

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Radiation image reconstruction and uncertainty quantification using a Gaussian process prior

### Permalink

<https://escholarship.org/uc/item/4xd4z06c>

### Journal

Scientific Reports, 14(1)

### ISSN

2045-2322

### Authors

Lee, Jaewon  
Joshi, Tenzing H  
Bandstra, Mark S  
et al.

### Publication Date

2024

### DOI

10.1038/s41598-024-71336-z

Peer reviewed



OPEN

# Radiation image reconstruction and uncertainty quantification using a Gaussian process prior

Jaewon Lee<sup>1✉</sup>, Tenzing H. Joshi<sup>2</sup>, Mark S. Bandstra<sup>2</sup>, Donald L. Gunter<sup>3</sup>, Brian J. Quiter<sup>2</sup>, Reynold J. Cooper<sup>2</sup> & Kai Vetter<sup>1,2</sup>

We propose a complete framework for Bayesian image reconstruction and uncertainty quantification based on a Gaussian process prior (GPP) to overcome limitations of maximum likelihood expectation maximization (ML-EM) image reconstruction algorithm. The prior distribution is constructed with a zero-mean Gaussian process (GP) with a choice of a covariance function, and a link function is used to map the Gaussian process to an image. Unlike many other maximum a posteriori approaches, our method offers highly interpretable hyperparameters that are selected automatically with the empirical Bayes method. Furthermore, the GP covariance function can be modified to incorporate a priori structural priors, enabling multi-modality imaging or contextual data fusion. Lastly, we illustrate that our approach lends itself to Bayesian uncertainty quantification techniques, such as the preconditioned Crank–Nicolson method and the Laplace approximation. The proposed framework is general and can be employed in most radiation image reconstruction problems, and we demonstrate it with simulated free-moving single detector radiation source imaging scenarios. We compare the reconstruction results from GPP and ML-EM, and show that the proposed method can significantly improve the image quality over ML-EM, all the while providing greater understanding of the source distribution via the uncertainty quantification capability. Furthermore, significant improvement of the image quality by incorporating a structural prior is illustrated.

Radiation imaging plays a crucial role in numerous applications including medical diagnostic imaging<sup>1</sup>, non-destructive testing<sup>2</sup>, nuclear contamination remediation, and nuclear safeguards<sup>3</sup>. To this end, various radiation imaging modalities have been proposed, developed, and used in the past, with the goal of imaging the spatial variation of relevant properties such as the radioactivity distribution or attenuation coefficients of intervening materials. However, these properties are difficult to measure directly without physical modulations and subject to statistical noise. As a result, image reconstruction, in which the image of interest is recovered from indirect, noisy measurements, is a crucial part of radiation imaging.

Image reconstruction is often formulated as a statistical inverse problem and conventionally, it is solved with the maximum likelihood estimation (MLE) method<sup>4</sup>. However, the noisy nature of the measurements, combined with various resource constraints in the measurement process often leads to solutions that are unstable and non-unique (i.e., it is an ill-posed inverse problem), adding complexity to the task of accurately interpreting the reconstructed images. Hence, an image reconstruction technique capable of improving the MLE solution and estimating uncertainties in the reconstructed images is of significant value.

Bayesian image reconstruction, in which a priori information is incorporated into an image reconstruction problem, is an attractive alternative to the MLE approach<sup>5</sup>. In the Bayesian paradigm, image reconstruction is often posed as a maximum a posteriori (MAP) problem, where the image is recovered by locating the mode of the posterior distribution. Various MAP techniques have been studied extensively in the context of image reconstruction, ranging from Tikhonov regularization<sup>6</sup>, total variation (TV)<sup>7</sup> regularization to plug-and-play priors<sup>8</sup> that utilize pre-trained deep neural networks. It has been shown that the MAP approaches can significantly improve the quality of reconstructed images compared to the MLE, while alleviating the ill-posedness of the image reconstruction problem<sup>9</sup>. Moreover, unlike frequentist approaches such as MLE, the Bayesian paradigm offers a straightforward interpretation of uncertainties, including both aleatoric uncertainties, arising from the inherent randomness in data, and epistemic uncertainties, resulting from a lack of data<sup>10</sup>.

<sup>1</sup>Department of Nuclear Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. <sup>2</sup>Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>3</sup>Gunter Physics, Inc., Lisle, IL 60532, USA. ✉email: jwonlee@berkeley.edu

However, the Bayesian image reconstruction paradigm often faces several challenges. First, the hyperparameters (i.e., regularization parameters) reflecting a priori knowledge are often difficult to choose in a principled manner<sup>11</sup>. Secondly, even with carefully selected hyperparameters, the prior model itself may not be flexible enough to model the a priori knowledge that needs to be incorporated. Lastly, Bayesian uncertainty quantification of reconstructed images require very high dimensional integration, and analytical solutions are generally unavailable unless conjugate priors are employed<sup>12</sup>.

To tackle the aforementioned challenges, we propose the use of a Gaussian process (GP)-based prior for image reconstruction and uncertainty quantification tasks. A GP is a stochastic process where the constituent random variables jointly follow a multivariate Gaussian distribution<sup>13</sup>. The GPs can be made very flexible with a choice of mean and covariance functions, and they have been extensively studied in the context of regression tasks<sup>14</sup>. In this work, we demonstrate that GPs can also serve as a powerful prior for Bayesian image reconstruction tasks. The proposed approach to use a GP prior will be referred to as Gaussian process prior (GPP) henceforth. The following are the contributions of our work:

- First, we demonstrate that the MAP estimate with the proposed GPP can significantly enhance the quality of the reconstructed radiation images compared to the conventional MLE-based approach. The computational cost in using the GP is greatly eased by exploiting the Kronecker product structure in the GP covariance matrix.
- We then show that uncertainties in the reconstructed images can be quantified using the preconditioned Crank–Nicolson (pCN) algorithm<sup>15</sup> and the Laplace approximation<sup>16,17</sup>.
- We illustrate that the flexibility of the GPP allows for incorporation of structural prior, which can dramatically improve the reconstruction in some imaging scenarios.
- Lastly, we lay out a principled method for choosing the hyperparameters of the GPP, using the empirical Bayes method and Laplace approximation.

We note that a Bayesian approach using a GP prior for image reconstruction and uncertainty quantification was recently introduced by Zhou et al.<sup>18</sup>, where a combination of a TV and a GP prior and the pCN-based uncertainty quantification method were proposed. However, due to the non-differentiable nature of the TV prior, it limits the use of gradient-based optimization and Hessian information. In contrast, in this work, we demonstrate that the Laplace approximation, which leverages the Hessian of the posterior, enables scalable, yet effective uncertainty quantification and hyperparameter selection.

The proposed framework is general and can be used for most radiation imaging modalities with Poisson likelihood and (with a small modification) Gaussian likelihood. In this work, we show the effectiveness of the proposed method with free-moving single detector imaging scenarios, which is based on the recently developed scene data fusion (SDF) concept<sup>3,19,20</sup>. We chose this imaging modality for two reasons: firstly, due to the insufficient angular modulation from a single detector and vast imaging space, the imaging problem is highly ill-posed. As a result, the solutions found from the ML-EM algorithm are often undesirable<sup>21</sup>, highlighting the benefits of the GPP image reconstruction and uncertainty quantification methods. Secondly, there has been growing interest in the free-moving imaging modalities due to its significant impact in nuclear safety and security applications<sup>22</sup>; however, the concept is often challenged by the absence of reliable image reconstruction methods. Hence the proposed approach can provide accurate reconstruction of radiation fields for a wide range of applications, thereby addressing the current needs of the research field.

## Results

In this section, we demonstrate the effectiveness of the proposed GPP MAP image reconstruction and uncertainty quantification techniques using simulated free-moving single detector imaging scenarios.

Each free-moving measurement scenario was simulated with a 10-cm diameter spherical detector (i.e., isotropic response) in a 20 m × 20 m 2-D imaging space. The detector intrinsic efficiency was assumed to be 10% and the detector moving paths were simulated on the x-y plane. In all measurement scenarios, the z-coordinates of the pixel centers and detector paths were fixed at 0 m and 0.5 m, respectively. Along the detector paths, recorded counts were binned with the integration time of 0.1 s.

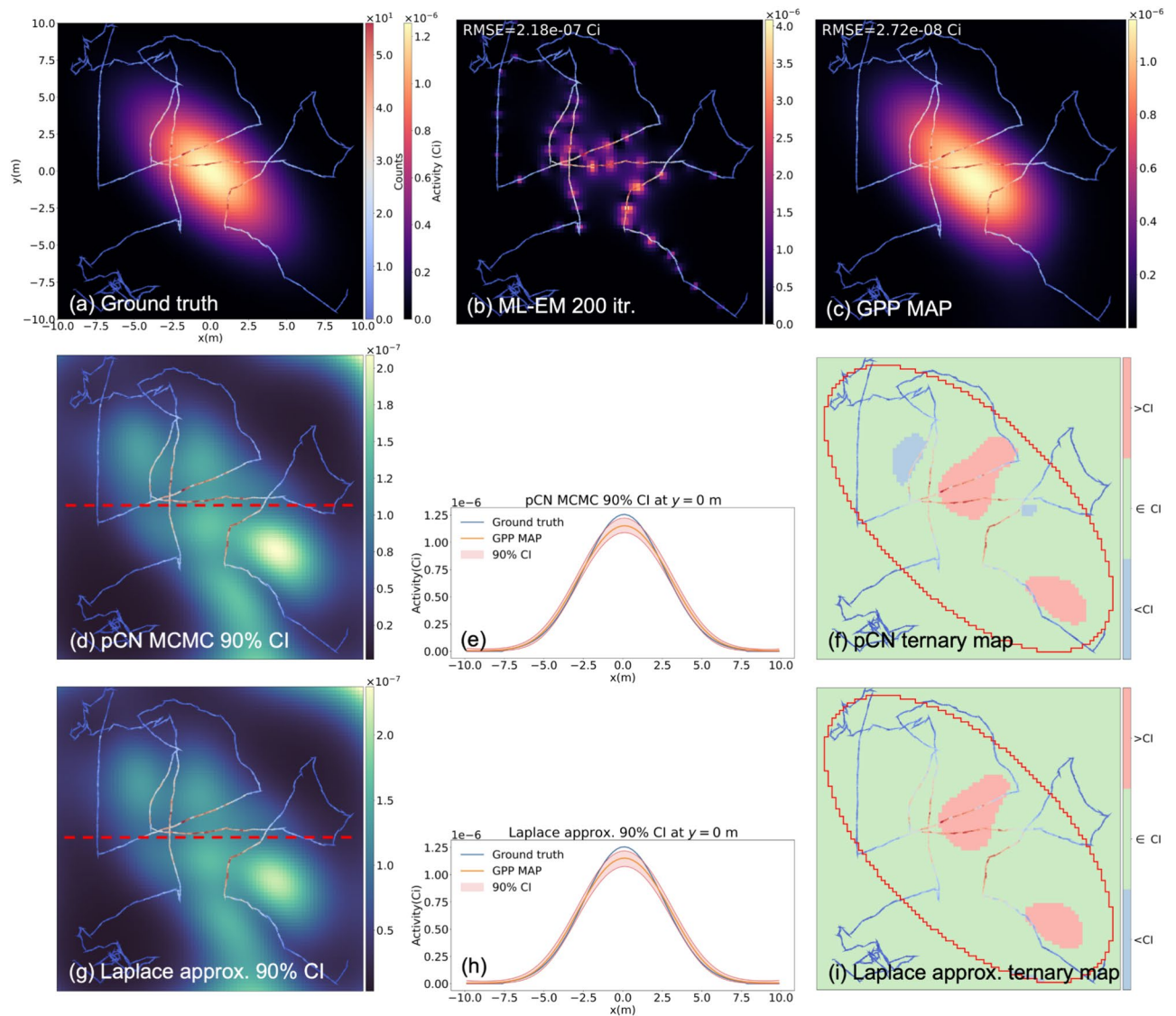
### Anisotropic Gaussian Source

First, a measurement scenario with an anisotropic Gaussian source was simulated to demonstrate the effectiveness of the proposed algorithm. Gaussian source distributions are of particular interest because radioactive material dispersion often exhibits Gaussian-like distributional features<sup>23</sup>.

Fig. 1a illustrates the simulated 1-mCi anisotropic Gaussian distributed source positioned at the center of the imaging space and the 150-m path of the moving detector. Assuming the average human walking speed of 1.5 m/s, the total measurement time amounted to 98 s.

Fig. 1b,c show the image reconstruction results of the ML-EM after 200 iterations and the proposed GPP algorithm. As can be evidently seen, the ML-EM reconstructed image exhibits high frequency features clustered around the path, and we observed such patterns even after 20 iterations. On the other hand, the image reconstructed with the GPP MAP successfully localizes and maps the Gaussian distribution.

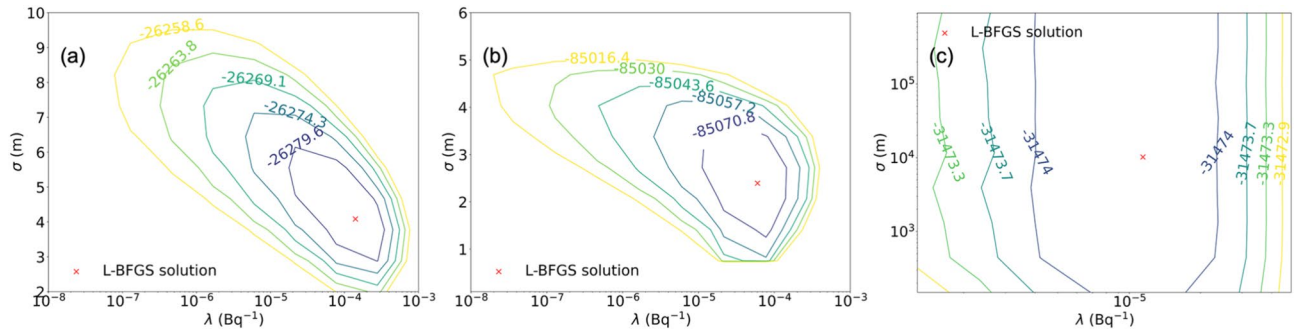
The hyperparameters of the GPP algorithm ( $\sigma = 4.1$  m and  $\lambda = 1.4 \times 10^{-4}$  Bq<sup>-1</sup>) were chosen with Algorithm 2. Fig. 2a shows the contour plot of the negative log-marginal likelihood as a function of  $\sigma$  and  $\lambda$ , together with the optimal hyperparameters selected from Algorithm 2. To investigate the impact of hyperparameter choices on image reconstruction, Fig. 3 shows the GPP MAP reconstructed images with various hyperparameters. As illustrated, the algorithm performs well over a large range of hyperparameter choices.



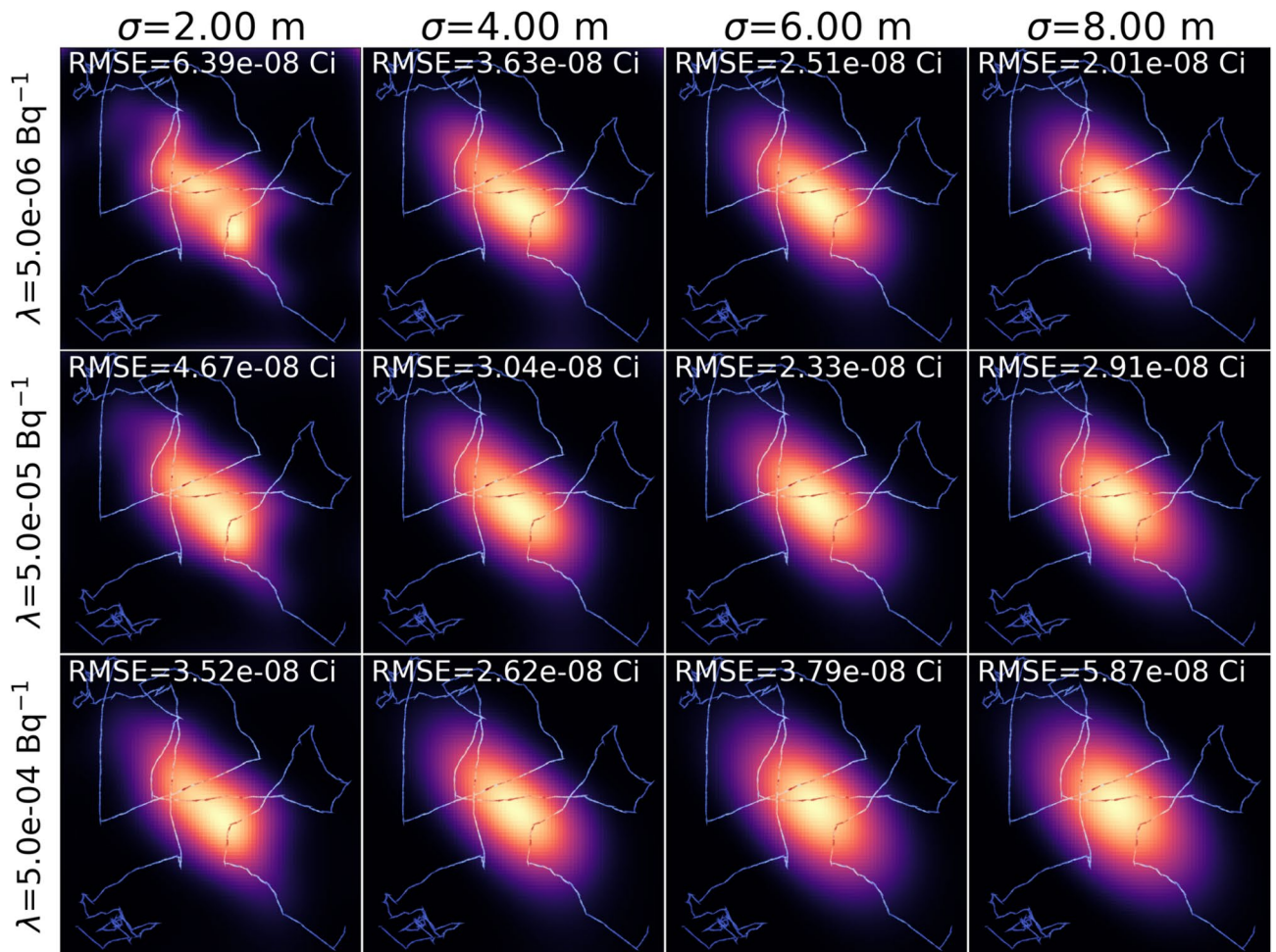
**Fig. 1.** (a) A simulated measurement scenario with a 1-mCi anisotropic Gaussian distributed source. The human-walked path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters  $\sigma=4.1$  m and  $\lambda = 1.4 \times 10^{-4} \text{ Bq}^{-1}$  found from the empirical Bayes method were used for reconstruction. (d, g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e, h) The credible intervals and the ground truth at  $y=0$  m (red dotted lines in (d) and (g)). (f, i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

Fig. 1d,g show the pixel-wise 90% credible interval maps from the pCN and Laplace approximation, respectively. For the pCN MCMC,  $10^6$  samples were collected and the pixel-wise credible intervals were computed. In both cases, there exists about 10% of uncertainty around the source region, due to the ambiguity in the reconstructed images. Interestingly, the uncertainty is lower in those pixels close to the path, reflecting the fact that the source activity can be inferred with more confidence in the pixels with large sensitivity. Conversely, the unexplored region in the upper right corner exhibits high uncertainties, because the measurement is not sensitive to the pixels in the region. Note that with the Laplace approximation, the uncertainties in the upper right corner are slightly overestimated. We suspect that the marginal pixel posteriors in the upper right corner are exponentially distributed due to the lack of measurements, (i.e., prior  $\sim$  posterior). Capturing the skewness of the exponential distribution is challenging with the Laplace approximation because of the Gaussian posterior assumption. However, from Fig. 1e,h, which depict the slices of the credible interval maps at  $y = 0$  m, it can be observed that both uncertainty quantification methods produce very similar results at the center of the imaging space where the source is distributed.

Fig. 1f,i display the ternary maps, where 3 different colors are used depending on the ground truth pixel values relative to the estimated credible intervals. The pixels where the ground truth values are bounded by the



**Fig. 2.** Contour plots of the negative log-marginal likelihood as a function of the hyperparameters  $\sigma$  and  $\lambda$ . The hyperparameters selected from Algorithm 2 are marked with red Xs. (a) The anisotropic Gaussian source measurement scenario. (b) The ring and Gaussian source measurement scenario. (c) The uniform square source measurement scenario with a structural prior. When using the structural prior, 2 different sets of hyperparameters were used for in-square and out-of-square. Therefore, only the in-square hyperparameters were used to create the contour plot, while the out-of-square hyperparameters were fixed at the values found from Algorithm 2.



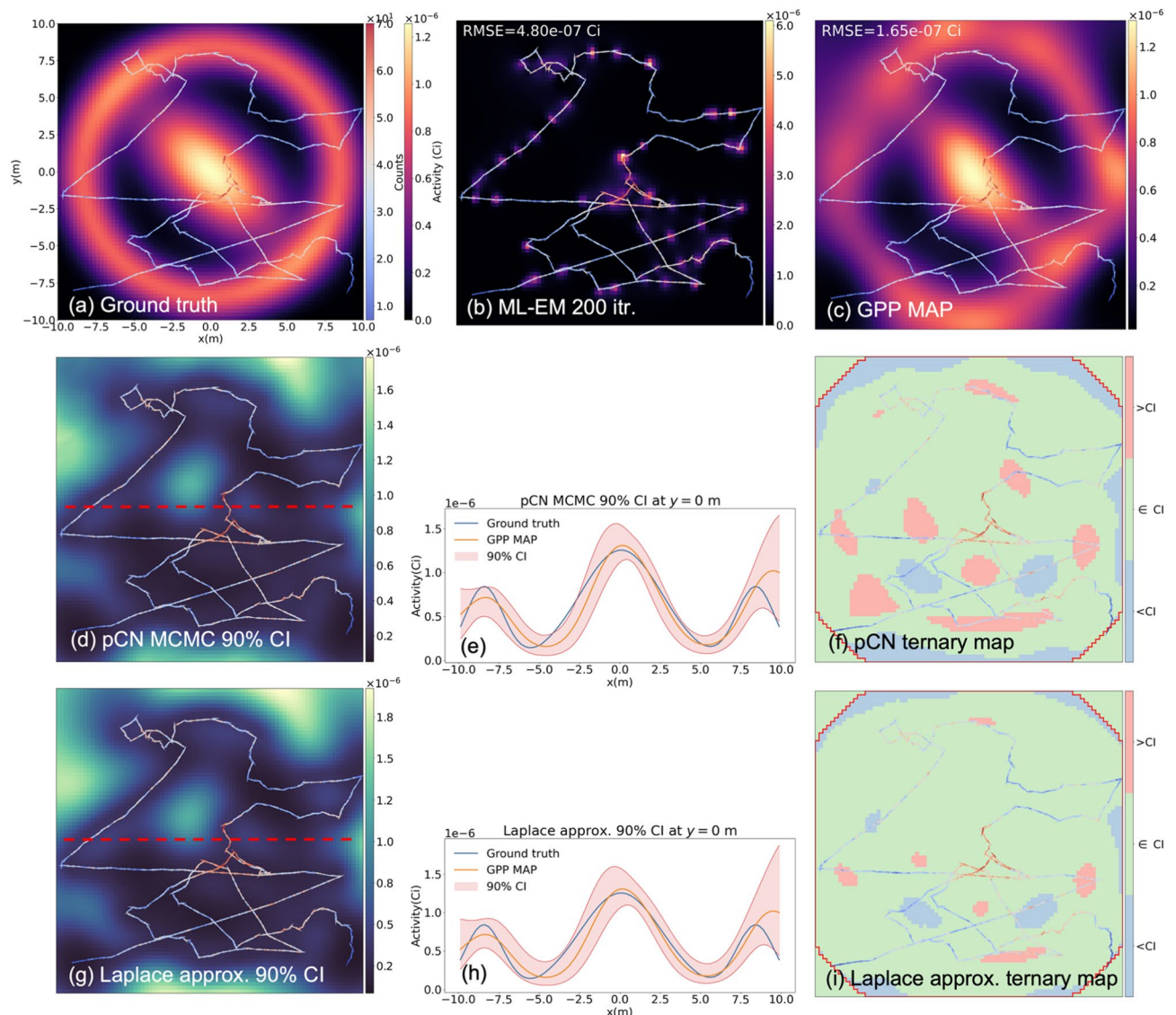
**Fig. 3.** To better understand the impact of the hyperparameters, the anisotropic Gaussian source images reconstructed from the GPP MAP with different choices of the hyperparameters  $\sigma$  and  $\lambda$  are displayed. With larger  $\sigma$  and  $\lambda$ , the prior distribution is spatially more correlated and the pixel values are pushed towards zeros. As a result, the reconstructed image with large hyperparameters (bottom right) tend to spread out.

credible intervals are colored green, while red and blue indicate the ground truth values exceeding and falling below the credible intervals, respectively. As illustrated, in most pixels the ground truth is well bounded by the credible intervals estimated from both the pCN and Laplace approximation.

### Ring and Gaussian source

A more challenging measurement scenario involving a 2 mCi ring-shaped source and a 1 mCi anisotropic Gaussian source was simulated as depicted in Fig. 4a along with the detector moving path. The total measurement duration was approximately 100 s. As can be seen in Fig. 4b,c, while the ML-EM produces source distribution concentrated around the measurement path, the GPP MAP estimate offers significantly more improved reconstruction results. The GPP hyperparameters were set to  $\sigma = 2.4$  m and  $\lambda = 6.0 \times 10^{-5} \text{ Bq}^{-1}$ , which were selected with Algorithm 2. The contour plot of the negative marginal log-likelihood as a function of the hyperparameters can be found in Fig. 2b.

Fig. 4d,g present the results of the uncertainty quantification using the pCN and Laplace approximation, respectively. Similarly to the findings in Sect. 2.1, the pixel uncertainties are predominantly influenced by the

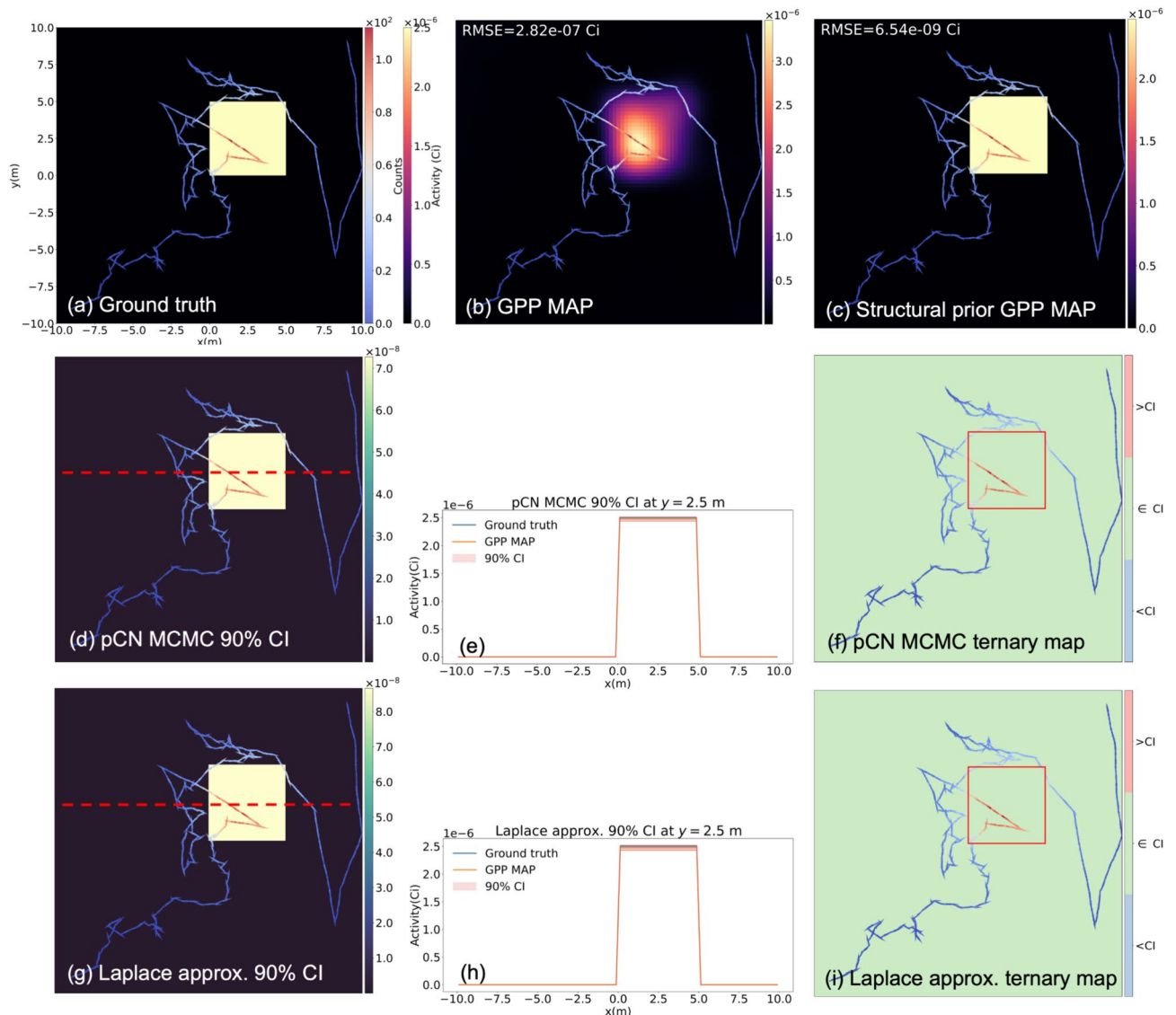


**Fig. 4.** (a) A simulated measurement scenario with a 2-mCi ring-shaped source and a 1-mCi anisotropic Gaussian distributed source. The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters  $\sigma=2.4$  m and  $\lambda = 6.0 \times 10^{-5} \text{ Bq}^{-1}$  found from the empirical Bayes method were used for reconstruction. (d, g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e, h) The credible intervals and the ground truth at  $y=0$  m (red dotted lines in (d) and (g)). (f, i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

sensitivities of the pixels; uncertainties are lower around the detector path and elevated in those unvisited regions. Also, the Laplace approximation slightly overestimates the uncertainties, yet overall it remains largely comparable to the pCN as can be seen in Fig. 4e,h. In addition, the ternary maps Fig. 4f,i demonstrate that the ground truths are mostly bounded by the credible intervals estimated from the both methods.

### Uniform square source with a structural prior

We also demonstrate that incorporation of a structural prior using the GPP algorithm could significantly improve the image quality. Fig 5a show another measurement scenario with a 1 mCi uniform square source locate at  $(x, y) = (2.5 \text{ m}, 2.5 \text{ m})$ . The total measurement duration was approximately 70 s. The reconstruction result using the GPP MAP is shown in Fig. 5b. Although the GPP MAP reconstruction successfully localized the source, the high frequency edges could not be recovered for two reasons: first, due to the lack of modulations in the single detector measurements, the forward projected measurements are blurred and noisy, leading to some ambiguity



**Figure 5.** (a) A simulated measurement scenario with a 1-mCi uniform located at  $(x, y) = (2.5 \text{ m}, 2.5 \text{ m})$ . The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the GPP MAP without incorporating the structural information. The hyperparameters found from Algorithm 2,  $\sigma = 2.03 \text{ m}$  and  $\lambda = 8.10 \text{ Bq}^{-1}$  were used. (c) The reconstructed from the proposed GPP MAP with the structural prior.  $\sigma = 1.0 \times 10^4 \text{ m}$  and  $\lambda = 1.1 \times 10^{-5} \text{ Bq}^{-1}$  for the in-square hyperparameters, and  $\sigma = 9.8 \times 10^3 \text{ m}$  and  $\lambda = 7.7 \times 10^4 \text{ Bq}^{-1}$  for the out-of-square hyperparameters. (d, g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e, h) The credible intervals and the ground truth at  $y = 2.5 \text{ m}$  (red dotted lines in (d) and (g)). (f, i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

in the reconstructed images. Second, the GP prior with the SE kernel is not apt to model the high frequency edges of the ground truth distribution.

Here we show that the limitations can be overcome by incorporating a priori structural information. The prior covariances between the pixels within and outside the source square can be zeroed to enforce independence between the two pixel clusters. The reconstruction result with the structural prior is presented in Fig. 5c. As can be observed, the reconstruction successfully recovers the uniform ground truth distribution with the edges. Note that two different sets of hyperparameters were used for the two pixel clusters, namely the in-square hyperparameters and the out-of-square hyperparameters. Using Algorithm 2, the hyperparameters were set at  $\sigma = 1.0 \times 10^4$  m and  $\lambda = 1.1 \times 10^{-5}$  Bq<sup>-1</sup> for the in-square hyperparameters, and  $\sigma = 9.8 \times 10^3$  m and  $\lambda = 7.7 \times 10^4$  Bq<sup>-1</sup> for the out-of-square hyperparameters. The very large  $\sigma$  hyperparameters for both pixel clusters indicate that a piecewise constant function with the discontinuities at the pixel cluster boundaries well-explain the measurement data. Also the large difference in the  $\lambda$  parameters between the pixel clusters reflects the fact that the source is only present in the in-square pixel clusters. Fig. 2c shows a contour plot illustrating the negative log-marginal likelihood as a function of in-square hyperparameters.

The 90% credible interval maps found from the pCN and Laplace approximation are displayed in Fig. 5d,g, respectively. Note that due to the strong correlations between the pixel clusters, the marginal credible interval maps are also step-function like. Also, similarly to the findings from the results presented in Sects. 2.1 and 2.2, a slight overestimation of uncertainties was observed with the Laplace approximation. However, both methods produce comparable results as can be observed from Fig. 5e,h, where the slices of the uncertainty maps at  $y = 2.5$  m are shown. Also, the ternary maps in Fig. 5f,i indicate that the ground truth distribution is well-bounded by the estimated uncertainties.

## Discussion

In this section, we discuss the implications of the proposed methods and potential future research directions.

### Implications of the proposed approach

First, the proposed GP prior can be made highly flexible with a choice of a covariance function and a link function. Therefore, the GP prior can effectively model desired properties of the reconstructed images, and is directly applicable to a wide range of highly ill-posed image reconstruction tasks. Such tasks arise from various radiation imaging modalities including Compton cameras<sup>24</sup>, coded aperture imaging<sup>25</sup>, limited angle computed tomography (CT)<sup>26</sup>, low-dose positron emission tomography (PET)<sup>27</sup>, among others.

Secondly, the uncertainty quantification capability of the proposed framework offers valuable insights into the reconstructed images, and it becomes particularly important when critical decisions are made based on the reconstructed images. Such problems commonly arise in many healthcare<sup>28</sup> and nuclear security applications<sup>29</sup>. Moreover, posterior distributions estimated from the pCN and the Laplace approximation can be employed for further inferences on quantities derived from the reconstructed images. For instance, suppose a quantity of interest  $u$  can be derived from an image  $\mathbf{x}$  as  $u = g(\mathbf{x})$ . From a set of posterior samples obtained from the pCN,  $\{\xi_n\}_{n=1}^J$ , the samples of the derived quantity  $\{u_n\}_{n=1}^J$  can be obtained as  $u_n = g(f^{-1}(\xi_n))$ . In the context of the Laplace approximation, samples of  $\xi$  can be obtained from the approximating Gaussian distribution in (29). Some examples of such derived quantities are total dose registered in radiopharmaceutical therapy<sup>30</sup>, in-beam ion ranges in ion cancer therapy, and in-vivo proton range verification in ion cancer therapy<sup>31</sup>.

### Future works

#### *Incorporation of background contribution*

The results presented in section demonstrate that the proposed GPP algorithm can significantly improve the reconstruction quality where the ML-EM may be inadequate. However, the simulated measurement scenarios assume no present background contributions, which may be unrealistic in some radiation imaging applications. When the background contribution is known, it can be straightforwardly accounted for by changing the forward projection  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$  to  $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{b}$ , where the elements of the vector  $\mathbf{b}$  are the known background contribution to the measurements. When the background rate is unknown, it has to be jointly estimated with the image  $\mathbf{x}$ , and it can be done by augmenting the system matrix  $\mathbf{A}$  and the image vector  $\mathbf{x}$  as  $\tilde{\mathbf{A}} = [\mathbf{A} | \mathbf{a}_b]$  and  $\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix}$ , where the vector  $\mathbf{a}_b$  is assumed to be known (e.g., in the case of a constant background rate, a vector of integration time  $\mathbf{a}_b = \Delta t \mathbf{1}$ ). Subsequently, the prior covariance matrix  $\Sigma$  can be augmented as

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Sigma_b \end{bmatrix}, \quad (1)$$

where the  $\Sigma_b$  accounts for the variance of the background rate.

The use of the Gaussian-to-exponential link function may be inadequate as the background rate can be typically well specified within an order of magnitude. In such case, a different link function, such as the truncated Gaussian-to-Gaussian link function<sup>32</sup>, can be used to effectively model the background prior distribution.

#### *List-mode data and Gaussian likelihood*

Although the work presented here assumes bin-mode data with the Poisson likelihood model, it should be emphasized that the proposed framework can be trivially modified to accommodate list-mode data<sup>33</sup> or Gaussian likelihood model. Essentially, the only differences in the formulation will be the likelihood function formulation in (4), and subsequently its gradient, Hessian expressions.



For example, when the measurements are given in terms of dose rates, Gaussian likelihood model can more aptly describe the measurement noise. In such cases, the likelihood function  $l(\mathbf{x}; \mathbf{y})$  in (5) can be modified as follows:

$$l(\mathbf{x}; \mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{A}\mathbf{x})^\top \mathbf{W}(\mathbf{y} - \mathbf{A}\mathbf{x}), \quad (2)$$

where the matrix  $\mathbf{W} \in \mathbb{R}^{M \times M}$  is the inverse of the measurement covariance matrix (i.e. a precision matrix). When independent and identical noise distribution is assumed,  $\mathbf{W}$  is an identity matrix. With the Gaussian likelihood model, the GPP image reconstruction and uncertainty quantification presented in this paper can be similarly applied.

By accommodating various data types, the proposed framework is applicable to a wide range of radiation imaging applications.

#### Scalable implementation

In Sect. 4.1, we illustrated how the Kronecker product structure in the covariance matrix  $\Sigma$  can be exploited to avoid the computational challenges involving the Cholesky factor  $\mathbf{L}$ . The approach is particularly attractive as it provides exact representation of the covariance matrix; however, the method is not readily applicable when the imaging space is not in a dense Cartesian grid. For example, Compton cameras or active coded apertures imagers often employ a  $4\pi$  spherical imaging space<sup>34,35</sup>, and free-moving imaging scenarios often assume sparse voxel grid imaging spaces<sup>36</sup>. Thus, a different strategy is needed when more diverse imaging spaces are considered.

Several recent advances from the GP regression research community could potentially be employed to tackle the challenges. For example, in<sup>37</sup>, Willson et al. introduced the structured kernel interpolation (SKI) scheme, where the covariance matrix is interpolated from a Kronecker product represented sparse inducing points. The resulting approximate covariance matrix can be handled efficiently thanks to the sparsity of the interpolation coefficient matrix and the Kronecker product structure. Such formulation could be applied to imaging spaces on a non-structured grid.

Another interesting approach is Gaussian Markov random field (GMRF), where a sparse precision matrix  $\Sigma^{-1}$  is directly constructed and utilized. Such precision matrices admit rapid realization of GPs, removing the computational bottlenecks involving  $\mathbf{L}\xi_w$ . Moreover, Lindgren et al. showed that a GMRF can directly approximate GPs with a Matérn covariance kernel, and it can be efficiently applied to meshed imaging spaces<sup>38</sup>.

#### Covariance Kernels

The SE kernel in (11) is a popular choice in many GP regression tasks, and in this work we demonstrated that it can be used to effectively model the prior covariance. In many radiation imaging modalities, the point spread functions (PSF) have smoothly falling edges and therefore, the use of the SE kernel could be well-justified. However, when it is required to recover high frequency features (e.g., edges or point sources), more expressive kernels may be desired. Rational quadratic kernel or Matérn covariance kernels could be readily used to provide more flexibility for better high frequency feature resolving capability. Moreover, recent works from the GP regression community suggest the use of spectral mixture (SM) kernel<sup>39</sup> or deep kernel learning framework<sup>40</sup>, where a deep neural network is used to construct a highly expressive covariance kernel. It was demonstrated in<sup>40</sup> that such kernels can successfully recover piecewise constant functions with sharp falling edges, which is attractive for imaging applications.

## Methods

### Image reconstruction with GPP

#### GPP MAP image reconstruction

The goal of radiation imaging is to recover an  $N$ -dimensional image vector  $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$  from an  $M$ -dimensional non-negative measurement vector  $\mathbf{y} \in \mathbb{Z}_{\geq 0}^M$ . The two vectors are related by a system matrix  $\mathbf{A} \in \mathbb{R}_{\geq 0}^{M \times N}$  such that

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}), \quad \text{where } \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}. \quad (3)$$

Hence the system matrix contains the probabilities of radiation emitted from a voxel being detected in a measurement bin. With the Poisson measurement model, the likelihood function of an image vector  $\mathbf{x}$  given the measurement vector  $\mathbf{y}$  can be formulated as

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^M \frac{e^{-\bar{y}_i} \bar{y}_i^{y_i}}{y_i!}. \quad (4)$$

The well-established maximum likelihood (ML) image reconstruction approach seeks a solution  $\hat{\mathbf{x}}_{\text{ML}}$ , which minimizes the negative log-likelihood function  $l(\mathbf{x}; \mathbf{y}) = -\log p(\mathbf{y}|\mathbf{x})$  up to a constant as follows:

$$\hat{\mathbf{x}}_{\text{ML}} = \underset{\mathbf{x}, \mathbf{x} \geq 0}{\text{argmin}} l(\mathbf{x}; \mathbf{y}) \quad (5)$$

$$\stackrel{c}{=} \underset{\mathbf{x}, \mathbf{x} \geq 0}{\text{argmin}} \sum_{i=1}^M \bar{y}_i - y_i \log \bar{y}_i \quad (6)$$

$$= \operatorname{argmin}_{\mathbf{x}, \mathbf{x} \geq 0} \sum_{i=1}^M [\mathbf{Ax}]_i - y_i \log [\mathbf{Ax}]_i, \quad (7)$$

where  $\mathbf{x} \geq 0$  imposes the non-negativity constraints and  $\stackrel{c}{=}$  denotes equivalence up to a constant. It is well-known that the above optimization problem can be iteratively solved using the expectation maximization (EM) algorithm, leading to the popular maximum likelihood-expectation maximization (ML-EM) algorithm<sup>41</sup>.

Although ML-EM has become a standard approach for image reconstruction tasks, for many radiation imaging problems the system matrix  $\mathbf{A}$  is often not well-conditioned. Hence, the optimization problem (5) is ill-posed and the ML-EM algorithm is susceptible to overfitting.

A popular approach to combat the ill-posedness is to incorporate a prior distribution  $p(\mathbf{x})$  to make use of Bayesian formalism. Using Bayes' rule<sup>42</sup>, the posterior distribution  $p(\mathbf{x}|\mathbf{y})$  is given by

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (8)$$

In Bayesian image reconstruction, often the image vector is recovered by locating the mode of the posterior distribution, leading to the Maximum a Posteriori (MAP) estimation as follows:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}, \mathbf{x} \geq 0} l(\mathbf{x}; \mathbf{y}) + l_p(\mathbf{x}), \quad (9)$$

where  $l_p(\mathbf{x}) = -\log p(\mathbf{x})$  is the negative log prior distribution. An appropriate choice of a prior distribution is crucial for reliable image reconstruction, especially when the system matrix  $\mathbf{A}$  is ill-conditioned.

In this work, for prior model construction, we assume that the image vector  $\mathbf{x}$  is governed by a latent variable  $\xi$ , which follows a zero-mean Gaussian process (GP),

$$\xi \sim \mathcal{G} \mathcal{P}(\mathbf{0}, k(\mathbf{r}, \mathbf{r}')), \quad (10)$$

where  $k(\mathbf{r}, \mathbf{r}')$  is a covariance kernel, and  $\mathbf{r}, \mathbf{r}' \in \mathbb{R}^3$  represents vectors containing the Cartesian coordinates of two points in the imaging space.

The choice of a covariance kernel  $k(\mathbf{r}, \mathbf{r}')$  is crucial as it dictates the underlying spatial correlation of the image. In this work, we use the unit-variance squared exponential (SE) kernel to encode spatial correlation between pixels as follows:

$$k(\mathbf{r}, \mathbf{r}') = \exp\left(-\frac{\|\mathbf{r} - \mathbf{r}'\|^2}{2\sigma^2}\right), \quad (11)$$

where  $\sigma$  is the characteristic correlation length scale parameter of the prior distribution with the dimension of distance. It should be noted that various covariance kernels have been explored in the context of GPs, and the optimal covariance kernel may vary depending on applications<sup>43</sup>.

For practical purposes, the imaging space is discretized with pixels and the GP in (10) is evaluated at the pixel centers. Then, the GP reduces to a zero-mean multivariate Gaussian distribution,

$$\xi \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (12)$$

$$\text{where } \Sigma_{ij} = k(\mathbf{r}_i, \mathbf{r}_j). \quad (13)$$

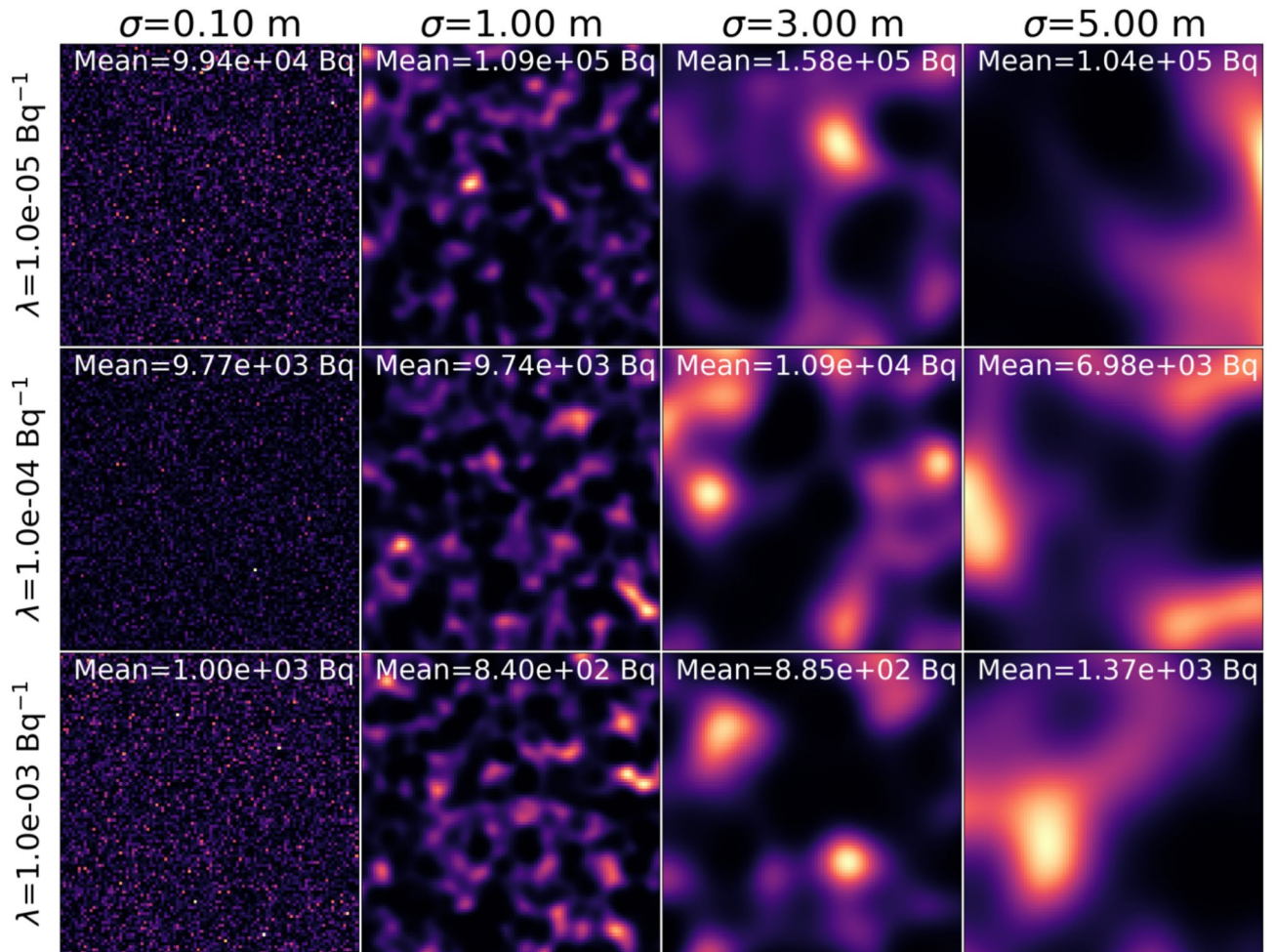
Hence,

$$p(\xi) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} \xi^\top \Sigma^{-1} \xi\right). \quad (14)$$

Once the prior  $p(\xi)$  is defined, a link function  $f: \mathbb{R}_{\geq 0}^N \rightarrow \mathbb{R}^N$  is used to relate  $\xi$  and  $\mathbf{x}$ , keeping the positivity constraints imposed on  $\mathbf{x}$ . There are a variety of choices for the link function; for example, a simple logarithmic function,  $f(x_i) = \log x_i = \xi_i$  could be used as a link function satisfying the non-negativity constraint on  $\mathbf{x}$ . However, for simpler statistical interpretation, we use the exponential-to-Gaussian link function<sup>32</sup>, and its inverse can be used to recover  $\mathbf{x}$  as follows:

$$\mathbf{x} = \mathbf{f}^{-1}(\xi) = -\frac{1}{\lambda} \log\left(\frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\xi \oslash (\sqrt{2} \operatorname{diag}(\Sigma))\right)\right), \quad (15)$$

where  $\lambda$  is the scale parameter for the exponential distribution and erf is the error function.  $\oslash$  denotes the element-wise division operator. The link function converts a zero-mean Gaussian distributed random variable  $\xi_i$  with the variance  $\Sigma_{ii}$  into a random variable following the exponential distribution with the scale parameter  $\lambda$ . As the reciprocal of  $\lambda$  defines the mean value of the exponential distributed random variable, it can be used to incorporate *a priori* information about the average pixel value, allowing for more statistically interpretable hyperparameter selection. Furthermore, due to the long-tailed nature of the exponential distribution, the link function can effectively model radiation source strength that varies over several orders of magnitudes. Samples from the prior distribution  $p(\mathbf{x})$  with different hyperparameters  $\sigma$  and  $\lambda$  are shown in Fig. 6.



**Fig. 6.** Random samples from the prior distribution  $p(\mathbf{x})$  on a  $20 \times 20$  m 2-D imaging space with different sets of hyperparameters,  $\sigma$  and  $\lambda$ . Each sample was realized by first drawing a sample from  $\xi_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  and then transforming it by  $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{L}\xi_w)$ . The properties and characteristics of the prior distribution can be better understood by examining random samples from the prior. Note that the  $\sigma$  parameter defines the characteristic correlation length, while the reciprocal of  $\lambda$  parameter determines the mean pixel values.

It should be emphasized that the appropriate choice of a link function may vary depending on the application, and the methods presented in this work are applicable with various link functions.

From (14), the negative log-prior for  $\xi$  can be written as follows:

$$\begin{aligned}
 l_p(\xi) &= -\log p(\xi) \\
 &= \frac{1}{2} \log \sqrt{(2\pi)^n |\Sigma|} + \frac{1}{2} \xi^T \Sigma^{-1} \xi
 \end{aligned}
 \tag{16}$$

Combining (15) and (16), the MAP solution in terms of  $\xi$  can be obtained by minimizing the unnormalized negative log posterior  $\Psi(\xi) = l(\xi; \mathbf{y}) + l_p(\xi)$  as follows:

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \Psi(\xi)
 \tag{17}$$

$$= \underset{\xi}{\operatorname{argmin}} \sum_i^N [A\mathbf{f}^{-1}(\xi) - \mathbf{y} \odot \log(A\mathbf{f}^{-1}(\xi))]_i + \frac{1}{2} \xi^T \Sigma^{-1} \xi.
 \tag{18}$$

The optimization problem (18) can be solved in a more efficient and numerically stable manner by introducing an uncorrelated multivariate unit Gaussian variable  $\xi_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ . Hence, we use a change of variable  $\mathbf{L}\xi_w = \xi$ , where  $\mathbf{L}$  is the Cholesky factor of  $\Sigma$ . The final optimization problem in terms of  $\xi_w$  is

$$\hat{\xi}_w = \underset{\xi_w}{\operatorname{argmin}} \Psi(\mathbf{L}\xi_w) \quad (19)$$

$$= \underset{\xi_w}{\operatorname{argmin}} \sum_i^N [\mathbf{A}f^{-1}(\mathbf{L}\xi_w) - \mathbf{y} \odot \log(\mathbf{A}f^{-1}(\mathbf{L}\xi_w))]_i + \frac{1}{2} \xi_w^T \xi_w, \quad (20)$$

where  $\odot$  is the element-wise product operation. Note that (19) is an unconstrained optimization problem, and it can be solved with a general purpose global optimization algorithms. We find that quasi-Newton algorithms, such as the Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS)<sup>44</sup> achieves swift convergence. The gradient of (19), which is required for many optimization algorithms, can be analytically found as follows:

$$\nabla_{\xi_w} \Psi(\mathbf{L}\xi_w) = \mathbf{L}^T \left( \left( \mathbf{A}^T - \mathbf{A}^T \mathcal{K}(\mathbf{y} \odot (\mathbf{A}\mathbf{x})) \right) \odot \left( \mathcal{K} \odot \left( \sqrt{2\pi} \lambda \operatorname{diag}(\Sigma) \right) \odot \exp(\lambda \mathbf{x} - \xi^{\circ 2} \odot (2\operatorname{diag}(\Sigma))) \right) \right) + \xi_w, \quad (21)$$

where  $\circ$  is the element-wise power operations. Also, the  $\operatorname{diag}(\cdot)$  operator extracts the main diagonal elements of a matrix into a vector. The derivation of the gradient expression (21) can be found in the "Supplementary information". Once the optimization problem is solved and  $\hat{\xi}_w$  is obtained, the image  $\hat{\mathbf{x}}$  can be obtained by  $\hat{\mathbf{x}} = f^{-1}(\mathbf{L}\hat{\xi}_w)$ . The pseudocode for the GPP MAP estimation is shown in Algorithm 1.

---

**Require:**  $\mathbf{A}$ ,  $\mathbf{y}$ ,  $\sigma$ ,  $\lambda$ ,  $\xi_w^{\text{init}}$ ,  $\varepsilon$  (Convergence criterion)

$\Sigma \leftarrow (13)$  ▷ Prior covariance matrix construction

$\mathbf{L} \leftarrow \text{Cholesky}(\Sigma)$  ▷ Cholesky decomposition

$k \leftarrow 0$

$\xi_w^k \leftarrow \xi_w^{\text{init}}$  ▷ Initialization

$\Psi^k \leftarrow \Psi(\mathbf{L}\xi_w^k)$  ▷ Initialization

Convergence  $\leftarrow$  False

**while** Convergence is False **do**

$\xi_w^{k+1} \leftarrow \text{L-BFGS}(\xi_w^k, (19), (21))$  ▷ L-BFGS iteration

$\Psi^{k+1} \leftarrow \Psi(\mathbf{L}\xi_w^{k+1})$

**if**  $|\Psi^k - \Psi^{k+1}| < \varepsilon$  **then**

Convergence  $\leftarrow$  True

**else**

$k \leftarrow k + 1$

**end if**

**end while**

$\hat{\xi}_w \leftarrow \xi_w^{k+1}$

$\hat{\xi} \leftarrow \mathbf{L}\hat{\xi}_w$

$\hat{\mathbf{x}} \leftarrow (15)$  ▷ Reconstructed image

$\hat{\Psi} \leftarrow \Psi^{k+1}$  ▷ Negative log-posterior at the mode

---

### Algorithm 1. GPP MAP estimation

#### *Kronecker product representation of the covariance matrix*

With a large number of voxels  $N$ , setting up and solving (19) is a burdensome task due mainly to the computational costs involved in the Cholesky factor  $\mathbf{L} \in \mathbb{R}^{N \times N}$ . For example, storing and computing the Cholesky factor requires  $\mathcal{O}(N^2)$  memory and  $\mathcal{O}(N^3)$  time, which can be prohibitively expensive for a large  $N$ .

Computation of the Cholesky decomposition is also the main computational bottleneck in the GP regression problems. To overcome the challenge, in<sup>45</sup> Saatci et al. showed that when the GP regression input variables lie on a multi-dimensional Cartesian grid, the Kronecker structure in the covariance matrix  $\Sigma$  can be exploited to significantly lessen the computational costs involving the Cholesky factor  $\mathbf{L}$ . As many imaging problems employ Cartesian grid imaging spaces, here the strategy can be readily extended to the GPP image reconstruction.

When a covariance kernel function is the product of kernels across each Cartesian input dimension, such that  $k(\mathbf{r}_i, \mathbf{r}_j) = \prod_{d=1}^D k(r_i^d, r_j^d)$ , the covariance matrix  $\Sigma$  constructed for a multi-dimensional lattice imaging space can be expressed as a Kronecker product of much smaller covariance matrices for each input dimension  $\Sigma_d$  (i.e.,  $x$ -,  $y$ - and  $z$ -dimensions), such that  $\Sigma = \bigotimes_{d=1}^D \Sigma_d$ . Furthermore, the Cholesky factor of a such covariance matrix is also a Kronecker product of Cholesky factor in each dimension,  $\mathbf{L} = \bigotimes_{d=1}^D \mathbf{L}_d$ . Hence, assuming the same number of voxels in each dimension, the memory and time complexity of the Cholesky factor computation are eased with  $\mathcal{O}(DN^{2/D})$  and  $\mathcal{O}(DN^{3/D})$ .

Even with the obtained Kronecker product of the Cholesky factor, a naive matrix-vector multiplication  $\mathbf{L}\xi_w$  in (19) and (21) limits the scalability of the algorithm. Surprisingly, multiplication of a Kronecker product matrix on a vector can be cast as a tensor outer product and performed in  $\mathcal{O}(DN^{\frac{1+D}{D}})$  time<sup>45,46</sup>.

Equipped with the Kronecker product covariance matrix and its Cholesky factor, the overall time complexity of computing the cost function (19) and its gradient (21) is determined by the forward projection  $\mathbf{A}f^{-1}(\mathbf{L}\xi_w)$ , which is computed in  $\mathcal{O}(MN)$ . Note that the ML-EM has the same complexity of  $\mathcal{O}(MN)$  per iteration.

### Incorporation of structural prior

With the growing interest in contextual data fusion and multi-modality radiation imaging<sup>47–49</sup>, we discuss here the incorporation of a priori structural information using the GPP. Such a priori structural information can be obtained from other radiation imaging modalities (e.g., MRI, CT), commercial sensors (e.g., LiDAR, visual camera), or object detection algorithms.

We consider an imaging space  $\mathcal{X}$  with  $N$  voxels  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$  and further,  $L$  voxel clusters  $C_l \subseteq \mathcal{X}$  for  $l = 1, 2, \dots, L$ , each belonging to different objects or structures in the imaging space  $\mathcal{X}$ . This structural information can be incorporated by modifying the covariance matrix of the Gaussian prior  $\Sigma$  (11). The modified covariance matrix is given by:

$$\Sigma_{ij} = \begin{cases} k(\mathbf{r}_i, \mathbf{r}_j), & \text{if } x_i, x_j \in C_l \quad \forall l = 1, 2, \dots, L \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

In other words, spatial correlation is defined only within the same object  $C_l$ . Note that different hyperparameters  $\sigma$  and  $\lambda$  can be applied to each voxel cluster  $C_l$ . The benefit of incorporating a structural prior is that it introduces correlations within salient features in the imaging space, leading to improved image quality and interpretability of the reconstruction results.

Note that, with the use of structural prior, the covariance matrix cannot be represented as a Kronecker product; however, by assigning zeroes for the covariances between different voxel clusters, the covariance matrix becomes a sparse block diagonal matrix, whose Cholesky decomposition can be computed and stored efficiently.

### Bayesian uncertainty quantification

The use of a Gaussian process prior in this work furnishes the basis for further Bayesian inference such as Bayesian uncertainty quantification. In Bayesian statistics, the uncertainty in the posterior distribution  $p(\mathbf{x}|\mathbf{y})$  is often expressed in terms of a Bayesian credible interval. For a marginalized posterior variable  $x_j$ , the  $(1 - \alpha)100\%$  Bayesian credible interval is defined as the range  $[x_{j-}, x_{j+}]$  within which  $x_j$  falls with a probability of  $1 - \alpha$ . In other words,  $p(x_j \in [x_{j-}, x_{j+}]|\mathbf{y}) = 1 - \alpha$ . The choice of such interval is not unique; therefore, we use an equal-tailed interval to define  $x_{j-}$  and  $x_{j+}$  as follows:

$$\int_{-\infty}^{x_{j-}} p(x_j|\mathbf{y}) dx_j = \frac{\alpha}{2} \quad (23)$$

$$\int_{x_{j+}}^{\infty} p(x_j|\mathbf{y}) dx_j = \frac{\alpha}{2}. \quad (24)$$

However, computation of the credible intervals is a daunting task, as it requires the integration of a very high dimensional distribution,  $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) d\mathbf{x}$  (i.e., the marginal probability). Therefore, in this work, we investigate two different posterior approximation methods for credible interval estimation.

### Preconditioned Crank–Nicolson (pCN) algorithm

In Bayesian statistics, Markov chain Monte Carlo (MCMC) algorithms have been widely used to generate a sequence of samples from an analytically intractable posterior distribution. However, for high dimensional posterior distributions (e.g., images), the MCMC sample convergence rate becomes very slow<sup>50</sup>.

In our approach, the use of GP prior in (19) lends itself to the use of the pCN algorithm<sup>15</sup>, a dimension-robust MCMC algorithm specifically designed for posterior distributions with a Gaussian prior. The pCN convergence rate is independent of discretization of a parameter space, which is an attractive feature for high-dimensional inverse problems such as image reconstruction.

With the pCN algorithm, a chain of  $J$  samples  $\{\xi_n\}_{n=1}^J$  is obtained by following the proposal and acceptance/rejection steps. First, the pCN proposal in terms of  $\xi$  is given by,

$$\xi'_{n+1} = \sqrt{1 - \beta^2} \xi_n + \beta \Xi_{n+1}, \quad (25)$$

where  $\Xi_{n+1}$  is a sample from  $\mathcal{N}(\mathbf{0}, \Sigma)$  and the hyperparameter  $0 < \beta < 1$  is a step size which is chosen manually. With the obtained  $\xi'_{n+1}$ , the acceptance probability is computed as

$$\alpha(\xi_n, \xi'_{n+1}) = \min(1, \exp(l(\xi_n; \mathbf{y}) - l(\xi'_{n+1}; \mathbf{y}))). \quad (26)$$

Subsequently, the acceptance probability  $\alpha(\xi_n, \xi'_{n+1})$  is compared to a random number  $Z_{n+1}$  drawn from a uniform distribution  $Z_{n+1} \sim \text{Unif}([0, 1])$  and sets

$$\xi_{n+1} = \begin{cases} \xi'_{n+1}, & \text{if } Z_{n+1} < \alpha(\xi_n, \xi'_{n+1}) \\ \xi_n, & \text{otherwise.} \end{cases} \quad (27)$$

In our experiments, the MCMC sample chains were started from the reconstructed image (i.e., the mode of the posterior). Once a sufficient number of samples are obtained, the pixelwise  $(1 - \alpha)100\%$  confidence intervals can be found as follows. First, compute the pixelwise  $(\alpha/2)100\%$  and  $(1 - \alpha/2)100\%$  quantiles  $\xi_-$  and  $\xi_+$  from the obtained samples  $\{\xi_n\}_{n=1}^J$ . Then, the quantiles in  $\mathbf{x}$  are obtain as  $\mathbf{x}_- = \mathbf{f}^{-1}(\xi_-)$  and  $\mathbf{x}_+ = \mathbf{f}^{-1}(\xi_+)$ . As the inverse link function  $\mathbf{f}^{-1}$  is monotonically increasing, it ensures that the quantiles of  $\mathbf{x}$  can be directly found

by transforming the quantiles of  $\xi$ . Lastly, the pixelwise  $(1 - \alpha)100\%$  credible interval map  $\mathbf{x}_{CI}$  can be found as  $\mathbf{x}_{CI} = \mathbf{x}_+ - \mathbf{x}_-$ .

Collecting  $S$  samples, the time complexity of the pCN method is  $\mathcal{O}(MNS)$ . Although each sample can be taken efficiently, ensuring convergence of the MCMC chain requires a large  $S$ , on the order of  $10^5$  to  $10^6$  samples. In the next section, we explore an alternative approach to posterior approximation, which typically runs much faster than the pCN method.

*Laplace approximation*

Another posterior estimation method, the Laplace approximation, uses a multivariate Gaussian distribution with the mean located at the MAP estimate  $\hat{\xi}$  found from (19) to directly approximate a posterior distribution.

Using second-order Taylor expansion of the unnormalized log posterior  $-\Psi(\xi)$  at the mode  $\hat{\xi}$ , the posterior distribution can be approximated with  $q(\xi|\mathbf{y})$  as follows:

$$p(\xi|\mathbf{y}) \approx q(\xi|\mathbf{y}), \quad \text{where} \tag{28}$$

$$q(\xi|\mathbf{y}) \sim \mathcal{N}\left(\hat{\xi}, \left(\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}\right)^{-1} = \left(\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} + \Sigma^{-1}\right)^{-1}\right), \tag{29}$$

A more detailed derivation as well as the expression for  $\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}$  is presented in the "Supplementary information". Once the approximate posterior distribution  $q(\xi|\mathbf{y})$  is obtained, the pixelwise  $(\alpha/2)100\%$  and  $(1 - \alpha/2)100\%$  quantiles  $\xi_-$  and  $\xi_+$  are found directly from the Gaussian distribution quantile functions. Then, the  $(1 - \alpha)100\%$  pixelwise credible interval map  $\mathbf{x}_{CI}$  can be obtained as  $\mathbf{x}_{CI} = \mathbf{f}^{-1}(\xi_+) - \mathbf{f}^{-1}(\xi_-)$ . Note that the matrix  $\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}$  has the dimension of  $N \times N$ . With moderate number of voxles (i.e.,  $N < 10^4$ ), the inversion in (29) can be directly performed. However, with a large number of voxels, instantiating the covariance matrix and computing the inversion in (29) can be computationally very expensive.

In order for more efficient covariance matrix computation, the Hessian of the log-likelihood can be approximated with the Fisher information matrix at the mode  $\mathbf{F}(\hat{\xi})$ ,

$$\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \approx \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[ \left( \nabla_{\xi} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right) \left( \nabla_{\xi} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right)^{\top} \right] \tag{30}$$

$$= \left( \mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}} \right)^{\top} \mathbf{A}^{\top} \text{diag}\left( (\mathbf{A} \mathbf{f}^{-1}(\xi))^{\circ-1} \right) \mathbf{A} \left( \mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}} \right) \tag{31}$$

$$= \mathbf{B} \mathbf{B}^{\top}, \tag{32}$$

where we define  $\mathbf{B} \equiv \left( \mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}} \right)^{\top} \mathbf{A}^{\top} \text{diag}\left( (\mathbf{A} \mathbf{f}^{-1}(\xi))^{\circ-1} \right)$ . The detailed derivation of (31) and the expression for  $\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)$  is presented in the accompanying "Supplementary information". Fisher information has been studied extensively for neural network Hessian approximation, and it is well-justified as it is the expected Hessian of log-likelihood under  $p(\mathbf{y}|\bar{\mathbf{y}})$ . Equipped with the Hessian of the log-likelihood approximation, the expression for the covariance matrix in (29) becomes

$$\left( \nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}} \right)^{-1} \approx (\mathbf{B} \mathbf{B}^{\top} + \Sigma^{-1})^{-1} \tag{33}$$

$$= \mathbf{L} (\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L} + \mathbf{I})^{-1} \mathbf{L}^{\top}. \tag{34}$$

Next, we perform rank- $K$  truncated eigendecomposition on  $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L}$ , such that  $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L} \approx \mathbf{Q} \mathbf{S} \mathbf{Q}^{\top}$  where  $\mathbf{Q} \in \mathbb{R}^{N \times K}$  is a matrix containing eigenvectors and  $\mathbf{S} \in \mathbb{R}^{K \times K}$  is a diagonal matrix whose elements contain  $K$  largest eigenvalues of  $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L}$ . Such decomposition can be efficiently obtained from randomized eigendecomposition algorithms. Randomized decomposition methods first proceed by forming a matrix  $\mathbf{Y} \in \mathbb{R}^{N \times K}$  by multiplying a random matrix  $\mathbf{P} \in \mathbb{R}^{N \times K}$  such that

$$\mathbf{Y} = (\mathbf{L}^{\top} \mathbf{B} (\mathbf{B}^{\top} \mathbf{L} \mathbf{P})). \tag{35}$$

The matrix  $\mathbf{Y}$  captures most of the column space in  $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L}$  yet considerably smaller with  $K \ll \min(M, N)$ . Therefore, it can be further processed to efficiently estimate the eigendecomposition of  $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L}$ . More details on the randomized eigendecomposition methods can be found in Halko et al.<sup>51</sup>. When the system matrix  $\mathbf{A}$  is computed on-the-fly or too large to be loaded into core memory, the decomposition can be obtained efficiently with only single-pass of  $\mathbf{A}$  using the matrix sketch algorithm<sup>52</sup>. Also, note that by computing the matrix-matrix multiplication in the order specified by the parentheses in (35), we never instantiate an  $N \times N$  matrix, which is potentially too large to be stored in memory.

Once the eigendecomposition is obtained, the covariance matrix can be computed as follows:

$$\left(\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}\right)^{-1} \approx \mathbf{L}(\mathbf{Q}\mathbf{S}\mathbf{Q}^{\top} + \mathbf{I})^{-1}\mathbf{L}^{\top} \quad (36)$$

$$= \mathbf{L}(\mathbf{I} - \mathbf{Q}(\mathbf{S}^{-1} + \mathbf{Q}^{\top}\mathbf{Q})^{-1}\mathbf{Q}^{\top})\mathbf{L}^{\top} \quad (37)$$

$$= \mathbf{L}(\mathbf{I} - \mathbf{Q}(\mathbf{S} \oslash (\mathbf{S} + \mathbf{I}))\mathbf{Q}^{\top})\mathbf{L}^{\top} \quad (38)$$

$$= \mathbf{L}\mathbf{L}^{\top} - \mathbf{L}\mathbf{Q}(\mathbf{S} \oslash (\mathbf{S} + \mathbf{I}))\mathbf{Q}^{\top}\mathbf{L}^{\top} \quad (39)$$

The second equality follows from the Woodbury matrix identity<sup>53</sup>. The resulting expression (39) for the covariance matrix can be partially computed to obtain necessary covariance elements for uncertainty quantification. For example, to create a pixel-wise uncertainty map, only the diagonal elements can be computed without forming an  $N \times N$  matrix.

Due to the unimodality of the approximating Gaussian distribution, the Laplace approximation is well-justified when the posterior density is concentrated around the mode of the posterior,  $\hat{\xi}$ . The so-called ‘‘Bayesian central limit theorem’’ (i.e., Bernstein-von Mises theorem) states that with increasing number of conditionally independent measurements, posterior distributions tend toward Gaussian distributions under certain conditions<sup>54</sup>. Hence, with a sufficient number of measurements, the Laplace approximation provides fast, yet reasonably accurate approximation of the posterior distribution.

### Hyperparameter selection with empirical Bayes

Although the hyperparameters of the GPP offers good physical and statistical interpretability, a principled method to infer the hyperparameters from the measurement is highly desirable. Here we show that the hyperparameters  $\sigma$  and  $\lambda$  can be systematically determined with the empirical Bayes method.

Including the hyperparameter vector  $\theta = [\sigma, \lambda]^{\top}$ , the Bayes rule (8) can be re-written in terms of  $\xi$  as

$$p(\xi|y, \theta) = \frac{p(y|\xi, \theta)p(\xi|\theta)}{p(y|\theta)}, \quad \text{where} \quad (40)$$

$$p(y|\theta) = \int p(y|\xi, \theta)p(\xi|\theta) d\xi. \quad (41)$$

The empirical Bayes approach finds the hyperparameters  $\hat{\theta}$  maximizing the marginal likelihood as follows:

$$\hat{\theta} = \underset{\theta, \theta > 0}{\operatorname{argmin}} -\log p(y|\theta) = \underset{\theta, \theta > 0}{\operatorname{argmin}} \Phi(\theta), \quad (42)$$

where  $\Phi(\theta) \stackrel{c}{=} -\log p(y|\theta)$  is the unnormalized log-marginal likelihood. However, as the computation of  $\Phi(\theta)$  is intractable due to the high dimensional integration in (40), we use the Laplace approximation to estimate  $\Phi(\theta)$ <sup>55</sup>. With the Laplace approximation, the unnormalized negative log-marginal likelihood  $\Phi(\theta)$  at the posterior mode  $\hat{\xi}$  is given as

$$\Phi(\theta) \approx \frac{1}{2} \log |\mathbf{I} + \nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \Sigma| + \Psi(\hat{\xi}). \quad (43)$$

The derivation can be found in the ‘‘Supplementary information’’. Note that in (43), the  $\Psi(\hat{\xi})$  term is an output of Algorithm 1. The computation of the log-determinant term has the complexity of  $\mathcal{O}(N^3)$ , which prevents scalable computation of the log-marginal likelihood when  $N$  is large. Therefore, we exploit the low-rank structure of the matrix  $\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \Sigma$  to accelerate the computation of the log determinant.

As in (31), using the Fisher information matrix approximation to  $\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}}$ , the determinant can be approximated as,

$$\Phi(\theta) \approx \frac{1}{2} \log |\mathbf{I} + \mathbf{B}\mathbf{B}^{\top} \Sigma| + \Psi(\hat{\xi}). \quad (44)$$

We also observe that using the precomputed Cholesky decomposition  $\Sigma = \mathbf{L}\mathbf{L}^{\top}$ , the matrix determinant lemma enables the following reformulation of the log-determinant term,

$$\frac{1}{2} \log |\mathbf{I} + \mathbf{B}\mathbf{B}^{\top} \Sigma| = \frac{1}{2} \log |\mathbf{I} + \mathbf{B}^{\top} \mathbf{L}\mathbf{L}^{\top} \mathbf{B}| \quad (45)$$

$$= \frac{1}{2} \log |\mathbf{I} + \mathbf{L}^{\top} \mathbf{B}\mathbf{B}^{\top} \mathbf{L}|. \quad (46)$$

The Kronecker structure of the Cholesky factor  $\mathbf{L}$  allows for very efficient matrix-matrix multiplication  $\mathbf{B}^{\top} \mathbf{L}$  or  $\mathbf{L}^{\top} \mathbf{B}$  in (45) and (46). Furthermore, depending on the dimensions of the system matrix  $\mathbf{A}$ , (45) (when  $M \ll N$ ) or (46) (when  $N \ll M$ ) can be used to compute the log determinant efficiently.

---

**Require:**  $\mathbf{A}, \mathbf{y}, \xi_w^{\text{init}}, \theta^{\text{init}}, \varepsilon_1, \varepsilon_2$

$k \leftarrow 0$

$\xi_w^k \leftarrow \xi_w^{\text{init}}$

$\theta^k \leftarrow \theta^{\text{init}}$  ▷ Initialization

Convergence  $\leftarrow$  False

**while** Convergence is False **do**

$\xi_w^{k+1}, \hat{\Psi}^k \leftarrow$  Algorithm 1 ( $\mathbf{A}, \mathbf{y}, \theta^k, \xi_w^k, \varepsilon_1$ )

$\Phi(\theta^k) \leftarrow$  (44) ▷ (45), (46), or (47) for the determinant

$\log \theta^{k+1} \leftarrow$  L-BFGS ( $\log \theta^k, \Phi(\theta^k)$ ) ▷ L-BFGS iteration, logarithm for the positivity of the hyperparameters

**if**  $|\Phi(\theta^k) - \Phi(\theta^{k+1})| < \varepsilon_2$  **then**

        Convergence  $\leftarrow$  True

**else**

$k \leftarrow k + 1$

**end if**

**end while**

$\hat{\theta} \leftarrow \theta^{k+1}$  ▷ Selected hyperparameters

---

### Algorithm 2. Hyperparameter selection with empirical Bayes

When both the  $M$  and  $N$  are too large, a low rank approximation of the system matrix  $\mathbf{A} \approx \mathbf{U}_K \mathbf{V}_K^T$  where  $\mathbf{U}_K \in \mathbb{R}^{M \times K}$  and  $\mathbf{V}_K \in \mathbb{R}^{N \times K}$  with  $K \ll M, N$  can be used to accelerate the computation. Defining  $\mathbf{C} = \mathbf{V}_K^T \mathbf{J}_\xi f^{-1}(\xi)|_{\xi=\hat{\xi}} \mathbf{L}$  and  $\mathbf{D} = \mathbf{U}_K^T \text{diag}(\mathcal{K} \oslash \bar{\mathbf{y}}^{\circ \frac{1}{2}})$ , the log-determinant term can be approximated as,

$$\frac{1}{2} \log |\mathbf{I} + \mathbf{W} \Sigma| \approx \frac{1}{2} \log |\mathbf{I} + \mathbf{C} \mathbf{C}^T \mathbf{D} \mathbf{D}^T|. \quad (47)$$

Hence using the rank- $K$  approximation, the log-determinant computation has  $\mathcal{O}(K^3)$  time complexity. The low-rank approximation of  $\mathbf{A}$  needs to be computed only once using various techniques, and we found that randomized singular value decomposition (rSVD) provides a good approximation with the time complexity of  $\mathcal{O}(MN \log K + (M + N)K^2)^{51}$ . Such approximation of  $\mathbf{A}$  may sacrifice some accuracy in image reconstruction; however, we found that the approximation has negligible impact for hyperparameter selection and once the hyperparameters are chosen, the full-rank system matrix can be used for more accurate image reconstruction. Additionally, the low rank approximation can also reduce the computation of forward projection  $\mathbf{A} \mathbf{f}^{-1}(\xi)$ , which helps to accelerate the overall optimization process.

Once an adequate log-marginal likelihood computation scheme is selected, the L-BFGS can be used to solve the optimization problem (42). The necessary gradient of  $\Phi(\theta)$  with respect to  $\theta$  can be found with finite difference methods. The entire marginal likelihood maximization scheme is illustrated in Algorithm 2.

### Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

### Code availability

The code used for analysis can be found at [https://github.com/jwonl/GPP\\_public](https://github.com/jwonl/GPP_public).

Received: 30 January 2024; Accepted: 27 August 2024

Published online: 03 October 2024

### References

- Hussain, S. *et al.* Modern diagnostic imaging technique applications and risk factors in the medical field: A review. *Biomed. Res. Int.* **2022**, 5164970. <https://doi.org/10.1155/2022/5164970> (2022).
- Wang, B., Zhong, S., Lee, T.-L., Fancey, K. S. & Mi, J. Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review. *Adv. Mech. Eng.* **12**, 1687814020913761. <https://doi.org/10.1177/1687814020913761> (2020).
- Vetter, K. *et al.* Gamma-Ray imaging for nuclear security and safety: Towards 3-D gamma-ray vision. *Nucl. Instrum. Methods Phys. Res., Sect. A* **878**, 159–168 (2018).
- Rockmore, A. J. & Macovski, A. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nucl. Sci.* **23**, 1428–1432. <https://doi.org/10.1109/TNS.1976.4328496> (1976).
- Calvetti, D. & Somersalo, E. Inverse problems: From regularization to Bayesian inference. *WIREs Comput. Stat.* **10**, e1427. <https://doi.org/10.1002/wics.1427> (2018).
- Gockenbach, M. S. Linear Inverse Problems and Tikhonov Regularization (American Mathematical Soc., 2016).
- Panin, V., Zeng, G. & Gullberg, G. Total variation regulated EM algorithm. In 1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No.98CH36255), vol. 3, 1562–1566. <https://doi.org/10.1109/NSSMIC.1998.773841> (IEEE, Toronto, Ont., Canada, 1998).
- Zhang, K. *et al.* Plug-and-Play Image Restoration with Deep Denoiser Prior. <https://doi.org/10.48550/arXiv.2008.13751> (2021).
- Stuart, A. M. Inverse problems: A Bayesian perspective. *Acta Numer.* **19**, 451–559. <https://doi.org/10.1017/S0962492910000061> (2010).
- Gal, Y., Koumoutsakos, P., Lanusse, F., Louppe, G. & Papadimitriou, C. Bayesian uncertainty quantification for machine-learned models in physics. *Nat. Rev. Phys.* **4**, 573–577. <https://doi.org/10.1038/s42254-022-00498-4> (2022).



11. Bardsley, J. M. & Goldes, J. Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Prob.* **25**, 095005. <https://doi.org/10.1088/0266-5611/25/9/095005> (2009).
12. Gelman, A. *et al.* *Bayesian Data Analysis* 3rd edn. (CRC Press, Boca Raton, 2013).
13. MacKay, D. J. Introduction to Gaussian processes. *NATO ASI Ser. F Comput. Syst. Sci.* **168**, 133–166 (1998).
14. Gaussian Processes for Machine Learning: Contents. <https://gaussianprocess.org/gpml/chapters/>.
15. Cotter, S. L., Roberts, G. O., Stuart, A. M. & White, D. MCMC methods for functions: Modifying old algorithms to make them faster. *Stat. Sci.[SPACE]* <https://doi.org/10.1214/13-ST5421> (2013).
16. MacKay, D. J. C. A practical Bayesian framework for backpropagation networks. *Neural Comput.* **4**, 448–472. <https://doi.org/10.1162/neco.1992.4.3.448> (1992).
17. Daxberger, E. *et al.* Laplace redux – effortless Bayesian deep learning. <https://doi.org/10.48550/arXiv.2106.14806> (2022).
18. Zhou, Q., Yu, T., Zhang, X. & Li, J. Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data. *SIAM J. Imag. Sci.* **13**, 29–52. <https://doi.org/10.1137/19M1248352> (2020).
19. Vetter, K. *et al.* Advances in nuclear radiation sensing: Enabling 3-D gamma-ray vision. *Sensors* **19**, 2541. <https://doi.org/10.3390/s19112541> (2019).
20. Barnowski, R., Haefner, A., Mihailescu, L. & Vetter, K. Scene data fusion: Real-time standoff volumetric gamma-ray imaging. *Nucl. Instrum. Methods Phys. Res., Sect. A* **800**, 65–69. <https://doi.org/10.1016/j.nima.2015.08.016> (2015).
21. Hellfeld, D. *et al.* Gamma-ray point-source localization and sparse image reconstruction using Poisson likelihood. *IEEE Trans. Nucl. Sci.* **66**, 2088–2099. <https://doi.org/10.1109/TNS.2019.2930294> (2019).
22. Quiter, B. J. *et al.* Ongoing advancement of free-moving radiation imaging and mapping. in *Hard X-Ray, Gamma-Ray, and Neutron Detector Physics XXIV* Vol. 12241, pp. 131–137. <https://doi.org/10.1117/12.2635752> (2022).
23. Sutton, O. G. The problem of diffusion in the lower atmosphere. *Q. J. R. Meteorol. Soc.* **73**, 257–281. <https://doi.org/10.1002/qj.49707331704> (1947).
24. Parajuli, R. K., Sakai, M., Parajuli, R. & Tashiro, M. Development and applications of Compton camera—a review. *Sensors (Basel, Switzerland)* **22**, 7374. <https://doi.org/10.3390/s22197374> (2022).
25. Cieślak, M. J., Gamage, K. A. A. & Glover, R. Coded-aperture imaging systems: Past, present and future development —a review. *Radiat. Meas.* **92**, 59–71. <https://doi.org/10.1016/j.radmeas.2016.08.002> (2016).
26. Hanson, K. M. & Wecksung, G. W. Bayesian approach to limited-angle reconstruction in computed tomography. *JOSA* **73**, 1501–1509. <https://doi.org/10.1364/JOSA.73.001501> (1983).
27. Catana, C. The dawn of a new era in low-dose PET imaging. *Radiology* **290**, 657–658. <https://doi.org/10.1148/radiol.2018182573> (2019).
28. Barbano, R., Arridge, S., Jin, B. & Tanno, R. Chapter 26 - Uncertainty quantification in medical image synthesis. In Burgos, N. & Svoboda, D. (eds.) *Biomedical Image Synthesis and Simulation*, The MICCAI Society Book Series, 601–641. <https://doi.org/10.1016/B978-0-12-824349-7.00033-5> (Academic Press, 2022).
29. Quantitative Compton Imaging in 3D | INMM Resources. <https://resources.inmm.org/annual-meeting-proceedings/quantitative-compton-imaging-3d>.
30. Sgouros, G. & Hobbs, R. F. Dosimetry for radiopharmaceutical therapy. *Semin. Nucl. Med.* **44**, 172–178. <https://doi.org/10.1053/j.semnuclmed.2014.03.007> (2014).
31. Knopf, A.-C. & Lomax, A. In vivo proton range verification: A review. *Phys. Med. Biol.* **58**, R131–160. <https://doi.org/10.1088/0031-9155/58/15/R131> (2013).
32. Schmidt, M. N. & Laurberg, H. Nonnegative matrix factorization with Gaussian process priors. *Comput. Intell. Neurosci.* **1–10**, 2008. <https://doi.org/10.1155/2008/361705> (2008).
33. Barrett, H. H., White, T. & Parra, L. C. List-mode likelihood. *J. Opt. Soc. Am. A, Opt., Image Sci., Vis.* **14**, 2914–2923 (1997).
34. 4- $\pi$  Compton imaging with single 3D position-sensitive CdZnTe detector. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5540/0000/4-pi-Compton-imaging-with-single-3D-position-sensitive-CdZnTe/10.1117/12.563905.short?SSO=1>
35. Hellfeld, D., Barton, P., Gunter, D., Mihailescu, L. & Vetter, K. A spherical active coded aperture for 4 $\pi$  gamma-ray imaging. *IEEE Trans. Nucl. Sci.* **64**, 2837–2842. <https://doi.org/10.1109/TNS.2017.2755982> (2017).
36. Hellfeld, D. *et al.* Real-time free-moving active coded mask 3D gamma-ray imaging. *IEEE Trans. Nucl. Sci.* **66**, 2252–2260. <https://doi.org/10.1109/TNS.2019.2939948> (2019).
37. Wilson, A. G. & Nickisch, H. Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP). <https://doi.org/10.48550/arXiv.1503.01057> (2015).
38. Lindgren, F., Rue, H. & Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **73**, 423–498. <https://doi.org/10.1111/j.1467-9868.2011.00777.x> (2011).
39. Wilson, A. G. & Adams, R. P. Gaussian process kernels for pattern discovery and extrapolation. <https://doi.org/10.48550/arXiv.1302.4245> (2013).
40. Wilson, A. G., Hu, Z., Salakhutdinov, R. & Xing, E. P. Deep Kernel learning. <https://doi.org/10.48550/arXiv.1511.02222> (2015).
41. Shepp, L. A. & Vardi, Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**, 113–122. <https://doi.org/10.1109/TMI.1982.4307558> (1982).
42. Bayes, T. LII. An essay towards solving a problem in the doctrine of chances By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton A. M. F. R. S.. *Philos. Trans. R. Soc. Lond.* **53**, 370–418. <https://doi.org/10.1098/rstl.1763.0053> (1997).
43. Duvenaud, D. Automatic model construction with Gaussian processes. Ph.D. thesis, University of Cambridge (2014).
44. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528. <https://doi.org/10.1007/BF01589116> (1989).
45. Saatçi, Y. Scalable inference for structured Gaussian process models. Ph.D. thesis, Citeseer (2012).
46. Pereyra, V. & Scherer, G. Efficient computer manipulation of tensor products with applications to multidimensional approximation. *Math. Comput.* **27**, 595–605. <https://doi.org/10.2307/2005663> (1973).
47. Ouyang, X., Wong, W., Johnson, V., Hu, X. & Chen, C.-T. Incorporation of correlated structural images in PET image reconstruction. *IEEE Trans. Med. Imaging* **13**, 627–640. <https://doi.org/10.1109/42.363105> (1994).
48. Bathke, C., Kluth, T., Brandt, C. & Maaß, P. Improved image reconstruction in magnetic particle imaging using structural a priori information. *Int. J. Mag. Part. Imag. JMRI* **3**, 1703015 (2017).
49. Wang, G. & Qi, J. PET image reconstruction using Kernel method. *IEEE Trans. Med. Imag.* **34**, 61–71. <https://doi.org/10.1109/TMI.2014.2343916> (2015).
50. Speagle, J. S. A Conceptual introduction to Markov Chain Monte Carlo methods. <https://doi.org/10.48550/arXiv.1909.12313> (2020).
51. Halko, N., Martinsson, P.-G. & Tropp, J. A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. <https://doi.org/10.48550/arXiv.0909.4061> (2010).
52. Tropp, J. A., Yurtsever, A., Udell, M. & Cevher, V. Practical sketching algorithms for low-rank matrix approximation. *SIAM J. Matrix Anal. Appl.* **38**, 1454–1485. <https://doi.org/10.1137/17M1111590> (2017).
53. Woodbury, M. A. *Inverting modified matrices* (Princeton University, Princeton, NJ, 1950).
54. van der Vaart, A. W. *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 1998).

55. Pierre-Simon Laplace - Œuvres complètes, tome 6. [http://sites.mathdoc.fr/cgi-bin/oetoc?id=OE\\_LAPLACE\\_\\_6](http://sites.mathdoc.fr/cgi-bin/oetoc?id=OE_LAPLACE__6).

### Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Berkeley National Laboratory under Contract DE-AC02-05CH11231.

### Author contributions

J.L. conceptualized and implement the methods presented. T.H.J., M.S.B., D.L.G., B.J.Q., R.J.C., and K.V. provided supervision and reviewed the methods and the results presented. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-71336-z>.

**Correspondence** and requests for materials should be addressed to J.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024