Comparison of SCIPUFF Plume Prediction with Particle Filter Assimilated Prediction for Dipole Pride 26 Data

Gabriel Terejanu  
Department of Computer Science and Engineering  
University at Buffalo  
Buffalo, NY 14260  
terejanu@buffalo.edu

Yang Cheng  
Department of Mechanical and Aerospace Engineering  
University at Buffalo  
Buffalo, NY 14260  
cheng3@buffalo.edu

Tarunraj Singh  
Department of Mechanical and Aerospace Engineering  
University at Buffalo  
Buffalo, NY 14260  
tsingh@eng.buffalo.edu

Peter D. Scott  
Department of Computer Science and Engineering  
University at Buffalo  
Buffalo, NY 14260  
peter@buffalo.edu

Abstract—This paper presents the application of a particle filter for data assimilation in the context of puff-based dispersion models. Particle filters provide estimates of the higher moments, and are well suited for strongly nonlinear and/or non-Gaussian models. The Gaussian puff model SCIPUFF, is used in predicting the chemical concentration field after a chemical incident. This model is highly nonlinear and evolves with variable state dimension and, after sufficient time, high dimensionality. While the particle filter formalism naturally supports variable state dimensionality high dimensionality represents a challenge in selecting an adequate number of particles, especially for the Bootstrap version. We present an implementation of the Bootstrap particle filter and compare its performance with the SCIPUFF predictions. Both the model and the Particle Filter are evaluated on the Dipole Pride 26 experimental data. Since there is no available ground truth, the data has been divided in two sets: training and testing. We show that even with a modest number of particles, the Bootstrap particle filter provides better estimates of the concentration field compared with the process model, without excessive increase in computational complexity.

Keywords: Data Assimilation, Particle Filter, Chem-Bio Field Test, Dispersion Model.

I. INTRODUCTION

There is an increase in requirement for accuracy and computational performance in atmospheric transport and diffusion models used in critical decision making in the context of chemical, biological, radiological, and nuclear (CBRN) incidents. The output (field concentrations and dosages) of the dispersion models is used directly to guide decision-makers, and as an input for higher fusion levels, such as situation and threat assessment. Therefore the accuracy of the models as well as the time of delivery of the forecasts plays an important part in decision making.

Gaussian dispersion models have been extensively studied and used in assessing the impact of CBRN incidents. They have gained popularity due to their straightforward theoretical approach, and due to their relatively low computational complexity [1].

For accurate CBRN output, one cannot rely solely on mathematical models or on measurements recorded by the sensors on the field, because of their uncertainties. Thus, for better estimates and lower uncertainty, a fusion step, called Data Assimilation, is necessary to combine the model forecasts and the measurements.

In Data Assimilation, the estimation of the unknown system states given the underlying dynamics of the system and a set of observations may be framed as a filtering, smoothing or prediction problem. Given a fixed discrete time interval, \{t_1, t_2, \ldots, t_N\}, over which observations are available, the problem of filtering is to find the best state at time \(t_k\) given all the observations prior to and including \(t_k\). The smoothing problem is to find the best state at time \(t_k\) given all the observations up to time \(t_N\), where \(t_k \leq t_N\). For \(t_k > t_N\) the prediction problem is to forecast the state of the system at time \(t_k\) using all the observations in the given interval.

For a linear system, under the assumption of Gaussian probability distributions, the problem of estimating the states of the system has an exact closed form solution given by the Kalman Filter [2]. If the probability distributions are non-Gaussian or the system is nonlinear, in general no closed-form solution is available and different assumptions and approximations have been made for quasi-optimal solutions maintaining both accuracy and tractability. The systems considered here are in general nonlinear, but the assumption that the process and observation uncertainties can be adequately modeled as Gaussian is made.

The nonlinear filtering problem has been extensively studied and successfully employed in many applications, with various methods provided in the literature. Among the best understood and most frequently cited are the Extended Kalman Filter (EKF), the Ensemble Kalman Filter (EnKF), and more recently the Unscented Kalman Filter (UKF), and the Particle Filter (PF). The EKF is historically the first, and still the most widely adopted approach to nonlinear estimation problems. It is based on the assumption over small time increments, nonlinear system dynamics can be accurately modeled by a first-order Taylor series expansion [3].

The PF uses a sampling approach, estimating the posterior
probability distribution, including its higher order moments, by propagating and updating a number of particles, without the assumption of Gaussian statistics [4]. The variable and high dimensionality of the state vector, which poses a problem to the standard nonlinear assimilation techniques, can be dealt with using Particle Filters [5]. Daum, et al. [6] showed that a carefully designed Particle Filter should mitigate the curse of dimensionality for certain filtering problems.

This paper presents an implementation of the Bootstrap Particle Filter to assimilate chemical concentration readings of the Dipole Pride 26 experiment [7] into the SCIPUFF dispersion model. The Particle Filter takes into account the uncertainty due to the data errors in the observed meteorology. The results show that the particle filter, with a modest number of particles, provides better estimates of the concentration field compared with the process model, underlying the importance of meteorological data accuracy in CBRN incidents.

Data assimilation based on sampling techniques for CBRN incidents and numerical weather prediction (NWP), such as Particle Filter [5], Unscented Kalman Filter [8], [9] or Ensemble Kalman Filter [10], [11], become more and more accessible as parallel computing becomes mainstream with the introduction of multi-core processors and multiple CPU computers. Sampling techniques such as Monte Carlo analysis [12], [13] or ensemble method [14], [15] have been used before to account for uncertainty due to data errors.

The Dipole Pride 26 (DP26) field experiment and the SCIPUFF dispersion model are presented in Section II. The Bootstrap Particle Filter is introduced in Section III and its implementation for this particular problem is described in Section IV. Numerical results on the CBRN scenario are presented in Section V and the conclusions and future work are discussed in Section VI.

II. DIPOLE PRIDE 26 AND SCIPUFF

The Dipole Pride 26 field experiment has been designed to validate transport and diffusion models at mesoscale distances [7]. The experiment has been conducted at Yucca Flat, Nevada where gaseous sulfur hexafluoride (SF6) has been released in a series of instrumented trials, of which only 17 provide useful puff dimension information.

Three lines of sensors, Fig. 1, each with 30 whole air-samplers have been used to record average concentrations every 15 min. The chemical sensors, known also as sequential bag samplers (12 bags per sensor), are placed at 1.5 m above the ground and spacing along lines is about 250 m. The total sampling time of the chemical sensor is 3 hr, hence total experimental duration for each trial to be monitored is 3.5 hr. This is achieved by delaying the acquisition of the last line of sensors with 30 min. Six continuous SF6 analyzers, TGA-4000, have been used to record high-frequency variations of the gas concentration field, but their placement does not offer enough resolution to capture the crosswind structure of the chemical plume, and for the purpose of this paper these readings have been excluded from the study.

Eight Meteorological Data (MEDA) stations were used to provide surface-based meteorological measurements and two pilot balloon stations and one radiosonde provided the upper-air meteorological profiles. The meteorological measurements recorded provide information about the wind direction and speed, temperature, pressure and humidity. The variation of the wind field is given by the standard deviation of hourly wind speeds and directions recorded at the MEDA stations which are about 0.5 − 2 m s$^{-1}$ and 10$^0$ − 30$^0$ respectively [16].

In this paper we are focusing only of trial number six, which took place in Nov 12, 1996 and where a mass of 11.6 kg of SF6 had been release from 6 m height at the North dissemination site N2 as in Fig. 1. The dispersion model used to predict the chemical concentrations and dose at the sensor locations is SCIPUFF [17]. SCIPUFF is a Lagrangian puff dispersion model that outputs the chemical concentrations at specified locations as cumulative contributions of Gaussian puffs. SCIPUFF is the dispersion engine incorporated into Hazard Prediction and Assessment Capability (HPAC) tool, used by the Defense Threat Reduction Agency (DTRA) for situation assessment of CBRN incidents. Besides the mean concentration field, SCIPUFF is also providing the uncertainty in the concentration value due to the stochastic nature of the turbulent diffusion process. The simulation is driven by the meteorological data, which in this case is provided by the MEDA stations, from which a wind field is created.

The observed wind data are fit in a least square sense, using variational methodology. An initial gridded wind field is constructed from the observation data by interpolation. Adjustments are then made to the initial 3D interpolated wind field vectors so as to satisfy conservation of mass in a way that also minimizes an integral function of the difference between
the initial and adjusted fields.

Existing literature provides evaluation studies of SCIPUFF with Dipole Pride 26 [16], [18], [19]. The results reported show that SCIPUFF predictions are within a factor of 2 of observations about 50\%, where the model evaluation was based on maximum dosage anywhere on the sampling lines [16]. The studies emphasize the importance of wind field in the chemical concentration prediction accuracy and conclude that SCIPUFF prediction performance is comparable or better than other dispersion models.

III. Bootstrap Particle Filter

In this section we focus our attention on sequential state estimation using sequential Monte Carlo (SMC). SMC is known also as bootstrap filtering, particle filtering, the condensation algorithm, interacting particle approximations and survival of the fittest [20].

Consider the following nonlinear system, described by the difference equation and the observation model:

\[ x_{k+1} = f(x_k) + w_k \]  
\[ z_k = h(x_k) + v_k \]

(1) \hspace{1cm} (2)

Denote by \( Z_k = \{ z_i | 1 \leq i \leq k \} \) the set of all observations up to time \( k \), conditionally independent given the process with distribution \( p(z_k | x_k) \). Also, assume that the state sequence \( x_k \) is an unobserved (hidden) Markov process with initial distribution \( p(x_0) \) and transition distribution \( p(x_{k+1} | x_k) \).

Our aim is to estimate recursively the posterior distribution \( p(x_k | Z_k) \) and expectations of the form [21]:

\[ E[Z_k] = E[h(x_k)] = \int h(x_k)p(x_k | Z_k)dx_k \]

(3)

In particle filters, the posterior distribution \( p(x_k | Z_k) \) is approximated with \( N \) weighted particles \( \{ x_{k,i}^{(i)}, w_{k,i}^{(i)} \}_{i=1}^{N} \), given by

\[ p(x_k | Z_k) \approx \sum_{i=1}^{N} w_{k,i}^{(i)} \delta_{x_{k,i}^{(i)}}(x_k) \]

(4)

where \( x_{k,i}^{(i)} \) are the particles drawn from the importance function or proposal distribution, \( w_{k,i}^{(i)} \) are the normalized importance weights that sum up to one and \( \delta_{x_{k,i}^{(i)}}(x_k) \) denotes the Dirac-delta mass located in \( x_{k,i}^{(i)} \). Thus the expectation of a known function \( h(x_k) \) with respect to \( p(x_k | Z_k) \) is then approximated by

\[ \int h(x_k)p(x_k | Z_k)dx_k \approx \sum_{i=1}^{N} w_{k,i}^{(i)} h(x_{k,i}^{(i)}) \]

(5)

Suppose that we cannot sample from the posterior distribution and use an importance sampling approach to sample from a proposal distribution \( q(x_k | Z_k) \). Hence, we can recursively update the weights:

\[ w_{k+1}^{(i)} = w_{k}^{(i)} \frac{p(z_{k+1} | x_{k,i}^{(i)})p(x_{k,i}^{(i)} | x_k)}{q(x_{k,i}^{(i)} | x_k, Z_{k+1})} \]

(6)

After a few iterations all but one particle will have negligible weights. Hence the algorithms allows time to update a large number of weights with no effect in our sampling, effect known as degeneracy problem. This problem can be overcome by adding a resampling strategy to Sequential Importance Sampling (SIS) Algorithm [1].

Algorithm 1 Sequential Importance Sampling (SIS) algorithm

1. Draw \( x_k^{(i)} \sim q(x_k | x_{k-1}, Z_k) \) for \( i = 1 \ldots N \)
2. Compute the importance weights
\[ w_{k+1}^{(i)} = w_{k}^{(i)} \frac{p(z_{k+1} | x_{k,i}^{(i)})p(x_{k,i}^{(i)} | x_k)}{q(x_{k,i}^{(i)} | x_k, Z_{k+1})} \]
3. Normalize the importance weights \( \bar{w}_{k}^{(i)} = \frac{w_{k}^{(i)}}{\sum_{i=1}^{N} w_{k}^{(i)}} \)
4. Multiply/Discard particles \( \{ x_{k+1}^{(i)} \}_{i=1}^{N} \) with respect to high/low importance weights \( w_{k+1}^{(i)} \) to obtain \( N \) new particles \( \{ x_{k+1,i}^{(i)} \}_{i=1}^{N} \) with equal weights.

The importance function plays a significant role in the particle filter. Usually, it is difficult to find a good proposal distribution especially in a high dimensional space. One may choose to approximate it using different methods, thus different flavor of particle filter. One of the simplest importance function is given by

\[ q(x_{k+1}^{(i)} | x_k^{(i)}, Z_{k+1}) = p(x_{k+1}^{(i)} | x_k^{(i)}) \]

(7)

This implementation is called the Bootstrap Particle filter. By substituting (7) back into (6) the new weight update equation becomes:

\[ w_{k+1}^{(i)} \propto w_{k}^{(i)} p(z_{k+1} | x_{k,i}^{(i)}) \]

(8)

The resampling step reduces the sample impoverishment effect but introduces new practical problems: it limits the opportunity to have an efficient parallel algorithm, and particles with high weights are statistically selected many times and leads to a loss of diversity among the particles [20].

IV. PARTICLE FILTER IMPLEMENTATION

Three types of uncertainties [22] are present in the model predictions: model uncertainty due to the inaccurate representation of the chemical and dynamical processes, data uncertainty due to the errors in data used to drive the model and random turbulence of the atmosphere.

The model uncertainty is not completely reducible, and in the SCIPUFF prediction case it is estimated. The uncertainty due to the variability of the atmosphere cannot be further reduced and the uncertainty due to data errors it is usually high, more than 50\% of the total uncertainty [23], but it can be minimized. The data errors considered in this paper are meteorological data: wind speed and wind direction. The errors are due to the unrepresentative sitting of the wind sensors in the field and sensor accuracy and calibration [19]. This information is rarely available and for this study a standard deviation of 0.5 m s\(^{-1}\) for the wind speed and 5° for the wind direction has been considered.
A FORTRAN implementation of the Particle Filter has been specially created for SCIPUFF to run in parallel on the cluster hosted at the Center for Computational Research at University at Buffalo. This implementation of the particle filter is designed to account for the uncertainty in the wind sensor readings while coping with the challenges of the data assimilation for the CBRN incidents using Gaussian puff models: variable dimensionality and high dimensionality [5].

Compared with the maximum dosage approach [16], the evaluation method used in this paper is to compare the predicted dosage after every 15 min with the corresponding observed dosage.

Since there is no ground truth, the samples provided in the DP26 have been divided to two sets: the training set used to perform the data assimilation, composed of Line 1 and Line 2 of sensors, and the testing set. Line 3 of sensors, used for performance evaluation.

Due to the spatial and temporal distance between the wind readings all the considered independent random variables described by a Gaussian distribution with mean given by the nominal sensor reading and uncertainty given by the standard deviation mentioned above. Hence the particles, each representing a SCIPUFF instance, are propagated using wind field generated from data sampled from this distribution. Each particle outputs a dose field \( d_j^i(k) \) which depends on the wind field; here \( j \) is the sensor number, \( i \) is the particle number and \( k \) is the time step. The estimated dosage \( \hat{d}_j(k) \) is given by the following relation:

\[
\hat{d}_j(k) = \sum_{i=1}^{N} w_i^j d_j^i(k) \tag{9}
\]

Here \( w_i^j \) is given by Eq.\( (8) \). The conditional probability present in Eq.\( (8) \) is unknown since the authors could not find information regarding the uncertainty in the concentration readings for the whole air-samplers. The concentrations readings of the sensors have been assumed to be independent random variables due to the spatial and temporal distances. The likelihood have been approximated with a Gaussian function given by:

\[
p(d_j(k)|d_j^i(k)) = \mathcal{N}(d_j^i(k)|d_j(k), \sigma(k)) \tag{10}
\]

Here \( d_j(k) \) is the observed dosage at the \( j^{th} \) sensor at time \( k \) and \( \sigma(k) \) is the sample variance of the difference between observed dosages and predicted ones over all the particles. Any predicted dosage less than 1 ppt-hr (including zero) has been set to 1 ppt-hr and all the observed dosages less than 10 ppt-hr have been ignored. Hence the total number of dosage values to be compared is 47.

V. NUMERICAL RESULTS

Both the process model predictions, using nominal wind sensor readings, and the particle filter predictions with perturbed wind field have been compared against the observed dosages on the third line of sensors.

To evaluate the performance of the particle filter compared with the process model, we consider the following statistical measures [24]: FB - fractional bias, MG - geometric mean bias, NMSE - normalized mean square error, VG - geometric variance, FAC2 - fraction of predictions within a factor 2 of observations and FAC3 - fraction of predictions within a factor 3 of observations.

\[
FB = \frac{\overline{D_o} - \overline{D_p}}{0.5(\overline{D_o} + \overline{D_p})} \tag{11}
\]

\[
MG = \exp(\ln(D_o) - \ln(D_p)) \tag{12}
\]

\[
NMSE = \frac{(D_o - D_p)^2}{D_o D_p} \tag{13}
\]

\[
VG = \exp\left(\frac{(\ln(D_o) - \ln(D_p))^2}{\ln(D_o) - \ln(D_p)}\right) \tag{14}
\]

\[
FAC2 = \#D_p \text{ such that } \frac{1}{2} \leq \frac{D_p}{D_o} \leq 2 \tag{15}
\]

\[
FAC3 = \#D_p \text{ such that } \frac{1}{3} \leq \frac{D_p}{D_o} \leq 3 \tag{16}
\]

Here \( D_o \) represents the observed dosage and \( D_p \) the predicted dosage. Since we are dealing with random samples, 50 Monte Carlo runs have been performed in assessing the performance of the particle filter. The numerical results based on the performance metrics have been tabulated in Table I.

Overall the particle filter provides improved estimates compared to the process model and all the performance measures are better on average. Since the predicted and observed values vary by several orders of magnitude, MG, VG, FAC2 and FAC3 are more appropriate. While we do not see a significant

| Process Model | Particle Filter | PF 95% CI |
|---------------|----------------|-----------|
| FB 1.426      | 1.405          | 1.390 – 1.419 |
| MG 0.456      | 0.443          | 0.402 – 0.484 |
| NMSE 8.658    | 8.466          | 8.249 – 8.683 |
| VG 86.75      | 61.20          | 53.65 – 68.74 |
| FAC2 6.38%    | 11.45%         | 9.31% – 13.58% |
| FAC3 6.38%    | 22.17%         | 19.45% – 24.89% |
using concentration measurements provided by the chemical sensors deployed in the field. Due to the high uncertainty in the meteorological input, the CBRN dispersion models should be accompanied by a data assimilation step to account for this uncertainty and improve the predictions. For the Dipole Pride 26 the particle filter has doubled the number of predictions within a factor of 2 of the observations and it has tripled the ones within a factor of 3 of the observations.

A complete evaluation of the particle filter on all the Dipole Pride 26 trials and simulations with uncertainty for also the future work.

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. The authors would also like to thank Dr. Joseph C. Chang and Dr. Steve Hanna for their input regarding the Dipole Pride 26 dataset.