APPLICATION OF THE RANDOM FIELD THEORY IN PET IMAGING – INJECTION DOSE OPTIMIZATION

Jiří Dvořák, Jiří Boldyš, Magdaléna Skopalová and Otakar Bělohlávek

This work presents new application of the random field theory in medical imaging. Results from both integral geometry and random field theory can be used to detect locations with significantly increased radiotracer uptake in images from positron emission tomography (PET). The assumptions needed to use these results are verified on a set of real and simulated phantom images.

The proposed method of detecting activation (locations with increased radiotracer concentration) is used to quantify the quality of simulated PET images. Dependence of the quality on the injection dose (amount of applied radiotracer) and patient's body parameters is estimated. It is used to derive curves of constant quality determining the injection dose needed to achieve desired quality of the resulting images. The curves are compared with the formula currently used in medical practice.

Keywords: random field theory, Euler characteristic, PET imaging, PET image quality

Classification: 60G35, 60G60

1. INTRODUCTION

Random field theory has many exciting applications for example in cosmology or neuroscience, see [25] and references therein. This work presents new application of the random field theory in medical imaging – particularly in positron emission tomography (PET). It is based on the results concerning exceedance probabilities of smooth random fields.

PET is a medical imaging technique designed to provide functional information about the processes in human body [16], as opposed to the anatomical information provided by other imaging modalities such as computed tomography (CT). It estimates spatial distribution of the radiotracer in the patient's body. Figure 1 shows examples of sections of the corresponding CT and PET images of the same patient.

All radiotracers used in PET produce positrons during their decay. Specific radiotracer is chosen according to the diagnostic task. The most common application of PET in medical practice is examination of oncologic patients and detection of possible tumors. In the case of tumor detection the typical choice of radiotracer is ^{18}F-FDG, a radioactive analogue of glucose, because tumors display higher glucose uptake than normal tissues.
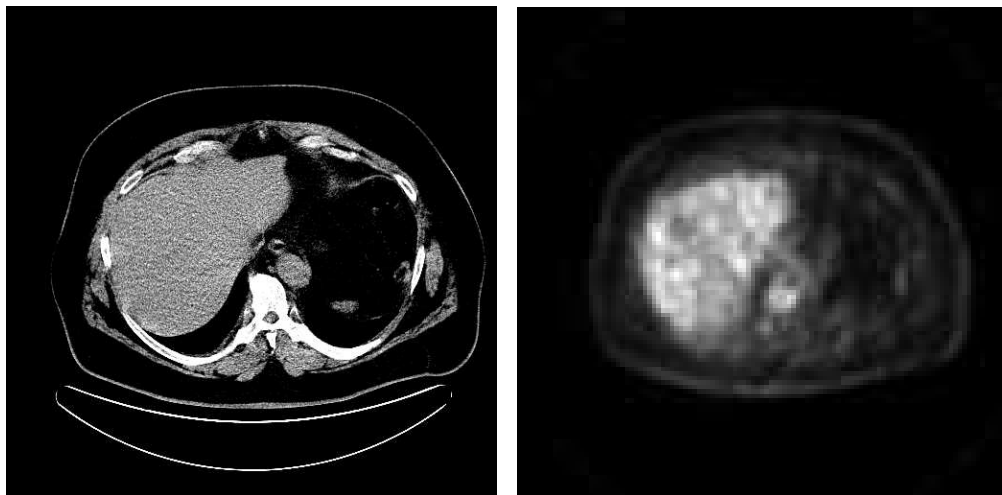
Fig. 1. Examples of CT (left) and PET images (right) of the corresponding section of patient's liver.

This leads to increased concentration of the radiotracer in the tumor compared to the background.

Quality of the resulting three-dimensional (3D) image depends strongly on the amount of radiotracer applied (and its activity), duration of the scan and physical parameters of the patient's body, besides scanner quality and other scanning and reconstruction parameters. The need to optimize the injected dose of radiotracer arises from the contradictory demands for sufficient image quality and minimization of the radiation dose received by the patient.

The radiotracer dosage recommendation used nowadays in practice depends linearly on the patient's weight (EANM guidelines [5]). An exception is the specific recommendation for children which tries to equalize effective radiation dose among the patients (the EANM pediatric dosage card [13]). However, there is a possibility to improve the recommendation so that it produces images of comparable diagnostic quality for different patients.

Standard metrics of PET image quality are based on statistical properties of the acquired data [2, 9, 15, 22], the amount of noise in the resulting image [15, 21] or its subjective visual quality [10, 12].

The ultimate measure of PET image quality is the performance of human observer in a given diagnostic task, such as detection of lesions. These are seen in the image as small foci of increased radiotracer concentration, suggesting the presence of a tumor [11, 12]. Studies involving human observers are very time-consuming and thus mathematical model observers were developed to simulate performance of human observers [1, 11].

In certain situations PET images can be regarded as realizations of a smooth random field sampled on a rectangular lattice of discrete points [19, 24, 26, 27]. The diagnostic task of detecting significant increase of the random field's values in a given homogeneous

region can be treated as a problem of testing null hypothesis that the random field has a constant mean. Example of such a task is detecting tumors in patient's liver.

Under the assumptions of Gaussian distribution, stationarity and isotropy threshold can be determined which is exceeded by the values of the random field only with a given probability (e. g. 0.05) [26]. Then, following the standard procedure of hypothesis testing, one can reject the null hypothesis of constant mean if the values of the random field exceed the threshold at any point. This leads to the conclusion that significant activation (increase of the random field's values) is present in the image. What's more, the activation can be easily localized – it is situated at the points at which the values exceed the threshold.

This work presents alternative method of expressing quality of (simulated) PET images in terms of lesion detectability, see Section 4. The method makes use of the results of the random field theory and integral geometry. For comparison, a standard metric of PET image quality – NEC, see Section 2 or [9, 15, 18] – is calculated as well.

Relationship between the quality of the images, the amount of applied radiotracer and patient's body parameters is estimated. Curves of constant quality are derived which determine the amount of radiotracer needed to achieve required quality of the resulting image. The curves are compared with the formula currently used in medical practice [13].

The paper is organized as follows. Section 2 provides necessary information about PET. Section 3 formulates the task of lesion detection as a hypotheses testing problem. Exceedance probabilities of random fields and their approximations are discussed in Section 4 and 5. Details of the PET images simulation are given in Section 6. In Section 7 statistical properties of real and simulated PET images are assessed. The curves of constant quality are derived in Section 8 and the results are summarized in Section 9. Finally, concluding remarks are given in Section 10.

2. POSITRON EMISSION TOMOGRAPHY

In a typical case of oncologic PET examination the patient is injected some amount of a radiotracer (e. g. ^{18}F-FDG, as stated above) and rests for $60-90$ minutes so that the radiotracer is distributed along the patient's body in the time of examination.

^{18}F is an unstable isotope of fluorine and it decays with half-life of approximately 110 minutes while emitting a positron. The positron almost immediately annihilates and a pair of annihilation photons with high energy moving in opposite directions is emitted.

If the annihilation photon hits the detector surrounding the patient's body it is detected by a scintillation crystal. If two photons are registered within a short time interval called coincidence window (typically a few nanoseconds long) it is assumed that they originated in the same annihilation. The event of detecting such a pair of annihilation photons is called coincidence.

There are three different types of coincidences (see also Figure 2):

- true – two photons from the same annihilation are detected and none of them changed its direction (no scatter occurred),

- scattered – one or both photons from the same annihilation interacted with material and underwent a scatter (change of direction),
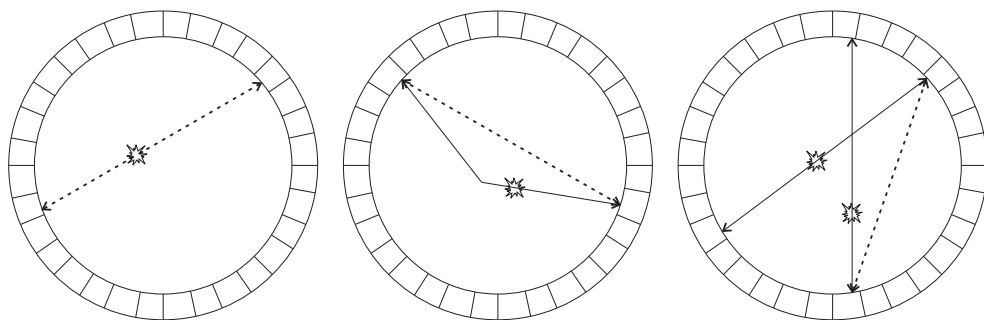
Fig. 2. Three types of events which give rise to different types of coincidences: true (left), scattered (middle) and random coincidences (right), with corresponding lines of response (dashed lines).

- random – two photons from different annihilations are detected in the given coincidence window.

Only the true coincidences bring valid information about the radiotracer distribution in the patient's body. Scattered coincidences can be thought of as a blur and random coincidences represent noise in the data.

To every recorded coincidence so-called line of response (LOR) can be assigned. It is a line joining the two detectors which detected the two photons in coincidence (see also Figure 2). It is assumed that the annihilation during which the photons originated took place somewhere along this LOR.

True coincidences are assigned a correct LOR while scattered and random coincidences are assigned a wrong LOR (see again Figure 2) which degrades the data quality. The acquisition process results in counts of coincidences which occur along individual LORs. This form of output is called sinogram and it is used to estimate the radiotracer distribution in the field of view [16].

The undesirable effect of scattered and random coincidences can be reduced by application of different corresponding correction methods. They are followed by a reconstruction process during which a 3D image of radiotracer concentration is calculated from the sinogram.

So called noise equivalent count (NEC) calculated from the coincidence counts is a standard measure of the quality of the sinogram. It quantifies signal strength in noisy data. Let T, S and R, respectively, denote the number of true, scattered and random coincidences recorded during the acquisition. Then NEC can be calculated as follows:

$$NEC = \frac{T^2}{T + S + kfR},\tag{1}$$

where f is fraction of the patient's volume in the scanner field of view and k is either 1 or 2, depending on the type of applied random coincidence correction.

NEC describes statistical quality of the sinogram and it is not affected by the reconstruction process. This makes it very useful when comparing data obtained from different scanners and under different acquisition protocols.

In practice the number and distribution of the scatter coincidences can be estimated e. g. by a model-based scatter correction algorithm using a single-scatter simulation technique described in [23]. Such kind of algorithm is implemented in many scanners.

The most common way of estimating the random coincidence counts is using delayed coincidence channel, where one of the single-photon events (registration of a photon in a detector) has an arbitrarily large time delay. Thus, no coincidence recorded in this channel can be true or scattered. Number of these delayed coincidences equals, on average, number of actual random coincidences [7].

These estimates of numbers of scattered and random coincidences can be subtracted from the total number of recorded coincidences (so-called prompt coincidences) in order to estimate the number of true coincidences and calculate the NEC value. In the simulations described in Section 6 data related to the true, scattered and random coincidences are directly available. This enables simple and efficient calculation of NEC.

From finite amount of data one cannot estimate whole spatial distribution of the radiotracer and thus some discretization is needed. The field of view is divided into a finite number of voxels (3D analogue of a two-dimensional pixel). Output of the reconstruction algorithm is the estimate of the number of annihilations in each voxel. High value in a voxel indicates high concentration of the radiotracer in the given area.

The resulting image is finally smoothed by a 3D Gaussian filter with a full width at half maximum (FWHM) of 5 mm. The reason is to reduce noise and to improve the visual image quality prior to the diagnostic process.

3. LESION DETECTION AS A HYPOTHESES TESTING PROBLEM

We regard concentration of the radiotracer in the patient's body as a realization of a smooth random field and values recorded in the PET image as the values of the random field on a regular grid of points. We will consider the task of lesion detection – finding locations with significantly increased values that should not be exhibited by background with no pathological changes. Such areas will be attributed to an "activation".

Both real and simulated PET images contain large amount of noise and it is difficult to distinguish whether certain peak in the data occurs by chance or indicates presence of a lesion, see Figure 1 (right).

Random field theory can help solve this problem. It has been successfully used in the past years in interpretation of brain imaging experiments (e. g. [27] and references therein). When studying which parts of human brain are involved in solving given tasks two brain PET images are acquired – one during the task and the other when the patient is at rest.

Parts of the brain involved in solving the task exhibit increased blood flow due to increased neuron activity. This leads to increased image intensities compared to those recorded in parts of the brain not involved in solving the task or to values recorded under the rest condition. The difference of the two images is then analyzed. It is assumed that the values of this difference image are centered around zero except of the parts of the brain activated by the task.

To decrease the effect of noise the same study is performed for several volunteers. The average of the difference images is then analyzed to locate parts of the brain activated by the task.

Our situation is different. We want to inspect PET image of the liver area in order to locate possible tumors and we work with a single image of a single patient. There is no way we could obtain two images of the patient in "activated/non-activated" conditions. Similarly, we cannot obtain multiple images to be averaged in order to increase the signal-to-noise ratio as this would extend scan duration and accentuate problems related to the patient's movement. However, with some caution we can utilize the same methods of the random field theory even in this case.

PET image model used in the following is a smooth real-valued random field $(X_s, s \in \mathbb{R}^3)$. We are interested in its properties in the set $S \subset \mathbb{R}^3$, e.g. in part of the patient's body. We take the values of the PET images in individual voxels for the values of the random field in appropriate lattice points.

Applying the classical scheme of hypothesis testing we formulate the null hypothesis H_0 that no lesion ("activation") is present in the region of interest S (just as in the standard method used in [26] or [27]). This means that under H_0 the random field $(X_s, s \in S)$ has a constant mean. If the values of the random field exceed an appropriate threshold t_0 in any point of S we reject the null hypothesis of no activation.

Important consequence of this procedure is that rejection of the null hypothesis implies localization of the activation – it occurs in those points of S in which the random field's values exceed the threshold t_0.

In the context of the brain imaging experiment described above the choice of the null hypothesis (zero mean of the random field) was motivated by the subtraction of the two images. In our situation we restrict ourselves to regions S with homogeneous radiotracer distribution (under H_0), such as the liver area. This means that the random field $(X_s, s \in S)$ has a constant mean which can be estimated and subtracted from X_s. Thus we can assume that the random field has a zero mean under the null hypothesis and in what follows we can work along the lines of the standard method used in [26] or [27] (where the zero mean of the random field, under H_0, was ensured by subtracting two independent copies of the random field).

When constructing a test of the hypothesis of no activation a classical problem of multiple comparison arises. Typical PET image of liver comprises of 10 to 30 thousand voxels. If we compared each value with a threshold t_0 set to e.g. a 95%-quantile of its distribution, several hundred voxels would exceed this threshold only by chance even under the null hypothesis. We would find activation in hundreds of voxels even though no activation was actually present in the image.

Clearly another method of choosing t_0 must be applied. One option is to use the Bonferroni correction. Suppose we have n identically distributed observations (this would be the case if our random field was weakly stationary and Gaussian). We want to reject H_0 if *any* of the observations exceeds a given threshold t_0. If we set $\alpha_0 = \alpha/n$ and find the threshold t_0 such that it is exceeded by a single observation only with the probability α_0, the significance level of the test will be less than or equal to α.

This method is very useful if the observations are independent or nearly independent – the actual significance level of the test is close to α. If the observations are strongly

correlated (as could be expected in the case of a smooth random field), this test is unnecessarily conservative and its significance level is much lower than α.

Another option is to slightly restate the problem. We want to reject the null hypothesis if the maximum of the observed values exceeds t_0, i.e. if $\sup_{s \in S} X_s \geq t_0$. We set the significance level $\alpha \in (0, 1)$. Then we can determine the value of the threshold t_0 from the equation

$$\mathbb{P}\left(\sup_{s \in S} X_s \geq t_0\right) = \alpha. \tag{2}$$

The only flaw of this procedure is the necessity to know the distribution of $\sup_{s \in S} X_s$ under H_0 or at least good approximations of $\mathbb{P}(\sup_{s \in S} X_s \geq t)$ for high values of t.

4. EXCEEDANCE PROBABILITIES OF RANDOM FIELDS

Surprising result of the random field theory is that approximation of $\mathbb{P}(\sup_{s \in S} X_s \geq t)$ is provided by the expected value of the Euler characteristic of the set of points where the values of the random field exceed the threshold t [25]:

$$\mathbb{P}\left(\sup_{s \in S} X_s \geq t\right) \approx \mathbb{E}\,\chi(\{s \in S : X_s \geq t\}). \tag{3}$$

Here we call $E_t = \{s \in S : X_s \geq t\}$ the excursion set of the random field X above the threshold t and $\chi(A)$ denotes the Euler characteristic of the set A.

We consider the definition of the Euler characteristic as one of the intrinsic volumes. They are a class of additive functionals that can be defined on a fairly general class of sets called the basic complexes. We do not need that level of generality so we first consider only the class of compact convex sets $\mathcal{K}$. For the following definition and more thorough discussion on intrinsic volumes see [4].

Definition 4.1. Let K be a convex body in $\mathbb{R}^N$ and K_r its dilation by distance $r > 0$:

$$K_r = \left\{x \in \mathbb{R}^N : d(x, K) \leq r\right\}, \tag{4}$$

where $d(x, K) = \inf_{y \in K} |x - y|$ is the Euclidean distance of the point x and the set K. Let λ_N denote the Lebesgue measure in $\mathbb{R}^N$. The intrinsic volumes $\mu_j, j = 0, 1, \ldots, N$, are the only additive functionals on the class of compact convex sets $\mathcal{K}$ fulfilling the Steiner's formula

$$\lambda_N(K_r) = \sum_{j=0}^{N} \omega_{N-j} r^{N-j} \mu_j(K), \tag{5}$$

where ω_j is volume of the unit ball in $\mathbb{R}^j$ (see e.g. [4]).

$\mu_0(K)$ is the Euler characteristic of K and $\mu_N(K)$ its volume. In $\mathbb{R}^3$ ($N = 3$) $\mu_2(K)$ and $\mu_1(K)$ express (up to a scaling factor) surface area of K and its cross-sectional diameter, respectively.

Definition of the intrinsic volumes can be extended to the class of finite unions of compact convex sets using the following additivity property:

$$\mu_j(K \cup L) = \mu_j(K) + \mu_j(L) - \mu_j(K \cap L) \tag{6}$$

for any $K, L \in \mathcal{K}$ and $j = 0, 1, \ldots, N$.

For further extension of the definition of Euler characteristic to the class of basic complexes see Ch. 6 of [4].

Ch. 14 of [4] discusses the accuracy of the approximation (3) and it derives an upper bound on $|\mathbb{E}\left[\chi(E_t)\right] - \mathbb{P}\left(\sup_{s \in S} X_s \geq t\right)|$.

Note that for values of t close to the global maximum of (suitably regular) random field X it can be seen that $\chi(E_t) = 1$ if $\sup_{s \in S} X_s \geq t$ (the excursion set E_t consists of a single connected component with no holes or hollows), and $\chi(E_t) = 0$ for $\sup_{s \in S} X_s < t$ (E_t is empty). Calculating the expectation gives (3) and intuitively justifies the approximation, at least for high values of t.

Important result of integral geometry and differential topology is an explicit formula for calculating $\mathbb{E}\left[\chi(E_t)\right]$ for certain classes of random fields (particularly Gaussian and related fields). This enables us to solve the equation

$$\mathbb{P}\left(\sup_{s \in S} X_s \geq t_0\right) \approx \mathbb{E}\left[\chi(E_{t_0})\right] = \alpha \tag{7}$$

for small values of α. From (7) we determine the desired value of t_0 needed to perform the test of the null hypothesis of no activation. If the values of the random field exceed t_0 in any point of S we reject the null hypothesis. Moreover, the activation is localized in E_{t_0}.

5. CALCULATING THE EXPECTED EULER CHARACTERISTIC

The crucial formulae for calculating $\mathbb{E}\left[\chi(E_t)\right]$ appear in literature in many forms with different context and level of generality. Very general results of this type can be found in [4]. We present more specific version from the paper [25] (for notation and overview of the results for different types of random fields see also [8]).

Suppose that $S \subset \mathbb{R}^N$ is a compact convex set with a smooth boundary ∂S that is a regular $(N-1)$-dimensional C^2-manifold. Following results hold for random fields defined in all points of S. In the case of a discrete random field with values defined only in a set of given lattice points the results can be used as approximations if the random field is smooth enough (see e.g. discussion in [19]).

Let $X = X(s)$, $s = (s_1, \ldots, s_N) \in \mathbb{R}^N$, be a weakly stationary random field in $\mathbb{R}^N$ and $\dot{X}_j = \dot{X}_j(s) = \partial X(s)/\partial s_j$, $\ddot{X}_{jk} = \ddot{X}_{jk}(s) = \partial^2 X(s)/\partial s_j \partial s_k$, $j, k = 1, \ldots, N$. The moduli of continuity of $\dot{X}_j$ and $\ddot{X}_{jk}$ inside S are defined as follows:

$$\omega_j(h) = \sup_{\substack{s, t \in S, \\ \|s-t\| < h}} \left|\dot{X}_j(s) - \dot{X}_j(t)\right|, j = 1, \ldots, N, \tag{8}$$

$$\omega_{jk}(h) = \sup_{\substack{s, t \in S, \\ \|s-t\| < h}} \left|\ddot{X}_{jk}(s) - \ddot{X}_{jk}(t)\right|, j, k = 1, \ldots, N. \tag{9}$$

Let $\dot{X} = \partial X/\partial s$ be the gradient vector of X and $\ddot{X} = \partial^2 X/\partial s \partial s'$ its $N \times N$ Hessian matrix.

Theorem 5.1. (Worsley [25])

Let $S \subset \mathbb{R}^N$ be as above and $X(s)$ a weakly stationary isotropic random field satisfying the following conditions:

C1. for any $\epsilon > 0$

$$\mathbb{P}\left(\max_{j,k}\{\omega_j(h), \omega_{jk}(h)\} > \epsilon\right) = o(h^N), h \searrow 0, \qquad (10)$$

C2. $\ddot{X}$ has finite variance conditional on $(X, \dot{X})$,

C3. the density of $(X, \dot{X})$ is bounded from above, uniformly for all $s \in S$.

Then

$$\mathbb{E}\left[\chi(E_t)\right] = \sum_{i=0}^{N} \mu_i(S)\rho_i(t), \qquad (11)$$

where $\rho_i(t), i = 1, \ldots, N$, are so-called i-dimensional intensities of the Euler characteristic of the excursion set E_t (for precise definition see [25]).

It is clear from (11) that the expected value of $\chi(E_t)$ depends on the properties of S only by means of its intrinsic volumes $\mu_i(S)$ and on the properties of the random field only by means of the intensities $\rho_i(t)$.

The considered conditions on smoothness of ∂S are too restrictive in many applications. However, [26] shows that (11) holds in $\mathbb{R}^2$ and $\mathbb{R}^3$ for sets S with a piecewise smooth boundary such as cube and other polyhedra.

In the following we consider a zero-mean stationary isotropic Gaussian random field X with unit variance. Components of the vector $\dot{X}$ are in this case independent (see Ch. 5 of [4]). The intensities of the Euler characteristic are in this situation given by

$$\rho_0(t) \quad = \quad \frac{1}{(2\pi)^{\frac{1}{2}}} \int_t^\infty e^{-u^2/2} \, \mathrm{d}u, \qquad (12)$$

$$\rho_n(t) \quad = \quad \frac{\lambda^{\frac{n}{2}}}{(2\pi)^{\frac{n+1}{2}}} H_{n-1}(t)e^{-t^2/2}, n = 1, \ldots, N, \qquad (13)$$

(see [8, 25] and consider the independence of the components of $\dot{X}$). The parameter λ expresses the roughness of the random field and it is defined as the variance of one component of the gradient vector $\dot{X}$. Note that due to the stationarity and isotropy of the random field X all N components of $\dot{X}$ have the same variance. In practice λ is estimated by empirical variance of properly scaled numerical differences of the values in neighbouring voxels as suggested in [24]. $H_m(t)$ denotes the Hermite polynomial of degree m in t given by

$$H_m(t) = m! \sum_{j=0}^{\lfloor m/2 \rfloor} \frac{(-1)^j t^{m-2j}}{j!(m - 2j)!2^j}, m \geq 0, t \in \mathbb{R}, \qquad (14)$$

where $\lfloor x \rfloor$ is the largest integer not greater than x.

For Gaussian random fields Theorem 5.2.2 of [3] presents simpler conditions which assure that C1 − C3 from the Theorem 5.1 are satisfied. Also, a sufficient condition on the covariance function can be obtained from Theorem 3.4.1 of [3]. For example, it is sufficient that the third derivatives of $X(s)$ have finite variances or that the correlation function of X is of the form $r(h) = \exp\{-\rho||h||^2/2\}$ for some $\rho > 0$ [25].

Theorem 4.3.1 of [3] shows that under certain conditions on regularity of the random field X the excursion set E_t is a basic complex and thus its Euler characteristic is properly defined. These conditions can be verified using Theorem 3.2.4 of [3]. For detailed discussion of these conditions for Gaussian random fields see Section 3.3 of [3].

The proposed method of lesion detection will be used in the following sections to quantify the quality of a set of simulated PET images in terms of detectability of liver lesions with different contrast against the background.

6. SIMULATION OF PET IMAGES

Computer simulations can be used to generate artificial PET images, mainly by Monte Carlo methods. For this purpose we used the open-source software GATE dedicated to numerical simulations in medical imaging and radiotherapy [14]. It simulates the physical processes that give rise to PET images (radioactive decay, positron-electron annihilation, photon interactions in different materials, etc.).

We prepared model of the Siemens Biograph 40 TruePoint TrueV HD PET/CT scanner. To assess the statistical properties of PET images we modeled a simple cylindrical phantom (artificial object representing the patient's body) with a homogeneous distribution of the radiotracer. The height of the phantom is 250 mm, diameter of its base is 210 mm. Real images of the same phantom from the Siemens Biograph scanner are available for comparison (see Section 7).

We chose liver area as the main region of interest in this study. It is important region for PET imaging with frequent tumor occurrence and it exhibits a homogeneous distribution of ^{18}F-FDG. We want to quantify PET image quality in terms of detectability of simulated liver lesions.

For assessing the relationship between the image quality, the amount of radiotracer applied and the patient's physical parameters we constructed another cylindrical phantom that approximates human abdomen. The phantom's height is 250 mm and its structure can be described as follows: a cylinder with base radius r_1 represents the patient's body; it contains another cylinder with base radius r_2 which represents the internal organs and contains a sphere with radius $r_3 = r_2/2$ which approximates the liver.

This structure gives rise to three different layers of the phantom – liver, other internal organs and subcutaneous adipose tissue. A central section of the phantom which illustrates these layers is shown in Figure 3. By adjusting the values of r_1 and r_2 we can simulate images of patients with different body parameters (most importantly weight). Construction of the phantom is described in more detail in [6].

Distribution of the radiotracer is homogeneous in each layer (liver, internal organs, adipose tissue). The radiotracer concentrations are chosen so that their ratios approximately match the ratios determined empirically from a set of real PET images.

Six spherical lesions were added to the liver area of the phantom for the intended lesion detection experiment. Their radius is 10 mm and they are positioned symmetrically around the liver's center, see Figure 3. The concentration of the radiotracer is different in each lesion and it is higher than the background (liver) concentration, see Section 9.

A method of detecting the lesions in noisy PET images was proposed in Section 4. The method was tested on a set of simulated images of this phantom with different radii
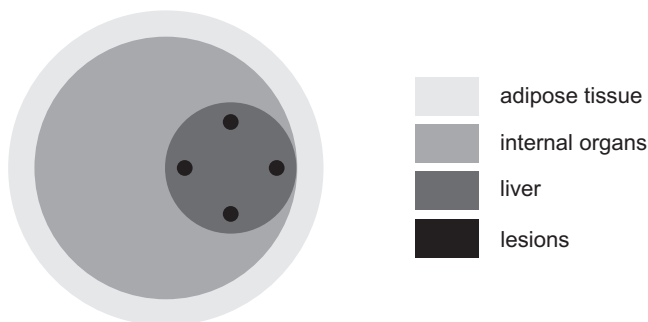
Fig. 3. Central section of the cylindrical phantom approximating
patient's abdomen.

r_1 and r_2 and different amounts of the radiotracer.

Different types of output formats are available in GATE. At this early stage of the research a simple format was chosen. This format includes information about locations of the annihilations which give rise to true and scattered coincidences. The number of random coincidences is also recorded. The result can be viewed as an image after ideal reconstruction which is in practice estimated by intricate reconstruction algorithms.

Also, simple attenuation correction was applied to the images. The correction factors were determined by simulation of a homogeneous phantom with appropriate size. Voxel size was chosen to be $4.07 \times 4.07 \times 5$ mm which is consistent with the voxel size in available real PET images of patients.

7. STATISTICAL PROPERTIES OF PET IMAGES

To assess the properties of PET images we used 15 real images of a cylindrical phantom (diameter 210 mm, height 250 mm) filled with a homogeneous solution of ^{18}F-FDG. These images were acquired during regular quality checks in Na Homolce Hospital, Prague, Czech Republic. After reconstruction the images were smoothed using a Gaussian kernel with FWHM 2 mm. 20 artificial images of the same phantom (not smoothed prior to the analysis) were simulated to compare the statistical properties of real and simulated PET images.

Voxel dimensions in the real phantom images were $5.35 \times 5.35 \times 2.03$ mm. Note that they are different from the voxel dimensions in the real patients images, see Section 6. In the simulated phantom images the voxel dimensions were chosen to be $4 \times 4 \times 4$ mm for simplicity (such a cubic voxel has approx. the same volume as the voxel in real phantom images). Central part of the phantom covering approx. 2000 cm^3 was chosen for the analysis. It also covers parts of the phantom just under its surface as well as those near its base. All images (of the central part) were standardized to have zero mean and unit variance.

To assess the range of correlation in the data we chose a voxel V_0 in the centre of the phantom and estimated the Pearson correlation coefficients of the values observed in V_0

Location	Real data		Simulated data	
	Rejected	Fraction	Rejected	Fraction
centre	0/26	0.00 %	1/26	3.85 %
surface	2/26	7.69 %	0/26	0.00 %
base	3/26	11.54 %	3/26	11.54 %

Tab. 1. Results – testing if the true correlation in neighbouring voxels is zero. In each location 26 tests at 5 % significance level were performed.

and the values observed in voxels shifted from V_0 by different amount in the direction of one of the coordinate axes.

The estimated values of the correlation coefficient do not suggest stronger correlation between pairs of nearby voxels compared to the distant pairs. We performed tests at 5% significance level if the true correlation coefficients are zero. For real images 5 out of 112 tests (4.46 %) rejected the null hypothesis of no correlation, for simulated images 4 out of 93 tests (4.30 %) rejected the null hypothesis. However, the pairs of voxels for which the null hypothesis was rejected do not tend to be close together and are rather uniformly distributed among all pairs. Thus the range of correlation could not be estimated by this analysis – it seems to be smaller than the voxel dimensions. Different number of voxel pairs was analyzed (and number of tests performed) for real and simulated images due to different voxel dimensions in order to cover the same distance between voxels in the direction of each coordinate axis.

We also chose three voxels in different parts of the phantom: V_0 in the centre of the phantom, V_1 near its surface and V_2 near its base. We estimated the Pearson correlation coefficient of values observed in V_i's and their 26 respective neighbours and tested if the true correlation is zero at 5 % significance level. We performed 3 sets of 26 tests separately for real and simulated images.

The results are summarized in Table 1. The number of rejections is higher than expected near the base for both real and simulated data. However, after Bonferroni correction (for 26 tests in one set) no test rejected the null hypothesis. Therefore we consider the values in different voxels uncorrelated.

Considering the Poisson character of the underlying physical processes (decay of unstable radionuclides) one could expect the observed values to be Poisson distributed. Since the Poisson distribution can be approximated by a Gaussian distribution for large values of the mean we performed tests for the normality of distribution.

To test univariate normality of the observed values in individual voxels we chose two different types of tests: Shapiro–Wilk and Anderson–Darling test (see e. g. [20]). For the real data we chose a large rectangular block of voxels fully covered by the phantom ($25 \times 25 \times 65 = 40,625$ voxels). For each voxel we have 15 independent observed values from 15 images and we test the normality of these values. Altogether we performed 40,625 tests for the real images.

Similarly, for the simulated data we selected a block of $32^3 = 32,768$ voxels covering approx. the same volume as the block chosen for the real images. For each voxel we have 20 independent observed values from the 20 simulated images. Altogether we performed 32,768 tests for the simulated images.

At 5 % significance level the null hypothesis of the Gaussian distribution was rejected in approx. 5.2 % of the tests (of both types) performed for different voxels for the real data and approx. 5.6 % for the simulated data. When using the Bonferroni correction no test rejected the null hypothesis. This means that the values observed in individual voxels can be considered normally distributed.

Multivariate normality follows immediately from the univariate normality and the correlation structure of the data (values in different voxels are uncorrelated). Formal test of multivariate normality on a smaller portion of the data was performed. We chose the test proposed in [17]. We tested the null hypothesis of multivariate normality of the vector of 8 neighbouring values in a cube of $2 \times 2 \times 2$ voxels. At 5% significance level the null hypothesis was rejected in approx. 5.9 % of the performed tests for the real data and approx. 4.4 % for the simulated data. When using the Bonferroni correction no test rejected the null hypothesis of multivariate normality.

To use Theorem 5.1 we need also weak stationarity and isotropy of the considered random field. To assess these properties of real and simulated data we estimated, individually for each voxel, the mean value and standard deviation (s.d.) of the recorded values as sample mean and sample s.d. By this we obtained 3D maps of the sample means and sample s.d.'s corresponding to individual voxels. We performed the estimation separately for real and for simulated images of the phantom.

Since the values in different voxels are uncorrelated it is sufficient for the random field to be stationary that the mean values and standard deviations are constant.

Real data exhibit visible spatial trends in mean values despite the fact that the distribution of radiotracer is homogeneous in the whole phantom and attenuation correction was applied during the reconstruction process. Values in voxels near the axis of the cylindrical phantom are lower (on average) than in the voxels closer to surface. On the other hand, simulated data show only small random deviations of the mean values in individual voxels from the overall average without any spatial trends.

Inspection of the standard deviations of the voxel values in simulated data showed mild global trends. Values in voxels near the axis and near the base of the cylinder exhibit higher standard deviations than voxels near the surface. This behavior reflects the different values of the attenuation correction coefficients, see Section 6.

As a result, we cannot consider the real data to be realizations of a stationary random field and we cannot apply the method described in Section 4. In the case of simulated data, the random field in question can be considered stationary only locally (in parts of the phantom where the standard deviations are constant or almost constant) rather than in the whole volume of the phantom.

From the correlation structure of the simulated data we can expect locally isotropic behavior of the random field in the parts of the phantom where we consider it stationary. Informal method for assessing isotropy of a random field observed in discrete lattice points proposed in [28] confirmed that the simulated data can be considered isotropic in smaller parts of the phantom.

Theorem 5.1 assumes that the random field is sufficiently smooth. We can expect that the simulated data require additional smoothing before calculating the expected value of Euler characteristic of an excursion set due to its correlation structure (see [19] for a relevant simulation study in 2D).

We applied isotropic Gaussian filters with different widths to the 20 simulated images of cylindrical phantom to obtain images with different smoothness. Application of the filters to uncorrelated data results in a correlation function of the form $r(h) = \exp\{-\rho||h||^2/2\}$ for some $\rho > 0$. This justifies the use of Theorem 5.1, see Section 5 above. Since the filtering consists of forming linear combinations of the (uncorrelated) values of the random field the normal distribution of the values is not affected by this process.

We calculated mean Euler characteristic of excursion sets E_t of the smoothed simulated data for a wide range of values of threshold t. Estimation of $\chi(E_t)$ from discretely sampled values of the random field was performed using a method described in [26]. Then we compared the mean values of $\chi(E_t)$ with the theoretical values of $\mathbb{E}\left[\chi(E_t)\right]$, see Figure 4.

The results lead to conclusion that the empirical behavior of $\chi(E_t)$ corresponds well to the theoretical expressions for $\mathbb{E}\left[\chi(E_t)\right]$, particularly in the tails, if the simulated data were smoothed by isotropic Gaussian filter with FWHM larger than 5 times the length of voxel edge. This is consistent with the conclusions in [19].

The above analysis showed that the formula (11) can be applied to the simulated data after sufficient smoothing and in smaller parts of the image rather than in the whole volume of the phantom.

8. QUALITY OF SIMULATED PET IMAGES

To determine the relationship of PET image quality, amount of the applied radiotracer and patient's body parameters we simulated images of phantom approximating the abdomen of patients with different weight, see Section 6 and [6] for details. We chose six weights varying from 48.6 kg to 129.6 kg which result in BMI from 15.0 to 40.0 for a patient with height 1.80 m.

For each weight we simulated a set of 15 images with different amount of applied radiotracer. Its activity was chosen randomly between 180 and 530 MBq. This interval covers the activities that would be administered in practice to patients with weights considered above.

To detect simulated lesions in the liver we applied the method using Euler characteristic described in Section 4 and also the method using Bonferroni correction. Measure of quality of the simulated images based on the lesion detectability should take into account two factors: how many lesions were detected by the given method and the degree of certainty (expressed by the number of voxels marked as significant by the given method; the idea is that larger number of significant voxels in a small area brings more information about the presence of a lesion and its location).

To evaluate the quality of simulated images we used the following formula:

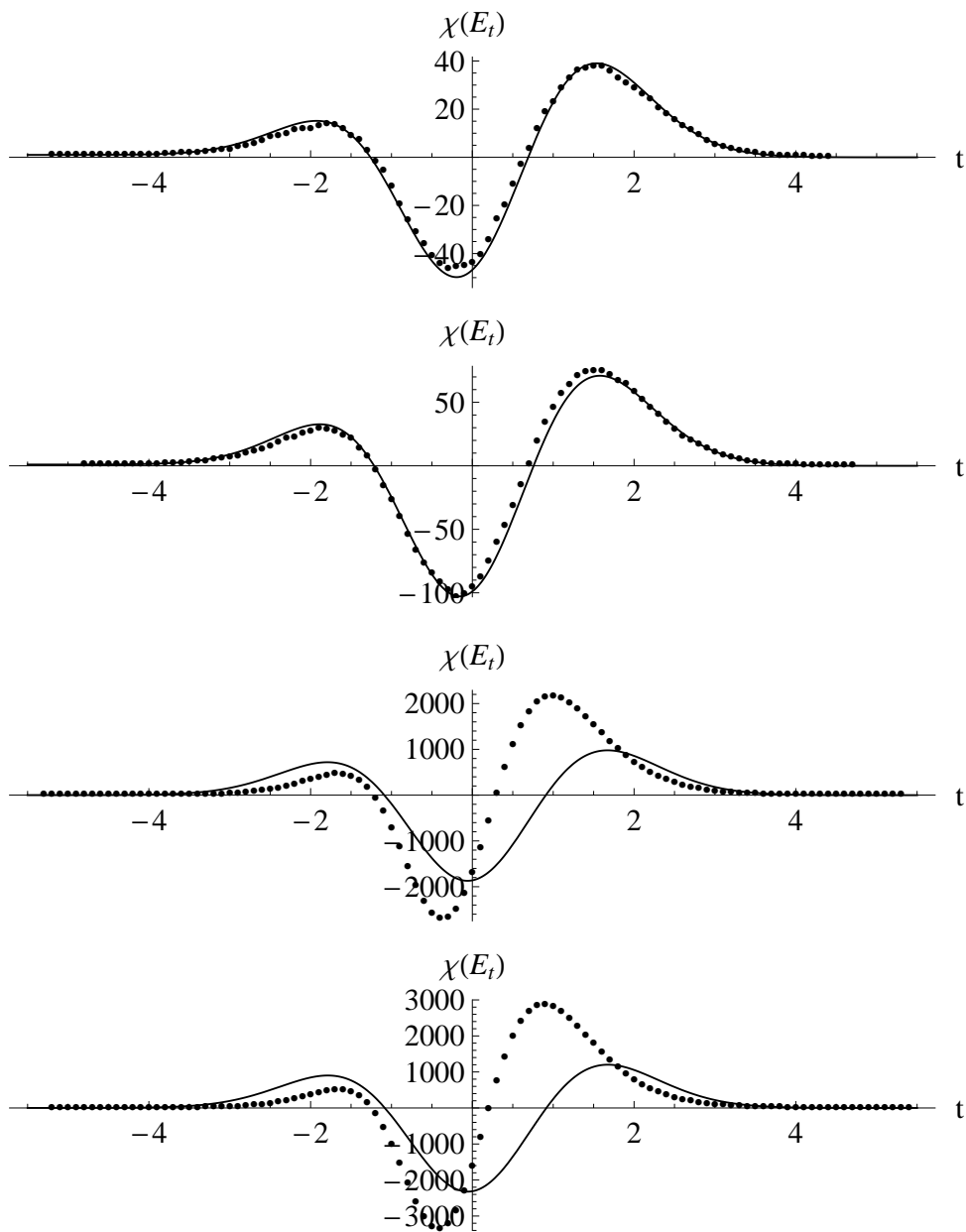$$Q_{m,i} = k \cdot n_{m,i} + v_{m,i}, \qquad m = 1, \ldots, 6, \quad i = 1, \ldots, 15, \tag{15}$$

Fig. 4. Simulated data: comparison of theoretical value of $\mathbb{E}\left[\chi(E_t)\right]$ (solid curve) and mean values $\widehat{\chi}(E_t)$ (points) calculated for different values of threshold t. Top to bottom: simulated images smoothed with isotropic Gaussian filter with FWHM 25 mm, 20 mm, 5 mm and 0 mm (unsmoothed images), respectively.

where $n_{m,i}$ is the number of detected lesions in the ith image in the mth weight category, $v_{m,i}$ is the number of voxels marked as significant and k is the mean number of significant voxels per detected lesion in the whole study (calculated by averaging the values $v_{m,i}/n_{m,i}$, $m = 1, \ldots, 6$, $i = 1, \ldots, 15$, i.e. averaging across images and weight categories). Of course the choice of the constant k is arbitrary and different values put more emphasis on the first or the second factor (number of detected lesions or number of significant voxels). However, it turned out that the above-mentioned choice of k is a good compromise incorporating all the available information.

We also calculated values of NEC in each simulated image according to the formula (1). It is a standard metric of PET image quality in medical practice and it enables us to assess clinical relevance of the results obtained by the previous methods.

It is difficult to find a suitable model for the image quality $Q_{m,i}$ depending on weight and activity of the applied radiotracer. We attempted to fit simple linear additive or multiplicative models to the data (quality of simulated images expressed in terms of lesion detectability). However, these models did not fit the data satisfactorily and more complex model needs to be considered. Instead, we concentrate on the dependence of image quality on the activity for each weight category separately. For a fixed weight category m we fit the following curve to the data by least squares

$$Q_{m,i} = a_m + b_m \cdot (A_{m,i})^{c_m}, \qquad i = 1, \ldots, 15, \qquad (16)$$

where $A_{m,i}$ is the activity of the radiotracer applied in the ith simulation in the weight category m and a_m, b_m and c_m are parameters describing the dependence of image quality on applied activity in this weight category.

Finally, we invert formula (16) to obtain curves of constant quality (CCQ), i.e. curves determining the activity of radiotracer needed to obtain image of the same (given) quality in each weight category. By this procedure we obtain three sets of CCQs, one for quality measure based on Euler characteristic, one for Bonferroni correction and one for NEC.

9. RESULTS

In this section we present the curves of constant quality derived from the simulated images. Note that we cannot do the same for real images – the real images of phantom contain no lesions. Applicability of the method of lesion detection as described in Section 4 to images of patients with possible lesions is to be further investigated.

First we performed simulations with the concentration of radiotracer in the lesions being 120, 135, 150, 165, 180 and 195 % of the background (liver) concentration, respectively. This choice was suitable for using the Bonferroni correction but not for the method based on Euler characteristic (using the latter method all lesions were detected in nearly all images and thus we obtained only little information).

For this reason we repeated the simulations with the same amounts of applied activity and concentrations of the radiotracer in the lesions being 105, 110, 115, 120, 125 and 130 % of the background concentration, respectively. This time the number of lesions detected by the method using Euler characteristic was more affected by the applied activity and weight category. Method of Bonferroni correction showed no trend in the number of detected lesions with increasing activity in each weight category and thus was

not used to express quality of images in this case. Differences between values of NEC calculated in each set of simulations were negligible.

The obtained CCQs are shown in Figure 5 together with the prescription used nowadays in clinical practice which is based on recommendations in [13]. In three cases (the highest weight category for method using Bonferroni correction and two highest weight categories for method using Euler characteristic) the calculated amounts of activity were considerably higher than amounts applicable in medical practice and thus were not taken into account.

It is clear that the shapes of CCQs are qualitatively different from the shape of the curve given by the current prescription (note particularly convexity of most of the CCQs and nearly linear shape of the curve used nowadays). This supports the empirical experience that radiotracer dosage based on [13] results in images of different quality for patients with varying weight and it suggests the possibility of dosage optimization.

To estimate the significance level of the suggested testing procedure (of the null hypothesis that no lesions are present in the image) the method using Euler characteristic was applied also to the set of 90 images with no lesions. Due to the specific structure of the simulation output it was possible to use the same images as in the study described above but with the lesions removed. For 5 of these lesion-free images the null hypothesis was rejected. Thus the estimated significance level of the test is $5.55\,\%$. This is in reasonable agreement with the nominal significance level of $5\,\%$.

The same analysis was performed also for the method using Bonferroni correction. In this case the null hypothesis was rejected for 45 images. This indicates that the assumptions of the method are not fulfilled in this case and it should not be used for hypothesis testing. In particular, the assumption of the same distribution of values in all analyzed voxels turned out to be violated.

The previous statistical analysis was based on properties of values in individual voxels (e.g. for a given voxel (location), normality of the values was assessed based on a sample of values in this voxel in 20 independent images). In the method using Bonferroni correction values in thousands of voxels covering the liver are analyzed in a single image, where the values in different voxels are treated as independent. In a sample of this size even slight deviations from normality (or generally the "common distribution" of the values) are exposed which may have been overlooked in the previous analysis. The method using Euler characteristic requires substantial smoothing, thus reducing the differences in distribution in individual voxel and making the method suitable for hypothesis testing.

Even though the method using Bonferroni correction suffers from this problem and cannot be used to test the hypothesis of no activation in the image, it can be used to describe the image quality as suggested in Section 8. In the case of false positive discovery (i.e. when estimating the significance level on images with no lesions) typically only one or two voxels were marked as significant. On the other hand, tens of voxels were marked as significant in images actually containing lesions. Thus, the total number of significant voxels in this case is affected by the false positives only marginally.

Also, only the number of simulated lesions detected by this method was recorded in the study described above, disregarding the regions of activation that do not coincide with positions of any of the simulated lesions (i.e. disregardning the false positives). As

a consequence the image quality based on the Bonferroni correction, calculated according to (15), is still a useful tool for describing the dependence of the image quality on the patient body parameters and the activity of applied radiotracer.

10. CONCLUSION

We described a method of lesion detection in PET images which takes advantage of important results from the geometry of random fields. The method requires several nontrivial assumptions about the distribution and probabilistic properties of the considered random field.

These assumptions were fulfilled for the set of simulated images. However, the real images of phantom inspected in this study violate the assumption of stationarity even though the distribution of the radiotracer was homogeneous. This is a serious drawback and the method needs to be adapted if it is to be used in the medical practice.

We applied the method to a set of simulated data to quantify the quality of simulated PET images in terms of lesion detectability and to derive curves of constant quality. Other measures of PET image quality were used as well.

The method using Euler characteristic was able to detect simulated lesions with very low contrast against the background. On the other hand it required substantial smoothing (with isotropic Gaussian filter with a FWHM of 25 mm, i. e. 5 times the length of a voxel edge). That means the lesions could not be localized correctly and appreciable number of voxels were marked as significant, even though they are outside the locations of the simulated lesions.

The second method using Bonferroni correction performed well for simulated lesions with higher contrast against background and could be used without any data smoothing. The lesions were localized very precisely because virtually every voxel marked as significant indeed covered (part of) a simulated lesion.

The last measure of PET image quality, NEC, is based on the coincidence counts and not on lesion detection. It was affected only marginally by the contrast of simulated lesions.

Shape of the derived curves of constant quality differs from the shape of the curve given by the current prescription (see Figure 5). We see the aim of dosage optimization for PET imaging in obtaining images of the same quality for different patients, thus making the diagnostic process more consistent. It is clear that patients with low weight and children are currently being given unnecessary amounts of the radiotracer, whereas the overweight and obese patients need a higher dose in order to obtain image of sufficient quality.

This shows that further optimization of the radiotracer dosage is desirable and it has potential to reduce radiation burden to children and patients with low weight and to improve diagnostic value of PET images of overweight and obese patients.
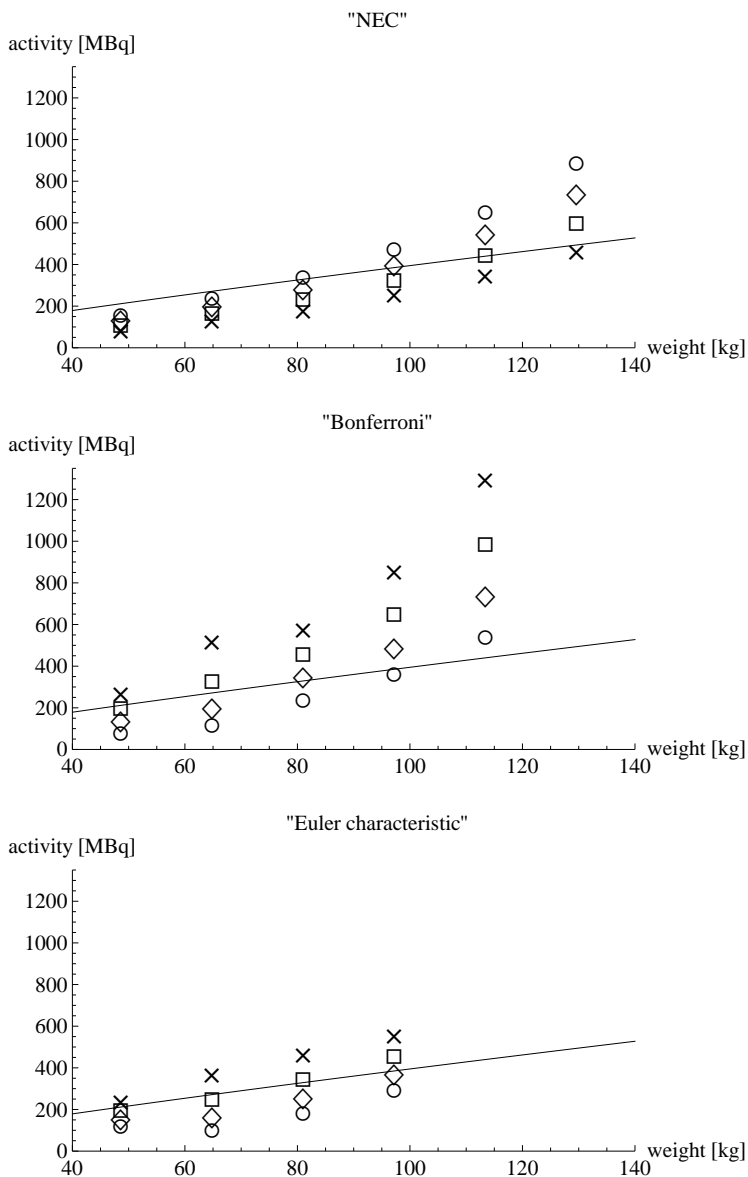
Fig. 5. Derived curves of constant quality for different measures of quality and different levels of required quality. Solid lines – activity injected in Na Homolce Hospital, Prague (EANM pediatric dosage card [13] extrapolated to the whole range of patient weight). Top: quality levels are chosen as $1.5 \cdot 10^6, 1.8 \cdot 10^6, 2.1 \cdot 10^6$ and $2.4 \cdot 10^6$, respectively. Middle: quality levels are chosen as 300, 330, 360 and 390, respectively. Bottom: quality levels are chosen as 1000, 1100, 1200 and 1300, respectively.

R E F E R E N C E S

[1] 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. Amer. A *18* (2001), 473–488.

[2] R. Accorsi, J. S. Karp, and S. Surti: Improved dose regimen in pediatric PET. J. Nucl. Med. *51* (2010), 293–300.

[3] R. J. Adler: The Geometry of Random Fields. Wiley, London 1981.

[4] R. J. Adler and J. E. Taylor: Random Fields and Geometry. Springer, New York 2007.

[5] R. Boellaard et al.: FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. European J. Nucl. Med. Mol. Imaging *37* (2010), 181–200.

[6] J. Boldyš et al.: Monte Carlo simulation of PET images for injection dose optimization. In: Proc. III ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing: VipIMAGE 2011. Taylor and Francis, London 2012.

[7] D. Brasse et al.: Correction methods for random coincidences in fully 3D whole-nody PET: Impact on data and image quality. J. Nucl. Med. *46* (2005), 859–867.

[8] J. Cao and K. J. Worsley: Applications of random fields in human brain mapping. In: Spatial Statistics: Methodological Aspects and Applications. Springer Lecture Notes in Statistics *169* (2001), pp. 169–182.

[9] M. Danna et al.: Optimization of tracer injection for 3D ^{18}F-FDG whole body (WB) PET studies using an acquisition-specific NEC (AS-NEC) curve generation. IEEE Nucl. Sci. Conf. R. (2004), 2615–2619.

[10] H. Everaert et al.: Optimal dose of ^{18}F-FDG required for whole-body PET using an LSO PET camera. European J. Nucl. Med. Mol. Imaging *30* (2003), 1615–1619.

[11] H. C. Gifford et al.: Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. J. Nucl. Med. *41* (2000), 514–521.

[12] B. S. Halpern et al.: Optimizing imaging protocols for overweight and obese patients: a lutetium orthosilicate PET/CT study. J. Nucl. Med. *46* (2005), 603–607.

[13] F. Jacobs et al.: Optimised tracer-dependent dosage cards to obtain weight-independent effective doses. European J. Nucl. Med. Mol. Imaging *32* (2005), 581–588.

[14] S. Jan et al.: GATE: a simulation toolkit for PET and SPECT. Phys. Med. Biol. *49* (2004), 4543–4561.

[15] T. Mizuta et al.: NEC density and liver ROI S/N ratio for image quality control of whole-body FDG-PET scans: comparison with visual assessment. Mol. Imaging Biol. *11* (2009), 480–486.

[16] R. A. Powsner and E. R. Powsner: Essential Nuclear Medicine Physics. Second edition. Wiley-Blackwell, 2006.

[17] G. J. Székely and M. L. Rizzo: A new test for multivariate normality. J. Multivariate Anal. *93* (2005), 58–80.

[18] S. C. Strother, M. E. Casey, and E. J. Hoffman: Measuring PET scanner sensitivity: relat-ing countrates to image signal-to-noise ratios using noise equivalents counts. IEEE Trans. Nucl. Sci. *37* (1990), 783–788.

[19] J. E. Taylor, K. J. Worsley, and F. Gosselin: Maxima of discretely sampled random fields, with an application to 'bubbles'. Biometrika *94* (2007), 1–18.

[20] H. C. Thode: Testing for Normality. Marcel Dekker, New York 2002.

[21] C. C. Watson: Count rate dependence of local signal-to-noise ratio in positron emission tomography. IEEE Trans. Nucl. Sci. *51* (2004), 2670–2680.

[22] C. C. Watson et al.: Optimizing injected dose in clinical PET by accurately modeling the counting-rate response functions specific to individual patient scans. J. Nucl. Med. *46* (2005), 1825–1834.

[23] C. C. Watson, D. Newport, and M. E. Casey: Evaluation of simulation-based scatter correction for 3D PET cardiac imaging. IEEE Trans. Nucl. Sci. *44* (1997), 90–97.

[24] K. J. Worsley et al.: A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. *12* (1992), 900–918.

[25] K. J. Worsley: Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. Adv. in Appl. Probab. *27* (1995), 943–959.

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

[27] K. J. Worsley et al.: Searching scale space for activation in PET images. Hum. Brain Mapp. *4* (1996), 74–90.

[28] K.J. Worsley et al.: Detecting changes in non-isotropic images. Hum. Brain Mapp. *8* (1999), 98–101.

Jiří Dvořák, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
   e-mail: dvorak@utia.cas.cz

Jiří Boldyš, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
   e-mail: boldys@utia.cas.cz

Magdaléna Skopalová, Na Homolce Hospital, PET-Center, Roentgenova 2/37, 150 30 Praha 5. Czech Republic.
   e-mail: magdalena.skopalova@homolka.cz

Otakar Bělohlávek, Na Homolce Hospital, PET-Center, Roentgenova 2/37, 150 30 Praha 5. Czech Republic.
   e-mail: otakar.belohlavek@homolka.cz