A Robust Data Assimilation Approach in the Absence of Sensor Statistical Properties

Reza Madankan[1] and Puneet Singla[2] and Tarunraj Singh[3]

***Abstract*—**
A convex optimization based approach is presented to perform model-data assimilation of spatial temporal dynamical systems where sensor error characteristics are not available. The key idea of the proposed technique is that one should not make any assumption regarding the statistical properties of sensor data when they are not available. Recently developed quadrature scheme, *Conjugate Unscented Transformation* in conjunction with convex optimization tools is used to obtain an approximation of posterior density function. The proposed approach is validated by considering the problem of source parameter estimation for toxic material release in the atmosphere. The numerical experiments provides a basis for optimism for the robustness of the proposed methodology.

I. INTRODUCTION

A major challenge in estimation of dynamic systems is when information regarding the statistical properties of sensor data is not available or partially available. This frequently happens in data assimilation of atmospheric data by using satellite imagery. This is due to the fact that satellite imagery data can be polluted with noise, depending on weather conditions, clouds, humidity, etc. Unfortunately, there is no accurate procedure to quantify the error due to these factors on the output of satellite data. Hence, use of the classical data assimilation methods in this situation is not straight forward.

There exist numerous efforts [1]–[13] regarding data assimilation in absence of sensor error characteristics. The essence of these works is about the estimation of the noise statistics along with model-data fusion of dynamical systems. These works can be divided into two different categories: i) covariance estimation methods, ii) methods that estimate distribution of the associated noise in measurement data.

The key idea of covariance estimation methods is to estimate the covariance of associated measurement noise along with other unknown parameters [1]–[10] . Note that most of the these techniques assume the associated noise signal to be Gaussian, which could be a restrictive assumption for some practical applications. In addition, they only concentrate on estimating covariance matrix of associated noise signal, while knowing only covariance of the noise signal may not be enough. One should also consider that all the current approaches for covariance estimation of noise signal increase computational complexity of the whole model-data fusion process due to the introduction of new uncertain parameters.

Besides covariance estimation technique, there exist few recent efforts [11], [13] that focus on estimation of the probability density function (pdf) of the existing noise signal in observations. Note that in spite of most of the covariance estimation methods, all the density estimation methods are applicable to nonlinear time varying dynamical systems. In addition, they don't consider any restrictive assumption regarding distribution of the noise signal. However, similar to covariance estimation techniques, density estimation methods significantly increase the computational complexity involved in the whole model-data fusion process by introducing new uncertain parameters that need to be estimated.

Our Approach: The key contribution of this article is to develop a new model-data fusion method, that does not require any information regarding statistics of the existing noise signal in measurement data. Hence, avoiding the computational complexities involved in current methods. The key idea of this paper is to not make any assumption regarding the statistical properties of sensor data that are not available. In other words, in the absence of statistical information of the sensor data, we maintain the higher order prior statistics but update the posterior mean to be in compliance with sensor data. Schematic view of the proposed methodology is shown in Fig. 1. As Fig. 1 illustrates, our approach consists of four different components that are combined together to perform the task of parameter estimation and model forecasting. These components are i) Uncertainty Quantification, ii) Error Evaluation, iii) Optimization, and iv) Reconstruction of Posterior Quadrature Points.

The structure of this paper is as follows. First, problem statement is described in section II. We briefly discuss the method of quadrature points, used for uncertainty quantification, in section III. In section IV, we will discuss about the theoretical basis of evaluation of error ensembles and error statistics. Then in section V, we construct a convex-quadratic optimization problem whose solution results in posterior value of quadrature weights. We will propose a scheme for reconstruction of quadrature points of uncertain parameters in section VI. Performance of the proposed approach is then demonstrated by using a simple numerical example in section VII. Finally, conclusion is presented in section VIII.

*This work was supported by the National Science Foundation under award number CMMI-534 1054759 and AFOSR grant number FA9550-11-1-0012.

[1]Reza Madankan is with the Department of Imaging Physics, University of Texas MD Anderson cancer research center rm93@buffalo.edu

[2]Puneet singla is with the Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, 14260 psingla@buffalo.edu

[3]Tarunraj Singh is with the Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, 14260 tsingh@buffalo.edu

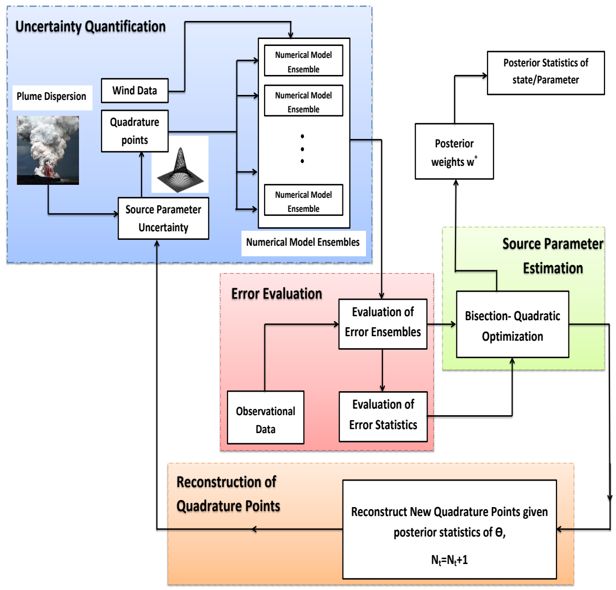


Fig. 1. Schematic view of estimation process. N_t represents the time step. w^- and w^+ denote prior and posterior weights

II. Problem Statement

Let us consider a dynamic system, given by

$$\dot{\mathbf{x}} = f(t, \mathbf{x}, \Theta) \quad (1)$$

where, t is time, and $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\Theta \in \mathbb{R}^{m\times 1}$ represent states and parameters of the system, respectively. Parameter Θ is assumed to be uncertain, defined by some prior distribution $p(\Theta)$.Also, measurement data is given as:

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \Theta, \nu) \quad (2)$$

where, $\mathbf{y} \in \mathbb{R}^{s\times 1}$ denotes observation data, $\mathbf{h}(.)$ is the observation operator, and ν is the associated noise. Note that no information is available regarding the statistical properties of ν. Our goal is to have a reliable estimate of parameter $\boldsymbol{\Theta}$, by assimilating measurement data with model forecasts.

III. Uncertainty Quantification

The first step to perform estimation process is to quantify the uncertainty associated with forward model output during the time propagation. In here, the method of quadrature points is used to perform this task, i.e. N quadrature points for parameter Θ are generated based on prior pdf, $p(\Theta)$ and Eq. (1) is simulated for each of these realizations. The k^{th} order moment of the model output then can be evaluated as

$$\mathcal{E}[\mathbf{x}^k] = \int_{\boldsymbol{\xi}} \mathbf{x}^k(\boldsymbol{\Theta}, t) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta} \simeq \sum_{q}^{N} w_q \mathbf{x}^k(\boldsymbol{\Theta}(\boldsymbol{\xi}^q), t) \quad (3)$$

where, $\Theta(\boldsymbol{\xi}^q)$s are the quadrature points, generated based on associated pdf of uncertain parameter Θ, and w_qs are their corresponding quadrature weights. Different types of quadrature schemes like classical Gaussian quadrature rule can be used to evaluate above integral. For a generic m-dimensional integral, the tensor product of 1-dimensional Gaussian quadrature points results in an undesirable exponential growth of the number of points. Hence, using regular Gaussian quadrature points will be computationally expensive in higher dimensions. In here, we have used recently developed Conjugate Unscented Transform (CUT) [14], to overcome this drawback of regular quadrature points. The proposed CUT points are efficient in terms of accuracy while integrating polynomials and yet just employ a small fraction of the number of points used by the traditional Gaussian quadrature scheme. For instance, only 161 CUT quadrature points are required to satisfy the 8^{th} order moments in $4-$dimensional space, while 625 quadrature points are needed based on Gauss-Legendre quadrature scheme.

IV. Error Evaluation

After precise uncertainty quantification, in order to compare model outputs with measurement data, one needs to provide a measure to compute the difference (error) between model outputs and the measurement data. To describe this in more detail, let us assume the error between model predictions and given measurement data to be given as:

$$e = d(\mathcal{M}(\boldsymbol{\Theta}), \mathbf{y}) \quad (4)$$

where, error $e \in \mathbb{R}^+$, and $d(\mathcal{M}(\boldsymbol{\Theta}), \mathbf{y}) \geq 0$ is a metric operator that is used to calculate the difference between model forecast $\mathcal{M}(\boldsymbol{\Theta}) = \mathbf{h}(\mathbf{x}, \Theta, \mathbf{0})$ and observation data $\mathbf{y}$. A wide variety of error metrics [15]–[17] can be used for measuring the difference between model forecast ensembles and observation data. n the following, we describe few of these metrics which we have used in this manuscript.

Hausdorff Metric: is defined as the *maximum distance of a set to the nearest point in the other set*. Mathematically, Hausdorff distance from set A to set B is defined as:

$$h(A, B) = \max_{a\in A}\left\{\min_{b\in B}\{dist(a, b)\}\right\} \quad (5)$$

where, a and b are points of sets A and B, respectively, and $dist(a, b)$ is any metric distance between these points. Note that $h(A, B) \neq h(B, A)$ in general. Hence, Hausdorff distance can not be used as a metric. To overcome this drawback, a more general definition of Hausdorff distance is given as

$$H(A, B) = \max\{h(A, B), h(B, A)\} \quad (6)$$

Note Eq. (6) represents a metric and can be used to quantify the error between model forecasts and observation, i.e.

$$e(\Theta_q) = H(\mathcal{M}(\Theta_q), \mathbf{y}), \quad q = 1, 2, \cdots, N$$

Euclidean Distance: Given two image A and B, Euclidean metric is defined as:

$$d_E(A, B) = \sqrt{\sum_{i,j}(a_{i,j} - b_{i,j})^2} \quad (7)$$

where, $a_{i,j}$ and $b_{i,j}$ are the intensity of images A and B, respectively, at grid point (i, j).

A. Error Statistics

Once the error between each of the model forecast ensembles and measurement data is calculated, one can find its statistics as a weighted average of error ensembles. Note that given N quadrature points for model parameter Θ, there will be N realizations for the error. Statistics of the error can be defined using these quadrature values. For instance, expected value of the error $e \in \mathbb{R}^+$ can be written as:

$$m_1 = \mathcal{E}[e] = \sum_{q=1}^{N} w_q^- e_k(\Theta_q) \tag{8}$$

where, w_q^- is the corresponding weight for Θ_q which is q^{th} quadrature value of parameter Θ. Similarly, higher order central moments of error e are defined as:

$$m_k = \sum_{q=1}^{N} w_q^- \left(e_k(\Theta_q) - m_1\right)^k, \quad k = 2, 3, \cdots, N_m \tag{9}$$

where, $N_m \in \mathbb{N}$ is the highest order of central moments calculated.

V. Source Parameter Estimation

The key idea of parameter estimation is to find posterior values of quadrature weights w_qs to minimize the expected value of the error, denoted by $\mathcal{E}[e]$, while preserving its higher order central moments. This can be formulated as:

$$\min_{\mathbf{w}} \mathcal{E}[e] = \min_{\mathbf{w}} \sum_{q=1}^{N} w_q e(\Theta_q) \tag{10}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} \right)^k = m_k, \quad 2 \leq k \leq N_m \tag{11}$$

$$\sum_{q=1}^{N} w_q = 1, \quad 0 \leq w_q \leq 1 \tag{12}$$

where, $\mathbf{w} = [w_1, w_2, \cdots, w_N]^T$. Note that Eq. (10) is an ill-posed optimization problem and minimizing only $\mathcal{E}[e]$, given Eq. (11) and Eq. (12) has infinite solution. To overcome this, one can add a penalty term to Eq. (10) to minimize the difference between prior and posterior weights, along with minimizing $\mathcal{E}[e]$. In this way, modified optimization problem can be written as:

$$\min_{\mathbf{w}} \lambda_1 \mathcal{E}[e] + \lambda_2 |\mathbf{w} - \mathbf{w}^-|_2 =$$
$$\min_{w} \lambda_1 \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} + \lambda_2 \sum_{q=1}^{N} (w_q - w_q^-)^2 \tag{13}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} \right)^{n_k} = m_k \tag{14}$$

$$\sum_{q=1}^{N} w_q = 1, \quad 0 \leq w_q \leq 1 \tag{15}$$

where, $2 \leq k \leq N_m$ and m_1 is given by Eq. (8), $\lambda_1, \lambda_2 > 0$ are given constants, and $\mathbf{w}^-$ denotes *prior* value of quadrature weights. Coefficients λ_1 and λ_2 represent the relative weights to the two terms in the cost function. Intuitively, when $\lambda_2 \ll \lambda_1$ Eq. (13) focuses on minimizing $\mathcal{E}[e]$, rather than the difference of $\mathbf{w}$ and $\mathbf{w}^-$. On the other hand, when $\lambda_1 \ll \lambda_2$, Eq. (13) returns the same values of $\mathbf{w}^-$ for $\mathbf{w}$.

As one can see, proposed approach minimizes the expected value of the difference between model forecasts and measurement data while preserving higher order central moments of error. Schematic view of this process is illustrated in Fig. 2. As illustrated, proposed technique minimizes expected value of the error while preserving shape of the error distribution, i.e. higher order statistics of error distribution. The intuition behind preserving higher order central moments of the error is to account for possible inaccuracies of measurement data.

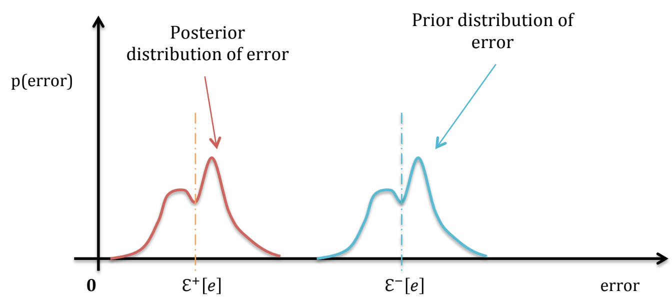


Fig. 2. Schematic view of the proposed optimization method for finding the posterior quadrature weights. As illustrated, proposed method minimizes expected value of the error while preserving the shape of error distribution, i.e. higher order statistics of error. Prior ($\mathcal{E}^-[e]$) and posterior ($\mathcal{E}^+[e]$) expected value of error ensembles are also shown.

Convex Quadratic Optimization: Eq. (13), along with Eq. (14) and Eq. (15) is a nonlinear optimization problem due to presence of equality constraints for higher order central moments. One can utilize an iterative approach to transfer this nonlinear optimization problem into a convex quadratic optimization problem, which is easier to solve. This has been achieved by making use of the bisection method. In detail, let us assume that the value of m_1 is known, then the nonlinear optimization problem is converted to the following quadratic optimization problem.

$$\min_{\mathbf{w}} |\mathbf{w} - \mathbf{w}^-|_2 = \sum_{q=1}^{N} (w_q - w_q^-)^2 \tag{16}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - m_1\right)^k = \quad m_k, \quad 2 \leq k \leq N_m \qquad (17)$$

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) = \qquad m_1 - e(\Theta_1) \qquad (18)$$

$$0 \leq w_q \leq 1 \qquad (19)$$

As one can see, Eq. (16) along Eq. (17) to Eq. (19) is a quadratic optimization problem which can be solved much easier than the original optimization problem described by Eq. (13) to Eq. (15). One can iteratively solve resulted quadratic optimization problem for different values of m_1 to minimize the original optimization problem. This can be achieved with the help of bisection method. In detail, we first assume lower bound and upper bound of m_1 to be zero and prior mean, $\sum_{q=1}^{N} w_q^- e(\Theta_q)$, respectively. Then, we assign m_1 to be the average of considered lower and upper bounds and solve the quadratic convex optimization problem in Eq. (16)-Eq. (19). If the optimization was not successful, then we substitute m_1 with the new value which is given from bisection method. This process is repeated iteratively, until convergence of bisection method is ensured. The pseudo-code for this procedure is described in the following:

Iterative algorithm for solving Eq. (16) to Eq. (19)

```
assign upper bound (ub) of m_1 to be the prior
expected value of error, i.e.
    ub = sum_{q=1}^{N} w_q^- e(Θ_q)
assign lower bound (lb) of m_1 to be zero.
while (ub − lb) ≥ threshold
        m_1 = 0.5 * (ub + lb);
        Perform quadratic optimization to
        solve Eq. (16) to Eq. (19)
        if optimization is feasible
              ub = m_1;
        else
              lb = m_1;
        end
end
```

The following theorems ensure that the obtained value for m_1 by the above convex optimization is in fact the smallest possible value of m_1 that can be found by solving the original optimization problem defined by Eq. (13) to Eq. (15).

Theorem 1: Assume that for a given value of $m_1 = m_1^c$, there exists no solution for the optimization problem defined by Eq. (16) to Eq. (19). Then the optimization problem defined by Eq. (16) to Eq. (19) doesn't have any solution for any values of $m_1 < m_1^c$.

Theorem 2: For a given value of $m_1 = m_1^c$ and $\forall\ 0 \leq w_q \leq 1$, the optimization problem defined by Eq. (13) to Eq. (15) returns an infeasible solution if and only if the optimization problem defined by Eq. (16) to Eq. (19) returns an infeasible solution for $m_1 = m_1^c$ and $\forall\ 0 \leq w_q \leq 1$.

We refer the reader to [18] for detailed proof of aforementioned theorems.

A. *Posterior Statistics of Parameter Θ and State $\mathbf{x}$*

After finding posterior values of quadrature weights, one can use these weights to evaluate posterior statistics of Θ.

$$m_k^+(\Theta) = \sum_{i=1}^{N} w_i^+ \left(\prod_{j=1}^{m} \Theta_{i_j}^{n_j}\right), \quad \sum_{j=1}^{m} n_j = k, \qquad (20)$$

where, $m_k^+(\Theta)$ denotes k^{th} order posterior moment of parameter Θ and Θ_{i_j} is j^{th} element of i^{th} quadrature realization of vector Θ. Similar procedure can be used for finding posterior statistics of model output, i.e.

$$m_k^+(\mathbf{x}) = \sum_{i=1}^{N} w_i^+ \left(\prod_{j=1}^{n} x_{i_j}^{n_j}\right), \quad \sum_{j=1}^{m} n_j = k \qquad (21)$$

VI. Reconstruction of Quadrature Points

The major drawback of proposed approach for data assimilation lies in possible degeneracy of weights after few time steps. In other words, most of the weights reduces to zero after few measurement updates and the proposed method can result in a erroneous estimate. To surpass this issue, a resampling approach is considered to update the weights and model forecast after each measurement update. Hence, after each measurement update, new quadrature points are reconstructed based on obtained posterior statistics from data assimilation. Numerical model is then simulated to reproduce model forecasts using these new quadrature points.

In here, we have used a simple approximation technique that *approximates* the posterior pdf of uncertain parameter Θ with a uniform pdf. Note that the mean and covariance of approximated uniform pdf are equal to posterior statistics obtained from data assimilation. New quadrature points are then generated based on this approximated uniform pdf.

VII. Numerical Simulations

To verify performance of the proposed approach, we consider the dispersion of Propane over Manhattan island. The domain of interest and applied wind field (at one specific time) are shown in Fig. 3. Simulation time is considered to be 10 $hrs.$ starting from $00:00$ of September 1^{st}, 2013. North American Regional Reanalysis wind data at pressure level 100 kpa (height $\simeq$ 100 $m.$) is used as the wind-field for simulation. Four instantaneous mass releases, as shown in Fig. 3, are considered where their locations are known and the only uncertain parameters are their amount of mass release. It is assumed that releases happen at the same time, i.e. all source releases happen at $00:00$ of September 1^{st}. All mass releases are assumed to be uniformly distributed between 50 kg and 150 kg. A set of 161 CUT8 quadrature points (i.e. satisfying upto the 8^{th} order moments) are used to quantify the uncertainty involved in concentration of propane. Simulation of dispersion/advection has been performed using SCIPUFF [19] numerical model, where concentration of propane is recorded every 30 $mins.$

Measurement data is available at every 30 $mins.$, starting from $02:00$ of September 1^{st}. A random realization of mass $[m_1, m_2, m_3, m_4] = [66.2, 137.04, 83.8, 122.1]^T\ kg.$ was used for simulation of dispersion incident. Then, obtained concentration field was discretized and polluted with a uniformly distributed integer random field $\omega(lon, lat, t)$ to generate the measurement data. The magnitude of ω at each spatio-temporal location is between -5 and $+5$, i.e. $\omega(lon, lat, t) \in \{-5, -4, \cdots, +5\}$. Note that resampling is performed after each weight update, as explained in Section VI. In addition, $N_m = 3$ in our simulations.

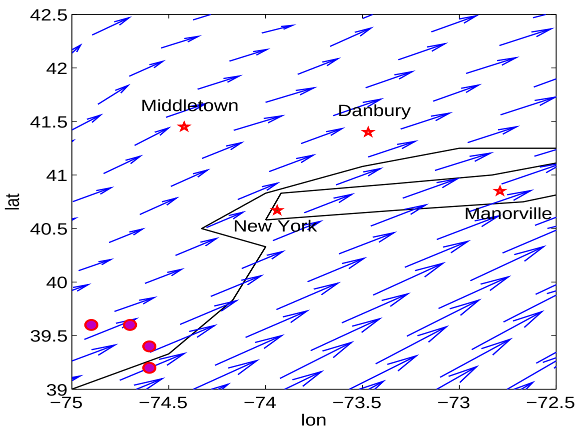


Fig. 3. Schematic layout of Propane release over New York, source locations are shown with circle markers, the wind-field (at $t = 0\ hr$) and different cities can be seen in the background.

Fig. 4(a) represent statistics (mean and variance) of m_2 estimate using Hausdorff metric (HD) and Euclidean distance. To be concise in writing, we have only shown statistics of parameters m_2. From Fig. 4(a) it is clear that mean estimate of m_2 converges to its actual value and it shows similar convergence behavior irrespective of the use of the metrics. However, Fig. 4(b) shows a significant difference between variance of m_2 as computed with the application of HD and Euclidian metrics. In particular, using Hausdorff metric consistently results in greater variance for source parameter estimates, as compared to the Euclidean metric. This behavior can be easily understood by considering Eq. (6). As Eq. (6) shows, Hausdorff metric is constructed based on highly nonlinear operators (minimum and maximum). Hence, it is possible that two very similar image have large value of Hausdorff distance. On the other hand, Euclidean distance is a smooth function of difference between model forecast and observational data and small changes in either of these images results in small changes in the value of Euclidean distance. Consequently, these inherent properties of applied metrics affect the statistics of source parameter estimates and result in completely different statistics for the variance of source parameters.

Comparison with Minimum Variance Estimation: To compare the performance of the proposed methodology, we have also shown performance of the minimum variance estimation framework in Fig. 4. Note that in the minimum variance framework, we need to *assume* some value for the covariance of associated noise in measurement data. We have shown simulation results for the minimum variance framework in Fig. 4(a), while considering four different values for the covariance of associated noise in measurement data. As one can see, performance of the minimum variance framework crucially depends on the value of covariance matrix $\mathbf{R}$. For instance, when value of $\mathbf{R}$ is underestimated, the minimum variance framework results in inaccurate estimate of m_2. On the other hand, overestimation of the covariance matrix, $\mathbf{R}$ leads to less confident estimate, as it can be seen in Fig. 4(b).

TABLE I

OVERALL RMSE BETWEEN MEAN ESTIMATE OF THE SOURCE PARAMETERS AND THEIR ACTUAL PARAMETERS, OBTAINED BY THE PROPOSED APPROACH AND THE MINIMUM VARIANCE FRAMEWORK.

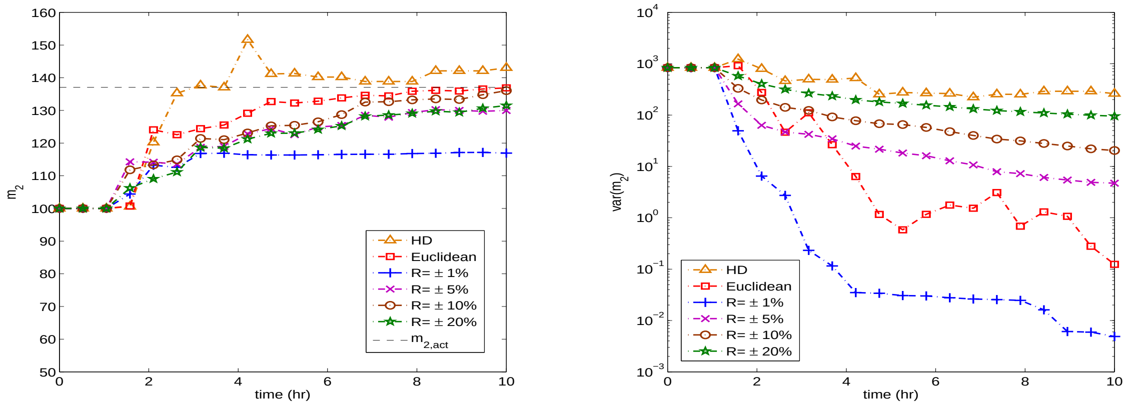
Method	RMSE
proposed approach (using Hausdorff metric)	3.1715
proposed approach (using Euclidean distance)	2.4510
minimum variance framework ($R = \pm 1\%$)	8.5793
minimum variance framework ($R = \pm 5\%$)	3.5670
minimum variance framework ($R = \pm 20\%$)	3.7781

To summarize, Table I shows the Root Mean Square Error (RMSE) between actual values of source parameters and their estimates, obtained by the proposed approach and the minimum variance framework. Comparison of RMSE for the minimum variance estimation method shows that whenever statistics of the noise is accurately known, the minimum variance framework results in very reliable results with minimal error. But whenever inaccurate estimation of noise statistics is used, the minimum variance framework results in poor performance. On the other hand, comparing the RMSE of the proposed approach with the minimum variance framework shows that its performance is comparable with performance of the minimum variance framework at its best performance, while it doesn't make use of any information regarding statistics of associated noise in measurement data.

Forecast of Concentration Field

The quality of the parameter estimates can be verified by performing a single deterministic run of SCIPUFF numerical model corresponding to posterior mean of parameters and comparing the model forecast against the observation data.

The discrepancy between forecast and true concentration fields is shown in Table II. We have also shown prior error in this Table, where prior error denotes the discrepancy between forecast concentration field, obtained using prior mean estimates of source parameters, and true concentration field. Note that corresponding error metric is used for each of the methods to evaluate the discrepancy between model forecast and true concentration field. As Table II represents, the error of model forecast,obtained by posterior mean estimate of parameters, is always less than the error of model forecast obtained by prior mean estimate of source parameters.



(a) $\mathcal{E}[m_2]$

(b) $var(m_2)$

Fig. 4. a) Mean estimate of parameter m_2 over time using Hausdorff (HD) and Euclidean metrics. Dashed black line corresponds with actual value of m_2. b) Variance of m_2 estimates over time. Mean and variance of m_2 estimate using the minimum variance framework are also illustrated in both figures.

TABLE II

COMPARISON OF THE ERROR BETWEEN THE FORECAST AND TRUE CONCENTRATION FIELDS, AT $t = 9:30\ hrs$.

	Prior Error	Posterior Error
Euclidean	1079	67
Hausdorff	0.04	0.02

VIII. CONCLUSION

In this paper, a novel probabilistic approach is developed for accurate parameter estimation and forecasting of large scale systems in the absence of any knowledge regarding the sensor error characteristics. The key idea of this approach is to minimize the expected value of the error between model forecasts and measurement data while preserving higher order prior central moments of the error distribution. Furthermore, an iterative approach is presented to convert the original nonlinear optimization problem to a convex quadratic optimization problem with guaranteed global optimal solution. The performance of the proposed approach is demonstrated by simulation of an atmospheric release incident over New York region.

Note that the proposed approach does not require any information regarding statistics of measurement. Hence, it avoids the computational and observability complexities associated with measurement noise covariance estimation methods. These properties make the proposed approach a powerful tool for applications where no accurate information regarding associated noise in measurement data is available.

REFERENCES

[1] R. K. Mehra, "On the identification of variances and adaptive kalman filtering," *Automatic Control, IEEE Transactions on*, vol. 15, no. 2, pp. 175–184, 1970.

[2] A. Bagchi, "Continuous time systems identification with unknown noise covariance," *Automatica*, vol. 11, no. 5, pp. 533–536, 1975.

[3] M. Oussalah and J. De Schutter, "Adaptive kalman filter for noise identification," in *PROCEEDINGS OF THE INTERNATIONAL SEMINAR ON MODAL ANALYSIS*, vol. 3. KU Leuven; 1998, 2001, pp. 1225–1232.

[4] V. Smídl, "On estimation of unknown disturbances of non-linear state-space model using marginalized particle filter," Citeseer, Tech. Rep., 2008.

[5] S. Saha, E. Ozkan, F. Gustafsson, and V. Smidl, "Marginalized particle filters for bayesian estimation of gaussian noise parameters," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[6] E. Özkan, V. Smídl, S. Saha, C. Lundquist, and F. Gustafsson, "Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters," *Automatica*, vol. 49, no. 6, pp. 1566–1575, 2013.

[7] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," *Automatica*, vol. 42, no. 2, pp. 303–308, 2006.

[8] M. R. Rajamani and J. B. Rawlings, "Estimation of the disturbance structure from data using semidefinite programming and optimal weighting," *Automatica*, vol. 45, no. 1, pp. 142–148, 2009.

[9] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive kalman filtering by variational bayesian approximations," *Automatic Control, IEEE Transactions on*, vol. 54, no. 3, pp. 596–600, 2009.

[10] T. T. Cai, C.-H. Zhang, H. H. Zhou *et al.*, "Optimal rates of convergence for covariance matrix estimation," *The Annals of Statistics*, vol. 38, no. 4, pp. 2118–2144, 2010.

[11] N. Jaoua, F. Septier, E. Duflos, and P. Vanheeghe, "State and impulsive time-varying measurement noise density estimation in nonlinear dynamic systems using dirichlet process mixtures," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 330–334.

[12] M. Darouach, M. Zasadzinski, and M. Boutayeb, "Extension of minimum variance estimation for systems with unknown inputs," *Automatica*, vol. 39, no. 5, pp. 867–876, 2003.

[13] E. Ozkan, S. Saha, F. Gustafsson, and V. Smidl, "Non-parametric bayesian measurement noise density estimation in non-linear filtering," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5924–5927.

[14] A. Nagavenkat, P. Singla, and T. Singh, "Conjugate unscented transform rules for uniform probability density functions." Proceedings of the American Control Conference, June 2013.

[15] V. Di Gesu and V. Starovoitov, "Distance-based functions for image comparison," *Pattern Recognition Letters*, vol. 20, no. 2, pp. 207–214, 1999.

[16] H. Zhou, M. Chen, and M. F. Webster, "Comparative evaluation of visualization and experimental results using image comparison metrics," in *Proceedings of the conference on Visualization'02*. IEEE Computer Society, 2002, pp. 315–322.

[17] L. Wang, Y. Zhang, and J. Feng, "On the euclidean distance of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1334–1339, 2005.

[18] R. Madankan, "Model-Data Fusion and Adaptive Sensing of Large Scale Systems: Applications to Atmospheric Incidents," Ph.D. dissertation, State University of New York at Buffalo, 2014.

[19] R. I. Sykes, S. F. Parker, and D. S. Henn, "SCIPUFF Version 2.2 Technical Documentation," no. 729, 2006.