

ORIGINAL RESEARCH

Open Access



Algorithms for joint activity–attenuation estimation from positron emission tomography scatter

Yannick Berker^{1,2,3*} , Volkmar Schulz^{2,4,5} and Joel S. Karp³

*Correspondence:

yannick.berker@dkfz.de

¹Division of X-ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

²Department of Physics of Molecular Imaging Systems, Institute for Experimental Molecular Imaging, RWTH Aachen University, Forckenbeckstraße 55, 52074 Aachen, Germany

Full list of author information is available at the end of the article

Abstract

Background: Attenuation correction in positron emission tomography remains challenging in the absence of measured transmission data. Scattered emission data may contribute missing information, but quantitative scatter-to-attenuation (S2A) reconstruction needs to input the reconstructed activity image. Here, we study S2A reconstruction as a building block for joint estimation of activity and attenuation.

Methods: We study two S2A reconstruction algorithms, maximum-likelihood expectation maximization (MLEM) with one-step-late attenuation (MLEM-OSL) and a maximum-likelihood gradient ascent (MLGA). We study theoretical properties of these algorithms with a focus on convergence and convergence speed and compare convergence speeds and the impact of object size in simulations using different spatial scale factors. Then, we propose joint estimation of activity and attenuation from scattered and nonscattered (true) emission data, combining MLEM-OSL or MLGA with scatter-MLEM as well as trues-MLEM and the maximum-likelihood transmission (MLTR) algorithm.

Results: Shortcomings of MLEM-OSL inhibit convergence to the true solution with high attenuation; these shortcomings are related to the linearization of a nonlinear measurement equation and can be linked to a new numerical criterion allowing geometrical interpretations in terms of low and high attenuation. Comparisons using simulated data confirm that while MLGA converges largely independent of the attenuation scale, MLEM-OSL converges if low-attenuation data dominate, but not with high attenuation. Convergence of MLEM-OSL can be improved by isolating data satisfying the aforementioned low-attenuation criterion. In joint estimation of activity and attenuation, scattered data helps avoid local minima that nonscattered data alone cannot. Combining MLEM-OSL with trues-MLEM may be sufficient for low-attenuation objects, while MLGA, scatter-MLEM, and MLTR may additionally be needed with higher attenuation.

Conclusions: The performance of S2A algorithms depends on spatial scales. MLGA provides lower computational complexity and convergence in more diverse setups than MLEM-OSL. Finally, scattered data may provide additional information to joint estimation of activity and attenuation through S2A reconstruction.

Keywords: Positron emission tomography, Image reconstruction, Attenuation correction, Compton scattering

Introduction

Positron emission tomography (PET) is an important noninvasive medical imaging modality for clinical and research applications [1], with particular strengths in sensitive detection of photon pairs emitted by a radiotracer and quantitative reconstruction of the radiotracer activity image λ . PET image reconstruction is usually based on linear models, involving the Radon transform $\mathcal{R}\lambda$ in the analytic case or discrete mappings of a vector $\vec{\lambda}$ in the numerical case, respectively.

For quantitative reconstruction of the activity image, attenuation correction (AC) is essential, compensating for a lack of detected photon pairs along lines of response (LORs) due to the photoelectric effect and Compton scattering in the patient. A complementary step, scatter correction (SC), computes an estimate of extraneous photon pairs along *broken* LORs, which are generated through Compton scattering. Both corrections usually input the spatial distribution of the electron density ρ in the form of a map of linear attenuation coefficients μ or, for AC purposes, the so-called attenuation *sinogram* $\mathcal{R}\mu$. Given μ , both AC and SC are state of the art using well-validated algorithms [2, 3], but vast research efforts had to be—and still are—directed to the determination of μ .

Determination of an attenuation map

Depending on the level of integration of PET with other modalities (standalone or multi-modality PET), information from radionuclide transmission sources [4, 5], X-ray computed tomography (CT) [6], or magnetic resonance imaging (MRI) [7] can be used. However, radionuclide transmission data suffer from low signal-to-noise ratio, necessitating segmentation to prevent noise in the transmission data from impacting activity images. In PET/CT, 4-D attenuation correction of PET data acquired from a moving subject remains limited due to concerns over radiation doses induced by cine CT imaging. In PET/MRI outside of the head/neck area, MRI is often incapable of distinguishing bone from air in reasonable scan times [8].

More universal approaches to determine μ do not depend on multi-modality information. Popularized through maximum-likelihood reconstruction of attenuation and activity (MLAA, [9–11]), these algorithms use only PET emission data, replacing the optimization problem in λ by a joint problem in (λ, μ) . A recently proposed group of algorithms jointly estimate the activity image and the attenuation sinogram $\mathcal{R}\mu$, either using alternation [12, 13] or simultaneous updates [14].

Time-of-flight (TOF) PET emission data determine the attenuation sinogram $\mathcal{R}\mu$, but only on LORs with activity ($\mathcal{R}\lambda > 0$) and only up to an unknown offset [15]. The former limitation is not a severe issue for AC, where other values of $\mathcal{R}\mu$ are not needed. However, it complicates reconstruction of μ from $\mathcal{R}\mu$ and therefore is a problem for SC, where an image-space μ -map is usually required. The latter limitation translates into an unknown scaling factor in the reconstructed λ . For these reasons, AC and SC using only PET emission data are still impractical [16].

Another type of available data is low-energy, object-scattered PET emission data, which may contain enough additional information to address both aforementioned limitations [17]: particularly, in a joint reconstruction scheme [18]. Similar opportunities arise in single-photon emission computed tomography [19–22]. Unfortunately, the model of the measured PET scatter data is neither based on the regular Radon transform nor linear in μ . A maximum-likelihood gradient ascent algorithm for scatter-to-attenuation (S2A)

reconstruction has therefore been proposed [23, 24] but, so far, not been used in joint estimation. Most recently, a Broyden–Fletcher–Goldfarb–Shanno (BFGS)-based algorithm has been proposed for attenuation reconstruction from coincidences in a lower energy window [25, 26].

The problem of estimating attenuation from scattered PET photons shares similarities with Compton scatter imaging, in which external Gamma sources are used to probe an object's electron density for medical [27] or industrial [28] applications. While it is known from the latter that the nonlinearity of the problem favors thin, low-density objects, the impact of object size in scatter-based PET attenuation correction remains to be studied.

Objectives

This paper is thus concerned with characterizing S2A reconstruction as a building block in joint estimation of activity and attenuation (*joint estimation*). We follow three objectives: (1) further understand fundamental properties impacting convergence and convergence speed of S2A algorithms; (2) compare S2A algorithms using simulated data, specifically, in terms of convergence speed, the impact of object size, and improved performance of one algorithm by reducing its input data; and (3) study joint estimation, which implies dropping the assumption of known radiotracer activity images in S2A reconstruction [17, 24]. Therefore, we integrate scatter data into joint estimation by interleaving S2A reconstruction with trues-to-activity reconstruction, as proposed before [18, 29], as well as with trues-to-attenuation and scatter-to-activity reconstruction.

In this algorithmically oriented proof of concept, studies are carried out using 2-D digital phantoms and simulations restricted to single scattering without TOF information. Furthermore, we assume perfect energy resolution that enables ideal separation of scattered and nonscattered events and noise-free data.

After statement of the problem, introducing required imaging models for use in S2A reconstruction and joint estimation, we summarize and propose algorithms for both and describe the evaluation data used and the experiments carried out, before presenting and discussing our results.

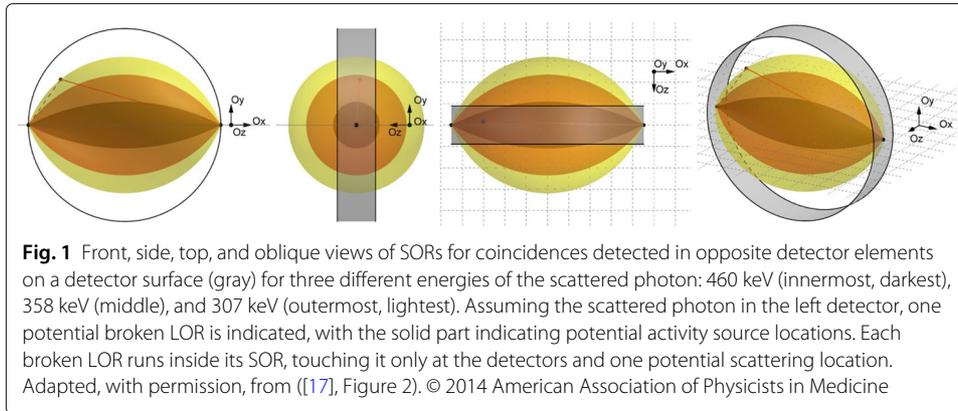
Problem statement

This section summarizes notation and models for scattered and unscattered data.

Scattered data for S2A reconstruction

Scatter-to-attenuation reconstruction requires a model of the low-energy, scattered data. Therefore, we identify the coincident detection of two photons along an LOR by the involved detector pair. If exactly one of two detected photons has been object-scattered exactly once, the coincidence is said to be single-scattered and the energy of that photon is denoted E . We denote the respective detector d_s (*scattered*) and the other d_n (*nonscattered*). Thus, a tuple $i = (d_s, d_n, E)$ comprises all properties of a single-scattered coincidence used in this work; $l = (d_s, d_n)$ denotes a regular LOR.

The trajectory of both photons is a broken LOR as shown in Fig. 1, connecting a scattering location \vec{x}_s with both detector locations. Unfortunately, many different broken LORs, in particular, having different scattering locations, yield the same apparent LOR l , so that the photon trajectory, and in particular, the scattering location,



cannot be determined from l . It is known, however, that the true \vec{x}_s lies on an American-football-shaped surface of revolution, with the pointed ends in the detector locations and the radius determined by E . We use i to denote this *surface of response* (SOR), comprising all possible scattering locations for a single-scattered coincidence¹. For each scattering location \vec{x}_s in SOR i , (i, s) describes one potential broken LOR.

In list-mode acquisition, the raw scatter data is a sequence (i_1, i_2, \dots) ; after histogramming, the data is the number of detected single-scattered coincidences for each possible i . Here, y^i denotes the simulated or measured data on SOR i , while \bar{y}^i is used for the expected data. The dimension of the data space is $N_i \leq N_d^2 \times N_E$, with N_d detectors and N_E energy bins (or equivalently, energy windows)².

Voxels are indexed according to their physical roles using e (*emitting*), s (*scattering*), and t (*transmitting*). A 2-D matrix A_λ (with entries a_s^i) describes the sensitivity of the PET camera on SOR i for radiation scattered in a voxel s in the absence of attenuation; it integrates both normalized camera sensitivity (scatter geometry, photon detection efficiency) and the object's source density λ , as detailed in the [Appendix](#). A 3-D tensor \underline{K} (with entries $k_{s,t}^i$) represents the attenuating path length of that radiation through a voxel t , independent of the object.

The expected number of low-energy scatter coincidences \bar{y} , which is linear in the activity $\bar{\lambda}$, is modeled according to a discretized variant of the scatter-measurement equation (23), a generalization of the single scatter simulation (SSS) equation [3]. Using the notation in Table 1, we write the discrete measurement equation as:

$$\bar{y}^i = \sum_s \left(\sum_e b_{s,e}^i \lambda^e \right) \exp \left(- \sum_t k_{s,t}^i \rho^t \right) \rho^s = \sum_s a_s^i \exp \left(- \sum_t k_{s,t}^i \rho^t \right) \rho^s \quad (1)$$

or, in matrix notation, denoting the element-wise operations by \odot and \exp° :

$$\bar{y}(\bar{\lambda}, \bar{\rho}) = \left(A_\lambda \odot \exp^\circ(-\underline{K}\bar{\rho}) \right) \bar{\rho}. \quad (2)$$

The above expressions are equivalent to (41) and (42), respectively. Their derivation is subject to the following assumptions:

¹SORs are single-scatter analogs of LORs, which comprise potential *emission* locations of *nonscattered* coincidences.

²While *energy bin* may be a more appropriate term for list-mode data with energy information, and *energy window* more appropriate for list-mode data without energy information or histogrammed data, we use these terms interchangeably as we are only concerned with the maximum granularity of the energy information that can be obtained from the data.

Table 1 Notation used for mathematics, image space, measurement space, and physics

Symbol	Description
$\underline{\mathbf{T}}, \underline{\mathbf{M}}, \vec{v}, \vec{1}_{[i]}$	Rank-3 tensor, matrix, vector, vector composed of ones
$\otimes; \odot, \oslash, \exp$	Outer product; element-wise multiplication, division, exp,
$\vec{\lambda}, \vec{\mu}, \vec{\rho}$	Spatial distribution of activity, linear attenuation coefficient, electron density
j, e, s, t	Indices of all, emitting, scattering, transmitting voxels
$\vec{y}, \vec{y}^i; \vec{z}, \vec{z}^i$	Measured/simulated, expected scatter (y) or trues (z) data
d_s, d_n	Detectors of scattered, nonscattered photons of a single-scatter coincidence
E, θ	Energy of scattered photon, associated scattering angle
l	$= (d_s, d_n)$, index of an LOR
i	$= (d_s, d_n, E)$, index of an SOR
(i, s)	$= (d_s, d_n, E, s)$, index of a broken LOR
N_d, N_E, N_l, N_i	Numbers of detectors, energy bins, LORs, SORs
$\mathbf{A}, \mathbf{A}_\omega, \tilde{\mathbf{A}}$	SOR system matrices: (\mathbf{A}_ω , function of parameters; ^a $\tilde{\mathbf{A}}$ with attenuation)
$\mathbf{U}, \tilde{\mathbf{U}}_\rho$	LOR system matrices (as above)
$\underline{\mathbf{B}}, b_{s,e}^i$	Probability that radiation emitted in e is detected along the broken LOR (i, s) , disregarding attenuation, per unit electron density in s , per unit activity in e
$\underline{\mathbf{L}}, l_{s,t}^i$	Effective intersection length of photon path along the broken LOR (i, s) with the transmitting (or attenuating) voxel t , taking photon energy into account
$\underline{\mathbf{K}}, k_{s,t}^i$	$= l_{s,t}^i \cdot \mu^t / \rho^t$, effective intersection length times LAC–electron-density ratio
$\mathcal{L}_y(\bullet)$	Poisson log-likelihood (LL) given the measured data \vec{y}
$\mathcal{S}(\bullet, \vec{x}^{\text{true}})$	Normalized mean squared error (NMSE) with respect to a reference \vec{x}^{true}

^aThese parameters are assumed to be constant within one iteration, but can be updated between iterations

$k_{s,t}^{i,e} = k_{s,t}^i$ Effective attenuation lengths of a voxel t seen by photons along a broken LOR (i, s) are independent of the point of emission e along that broken LOR: this is a common assumption in PET that has been fundamental in showing that TOF PET data determine the attenuation sinogram up to a constant [15].

$\mu \propto \rho$ The linear attenuation coefficient μ is proportional to the electron density ρ : approximately, this is true because in biological (low-Z) materials at PET energies, Compton scattering is the dominant interaction preventing gamma photon pairs from being detected. At fixed energy, the ratio μ/ρ can be formulated to depend on the mass attenuation coefficient (μ/ρ_m) and the quotient of mass density and electron density (ρ_m/ρ). The former is fairly constant across human tissues at PET energies ([30], Fig. 3), and the latter is almost perfectly constant for materials less dense than water and deviates a maximum of 10% for materials three times as dense ([31], Fig. 1). Note that μ/ρ may well depend on the photon energy; no assumption about the energy dependence of μ is implied (see the discussion around (35)).

If we assume that the electron density $\vec{\rho}$ is known accurately enough to approximate attenuation effects, as we will for one algorithm, we can simplify (2) further. That is, with an estimate $\vec{\rho}^{\text{est}}$ and using the abbreviation $\tilde{\mathbf{A}}_{\lambda,\rho} := \mathbf{A}_\lambda \odot \exp(-\underline{\mathbf{K}}\vec{\rho})$, we find the linear mapping:

$$\vec{y}^i = \tilde{\mathbf{A}}_{\lambda,\rho^{\text{est}}} \vec{\rho} \iff \vec{y}^i = \sum_j \tilde{a}_j^i \rho^j. \quad (3)$$

We denote attenuated system matrices by a tilde (see Table 1; \tilde{a} for components) and refer to (3) as the linearization of the scatter measurement equation.

Unscattered data for joint estimation

Our joint estimation approach requires, additionally, a model of the nonscattered data and three well-known algorithms. We assume the LOR model with:

$$\begin{aligned} \vec{z} &= (\mathbf{U}\vec{\lambda}) \odot \exp(-\mathbf{U}\vec{\rho}) = \tilde{\mathbf{U}}_{\rho}\vec{\lambda} \\ \iff \quad z^l &= \left(\sum_j u_j^l \lambda^j \right) \exp \left(- \sum_j u_j^l \rho^j \right) = \sum_j \tilde{u}_j^l \lambda^j, \end{aligned} \quad (4)$$

where \vec{z} is the expected nonscattered data, \mathbf{U} is the LOR system matrix without attenuation, i.e., the usual system matrix applied for the usual PET reconstruction (see [Appendix](#)), and $\tilde{\mathbf{U}}_{\rho}$ the attenuated one; u_j^l and \tilde{u}_j^l represent entries of \mathbf{U} and $\tilde{\mathbf{U}}_{\rho}$, respectively, for LOR l and voxel j .

These nonscattered (true) data are used by maximum-likelihood expectation-maximization [32], which we refer to as trues-MLEM:

$$\vec{\lambda}^{\text{new}} = \vec{\lambda} \odot \left(\tilde{\mathbf{U}}_{\rho}^{\top} (\vec{z} \oslash \vec{\bar{z}}) \right) \oslash \left(\tilde{\mathbf{U}}_{\rho}^{\top} \vec{1}_{[l]} \right), \quad (5)$$

and by the relaxed maximum-likelihood transmission algorithm [33] (trues-MLTR):

$$\vec{\rho}^{\text{new}} = \vec{\rho} + \eta \cdot \left(\mathbf{1} - \left(\mathbf{U}^{\top} \vec{z} \right) \oslash \left(\mathbf{U}^{\top} \vec{\bar{z}} \right) \right). \quad (6)$$

In addition, scatter-MLEM [34] will be used, of which a brief derivation is given in the [Appendix](#). This algorithm's update equation reads:

$$\vec{\lambda}^{\text{new}} = \vec{\lambda} \odot \left(\tilde{\mathbf{A}}_{\rho}^{\top} (\vec{y} \oslash \vec{\bar{y}}) \right) \oslash \left(\tilde{\mathbf{A}}_{\rho}^{\top} \vec{1}_{[l]} \right). \quad (7)$$

Methods and materials

In this section, we summarize two recent S2A algorithms and then look at fundamental differences between them. We then propose two novel joint estimation approaches using these algorithms and present our evaluation strategy.

Gradient-based algorithms for S2A reconstruction

In previous work, we introduced the two-branch back-projection (2BP) algorithm [17] which chooses between a positive and a negative update of ρ in a binary random-walk fashion. Since we found this algorithm to be impractical for most applications [24], we focus on two gradient-ascent-based algorithms here.

The Poisson log-likelihood (LL) of some expected data $\vec{\bar{y}}$, given the data \vec{y} and omitting terms that do not depend on \vec{y} , reads:

$$\mathcal{L}_y(\vec{y}) = \sum_i (y^i \log \bar{y}^i - \bar{y}^i), \quad (8)$$

with its gradient with respect to a vector $\vec{\rho}$

$$\vec{\nabla}_{\vec{\rho}} \mathcal{L}_y = \left(\vec{\nabla}_{\vec{\rho}} \otimes \vec{\bar{y}} \right) (\vec{y} \oslash \vec{\bar{y}} - \vec{1}_{[l]}) \iff \frac{\partial \mathcal{L}_y}{\partial \rho^j} = \sum_i \frac{\partial \bar{y}^i}{\partial \rho^j} \left(\frac{y^i}{\bar{y}^i} - 1 \right). \quad (9)$$

For the linearization (3), since $\tilde{\mathbf{A}}_{\lambda, \rho^{\text{est}}}$ does not depend on $\vec{\rho}$, we find the gradient of the expected data to be:

$$\vec{\nabla}_{\vec{\rho}} \otimes \vec{\bar{y}} = \tilde{\mathbf{A}}_{\lambda, \rho^{\text{est}}}^{\top} \iff \frac{\partial \bar{y}^i}{\partial \rho^j} = \tilde{a}_j^i. \quad (10)$$

By contrast, observing the double dependence of (2) on $\vec{\rho}$, one finds:

$$\vec{\nabla}_{\vec{\rho}} \otimes \vec{y} = \left[\mathbf{A}_{\lambda} \odot \exp(-\mathbf{K}\vec{\rho}) - \vec{\rho}^{\top} \left\{ \mathbf{K} \odot \left(\left[\mathbf{A}_{\lambda} \odot \exp(-\mathbf{K}\vec{\rho}) \right] \otimes \vec{\mathbf{1}}_{[t]} \right) \right\} \right]^{\top} \quad (11a)$$

$$= \left[\tilde{\mathbf{A}}_{\lambda, \rho} - \vec{\rho}^{\top} \left\{ \mathbf{K} \odot \left(\tilde{\mathbf{A}}_{\lambda, \rho} \otimes \vec{\mathbf{1}}_{[t]} \right) \right\} \right]^{\top} \quad (11b)$$

instead, which simplifies to (10) only under $\vec{\rho} = 0$ or $\mathbf{K} = 0$ (nonscattering or nonattenuating object) and $\vec{\rho}^{\text{est}} = \vec{\rho}$. This vectorial expression lends itself particularly well to an implementation in MATLAB (The MathWorks, Natick, MA); note that the multiplication with $\vec{\rho}^{\top}$ (from the left) denotes a summation over the scattering voxels s , as indicated in the component-wise expression:

$$\frac{\partial \bar{y}^i}{\partial \rho^j} = a_j^i \exp\left(-\sum_t k_{j,t}^i \rho^t\right) - \sum_s \rho^s k_{s,j}^i a_s^i \exp\left(-\sum_t k_{s,t}^i \rho^t\right) \quad (12a)$$

$$= \tilde{a}_j^i - \sum_s \rho^s k_{s,j}^i \tilde{a}_s^i. \quad (12b)$$

Scatter-to-attenuation MLEM with one-step-late attenuation (MLEM-OSL)

This algorithm, called MLEM by its authors [18], is based on subsuming attenuation effects under the system matrix in the linearized measurement equation (3), yielding the MLEM update [32]:

$$\vec{\rho}^{\text{new}} = \vec{\rho} \odot \left(\tilde{\mathbf{A}}_{\lambda, \rho}^{\top} \left(\vec{y} \otimes \vec{y} \right) \right) \odot \left(\tilde{\mathbf{A}}_{\lambda, \rho}^{\top} \vec{\mathbf{1}}_{[t]} \right) \iff \rho^{\text{new}, j} = \rho^j \frac{\sum_i \tilde{a}_j^i y^i / \bar{y}^i}{\sum_i \tilde{a}_j^i}. \quad (13)$$

However, this update ignores the fact that $\tilde{\mathbf{A}}_{\lambda, \rho}$ depends on $\vec{\rho}$, and $\tilde{\mathbf{A}}_{\lambda, \rho}$ has to be updated *after* every iteration: Eq. (13) follows the spirit of the so-called one-step-late (OSL) algorithms [35], and we will refer to it as MLEM-OSL here.

Maximum-likelihood gradient ascent (MLGA)

The MLEM(-OSL) update (13) can be written as a scaled gradient ascent, with the gradient given by Eqs. (9 and 10) and a vector-valued step size [36]:

$$\vec{\rho}^{\text{new}} = \vec{\rho} + \underbrace{\vec{\rho} \odot \left(\tilde{\mathbf{A}}_{\lambda, \rho}^{\top} \vec{\mathbf{1}}_{[t]} \right)}_{\text{step size}} \odot \vec{\nabla}_{\vec{\rho}} \mathcal{L}_y \iff \rho^{\text{new}, j} = \rho^j + \underbrace{\frac{\rho^j}{\sum_i \tilde{a}_j^i}}_{\text{step size}} \frac{\partial \mathcal{L}_y}{\partial \rho^j} \quad (14a)$$

$$= \vec{\rho} + \vec{s} \odot \vec{\nabla}_{\vec{\rho}} \mathcal{L}_y \quad = \rho^j + s^j \frac{\partial \mathcal{L}_y}{\partial \rho^j}. \quad (14b)$$

The closed-form expression of MLGA [23, 24] is obtained from (14b) by inserting the full log-likelihood gradient (9 and 11) and choosing a step size. In this work, we focus on a step size inspired by the MLEM update equation (14a):

$$\vec{s} = \gamma \cdot \vec{\rho} \odot \left(\tilde{\mathbf{A}}_{\lambda, \rho}^{\top} \vec{\mathbf{1}}_{[t]} \right) \iff s^j = \gamma \frac{\rho^j}{\sum_i \tilde{a}_j^i}. \quad (15)$$

In addition to this MLEM-like step size, two additional step sizes have been tested: the constant step size, proposed before [24], and the scaled nonuniform step size:

$$\vec{s}' = \alpha \cdot \vec{\mathbf{1}}_{[t]} \quad \text{and} \quad \vec{s}'' = \beta \cdot \vec{\rho}. \quad (16)$$

Step-size constants α , β , and γ have been optimized empirically for fastest, yet stable convergence with our data.

Validity of linearizing the scatter measurement equation

The amount of scatter as a function of electron density may not be sufficiently well represented by a linearized measurement equation, and (2) may require more careful treatment. To explore the limits of the linearization, we derive a geometrical interpretation as well as a numerical criterion. This criterion is used to distinguish data that can be used for algorithms based on linearization (here, MLEM-OSL) from data that cannot; further, it is linked to the log-likelihood gradient (11).

The most basic, one-dimensional (1-D) simplification of (2):

$$\bar{y} = a_{\lambda} \exp(-k\rho^{\text{est}})\rho \quad (a_{\lambda}, k, \rho^{\text{est}}, \rho > 0), \quad (17)$$

confirms that no possible linearization in the form of (3):

$$\bar{y}' = a_{\lambda, \rho^{\text{est}}}\rho, \quad (18)$$

reflects the behavior of (17) with high attenuation (see Fig. 2): in particular, the derivative of (18) misrepresents the sign of the derivative of (17) for $k\rho > 1$. This may have drastic implications for gradient-based algorithms using the linearization to compute the gradient: in particular, if $\rho^n > \rho^{\text{true}} > 1/k$, an iteration of MLEM-OSL yields $\rho^{n+1} > \rho^n$ regardless of the value of ρ^{est} .

Comparison of the gradient of the linearization (10) with the full gradient (11) reveals the advantage of MLGA over MLEM-OSL; the difference term $-\bar{\rho}^{\top}\{\underline{\mathbf{K}}\odot[\dots]\}$ reverses the direction of the full gradient (only) with high attenuation, all components of $\bar{\rho}$ and $\underline{\mathbf{K}}$ being nonnegative³.

A multi-voxel interpretation of the high-attenuation situation is presented in the Appendix: it is of importance in patients with great attenuation-length–electron-density products $\bar{\rho}^{\top}\underline{\mathbf{K}}$ ⁴. One conclusion from the arguments in the Appendix is that it is not straightforward to downsample (or downsize) S2A experiments, as that can transform low attenuation into high attenuation (by downsampling), or vice versa (by downsizing)⁵.

For MLEM-OSL, the linearization of the measurement equation may only be appropriate whenever attenuation effects do not reverse the sign of $\partial\bar{y}^i/\partial\rho^j$. Since this may be true for some SORs i , but not for others, it may be appropriate to remove the latter from the data and apply MLEM-OSL to the reduced data set: (46) represents an approximate inclusion criterion used later.

n -algorithms for joint estimation

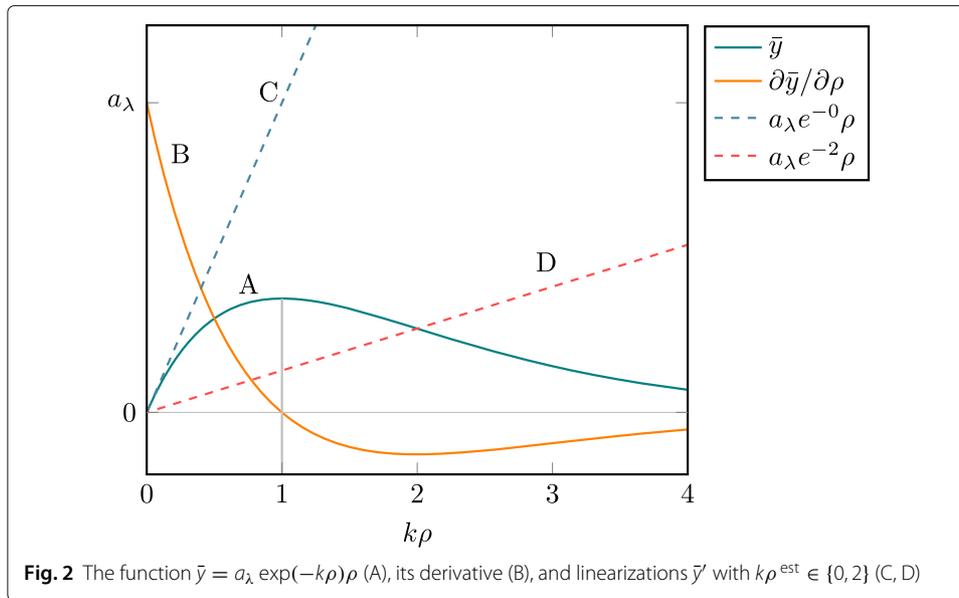
Up to this point, we have focused on dedicated S2A reconstruction algorithms, assuming knowledge of the activity distribution; in this section, we drop this assumption and extend our studies to joint estimation of activity and attenuation using scattered as well as nonscattered data. The added value of combining scattered and nonscattered data is visualized in the Appendix (Fig. 16).

We use five building blocks: the aforementioned MLGA with MLEM-like step size for S2A reconstruction, henceforth referred to as *scatter-MLGA*; *scatter-MLEM-OSL* as an alternative; *scatter-MLEM* for scatter-to-activity reconstruction; *trues-MLEM* for

³This update using both terms is also responsible for vastly improved performance of MLGA compared to 2BP [24]: the full gradient of the data (11) is effectively a weighted sum of the 2BP update branches, where the weights ensure maximization of the log-likelihood.

⁴The use of an elliptical phantom with relatively short, 12 and 15 cm long axes may have prevented this behavior from manifesting in an earlier study [18].

⁵By contrast, downsampling or downsizing activity reconstruction is possible without such considerations.

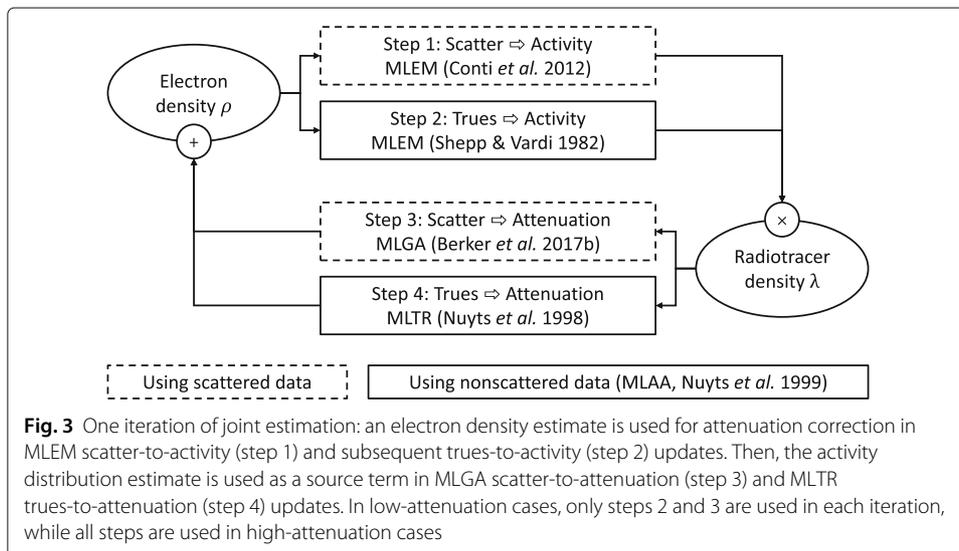


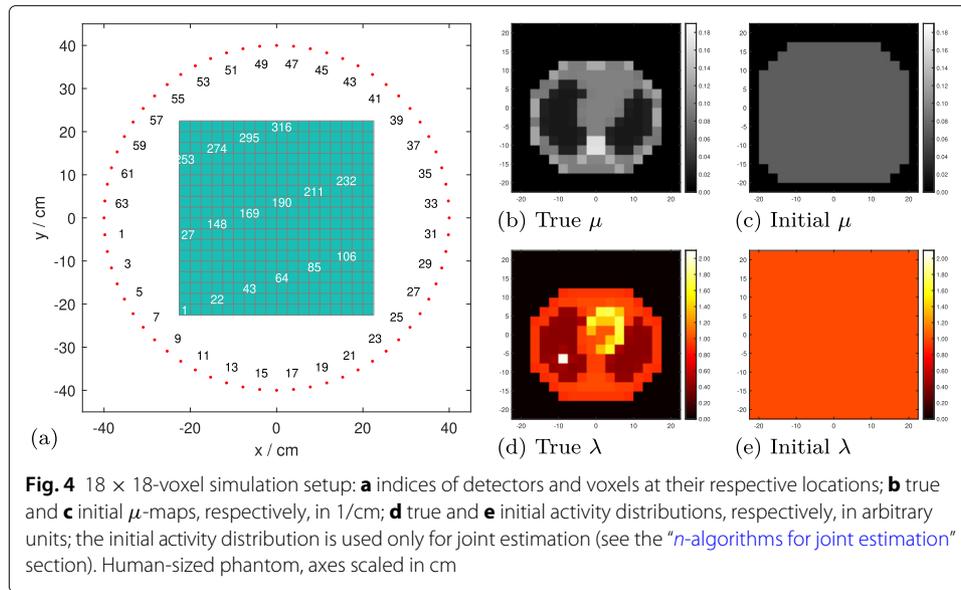
trues-to-activity reconstruction; and trues-MLTR for trues-to-attenuation reconstruction. Combinations of n individual algorithms form n -algorithms.

As for S2A reconstruction, we distinguish two main cases: low and high attenuation. The general data flow per iteration is similar for both cases and is visualized in Fig. 3; radiotracer activity distribution λ and electron density ρ are repeatedly updated using the current estimate of the respective other quantity. For both cases, we re-optimized the MLGA step sizes to achieve stability.

2-algorithms for low attenuation

The idea of this subsection has been presented before [18, 29]. For low attenuation situations (e.g., with a spatial scale factor of 0.2 in Fig. 4), we interleave trues-MLEM with scatter-MLGA. The data flow in this part is similar to that proposed earlier [18]. We start with initial guesses for ρ and λ . In each iteration, plugging the current electron-density estimate $\tilde{\rho}$ into $\tilde{\mathbf{U}}_\rho$, we use trues-MLEM to update the current activity estimate





$\vec{\lambda}$ using the nonscattered data \vec{z} ; then, we use the updated activity estimate to compute scatter-MLGA updates of $\vec{\rho}$.

In this part of the study, we aim to minimize the number of computationally expensive updates of the system matrix $\vec{A}\lambda$. Therefore, we run 10 iterations of attenuation-corrected (using the current $\vec{\rho}$) trues-OSEM (with 4 data subsets) at a time, followed by 10 iterations of scatter-MLGA with 4 data subsets (the use of subsets in scatter-MLGA being studied in detail elsewhere [24]). This low-attenuation 2-algorithm is summarized as (trues-OSEM₄¹⁰) + (scatter-MLGA₄¹⁰), with a total of 20 sub-iterations per iteration. Since MLEM-OSL can replace scatter-MLGA for low attenuation, we also run (trues-OSEM₄¹⁰) + (MLEM-OSL¹⁰) on the same data for comparison.

4-algorithms for low or high attenuation

For high attenuation (e.g., Fig. 4 at the original, that is, human spatial scale), we find it necessary to further consider activity information contained in scattered coincidences, as well as attenuation information contained in true coincidences. The former is achieved by the scatter-MLEM algorithm, the latter with the trues-MLTR algorithm with a relaxation factor of $\eta = 0.03$. Iterations of different algorithms updating the same quantity are considered as one sub-iteration; all updates are applied subsequently (e.g., the trues-MLEM update used the estimate of λ as updated by the previous scatter-MLEM update; see Algorithm 1).

Not using any subsets, the high-attenuation 4-algorithm is noted (scatter-MLEM + trues-MLEM) + (scatter-MLGA + trues-MLTR), with two sub-iterations per iteration.

Evaluation strategy

Evaluation data

We simulate data based on an 18 × 18-voxel version (*high resolution*, Fig. 4) of the human-sized chest cross-section phantom used previously [24], as well as the original one (9 × 9 voxels, *low resolution*, Fig. 14a). For the former, the voxel size is 25 × 25 mm² and the radius of the 2-D PET scanner used to simulate a PET acquisition is 40 cm. For a rat-sized field of view (FOV), the phantom (and the scanner geometry) are uniformly scaled down

Algorithm 1 The joint 4-algorithm with updates in simplified notation

1:	$\lambda \leftarrow \lambda^{\text{init}}, \mu \leftarrow \mu^{\text{init}}$	//Initialization (Fig. 4(c) and 4(e))
2:	for each iteration do	
3:	$\lambda \leftarrow \lambda \times (A(y/\bar{y}))/A1)$	//Sub-iteration 1: scatter-MLEM (7)
4:	$\lambda \leftarrow \lambda \times (U(z/\bar{z}))/U1)$	//Sub-iteration 1: trues-MLEM (5)
5:	$\rho \leftarrow \rho + \gamma\rho((A - \rho KA)(y/\bar{y} - 1))/A1)$	//Sub-iteration 2: scatter-MLGA (14)
6:	$\rho \leftarrow \rho + \eta(1 - (Uz)/(U\bar{z}))$	//Sub-iteration 2: trues-MLTR (6)

by a factor of 0.2, all size relations remaining identical ($5 \times 5 \text{ mm}^2$ pixel size, 8-cm detector radius)⁶. An intermediate, rabbit-sized FOV is obtained using a linear downscaling factor of 0.35. At all scales, the scanner is equipped with $N_d = 64$ equidistant detectors having $N_E = 10$ energy bins, of which 7 are effectively used: (511 keV, 460 keV] down to (204 keV, 153 keV].

Single-scattered data is simulated by evaluating (2). Nonscattered data is simulated by (4) using a system matrix \mathbf{U} , each column j of which is constructed from the result of the MATLAB `radon` function [37] for a unity point source in j .

For all algorithms, the initial guess of ρ generously bounds the true object and is filled with the equivalent of $\mu = 0.07/\text{cm}$ (Fig. 4c): this value ensures approximately correct attenuation correction factors for the first iteration of trues-to-activity reconstruction. For joint estimation, the initial activity is homogeneous throughout the FOV (Fig. 4e).

S2A reconstruction

The first part of this comparison of MLGA and MLEM-OSL is along the lines of earlier work comparing MLGA with 2BP [24], using additional simulation data with higher numbers of voxels than before. Therefore, both algorithms are applied to the (low and high resolution) data described above. Due to the small number of voxels, specific features of reconstructed images are of less interest; for the agreement between reconstructed images \vec{x} with their respective references \vec{x}^{true} , we therefore report normalized mean squared errors (NMSE):

$$\mathcal{S}(\vec{x}, \vec{x}^{\text{true}}) = \sum_i (x^i - x^{\text{true},i})^2 / \sum_i (x^{\text{true},i})^2. \quad (19)$$

FOV size variations Both algorithms are applied to the data simulated at all three spatial scales (human: scale 1; rabbit: scale 0.35; rat: scale 0.2).

Reduced data For MLEM-OSL, data is reduced by separating SORs into useful and less useful ones based on the aforementioned criterion, useful ones fulfilling:

$$\max_j \sum_s \rho^s k_{s,j}^i \leq 1. \quad (20)$$

This criterion is evaluated using the current estimate of $\vec{\rho}$ in every iteration. SORs i which are to be left out are removed both from the data \vec{y} (removing single data points) and the system matrix \mathbf{A}_λ (removing whole rows), and all computations are carried out with these reduced variables when working with reduced data.

⁶This spatial scaling is identical, up to a global factor in the measurement data, to scaling the μ -map intensities by the same factor while maintaining all spatial dimensions—hence the interpretation in terms of high and low attenuation.

Computational complexity and sparsity Computational complexity of both algorithms is assessed by measuring run times on a consumer-grade laptop (Intel Core 2 Duo 2.8 GHz processor, 4 GB memory). Therefore, the simulation parameters are varied in two ways. First, with the number of voxels fixed at low resolution, we vary the number of detectors following $N_d = 2^n$ with $n \in \{1, \dots, 7\}$. Second, with the number of detectors fixed (at $N_d = 32$), we vary the number of voxels following $2^n \times 2^n$ with $n \in \{1, \dots, 5\}$; in terms of vector lengths, that corresponds to $N_e = N_s = N_t = 4^n$. When varying the number of voxels, the voxel dimensions are adapted to maintain a constant spatial extent of the phantom.

For this part of the study, we choose constant activity and attenuation distributions ($\lambda_j = 1, \mu_j = 0.1/\text{cm}$), with an initial $\mu_j = 0.05/\text{cm}$. Since this choice implies a maximum population of the system matrix A_λ , we also determine what we term the *geometrical density* (fraction of non-null entries with flat activity) of A_λ and \mathbf{K} , respectively, which represent upper bounds for cases with less extended activity distributions.

Joint estimation

In joint estimation, in addition to computing NMSEs, we are interested in the evolution of several likelihood values. Attenuation- and activity-reconstruction algorithms are designed to maximize likelihoods given the true value of all other quantities. For scattered data, these are:

$$\mathcal{L}_{\text{scatt}}^{\text{att}}(\rho^{\text{est}}) = \mathcal{L}_y(\bar{\bar{y}}(\lambda^{\text{true}}, \rho^{\text{est}})) \quad (21a)$$

$$\mathcal{L}_{\text{scatt}}^{\text{act}}(\lambda^{\text{est}}) = \mathcal{L}_y(\bar{\bar{y}}(\lambda^{\text{est}}, \rho^{\text{true}})). \quad (21b)$$

However, in a joint-estimation setting, λ^{true} and ρ^{true} are generally not available. For the scattered and the true data, respectively, we therefore also track the *apparent* likelihoods, which are the quantities as seen by the optimization algorithms:

$$\mathcal{L}_{\text{scatt}}^{\text{app}} = \mathcal{L}_y(\bar{\bar{y}}(\lambda^{\text{est}}, \rho^{\text{est}})) \quad (22a)$$

$$\mathcal{L}_{\text{trues}}^{\text{app}} = \mathcal{L}_z(\bar{\bar{z}}(\lambda^{\text{est}}, \rho^{\text{est}})). \quad (22b)$$

We then study the following combinations of data and algorithms.

2-algorithms for low attenuation We apply (trues-OSEM₄¹⁰) + (scatter-MLGA₄¹⁰) to the high-resolution, low-attenuation data. To verify that MLEM-OSL can replace MLGA, we also apply (trues-OSEM₄¹⁰) + (MLEM-OSL¹⁰) to the same data.

4-algorithms for low and high attenuation We first compare the 4-algorithm, (scatter-MLEM + trues-MLEM) + (scatter-MLGA + trues-MLTR), to the 2-algorithm in terms of performance on the high-resolution, low-attenuation data; then, we apply only the 4-algorithm to high-resolution, high-attenuation data.

4-algorithm to resolve MLAA crosstalk During initial studies with a low-resolution object at the human scale (Fig. 14a), traditional MLAA (trues-MLEM₁¹ + trues-MLTR₁¹) converged to an apparent local maximum of the likelihood (Fig. 14c). Therefore, we use the values of ρ and λ at this point to initialize the 4-algorithm.

Implementation

All algorithms are implemented in MATLAB (R2019a; The MathWorks, Natick, MA, USA). The [Appendix](#) describes the use of sparse matrices in evaluating likelihoods and gradients. In trues- and scatter-to-attenuation algorithms, instead of adding nonnegativity constraints, we set $\rho^j \leftarrow \max\{0, \rho^j\}$ after each update.

Results

S2A reconstruction

In this section, we verify the theoretical findings using NMSEs of reconstructed ρ -maps, of which we present some examples in [Fig. 5](#).

[Figure 6a](#) shows the NMSEs of ρ for the human-sized phantom and system, for both low and higher resolution. All variants of MLGA converge to the correct solution as all NMSE curves tend to zero, while MLEM-OSL does not; MLGA with the MLEM-like step size is the fastest algorithm in both cases.

FOV size variations

[Figure 6b](#) and [c](#) show the data for the same phantom and system at rabbit and rat sizes, respectively. In these cases, all algorithms converge to the correct solution; generally, MLGA with the MLEM-like step size is among the fastest.

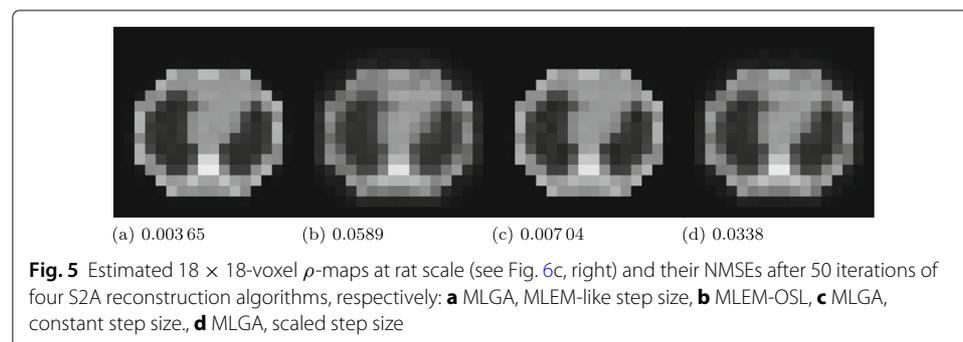
As summarized in [Fig. 7a](#), MLEM-OSL converges well with a rat-sized FOV, less rapidly (and nonmonotonously) so with a rabbit-sized FOV, and not at all with a human-sized FOV despite otherwise identical simulations. This dependence of convergence, and convergence rates, on the spatial scale of the simulated phantom is less pronounced with MLGA which, even with a constant step size as an example, converges faster than MLEM-OSL in most cases.

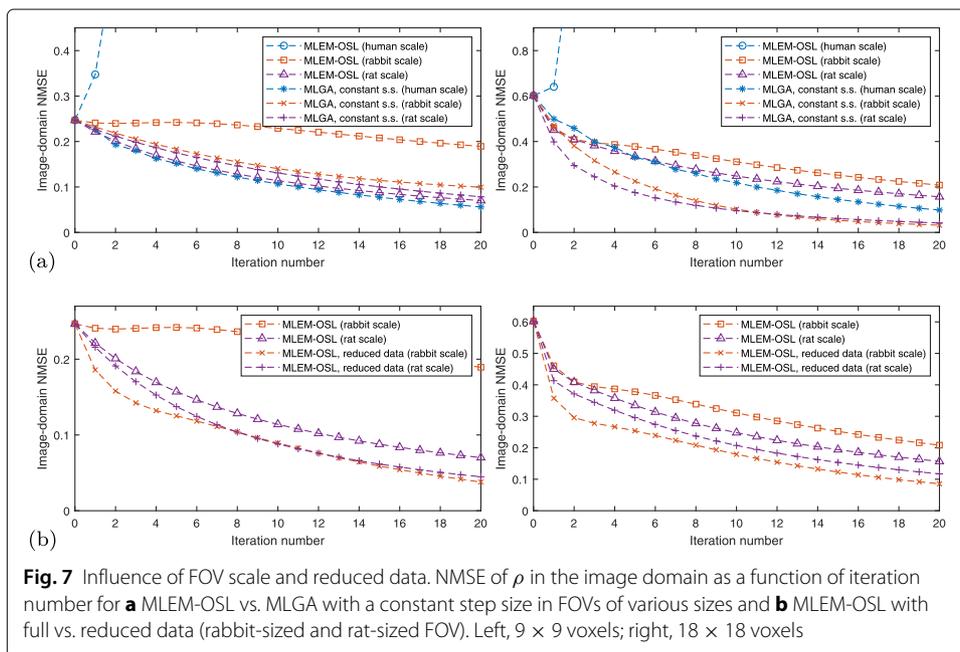
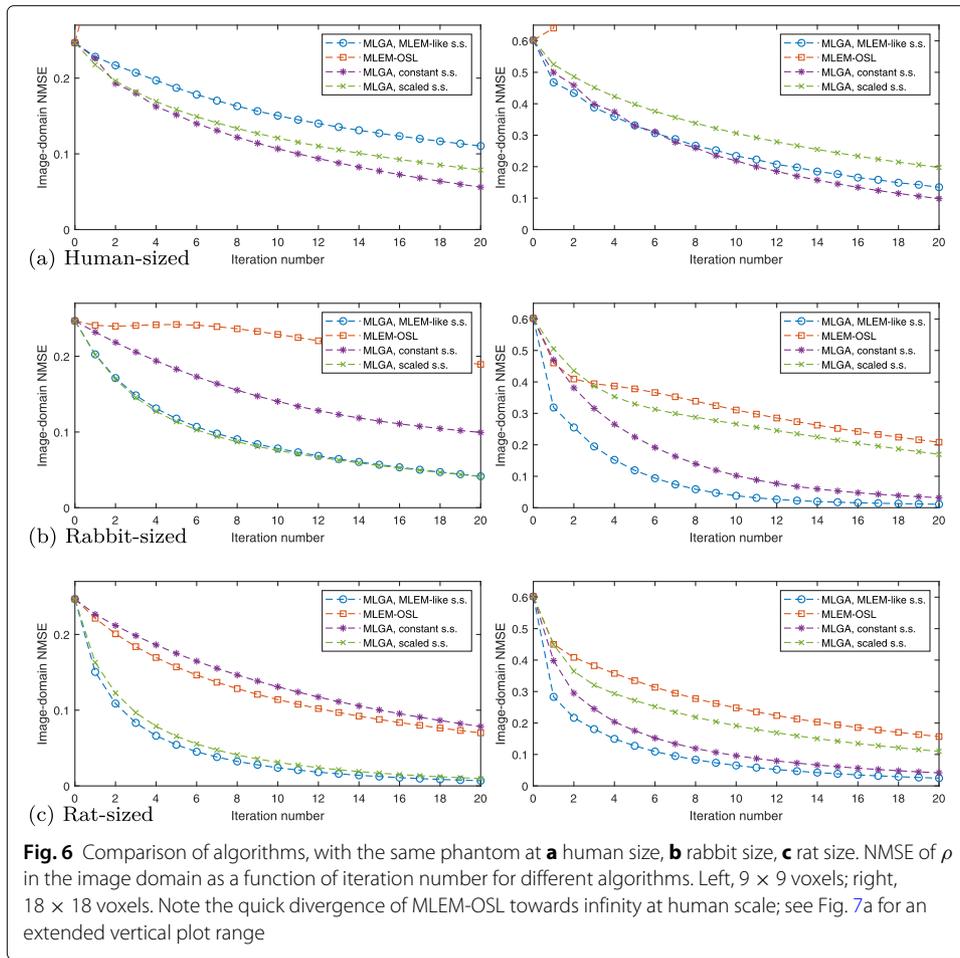
Reduced data

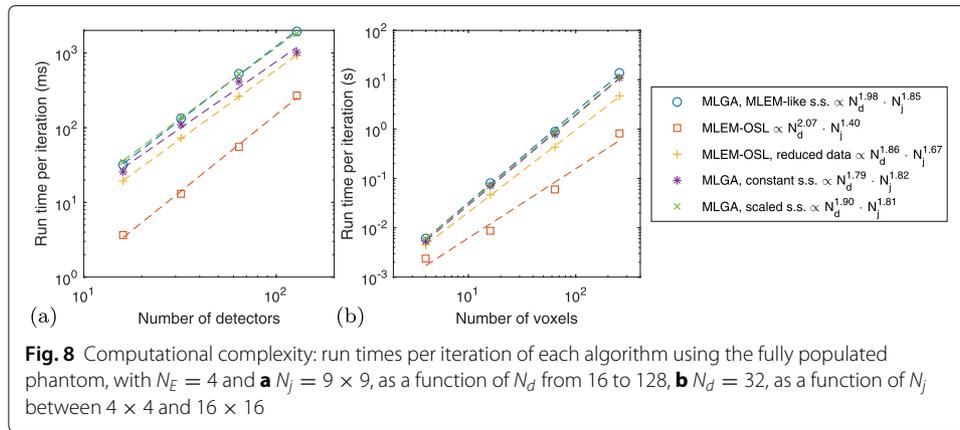
[Figure 7b](#) verifies the hypothesis that MLEM-OSL scale dependence (and hence convergence) is improved by ignoring high-attenuation SORs in the data and the system matrix using [\(20\)](#). In fact, decreasing the dimensionality of the problem in this way leads to an increase in convergence speed for MLEM-OSL.

Computational complexity and sparsity

[Figure 8](#) visualizes the run times per iteration of all algorithms in the fully populated geometry as a function of the number of detectors N_d and the number of voxels N_j , respectively. The figure legends also include fitted power laws $(N_d^a N_j^b)$. While







most algorithms show $\mathcal{O}(N_d^2)$ behavior, in terms of N_j , the exponents range from 1.40 (MLEM-OSL) to 1.85 (MLGA, MLEM-like step size).

Tables 2 and 3 summarize the density of \underline{K} and A_λ for the experiments shown in Fig. 8. Generally, the fraction of nonzero entries in both quantities decreases with increasing numbers of detectors or voxels.

Joint estimation

2-algorithms for low attenuation

The resulting images of the 2-algorithm for the low-attenuation phantom are shown after 100 and 1000 sub-iterations, respectively, in Fig. 9a and b. For the evolution of true and apparent likelihoods, we refer to Fig. 10: this plot shows the NMSE of μ and λ as a function of sub-iterations, indicating the alternating updates of activity and attenuation, and similarly, the ideal log-likelihoods (LL) using the current estimate of one quantity and the true value of the respective other quantity, respectively. Finally, the apparent LLs of scattered and nonscattered data, based on both estimated activity and estimated attenuation, are plotted.

In summary, the 2-algorithm converges towards the true activity and attenuation, even though true NMSE and apparent LL curves are nonmonotonous in parts. Figure 11 confirms that MLEM-OSL can replace scatter-MLGA in joint estimation, at low-attenuation and at the cost of reduced convergence speed.

4-algorithm for low or high attenuation

Figure 12 shows the results for the 4-algorithm in the low-attenuation case. Due to the fact that each iteration consists of only 2 sub-iterations (2 updates each), the LL and NMSE curves appear smoother than the same curves for the 2-algorithm.

Applied to the same phantom at high attenuation, the 4-algorithm converges slower, but in a similarly smooth way as for low attenuation to the true activity and attenuation

Table 2 Geometrical density (in %) of \underline{K} and A_λ with $N_j = 9 \times 9, N_E = 4$

N_d	2	4	8	16	32	64	128
\underline{K}	4.30	4.30	3.80	3.16	2.78	2.15	1.42
A_λ	33.6	34.5	26.8	22.1	18.9	14.3	9.19

Flat activity sources λ covering the complete FOV were used for A_λ to establish an upper bound

Table 3 Geometrical density as in Table 2 for $N_d = 32$ and $N_E = 4$

N_j	2×2	4×4	8×8	9×9	16×16
\mathcal{K}	1.31	2.41	2.79	2.78	1.85
A_λ	2.25	7.76	17.4	18.9	23.0

(Figs. 9c, d and 13): some nonmonotonicity remains, both at the sub-iteration level (for apparent LLs) and at the scale of dozens of sub-iterations (e.g., the true-activity LL early on or the true-attenuation LL, for which we note that it increases for later sub-iterations).

4-algorithm to resolve MLAA crosstalk

Finally, Fig. 14 shows the results of the crosstalk study. While the MLAA 2-algorithm is stuck in an apparent local maximum (Fig. 14c),⁷ the proposed 4-algorithm is able to not only avoid, but escape from this local maximum and converges towards the true solution (Fig. 14d).

Discussion

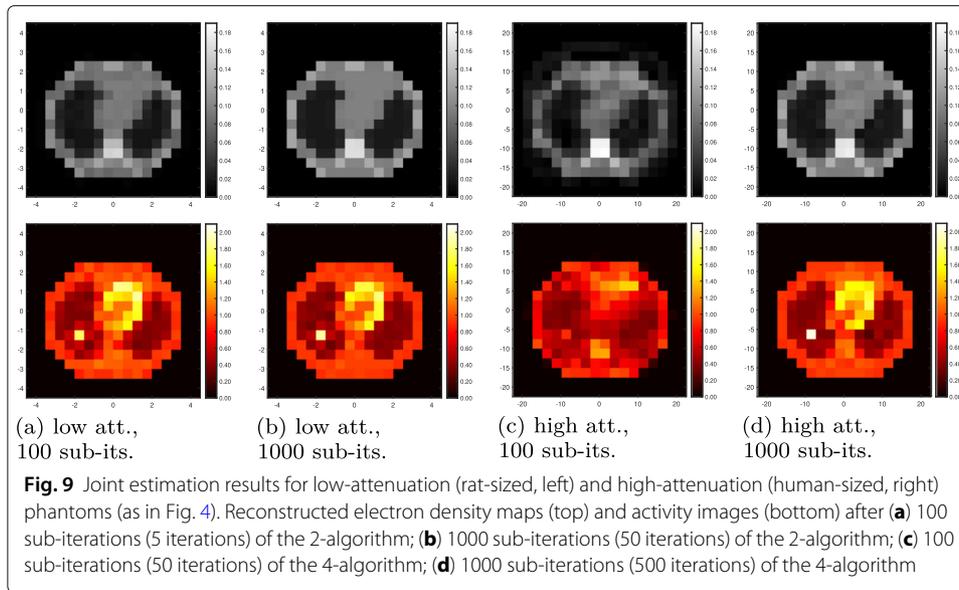
We have studied reconstruction of attenuation information from scattered coincidences in PET. Unlike other problems regarding reconstruction of activity or attenuation from nonscattered or scattered coincidences, the problem at hand is unique in that the reconstructed quantity appears twice in the measurement equation (2). The problem is therefore nonlinear, with the degree of nonlinearity depending on attenuation, which in mostly water-like objects implies spatial scale⁶.

S2A reconstruction

We have interpreted a recent take on the problem [18] as MLEM with a one-step-late update of the attenuated system matrix. MLEM-OSL, ignoring the dependence of $\tilde{A}_{\lambda,\rho}$ on ρ in computing a new estimate, has been derived by linearizing the nonlinear measurement equation; however, the fact that MLEM-OSL thus relies on a linear relationship between electron density and scattered coincidences impacts performance with high attenuation (Fig. 7a). An important result is therefore the characterization of the spatial-scale problem, which is complementary to the intensity-scale problem described earlier for joint estimation of activity and attenuation from only nonscattered data [15, 39]. Another result is a potential nonuniqueness of the isolated S2A problem indicated in Fig. 16: fortunately, the same figure indicates that combined scattered and nonscattered data do not necessarily feature the same nonuniqueness. Also, we hypothesize that additional voxels and detectors help further resolve nonuniqueness.

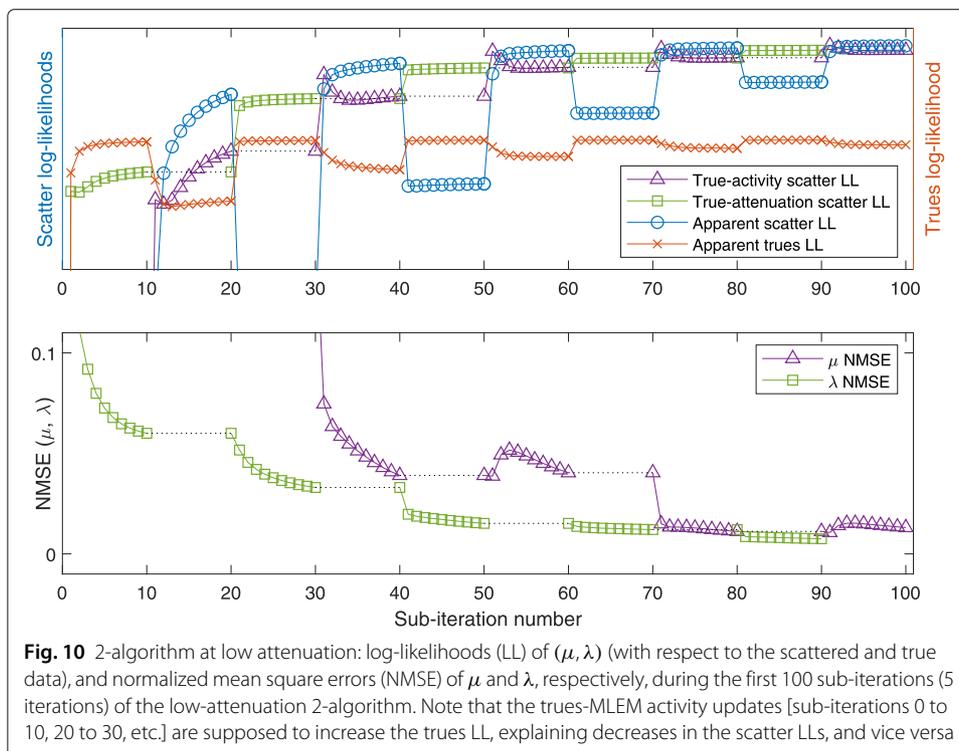
We have studied a maximum-likelihood gradient-ascent method for attenuation-map reconstruction based on the full, nonlinear data log-likelihood. Since the step size ignores the dependence of $\tilde{A}_{\lambda,\rho}$ on ρ , MLGA does not feature the provable monotonicity of the likelihood that MLEM offers in activity reconstruction from nonscattered coincidences. However, MLGA has advantages over MLEM-OSL, where ignoring said dependence leads to a wrong *direction* of the update and results in instabilities: the impact on MLGA is noticeably smaller.

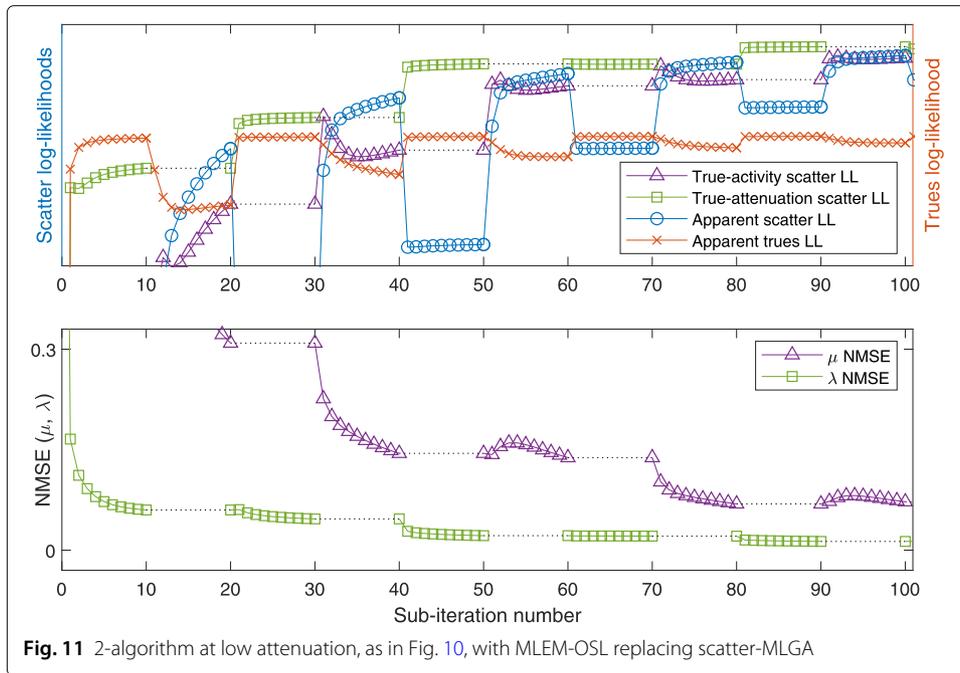
⁷Note that neither the true solution nor the limit distribution are radially symmetric; the existence of nonradial nonuniqueness in 2D non-TOF PET has recently been proven [38] and may represent a practical problem.



Application to simulated data confirmed that in larger objects, MLGA outperforms MLEM-OSL. Nonetheless, MLEM-OSL is a simple and fast algorithm for rabbit- or rat-sized objects, while MLGA may require additional speedup [24].

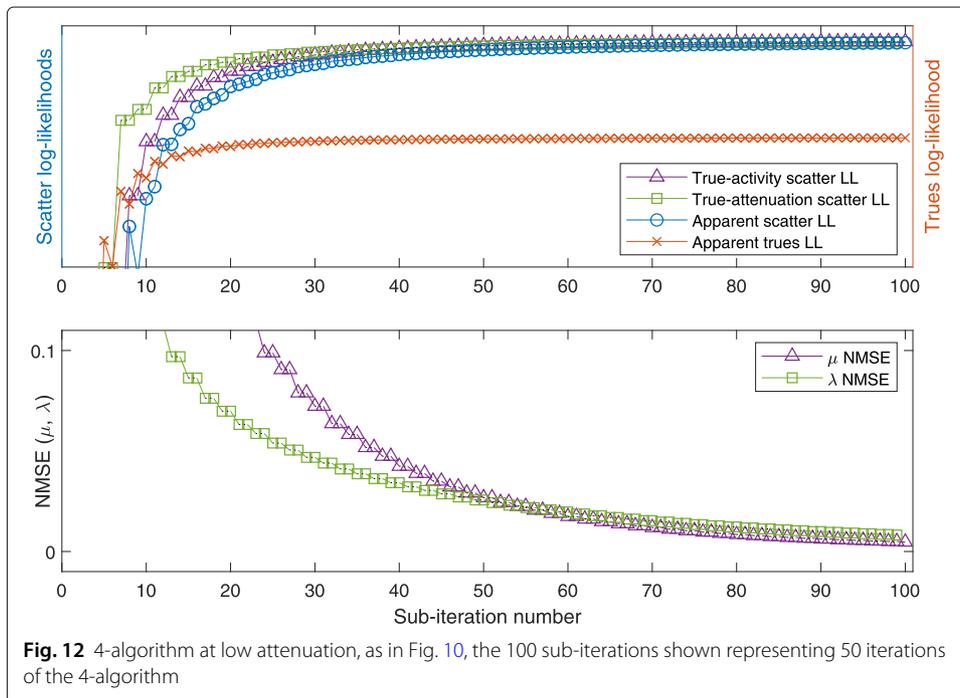
By characterizing high attenuation, we have found a criterion to separate low- from high-attenuation data and improve MLEM-OSL convergence speed (Fig. 7b). Thus, one strategy to decrease the size of system matrices and tensors, and thus computational complexity, lies in choosing the most useful SORs from the full data. Here, we only briefly

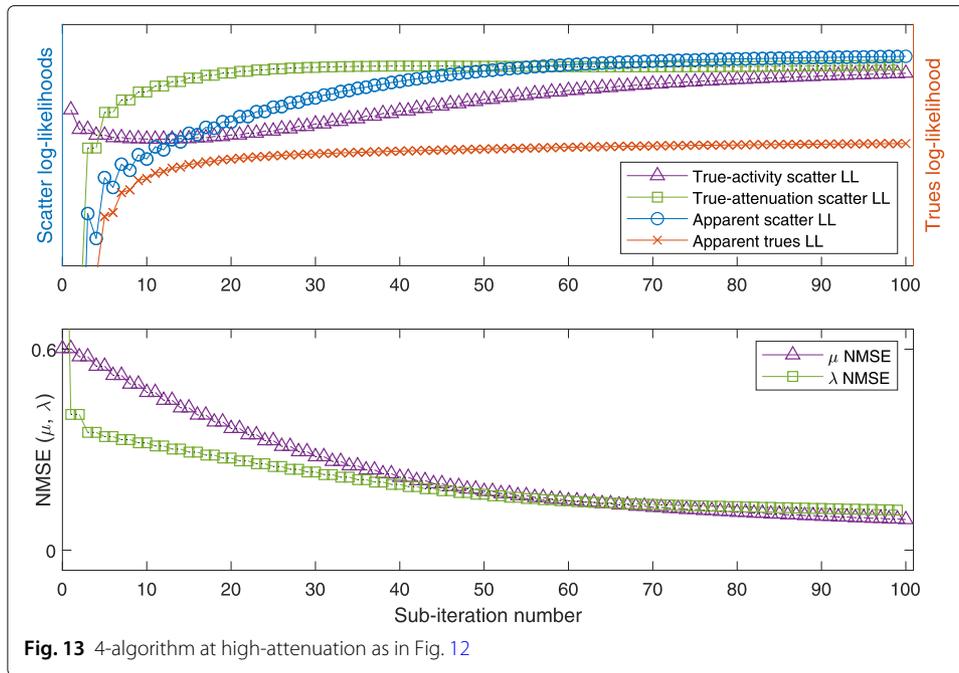




mention the possibilities of optimizing lower and upper energy photon thresholds in the detectors, redefining the detector’s transaxial acceptance angle, or selecting SORs intersecting specific parts of the subject [17, 24].

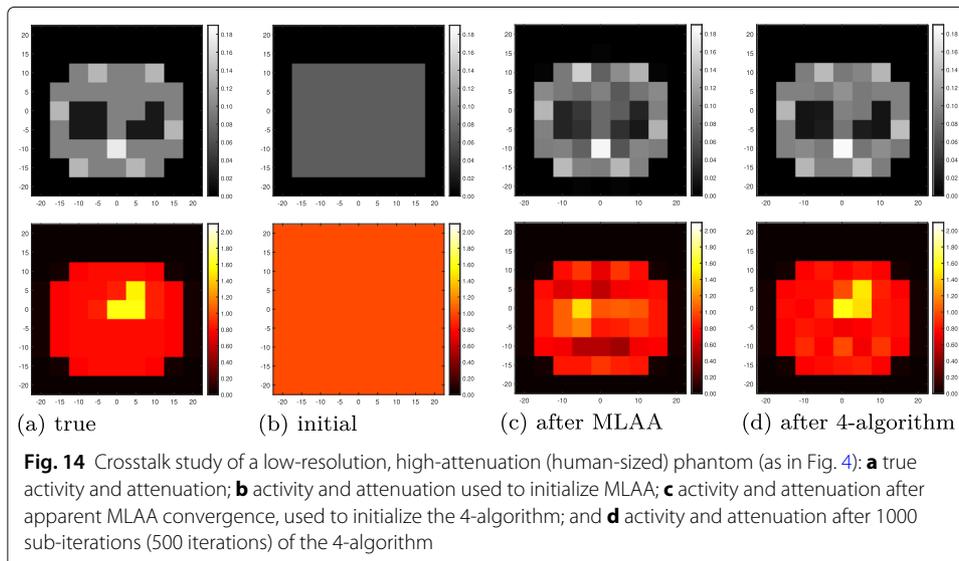
While MLGA converges, we find it to be more computationally complex than MLEM-OSL (Fig. 8). Also, MLEM-OSL with reduced data is more complex than using the full data: this could be remedied by stopping to re-evaluate (20) after some iterations.





Joint estimation

MLGA, or MLEM-OSL, are not primarily meant as stand-alone algorithms, as they assume knowledge of the unknown λ , mandating a joint (λ, ρ) estimation scheme. This scheme is similar to traditional MLAA for estimation of (λ, μ) from nonscattered coincidences, where knowledge of λ is assumed by MLTR. Along the same lines, we have focused on using MLGA in joint n -algorithms. Just as MLAA iterates back and forth between MLEM and MLTR, reconstructing one quantity (λ or μ) while keeping the other (μ or λ , respectively) fixed, our proposed 2-algorithm iterates back and forth between trues-MLEM and scatter-MLGA. Another viable scheme encompassing all available data has been presented in the form of the scatter-MLEM/trues-MLEM/MLGA/MLTR



4-algorithm, using both true and scattered coincidences for estimation of both λ and ρ .

For low-attenuation data, the simple 2-algorithm may be sufficient: in this case, MLEM-OSL may serve as a drop-in replacement for MLGA, however with decreased convergence speed (compare Figs. 10 and 11). This result is compatible with the results for S2A reconstruction (Fig. 6b and c). The more sophisticated 4-algorithm enables joint reconstruction with high-attenuation as well as low-attenuation data. In addition, this 4-algorithm can employ the scatter information to escape from a nonoptimal fix point of MLAA (Fig. 14d).

The plots of apparent (using estimated activity and attenuation) and ideal likelihoods indicate several nonmonotonicities that are overcome by the combination of algorithms. In particular, in the 2-algorithm, increasing the apparent scatter likelihood with MLGA updates of the attenuation often decreases the true-activity scatter likelihood, limiting the number of repeated MLGA updates that can be concatenated. This observation is one reason for choosing only a single update of each algorithm in each iteration of the 4-algorithm (Algorithm 1).

Limitations

This study has several limitations in the simplicity of the simulations: in particular, in neglecting detector scatter, multiple scatter, and energy-measurement uncertainties; in using the same forward model for the simulation as for the reconstruction; and in using low-dimensional objects and scanners.

In reality, the use of scattered photons is complicated by the fact that the detected signal of a nonscattered 511 keV photon, when it deposits only part of its energy in the detector, resembles that of a lower-energy, object-scattered photon [34]. One solution may be the use of an object-scatter energy window above the Compton edge (at 341 keV) and below the photopeak, which is virtually free of detector-scattered photons ([40], Fig. 1). This highest-possible energy window also has a lower contribution of multiply scattered photons [34].

Furthermore, photon energy measurements suffer from uncertainties in the range of 10% FWHM in state-of-the-art PET scanners. This uncertainty leads to blurred estimation of potential scattering locations in the object. So far, it is unclear exactly which energy resolution is required to successfully use this approach in practice, although some comparisons have been made for S2A reconstruction ([17], Fig. 11). In joint n -algorithms, separation of scattered and unscattered photons will be of importance in all sub-algorithms. New detector materials, such as LaBr₃ [41] or cadmium zinc telluride [42], might be needed.

Another limitation is that specific findings may not be generalizable to arbitrary scanner geometries; for example, 3-D geometries may exhibit different, presumably much sparser system quantities A_λ and \mathbf{K} . We expect that this increasing sparsity partially offsets the (otherwise unmanageable) size increase of these quantities with growing number of voxels, detectors, and energy bins.

It should be noted that due to computational complexity, 2-D considerations are not uncommon in recent studies regarding image reconstruction from scattered photons [20, 26, 43]. Furthermore, a more sophisticated imaging model that offers more realistic system matrix components A_λ and \mathbf{K} would be subject to the same measurement equations and lead to the same derivation of MLGA. So while specific convergence rates may vary

with the density and condition number of those quantities, we expect the overall conclusions to prevail in more realistic settings. Finally, noise will have to be considered in future studies; currently, it is challenging to determine realistic noise levels for these nonrealistic types of objects and scanners.

Outlook

In this paper, we have used MLGA as a S2A building block in the context of joint estimation. It might be possible to find improved algorithms: for example, one might pursue one of the many paths that lead to the MLEM update equation for an algorithm which features more of the well-known properties of MLEM. This may include the minorize/maximize (MM) algorithm [44], of which regular MLEM is one special case. Following earlier incomplete-data formulations [32], one might define complete data that involve not only the emission location, but also the scattering location of every coincidence; this may result in a formulation similar to that for joint estimation from nonscattered data [45]. Algorithms that use the formulation of the Hessian of the log-likelihood may also be of value without requiring inversion of the full Hessian during image reconstruction, as has been shown recently [25, 26].

TOF information might improve S2A reconstruction by increasing the sparsity of \mathbf{K} (as some scattering locations on the surface of the football may not be compatible with the emission locations indicated by a TOF measurement) and further improving the condition of A_λ (by reducing the number of emission voxels over which to compute $\sum_e b_{s,e}^i \lambda^e$). Our study does not simulate, or incorporate, TOF measurements, as the amount of additional information from TOF is nonetheless limited in attenuation reconstruction compared to activity reconstruction: even with perfect TOF information, the surface of potential scattering locations (and hence the density of \mathbf{K}) will hardly be reduced by more than a few times, on average, as most broken LORs compatible with a non-TOF coincidence will also be compatible with the TOF coincidence. Therefore, the primary way for TOF information to find its way into this problem may be through activity-reconstruction building blocks (TOF-trues-MLEM and TOF-scatter-MLEM) and the estimate of the activity distribution they provide—similar to how TOF-MLAA benefits from TOF information without the MLTR algorithm using it explicitly.

Regarding the impact of the results outside of PET imaging, we have achieved a definition of data being more or less compatible with a linearization of the measurement equation that may be applied in external Compton scatter imaging, in which a number of ways have been tried to solve a structurally similar measurement equation [28]—compare in particular (2) to ([46], Eq. 3) or ([47], Eq. 1). Furthermore, full knowledge of the “source” distribution is given in CT and other transmission imaging modalities, where scattered radiation could be similarly exploited if discerned by energy measurements, such as in multi-energy (spectral) CT [48].

Conclusion

In reconstruction of attenuation information from scattered PET coincidences, maximum-likelihood gradient-ascent algorithms provide faster convergence and convergence in more diverse setups than MLEM-OSL, for which we have presented both analytic and experimental evidence: MLGA converges across all spatial scales, while

MLEM-OSL may only converge with smaller objects. Nonetheless, MLEM-OSL can be a lower-complexity alternative to MLGA. We have defined a numerical criterion to determine when the simpler and more efficient MLEM-OSL can be used and described how its performance can be improved by reducing data based on said criterion. Finally, joint estimation of activity and attenuation from scattered and nonscattered coincidences has been presented using either MLGA or MLEM-OSL, in particular, in an example where MLAA fails to converge to the correct solution.

Appendix

Derivation of discrete measurement equations

In the previous notation of the measurement equation ([17], Eq. 5) and its derivation ([49], Eq. 11), the continuous measurement equation is:

$$\begin{aligned} \bar{N}^S(d_1, E_1, d_2) \approx & T \cdot \Delta E_1 \cdot \left(\frac{b}{2}\right)^3 \cdot \frac{(\cos \theta_1 - 2)^2}{E_0 \cdot \sin \theta_1} \cdot \left. \frac{d^2 \sigma_{\text{KN}}}{d^2 \Omega} \right|_{\theta_1} \cdot \int_{-\theta_1}^{+\theta_1} d\omega_S \int_0^{2\pi} d\varphi_S \\ & \left[\frac{(\cos \omega_S - \cos \theta_1)^2}{\sin^4 \theta_1} \cdot \rho_e(\vec{x}_S) \cdot \int_{L(\vec{x}_S, \vec{x}_2)} \lambda(\vec{x}_\lambda) d\vec{r}_\lambda^S \cdot \Omega_1(\vec{x}_S) \cdot \Omega_2(\vec{x}_S) \right. \\ & \cdot \exp \left(- \int_{L(\vec{x}_2, \vec{x}_S)} \mu(\rho_e(\vec{x})) d\vec{x} - \int_{L(\vec{x}_S, \vec{x}_1)} \mu_{E_1}(\rho_e(\vec{x})) d\vec{x} \right) \\ & \left. \cdot \mathcal{E}(d_1, E_1, \vec{x}_1 - \vec{x}_S) \cdot \mathcal{E}(d_2, E_0, \vec{x}_2 - \vec{x}_S) \right]. \end{aligned} \quad (23)$$

We refer to the original publications for more context of this equation: $\bar{N}^S(d_1, E_1, d_2)$, number of scattered coincidences detected in detectors d_1 (with energy E_1) and d_2 ; T , acquisition time; ΔE_1 , energy bin width; b , length of baseline connecting d_1 and d_2 ; θ_1 , scattering angle associated with E_1 ; E_0 , 511 keV; σ_{KN} , Klein-Nishina scattering cross-section; ω_S, φ_S , angles parameterizing the surface of scattering locations \vec{x}_S ; ρ_e , electron density; λ , radiotracer density; L , a line segment connecting two points; Ω_1, Ω_2 , detector solid angles seen from \vec{x}_S ; μ, μ_{E_1} , linear attenuation coefficient at E_0 and E_1 ; \mathcal{E} , photon detection sensitivity as a function of detector element, photon energy, and angle of incidence. While most terms have direct physical interpretations:

$$\left(\frac{b}{2}\right)^3 \cdot \frac{(\cos \omega_S - \cos \theta_1)^2}{\sin^4 \theta_1} \quad \text{and} \quad \frac{(\cos \theta_1 - 2)^2}{E_0 \cdot \sin \theta_1} \quad (24)$$

arise from the Jacobian determinant of the coordinate transformation from Cartesian coordinates to surface parameter angles and account for a change of variables from polar scatter angle to photon energy, respectively [49].

In discretizing and changing notation, we apply the following mapping:

$$i = (d_s, d_n, E) \longleftarrow (d_1, E_1, d_2) \quad (25)$$

$$\bar{y} \longleftarrow \bar{N}^S \quad (26)$$

$$c^i \longleftarrow T \cdot \Delta E_1 \cdot \left(\frac{b}{2}\right)^3 \cdot \frac{(\cos \theta_1 - 2)^2}{E_0 \cdot \sin \theta_1} \cdot \frac{d^2 \sigma_{\text{KN}}}{d^2 \Omega} \Big|_{\theta_1} \quad (27)$$

$$\sum_s \hat{c}_s^i \longleftarrow \int_{-\theta_1}^{+\theta_1} d\omega_S \int_0^{2\pi} d\varphi_S \quad (28)$$

$$\hat{c}_s^i \longleftarrow \frac{(\cos \omega_S - \cos \theta_1)^2}{\sin^4 \theta_1} \cdot \Omega_1(\vec{x}_S) \cdot \Omega_2(\vec{x}_S) \cdot \mathcal{E}(d_1, E_1, \vec{x}_1 - \vec{x}_S) \cdot \mathcal{E}(d_2, E_0, \vec{x}_2 - \vec{x}_S) \quad (29)$$

$$\rho^s \longleftarrow \rho_e(\vec{x}_S) \quad (30)$$

$$\sum_e \tilde{c}_{s,e}^i \longleftarrow \int_{L(\vec{x}_S, \vec{x}_2)} dr_{\lambda}^S \quad (31)$$

$$\lambda^e \longleftarrow \lambda(\vec{x}_\lambda) \quad (32)$$

$$\sum_t l_{s,t}^i \longleftarrow \int_{L(\vec{x}_2, \vec{x}_S)} dx + \int_{L(\vec{x}_S, \vec{x}_1)} (\mu_{E_1}(\rho_e(\vec{x})) / \mu(\rho_e(\vec{x}))) dx \quad (33)$$

$$\mu^t \longleftarrow \mu(\rho_e(\vec{x})) \quad (34)$$

Two remarks are in order. First, \hat{c}_s^i serves a double purpose of both defining the domain of summation over potential scattering locations \vec{x}_s (28) and including weighting factors (29). Due to the former, it can be stored efficiently as a sparse matrix; this holds true for $c_{s,e}^i$ and $l_{s,t}^i$ as well. Second, in the mapping (33), $l_{s,t}^i$ represents an *effective* attenuation length by encoding the differences in attenuation coefficients seen by scattered photons of different energies (μ_E at lower energy E) compared to nonscattered photons ($\mu = \mu_{511\text{keV}}$). In other words, a voxel can have different lengths along different broken LORs. This eliminates the need to store different attenuation coefficients for the same voxel, assuming that μ_E/μ is known. If one has $\mu \propto \rho$, that is, $\mu_E(\rho) = \alpha(E) \cdot \rho$, one finds

$$\frac{\mu_E}{\mu} = \frac{\alpha(E)}{\alpha(511 \text{ keV})}, \quad (35)$$

which can be determined knowing only E . Since i comprises the scattered photon energy E (25), and each broken LOR (i, s) defines which voxels t are part of the respective lower-energy section (Fig. 15a), the effective attenuation lengths $l_{s,t}^i$ can be computed. While this reasoning confirms that the shape of (2) does not impede taking into account energy-dependent attenuation, we did not do so here.

Compression and simplification

In the above way, fully accounting for all terms and operators in (23), we obtain:

$$\bar{y}^i = c^i \sum_s \hat{c}_s^i \cdot \rho^s \cdot \left(\sum_e \tilde{c}_{s,e}^i \cdot \lambda^e \right) \cdot \exp \left(- \sum_t l_{s,t}^i \cdot \mu^t \right). \quad (36)$$

While in terms of computational complexity, it may be beneficial to compute c^i , \hat{c}_s^i , and $\tilde{c}_{s,e}^i$ separately; with regard to storage and algorithmic simplicity, one may prefer combining all these terms into $b_{s,e}^i = c^i \cdot \hat{c}_s^i \cdot \tilde{c}_{s,e}^i$. Continuing to assume that $\mu \propto \rho$, we set

$$k_{s,t}^i := l_{s,t}^i \cdot \mu^t / \rho^t = l_{s,t}^i \cdot \alpha (511 \text{ keV}) \quad (37)$$

to find

$$\bar{y}^j = \sum_s \rho^s \cdot \left(\sum_e b_{s,e}^i \cdot \lambda^e \right) \cdot \exp \left(- \sum_t k_{s,t}^i \cdot \rho^t \right). \quad (38)$$

This equation forms the basis of both activity and attenuation reconstruction.

Activity reconstruction

For the simpler case of activity reconstruction [34], we consider $\vec{\rho}$ as a parameter for attenuation correction. Then, changing the order of summation in (38):

$$\bar{y}^j = \sum_e \lambda^e \cdot \left(\sum_s \rho^s \cdot b_{s,e}^i \cdot \exp \left(- \sum_t k_{s,t}^i \cdot \rho^t \right) \right), \quad (39)$$

we recognize that (39) can be written as a matrix-vector product:

$$\vec{\bar{y}} = \tilde{A}_\rho \vec{\lambda}, \quad (40)$$

where \tilde{A}_ρ is the matrix with entries $\tilde{a}_e^i := \sum_s \rho^s \cdot b_{s,e}^i \cdot \exp \left(- \sum_t k_{s,t}^i \cdot \rho^t \right)$. Knowing (an estimate of) ρ , the linear nature of (40) allows application of the regular MLEM algorithm [32] to reconstruct or update λ , as has been described by [34].

Attenuation reconstruction

Focusing on attenuation reconstruction, we consider $\vec{\lambda}$ as a parameter instead. Then, $b_{s,e}^i$ and λ^e can be combined into A_λ with entries $a_s^i := \sum_e b_{s,e}^i \cdot \lambda^e$, yielding

$$\bar{y}^j = \sum_s \rho^s \cdot a_s^i \cdot \exp \left(- \sum_t k_{s,t}^i \cdot \rho^t \right). \quad (41)$$

By using element-wise operations (\odot and $\overset{\circ}{\exp}$) and by defining 3rd-order tensors \underline{B} and \underline{K} and the matrix $A_\lambda := \underline{B} \vec{\lambda}$, (38) and (41) can also be written as:

$$\vec{\bar{y}} = \left(\left(\underline{B} \vec{\lambda} \right) \odot \overset{\circ}{\exp} \left(- \underline{K} \vec{\rho} \right) \right) \vec{\rho} = \left(A_\lambda \odot \overset{\circ}{\exp} \left(- \underline{K} \vec{\rho} \right) \right) \vec{\rho}, \quad (42)$$

respectively, where the latter is (2) as was to show.

Unscattered coincidences

For (mostly) unscattered data, where the one value of E represents the photopeak energy window, an SOR reduces to an LOR, and \tilde{a}_e^i is replaced by \tilde{u}_j^i for activity reconstruction; similarly, in attenuation reconstruction, a_s^i is replaced by u_j^i .

Implementation using sparse matrices

\underline{K} and A_λ are highly sparse, and so is $\underline{K} \vec{\rho}$. However, since $\exp 0 = 1$, $\exp(-\underline{K} \vec{\rho})$ is not sparse, which impedes storage of intermediate results as sparse matrices. Hence, we rewrite:

$$A_\lambda \odot \overset{\circ}{\exp} \left(- \underline{K} \vec{\rho} \right) = A_\lambda \odot \left(\overset{\circ}{\exp} \left(- \underline{K} \vec{\rho} \right) - \vec{1}_{[i]} \otimes \vec{1}_{[s]} \right) + A_\lambda \quad (43)$$

and use the `expm1` function [50] to compute the sparse $\mathbf{S} := \exp(-\mathbf{K}\bar{\rho}) - \bar{\mathbf{1}}_{[i]} \otimes \bar{\mathbf{1}}_{[s]}$, followed by $\mathbf{A}_\lambda \odot \mathbf{S} + \mathbf{A}_\lambda$. This maintains the sparsity in all intermediate results along the memory-efficient evaluation of (42).

Multi-voxel interpretations of high attenuation

Motivated by the 1-D results around (17), we are interested in the sign of the gradient of the expected data. In component-wise notation, (12) reads:

$$\partial \bar{y}^i / \partial \rho^j = \sum_s \left(\delta_{sj} - \rho^s k_{s,j}^i \right) a_s^i \exp \left(- \sum_t k_{s,t}^i \rho^t \right), \quad (44)$$

with δ_{sj} the Kronecker delta. Based on the sign of this expression, we can identify distinct regions of the football-shaped SOR in Fig. 1. Figure 15a shows half a cross-sectional plane through the football,⁸ determined by d_s , s , and d_n . On the inside and outside of the football, the sign is determined mainly by the quantities a_j^i and $k_{s,j}^i$, which represent the scattering contributions of a voxel j along an SOR i and the attenuating contributions of a voxel j along broken LORs (i, s) , respectively.

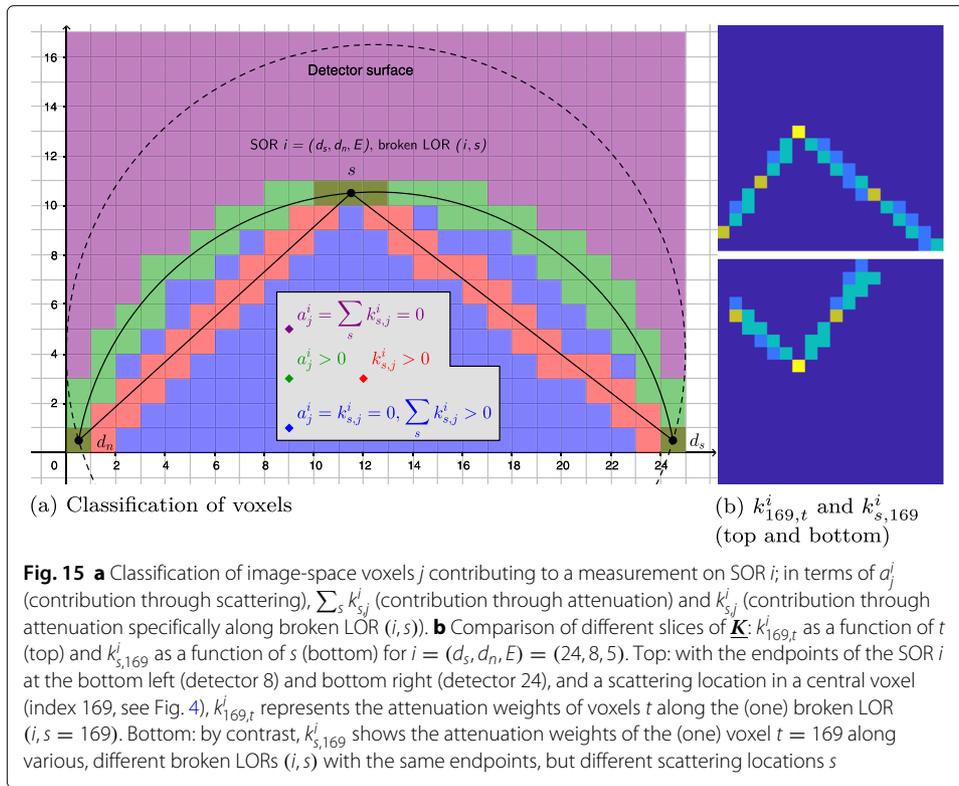
Outside When both $a_j^i = 0$ and $\sum_s k_{s,j}^i = 0$ ($\Leftrightarrow k_{s,j}^i = 0 \forall s$), we find $\partial \bar{y}^i / \partial \rho^j = 0$. The voxel j does not contribute to a measurement along i , neither through scattering nor through attenuation; in Fig. 15a, this region of voxels j corresponds to the violet area outside the football. Changing ρ^j does not influence \bar{y}^i .

Inside The football's strict inside contributes to a measurement along SOR i only through attenuation. We have $a_j^i = 0$ (voxel j not contributing scatter to this SOR), but $\sum_s k_{s,j}^i > 0$ (voxel j attenuating on at least one broken LOR of this SOR). These are red and blue areas in Fig. 15a: red highlights where $k_{s,j}^i > 0$ for one particular s ; blue areas where $k_{s,j}^i = 0$, but $k_{s',j}^i > 0$ for at least one other $s' \neq s$. The result is $\partial \bar{y}^i / \partial \rho^j < 0$: increasing ρ in j decreases the expected number of coincidences on i .

Thus, when taking these attenuating effects into account, e.g., in an iterative update, it would be more appropriate to speak of a *volume* rather than a surface of response; we still hold on to the term SOR here.

Surface Defined by $a_j^i > 0$ and marked green in Fig. 15a, the football's hull is the most interesting segment. These voxels are the only ones with scattering contributions, with potentially additional attenuating contributions (overlap between green and red regions). If we had $\sum_s \rho^s k_{s,j}^i = 0$, then $\partial \bar{y}^i / \partial \rho^j \geq 0$ since j would only contribute through scattering. However, with $a_j^i > 0$, we certainly have $k_{j,j}^i > 0$: a scattering voxel is always on the broken LOR through it, and hence attenuating. Thus, if $\rho^j > 0$ (which means j lies within the object), the sign of $\partial \bar{y}^i / \partial \rho^j$ is more complex to determine. Besides that, we have only little general information about $k_{s,j}^i$ for $s \neq j$; this greatly depends on the object (low vs. high attenuation), the geometry of i and (i, s) , as well as the discretization strategy.

⁸This segmentation of the image space offers a geometric interpretation of the different back-projection kernels in the two branches of the 2BP update ([24], Eq. 17): the positive branch distributes an excess of measured coincidences (compared to expected ones) on SOR i by increasing ρ on the surface of the football (in voxels j where $a_j^i > 0$) and by decreasing ρ there for a lack of measured coincidences. By contrast, the negative branch decreases ρ (and hence μ) on the inside of the football (where $\sum_s k_{s,j}^i > 0$) for an excess of measured coincidences, and vice versa.



We therefore aim for a definition of “high attenuation.” For now, we will therefore ignore differences between voxels that are both inside the object ($\rho^j > 0$) and on the surface of SOR i ($a_j^i > 0$), and refer to them by their number n and averages \bar{a} , \bar{k} , and $\bar{\rho}$. This way, we simplify (44) to:

$$\partial \bar{y}^i / \partial \rho^j \approx \bar{a} \exp(-n \bar{k} \bar{\rho}) \sum_{s: a_s^i > 0} (\delta_{sj} - \rho^s k_{s,j}^i). \quad (45)$$

By setting $k_{s,j}^i = 0 \forall j$ when $a_s^i = 0$ (no attenuation along broken LORs through scattering voxels s geometrically incompatible with an SOR i), we simplify the sum over s : on the hull ($a_j^i > 0$), the expression determining the sign of (44) is then:

$$1 - \sum_s \rho^s k_{s,j}^i \iff \bar{\mathbf{I}}_{[i]} \otimes \bar{\mathbf{I}}_{[t]} - \bar{\rho}^\top \underline{K}, \quad (46)$$

the multidimensional (size of a system matrix) analog of $1 - k\rho$.

Note that $\underline{K} \bar{\rho}$ (or $\sum_t k_{j,t}^i \rho^t$; used in (2) and the attenuation-factor expressions of (11 and 12)) is a matrix of radiological paths, so in terms of units of measurement, so is $\bar{\rho}^\top \underline{K}$ (or $\sum_s \rho^s k_{s,j}^i$; used in (11) and, in the following, (20)). However, there are fundamental differences: the former is the weighted sum of the red voxels in Fig. 15a [over all voxels t that attenuate for one specific broken LOR (i, s) , weighted by ρ^t], and as such, is the line integral of $\bar{\rho}$ (discrete) attenuation coefficient along the broken LOR. By contrast, the latter has a more complex geometrical interpretation (see Fig. 15b, bottom): it is the sum of the attenuating contributions of the *same* voxel t to all possible broken LORs (i, s) , weighted by ρ^s , the electron density in each broken LOR’s scattering voxel s . Low attenuation, in this case, corresponds to a small number of scattering voxels which are impacted by one attenuating voxel j , and thus a small overlap

between scattering (green) and attenuating (red, blue) voxels across all broken LORs of an SOR.

In the minimal overlap case depicted in Fig. 15a, a voxel j on the football's surface can have attenuating contributions to one of only three locations of each broken LOR (i, s) , namely at d_s , s , and d_n (where d_s and d_n will be outside of the unknown object). However, significant additional overlap may be the result of a number of factors: SORs with low curvature, that is, small scattering angle θ and large SOR radius $R = b/(2 \sin \theta)$; scattering points s close to d_s or d_n ; consideration of energy uncertainty in A_λ ; large subjects; and nonrectangular grids and large voxel sizes. An extreme example is a barely scattered coincidence ($E \approx 511$ keV, $\theta \approx 0$, $R \rightarrow \infty$) on an LOR-like SOR i , along which, when $a_j^i > 0$, we find that $a_s^i > 0$ implies $k_{s,j}^i > 0$. So each voxel j attenuates (with its full effective length, $k_{s,j}^i \approx k_{j,j}^i$) contributions from every scattering voxel; in other words, contributions from any scattering voxel s are attenuated by a voxel j . In this case, (46) reads:

$$1 - k_{j,j}^i \sum_{s: a_s^i > 0} \rho^s = 1 - l_{j,j}^i \sum_{s: a_s^i > 0} \mu^s = 1 - l_{j,j}^i n_i \bar{\mu}, \quad (47)$$

where n_i is the number of object voxels part of SOR i . With $l_{j,j}^i$ on the order of the voxel size and $\bar{\mu} \approx 0.1/\text{cm}$ in water, (47) has the same sign as $(10\text{cm} - L_i)$, where L_i is the path length through the object. So for very tight SORs, the interpretations of $\bar{\rho}^\top \underline{K}$ and $\underline{K} \bar{\rho}$ are similar, and the SOR through a water-like body of more than 10-cm thickness has $\partial \bar{y}^i / \partial \rho^j < 0$; this is where the linear approximation underlying MLEM-OSL fails. The situation is more complex for other SORs.

Added value of combining nonscattered and scattered data

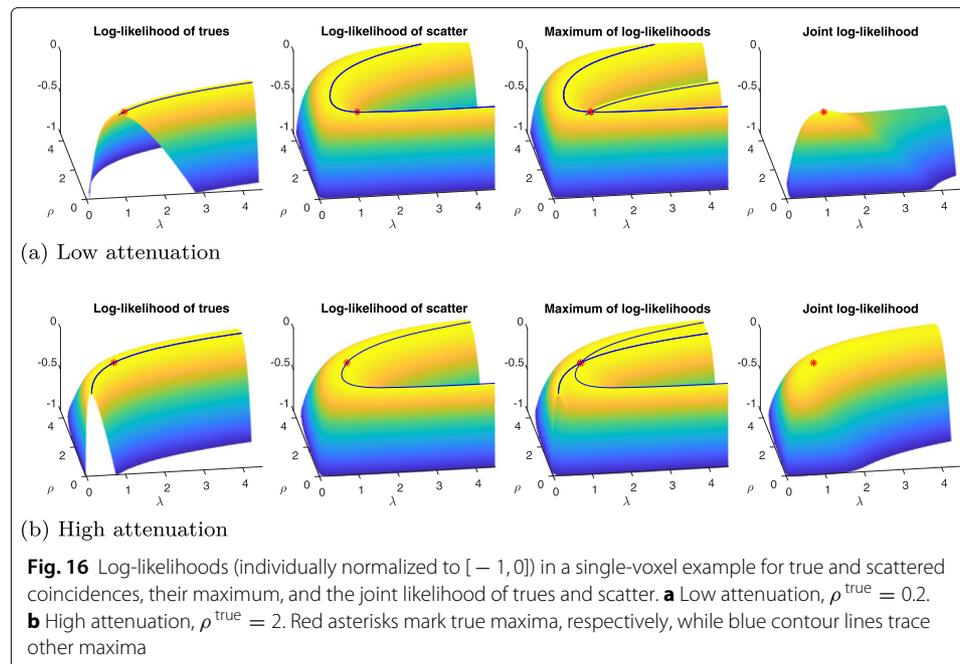
We focus on an artificial single-voxel/single-detector-pair problem, with the expected true and scattered data:

$$\bar{z} = (u\lambda) \exp(-u\rho) \quad \text{and} \quad \bar{y} = (b\lambda) \exp(-k\rho)\rho, \quad (48)$$

respectively (compare (4) and (2) for multi-voxel variants). For a set of (λ, ρ) candidates, we plot the log-likelihood of trues and scatter as well as the joint log-likelihood, using system values $b = k = u = 1$ and true values $\lambda^{\text{true}} = 1$ and $\rho^{\text{true}} \in \{0.2, 2\}$ (low and high attenuation, respectively), chosen such that $0.2 \exp(-0.2) \approx 0.16$ is in the range of $2 \exp(-2) \approx 0.27$.

Figure 16 illustrates the added value of combining nonscattered (true) and scattered data, as compared to using only trues (MLAA) or only scatter (MLGA and scatter-MLEM). In both the low and the high attenuation example, the likelihood of trues exhibits the well-known scaling issue that can be appreciated in the form of an extended nonunique maximum; a similar effect is seen with the likelihood of scatter⁹. In this single-voxel example, the likelihood of scatter exhibits an additional nonuniqueness: for each value of λ , two values of ρ yield the same scatter likelihood, in line with two solutions of the 1-D scatter measurement equation (see Fig. 2); however, this nonuniqueness is more difficult to characterize with multiple voxels.

⁹Note that the scatter log-likelihoods are different between low and high attenuation; they appear similar to each other due to the choice of ρ^{true} values.



The intersection of the curves that follow the maxima features a unique intersection, which coincides with the maximum of the joint likelihood of trues and scatter data. The peak around this maximum is broader for high attenuation, in line with the lower angle of intersection between the maxima of the individual likelihoods.

Acknowledgements

We thank the anonymous reviewers for their insight and help in improving the focus of this paper.

Authors' contributions

YB designed the study, selected and implemented the algorithms, implemented the data simulation, analyzed and interpreted the data, and drafted the manuscript. VS and JSK critically revised the manuscript. All authors read and approved the final manuscript.

Funding

The work leading to this publication was supported by the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research (BMBF) and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 605728 (P.R.I.M.E. – Postdoctoral Researchers International Mobility Experience), by the National Cancer Institute of the National Institutes of Health under award number R01CA113941 and by a German Cancer Research Center (DKFZ) Postdoctoral Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Two patent applications relevant to this work have been filed by Koninklijke Philips Electronics N.V. with YB and VS as inventors, both of which have been granted.

Author details

¹Division of X-ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. ²Department of Physics of Molecular Imaging Systems, Institute for Experimental Molecular Imaging, RWTH Aachen University, Forckenbeckstraße 55, 52074 Aachen, Germany. ³Department of Radiology,

University of Pennsylvania, 3620 Hamilton Walk, Philadelphia, PA 19104, USA. ⁴III. Physikalisches Institut B, RWTH Aachen University, Otto-Blumenthal-Straße, Aachen, Germany. ⁵Fraunhofer Institute for Digital Medicine MEVIS, Forckenbeckstraße 55, 52074 Aachen, Germany.

Received: 11 February 2019 Accepted: 10 September 2019

Published online: 28 October 2019

References

1. Phelps ME. PET: the merging of biology and imaging into molecular imaging. *J Nucl Med.* 2000;41(4):661–81.
2. Phelps ME, Hoffman EJ, Mullani NA, Ter-Pogossian MM. Application of annihilation coincidence detection to transaxial reconstruction tomography. *J Nucl Med.* 1975;16(3):210–24.
3. Watson CC. New, faster, image-based scatter correction for 3D PET. *IEEE Trans Nucl Sci.* 2000;47(4):1587–94. <https://doi.org/10.1109/23.873020>.
4. Carson RE, Daube-Witherspoon ME, Green MV. A method for postinjection PET transmission measurements with a rotating source. *J Nucl Med.* 1988;29(9):1558–67.
5. Karp JS, Muehllehner G, Qu H, Yan X-H. Singles transmission in volume-imaging PET with a ¹³⁷Cs source. *Phys Med Biol.* 1995;40(5):929–44. <https://doi.org/10.1088/0031-9155/40/5/014>.
6. Kinahan PE, Townsend DW, Beyer T, Sashin D. Attenuation correction for a combined 3D PET/CT scanner. *Med Phys.* 1998;25(10):2046–53. <https://doi.org/10.1118/1.598392>.
7. Keereman V, Mollet P, Berker Y, Schulz V, Vandenberghe S. Challenges and current methods for attenuation correction in PET/MR. *Magn Reson Mater Phys Biol Med.* 2013;26(1):81–98. <https://doi.org/10.1007/s10334-012-0334-7>.
8. Berker Y, Franke J, Salomon A, Palmowski M, Donker HCW, Temur Y, Mottaghy FM, Kuhl C, Izquierdo-Garcia D, Fayad ZA, Kiessling F, Schulz V. MRI-based attenuation correction for hybrid PET/MRI systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence. *J Nucl Med.* 2012;53(5):796–804. <https://doi.org/10.2967/jnumed.111.092577>.
9. Nuyts J, Dupont P, Stroobants S, Banninck R, Mortelmans L, Suetens P. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Trans Med Imaging.* 1999;18(5):393–403. <https://doi.org/10.1109/42.774167>.
10. Rezaei A, Defrise M, Bal G, Michel C, Conti M, Watson C, Nuyts J. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. *IEEE Trans Med Imaging.* 2012;31(12):2224–33. <https://doi.org/10.1109/TMI.2012.2212719>.
11. Rezaei A, Deroose CM, Vahle T, Boada F, Nuyts J. Joint reconstruction of activity and attenuation in time-of-flight PET: a quantitative analysis. *J Nucl Med.* 2018;59(10):1630–5. <https://doi.org/10.2967/jnumed.117.204156>.
12. Defrise M, Rezaei A, Nuyts J. Transmission-less attenuation correction in time-of-flight PET: analysis of a discrete iterative algorithm. *Phys Med Biol.* 2014;59(4):1073–95. <https://doi.org/10.1088/0031-9155/59/4/1073>.
13. Rezaei A, Defrise M, Nuyts J. ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors. *IEEE Trans Med Imaging.* 2014;33(7):1563–72. <https://doi.org/10.1109/TMI.2014.2318175>.
14. Salvo K, Defrise M. sMLACF: a generalized expectation-maximization algorithm for TOF-PET to reconstruct the activity and attenuation simultaneously. *Phys Med Biol.* 2017;62(21):8283–313. <https://doi.org/10.1088/1361-6560/aa82ea>.
15. Defrise M, Rezaei A, Nuyts J. Time-of-flight PET data determine the attenuation sinogram up to a constant. *Phys Med Biol.* 2012;57(4):885–99. <https://doi.org/10.1088/0031-9155/57/4/885>.
16. Berker Y, Li Y. Attenuation correction in emission tomography using the emission data—a review. *Med Phys.* 2016;43(2):807–32. <https://doi.org/10.1118/1.4938264>.
17. Berker Y, Kiessling F, Schulz V. Scattered PET data for attenuation-map reconstruction in PET/MRI. *Med Phys.* 2014;41(10):102502. <https://doi.org/10.1118/1.4894818>.
18. Zhang G, Sun H, Pistorius S. Electron density reconstruction from scattered coincidences for attenuation correction in positron emission tomography. In: Proc 13th Intern Meeting Fully Three-Dimens Image Reconstr Radiol Nucl Med; 2015. p. 526–9.
19. Cade SC, Arridge S, Evans MJ, Hutton BF. Use of measured scatter data for the attenuation correction of single photon emission tomography without transmission scanning. *Med Phys.* 2013;40(8):. <https://doi.org/10.1118/1.4812686>.
20. Courdurier M, Monard F, Osses A, Romero F. Simultaneous source and attenuation reconstruction in SPECT using ballistic and single scattering data. *Inverse Probl.* 2015;31(9):095002. <https://doi.org/10.1088/0266-5611/31/9/095002>.
21. Bousse A, Sidlesky A, Roth N, Rashidnasab A, Thielemans K, Hutton BF. Joint activity/attenuation reconstruction in SPECT using photopeak and scatter sinograms. 2016 IEEE Nucl Sci Symp Med Imaging Conf. 2016. <https://doi.org/10.1109/NSSMIC.2016.8069448>.
22. Cueva E, Osses A, Quintana JC, Tejos C, Courdurier M, Irarrazaval P. Algebraic reconstruction of source and attenuation in SPECT using first scattering measurements. In: Hofmann B, Leitão A, Zubelli J, editors. *New Trends in Parameter Identification for Mathematical Models. Trends in Mathematics.* Birkhäuser, Cham. p. 53–66. https://doi.org/10.1007/978-3-319-70824-9_3.
23. Berker Y, Schulz V, Karp JS. Discrete iterative algorithms for scatter-to-attenuation reconstruction in PET. 2016 IEEE Nucl Sci Symp Med Imaging Conf. 2016. <https://doi.org/10.1109/NSSMIC.2016.8069455>.
24. Berker Y, Karp JS, Schulz V. Numerical algorithms for scatter-to-attenuation reconstruction in PET: empirical comparison of convergence, acceleration, and the effect of subsets. *IEEE Trans Radiat Plasma Med Sci.* 2017;1(5):426–34. <https://doi.org/10.1109/TNS.2017.2713521>.
25. Brusaferrri L, Bousse A, Efthimiou N, Emond E, Atkinson D, Ourselin S, Hutton BF, Arridge S, Thielemans K. Potential benefits of incorporating energy information when estimating attenuation from PET data. 2017 IEEE Nucl Sci Symp Med Imaging Conf. 2017. <https://doi.org/10.1109/NSSMIC.2017.8532765>.

26. Brusaferrri L, Bousse A, Tsai Y.-J., Atkinson D, Ourselin S, Hutton BF, Arridge S, Thielemans K. Maximum-likelihood estimation of emission and attenuation images in 3D PET from multiple energy window measurements. 2018 IEEE Nucl Sci Symp Med Imaging Conf. 2018. <https://doi.org/10.1109/nssmic.2018.8824557>.
27. Redler G, Jones KC, Templeton A, Bernard D, Turian J, Chu JCH. Compton scatter imaging: a promising modality for image guidance in lung stereotactic body radiation therapy. *Med Phys*. 2018;45(3):1233–40. <https://doi.org/10.1002/mp.12755>.
28. Hussein EMA. On the intricacy of imaging with incoherently-scattered radiation. *Nucl Instrum Methods Phys Res Sect B*. 2007;263(1):27–31. <https://doi.org/10.1016/j.nimb.2007.04.037>.
29. Berker Y, Karp JS, Schulz V. Joint reconstruction of PET attenuation and activity from scattered and unscattered data. 2017 IEEE Nucl Sci Symp Med Imaging Conf. 2017. <https://doi.org/10.1109/NSSMIC.2017.8532746>.
30. Kinahan PE, Hasegawa BH, Beyer T. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Semin Nucl Med*. 2003;33(3):166–79. <https://doi.org/10.1053/snuc.2003.127307>.
31. Kanematsu N, Inaniwa T, Nakao M. Modeling of body tissues for Monte Carlo simulation of radiotherapy treatments planned with conventional x-ray CT systems. *Phys Med Biol*. 2016;61(13):5037–50. <https://doi.org/10.1088/0031-9155/61/13/5037>.
32. Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging*. 1982;1(2):113–22. <https://doi.org/10.1109/TMI.1982.4307558>.
33. Nuyts J, De Man B, Dupont P, Defrise M, Suetens P, Mortelmans L. Iterative reconstruction for helical CT: a simulation study. *Phys Med Biol*. 1998;43(4):729–37. <https://doi.org/10.1088/0031-9155/43/4/003>.
34. Conti M, Hong I, Michel C. Reconstruction of scattered and unscattered PET coincidences using TOF and energy information. *Phys Med Biol*. 2012;57(15):307–17. <https://doi.org/10.1088/0031-9155/57/15/N307>.
35. Green PJ. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans Med Imaging*. 1990;9(1):84–93. <https://doi.org/10.1109/42.52985>.
36. Kaufman L. Maximum likelihood, least squares, and penalized least squares for PET. *IEEE Trans Med Imaging*. 1993;12(2):200–14. <https://doi.org/10.1109/42.232249>.
37. The MathWorks: radon. In: *Image Processing Toolbox™ Reference*. R2019a edn. Natick: The MathWorks; 2019. p. 2672–6.
38. Defrise M. Simultaneous estimation of activity and attenuation in classical PET: uniqueness issues. 2016 IEEE Nucl Sci Symp Med Imaging Conf. 2016. <https://doi.org/10.1109/NSSMIC.2016.8069452>.
39. Natterer F. Computerized tomography with unknown sources. *SIAM J Appl Math*. 1983;43(5):1201–12. <https://doi.org/10.1137/0143079>.
40. Popescu LM. PET energy-based scatter estimation in the presence of randoms, and image reconstruction with energy-dependent scatter and randoms corrections. *IEEE Trans Nucl Sci*. 2012;59(5):1958–66. <https://doi.org/10.1109/TNS.2012.2201170>.
41. Daube-Witherspoon ME, Surti S, Perkins A, Kyba CCM, Wiener R, Werner ME, Kulp R, Karp JS. The imaging performance of a LaBr₃-based PET scanner. *Phys Med Biol*. 2010;55(1):45–64. <https://doi.org/10.1088/0031-9155/55/1/004>.
42. Gu Y, Matteson JL, Skelton RT, Deal AC, Stephan EA, Duttweiler F, Gasaway TM, Levin CS. Study of a high-resolution, 3D positioning cadmium zinc telluride detector for PET. *Phys Med Biol*. 2011;56(6):1563–84. <https://doi.org/10.1088/0031-9155/56/6/004>.
43. Hemmati H, Kamali-Asl A, Ay M, Ghafarian P. Compton scatter tomography in TOF-PET. *Phys Med Biol*. 2017;62(19):7641–58. <https://doi.org/10.1088/1361-6560/aa82ab>.
44. Lange K. The MM algorithm. In: *Optimization*. Springer Texts in Statistics, vol. 95. 2nd edn. New York: Springer; 2013. p. 185–219. https://doi.org/10.1007/978-1-4614-5838-8_8.
45. Mihlin A, Levin CS. An expectation maximization method for joint estimation of emission activity distribution and photon attenuation map in PET. *IEEE Trans Med Imaging*. 2017;36(1):214–24. <https://doi.org/10.1109/TMI.2016.2602339>.
46. Hussein EMA, Bowles JTC. Expanding the domain of contraction mapping in the inverse problem of imaging with incoherently scattered radiation. *Inverse Probl*. 2007;23(3):1289–309. <https://doi.org/10.1088/0266-5611/23/3/026>.
47. Hussein EMA, Enjilela E. Transmission-like calibration-free tomographic reconstruction with Compton-scattered photons. *IEEE Trans Nucl Sci*. 2016;63(5):2586–98. <https://doi.org/10.1109/TNS.2016.2599152>.
48. Danad I, Fayad ZA, Willemink MJ, Min JK. New applications of cardiac computed tomography: Dual-energy, spectral, and molecular CT imaging. *JACC: Cardiovasc Imaging*. 2015;8(6):710–23. <https://doi.org/10.1016/j.jcmg.2015.03.005>.
49. Berker Y, Schulz V. Scattered PET data for attenuation-map reconstruction in PET/MRI: Fundamentals. 2014 IEEE Nucl Sci Symp Med Imaging Conf. 2014. <https://doi.org/10.1109/NSSMIC.2014.7430785>.
50. The MathWorks: expm1. In: *MATLAB® Function Reference*. R2019a edn. Natick: The MathWorks; 2019. p. 3712–3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.