

# Approximate Bayesian Computation for Source Term Estimation

R. O. Lane, M. Briers, K. Copsey

{rlane1, mbriers, kkcopsey}@QinetiQ.com

QinetiQ, Malvern Technology Centre, St. Andrews Road, Great Malvern, UK, WR14 3PS

## Abstract

Bayesian inference is a vital tool for consistent manipulation of the uncertainty that is present in many military scenarios. However, in some highly complex environments, it is hard to write down an analytic form for the likelihood function that underlies Bayesian inference. Approximate Bayesian computation (ABC) algorithms address this difficulty by enabling one to proceed without analytically specifying or evaluating the likelihood distribution. This is achieved through the use of computer simulation models that stochastically simulate measurements for a given set of parameter values. This paper gives an overview of standard ABC methods such as rejection and Markov chain Monte Carlo (MCMC) sampling. It then goes on to discuss ABC versions of sequential Monte Carlo (SMC) samplers, which are the recently-developed next-generation particle filters for Bayesian sampling. SMC Samplers have properties that make them highly suitable for complex estimation and inference problems in the presence of uncertainty. This includes an ability to efficiently explore multi-modal distributions.

One application of SMC-ABC algorithms is source term estimation for chemical, biological, radiological and nuclear (CBRN) defence when the number of releases is unknown. A proof-of-principle study has been conducted using a bar-sensor model and a Gaussian dispersion model for agent behaviour, including the effect of wind. The outcome is that the algorithms are able to estimate model parameters to a reasonable degree of accuracy, but there is some ambiguity between the location and time of releases due to wind effects.

## 1. Introduction

In many scientific problems, data  $x$  is distributed according to the probability distribution  $p(x|\theta)$ , where  $\theta$  is a set of parameters that describes the data model. We are often interested in making inference on those parameters through the posterior distribution  $p(\theta|x)$ . This distribution may be calculated using Bayes' theorem

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{p(x)} = \frac{p(x | \theta)p(\theta)}{\int p(x | \theta)p(\theta)d\theta},$$

where  $p(\theta)$  is the prior distribution of  $\theta$ . Usually the functional form of the likelihood  $p(x|\theta)$  is known and may be evaluated directly using appropriate values of  $x$  and  $\theta$ . However, in some situations the likelihood is either analytically or computationally intractable (Bortot *et al.*, 2007). If that is the case then analytic Bayesian inference cannot be used. Furthermore, neither can numeric Bayesian sampling methods be used since they require knowledge or evaluation of the likelihood function. Typical numerical methods include rejection sampling, Markov chain Monte Carlo (MCMC) (Robert & Casella, 1999), particle filters (Arulampalam *et al.*, 2002) and sequential Monte Carlo (SMC) samplers (Del Moral *et al.*, 2006).

The problem of intractable likelihood calculations can be avoided if it is possible to use model simulation in place of likelihood evaluation. Such techniques are referred to as approximate Bayesian computation (ABC) and were originally proposed in the statistical genetics literature by Beaumont *et al.* (2002). In complex environments, it is generally much

easier to sample from simulated models than it is to write down or compute the likelihood function. ABC algorithms therefore open the possibility of dealing with highly complex situations, which were hitherto considered analytically or computationally intractable for standard Bayesian techniques. We now describe two initial ABC algorithms from the literature, rejection ABC sampling and MCMC ABC.

In standard rejection sampling, where the likelihood is known and the data  $D$  is discrete, the algorithm generates  $\theta$  from  $p(\theta)$  and accepts  $\theta$  with probability  $p(D|\theta)$ . A continuous-data version of rejection sampling is given by Chen (2005). If the likelihood  $p(D|\theta)$  cannot be calculated then a likelihood-free version of the algorithm is required. In this version  $\theta$  is generated from  $p(\theta)$ ,  $D'$  is simulated from the model using the parameter  $\theta$ , and  $\theta$  is accepted if  $D'=D$ . A problem with this algorithm is that when there are lots of data, many simulated data sets will be rejected. Another disadvantage is that this method can only be applied to discrete data because the simulated data must exactly equal the measured data, which is not possible with continuous data. These are strong motivations for some form of approximation and are the reason ABC was first introduced.

The simplest ABC algorithm is often referred to as the ‘‘vanilla’’ algorithm (Butler *et al.*, 2007). It is a likelihood-free rejection sampling algorithm with data approximation. For each sample, the following steps are followed: generate  $\theta$  from  $p(\theta)$ ; simulate  $D'$  from the model using the parameter  $\theta$ , accept  $\theta$  if  $\rho(D, D') \leq \varepsilon$ . In this algorithm  $\rho$  is some form of distance metric and  $\varepsilon$  is a suitable tolerance threshold. The choice of  $\varepsilon$  is a trade-off between accuracy and computational effort. If  $\varepsilon=0$  then no approximation is involved and we have basic likelihood-free rejection sampling.

In the vanilla ABC algorithm the rejection rate will be high if the data is high-dimensional. It is possible to reduce the dimensionality of the problem by comparing lower-dimensional summary statistics of the data instead of the data themselves. If the statistics are sufficient then this is equivalent to the data approximation method. However, even if the statistics are not sufficient this method should provide a reasonable approximation to the posterior distribution, provided the statistics are near-sufficient and the error threshold  $\varepsilon$  is small.

The above rejection algorithms are easy to implement, amenable to parallelisation, and produce independent samples. However, the rejection rate is likely to be high when the prior and posterior distributions are not very similar. The ABC rejection sampling algorithm suffers from all the disadvantages of standard rejection sampling, as well as the extra approximations introduced by ABC.

The inefficiency of rejection sampling may be avoided by using MCMC methods such as the Metropolis-Hastings algorithm (Chib & Greenberg, 1995). A likelihood-free version of MCMC was proposed by Marjoram *et al.* (2003). The algorithm proceeds in much the same way as rejection ABC but proposes changes in  $\theta$  according to a proposal distribution and accepts or rejects new values depending on the ABC distance metric and the standard MCMC acceptance probability with the likelihood functions removed. This process is more efficient than rejection ABC but still has inefficiencies associated with the chain getting stuck in tails of the distribution where the simulated data is less likely to match the measured data. Since SMC samplers have advantages over MCMC samplers, such as more efficient exploration of complex multimodal distributions, reduced computational cost, and no need to develop and assess convergence diagnostics, the combination of SMC samplers and ABC has been investigated here.

## 2. Sequential Monte Carlo ABC

Particle filters have achieved much success in dynamic estimation problems (Arulampalam *et al.*, 2002). A key aspect to these problems is that the state space dimensions grow with time. SMC samplers can be considered to be the next generation of particle filters, allowing a

similar underlying methodology to be applied to problems with static parameters, where the state space remains fixed with time. We now outline the SMC sampler methodology. The reader is referred to Peters (2005) and Del Moral *et al.* (2006) for full treatments with justifications.

Suppose we wish to draw samples from the distribution  $\pi(\theta)$  on the parameter space  $\theta \in \mathfrak{R}^n$ . The first step in defining an SMC sampler is to define a sequence of target distributions  $\pi_1(\theta), \dots, \pi_T(\theta)$ , all on the same parameter space, with the properties that:  $\pi_T(\theta) = \pi(\theta)$ ;  $\pi_1(\theta)$  is an easy to sample from distribution; and the difference between distributions  $\pi_i(\theta)$  and  $\pi_{i+1}(\theta)$  is not too large. The basic premise behind the SMC sampler is to draw samples from  $\pi(\theta)$  by sampling first from the easy-to-sample initial distribution  $\pi_1(\theta)$ , and then moving towards the distribution  $\pi(\theta)$  in a sequence of  $T$  easy steps by adjusting and weighting the samples from step to step.

Trying to apply standard importance sampling to such a sequence of distributions requires the evaluation of typically intractable integrals. To avoid this, SMC samplers augment the target distributions by defining a new distribution:

$$\tilde{\pi}_t(\theta_1, \dots, \theta_t) = \pi_t(\theta_t) \prod_{s=1}^{t-1} L_s(\theta_{s+1}, \theta_s), \quad t = 1, \dots, T.$$

Here, each  $\theta_s \in \mathfrak{R}^n$ , and the  $L_s(\theta_{s+1}, \theta_s)$  are backwards kernels, generally termed L-kernels. From one distribution to the next, the parameter space increases by  $\mathfrak{R}^n$ . The original target distributions  $\pi_t(\theta_t)$  can be obtained by marginalising over (equivalent to ignoring) the augmented part of the distributions. The significance of the augmentation is that particle filtering ideas can be used to draw samples from the increasing-dimension distributions  $\tilde{\pi}_t(\theta_1, \dots, \theta_t)$  in sequence, starting from  $\tilde{\pi}_1(\theta_1)$  and ending at the distribution  $\tilde{\pi}_T(\theta_1, \dots, \theta_T)$ .

ABC-SMC samplers combine the ideas of ABC and SMC samplers. In this paper we investigate the use of three types of sampler: fixed dimension ABC-SMC samplers, trans-dimensional ABC-SMC samplers and an ABC population Monte Carlo (PMC) scheme. An overview of the algorithms is given here; further details are given by Briers *et al.* (2009).

The fixed dimension ABC-SMC sampler embeds standard ABC-MCMC steps into the SMC sampler framework. Sample weights are calculated using the ABC simulated-measurement augmented-target formulation of Del Moral *et al.* (2008). The sequence of underlying SMC target distributions  $\pi_1(\theta), \dots, \pi_T(\theta)$  are defined so that  $\pi_i(\theta)$  is the posterior distribution with distance metric tolerance specified to be  $\varepsilon_i > 0$ . The  $\varepsilon_i > 0$  are defined as a strictly decreasing sequence  $\varepsilon_1 > \varepsilon_2 > \dots > \varepsilon_T > 0$  and  $\varepsilon_T$  is set sufficiently small that the ABC posterior distribution defined by it should be a good approximation to the actual posterior distribution. Performance of the algorithm is enhanced by simulating multiple ABC measurements and carrying out several MCMC steps per distribution in the SMC sequence.

The ABC-SMC sampler described above concerns parameter estimation when the model order is known. However, in many military problems the model order is unknown and needs to be determined at the same time as estimating the parameters. To address this, a joint model-order and parameter-estimation version of the ABC-SMC algorithm has been developed. The approach is termed trans-dimensional ABC-SMC. In this algorithm the MCMC forward kernel within our ABC-SMC sampler is replaced with a reversible jump MCMC kernel (Green, 1995).

The third ABC-SMC sampler is the ABC-PMC approach proposed by Beaumont *et al.* (2009). For each distance metric tolerance, this algorithm picks samples at random according to their weight, generates a new sample from a proposal distribution based on the old sample, and simulates a measurement from the model given the new sample. This process is repeated until the simulated-to-measured-data distance metric tolerance is passed for the required

number of samples. Weights are recalculated based on the proposal and the desired distribution, and a decreasing tolerance schedule is used. ABC-PMC is an  $O(N^2)$  algorithm, where  $N$  is the number of particles, whereas the preceding ABC-SMC algorithms have lower computational cost at  $O(N)$ . In our experiments, the ABC-PMC approach has tended to have a high computational load, although it can give good estimation performance. Model-order selection versions of ABC-PMC have been proposed by Grelaud *et al.* (2008) and Toni *et al.* (2008) but these are not considered here due to space limitations.

### 3. Source Term Estimation

In this section we demonstrate the proof-of-principle use of ABC-SMC samplers for source term estimation (STE). STE refers to the use of sensor measurements to estimate the locations, times, amounts and types of material in multiple chemical, biological, radiological and nuclear (CBRN) releases (Green, 2008b; Robins, 2005; Robins, 2009). Algorithms have been developed here for the case where the number of releases is unknown, addressing a previously unsolved problem.

The proof-of-principle experiment is conducted using synthetic data. The first aspect affecting the sensor measurements concerns the dispersion of the released material. This can be a complex stochastic process that depends on the atmospheric conditions, such as wind profiles, and the environment, including buildings and other obstacles. For the purposes of this study we consider a simple Gaussian dispersion model (Green, 2008b) where the concentration is specified to be given by a Gaussian distribution clipped at a concentration of zero. The Gaussian distribution is defined by a mean and variance, both of which vary with the location and time. The key aspect of this dispersion model is the specification of the Gaussian means and variances according to a simple Gaussian puff model, obtained from Green (2008a). This model takes as inputs: release parameters (mass, location, time); Cartesian-coordinate velocity of the wind; a background concentration; and additional dispersion factors. An example of the model output for a rectangular array of measurement points is shown in Figure 1. The major parameters of a specific simulation involving three releases are shown in Table 1 using arbitrary units. The wind speed was -11 in the  $x$  direction.

The second aspect of the sensor measurement distribution is the sensor model. The sensor model relates the actual agent concentration (as shown in Figure 1) to the measurements made by the sensors. Following Green (2008b) we consider arrays of man-portable chemical agent detector (MCAD) bar sensors. These sensors output a bar reading that is related to the concentration of a CBRN agent in the atmosphere over a given time interval. The reading is a non-linear function of the concentration, and for a given concentration more than one bar reading is possible, with the different possibilities modeled by known probability distributions.

The sensor grid used to test the ABC algorithms consists of a 4-by-4 array of sensors, which is a sparser sampling than that shown in Figure 1. Normal distributions were used as prior distributions for the time, location and mass parameters. A uniform distribution on [1,3] was used for the number of releases. The ABC comparison function was the error between measured and simulated agent concentrations averaged over all measurement times for all sensors. The ABC tolerances were set to a linearly descending sequence. More details of the experiment are given by Briers *et al.* (2009).

Results of applying the three ABC parameter estimation algorithms are summarized in Table 1. The table records the mean parameter values for each estimated release. It is seen that generally the estimated parameters are close to the true values, with some variations, and the trans-dimensional algorithm has correctly estimated the number of releases. However, for all algorithms, a noticeable bias is that the time of the first release is generally estimated to occur earlier than the true time and at an  $x$  coordinate to the right of the true position. This is an effect caused by the wind: by the time of the true release, the estimated release has drifted via dispersion to the vicinity of the true release location. This ambiguity between time and

location is an inherent part of the problem and can only be removed through the use of a greater number of more sensitive sensors. In comparing the three algorithms, the fixed-dimension ABC was best at estimating the mass of each release. The results for the other parameters showed no strong pattern, with each algorithm having best performance only for a subset of the parameters. A graphical representation of estimated parameter distributions for the fixed-dimension ABC-SMC algorithm is shown in Figures 2, 3 and 4. Clustering near the true values of the estimated location and mass of releases is evident from Figures 2 and 4. In Figure 3 the effect of wind has shifted the estimated release times back in time from the true values.

Algorithm	No. of releases	Time	X coord.	Y coord.	Mass
Ground-truth	3	0	40	15	250
		0	30	5	300
		0.8	40	15	200
Fixed-dimension ABC-SMC	Fixed = 3	-0.43	45.0	15.7	205.7
		-0.21	32.0	5.2	263.0
		0.1	45.9	16.0	198.7
Trans-dimensional ABC-SMC	Most likely = 3	-0.97	53.4	16.7	302.9
		-0.03	30.4	5.3	261.6
		-0.36	49.6	15.2	162.5
ABC-PMC	Fixed = 3	-0.41	44.7	15.5	196.0
		-0.26	33.0	5.2	203.5
		0.33	44.1	15.2	196.9

TABLE 1: Source release details and parameter estimates

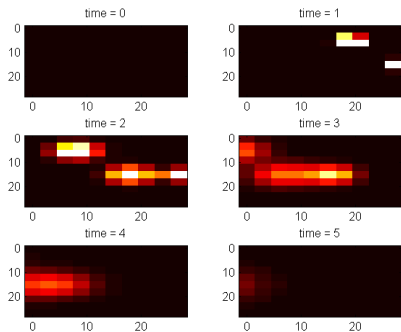


FIGURE 1: Example multi-release dispersal density plot for a sensor grid at six different times.

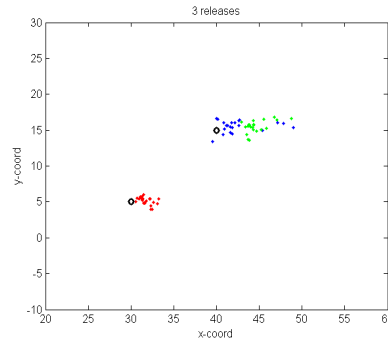


FIGURE 2: Scatter-plot of estimated release coordinates for ABC-SMC. Ground-truth releases are marked by black rings

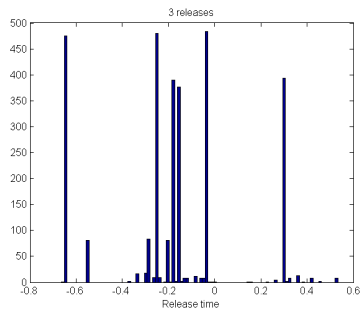


FIGURE 3: Release-time histogram for fixed-dimension ABC-SMC

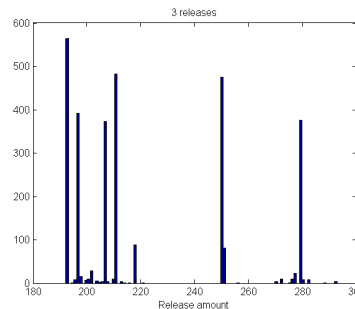


FIGURE 4: Release-mass histogram for fixed-dimension ABC-SMC

#### 4. Conclusions

We have introduced sampling algorithms that combine the properties of approximate Bayesian computation and sequential Monte Carlo methods. Such algorithms are useful in parameter estimation problems where it is not possible to specify the Bayesian likelihood function analytically. A further step has been to introduce a trans-dimensional version of ABC-SMC algorithms where the number of model parameters is not required to be known. The algorithms have been applied to a source term estimation problem in CBRN defence. The results show some ambiguity in the location and time of releases due to wind effects, which is inherent to the problem being solved. However, despite these difficulties, the algorithms are able to estimate model parameters to a reasonable degree of accuracy.

#### REFERENCES

- Arulampalam, S., Maskell, S., Gordon, N. & Clapp, T. 2002. A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking. *IEEE Trans. on Sig. Proc.*, 50(2):174-188.
- Beaumont, M.A., Zhang, W., & Balding, D.J. 2002. Approximate Bayesian computation in population genetics, *Genetics*, 162(4):2025-2035, December 2002.
- Beaumont, M.A., Corneut, J.-M., Marin, J.-M., & Robert, C.P. 2009. Adaptivity for ABC algorithms: the ABC-PMC scheme, to appear in *Biometrika*.
- Briers, M., Copsey, K. & Lane, R.O. 2009. Approximate Bayesian Computation for Scalable Manipulation of Belief: Final Report, *QinetiQ Report: QinetiQ/09/00507*, March 2009.
- Bortot, P., Coles, S.G., & Sisson, S.A. 2007. Inference for stereological extremes, *Journal of the American Statistical Association*, 102(477):84-92, March 2007.
- Butler, A., Glasbey, C. & Wanless, S. 2007. Approximate Bayesian inference in a latent Gaussian model, *Biomathematics and Statistics Scotland Technical Report*.
- Chen, Y. 2005. Another look at rejection sampling through importance sampling, *Statistics & Probability Letters*, 72:277-283, May 2005.
- Chib, S. & Greenberg, E. 1995. Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, 49(4):327-335.
- Green, P.J. 1995. Reversible jump MCMC computation and Bayesian model determination, *Biometrika*, 82:711-732.
- Green, N. 2008a. Simple Gaussian Dispersion Model, *Dstl Technical Note*, September 2008.
- Green, N. 2008b. Description of Chemical Sensor Likelihoods for Source Term Estimation and Plume Tracking, *Dstl Technical Note*, December 2008.
- Grelaud, A., Robert, C. P. & Marin, J.-M. 2009. ABC methods for model choice in Gibbs random fields, *C. R. Acad. Sci. Paris, Ser. I 347*, pp. 205-210.
- Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. 2003. Markov chain Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci. USA*, 100(26):15324-15328, December 2003.
- Del Moral, P., Doucet A., & Jasra, A. 2006. Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411-436.
- Del Moral, P., Doucet A., & Jasra, A. 2008. Adaptive Sequential Monte Carlo Samplers, *Journal Submission*, July 2008.
- Peters, G. 2005. Topics in Sequential Monte Carlo Samplers, *MSc Dissertation*, Cambridge University, January 2005.
- Robert, C.P. & Casella, G. 1999. *Monte Carlo Statistical Methods*. Springer.
- Robins, P. & Thomas, P. 2005. Non-linear Bayesian CBRN source term estimation, *Proc. of 2005 8th IEEE International Conference on Information Fusion*, Vol. 2.
- Robins, P., Rapley, V.E. & Green N. 2009. Real-Time Sequential Inference of Static Parameters with Expensive Likelihood Complications, *To appear in J. Royal Stat. Soc. Series C*.
- Sisson, S.A., Peters, G.W., Fan, Y., & Briers, M. 2008. Likelihood-free samplers, *Journal Submission*, December 2008.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M.P.H. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *J. Royal Soc. Interface*, 6(31):187-202, February 2009.