Bayesian methodology in the stochastic event reconstruction problems

Anna Wawrzynczak[1], Piotr Kopka[1,2], Mieczyslaw Borysiewicz[2]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 42

[1] National Centre for Nuclear Research, Świerk-Otwock, Poland
a.wawrzynczak@ncbj.gov.pl
[2] Institute of Computer Science, Polish Academy of Sciences

Abstract

In many areas of application it is important to estimate unknown model parameters in order to model precisely the underlying dynamics of a physical system. In this context the Bayesian approach is a powerful tool to combine observed data along with prior knowledge to gain a current (probabilistic) understanding of unknown model parameters. We have applied the methodology combining Bayesian inference with Sequential Monte Carlo (SMC) to the problem of the atmospheric contaminant source localization. The algorithm input data are the on-line arriving information about concentration of given substance registered by the downwind distributed sensor's network. We have proposed the different version of the Hybrid SMC along with Markov Chain Monte Carlo (MCMC) algorithms and examined its effectiveness to estimate the probabilistic distributions of atmospheric release parameters.

Keywords: Bayesian inference, stochastic reconstruction, MCMC methods, SMC methods.

1 Introduction

Accidental atmospheric releases of hazardous material pose great risks to human health and the environment. Examples, like Chernobyl nuclear power plant accident in 1986 in Ukraine, chemical plants producing, or storing dangerous materials (e.g. Seveso disaster in 1978) or transportation accidents (bromine release on the train in Chelyabinsk in 2011), prove that it is necessary to have properly

fast response to such incidents. In this context it is valuable to develop the emergency action support system, which based on the concentration measurement of dangerous substance by the network of sensors, can identify probable location and characteristics of the release source.
It is obvious that if we are able to create the model giving the same point concentration of registered substance, as we get from the sensors' network, we could say that we understand the situation we face up. However, to create the model realistically reflecting the real situation based only on a sparse point-concentration data is not trivial. This task requires specification of set of models' parameters, which depends on the applied dispersion model's characteristics.

In general, the stated inverse problem for the dispersion of released materials in the air is ill-posed. 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. This problem has no unique solution and can be considered only in the probabilistic frameworks. In the case of gas dispersion, the unknowns to be determined are the gas source distribution of strengths and locations; given the measured gas concentrations at measurement locations for the associated wind field and other weather data (e.g. weather stability pattern). In fact, our aim is to find the source parameter's distributions that will generate predicted concentrations closest to those actually measured.

In this paper we present the developed stochastic dynamic data-driven event reconstruction model which couples data and predictive models through Bayesian inference to obtain a solution to the inverse problem i.e. based on the successively arriving information about concentration of given substance registered by distributed sensor network find the most probable source location and its strength.

2 Theoretical preliminaries

Bayes' theorem, as applied to an emergency release problem, can be stated as follows:

$$P(M|D) \propto P(D|M)P(M) \tag{1}$$

where M represents possible model configurations or parameters and D are observed data. For our problem, Bayes' theorem describes the conditional probability $P(M|D)$ of certain source parameters (model configuration M) given observed measurements of concentration at sensor locations (D). This conditional probability $P(M|D)$ is also known as the posterior distribution and is related to the probability of the data conforming to a given model configuration $P(D|M)$, and to the possible model configurations $P(M)$, before taking into account the sensors' measurements. The probability $P(D|M)$, for fixed D, is called the likelihood function, while $P(M)$ is the prior distribution[1].
Value of likelihood for a sample is computed by running a forward dispersion model with the given source parameters M. To achieve the rapid-response event

reconstructions and limit the computational time we have adopted the fast-running Gaussian plume dispersion model [2] as the forward dispersion model. The model predicted concentrations M in the points of sensors location are compared with actual data D. The closer the predicted values are to the measured ones, the higher is the likelihood of the sampled source parameters. This function is taken as:

$$ln[P(D|M)] = ln[\lambda(M)] = -\frac{\sum_{i=1}^{N}[log(C_i^M) - log(C_i^E)]^2}{2\sigma_{rel}^2} \quad (2)$$

where λ is the likelihood function, C_i^M are the predicted by the forward model concentrations at the sensor locations i, C_i^E are the sensor measurements, σ_{rel}^2 is the standard deviation of the combined forward model and measurement errors, N is the number of sensors.

We use a sampling procedure with the Metropolis-Hastings algorithm to obtain the posterior distribution $P(M|D)$ of the source term parameters given the concentration measurements at sensor locations. This way we completely replace the Bayesian formulation with a stochastic sampling procedure to explore the model parameters' space and to obtain a probability distribution for the source location [3, 4]. The scanned model's parameters' space is

$$M \equiv M(x, y, q, \zeta_1, \zeta_2) \quad (3)$$

where x and y are spatial location of the release, q release rate and ζ_1, ζ_2 are stochastic terms in the turbulent diffusion parameters.

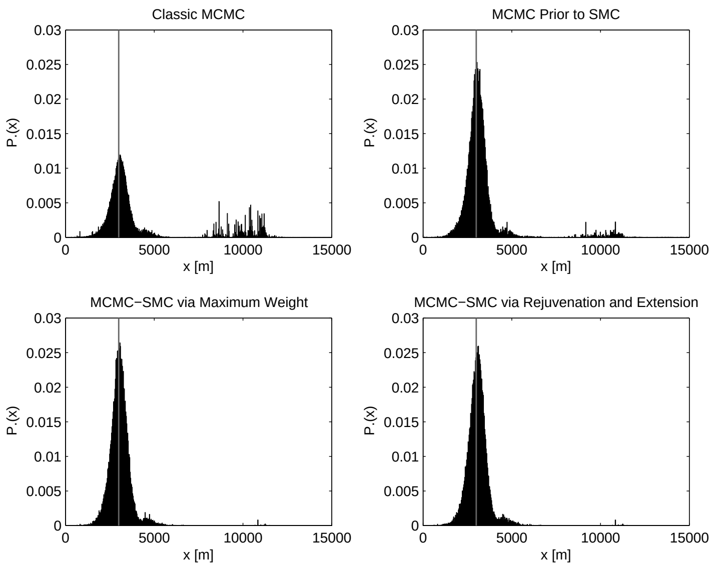

Figure 1: Posterior distribution as inferred by the Bayesian event reconstruction for all applied algorithms for x parameter. Vertical lines represent the target x value.

3 Summary and Results

In this paper we examine 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. We present the possibility to connect MCMC and SMC to provide additional benefit in the process of event reconstruction. Based on the synthetic release experiment we have proposed and tested various version of the Hybrid SMC with MCMC algorithms i.e. classic MCMC, MCMC prior to SMC, MCMC prior to SMC via Rejuvenation and Extension, MCMC prior to SMC via Maximal Weights in effectiveness to estimate the probabilistic distributions of searched parameters. We have shown the advantage of the algorithms that in different ways use the source location parameters probability distributions obtained basing on available measurements to update the marginal probability distribution. As the most effective we pointed the modifications of MCMC prior to SMC (see Figure 1).

References

[1] Gilks, W., Richardson, S., Spiegelhalter, D. **: Markov Chain Monte Carlo in Practice**, 1996; Chapman & Hall/CRC.

[2] Panofsky, H.A., Dutton, J.A., (1984).**: Atmospheric Turbulence**, 1984; John Wiley.

[3] Borysiewicz, M., Wawrzynczak A., Kopka P.**Stochastic algorithm for estimation of the model's unknown parameters via Bayesian inference**, *Proceedings of the Federated Conference on Computer Science and Information Systems*; 2012; pp. 501-508.

[4] Borysiewicz, M., Wawrzynczak A., Kopka P. **S Bayesian-Based Methods for the Estimation of the Unknown Model's Parameters in the Case of the Localization of the Atmospheric Contamination Source**, *Foundations of Computing and Decision Sciences*; 2012; 37(4); pp.253-270.