Atmospheric Environment 145 (2016) 201–212

Contents lists available at ScienceDirect

Atmospheric Environment

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

Application of the Approximate Bayesian Computation methods in the stochastic estimation of atmospheric contamination parameters for mobile sources

Piotr Kopka [a], Anna Wawrzynczak [a, b, *], Mieczyslaw Borysiewicz [a]

[a] National Centre for Nuclear Research, Świerk-Otwock, Poland
[b] Institute of Computer Science, Siedlce University, Poland

HIGHLIGHTS

- Application of the Approximate Bayesian Computation to the problem of the atmospheric contamination source identification.
- Stochastic reconstruction of the mobile contamination source.
- Verification of the proposed algorithm by the data from Over-Land Atmospheric Diffusion (OLAD) field tracer experiment.
- Estimation of seven parameters characterizing the contamination source i.e.: (x, y, d, v, q, ts, td).

ARTICLE INFO

Article history:
Received 16 May 2016
Received in revised form
25 August 2016
Accepted 13 September 2016
Available online 16 September 2016

Keywords:
Bayesian inference
Stochastic reconstruction
Approximate Bayesian Computation
Sequential Monte Carlo
OLAD field tracer experiment
Reconstruction of the mobile contamination source

ABSTRACT

In this paper the Bayesian methodology, known as Approximate Bayesian Computation (ABC), is applied to the problem of the atmospheric contamination source identification. The algorithm input data are online arriving concentrations of the released substance registered by the distributed sensors network. This paper presents the Sequential ABC algorithm in detail and tests its efficiency in estimation of probabilistic distributions of atmospheric release parameters of a mobile contamination source. The developed algorithms are tested using the data from Over-Land Atmospheric Diffusion (OLAD) field tracer experiment. The paper demonstrates estimation of seven parameters characterizing the contamination source, i.e.: contamination source starting position (x,y), the direction of the motion of the source (d), its velocity (v), release rate (q), start time of release (ts) and its duration (td). The online-arriving new concentrations dynamically update the probability distributions of search parameters. The atmospheric dispersion Second-order Closure Integrated PUFF (SCIPUFF) Model is used as the forward model to predict the concentrations at the sensors locations.

1. Introduction

Instantaneous releases of harmful material into the atmosphere pose a significant risk to human health and the environment. A particularly high risk is involved with the transport of such substances in tankers or by rail. In the case of a sudden atmospheric release of chemical, radioactive or biological material, emergency responders need to determine the location of the source of the dispersed substance quickly. Such information helps responders to make time-critical decisions regarding precautions for people's safety, plans for evacuation and management of emergency services. In this context, it is important to develop an emergency system capable of estimating the most probable location and other characteristics of the atmospheric contamination source. This estimation has to be done based on the concentration of the released substance registered by the sensors network.

Given the gas source and wind field we can apply the appropriate atmospheric dispersion model to calculate the expected gas concentration for any downwind location. However, concentration measurements, knowledge of the wind field and other atmospheric air parameters are not enough to determine the actual location of the release source and its parameters. This problem has no unique solution and can be considered in the probabilistic frameworks. The

* Corresponding author. National Centre for Nuclear Research, Świerk-Otwock, Poland.
E-mail address: a.wawrzynczak@ncbj.gov.pl (A. Wawrzynczak).

http://dx.doi.org/10.1016/j.atmosenv.2016.09.029
1352-2310/

issue boils down to the creation of the atmospheric dispersion model realistically reflecting the encountered situation based only on a sparse point-concentration data. This task requires specification of a set of model parameters. In the framework of Bayesian statistics, all quantities included in the mathematical model are modeled as random variables with joint probability distributions. This randomness can be interpreted as parameter variability, and is reflected in the uncertainty of the true values expressed regarding probability distributions. Bayesian methods reformulate the problem into searching for a solution based on efficient sampling of an ensemble of simulations, guided by comparisons with data.

The problem of the source term estimation was studied in literature grounded both on the deterministic and probabilistic approach. Pudykiewicz (1998) implemented an algorithm based on integrating the adjoint of a linear dispersion model backward in time to solve a reconstruction problem. Johannesson et al. (2004, 2005) introduced dynamic Bayesian modeling and the Markov Chain Monte Carlo (MCMC) sampling approaches to reconstruct a contaminant source for synthetic data. Source reconstruction at urban scales using building resolving models was reported by Keats et al. (2007) and Chow et al. (2008). Keats et al. (2007) used an adjoint representation of the source-receptor relationship. The authors used Bayesian inference methodology in conjunction with MCMC sampling procedures. This approach was validated using data from water channel simulations and a field experiment (Joint Urban 2003) in Oklahoma City. Chow et al. (2008) applied the methodology presented in Johannesson et al. (2004) to the reconstruction of the flow around an isolated building and to the reconstruction of the flow during IOP3 and IOP9 of the Joint Urban 2003 Oklahoma City experiment. In these experiments the authors found the source location ~70 m from the actual location for IOP3 (within the domain ~400 $m \times 400\ m$) while errors and other uncertainties limited the ability to pinpoint the source location for the IOP9 model. In this reconstruction the FEM3MP (Chan and Leach (2007)) model was applied to predict the atmospheric dispersion of the released substance. Delle Monache et al. (2008) applied the inversion algorithm presented in Johannesson et al. (2004) to a continental-scale accidental release of radioactive material from near Algeciras, Spain in May 1998. The forward dispersion simulation conducted in this study used the Lagrangian Operational Dispersion Integrator (LODI) model Ermak and Nasstrom (2000). The source location was reconstructed as a roughly bimodal distribution, with modes located a few dozen kilometers north of Algeciras and within ~100 km downwind of the real source location. Senocak et al. (2008) presented the feedback of the likelihood function taking into account both zero and non-zero concentration measurements that can be reported by the sensor network. The algorithm applying the Gaussian plume dispersion model was validated against the synthetic and real Copenhagen tracer experiment. In Wade and Senocak (2013) this methodology was applied to the reconstruction of the multi-source atmospheric releases. Platt and DeRiggi (2012) presented the comparative investigation of source term estimation (STE) algorithms using the FUSION Field Trials of 2007 (FFT-07) dataset (Storwold (2007)). The linear regression analysis models were applied to determine the significant factors that affected the reconstruction results and to compare STE tools such as stochastic event reconstruction tools. The goal of these evaluations was not to declare a 'winning' algorithm, but rather to examine the strengths and weaknesses of each of the proposed methodologies because different approaches may best apply to various sets of tracer release scenarios.

In our previous papers, we have tested the methodology combining the Bayesian inference with MCMC methods in the task of the dynamic data-driven contaminant source localization. The tests were based on the synthetic experiment data. In Borysiewicz et al. (2012b, 2012a) various modifications of the MCMC algorithm to estimate the probability distributions of searched parameters were examined. We have shown the advantage of the algorithms that in different ways use the probability distributions of source location parameters obtained based on available measurements to update the marginal probability distribution of the considered parameters by the newly received concentrations. We have also presented the application of the Sequential Monte Carlo (SMC) methods combined with the Bayesian inference to the problem of the localization of the atmospheric contamination source. The SMC methods were applied to the contamination source localization based both on the synthetic experiment data (Wawrzynczak et al. (2014c)) and data from the real Kori field tracer experiment (Kopka et al. (2015)). In Wawrzynczak et al. (2014a) we have proposed the methodology combining Bayesian inference with Genetic algorithm (GA) to localize atmospheric contaminant source. The effectiveness of GA was compared with the SMC algorithm (Wawrzynczak et al. (2014b)) and tested using the synthetic experiment data and data from the real Kori field tracer experiment (Wawrzynczak et al. (2016)).

In this paper for the first time, we propose the application of the approximate Bayesian computation (ABC) methods to the estimation unknown contamination source parameters. In general, the ABC methods are particularly useful when the likelihood function is analytically intractable or expensive to compute. We utilize the modification of the ABC method by using SMC to automatically, sequentially refine posterior approximations to be used to generate proposals for further steps. The original version of the ABC SMC algorithm was proposed in Sisson et al. (2007). Applications of this algorithm have been presented in a multiplicity of areas including population biology (e.g., Toni et al. (2009)), genetics (e.g., Beaumont et al. (2002)) and psychology (e.g.,Turner and Van Zandt (2012)). In recent years an increasing interest in extensions and improvements of this algorithm has been observed, as demonstrated e.g. in Lenormand et al. (2013), Filippi et al. (2013), Silk et al. (2012), Del Moral et al. (2012). One of the modifications proposed in Bonassi et al. (2015) uses data-based adaptive weights. This modification is called the ABC SMC with Adaptive Weights (ABC SMC AW) and is adopted in this paper to estimate the atmospheric contamination source parameters. The proposed algorithm is employed to localize the source of contamination based on the data from Over-Land Alongwind Dispersion (OLAD) field tracer experiment (Biltoft et al. (1999)). The reconstruction of the contamination source for this experiment is challenging because we are dealing with continuous mobile release. This characteristic implies the need to estimate numerous parameters of release such as contamination source starting position (x,y), the source motion direction (d), its velocity (v), release rate (q), start time of release (ts) and its duration (td). The release complexity entails the use of a more advanced atmospheric dispersion model than Gaussian plume model usually applied in the literature. Thus, in the reconstruction, we apply the Second-order Closure Integrated PUFF Model (SCIPUFF) (Sykes et al. (2000)). Moreover, the high frequency of data incoming from the sensors network and the large size of the search parameters space caused the need for adaptation of the effective space scanning algorithm. In the result we decided to employ the modern algorithm from the class of likelihood-free Bayesian methods (Bonassi et al. (2015)) with some extension described in detail in Section 4.

2. Description of the OLAD experiment

The Over-Land Alongwind Dispersion (OLAD) Experiment was conducted on 8–25 September 1997 at the U.S. Army Dugway Proving Ground West Desert Test Center Target S Grid (Biltoft et al. (1999), Watson (2000)). The test domain was mostly mud flat, with