



PAPER

Interval-based reconstruction for uncertainty quantification in PET

RECEIVED
26 May 2017REVISED
29 November 2017ACCEPTED FOR PUBLICATION
1 December 2017PUBLISHED
25 January 2018Florentin Kucharczak^{1,2,4}, Kevin Loquin¹, Irène Buvat³, Olivier Strauss¹ and Denis Mariano-Goulart⁴¹ LIRMM, Univ. Montpellier, CNRS, France² Siemens Healthineers, Saint-Denis, France³ Imagerie Moléculaire *In Vivo*, UMR 1023 Inserm/CEA/Université Paris Sud - ERL 9218 CNRS, CEA Service Hospitalier Frédéric Joliot, Orsay, France⁴ Department of Nuclear Medicine, Montpellier University Hospital and PhyMedExp, University of Montpellier, INSERM U1046, CNRS UMR 9214, Montpellier, FranceE-mail: florentin.kucharczak@lirmm.fr**Keywords:** uncertainty quantification, statistical estimation, image reconstruction, non-additive modeling, emission tomography, system matrix**Abstract**

A new directed interval-based tomographic reconstruction algorithm, called non-additive interval based expectation maximization (NIBEM) is presented. It uses non-additive modeling of the forward operator that provides intervals instead of single-valued projections. The detailed approach is an extension of the maximum likelihood—expectation maximization algorithm based on intervals. The main motivation for this extension is that the resulting intervals have appealing properties for estimating the statistical uncertainty associated with the reconstructed activity values. After reviewing previously published theoretical concepts related to interval-based projectors, this paper describes the NIBEM algorithm and gives examples that highlight the properties and advantages of this interval valued reconstruction.

1. Introduction

In emission tomography, and particularly in positron emission tomography (PET), tomographic reconstruction algorithms implemented by medical imaging manufacturers estimate the 3D map of a radio-tracer concentration without systematic estimation of the uncertainty associated with the reconstructed voxel activities. Yet decision procedures based on quantitative analysis of images would benefit from such uncertainty estimates. The methods currently used for comparing activity values in different regions of interest (ROI) are mostly empirical. Associating a confidence interval with each estimated value would enable a statistical comparison based on hypothesis testing. For instance, in the context of patient monitoring, a physician would be able to perform an objective statistical analysis to compare two ROIs in the same image or in two images reconstructed at two different time points.

Uncertainties associated with reconstructed activity values can actually be estimated using numerous approaches. The first set of approaches consists of propagating the known variance of measurements to the variance and covariance of the reconstructed values. This is usually based on the assumption that the measurements follow a Poisson distribution. Methods have been described in Fessler (1996), Qi and Leahy (2000), Stayman and Fessler (2004) and Li (2011) for the penalized likelihood maximization (PLM) algorithm, or equivalently for the maximum *a posteriori* (MAP) algorithm; Barrett *et al* (1994a) and Barrett *et al* (1994b) for the maximum likelihood—expectation maximization (ML-EM) method; Higdon *et al* (1997) and Sitek (2012) for Bayesian estimates; Sitek (2008) for the origin ensembles method and (Soares *et al* 2000, 2005) for block-iterative methods. The second set of approaches does not require any assumption regarding the statistical properties of the measured data. It uses bootstrap resampling to produce replicates that are then used to determine the statistical properties of the activity values (Buvat 2002, Dahlborn 2001, Lartzien *et al* 2010). The common point between these methods is that they usually describe the statistical variability using the variance or standard deviation estimate. As the statistical distribution of the reconstructed image is not necessarily Gaussian, confidence intervals cannot be constructed in a straightforward way from these estimates. Moreover, all of these approaches are computationally demanding and difficult to implement in routine clinical applications.

The approach presented in this paper involves direct estimation of confidence intervals associated with activity values. Note that here we only focus on providing an estimate of the uncertainty induced by statistical variations in the measurements. Other sources of uncertainty (eg, associated with the scatter estimate) are not addressed in this paper. We describe a new reconstruction paradigm in which the error is estimated as part of the reconstruction process. Our algorithm, called non-additive interval based expectation maximization (NIBEM), is an extension of the usual ML-EM algorithm. It uses non-additive modeling of the measuring process to build a forward projection operator that generates intervals instead of precise values. This operator has already been described in Rico *et al* (2009). In the aforementioned paper, the authors stressed that using this kind of projection operator leads to projected intervals whose width is closely correlated with the level of noise of the projected central image. Here we extend the work of Rico *et al* (2009) to make those operators compatible with a multiplicative reconstruction algorithm such as ML-EM, which is, with its variants, the most widespread algorithm in PET clinical routine, by using directed interval arithmetic (Markov 1996). Some very preliminary results were briefly presented in the abstract (Loquin *et al* 2014). In this paper, using analytical simulations and GATE (Jan *et al* 2004) simulations, we show that NIBEM provides a practical approximation of confidence intervals associated with reconstructed values.

2. Radon matrix modeling and imprecision

The approach we present here is formulated for 2D reconstructions. The extension to 3D is straightforward when ignoring border effects due to the limited field of view of the tomograph.

Reconstructing a tomographic image involves inverting a model that projects the values of each pixel of the unknown image to the measured projections. As this projection model is usually supposed to be linear, this involves a system matrix R , also called a Radon matrix, whose (i, j) th element can be interpreted as *the probability* of a photon emitted in the i th pixel to reach the j th detector⁶. Let f be the $n \times 1$ vector of the pixel values of the image to be reconstructed and let p be the $m \times 1$ measurement vector, then the linear projection operator \mathcal{P} based on the system matrix R is defined by:

$$\hat{p}_j = \mathcal{P}(f)_j = \sum_{i=1}^n R_{i,j} f_i. \quad (1)$$

Most reconstruction methods make use of the dual operator of the projection operator called the back-projection operator \mathcal{B} defined by:

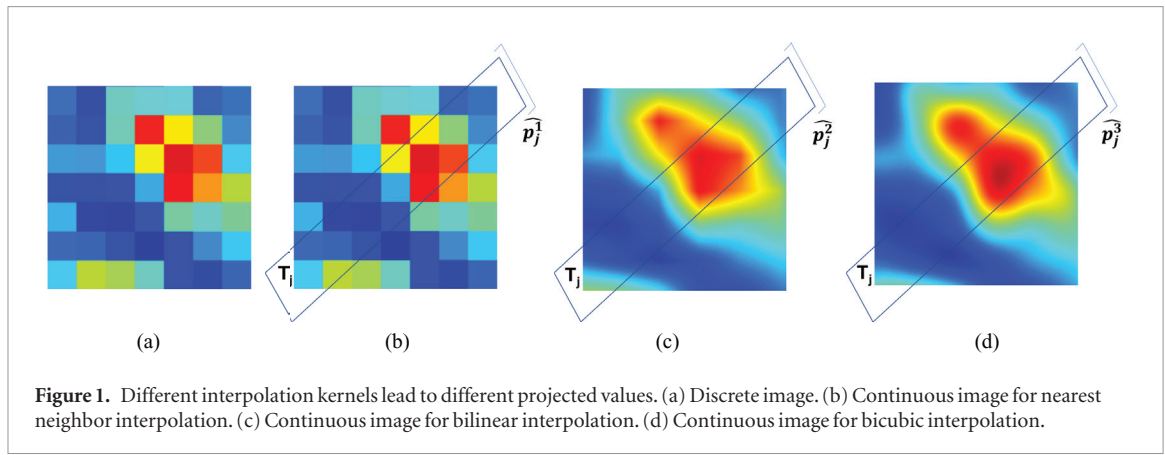
$$\hat{f}_i = \mathcal{B}(p)_i = \sum_{j=1}^m R_{i,j} p_j. \quad (2)$$

Many different models of the system matrix have been proposed in the relevant literature that account for different ways of modeling the interplay (Unser *et al* 1993) between the discrete reconstruction space and the continuous space, where the problem can be formulated. On one hand, continuous-to-discrete interplay is modeled by describing how the detector measures the projected activity. We call this the *measurement model* (MM). On the other hand, discrete-to-continuous interplay involves interpolation kernels. For example, a nearest neighbor interpolation kernel associated with a Dirac MM leads to the line-length model (Wernick and Aarsvold 2004), while rectangular modeling of the measurement with the same interpolation leads to the concave disk model (Huesman *et al* 1977).

In this section, we highlight how the choice of the interplay modeling between the discrete image and its underlying continuous image impacts the resulting projected values. For the sake of simplicity, we choose a rectangular MM to illustrate the proposed method. More complex MM, e.g. taking the detector point spread function (PSF) into account, could also be used with the new framework we present here. Projecting the activity value associated with each pixel on the j th dixel requires computation of the value p_j as a weighted sum (1). This weighted sum involves an interpolation kernel κ . Figure 1 illustrates the fact that using the same rectangular MM with three different interpolation kernels leads to three different projected values (\hat{p}_j^1 , \hat{p}_j^2 and \hat{p}_j^3), i.e. a projected value highly depends on the choice of the interpolation method.

In Rico *et al* (2009), the authors assumed that the appropriate model (i.e. the appropriate kernel) to be used to ensure the discrete-to-continuous interplay is usually unknown and therefore used a convex set \mathcal{K} of bounded interpolation kernels. They thus propose to use this convex set \mathcal{K} to define a new interval-valued projection operator that can be expressed as: $\bar{p}_j = [\min\{p_j^{\kappa \in \mathcal{K}}\}; \max\{p_j^{\kappa \in \mathcal{K}}\}]$. One appealing property is that, since the spread of the resulting intervals reflects the consensus between the different possible projections, it is thus linked to local variations in the projections. As pointed out in Loquin *et al* (2010), poor consensus in the projected values, i.e. a wide spread in the intervals, is a marker of the effect of noise in the projections.

⁶by analogy with the term *pixel* (for picture element) we call *dixel* a detector element.



Computing this interval-valued projection can appear to be computationally complex since no pair of interpolation kernels can provide the extreme values of every set \bar{p}_j . A Monte-Carlo-like approach could be used, however, this would be very time consuming and would only provide an inner approximation of the two extreme points. The solution proposed in Rico *et al* (2009) uses the notion of concave capacities. We explain in section 3 how this solution can be used to define an interval valued projection operator that accounts for the supposed imprecise knowledge on this appropriate interpolation kernel.

3. Non-additive construction of the forward projection operator

This section explains how a concave capacity (Denneberg 1994) can be used to model imprecise knowledge regarding the most appropriate interpolation model (Rico *et al* 2009). As previously, we choose a rectangular MM for didactic purposes and show that the method can be seen as a non-additive extension of the concave disk model.

3.1. A concave capacity to define the geometric interaction between pixels and dexels

Let Ω be a reference set (e.g. $\{1, \dots, n\}$). A concave capacity ν is a non-additive confidence measure, i.e. a weight associated with any subset of Ω . It extends the notion of probability measure since it does not follow the additive axiom, i.e. $\forall A, B \subseteq \Omega, \nu(A) + \nu(B) \geq \nu(A \cup B) + \nu(A \cap B)$, while a probability measure P would verify $P(A) + P(B) = P(A \cup B) + P(A \cap B)$: a probability measure is an additive confidence measure. Let $\Lambda(\Omega)$ be the set of all probability measures on Ω . It has been shown in Denneberg (1994) and Schmeidler (1986) that a concave capacity ν encodes a convex subset of $\Lambda(\Omega)$ called the core of ν and is denoted $\mathcal{M}(\nu)$:

$$\mathcal{M}(\nu) = \{P \in \Lambda(\Omega) \mid \forall A \subseteq \Omega, P(A) \in [\nu^c(A), \nu(A)]\}, \quad (3)$$

where ν^c is the dual capacity of ν defined by: $\forall A \subseteq \Omega, \nu^c(A) = \nu(\Omega) - \nu(A^c)$, with A^c being the complementary set of A in Ω . By construction, ν^c is convex (Rico and Strauss 2010) (i.e. $\nu^c(A) + \nu^c(B) \leq \nu^c(A \cup B) + \nu^c(A \cap B)$).

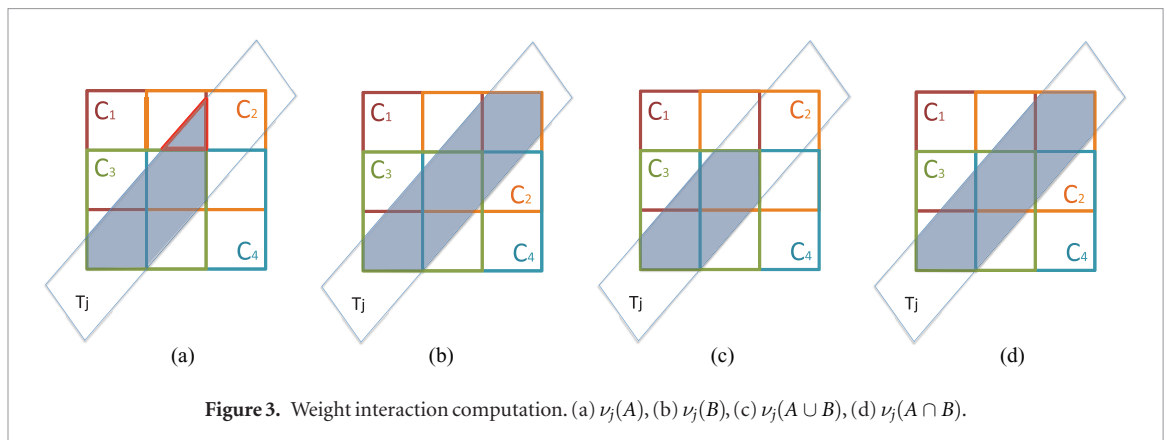
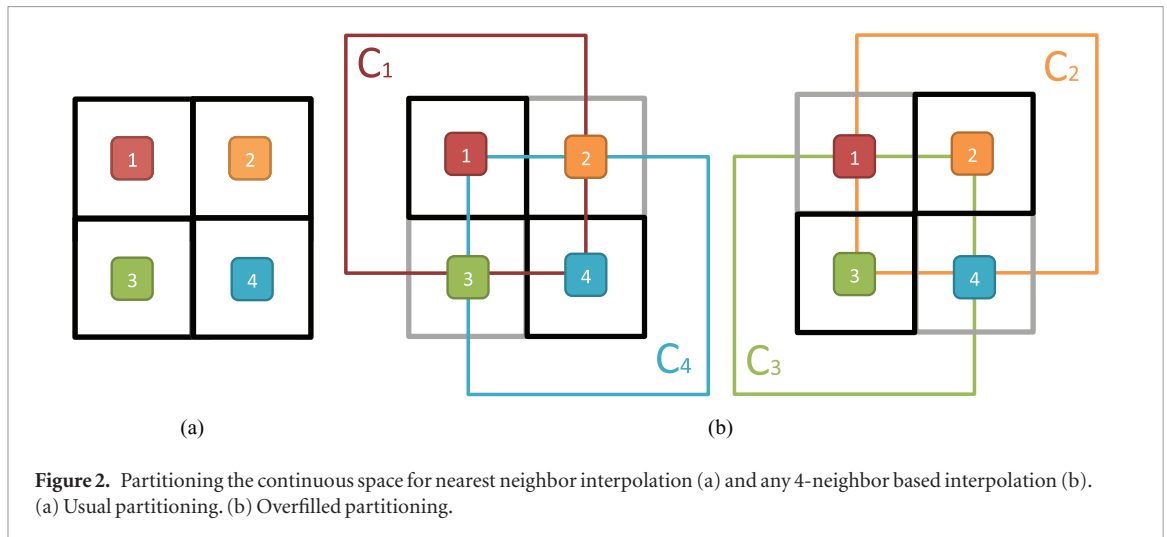
Let us now call T_j the tube of response (TOR) associated with the j th dexel of the measurement vector (i.e. the MM is assumed to be a rectangular function). Then, the confidence measure (i.e. the weight) associated with the interaction between the i th pixel and the j th dexel can be expressed as:

$$R_{ij} = |C_i \cap T_j|, \quad (4)$$

where $|\cdot|$ is the Lebesgue measure. This is the concave disk model. Let A be a subset of Ω . The confidence measure $P_j(A)$ associated with the interaction between A and the j th dexel is defined by $P_j(A) = \sum_{i \in A} R_{ij}$. The thus-defined confidence measure is additive, since $\forall i \neq i', C_i \cap C_{i'} = \emptyset$.

Let us consider the four pixel discrete image depicted in figure 2(a) with $\Omega = \{1, 2, 3, 4\}$. With the usual (additive) modeling of the interaction between pixels and dexels, the domain C_i associated with the i th pixel is obtained by partitioning the continuous domain as illustrated in figure 2(a). With this partitioning, $\forall i \neq i', C_i \cap C_{i'} = \emptyset$ and $\bigcup_{i \in \Omega} C_i$ covers the entire continuous domain. Each domain C_i can be considered as being generated by the rectangular kernel κ_r that defines the nearest neighbor interpolation: let ω_i be the center of C_i , then $C_i = \{\omega \in \mathbb{R}^2, \kappa_r(\omega - \omega_i) = 1\}$. Thus, the continuous value $\hat{f}(\omega)$ associated with a location ω in the continuous domain is defined by $\hat{f}(\omega) = f_i$, where i is the only value of Ω such that $\omega \in C_i$ (i.e. $\kappa_r(\omega - \omega_i) = 1$).

We now show how a concave capacity can be used to compute the bounds of the j th interval-valued projection \bar{p}_j defined in section 2. The approach proposed in Rico *et al* (2009) allows us to define an interval-valued extension of the concave disk model. Let us replace the usual partitioning of figure 2(a) by a partitioning including overlaps as depicted in figure 2(b). We still have $\bigcup_{i \in \Omega} C_i$ covering the entire continuous domain, but now



$\exists i \neq i', C_i \cap C_{i'} \neq \emptyset$. This partitioning can be considered as being generated by a rectangular kernel η_r that is larger than κ_r . This leads to an imprecise extension of the nearest neighbor interpolation: the interpolated continuous activity values at location ω is interval-valued and defined by $\bar{f}(\omega) = [\min_{i \in \mathcal{N}(\omega)} \{f_i\}; \max_{i \in \mathcal{N}(\omega)} \{f_i\}]$, where $\mathcal{N}(\omega) \subseteq \Omega$ is the set of neighbors of ω defined by: $\mathcal{N}(\omega) = \{i \in \Omega \mid \eta_r(\omega - \omega_i) = 1\}$. $\bar{f}(\omega)$ is the set of all the values that can be obtained by a normalized weighted sum of the nearest neighbors of ω . The proposed model no longer coincides with the concave disk model but defines a concave capacity ν_j that can be considered as its imprecise extension. The core $\mathcal{M}(\nu_j)$ of ν_j contains any classical (Radon-matrix based) model that could have been defined by considering any four-neighbor weighted based linear interpolation operator.

We now graphically illustrate that the above defined capacity is a concave capacity (this property has been proved in Rico *et al* (2009)). Let us consider $A = \{1, 3\}$ and $B = \{2, 3\}$ and the j th TOR T_j illustrated in figure 3. Computation of the values $\nu_j(A)$, $\nu_j(B)$, $\nu_j(A \cup B)$ and $\nu_j(A \cap B)$ is based on the Lebesgue measures of the blue areas depicted in figures 3(a)–(d).

In this particular example,

$$\nu_j(A) + \nu_j(B) \geq \nu_j(A \cup B) + \nu_j(A \cap B). \quad (5)$$

It is easy to check that the difference between $\nu_j(A) + \nu_j(B)$ and $\nu_j(A \cup B) + \nu_j(A \cap B)$ is the area of the triangle with red borders in figure 3(a).

3.2. Interval-based forward projection operator construction

A concave capacity, as defined in section 3.1, can be used to define an interval-valued projection operator as defined in section 2. This is illustrated here using a concave disk model but can be performed with any other model constructed in the same way (i.e. involving an MM and an interpolation kernel).

The conventional projection of pixel activities on the j th dixel (1) can be rewritten as:

$$\hat{p}_j = \mathcal{P}(f)_j = \mathbb{E}_{P_j}(f), \quad (6)$$

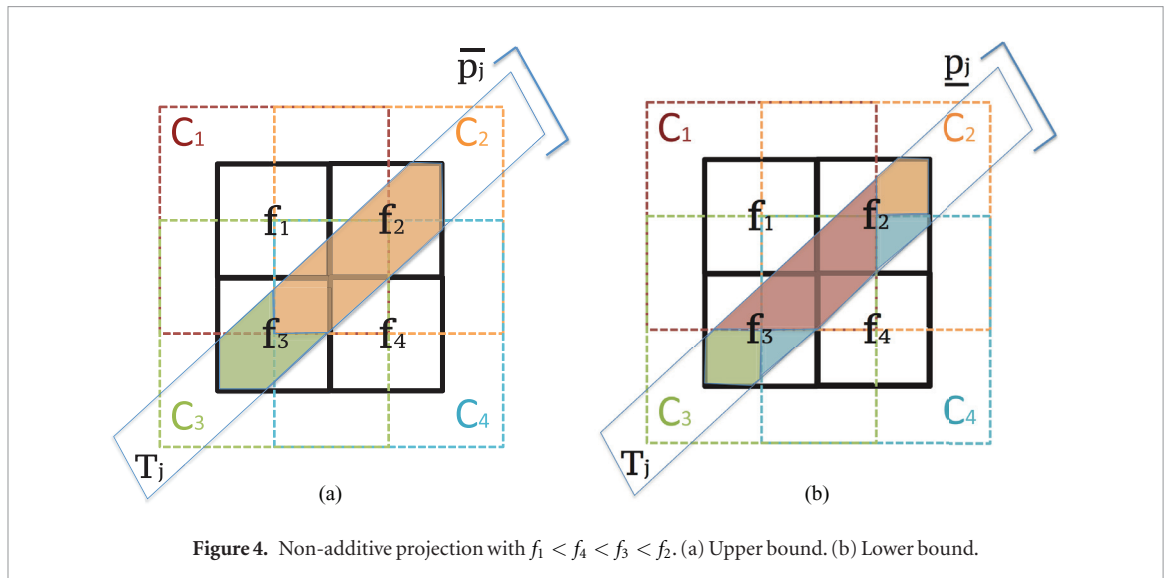


Figure 4. Non-additive projection with $f_1 < f_4 < f_3 < f_2$. (a) Upper bound. (b) Lower bound.

with \mathbb{E} being the linear expectation operator and P_j being the j th additive confidence measure associated with the Radon matrix defined by equation (4).

Now let us consider a concave capacity measure ν_j as defined in section 3.1 with an overlapping partition as depicted in figure 2(b). The concave capacity theoretical framework allows us to keep the computational complexity as low as possible. It consists of replacing the linear expectation operator \mathbb{E} by its interval-valued extension $\overline{\mathbb{E}}$, as defined in Rico and Strauss (2010), based on the use of the Choquet integral (Denneberg 1994):

$$[\underline{p}_j, \overline{p}_j] = \overline{\mathcal{P}}(f)_j = \overline{\mathbb{E}}_{\nu_j}(f) = [\underline{\mathbb{E}}_{\nu_j}(f), \overline{\mathbb{E}}_{\nu_j}(f)], \quad (7)$$

with $\underline{\mathbb{E}}_{\nu_j}(f) = \mathbb{C}_{\nu_j^c}(f)$ and $\overline{\mathbb{E}}_{\nu_j}(f) = \mathbb{C}_{\nu_j}(f)$, $\mathbb{C}_{\nu_j}(f)$ being the discrete Choquet integral with respect to the capacity ν defined by:

$$\mathbb{C}_{\nu_j}(f) = \sum_{i=1}^n f_{(i)} (\nu_j(A_{(i)}) - \nu_j(A_{(i+1)})), \quad (8)$$

where (\cdot) is the permutation of n pixel values, such that $f_{(1)} \leq \dots \leq f_{(n)}$. $A_{(i)} = \{(i), \dots, (n)\}$ is the set of pixels whose values are greater than or equal to $f_{(i)}$. By convention $A_{(n+1)} = \emptyset$. $\overline{\mathcal{P}}$ defined by (7) is an interval-valued extension of the projection operator defined by equation (1).

The fact that the obtained interval $[\underline{p}_j, \overline{p}_j]$ is exactly the interval of all projected values that could have been obtained by considering the convex set of all four-neighbor based linear interpolation kernels, derives from the property proved in Rico and Strauss (2010): if $P \in \mathcal{M}(\nu_j)$, then $\mathbb{E}_P(f) \in \overline{\mathbb{E}}_{\nu_j}(f)$ and conversely if $g \in \overline{\mathbb{E}}_{\nu_j}(f)$, then $\exists P \in \mathcal{M}(\nu_j)$ such that $g = \mathbb{E}_P(f)$.

Moreover, it has been proven in Strauss and Rico (2012) that this imprecise projection operator can easily be extended to an interval-valued function $[f] = [f, \overline{f}]$. We can thus define the imprecise projection of an imprecise image by:

$$[\underline{p}_j, \overline{p}_j] = \overline{\mathcal{P}}([f])_j = \overline{\mathbb{E}}_{\nu_j}([f]) = [\underline{\mathbb{E}}_{\nu_j}(f), \overline{\mathbb{E}}_{\nu_j}(\overline{f})]. \quad (9)$$

Finally we denote $\overline{\mathcal{P}}([f])$ the interval-valued vector of all the interval-valued projections of the interval-valued image $[f]$. This vector pools all projections that would have been obtained by using all possible four-neighbor based interpolations.

3.3. Calculation of a non-additive projection

This section gives an example of how the Choquet integral defined by equation (8) can be used to compute the upper projection of a four-pixel image. The capacity that we use in this example is the imprecise extension of the concave disk model presented in section 3.1. The different values of the capacity are illustrated in figure 4. The intersection of each pixel group with the TOR T_j is calculated as a pseudo-code and represented by the same color in figure 4 and procedure 1. The output of procedure 1 is the upper bound \overline{p}_j of the interval-valued estimate. The lower bound can be computed in a similar way.

Procedure 1. Computation of the upper projection \bar{p}_j .

Input: Image f considering $f_1 < f_4 < f_3 < f_2$

$$\nu_j(A_{(1)}) = | (C_1 \cup C_4 \cup C_3 \cup C_2) \cap T_j |$$

$$\nu_j(A_{(2)}) = | (C_4 \cup C_3 \cup C_2) \cap T_j |$$

$$\nu_j(A_{(3)}) = | (C_3 \cup C_2) \cap T_j |$$

$$\nu_j(A_{(4)}) = | C_2 \cap T_j |$$

$$\begin{aligned} \bar{p}_j &= \mathbb{C}_{\nu_j}(f) = \sum_{i=1}^4 f_{(i)} (\nu_j(A_{(i)}) - \nu_j(A_{(i+1)})) \\ &= f_1 (\nu_j(A_{(1)}) - \nu_j(A_{(2)})) + f_4 (\nu_j(A_{(2)}) - \nu_j(A_{(3)})) + f_3 (\nu_j(A_{(3)}) - \nu_j(A_{(4)})) \\ &\quad + f_2 (\nu_j(A_{(4)})) \\ &= 0 + 0 + f_3 (| (C_3 \cup C_2) \cap T_j | - | C_2 \cap T_j |) + f_2 | C_2 \cap T_j | \end{aligned}$$

Output: \bar{p}_j

4. Generalization of ML-EM to intervals

The NIBEM algorithm can be seen as an interval-valued extension of the ML-EM iterative algorithm (Shepp and Vardi 1982). Each iteration of this reconstruction algorithm can be written as:

$$f_i^{k+1} = f_i^k \times \frac{1}{\sum_{j=1}^m R_{i,j}} \times \sum_{j=1}^m \frac{P_j}{\sum_{l=1}^n R_{l,j} f_l^k} R_{i,j}, \quad (10)$$

where f^k is the image reconstructed at iteration k and \times stands for the term by term product.

To facilitate the extension of ML-EM to intervals, let us rewrite the iterative scheme of equation (10) as follows:

$$f^{k+1} / f^k = \mathcal{B}^* \left(\frac{P}{\mathcal{P}(f^k)} \right) \quad \text{with} \quad \mathcal{B}^*(e) = \frac{1}{\mathcal{B}(\mathbf{1})} \times \mathcal{B}(e), \quad (11)$$

where $/$ stands for the term by term division and \mathcal{B}^* is a normalized back-projection operator defined for any vector e of \mathbb{R}^m , with $\mathbf{1}$ being the unit vector of \mathbb{R}^m . Thus $f^{k+1} = f^k \times \mathcal{B}^* \left(\frac{P}{\mathcal{P}(f^k)} \right)$ is the iterative solution of the iterative gradient descent expressed by equation (11). In this solution, the role of the back-projection operator is to transfer the error from the projection space to the image space. Its normalization ensures energy conservation throughout the reconstruction process.

4.1. Directed intervals

The ML-EM tomographic reconstruction algorithm requires vector element-wise multiplications and divisions. Since the operator \mathcal{P} is replaced by its imprecise extension $\bar{\mathcal{P}}$, extending the MLEM algorithm requires interval-valued arithmetical operations.

Like the extension proposed in Strauss *et al* (2009), we worked in the framework of directed intervals (Markov 1996) to ensure the consistency of this extension. A directed interval $[a] = [\underline{a}, \bar{a}]$ is an interval whose bounds are not always properly directed. An interval is properly directed if $\underline{a} \leq \bar{a}$. Such an interval is called a *real interval*. An interval is not properly directed if $\underline{a} > \bar{a}$.

Since the values involved in the reconstruction process are always positive, we considered positive intervals. In the directed interval framework, each arithmetic operation has two Minkowsky extensions.

The first Minkowsky extensions of product (\otimes) and division (\oslash) are defined for any strictly positive intervals $[a] = [\underline{a}, \bar{a}]$ and $[b] = [\underline{b}, \bar{b}]$ by

$$\begin{aligned} [a] \otimes [b] &= [\underline{ab}, \bar{ab}], \\ [a] \oslash [b] &= [\underline{a/\bar{b}}, \bar{a/b}]. \end{aligned} \quad (12)$$

If $[a]$ and $[b]$ are proper intervals, $[a] \otimes [b]$ and $[a] \oslash [b]$ are also proper intervals. Moreover $[a] \oslash [b] \otimes [b] = [\underline{ab/\bar{b}}, \bar{ab/b}] \neq [a] \otimes [b] \oslash [b] = [\underline{a/\bar{b}}, \bar{a/b}] \neq [a]$ in general, and the following inclusion is always true: $[a] \subseteq ([a] \oslash [b] \otimes [b])$.

The second Minkowsky extensions of product (\boxtimes) and division (\boxdiv) are defined by:

$$\begin{aligned} [a] \boxtimes [b] &= [a] \otimes [b]_- = [\underline{a\bar{b}}, \bar{ab}], \\ [a] \boxdiv [b] &= [a] \oslash [b]_- = [\underline{a/\bar{b}}, \bar{a/b}]. \end{aligned} \quad (13)$$

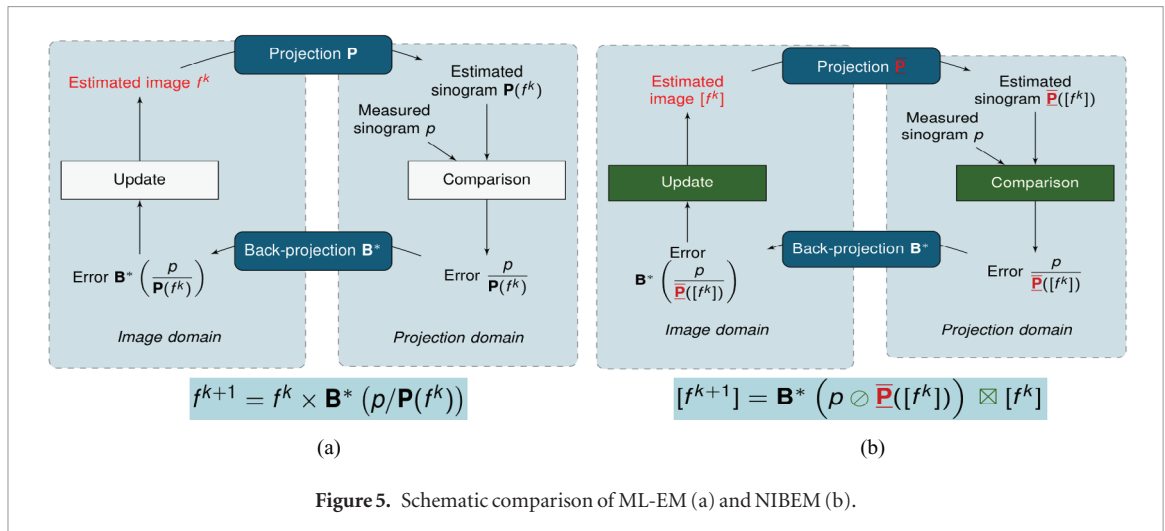


Figure 5. Schematic comparison of ML-EM (a) and NIBEM (b).

Those operators are built to be the dual operators of the first extensions. We have: $[a] \oslash [b] \boxtimes [b] = [a]$ and $[a] \otimes [b] \boxdot [b] = [a]$. Those operators do not fit in the real interval framework since, if $[a]$ and $[b]$ are proper intervals, $[a] \boxtimes [b]$ and $[a] \boxdot [b]$ are not always proper intervals.

Thus, in the directed interval framework, the following interval-valued equations:

$$[x] \otimes [a] = [b], \quad (14)$$

$$[x] \oslash [a] = [b], \quad (15)$$

where $[a]$ and $[b]$ are given and $[x]$ is the unknown, always have a solution:

$$\begin{aligned} [x] \otimes [a] = [b] &\implies [x] = [b] \boxtimes [a], \\ [x] \oslash [a] = [b] &\implies [x] = [b] \boxdot [a]. \end{aligned} \quad (16)$$

Interpreting those solutions is not straightforward. In the positive real interval framework, let $[a]$ and $[b]$ be two positive real intervals and let us consider the solution of equation (14). If $[x] = [b] \boxtimes [a]$ is proper, then $\forall a \in [a], \exists b \in [b], \forall x \in [x]$ such that $a \cdot x = b$. If $[x]$ is improper, then $\exists a \in [a], \forall b \in [b], \forall x \in [x]_-$ such that $a \cdot x = b$, $[x]_-$ is the dual of $[x]$ defined by $[x]_- = [\bar{x}, x]$. When $[x]$ is proper $[x]_-$ is improper and vice versa. These semantics were formalized in Gardenes *et al* (2001). All those Minkowsky extensions of product and division can easily be extended to interval-valued vectors like any other arithmetic operators (see e.g. Strauss and Rico (2012)).

4.2. NIBEM algorithm

NIBEM can be viewed as a straightforward interval-based extension of the iterative ML-EM algorithm corresponding to equation (11). Shifting directly from the precise framework to an interval-valued framework using the projection \bar{P} and back-projection B^* operators, (15) gives:

$$[f^{k+1}] \oslash [f^k] = B^*\left(p \oslash \bar{P}([f^k])\right), \quad (17)$$

where applying the linear operator B^* to the interval $(p \oslash \bar{P}([f^k]))$ simply consists of applying this operator to each of its bounds (see figure 5).

Solving equation (17) is similar to solving equation (15). The final expression of the NIBEM iterative scheme is:

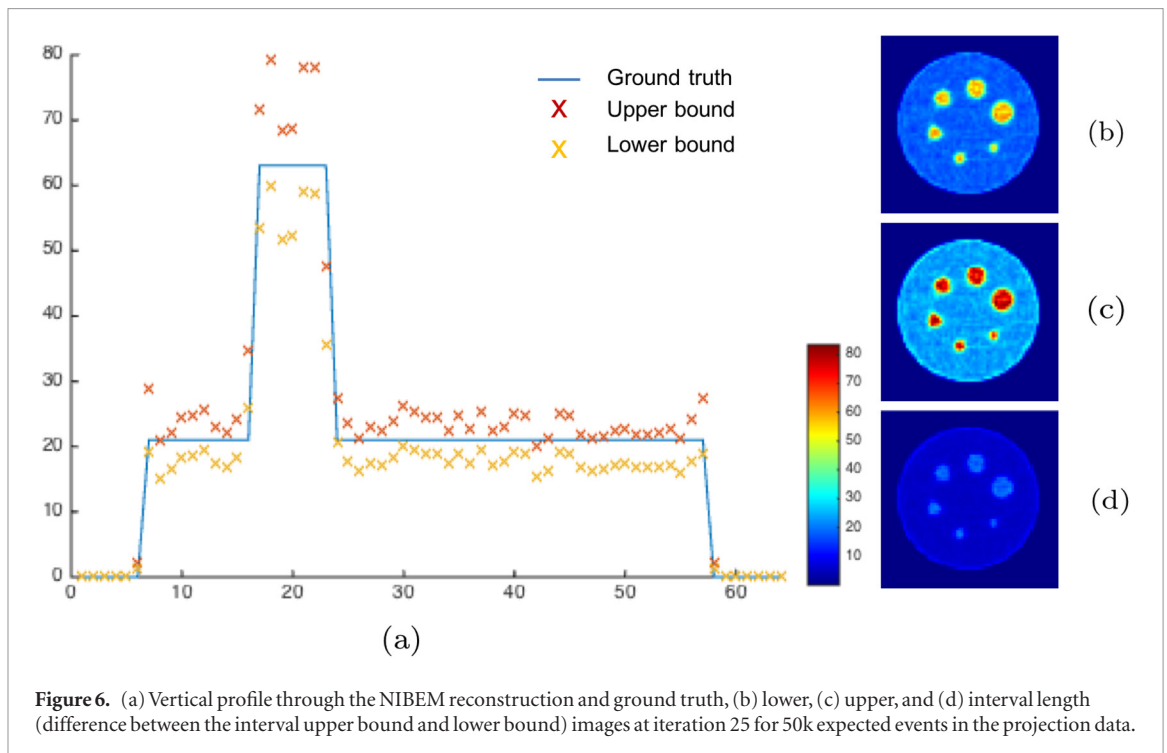
$$[f^{k+1}] = B^*\left(p \oslash \bar{P}([f^k])\right) \boxtimes [f^k], \quad (18)$$

where p are the measured projections and k the current iteration. Note that, even if image $[f^0]$ used for initialization is precise, image $[f^1]$ is likely to be imprecise.

5. Experimental assessment of NIBEM

This section aims at testing the ability of the NIBEM algorithm to estimate the statistical variability of the measurements in the reconstructed images.

The first experiment, based on simulated statistical repetitions, aims at testing whether intervals produced by the NIBEM algorithm could be considered as statistical confidence intervals (Li 2011). This experiment, presented in section 5.1, uses noisy realizations of a simulated Jaszczak-like phantom. The second



experiment, presented in section 5.2, aims at comparing the interval-based statistical estimation produced by NIBEM to the one relying on the bootstrap approach described in Buvat (2002). Finally, the third experiment, presented in section 5.4, aims at investigating the estimated uncertainty as a function of the number of counts.

5.1. NIBEM intervals as confidence intervals

5.1.1. Simulation setup

A modified Jaszczak-like phantom with a similar setup like that used in Li (2011) was used. It consisted of a 2D uniform disk of diameter 160 mm including six hot regions with diameters of 9.5, 11.1, 12.7, 15.9, 19.1 and 25.4 mm. The hot region concentration was 3-fold greater than the concentration of the background. The phantom was digitalized into a 64×64 image with 3.125 mm pixel size. The sinograms were simulated with 64 linearly sampled detector bins and 64 angular views evenly spaced over 180° . The Radon matrix was computed using the rectangular MM model, implemented as described in Fessler (1995). Photon attenuation and scatter were not simulated. Three different noise levels (high, medium and low) were simulated using a Poisson random generator on noise-free projections (50 k, 250 k and 1250 k expected events in the projection data). Typical lower, upper, interval length reconstructed images and vertical profile through NIBEM reconstruction are presented in figure 6.

5.1.2. Data analysis

We performed NIBEM reconstructions of $q = 1000$ simulations for the three noise levels. For each pixel i , we computed how many times the true activity A_i was included in the r th estimated interval $[f_i^r]$ corresponding to pixel i . The Confidence Level (CL) of the interval corresponding to the i th pixel is defined by: $CL_i = \frac{1}{q} \sum_{r=1}^q \mathbb{1}_{[f_i^r]}(A_i)$. The mean CL of a region (hot regions or background) is estimated by averaging the CL of all pixels belonging to it.

5.1.3. Results

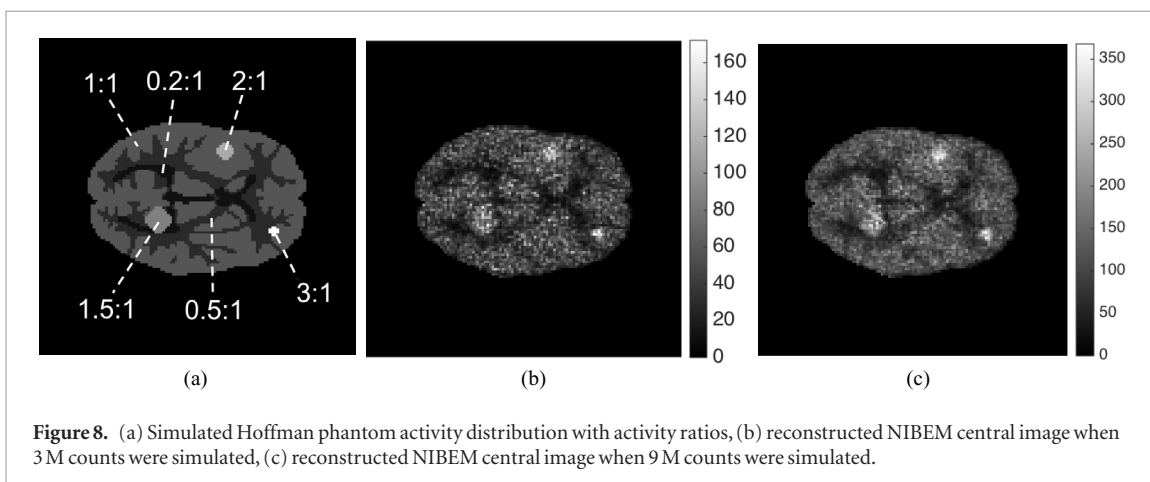
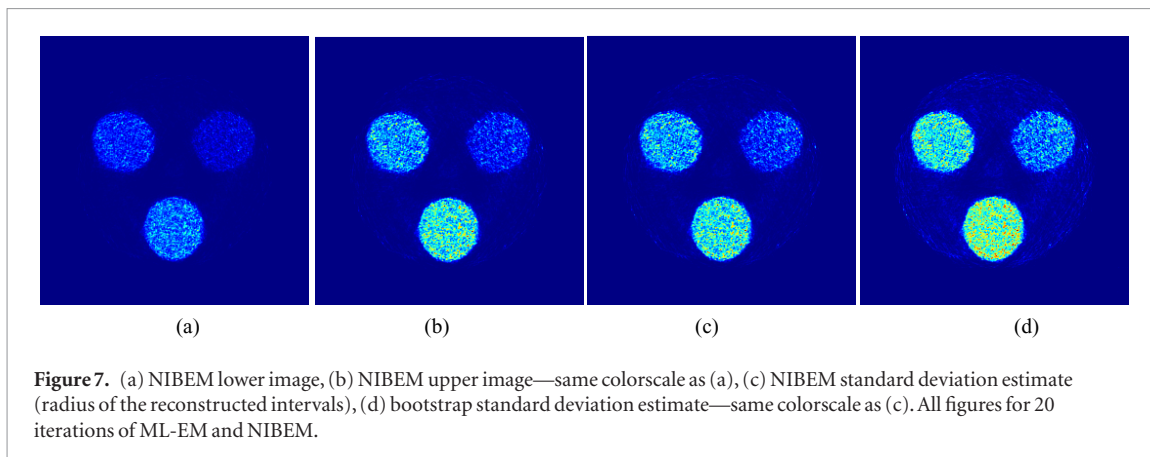
Table 1 shows the mean CL of all interval-valued activity in each of the two considered regions (background and hot spots). CL of the background and the hot regions was close to 0.9. Whatever the noise level, the confidence level did not depend on the reconstructed value or on the level of noise. This experiment showed that the reconstructed intervals could be considered as having the same CL.

5.2. NIBEM intervals as a statistical variability estimator

This experiment aimed at testing whether the information contained in the interval-valued images reconstructed by the NIBEM algorithm could be considered as a statistical variability estimator or not. The NIBEM image statistical properties were compared with those obtained using the bootstrap approach (Buvat 2002).

Table 1. Quantitative assessment of NIBEM—the mean CL is the average CL value over all hot regions or over the whole background region in the phantom image.

Region	Background			Hot regions		
	High	Medium	Low	High	Medium	Low
Noise level	High	Medium	Low	High	Medium	Low
Mean CL	0.868	0.897	0.899	0.919	0.932	0.937



5.2.1. Cylinder phantom

PET data were simulated using a GATE model of the Siemens Biograph PET scanner (Jan *et al* 2004). The PET acquisition of a 25 cm high cylindrical phantom (50 cm diameter) including three smaller cylinders (16 cm high, 8 cm in diameter) filled of water was simulated. Each cylinder was simulated with different fluoro-deoxyglucose (^{18}F FDG) concentrations. The activity ratios in the three cylinders $c_{1,2,3}$ with respect to c_1 were 2:1 (200 kBq for c_2) and 3:1 (300 kBq for c_3). The activity ratio of the background cylinder c_{bg} was 0.5:1 with respect to c_1 . Only one cross-section was considered for this experiment and only coincidences located within that cross-section were registered. Scatter was not simulated. An acquisition of 180 s was simulated and detected coincidences (about 217 k) were sorted to obtain 30 sub-sinograms, where each sub-sinogram corresponded to a 6 s acquisition. For the bootstrap approach, we created 15 k bootstrap sub-sinograms based on the 30 sub-sinograms of 6s, that we grouped 30 by 30 to obtain 500 bootstrap sinograms (Buvat 2002). Sinograms had 276 bins of projections and 276 angular views. The reconstructed images were 276×276 . As visual inspection of the ML-EM and NIBEM reconstructed images suggested a similar convergence rate of the two algorithms, an implicit regularization was obtained by stopping ML-EM and NIBEM iterative processes at the same 20 iteration for the cylinder phantom reconstruction. A typical reconstructed slice is presented in figure 7.

5.2.2. Hoffman phantom

To test the ability of the proposed method to estimate statistical variability in more complex objects, we performed a similar experiment on 128×128 sinograms of a slice of a Hoffman phantom simulated using GATE (figure 8(a)). For the bootstrap approach, we created two series of 15 k bootstrap sub-sinograms based on the 30 sub-sinograms of approximately 100 k and 300 k counts. We grouped those sub-sinograms 30 by 30 to obtain

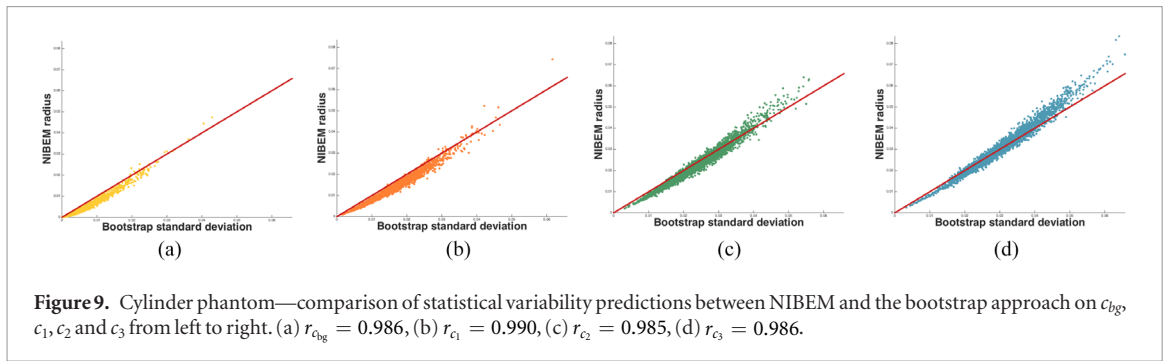


Figure 9. Cylinder phantom—comparison of statistical variability predictions between NIBEM and the bootstrap approach on c_{bg} , c_1 , c_2 and c_3 from left to right. (a) $r_{c_{bg}} = 0.986$, (b) $r_{c_1} = 0.990$, (c) $r_{c_2} = 0.985$, (d) $r_{c_3} = 0.986$.

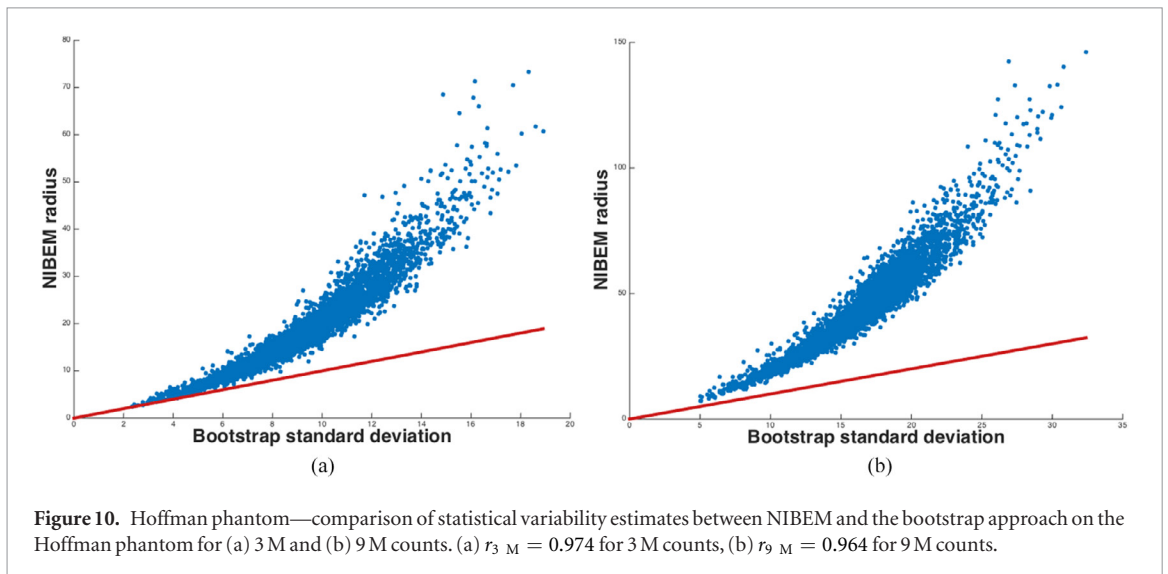


Figure 10. Hoffman phantom—comparison of statistical variability estimates between NIBEM and the bootstrap approach on the Hoffman phantom for (a) 3 M and (b) 9 M counts. (a) $r_{3M} = 0.974$ for 3 M counts, (b) $r_{9M} = 0.964$ for 9 M counts.

500 bootstrap sinograms (Buvat 2002) for each count level (respectively 3 M and 9 M counts). Acquisitions were pre-corrected for attenuation, normalization, scatter and random. We performed 120 iterations for NIBEM and ML-EM reconstructions. Activity ratios and typical reconstructed images of the Hoffman phantom for 3 M and 9 M counts are presented in figure 8.

5.2.3. Results

In this experiment, we investigated whether the NIBEM radius $r_i = \frac{1}{2} |\bar{f}_i - f_i|$ of the i th reconstructed interval $[f_i; \bar{f}_i]$ was related to the variability estimated by the bootstrap standard deviation or not. For the cylinder phantom, we computed the Spearman correlation coefficient r in each cylinder between the NIBEM and bootstrap statistical variation estimates. The corresponding scatter plots are shown in figure 9. The Hoffman phantom results are displayed in figure 10.

In the cylinder phantom experiment, we observed that the NIBEM statistical variability estimate was highly correlated with that obtained with the bootstrap approach, independently of the activity level in the considered regions. Indeed, figure 9 shows that the correlation coefficient was around 0.99 for the four cylinders of the phantom. The amplitudes of the bootstrap standard deviation and NIBEM radius were consistent, as highlighted by the identity line shown in red in figure 9. This experiment suggests that the NIBEM statistical variability estimated by the reconstructed radius matches the bootstrap standard deviation.

In the Hoffman phantom experiment, the Spearman correlation coefficient remained high (around 0.97) for both 3 M and 9 M counts. However the plots in figure 10 highlighted that the dependency was not linear. We also observed that the amplitudes of the NIBEM and bootstrap statistical variation estimates were different and dependent on the number of counts. This result suggests that the NIBEM interval radii cannot be considered as direct estimates of the standard deviation. In fact, bootstrap and NIBEM estimates are correlated because they both provide an estimate of the statistical variability. But while bootstrap technique yields an estimate of the standard deviation of this variability, the NIBEM approach rather yields an estimate of confidence intervals. Confidence intervals could be easily constructed by using the mean and standard deviation if the statistical distribution is Gaussian. Yet, in our Hoffman phantom experiment, we tested whether the statistical distribution computed from the bootstrap replicates followed a Gaussian distribution or not, for both the 3M and 9M counts simulations. The null hypothesis of normality of the distribution at a 5% significance level was always rejected by the Kolmogorov Smirnov test whatever the considered region of interest and number of counts.

Table 2. Quantitative comparison of mean *CL* of NIBEM, mean *IR* from NIBEM intervals set on the bootstrap estimated distribution and mean *CL* of naive intervals calculated as mean ± 1.96 standard deviations estimated from the bootstrap replicates.

Number of counts	NIBEM CL		$IR_{\text{bootstrap}}$		Naive CL	
	3 M	9 M	3 M	9 M	3 M	9 M
Region 1:1	0.940	0.920	0.920	0.979	0.761	0.669
Region 1.5:1	0.963	0.975	0.932	0.986	0.716	0.580
Region 2:1	1.0	1.0	0.965	0.989	0.757	0.595

Both information items—NIBEM confidence intervals and bootstrap standard deviations—can potentially lead to more reliable comparisons between regions of interest, and thus to more reliable diagnosis based on this comparison. But while the bootstrap approach requires generation of 15k sub-sinograms in order to make 500 bootstrap sinograms, the algorithmic complexity of the NIBEM approach is comparable to that of the MLEM approach. For example, a bootstrap reconstruction using 120 iterations of an MLEM algorithm lasted 1989 s, while, with the same setup and the same number of iterations, the NIBEM algorithm lasted only 18.4 s. i.e. 100 times faster⁷. Note also that the NIBEM computation time could still be reduced by paralleling some operations. In addition, as NIBEM is an extension of ML-EM, an accelerated ordered subset version of NIBEM can be straightforwardly deduced from OSEM (Hudson and Larkin 2002). Preliminary tests (results not shown) showed that the ordered subset NIBEM reconstructions had properties very similar to those of NIBEM in terms of computation time and statistical variability estimation.

5.3. Statistical uncertainty estimation as confidence level estimator

What NIBEM provides is an estimation of a confidence interval associated with the statistical variability of the reconstructed activity, while state of the art methods are devoted to estimating the standard deviation associated with this variability. Since it has been proven that this variability cannot be modeled by a normal variable, comparing the two approaches can be challenging (i.e. we cannot assume that the confidence of an interval centered on the mean and whose radius is proportional to the standard deviation has a known confidence). In this section, we propose to investigate the consistency between the bootstrap approach (considered as a gold standard (Buvat 2002, Lartizien *et al* 2010)) and the NIBEM approach, by considering the ratio of bootstrap values included in the NIBEM intervals. This Inclusion Rate (IR) is known to be highly dependent on the confidence level of the considered interval.

5.3.1. Data analysis

A setup as presented in section 5.2.2 was used, where a Hoffman phantom was simulated using GATE for 3 M and 9 M counts. For both 3 M and 9 M counts, we considered the 3 regions of interest of ratio 1:1, 1.5:1 and 2:1 of figure 8(a). On one hand, for each 3M and 9M count and each region, we estimated the mean CL for the NIBEM approach assuming that all pixels in the considered region had the same statistics. This estimate was computed with the ratio of NIBEM interval-valued estimates that contained the true value. This result is reported in the first two columns of table 2 (NIBEM CL). On the other hand, we computed, for each pixel of each region, the proportion of the 500 bootstrap values included in the corresponding interval-valued NIBEM estimate. The average of these values within a region is reported in columns 4 and 5 of table 2 ($IR_{\text{bootstrap}}$). The statistical distribution is not Gaussian, but since it is unimodal it could be naively tempting to build confidence intervals using the mean and standard deviation estimated with bootstrap approach. We tested, whether or not, intervals constructed as mean ± 1.96 standard deviations estimated from the bootstrap replicates could be considered as confidence intervals. In the last two columns of table 2, we reported the CL of these intervals (Naive CL).

5.3.2. Results

Table 2 shows that the mean *CL* estimated by NIBEM was very stable and independent of the overall number of counts. This statement is also true for the ratio of the bootstrap estimated values included in the NIBEM estimated interval. $IR_{\text{bootstrap}}$ is over 0.9 for the three investigated regions and both 3 M and 9 M cases. Those results highlight the link between the NIBEM and bootstrap statistical variability estimates. As expected, we also found that intervals built on the basis of the Gaussian hypothesis had unstable *CL* depending on the overall number of counts, so they could not be considered as confidence intervals. NIBEM does not produce direct estimates of standard deviation like bootstrap does, but instead provides direct stable and reliable estimates of confidence intervals, which is not possible with other state of the art methods. This property is particularly relevant for ROI comparison.

⁷Experiments have been carried out on a MacBook Pro, 2,8 GHz Intel Core i7.

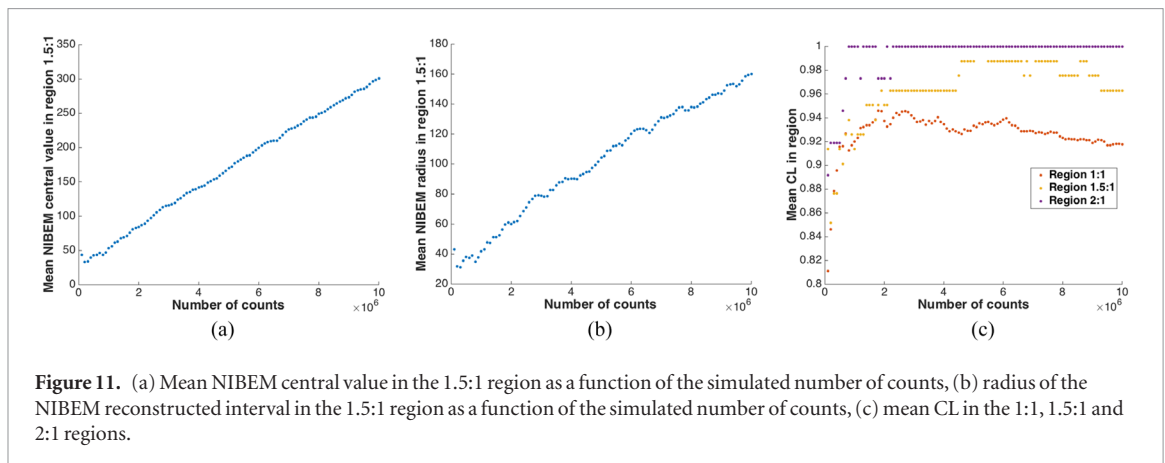


Figure 11. (a) Mean NIBEM central value in the 1.5:1 region as a function of the simulated number of counts, (b) radius of the NIBEM reconstructed interval in the 1.5:1 region as a function of the simulated number of counts, (c) mean CL in the 1:1, 1.5:1 and 2:1 regions.

5.4. Estimated statistical variability as a function of the number of counts

This experiment illustrates the behavior of NIBEM confidence intervals for different number of counts.

5.4.1. Experimental setup

We simulated 100 acquisitions of the Hoffman phantom presented in figure 8(a) using GATE. Simulations corresponding to different number of counts were performed, from 100 k to 10 M with a step of 100 k counts between each. The simulated data were pre-corrected for attenuation, normalization, scatter and random. Sinograms had 128 bins of projections and 128 angular views. Reconstructions of 128×128 images were performed using 120 iterations of NIBEM.

5.4.2. Data analysis

Figure 11(a) plots the mean NIBEM central value versus the number of counts and figure 11(b) plots the average NIBEM radius versus the numbers of counts, both for the region with a 1.5:1 activity ratio. Figure 11(c) plots the mean CL in 3 regions with 1:1, 1.5:1 and 2:1 activity ratios versus the number of counts.

5.4.3. Results

Figure 11(a) suggests that the mean central value in the 1.5:1 region increased linearly with the number of counts. Figure 11(b) suggests that the mean NIBEM statistical variation estimate (radius) follows the same trend. Figure 11(c) highlights the fact that the NIBEM reconstructed intervals estimated stable and reliable confidence intervals for the 3 investigated regions. However, figure 11(c) shows that if the number of counts (i.e. the signal-to-noise ratio) is high enough (approximately 1 M here), then the mean CL remains stable and higher than 90%, whatever the activity level and size of the ROI in the object.

6. Discussion

We have proposed here a new original reconstruction algorithm for Positron Emission Tomography. NIBEM is an adaption of the widely used iterative reconstruction algorithm ML-EM based on a directed interval arithmetic. One of its specificities lies in the fact that the projection operator used is interval-valued because it is based on a non-additive aggregation operator.

The main advantages of reconstructing interval-valued instead of precise activity values concern the properties of the reconstructed intervals. Indeed, we showed that the radius of these intervals can be used as a statistical variation predictor. In other words, it is now possible to replace the high number of statistical repetitions of ML-EM needed to estimate the reconstruction uncertainty by only one NIBEM reconstruction. The simulations also suggest that the reconstructed intervals might be interpreted as confidence intervals associated with the true value of the activity of each pixel. It is also important to underline that the presented method only focuses on providing an estimate of the statistical uncertainty. However, if uncertainties associated with, for instance, injected activity, scanner sensitivity, patient physiology, scatter estimate, can be quantified, they could also be integrated in the proposed framework. NIBEM is also fully compatible with time-of-flight (ToF) framework (Lewellen 1998). This would result in a reduction in the estimated uncertainty represented by the interval lengths.

NIBEM has a potential scope of applications in patient monitoring in oncology and in neurodegenerative disease early diagnosis. Indeed, comparing two image values at two different times, or comparing the reconstructed activities of two ROI in the same image would be greatly facilitated by the availability of confidence interval associated with each reconstructed voxel value, combined with a procedure for comparing intervals.

Such procedures have already been proposed in the relevant literature (Perolat *et al* 2015, Denœux *et al* 2005). Further studies are now necessary to thoroughly investigate the behavior of reconstructed intervals provided by NIBEM, to develop appropriate statistical tests in order to decide whether two sets of intervals measured in two different ROI are different, and to model effects beyond noise and image discretization in the NIBEM framework.

Acknowledgment

This work was partially supported by the INSERM Physicancer 2012 STIPPI project and Siemens Healthineers. The authors would like to kindly thank Fayçal Ben Bouallegue for his help on the simulations, and the anonymous reviewers for their careful and meticulous reading that helped us improve this paper.

References

- Barrett H, Wilson D and Tsui B 1994a Noise properties of the EM algorithm: I. Theory *Phys. Med. Biol.* **39** 833–46
- Barrett H, Wilson D and Tsui B 1994b Noise properties of the EM algorithm: II. Monte Carlo simulations *Phys. Med. Biol.* **39** 847–71
- Buvat I 2002 A non-parametric bootstrap approach for analyzing the statistical properties of SPECT and PET images *Phys. Med. Biol.* **47** 1761–75
- Dahlbom M 2002 Estimation of image noise in PET using the bootstrap method *IEEE Trans. Med. Imag.* **49** 2062–6
- Denneberg D 1994 *Non-Additive Measure and Integral* (Dordrecht: Kluwer) (<https://doi.org/10.1007/978-94-017-2434-0>)
- Denœux T, Massonand M and Hébert P 2005 Nonparametric rank-based statistics and significance tests for fuzzy data *Fuzzy Sets Syst.* **153** 1–28
- Fessler J 1995 *Aspire 3.0 user's guide: a sparse iterative reconstruction library Technical Report 293* Department EECS, University of Michigan, Ann Arbor
- Fessler J 1996 Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography *IEEE Trans. Image Process.* **5** 493–506
- Gardenes E, Sainz M, Jorba I, Calm R, Estela R, Mielgo H and Trepas A 2001 Modal intervals *Reliable Comput.* **7** 77–111
- Higdon D, Bowsher J, Johnson V, Turkington T, Gilland D and Jaszczak R 1997 Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data *IEEE Trans. Image Process.* **16** 516–26
- Hudson H and Larkin R 2002 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9
- Huesman R H, Gullberg G T, Greenberg W L and Budinger T F 1977 *RECBL Library Users Manual: Donner Algorithms for Reconstruction Tomography* (Berkeley: Lawrence Berkeley Laboratory)
- Jan S *et al* 2004 GATE: a simulation toolkit for PET and SPECT *Phys. Med. Biol.* **49** 4543–61
- Lartizien C, Aubin J and Buvat I 2010 Comparison of bootstrap resampling methods for 3D PET imaging *IEEE Trans. Med. Imaging* **29** 1442–54
- Lewellen T 1998 Time-of-flight PET *Semin. Nucl. Med.* **28** 268–75
- Li Y 2011 Noise propagation for iterative penalized-likelihood image reconstruction based on Fisher information *Phys. Med. Biol.* **56** 1083–103
- Loquin K, Strauss O and Crouzet J 2010 Possibilistic signal processing: how to handle noise? *Int. J. Approx. Reason.* **51** 1129–44
- Loquin K, Strauss O, Mariano-Goulart D and Buvat I 2014 A non-additive interval-valued reconstruction method in emission tomography *J. Nucl. Med.* **55** 270
- Markov S 1996 On directed interval arithmetic and its applications *J. Universal Comput. Sci.* pp 514–26
- Perolat J, Couso I, Loquin K and Strauss O 2015 Generalizing the Wilcoxon rank-sum test for interval data *Int. J. Approx. Reason.* **56** 108–21
- Qi J and Leahy R 2000 Resolution and noise properties of MAP reconstruction for fully 3D PET *IEEE Trans. Med. Imaging* **19** 493–506
- Rico A and Strauss O 2010 Imprecise expectations for imprecise linear filtering *Int. J. Approx. Reason.* **51** 933–47
- Rico A, Strauss O and Mariano-Goulart D 2009 Choquet integrals as projection operators for quantified tomographic reconstruction *Fuzzy Sets Syst.* **160** 198–211
- Schmeidler D 1986 Integral representation without additivity *Proc. AMS* **97** 255–61
- Shepp L and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22
- Sitek A 2008 Representation of photon limited data in emission tomography using origin ensembles *Phys. Med. Biol.* **53** 3201–16
- Sitek A 2012 Data analysis in emission tomography using emission-count posteriors *Phys. Med. Biol.* **57** 6779–95
- Soares E, Byrne C and Glick S 2000 Noise characterization of block-iterative reconstruction algorithms: I. Theory *IEEE Trans. Med. Imaging* **19** 261–70
- Soares E, Glick S and Hoppin J 2005 Noise characterization of block-iterative reconstruction algorithms: II. Monte Carlo simulations *IEEE Trans. Med. Imaging* **24** 112–21
- Stayman J and Fessler J 2004 Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in full 3D SPECT *IEEE Trans. Med. Imaging* **23** 1543–56
- Strauss O, Lahrech A, Rico A, Mariano-Goulart D and Telle B 2009 NIBART: A new interval based algebraic reconstruction technique for error quantification of emission tomography images *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 148–55
- Strauss O and Rico A 2012 Toward interval-based non-additive deconvolution in signal processing *Soft Comput.* **16** 809–20
- Unser M, Aldroubi A and Eden M 1993 B-spline signal processing: part I—theory *IEEE Trans. Signal Process.* **41** 821–33
- Wernick M N and Aarsvold J N 2004 *The Fundamentals of PET and SPECT* (Elsevier: Academic)