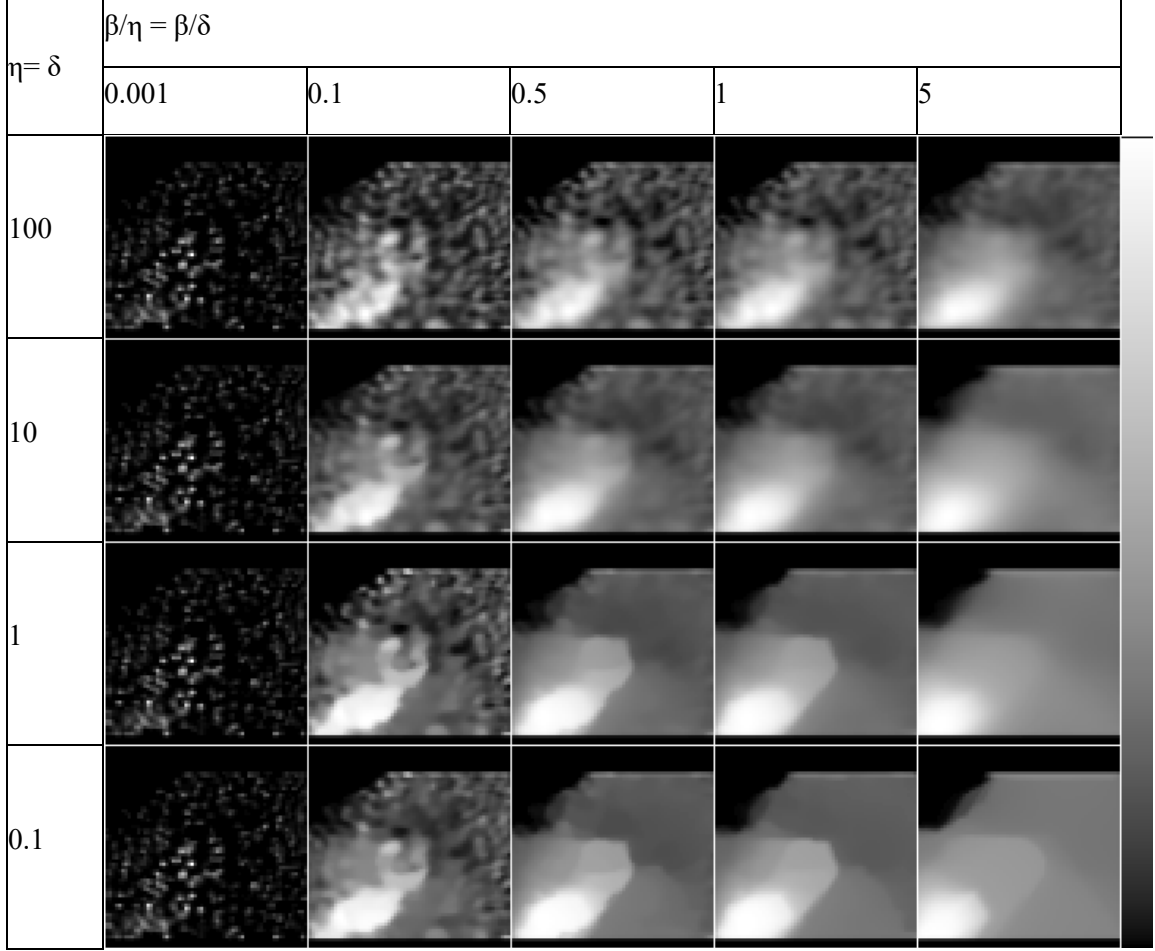


Figure 6-8. Rest images of the phantom F.BL.HL.FL (Female, large body core size, large heart size, and large Subcutaneous adipose tissue level) with the $\beta/\eta=\beta/\delta$ ratios and $\eta=\delta$ values indicated.

Since there are only two parameters, β/η and β/δ , to be determined, we searched for the optimal values of these ratios using a grid search method.

We generated the DTMAP images with η equal to 1 (sufficiently small compared to the noise in the myocardium), β with values equal to 0.1, 0.3, 0.5, 0.8, and 1, and δ with values equal

to $\sqrt{10/3}$, $\sqrt{2}$, and 1. These values were chosen so that $\left(\frac{\Delta \mathbf{f}_r}{\delta}\right)^2 + \left(\frac{\Delta \mathbf{f}_s}{\eta}\right)^2 \gg 1$ and so that sample

reconstructed images qualitatively covered a wide range of visual image noise and resolution tradeoffs. We generated reconstructions for 10 uptake realizations for each anatomy and defect type. We optimized β and δ in terms of the VUS for the 3-class task for each anatomy group (HRLN, LRHN, and LRLN) as discussed in Chapter 5. The combination of parameters that

achieved the maximum 3-class task performance for each anatomy group was deemed to be optimal for that group.

6.2.4 Image quality evaluation

In order to thoroughly compare the quality of images reconstructed using the DTMAP and OS-EM methods, we computed the VUS as a metric of 3-class task-based performance as well as values of some typical physical image quality metrics. We computed three-class task performance in terms of VUS using the same observer method illustrated in Section 5.2.2.

To calculate noise metrics, we generated 100 noise realizations for each uptake realization for each phantom. We calculated two noise metrics. One is the average coefficient of variation (COV) in the myocardium over all noise realizations. To calculate this, we first computed the standard deviation and mean image over the 100 noise realizations. Each voxel value in the standard deviation image was equal to the standard deviation of that voxel over all the noise realizations. Each voxel in the mean image was equal to the mean of that voxel over all noise realizations. The voxel value in the COV image was equal to the ratio of the voxel in the standard deviation image to the voxel in the mean image. The average COV in the left ventricular myocardium was computed by averaging the COV over the normal left ventricular myocardium. We refer to this value as the average COV. The second noise metric was computed by calculating the standard deviation over all the voxels in the normal left ventricular myocardium in one image, and then averaging this over all noise realizations. We refer to this average standard deviation as the average noise fluctuations since this tells us something about the fluctuation of voxel values in the myocardium.

We also compared the resolution properties of the OS-EM and DTMAP reconstructed images using the local point spread function (PSF). Because both reconstruction algorithms are nonlinear, the local PSFs were generated using perturbation methods. In this method, the PSF is the difference in the reconstructed images from projections with and without a voxel source

perturbation at the point of interest in the object. In this study, we calculated the local PSF at the centroid of the defect position with a voxel source having an activity concentration of 3% of the average activity concentration in the myocardium.

The local PSFs of images reconstructed with the two reconstruction methods were compared visually and quantitatively using FWHMs defined as follows. Please note that since the shapes of the local PSFs were not the same, the FWHM does not fully characterize the PSF. However, it can still be used as a partial measure of the resolution.

The local PSF for DTMAP was observed to be asymmetric, i.e., the 1D FWHMs were different along different radial directions. To calculate the mean resolution in all directions in 2D and 3D image space, we calculated 2D and 3D FWHMs of the local PSF. The 2D FWHM was defined the same as in [110], as:

$$\text{2D FWHM} = \sqrt{\frac{\text{area of the 2D PSF contour at half maximum}}{\pi / 4}}. \quad (6-9)$$

We defined the 3D FWHM similarly, as:

$$\text{3D FWHM} = \sqrt[3]{\frac{\text{volume of the 3D PSF contour at half maximum}}{\pi / 6}}. \quad (6-10)$$

To study the resolution properties of the two reconstruction algorithms, we studied the dependence of 2D and 3D FWHMs on location, phantom type, and tracer uptake realization.

6.3 Results

6.3.1 Optimal parameters

6.3.1.1 In the noise-smoothing-edge-preserving range

The optimal values of k_s in the noise-smoothing-edge-preserving range are shown in Table 6-4. Note that the HRLN and LRLN groups had similar values of optimal k_s , while the LRHN group had a smaller optimal k_s . This observation can be explained as follows. As seen from Section 5.2.3, the LHRN group had the smallest defect-to-FWHM ratio, resulting in lower

defect contrast; as a result, this group may require smaller values of k_s to retain the defect.

Table 6-4. Optimal values of k_s for different anatomy groups obtained in the noise-smoothing-edge-preserving range.

Anatomy group	Optimal k_s
HRLN	0.05-0.1
LRHN	0.03
LRLN	0.05-0.1

* Ranges of optimal k_s indicate the ranges of values where the difference in VUS was not statistically significant.

The optimal values of k_r obtained while using the optimal k_s values in Table 6-4 are summarized in Table 6-5.

The values of k_s and k_r in Table 6-4 and Table 6-5 are around 0.1, except for the LRHN group. Based on the discussions in 6.2.2, if k_s and k_r are smaller than 0.1, the prior is not nearly quadratic inside the myocardium. As a result, we may expect the images to be noisy inside the myocardium for the LRHN group, especially for the rest images.

Table 6-5. The mean, minimum, and maximum values of k_r over all the population and uptake realizations for each anatomy groups obtained using the optimal k_s values.

Anatomy group	Mean	Minimum	Maximum
HRLN	0.15	0.07	0.25
LRHN	0.02	0.01	0.05
LRLN	0.07	0.04	0.12

6.3.1.2 In the edge-preserving range

The optimal hyperparameters obtained in the edge-preserving range are summarized in Table 6-6 for $\eta=1$.

Table 6-6. Optimal parameters for different anatomy groups obtained in the edge-preserving range.

Anatomy group	Optimal β	Optimal $1/\delta^2$
HRLN	0.5	0.5-1*
LRHN	0.3	0.3-0.5*
LRLN	0.5	0.5-1*

* In the ranges of optimal δ , the differences in the VUS values were either not statistically significant or less than 0.01.

6.3.1.3 The optimal hyperparameter in the two search ranges

The 3-class task performance for each anatomy group and the whole population obtained using the optimal hyperparameters in the two search ranges for DTMAP is given in Table 6-7 .

Table 6-7. Three-class task performance obtained using the optimal hyperparameters in the two search ranges for DTMAP

	DTMAP1 ^a	DTMAP2 ^b
HRLN	0.809±0.008	0.823±0.008
LRHN	0.45±0.01	0.52±0.01
LRLN	0.59±0.01	0.60±0.01
Whole population	0.68±0.01	0.71±0.01

^a Use optimal hyperparameters in the noise-smoothing-edge-preserving range.

^b Use optimal hyperparameters in edge-preserving range.

^c The values after \pm are the standard deviation of the corresponding VUS values

From the results in Table 6-7, it is evident that the optimal hyperparameters obtained in the edge-preserving range provided higher values of the VUS, and they are thus deemed as the optimal hyperparameters for the overall optimization problem.

6.3.1.4 Properties of optimal hyperparameters

Although the optimal parameters obtained in the two search ranges may not be the globally optimal parameters, the results illustrate some properties of hyperparameters and provide insight into methods for improving the selection of hyperparameters in the future. We found several interesting properties of the optimal parameters from the results in 6.3.1.1 and 6.3.1.2.

First, the optimal ratio, η/δ , based on the result for the edge-preserving range in 6.3.1.2,

was less than ratio of the average myocardial count level between the stress and rest images (which was approximately 2 in this study). Based on Equations (6-6) and (6-7), this means $\frac{k_s}{k_r} < 1$. This can be explained as follows. In rest images, there was a lower count level (resulting in higher noise) in the myocardium. As a result, the optimal value of k_r was greater than the optimal value of k_s . Based on (6-6) and (6-7) and the fact that greater values of η and δ result in more smoothing images, $k_r > k_s$ means greater smoothing of the rest images.

The ratios of k_s to k_r in the noise-smoothing-edge-preserving range are given in Table 6-8. These data show that: for the HRLN group, k_r was greater than k_s on average; for the LRLN group, k_r was smaller than k_s on average with some values of k_r greater than the corresponding values of k_s ; for the LRHN group, k_r was always smaller than k_s . Since there is higher noise in the rest images, it is desirable if we have $k_r > k_s$ so that there is greater smoothing in the rest images. Thus, we expect that the rest images did not have the optimal amount of smoothing for some objects in the LRLN group and for all objects in the LRHN group using the parameters obtained in the noise-smoothing-edge-preserving range. Based on the discussion in this and the previous paragraph, we expect the optimal parameters obtained in the edge-preserving range are more nearly optimal. The results in 6.3.1.3 that compare the task performance achieved by the optimal hyperparameters in the two search ranges confirmed this conjecture.

Table 6-8. The mean, minimum, and maximum values of k_s/k_r for different anatomy groups obtained using the matched resolution method.

Anatomy group	Mean	Minimum	Maximum
HRLN	0.65	0.41	1.44
LRHN	5.05	2.21	12.33
LRLN	1.39	0.84	2.70

Second, the ratio of η/δ depended not only on the count level of the two images, but also

depended on the choice of β , as shown in Figure 6-9. We observed that when β is larger, the optimal value of η/δ tends to be smaller.

Third, the optimal β was also dependent on η/δ , as shown in Figure 6-10.

To summarize the properties of the hyperparameters, we found that the optimal choices of parameter k_s and k_r were also count level dependent. Thus, the optimal choices of k_s and k_r depended on both the defect contrast and the image count level. In addition, the optimal value of each of the three hyperparameters was dependent on the values of the other two. This strongly suggests that the hyperparameters cannot be optimized independently.

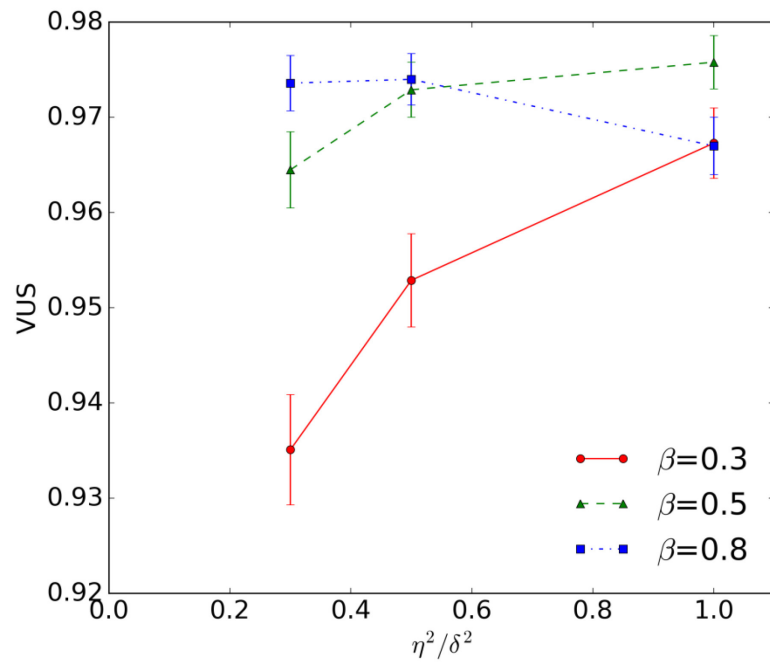


Figure 6-9. Plot of VUS versus δ for different values of β for the HRLN group and the anterior defect with 5% extent and 50% severity.

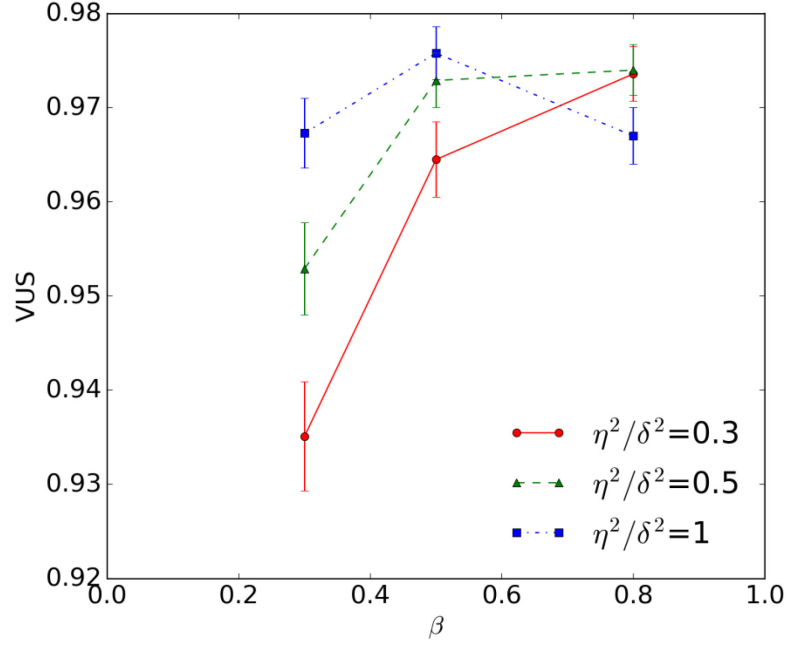


Figure 6-10. Plot of VUS versus β for different values of δ for the HRLN group defect and the anterior location with 5% extent and 50% severity.

6.3.2 Image quality evaluation based on task performance

The 3-class task performance is summarized in Table 6-9 for DTMAP using the optimal parameters obtained by our proposed method and optimized post-filtered OS-EM. The results show that optimized OS-EM had better performance in terms of VUS compared to the optimized DTMAP.

Table 6-9. Task performance of optimized post-filtered OS-EM and DTMAP.

	OS-EM	DTMAP
HRLN	0.844±0.005	0.823±0.008
LRHN	0.593±0.005	0.52±0.01
LRLN	0.640±0.006	0.60±0.01
Whole anatomy	0.74±0.01	0.71±0.01

The values after \pm are the standard deviation of the corresponding VUS values

6.3.3 Image quality evaluation based on physical properties

To shed light on the reason for the poorer performance of DTMAP, we compared the edge sharpness and noise in the myocardium in images obtained using the optimized OS-EM and

DTMAP methods for several phantoms. The parameters of these phantoms are shown in Table 6-10.

Table 6-10. Parameters of phantoms used to study the image quality through physical properties.

Phantom notation	Gender	Body core size	Heart size	Subcutaneous adipose tissue level	Anatomy group
M.BS.HS.FS	Male	Small	Small	Small	LRLN
M.BL.HL.FL	Male	Large	Large	Large	HRLN
F.BL.HS.FM	Female	Large	Small	Medium	LRHN

Since the quantification of edge sharpness is difficult due to the irregular shape and position dependence of the PSF, we present only a subjective evaluation of image sharpness. The optimized DTMAP and OS-EM reconstructed images for the three phantoms are shown in Figure 6-11 to Figure 6-13. From these images, we observed that the optimized DTMAP images have sharper edges compared to optimized OS-EM.

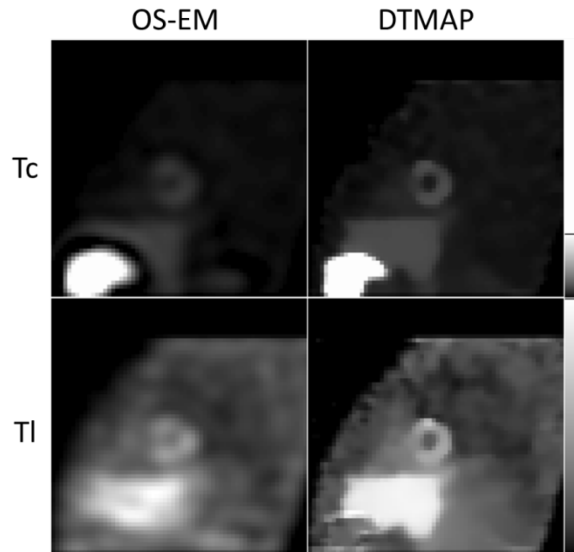


Figure 6-11. Sample Tc (top) and Tl (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the parameters obtained in the edge-preserving range) for phantom M.BS.HS.FS with defect type 1 with 5% extent and 50% severity at the anterior location.

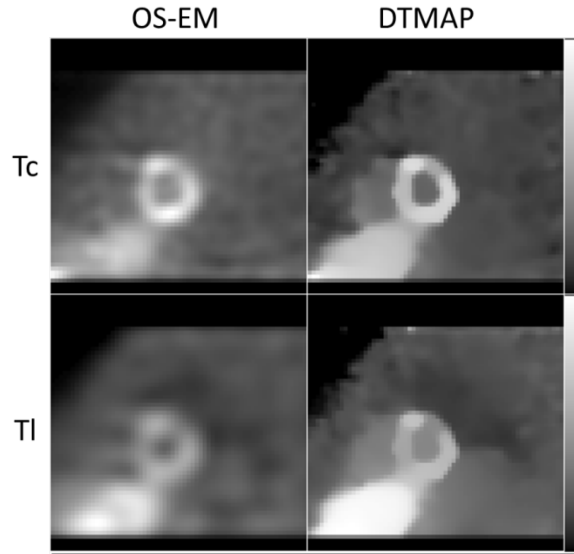


Figure 6-12. Sample Tc (top) and Tl (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the optimal parameters obtained in the edge-preserving range) for the phantom M.BL.HL.FL with defect type 1 with 5% extent and 50% severity at the anterior location.

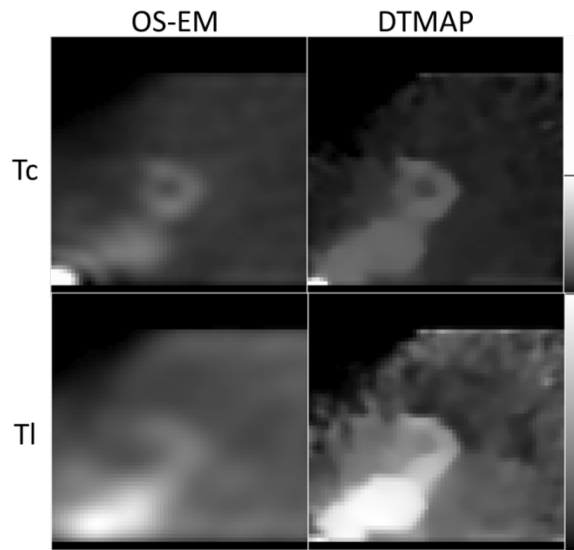


Figure 6-13. Sample Tc (top) and Tl (bottom) reconstructed images using optimal OS-EM (left), and DTMAP (right, DTMAP using the parameters obtained in the edge-preserving range) for the phantom F.BL.HS.FM with defect type 1 with 5% extent and 50% severity at the anterior location.

We also compared the noise level between the optimized OS-EM and DTMAP images. The mean, standard deviation, and COV images for phantom M.BL.HL.FM are shown in Figure 6-14 and Figure 6-15 for Tc and Tl, respectively. These show that the COV for DTMAP is very position-dependent. In particular, the COV for DTMAP was larger at the edges than in uniform

regions; for OS-EM, the COV was more uniform. This is because the resolution for DTMAP is more object-dependent, as shown in Section 6.3.4. This may result from the fact that, though edges in 1 image are very sharp, the position of the edges is different for different noise realizations, as shown in Figure 6-17. In that figure, we drew a profile of a vertical line passing through the center of the defect for the phantom in Figure 6-15 for four noise realizations; note that the position of the edge is different for DTMAP between different noise realizations but not for OS-EM. The edges of the mean images were sharper for DTMAP than OS-EM. The ratios of the average COV in the myocardium for optimized DTMAP divided by that of OS-EM are summarized in Table 6-11. The results in this table indicated that DTMAP has a larger COV than OS-EM in the myocardium.

The ratios of the average fluctuations in the myocardium for optimized DTMAP divided by that of OS-EM are summarized in Table 6-12. The results indicate that the fluctuations inside the myocardium are comparable for the two algorithms.

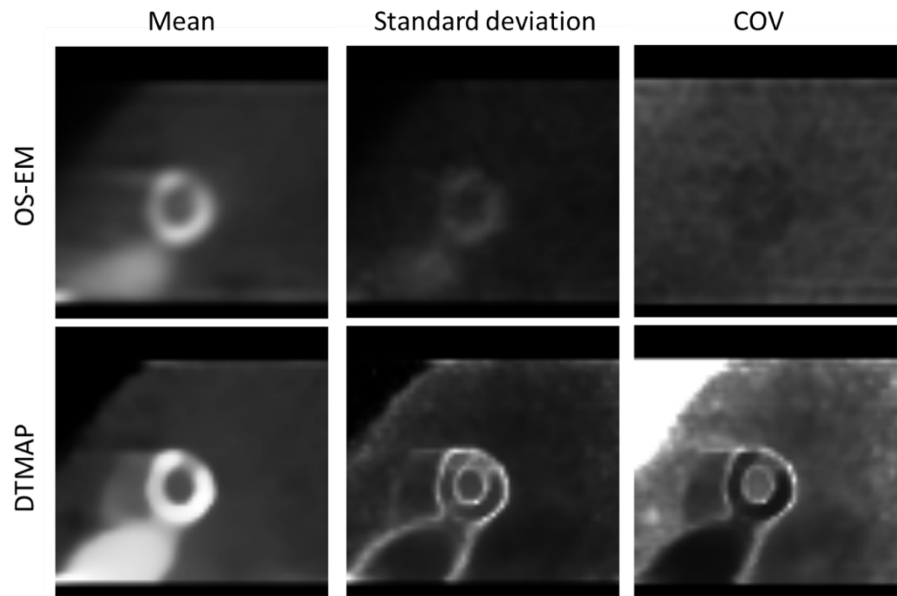


Figure 6-14. Mean, standard deviation, and COV images (generated using 100 noise realizations) of Tc images for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with a 5% extent and 50% severity at the anterior location. Note that the images in each column have the same display range [min, max], with min set to the smallest voxel value of the two images and max set to the maximum voxel value of the two images in a region that is defined by the myocardium dilated by 5 voxels.

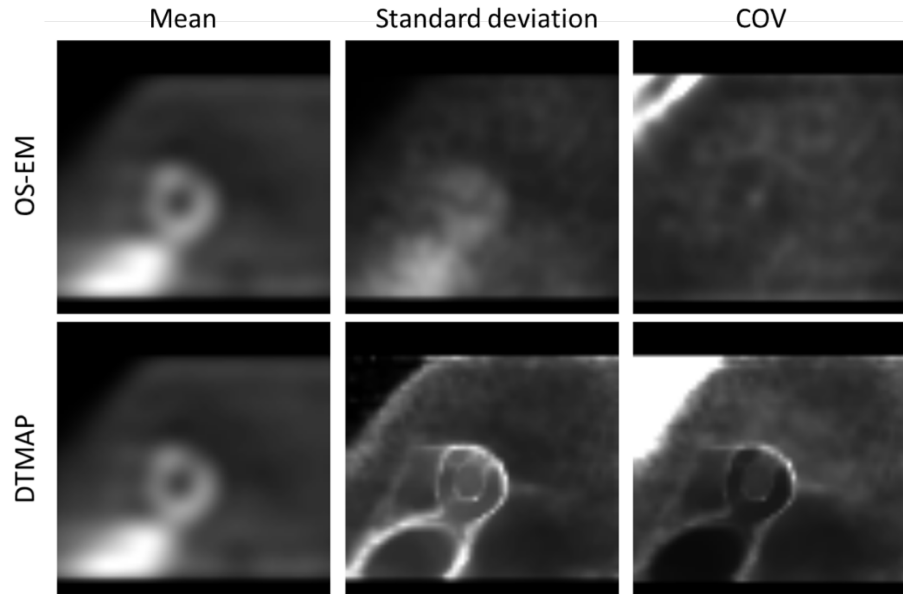


Figure 6-15. Mean, Standard deviation, COV images (generated using 100 noise realizations) of Tl images for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with 5% extent and 50% severity at anterior location. Note that the images in each column have the same display range [min, max], with min set to the smallest voxel value of the two images and max set to the maximum voxel value of the two images in a region that is defined by the myocardium dilated by 5 voxels.

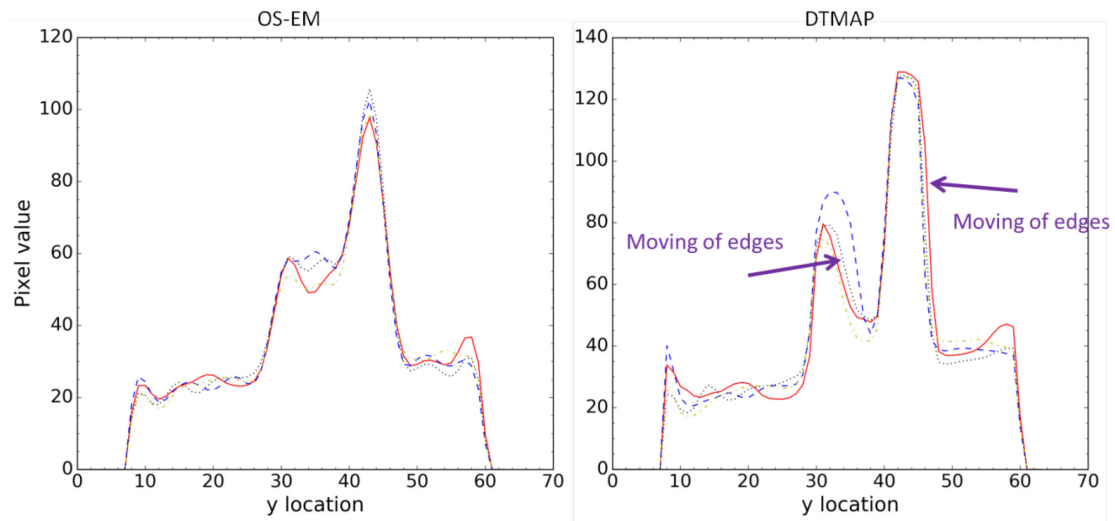


Figure 6-16. Profile of a vertical line passing through the center of the defect for optimized OS-EM and DTMAP for the phantom M.BL.HL.FL with defect type 1 with a 5% extent and 50% severity at the anterior location. The four different line styles represent four different noise realizations.

Table 6-11. The ratio of the average COV in the myocardium for DTMAP divided by that of OS-EM.

Phantom	Tc	Tl
M.BS.HS.FS	1.42	1.35
M.BL.HL.FL	1.19	0.87
F.BL.HS.FM	1.77	2.30

Table 6-12. The ratio of the average fluctuations in the myocardium for DTMAP divided by that of OS-EM.

Phantom	Tc	Tl
M.BS.HS.FS	1.04	1.03
M.BL.HL.FL	0.98	1.00
F.BL.HS.FM	1.02	1.06

6.3.4 Resolution Properties

We studied the resolution properties of un-filtered OS-EM, post-filtered OS-EM, and DTMAP. Sample images of the local PSFs for these algorithms are shown in Figure 6-16.

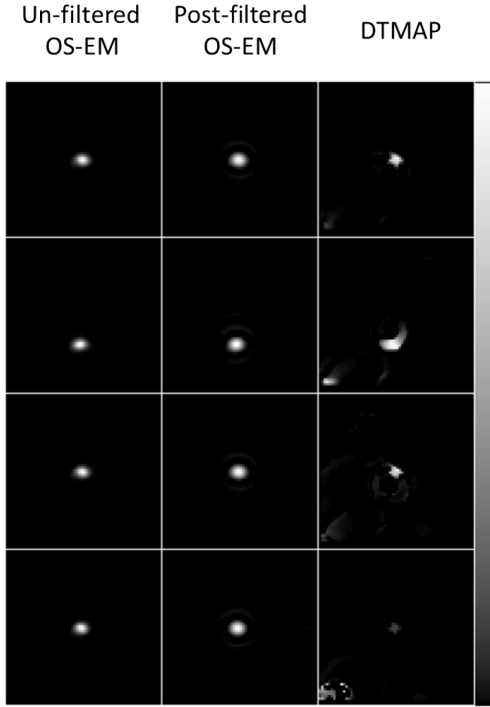


Figure 6-17. The local PSFs at the center of the defect obtained with (from left to right): unfiltered OS-EM (2 iterations with 4 subsets/iteration), post-filtered OS-EM (2 iterations and cutoff frequency of 0.2 pixels^{-1}), and DTMAP (DTMAP with hyperparameters obtained in the edge-preserving range) for 1) the male phantom with the large core body size, heart size and subcutaneous adipose tissue thickness with defect at an anterior position, 5% extent and 50% severity and a single uptake realization (first row); 2) same as 1) except the defect was at an inferior position (second row); 3) same as 1) except at for a different uptake realization (third row); 4) same as 1) except that the phantom had a medium body core size, small heart size and subcutaneous adipose tissue thickness (same anatomy group as 1), fourth row).

For un-filtered and post-filtered OS-EM, the PSF is anisotropic due to the depth-dependent resolution. For the DTMAP algorithm, the shape of the PSF is very irregular: it is very asymmetric and not compact (there are nonzero values far from the center of the PSF). For un-filtered and post-filtered OS-EM, the PSFs do not vary significantly as functions of location, uptake level, and phantom. For the DTMAP algorithm, the PSFs varied significantly for these object-dependent factors. This has been observed before for the QMAP algorithm [111] and is a general property of MAP algorithms with spatially-invariant priors. This results from the fact that the likelihood term of the objective function is object-dependent and the prior is not: when the count level is higher, which results in higher absolute variance, the data certainty is smaller, the

weight of the prior would become higher, resulting in and more smoothing.

To quantify the variations in the PSFs, we calculated the 2D and 3D FWHMs of the PSFs in Figure 6-16. The 2D and 3D FWHMs are shown in Table 6-13 and Table 6-14, respectively. The FWHM values confirm the visual observations about the degree of object-dependency for DTMAP and OS-EM algorithms.

Table 6-13. 2D FWHM of PSFs for 4 reconstruction algorithms and different objects.

Object*	Un-filtered OS-EM	Post-filtered OS-EM	DTMAP
1	3.42	4.25	3.77
2	3.40	4.18	5.62
3	3.37	4.21	3.54
4	3.46	4.25	3.28

*Object 1: Male phantom with large body core size, heart size and subcutaneous adipose tissue thickness with defect at an anterior position with a 5% extent and 50% severity and a specific uptake realization.

Object 2: same as 1 except the defect was at an inferior position.

Object 3: same as 1 except for a different uptake realization

Object 4: same as 1 except the phantom had a medium body core size, small heart size, and subcutaneous adipose tissue thickness.

Table 6-14. 3D FWHM of PSFs for 4 reconstruction algorithms and different objects.

Object*	Un-filtered OS-EM	Post-filtered OS-EM	DTMAP
1	3.82	4.78	3.68
2	3.70	4.72	3.29
3	3.79	4.72	2.78
4	3.89	4.66	3.57

*Object 1: Male phantom with large size of body core size, heart size and subcutaneous adipose tissue thickness with defect at an anterior position with a 5% extent and 50% severity and a specific uptake realization

Object 2: same as 1 except the defect was at an inferior position.

Object 3: same as 1 except at a different uptake realization

Object 4: same as 1 except the phantom had a medium body core size, small heart size, and subcutaneous adipose tissue thickness.

The object-dependent resolution of DTMAP could result in object-dependent bias, which

might affect the task performance. An example of how location-dependent resolution might affect the task performance is shown below. In Figure 6-17, sample images of the phantom, optimized OS-EM and optimized DTMAP images are shown for one phantom and uptake realization. Note that for OS-EM the inferior region looks brighter than the anterior region because of activity spill-in from the liver due to limited spatial resolution. However, for the two DTMAP images, the inferior region looks dimmer than the anterior region. This is shown more clearly from the profile plot of a vertical line passing the center of the anterior defect location (drawn in the phantom image in Figure 6-17), as shown in Figure 6-18. How the object-dependency of resolution affect task performance still need further investigation.

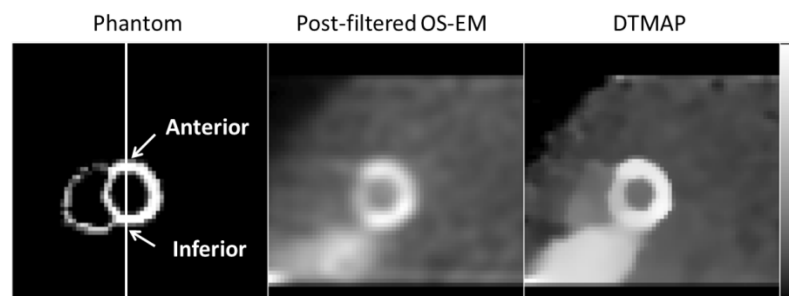


Figure 6-18. Sample images of the phantom (just the myocardium is shown) and from optimal post-filtered OS-EM and DTMAP with parameters obtained in the edge-preserving range (from left to right). The phantom did not have a perfusion defect.

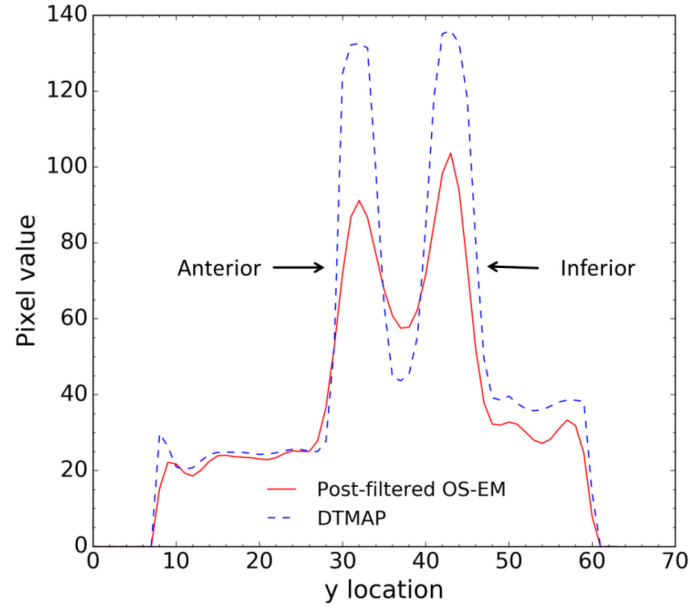


Figure 6-19. Profile from a vertical line passing through the center of the images reconstructed with optimal post-filtered OS-EM, DTMAP with parameters obtained in the edge-preserving range . The phantom did not have a perfusion defect.

6.4 Discussion

Because of the large number of hyperparameters and cost of optimizing all of them, we proposed ways to narrow down the search range. We searched for the optimal patient-specific hyperparameters for DTMAP in two hyperparameter ranges: in one range the prior was approximately quadratic inside the myocardium and edge-preserving at the expected boundaries of defects; in the other range the prior was edge-preserving, even inside myocardium. For the data in this study, the optimal hyperparameters for the two ranges were found in the second (edge-preserving) range. This might be because the differences between the amplitude of the noise and defect edges (as shown in Table 6-1 and Table 6-2) were small in this study. Thus, it is difficult for the priors to be simultaneously quadratic in the myocardium and edge-preserving enough at the defect edges. For cases where the noise level was lower or the defect contrast was higher the optimal values could be found in the first range.

The optimized DTMAP images resulted in inferior task performance compared to the

optimized post-filtered OS-EM images. This is despite the fact that, at least qualitatively and anecdotally, the DTMAP images appear better. There are several plausible reasons for this. The first reason might be that the hyperparameter search method used here failed to find the global hyperparameters. For example, it might be the case that the optimal hyperparameters for DTMAP are far from those obtained by quadratic resolution matching with the optimized OS-EM in the noise-smoothing-edge preserving range. A second reason is that the globally optimal hyperparameters for DTMAP were not within the selected search ranges. The third reason is that DTMAP is inferior to OS-EM for the task investigated here. Finally, the inferiority might be due to the observer strategies used in this study, and DTMAP might be superior for human observers. For example, it is possible that the observer strategies in this study do not predict human observer performance, or does not predict human performance for algorithms like DTMAP that have object-dependent PSFs. Further study is needed to develop more general optimization methods for DTMAP, compare DTMAP to OS-EM, and study the observer strategies.

This study also revealed several properties of DTMAP that must be considered in future studies and applications. First, DTMAP has a very irregularly-shaped PSF and object-dependent resolution and bias. As a result, the shape of the object as well as quantitative information, such as contrast of defects in the image, is affected in ways that are somewhat hard to appreciate. It is unclear how these properties affect either human or model observer performance. More importantly, these resolution properties are difficult to control.

6.5 Conclusions

In this chapter, we proposed a practical method to obtain patient-specific hyperparameters for DTMAP by searching the hyperparameters in two ranges that have different properties of the prior function. One range is based on the theory that, with appropriate selection of hyperparameters, DTMAP behaves like MAP with a quadratic prior in uniform regions and is more edge-preserving in regions, such as at the edges of defects, where the activity concentration

changes rapidly. The other range is based on the theory that, with a different choice of parameters, the prior is more edge-preserving in both the uniform regions and at defect edges.

The optimized DTMAP had, as expected, sharper edges, but poorer noise properties, as measured by the coefficient of variation over multiple noise realizations averaged over the myocardium. This seemed to result from changes in the position of the edges for different noise realizations. In addition, DTMAP resulted in poorer 3-class task performance compared to the optimized OS-EM method. However, it should be noted that the proposed optimization methods were not guaranteed to find the global optimum values of the hyperparameters, and so the possibility remains that DTMAP could provide improved performance with appropriate combinations of hyperparameters. We studied the image quality as a function of the hyperparameter in order to shed light on properties that might be useful in developing future methods for selecting optimal hyperparameters. We also demonstrated some limitations of the DTMAP algorithm: it has unusual resolution properties with a very irregular-shaped and object-dependent point spread functions. As a result, the appearance of the objects and quantitative information about the myocardium and defects can be distorted, potentially resulting in degraded detection performance. These properties complicate the optimization and interpretation of images reconstructed with DTMAP.

7. Conclusions

This chapter reviews the results of this work, summarizes its technical contributions, and discusses some future work directions.

7.1 Summary of results

Myocardial perfusion SPECT (MPS) is currently the most cost-effective and well-developed non-invasive imaging technique to diagnose Coronary Artery Disease (CAD) [3]. The quality of MPS images is determined by the system hardware, the imaging protocol, and the image reconstruction and processing methods. In this dissertation, we aimed at rigorous optimization of the reconstruction methods for dual-isotope MPS by exploring patient-specific optimization of advanced reconstruction methods. These studies were performed using simulations of a realistic population of patients having variations in anatomy, organ uptake, heart size, and defect position, size, and contrast. The simulation was based on a previously optimized imaging protocol.

To optimize the reconstruction methods, it is desirable to employ an appropriate model observer to evaluate the image quality in terms of performance on a clinically relevant task. We first proposed and investigated a novel multi-template strategy based on the linear discriminate and sub-ensemble- analysis of the data. This strategy was required because the clinically realistic simulated data resulted in feature vectors that were not multivariate-normally (MVN) distributed, partly due to the variations in the signal (i.e., the defect size, position, and contrast). The method, underlying justification and theory, and results and numerical validation are summarized in Section 7.1.1, below. We then proposed a practical patient-specific parameter optimization method for the widely-used OS-EM reconstruction algorithm and validated the hypothesis that patient-specific reconstruction parameters result in better task performance. The methods and results of that study are summarized in Section 7.1.2. In addition, we proposed two practical

patient-specific parameter optimization methods for a dual-tracer MAP (DTMAP) algorithm. We compared the performance of optimized DTMAP to that of optimized OS-EM and studied some limitations of the DTMAP algorithm. The methods and results of that study are summarized in Section 7.1.3.

7.1.1 A sub-ensemble-based approach to evaluate detection task

performance for non-MVN distributed data using multi-template

linear observer strategies

Applying model observers to clinically-realistic signal-known-statistically (SKS) tasks with realistic background variability is generally mathematically difficult and computationally expensive. Applying the widely-used Hotelling Observer (HO), which is the optimal linear observer when the distributions for the two classes are homoscedastic and MVN, is mathematically simple. However, it is not optimal when the data are MVN distributed. In Chapter 4, we proposed a novel multi-template strategy that uses the linear discriminant to handle non-MVN data. The method is based on a sub-ensemble approach where the entire data ensemble is divided into multiple sub-ensembles that are exactly or approximately MVN and homoscedastic (Section 4.2.1). The Linear Discriminant (LD) is then applied to the data from the different sub-ensembles to estimate test statistics for each sub-ensemble, and the test statistics from all the sub-ensembles are pooled to calculate a single area under the receiver operating characteristics (ROC) curve (AUC) (Section 4.2.3). The test statistics for the LD and HO differ by a term that is constant for each sub-ensemble. We proved that this constant shifted the distribution of test statistics for each sub-ensemble so that the AUC was maximized as compared to shifting by any other set of constants (Section 4.2.3.3). For comparison, we also investigated a strategy, adapted from previous work, that applies the HO to and estimates an AUC for each sub-ensemble and averages these AUCs to produce an overall AUC (Section 4.2.2). We compared these strategies to the use of a single HO template for the entire data ensemble (Section 4.3 and 4.4). The strategies

were evaluated by applying them to reconstructed images obtained from realistic simulated myocardial perfusion SPECT studies of a population of phantoms, which produced feature vectors that were non-MVN distributed. This evaluation was in the context of optimizing reconstruction parameters. Of the strategies investigated, the multi-template LD strategy yielded the highest AUC for any given set of reconstruction parameters. The optimal reconstruction parameters obtained by the two multi-template strategies were comparable, and produced higher AUCs for each sub-ensemble than the single template HO strategy. The novel multi-template LD strategy has a stronger theoretical justification compared to the adapted multi-template HO strategy. Based on the presented theory and results, there is strong evidence in favor of using the proposed multi-template LD strategy to classify the non-MVN data from clinically realistic SKS tasks such as that used in this study. This strategy was adopted for the studies performed in the remainder of the dissertation.

7.1.2 Toward patient-specific optimization of post-filtered OS-EM reconstruction for simultaneous-acquisition dual-isotope myocardial perfusion SPECT

Post-filtered OS-EM is a regularized statistical reconstruction method that is widely used in nuclear medicine. It has the advantage of allowing modeling of and compensating for multiple image degrading factors, including attenuation, scatter, collimator and detector response, and cross-talk in the case of simultaneously acquired dual isotope data. It is relatively fast and provides images with good quality if the reconstruction parameters are chosen appropriately. Four adjustable parameters control the image quality in this study: the number of updates and the cutoff frequencies of the post-reconstruction low-pass filter applied to the rest and stress images. It is desirable to find the values of parameters that maximize the image quality in terms of task performance on a 3-class classification task. Previous authors have used an anthropomorphic channelized Hotelling Observer (CHO) to optimize the parameters in similar studies [72, 91, 97].

However, their work was aimed at obtaining the single set of parameters that produced the highest VUS for the whole phantom population. There have been previous studies showing that the optimal reconstruction parameters depend on characteristics of the imaging system and object imaged, such as the system modulation transfer function or the count level [99, 100]. Thus, we hypothesized that patient-specific reconstruction parameters would result in better task performance. In Chapter 5, we proposed the use of patient-specific reconstruction parameters in the context of simultaneous-acquisition dual-isotope MPS. In Section 5.2.3, we proposed a practical group-wise optimization method to obtain more patient-specific reconstruction parameters. In this method, the clinically representative XCAT phantom population was divided into three anatomy groups according to parameters that could be estimated from the measured data without knowledge of the defect status and were hypothesized to affect the optimal parameters: the estimated image resolution at the center of the heart, the heart size and image noise as estimated from the number of photons detected from the myocardial wall. Optimal reconstruction parameters were determined for each anatomy group. We also studied optimization of the parameters for different defect types (i.e., defect position, size, and contrast). We applied the proposed multi-template Linear Discriminant (LD) with pooled test statistics strategy (Chapter 4) to the channelized data to obtain the optimal reconstruction parameters for the rest and stress images in terms of performance on a 2-class defect detection task (Section 5.2.2). The results in Section 5.3 showed that the patient-specific parameters resulted in better task performance. Parameters optimized by patient anatomy group or by defect type, individually or combined, resulted in better overall 2- and 3-class task performance. Since the defect type (position, extent and severity) is difficult to estimate from the data, optimization by patient anatomy groups is more practical and provided modestly better two- and three- class performance than the use of a single set of parameters for all the phantoms. However, these results indicate that optimization based on defect type could have an important benefit in terms of improving task

performance, and methods for achieving this are worth pursuing.

7.1.3 Toward patient-specific optimization of a maximum *a posteriori* reconstruction algorithm with a dual-tracer prior for simultaneous-acquisition dual-isotope myocardial perfusion SPECT

MAP algorithms have been widely used to regularize the ill-posed image emission computed tomography reconstruction problem. Various priors, each with its specific properties, have been proposed. He et al. have previously developed a MAP algorithm with dual-tracer prior (DTMAP) for simultaneous-acquisition dual-isotope myocardial perfusion SPECT [1]. The dual-tracer prior in [1] can couple the information from the images of the two isotopes, based on the fact that the two images are perfectly registered in space, to improve the quality of both images. The DTMAP algorithm was reviewed in section 2.4. Although this algorithm has the advantages of preserving edges and reducing noise in uniform regions, three hyperparameters are used that affect these properties; finding optimal values for these parameters is technically and practically challenging. Previously, partial optimization of the hyperparameters for this DTMAP algorithm provided better image quality compared to optimized OS-EM in terms of three-class classification task performance for a simulated simultaneous-acquisition dual-isotope MPS study based on an NCAT [106] phantom population [1]. However, this previous work had some limitations. First, the NCAT phantom is less realistic for modeling human anatomies than the state-of-the-art XCAT phantom. Second, the defect size and contrast used in that study were less challenging than those used in this work. As a result, it is unclear that the method would improve performance for the difficult defects most likely to result in improved overall performance. Finally, the study did not have crosstalk present in the projection data, which is always present in simultaneously acquired dual isotope imaging. As a result of these limitations, the previous study did not provide sufficient evidence that DTMAP is superior to OS-EM to justify clinical evaluation. In Chapter 6, we investigated a more rigorous and patient-specific method for optimizing the hyperparameters

of the DTMAP algorithm. These methods were evaluated in data that included crosstalk (and crosstalk compensation), based on a more realistic simulation of an XCAT phantom population, and used clinically-relevant and challenging defect sizes and contrasts.

We searched for the optimal patient-specific hyperparameters for DTMAP in two ranges, based on the behavior of the prior as a function its hyperparameters. In one hyperparameter range, DTMAP behaves like MAP with a quadratic prior in uniform regions and is more edge-preserving at object boundaries (Section 6.2.2). In the second range, the prior is edge-preserving in both the uniform regions and at edges at the myocardium region (Section 6.2.3). Patient-specific optimization was achieved using the same group-wise optimization approach described in Chapter 5.

The optimal DTMAP parameters obtained using the proposed optimization methods resulted in images with sharper edges, poorer noise properties, and poorer task performance compared to the images obtained using the optimized OS-EM method. We studied the image quality as a function of the hyperparameter in order to shed light on properties that might be useful in developing future methods for selecting optimal hyperparameters. We also demonstrated some limitations of the DTMAP method: DTMAP results in a very irregular-shaped and object-dependent point spread function, which is also difficult to control by adjusting the hyperparameters. As a result, the appearance of the objects and quantitative information about the myocardium and defects can be distorted, potentially resulting in degraded detection performance. As a result of these features of the PSF, extra caution should be taken when using DTMAP. It should also be noted that these unusual point spread functions might result in images where conventional anthropomorphic channels do not model human observer performance, and it is possible that human observers would perform better with these images than the mathematical observers used in this work.

7.2 Summary of technical and methodological contributions

The methodologies and results in this dissertation could have a significant clinical impact for simultaneous-acquisition dual-isotope MPS imaging. In addition, the methodologies presented in this dissertation may also contribute to other relevant fields, as discussed in the following.

First, the multi-template LD with pooled test statistic strategy proposed in Chapter 4 is a general strategy for handling non-MVN data in classification tasks. The input data is not restricted to channelized data from SPECT images. The method could be applied to any non-MVN distributed feature vectors or image data, whether from another modality, disease, or even in an area other than medical imaging.

Second, we verified the hypothesis that patient-specific optimization of reconstruction parameters can provide better 2- and 3-class classification task performance. This idea of patient-specific optimization can be applied to other medical imaging modalities and other applications of SPECT imaging. The approach to grouping patients based on factors that are expected to affect image quality prior to optimizing image processing parameters could also be adapted to apply to other modalities. The specific metrics proposed could easily be adapted to other applications of SPECT or PET imaging.

Finally, the DTMAP is a general method that can be applied to all simultaneous-acquisition dual-isotope imaging applications. In addition, the edge-preserving prior used in DTMAP also has applications to other modalities and diseases. The studies of the hyperparameter properties and limitations of DTMAP may benefit the future exploration and application of this algorithm.

7.3 Future work

This dissertation suggests some possible avenues for future work, such as evaluation of the proposed model observer strategy by comparing to human observers, studying the advantages and disadvantages of the DTMAP algorithm, and improved methods for selecting optimal parameters for the DTMAP algorithm.

7.3.1 Multi-template LD with pooled test statistic strategy evaluation

In order to apply the proposed model observer strategy to rank different systems for SKS tasks in place of humans, human observer studies for relevant tasks must be conducted. In particular, it would be interesting to study how well the proposed model observer strategies model human observers for algorithms like DTMAP algorithm that have irregular-shaped and object-dependent point spread functions. When designing the human observer studies to validate the proposed model observer strategies, two factors that have not been considered in the past should be kept in mind.

The first factor is that the design of the human observer studies should ensure that human observers make rational choices, as proposed in [112]. Non-rational choice may result in non-convex ROC curves, which indicate poorly designed human observer studies. Please refer to [112] for all these conditions that promote rationality.

The second factor that must be considered is that the correlation between the different observers in a one-dimensional parameter space does not necessarily indicate good correlation in a multi-dimensional space, as shown in Section 4.5. So, it is important to compare the model and human observers in a multi-dimensional space, especially since the application here involved using the model observers to optimize more than one parameter at the same time.

7.3.2 Research on the advantages and disadvantages of the DTMAP algorithm

The results in Chapter 6 and in previous paper [1] are different in terms of the superiority of the DTMAP vs. OS-EM algorithms. We suspect that the observed difference has to do with the interactions of the properties of DTMAP and the task and object. This is because how well the proposed DTMAP optimization method performs depends on the task and the object. For example, finding the optimum in the noise-smoothing-edge-preserving range is only possible

when the differences between the amplitude of the noise and defect edges are large enough. Then, the advantage of DTMAP that it produces sharp edges and reduces noise in uniform regions would be more prominent. In addition, we have demonstrated the unusual resolution properties of DTMAP in Chapter 6. How these properties affect DTMAP is likely also be task-dependent. Thus, in the future, it would be interesting to compare DTMAP and OS-EM for different kinds of tasks and determine the kinds of tasks for which DTMAP is preferable.

7.3.3 Hyperparameter optimization for DTMAP

To better study the DTMAP algorithm in the future, it is necessary to develop a complete hyperparameter optimization method for DTMAP. Even though one can always use a grid search method to search for the optimal the parameters, it can be too computationally expensive to implement in most studies since the DTMAP reconstruction is quite computationally expensive. It is necessary to propose a practical optimization method. The ultimate approach, which can be very challenging, is to express the task performance as a function of the hyperparameters and then directly optimize the hyperparameters to achieve the best task performance. Previous work by Yang et al. [113] may provide some guidance on how to do this. In that previous work, they expressed the task performance in terms of SNR using the hyperparameters for a MAP algorithm with a quadratic prior. In that work, they assumed the defect was small so that the MAP algorithm was approximately linear. They then developed an analytic expression for the SNR based on previous analytic expressions of the resolution and noise properties for the MAP algorithm with quadratic prior [109, 114]. To develop similar expressions as a function of the hyperparameters for DTMAP, it is necessary to first study whether or under what conditions we can treat DTMAP as a linear algorithm. If the algorithm is approximately linear, we can then try to develop analytic expressions of the resolution and noise properties for DTMAP using the similar method as in previous work [109, 114]. Based on this previous works, the key step is to develop an analytic expression for the Hessian matrix of the prior term. Further study is needed in this since this

Hessian matrix is object-dependent for the prior in DTMAP and approximations may be needed to express it in an analytic form.

7.4 Conclusions

In this work, we further optimized two statistical image reconstruction methods for dual-radionuclide myocardial perfusion SPECT (MPS): post-filtered ordered subsets-expectation maximization (OS-EM) method and a maximum a *posteriori* (MAP) with dual-tracer prior (DTMAP) method. We first developed and evaluated a model observer strategy to evaluate image quality. The strategy was a multi-template Linear Discriminant strategy based on sub-ensemble analysis of the image data. We proposed to optimize the parameters for the two reconstruction methods by patient anatomy groups that was divided based on metrics that affect the optimal reconstruction parameters. We then compared the two optimized reconstruction methods. The optimal reconstruction parameters could be applied to clinical dual-radionuclide MPS studies. The patient-specific optimization approach could be applied to similar imaging fields. The model observer strategy, although still needs further evaluation to study correlation with human observers, is a general classification strategy that can be applied to many other fields. Since the optimization of DTMAP in this work is restricted, future work is suggested to propose the complete optimization method for DTMAP and compare it with OS-EM more comprehensively.

REFERENCES

- [1] X. He, *et al.*, "Regularized image reconstruction algorithms for dual-isotope myocardial perfusion SPECT (MPS) imaging using a cross-tracer prior," *IEEE Trans Med Imaging*, vol. 30, pp. 1169-83, Jun 2011.
- [2] World Health Organization. (2015, January). *Cardiovascular diseases (CVDs)*. Available: <http://www.who.int/mediacentre/factsheets/fs317/en/>
- [3] A. Cuocolo, *et al.*, "The many ways to myocardial perfusion imaging," *The quarterly Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, p. 4, 2005.
- [4] K. L. Gould and K. Lipscomb, "Effects of coronary stenoses on coronary flow reserve and resistance," *The American Journal of Cardiology*, vol. 34, pp. 48-55, 1974.
- [5] M. J. O'Hara, *et al.*, "Detection of high risk coronary artery disease by thallium imaging," *British Heart Journal*, vol. 53, pp. 616-623, 1985.
- [6] C. Marcassa, *et al.*, *Clinical value, cost-effectiveness, and safety of myocardial perfusion scintigraphy: a position statement*, 2008.
- [7] L. J. Shaw and A. E. Iskandrian, "Prognostic value of gated myocardial perfusion SPECT," *Journal of Nuclear Cardiology*, vol. 11, pp. 171-185, 2004.
- [8] S. R. Underwood, *et al.*, "Myocardial perfusion scintigraphy: the evidence," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 31, pp. 261-291, 2004/02/01 2004.
- [9] X. He, *et al.*, "Regularized image reconstruction algorithms for dual-isotope myocardial perfusion SPECT (MPS) imaging using a cross-tracer prior," *Medical Imaging, IEEE Transactions on*, vol. 30, pp. 1169-1183, 2011.
- [10] A. Cuocolo, *et al.*, "The many ways to myocardial perfusion imaging," *QJ Nucl Med Mol Imaging*, vol. 49, pp. 4-18, 2005.
- [11] E. G. DePuey, *et al.*, *Cardiac SPECT imaging*: Lippincott Williams & Wilkins, 2001.
- [12] J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*: Pearson Education, 2014.
- [13] I. A. E. AGENCY, *Nuclear Medicine Physics*. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2015.
- [14] G. A. Beller and S. R. Bergmann, "Myocardial perfusion imaging agents: SPECT and PET," *Journal of Nuclear Cardiology*, vol. 11, pp. 71-86, 2004.
- [15] J. A. Case, *et al.*, "Critical evaluation of a simultaneous acquisition dual Tl-201/sestamibi imaging protocol in patients with known CAD," *Journal of Nuclear Cardiology*, vol. 6, p. S115, 1999.
- [16] C. G. Kwok, *et al.*, "Feasibility of simultaneous dual-isotope myocardial perfusion acquisition using a lower dose of sestamibi," *European journal of nuclear medicine*, vol. 24, pp. 281-285, 1997.
- [17] S. C. Moore, *et al.*, "Simultaneous Tc-99m/Tl-201 imaging using energy-based estimation of the spatial distributions of contaminant photons," *Nuclear Science, IEEE Transactions on*, vol. 42, pp. 1189-1195, 1995.
- [18] M. Ghaly, *et al.*, "Optimization and comparison of simultaneous and separate acquisition protocols for dual isotope myocardial perfusion SPECT," *Physics in Medicine and Biology*, vol. 60, p. 5083, 2015.
- [19] E. C. Frey and B. M. W. Tsui, "A new method for modeling the spatially-variant, object-dependent scatter response function in SPECT," in *Nuclear Science Symposium, 1996. Conference Record., 1996 IEEE*, 1996, pp. 1082-1086 vol.2.
- [20] X. Song, *et al.*, "Fast modelling of the collimator-detector response in Monte Carlo simulation of SPECT imaging using the angular response function," *Physics in Medicine and Biology*, vol. 50, p. 1791, 2005.

- [21] X. Song, *et al.*, "Validation and evaluation of model-based crosstalk compensation method in simultaneous 99m Tc stress and 201 Tl rest myocardial perfusion SPECT," *Nuclear Science, IEEE Transactions on*, vol. 51, pp. 72-79, 2004.
- [22] W.-T. Wang, *et al.*, "Parameterization of Pb X-ray contamination in simultaneous Tl-201 and Tc-99m dual-isotope imaging," *Nuclear Science, IEEE Transactions on*, vol. 49, pp. 680-692, 2002.
- [23] J. T. Bushberg, *The Essential Physics of Medical Imaging*: Lippincott Williams & Wilkins, 2002.
- [24] H. Zaidi, *Quantitative analysis in nuclear medicine imaging* vol. 368: Springer, 2006.
- [25] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *Medical Imaging, IEEE Transactions on*, vol. 1, pp. 113-122, 1982.
- [26] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *Medical Imaging, IEEE Transactions on*, vol. 13, pp. 601-609, 1994.
- [27] H. Barrett, "Model observers for assessment of image quality," in *Nuclear Science Symposium Conference Record, 2002 IEEE*, 2002, p. 652 vol. 2.
- [28] P. Khurd, *et al.*, "Ideal AFROC and FROC observers," *Medical Imaging, IEEE Transactions on*, vol. 29, pp. 375-386, 2010.
- [29] X. He, *et al.*, "Discovering intrinsic properties of human observers' visual search and mathematical observers' scanning," *JOSA A*, vol. 31, pp. 2495-2510, 2014.
- [30] C. E. Metz and J.-H. Shen, "Gains in Accuracy from Replicated Readings of Diagnostic Images Prediction and Assessment in Terms of ROC Analysis," *Medical Decision Making*, vol. 12, pp. 60-75, 1992.
- [31] H. H. Barrett and K. J. Myers, "Foundations of image science," *Foundations of Image Science, by Harrison H. Barrett, Kyle J. Myers, pp. 1584. ISBN 0-471-15300-1. Wiley-VCH, October 2003.*, vol. 1, 2003.
- [32] H. H. Barrett, *Foundations of image science*. Wiley-Interscience: Hoboken, NJ, 2004.
- [33] D. C. Edwards, *et al.*, "Ideal observers and optimal ROC hypersurfaces in N-class classification," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 891-895, 2004.
- [34] H. H. Barrett and K. J. Myers, *Foundations of image science*: John Wiley & Sons, 2013.
- [35] X. He, *et al.*, "Three-class ROC analysis-A decision theoretic approach under the ideal observer framework," *Medical Imaging, IEEE Transactions on*, vol. 25, pp. 571-581, 2006.
- [36] H. H. Barrett, *et al.*, "Linear discriminants and image quality," *Image and Vision Computing*, vol. 10, pp. 451-460, 1992.
- [37] R. FUKUNAGA, "Statistical pattern recognition," 1990.
- [38] X. He and E. C. Frey, "An Optimal Three-Class Linear Observer Derived From Decision Theory," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 77-83, 2007.
- [39] A. E. Burgess, *et al.*, "Visual signal detectability with two noise components: anomalous masking effects," *JOSA A*, vol. 14, pp. 2420-2442, 1997.
- [40] C. K. Abbey and H. H. Barrett, "Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *JOSA A*, vol. 18, pp. 473-488, 2001.
- [41] Y. Zhang, *et al.*, "The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds," *Medical Imaging, IEEE Transactions on*, vol. 25, pp. 1348-1362, 2006.
- [42] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Am. A*, vol. 4, pp. 2447-2457, 1987.
- [43] H. C. Gifford, *et al.*, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 41, pp. 514-521, 2000.

- [44] J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," 1992, pp. 161-168.
- [45] M. P. Eckstein, *et al.*, "Human vs model observers in anatomic backgrounds," 1998, pp. 16-26.
- [46] X. He, *et al.*, "An investigation of the trade-off between the count level and image quality in myocardial perfusion SPECT using simulated images: the effects of statistical noise and object variability on defect detectability," *Physics in Medicine and Biology*, vol. 55, p. 4949, 2010.
- [47] S. D. Wollenweber, *et al.*, "Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging," *Nuclear Science, IEEE Transactions on*, vol. 46, pp. 2098-2103, 1999.
- [48] S. Park, *et al.*, "Efficiency of the human observer detecting random signals in random backgrounds," *Journal of the Optical Society of America A*, vol. 22, pp. 3-16, 2005/01/01 2005.
- [49] S. Sankaran, *et al.*, "Optimum Compensation Method and Filter Cutoff Frequency in Myocardial SPECT: A Human Observer Study," *Journal of Nuclear Medicine*, vol. 43, pp. 432-438, March 1, 2002 2002.
- [50] B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *Journal of the Optical Society of America A*, vol. 20, pp. 1725-1738, 2003/09/01 2003.
- [51] F. E. Elshahaby, *et al.*, "The effect of signal variability on the histograms of anthropomorphic channel outputs: factors resulting in non-normally distributed data," in *SPIE Medical Imaging*, 2015, pp. 94160P-94160P-6.
- [52] M. P. Eckstein and C. K. Abbey, "Model observers for signal-known-statistically tasks (SKS)," 2001, pp. 91-102.
- [53] M. P. Eckstein, *et al.*, "Effect of image compression for model and human observers in signal-known-statistically tasks," 2002, pp. 13-24.
- [54] M. P. Eckstein, *et al.*, "Optimization of model observer performance for signal known exactly but variable tasks leads to optimized performance in signal known statistically tasks," 2003, pp. 123-134.
- [55] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative radiology*, vol. 21, pp. 720-733, 1986.
- [56] D. R. Gilland, *et al.*, "Determination of the optimum filter function for SPECT imaging," *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, vol. 29, pp. 643-650, 1988.
- [57] W. Segars, *et al.*, "4D XCAT phantom for multimodality imaging research," *Medical Physics*, vol. 37, pp. 4902-4915, 2010.
- [58] W. P. Segars, *et al.*, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, pp. 3800-3808, 2008.
- [59] J. Tabary, *et al.*, "Realistic X-Ray CT simulation of the XCAT phantom with SINDBAD," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009, pp. 3980-3983.
- [60] X. Rong, *et al.*, "Development and evaluation of an improved quantitative 90Y bremsstrahlung SPECT method," *Medical Physics*, vol. 39, pp. 2346-2358, 2012.
- [61] G. S. Fung, *et al.*, "Development of a model of the coronary arterial tree for the 4D XCAT phantom," *Physics in Medicine and Biology*, vol. 56, p. 5651, 2011.
- [62] N. L. o. Medicine. *Visible Human Male and Female datasets*. Available: http://www.nlm.nih.gov/research/visible/visible_human.html
- [63] M. Ghaly, *et al.*, "Design of a digital phantom population for myocardial perfusion SPECT imaging research," *Physics in Medicine and Biology*, vol. 59, p. 2935, 2014.

- [64] M. Ghaly, *et al.*, "Optimization of SPECT using variable acquisition duration," *Journal of Nuclear Medicine*, vol. 53, pp. 2411-2411, 2012.
- [65] M. Ghaly, *et al.*, "Design of a digital phantom population for myocardial perfusion SPECT imaging research," *Physics in Medicine and Biology*, vol. 59, p. 2935, 2014.
- [66] O. Ergonomics. *PeopleSize 2008*. Available: <http://www.openenerg.com/psz/index.html>
- [67] R. L. E. a. E. V. D. A. B. Barclay. PET thorax model database [Online]. Available: <http://www.emory.edu/CRL/abb/thoraxmodel/Contents.html>
- [68] M. Ghaly, *et al.*, "Model mismatch and the ideal observer in SPECT," 2013, pp. 86730K-86730K-9.
- [69] R. L. Harrison and T. K. Lewellen, "The SimSET Program," *Monte Carlo Calculations in Nuclear Medicine: Applications in Diagnostic Imaging*, p. 87, 2012.
- [70] H. H. Barrett, *et al.*, "Model observers for assessment of image quality," *Proceedings of the National Academy of Sciences*, vol. 90, pp. 9758-9765, November 1, 1993 1993.
- [71] A. K. Jha, *et al.*, "An ideal-observer framework to investigate signal detectability in diffuse optical imaging," *Biomedical optics express*, vol. 4, pp. 2107-2123, 2013.
- [72] M. V. Narayanan, *et al.*, "Optimization of iterative reconstructions of ^{99m}Tc cardiac SPECT studies using numerical observers," *Nuclear Science, IEEE Transactions on*, vol. 49, pp. 2355-2360, 2002.
- [73] S. Kulkarni, *et al.*, "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors," *Physics in Medicine and Biology*, vol. 52, p. 3601, 2007.
- [74] S. Sankaran, *et al.*, "Optimum compensation method and filter cutoff frequency in myocardial SPECT: a human observer study," *Journal of Nuclear Medicine*, vol. 43, pp. 432-438, 2002.
- [75] L. Yu, *et al.*, "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: Impact of radiation dose and reconstruction algorithms," *Medical Physics*, vol. 40, p. 041908, 2013.
- [76] E. C. Frey, *et al.*, "Application of task-based measures of image quality to optimization and evaluation of three-dimensional reconstruction-based compensation methods in myocardial perfusion SPECT," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 1040-1050, 2002.
- [77] M. Ghaly, *et al.*, "Optimization of energy window and evaluation of scatter compensation methods in myocardial perfusion SPECT using the ideal observer with and without model mismatch and an anthropomorphic model observer," *Journal of Medical Imaging*, vol. 2, pp. 015502-015502, 2015.
- [78] F. E. Elshahaby, *et al.*, "Factors affecting the normality of channel outputs of channelized model observers: an investigation using realistic myocardial perfusion SPECT images," *Journal of Medical Imaging*, vol. 3, pp. 015503-015503, 2016.
- [79] W. Braje, *et al.*, "Human efficiency for recognizing and detecting low-pass filtered objects," *Ophthalmic Literature*, vol. 3, p. 215, 1996.
- [80] Y. Zhang, *et al.*, "Automated optimization of JPEG 2000 encoder options based on model observer performance for detecting variable signals in X-ray coronary angiograms," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 459-474, 2004.
- [81] C. Castella, *et al.*, "Mass detection on mammograms: influence of signal shape uncertainty on human and model observers," *JOSA A*, vol. 26, pp. 425-436, 2009.
- [82] W. W. Peterson, *et al.*, "The theory of signal detectability," *Information Theory, Transactions of the IRE Professional Group on*, vol. 4, pp. 171-212, 1954.
- [83] H. L. Van Trees, *Detection, estimation, and modulation theory*: John Wiley & Sons, 2004.
- [84] J. P. Egan, "Signal detection theory and {ROC} analysis," 1975.

- [85] I. Narsky and F. C. Porter, *Statistical Analysis Techniques in Particle Physics*: John Wiley & Sons, 2013.
- [86] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, pp. 179-188, 1936.
- [87] K. Fukunaga, "Introduction to statistical pattern recognition. 1990," *Ch*, vol. 9, pp. 401-405.
- [88] P. A. Lachenbruch and M. Goldstein, "Discriminant analysis," *Biometrics*, pp. 69-85, 1979.
- [89] H. H. Barrett, *et al.*, "Stabilized estimates of Hotelling-observer detection performance in patient-structured noise," 1998, pp. 27-43.
- [90] M. A. Kupinski, *et al.*, "Bias in Hotelling observer performance computed from finite data," 2007, pp. 65150S-65150S-7.
- [91] E. C. Frey, *et al.*, "Application of task-based measures of image quality to optimization and evaluation of three-dimensional reconstruction-based compensation methods in myocardial perfusion SPECT," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 1040-1050, 2002.
- [92] K. Fukunaga, *Introduction to statistical pattern recognition*: Academic press, 2013.
- [93] C. E. Metz, *et al.*, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously - distributed data," *Statistics in medicine*, vol. 17, pp. 1033-1053, 1998.
- [94] C. Metz, *et al.*, "CLABROC, LABROC," *Chicago: University of Chicago, Department of Radiology and the Franklin McLean Memorial Research Institute*, 1988.
- [95] M. Herdin, *et al.*, "Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels," in *2005 IEEE 61st Vehicular Technology Conference*, 2005, pp. 136-140.
- [96] P. D. Lax, *Linear Algebra and Its Applications*: Wiley, 2007.
- [97] H. Xin, *et al.*, "A mathematical observer study for the evaluation and optimization of compensation methods for myocardial SPECT using a phantom population that realistically models patient variability," *Nuclear Science, IEEE Transactions on*, vol. 51, pp. 218-224, 2004.
- [98] E. W. Ng and M. Geller, "A table of integrals of the error functions," *Journal of Research of the National Bureau of Standards B*, vol. 73, pp. 1-20, 1969.
- [99] M. A. King, *et al.*, "Variation of the count - dependent Metz filter with imaging system modulation transfer function," *Medical Physics*, vol. 13, pp. 139-149, 1986.
- [100] D. R. Gilland, *et al.*, "Determination of the Optimum Filter Function for SPECT Imaging," *Journal of Nuclear Medicine*, vol. 29, pp. 643-650, May 1, 1988 1988.
- [101] H. Xin, *et al.*, "Application of Three-Class ROC Analysis to Task-Based Image Quality Assessment of Simultaneous Dual-Isotope Myocardial Perfusion SPECT (MPS)," *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1556-1567, 2008.
- [102] X. He, "The meaning and use of the volume under a three-class ROC surface (VUS)," *IEEE Trans. Med. Imaging*, vol. 27, p. 577, 2008.
- [103] M. Hamon, *et al.*, "Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease," *Journal of Cardiovascular Magnetic Resonance*, vol. 12, p. 1, 2010.
- [104] A. Schinkel, *et al.*, "Noninvasive evaluation of ischaemic heart disease: myocardial perfusion imaging or stress echocardiography?," *European heart journal*, vol. 24, pp. 789-800, 2003.
- [105] T. Sharir, *et al.*, "A new algorithm for the quantitation of myocardial perfusion SPECT. II: validation and diagnostic yield," *Journal of Nuclear Medicine*, vol. 41, pp. 720-727, 2000.

- [106] W. P. Segars, *et al.*, "Development and application of the new dynamic Nurbs-based Cardiac-Torso (NCAT) phantom," 2001.
- [107] V. Panin, *et al.*, "Total variation regulated EM algorithm [SPECT reconstruction]," *IEEE Transactions on Nuclear Science*, vol. 46, pp. 2202-2210, 1999.
- [108] J. Nuyts and J. A. Fessler, "A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 1042-1052, 2003.
- [109] J. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *Image Processing, IEEE Transactions on*, vol. 5, pp. 1346-1358, 1996.
- [110] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *Medical Imaging, IEEE Transactions on*, vol. 19, pp. 493-506, 2000.
- [111] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *Image Processing, IEEE Transactions on*, vol. 5, pp. 1346-1358, 1996.
- [112] X. He, *et al.*, "The equivalence of a human observer and an ideal observer in binary diagnostic tasks," 2013, pp. 86730E-86730E-8.
- [113] L. Yang, *et al.*, "Regularization design for breast lesion detection in penalized maximum likelihood image reconstruction," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012, pp. 626-629.
- [114] J. Qi, "Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 321-329, 2004.

CURRICULUM VITAE

Xin Li was born on March 13th, 1983 in Jilin, China. She completed her Bachelor degree in 2006 in Engineering Physics from Tsinghua University in Beijing. She then received a Master degree in 2008 in Engineering Physics from Tsinghua University with research focused on the development of a coded aperture environment radiation monitoring system, which later became a produce in Beijing NucMed, Ltd. She then came to the U.S to start a Ph.D. program in Electrical and Computer Engineering and worked as a Research Assistant in the Division of Medical Imaging Physics. She was under the supervision of Professor Benjamin M.W. Tsui from September 2008 to January 2011 and under the supervision of Professor Eric C. Frey from February 2011 to Oct 2016. Her Ph.D. dissertation was focused on the optimization of reconstruction methods for dual-radionuclide myocardial perfusion SPECT.

Peer-reviewed Journal Publications:

1. **X. Li**, A. K. Jha, M. Ghaly, F. E. Elshahaby, J. M. Links and E. C. Frey, “A sub-ensemble-based approach to evaluate detection task performance for non-multivariate-normal data using multi-template linear observer strategies”, submitted to *IEEE transactions on Medical Imaging*
2. **T. Feng**, **B. M. W. Tsui**, **X. Li**, **M. Vranesic**, **M. A. Lodge**, **N. C. M. Gulaldi** and **Z. Szabo**. “Image-derived and arterial blood sampled input functions for quantitative PET imaging of the angiotensin II subtype 1 receptor in the kidney”, *Medical Physics*, vol. 42, pp. 6736-6744, 2015.

Conference Proceedings and Abstracts:

1. F. E. Elshahaby, M. Ghaly, **X. Li**, A. K. Jha, E. C. Frey. “Estimating model observer performance with small image ensembles”, *Society of Nuclear Medicine conference*, **2015**
2. **X. Li**, X. Wang, Y. Zhou. “X-ray tube potential modulation in spectral CT”, *the Third International Conference on Image Formation in X-ray Computed Tomography*, 2014
3. **X. Li**, M. Ghaly, J. M. Links and E. C. Frey. “Patient specific optimization of reconstruction methods for dual isotope myocardial perfusion SPECT”, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2013
4. **X. Li**, M. Ghaly, L. Cheng and E. C. Frey. “Patient specific method to determine maximum a posteriori reconstruction hyperparameters for myocardial perfusion SPECT”, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2012

5. T. Feng, B. M. W Tsui, **X. Li**, M. Lodge, N. Gulaldi and Z. Szabo, "Evaluation of image-derived and arterial blood sampled input functions for PET imaging of angiotensin II subtype I receptor", *Society of Nuclear Medicine conference*, **2012**
6. **X. Li**, B.M.W Tsui, N. Gulaldi, J. Xia and Z. Szabo, "Evaluation of image-derived and arterial blood sampled input functions in a PET angiotensin II subtype I receptor study", *Society of Nuclear Medicine conference*, **2010**
7. A. J. Rittenbach, **X. Li**, J. Xu, J. W. Hugg, B. M. W. Tsui, "The design and evaluation of optimal multipinhole collimators for two ring-type small animal SPECT systems", *Workshop on Small-Animal SPECT Imaging*, **2010**
8. **X. Li**, Z. Wu, Y. Liu, T. Ma and Y. Jin, "Design of wide energy range coded aperture mask of an environmental radiation monitoring system", *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2007
9. X. Sun, S. Wang, T. Ma, R. Zhou, **X. Li**, Z. Zhang, Y. Liu, Z. Wu and Y. Jin, "A high resolution and high sensitivity small animal SPECT system based on H8500", *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2007