The Approximate Bayesian Computation methods in the localization of the atmospheric contamination source

P Kopka¹, A Wawrzynczak¹,², M Borysiewicz¹

¹ National Centre for Nuclear Research, Andrzeja Soltana 7, 05-400 Otwock, Poland
² Institute of Computer Sciences, Siedlce University, Poland

E-mail: piotr.kopka@ncbj.gov.pl

Abstract. In many areas of application, a central problem is a solution to the inverse problem, especially estimation of the unknown model parameters to model the underlying dynamics of a physical system precisely. In this situation, the Bayesian inference is a powerful tool to combine observed data with prior knowledge to gain the probability distribution of searched parameters. We have applied the modern methodology named Sequential Approximate Bayesian Computation (S-ABC) to the problem of tracing the atmospheric contaminant source. The ABC is technique commonly used in the Bayesian analysis of complex models and dynamic system. Sequential methods can significantly increase the efficiency of the ABC. In the presented algorithm, the input data are the on-line arriving concentrations of released substance registered by distributed sensor network from OVER-LAND ATMOSPHERIC DISPERSION (OLAD) experiment. The algorithm output are the probability distributions of a contamination source parameters i.e. its particular location, release rate, speed and direction of the movement, start time and duration. The stochastic approach presented in this paper is completely general and can be used in other fields where the parameters of the model bet fitted to the observable data should be found.

1. Introduction and problem setup

Accidental atmospheric releases of hazardous material pose significant risks to human health and the environment. In this context, it is valuable to develop the emergency techniques able to identify the probable location and characteristics of the release source using only the concentrations of released substance by sensors ([1], [2]). Given concentration measurements and knowledge of the wind field and other atmospheric air parameters, finding the location of the source and its parameters is ambiguous ([3], [4]). In this paper, we present the stochastic dynamic data-driven event reconstruction model that through Sequential Approximate Bayesian Computation (S-ABC) obtains a solution to the inverse problem. This algorithm based on the successively arriving concentrations of given substance finds the most probable values of contamination source parameters. The developed algorithm is tested using the data from the Over-Land Alongwind Dispersion (OLAD) experiment [5]. The OLAD Experiment was conducted during 8-25 September 1997 at the U.S. Army Dugway Proving Ground. The experimental objective was to acquire information on alongwind dispersion of passive contaminants over distances from 2 to 20 km for evaluation of atmospheric dispersion models.
Meteorological data, including near-surface winds and temperatures at locations across the test site, vertical wind and thermodynamic profiles, were collected to supplement the tracer concentration data. In the reconstruction procedure, we used second experiment data *(Trial252–2)*. The domain area is 25 km × 25 km (see fig. 1); contamination source moves uniformly in a certain direction (see fig. 1 black arrow); the gas samplers have registered the tracer concentration in 45 one-minute time intervals since 06:45 up to 07:30; six meteorological towers registered the wind speed and its direction (see fig. 2 and fig. 1 green points).

![Figure 1](image1.png)

**Figure 1:** Domain of the OLAD Trial-252-2 experiment. The S and F denotes the starting and final point of the release, respectively. The locations of the sampling points are marked in red while the meteorological towers in green.

**Figure 2:** The wind speed and direction measured on 9 September since 6:45 up to 7:30. The weather stations numbers corresponds to the meteorological towers marked in Fig. 1 by the green dots.

### 2. Sequential Approximate Bayesian Computation

Let \( \theta \) be a parameter vector, given the prior distribution \( \pi(\theta) \). The goal of Bayesian inference is to approximate the posterior distribution, \( \pi(\theta|x) \propto \pi(x|\theta)\pi(\theta) \), where \( \pi(x|\theta) \) is the likelihood of \( \theta \) given the data \( x \). The main idea of the Approximate Bayesian Computation (ABC) methods is to accept \( \theta \) as an approximate posterior draw if its associate data \( x \) are close enough to the observed one \( x_{\text{obs}} \). Accepted parameters are sampled from \( \pi(\theta|x_{\text{obs}} < \epsilon) \), where the \( \rho(x, x_{\text{obs}}) \) is the chosen measure of discrepancy, and \( \epsilon \) is a threshold defining 'closeness margin'.

In ABC methods, Sequential Monte Carlo (SMC) [2] is used to automatically sequentially 'clean' approximation of the posterior distribution used to generate proposals for further stages. In ABC – SMC methods, the set of samples with weights, called particles sampled from the prior distribution \( \pi(\theta) \) are propagated through a sequence of intermediate posterior distributions \( \pi(\theta|x_{\text{obs}} < \epsilon_i), t = 1, \ldots, T \). Process is repeated until the set represents a sample from the target distribution, \( \pi(\theta|x_{\text{obs}} < \epsilon_T) \). In [6] authors proposes strategies called ABC SMC with Adaptive Weights. This methods includes a stage where the weights are modified according to the respective values of \( x \). The algorithm 1 presents the ABC – SMC – AW procedure [6].

After initialization of the threshold schedule, first \( N \) samples are simulated based on the predefined priori \( \pi(\theta) \) and the corresponding acceptance condition \( \rho(x, x_{\text{obs}}) < \epsilon_1 \). In time step \( t = 2 \) simple uniform weights are changed based on additional kernel \( K_{x,t}^i(x_{\text{obs}}|x_{i}^{t-1}) \) proposed in [6]. Samples, denoted by a tilde are drawn from the previous generation with probabilities \( v_i^{t-1} \). Using perturbation kernel \( K_{\theta,t}^i(\theta_j|\theta_i) \) new 'fresh' samples \( \theta_i^t \) are obtained, with the veracity of the condition \( \rho(x, x_{\text{obs}}) < \epsilon_i \). The weights are calculated according to the formula in stage (11); in stage (12) weights are normalized and the time step is increased \( t = t + 1 \). Procedure is
under the assumption, that when the sensors locations (hereafter called time steps. It is important to know that for time step registered by the sensors (1).

The reconstruction algorithm starts to search for the starting source location (x,y), speed and direction of movement (s,d), release rate (q), duration of release and start time (u,t) after 10 minutes from the first sensors measurements (east→west direction line correspond to 0° and t = 0 refers to 06:45). Consequently, the algorithm is run with obtaining the substance concentrations and continues working, as long as sensors report a new data. The scanned parameters space is \( \theta \equiv \{ x, y, d, s, q, u, t \} \). Fig. 3 presents the posterior distributions of all parameters estimated during the reconstruction employing \( S-ABC \) method after all time steps.

Algorithm 1 Approximate Bayesian Computation-Sequential Monte Carlo-Adaptive Weights

1. Initialize threshold schedule \( \epsilon_1 < \epsilon_2 < \ldots < \epsilon_T \)
2. Set \( t = 1 \)
3. \( i = 1 \) to \( N \)
   3. Simulate \( \theta_i^t \sim \pi(\theta) \) and \( x \sim \pi(x|\theta_i^t) \)
   4. Until \( p(x,x_{obs}) < \epsilon_t \)
   5. Set \( w_i^t = 1 \)
   end for
4. For \( i = 2 \) to \( T \)
   6. Compute new weights \( v_i^t = \) \( v_i^{t-1} K_{x,t}(x_{obs}|x_i^{t-1}) \)
   7. Normalizes weights \( v_i^t \)
   8. Pick \( \theta_i \) from the \( \theta_i^{t-1} \) set with probabilities \( v_i^{t-1} \)
   9. Draw \( \theta_i^t \sim K_{\theta,t}(\theta_i^t|\theta_i) \) and \( x \sim \pi(x|\theta_i^t) \)
   10. Until \( p(x,x_{obs}) < \epsilon_t \)
   11. Compute new weights as \( \pi(x(t)) \) \( v_i^t \) \( \sum_j v_j^{t-1} K_{x,t}(\theta_j^t|\theta_j^{t-1}) \)
   12. Normalizes weights \( w_i^t \) for \( i = 1 : N \)
end for

repeated until \( t \leq T \). In the presented reconstruction we assume that the substance concentration registered by the sensors \( \{ S_1, S_2, S_3, S_4, S_5 \} \) (Fig. 1) arrive subsequently in time intervals, hereafter called time steps. It is important to know that for time step \( t \) only data \( \hat{C}_i^{S_j} \) \( \hat{C}_{S_2} \) \( \ldots \) \( \hat{C}_{S_j} \) are available. Finally we have forty-five time steps (\( T = 45 \)). To get the predicted concentration a Scipuff dispersion model [7] model was run; it refers to the procedure \( x \sim \pi(x|\theta_i^t) \) in algorithm 1. To run dispersion model and obtain data \( x \) we use source parameter vector \( \theta_i^t \) and hitherto known weather data \( I \). The simulated concentrations \( C_{t_i}^{S_j} \) for time step \( t \) correspond to five sensors locations \( \{ S_1, S_3, S_3, S_4, S_5 \} \).

The choice of distance measure or summary statistics is crucial in \( ABC \). For number of sensors greater than one, overall measure of \( \rho(.,.) \) will be simple mean of distances from all sensors:

\[
\rho(x,x_{obs}) = \frac{1}{5} \sum_{j=1}^{5} \frac{1}{t} \sum_{i=1}^{t} \frac{|C_{i}^{S_j} - \hat{C}_{i}^{S_j}|}{C_{i}^{S_j} + \hat{C}_{i}^{S_j}},
\] (1)

under the assumption, that when \( C_{i}^{S_j} = \hat{C}_{i}^{S_j} = 0 \) then \( \frac{|C_{i}^{S_j} - \hat{C}_{i}^{S_j}|}{C_{i}^{S_j} + \hat{C}_{i}^{S_j}} = 0 \). Adaptive scheme for the threshold is based on the quantile of the empirical distribution of the distances between the simulated data and observations from the previous population (see [8]). We selected transition kernel \( K_{\theta,t}(\cdot|\cdot) \) to be a multivariate normal distribution. Unfortunately, in the type of inverse problems presented here the parameters often are highly correlated, and multimodality is very common. In [9] one the proposed methods is to use the multivariate normal kernel based on the \( M \) neighbors. Based on some pre-processed experiment we determined the optimal number of neighbors \( M = 30 \).

3. Results
The reconstruction algorithm starts to search for the starting source location \( (x,y) \), speed and direction of movement \( (s,d) \), release rate \( (q) \), duration of release and start time \( (u,t) \) after 10 minutes from the first sensors measurements (east→west direction line correspond to 0° and \( t = 0 \) refers to 06:45). Consequently, the algorithm is run with obtaining the substance concentrations and continues working, as long as sensors report a new data. The scanned parameters space is \( \theta \equiv \{ x, y, d, s, q, u, t \} \). Fig. 3 presents the posterior distributions of all parameters estimated during the reconstruction employing \( S-ABC \) method after all time steps.
The corresponding most probable values (MPV) of parameters are listed in Table 1. The most probable release source initial coordinates were estimated as: \( P(x = 16775m \pm 65) = 0.058 \) and \( P(y = 12565m \pm 65) = 0.033 \), which is very neighbouring to the real location (RV) i.e. \( x = 16369m \) and \( y = 12052m \). In Fig. 3 and Table 1 we can see that all parameters were estimated very close to its actual values.

This results confirm that the presented \( S - ABC \) method can be successfully applied in the problem of the localization of the atmospheric contamination source.

### Table 1: The real values (RV) and estimated during the reconstruction most probable values (MPV) of searched parameters and its probabilities.

| Parameter | RV            | MPV            | Prob.  |
|-----------|---------------|----------------|--------|
| \( x \) [m] | 16369         | 16775 ± 65     | 0.058  |
| \( y \) [m] | 12052         | 12565 ± 65     | 0.033  |
| \( d \) [°] | 40.96         | 40.75 ± 0.9    | 0.096  |
| \( s \) [m/s] | 16.8          | 22.23 ± 0.13   | 0.021  |
| \( q \) [g/min] | 1.51         | 1.90 ± 0.03    | 0.028  |
| \( u \) [min] | 8             | 8.52 ± 0.06    | 0.018  |
| \( t \) [min] | 0             | 1 ± 0.03       | 0.011  |

### 4. Acknowledgments and References

This work was supported by the Welcome Programme of the Foundation for Polish Science operated within the European Union Innovative Economy Operational Programme 2007-2013 and by the EU and MSHE grant nr P OIG.02.03.00-00-013/09.

[1] Borysiewicz M, Wawrzynczak A and Kopka P. 2012 Stochastic algorithm for estimation of the model’s unknown parameters via Bayesian inference Proceedings of the Federate Conference on Computer Science and Information Systems 501–508

[2] Wawrzynczak A, Borysiewicz M and Kopka P. 2014 Sequential Monte Carlo in Bayesian assessment of contaminant source localization based on the distributed sensors measurements Lecture Notes in Computer Sciences 8385

[3] Wade D and Senocak I. 2013 Stochastic reconstruction of multiple source atmospheric contaminant dispersion events Atmospheric Environment 74 45–51

[4] Senocak I, Hengartner N.W, Short M. B., et al. 2008 Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference, Atmospheric Environment 42 7718–7727

[5] Biltoft C A, Turley S D et al. Final test report of the over-land alongwind dispersion field trials Report No. WDTC/JCP-00/004

[6] Bonassi F V, and West M 2014 Sequential Monte Carlo with Adaptive Weights for Approximate Bayesian Computation Bayesian Analysis 10(1) 71–187

[7] Sylves R I et al. 1998 PC-SCIPUFF Version 1.2PD Technical Documentation ARAP Report 718

[8] Del Moral P, Doucet A, and Jasra A. 2012 An adaptive sequential Monte Carlo method for approximate Bayesian computation Statistics and Computing 22(5) 1009–1020

[9] Filippi S, Barnes C P, Cornebise J and Stumpf M P 2013 On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo Statistical applications in genetics and molecular biology 12(1) 87–107