

Probability density function estimation of laser light scintillation via Bayesian mixtures

Eric X. Wang,^{1,*} Svetlana Avramov-Zamurovic,² Richard J. Watkins,³
Charles Nelson,⁴ and Reza Malek-Madani¹

¹Mathematics Department, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

²Weapons and Systems Engineering, United States Naval Academy, Annapolis, Maryland 21402, USA

³Mechanical Engineering Department, United States Naval Academy, Annapolis, Maryland 21402, USA

⁴Electrical Engineering Department, United States Naval Academy, Annapolis, Maryland 21402, USA

*Corresponding author: wang73@llnl.gov

Received August 29, 2013; accepted November 18, 2013;
posted December 24, 2013 (Doc. ID 195830); published February 14, 2014

A method for probability density function (PDF) estimation using Bayesian mixtures of weighted gamma distributions, called the Dirichlet process gamma mixture model (DP-GaMM), is presented and applied to the analysis of a laser beam in turbulence. The problem is cast in a Bayesian setting, with the mixture model itself treated as a random process. A stick-breaking interpretation of the Dirichlet process is employed as the prior distribution over the random mixture model. The number and underlying parameters of the gamma distribution mixture components as well as the associated mixture weights are learned directly from the data during model inference. A hybrid Metropolis–Hastings and Gibbs sampling parameter inference algorithm is developed and presented in its entirety. Results on several sets of controlled data are shown, and comparisons of PDF estimation fidelity are conducted with favorable results. © 2014 Optical Society of America

OCIS codes: (000.5490) Probability theory, stochastic processes, and statistics; (010.1300) Atmospheric propagation; (010.1330) Atmospheric turbulence; (010.7060) Turbulence; (290.5930) Scintillation; (150.1135) Algorithms.

<http://dx.doi.org/10.1364/JOSAA.31.000580>

1. INTRODUCTION

The estimation of probability density functions (PDFs) to characterize light intensity in a turbulent atmosphere is a long-standing problem in optics [1–7]. Understanding the probability distribution of the light intensity is important in many areas, including optical communications, lidar systems, and directed energy weaponry. This paper presents a method of density estimation using hierarchical nonparametric Bayesian methods, specifically Dirichlet-process-distributed Bayesian mixtures. Nonparametric Bayesian models are ones that adapt their complexity to the data within a model class without relying on the modeler to define model complexity. This allows such models to be less prone to over- and under-fitting. For completeness, this paper presents our model in full detail and assumes a familiarity with statistical methods on the part of the reader. For a thorough review on Bayesian modeling and methods, see [8].

Traditionally, PDF estimates of laser light intensity in turbulent atmosphere are based on stochastic physics-driven models, heuristic concepts, or fitting via lower order moments of the data [9]. Several well-known and widely used examples of PDF estimation algorithms for laser light intensity include the gamma distribution [10], the gamma-gamma model [11], and the log-normal distribution [7]. Significant ongoing research is devoted to the merits of the various approaches and is summarized well in a recent survey paper [9]. Physics-driven models and heuristic approaches are parametric in that they attempt to fit a distribution of known form to the observed scattered light, with goodness-of-fit assessed by

computing the root mean square (RMS) error of the estimated PDF to a discrete histogram.

An exception to the highly parametric approach to PDF estimation in optics has been a series of papers that proposed estimating the PDF of the scattered light intensity in turbulence from the first several moments of the data (usually up to the fifth moment) [12,13]. The novelty of this approach is its nonparametric nature. That is, it does not assume that the PDF of the observed data must follow a basic functional form of the PDF, but instead creates highly flexible PDFs by modulating a gamma distribution using weighted Laguerre polynomials. This approach has been used recently in [14].

While the gamma-Laguerre method of [13] is nonparametric and more data-driven than previous approaches in optics, it has several drawbacks. First, the estimated PDFs cannot be guaranteed to be nonnegative everywhere [15,16]. Second, the approach is highly sensitive to outliers in the higher order moments, meaning that the number of moments considered is usually driven by computational considerations [15]. This sensitivity to outliers also causes significant oscillations of the PDF [14,17]. Finally, the number of moments considered has a significant effect on the resultant shape of the PDF [13].

This paper explores an alternative avenue of nonparametric PDF estimation: mixture models. Mixture models are popular in the statistical and machine learning fields but to the best of the authors' knowledge are not widely used in the optics community. PDFs generated by mixture model approaches are constructed as linear combinations of overlapping probability distributions on a shared sample space. In [18], the authors

showed that mixture models are generalizations of the well-known kernel density estimators (KDE) [19,20]. KDEs place a smooth kernel with normalized area at each observation with the resultant PDF generated by summing the kernels. Mixture models take a similar approach to kernel estimators but allow subpopulations of the data to cluster into separate overlapping distributions called mixture components. These mixture components are then appropriately weighted and summed such that the resultant PDF integrates to 1.

By construction, kernel and mixture models yield proper PDFs. Moreover, such methods do not rely on the estimation or computation of lower order moments from the data. In particular, mixture models are an attractive solution to density estimation due to the small set of parameters needed to characterize a PDF. However, a drawback of the classical mixture model is the need to choose the number of mixture components. This choice can significantly affect the resulting PDF.

To address the issue of choosing the number of mixture components, [18] proposed Bayesian mixtures for density estimation. This approach casts the PDF estimation problem in a Bayesian statistical context and treats the mixture model itself (and hence the PDF) as random, endowed with a Dirichlet process (DP) prior distribution [21]. DP is the natural choice for a prior distribution over the mixture model due to computational convenience. DP-distributed mixture models, often referred to as DP mixture models, have been widely used in Bayesian machine learning not only as PDF estimators but as powerful nonparametric clustering tools with applications in image processing [22], language and topic modeling [23], and analysis of legislative roll call votes [24], among many others. Previously, [21] and [18] showed that PDF estimation via DP mixture models can be viewed as a data clustering problem.

We adopt the framework of the DP mixture model for PDF estimation but tailor its details for the specific application of estimating the PDF of light intensity in turbulent atmosphere. In doing so, we differ from previous work, specifically [21] and [18], not only in application but also in several important technical ways. Both [21] and [18] used a parameter inference scheme that iterates over each data point sequentially, while we adopt the mathematically equivalent but computationally more efficient “stick-breaking” scheme of [25] that allows us to consider all the data simultaneously. Moreover, [18] considered Gaussian distributions as their mixture components while our approach uses gamma distributions as mixture components. This model design choice was for consistency with the gamma-distribution-based methods in previous work such as [3] and [13]. In this spirit, we also follow the data preprocessing in [13]. We note, however, that our approach is general and requires neither normalization of the data to a specific mean or limiting the mixture components to be gamma distributions. Finally, our work differs from previous work using mixtures of gamma distributions such as [26], [27], and [28] in that none of these approaches adopts the aforementioned stick-breaking construction of the DP. We will henceforth refer to our model as the DP gamma mixture model (DP-GaMM).

We also address the issue of evaluating goodness-of-fit of the PDF to the observed data. The primary approach to-date within the field of optics has been to compute the root mean square error between the estimated PDF at discrete points to a

normalized frequency plot of the observations [6,9,29]. Such a metric is inherently dependent on and affected by the choice of bin widths used to construct the frequency plot. Furthermore, it does not address the innate stochastic nature of a PDF. We present an alternative method of evaluating goodness-of-fit of held-out data to the estimated PDF using log-likelihood. This approach is widely used in statistics and machine learning [23,30] as a reliable and stochastic metric for deciding which model (from a selection of models) was most likely—in a probabilistic sense—to have generated the data [31]. The most likely model is deemed to have the most predictive skill in describing future observations under similar atmospheric and experimental conditions. The advantage of the held-out likelihood test is that it assesses the ability of the model to generalize to data that was previously unseen but statistically related to the observed training data.

The rest of this paper is organized as follows. Section 2 presents the results of applying the DP-GaMM to laser beam propagation data collected in a lab setting at the United States Naval Academy (USNA). Section 3 discusses the mathematical development of the DP-GaMM. Parameter inference is discussed in Section 4. We compare the PDF learned via the DP-GaMM to classical PDF fitting methods via held-out log-likelihood in Section 5. Additionally, in the Section 5.B we demonstrate the DP-GaMM’s versatility by modeling raw, unprocessed data collected in a maritime environment. Finally, we conclude in Section 6.

2. BAYESIAN MIXTURES

Mixture models are probability distributions comprised of a weighted sum of parametrically known distributions, as shown in Eq. (1) below. Each one of the individual distributions represents the underlying distribution for a subpopulation of the observed data and is called a mixture component. In the case of observing light intensity in turbulent atmosphere, the data we consider is a time series of light intensity at a single pixel. Thus, subpopulations in this case are clusters of similarly valued intensities. The advantage of a mixture model approach to PDF estimation is in its computational simplicity relative to traditional methods such as [3] and [13]. It requires neither moments of the data to be computed, nor environmental or physical factors to be known. Compared to the conceptually similar KDE, mixture models are characterized by a significantly smaller set of parameters. These parameters are learned either via expectation maximization or via Bayesian inference methods. Bayesian approaches such as variational Bayes [32] and the Gibbs sampler [33] offer the additional advantage of full posterior distributions on the model parameters, helping to quantify model parameter uncertainty.

For a random variable x , a mixture model is defined as

$$p(x|\pi, \Phi) = \pi_1 f(x|\phi_1) + \pi_2 f(x|\phi_2) + \pi_3 f(x|\phi_3) + \dots, \quad (1)$$

where $\pi = \{\pi_k\}_{k=1,2,3,\dots}$ is a set of mixing weights, $\Phi = \{\phi_k\}_{k=1,2,3,\dots}$ is the set of parameters of the mixture components, with ϕ_k defined as the set of parameters of the k th mixture component $f(x|\phi_k)$. The mixing weights are constrained to sum to 1, $\sum_k \pi_k = 1$, and the mixture components are themselves probability distributions of a known form, for example gamma distributions. Notationally, K denotes the choice of

the number of mixture components, i.e., $k = 1, 2, 3, \dots, K$. Notice that K is intentionally omitted in Eq. (1) to make the expression as general as possible. As discussed in [28], a key advantage of mixture models is their ability to represent any distribution within the support space of the mixture component as K approaches ∞ .

In this paper, for the specific application of modeling the stochastic behavior of light propagating in turbulence, the mixture components are chosen to be gamma distributions, following the approach of [3] and [13]. That is, we let $\phi_k = \{\eta_k, \mu_k\}$ and

$$f(x|\phi_k) = \text{Gamma}(x; \eta_k, \eta_k/\mu_k) = \frac{(\eta_k/\mu_k)^{\eta_k}}{\Gamma(\eta_k)} x^{\eta_k-1} \exp\left(-\frac{\eta_k}{\mu_k} x\right), \quad (2)$$

where η_k is the shape parameter and μ_k is the mean of the k th mixture component. It is straightforward to incorporate different, i.e., non-gamma, distributions as mixture components in the mixture model, and future work will explore the use of different distributions as mixture components in the analysis of light scattered in turbulence. We discuss our choice of mixture components more completely in Section 3.B.

Traditionally, estimation of the model parameters π and Φ is done using the expectation-maximization (EM) framework

[34]. In the mixture model setting, an auxiliary cluster assignment variable z is associated with each observation x and denotes the subpopulation x belongs to, i.e., $z \in [1, K]$ [35]. The EM algorithm iteratively alternates between finding the expectation (E) step where the observations are assigned to subpopulations, and the maximization (M) step where, conditioned on the assignments, the model parameters π and Φ are fitted. The EM mixture model is widely used across many disciplines as a *de facto* analysis tool for data clustering and PDF estimation. For an excellent survey of the field, see [36].

Despite the popularity of EM mixture models, a notable drawback is the need to manually set K *a priori*. This choice has a significant effect on the resulting behavior of the model. The affect this choice has is demonstrated on data collected at the United States Naval Academy in Fig. 1.

Before proceeding, it is worthwhile to give the details of the data. The data was collected in a controlled turbulence tunnel at the United States Naval Academy. The tunnel has six inch square holes in the top and bottom through which air (heated and unheated) can be blown by a variable speed fan. The beam is centered over the turbulence. The tunnel is otherwise closed from the outside. A 5 mm diameter 2 mW 633 nm He-Ne laser was used, with the distance from transmitter to the square hole set to 2.14 m. Total distance from transmitter

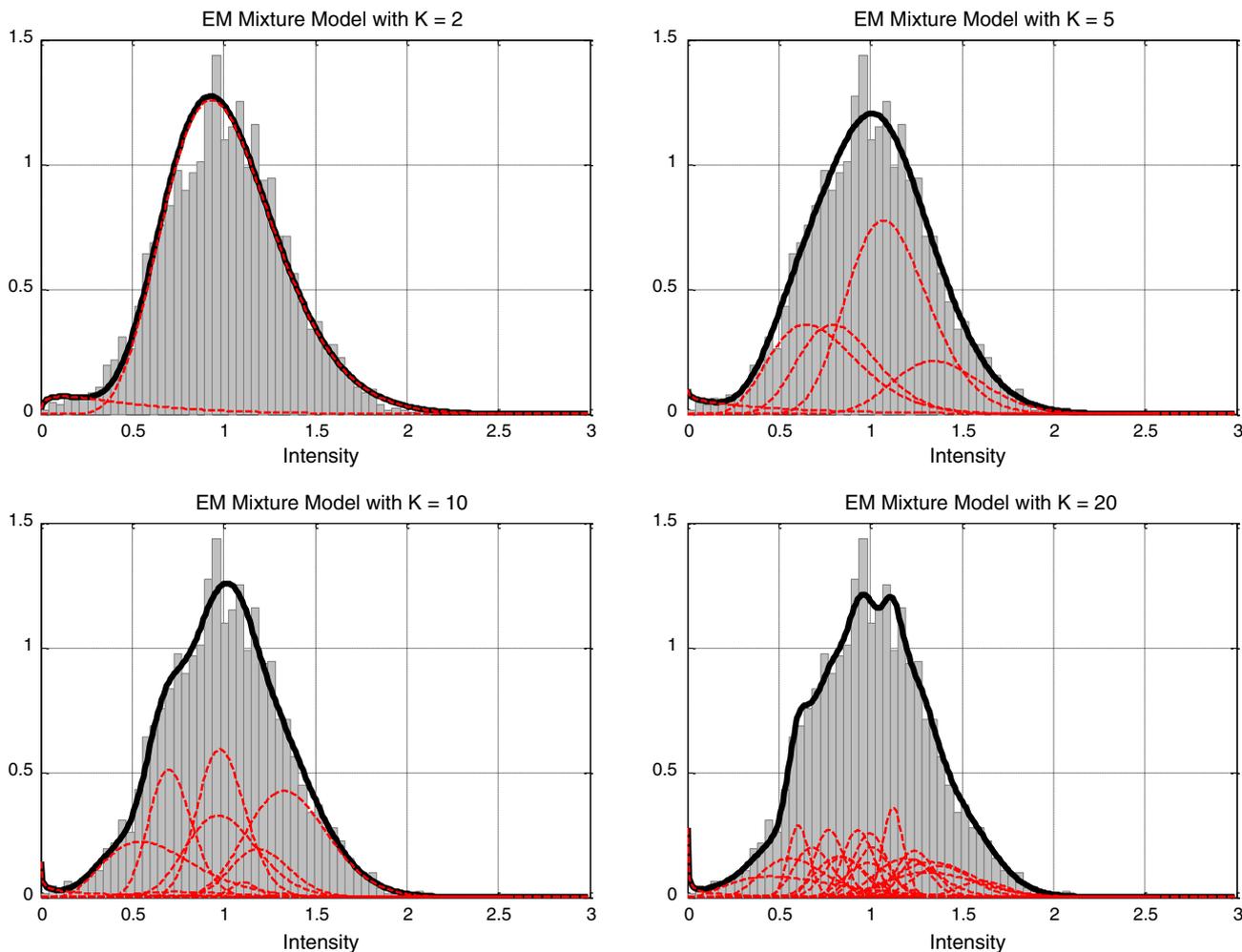


Fig. 1. Mixture model with different numbers of mixture components (clockwise from top left $K = 2, 5, 20, 10$). Mixture components shown in red and estimated PDF shown in black for Run 1, $SI = 7.99 \times 10^{-2}$. The normalized histogram in gray has area 1.

to receiver was 4.5 m. The measured temperature inside the tunnel was 23.6°C, and no additional heat was applied for this set of experiments. However, the air drawn into the tunnel was warmer than the air in the tunnel (as desired to cause some turbulence), due to the equipment operating on the floor near the tunnel slot. Each run consisted of 3636 frames, taken using a Ophir-Spiricon CCD Laser Beam Profiler, Model BG-USB-SP620, operating at 30 Hz, from which we computed the average hot pixel and extracted our time series from that location. For consistency with previous work, we normalized the data as in [13] and [14] by first subtracting the minimum intensity of the dataset and then dividing by the mean. As in [14], we choose to model the average hot pixel. We acknowledge that other methods of averaging or aggregation exist. However, since the goal of this paper is to discuss a new tool for estimating PDFs, we have decided to stay consistent with the methods presented in previous work. Three runs were conducted in this tunnel with varying fan speeds. For each run, we computed the scintillation index (SI) as

$$\text{SI} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} - 1, \quad (3)$$

where I is the intensity of the observed laser light, and $\langle \cdot \rangle$ denotes the empirical mean. To compute the SI and prepare the data for analysis, the data was normalized by removing the floor (minimum value) and dividing by the resultant mean. The three runs yielded SI values of 7.99×10^{-2} (Run 1), 8.42×10^{-2} (Run 2), and 10.75×10^{-2} (Run 3). We will refer to these experiments by these names henceforth.

The EM mixture models were trained on the data from Run 1 with several different choices for K ($K = 2, 5, 10, 20$), and the resultant PDFs are shown Fig. 1. The behavior of the model changes from a smooth PDF to a more jagged PDF as the number of mixture components increases. In theory, all of these PDFs are representative PDFs of the data. However, since the choice of K is entirely in the hands of the modeler, it is often difficult to decide which choice is best.

To address the issue of finding an ideal K , [18] proposed Bayesian mixtures for density estimation where the PDF in Eq. (1) is treated as a random mixture by modeling the mixture model as being DP distributed. Such a construction is denoted hierarchically as

$$x \sim G, G \sim \text{DP}(\alpha, G_0), \quad (4)$$

where the symbol \sim denotes “drawn from,” and $G = p(x|\pi, \Phi)$ is the mixture model defined in Eq. (1) and modeled as a DP-distributed mixture distribution. Note that by this notation, $x \sim G$ is equivalent to $p(x|\pi, \Phi)$. The DP-GaMM is described by Eqs. (1), (2), and (4), with the details of Eq. (4) discussed in the following sections.

In the Bayesian framework, the parameters of the mixture G , namely π and Φ , are themselves treated as random variables. Instead of learning a single maximally likely value for each model parameter, posterior distributions on both π and Φ are inferred through a stochastic random walk process discussed in the following section. As mentioned in the introduction, an attractive property of the DP-GaMM (and indeed, any DP mixture model) is that the number of mixture components in any mixture model is learned from the data during the parameter inference process, rather than set *a priori* by the modeler.

In Fig. 2, we show the mixture models learned using the construction sketched in Eq. (4) for the three different datasets. In each figure, we show a normalized histogram of the data (normalized such that the total area under the histogram is unity) in gray, the PDF $p(x|\pi, \Phi)$ in black, and the individual mixture components multiplied by their respective mixing weights $\{\pi_k \text{Gamma}(x_n; \eta_k, \eta_k / \mu_k)\}_{k=1,2,3,\dots,K}$ in the red dashed lines. It should be noted that the parameter learning and sampling algorithm for Bayesian models is iterative in nature. The mixtures shown are examples representing only a single iteration of the model. Specifically, we show the sample with the highest training data likelihood for each run. To fully characterize the posterior distribution on the set of all possible models, many samples (of the order of 10^4 iterations or more) need to be realized. The set of these samples then approximately characterize the posterior distribution on the space of all possible solutions and yields valuable information on the uncertainty of the underlying analysis. Application of the inferred model to predictive tasks is open-ended, depending on the needs of application. One could, for example, choose the sample with the highest training set log-likelihood (discussed later) to use as the PDF. However, this approach discards the notion of model uncertainty mentioned previously. A more rigorous method using average held-out log-likelihood is described in Section 5.

In all experiments within this paper, we set the maximum number of mixture components, called the truncation level, to 30 and allowed the model to prune out unneeded mixture components. Notice that the DP-GaMM encourages the resulting model to have only a few occupied mixture components through imposing a clustering on the data. An interesting observation is that in all experiments the DP-GaMM finds a small cluster with low intensity (far left side of the PDF's main lobe). In the case of Run 3, the model used four mixture components, while in Run 1 and Run 2 only three mixture components are used to describe the data. Finally, although it is unnoticeable in the figures, each PDF includes a small additional mixture component with mean near 0. While we do not as yet understand the physical phenomenon behind this cluster, its appearance in the analysis is consistent and suggests that our model may be useful in identifying physical phenomena that are not easily described by traditional heuristic methods. It is important to note that this truncation does not force the model to use every mixture component and is simply for computational purposes. It has been shown that the truncation level has no effect on the result if it is set sufficiently large [37].

To demonstrate the flexibility of the DP-GaMM, we found PDFs of Run 1 using three popular models: gamma-gamma (GG) [3], gamma (G) [10], and log-normal (LN) [7]. We then sampled 10,000 observations from each PDF and ran our DP-GaMM on each dataset. Below in Fig. 3, we demonstrate that the DP-GaMM can accurately capture the behavior of the three models by plotting the true distribution (in red) and the DP-GaMM's inferred distribution (in black). The gray bars denote a normalized frequency plot of the data sampled from the underlying distribution in each case, where the area of the gray plot has been normalized to 1. These results show that the GG, G, and LN PDFs can all be viewed as special cases of the DP-GaMM.

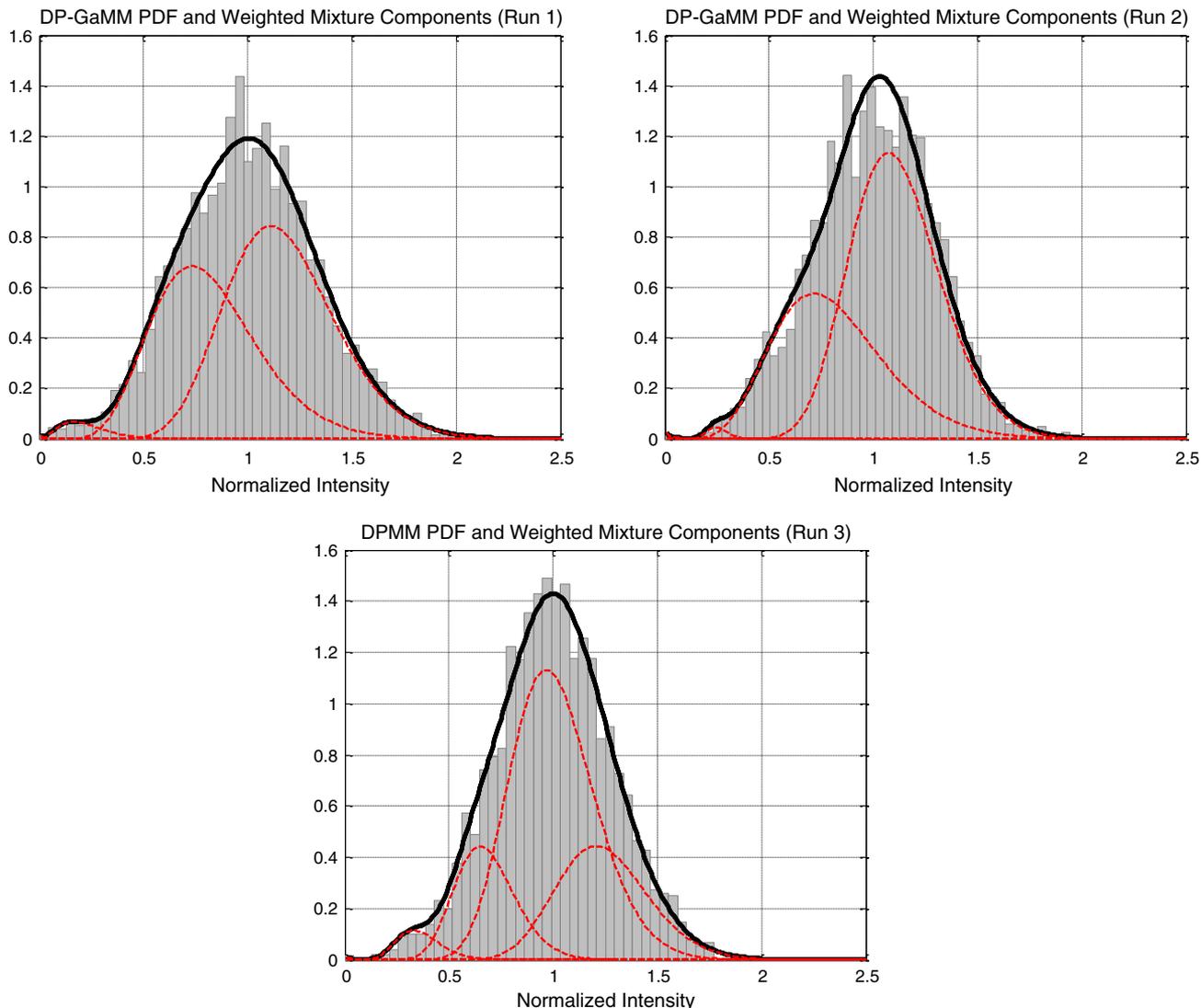


Fig. 2. Estimated DP-GaMM PDFs for three different runs. Mixture components shown in red and estimated PDF shown in black. The normalized histogram in gray has area 1. Top left: Run 1, $SI = 7.99 \times 10^{-2}$. Top right: Run 2, $SI = 8.42 \times 10^{-2}$. Bottom: Run 3, $SI = 10.75 \times 10^{-2}$.

3. DP-GaMM

We now present the details of the DP-GaMM. Consider the mixture model in Eq. (1). A DP, denoted by $DP(\alpha, G_0)$, is characterized by a concentration parameter α and a base measure G_0 . Draws from a DP are random mixture models, which we denote by G . Let $G = \sum_k \pi_k \delta_{\phi_k}$ define a random mixture model whose full form is defined in Eq. (1). The δ_{ϕ_k} is a Dirac delta supported at ϕ_k . Since, in a mixture model, the form of the mixture components f is defined ahead of time, the mixing weights π and the Φ are sufficient to define a mixture model.

Recall from the previous section that K , the number of mixture components in G , is assumed infinite in a DP mixture model. The requirement to have $K \rightarrow \infty$ means that the resulting random mixture model G has an infinite number of mixture components. In theory, because it has an infinite number of mixture components, a DP-distributed mixture model can model any distribution, provided that it has the same support as the mixture components. It is important to note that while the theoretical number of mixture components is infinite, only a small, finite number of those mixture components will have associated weights with appreciable values.

Sampling a random G from a DP is described in the next section.

A. Stick Breaking Construction of the Dirichlet Process

The task of actually sampling G from $DP(\alpha, G_0)$ has been the subject of much research [23]. A popular method is the constructive “stick-breaking” definition of a DP proposed in [25]. This construction utilizes successive “breaks” of a unit length stick to generate the mixing weights. Hierarchically, this is written as

$$\begin{aligned}
 x &\sim f(\phi_z), \\
 z &\sim \text{Mult}(\pi), \\
 G &= \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}, \\
 \pi_k &= V_k \prod_{l < k} (1 - V_l), \\
 V_k &\sim \text{Beta}(1, \alpha), \\
 \phi_k &\sim G_0,
 \end{aligned} \tag{5}$$

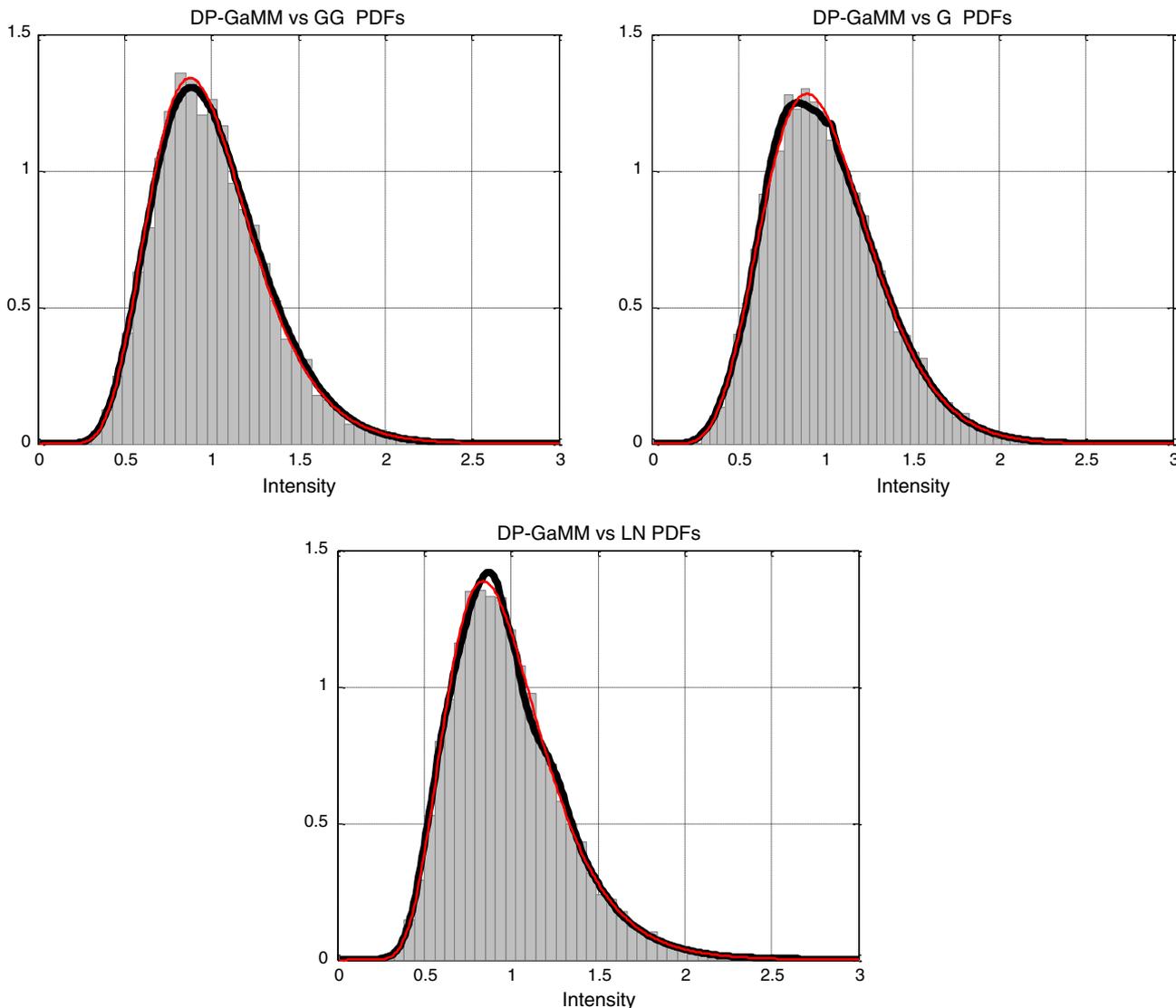


Fig. 3. Estimated DP-GaMM PDFs using data sampled from gamma–gamma (top left), gamma (top right), and log–normal (bottom). DP-GaMM PDF is shown in black, sampling PDF is shown in red, and normalized frequency (normalized to area 1) of the sampled data is shown in gray.

where V_k is a beta-distributed random variable that denotes the proportion of the remaining stick to be “broken,” and $\prod_{l < k} (1 - V_l)$ denotes the amount of the originally unit length stick remaining after k breaks. $\text{Beta}(1, \alpha)$ is a beta distribution, and α is the concentration parameter of the DP. The selection of $\text{Beta}(1, \alpha)$ is standard for the stick-breaking representation of DP [25]. Note that Eqs. (4) and (5) are equivalent, with $G \sim \text{DP}(\alpha, G_0)$ replaced by the specific construction for the stick-breaking representation of DP. Hierarchical Bayesian representations such as (5) are read from bottom to top, sampling the random variables from the distributions with hyperparameters (α , and the parameters of G_0) set by the modeler at the bottom to the observed data at the top. In this case, the base measure G_0 is a distribution whose form and parameters (the hyperparameters) are set by the modeler. The mixture component parameters ϕ_k are drawn independently and identically distributed from base measure G_0 . Finally, in order to sample data x from G , an auxiliary random positive integer z is introduced where $z = k$ means that x is drawn from the k th mixture component and $\text{Mult}(\pi)$ is a multinomial

distribution and $\pi = \{\pi_k\}_{k=1:K}$ are the mixing weights drawn from the stick-breaking construction.

In summary, the generative procedure for the observed data x via a stick-breaking construction of DP is as follows [notice that this follows the relations in (5) from bottom to top]:

- Independently and identically sample ϕ_k from G_0 , for $k = 1, \dots, K$.
- Generate the mixing weights by sampling V_k from $\text{Beta}(1, \alpha)$ for $k = 1, \dots, K$ and constructing the weights $\pi_k = V_k \prod_{l < k} (1 - V_l)$.
- For each observation index $n = 1, \dots, N$, sample z_n from $\text{Mult}(\{\pi_k\}_{k=1:K})$.
- For each observation index $n = 1, \dots, N$ and given z_n , sample the data x_n from the distribution $f(x_n | \phi_{z_n})$.

The stick-breaking construction of DP makes clear the parsimonious nature of DP as the π_k decreases quickly as k increases. Thus, most of the mass of the model will reside in a few mixture components with relatively small mixture index

k . In practice, as long as K is set to be sufficiently large such that unoccupied clusters result after parameter inference, the model can learn the number of mixture components necessary to represent the data.

Up to this point, the concentration parameter α has been treated as a constant that is set by the modeler *a priori*. Tuning of α gives the modeler some control over the relative complexity of the resultant model. A larger α results in more mixture components with appreciable mixing weights and therefore a more complex model. Conversely, a smaller α results in a less complex model. In some situations, it is desirable to infer α rather than set it. In this case, each mixture component has an associated α_k rather than setting a single α for all k . The mixture component specific α_k is drawn from a gamma distribution, denoted by $\alpha_k \sim \text{Gamma}(d, e)$. The setting of d and e is chosen by the modeler. In this case, we set $d = e = 10^{-6}$. We observe consistent model behavior for $10^{-6} < d, e < 1$, indicating that the model is relatively robust to choices of d and e .

It is important to note that in this section we outlined the generative process of a DP mixture model and have not yet discussed how to actually learn the model parameters. The key point here is that every variable in the generative process, aside from x_n , is latent and must be uncovered through Bayesian inference methods such as variational Bayes [32] or the Gibbs sampler [33]. Moreover, this section presented a general structure of the DP mixture model, without specifying a specific distribution form of $f(x_n|\phi_k)$. The following section discusses the choice of defining $f(x_n|\phi_k)$ as in Eq. (2).

B. Application to Laser Beams Propagating in Turbulence

Let $I = \{I_n\}_{n=1:N}$ denote the set of N observations of a photo sensor measuring beam intensity in turbulence. In order to be consistent with most popular methods of PDF estimation for laser light in turbulence [3,13,14], the following normalization is adopted

$$x_n = \frac{I_n}{\frac{1}{N} \sum_{n=1}^N I_n}, \tag{6}$$

although such normalization is not required for the DP-GaMM. As previously stated, we make the assumption that the normalized observations x_n are gamma distributed, thus $\phi_k = \{\eta_k, \mu_k\}$, where η_k is the shape parameter and μ_k is the mean of mixture component k ,

$$f(x_n|\phi_k) = \frac{(\eta_k/\mu_k)^{\eta_k}}{\Gamma(\eta_k)} x_n^{\eta_k-1} \exp\left(-\frac{\eta_k}{\mu_k} x_n\right). \tag{7}$$

As a notational point, note that $f(x_n|\phi_k)$ is equivalent to $x_n \sim f(\phi_k)$. To complete the Bayesian specification of the mixture model, we follow [28] and assume that each mixture component's shape parameter η_k is exponentially distributed while the mean μ_k is inverse-gamma distributed. These choices are to allow efficient learning of the posterior distributions on μ_k and η_k and are discussed in detail in the following section. The full hierarchical form of DP-GaMM is given as

$$\begin{aligned} x_n &\sim \text{Gamma}(\eta_k, \eta_k/\mu_k), \\ z_n &\sim \text{Mult}(\pi), \\ \pi_k &= V_k \prod_{l < k} (1 - V_l), \\ V_k &\sim \text{Beta}(1, \alpha), \\ \eta_k &\sim \text{Exp}(a), \\ \mu_k &\sim \text{InvGamma}(b, c), \end{aligned} \tag{8}$$

where $\text{Exp}(\cdot)$ is the exponential distribution and $\text{InvGamma}(\cdot, \cdot)$ is the inverse-gamma distribution. A random variable that is inverse-gamma distributed is one whose reciprocal is gamma distributed. In this model, the base measure G_0 is defined as

$$\begin{aligned} G_0 = p(\phi_k) &= p(\eta_k, \mu_k|a, b, c) = p(\eta_k|a)p(\mu_k|b, c) \\ &= \text{Exp}(\eta_k; a)\text{InvGamma}(\mu_k; b, c), \end{aligned} \tag{9}$$

thus, the base measure G_0 is actually a product of two distributions $\text{Exp}(\eta_k; a)\text{InvGamma}(\mu_k; b, c)$.

This construction casts the estimation of the PDF of the x_n in a fully Bayesian context: only the x_n are observed, the parameters $\{z_n\}_{n=1:N}$ and $\{\gamma_k, V_k, \eta_k, \mu_k\}_{k=1:K}$ are all treated as latent random variables whose posterior distributions (and not simply an optimal value) are estimated during parameter inference. By estimating full posterior distributions on the model parameters, the Bayesian approach naturally quantifies uncertainty about the model parameters.

4. PARAMETER INFERENCE

Bayesian inference proceeds following Bayes' rule. In general, if X are observations distributed as $p(X|A)$, called the likelihood, and $p(A)$ is a prior distribution on parameters A , then the posterior distribution of the parameters of interest $p(A|X)$ is

$$p(A|X) = \frac{p(X|A)p(A)}{\int p(X|A')p(A')dA'}, \tag{10}$$

where the denominator $\int p(X|A')p(A')dA'$ is the normalizing constant. Computation of this integral is intractable for most cases. A variety of methods, often based on rejection methods, have been developed to draw samples from $p(A|X)$ (for a good review of such methods, see [38]). However, if $p(X|A)$ and $p(A)$ are selected to be conjugate in nature, then the posterior $p(A|X)$ will have the same form as the prior $p(A)$ but with updated parameters, allowing analytic iterative updating of the model parameters. We adopt this conjugacy in most cases of our parameter inference. In the step where conjugacy cannot be achieved in our model, we adopt a form of rejection sampling called Metropolis–Hastings sampling [39,40].

In fully conjugate situations, inference can be performed via approximate variational methods [32] or via the Gibbs sampler [33]. Variational methods rely on an approximate factorization of the model parameters and have the advantage of high convergence rate owing to it being a stochastic gradient ascent algorithm that does not rely on sampling. Gibbs sampling is a stochastic random walk algorithm based on Markov–Chain Monte Carlo and is (at the limit as the number of steps approaches infinity) a representation of the true posterior. We

adopt the Gibbs sampler as it is more commonly used in the statistical community. The full model likelihood of DP-GaMM is

$$\begin{aligned} & \prod_{n=1}^N p(x_n | z_n, \{\gamma_k, V_k, \eta_k, \mu_k\}_{k=1:K}) \\ &= \prod_{n=1}^N \prod_{k=1}^K \{p(x_n | z_n, \eta_k, \mu_k) p(z_n | V_k) p(V_k | \alpha) \dots \\ & \quad \times p(\eta_k | a) p(\mu_k | b, c)\}. \end{aligned} \quad (11)$$

Inference is performed sequentially by cycling through posterior updates and sampling of the latent parameters in order. We began every run of the DP-GaMM by randomly sampling μ_k from $\text{InvGamma}(\mu_k; 10^{-3}, 10^{-3})$ and η_k from $\text{Exp}(\eta_k; 1)$. Various initializations were tried, and we found that the model is relatively robust to the choice of initialization settings. The mixing weights π were set to be uniform $\pi_k = 1/K$ for all k . Next, we iteratively stepped through the following steps:

- For every observation x_n , $n = 1, \dots, N$, sample the cluster indicator z_n as

$$z_n \sim \text{Mult}(\nu_1, \nu_2, \dots, \nu_K), \quad (12)$$

where

$$\nu_k = \frac{\pi_k \text{Gamma}(x_n; \eta_k, \eta_k / \mu_k)}{\sum_{k'=1}^K \pi_{k'} \text{Gamma}(x_n; \eta_{k'}, \eta_{k'} / \mu_{k'})}. \quad (13)$$

- Sample the V_k from its posterior beta distribution, $k = 1, \dots, K$

$$V_k \sim \text{Beta}\left(1 + N_k, \alpha + \sum_{k'=k+1}^K N_{k'}\right), \quad (14)$$

where $N_k = \sum_{n=1}^N \mathbf{1}(z_n = k)$ is the number of data points associated with mixture component k and $\mathbf{1}(\cdot)$ denotes an indicator that is 1 if the argument is true and 0 otherwise. The mixing weights π are then fully defined as in Eq. (8). If the concentration parameter is treated as a random variable, then replace α with α_k in the preceding posterior distribution.

- For each mixture component k , $k = 1, \dots, K$, sample the mean μ_k from its posterior inverse-gamma distribution as

$$\mu_k \sim \text{InvGamma}\left(b + \sum_{n=1}^N \mathbf{1}(z_n = k), c + \sum_{n=1}^N \mathbf{1}(z_n = k)x_n\right). \quad (15)$$

- The posterior of the shape parameter η_k can be shown to be

$$\begin{aligned} & p(\eta_k | \{x_n, z_n\}_{n=1:N}, \{\mu_k\}_{k=1:K}) \\ & \propto \frac{\eta_k^{\sum_{n=1}^N \mathbf{1}(z_n=k)\eta_k}}{\Gamma(\eta_k)^{\sum_{n=1}^N \mathbf{1}(z_n=k)}} \exp\left(-\eta_k \left(a + \frac{\sum_{n=1}^N \mathbf{1}(z_n=k)x_n}{\mu_k}\right)\right) \\ & \quad + \eta_k \log \mu_k - \log \prod_{n=1}^N \mathbf{1}(z_n = k)x_n \Bigg), \end{aligned} \quad (16)$$

however, since this distribution does not fall into any known form, sampling from it must be done using Metropolis–Hastings sampling [39,40]. Following the approach of [28], we

note that for large η_k , Eq. (16) is similar to a gamma density and therefore adopt a gamma distribution as the proposal distribution to generate candidate values. In this case, the proposal is generated as $\tilde{\eta}_k \sim \text{Gamma}(r, r/\eta_k)$, where η_k is the current value of the shape parameter, and r is a tuning parameter that is set *a priori* by the modeler. It should be noted that the choice of r does not directly affect the posterior distribution of η_k , since it only controls the statistics of the proposals $\tilde{\eta}_k$. However, the choice can potentially have a significant impact on the acceptance rate and degree to which the model mixes. In our experiments, we found a setting of $r = 10$ to work well and offer consistently high acceptance rates (above 0.5). Note that the choice of r is a choice made by the modeler through empirical testing.

The proposed value of η_k is then accepted with probability

$$\min\left\{1, \frac{p(\tilde{\eta}_k | \{x_n, z_n\}_{n=1:N}, \{\mu_k\}_{k=1:K}) p(\tilde{\eta}_k | r, \eta_k)}{p(\eta_k | \{x_n, z_n\}_{n=1:N}, \{\mu_k\}_{k=1:K}) p(\eta_k | r, \tilde{\eta}_k)}\right\}, \quad (17)$$

where $p(\tilde{\eta}_k | r, \eta_k) = \text{Gamma}(\tilde{\eta}_k; r, r/\eta_k)$ and $p(\eta_k | r, \tilde{\eta}_k) = \text{Gamma}(\eta_k; r, r/\tilde{\eta}_k)$.

- If the concentration parameter α_k is treated as a random variable, then the posterior is

$$\alpha_k \sim \text{Gamma}(d + 1, e - \log(1 - V_k)), \quad (18)$$

where $\log(\cdot)$ denotes the natural log. Note that this step is optional. If greater control over the model's complexity is desired by the modeler, α_k can be set to a chosen α determined by the modeler.

The above steps are repeated iteratively until the model settles into a steady state which can be evaluated by computing the model likelihood defined in Eq. (11) and observed when the value stabilizes. It is common practice to discard the first several thousand samples as a “burn-in” phase to allow the model to find a suitable stationary distribution on the set of all G . The subsequent samples are recorded and are referred to as “collection” samples. Collection samples are then assumed to be samples drawn from the true posterior distribution of the model. For the results in this paper, we used 5000 iterations as burn-in and 5000 iterations as collection.

5. COMPARISON RESULTS AND APPLICATION TO MARITIME DATA

A. Comparison Results

In this section, we describe the results of comparing—both qualitatively and quantitatively via the log-likelihood metric—the performance of the DP-GaMM against the gamma-gamma model of [3] (GG), the log-normal distribution of [7] (LN), and the gamma distribution considered in [10] (G). All fitting was done as described in [9]. The approach proposed by [13] cannot guarantee non-negativity in the domain $[0, \infty)$, and so was not considered in this comparison because the log-likelihood metric is only valid for functions that are non-negative everywhere and integrate to 1.

The PDFs for the four methods are shown against a 50 bin normalized histogram in Figs. 4 (Run 1), 5 (Run 2), and 6 (Run 3). As before, the normalization is such that the histogram has area 1 under it. In all three runs, the DP-GaMM offers better models of the tails than the other three models. The log-normal distribution in particular seemed to consistently

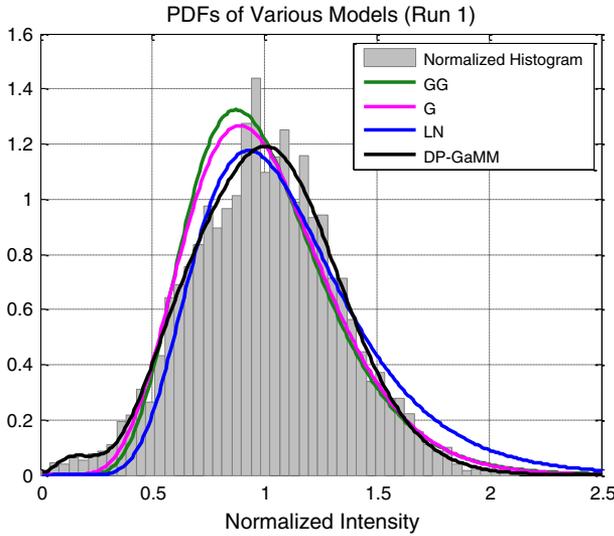


Fig. 4. PDFs from various models for Run 1. The normalized histogram in gray has area 1.

overestimate the right-hand (upper) tail of the data, while GG, gamma, and log-normal all seem to underestimate the left-hand (lower) tail that the DP-GaMM captures. Additionally, we note that the GG and gamma models behave very similarly for all three runs. It seems that in the high SI case (Run 3), the PDFs seem to agree more with one another than in Runs 1 and 2 (lower SI).

In order to quantitatively compare the performance of different estimators the three runs, we used sixfold held-out log-likelihood. Held-out log-likelihood is widely accepted in machine learning and statistics [23,30]. To do so, for each run we uniformly randomly partition each run (time series of 3636 intensity measurements) into one of six (approximately) equal portions called “folds.” For each round, we learned a PDF on five of the six folds and computed the log-likelihood on the remaining held-out fold (for each model considered). During each round, a different fold was considered as the testing set until each fold had been treated as the testing set

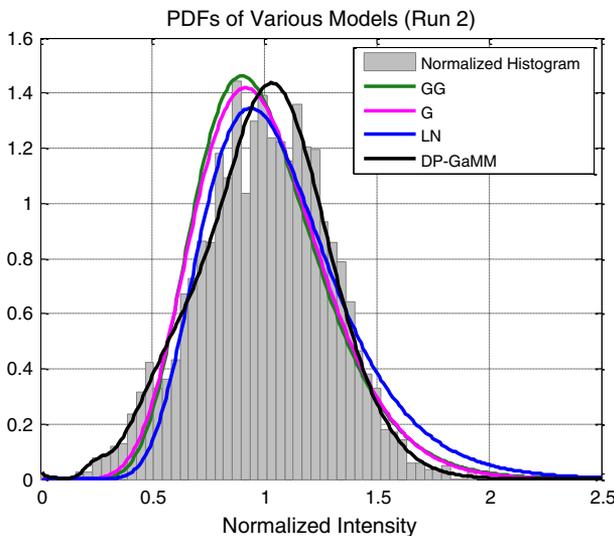


Fig. 5. PDFs from various models for Run 2. The normalized histogram in gray has area 1.

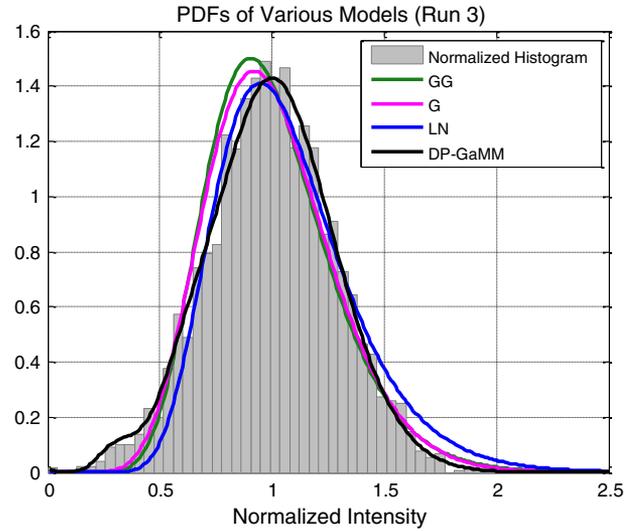


Fig. 6. PDFs from various models for Run 3. The normalized histogram in gray has area 1.

exactly once. We then averaged the held-out log-likelihood values across all six rounds together to report.

The log-likelihood measures the likelihood that the observed data was actually generated from a particular PDF. A higher (generally less-negative) held-out log-likelihood score means better generalization to previously unseen (but statistically identical) data. The held-out log-likelihood H is computed as

$$H = \sum_{m=1}^M \log p(x_m^{(\text{held-out})} | \Omega), \quad (19)$$

where Ω is the set of parameters needed to define a PDF, and m is used to index the set of held-out data. In the case of the DPMM, $\Omega = \{\pi, \Phi\}$. For the gamma distribution, Ω is the shape and scale parameters defined in [9]. For the log-normal distribution, Ω contains the mean and variance of the log-intensity. Finally, for GG, Ω is the scale and rate parameters denoted in [3].

Quantitatively, the held-out log-likelihoods of the four models across the three datasets is shown below in Table 1. A PDF with highest (least negative) log-likelihood is considered a better fit for the held-out data than the others. As mentioned previously, held-out log-likelihood is a widely used metric in machine learning and statistics for the quantitative comparison of different models. A model that has a high held-out log-likelihood value is said to generalize well to unseen data and is therefore regarded as the one best suited to make predictions on such unseen data in the future. Log-likelihood values are relative and dimensionless. They only have meaning when the

Table 1. Log-Likelihood Comparison of Various Models on the Three Runs^a

Run	GG	G	LN	DP-GaMM
1	-252.5185	-315.1688	-337.5740	-184.7727
2	-186.2339	-250.6055	-215.1999	-109.8983
3	-166.6684	-230.8744	-201.3958	-96.204

^aValues are relative (less negative is better) only within a single run and not across runs. Boldface indicates the best performance.

partitioned training and testing data is the same across all models under comparison.

In practice, H is computed as follows. Let Ω_t denote the model parameters at the t th iteration of an MCMC chain, then

$$H = \sum_{t=1}^T \sum_{m=1}^M \log p(x_m^{(\text{held-out})} | \Omega_t) / T \quad (20)$$

is called the mean held-out log-likelihood. This quantity is what we report for the DP-GaMM, as it considers the held-out log-likelihood computed at many samples and therefore characterizes how well the estimated random model generalizes to unseen data. We do not compute this mean held-out log-likelihood for the other models in this section as their parameters are point-estimates.

Table 1 shows that the DP-GaMM outperforms all three other methods in all runs considered, with the GG model coming in second. We note that the performance improvement of the GG model over the gamma and log-normal models is most likely due to the additional flexibility imposed by the second gamma distribution. The DP-GaMM offers even more flexibility than the other models considered and is therefore able to best represent the data. It is important to mention that an overly flexible model is also undesirable due to its tendency to overfit the training data. Such a situation also reduces the model's ability to characterize and generalize to unseen but statistically similar data. The DP-GaMM is, in a sense, balancing the opposing forces of fitting the training data well but smoothing the PDF enough that it generalizes well to unseen data [23].

B. Maritime USNA Data

We also apply DP-GaMM to maritime data collected at the USNA. In this experiment, a 632.8 nm He-Ne laser was used over a 314 m laser beam propagation path over water. The beam was vertically polarized, passed through a beam expander, and reflected off a 7.68 mm \times 7.68 mm spatial light modulator (SLM). The SLM was used to collect data for future

work and allows for spatial phase modulation of the beam. For the data considered in this paper, the SLM was set to produce a fully spatially coherent beam. An amplified photodetector and data acquisition device were used to collect data at 1×10^4 Hz. The data considered was collected over approximately two minutes. A scintillometer was used to estimate the value of C_n^2 over the propagation path for the test and recorded values on the order of $10^{-14} \text{ m}^{-2/3}$. The DP-GaMM is applied to the raw voltage readings from the sensor without normalization or preprocessing. We do this to demonstrate the flexible nonparametric nature of the DP-GaMM.

In Fig. 7, we show the PDF (in blue) of the iteration with the highest training set likelihood. In addition, we show (in red) 100 samples from the collection. Each sample is a PDF, and the set of all sample PDFs show the posterior distribution of the PDFs. Notice that the set of sample PDFs seem to suggest that there is greater uncertainty near the mode (peak) of the PDF and less uncertainty in the tails. Moreover, the sharp peaks near 1.15 V occur in only a small number of iterations, indicating that such PDFs have corresponding low probability of being the underlying generative model for the observed data.

It is important to note that while the sample with the highest training set likelihood is suitable for choosing a representative PDF to display, the advantage of the posterior distribution on the space of all possible PDFs can offer significant insight into the resultant model.

6. CONCLUSION AND FUTURE WORK

The DP gamma mixture model was presented as a new method for modeling the PDF of light scattered in turbulence. The DP-GaMM directly addresses the issues of parameter learning and selection through a hierarchical Bayesian construction. Gamma distributions were used for mixture components, and a hybrid Metropolis-Hastings/Gibbs sampling algorithm was developed and presented for parameter inference. The DP-GaMM was compared to several benchmark PDF estimation algorithms with favorable results. We also demonstrated the DP-GaMM on maritime data collected at the USNA without any normalization or preprocessing, yielding encouraging qualitative results.

Future work using the DP-GaMM will involve incorporating measurable environmental parameters via a kernel stick-breaking construction [41] or Bayesian density regression [42], as well as incorporating time evolution of the mixing weights [43] to better model changes in the atmosphere. The goal of this future research will be to generate PDFs that are functions of measurable environmental parameters. Additionally, we plan to take advantage of the DP-GaMM's ability to estimate high-dimensional PDFs in order to examine multiple pixels within each frame.

A related and equally important direction of research is in characterizing the quality of fit and physical underpinnings of the DP-GaMM PDFs in a rigorous manner. Tests such as the Kolmogorov-Smirnov test would complement the held-out log-likelihood test presented in this paper to give some measure of goodness-of-fit to the estimated PDFs. Additionally, such tests could potentially give insight into the physical mechanisms that cause discrepancies between learned PDFs and the empirical data, link DP-GaMM's mixture components to known physical phenomena and helping to provide

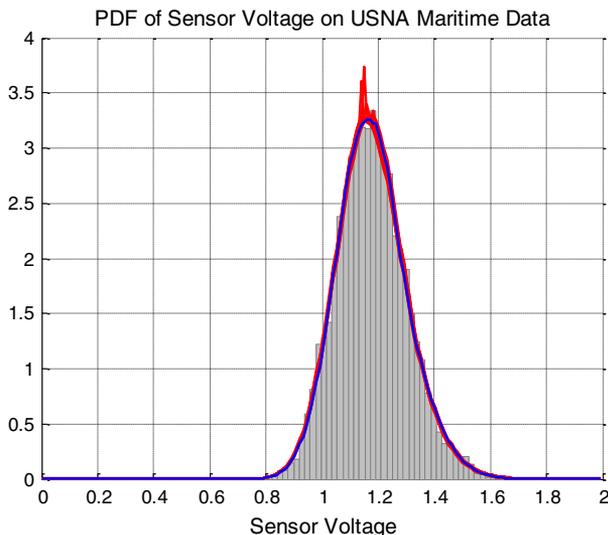


Fig. 7. PDF of raw voltage readings from sensors. The red PDFs are samples from the collection iterations, and the blue PDF is the sample with highest training set likelihood. The normalized histogram in gray has area 1.

statistical justification for existing or new families of physics based models.

REFERENCES

1. E. Jakeman and P. Pusey, "Significance of the k-distribution in scattering experiments," *Phys. Rev. Lett.* **40**, 546–550 (1978).
2. V. Gudimeta and J. Holmes, "Probability density function of the intensity for a laser-generated speckle field after propagation through the turbulent atmosphere," *J. Opt. Soc. Am.* **72**, 1213–1218 (1982).
3. R. Phillips and L. Andrews, "Universal statistical model for irradiance fluctuations in a turbulent medium," *J. Opt. Soc. Am.* **72**, 864–870 (1982).
4. L. Bissonette and P. Wizniowich, "Probability distribution of turbulent irradiance in a saturation regime," *Appl. Opt.* **18**, 1590–1599 (1979).
5. W. Strohbein, T. Wang, and J. Speck, "On the probability distribution of line-of-sight fluctuations for optical signals," *Radio Sci.* **10**, 59–70 (1975).
6. J. Churnside and R. Hill, "Probability density of irradiance scintillations for strong path-integrated refractive turbulence," *J. Opt. Soc. Am.* **4**, 727–733 (1987).
7. D. Mudge, A. Wedd, J. Craig, and J. Thomas, "Statistical measurements of irradiance fluctuations produced by a reflective membrane optical scintillator," *J. Opt. Laser Technol.* **28**, 381–387 (1996).
8. P. Orbanz and Y. Teh, "Bayesian nonparametric models," in *Encyclopedia of Machine Learning* (Springer, 2010), pp. 81–89.
9. J. McLaren, J. Thomas, J. Mackintosh, K. Mudge, K. Grant, B. Clare, and W. Cowley, "Comparison of probability density functions for analyzing irradiance statistics due to atmospheric turbulence," *J. Appl. Opt.* **51**, 5996–6002 (2012).
10. Y. Jiang, J. Ma, L. Tan, S. Yu, and W. Du, "Measurement of optical intensity fluctuation over an 11.8 km turbulent path," *Opt. Express* **16**, 6963–6973 (2008).
11. M. Al-Habash, L. Andrews, and R. Phillips, "Mathematical model for the irradiance probability density function of a laser beam propagating through turbulent media," *J. Opt. Eng.* **40**, 1554–1562 (2001).
12. R. Barakat, "Second-order statistics of integrated intensities and detected photons, the exact analysis," *J. Mod. Opt.* **43**, 1237–1252 (1996).
13. R. Barakat, "First-order intensity and log-intensity probability density functions of light scattered by the turbulent atmosphere in terms of lower-order moments," *J. Opt. Soc. Am. A* **16**, 2269–2274 (1999).
14. O. Korotkova, S. Avramov-Zamurovic, R. Malek-Madani, and C. Nelson, "Probability density function of the intensity of a laser beam propagating the maritime environment," *Opt. Express* **19**, 20322–20331 (2011).
15. M. Welling, "Robust series expansions for probability density estimation," Technical Note (Department of Electrical and Computer Engineering, California Institute of Technology, 2001).
16. B. Silverman, "Survey of existing methods," in *Density Estimates for Statistics and Data Analysis, Monographs on Statistics and Applied Probability* (Chapman & Hall, 1986), pp. 1–22.
17. C. Nelson, S. Avramov-Zamurovic, R. Malek-Madani, O. Korotkova, R. Sova, and F. Davidson, "Measurements and comparison of the probability density and covariance functions of laser beam intensity fluctuations in a hot-air turbulence emulator with the maritime atmospheric environment," *Proc. SPIE* **8517**, 851707 (2012).
18. M. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *J. Am. Stat. Assoc.* **90**, 577–588 (1995).
19. M. Rosenblatt, "Remarks on some nonparametric estimates of density function," *Ann. Math. Sci.* **27**, 832–837 (1956).
20. E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Sci.* **33**, 1065–1076 (1962).
21. T. Ferguson, "Bayesian density estimation by mixtures of normal distributions," *Recent Advances in Statistics* (Academic, 1983), pp. 287–302.
22. P. Orbanz and J. Buhmann, "Smooth image segmentation by nonparametric Bayesian inference," in *European Conference on Computer Vision* (Springer, 2006), Vol. 1, pp. 444–457.
23. Y. Teh, M. Jordan, M. Beal, and M. Jordan, "Hierarchical Dirichlet processes," *J. Am. Stat. Assoc.* **101**, 1566–1581 (2005).
24. E. Wang, D. Liu, J. Silva, and L. Carin, "Joint analysis of time-evolving binary matrices and associated documents," in *Advances in Neural Information Processing Systems* (Curran Associates, 2010), pp. 2370–2378.
25. J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica* **4**, 639–650 (1994).
26. A. Webb, "Gamma mixture models for target recognition," *Pattern Recogn.* **33**, 2045–2054 (2000).
27. K. Corsey and A. Webb, "Bayesian gamma mixture model approach to radar target recognition," *IEEE Trans. Aerosp. Electron. Syst.* **39**, 1201–1217 (2003).
28. M. Wiper, D. Insua, and F. Ruggeri, "Mixtures of gamma distributions with applications," *J. Comput. Graph. Stat.* **10**, 440–454 (2001).
29. R. Hill and J. Churnside, "Observational challenges of strong scintillations of irradiance," *J. Opt. Soc. Am. A* **5**, 445–447 (1988).
30. D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
31. J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypothesis," *Philos. Trans. R. Soc. London* **231**, 289–337 (1933).
32. M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. Thesis (University College London, 2003).
33. A. Gelfand and A. Smith, "Sampling-based approaches to calculating marginal densities," *J. Am. Stat. Assoc.* **85**, 398–409 (1990).
34. A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.* **39**, 1–38 (1977).
35. M. Aitkin and D. Rubin, "Estimation and hypothesis testing in finite mixture models," *J. R. Stat. Soc.* **47**, 67–75 (1985).
36. J. Marin, K. Mengersen, and C. Roberts, *Handbook of Statistics: Bayesian Thinking - Modeling and Computation* (Elsevier, 2011), Chap. 25.
37. H. Ishwaran and M. Zarepour, "Exact and approximate sum representations for the Dirichlet process," *Can. J. Stat.* **30**, 269–283 (2002).
38. C. Andrieu, N. de Freitas, A. Doucet, and M. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.* **50**, 5–43 (2003).
39. N. Metropolis, A. Rosenbluth, M. Rosenbluth, and A. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.* **21**, 1087–1092 (1953).
40. W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika* **57**, 97–109 (1970).
41. Q. An, C. Wang, I. Shterev, E. Wang, L. Carin, and D. Dunson, "Hierarchical kernel stick-breaking process for multi-task image analysis," in *International Conference on Machine Learning (ICML)* (Omnipress, 2008), pp. 17–24.
42. D. Dunson and N. Pillai, "Bayesian density regression," *J. R. Stat. Soc.* **69**, 163–183 (2007).
43. I. Pruteanu-Malcini, L. Ren, J. Paisley, E. Wang, and L. Carin, "Hierarchical Bayesian modeling of topics in time-stamped documents," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 996–1011 (2010).