Variance reduction in particle transport models

J.W. Stijnen[a*], A.W. Heemink[a], and H.X. Lin[a].

[a]Faculty of Information Technology and Systems (ITS)
Department of Applied Mathematics, Delft University of Technology
Mekelweg 4, P.O. Box 5031, 2600 GA Delft, The Netherlands
E-mail: j.w.stijnen@hkv.nl

Prediction of the transport of pollutants in coastal areas is often performed with the aid of a so-called particle model. These types of models are Lagrangian models based on the advection-diffusion equation, either in three dimensions, or in a depth-averaged form.

Although these types of models have very distinct advantages when compared to deterministic transport models based on a numerical approximation of the advection-diffusion equation, there are some disadvantages. One of these is that because the model is stochastic in nature, many particle tracks need to be simulated in order to make reliable predictions. In order to address this problem several variance reduction methods are available. With the aid of these methods it is possible to estimate an ensemble average of the concentration using significantly less particles.

In this study the efficiency of the method of control variates is first investigated in a one-dimensional particle transport problem. Because satisfactory results are obtained, it is then applied to a two-dimensional situation. Although the problem becomes more complicated, the concept remains the same. It is shown that such a variance reduction method is an attractive way to decrease the number of computations in particle transport problems, especially when accurate estimates are needed.

1. INTRODUCTION

In order for an environmental management system in a water area to be of any value, insight into the transport and dispersion of pollutants within that area is essential. Lagrangian transport models are a versatile tool with respect to the modelling of dispersion of pollutants in turbulent flow. Much research has been carried out for a great many different situations to calculate the Lagrangian trajectories of these particles. The behavior of the ensemble average concentration of a passive tracer is quite well understood, and the amount of knowledge about concentration fluctuations from two-particle (or more) models has grown as well.

The transport of pollutants in coastal areas is often performed with the aid of a so-called particle model. The concept of these types of models is quite simple: by representing the pollutant parcels as particles it is possible to track these particles over time. This way

*Current affiliation: HKV CONSULTANTS, Botter 11-29, P.O. Box 2120, 8203 AC, Lelystad, The Netherlands

the spreading of a cloud of pollutant can be monitored. There are many types of Lagrangian transport models, but those used in this application belong in the category of random-walk models. These models are based on the advection-diffusion equation, and they describe the behavior of a large number of particles, each representing a certain amount of pollutant mass. By updating the positions of these particles an ensemble mean concentration can be approximated. The movement of every particle is completely independent from other particles, and consists of a deterministic part as well as a stochastic part. The deterministic part represents the movement of the particle due to advection (flow and wind for example), while the turbulent structure of the fluid is modelled by the stochastic part (diffusion). When one thinks in terms of particles one must realize that the probability density of particles at a certain location is in fact a description of the concentration at that location. The concentration transport described by the advection-diffusion is consistent with the probability density evolution described by the Kolmogorov equations (see [1] for example). Following Einstein's explanation of observed Brownian motion during the first decade of the previous century, attempts were made by Langevin and others to formulate the motions of particles in flow in terms of stochastic differential equations.

2. THE CONCEPT OF VARIANCE REDUCTION

So far we have primarily looked at the simulation of many particles in order to make a prediction as to where a certain amount of pollution might end up. We have used the concept of an ensemble of particles to propagate the mean of the cloud. Suppose we are investigating the results of a spill at *a single location* in our model. Think about a toxic spill in a harbor area and suppose that we are only interested in the concentration of the toxic material at a specific location away from the initial spill area. To this end, we make use of a particle transport model that simulates the movement of the toxic pollutant. As explained before, the concentration of a pollutant is connected to the particles' positions through their probability density function. The number of particles in our model that ends up at our location of interest can be seen as a measure of the concentration at that location. The question of how many particles we need to use is an important one. As long as a significant number of particles ends up near the desired location, we can speak of a statistically reliable answer to our question, typically in the order of hundreds of particles. In the case described above, most likely only a few percent of all simulated trajectories will end up in the target area. Since we are interested in the tracks that arrive in the target area, almost all of our computation is wasted.

But what if we could "steer" the trajectories of the particles in the right direction? That way we would need less particles, since more particles would reach the target area. On the other hand it means that we would change the original stochastic differential equation (SDE) by which the particles are transported, since we would be adjusting the particle paths. We would of course have to take this into account and correct the answer in such a way that we would again obtain the sought-after concentration. The above concept is based on a variance reduction technique called *importance sampling* ([3], and [4]). With variance reduction we could use fewer particles to obtain the same result. Another variance reduction technique is the method of control variates, which we will

focus on in this paper.

3. VARIANCE REDUCTION WITH CONTROL VARIATES

We define the positions of a group of n particles at time t to be $X_t^{\tau,\xi}(i)$, with $1 \leq i \leq n$, where the particles' starting position at time τ was equal to ξ. Consider next the following SDE in the Itô sense, which describes the motion of the particles

$$\begin{aligned} dX_t &= f(t, X_t)\, dt \;+\; g(t, X_t)\, d\beta_t \\ X_\tau &= \xi, \quad 0 \leq \tau \leq t \leq T \end{aligned} \tag{1}$$

where X_t is a stochastic variable, f and g are arbitrary functions and $d\beta_t = d\beta(t)$ are random increments. The process β_t is a Gaussian process with independent increments, and is called Brownian motion, i.e. $E\{\beta_t\} = 0$, and $E\{(\beta_t - \beta_s)^2\} = (t - s)$ The function f represents the deterministic or averaged drift term, and g is the space-time dependent intensity factor for the noisy, diffusive term. From now on, we also assume that the starting time τ, the final time T, and the starting location ξ are fixed.

Suppose our interest is in the probability that the pollutant reaches a critical region in space. For this region we define a certain indicator (reward) function ϕ, which measures the impact a particle has on the concentration in the region. The function ϕ depends only on $X_T^{\tau,\xi}$. If we assume that this function is symmetric with respect to its center, which is our main point of interest, say y, it takes on its highest value at this location. The further away from this point a particle is at our time of interest T, the less impact it will have on the concentration at y. Point-spread functions are ideal functions for this (see [5]). The probability that a pollutant will reach the area at a future time T is then given by

$$v(\tau, \xi) \;=\; E\{\phi(X_T^{\tau,\xi})\} \;=\; E\{\phi(X_T) \mid X_\tau = \xi\} \tag{2}$$

In order to evaluate this expectation, several approaches are available. One possibility is to simply use a forward Monte Carlo method to obtain an answer for the desired functional (2). We are able to directly calculate the function $v(\tau, \xi)$. Since we are only interested in the expectation, higher-order numerical methods (especially the weak ones) for SDEs can be used to evaluate the functional (2), or more specifically to evaluate the random variable $X_T^{\tau,\xi}$. We simulate n independent trajectories $X_T^{\tau,\xi}(i)$ forward in time from $t = \tau$ to $t = T$ and get the approximate solution:

$$\tilde{v} \;=\; \frac{1}{n} \sum_{i=1}^{n} \phi\big(X_T^{\tau,\xi}(i)\big) \tag{3}$$

where $\phi\big(X_T^{\tau,\xi}(i)\big)$ stands for the contribution of the i^{th} realization.

The idea behind variance reduction is to replace the process $X_T^{\tau,\xi}$ by another process $\tilde{X}_T^{\tau,\xi}$ with the same mean, but with significantly smaller variance. In the case of control variates, this means that not the particle paths are changed, but only the functional (2) itself. Therefore we do not need to change the SDE (1). We introduce a correction process, which we shall call Λ_t:

$$d\Lambda_t \;=\; \sum_{k=1}^{m} J^k(t, X_t)\, d\beta_t^k \tag{4}$$

$$\Lambda_\tau \;=\; \lambda \tag{5}$$

The control function J is hereby chosen as

$$J_k(t, X_t) \;=\; \sum_{i=1}^{d} \frac{\partial \tilde{v}}{\partial \xi_i}\bigg|_{(t,X_t)} g_{ik}(t, X_t) \tag{6}$$

where $\tilde{v}$ is an *approximate* solution to the backward Kolmogorov equation ([1] and [4]). The method of control variates is now based on the observation that

$$E\left\{\phi(X_T^{\tau,\xi})\right\} \;=\; E\left\{\phi(X_T^{\tau,\xi}) - \Lambda_T + \Lambda_T\right\} \tag{7}$$

$$\;=\; E\left\{\phi(X_T^{\tau,\xi}) - \Lambda_T\right\} \;+\; E\left\{\Lambda_T\right\} \tag{8}$$

and the fact that the process Λ_t is a martingale. Therefore

$$E\left\{\phi(X_T^{\tau,\xi})\right\} \;=\; E\left\{\phi(X_T^{\tau,\xi}) - \Lambda_T\right\} \;+\; \Lambda_\tau \tag{9}$$

Since the value of Λ_τ does not affect the variance, one can take this to be zero. Instead of computing the left-hand side of expression (9), we calculate the right-hand side, which has the same expectation, but significantly smaller variance.

3.1. A one-dimensional experiment

To make the concept of control variates more transparent, we will show an example where we know the analytical solutions for both the forward and backward equations. We will take a closer look at particles that only move due to diffusion from time $t = 0$ to $t = T$. The driving motion is a simple Brownian motion process and the diffusion coefficient D is kept constant:

$$dX_t \;=\; \sqrt{2D}\, d\beta_t, \qquad X_0 \;=\; \xi \;=\; 0 \tag{10}$$

where D is a constant diffusion coefficient, and X_t and β_t as before. We would like to know the concentration at final time T at location y. As our "reward" function to estimate this concentration, we choose a Gaussian function which is centered around our point of interest y, with variance σ^2

$$\phi(z) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(z-y)^2}{2\sigma^2}} \tag{11}$$

Our goal is to estimate the functional $E\{\phi(X_T^{0,0})\}$. The forward Kolmogorov equation that is consistent with this Lagrangian model is

$$\frac{\partial C_f}{\partial t} \;=\; D\frac{\partial^2 C_f}{\partial x^2} \tag{12}$$

Starting with a δ-peak at starting time $\tau = 0$ at $\xi = 0$, the forward solution is known completely in time and space, and is easily calculated. The backward Kolmogorov equation for this case is

$$\frac{\partial C_b}{\partial \tau} \;=\; D\frac{\partial^2 C_b}{\partial \xi^2} \tag{13}$$

Practically speaking, the most critical part of variance reduction with control variates is the fact that some backward solution is required to introduce corrections that reduce the variance. In general this backward solution is not known exactly and we are required to resort to numerical techniques. The complexity and computational efforts of methods such as finite differences become high rather quickly. Deterministic methods require computations in the entire domain, even though maybe only a small portion of the model is of interest.

An alternative would be to use a particle model for the backward calculations. It is similar to a forward particle model, but with different starting conditions and propagation equation. The required probability density for variance reduction can be calculated with the aid of point-spread functions (see [5]). For *each time instant* the variance-reduced forward calculation requires information about the backward solution; a probability density must be generated. Fortunately only a rough estimate of the probability density is needed, although the variance reduction is much better when the simulated density is close to the true one. To be more precise, it is the derivative of the backward solution that is required. When the point-spread function for the evaluation of the backward solution is chosen to be Gaussian, it is possible to directly generate this derivative. The more particles are used to generate the backward trajectories, the more accurate the approximation of the backward solution will be.

In our one-dimensional experiment the particles are traced backward from time $\tau = T$ ($t = 0$) to time $\tau = 0$ ($t = T$). Figures (1) and (2) give a graphical representation of the situation. The desired functional can be calculated through the particle model and then be compared to the true value

$$C_b(T,0) \;=\; \frac{1}{\sqrt{4\pi DT + 2\pi\sigma^2}} e^{\frac{-y^2}{4DT+2\sigma^2}} \tag{14}$$

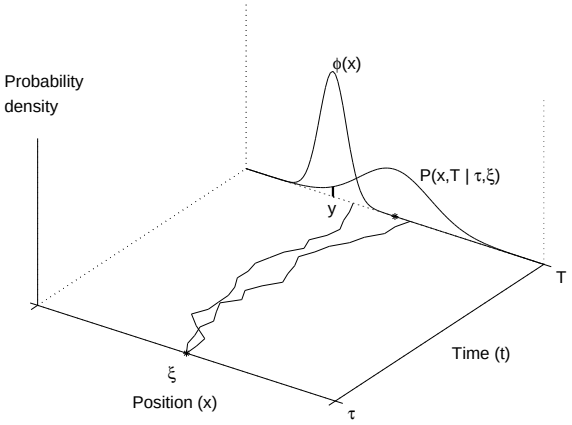

Figure 1. The forward Kolmogorov equation (τ and ξ are fixed).

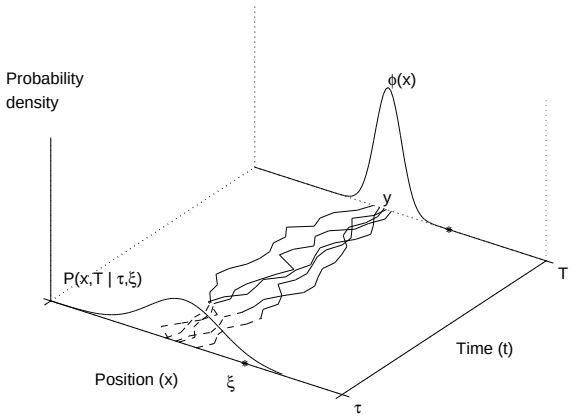

Figure 2. The backward Kolmogorov equation (x and t are fixed)

Note as well that

$$E\{\phi(X_T^{\tau,\xi})\} \;=\; \int_{-\infty}^{\infty} \phi(x)\; C_f(t,x)\; dx \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-y)^2}{2\sigma^2}} \;=\; C_b(T,0) \qquad (15)$$

Now, for the variance reduction part we write down the following system of equations:

$$\begin{bmatrix} dX_t \\ d\Lambda_t \end{bmatrix} \;=\; \begin{bmatrix} \sqrt{2D} \\ J(t,X_t) \end{bmatrix} d\beta_t \qquad (16)$$

with $X_0 = \xi = 0$ and $\Lambda_0 = \lambda = 0$, and

$$J(t,X_t) \;=\; \sqrt{2D}\; \frac{\partial C_b}{\partial \xi}\bigg|_{(t,X_t)} \qquad (17)$$

$$\;=\; \sqrt{2D}\; \left(\frac{-(X_t - y)}{2DT + \sigma^2}\right)\; \frac{1}{\sqrt{4\pi D\tau + 2\pi\sigma^2}} e^{\frac{-(X_t-y)^2}{4D\tau+2\sigma^2}} \qquad (18)$$

To gain some more insight in the behavior of the different equations and correction processes, we carried out some computer experiments with the following choices for our parameters: starting time $\tau = 0$, final time $T = 1$ and starting location $X_0 = \xi = 0$. Our point of interest, the location where we wish to evaluate the functional $E\{\phi(X_T^{0,0})\}$ was set to $y = \frac{1}{2}$. Finally we chose the diffusion coefficient $D = \frac{1}{2}$, and $\sigma^2 = \frac{1}{10}$ in our function ϕ. The number of time steps was varied as was the number of particles. The results from the experiments are shown in table (1), together with the exact solution. Clearly there is a large amount of gain in the accuracy of the results when using control variates. Also note that this gain depends on the amount of time steps (i.e. the accuracy of the numerical approximation) as well as the number of particles.

3.2. A two-dimensional experiment

Next, a similar experiment was performed in two dimensions. We take the same example as the one-dimensional example in section 3.1, but we extend it to two dimensions. We are again interested in the concentration at final time T at location $y = (y_1, y_2)$. As our "reward" function to estimate this concentration we choose a bi-variate Gaussian point-spread function which is centered around our point of interest (y_1, y_2), with variance σ^2:

$$\phi(z_1, z_2) \;=\; \frac{1}{2\pi\sigma^2} e^{\frac{-\left((z_1-y_1)^2+(z_2-y_2)^2\right)}{2\sigma^2}} \qquad (19)$$

Table 1
Results with control variates for the value of the functional $E\{\phi(X_T^{0,0})\}$.

Particles	$N_p = 10$	$N_p = 10$	$N_p = 100$	$N_p = 100$
Time steps	$N_t = 1000$	$N_t = 10000$	$N_t = 1000$	$N_t = 10000$
Exact	0.3395	0.3395	0.3395	0.3395
Plain Monte Carlo	0.3547 ± 0.1385	0.2684 ± 0.1181	0.3500 ± 0.0422	0.3695 ± 0.0446
Control variates	0.3468 ± 0.0091	0.3403 ± 0.0018	0.3373 ± 0.0025	0.3393 ± 0.0009

Notice that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(z_1,z_2)\;dz_1\;dz_2 \;=\; 1$. The particle tracks are generated according to the following stochastic differential equations:

$$dX_t \;=\; \sqrt{2D}d\beta_t^1 \tag{20}$$

$$dY_t \;=\; \sqrt{2D}d\beta_t^2 \tag{21}$$

where β_t^1 and β_t^2 are two independently Gaussian distributed random variables, and D and X_t as before. Our goal is to estimate the functional $E\{\phi(X_T^{0,0},Y_T^{0,0})\}$. The forward Kolmogorov equation that is consistent with this Lagrangian model is

$$\frac{\partial C_f}{\partial t} \;=\; D\left(\frac{\partial^2 C_f}{\partial x^2} \;+\; \frac{\partial^2 C_f}{\partial y^2}\right) \tag{22}$$

Starting with a δ-peak at starting time $\tau = 0$ at $(\xi,\eta) = (0,0)$ the forward solution is known completely in time and space. The backward Kolmogorov equation for this case is

$$\frac{\partial C_b}{\partial \tau} \;=\; D\left(\frac{\partial^2 C_b}{\partial \xi^2} \;+\frac{\partial^2 C_b}{\partial \eta^2}\right) \tag{23}$$

We apply the method of control variates that we used in the one-dimensional example once more. There are various ways to improve on the efficiency of the calculations, but these would be to extensive to discuss here. Several factors are of importance in the experiments: the number of particles in the forward part of the calculations, the number of particles in the backward part, the amount of CPU time for the plain Monte Carlo method, the CPU time for the control variates, and the amount of variance reduction that is obtained.

Because the problem is now two dimensional, we need more particles in the forward simulation to obtain a certain amount of confidence in the estimation of the mean. With 100 particles in the forward simulation the plain Monte Carlo result in this specific example will give a mean of around 0.112 and a variance of 0.027. Although the variance seems small, this is still an error of about 25% with respect to the mean. With 1000 particles this variance reduces to about 0.0089 (approximately 8%). Remember that in order to decrease the size of the error with a certain factor with plain Monte Carlo, we must increase the number of samples to the *square* of that factor. In the one-dimensional case this number was much less (just 100 particles for an error of 12% and 1000 particles for an error of 4%). It is obvious that in two-dimensional problems the number of particles for both the forward and backward simulations must be larger to find a satisfactory variance.

4. CONCLUSIONS

When is a variance reduction method useful? Obviously the method itself requires additional computations. Especially when the number of particles in a forward simulation is relatively small, this becomes visible. In a specific scenario, control variates may take many times longer to calculate. Fortunately, there is also a substantial reduction in variance in that case. Most notable are three things. The first is that the computation time for the plain Monte Carlo method scales linearly with the number of particles that is used in the forward simulation. Two times more particles will double the simulation time. The second observation is that the computation times for the control variates scale

roughly linearly with the number of particles in the backward simulation. Most important, however, is that the reduction in variance remains approximately constant, independent of the desired accuracy (number of samples in the forward simulation). The computational cost for the generation of the derivative of the backward solution remains the same, no matter how many particles are used in the forward model.

As it turns out, even with these additional calculations the variance reduction methods are useful. A generic variance reduction method was developed which, instead of using the more common methods of numerical approximations or an analytical solution, uses a slight adaption of the forward particle model. This generic method uses a particle method in combination with point-spread functions and was successfully applied in order to generate the desired derivative of the solution of the backward problem, both in one and two dimensions. The convergence rate of a plain Monte Carlo method is proportional to the square root of the number of particles. The negative impact of the high initial computation time for control variates becomes smaller as soon as more than a certain number of particles is used in the forward computations. This means that in situations where a certain amount of accuracy is required, it is beneficial to use a variance reduction technique. In case the simulated time in a scenario is long, these methods become attractive as well. Long time spans mean that the resulting variances are large, requiring large numbers of particles for a plain Monte Carlo method to reduce the variance. Recent developments have shown that it is also possible to generalize variance reduction methods to more arbitrary advection-diffusion type problems [2]. Generally speaking the variance reduction methods are very attractive in scenarios with either a large variance or in scenarios where an accurate estimate is needed. Another example where variance reduction can be used is when determining a probability of exceedence. In the case that the exceedence probability is small, many realizations of the process have to be generated to obtain a reliable estimate (see [3], or [4]).

REFERENCES

1. P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1995, 0-387-54062-8
2. G.N. Milstein and J.G.M. Schoenmakers and V. Spokoiny, *Transition density estimation for stochastic differential equations via forward-reverse representations*, Weierstraß-Institut für Angewandte Analysis und Stochastik, Preprint 680, 1–29, 2001
3. K. Ponnambalam, A.W. Heemink and J.G.M. Schoenmakers, *Estimation of Risk in Environmental Systems*, First International Conference on Measurements and Modelling in Environmental Pollution, R. San José and C.A. Brebbia, Computational Mechanics Publications, Southampton, U.K., 1997, 1-85312-461-3
4. J.G.M. Schoenmakers and A.W. Heemink, *Fast valuation of financial derivatives*, Risk Publications, 104-112 Marylebone Lane, London W1M 5FU, Journal of Computational Finance, Vol. 1, Nr. 1, pp. 47–67, 1997
5. J.W. Stijnen, A.W. Heemink, H.F.P. van den Boogaard and Lin, H.X.,*Optimal estimation of concentrations in particle-type transport models*, Environmental Coastal Regions, C.A. Brebbia, Computational Mechanics Publications, Southampton, UK, pp. –, 5312 527 X, 1998