

Model observers to predict human performance in LROC studies of SPECT reconstruction using anatomical priors

Andre Lehovich, Howard C. Gifford, and Michael A. King

University of Massachusetts Medical School, 55 Lake Ave N, Worcester, MA, USA, 01655

ABSTRACT

We investigate the use of linear model observers to predict human performance in a localization ROC (LROC) study. The task is to locate gallium-avid tumors in simulated SPECT images of a digital phantom. Our study is intended to find the optimal strength of smoothing priors incorporating various degrees of anatomical knowledge. Although humans reading the images must perform a search task, our models ignore search by assuming the lesion location is known. We use area under the model ROC curve to predict human area under the LROC curve. We used three models, the non-prewhitening matched filter (NPWMF), the channelized nonprewhitening (CNPW), and the channelized Hotelling observer (CHO). All models have access to noise-free reconstructions, which are used to compute the signal template. The NPWMF model does a poor job of predicting human performance. The CNPW and CHO model do a somewhat better job, but still do not qualitatively capture the human results. None of the models accurately predicts the smoothing strength which maximizes human performance.

Keywords: LROC, model observer, human-observer performance, SPECT reconstruction, anatomical priors

1. INTRODUCTION

We have been developing new emission-tomography reconstruction techniques which incorporate prior knowledge about the anatomy.¹ We believe that the use of anatomical priors will improve image quality, as measured by human-observer performance on a clinically relevant task. We have had human observers read simulated ⁶⁷Ga SPECT images reconstructed using various algorithms in order to measure the improvement due to anatomical priors.² One of the things we have studied is how human performance varies as a function of the strength of the prior.³ This paper describes our attempts to use model observers to predict human performance as a function of prior strength on lesion search-detection tasks.

Reconstruction algorithms typically have one or more tunable parameters. For example, one may be able to set the level of smoothing or noise suppression. We would like to set these parameters to maximize the performance of humans reading the reconstructed images. One way of doing so is setting up a human-observer study in which reader performance on a task is measured as the tunable parameters are varied. Though effective, this process is expensive. A good model observer would be able to predict which value of the parameter maximizes human performance, and also give us a feel for how quickly performance drops off around the peak. Thus a good model observer gives us a *qualitative* feel for how performance varies with the parameter. It is less important for the model to give accurate *quantitative* predictions of observer performance.

2. MATERIALS & METHODS

2.1 Image reconstruction

We simulated ⁶⁷Ga-citrate SPECT scans using the MCACT phantom and 1-cm spherical lesions at a variety of contrasts and locations in the mediastinum as input to the SIMIND Monte Carlo simulator. Simulated data were then reconstructed using De Pierro's maximum-a-posteriori (MAP) algorithm⁴ iterated to convergence. In

Further author information: (Send correspondence to A.L.)

A.L.: E-mail: Andre.Lehovich@umassmed.edu

H.C.G.: E-mail: Howard.Gifford@umassmed.edu

M.A.K.: Michael.King@umassmed.edu

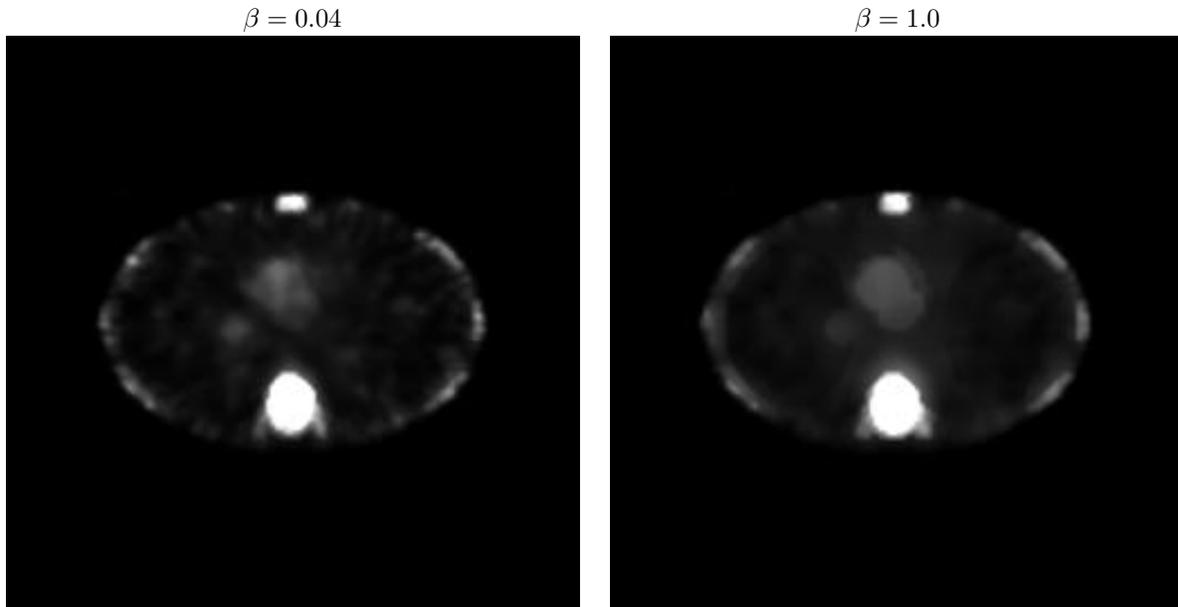


Figure 1. A transaxial slice reconstructed with two different prior strengths. The spine is at the bottom of the image; the sternum at the top. A lesion is visible to the lower-left of the heart.

in addition to the three different forms of prior knowledge about the anatomy described below, we also generated noise-free reconstructions. Transaxial slices were then extracted from the reconstructed volumes.

We considered three types of anatomical prior. The first prior used knowledge of the organ boundaries (but not lesion boundaries) to define regions of quadratic smoothing. The second prior had knowledge of both the organ boundaries, as well as lesion boundaries. The third prior had knowledge of the organ and lesion boundaries, but also had pseudo-lesion boundaries when a lesion was not actually present. Including pseudo lesions allows us to test whether or not extra boundaries increase the false-alarm rate. Full details on computing the priors are given by Bruyant et al.¹ Smoothing was performed within the anatomical regions defined by the prior, but not across region boundaries.

Prior strength is determined by a parameter β , with $\beta = 0.0$ corresponding to maximum-likelihood reconstruction with no prior, and $\beta = \infty$ ignoring the data in favor of the prior. For this study we computed reconstructions at a variety of prior strengths, $\beta \in \{0.02, 0.04, 0.06, 0.1, 0.2, 1.0\}$.

Figure 1 shows a sample data set reconstructed using the organ prior at two different β values. As the prior is strengthened the anatomical regions become smoother, however boundaries remain sharp.

2.2 Human observer study

The reconstructed noisy images were read by three members of our medical physics lab. After reading the image the observer used the mouse to identify the most-likely lesion location within the image, and indicated his confidence that a lesion was present at that location. Reading was done in a darkened room using a monitor with a perceptually-linearized greyscale.⁵

We started the study with an initial training period in which each of the observers read 100 training cases reconstructed by each prior at each of the strengths (a total of 18 training sets). After reading each training image the observer was given feedback about the truth. Reading one training set took about 20 minutes, though readers were given as much time as they desired.

During the study we had the observers warm up by re-reading 50 of the training images, again receiving feedback about the truth after each image. Then we had the observers read 100 testing images with no feedback. Rereading the training images and then reading the testing images took about 20 minutes, though observers were free to take as long as they wanted. Half of the images in each set were lesion present, the other half absent.

The observers were three scientists in our medical-physics research group. We then pooled two of the testing studies (total of 200 testing images) to compute a figure of merit for each reader/prior/beta combination. We used area under the LROC curve⁶ as the figure of merit.

2.3 Model observers

Although the task performed by the humans included a search component, we did not include search in the model. Instead we modeled reading the images as a signal-known-exactly (SKE) task. The model observer thus knew the possible lesion location and shape. We omitted search for two reasons: First, it greatly simplifies the model observer, as one does not need to specify a search region. Second, most reconstruction researchers are using SKE model observers during algorithm tuning, and we wanted to see how well such observers predict performance on the search task.

Our model observers were provided with the noisy images presented to human observers, as well as the noise-free signal-present (NFpresent) and signal-absent (NFabsent) images. These were used to compute the signal template,

$$\text{signal} = \text{NFpresent} - \text{NFabsent}. \quad (1)$$

By using the noise-free reconstructions the models know about the slight variations in reconstructed lesion shape at different locations.

2.3.1 Non-prewhitening matched filter

The first model we considered was the non-prewhitening matched filter (NPWMF). This observer computes the inner product of the noise-free signal template with the image,

$$t = \text{signal}^t \text{image}, \quad (2)$$

where t is the test statistic used to decide if a lesion is present. Note that this model does not subtract off the background. (We found that subtracting off the background gave essentially perfect performance at all values of β .)

The test statistics are then used to produce a receiver operating characteristic (ROC) curve for the model. We use area under the model's ROC curve for this SKE task as our predictor of area under the LROC curve for humans performing the search task.

2.3.2 Channelized non-prewhitening matched filter

The second model is a channelized non-prewhitening matched filter (CNPW). We use the radially-symmetric sparse difference of Gaussians (S-DOG) channels,⁷ defined in the frequency domain by

$$C_j(\rho) = \exp \left[-\frac{1}{2} \left(\frac{\rho}{Q\sigma_j} \right)^2 \right] - \exp \left[-\frac{1}{2} \left(\frac{\rho}{\sigma_j} \right)^2 \right], \quad (3)$$

with $Q = 2.0$, $\sigma_0 = 0.015$, $\sigma_j = \sigma_0 2^j$. The sparse channel set consists of the first three channels, $j \in \{0, 1, 2\}$.

The CNPW observer computes the inner product

$$t = \text{channelize}(\text{signal})^t \text{channelize}(\text{image}). \quad (4)$$

The channelization operator produces a vector containing the three channel outputs. Channel output consists of the inner product of the channel template defined in equation 3 with the image.

Again we use area under the ROC curve for the CNPW model to predict human area under the LROC curve.

2.3.3 Channelized Hotelling observer

The channelized Hotelling observer (CHO) is similar to the CNPW, however it uses the channel covariance matrix K to do prewhitening step:

$$t = \text{channelize}(\text{signal})^t K^{-1} \text{channelize}(\text{image}). \quad (5)$$

We estimate the channel covariance matrix K using the reconstructed images.

As before, we use area under the model's ROC curve to predict human area under the LROC curve.

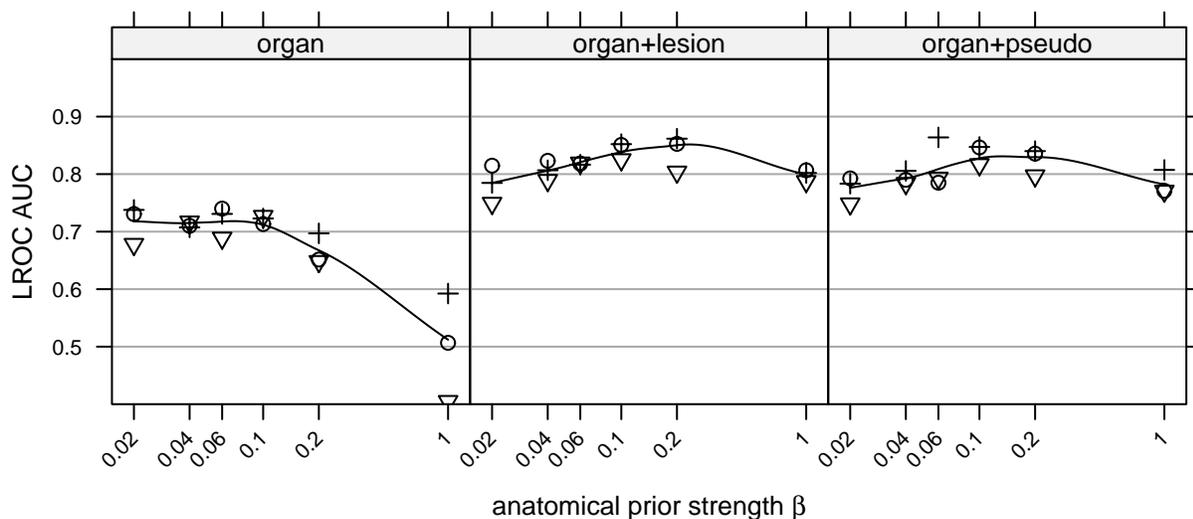


Figure 2. Results from the human observer study. Each symbol represents area under the LROC curve for one observer. The solid line is a LOESS fit. Note logarithmic scale on the horizontal axis.

3. RESULTS & DISCUSSION

3.1 Human observer results

Figure 2 shows results from the human observer study.³ The symbols on the graph indicate results for each observer at a variety of β values for each of the three types of prior information. Overlaid on top of the data is a LOESS fit. Because of the logarithmic scale on the horizontal axis the graphs span almost two orders of magnitude in prior strength.

The organ prior shows a broad plateau between $\beta = 0.02$ and $\beta = 0.1$, with no obvious peak. At larger values of β performance begins to dramatically drop off. These are the features we would like the models to predict.

The organ+lesion and organ+pseudo lesion both show a peak value somewhere between $\beta = 0.1$ and $\beta = 1.0$. (The human observer study undersampled this region, so we don't know exactly where the peak occurs.) Performance for all values of β is higher than on the lesion-only prior. We would like the models to tell us in which region human performance is optimized.

3.2 Model observer predictions

3.2.1 Non-prewhitening matched filter

Figure 3 shows predictions of the NPWMF model, with the human data overlaid. The model doesn't do a good job for any of the three prior types. Although the NPWMF model captures the dropoff at high β values of the organ prior, it does not show the broad plateau at lower β values seen in the human data.

The NPWMF model predicts that organ+lesion and organ+pseudo priors will have peak value at $\beta = 0.02$, at least an order of magnitude away from the actual peak in the human data which is in the region $0.1 < \beta < 1.0$. The model also predicts a more dramatic difference between performance at very high and very low β values than was found in the human data.

3.2.2 Channelized non-prewhitening matched filter

Figure 4 shows results for the CNPW model. As one might expect, including channels in the model lowers the predicted observer performance. The curves predicted by the CNPW model are a bit closer to those of the human than those of the NPWMF. However the CNPW still does not predict the plateau on the organ prior, nor does it accurately predict in which region the peak value for the other two priors occurs.

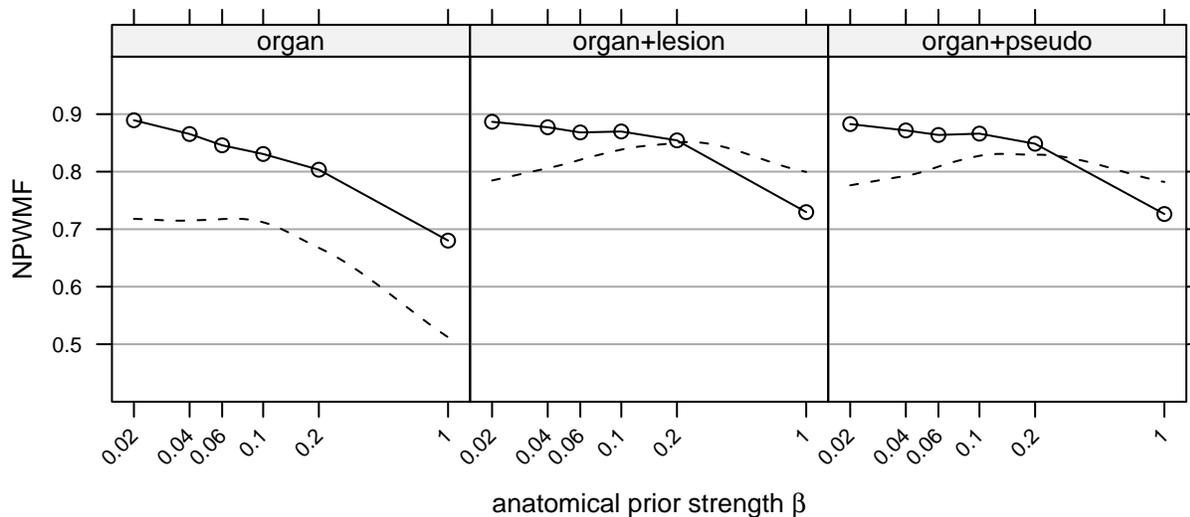


Figure 3. Model observer predictions from the nonprewhitening-matched filter. The dashed curve shows the human data.

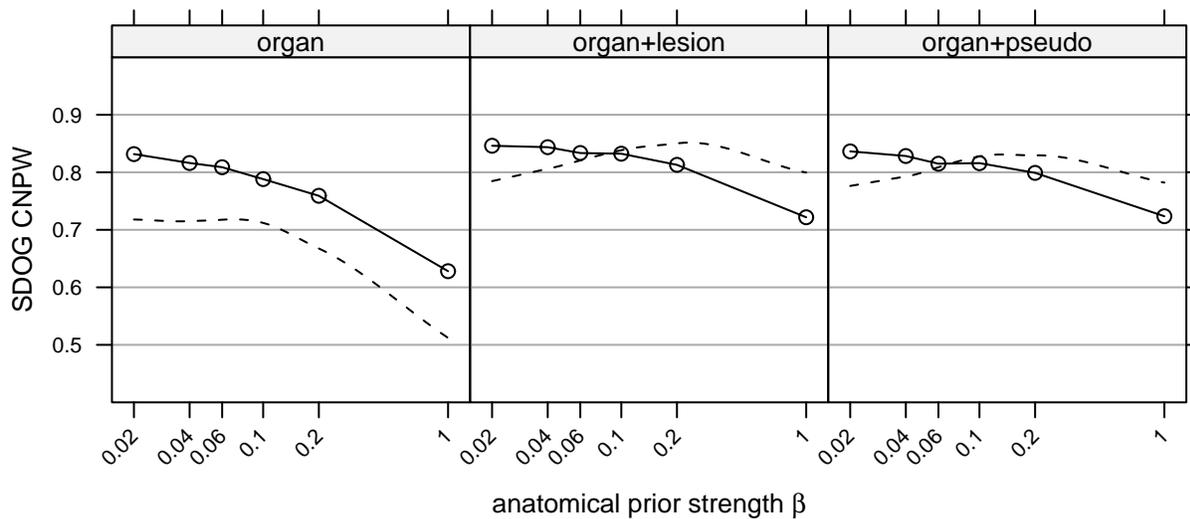


Figure 4. Model observer predictions from the channelized nonprewhitening observer, using the S-DOG channels. The dashed curve shows the human data.

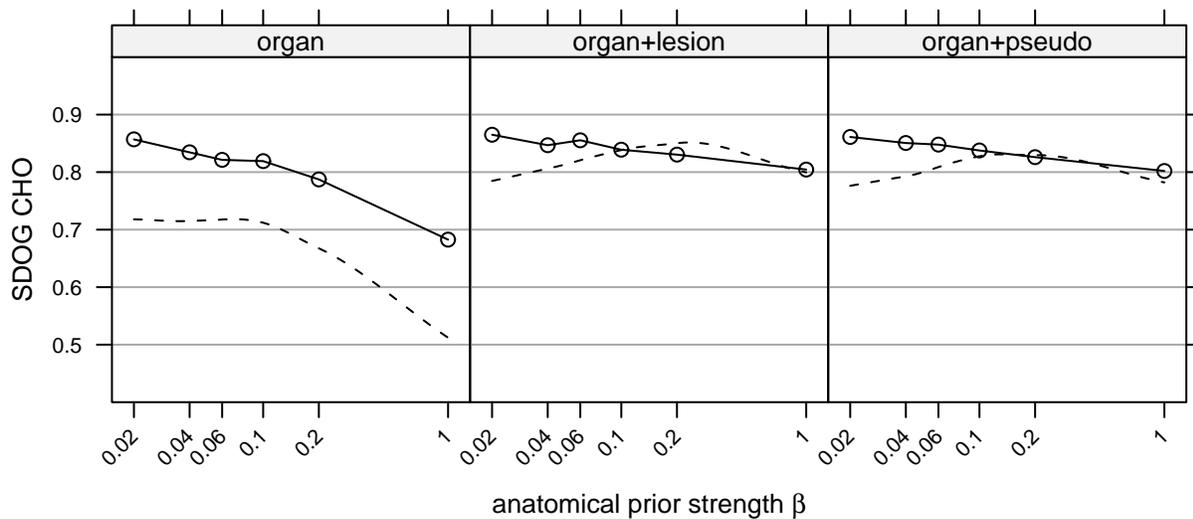


Figure 5. Model observer predictions from the channelized Hotelling observer, using the S-DOG channels. The dashed curve shows the human data.

3.2.3 Channelized Hotelling observer

Figure 5 shows results for the CHO model. As one might expect, using the covariance matrix to prewhiten the channel outputs increases the predicted observer performance when compared with the CNPW. Prewhitening also flattens the curves. However the including prewhitening in the CHO model does not predict the plateau in the organ prior, nor does it correctly predict the peak value of the other two priors.

4. CONCLUSION

The results of the human observer study presented in figure 2 show that observer performance varies as a function of prior strength β . The optimal value of β depends on the level of prior knowledge about the anatomy used during reconstruction.

None of the three models presented here, all of which ignore search, do a good job of predicting human performance. All of the models predict that $\beta = 0.02$ is the optimal prior strength for all three types of prior knowledge. This is off by over an order of magnitude for the organ+lesion and organ+pseudo priors. These models do not predict the shape of how the curve varies with β . This suggests that it is inappropriate to use an SKE task which ignores search to optimize prior strength in reconstruction algorithms. In future work we will include search in the models to see if that provides better predictions of human performance.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Institute of Biomedical Imaging and BioEngineering under grant R01 EB002798. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] Bruyant, P. P., Gifford, H. C., Gindi, G., and King, M. A., "Numerical observer study of MAP-OSEM regularization methods with anatomical priors for lesion detection in ^{67}Ga images," *IEEE Trans. Nucl. Sci.* **51**, 193–7 (Feb. 2004).

- [2] Lehovich, A., Bruyant, P. P., Gifford, H. C., Schneider, P. B., Squires, S., Licho, R., Gindi, G., and King, M. A., "Human-observer LROC study of lesion detection in Ga-67 SPECT images reconstructed using MAP with anatomical priors," in [2006 *IEEE Nuclear Science Symposium Conference Record*], 1699–702 (Oct. 2006).
- [3] Lehovich, A., Gifford, H. C., Schneider, P. B., and King, M. A., "Choosing anatomical-prior strength for MAP SPECT reconstruction to maximize lesion detectability," in [2007 *IEEE Nuclear Science Symposium Conference Record*], 4222–25 (Nov. 2007).
- [4] De Pierro, A. R., "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.* **14**, 132–7 (Mar. 1995).
- [5] Nawfel, R. D., Chan, K. H., Wagenaar, D. J., and Judy, P. F., "Evaluation of video gray-scale display," *Med. Phys.* **19**, 561–7 (May 1992).
- [6] Swensson, R. G., "Unified measurement of observer performance in detecting and localizing target objects on images," *Med. Phys.* **23**, 1709–25 (Oct. 1996).
- [7] Abbey, C. K. and Barrett, H. H., "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **18**, 473–488 (Mar. 2001).