

Journal of  
**Medical Imaging**

MedicalImaging.SPIEDigitalLibrary.org

**Accounting for anatomical noise in  
search-capable model observers for  
planar nuclear imaging**

Anando Sen  
Howard C. Gifford

# Accounting for anatomical noise in search-capable model observers for planar nuclear imaging

Anando Sen and Howard C. Gifford\*

University of Houston, Department of Biomedical Engineering, 3605 Cullen Boulevard, Houston, Texas 77004, United States

**Abstract.** Model observers intended to predict the diagnostic performance of human observers should account for the effects of both quantum and anatomical noise. We compared the abilities of several visual-search (VS) and scanning Hotelling-type models to account for anatomical noise in a localization receiver operating characteristic (LROC) study involving simulated nuclear medicine images. Our VS observer invoked a two-stage process of search and analysis. The images featured lesions in the prostate and pelvic lymph nodes. Lesion contrast and the geometric resolution and sensitivity of the imaging collimator were the study variables. A set of anthropomorphic mathematical phantoms was imaged with an analytic projector based on eight parallel-hole collimators with different sensitivity and resolution properties. The LROC study was conducted with human observers and the channelized nonprewhitening, channelized Hotelling (CH) and VS model observers. The CH observer was applied in a “background-known-statistically” protocol while the VS observer performed a quasi-background-known-exactly task. Both of these models were applied with and without internal noise in the decision variables. A perceptual search threshold was also tested with the VS observer. The model observers without inefficiencies failed to mimic the average performance trend for the humans. The CH and VS observers with internal noise matched the humans primarily at low collimator sensitivities. With both internal noise and the search threshold, the VS observer attained quantitative agreement with the human observers. Computational efficiency is an important advantage of the VS observer. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.3.1.015502](https://doi.org/10.1117/1.JMI.3.1.015502)]

Keywords: model observers; nuclear medicine; planar imaging; scanning observers; visual search; image quality; collimators; internal noise.

Paper 15064PRRR received Mar. 25, 2015; accepted for publication Dec. 23, 2015; published online Jan. 26, 2016.

## 1 Introduction

Medical images are routinely used to identify lesions within the human body. These images are read by radiologists, who generally make the final judgment about the presence of a lesion. Image quality is crucial for the radiologist to make the best possible judgment and may be improved by optimizing system designs and acquisition protocols. However, it is expensive and time consuming to conduct human-observer studies for this purpose at every stage of developmental research. This has led to the development of mathematical model observers as surrogates for humans in diagnostic imaging studies. Of particular interest are model observers which can predict human performance in clinically realistic tasks involving lesion search. Such tasks should probe how quantum and anatomical noise in the images affect observer performance. Anatomical noise is background structure that masquerades as a lesion or which obscures actual lesions. Together, quantum and anatomical noise comprise the image texture.

Popular model observers such as the channelized Hotelling (CH) observer<sup>1</sup> operate based on prior information in the form of image statistics. The statistics include covariance matrices that characterize the noise in an image set. The level of prior information afforded an observer is dictated by several well-known task paradigms. For signal-known-exactly detection tasks, the

observer need only decide lesion presence at a fixed location. Information about the background can be either exact [background-known-exactly (BKE)] or statistical [background-known-statistically (BKS)]. The task of lesion detection and localization is signal-known-statistically (SKS) in nature (e.g., the shape of the lesion profile may be known although its location is variable), but may additionally be either BKE or BKS. These search tasks can be performed by scanning versions of the CH observer and other Hotelling-type models.<sup>2</sup>

With a BKE search task, a scanning model observer has knowledge of the quantum-mean background corresponding to the given test image. This knowledge greatly reduces the impact of anatomical noise on task performance. For BKS search tasks, the scanning CH observer computes image class means and covariances that account for anatomical variations. This computation can be extensive since the covariance matrices must be computed at all possible lesion locations. Moreover, it is unclear whether, given the covariance information, the scanning CH observer can adequately quantify how human observers respond to anatomical noise in individual images.

As an alternate approach, we have been investigating observer models which use less statistical knowledge.<sup>3,4</sup> These models perform distinct target search and analysis procedures, thereby simulating the visual-search (VS) process that trained radiologists engage in to interpret medical images.<sup>5</sup> Our VS model observer applies a feature-guided search that identifies target-like structure for subsequent analysis. For this study, the observer identified hot “blobs” (regions of elevated tracer

\*Address all correspondence to: Howard C. Gifford, E-mail: [hgifford@uh.edu](mailto:hgifford@uh.edu)

uptake) as candidates, which may be actual lesions or false positives. Statistical information about the image background was only supplied for the candidate analysis, at which point some nonconspicuous lesions may have been ignored. The SKS task paradigm is a key constraint for the model, as many of the complex processes of human vision and cognition are not accounted for. The effects of image texture are nonetheless intrinsic to search performance with this relatively simple model.

We tested the VS and scanning CH observers as competing ways of predicting human-observer performance in a realistic search task involving both quantum and anatomical noise. The task was detection and localization of pelvic lesions in simulated In-111 planar nuclear medicine images. A localization receiver operating characteristic (LROC) study examined how observer performance was affected by the resolution-versus-sensitivity trade-offs among a family of medium-energy imaging collimators. The human observers in our study read a subset of the images considered by the model observers. Comparison was also made with a channelized nonprewhitening (CNPW) observer that performed both BKE and BKS search tasks.

Human observers are often modeled as ideal observers with added inefficiencies.<sup>6</sup> In our study, the CH and VS observers were applied with and without internal noise that was added to the observers' decision variables. In addition, the VS model was tested with a perceptual search threshold<sup>7</sup> that rejected relatively nonsuspicious search candidates. Various covariance-based internal-noise models have been proposed for the CH observer,<sup>8</sup> but the motivations for choosing one model or another are often unclear. The inefficiencies that were tested for the VS observer may afford a better intuition.

## 2 Model Observers

### 2.1 Framework for Localization Receiver Operating Characteristic Studies

All search tasks in our study used two-dimensional (2-D) slices extracted from reconstructed three-dimensional (3-D) volumes. A test image of size  $N \times N$  is represented by the  $N^2 \times 1$  vector  $\mathbf{g}$ . This image could contain a single lesion within a region of interest (ROI)  $\Omega$ . With the lesion centered at pixel  $j \in \Omega$ , the test image can be regarded as

$$\mathbf{g} = \mathbf{b} + \mathbf{s}_j + \mathbf{n}, \quad (1)$$

with anatomical background  $\mathbf{b}$ , lesion image  $\mathbf{s}_j$ , and zero-mean quantum noise  $\mathbf{n}$ . Lesion-absent images are comprised of only  $\mathbf{b}$  and  $\mathbf{n}$ . The background can vary with image. We denote the quantum-mean lesion-present image for fixed location  $j$  as  $\langle \mathbf{g} \rangle_{\mathbf{n}|\mathbf{b},j} = \mathbf{b} + \mathbf{s}_j$ , where the bracket notation indicates an average over  $\mathbf{n}$  with  $\mathbf{b}$  and  $j$  fixed. The corresponding mean lesion-absent image is  $\langle \mathbf{g} \rangle_{\mathbf{n}|\mathbf{b},0} = \mathbf{b}$ . With additional averaging over anatomical background, we have the mean images  $\langle \mathbf{g} \rangle_{\mathbf{n},\mathbf{b}|0} = \bar{\mathbf{b}}$  and  $\langle \mathbf{g} \rangle_{\mathbf{n},\mathbf{b}|j} = \bar{\mathbf{b}} + \mathbf{s}_j$ . Thus,  $\bar{\mathbf{b}}$  is the mean background image computed over all anatomical realizations of  $\mathbf{b}$ .

Given test image  $\mathbf{g}$  in a LROC study, a model observer first computes a perception measurement  $\lambda_j$  for every location  $j$  within  $\Omega$ , and then reports the most suspicious lesion location  $r$  and a confidence rating  $\lambda$  according to the formulas

$$\lambda = \max_{j \in \Omega} \lambda_j, \quad (2)$$

$$r = \arg \max_{j \in \Omega} \lambda_j. \quad (3)$$

An image with rating  $\lambda$  is read as lesion-present (or abnormal) at threshold  $\lambda_t$  if  $\lambda > \lambda_t$ . An abnormal image is said to have been localized correctly if the predicted location is within a threshold distance [the radius of correct localization (RCL)] of the true location. The LROC curve plots the probability of a true-positive response conditioned on correct localization against the probability of a false-positive response as  $\lambda_t$  varies. The area under the LROC curve is an accepted figure of merit for observer performance.

### 2.2 Scanning Observers for Detection-Localization Tasks

The Hotelling-type scanning observers in this work computed linear perception measurements for all  $j \in \Omega$  using the general formula

$$\lambda_j^{\text{obs}} = (\mathbf{w}_j^{\text{obs}})^t [\mathbf{g} - \mathbf{c}_j], \quad (4)$$

where the superscript  $t$  indicates transpose and  $\mathbf{w}_j$  is a location- and observer-specific scan template that is the same size as  $\mathbf{g}$ . The location-dependent reference image  $\mathbf{c}_j$  sets the measurement origin. With BKE tasks

$$\mathbf{c}_j = \frac{\langle \mathbf{g} \rangle_{\mathbf{n}|\mathbf{b},0} + \langle \mathbf{g} \rangle_{\mathbf{n}|\mathbf{b},j}}{2} = \mathbf{b} + \frac{1}{2} \mathbf{s}_j, \quad (5)$$

while  $\mathbf{b}$  is replaced by  $\bar{\mathbf{b}}$  for BKS tasks.<sup>9</sup>

For this work, the lesion profile was assumed to be shift-invariant. In fact, the lesion profile in our test images changed slightly with location due to attenuation effects, but the effect was largely imperceptible. We thus computed a location-averaged profile  $\bar{\mathbf{s}}$  for use by the model observers. The subscript on  $\bar{\mathbf{s}}_j$  implies a shift of the profile to the  $j$ th location. How this mean profile was computed is described in Sec. 3.5.

### 2.3 The Scanning Channelized Hotelling Observer

For BKS search tasks, the scanning CH observer uses the reference image

$$\mathbf{c}_j = \bar{\mathbf{b}} + \frac{1}{2} \bar{\mathbf{s}}_j, \quad (6)$$

in which we have accounted for the shift-invariant lesion profile. The scanning template for the CH observer takes the form

$$\mathbf{w}_j^{\text{CH}} = \mathbf{U}_j \mathbf{K}_j^{-1} \mathbf{U}_j^t \bar{\mathbf{s}}_j, \quad (7)$$

where  $\mathbf{U}_j$  is a shift-invariant  $N^2 \times C$  matrix whose columns contain the spatial responses at location  $j$  from a fixed set of  $C$  ( $C \ll N$ ) spatial-frequency channels, and  $\mathbf{K}_j$  is the location-specific  $C \times C$  channel covariance matrix. For this work, the matrix was computed as

$$\mathbf{K}_j = \langle \mathbf{U}_j^t (\mathbf{g} - \bar{\mathbf{b}}) (\mathbf{g} - \bar{\mathbf{b}})^t \mathbf{U}_j \rangle_{\mathbf{n},\mathbf{b}|0}, \quad (8)$$

where the bracket notation indicates only lesion-absent images were used based on the assumption that the presence of a low-contrast lesion would not substantially affect these calculations. The dependence on  $\mathbf{n}$  and  $\mathbf{b}$  in Eq. (8) is implicit in  $\mathbf{g}$  [see

Eq. (1)]. It is important to note that the CH template is shift-variant due to the location-specific covariances.

When  $\mathbf{K}_j$  includes the anatomical variations, the scanning CH observer operates under an SKS-BKS task paradigm. In that case,  $\mathbf{K}_j$  can be decomposed into a sum of quantum-noise and anatomical-noise components.<sup>10</sup> A mean quantum-noise covariance  $\mathbf{K}_j^{\text{quant}}$  is

$$\mathbf{K}_j^{\text{quant}} = \langle \langle \mathbf{U}_j^t (\mathbf{g} - \mathbf{b})(\mathbf{g} - \mathbf{b})^t \mathbf{U}_j \rangle_{\mathbf{n}|\mathbf{b},0} \rangle_{\mathbf{b}}. \quad (9)$$

The quantum-mean backgrounds are also used to compute the anatomical-noise covariance via the formula

$$\mathbf{K}_j^{\text{anat}} = \langle \mathbf{U}_j^t (\mathbf{b} - \bar{\mathbf{b}})(\mathbf{b} - \bar{\mathbf{b}})^t \mathbf{U}_j \rangle_{\mathbf{b}}. \quad (10)$$

## 2.4 Visual-Search Observers

The VS observer<sup>3</sup> offers an alternative approach to accounting for anatomical noise in model observer performance of lesion-search tasks. Based on the VS paradigm proposed by Kundel et al.<sup>5</sup> for how radiologists read images, the VS observer combines a front-end search for suspicious candidate locations with subsequent analysis of just those candidates. Lesion detection in nuclear medicine is frequently a hot-spot search, which our basic VS observer performs through segmentation of the test image into blobs. As background knowledge is not incorporated into the search, a given blob may be a lesion or an artifact of the image texture. At the same time, actual lesions may be masked by the texture. The search process is thereby implicitly affected by both quantum and anatomical noise. The pixel with the maximum grayscale intensity in a given blob constitutes the blob focal point. The set of focal points within the ROI  $\Omega$  comprises a relatively small set of candidate locations for the observer.

The search is followed by a directed analysis of these candidate locations. Various discriminants may be used for this analysis. Many of our earlier studies<sup>3,11</sup> have used the scanning CNPW discriminant. In scanning mode, the CNPW observer can be viewed as a nonprewhitening approximation to the CHO, with the shift-invariant template

$$\mathbf{w}_j^{\text{CNPW}} = \mathbf{U}_j \mathbf{U}_j^t \bar{\mathbf{s}}_j. \quad (11)$$

Herein, the CNPW discriminant was applied to the candidate locations under a BKE assumption. With this form of candidate analysis, the VS observer task as a whole is quasi-BKE. Given the shift-invariant lesion profile, the BKE reference image for Eq. (4) is  $\mathbf{c}_j = \mathbf{b} + \frac{1}{2}\bar{\mathbf{s}}_j$ , but the lesion term may be ignored since the CNPW template is also shift-invariant.

The CNPW discriminant was also tested separately as a scanning observer in our study, being applied for both BKE and BKS tasks. The BKE task with  $\mathbf{c}_j = \mathbf{b}$  represented lesion detection in reconstructed quantum noise. The reference image for the BKS task was  $\mathbf{c}_j = \bar{\mathbf{b}}$ . We shall use the notation CNPW-BKE and CNPW-BKS to differentiate these two observer/task combinations.

## 3 Methods

### 3.1 Model-Observer Specifications

Based on previous work,<sup>2,12</sup> a set of three difference-of-Gaussian channels was selected for both the CH and CNPW

observers. Defined in the frequency domain, the  $i$ th channel  $\mathbf{u}_i$  has elements

$$\mathbf{u}_i(\boldsymbol{\xi}) = \exp \left[ - \left( \frac{\|\boldsymbol{\xi}\|}{2^{i+1}\sigma_0} \right)^2 \right] - \exp \left[ - \left( \frac{\|\boldsymbol{\xi}\|}{2^i\sigma_0} \right)^2 \right], \quad i = 0, 1, 2, \quad (12)$$

with  $\boldsymbol{\xi}$  the 2-D spatial frequency in cycles per pixel and  $\sigma_0 = 0.015$ . The spatial response of the channel is the inverse Fourier transform of  $\mathbf{u}_i$ .

The image segmentation for the VS observer was performed with a watershed algorithm, a computationally efficient routine that is available in many software packages. A less-efficient gradient-ascent segmentation method yielded similar results in previous VS observer studies.<sup>4</sup> With the watershed algorithm, a 2-D test image is viewed as a topographic landscape with holes at the local minima.<sup>13</sup> When the landscape is lowered into a tank of water, water starts to fill the ‘‘catchment basins’’ around the holes. Dams (or watersheds) are constructed at points where water from two or more catchment basins meet. Once the entire landscape has been submerged, these catchment basins define the segmented blobs. The watershed algorithm is frequently applied to the gradient magnitude of an image. We instead applied the algorithm to the additive inverse of the test image in order to avoid the noise penalties involved in computing the gradient. With this approach, the minimum pixel within a given catchment basin represented a focal point.

Details about the training for the model observers are provided in Sec. 3.5.

## 3.2 Model-Observer Inefficiencies

### 3.2.1 Internal noise

Given the same image, human observers can make different decisions at different times. With the model-observer formulations presented above, the decision for a given image would never change. To address this issue, we experimented with adding internal noise to the CH and VS observers. The noise was implemented by adding a random Gaussian deviate to the perception measurements  $\lambda_j$ . The width of the Gaussian distribution was selected relative to the standard deviation  $\sigma_j$  of the set  $\{\lambda_j : j \in \Omega\}$  for a given set of training images. With  $N(0,1)$  representing a random deviate from a Gaussian distribution with zero mean and unit standard deviation, the noisy measurement was given by

$$\tilde{\lambda}_j = \lambda_j + N(0, \gamma\sigma_j). \quad (13)$$

Values of  $\gamma$  between 0.5 and 2.0 were tested.

For the CH observer, internal noise can also be implemented through the covariance matrices instead of the analysis statistic. Several models of this type are discussed by Zhang et al.<sup>8</sup> A corresponding process for the VS observer would be to add internal noise during the candidate search.<sup>7</sup>

### 3.2.2 Perceptual search threshold

A second type of model inefficiency applied in our study was specific to the VS observer. Recall from Sec. 2.4 that the VS observer identifies blob focal points as suspicious candidates by means of image segmentation. The segmentation can produce many candidates that human observers would ignore. The result,



particularly with high-count images, is that the BKE analysis with the CNPW discriminant will often locate the lesion. In such cases, the VS and CNPW-BKE observers may perform similarly. A solution is to limit the set of candidates by setting a threshold  $\tau$  such that a focal point with greyscale intensity less than  $\tau$  is disregarded. Parameter  $\tau$  is thus comparable to the drowning threshold described in the watershed literature.<sup>14</sup>

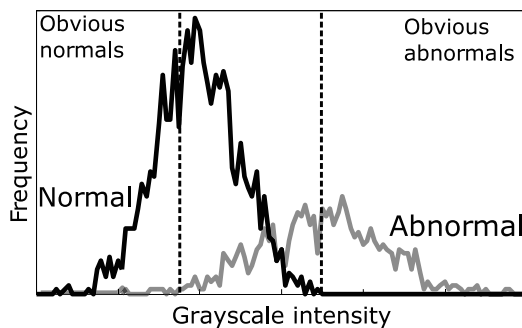
This threshold was calculated with reference to the focal points from a set of training images. Each training image yielded a number of focal points, some of which were lesion-absent. The set of greyscale intensities from all of the lesion-absent focal points in the training images produced a distribution with mean and standard deviation denoted by  $\mu_l$  and  $\sigma_l$ , respectively. We tested search thresholds of the form  $\tau = \mu_l + \beta\sigma_l$  for four different scaling factors  $\beta$  between 0.2 and 0.8.

Figure 1 shows the greyscale distributions of the normal and abnormal focal points obtained from a training set of images. The distributions are approximately Gaussian. Threshold values of particular interest might lie in the zone between the dashed vertical lines where the disease status is not obvious.

### 3.3 Planar Imaging Simulation

The 3-D XCAT anthropomorphic phantom<sup>15</sup> was used for our studies, with the volume of interest consisting of the pelvic and abdominal regions. This region was discretized to voxel dimensions  $128^3$ . Five In-111 biodistributions were simulated in the volume and then imaged with an analytic projector. Our simulation modeled a system with medium-energy parallel-hole collimators and accounted for nonhomogeneous photon attenuation and distance-dependent collimator blur. Eight collimators that varied in terms of spatial resolution and sensitivity were tested (see Sec. 3.4). Noise-free projections were calculated for the two principle In-111 energies (171 and 245 keV), scaled for the abundance and absorption coefficients of a NaI detector with 1-cm crystal thickness, and then added. Additional details about the system model can be found in our previous work.<sup>11</sup>

Spherical soft-tissue lesions of diameter 1 cm were considered. A lesion-placement map contained the prostate and lymph-node regions of the XCAT volume. The projection of this map represented the search region  $\Omega$  for the model observers in our study. A total of 225 distinct lesion locations were randomly generated from the map, with roughly equal distribution between the prostate and the lymph nodes. Two lesion-to-prostate relative activities (or contrasts) of 16:1 and 24:1 were used. Lesion projections were created separately, scaled for contrast and then added to the noise-free background projections to form noise-free lesion-present data.



**Fig. 1** Example greyscale intensity distributions for the normal and abnormal focal points identified by the VS observer in an image set.

The addition of Poisson noise to the projections was guided by photon count levels from clinical data. The mean number of counts in a projection varied with collimator as described next.

### 3.4 Collimators and System Resolution

Our LROC study compared observer performance with different collimators. This section reviews the properties of a gamma-camera collimator and how these properties affect image formation. The two main parameters for a parallel-hole collimator are the hole length  $L$  and the hole width  $d$ , both of which affect the collimator blur and count sensitivity. The number of counts per unit time  $\eta$  for regular hexagonal holes obeys the proportionality  $\eta \sim 0.65d^2$ .<sup>16</sup>

The system spatial resolution  $R_{\text{sys}}$  of a gamma camera has two components: intrinsic resolution and distance-dependent geometric resolution. Expressed in terms of the full-width at half-maximum of a Gaussian blur function, the resolution for photons emitted at a distance  $c$  from the face of the collimator is

$$R_{\text{sys}}(c) = \sqrt{R_{\text{int}}^2 + R_{\text{geom}}^2(c)}. \quad (14)$$

The geometric component of the system resolution is completely determined by  $L$  and  $d$  according to the formula

$$R_{\text{geom}}(c) = \frac{dc}{L}, \quad (15)$$

while the intrinsic resolution depends on the gamma-ray energy and the crystal material. For the In-111 energies and a NaI detector, typical intrinsic resolutions are between 0.3 and 0.4 cm.<sup>17</sup> Note that geometric resolution can also be defined in terms of distance from the face of the crystal.

For typical values of  $c$  in patient imaging, the system resolution may be expressed with the linear model

$$R_{\text{sys}}(c) = mc + R_0, \quad (16)$$

with slope  $m = d/L$  and the intercept  $R_0 \approx R_{\text{sys}}(c) - mc$  for a suitably large value of  $c$ .

Our study modeled variants of a typical medium-energy collimator with  $L = 4.06$  cm and  $d = 0.294$  cm. Eight collimators were defined by varying  $d$  between 0.044 and 0.394 cm in increments of 0.05 cm while keeping  $L$  constant. The intrinsic resolution was fixed at 0.38 cm. The relevant collimator parameters are listed in Table 1. The collimators are labeled using the notation C1, ..., C8, listed in order of increasing  $d$ . The resolutions given in the table were computed for a distance  $c = 4L$ . The sensitivities  $\eta$  have units of millions of counts per unit time.

As  $\eta$  and  $R_{\text{sys}}$  both increase monotonically with  $d$ , our study examined the trade-off between sensitivity and spatial resolution in terms of observer performance. Low sensitivity leads to relatively higher quantum noise in the images, whereas low resolution increases anatomical noise due to partial-volume and blurring effects. Sample images obtained with the collimator models for two lesion-present cases are shown in Fig. 2.

### 3.5 Observer Studies

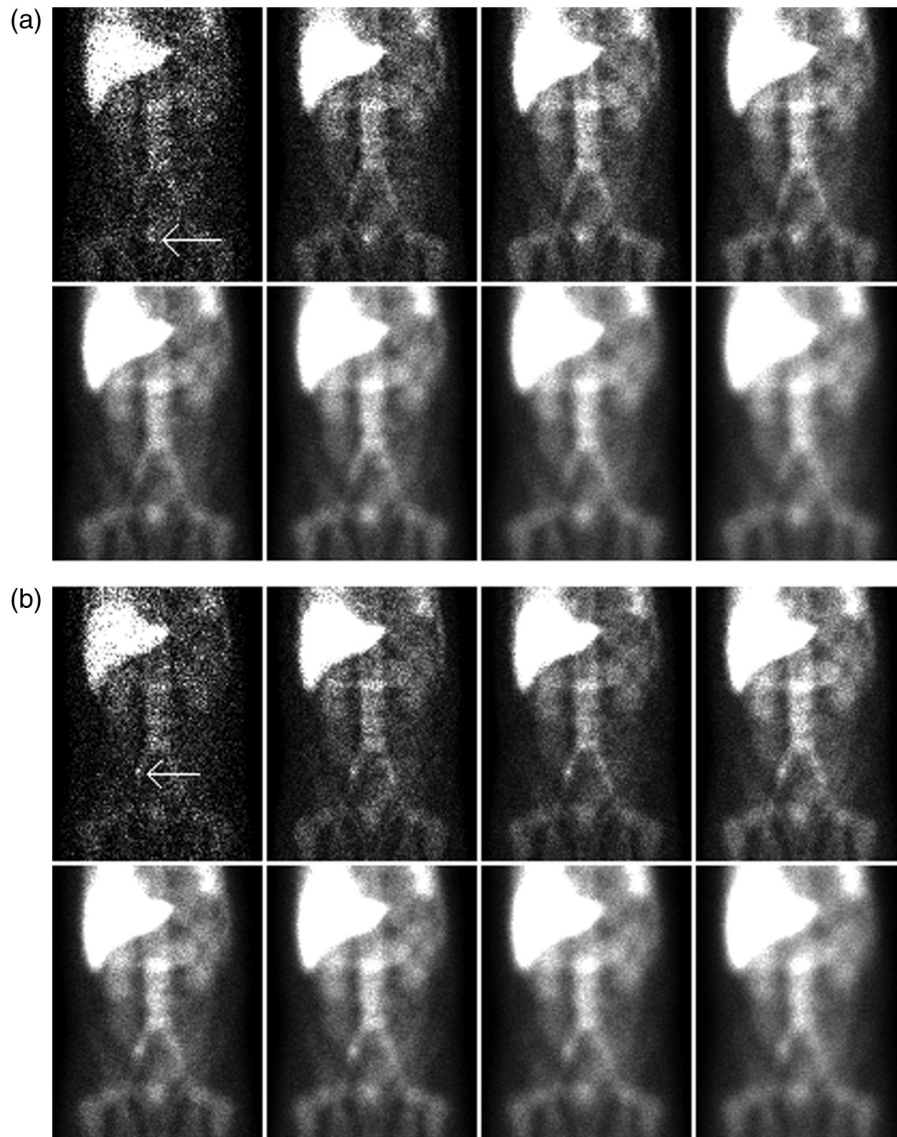
The model observers read images for each of the 16 combinations ( $2 \times 8 = 16$ ) of lesion contrast and collimator type in our study. For each combination, three image sets were prepared. A set of 450 test images consisted of a lesion-present/lesion-absent

**Table 1** Parameters for the eight collimator models. System resolution  $R_{\text{sys}}$  was measured at distance  $c = 4L$  from the collimator face. The count sensitivity  $\eta$  has units of millions of counts per unit time.

	Collimator							
	C1	C2	C3	C4	C5	C6	C7	C8
$d$ (cm)	0.044	0.094	0.144	0.194	0.244	0.294	0.344	0.394
$R_{\text{sys}}$ (cm)	0.402	0.473	0.575	0.695	0.824	0.960	1.100	1.240
$\eta$	0.04	0.19	0.45	0.81	1.23	1.87	2.55	3.35

pair for each of the 225 lesion locations. The remaining two sets were for model-observer training. One set of 225 images, representing different quantum-noise realizations for the 225 lesion-absent test images, was used to compute the channel covariance matrices for the CH observer. The channel covariance matrices were precomputed (but not preinverted) for

the study. The second training set contained 25 lesion-present/lesion-absent image pairs that were generated from distinct quantum-noise realizations for a subset of the 225 lesion locations. This latter set was used to estimate the mean lesion profile  $\bar{s}$ . Taking the difference between a given image pair yielded a noisy lesion profile centered at one location. This profile was

**Fig. 2** Sample study images obtained for a pair of high-contrast lesion cases: (a) a prostate lesion; (b) a lymph-node lesion. For both cases the first row consists of images from collimators C1 to C4 and the bottom row C5 to C8. The lesion position for each case is indicated by the white arrow in the first image.

shifted to the center of the field of view. This process was repeated for all the image pairs and the resultant center-shifted profiles were averaged to get  $\bar{s}$ . Note that the quantum-mean backgrounds used by the model observers were approximated by noise-free reconstructions.

Institutional approval was obtained to allow the participation of human observers and informed consent was obtained from each observer prior to the study. The human-observer study was restricted to the high-contrast image sets from the five collimators with the highest model-observer performances. Three nonradiologist imaging scientists participated in the study. Each observer read 25 training images and 50 test images per collimator. The observers provided confidence ratings on a four-point scale, with a rating of four implying high confidence that the image contains a lesion.

Correct localizations for all observers were scored based on an RCL of five pixels. This threshold radius was selected by following the empirical graphing process described in Wells et al.<sup>18</sup> For each observer, a Wilcoxon estimate<sup>19</sup> of the area under the LROC curve ( $A_L$ ) was computed. The average human performance for a given collimator was computed as the average of  $A_L$  from the three observers. Uncertainties are expressed as  $\pm$ one standard error in these averages.<sup>20</sup> For a model observer featuring internal noise,  $A_L$  was computed as the average performance based on three study realizations. The computational expense of running the CH observer was the reason for using only three realizations.

Separate analyses of variance (ANOVA) were conducted to test the statistical significance of the human and model-observer results. Collimator model and observer were fixed factors for the two-way ANOVA based on the human data. A three-way ANOVA with the model-observer data included lesion contrast as a main factor and the relevant two-way interactions. Significance was evaluated at the  $\alpha = 0.05$  level.

### 4 Results

The human-observer results from our study are summarized in Table 2. The uncertainties in individual observer performance were on the order of  $\pm 0.06$ . The two-way ANOVA indicated a significant collimator effect ( $p = 8.6E-4$ ) but the observer effect was not significant ( $p = 0.075$ ). A subsequent Tukey multiple comparisons test<sup>21</sup> found significant differences between C6 and the group {C2, C3, C4}. Collimator C5 was positioned in-between, with the data unable to distinguish it from C4 or C6.

Figure 3 shows how the model observers without internal noise or search threshold performed as a function of collimator and lesion contrast. The average human-observer results are included in Fig. 3(b). The uncertainties in the  $A_L$  estimates

for each of the model observers did not exceed  $\pm 0.02$ . With the high-contrast lesions, the model observers showed some qualitative similarities with the humans for the low-sensitivity collimators, but did not match the substantial drop in human performance that occurred with increased sensitivity.

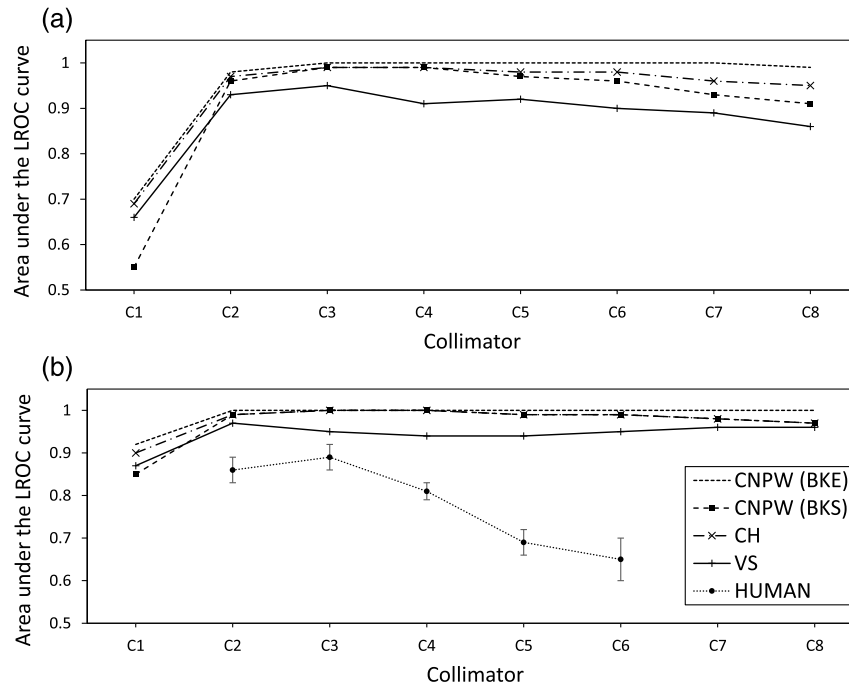
The ANOVA (Table 3) and multiple-comparisons testing applied to the model-observer data indicated that collimator C1 performed significantly worse than the other collimators. The differences between each observer were also significant. Affected almost entirely by quantum noise, the CNPW-BKE observer consistently outperformed the other three model observers, each of which also contended to some degree with anatomical noise. In particular, the CNPW-BKE observer performed at a high level with collimators C2 to C8 regardless of lesion contrast. The differences in performance that occurred between the CNPW-BKE and VS observers are attributable solely to the initial candidate search performed by the latter. The greatest deviations from CNPW-BKE performance occurred with the CNPW-BKS observer for collimator C1. The two CNPW observers otherwise performed similarly for collimators C2 to C4, only beginning to diverge once more with increased anatomical noise. The performance trends for the CNPW-BKS and CH observers were also similar, although the covariance noise modeling for the CH observer substantially moderated the effects of the reference-image subtraction for some collimators. The CH observer consistently outperformed the VS observer.

The performance trends in Fig. 3 for the scanning observers were largely independent of lesion contrast. This was not entirely so for the VS observer, for which Fig. 3(b) demonstrates an upward deflection in performance with the higher collimator sensitivities that does not show in Fig. 3(a). The deflection is small, with  $A_L$  increasing by 0.02 between C4 and C8, while performance with the low-contrast lesions decreased over that span. The BKE analysis is at the root of this variable behavior, amplifying the image quality effects of the higher lesion contrast and increased sensitivity. As shown below, some of the effects of the BKE assumption can be mitigated by the addition of perceptual inefficiencies to the observer.

Of course, the scanning models also relied on some prior knowledge of the image backgrounds that the human observers did not have. Internal noise is routinely added to Hotelling-type observers to compensate for this knowledge. The results from adding Gaussian noise to the perception measurements for the high-contrast lesions with the CH observer are shown in Fig. 4. The standard errors in the estimates of  $A_L$  for the model observers with inefficiencies were  $\sim \pm 0.03$ . Within the range of  $\gamma$  values tested, the added noise steadily reduced performance at all collimator sensitivities, but tended to penalize

**Table 2** Individual and average human-observer performances for the five tested collimator models C2–C6. The uncertainties in  $A_L$  for the individual observers are  $\pm 0.06$ . The uncertainties in average performance represent one standard error in the mean.

Observer	Collimator				
	C2	C3	C4	C5	C6
Human #1	0.91	0.91	0.84	0.74	0.72
Human #2	0.80	0.83	0.78	0.69	0.67
Human #3	0.87	0.92	0.80	0.63	0.55
Average	$0.86 \pm 0.03$	$0.89 \pm 0.03$	$0.81 \pm 0.02$	$0.69 \pm 0.03$	$0.65 \pm 0.05$



**Fig. 3** Model-observer and average human-observer performance as a function of collimator and lesion contrast. The model observers were applied without inefficiencies. (a) Model-observer performances with the low-contrast lesions; (b) human and model-observer performances with the high-contrast lesions. Uncertainties for the model observers did not exceed  $\pm 0.02$ .

**Table 3** Results from the three-way ANOVA conducted with the model-observer scores. The analysis tested collimator, observer and lesion contrast as factors. All three effects and two-way interactions were significant at the  $\alpha = 0.05$  level.

Factor	df	ss	F	Pr (>F)
Collimator	7	0.30	193.94	<2.2E-16
Contrast	1	0.040	181.50	<8.4E-12
Observer	3	0.037	56.14	<3.4E-10
Collimator:Contrast	7	0.081	53.24	<6.0E-12
Collimator:Observer	21	0.016	3.44	0.003
Contrast:Observer	3	0.0036	5.46	0.006
Residuals	21	0.0046		

performance with the lower sensitivities most. Thus, none of the model observer curves in Fig. 4 reproduced the performance trend displayed in the average human observer data. The left and right ends of the human performance curve were quantitatively fit with different values of  $\gamma$ .

Figure 5 shows the individual effects of internal noise and search thresholding for the VS observer. Figure 5(a) pertains to the internal-noise results. As with the CH observer, changes in  $\gamma$  generated performance changes that were fairly uniform with sensitivity. However, the VS observer was less affected by a given value of  $\gamma$  compared to the CH observer. An interesting outcome for  $\gamma \geq 1.5$  was the establishment of a nominal performance peak for C3 that mirrored the average human observer results.

With the search threshold [Fig. 5(b)], the VS observer was able to duplicate the considerable drop in average human performance that occurred at the higher sensitivities. However, the threshold had a negligible effect on observer performance at the lower sensitivities. The relatively high system resolutions with those collimator models ensured that actual lesions would generally be associated with focal points having greyscale maxima that exceeded the threshold.

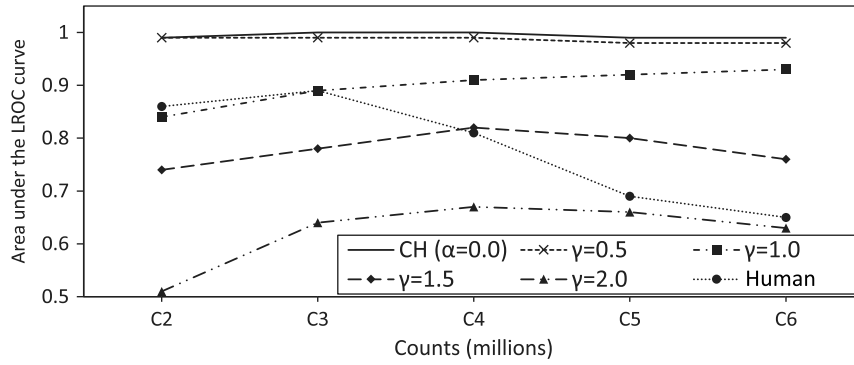
We also investigated how the VS observer responded with internal noise and the search threshold together. There were 16 inefficiency models in all, corresponding to the four values of  $\beta$  and four nonzero values of  $\gamma$  that were used for the plots in Fig. 5. Figure 6 compares the average human performance with the best-fitting results from the VS observer that used  $\gamma = 1.5$  and  $\beta = 0.6$ .

The computation times for the scanning and VS model observers in this study were disparate. Recall that the channel covariance matrices for the CH observer were computed (but not inverted) prior to the study. Despite this preprocessing, the CH computation times were relatively high:  $\sim 90$  min in IDL (Interactive Data Language, Exelis Visual Information Solutions, Boulder, Colorado) to read 450 test images compared to less than a minute for the VS and CNPW observers. These timings are for the model observers without internal noise or search thresholding. The lesion search region for the model observers contained 380 locations. From this, the VS observer averaged between 20 and 40 focal points per image, with the higher numbers associated with lower collimator sensitivities. However, the code implementation of the candidate analysis did not exploit this search reduction.

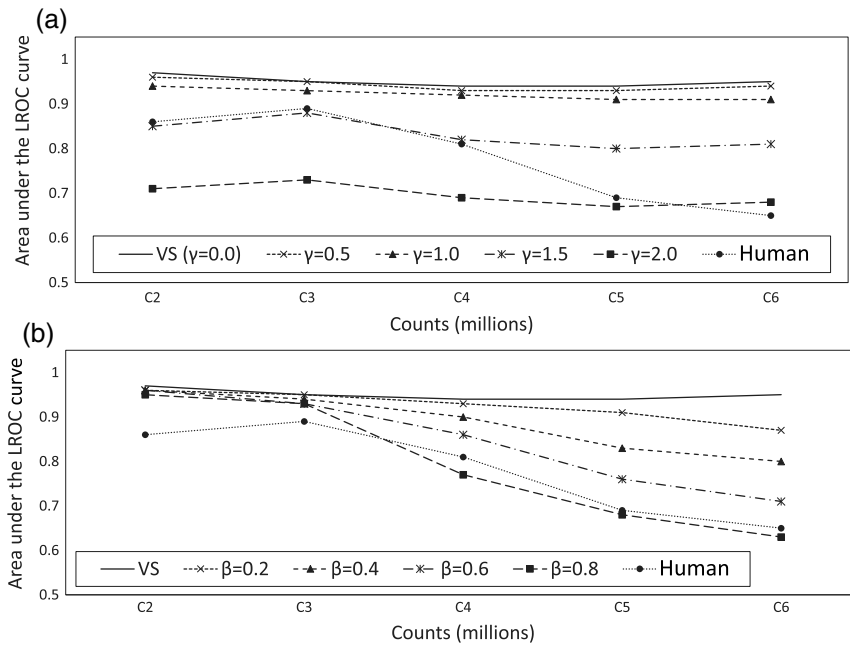
## 5 Discussion

Our objective with this work was to compare the VS and scanning CH observers as surrogates for human observers in task-

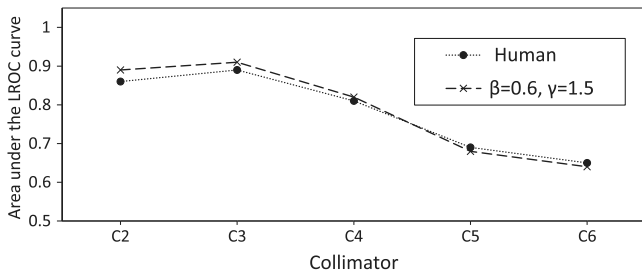




**Fig. 4** Effects of internal noise on CH observer performance with the high-contrast lesions. The noise parameter  $\gamma$  varied from 0 to 2.0, with  $\gamma = 0$  indicating no noise. Also shown is average human performance. The uncertainties in model-observer performance were  $\pm 0.03$ .



**Fig. 5** Individual effects of internal noise and search thresholding on VS observer performance with the high-contrast lesions. (a) Effects of internal noise; the noise parameter  $\gamma$  varied from 0 to 2.0, with  $\gamma = 0$  indicating no noise. (b) Threshold effects; the threshold parameter  $\beta$  varied from 0.2 to 0.8. The uncertainties in model-observer performance were  $\pm 0.03$ .



**Fig. 6** Effects of applying the VS observer with multiple inefficiencies. Average human performance is compared with the performance from the VS observer corresponding to  $\gamma = 1.5$  and  $\beta = 0.6$ . The uncertainties in model-observer performance were  $\pm 0.03$ .

based assessments. A focus for the comparison was how these model observers respond to anatomical noise. The image sets in the planar imaging study presented varying proportions of quantum and anatomical noise as dictated by the collimator sensitivity and spatial resolution. Without inefficiencies, neither model observer matched the magnitudes or general trend in average human performance as anatomical noise increased in the images. The results with the observer inefficiencies show that the two models admit different sets of mechanisms for improving the agreement with the humans.

One must keep in mind the different task paradigms used with the CH and VS observers. With the CH observer, anatomical noise is equated with background variability as quantified by the image class statistics. Adding internal noise terms to the channel covariance matrices is a common means of reconciling

differences with human observers. One may also inject greater anatomical noise into the model observer by modifying the reference image  $c_j$  [Eq. (6)], although comparison of the CNPW-BKE and CNPW-BKS results suggests this approach primarily affects observer performance with the higher-resolution collimators. Previous studies have also shown that excessive modification of the reference image can lead to artifacts that create persistent false positives for the observer.<sup>3</sup>

For the VS observer tested herein, anatomical noise only affects the initial candidate search since the analysis is a BKE process. Furthermore, the separate influences of quantum and anatomical noise are not discerned in the search. Instead, image texture as a whole impedes the identification of lesions as focal points. Accounting for noise in this manner effectively expanded the dynamic range in observer performance across the full set of collimators compared to the scanning models. Translating this expansion into greater statistical power will depend on developing methods of candidate analysis that are independent of the BKE and BKS paradigms. The use of greyscale intensity as the sole search feature for the VS observers in our study was only possible because of the BKE analysis. One promising alternative is based on the extraction of multiple morphological features.<sup>22</sup>

Overall, the VS framework provides a relatively flexible environment for studying observer inefficiencies. As shown in Fig. 5, different inefficiencies can influence different parts of parameter space, offering an intuition that is not always available with covariance modeling of internal noise. The VS observer attained quantitative agreement with the human observers, but this was the result of ad hoc fitting of the internal noise and search threshold parameters ( $\gamma$  and  $\beta$ , respectively). A version of the VS observer that can be reliably applied without the BKE or BKS background paradigms may require smaller amounts of internal noise and other inefficiencies to approximate human-observer performance, and discussion of how to set the inefficiency and noise parameters should await development of such a model.

Computational efficiency is another important point of comparison for the CH and VS observers. Runtimes for the CH observer were substantially longer than for the VS observer. Preinverting the covariance matrices would have improved efficiency, but the main factor in the longer times was the shift-variant CH observer template. With regard to the amount of preprocessing required for the CH observer, one must recognize that the background variations in this study were restricted to modeling different In-111 biodistributions in a single geometry of the XCAT phantom, so that the computation of  $\mathbf{K}_j^{\text{anat}}$  [Eq. (10)] and  $\bar{\mathbf{b}}$  was relatively simple. A more extensive study involving multiple phantom geometries could complicate the BKS computations by involving intensive Monte Carlo simulations as demonstrated for nonsearch tasks by Kupinski et al.<sup>23</sup> Observer studies with reconstructed images would substantially increase the computing time,<sup>24</sup> as would studies involving larger regions of interest, high-resolution imaging or image volumes.

Finally, there were a number of basic limitations with this study for purposes of collimator optimization, including the use of an analytic projector that disregarded the considerable scatter and penetration potential of In-111. Still, the results are relevant to the practice of task-based collimator optimization. The accepted approach to hardware optimization is to apply ideal observers in projection space; the scanning CH observer has been used previously as an approximate ideal

observer (although with channels other than what was used herein<sup>25</sup>). VS observer development has focused instead on modeling human observers as the gold standard for lesion-detection performance in much of diagnostic imaging. The planar imaging framework allows for an even comparison of the two approaches to optimization. Follow-up studies will analyze the prospects for optimization studies based on increasingly realistic detection tasks with the VS observer.

## 6 Conclusions

When applied without inefficiencies, the VS and CH observers failed to match either the magnitudes or the general trend in average human performance as anatomical noise increased in the images. The result for the VS observer is partly attributed to reliance on the quasi-BKE task, which artificially inflated observer performance in images with high anatomical noise. Overall, the VS model observer offers a flexible and computationally efficient framework for studying the combined effects of quantum and anatomical noise in lesion search tasks. Future studies will examine versions of the model that are independent of the standard BKE/BKS task paradigms.

## Acknowledgments

This research was supported by NIBIB Grant No. R01EB012070. The authors acknowledge the contributions of Nagarohit Katta in the development of the computer interface for the human-observer study and of Kheya Banerjee for participating in the study.

## References

1. H. H. Barrett et al., "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9758–9765 (1993).
2. H. C. Gifford, P. H. Pretorius, and M. A. King, "Comparison of human and model-observer LROC studies," *Proc SPIE* **5034**, 112–122 (2003).
3. H. C. Gifford, "A visual-search model observer for multislice-multiview SPECT images," *Med. Phys.* **40**, 092505 (2013).
4. H. C. Gifford, "Efficient visual-search model observers for PET," *Br. J. Radiol.* **87**, 20140017 (2014).
5. H. L. Kundel et al., "Holistic component of image perception in mammogram interpretation: gaze-tracking study," *Radiology* **242**, 396–402 (2007).
6. Z. L. Lu and B. A. Doshier, "Characterizing human perceptual inefficiencies with equivalent internal noise," *J. Opt. Soc. Am. A* **16**, 764–778 (1999).
7. A. Sen, F. Kalantari, and H. C. Gifford, "Impact of anatomical noise on model observers for prostate SPECT," in *IEEE NSS-MIC Record* (2014).
8. Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observer models," *Med. Phys.* **34**, 3312–3322 (2007).
9. L. Zhang et al., "Channelized model observer for the detection and estimation of signals with unknown amplitude, orientation, and size," *J. Opt. Soc. Am. A* **30**, 2422–2432 (2013).
10. M. K. Whitaker, E. Clarkson, and H. H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods," *Opt. Express* **16**, 8150–8173 (2008).
11. A. Sen, F. Kalantari, and H. C. Gifford, "Assessment of prostate cancer detection with a visual-search human model observer," *Proc SPIE* **9037**, 90370Q (2014).
12. C. K. Abbey and H. H. Barrett, "Human and model-observer performance in ramp spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A* **18**, 473–488 (2001).
13. J. B. Roerdink and A. Meijster, "The watershed transform: definitions, algorithms and parallelization strategies," *Fundamenta Informaticae* **41**, 187–228 (2001).

14. J. De Bock, P. De Smet, and W. Philips, "A fast sequential rainfalling watershed segmentation algorithm," *Lect. Notes Comput. Sci.* **3708**, 476–482 (2005).
15. W. P. Segars and B. M. W. Tsui, "MCAT to XCAT: the evolution of 4D computerized phantoms for imaging research," *Proc. IEEE* **97**, 1954–1968 (2009).
16. S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in Nuclear Medicine*, 4th ed., W.B. Saunders, Philadelphia (2012).
17. M. T. Madsen, "Recent advances in SPECT imaging," *J. Nucl. Med.* **48**, 661–673 (2007).
18. R. G. Wells et al., "Effect of filtering on the detection and localization of small Ga-67 lesions in thoracic single photon emission computed tomography images," *Med. Phys.* **26**, 1382–1388 (1999).
19. J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology* **143**, 29–36 (1982).
20. L. L. Tang and N. Balakrishnan, "A random-sum Wilcoxon statistic and its application to analysis of ROC and LROC data," *J. Stat. Plann. Inference* **141**, 335–344 (2011).
21. D. Montgomery, *Design and Analysis of Experiments*, 8th ed., Wiley, Hoboken, New Jersey (2013).
22. H. C. Gifford, A. Sen, and R. Azencott, "SVM-based visual-search model observers for PET tumor detection," *Proc SPIE* **9416**, 94160X (2015).
23. M. A. Kupinski et al., "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," *J. Opt. Soc. Am. A* **20**, 430–438 (2003).
24. X. He and E. Frey, "Towards realistic and practical ideal observer estimation for optimization of medical imaging systems," *IEEE Trans. Med. Imaging* **27**, 1535–1543 (2008).
25. C. J. Lee, M. A. Kupinski, and L. Volokh, "Assessment of cardiac single-photon emission computed tomography performance using a scanning linear observer," *Med. Phys.* **40**, 011906 (2013).

**Anando Sen** completed his PhD from the Department of Mathematics at the University of Houston in 2012. He is currently a postdoctoral research scientist in the Department of Biomedical Informatics at Columbia University and was previously a postdoctoral fellow in the Department of Biomedical Engineering at the University of Houston. His research interest include image quality assessment, model observers and tomographic reconstruction. He has presented his research at several national and international conferences.

**Howard C. Gifford** received his PhD in applied mathematics from the University of Arizona in 1997. He was a postdoctoral fellow in the Radiology Department at the University of Massachusetts Medical School (UMMS) until 2000 and a UMMS faculty member through 2011. He is an associate professor in the Department of Biomedical Engineering at the University of Houston. His research interests include task-based image quality assessment and optimization methods for nuclear medicine and x-ray imaging.