Adaptive Importance Sampling in Particle Filtering

Václav Šmídl
Institute of Information Theory and Automation,
Czech Republic
email: smidl@utia.cas.cz

Radek Hofman
Institute of Information Theory and Automation,
Czech Republic
email: hofman@utia.cas.cz

***Abstract*—Computational efficiency of the particle filter, as a method based on importance sampling, depends on the choice of the proposal density. Various default schemes, such as the bootstrap proposal, can be very inefficient in demanding applications. Adaptive particle filtering is a general class of algorithms that adapt the proposal function using the observed data. Adaptive importance sampling is a technique based on parametrization of the proposal and recursive estimation of the parameters. In this paper, we investigate the use of the adaptive importance sampling in the context of particle filtering. Specifically, we propose and test several options of parameter initialization and particle association. The technique is applied in a demanding scenario of tracking an atmospheric release of radiation. In this scenario, the likelihood of the observations is rather sharp and its evaluation is computationally expensive. Hence, the overhead of the adaptation procedure is negligible and the proposed adaptive technique clearly improves over non-adaptive methods.**

I. Introduction

The particle filter [1] is a popular method used in object tracking and state estimation problems in general. It is based on the importance sampling idea, where the common choice of the proposal density is the bootstrap proposal (i.e. the proposal is the transition density). The simplicity of the bootstrap proposal and parallel nature of the filter makes it suitable for implementation on parallel architectures [**?**], or field programmable logic arrays (FPGA) [**?**]. However, the bootstrap approach is inefficient especially for problems with sharp likelihood function. More efficient filters can be obtained when the proposal function is *adapted* in the sense of [2], i.e. it takes into account the latest observation.

Many techniques of adaptive particle filters has been proposed, e.g. [3], [4], [5]. However, their implementation in hardware is difficult since their operations can not be run in parallel or in a pipeline. Therefore, we seek an adaptation procedure that can be run as recursively as possible, which would allow efficient use of the pipeline in an FPGA.

In this paper, we investigate the use of adaptive importance sampling (AIS) approach [6]. It is based on a parametric form of the proposal density, the parameters of which are estimated from previously drawn particles. The key feature of this approach is its ability to update the parameters recursively for each realization of the particle. The problems that needs to be addressed in the context of particle filtering are initialization of the parameter statistics and the association of the particles from the previous step. The latter can be elegantly solved by the marginal particle filter approach [7], the former problem is more problem specific. Poor choice of the initial statistics may have significant impact on the filter performance and potentially lead to degenerate proposal. We use some concepts from information geometry [8] to prevent degeneration of the adaptation process.

Performance of the proposed AIS-PF algorithm is demonstrated on the problem of tracking of atmospheric pollution [**?**], with emphasis on tracking of radioactive pollutant [9]. In a realistic setup of the current measurement devices of the nuclear power plant, the observations are available only in the form of integrated gamma dose rates. This yields a model with a sharp likelihood and a transition model with high variance. These features make this application an excellent area where sophisticated adaptive proposals have significant impact. We show that the AIS-PF algorithms significantly increases computational efficiency of the estimation procedure.

II. Adaptive Importance Sampling

The idea of importance sampling is to approximate an unknown probability density function $p(\mathbf{x})$ by drawing samples from a proposal function $q(\mathbf{x})$ such that

$$p(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})} g(\mathbf{x}) \dot{\propto} \sum_{i=1}^{n} \frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})} \delta(\mathbf{x}-\mathbf{x}^{(i)}) = \sum_{i=1}^{n} w^{(i)} \delta(\mathbf{x}-\mathbf{x}^{(i)}), \tag{1}$$

where $\mathbf{x}^{(i)}$ are i.i.d. samples from $q(\mathbf{x})$, and $\tilde{w}^{(i)} \propto w^{(i)} = \frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})}$ are their associated weights. Symbol $\propto$ is used to denote equality up to normalization, i.e. the weights $\tilde{w}$ are multiplied by a constant such that $\sum_{i=1}^{n} \tilde{w}^{(i)} = 1$. The main advantage of this approach is that under mild conditions it converges to the unknown function $p(\mathbf{x})$ with probability one. However, the rate of convergence heavily depends on the chosen proposal, which is difficult to choose for a general problem.

An elegant solution is to choose the proposal function from a parametric family, $q(\mathbf{x}|\boldsymbol{\theta})$, and iteratively estimate the vector of parameters $\boldsymbol{\theta}$ from the sampled particles [**?**]. The approach was further elaborated by [6] to show

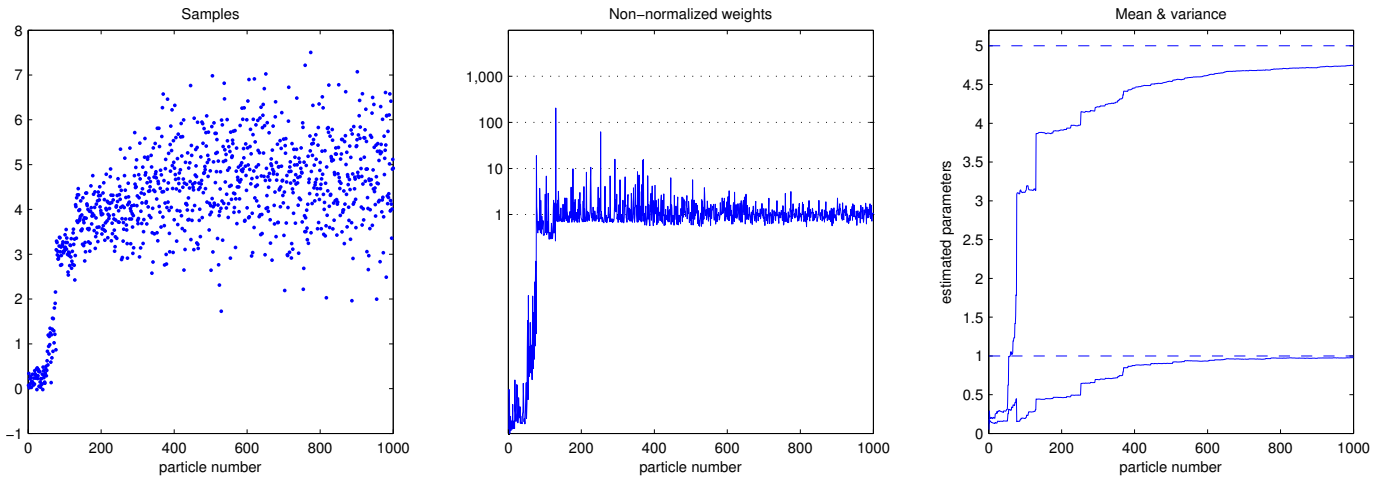

Figure 1. Illustration of the AIS algorithm for Example 1. Sequential generation of 1000 particles and recursive estimation of the parametric proposal. **Left**: the sequence of simulated particles, $\mathbf{x}^{(i)}$; **middle**: non-normalized weights $w^{(i)}$; **right**: evolution of the estimated parameters of the proposal $\hat{\mu}$ and $\hat{\sigma}$.

Algorithm 1 Adaptive Importance Sampling (AIS) algorithm.

Init: set batch counter $k = 0$, initial statistics $\mathbf{V}^{[0]}, \nu^{[0]}$, and number of samples in the batch n_k.

Repeat until convergence:

1) Compute estimate of the expectation parameter $\hat{\boldsymbol{\psi}}^{[k]} = \mathbf{V}^{[k]}/\nu^{[k]}$, and update the parameter estimates $\hat{\boldsymbol{\theta}}^{[k]}$.
2) Draw n_k samples from $q(\mathbf{x}|\hat{\boldsymbol{\theta}}^{[k]})$, and compute their non-normalized weights

$$w^{(i)} = \frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)}|\hat{\boldsymbol{\theta}}^{[k]})}, i = 1, \ldots, n_k.$$

3) Set batch counter $k = k + 1$ and evaluate statistics $\mathbf{T}^{[k]}, \nu^{[k]}$,

$$\begin{aligned} \mathbf{V}^{[k]} &= \mathbf{V}^{[k-1]} + \sum_{i=1}^{n_k} w^{(i)} \mathbf{T}(\mathbf{x}^{(i)}), \qquad (2) \\ \nu^{[k]} &= \nu^{[k-1]} + \sum_{i=1}^{n_k} w^{(i)}. \end{aligned}$$

4) Evaluate convergence criterion.

consistency of the adaptive importance sampling (AIS) scheme. The original algorithms is formalized for general statistics, however, we will present it in the special case of exponential families for clarity. The exponential family is defined as

$$q(\mathbf{x}|\theta) = h(\mathbf{x}) \exp\Big(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(\mathbf{x}) - A(\boldsymbol{\theta})\Big), \qquad (3)$$

$$p(\theta) = \exp\Big(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{V} - \nu A(\boldsymbol{\theta})\Big), \qquad (4)$$

where $\boldsymbol{\eta}(\boldsymbol{\theta})$ is known as the natural parameter and $\mathbf{T}(\mathbf{x})$ as the sufficient statistics. The likelihood function (3) is conjugate with the prior (4) with statistics $\mathbf{V}, \nu$. The expectation parameter of the exponential family is $\boldsymbol{\psi} = \mathsf{E}(\mathbf{T}(\mathbf{x})|\theta]$. Its significance is in the fact that it readily provides a transformation from sufficient statistics to maximum likelihood estimates [8].

The idea of the AIS is to draw samples in batches, indexed by $k = 1, \ldots, K$, of size n_k and after each batch compute the empirical sufficient statistics, $\mathbf{T}(\mathbf{x})$, of the proposal distribution. The statistics are used to update the estimate of parameter $\hat{\boldsymbol{\theta}}$ which is used for generation of the next batch. The full algorithm is adapted from [6] in Algorithm 1. An example for better intuition of the approach is now provided.

Example 1. Consider a proposal function in the form of Gaussian $q(x|\boldsymbol{\theta}) = \mathcal{N}(\mu, \sigma^2)$. It belongs to the exponential family (3) with assignments

$$\boldsymbol{\psi} = [\mu, \sigma^2 + \mu^2], \quad \mathbf{T}(x) = [x, x^2]$$

Thus, the update equation (2) computes weighted average of the particles and their squares using the non-normalized weights. The expectation parameter $\hat{\phi}$ is then only normalization of the statistics by the sum of the weights. The commonly used parameters of the Gaussian are $\hat{\mu} = \hat{\theta}_1$, $\hat{\sigma} = \sqrt{\psi_2 - \psi_1^2}$. The exponential family formulation is thus a generalization of the common approach of moment matching.

A simulation of the proposal from Example 1 with target distribution $p(x) = \mathcal{N}(5, 1)$ was estimated by the AIS (Algorithm 1) using $n_k = 1$ and $K = 1000$ with initial estimate $\mu^{[0]} = 0$, $\sigma^{[0]} = 0.4$. The results of the AIS are displayed in Figure 1. The initial variance of the proposal was chosen intentionally low to achieve slow convergence. Moreover, a realization with particularly slow convergence was chosen to demonstrate a potential drawback of the

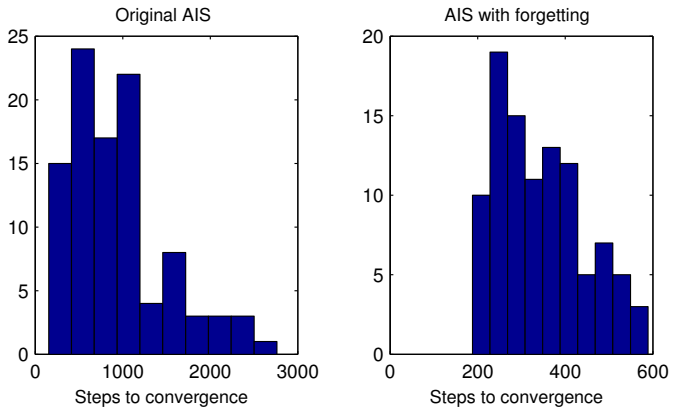

Figure 2. Histogram of convergence of two variants of the AIS algorithm for 100 runs. **Left**: original AIS. **Right**: AIS with forgetting $\lambda = 0.95$.

method. Note that the statistics (2) are weighted by non-normalized weights which can be rather high, e.g. around 200 in Figure 1, middle. The likelihood of the particle associated with such a weight is still very low. The significance of the contribution of this particle to the statistics creates a bias and slows down the convergence of the algorithm. With parameters of the proposal converging to the target density, the weights converge to one. In the long run, their contribution outweighs contributions of the particles with high weights. However, there may be ways how to speed up the convergence.

The key attribute of the AIS is the possibility to shorten the batch to $n_k = 1$ and run the algorithm in completely recursive manner. This is especially important in modern hardware implementation such as the field programmable logic arrays (FPGA). Therefore, we propose to use exponential discarding (forgetting) of the previous data to preserve recursivity.

Proposition 2 (Forgetting in AIS.)**.** *Influence of the poor initial conditions can be suppressed by exponential weighting of the older contributions. The update of statistics (2) in AIS is than:*

$$\begin{aligned} \mathbf{V}^{[k]} &= \lambda \mathbf{V}^{[k-1]} + \sum_{i=1}^{n_k} w^{(i)} \mathbf{T}(\mathbf{x}^{(i)}), \qquad (5) \\ \nu^{[k]} &= \lambda \nu^{[k-1]} + \sum_{i=1}^{n_k} w^{(i)}, \end{aligned}$$

where $0 < \lambda < 1$ *is the forgetting factor.*

Effectiveness of the Proposition 2 is tested in simulation on the system from Example 1 via the number of samples necessary to reach $|\hat{\mu}_k - 5| < 0.1$. Histograms of the steps to convergence from 100 samples are displayed in Figure 2.

Remark 3 (Population Monte Carlo). An alternative way of dealing with poor initial conditions is the population Monte Carlo approach [10], where only the particles from the last batch are considered. This may be computationally inefficient since the particles in the previous batches may be relevant. An improvement was proposed in [**?**] using Rao-Blackwellization. However, these techniques can not be implemented recursively and we do not consider them as competitors.

III. Particle Filtering

Consider a discrete-time stochastic process:

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}, \mathbf{v}_t), \qquad (6) \\ \mathbf{y}_t &= g(\mathbf{x}_t, \mathbf{w}_t), \qquad (7) \end{aligned}$$

where $\mathbf{x}_t$ is the state variable, $\mathbf{y}_t$ is the vector of observations and $f(\cdot)$ and $g(\cdot)$ are known functions transforming the state into the next time step, or observation, respectively. Both transformations are subject to disturbances $\mathbf{v}_t$ and $\mathbf{w}_t$ which are considered to be random samples from a known probability density.

Sequential estimation of the posterior state probability is based on recursive evaluation of the filtering density, $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, using Bayes rule [11]:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \qquad (8)$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}, \qquad (9)$$

where $p(\mathbf{x}_1|\mathbf{y}_0)$ is the prior density, and $\mathbf{y}_{1:t} = [\mathbf{y}_1, \ldots, \mathbf{y}_t]$ denotes the set of all observations. The integration in (9), and elsewhere in this paper, is over the whole support of the involved probability density functions.

Equations (8)–(9) are analytically tractable only for a limited set of models. The most notable example of an analytically tractable model is linear Gaussian for which (8)–(9) are equivalent to the Kalman filter. For other models, (8)–(9) need to be evaluated approximately.

A. Particle filter

The particle filter [1], [12] is based on approximation of the joint posterior density by a weighted empirical density

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}), \qquad (10)$$

where $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$ is the state trajectory, $\{\mathbf{x}_{1:t}^{(i)}\}_{i=1}^N$ are samples of the trajectory (the particles), $w_t^{(i)}$ is the weight of the ith sample, $\sum_{i=1}^N w_t^{(i)} = 1$, and $\delta(\cdot)$ denotes the Dirac δ-function.

The main appeal of sequential Monte Carlo methods is in the fact that this approximation can be evaluated for an arbitrary model (6)–(7) given a suitable proposal density, $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$, yielding

$$\tilde{w}_t^{(i)} \propto \frac{p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})}. \qquad (11)$$

For a choice of the proposal density in factorized form, $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \prod_{\tau=1}^{t} q(\mathbf{x}_\tau|\mathbf{x}_{\tau-1}, \mathbf{y}_{1:\tau})$, the weights (11) can be evaluated recursively via:

$$\tilde{w}_t \propto \frac{p(\mathbf{y}_t|\mathbf{x}_{1:t})p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)} w_{t-1}. \quad (12)$$

Common choice $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is known as the bootstrap proposal.

1) Marginal Particle filter: An alternative to sampling from a full particle trajectory (10) is sampling only from the predictive density (9). Under the empirical approximation of $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, (9) becomes

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \sum_{j=1}^{n} \tilde{w}_{t-1}^{(j)} p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(j)}). \quad (13)$$

The resulting filter has some theoretical advantages at the price of higher computational cost [7]. Applying the importance sampling idea, we obtain weights in the form

$$w_t \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t) \sum_{j=1}^{n} \tilde{w}_{t-1}^{(j)} p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(j)})}{q(\mathbf{x}_t|\mathbf{y}_{1:t})}. \quad (14)$$

Note, that in this case, the proposal density is not conditioned on the previous value of the state variable $\mathbf{x}_{t-1}$.

B. Adaptive particle filtering

Proposal density is often the main factor in computational efficiency of the particle filter and was heavily studied for this purpose. The optimal proposal density is [12]:

$$\begin{aligned} q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) &= q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) q(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}), \\ q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) &= \frac{p(x_t|x_{t-1})p(y_t|x_t)}{\int p(x_t|x_{t-1})p(y_t|x_t)dx_t}. \end{aligned} \quad (15)$$

Since evaluation of the integral in (15) is computationally intractable and (15) is helpful only as a theoretical concept. The goal is to approximate (15) as closely as possible, with many approaches how to achieve it.

The approach of adaptive particle filtering introduced in [2] is aiming to improve the proposal function by using the measured data. Apart from the auxiliary particle filter, it also advocates the use of Taylor series expansion of the models (6)–(7). Its application yields an approximation of the posterior density in the form of a Gaussian [13], and is closely related to the Laplace's approximation [14]. Specifically, it operates in two steps:

1) the maximum likelihood estimate $\hat{\mathbf{x}}_t^{(i)}$ is found for a given value of $\mathbf{x}_{t-1}^{(i)}$, and
2) the proposal is a Gaussian density with mean value $\hat{\mathbf{x}}_t^{(i)}$ and variance given by

$$\mathbf{\Sigma}_t^{(i)} = [\frac{\partial}{\partial \mathbf{x}_t} \log p(\mathbf{x}_t, \mathbf{y}_t|\mathbf{x}_{t-1}^{(i)})].$$

The list of techniques for potential proposal generation is rather long [15].

C. Adaptive Importance Sampling Particle Filtering (AIS-PF)

Two problems need to be addressed for application of the AIS algorithm in the context of particle filtering: (i) association of the generated particles with those from the previous step, and (ii) initialization of the proposal parameters. Note that the generated samples in the AIS procedure are no longer independent and the path sampling (10) is thus hard to achieve. Therefore, more natural candidate for extension of the particle filter is the marginal particle filter (Section III-A1). Note that in this case, the particles at time t have no relation to those in time $t-1$. However, this may be computationally inefficient since evaluation of the marginal (13) is a $O(n^2)$ operation. This approach will be denoted AIS MPF.

A computationally cheaper solution may be auxiliary sampling from the mixture (13), i.e. for each new particle an index j is sampled and the weight is

$$\tilde{w}_t \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t)\tilde{w}_{t-1}^{(j)} p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(j)})}{q(\mathbf{x}_t|\mathbf{y}_{1:t})}. \quad (16)$$

This approach was proposed in [16] and will be referred to as AIS auxiliary PF.

A more challenging problem is the choice of initial statistics $V^{[0]}, \nu^{[0]}$ and thus the initial values of the parameter $\hat{\theta}^{[0]}$. We may consider three basic scenarios: (i) naive, where the initial statistics are build using propagation of the previous parameter estimates $\hat{\theta}_{t-1}^{[K]}$, (ii) the Laplace approximation, where the initial statistics are build around the maximum likelihood estimate $\hat{\theta}_t$ found e.g. by a numerical method, and (iii) generation of a population of particles of size n_1 and estimation of the initial statistics from it; the particles are discarded afterwards.

Each of these options represents a compromise between computational simplicity and effectiveness. The main difficulty of the first two methods is the choice of the parameter $\nu^{[0]}$ since it represents the relative non-normalized weight of the initial statistics. Note that any unlikely realization from the proposal may increase the non-normalized weight significantly. In effect, the statistics are heavily biased towards this value and may even result in numerical instability in the case of the variance estimate.

Proposition 4 (Adaptive choice of the initial statistics)**.** *The recursive update (5) is initialized with zeros statistics and the initial statistics is added in the evaluation of the expectation parameter in Step 1 of the Algorithm 1 by*

$$\hat{\phi}^{[k]} = \frac{V^{[k]} + \kappa V^{[0]}}{\nu^{[k]} + \kappa \nu^{[0]}},$$

where κ is an adaptation factor (the original formula arise for $\kappa = 1$). A good heuristic choice is $\kappa = \nu^{[k]}/n_{eff}$, and $n_{eff} = 1/\sum_{i=1}^{n} (\tilde{w}_t^{(i)})^2$ is known as the number of effective particles.

The heuristics is based on an intuitively appealing assumption that the contribution from the previous particles

should be proportional to the number of effective particles. In the case of ideal sampling ($w = 1$), $n_{eff} = \nu^{[k]}$ and the original formula is recovered. In the degenerate case of $n_{eff} = 1$, $\hat{\psi}^{[k]} = (\mathbf{V}^{[0]} + T(\mathbf{x}^{(j)}))/2$, where j is the index of the particle with maximum weight.

IV. Tracking of an Atmospheric Release of Radiation

Release of a nuclear material into the atmosphere can have severe impact on the health of people living in the trajectory of the release. The most likely place for such an event is the nuclear power station. Hence all are equipped with radiation monitoring network. Since the network is relatively sparse, it is capable to provide enough information for estimation of the most important parameters of the accident. Particle filtering has already been applied to this problem [9], [17]. However, due to sharply peaked likelihood function, the bootstrap proposal is inefficient and adaptive particle filtering using the Laplace proposal significantly outperforms it in terms of effective sample size [18].

A. Atmospheric dispersion model

If we knew all parameters of the release and that of the atmospheric conditions, the spatial distribution of an instantaneous release of the pollutant can be described by a model known as the Gaussian puff. Under this model, the released material creates a symmetrical plume with center $\mathbf{l}$ that is moving in the atmosphere and growing in size.

The concentration of the pollutant C is

$$C(\mathbf{s},\tau) = \frac{Qe^{-\lambda\tau}}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3} \exp\left[-\frac{(s_1-l_{1,\tau})^2}{2\sigma_1^2} - \frac{(s_2-l_{2,\tau})^2}{2\sigma_2^2} - \frac{(s_3-l_{3,\tau})^2}{2\sigma_3^2}\right]. \quad (17)$$

where $\mathbf{s} = [s_1, s_2, s_3]$ are spatial coordinates in the Cartesian coordinate system, τ is the number of seconds since the release, Q is the magnitude of the instantaneous release assigned to the puff in becquerel $[Bq]$, $\mathbf{l}_t = [l_{1,t}, l_{2,t}, l_{3,t}]$ is the location of the center of the puff, λ denotes the decay constant of the modeled isotope, and $\sigma_\cdot$ are dispersion coefficients given by atmospheric conditions.

The released material is subject to the wind field yielding the following movement of the puff:

$$\begin{aligned} l_{1,t+1} &= l_{1,t} - \Delta t\, v_t(\mathbf{l}_t)\sin(\phi_t(\mathbf{l}_t)), \\ l_{2,t+1} &= l_{2,t} - \Delta t\, v_t(\mathbf{l}_t)\cos(\phi_t(\mathbf{l}_t)), \\ l_{3,t+1} &= l_{3,t}, \end{aligned} \quad (18)$$

where $v_t(\mathbf{l}_t)$ and $\phi_t(\mathbf{l}_t)$ are the wind speed and wind direction at location $\mathbf{l}_t$, respectively.

In the case of continuous release, the released material forms a plume, which can be approximated by a sequence of puffs, one per sampling period of the measurement network. See Figure 3 for illustration.

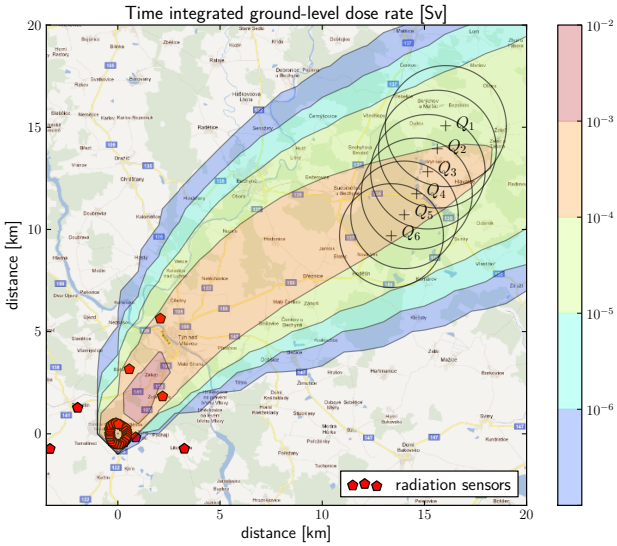

Figure 3. Illustration of the total absorbed dose of radiation from a short release modeled by six puff models.

B. State space model

The state space model for the problem is described in detail in[18] and is now briefly reviewed. The state variable is

$$\mathbf{x}_t = [a_t, b_t, Q_{t-M}, \dots Q_t, \mathbf{l}_{t-M}, \mathbf{l}_2 \dots, \mathbf{l}_t],$$

where $Q_{t-M}, \dots, Q_t$ are activities of the released puffs, and M is the number of puffs within the sensor range. The released activities are temporally independent with Gamma prior,

$$p(Q_t) = \mathcal{G}(\alpha_Q, \beta_Q), \quad (19)$$

with parameter α_Q, β_Q. Variables a_t, b_t are correction coefficients of the numerical wind field forecast, $\tilde{v}_t$ and $\tilde{\phi}_t$, such that the true wind speed and direction needed in (18) are

$$\begin{aligned} v_t(\mathbf{s}) &= \tilde{v}_t(\mathbf{s})a_t, & (20)\\ \phi_t(\mathbf{s}) &= \tilde{\phi}_t(\mathbf{s}) + b_t. & (21) \end{aligned}$$

Transition model for a_t and b_t is

$$\begin{aligned} p(a_t|a_{t-1}) &= \mathcal{G}(\gamma_a^{-2}, \gamma_a^2 a_{t-1}), & (22)\\ p(b_t|b_{t-1}) &= t\mathcal{N}(b_{t-1}, \sigma_b, \langle b_{t-1}-\pi, b_{t-1}+\pi\rangle). \end{aligned}$$

Here, $t\mathcal{N}$ denotes truncated Normal distribution. Parameters γ_a and σ_b govern variance of the transition model. The puff centers $\mathbf{l}_{t-M}, \dots, \mathbf{l}_t$ evolve deterministically according to (18).

C. Measurement model

The release of radiation is observable via radiation dose measurements from the radiation monitoring network (RMN) and an anemometer. Each sensor in the RMN provides measurements $y_{j,t}$ with inverse Gamma distribution

$$p(y_{j,t}|\mathbf{x}_{1:t}) = i\mathcal{G}(\gamma_y^{-2}+2,\,(\gamma_y^{-2}+1)(y_{nb,j}+y_{Q,j,t})). \quad (23)$$

where, γ_y is a parameter of relative variance of the distribution, $y_{nb,j}$ is a fixed value of typical radiation background at the jth sensor, and

$$y_{Q,j,t} = \sum_{m=1}^{M} c_{j,m,t}(\mathbf{l}_t)Q_{t-m}. \quad (24)$$

Here, $c_{j,m,t}$ is a complex function of the puff location obtained by 3D numerical integration.

V. Simulation Results

A. Simulation setup

The simulated accident was a release of radionuclide ^{41}Ar. Bayesian filtering is performed with a sampling period of 10 minutes, matching the sampling period of the radiation monitoring network which provides measurements of time integrated dose rate in 10-minute intervals. The same period was assumed for the anemometer.

Observations were sampled from distributions described in Section IV-C with mean values evaluated from the dispersion model. The accuracy of the radiation dose sensors was set to $\gamma_y = 0.2$. Accuracy of the anemometry was evaluated to correspond to parameters, $\gamma_a = 0.1$ and $\sigma_\phi = 5$ degrees. From historical data, we have estimated the variability of the transition model to be $\gamma_a = 0.2$ and $\sigma_b = 20$.

B. Tested methods

1) Conditionally independent proposal (Laplace): The proposal tailored for the problem of tracking atmospheric release of radiation was proposed in [18] as an improvement over the bootstrap proposal of [17].

The nature of the problem allows to choose a conditionally independent approximation

$$q(a_t, b_t, Q_t|\mathbf{y}_{1:t}) \approx q(a_t|v_t)q(b_t|\phi_t)q(Q_t|\mathbf{y}_{Q,t}), \quad (25)$$

where the first partitions are optimal proposals. For the inverse Gamma density of the likelihood (23), the Gamma transition model (22) is conjugate with posterior density in the form of Gamma density. So is the Normal likelihood of the wind direction and its Normal random walk model (22). The first two factorized densities in (25) can be thus solved analytically, using textbook results.

Derivation of $q(Q_t|y_{1:m}, \mathbf{s}_{1:t})$ is more demanding since the likelihood (23) is not conjugate. We apply the Laplace aproximation to obtain

$$q(Q_t|y_{1:m,t}, \mathbf{s}_{1:t}) = t\mathcal{N}(\mu_Q, \sigma_Q^2, \langle 0, \infty\rangle), \quad (26)$$

where algorithm for evaluation of moments μ_Q, σ_Q is described in [18].

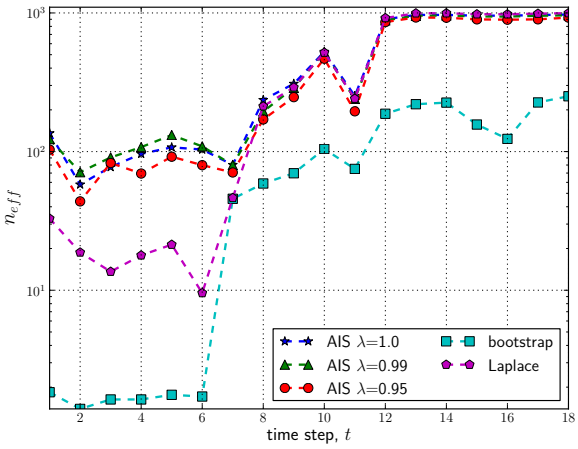

Figure 4. Efficiency of the proposal distributions for particle filtering for a release of radiation with known duration.

2) AIS PF algorithm: We choose the parametric form of the proposal to be:

$$q(\log a_t, b_t, \log Q_t|\mu_\theta, \Sigma_\theta) = \mathcal{N}(\mu_\theta, \Sigma_\theta), \quad (27)$$

which requires for an additional Jacobian in the evaluation of the likelihood function. The proposal is a multivariate extension of the Example 1. Hence, the natural parameter and the natural statistics are multivariate counterpart of those in Example 1, $\mathbf{T}(\mathbf{x}) = [\mathbf{x}, \mathbf{x}\mathbf{x}^T]$.

The initial statistics were chosen using a population of 100 particles using the conditionally independent Laplace proposal (Section V-B1). Statistics $\mathbf{V}^{[0]}$ was obtained by moment matching with the moments of the proposal for $\nu^{[0]} = 1$. All simple initializations failed in this case. The heuristics from Proposition 4 significantly improved performance of the method.

C. Efficiency of the methods

Efficiency of the proposals was tested on a simple scenario with a release of constant release rate Q_t from time $t = 1$, to $t = 6$, with $Q_{1:6} = [1, 1, 1, 1, 1, 1] \times 10^{16} Bq$. This scenario allows for comparison with previous approaches that were using the bootstrap proposal, e.g. [17]. The effective number of particles is displayed in Figure 4 for various values of the forgetting factor of AIS. The best results were obtained with $\lambda = 0.99$. Note that the proposed AIS approach significantly outperforms the conditionally independent Laplace proposal at the time of the release, at $t = 6$ the AIS PF with $\lambda = 0.99$ has by an order of magnitude higher n_{eff}. When the release is over and the cloud moves away from the sensor range (circa at time $t = 12$), the conditionally independent proposal becomes optimal. Note that in this case, the AIS result closely match those of the optimal proposal. This verifies that the parametric model is appropriate.

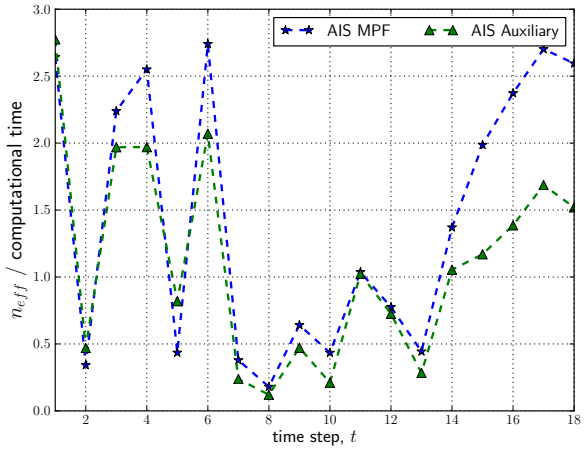

Figure 5. Efficiency of the AIS proposal distribution using marginalization or auxiliary sampling via the number of effective particles per one second of computational time.

In the first test, the difference in performance between the AIS MPF and AIS auxiliary PF was completely negligible. The improvement of explicit evaluation of the mixture (13) should be more apparent for narrow transition models [7]. To simulate such a scenario, we have run a more demanding experiment of unknown duration of the release (simulated at times $t = 7, \ldots, 12$, in stable weather ($\gamma_a = 0.05$ and $\sigma_b = 20$) and with higher measurement error $\gamma_y = 0.9$ (this can simulate the effect of model mismatch error). In this scenario, the AIS MPF outperformed the auxiliary sampling AIS in the number of effective particles per one second of execution time, Figure 5. All algorithms were implemented in the C language.

VI. Discussion

The proposed AIS-PF algorithm is especially suitable for implementation in FPGA pipeline. Note that in the recursive setup, the size of the batch n_k has the role of the delay between the time of computation of the statistics and its use for sampling. Thus the minimum value of n_k that can be chosen in the FPGA implementation is given by the length of the pipeline.

In the current form, the method is suitable mostly for unimodal distributions. It is relatively straightforward to extended it to the mixture proposals using a recursive method of estimation of their parameters [19]. However, such a method would be even more sensitive to the choice of its initial statistics. An alternative is to use the incremental mixture sampling [**?**] or the AMIS procedure [**?**], however, at the price of loosing the recursive evaluation.

VII. Conclusion

We have proposed to use the adaptive importance sampling (AIS) as a method of adaptive particle filtering. This method can be elegantly combined with the marginal particle filter, to yield a fully recursive adaptation. Thus, the algorithm is very attractive for implementation in field-programmable gate array. The price for recursivity is the sensitivity to the choice of initial statistics. We proposed several options and a heuristics based on the number of effective particles.

The method was tested on a challenging problem of tracking of atmospheric release of radiation. It was shown that the proposed method significantly improves the efficiency of particle generation at the demanding situations and matches the optimal proposal when available.

Acknowledgement

Support of grant MV ČR VG20102013018 is gratefully acknowledged.

References

[1] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," vol. 140, pp. 107–113, IEE, 1993.

[2] M. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.

[3] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, July 2000.

[4] J. Cornebise, E. Moulines, and J. Olsson, "Adaptive methods for sequential importance sampling with application to state space models," *Statistics and Computing*, vol. 18, no. 4, pp. 461–480, 2008.

[5] S. Saha, P. Mandal, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian proposal density using moment matching in smc methods," *Statistics and Computing*, vol. 19, no. 2, pp. 203–208, 2009.

[6] M. Oh and J. Berger, "Adaptive importance sampling in Monte Carlo integration," *Journal of Statistical Computation and Simulation*, vol. 41, no. 3, pp. 143–168, 1992.

[7] M. Klaas, N. De Freitas, and A. Doucet, "Toward practical N2 Monte Carlo: The marginal particle filter," in *Uncertainty in Artificial Intelligence*, pp. 308–315, 2005.

[8] R. Kulhavý, *Recursive nonlinear estimation: a geometric approach*. Springer, 1996.

[9] G. Johannesson, B. Hanley, and J. Nitao, "Dynamic bayesian models via monte carlo–an introduction with examples," tech. rep., Lawrence Livermore National Laboratory, 2004.

[10] O. Cappé, A. Guillin, J. Marin, and C. Robert, "Population monte carlo," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[11] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System identification* (P. Eykhoff, ed.), pp. 239–304, Oxford: Pergamon Press, 1981.

[12] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[13] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[14] R. E. Kass and A. E. Raftery, "Bayes factors," *J. of American Statistical Association*, vol. 90, pp. 773–795, 1995.

[15] M. Šimandl and O. Straka, "Sampling densities of particle filter: a survey and comparison," in *Proc. of the 26th American Control Conference (ACC)*, (New York City, USA), 2007.

[16] F. Lindsten, *Rao-Blackwellised particle methods for inference and identification*. Licentiate thesis no. 1480, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, June 2011.

[17] P. Hiemstra, D. Karssenberg, and A. van Dijk, "Assimilation of observations of radiation level into an atmospheric transport model: A case study with the particle filter and the etex tracer dataset," *Atmospheric Environment*, pp. 6149–6157, 2011.

[18] V. Šmídl and R. Hofman, "Application of sequential monte carlo estimation for early phase of radiation accident," Tech. Rep. 2322, UTIA, 2012.
[19] M. Sato, "Online model selection based on the variational Bayes," *Neural Computation*, vol. 13, pp. 1649–1681, 2001.