Bayesian Estimation for CBRN Sensors with Non-Gaussian Likelihood

Yang Cheng, K. V. Umamaheswara Reddy, Tarunraj Singh, and Peter Scott

Abstract

Many sensors in chemical, biological, radiological, and nuclear (CBRN) applications only provide very coarse, integer outputs. For example, chemical detectors based on ion mobility sensing typically have a total of eight outputs in the form of bar readings. Non-Gaussian likelihood functions are involved in the modeling and data fusion of those sensors. Under the assumption that the prior distribution is a Gaussian density or can be approximated by a Gaussian density, two methods are presented for approximating the posterior mean and variance. The Gaussian sum method first approximates the non-Gaussian likelihood function by a mixture of Gaussian components and then uses the Kalman filter formulae to compute the posterior mean and variance. The Gaussian-Hermite method computes the posterior mean and variance through three integrals defined over infinite intervals and approximated by Gaussian-Hermite quadrature.

Index Terms

data fusion, CBRN sensor, non-Gaussian likelihood, Gaussian sum, Gaussian-Hermite quadrature.

I. Introduction

Data fusion is the process of integrating information from sensor observations and mathematical models to form a single composite picture of the environment. It plays an important role in a wide range of applications including multi-sensor multi-target tracking [1]. In recent years, there has been an increasing interest in applying data fusion methods to dispersion forecasting and source estimation of chemical, biological, radiological, and nuclear (CBRN) processes. It has been recognized that the effectiveness of chemical hazard assessment and management relies on the capability of forecasting the dispersion and transport of chemical materials in the atmosphere in near real time. Given the vast complexity and uncertainty of the dispersion and transport process, forecasting models such as SCIPUFF [2] used in hazard assessment and management are only capable of capturing some of the main characteristics of the ensemble behavior of the process. As a result, information contained in chemical sensor observations needs to be fused with the model forecasts to ameliorate the model forecast errors. Sensor data fusion is also a new frontier

Authors' Addresses: Y. Cheng is with the Department of Aerospace Engineering, Mississippi State University, Mississippi State, MS 39762, E-mail: (cheng@ae.msstate.edu); K. V. U. Reddy, and T. Singh are with the Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY 14260, E-mail: ({venkatar, tsingh}@buffalo.edu); P. Scott is with the Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY 14260. E-mail: (peter@buffalo.edu).

of radiation detection. Recent research in this area includes detecting and estimating radioactive sources (locations and strengths) using distributed sensor networks ([3]–[5]).

One of the challenges in CBRN sensor data fusion comes from the fact that many CBRN sensors output integer measurements only, for example, in the form of bar readings or particle counts. This in general means coarse determination of the material concentration or radiation intensity. Certain very coarse sensors are binary, only capable of detecting whether the concentration exceeds a certain threshold or not. The chemical sensor (ion mobility sensor) model presented in [6], which is representative of many detectors using ion mobility sensing and other technologies, has eight integer outputs of the form of bar readings, quantizing the amounts of materials collected by the sensor. Any material concentration measurement (a continuous random variable) within an interval $[T_{\mathcal{I}}, T_{\mathcal{I}+1})$, where $T_{\mathcal{I}}$ and $T_{\mathcal{I}+1}$ are two consecutive concentration thresholds, is converted into the integer output $\mathcal{I}$. A standard radioactive source detector with discrete-valued outputs is the Geiger-Müller counter [7]. Its output represents the level of radiation and is a discrete random variable obeying the Poisson distribution. The likelihood functions of both classes of sensors are non-Gaussian.

This article is concerned with Bayesian estimation methods for non-Gaussian likelihood functions and Gaussian prior density functions. The main objective of this work is to compute the posterior mean and variance efficiently based on Bayes' rule. This work is related to estimation or data fusion of a dynamic system with coarse quantized measurements [8]–[14]. Because the CBRN sensor models usually involve randomness, quantization, and/or other nonlinearity, the likelihood models for the CBRN sensors can be more complex than those with the quantization effect only. An excellent survey of quantization is given in [15]. The monograph by Curray [8] is a classic reference on estimation with quantized measurements. Keenan and Lewis include a brief review of several important early works in the 1970s including Curray's [10]. Since the quantization effect is nonlinear, the estimation problem with quantized measurement is nonlinear and non-Gaussian. Exact solution to the problem, which involves computing the posterior density function, is intractable or impossible in most cases. Sviestins and Wigren derive an optimal recursive state estimator from the Fokker-Planck equation and Bayes' rule, but the results are only useful for special systems [11]. Most estimators of dynamical systems with quantized measurements are recursive, suboptimal, and based on the so-called Gaussian-fit approximation [8], which approximates the posterior probability density of the state vector given all the past measurements as a Gaussian density. Since a Gaussian density is determined by its mean and covariance, the inference of the posterior density reduces to the inference of the conditional mean and covariance, which are then computed using Kalman-like formulae. The Gaussian-fix approximation greatly reduces the complexity of the recursive algorithms. In addition to the conditional mean estimator [8], [9], the estimation problem with quantized measurements is also treated as the maximum likelihood or the conditional mode estimator [9]. Ruan *et al* present a particle filter (bootstrap filter) with straightforward insertion of the quantized measurement into the particle filter [14]. As an aside, the bootstrap filter only updates the weights of the particles but not the locations of the particles at the measurement update step. In contrast, particle filters such as the unscented particle filter [16] with higher sampling efficiency usually employ local data fusion techniques to drive the particles toward measurements or regions of high interest. Interesting, but only loosely related works on system identification and

target tracking with quantized measurements can be found in [17]–[20].

Two numerical solutions of the posterior mean and variance of a random variable are presented in this article. The Gaussian sum method approximates the given non-Gaussian likelihood function as a Gaussian sum and then computes the posterior mean and variance using the formulae of the Gaussian-sum Kalman filter. The Gaussian sum approximation of the non-Gaussian likelihood function is computationally expensive but can be done off-line. In the Gaussian-Hermite method, Gaussian-Hermite Quadrature, a numerical integration method for approximating integrals of the form $\int_{-\infty}^{\infty} g(x)\exp(-x^2)dx$ with $g(x)$ a general function (ideally a polynomial over $(-\infty,\infty)$), is used to approximate the three integrals defining the posterior mean and variance. Also presented in the article are two applications of the methods. The first application is about data fusion and involves measurements of a binary chemical detector, for which the likelihood function is strongly non-Gaussian. The second application is about radiation source parameter estimation in which the methods are used for efficient search of the parameter space and involves a Geiger-Müller counter, whose observations obey the Poisson distribution.

The remainder of the article is organized as follows. First, the estimation problem and the three integrals over infinite intervals are defined. Second, a brief review of numerical integration methods related to the estimation problem is given. Then, the Gaussian sum method and the Gaussian-Hermite method are presented and compared with the method of truncation, an existing numerical integration method for infinite intervals. Next, a chemical sensor model and a gamma radiation sensor model are reviewed and finally two examples of the methods in CBRN applications are given.

II. Problem Description

A. Basic Problem

The basic problem studied in this article is Bayesian estimation of a random variable with non-Gaussian likelihood functions. The likelihood function is determined by the sensor characteristics and can be of any form. As with the extended Kalman filter and many other suboptimal estimators, The estimation method is based on the Gaussian-fit approximation. That is, the prior distribution of the random variable is assumed to be or can be approximated as a Gaussian density. The problem in which the prior density is a clipped Gaussian density will be considered as well. The clipped Gaussian density has been used to characterize the distribution of chemical concentrations [2].

The mathematical description of the problem is as follows. Let the continuous-valued random variable (e.g., the concentration of a chemical material), the (integer) measurement (e.g., the bar readings of a chemical sensor) on it, the prior distribution of the random variable (assumed to be Gaussian), and the likelihood function (a model determined by the sensor characteristics) be denoted by c, $\mathcal{I}$, $p(c)$, and $p(\mathcal{I}|c)$, respectively. With the measurement $\mathcal{I}$ given, $p(\mathcal{I}|c)$ is a function of c. The mean and variance of the prior Gaussian distribution are denoted by $\hat{c}^-$ and P_c^-, respectively. The posterior density $p(c|\mathcal{I})$ is then determined by Bayes' rule:

$$p(c|\mathcal{I}) = \frac{p(\mathcal{I}|c)p(c)}{p(\mathcal{I})} \tag{1}$$

where

$$p(\mathcal{I}) \equiv S_0 = \int_{-\infty}^{\infty} p(\mathcal{I}|c)p(c)dc \tag{2}$$

Because estimating the non-Gaussian posterior density is computationally expensive or intractable, our objective is limited to the estimation of the posterior mean and variance accurately and efficiently. The posterior mean $\hat{c}^+$ and variance P_c^+ are defined in terms of $p(c|\mathcal{I})$:

$$\hat{c}^+ = \int_{-\infty}^{\infty} cp(c|\mathcal{I})dc \tag{3}$$

$$P_c^+ = \int_{-\infty}^{\infty} c^2 p(c|\mathcal{I})dc - (\hat{c}^+)^2 \tag{4}$$

Define

$$S_1 \equiv \int_{-\infty}^{\infty} cp(\mathcal{I}|c)p(c)dc \tag{5a}$$

$$S_2 \equiv \int_{-\infty}^{\infty} c^2 p(\mathcal{I}|c)p(c)dc \tag{5b}$$

Then, the posterior mean and variance can be rewritten as

$$\hat{c}^+ = \frac{S_1}{S_0} \tag{6}$$

$$P_c^+ = \frac{S_2}{S_0} - (\hat{c}^+)^2 = \frac{S_0 S_2 - S_1^2}{S_0^2} \tag{7}$$

The basic estimation problem is to determine $\hat{c}^+$ and P_c^+ from $\hat{c}^-$, P_c^-, and $\mathcal{I}$ (through S_0, S_1, and S_2).

B. Clipped Gaussian Density as Prior

The case in which c can only be nonnegative is considered now. An example of that is the concentration of a chemical material at the location of a sensor. Because the concentration c of the material is zero or positive, $p(c)$ as well as $p(\mathcal{I}|c)$ is only defined in $[0, \infty)$. In SCIPUFF, $p(c)$ is approximated as a clipped normal distribution [2]. More details of the distribution model and the comparisons between this model and the log-normal model and other models can be found in [21], [22]. If $\hat{c}$ and P_c are the mean and variance of a Gaussian distribution, the density function of the clipped Gaussian distribution derived from it is given by [2]

$$\begin{aligned} p_{\text{clipped}}(c|\hat{c}, P_c) = \frac{1}{2}\left[1 - \text{erf}(\frac{\hat{c}}{\sqrt{2P_c}})\right]\delta(c) \\ + \frac{1}{\sqrt{2\pi P_c}} \exp\left[-\frac{(c-\hat{c})^2}{2P_c}\right], \quad c \geq 0 \end{aligned} \tag{8}$$

where $\delta(c)$ is Dirac's δ function. The corresponding cumulative distribution function $F(c|, \hat{c}, P_c)$ is

$$F_{\text{clipped}}(c|, \hat{c}, P_c) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{c-\hat{c}}{\sqrt{2P_c}}\right)\right], \quad c \geq 0 \tag{9}$$

The cumulative distribution function $F(c|, \hat{c}, P_c)$ is right continuous for $c \geq 0$. A plot of $F(c|, \hat{c}, P_c)$ with $P_c = 1$ and $\hat{c} = 0, 1, 2, 3$ is shown in Figure. 1.

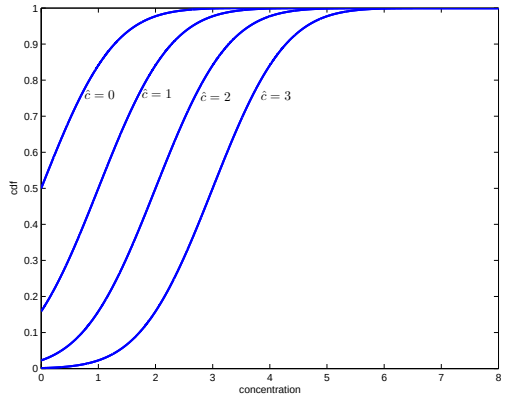

Figure 1. Cumulative Distribution Function

The mean $\hat{c}^*$ and variance P_c^* of the clipped Gaussian distribution are [2]

$$\hat{c}^* = \sqrt{\frac{P_c}{2\pi}} \exp\left[-\frac{\hat{c}^2}{2P_c}\right] + \frac{\hat{c}}{2}\left[1 + \mathrm{erf}\left(\frac{\hat{c}}{\sqrt{2P_c}}\right)\right] \tag{10}$$

and

$$P_c^* = -\hat{c}^2 + \frac{P_c}{2}\left[1 + \mathrm{erf}\left(\frac{\hat{c}}{\sqrt{2P_c}}\right)\right] + \hat{c}^* \cdot \hat{c} \tag{11}$$

where erf is the error function defined in a later section. SCIPUFF assumes that the concentration distribution is clipped Gaussian but outputs $\hat{c}$ and P_c instead of $\hat{c}^*$ and P_c^*. The difference between $\hat{c}$ and $\hat{c}^*$ and that between P_c and P_c^* become negligible only when $\hat{c} \gg 0$. For $\hat{c} = 0$, $\hat{c}^* = \sqrt{\frac{P_c}{2\pi}}$ and $P_c^* = P_c/2$.

When the prior density is a clipped Gaussian distribution parameterized by $\hat{c}^-$ and P_c^-, $p(c|\mathcal{I})$ and $p(\mathcal{I})$ are computed as

$$p(c|\mathcal{I}) = \frac{p(\mathcal{I}|c)p_{\text{clipped}}(c)}{p(\mathcal{I})} \tag{12}$$

$$p(\mathcal{I}) = \int_0^\infty p(\mathcal{I}|c)p_{\text{clipped}}(c)dc \tag{13}$$

where $p_{\text{clipped}}(c)$ is defined by Eq. (8). Before going further, we rewrite $p_{\text{clipped}}(c)$ as

$$p_{\text{clipped}}(c) = p_0\delta(c) + p(c) \tag{14}$$

with

$$p_0 = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\hat{c}^-}{\sqrt{2P_c^-}}\right)\right] \tag{15}$$

$$p(c) = \frac{1}{\sqrt{2\pi P_c^-}} \exp\left[-\frac{(c-\hat{c}^-)^2}{2P_c^-}\right] \tag{16}$$

Since $p(\mathcal{I}|c) = 0$ for any $c < 0$, we are allowed to replace the lower limits of the above integrals by $-\infty$ without altering the values of the integrals. That is,

$$\begin{aligned} p(\mathcal{I}) &= \int_0^\infty p(\mathcal{I}|c)p_{\text{clipped}}(c)dc \\ &= \int_{-\infty}^\infty p(\mathcal{I}|c)p_{\text{clipped}}(c)dc \\ &= \int_{-\infty}^\infty p(\mathcal{I}|c)[p_0\delta(c) + p(c)]dc \\ &= p(\mathcal{I}|0)p_0 + \int_{-\infty}^\infty p(\mathcal{I}|c)p(c)dc \\ &= p(\mathcal{I}|0)p_0 + S_0 \end{aligned} \tag{17}$$

and

$$\int_0^\infty cp(\mathcal{I}|c)p_{\text{clipped}}(c)dc = \int_{-\infty}^\infty cp(\mathcal{I}|c)p(c)dc = S_1 \tag{18}$$

$$\int_0^\infty c^2p(\mathcal{I}|c)p_{\text{clipped}}(c)dc = \int_{-\infty}^\infty c^2p(\mathcal{I}|c)p(c)dc = S_2 \tag{19}$$

Here S_0, S_1, and S_2 are defined by Eqs. (2) and (5). So, when the prior distribution $p(c)$ is a clipped Gaussian distribution and $p(\mathcal{I}|c) = 0$ for $c < 0$, the posterior mean and variance are still computed through S_0, S_1, and S_2:

$$\hat{c}^+ = \frac{S_1}{S_0 + p(\mathcal{I}|0)p_0} \tag{20}$$

$$P_c^+ = \frac{S_2}{S_0 + p(\mathcal{I}|0)p_0} - (\hat{c}^+)^2 \tag{21}$$

III. Methods for Computing Posterior Mean and Variance

In this section, two methods are presented for computing the posterior mean $\hat{c}^+$ and variance P_c^+ under the assumption that the prior distribution $p(c)$ is Gaussian (not clipped Gaussian) and the likelihood function satisfies $p(\mathcal{I}|c) = 0$ for $c < 0$. The first method approximates the non-Gaussian likelihood function $p(\mathcal{I}|c)$ by a Gaussian sum, which can be accomplished off-line, and then computes the posterior mean and variance using the formulae of the Gaussian-sum Kalman filter. No explicit evaluation of S_0, S_1, and S_2 is needed for the method. The second method uses Gaussian-Hermite quadrature to compute the integrals S_0, S_1, and S_2 and then computes the posterior mean and variance from S_0, S_1, and S_2. Before presenting the two methods, we will briefly review the numerical integration methods for one-dimensional integrals over $(-\infty, \infty)$.

A. Numerical Integration Methods

A key problem of posterior mean and variance computation is the approximation of the following integral:

$$S_n = \int_{-\infty}^\infty [c^n p(\mathcal{I}|c)]p(c)dc, \ n = 0, 1, 2. \tag{22}$$

where $p(c)$ is a Gaussian density with mean $\hat{c}^-$ and variance P_c^-. The integral is equivalent to

$$S = \int_{-\infty}^\infty f(c)dc = \int_{-\infty}^\infty g(c)e^{-c^2}dc \tag{23}$$

Several methods exist for approximate integration over infinite intervals, including change of variable, truncation of the infinite interval, the Fourier transform, the Laplace transform, the Gaussian quadratures, and so on [23]. Gaussian-Hermite quadrature belongs to the class of Gaussian quadratures. It is chosen over other Gaussian quadratures such as Gaussian-Laguerre quadrature to approximate S_n mainly because Gaussian-Hermite quadrature is immediately applicable to the integral given by Eq. (23). Gaussian-Laguerre quadrature is most suited to the integral $\int_0^\infty f(c)\exp(-c)dc$. In this article, only the method of truncation of the infinite interval is compared with the proposed methods. The idea of truncation of the infinite interval is to reduce the infinite interval to a finite interval by ignoring the tail of the integrand [23]. After the truncation, such standard methods for finite intervals as the trapezoidal rule or Simpson's rule can be applied. The trapezoidal rule with $n+1$ points takes the form

$$\int_a^b f(c)dc \approx \frac{h}{2}\left[f(c_0) + 2f(c_1) + \cdots + 2f(c_{n-1}) + f(c_n)\right] \tag{24}$$

The Simpson's rule with $(n+1)$ points (n is even) takes the form

$$\int_a^b f(c)dc \approx \frac{h}{3}\left[f(c_0) + 4f(c_1) + 2f(c_2) + 4f(c_3) + 2f(c_4) + \cdots + 4f(c_{n-1}) + f(c_n)\right] \tag{25}$$

In the above equations, $h = (b-a)/n$ and $c_k = a + kh$, $k = 0, \ldots, n$. In [23], it is proved that under certain circumstances, $h\sum_{k=-\infty}^{\infty} f(kh + c_0)$ with $-\infty < c_0 < \infty$ converges to the integral $\int_{-\infty}^{\infty} f(c)dc$ as h goes to 0.

B. Gaussian Sum Method

The Gaussian sum method begins with a Gaussian sum approximation of the likelihood function as a function of c, given by

$$P(\mathcal{I}|c) \approx \sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)}) \tag{26}$$

Any likelihood function $P(\mathcal{I}|c)$ satisfying $P(\mathcal{I}|c) \geq 0$ for all c and $-\infty < \int_0^\infty P(\mathcal{I}|c)dc < \infty$ and with at most a finite number of discontinuities can be approximated by a Gaussian sum [24]. In Eq. (26), the likelihood function $P(\mathcal{I}|c)$ is a function of c and $\mathcal{I}$ is the known measurement, α_l is the weight of the l-th of the L Gaussian components and $\mathcal{N}(c; c^{(l)}, R^{(l)})$ denotes a Gaussian density function of c with mean $c^{(l)}$ and variance $R^{(l)} > 0$:

$$\mathcal{N}(c; c^{(l)}, R^{(l)}) = \frac{1}{\sqrt{2\pi R_c^{(l)}}} \exp\left[-\frac{(c - c^{(l)})^2}{2R_c^{(l)}}\right] \tag{27}$$

The weights α_l may be negative as long as the Gaussian sum is nonnegative for all c.

The number L of components of the Gaussian sum depends on the similarity of the likelihood function to the Gaussian components, the length of the effective support of the likelihood function, and the accuracy and complexity requirements for the approximation. Because the Gaussian function is smooth, exponentially decaying, and localized, the Gaussian sum method is most suited to smooth localized likelihood functions. The number of the Gaussian components depend on the value of $R^{(l)}$, too. Large $R^{(l)}$ corresponds to large spread of the Gaussian component. If all $R^{(l)}$ are small compared to the effective support of the likelihood function, the number L of the Gaussian components for a good approximation of the likelihood function can only be large. For sake of simplicity, we

choose not to optimize L and use $L = 8$ in all examples in the article. An eight-component Gaussian sum have 24 free parameters (the variances $R^{(l)}$ are not necessarily identical) and can approximate a variety of functions to satisfactory accuracy and resolution.

Given the number L of Gaussian components, the Gaussian sum parameters α_l, $c^{(l)}$, and $R^{(l)}$ are obtained by solving a minimization problem of which the cost function is

$$\int_0^\infty \left(P(\mathcal{I}|c) - \sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)}) \right)^2 dc \tag{28}$$

The more general form of the cost function is [24]

$$\int_0^\infty \left| P(\mathcal{I}|c) - \sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)}) \right|^k dc \tag{29}$$

where $k \geq 1$. We will only consider the case of $k = 2$, however. Numerical integration methods (for example, the MATLAB function quadgk) over infinite intervals are used to evaluate the cost. The minimization problem is solved using the MATLAB function fmincon, which allows for inclusion of equality and inequality constraints and parameter bounds in the minimization problem. The constraints included are

$$\sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)}) \geq 0 \tag{30}$$

$$\int_0^\infty c^m \left(\sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)}) - P(\mathcal{I}|c) \right) dc = 0, m = 0, 1, 2. \tag{31}$$

The validity of the constraints requires that the integral is finite. If the likelihood function were a probability density function, $\int_0^\infty c^m P(\mathcal{I}|c) dc$ have the meaning of moments.

If the cost function is approximated by a finite sum over a large but finite interval of c based on the trapezoidal rule, given by

$$\sum_k h_k \left(P(\mathcal{I}|c_k) - \sum_{l=1}^{L} \alpha_l \mathcal{N}(c_k; c^{(l)}, R^{(l)}) \right)^2 \tag{32}$$

where h_k the step size and the points c_k are not necessarily uniformly spaced, the minimization problem has the interpretation of curve/surface fitting, which is rich in theory and methods. Tools for curve/surface fitting can be used to solve the minimization problem. The parameters of the Gaussian sum can be learned by a radial basis function network [25], too.

The Gaussian sum approximation of the likelihood function is the most computationally expensive part of the method. However, because the likelihood function is determined by the sensor characteristics and independent of the prior distribution, the Gaussian sum approximation procedure does not need to be carried out online. The total number of the optimization problems that need to be solved off-line is equal to the number of the possible values of the measurement $\mathcal{I}$. For example, a chemical sensor with eight possible readings requires eight optimization problems to be solved.

With the likelihood function a Gaussian sum and the prior distribution a Gaussian density, the posterior distribution of c is a Gaussian sum, too. Furthermore, the posterior mean $\hat{c}^+$ and variance P_c^+ can be obtained analytically.

The results are summarized below:

$$\hat{c}^+ = \sum_{l=1}^{L} \beta_l \hat{c}^{(l)+} \tag{33}$$

$$P_c^+ = \sum_{l=1}^{L} \beta_l \left[P_c^{(l)+} + (\hat{c}^{(l)+})^2 \right] - (\hat{c}^+)^2 \tag{34}$$

with

$$\hat{c}^{(l)+} = \frac{R^{(l)} \cdot \hat{c}^- + P_c^- \cdot c^{(l)}}{P_c^- + R^{(l)}} \tag{35}$$

$$P_c^{(l)+} = \frac{P_c^- \cdot R^{(l)}}{P_c^- + R^{(l)}} \tag{36}$$

$$\beta_l = \frac{\alpha_l \mathcal{N}\left(c^{(l)}; \hat{c}^-, P_c^- + R^{(l)}\right)}{\sum_{l=1}^{L} \alpha_l \mathcal{N}\left(c^{(l)}; \hat{c}^-, P_c^- + R^{(l)}\right)} \tag{37}$$

The closed-form expressions for S_0, S_1, and S_2 exist, too. They are

$$S_0 = \sum_{l=1}^{L} \alpha_l \mathcal{N}\left(c^{(l)}; \hat{c}^-, P_c^- + R^{(l)}\right) \tag{38}$$

$$S_1 = \sum_{l=1}^{L} \alpha_l \mathcal{N}\left(c^{(l)}; \hat{c}^-, P_c^- + R^{(l)}\right) \hat{c}^{(l)+} \tag{39}$$

$$S_2 = \sum_{l=1}^{L} \alpha_l \mathcal{N}\left(c^{(l)}; \hat{c}^-, P_c^- + R^{(l)}\right) \left[P^{(l)+} + \left(\hat{c}^{(l)+}\right)^2 \right] \tag{40}$$

Note that S_0, S_1, and S_2 are not needed in order to compute the posterior mean and variance.

The Gaussian sum approach is now summarized as follows:

1) Approximate the likelihood function $P(\mathcal{I}|c)$ as $\sum_{l=1}^{L} \alpha_l \mathcal{N}(c; c^{(l)}, R^{(l)})$. This is done off-line.
2) Compute β_l (Eq. (37)), $\hat{c}^{(l)+}$ (Eq. (35)), $P_c^{(l)+}$ (Eq. (36)).
3) Compute $\hat{c}^+$ (Eq. (33)) and P^+ (Eq. (34)).

The online computation for each update (Steps 2 and 3) involves approximately $4L$ multiplications, L divisions, and L evaluations of Gaussian functions.

C. Gaussian-Hermite Method

In the Gaussian-Hermite method, the integrals S_0, S_1, and S_2 are approximated using Gaussian-Hermite quadrature and the posterior mean and variance are computed using Eqs. (6) and (7). Gaussian-Hermite quadrature approximates $\int_{-\infty}^{\infty} f(c) \exp(-c^2) dc$ by $\sum_{l=1}^{L} \omega_l f(c_l)$. The points c_l are the L roots of the Hermite polynomial of degree L. The coefficients ω_l are the associated weights on the points. The points c_l are symmetric about 0 and ranges in $(-1.225, 1.225)$ when $L = 3$, $(-2.021, 2.021)$ when $L = 5$, $(-5.388, 5.388)$ when $L = 20$, and $(-13.407, 13.407)$ when $L = 100$. The weight decays as c_l moves away from 0. The quadrature is exact if $f(c)$ is a polynomial up to degree $(2L + 1)$ over $(-\infty, \infty)$. The Hermite polynomials are defined by [26]

$$H_L(x) = (-1)^n e^{x^2} \frac{d^L}{dx^L} e^{-x^2} \tag{41}$$

For example,

$$\begin{aligned} H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x \\ H_4(x) &= 14x^4 - 48x^2 + 12 \end{aligned} \tag{42}$$

The associated weights are given by [26]

$$\omega_l = \frac{2^{L-1} L! \sqrt{\pi}}{L^2 [H_{L-1}(x_l)]^2} \tag{43}$$

with $L!$ the factorial of L. For $L = 3$, the three points of c_l are 0 and $\pm\sqrt{3/2}$ and the associated weights are $2/3\sqrt{\pi}$ and $1/6\sqrt{\pi}$. With the three-point scheme, the expectation of $f(c)$ with respect to the Gaussian kernel $p_0(c)$ with mean $\hat{c}_0$ and variance P_{c0} is given by

$$\begin{aligned} &\int_{-\infty}^{\infty} f(c) p_0(c) dc \\ &= \frac{1}{\sqrt{2\pi P_{c0}}} \int_{-\infty}^{\infty} f(c) \exp\left[-\frac{(c - \hat{c}_0)^2}{2P_{c0}} \right] dc \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f\left(\hat{c}_0 + \sqrt{2P_{c0}} c \right) \exp\left[-c^2 \right] dc \\ &\approx \frac{1}{6} \left[4 f(\hat{c}_0) + f\left(\hat{c}_0 + \sqrt{3P_{c0}} \right) + f\left(\hat{c}_0 - \sqrt{3P_{c0}} \right) \right] \end{aligned} \tag{44}$$

Relations have been noted in [27], [28] between the Gaussian-Hermite quadrature and the Unscented Transformation [29] for approximating the expectation of a nonlinear function.

In the Gaussian-Hermite method, the integrals to be computed are

$$S_n = \int_{-\infty}^{\infty} [c^n p(\mathcal{I}|c)] p(c) dc = \int_{-\infty}^{\infty} \frac{c^n p(\mathcal{I}|c) p(c)}{p_0(c)} p_0(c) dc, \ n = 0, 1, 2. \tag{45}$$

where the prior distribution $p(c)$ is a Gaussian density with mean $\hat{c}^-$ and variance P_c^-. The straightforward application of Gaussian-Hermite quadrature with the points generated according to $p(c)$, that is, $p_0(c) = p(c)$, gives

$$S_n \approx \sum_{l=1}^{L} \frac{\omega_l}{\sqrt{\pi}} \cdot \left(\hat{c}^- + \sqrt{2P_c^-} c_l \right)^n \cdot p\left(\mathcal{I} | \hat{c}^- + \sqrt{2P_c^-} c_l \right), \ n = 0, 1, 2. \tag{46}$$

The Gaussian-Hermite method is now summarized as follows:

1) Determine the number L of nodes to be used and ω_l and c_l.
2) Compute S_0, S_1, and S_2 using Eq. (46).
3) Compute $\hat{c}^+$ and P_c^+ using Eqs. (6) and (7).

It involves approximately L likelihood function evaluations and $3L$ multiplications.

D. Examples

In this section, the prior density $p(c)$ is assumed to be Gaussian (not clipped Gaussian).

1) Uniform Likelihood: In this example, the likelihood function $p(\mathcal{I}|c)$ is a uniform distribution of c (up to a constant) over $[-1,1]$. It is shown in dashed line in Figure 2. The prior distribution $p(c)$ is a zero-mean ($\hat{c}^- = 0$) Gaussian distribution with variance $P_c^- = 1$. The centers of the prior distribution and the likelihood function coincide. Because the likelihood function has finite support and is flat over $[-1,1]$, the integrals S_n, $n = 0, 1, 2$, reduce to

$$S_n = \int_{-\infty}^{\infty} [c^n p(\mathcal{I}|c)] p(c) dc = \int_{-1}^{1} c^n p(c) dc = \frac{1}{\sqrt{2\pi P_c^-}} \int_{-1}^{1} c^n \exp\left[-\frac{(c-\hat{c}^-)^2}{2P_c^-}\right] dc \tag{47}$$

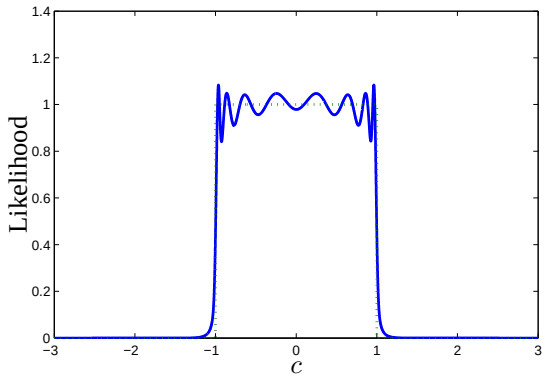

Figure 2. Uniform Likelihood Function over $[-1,1]$

This example is ideal for the method of truncation of the infinite interval because no approximation is introduced as the infinite interval $(-\infty, \infty)$ is replaced by $[-1,1]$ and the integrands involved can be well approximated in $[-1,1]$ by low-degree polynomials. The trapezoidal rule and Simpson's rule are used to compute the value of the integral over $[-1,1]$. Their results are summarized in Table I. It can be see from the table that Simpson's rule is more efficient for this example and the exact values of $\hat{c}^+$ and P_c^+ are 0.0000 and 0.2911, respectively.

Now an eight-component Gaussian sum is used to approximate the likelihood function. The results of the Gaussian sum approximation are given in Table II and plotted in solid line in Figure 2. Because the MATLAB function fmincon uses gradient-based search, there is no guarantee that the result is globally optimal. The Gaussian sum is approximately symmetric. The posterior mean and variance by the Gaussian sum method are 0.0000 and 0.2928, respectively. The error of the variance is about 0.6% of the exact value. The accuracy of the Gaussian sum method is low compared to the method of truncation because the likelihood function is dissimilar to the Gaussian function and the error in the approximation of the likelihood function cannot be made arbitrarily small with eight Gaussian components.

Additional simulation results indicate that while the Gaussian sum approximation is "independent of" the prior distribution, the accuracy of the posterior mean and variance does depend on the prior distribution. From Table II, the maximum spread (standard deviation) of the eight Gaussian components is about 0.26. When the standard

Table I

RESULTS OF TRAPEZOIDAL RULE AND SIMPSON'S RULE FOR UNIFORM LIKELIHOOD

	Trapezoidal Rule		Simpson's Rule	
n	$\hat{c}^+$	P_c^+	$\hat{c}^+$	P_c^+
3	0.0000	0.3775	0.0000	0.2327
5	0.0000	0.3108	0.0000	0.2899
11	0.0000	0.2942	0.0000	0.2911
21	0.0000	0.2919	0.0000	0.2911
31	0.0000	0.2915	0.0000	0.2911
51	0.0000	0.2912	0.0000	0.2911

Table II

GAUSSIAN SUM APPROXIMATION OF UNIFORM LIKELIHOOD

α_l	$c_v^{(l)}$	$R^{(l)}$
0.0377	-0.9695	0.0005
0.1027	-0.8835	0.0038
0.2343	-0.6979	0.0183
0.6312	-0.2987	0.0680
0.6240	0.3062	0.0670
0.2317	0.7011	0.0180
0.1006	0.8843	0.0037
0.0379	0.9696	0.0005

deviation of the prior distribution is much greater than 0.26, for example, several tens, the Gaussian sum method always yield good results. When the standard deviation of the prior distribution is close to or less than 0.26, the resolution of the eight-component Gaussian sum approximation may be too low and the accuracy of the computed mean and variance too poor. The problem becomes more severe when the overlap between the likelihood function and the prior distribution is small. For example, when $\hat{c}^- = 4$ and $P_c^- = 0.1$, the error with the eight-component Gaussian sum is about 100% for $\hat{c}^+$ and over 100% for P_c^+. Compare that when $\hat{c}^- = 4$ and $P_c^- = 1$, the error with the same Gaussian sum is less than 1% for $\hat{c}^+$ and less than 0.2% for P_c^+.

In the Gaussian-Hermite method, the points are generated according to a Gaussian distribution $p_0(c)$. The most straightforward choice for $p_0(c)$ is the prior distribution $p(c)$, which under our assumption is Gaussian. The results for $p_0(c) = p(c)$ are given in the columns under "$P_{c0} = 1$" in Table III. The accuracy of the Gaussian-Hermite method is the lowest of the three methods. The low accuracy of the method is due to the fact that most of the points are not in the intervals of interest. Take the 100-point scheme as an example. All but six points are between -1 and

Table III
RESULTS OF GAUSSIAN-HERMITE METHOD FOR UNIFORM LIKELIHOOD

	$P_{c0}=1$		$P_{c0}=0.06$	
L	$\hat{c}^+$	P_c^+	$\hat{c}^+$	P_c^+
3	0.0000	0.0000	0.0000	0.1209
5	0.0000	0.0000	0.0000	0.2385
10	0.0000	0.2352	0.0000	0.2991
20	0.0000	0.1204	0.0000	0.3209
30	0.0000	0.3535	0.0000	0.2894
100	0.0000	0.2574	0.0000	0.2934

1. Because the likelihood function equals zero outside the interval $[-1, 1]$, only these six points contribute to the computation of the posterior mean and variance. Higher accuracy is achieved, though not so good as that of the other methods, when $p_0(c)$ is a zero-mean Gaussian distribution with the variance $P_{c0} = 0.06$. The corresponding results are given under "$P_{c0} = 0.06$" in Table III.

2) Gaussian Likelihood: In this example, the prior distribution of c is a Gaussian distribution with mean -25 and variance 25 and the likelihood function is an exact Gaussian distribution with zero mean and variance 0.01. That is ideal for the Gaussian sum method because the error in the Gaussian sum representation of the likelihood function can be made arbitrarily small and there are no other approximation errors in the Gaussian sum method. The exact solution, computed using the formulae for the measurement update of the Kalman filter, is $\hat{c}^+ = -0.0100$ and $P_c^+ = 0.0100$. The result of the Gaussian sum method is identical to the exact solution.

To apply the method of truncation of the infinite interval, the finite interval over which the integrals are evaluated needs to be determined first. Noting that the 4-σ interval of the prior distribution and the likelihood function are $[-45, 5]$ and $[-0.4, 0.4]$, respectively, we choose $[-45, 0.4]$ as the finite interval over which the integrals are evaluated. The results of the trapezoidal rule and Simpson's rule are given in Table IV. Both integration rules need over two hundred evenly-spaced points to achieve the same accuracy as in the first example. Note that Simpson's rule is not more accurate than the trapezoidal rule in this example. The inefficiency of the method is due to the fact that the integration interval is large and the integrand over the interval is not similar to low-degree polynomials. If the interval were chosen as $[-0.4, 0.4]$, which encloses the significant interval of the posterior distribution, a Gaussian density with mean -0.0100 and variance 0.0100, the method would require only several points to achieve the same accuracy. However, the method of truncation usually does not have an "intelligent" way of determining the best finite interval to approximate the infinite interval without knowing the posterior mean and variance.

The results of the Gaussian-Hermite method are given in Table V. When the points are generated based on the prior distribution, the accuracy is extremely low. The "–" in the table stands for "not-a-number" in MATLAB. If the points are generated based on the likelihood function, with three points only, the method yields a result that is

Table IV
RESULTS OF TRAPEZOIDAL RULE AND SIMPSON'S RULE FOR GAUSSIAN LIKELIHOOD

	Trapezoidal Rule		Simpson's Rule	
n	$\hat{c}^+$	P_c^+	$\hat{c}^+$	P_c^+
3	0.4000	0.0000	0.4000	0.0000
5	0.4000	0.0000	0.4000	0.0000
11	0.4000	0.0000	0.4000	0.0000
21	0.4000	0.0000	0.4000	0.0000
31	0.4000	0.0000	0.4000	0.0000
51	0.3678	0.0282	0.3379	0.0526
101	-0.0539	0.0000	-0.0540	0.0000
201	-0.0212	0.0087	0.0012	0.0134
301	-0.0102	0.0100	-0.0223	0.0079
501	-0.0100	0.0100	-0.0106	0.0100
1001	-0.0100	0.0100	-0.0100	0.0100

Table V
RESULTS OF GAUSSIAN-HERMITE METHOD FOR GAUSSIAN LIKELIHOOD

	$c_0 = -25, P_{c0} = 25$		$c_0 = 0, P_{c0} = 0.01$	
L	$\hat{c}^+$	P_c^+	$\hat{c}^+$	P_c^+
3	–	–	-0.0100	0.0100
5	–	–	-0.0100	0.0100
10	-0.7027	0.0000	-0.0100	0.0100
20	-1.3271	0.0000	-0.0100	0.0100
30	0.0630	0.0000	-0.0100	0.0100
100	-0.4663	0.0000	-0.0100	0.0100

identical to the exact solution up to four decimal digits. These results highlight the importance of the selection of the Gaussian kernel in the Gaussian-Hermite method.

IV. MEASUREMENT UPDATE OF RANDOM VECTOR

So far, we have focused on methods for computing the posterior mean and variance of a random variable. Now we will consider the problem of updating the mean and covariance of a random vector $\mathbf{x}$ with the same kind of integer sensor observation $\mathcal{I}$. Here $\mathbf{x}$ may be a high-dimensional vector and $\mathcal{I}$ is a scalar. The two are connected by a continuous-valued random variable c. For example, $\mathbf{x}$ is the vector representation of the concentration levels of a

chemical material over a region of interest, c is the concentration at a sensor location, and $\mathcal{I}$ is the bar readings of the chemical sensor. Because the dimension of $\mathbf{x}$ may be high, in the order of hundreds or higher, the update method needs to be computationally efficient. A two-step procedure is proposed based on the methods for computing the posterior mean and variance in the previous section.

We assume that $\mathbf{x}$ and c are jointly Gaussian and that the corresponding mean and covariance are denoted by

$$\begin{bmatrix} \hat{\mathbf{x}}^- \\ \hat{c}^- \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_x^- & P_{xc}^- \\ (P_{xc}^-)^T & P_c^- \end{bmatrix}$$

respectively. The above quantities are computed from forward prediction models. The likelihood function $P(\mathcal{I}|\mathbf{x}) = P(\mathcal{I}|c)$ is non-Gaussian. Denote the posterior mean and covariance of the joint vector of $\mathbf{x}$ and c by

$$\begin{bmatrix} \hat{\mathbf{x}}^+ \\ \hat{c}^+ \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_x^+ & P_{xc}^+ \\ (P_{xc}^+)^T & P_c^+ \end{bmatrix}$$

Our objective is to compute $\hat{\mathbf{x}}^+$ and P_x^+.

The two-step procedure is as follows:

1) Update the mean and variance of c from $\hat{c}^-$ and P_c^- to $\hat{c}^+$ and P_c^+.
2) Update the mean and covariance of $\mathbf{x}$ from $\hat{\mathbf{x}}^-$ and P_x^- to $\hat{\mathbf{x}}^+$ and P_x^+ using the following formulae [30]:

$$\hat{\mathbf{x}}^+ = \hat{\mathbf{x}}^- + \mathcal{K}(\hat{c}^+ - \hat{c}^-) \tag{48}$$

$$P_x^+ = P_x^- + \mathcal{K}(P_c^+ - P_c^-)\mathcal{K}^T \tag{49}$$

with

$$\mathcal{K} = P_{xc}^-(P_c^-)^{-1} \tag{50}$$

The issue with non-Gaussian likelihood functions is isolated from the linear second step. The simple linear update method is used in the second step mainly because the dimension of $\mathbf{x}$ is high and higher-order methods are computationally expensive. The linear update method has the same form as the update of the estimator with quantized measurements in [8]. It is also identical to the measure update of the linear Kalman filter. The linear update method is optimal in the Bayesian sense when $\mathbf{x}$, c, and $\mathcal{I}$ are all jointly Gaussian and suboptimal in other cases.

V. Sensor Models

The chemical and Gamma radiation sensor models are described in [6] and [7] respectively and are briefly summarized in this section. These sensor models are used to determine $p(\mathcal{I}|c)$.

A. Chemical Sensor

The sensor model describes the relation between the concentration of a chemical material at the sensor location and the sensor readings. The output of the chemical sensor is a discrete number of bars. In [6], the number of bars ranges from zero to seven. These bar readings indicate the concentration magnitude at the sensor location at the

"instant"; the sensor displays $\mathcal{I} = 0, \ldots, 7$, bars when the internal continuous-valued concentration magnitude c_v is between thresholds $T_{\mathcal{I}}$ and $T_{\mathcal{I}+1}$, where $T_{\mathcal{I}+1} > T_{\mathcal{I}} \geq 0$ and $T_0 = 0$. The properties of the sensor are determined by the thresholds and the properties of c_v, which is assumed normally distributed about the true concentration c [6]. The measurement model is given by

$$c_v = c + v \tag{51}$$

where the noise v has mean zero and variance

$$R \triangleq R(c) = \alpha c + J \tag{52}$$

where α is the proportionality constant and J accounts for the thermal motion of the electrons in the components [6]. The noise v may be considered a combination of multiplicative noise and additive noise. The (unclipped) probability density function of c_v conditional on c is given by

$$p(c_v|c) = \frac{1}{\sqrt{2\pi R(c)}} \exp\left[-\frac{(c_v - c)^2}{2R(c)}\right] \tag{53}$$

Because the dependency of R on c, $p(c_v|c)$ is an asymmetric function of c (with c_v fixed). Following Eq. (53), the likelihood function of the concentration c given the output $\mathcal{I}$, or the probability of $\mathcal{I}$ conditioned on c, is determined by the following integrals

$$P(\mathcal{I}|c) = \int_{T_{\mathcal{I}}}^{T_{\mathcal{I}+1}} p(c_v|c)dc_v, \quad \mathcal{I} > 0 \tag{54}$$

$$P(0|c) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{c}{\sqrt{2R(c)}}\right)\right] + \int_0^{T_1} p(c_v|c)dc_v \tag{55}$$

It can be verified that $\sum_{\mathcal{I}} P(\mathcal{I}|c) = 1$. Note that the likelihood function is continuous in c. A plot of the likelihood function ($\alpha = 10^{-4}$, $J = 10^{-12}$, and $T_{\mathcal{I}}$ are logarithmically equally spaced in the range $[10^{-6}, 1]$) is given in Figure 3. Small bar readings correspond to large likelihood in the low concentration interval and small likelihood in the high concentration interval, while large bar readings correspond to small likelihood in the low concentration interval and large likelihood in the high concentration interval.

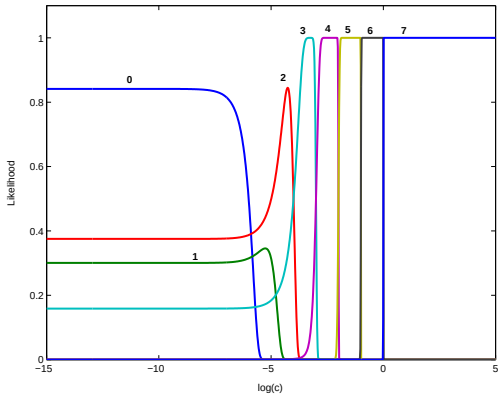

Figure 3. Likelihood Function of the Chemical Sensor

The uncertainty of the sensor output is due to both the quantization effect and the random noise v. When the number of thresholds goes to infinity and the length of the intervals between two thresholds goes to zero, the sensor output becomes continuous and the likelihood function reduces to the Gaussian distribution defined by Eq. (53). When the uncertainty due to the internal random variable vanishes, the likelihood function is flat over the interval $[T_{\mathcal{I}}, T_{\mathcal{I}+1})$ associated with $\mathcal{I}$ and zero elsewhere.

B. Gamma Radiation sensor

The radiation sensor measurement $\mathcal{I}$ in [7] is a Poisson random variable whose mean and variance are given by λ, which in turn is determined by the point radiation sources and the background radiation λ_b:

$$\lambda = \lambda_b + A \sum_{j=1}^{M} \frac{a_j}{d_j^2} \tag{56}$$

where a_j are the source intensities, d_j are the distances from the sources to the sensor, and A is a factor accounting for attenuation and other effects. Note that λ is a continuous-valued variable. The probability mass function of $\mathcal{I}$ is given by [7]

$$P(\mathcal{I}|\lambda) = \frac{e^{-\lambda}\lambda^{\mathcal{I}}}{\mathcal{I}!} \tag{57}$$

For large $\lambda > 20$, an approximation to the measurement model is $\mathcal{I} = \lambda + v$, where v is a zero-mean Gaussian random variable with variance λ [31]. That is to say, for large λ,

$$P(\mathcal{I}|\lambda) \approx \frac{1}{\sqrt{2\pi\lambda}} \exp\left[\frac{-(\mathcal{I}-\lambda)^2}{2\lambda}\right] \tag{58}$$

VI. Examples of CBRN Applications

A. Chemical Sensor Data Fusion

A fictitious binary sensor displays one bar when its internal measurement $c_v \geq 1$ and zero bar otherwise. The variance of c_v is $R = 0.09$. The objective of the data fusion problem is to combine the predicted concentration and the sensor observation. The prior $p(c)$ is a (clipped) Gaussian distribution with mean $\hat{c}^- = 1.5$ and variance $P_c^- = 4$. The reading of the sensor is one bar and the likelihood function over $[0, 20]$ is shown in Figure 4. Note that the integration of the likelihood function over $[0, \infty)$ is infinite. The posterior distribution, obtained using Bayes' rule and 500,001 uniformly-spaced points in $[0, 20]$, is shown in Figure 5. The posterior mean $\hat{c}^+$ and variance P^+ computed directly from the 500,001 points are 2.7806 and 1.7339, respectively. These values are used as the exact values.

1) Gaussian Sum: The Gaussian sum approximation to the likelihood function is obtained by minimizing the cost function defined over the internal $[0, 20]$ by Eq. (32). The weights, variances, and centers of the eight Gaussian components are given in Table VI. Note that some weights are negative but overall the Gaussian sum is nonnegative and provides a good approximation of the likelihood function over the interval $[0, 20]$. If the variances of the Gaussian components were small and identical, hundreds or thousands of Gaussian components would be needed to achieve good approximation of the likelihood function over $[0, 20]$. The approximation error is given in

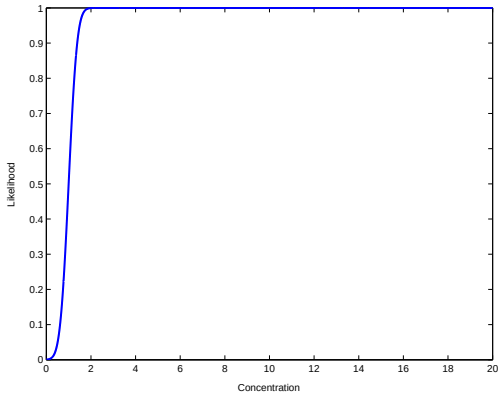

Figure 4. Likelihood Function

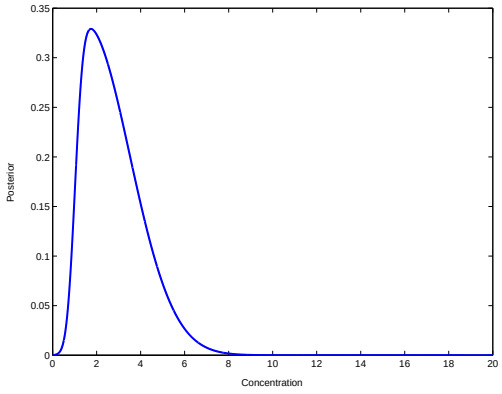

Figure 5. Posterior Density Function

Table VI

GAUSSIAN SUM APPROXIMATION PARAMETERS

	α_l	$c_v^{(l)}$	$R^{(l)}$
1	-1.0789	0.2530	3.4970
2	2.4323	0.3016	0.2509
3	1.4780	1.7946	1.8556
4	-5.4721	0.5602	0.3516
5	2.7054	0.9121	0.2982
6	3.8470	3.8091	8.1734
7	6.5080	9.0844	18.6352
8	16.9385	19.8336	48.9650

Figure 6. The maximum approximation error over the entire interval is 0.01. With the eight-component Gaussian sum approximation, the posterior mean and variance are 2.7804 and 1.7340, in good agreement with the true values. The Gaussian sum approximation of the likelihood function over $[0, 20]$ is more accurate than that of the uniform likelihood function in a previous section probably because the the former is continuous while the latter is discontinuous at ± 1. Note that the Gaussian sum is only good for $[0, 20]$ and is not a valid approximation of the likelihood function over $[0, \infty)$. Therefore, it should not be used with the prior distribution of which the effective support (e.g., $[-4\sqrt{P_c^-}, 4\sqrt{P_c^-}]$) are not within $[0, 20]$. On the other hand, when the prior distribution is such that the likelihood function over the effective support of it is flat, the posterior distribution is almost equal to the prior, and there is probably no need to apply the Gaussian sum method.

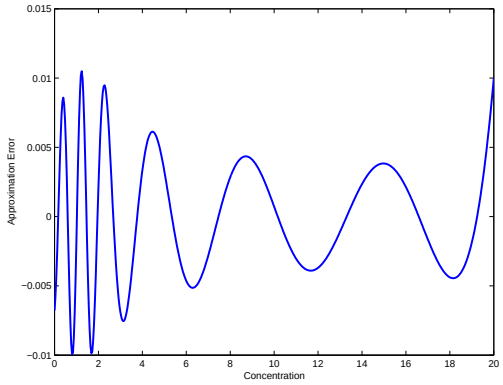

Figure 6. Approximation Error of the Likelihood Function

2) Gaussian-Hermite: The points are chosen based on the prior distribution and the likelihood function are evaluated at these points. The posterior means and variances with 3, 5, 10, 20, 30, and 100 points are given in Table VII. The results of the Gaussian-Hermite method contain large errors because the likelihood function has a sharp transition and few points are in the right interval where the sharp transition occurs. That is the same problem with the Gaussian-Hermite method identified in a previous section.

In addition to $R = 0.09$, the results for $R = 0.18$ and $R = 0.9$ are presented in the table. Larger R results in less sharp transition and higher accuracy in the computation of the posterior mean and variance. For $R = 0.09$, with 20 points the approximation error remains over 10%; for $R = 0.9$, just ten points are sufficient.

B. Source Estimation with Radiation Sensors

A two-dimensional source estimation example in [7] is considered. For the detail of the example see [7]. As shown in Figure 7, 60 radiation sensors are located in a circle of radius 200 centered at the origin. The sensor model is given by Eqs. (56) and (57). The sensor measurements are denoted by $\mathcal{I}_k$, $k = 1, \ldots, 60$. When the number of sources is unknown, source estimation involves a combinatorial problem that is difficult to solve. Because our purpose is to illustrate how the data fusion methods are used in a Monte-Carlo-based source parameter estimation

Table VII

POSTERIOR MEANS AND VARIANCES USING GAUSSIAN-HERMITE METHOD

	$R = 0.09$		$R = 0.18$		$R = 0.9$	
n	$\hat{\alpha}^+$	P^+	$\hat{\alpha}^+$	P^+	$\hat{\alpha}^+$	P^+
3	2.2204	1.9765	2.2659	2.0666	2.4097	2.3293
5	2.3991	1.8899	2.4479	1.9462	2.5837	2.1337
10	3.0615	1.4428	2.9412	1.6647	2.7031	2.1094
20	2.8527	1.7515	2.7941	1.8203	2.6924	2.0927
30	2.7869	1.7876	2.7687	1.8101	2.6924	2.0915
100	2.7793	1.7339	2.7697	1.7802	2.6924	2.0915
exact	2.7806	1.7340	2.7697	1.7803	2.6924	2.0915

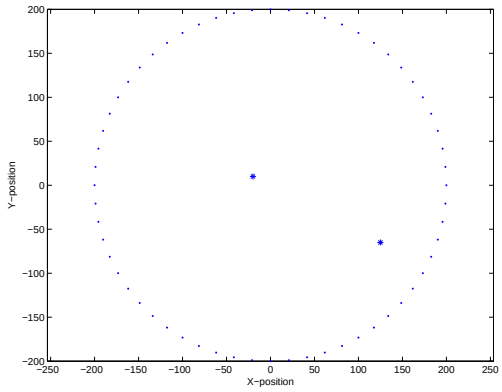

Figure 7. Source Estimation Scenario

method–the data fusion methods are used inside the Monte Carlo method to provide proposal distributions for generating new particles, we assume it is known that there are two identical static point sources in the horizontal plane. The unknown parameter vector $\mathbf{x}$ consists of the locations of the two sources and their common intensity. The true values of the parameter vector (consisting of the locations of the two sources and the intensity of them) are $(-20, 10)$, $(125, -65)$, and 360000. (The units are omitted.) In the Monte Carlo method, the parameter estimate is computed as $\hat{\mathbf{x}} = 1/N \sum_{i=1}^{N} \mathbf{x}^{(i)}$, where $N = 50$ and $\mathbf{x}^{(i)}$ denotes a set of N random samples (or particles) in the parameter space.

Of each particle, the initial locations of the two sources are randomly chosen in the circle, and the initial intensity is uniformly distributed in $[0, 400000]$. Because the initial particle set is unlikely to contain any particles that are close to the truth, iterations are needed to locally search the parameter space and gradually move the particles toward the true values. The search of the parameter space is based on the Metropolis-Hastings algorithm [32]. The proposed particle $\mathbf{x}^{\text{proposal}(i)}$ is drawn according to a proposal distribution $Q(\mathbf{x}^{\text{proposal}(i)}; \mathbf{x}^{(i)}, \{\mathcal{I}_k\}_{k=1}^{60})$. The

proposed particle is accepted or rejected based on the ratio of the likelihood functions of the proposed and present particles. The acceptance probability is computed as

$$\alpha = \min\left\{\frac{L(\mathbf{x}^{\text{proposal}(i)})\cdot Q(\mathbf{x}^{\text{proposal}(i)};\mathbf{x}^{(i)},\{\mathcal{I}_k\}_{k=1}^{60})}{L(\mathbf{x}^{(i)})\cdot Q(\mathbf{x}^{(i)};\mathbf{x}^{\text{proposal}(i)},\{\mathcal{I}_k\}_{k=1}^{60})},1\right\} \tag{59}$$

where $L(\mathbf{x}^{(i)}) = \prod_{k=1}^{60} P(\mathcal{I}_k|\lambda_k^{(i)})$ is the likelihood function of $\mathbf{x}^{(i)}$. The proposal distribution is assumed to be Gaussian. The mean and covariance of the proposal distribution are computed using the two-step data fusion method and then $\mathbf{x}^{\text{proposal}(i)}$ are sampled from the proposal distribution. The main steps of the Monte-Carlo parameter estimation algorithm are summarized below:

1) Let $P_x^{(i)} = Q_x$, where Q_x is a small constant covariance matrix, for example, $Q_x = \text{diag}[10^2, 10^2, 10^2, 10^2, 10^8]$.
2) For each measurement:
   a) Compute the intensity prediction $\hat{\lambda}_k^{(i)-}$ and $P_{\lambda_k}^{(i)-}$ corresponding to $\hat{\mathbf{x}}^{(i)}$ and $P_x^{(i)}$ using the extended Kalman filter. Also compute the cross covariance $P_{x\lambda_k}^{(i)-}$.
   b) Update the intensities with the sensor measurements $\mathcal{I}_k$ to $\hat{\lambda}_k^{(i)+}$ and $P_{\lambda_k}^{(i)+}$ using either the Gaussian sum method or the Gaussian-Hermite method.
   c) Update $\hat{\mathbf{x}}^{(i)}$ and $P_x^{(i)}$ based on Eqs. (48) to (50):

$$\begin{aligned}
\mathcal{K}_k^{(i)} &= P_{x\lambda_k}^{(i)-}\left(P_{\lambda_k}^{(i)-}\right)^{-1}\\
\hat{\mathbf{x}}^{(i)} &\leftarrow \hat{\mathbf{x}}^{(i)} + \mathcal{K}_k^{(i)}\left(\hat{\lambda}_k^{(i)+} - \hat{\lambda}_k^{(i)-}\right)\\
P_x^{(i)} &\leftarrow P_x^{(i)} + \mathcal{K}_k^{(i)}\left(P_{\lambda_k}^{(i)+} - P_{\lambda_k}^{(i)-}\right)\left(\mathcal{K}_k^{(i)}\right)^T
\end{aligned}$$

3) Draw $\mathbf{x}^{\text{proposal}(i)}$ randomly from the Gaussian distribution with mean $\hat{\mathbf{x}}^{(i)+}$ and covariance $P_x^{(i)+}$.
4) Accept $\mathbf{x}^{\text{proposal}(i)}$, that is, $\mathbf{x}^{(i)} \leftarrow \mathbf{x}^{\text{proposal}(i)}$, according to the acceptance probability α in Eq. (59).

Fifty Monte Carlo runs of the data fusion method using the three-point Gaussian-Hermite method are carried out. The parameters are iteratively updated 20 times. The final estimation result in a typical run is plotted in Figure 8, in which the sensor measurements, the average (true) and the estimated radiation intensities at the 60 sensor locations are shown. The estimated values of the parameters are $(-26, 17)$, $(125, -68)$, and 3.55×10^5, in good agreement with the truth and the results in [7]. Over the 50 runs, the RMS error of the first sensor location estimate is 11.05, the RMS error of the second sensor location estimate is 8.50, and the RMS error of the intensity estimate is 1.24×10^4. A simplified Gaussian sum method is also used. Because λ is large (greater than 20 at many sensor locations), the likelihood function is approximated by a single Gaussian with mean $\mathcal{I} + 0.5$ and variance $\mathcal{I} + 0.5$. The value $\mathcal{I} + 0.5$ is the best estimate of the unknown λ given a single measurement $\mathcal{I}$. For 50 Monte Carlo runs, the RMS error of the first sensor location estimate is 9.99, the RMS error of the second sensor location estimate is 9.58, and the RMS error of the intensity estimate is 1.31×10^4. The result is comparable to that with the Gaussian sum method.

The role of the data fusion method in the parameter estimation algorithm is to generate proposed particles at each iterations. Compared with measurement-independent random walk, the data fusion based proposal yields particles

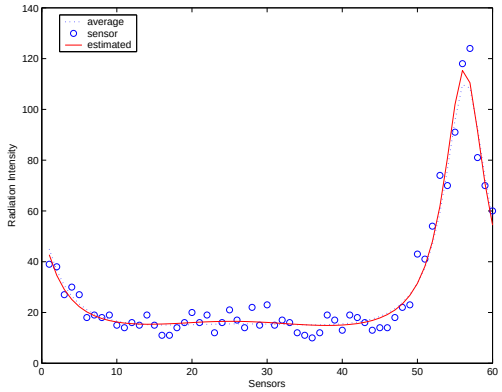

Figure 8. Radiation Intensities vs. Sensors

that are more likely to be consistent with the measurements, which is the same practice of the unscented particle filter.

VII. Conclusions

The problem of retrieving and fusing information from sensors for which the likelihood models are non-Gaussian was studied in this work. Examples of those sensors include chemical sensors and Gamma radiation sensors. Based on Bayes' rule, it was shown that the posterior mean and variance of the state variable are defined by three integrals over infinite intervals. Two general methods were developed for approximating the posterior mean and variance: the Gaussian sum method and the Gaussian-Hermite method.

The Gaussian sum method is based on the idea of approximating the likelihood function by a Gaussian sum. The method works best when the likelihood function is localized, smooth, and similar to Gaussian functions. Given the Gaussian sum approximation, the posterior mean and variance can be computed analytically using Kalman-like formulae. The Gaussian sum approximation of the likelihood function is computationally expensive when a large number of Gaussian components are needed to approximate the likelihood function accurately, but does not need to be done online. In many cases, the sensor likelihood function may be modeled as a Gaussian sum from the raw sensor data directly, for example, using the Expectation-Maximization algorithm.

The Gaussian-Hermite method is an efficient method based on Gaussian-Hermite quadrature, a numerical integration method closely related to the Unscented Kalman filter and Gaussian-Hermite filter. The method works best when the Gaussian kernel is chosen such that the integrand is approximately a low-degree polynomial over $(-\infty, \infty)$ (or at least over the effective support of the Gaussian kernel). When the Gaussian kernel is chosen as the prior distribution, which is the one of the most convenient choices, the performance of the method may be poor if the prior distribution is separated from the likelihood function and/or the likelihood function has a narrow support compared to the prior distribution.

The two methods are suited for CBRN sensor data fusion. Meanwhile, they are general enough for other data

fusion or estimation applications involving sensors with large uncertainty and non-Gaussian likelihood functions. In principle, both methods can be applied to multiple-dimensional problems, but as with other data fusion methods, they will suffer from the curse of dimensionality.

ACKNOWLEDGMENT

This work was supported by the Defense Threat Reduction Agency (DTRA) under Contract No. W911NF-06-C-0162. The authors gratefully acknowledge the support and constructive suggestions of Dr. John Hannan of DTRA.

REFERENCES

[1] S. Blackman and P. R., *Design and Analysis of Modern Tracking Systems*. Boston, MA: Artech House, 1999.

[2] R. I. Sykes, S. F. Parker, D. S. Henn, C. P. Cerasoli, and L. P. Santos, "PC-SCIPUFF version 1.1PD.1 technical documentation," Titan Corporation, Titan Research and Technology Division, ARAP Group, Tech. Rep., 1998.

[3] J. Daniel L. Stephens and A. J. Peurrung, "Detection of moving radioactive sources using sensor networks," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2273–2278, 2004.

[4] R. J. Nemzek, J. S. Dreicer, D. C. Torney, and T. T. Warnock, "Distributed sensor networks for detection of mobile radioactive sources," *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1693–1700, 2004.

[5] S. M. Brennan, A. M. Mielke, and D. C. Torney, "Radioactive source detection by sensor networks," *IEEE Transactions on Nuclear Science*, vol. 52, no. 3, pp. 813–819, 2005.

[6] P. Robins, V. Rapley, and P. Thomas, "A probabilistic chemical sensor model for data fusion," in *The Eighth International Conference on Information Fusion*, Philadelphia, PA, 2005.

[7] M. Morelande, B. Ristic, and A. Gunatilaka, "Detection and parameter estimation of multiple radioactive sources," in *The Tenth International Conference on Information Fusion*, Quebec City, Quebec, 2007.

[8] R. E. Curray, *Estimation and Control with Quantized Measurements*. Cambridge, MA: MIT Press, 1970, ch. 2.

[9] K. Clements and R. Haddad, "Approximate estimation for systems with quantized data," *IEEE Transactions on Automatic Control*, vol. 17, no. 2, pp. 235–239, Apr 1972.

[10] J. C. Keenan and J. B. Lewis, "Estimation with quantized measurements," *1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*, vol. 15, pp. 1284–1291, Dec. 1976.

[11] E. Sviestins and T. Wigren, "Optimal recursive state estimation with quantized measurements," *IEEE Transactions on Automatic Control*, vol. 45, no. 4, pp. 762–767, Apr 2000.

[12] A. Hodel and J. Hung, "A state estimator with reduced sensitivity to sensor quantization," *The 29th Annual Conference of the IEEE Industrial Electronics Society, 2003. IECON '03.*, vol. 1, pp. 586–590 vol.1, Nov. 2003.

[13] Z. Duan, V. P. Jilkov, and X. R. Li, "State estimation with quantized measurements: Approximate mmse approach," in *The Eleventh International Conference on Information Fusion*, Cologne, Germany, 2008.

[14] Y. Ruan, P. Willett, A. Marrs, F. Palmieri, and S. Marano, "Practical fusion of quantized measurements via particle filtering," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 15–29, January 2008.

[15] R. Gray and D. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.

[16] R. Merwe, A. Doucet, N. Freitas, and E. Wan, "The unscented particle filter," Cambridge University Engineering Department, Tech. Rep. Technical Report CUED/F-INFENG/TR 380, August, 2000.

[17] L. Y. Wang, J.-F. Zhang, and G. Yin, "System identification using binary sensors," *IEEE Transactions on Automatic Control*, vol. 48, no. 11, pp. 1892–1907, Nov. 2003.

[18] L. Y. Wang and G. G. Yin, "Asymptotically efficient parameter estimation using quantized output observations," *Automatica*, vol. 43, no. 7, pp. 1178–1191, July 2007.

[19] W. Kim, K. Mechitov, J.-Y. Choi, and S. Ham, "On target tracking with binary proximity sensors," *The Fourth International Symposium on Information Processing in Sensor Networks, 2005. IPSN 2005.*, pp. 301–308, April 2005.

[20] N. Shrivastava, R. Mudumbai, U. Madhow, and S. Suri, “Target tracking with binary proximity sensors: fundamental limits, minimal descriptions, and algorithms.” in *SenSys*, A. T. Campbell, P. Bonnet, and J. S. Heidemann, Eds. ACM, 2006, pp. 251–264. [Online]. Available: http://dblp.uni-trier.de/db/conf/sensys/sensys2006.html#ShrivastavaMMS06

[21] W. S. Lewellen and R. I. Sykes, “Analysis of concentration fluctuations from lidar observations of atmospheric plumes,” *Journal of Climate and Applied Meteorology*, vol. 25, pp. 1145–1154, 1986.

[22] E. Yee, “The shape of the probability density function of short-term concentration fluctuations of plumes in the atmospheric boundary layer,” *Boundary-Layer Meteorology*, vol. 51, pp. 269–298, 1990.

[23] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. New York, NY: Academic Press, 1975, ch. 3.

[24] H. W. Sorenson and D. L. Alspach, “Recursive bayesian estimation using gaussian sums,” *Automatica*, vol. 7, pp. 465–479, 1971.

[25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1998, ch. 5.

[26] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY: Dover, 1965, ch. 7, 22, 25.

[27] K. Ito and K. Xiong, “Gaussian filters for nonlinear filtering problems,” *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[28] O. Zoeter, A. Ypma, and T. Heskes, “Improved unscented kalman smoothing for stock volatility estimation,” in *2004 IEEE Workshop on Machine Learning for Signal Processing*, 2004, pp. 143–152.

[29] E. Wan and R. van der Merwe, “The unscented kalman filter,” in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. New York, NY: John Wiley & Sons, 2001, ch. 7.

[30] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2001.

[31] G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed. New York, NY: John Wiley & Sons, 1988, ch. 3.

[32] J. C. Spall, *Introduction to Stochastic Search and Optimization*. Hoboken, NJ: John Wiley & Sons, 2003, ch. 2.

PLACE
PHOTO
HERE

Yang Cheng

PLACE
PHOTO
HERE

K. V. Umamaheswara Reddy

Tarunraj Singh

Peter D. Scott