Atmospheric Environment 152 (2017) 34–50

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Uncertainty characterization in the retrieval of an atmospheric point release

Sarvesh Kumar Singh*, Pramod Kumar, Grégory Turbelin, Raj Rani

Laboratoire de Mécanique et d'Energétique d'Evry, Université d'Evry-val d'Essonne, 40 rue du Pelvoux, 91020 Evry cedex, France

HIGHLIGHTS

- Uncertainty estimation is proposed in renormalization inversion technique.
- Evaluation with Fusion Field and MUST field datasets.
- The present confidence estimates are comparable with bootstrap estimates.

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form
30 November 2016
Accepted 5 December 2016
Available online 10 December 2016

Keywords:
Confidence bounds
FFT07 field experiment
MUST field experiment
Renormalization
Source reconstruction
Uncertainty

ABSTRACT

The study proposes a methodology in a recent inversion technique, called as *Renormalization*, to characterize the uncertainties in the reconstruction of a point source. The estimates are derived for measuring the inversion error, the degree of model fit towards measurements (model determination coefficient) and the confidence intervals for the retrieved point source parameters (mainly, location and strength). The inversion error is reflected through an angular estimate which measures the deviation between the measured and predicted concentrations. The uncertainty estimation methodology is evaluated for point source reconstruction studies, using real measurements from two field experiments, known as Fusion Field Trials 2007 (FFT07) in flat terrain and Mock Urban Setting Test (MUST) in urban like terrain. In FFT07 and MUST experiments, the point source location is retrieved with an average Euclidean distance of 22 m and 15 m respectively. The source strength is retrieved, on average, within a factor of 1.5 in both the datasets. The inversion error is observed as 24^{o} and 21^{o} in FFT07 and MUST experiment, respectively. The 95% confidence interval estimates show that the uncertainty in the retrieved parameters is relatively large in approximately 50% FFT07 and 30% MUST trials in spite of their closeness towards true source parameters. For a comparative analysis, the interval estimates are also compared with a more general method of uncertainty estimation, Residual Bootstrap Sampling. In most of the trials, we observed that the intervals estimates with the present method are comparable (within 10–20% variations) to bootstrap estimates. The proposed methodology provides near accurate and computationally efficient uncertainty estimates in comparison to the methods based on Hessian and sampling procedures.

1. Introduction

Inversion of atmospheric trace gases has been an active research topic in inverse modelling and data assimilation. A particular interest for emergency response applications is the estimation of origin and mass of the unknown releases, occurred accidentally or deliberately into the atmosphere, from their limited measured concentrations. This is an ill-posed inverse problem often posed in a discretized space as a non-parametric or parametric estimation problem depending on the nature of the release. In particular, retrieval of a point type source is a parametric estimation problem which refers to the estimation of its location, strength, height and time of the release. The solution to the inverse problem is derived as a conditional estimate, called a posteriori, under an optimal integration of a priori information on release parameters and a misfit (or likelihood) function. The conditional estimate can be a functional form of the source parameters. The shape or distribution of the conditional estimate represents a limited amount of

---

* Corresponding author.
E-mail address: sarveshkumar.singh@univ-evry.fr (S.K. Singh).

http://dx.doi.org/10.1016/j.atmosenv.2016.12.016

information obtained from the data and gives the full picture of uncertainty in the parameters (Duijndam, 1988). Thus, a primary uncertainty in the source reconstruction can be characterized from the resolution analysis of the conditional estimate. The resolution is referred to the state of information gained by the measurements over a priori information. The shape of the distribution around the maximum determines how well the estimate is resolved from the information available. The estimation is said to be well resolved when the conditional estimate is sharply peaked. However, a perfect resolution for the estimate can not be achieved in reality since measurements are limited and noisy (Singh et al., 2015). Therefore, an analysis of uncertainties in the estimation process is considered important.

In recent years, inversion techniques are evolved within the framework of Bayesian inference and optimization to provide an optimal and stable solution in source retrieval problems. The true statistics of the source parameters are not known and can not be determined practically since only few realization of the measurements are available in a scenario. Thus, alternative approaches are required to determine uncertainty in the retrieved parameters. More commonly, the standard deviation or standard error of the estimate is derived (or considered as an asymptotic estimate) to approximate the uncertainty in the estimated parameter. For instance, in Bayesian framework coupled with Markov Chain Monte Carlo sampling, the mean estimate of the retrieved parameters along with the posterior uncertainties are directly derivable from the sampling statistics. However, the sampling process is tedious and computationally expensive. The uncertainty measures of the parameters are often computed in the form of standard deviations, when the features of the inverse solution are approximately Gaussian. The Bayesian framework requires a priori error statistics related to the measurement errors and background errors which are not known accurately in practice (Ganesan et al., 2014). These statistics are often hypothesized based on expert's opinion and, the source retrieval and uncertainty estimates are sensitive to the prescribed error statistics due to a limited set of measurements (Desroziers and Ivanov, 2001; Desroziers et al., 2005; Winiarek et al., 2012; Berchet et al., 2013, 2015).

Contrary to sampling framework, an estimation of posterior uncertainties is not straightforward in the optimization based techniques. To analyze uncertainties, the misfit/cost function can be inspected in the directions along which the parameters are poorly resolved or, simply, an overall uncertainty bound can be determined on the parameters. A common approach is the utilization of Fisher information matrix which measures the amount of information given by an observable random variable about the unknown parameters. Under mild regularity conditions, the information matrix can be obtained as negative of the expected value of the Hessian matrix (matrix of second order derivatives) of the log-likelihood/cost function (Thacker, 1989). The second order derivative indicates the extent to which the cost function is peaked or flat. From Cramer-Rao inequality, inverse of the Fisher information matrix provides a lower bound for the error covariance of any unbiased estimator of the true source parameters. Thus, the inverse of the Hessian matrix provides an approximation of the covariance matrix of the estimated parameters for Gaussian error assumptions in non-linear models. However, such estimation requires the derivative information and are subjected to the numerical approximations of derivatives and domain discretization.

Besides these approaches, there exists an alternative Monte Carlo resampling method, called *Bootstrap*, which can be applied based on the observed data to measure the statistical properties of the estimate from the sampled distribution (Efron, 1979). Bootstrap is useful when the theoretical distribution of a statistic of interest is complex or unknown. Since the bootstrapping procedure is distribution-independent, it provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution. Bootstrap is a straightforward way to (i) control and check the stability of the results and (ii) derive estimates of standard errors and confidence intervals for estimators of complex parameters of the distribution. Although for most problems, it is impossible to know the true confidence interval, bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality. However, the methods of standard error or asymptotic variance are computationally faster than bootstrapping. When the error statistics are provided correctly, the standard error estimates and the bootstrapping error estimates may be comparable.

The present study explores an analysis of uncertainty in a recent inversion technique, called *Renormalization* (Issartel et al., 2007), which has shown its applicability to retrieve both the areal (Issartel et al., 2007) as well as point source emissions (Issartel, 2005; Sharan et al., 2009, 2012). The technique is advantageous in the sense of not utilizing any hypothetical assumptions regarding error and background statistics. A priori information about the unknown emissions apparent to the monitoring network is derived in the form of weights according to the geometry of the monitoring network. Turbelin et al. (2014); Singh et al. (2015) have shown the optimality of the resolution features of the reconstructed source. However, the inversion technique lacks an inherent methodology to derive uncertainty in the retrieved source parameters.

As a further advancement in the inversion technique, the objective here is to propose a methodology to derive uncertainty in the retrieved parameters along with their estimation. In this study, the estimates are proposed for accounting the inversion error and uncertainty in the resolution features. The quality of the reconstructed parameters are assessed in terms of their ability to describe variability in the measurements. An asymptotic estimate of a posteriori standard error in the retrieved source estimate is approximated. Further, the interval estimates are determined for the retrieved point source parameters which describe the lower and upper bound of variations in the parameters due to inherent uncertainty. The uncertainty estimation methodology is illustrated here using real diffusion experiments, namely Fusion Field Trials 2007 (FFT07) (Storwald, 2007; Singh et al., 2015) and Mock Urban Setting Test (Biltoft, 2001; Kumar et al., 2015b) in flat and obstructed terrain, respectively. The accuracy of the uncertainty estimates is compared with the bootstrap methods.

2. Inversion framework

The inversion framework relies on processing the measured atmospheric concentrations in order to infer the origin and flux of the unknown tracer emissions. This is facilitated by the use of an atmospheric transport and dispersion model which can relate information in the model space to the measurements. Typically, limited measurements are sampled at discrete locations and thus, the physical problem needs to be formulated in a discretized space. We mention that the inversion framework is discussed here for a continuous ground level emission and thus, vertical and time dimensions will be ignored in the formulations. However, the technique is generalizable for elevated and time dependent releases as well (Issartel et al., 2007; Sharan et al., 2012). Assuming that the model is linear, the source-receptor relationship can be posed in a discretized space (composed of N cells) as,

$$\boldsymbol{\mu} = \mathbf{As} + \boldsymbol{\varepsilon} \tag{1}$$

in which $\boldsymbol{\mu} \in \mathbb{R}^m$ is a vector of m concentration measurements