Published in Phys. Med. Biol. 51 (2006) 6225-6244

# Small nodule detectability evaluation using a generalized scan statistic model

**Lucreţiu M. Popescu and Robert M. Lewitt**

University of Pennsylvania, Department of Radiology
423 Guardian Drive, 4-th floor Blockley Hall, Philadelphia, PA 19104-6021, USA

E-mail: popescu@mipg.upenn.edu

**Abstract.** In this paper is investigated the use of the scan statistic for evaluating the detectability of small nodules in medical images. The scan statistic method is often used in applications in which random fields must be searched for abnormal local features. Several results of the detection with localization theory are reviewed and a generalization is presented using the noise nodule distribution obtained by scanning arbitrary areas. One benefit of the noise nodule model is that it enables determination of the scan statistic distribution by using only a few image samples in a way suitable both for simulation or experimental setups. Also, based on the noise nodule model, the case of multiple targets per image is addressed and an image abnormality test using the likelihood ratio and an alternative test using multiple decision thresholds are derived. The results obtained reveal that in the case of low contrast nodules or multiple nodules the usual test strategy based on a single decision threshold underperforms compared with the alternative tests. That is a consequence of the fact that not only the contrast or the size, but also the number of suspicious nodules is a clue indicating the image abnormality. In the case of the likelihood ratio test, the multiple clues are unified in a single decision variable. Other tests that process multiple clues differently do not necessarily produce a unique ROC curve, as shown in examples using a test involving two decision thresholds. We present examples with two-dimensional time-of-flight (TOF) and non-TOF PET image sets analyzed using the scan statistic for different search areas, as well as the fixed position observer.

## 1. Introduction

Radiological diagnostic tasks often require the detection of features (also called targets or signals) consisting of small regions with higher than usual activity concentration. Due to the data noise and the limited detector resolution these features appear in the reconstructed images as small bumps of higher concentration than the surrounding background. Such features sometimes could be easily mistaken as arbitrary noise clumps that normally occur in the image background. Therefore distinct detection criteria have to be considered, depending on whether the feature position is a priori exactly known, or the feature must be searched for in a given image area (or volume).

In order to detect such small features we can scan the image by smoothly sliding a scanning window over the area/volume of interest, and for each window position compute some relevant local statistic of the image elements inside the scanning window. For instance the scanning window can be a disk (or a ball) of a fixed radius $R$ and the statistic can be simply the sum of the values of the pixels inside the disk. Or, in order to eliminate the influence of background nonuniformity, one can use an enlarged scanning window and use as the statistic the local contrast given by the ratio between the average image values inside the disk and the average over an exterior annulus. More complex scanning windows and statistics can be devised, where these can account for the searched features size variation, or their rotation asymmetry, or their known typical activity profiles. However, when the searched feature is small, due to the limited system resolution its shape and even its original size become less relevant, therefore using a scanning window with a fixed radius seems a reasonable approach.

If we expect to find only one feature per searched area, or we are interested only in the most suspicious one, a detection strategy would consist of scanning the interest area and selecting only the location with the maximum scan value as suspicious. Further in the text this value will be referred to as the maximum scan (or max-scan) result.

By determining the distribution $g(c)$ of such maximum scan returns $c$ for empty background cases (images with no features present) it can be assessed how usual or unusual a suspicious finding can be. Such tests are often called scan statistics tests and they are part of a statistics chapter actively studied (Parzen 1960, Naus 1965, Adler 2000, Glaz et al. 2001). These methods are preferred in applications when random fields must be searched for abnormally occurring local features. Scan statistic techniques are commonly used in biostatistics and epidemiology (Wallenstein 1980, Kulldorff 1997) and more recently have been also applied in astronomy (Orford 2000) and high energy physics (Terranova 2004).

In the medical image analysis literature the scan statistic is explicitly mentioned in only a few recent papers. However, evaluation techniques in the spirit of the scan statistic tests, or involving search models have a longer history in medical imaging. Kundel (1981) has performed detectability evaluations by determining the distribution of noise nodules and the distribution of target nodules in chest radiographs. Swensson (1996, 2001) has used a maximum search model as the basis for his localization receiver operating characteristic

(LROC) and non-degenerate ROC theory. Some of these results will be summarized below. Gifford et al. (2003, 2005) have used the Swensson model to analyze human and numerical observer image rating data. The application of the scan statistic method in medical image analysis is explicitly exemplified in (Naiman & Priebe 2001). A method to rapidly determine the LROC curves directly from the image reconstruction properties was proposed in (Khurd & Gindi 2005). A search model, related to our own result presented in this paper, has been recently published by Chakraborty (2006a, 2006b), which is also related to the model used in (Edwards et al. 2002). The relation between our approach and these works will be discussed in the conclusion section 7. An ideal observer that operates by using directly the likelihoods of the acquired data and not of the reconstructed images, is studied in (Park et al. 2005, Park et al. 2003) in conditions of signal localization uncertainty as well as background uncertainty.

The difficulty of the theoretical treatment of the problems involving the detection of small localized signals in noisy images (or more generally spatial extended random fields) stems from the marked difference in scale of the two entities to be compared. While the signal consists of only a few random variables and can be represented with even fewer parameters, the image is a very large lattice of random variables, that in many applications are not even independent of each other. If the problem is reduced to the detection of a signal at a known fixed position this difficulty is avoided, since from the whole image only a portion commensurate with the signal remains of interest and the classical signal detection formalism can be straightforwardly applied. Based on this fixed position approach many signal to noise ratio (SNR) (or detectability index) image quality metrics were proposed (Wagner & Brown 1985). In these methods in order to distinguish the signal from the background one can go to great lengths and apply a careful analysis by taking into account every detail of the shapes of the signal and of the restricted background region. For example we have the ideal observer when both the signal and the background are exactly known (from the statistical point of view) and the Hotelling observer with the signal and the background estimated statistically (from samples) (Barrett et al. 1993). However, if the signal is rather poor in details (a bump of activity determined more by the point spread function of the imaging system than the anatomical detail) and the fixed position restriction is removed, the differences due to the meticulous analysis of the signal and the background confined in a narrow region will be shadowed by the variability with the position of the background response. Therefore, in the scan statistic methodology the opposite approach is followed, and rather than comparing in full detail the differences between the signal and the background for a fixed position using a complicated statistic, a simplified statistic calculation procedure is adopted and its variation for the full spatial extent of the image background is determined.

The Hotelling observer mentioned above is based on the generalization to multidimensional random variables of Student's ratio (Hotelling 1931). However, there is another paper by Hotelling (1939) which proved more fruitful for the theoretical treatment of a whole class of problems involving the detection of a signal at an unknown location by means of finding the distribution of the maximum value of a stationary random field

(Knowles & Siegmund 1989, Johansen & Johnstone 1990, Adler 2000). In this class of approaches we have results developed and applied to functional MRI and PET brain imaging in (Siegmund & Worsley 1995, Worsley 1995, Worsley et al. 1996). Recently Yendiki and Fessler (2006b, 2005) have used the approach of Worsley (1996) for medical image quality analysis.

In this paper, instead of attempting to theoretically determine the scan statistic distribution from the random field properties as in the papers cited above, we present a scan statistic model based on certain general properties and experimental observations that are relatively easy to obtain. The model predicts the max-scan distribution for a given image area and provides a framework for treating the multiple signals cases.

## 2. Detection tests

If the search features are of known size we can determine the scan response distribution $f(c)$ from multiple realizations, or from the the image reconstruction algorithm behavior, if such theory is available. With both distributions $f$ and $g$ known we can study the properties of the detection tests and plot the ROC type curves.

We denote the cumulative distributions of $f$ and $g$ as $F(c) = \int_{c_0}^{c} f(c')\mathrm{d}c'$, $G(c) = \int_{c_0}^{c} g(c')\mathrm{d}c'$, where $c_0$ is the lower limit of the scan response value $c$.

### *2.1. Feature detection test*

If the scan procedure returns a value $c$ greater than a certain threshold value $d$ we declare the location corresponding to the value $c$ as a positive detection result, otherwise we declare that no feature/target/signal is detected.

If no feature is present the probability of the scan procedure to return a false positive result, that is a result $c$ greater than the threshold $d$, is

$$P_0(d) = \mathrm{Prob}(c > d) = \int_d^\infty g(c)\mathrm{d}c = 1 - G(d) \tag{1}$$

In the case when one feature is present, we have two random variables: $a$ the feature realization value, and $b$ the maximum scan obtained by searching the background. The scan will return the value $c = \max(a, b)$. The probability density that the scan procedure returns the value $c = a$ corresponding to the true target location (a true positive) is

$$h_{1\mathrm{L}}(c) = f(a = c) \cdot \mathrm{Prob}(b < c) = f(c)G(c)\,. \tag{2}$$

And the probability of detecting a true target location, that is $c > d$, is

$$P_{1\mathrm{L}}(d) = \int_d^\infty h_{1\mathrm{L}}(c)\mathrm{d}c = \int_d^\infty f(c)G(c)\mathrm{d}c\,. \tag{3}$$

It is assumed that the presence of a feature has negligible effect on the background and vice versa, which is a reasonable assumption if the feature is small and the background area is large compared to the feature size.

### *2.2. Image abnormality test*

In many image evaluation tests one is interested only in whether or not the image is abnormal without requiring the correct localization of the suspicious features. The test is: if the returned max-scan value $c$ exceeds the threshold $d$ then the image is abnormal (positive), otherwise we declare it normal (negative).

In this case the probability for a false positive result is the same as in previous case given by (1). While if a feature is present the probability density of the returned scan values is

$$\begin{aligned} h_1(c) &= f(a=c)\cdot \mathrm{Prob}(b<c) + g(b=c)\cdot \mathrm{Prob}(a<c) \\ &= f(c)G(c) + g(c)F(c) \end{aligned} \tag{4}$$

The probability of a true positive result (correct localization not required) is

$$\begin{aligned} P_1(d) &= \int_d^\infty h_1(c)\mathrm{d}c = \int_d^\infty f(c)G(c)\mathrm{d}c + \int_d^\infty g(c)F(c)\mathrm{d}c \\ &= 1 - F(d)G(d)\,. \end{aligned} \tag{5}$$

If $P_1(c)$ is plotted against $P_0(c)$ we have the receiver (or relative) operating characteristic (ROC) curve, while if $P_{1\mathrm{L}}(c)$ is plotted against $P_0(c)$ we obtain the localization receiver operating characteristic (LROC) curve.

### *2.3. Fixed position feature detection test*

In the situations when the feature position is known a priori, only limited or no search is required. That is often the case when the task is to confirm a previous finding or a finding obtained with a different imaging modality. We denote with $g_0(c)$ the background scan distribution for this case (and with $G_0(c)$ its cumulative distribution).

For the fixed position (or very small area search) detection we have the classical hypothesis test case. We have to determine whether the scan return random variable $c$ is due to the distribution $f(c)$ (the feature contrast variation) or $g_0(c)$ (normal background variation).

The probabilities for false positive and true positive results respectively, are

$$P_{0,\mathrm{fix}}(d) = \int_d^\infty g_0(c)\mathrm{d}c = 1 - G_0(d)\,, \tag{6}$$

$$P_{1,\mathrm{fix}}(d) = \int_d^\infty f(c)\mathrm{d}c = 1 - F(d)\,. \tag{7}$$

If $P_{1,\mathrm{fix}}(c)$ is plotted against $P_{0,\mathrm{fix}}(c)$ we have the particular case of the ROC curve for a priori known location case.

## 3. Generalization: noise nodule distribution model

In the scan procedure, instead of picking only the maximum contrast noise nodule from a scanned region of a given area, we can obtain a list of all noise nodules with values greater

than a certain limit $c_0$. We assume that such noise nodules are a relatively rare occurrence and they are independent of each other.

Let us assume that after scanning image regions of a total area $A_t$ we gather a list with a total $N_t$ noise nodules. The average density of noise nodules with $c > c_0$ is $n_0 = N_t/A_t$.

By histogramming (or using other density estimation techniques) we can obtain the distribution $p(c)$ of the noise nodules contrast (for $c > c_0$) in the scanned image set. We have $\int_{c_0}^{\infty} p(c)\mathrm{d}c = 1$.

For a region of a given size $A$, that may be different from the size of the scanned regions, we have $N = n_0 A$ for the average number of noise nodules with $c > c_0$.

The average number of noise nodules with contrast larger than $c$ (with $c > c_0$) in the area $A$ is $\nu(c) = Nq(c)$, where $q(c) = \int_c^{\infty} p(c')\mathrm{d}c'$. Since the number $k$ of noise nodules in a given image area is a random variable occurring at a constant rate, it is natural to assume (as in (Bunch et al. 1978)) that it follows the Poisson distribution

$$P(k;c) = \frac{\nu^k}{k!}\mathrm{e}^{-\nu}\,. \tag{8}$$

This is mainly valid for large enough areas $A$. Alternatively and with similar results the analysis can be based on the binomial distribution $P_{\mathrm{B}}(k;c) = \binom{N}{k}[1-q(c)]^{N-k}q^k(c)$ that converges to the Poisson distribution for large $N$ values.

The probability of having at least one noise nodule with contrast greater than $c$ is

$$Q(1;c) = 1 - P(0;c) = 1 - \mathrm{e}^{-\nu} \tag{9}$$

The likelihood of having $k$ nodules with contrast $c_1, \ldots, c_k$ greater than $c$ is

$$L(c_1, \ldots, c_k) = P(k;c)\prod_{j=1}^{k}\frac{p(c_j)}{q(c)} = \frac{N^k}{k!}\mathrm{e}^{-Nq(c)}\prod_{j=1}^{k}p(c_j)\,. \tag{10}$$

We can derive the max-scan distribution $g(c)$ for a given search area $A$, from the noise nodule distribution $p(c)$. The probability of obtaining a max-scan result with the value $c$ is given by the probability density of having one noise nodule with contrast $c$ and all the other noise nodules with contrast less than $c$.

$$g(c) = P(1;c)\frac{p(c)}{q(c)} = Np(c)\mathrm{e}^{-Nq(c)} \tag{11}$$

The cumulative distribution is

$$G(c) = \mathrm{e}^{-Nq(c)} \tag{12}$$

We assumed that the searched area $A$ is large enough so that $\mathrm{e}^{-N} \approx 0$.

Since the max-scan results $c_i$ obtained from $m$ distinct subregions each of area $A$ are independent, then the max-scan result for the total region of size $mA$ is $c = \max\{c_j | 1 \leq j \leq m\}$. If $g(c, A)$ is the scan statistic distribution for area size $A$ and $G(c, A)$ is the cumulative distribution, then the cumulative distribution of the scan statistic for area $mA$ is

$$G(c, mA) = G^m(c, A)\,. \tag{13}$$

and we have

$$g(c, mA) = mG^{m-1}(c, A)g(c, A) \tag{14}$$

The exponential (12) satisfies the property (13) in a straightforward manner.

## 4. The case of multiple features

In (Swensson 1996) the target localization model is extended to the case of multiple target images. Instead of picking the maximum suspicious feature from an image, the test can pick as suspicious all nodules with contrast greater than a certain decision threshold $c > d$. The analysis of the results can be performed using the free response receiver operating characteristic (FROC) graphs or the alternative free response (AFROC) graphs (Bunch et al. 1978, Chakraborty 1989, Chakraborty & Winter 1990).

In the FROC diagram are plotted the probability of detecting a true feature $P_{1,\text{fix}}(c) = 1 - F(c)$ against the mean of false-positive reports per image $\nu(c) = Nq(c)$. In the AFROC diagram are plotted probability of detecting a true feature $P_{1,\text{fix}}(c) = 1 - F(c)$ against the probability of reporting at least one false positive feature per image $P_0(c) = 1 - G(c)$.

All ROC, LROC, FROC and AFROC curves change if the search area varies, because of the variation of the $g(c)$ distribution with the image area. This fact makes it difficult to compare different data sets (or even different observers). However, if in the FROC diagram we plot the $P_{1,\text{fix}}(c)$ against the average number of false positive reports per image area $n(c) = \nu(c)/A = n_0 q(c)$, we obtain search area independent diagrams.

An image can be abnormal due to the presence of high contrast suspicious features, but it also can be abnormal due to the presence of an unusual number of features that otherwise taken individually would not look so suspicious. The probabilities of having at least $k$ nodules with contrast greater than $c$ are given by the equation

$$Q(k; c) = 1 - \sum_{i=0}^{k-1} P(i; c) \tag{15}$$

for $k > 0$ with $P(k; c)$ given by equation (8), and $Q(0; c) = 1$. In Figure 1 are plotted the $Q(k; c)$ curves for several $k$ values. We can see that having two or more nodules with contrast greater than $d_2$ is as unlikely as having only one nodule with contrast greater than $d_1$, while a single nodule with contrast greater than $d_2$ is well above the threshold that would qualify it as suspicious. The same can be observed for the groups of three or more nodules. In general, while the realization of $k$ nodules is unlikely the presence of $k - 1$ nodules is common, hence one nodule among these should be a true feature. In the case when all $k$ nodules have similar contrast the localization of the suspicious feature is ambiguous. In other words, in such situations we can declare with a certain confidence that the image is abnormal, but we cannot confidently indicate the localization of the suspicious feature. The best one can do is to give a list of the suspected locations and a probability score.

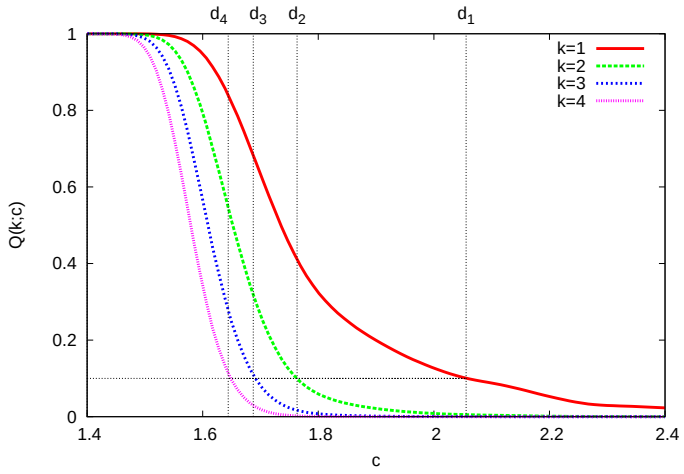


**Figure 1.** The realization probability of $k$ or more noise nodules with contrast greater than $c$ for several $k$ values.

### *4.1. The likelihood ratio test*

The likelihood of observing a set of $k$ nodules $\underline{c_k} \equiv \{c_1, \ldots, c_k\}$ with contrast larger than $c_0$, $l$ of them being true features is

$$L_k^l(\underline{c_k}) = P(k-l;c_0)\binom{k}{l}^{-1}\sum_{\mathcal{I}_k^l}\prod_{i\in\mathcal{I}_k^l} f(c_i)\prod_{j\in\mathcal{I}_k^k-\mathcal{I}_k^l} p(c_j) \tag{16}$$

where the $\mathcal{I}_k^l$ is a particular combination of $l$ indices $\{i_1, \ldots, i_l\}$ taken from the set of all $k$ indices $\{1, \ldots, k\}$. The sum in the above equation iterates over all $\binom{k}{l}$ possible combinations of this kind.

If the number of true features $l$ present in a positive image varies according to the distribution $\rho(l)$ then the likelihood of having a positive image with one or more features present is

$$L_k(\underline{c_k}) = \sum_l \rho(l) L_k^l(\underline{c_k})\,. \tag{17}$$

The likelihood ratio between the case when we have at least one true feature and the case with no true features present when a total of $k$ nodules are observed is

$$\lambda_k(\underline{c_k}) = \frac{L_k(\underline{c_k})}{L_k^0(\underline{c_k})} = \sum_l \rho(l)\lambda_k^l(\underline{c_k})\,, \tag{18}$$

where

$$\lambda_k^l(\underline{c_k}) = \frac{L_k^l(\underline{c_k})}{L_k^0(\underline{c_k})} = \frac{P(k-l;c_0)}{P(k;c_0)}\binom{k}{l}^{-1}\sum_{\mathcal{I}_k^l}\prod_{i\in\mathcal{I}_k^l}\frac{f(c_i)}{p(c_i)} \tag{19}$$

We can set a threshold $\lambda_d$ and use the following decision procedure

- If $\lambda_k(\underline{c_k}) \geq \lambda_d$ then we report a positive image (at least one target present).

- If $\lambda_k(\underline{c_k}) < \lambda_d$ then we report a negative image (no target present).

The probability of wrongly reporting an image as positive (false positive) is

$$P_0(\lambda_d) = \sum_k \int\limits_{\lambda_k(\underline{c_k}) \geq \lambda_d} L_k^0(\underline{c_k}) \mathrm{d}\underline{c_k} \tag{20}$$

The probability of correctly reporting a positive image (true positive) is

$$P_1(\lambda_d) = \sum_k \int\limits_{\lambda_k(\underline{c_k}) \geq \lambda_d} L_k(\underline{c_k}) \mathrm{d}\underline{c_k} \tag{21}$$

### *4.2. The multiple decision thresholds test*

A simpler test that does not require complicated mathematical calculations can be the following. Let $c_1 \geq c_2 \geq \ldots \geq c_m$ be the first $m$ nodules in the decreasing order of their contrast. We set a series of decision thresholds $d_1 > d_2 > \ldots > d_m$, and if $c_1 \geq d_1$ then the image is positive, otherwise if $c_2 \geq d_2$ the image is declared also positive and so on until if $c_m < d_m$ the image is negative.

We will consider here the case with only two decision thresholds $d_1, d_2$. The test is positive if there is one or more nodules with contrast greater than $d_1$ or two or more nodules with contrast greater than $d_2$, otherwise the test is negative.

By extending the notations from equations (8) and (15), we have the probability of obtaining $k$ noise nodules with contrast in the interval $(d_1, d_2)$ given by $P(k; d_2, d_1) = \frac{\nu_{21}^k}{k!} \mathrm{e}^{-\nu_{21}}$, with $\nu_{21} = N(q(d_2) - q(d_1))$, and the probability of having at least $k$ noise nodules with contrast in the same interval given by $Q(k; d_2, d_1) = 1 - \sum_{i=1}^{k-1} P(i; d_2, d_1)$.

The probability of a false positive result is

$$P_0(d_1, d_2) = Q(1; d_1) + [1 - Q(1; d_1)] Q(2; d_2, d_1) \tag{22}$$

In the cases when the image has one, two or three features, respectively, the probabilities for a true positive result are

$$\begin{aligned} P_{1,1}(d_1, d_2) = {} & Q(1; d_1) + [1 - Q(1; d_1)] \{1 - F(d_1) + \\ & [F(d_1) - F(d_2)] Q(1; d_2, d_1) + F(d_2) Q(2; d_2, d_1)\} \end{aligned} \tag{23}$$

$$\begin{aligned} P_{1,2}(d_1, d_2) = {} & Q(1; d_1) + [1 - Q(1; d_1)] \big\{1 - F^2(d_1) + \\ & [F(d_1) - F(d_2)]^2 + 2[F(d_1) - F(d_2)] F(d_2) Q(1; d_2, d_1) + \\ & F^2(d_2) Q(2; d_2, d_1)\big\} \end{aligned} \tag{24}$$

$$\begin{aligned} P_{1,3}(d_1, d_2) = {} & Q(1; d_1) + [1 - Q(1; d_1)] \big\{1 - F^3(d_1) + \\ & [F(d_1) - F(d_2)]^3 + 3[F(d_1) - F(d_2)]^2 F(d_2) + \\ & 3[F(d_1) - F(d_2)] F^2(d_2) Q(1; d_2, d_1) + F^3(d_2) Q(2; d_2, d_1)\big\} \end{aligned} \tag{25}$$

For comparison, when $m$ features are present the single decision level test has the following probability for reporting a true positive result

$$P_{1,m}(d_1) = 1 - F^m(d_1) G(d_1) = Q(1; d_1) + [1 - Q(1; d_1)][1 - F^m(d_1)]\,. \tag{26}$$

## 5. Simulation study

We have studied the above feature detection approaches for two dimensional PET image reconstruction configurations. That enabled us to conveniently produce large numbers of image reconstruction (100-400) realizations and determine the distributions $f$, $g_0$, and $g$ for different search area sizes by histogramming the results obtained by analyzing this large pool of images.

As phantoms we have considered activity distributions in the shape of disks of radius $R_p = 16\,\mathrm{cm}$. We have obtained hot feature phantoms by placing on the circle of radius $R_f = R_p/2$ a set of uniform circular features of radius $R = 0.5\,\mathrm{cm}$ and contrast 3:1 relative to the uniform background of the surrounding disk. The features were placed symmetrically at distances of minimum 6 cm one from another. Background only phantoms were obtained by omitting the hot features. In this study we have not considered any attenuation, scatter or randoms.

The data were generated (Popescu & Lewitt 2003) in list-mode with precise event positioning information. This fact enabled us, by randomly altering the longitudinal position information, to consider time-of-flight (TOF) with various timing precisions, as well as non-TOF image reconstruction situations.

The image reconstructions were performed using a list-mode ML-EM algorithm described in (Popescu et al. 2004), with images represented using blobs — smooth overlapping basis functions (Lewitt 1992) — on a rectangular grid (grid spacing $\Delta x = 0.4\,\mathrm{cm}$, Kaiser-Bessel blobs with radius $r_b = 2.4\Delta x$, and with parameters $m = 2$ and $\alpha = 6$). We used iterations over subsets made from consecutive events in the list. Further in the text a complete pass through the data will be referred to as an iteration. Multiple image realizations were obtained by using distinct data sets containing $6 \cdot 10^4$, $12 \cdot 10^4$ and $18 \cdot 10^4$ counts, split in 12 equal consecutive subsets, and using a relaxation parameter $\lambda = 0.8$ applied as an exponent to the update quantity (Popescu et al. 2004).

For image analysis we have considered as statistic the local contrast $c$ given by the ratio between the image average on a disk of radius $R_a = 0.5\,\mathrm{cm}$ and the image average on the annulus of inner and outer radii $R_{b1} = 0.6\,\mathrm{cm}$ and $R_{b2} = 2.0\,\mathrm{cm}$ concentric with the disk.

We have applied this local contrast calculation procedure on the hot features and background images in order to obtain the following results:

- The hot feature realizations contrast distribution $f(c)$;
- The max-scan distributions $g(c, A)$, obtained by scanning regions of area $A$ of the background-only images. We have considered an initial area $A_0$ in the form of a square of size $22 \times 22\,\mathrm{cm}^2$ and then we took regions of area size $A_0$ divided by 2, 4, 8, 16 and 32 respectively.
- The distribution $p(c)$ and the density $n_0$ of the noise nodules with $c \geq c_0$.
- The background variation $g_0(c)$ for a fixed position.

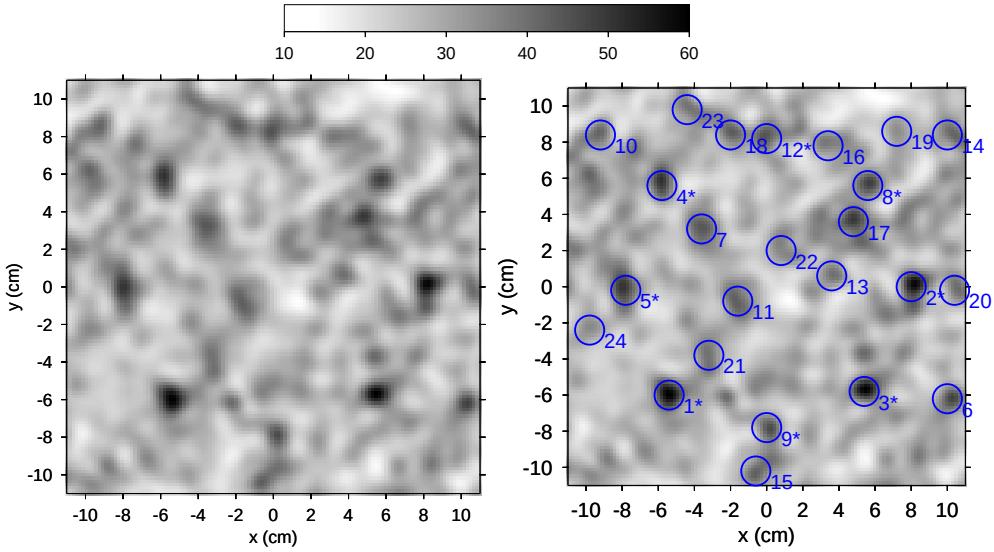


**Figure 2.** Example of image scan (image reconstructed from $4 \cdot 10^4$ counts, non-TOF, iteration 3). On the left side is the original unmarked image. On the right side the nodules returned by the scan procedure are marked and numbered in decreasing order of their contrast. All nodules with contrast value greater than $c_0 = 1.2$ are returned. The eight hot features present in the original phantom are marked with a '*' symbol and are returned with the following order numbers: 1, 2, 3, 4, 5, 8, 9, 12.

The scanning procedure comprises two steps. The first step is analogous with a filtering procedure and consists of computing an auxiliary scan image in which each point has the value of the scan result obtained with the scanning window centered on that point. In the second step we start by determining the maximum point of the scan image. Once this point is found the value is entered in a list and the pixels in the disk of radius $R_a$ corresponding to the nodule at that position are masked. The procedure continues with the rest of the unmasked image pixels as long as the maximum values found exceed the lower limit $c_0$. In this manner a list is produced with the non-overlapping nodules in the decreasing order of their contrast values. An example is shown in Figure 2.

## 6. Results

### *6.1. Nodule detectability evaluation using scan statistic*

In Figure 3 we show comparisons between the local contrast distribution for the background at fixed positions $g_0(c)$, the background max-scan (or scan statistic) distribution obtained for several search area sizes $g(c, A)$, and the true feature contrast distribution $f(c)$. The plotted distributions have been obtained in a smooth form over a refined grid by using the density kernel estimation technique with the kernel $w(\xi) = (1 - \xi^2)^2$, where $\xi = (c - c')/h$ with a window width $h = 0.08$. In Figure 4 are compared the ROC curves for the feature detection tests for the fixed location test and max scan test for several search area sizes.

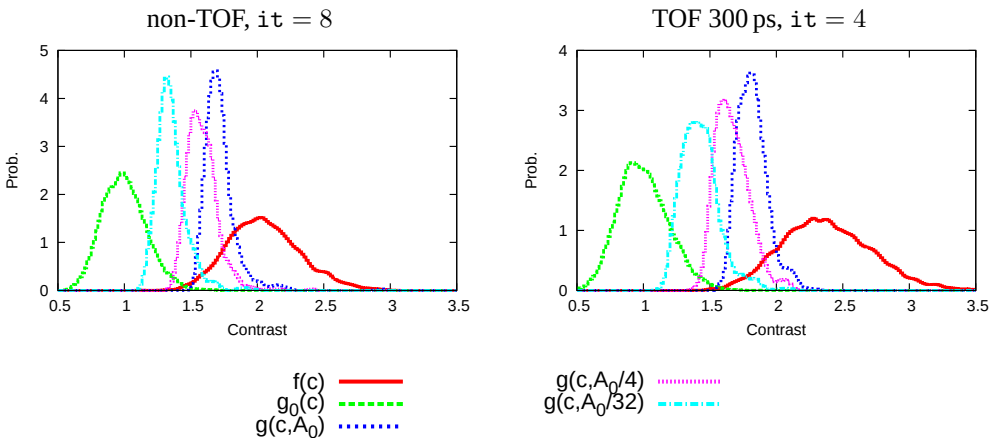


**Figure 3.** Comparison between the local contrast distribution for the background at fixed positions ($g_0(c)$), background max-scan for several search area sizes ($g(c, A), A = A_0, A_0/4, A_0/32$) and the feature realization ($f(c)$). Reconstructions using $N = 6 \cdot 10^4$ counts, for non-TOF case (left) and TOF case for 300 ps timing resolution (right).

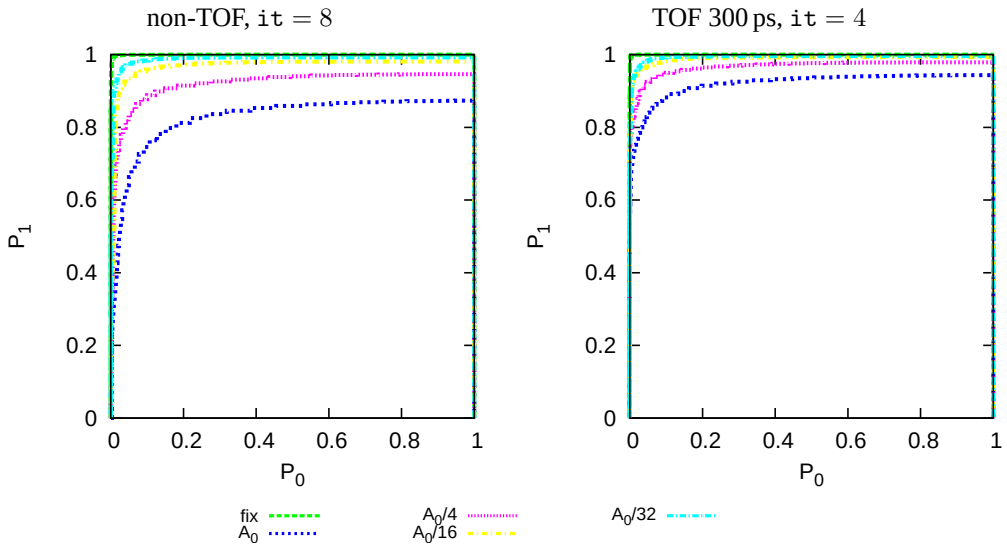


**Figure 4.** Comparison between the LROC curves for the fixed position case versus background max-scan for several search area sizes $A$ (same cases as in Figure 3).

In Figure 5 we compare the max-scan distributions for search area size $A_0$, as well as for the background fixed position variation, and feature contrast variation for the non-TOF and the 300 ps TOF case. We show results obtained at iteration 8 in the case of non-TOF, and iteration 4 in the TOF case, results that are close to the convergence point of each case, respectively.

We can notice that there is little difference between the fixed position background variation curves for TOF and non-TOF. Some improvement can only be seen in the high count case $N = 18 \cdot 10^4$. The background max-scan distributions show that for the low counts

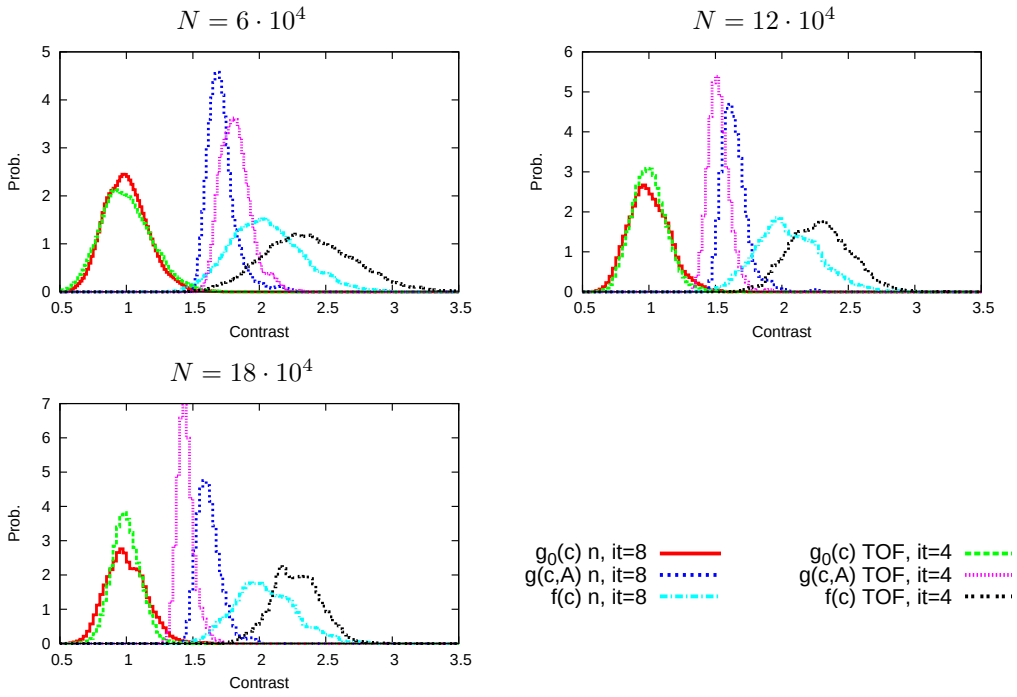


**Figure 5.** Comparison between the TOF (300 ps) and non TOF local contrast distribution for the background at fixed positions ($g_0(c)$ n, $g_0(c)$ TOF), max-scan for a search area of size $A = A_0$ ($g(c, A)$ n, $g(c, A)$ TOF), and the feature contrast variation ($f(c)$ n, $f(c)$ TOF). We show here results for $N = 6 \cdot 10^4$, $12 \cdot 10^4$, and $18 \cdot 10^4$ counts.

case $N = 6 \cdot 10^4$ the TOF reconstructions produce noisier backgrounds than the non-TOF. However, as the number of counts increases to $N = 12 \cdot 10^4$ and $N = 18 \cdot 10^4$ this situation is reversed, and the TOF reconstruction tends to produce noise nodules with smaller contrast than the non-TOF case. Also the TOF reconstructions systematically produce higher contrast features, leading to the difference in detectability, as shown by the LROC curves in Figure 6.

### *6.2. Experimental max-scan distribution versus estimated max-scan distribution*

The main drawback of the scan-statistic approach, as it was applied in the previous examples, is that it requires a large number of image samples in order to determine the distribution $g(c)$. One of the benefits of the generalized approach presented in section 3 is its more economic way of using the sample images, since from one scan are obtained multiple values for the determination of the noise nodule distribution $p(c)$, while with the max-scan procedure only one value per scan is obtained. In Figure 7 we show comparisons between the noise nodule distributions $p(c)$ obtained for a reduced set of only 8 images and the full set of 400 images. As shown in the figure, both experimental histograms are well fitted by the tail of a Gaussian distribution $a \exp\left[-\frac{(c-\mu)^2}{2\sigma^2}\right]$ with mean $\mu = 1$.

In Figure 8 and Figure 9 are shown comparisons between the experimental max-scan distributions $g(c)$ obtained for different search area sizes and the estimates of $g(c)$ obtained

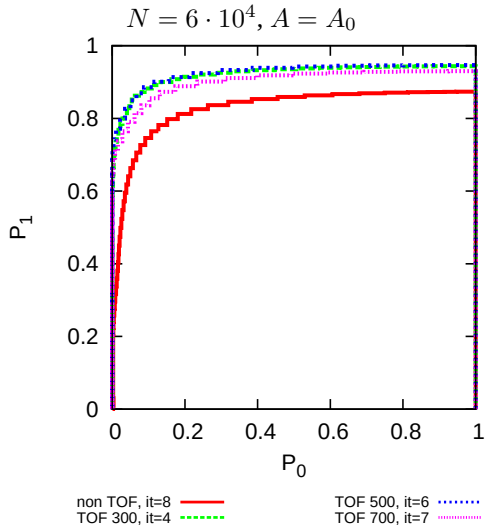


**Figure 6.** Comparison between the LROC curves for images for different TOF resolutions (300, 500 and 700 ps and non-TOF) for a search area of size $A_0$. Results shown are obtained for $N = 6 \cdot 10^4$ counts.

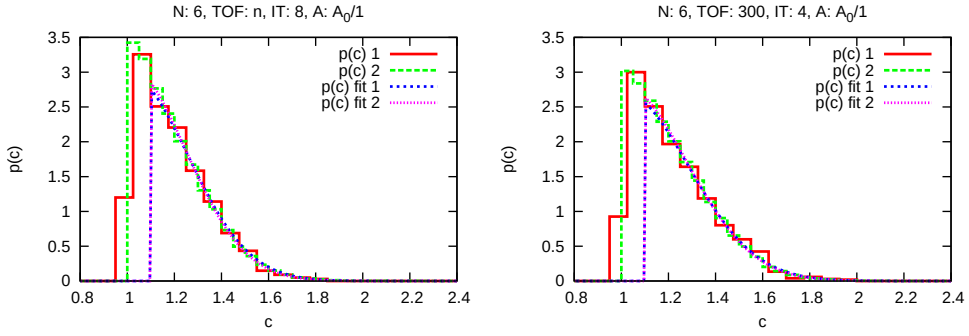


**Figure 7.** Comparison between the noise nodule distributions $p(c)$ obtained for a reduced set of 8 images (1) and the full set of 400 images (2). Both reduced set and full set histograms are well fitted by the tail of a Gaussian distribution with mean set to unity. On the left is a non TOF case ($6 \cdot 10^4$ counts, iteration 8). On the right is a TOF case ($6 \cdot 10^4$ counts, 300 ps timing resolution, iteration 4.)

from equation (11) using the curves fitting the experimental noise nodule distribution $p(c)$ from scanning areas of size $A_0$ with $c > 1.1$. The right column of each figure shows comparisons between experimental cumulative distributions $1 - G(c)$ and the corresponding estimates from equation (12). The plots reveal a good agreement especially for the tail of the distributions and for the large search area cases.

*6.2.1. Practical image evaluation procedure.* In the case of 3D PET there are two symmetrical slices per frame (axially symmetric about the scanner center) for which we have similar data acquisition and image reconstruction conditions. With about 4–5 reconstructed

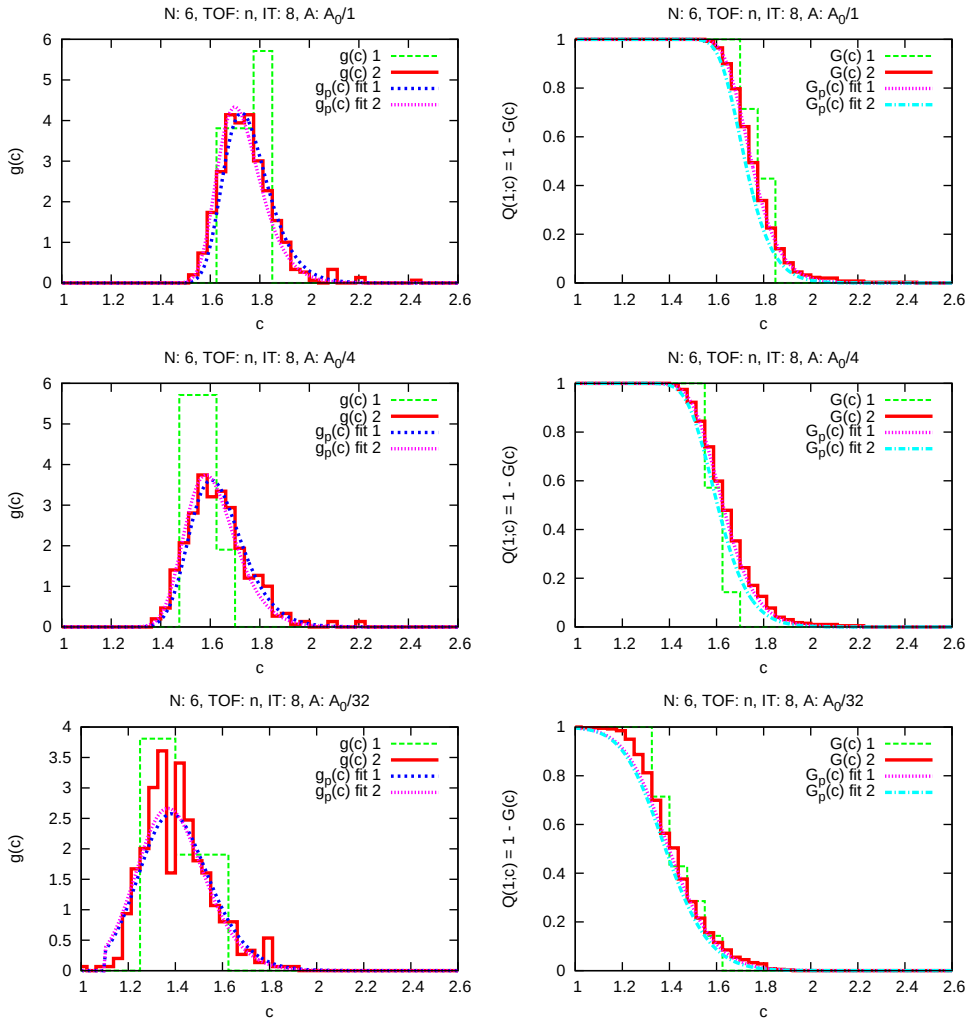


**Figure 8.** Comparisons between the experimental and the estimated (through the $p(c)$ fit) max-scan distributions $g(c)$ (left) and the cumulative distributions $1 - G(c)$ (right) obtained for different search areas. Pairs of curves show results obtained from both the reduced set of only 8 images (1) and the full set of 400 images (2). The non-TOF case ($6 \cdot 10^4$ counts, non-TOF, iteration 8).

images we obtain about 8–10 image realizations. As shown by the above examples this number seems enough for determination of the noise nodule distribution $p(c)$ necessary for the estimation of the scan statistic distribution $g(c)$ according to equation (11). Also we can create phantoms with a number of hot features placed in the same pair of slice positions, as shown in our example in Figure 2. With about 8 features per slice and about 16 in one image frame, only a few image reconstructions are necessary to determine the feature contrast distribution $f(c)$, which is usually assumed to fit a Gaussian. Moreover the hot features can be placed in the same slices used for the estimation of the noise nodule distribution. The scan procedure returns the list of all relevant nodules per scanned area, and from this list

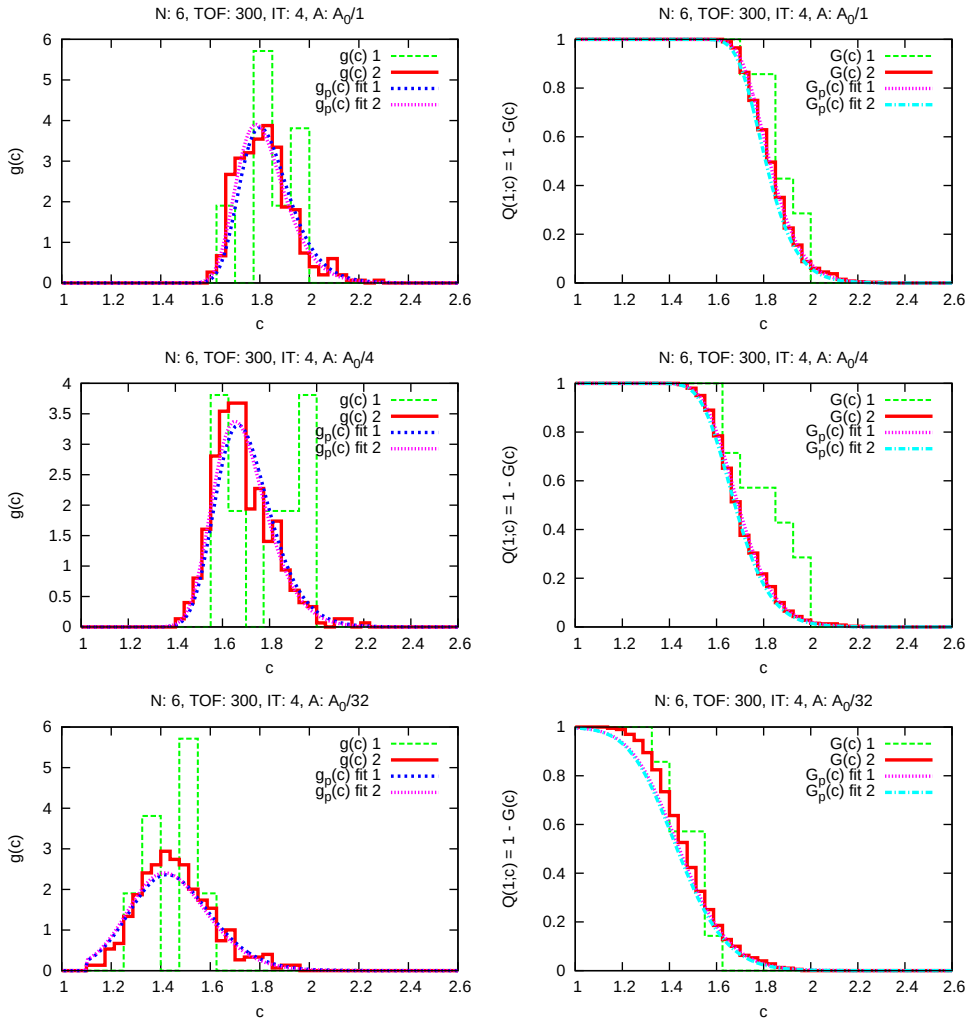


**Figure 9.** Comparisons between the experimental and the estimated max-scan distributions $g(c)$ (left) and the cumulative distributions $1 - G(c)$ (right) obtained for different search areas. Pairs of curves show results obtained from both the reduced set of only 8 images (1) and the full set of 400 images (2). The TOF case ($6 \cdot 10^4$ counts, 300 ps timing resolution, iteration 4)

the true hot features can be distinguished from the noise nodules by their position. Using the noise nodule model proposed and following the procedure outlined above, otherwise laborious detectability studies can be performed with only a few image realizations obtainable either from experiments or simulations.

### *6.3. Image abnormality tests in the case of multiple suspicious features*

In previous sections 2.2, 4.1 and 4.2 several approaches have been presented for testing the image abnormality. We have the usual test using a single decision threshold, and also the two decision thresholds test, and the likelihood ratio. In order to compare these tests we have chosen the image background noise model derived from images reconstructed from

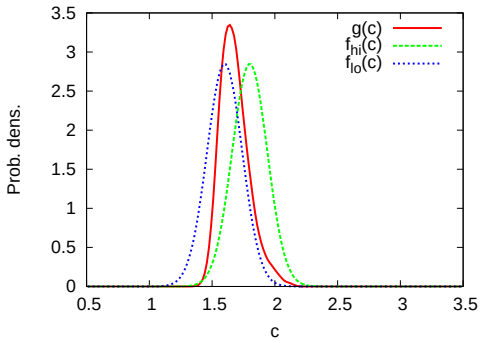


**Figure 10.** Comparison of the feature realization distribution ($f_{\rm hi}(c)$ for the high contrast feature, and $f_{\rm lo}(c)$ for the low contrast feature) with the max-scan distribution $g(c)$ determined for the images reconstructed from $N = 6 \cdot 10^4$ counts, TOF 300 ps data, iteration 4 and search area $A_0/4$.

$N = 6 \cdot 10^4$ counts and TOF 300 ps data, at iteration 4, and we considered a search area $A = A_0/4$. As signals we used two Gaussian distributions, one $f_{\rm hi}(c)$ that partially exceeds the max-scan distribution $g(c)$, and the second $f_{\rm lo}(c)$ that is only a little different from the max-scan distribution (see Figure 10). We considered cases with a single, two, or three signals per feature present image, as well as a combination $\rho(l)$ of these.

In the case of the likelihood ratio test the probabilities for the false $P_0$, respectively, true $P_1$ positive decisions given by equations (20) and (21) were computed using the Monte Carlo method. Random sets $\underline{c_k} \equiv \{c_1, \ldots, c_k\}$ of $k$ nodules were generated separately for the background only case (containing only noise nodules), and for the signal present cases (containing signals and noise nodules). In each situation the likelihood ratio $\lambda(\underline{c_k})$ given by equation (18) was computed and the result binned in the corresponding $L_k^0(\lambda)$ or $L_k(\lambda)$ histogram. We used a logarithmic scale for spacing of these histograms. In the end the reverse cumulative quantities, $P_0(\lambda)$ and $P_1(\lambda)$ respectively, were computed and normalized for the total number of cases generated ($\approx 10^5$). The random sets $\underline{c_k}$ of $k$ nodules, out of which $l$ are true features, were generated by first selecting from the Poisson distribution (8) the number $k - l$ of background noise nodules and then from the $f(c)$ distribution the signal nodules $\{c_1, \ldots, c_l\}$ were sampled, and from the $p(c)$ distribution the remaining $\{c_{l+1}, \ldots, c_k\}$ noise nodules were generated. The Poisson distribution, and the $f(c)$ and $p(c)$ distributions were sampled using the efficient alias sampling technique (Walker 1977, Popescu 2000).

In Figure 11 and Figure 12 we present comparisons between the ROC curves for the likelihood ratio test, a single contrast threshold test, and the family of ROC curves of the two contrast thresholds test. The plots shown in Figure 11 correspond to the high contrast feature case ($f_{\rm hi}$ in Figure 10) for a single feature and two features, respectively, present in the images. The graphs shown in Figure 12 correspond to the low contrast feature case ($f_{\rm lo}$ in Figure 10) for a single, two and three, respectively, features present in the images. We also show the results for the case of a distribution of features $\rho(l)$ with 60 % for $l = 1$, 30 % for

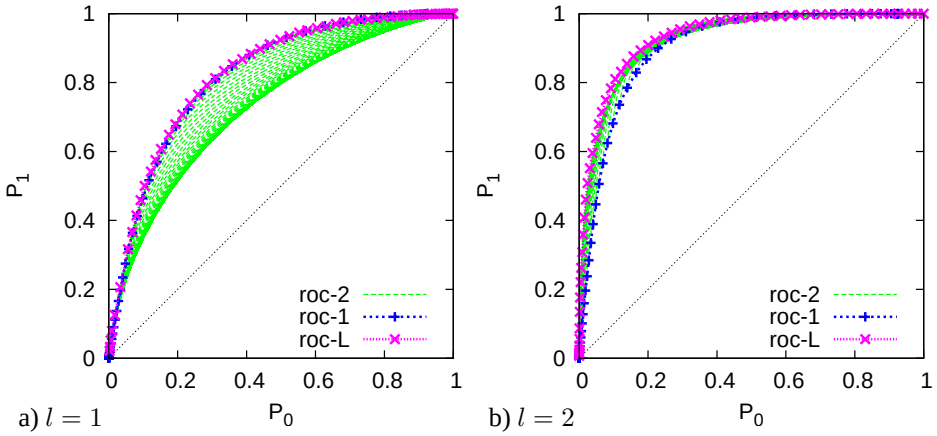


**Figure 11.** Comparisons between the ROC curves for likelihood ratio test (roc-L), a single contrast threshold test (roc-1), and the family of ROC curves of the two contrast thresholds test (roc-2). The plots correspond to the high contrast feature case $f_{\mathrm{hi}}$ (see Figure 10) for a) a single feature, and b) two features, respectively, present in the images.

$l = 2$ and, 10 % for $l = 3$.

The ROC plots show that the common test using only a single decision threshold performs optimally only in the case of the single high contrast feature, for which it is virtually identical to the likelihood ratio test. In the same situation (Figure 11 (a)) the two decision threshold test can lead to worse results as shown by the family of curves below the single threshold and likelihood ratio tests' curves. In the case of multiple features per image the two decision thresholds test represents a better strategy than the single decision threshold both in high contrast and low contrast signal cases. In the case of the single low contrast feature per image (Figure 12 (a)) the two decision level test performs better for most parts of the ROC domain and descends slightly under the single decision threshold test's curve in the high $P_0$ and $P_1$ region. Summarizing, in the cases when the number of nodules rather than the contrast of individual nodules represent a significant indicator of image abnormality, the alternatives to the usual single threshold test perform better. However, unlike the likelihood ratio test, the two decision thresholds test can lead to worse performance in the case of a single high contrast feature per image.

### *6.4. Relation with the human observers*

We can hypothesize, and eye movement studies (Kundel et al. 1989) indicate this, that the detection by the human observers consists also in a scanning process that identifies a number of suspicious locations (a short list) that is followed by a decision process that establishes whether these are abnormal or not. We can assume that the decision strategy is based on the size and contrast of the nodules, as well as their number and spatial distribution. In order to establish a correspondence between the numerical observer and the human observer we need first to determine the correspondence between the human perception of the nodules size and

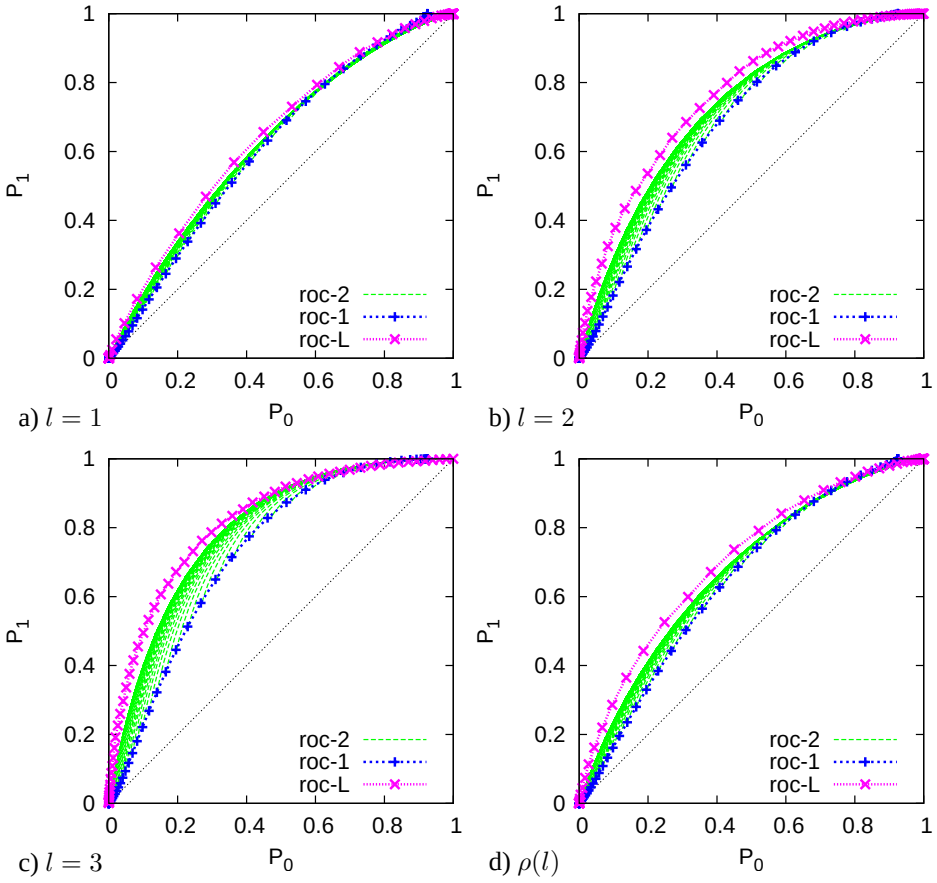


**Figure 12.** Comparisons between the ROC curves for likelihood ratio test (roc-L) and single contrast threshold test (roc-1) and the family of curves for the two contrast threshold test (roc-2). The graphs correspond to the low contrast feature case $f_{\text{lo}}$ (see Figure 10) for: a) single, b) two, c) three features present in the images. The case d) shows results for a distribution of true features per image $\rho(l)$ with 60 % for $l = 1$, 30 % for $l = 2$ and, 10 % for $l = 3$.

contrast and the numerical procedure, and second to determine (or at least approximate) the human observer decision strategy.

The correspondence between the numerical scanning procedure, such as the one used in our examples, and the human visual scanning of the image (e.g. evaluation of the contrast and size of a suspected nodule) can be established if the numerical procedure allows for a parameterization that determines a wide variation of its response $S(\theta)$ from very good to poor performance. This variation of the scanning procedure response can be achieved for example by coupling it with a filter. With the human observer response $S_h$ returning values somewhere in between the lowest and the highest response of the numerical procedure, there is always a point $\theta_h$ so that $S_h = S(\theta_h)$. With $\theta$ standing for multiple parameters, a fit could return values that will approximate the human observer for a wider range of operating situations.

A more difficult task is the determination of the decision strategy used by humans, or at least devising a procedure that imitates its results. Recent studies have revealed the ability

of humans to apply the correct underlying probability distribution when asked to predict outcomes of various random variables, if they had prior experience with them (Griffiths & Tenenbaum 2006). Similarly we can assume that when asked to detect small nodules on noisy backgrounds the human observers will also tend to adopt the optimal strategy corresponding to their prior assumptions. Our results indicate that from a theoretical point of view the optimal decision strategy depends on the contrast of the features, as well as how many features may be present. For an observer using a multiple decision threshold test approach, that corresponds to selecting the decision threshold combinations that produce the optimum operating curve from a family of ROCs. As previously shown, using a two decision thresholds test in the case with a relatively high feature contrast and only one feature present can lead to a suboptimal strategy, while it is always a better strategy if two or more features are present. If the human observers use a similar decision mechanism, using multiple clues for image abnormality (i.e. the individual nodule size and contrast, as well as their number and spatial distribution) then instead of a single ROC we also could have a family of ROC curves. The selection of the optimal strategy will depend on, and sometimes be very sensitive to, the accuracy of the assumptions made by the observer about the noise structure, target contrast and size distribution, as well as the likely number of targets per image.

## 7. Conclusions

The comparisons between scan-statistic based image evaluations and the fixed position evaluations reveal a reduced sensitivity of the latter with the TOF timing improvement. This is in agreement with similar conclusions reached in (Yendiki & Fessler 2006a) where various image regularization techniques were tested. In general the detectability of the features at unknown positions depends on the search area size. Therefore the search area size is a factor that should be carefully considered when the image evaluations are meant to produce results proportional to the real performance of the evaluated methods, as is required for cost-benefit evaluations. If only the ranking order is of interest this factor is less important, yet the area scan methods are more sensitive than the fixed position evaluations.

We present a generalized scan-statistic approach that has several advantages compared with the direct determination of the scan statistic (or max-scan) distribution from multiple realizations. One advantage is the greater efficiency, since one scan search obtains multiple values for the determination of the noise nodules distribution $p(c)$, while the max-scan procedure obtains only one value per scan. A further advantage is the greater flexibility, since the $p(c)$ distribution can be determined by searching areas of arbitrary size, and then the theory can be applied to a certain reference area size. The model proposed here is similar to the model recently published by Chakraborty (2006b, 2006a) as well as the model used in (Edwards et al. 2002), all relying on assumptions originally made in (Bunch et al. 1978) and later developed in (Swensson 1996). What differentiates our approach is the fact that here we focus on the derivation of the scan statistic distribution $g(c)$ from the distribution

of noise nodules $p(c)$, relating the expected number of nodules per image with the search area and the nodules density $\nu(c) = An_0q(c)$. Also, based on the noise model proposed we address the multiple targets per image case and derive image abnormality tests using the likelihood ratio and an alternative multiple decision thresholds approach. The results obtained reveal that in the case of low contrast nodules or multiple nodules the commonly assumed test strategy based on a single decision threshold underperforms compared with the alternative tests, showing that not only the contrast or size but also the number of suspicious nodules is a clue indicating the image abnormality. It must be noted that the tests that use multiple clues, that are not unified in a single decision variable as in the case of the likelihood ratio test, do not necessarily produce a unique ROC curve, as we show in the examples with the two decision thresholds test.

These theoretical results add to results of previous works pointing out that the evaluation of the detectability of small features at unknown locations in noisy images requires more complex underlying models than the binormal model usually assumed in ROC analysis (Dorfman & Alf 1969, Metz 1986) and other related works. The binormal model seems to apply as a realistic detection mechanism model only for the fixed position detection tasks, while for the other situations it can be seen rather as a pragmatic data modeling approach without basis on a specific detection mechanism model.

## Acknowledgment

This work was supported by the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health under grants R21-EB005434 and R33-EB001684.

## References

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields, *Ann. Appl. Probab.* **10**(1): 1–74.

Barrett, H. H., Yao, J., Rolland, J. P. & Myers, K. J. (1993). Model observers for assessment of image quality, *Proc. Nat. Acad. Sci. USA* **90**(21): 9758–9765.

Bunch, P. C., Hamilton, J. F., Sanderson, G. K. & Simmons, A. H. (1978). Free-response approach to the measurement and characterization of radiographic-observer performance, *Journal of Applied Photographic Engineering* **4**(4): 166–171.

Chakraborty, D. P. (1989). Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data, *Phys. Med. Biol.* **16**(4): 561–568.

Chakraborty, D. P. (2006a). ROC curves predicted by a model of visual search, *Phys. Med. Biol.* **51**: 3463–3482.

Chakraborty, D. P. (2006b). A search model and figure of merit for observer data acquired according to the free-response paradigm, *Phys. Med. Biol.* **51**: 3449–3462.

Chakraborty, D. P. & Winter, L. H. L. (1990). Free-response methodology: alternate analysis and a new observer-performance experiment, *Radiology* **174**: 873–881.

Dorfman, D. D. & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals — rating-method data, *J. Math. Psychol.* **6**: 487–496.

Edwards, D. C., Kupinski, M. A., Metz, C. E. & Nishikawa, R. M. (2002). Maximum likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model, *Med. Phys.* **29**(12): 2861–2870.

Gifford, H. C., King, M. A., Pretorius, P. H. & Wells, R. G. (2005). A comparison of human and model observers in multislice LROC studies, *IEEE Trans. Med. Imag.* **24**(2): 160–169.

Gifford, H. C., Pretorious, P. H. & King, M. A. (2003). Comparison of human- and model-observer LROC studies, *in* D. P. Chakraborty & E. A. Krupinski (eds), *Proc. SPIE*, Vol. 5034 of *Medical Imaging: Image Preception, Observer Performance, and Technology Assesment*, pp. 112–122.

Glaz, J., Naus, J. & Wallenstein, S. (2001). *Scan Statistics*, Springer, New York.

Griffiths, T. L. & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition, *Psychological Science* **17**(9): 767–773.

Hotelling, H. (1931). The generalization of Student's ratio, *Ann. Math. Stat.* **2**(3): 360–378.

Hotelling, H. (1939). Tubes and spheres in $n$-space and a class of statistical problems, *Am. J. Math.* **61**(2): 440–460.

Johansen, S. & Johnstone, I. M. (1990). Hotelling's theorem on the volumes of tubes: some illustrations in simultaneous inference and data analysis, *Ann. Statist.* **18**(2): 652–684.

Khurd, P. & Gindi, G. (2005). Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers, *Phys. Med. Biol.* **50**: 1519–1532.

Knowles, M. & Siegmund, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression, *Intl. Stat. Rev.* **57**(3): 205–220.

Kulldorff, M. (1997). A spatial scan statistic, *Communications in statistics — Theory and Methods* **26**(6): 1481–1496.

Kundel, H. L. (1981). Predictive value and threshold detectability of lung tumors, *Radiology* **139**: 25–29.

Kundel, H. L., Nodine, C. F. & Krupinski, E. A. (1989). Searching for lung nodules: visual dwell indicates location of false-positive and false-negative decisions, *Invest. Radiol.* **4**: 472–478.

Lewitt, R. M. (1992). Alternatives to voxels for image representation in iterative reconstruction algorithms, *Phys. Med. Biol.* **37**(3): 705–716.

Metz, C. E. (1986). ROC methodology in radiologic imaging, *Invest. Radiol.* **21**: 720–733.

Naiman, D. Q. & Priebe, C. E. (2001). Computing scan statistic $p$ values using importance sampling, with applications to genetics and medical image analysis, *Journal of Computational and Graphical Statistics* **10**(2): 296–328.

Naus, J. I. (1965). Clustering of random points in two dimensions, *Biometrika* **52**: 263–267.

Orford, K. J. (2000). The analysis of cosmic ray data, *J. Phys. G: Nucl. Part. Phys.* **26**: R1–R26.

Park, S., Clarkson, E., Kupinski, M. A. & Barrett, H. H. (2005). Efficiency of the human observer detecting random signals in random backgrounds, *J. Opt. Soc. Am. A* **22**(1): 3–15.

Park, S., Kupinski, M. A., Clarkson, E. & Barrett, H. H. (2003). Ideal-observer performance under signal and background uncertainty, *in* C. J. Taylor & J. A. Noble (eds), *Information Processing in Medical Imaging*, Vol. 2732 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, pp. 342–353.

Parzen, E. (1960). *Modern Probability Theory and Its Applications*, John Wiley and Sons, New York.

Popescu, L. M. (2000). An extension of alias sampling method for parametrized probability distributions, *J. Comp. Phys.* **160**: 612–622.

Popescu, L. M. & Lewitt, R. M. (2003). A versatile approach for Monte Carlo simulation of tomographic systems, *IEEE Nucl. Sci. Symp. Conf. Rec.*, Vol. 4, pp. 2785 – 2788.

Popescu, L. M., Matej, S. & Lewitt, R. M. (2004). Iterative image reconstruction using geometrically ordered subsets with list-mode data, *IEEE Nucl. Sci. Symp. Conf. Rec.*, Vol. 6, pp. 3536 – 3540.

Siegmund, D. O. & Worsley, K. J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field, *Ann. Statist.* **23**(2): 608–639.

Swensson, R. G. (1996). Unified measurement of observer performance in detecting and localizing target objects on images, *Med. Phys.* **23**(10): 1709–1725.

Swensson, R. G., King, J. L. & Gur, D. (2001). A constrained formulation for the receiver operating characteristic (ROC) curve based on probability summation, *Med. Phys.* **28**(8): 1597–1609.

Terranova, F. (2004). Peak finding through Scan Statistic, *Nucl. Instr. Meth. Phys. Res. A* **519**: 659–666.

Wagner, R. F. & Brown, D. G. (1985). Unified SNR analysis of medical imaging systems, *Phys. Med. Biol.* **30**(6): 489–518.

Walker, A. J. (1977). An efficient method for generating discrete random variables with general distributions, *ACM Transactions on Mathematical Software* **3**(3): 253–256.

Wallenstein, S. (1980). Test for detection of clustering over time, *American Journal of Epidemiology* **111**(3): 367–372.

Worsley, K. J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images, *Ann. Statist.* **23**(2): 640–669.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J. & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation, *Human Brain Mapping* **4**: 58–73.

Yendiki, A. (2005). *Analysis of signal detectability in statistically reconstructed tomographic images*, Ph.D. thesis, University of Michigan, Ann Arbor.

Yendiki, A. & Fessler, J. A. (2006a). Analysis of observer performance in known-location tasks for tomographic image reconstruction, *IEEE Trans. Med. Imag.* **25**(1): 28–40.

Yendiki, A. & Fessler, J. A. (2006b). Analysis of unknown-location signal detectability for regularized tomographic image reconstruction, *Proc. IEEE Intl. Symp. Biomed. Imag.*, pp. 279–283.