Spatial-Temporal Nonlinear Filtering in Command and Control (C2)

M. E. Irwin, *The Ohio State University*
N. Cressie, *The Ohio State University*
G. Johannesson, *The Ohio State University*

Technical Report No. 697

October, 2002

Department of Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43210-1247

Spatial-Temporal Nonlinear Filtering in Command and Control (C2)

Mark E. Irwin, Noel Cressie, and Gardar Johannesson
Department of Statistics
The Ohio State University

Abstract

Assimilating data in a highly changing dynamic battlespace, which allows battle commanders to make timely, informed decisions, is a difficult and challenging problem. In this paper, a spatial-temporal statistical approach to examining the battlespace is taken, based on noisy data from multiple signals. We examine the danger-potential field (or danger field) generated by the positions of an enemy's weapons in the battlespace. The incoming noisy data is filtered to update the weapons' positions and the danger field. Sequential Importance Sampling, sometimes known as particle filtering, is used to generate realizations from the posterior distribution of the spatial-temporal danger field. Based on these realizations, non-linear questions such as the locations of maximum and minimum danger, the extent of regions exceeding certain danger thresholds, and changes in the danger field over time can be addressed. One particular Sequential Importance Sampler, known as the Unscented Particle filter, is compared to faster but less-accurate approximations based on the Kalman filter, using data generated from an object-oriented combat-simulation program.

Key Words: battlespace, danger-potential field, Kalman filter, particle filter, resampling, Scaled Unscented Transformation, Sequential Importance Sampler, Unscented Particle filter.
AMS Subject Classification(s): 62M20, 62M30

1 Introduction

Command and Control (C2) involves a body of applications used by all branches of the armed forces. These include, but are not limited to, command applications, operations applications, intelligence applications, fire-support applications, logistics applications, and communications applications. As C2 needs are shared across all branches of the armed services, tools developed should be sufficiently flexible to work across services and across allied forces, as needed.

In developing tools for C2 applications, three goals needs to be considered: analyzing the battlespace, visualizing the battlespace, and understanding the state of the battlespace. When examining the battlespace, commanders need a wide variety of information available with critical information highlighted so rapid decisions may be made. Great flexibility is required, as the questions a commander needs answered during a battle may change quickly. Information needs to be updated in a timely fashion, and the analysis should be scalable so that consistent information is delivered for large-scale operations as well as unit tactics.

In a battle situation, a flood of data is often available. Data may arrive from multiple sources in a wide range of formats such as visual, verbal, and from sensors. C2 tools need to be able to combine these data into highly useful information. The tools also need to be able to deal with missing and corrupted data.

One tool for describing the state of the battlespace is mapping of the *danger-potential field* [6]. The danger potential of a single weapon depends on two major factors, the location of the weapon and the weapon's damage potential. In a battlespace with multiple weapons, the danger-potential field (or, for short, danger field) is assumed here to be a sum of the danger potentials for the individual weapons.

In our development of the notion of danger potential, weapon damage is assumed to be of an explosive type; that is, assume that an ordinance from a weapon affects a continuous region and is a non-increasing function of distance from its impact point. The following formula describes one possible form of *damage potential* at a distance r from the impact point:

$$\delta(r) = \begin{cases} \alpha(1 - (r/R)^{p_1})^{p_2}, & \text{for } 0 \leq r \leq R \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where α, R, p_1, and p_2 are all explosion parameters defined for the weapon of interest. With this definition, a single location in the battlespace can be affected by damage resulting from nearby impacts in the space, and the damage potential will vary with distance from the impact.

Before continuing with the technical definition of the danger field, consider the following notation. Let $\mathbf{w}$ denote the location impacted by an ordinance from a weapon that is positioned at $\mathbf{Y} = (y_1, y_2)$ in the battlespace D. Of course, battlespaces are not static, so we introduce the time component t and let $\mathbf{X}_t$ denote the state of the weapon at time t, which includes its location $\mathbf{Y}_t$ plus possibly other information, such as the weapon's speed and direction. Further, let $f(\mathbf{w}|\mathbf{s}, \mathbf{X}_t)$ denote the probability density function of an impact at $\mathbf{w}$ given the weapon's state $\mathbf{X}_t$, and given it is aiming at a location $\mathbf{s}$ in D. Often the density function f will only depend on the weapon's state through its location $\mathbf{Y}_t$, which is assumed here. Finally, let $\mathbf{Z}_{it}$ denote the i-th observer's *reported* location of the enemy weapon at time t. For example, $\mathbf{Z}_{1t}$ might correspond to a radar observation, and $\mathbf{Z}_{2t}$ might correspond to a satellite observation.

The danger potential at time t, generated by a single weapon element in state $\mathbf{X}_t$, is defined as the expected damage at any location $\mathbf{s}$:

$$g(\mathbf{s}, t; \mathbf{X}_t) = \int \delta(r_{\mathbf{s},\mathbf{w}}) f(\mathbf{w}|\mathbf{s}, \mathbf{X}_t) d\mathbf{w}; \quad \mathbf{s} \in D, \tag{2}$$

where $r_{\mathbf{s},\mathbf{w}}$ is the distance between the target location $\mathbf{s}$ and the impact point $\mathbf{w}$, and the other terms are as described above; also see [6]. The aiming-accuracy distribution, $f(\mathbf{w}|\mathbf{s}, \mathbf{X}_t)$, is assumed to be based on a lognormal/normal cone shooting probability distribution, which is described in Section 3.

It should be noted that the danger field $g(\cdot, t; \mathbf{X}_t)$ can be computed in advance for various possible values of $\mathbf{X}_t$. Then, when a filter produces $\mathbf{X}_t^*$ based on present and past data, the "plug-in" estimate of the danger at $\mathbf{s}$, $g(\mathbf{s}, t; \mathbf{X}_t^*)$, is readily available.

In this paper, it is assumed that the danger potential is summable. That is, the danger potential at $\mathbf{s}$ from the k-th enemy weapon, which is in state $\mathbf{X}_{kt}$, contributes to the total danger field as follows:

$$g(\mathbf{s}, t; \{\mathbf{X}_{kt}\}) = \sum_k \int \delta(r_{\mathbf{s},\mathbf{w}_k}) f(\mathbf{w}_k|\mathbf{s}, \mathbf{X}_{kt}) d\mathbf{w}_k, \tag{3}$$

where $\mathbf{w}_k$ is the location impacted by an ordinance from the k-th weapon aiming at $\mathbf{s}$. More generally, if there are weapons of different types, f might depend on k also.

For the purpose of analysis, it is sufficient to consider the single-weapon case, although in the application in Section 4 we obtain the total danger field (3) for five tanks. The only unknown in (2) is the state $\mathbf{X}_t$. In this paper, its estimation, and the consequences of that estimation on making inference on its associated danger potential, are studied. Let the set of all observations on the weapon up to and including time t be denoted by $\mathbf{Z}_{1:t} \equiv (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_t)$, where for the convenience of presentation, observation times are equally spaced. Then one possible estimate of the danger field is

$$\begin{aligned} \hat{g}(\mathbf{s}, t; \mathbf{Z}_{1:t}) &\equiv E[g(\mathbf{s}, t; \mathbf{X}_t)|\mathbf{Z}_{1:t}] \\ &= \int g(\mathbf{s}, t; \mathbf{X}_t) p(\mathbf{X}_t|\mathbf{Z}_{1:t}) d\mathbf{X}_t, \end{aligned} \tag{4}$$

where $p(\mathbf{X}_t|\mathbf{Z}_{1:t})$ is the conditional density of the weapon's state at time t given the current and past data. The estimate (4) minimizes the Bayes risk based on the squared-error-loss function [11].

The state of the weapon at time t can be estimated by

$$\hat{\mathbf{X}}_t = E[\mathbf{X}_t|\mathbf{Z}_{1:t}]. \tag{5}$$

Hence, an alternative estimator of the danger field would be to use the "plug-in" method mentioned earlier and substitute (5) into (2) to obtain,

$$\tilde{g}(\mathbf{s}, t; \mathbf{Z}_{1:t}) \equiv g(\mathbf{s}, t; \hat{\mathbf{X}}_t). \tag{6}$$

Note that this estimator $\tilde{g}$ is different from $\hat{g}$ given by (4). While (4) is an unbiased estimator of the danger field in the sense that $E[\hat{g}] = E[g]$, the estimator (6) will generally be biased. The size of the bias depends on the degree of non-linearity of the damage-potential function $\delta(\cdot)$; see (1) for an example of a non-linear $\delta(\cdot)$.

While (4) has many nice properties, such as optimality under squared-error loss, calculating this estimator usually requires non-trivial computation. At the heart of the problem of estimating the danger field using (4), is the evaluation of $p(\cdot|\mathbf{Z}_{1:t})$, the conditional density of $\mathbf{X}_t$ given all of the data up to and including time t. Since the movement of the weapons may be given by highly non-linear dynamic models, standard Kalman filter (KF) approaches based only on second moments may result in poor approximations to this conditional distribution. This is further complicated by the danger potential being a non-linear function of the weapon's state.

There are a number of approaches for analyzing non-linear systems that go beyond the standard Kalman filter. A widely used approach is the extended Kalman filter (EKF) [1]. In this approach, the non-linear system is linearized, and then the standard KF is applied to this linearized system. Unfortunately, for many problems, the EKF does not give an accurate approximation to the posterior mean, variances, and covariances of the unknown state variables in $\mathbf{X}_t$.

Another approach is the Unscented Kalman filter (UKF), due to Julier and Ulhmann [19]; see also van der Merwe et al. [27]. This is based on the Unscented Transformation and the Scaled Unscented Tranformation [18]. Here, instead of linearizing the system, specially chosen realizations of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ are determined. These realizations are based on all the eigenvectors of the variance matrix of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$. The data $\mathbf{Z}_{1:t}$ are then combined with these realizations to give approximations to the posterior mean and variance of $\mathbf{X}_t$. This approach has a number of advantages over the EKF. First, the distribution of the underlying state process $\mathbf{X}_t$ is being approximated, not the non-linear function describing the evolution of $\mathbf{X}_t$ from $\mathbf{X}_{t-1}$, as with the EKF. This allows the UKF to partially incorporate information about skewness and kurtosis of the distribution, improving the accuracy. Also, the posterior means and variances of $\mathbf{X}_t$ can be calculated using standard vector and matrix operations, and no Jacobians are needed, unlike for the EKF. This suggests that the UKF can be much faster to compute than the EKF. A description of the UKF is given in Appendix A.

A similar approach to the UKF is the Ensemble Kalman filter (EnKF) [10, 15]. Instead of a deterministic scheme for generating realizations of $\mathbf{X}_t$, a Monte Carlo approach is taken. This approach has similar advantages to the UKF, as it approximates the distribution of $\mathbf{X}_t$ and not the non-linear function describing the evolution of $\mathbf{X}_t$ from $\mathbf{X}_{t-1}$. However, as the errors in the approximation are statistical in nature, the EnKF tends not to be as accurate as the UKF when the same number of chosen realizations are generated.

Another approach that is similar to the UKF is the Reduced Rank Square Root filter (RRSRF) [5, 15]. It is similar in that the realizations of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ are generated in a deterministic fashion. However, they are only generated in the directions of the q leading eigenvectors of the variance matrix of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$, where q is less than or equal to the dimension of the state space. The choice of q is important as it will affect the accuracy of the approximation. In fact, the UKF can be thought of as a generalization of this approach with all eigenvector directions being chosen, whereas the RRSRF appears to go only in the positive eigenvector direction. While this may speed up calculations, it could lead to a poor approximation as it is unable to handle skewness in the distribution of $\mathbf{X}_t$, given $\mathbf{X}_{t-1}$. A combination of EnKF and RRSRF, the Partially Orthogonal Ensemble Kalman filter (POEnKF) attempts to combine the advantages of the individual approaches [15]. However, as for the EnKF, its approximations are statistical in nature and it tends not to be as accurate as the UKF.

In all of these Kalman-filter type approaches, only approximations to the first two moments of the posterior distribution of the state variables are obtained. Unless the model can be described by a linear Gaussian process, these two moments do not completely describe the distribution and thus may lead to poor estimates of many (typically non-linear) properties of the state process.

To overcome these problems, sequential Monte Carlo approaches known as *Sequential Importance Samplers* (SIS), or *particle filters*, have been proposed [27, 24, 14]. SIS has been used successfully to analyze data from terrain navigation [2], genetics [17], and image analysis [4], for example. Another Monte Carlo approach, known as Markov Chain Monte Carlo, or MCMC [26, 12], is also possible, although it is less suitable to the sequential (or dynamic) nature of this problem; see the discussion in Section 6. These simulation approaches give many thousands of realizations$\{\mathbf{X}_t^{(i)} : i = 1, \dots, N\}$, from the complete posterior distribution of the state process, which allows inference on any functional of the posterior distribution, not just the first two moments. Thus, SIS can be used for estimating the danger field using (4), by approximating it with the (weighted) average of the associated danger fields $\{g(\mathbf{s}, t; \mathbf{X}_t^{(i)}) : i = 1, \dots, N\}$. Clearly, the main disadvantage when compared to the various forms of Kalman filtering, is the increased amount of computational time involved, something that will be addressed in this paper. A detailed presentation of SIS is given in Section 2.

Statistical inference on functionals of the danger field is straightforward using SIS. Suppose that a battle commander has a number of queries about the danger field, which may include the following.

1. Minimum and maximum danger: The locations of the minimum and maximum dangers could be of great use to a battle commander, suggesting areas that may need to be supported further or avoided, or regions that could be attacked.

2. Danger thresholds: Regions that need to be supported further or to be avoided can also be examined by investigating $I(g(\mathbf{s}, t; \mathbf{X}_t) \geq G)$, the indicator of whether the danger exceeds a given threshold G.

3. Changes over time: Given $\mathbf{s}$ and $t_2 > t_1$, evaluate
$$g(\mathbf{s}, t_2; \mathbf{X}_{t_2}) - g(\mathbf{s}, t_1; \mathbf{X}_{t_1}).$$
4. Regional danger: A battle commander may be interested in danger over a set of disjoint regions $B_1, B_2, \ldots, B_m$. The regions of interest may be arbitrary in terms of shape and size.

When there are potentially many questions or summaries of interest to a battle commander, a particle-filtering approach has an important advantage. The realizations from the simulation can be re-used to answer the different questions of interest. However, for the various forms of Kalman filtering reviewed earlier, a different filter needs to be run for each of the commander's queries.

While the Kalman-filter methods may not give optimum answers to solving nonlinear problems, they can in fact be used to improve SIS. In Section 2, it is seen that SIS relies on the specification of a proposal distribution, the optimal choice of which is not always computationally feasible. One approach is to use the Extended Kalman filter to give the proposal distribution, which is known as the extended Kalman particle filter [8, 25]. A more promising suggestion is the Unscented particle filter (UPF), where the Unscented Kalman filter is used to generate the proposal distribution [27, 28]; see Appendix B. The advantages that the Unscented Kalman filter has over the Extended Kalman filter, carry through to their associated particle filters.

In Section 2, an SIS is developed, resulting in the UPF, which can also be used for inference on the danger field. In Section 3, a description of a battlespace is given, and in Section 4, a simulated battlespace dataset is described. Analyses of this dataset by the Unscented particle filter (UPF), the Unscented Kalman filter (UKF) and the Extended Kalman filter (EKF), follow in Section 5. Finally, in Section 6, possible future directions are discussed.

2 Sequential Importance Sampling

In the three subsections that follow, we give a generic description of sequential importance sampling. Application of this methodology is found in Section 5.

2.1 Basic sequential importance sampler

Recall the definition of $\mathbf{Z}_{1:t}$; in a like manner, define $\mathbf{X}_{1:t}$ to be the sequence of state variables up to and including time t. In an ideal situation, it would be possible to simulate directly from $p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})$, the posterior distribution of $\mathbf{X}_{1:t}$ given the data $\mathbf{Z}_{1:t}$, where
$$\begin{aligned} p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t}) &= \frac{p(\mathbf{X}_{1:t})p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t})}{p(\mathbf{Z}_{1:t})} \\ &\propto p(\mathbf{X}_{1:t})p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t}), \end{aligned} \tag{7}$$

by Bayes Theorem. Both the state-variable density $p(\mathbf{X}_{1:t})$, and the measurement model $p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t})$, may have parameters associated with them. When implementing the methodology presented in the paper, these parameters will have to be estimated, leading to what is sometimes called an empirical Bayesian methodology. Alternatively, a fully Bayesian approach could be taken, where these parameters are given prior distributions and then averaged out [12].

Regardless of the method chosen to deal with the parameters, in many situations it is possible to use importance sampling techniques to simulate from the desired posterior distribution. The basic idea is as follows. Instead of simulating directly from the posterior distribution, a tractible approximation, $q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})$, is used as the proposal distribution for sampling, and the realizations are reweighted as follows:

$$\begin{aligned} E_p[h(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}] &= \int h(\mathbf{X}_{1:t})p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= \int h(\mathbf{X}_{1:t})\frac{p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= \int h(\mathbf{X}_{1:t})w_t(\mathbf{X}_{1:t})q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= E_q[h(\mathbf{X}_{1:t})w_t(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}], \end{aligned} \tag{8}$$

where h is a function of interest on the state variables, and

$$w_t(\mathbf{X}_{1:t}) \equiv \frac{p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}$$

is known as the *unnormalized importance sampling weight.* Furthermore, the proposal distribution q is known as the *importance sampling distribution.* Thus, given an importance sample $\{\mathbf{X}_{1:t}^{(1)}, \ldots, \mathbf{X}_{1:t}^{(N)}\}$ generated from $q(\cdot|\mathbf{Z}_{1:t})$ with associated importance sampling weights $\{w_t(\mathbf{X}_{1:t}^{(1)}), \ldots, w_t(\mathbf{X}_{1:t}^{(N)})\}$, the quantity of interest, $E_p[h(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}]$, can be consistently estimated by

$$\begin{aligned} \hat{E}_p[h(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}] &\equiv \frac{\sum_{i=1}^N h(\mathbf{X}_{1:t}^{(i)})w_t(\mathbf{X}_{1:t}^{(i)})}{\sum_{i=1}^N w_t(\mathbf{X}_{1:t}^{(i)})} \\ &= \sum_{i=1}^N h(\mathbf{X}_{1:t}^{(i)})\tilde{w}_t(\mathbf{X}_{1:t}^{(i)}), \end{aligned} \tag{9}$$

where $\tilde{w}_t(\mathbf{X}_{1:t}^{(i)})$ are the *normalized importance sampling weights* given by,

$$\tilde{w}_t(\mathbf{X}_{1:t}^{(i)}) \equiv \frac{w_t(\mathbf{X}_{1:t}^{(i)})}{\sum_{j=1}^N w_t(\mathbf{X}_{1:t}^{(j)})}; i = 1, \ldots, N.$$

There are many possible choices for the proposal distribution q. One form that is useful for many filtering problems is the Sequential Importance Sampler (SIS), where the importance sampling distribution has the form,

$$q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t}) = q(\mathbf{X}_1|\mathbf{Z}_1)\prod_{k=2}^{t} q(\mathbf{X}_k|\mathbf{X}_{1:k-1}, \mathbf{Z}_{1:k}). \tag{10}$$

Thus, to implement this sampler, a choice must be made for $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$ for each time t. As has been shown by Doucet, Gordon, and Krishnamurthy [9], an optimal choice for q is given by $q^*(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$. This choice is optimal in that it minimizes the variance of the importance sampling weights conditional on $\mathbf{X}_{1:t-1}$ and $\mathbf{Z}_{1:t}$. This possibility has also been advocated by others [21, 17, 23]. Another popular choice is to set $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t|\mathbf{X}_{1:t-1})$, the conditional distribution of the current state $\mathbf{X}_t$ given its past states. This proposal has the advantage that it is often easy to implement, but the disadvantage of having potentially much higher Monte Carlo variation as it does not incorporate the current and past observations $\mathbf{Z}_{1:t}$.

Note that with SIS, the set of unnormalized importance sampling weights $\{w_t(\mathbf{X}_{1:t}^{(i)})\}$ can be decomposed as,

$$w_t(\mathbf{X}_{1:t}^{(i)}) = w_{t-1}(\mathbf{X}_{1:t-1}^{(i)})\frac{p(\mathbf{X}_{1:t}^{(i)}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{1:t-1}^{(i)}, \mathbf{Z}_{1:t})p(\mathbf{X}_{1:t-1}^{(i)}|\mathbf{Z}_{1:t-1})}. \tag{11}$$

For many models, the multiplicative factor,

$$\frac{p(\mathbf{X}_{1:t}^{(i)}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{1:t-1}^{(i)}, \mathbf{Z}_{1:t})p(\mathbf{X}_{1:t-1}^{(i)}|\mathbf{Z}_{1:t-1})},$$

is easy to compute.

Consider the case where the state process is described by a first-order Markov process, so that

$$p(\mathbf{X}_{1:t}) = p(\mathbf{X}_1)\prod_{k=2}^{t} p(\mathbf{X}_k|\mathbf{X}_{k-1}); \tag{12}$$

and suppose that the measurements are conditionally independent given the states, so that

$$p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t}) = \prod_{k=1}^{t} p(\mathbf{Z}_k|\mathbf{X}_k). \tag{13}$$

Then recall that the optimal choice for q is,

$$q^*(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) \propto p(\mathbf{X}_t|\mathbf{X}_{t-1})p(\mathbf{Z}_t|\mathbf{X}_t). \tag{14}$$

However, even though the structure is relatively simple, (14) may still not be tractable for easy generation of importance samples; the model to be presented in Section 4 is one such example. Often, the importance sampling distribution (14) can be well approximated by a Gaussian distribution, such as when the measurement-model components, $\{p(\mathbf{Z}_k|\mathbf{X}_k) : k = 1, \ldots, t\}$, are Gaussian. Then a Gaussian approximation to (14) is, up to a normalizating constant, $\tilde{p}(\mathbf{X}_t|\mathbf{X}_{t-1})p(\mathbf{Z}_t|\mathbf{X}_t)$, where $\tilde{p}(\mathbf{X}_t|\mathbf{X}_{t-1})$ is a Gaussian approximation to the state-transition distribution $p(\mathbf{X}_t|\mathbf{X}_{t-1})$. One possible way to obtain $\tilde{p}(\mathbf{X}_t|\mathbf{X}_{t-1})$ is to use the Scaled Unscented Transformation [27] on $p(\mathbf{X}_t|\mathbf{X}_{t-1})$, leading to the Unscented particle filter (UPF). This is the principal method we use for analyzing the danger field in Section 5. Further details, with pseudo code for this situation, are given in Appendix B. Alternatively, one might consider a Taylor-series approximation, yielding the Extended particle filter, however this approximation will tend to break down as $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ deviates from a linear process.

In the case where $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$ in (10) depends on $(\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$ through only $(\mathbf{X}_{t-1}, \mathbf{Z}_t)$, and assuming the first-order Markov model (13), the update formula (11) for the unnormalized importance sampling weights satisfies,

$$w_t(\mathbf{X}_{1:t}^{(i)}) \propto w_{t-1}(\mathbf{X}_{1:t-1}^{(i)}) \frac{p(\mathbf{X}_t^{(i)}|\mathbf{X}_{t-1}^{(i)})p(\mathbf{Z}_t|\mathbf{X}_t^{(i)})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{t-1}^{(i)}, \mathbf{Z}_t)}. \tag{15}$$

Thus, for most problems of this type, the multiplicative factor that updates the weights is easy to compute. For the UPF, the multiplicative factor will often not vary greatly. However, in the case where the process model is used for the proposal (i.e., $q(\mathbf{X}_t|\mathbf{X}_{t-1}, \mathbf{Z}_t) = p(\mathbf{X}_t|\mathbf{X}_{t-1})$), the multiplicative factor for the weight update is $p(\mathbf{Z}_t|\mathbf{X}_t)$. This will often lead to very variable weights, yielding a much less efficient filter.

2.2 Resampling adaptations

One potential problem with the SIS algorithm is that the variance of the importance sampling weights increases over time [21]. This implies that, as t increases, more and more realizations will have normalized importance sampling weights close to zero. Thus, only one, or very few realizations will get the bulk of the weight in the approximation. This can occur, even if the optimal choice of the proposal distribution is used.

To avoid these problems, resampling approaches have been proposed [24, 27]. In these approaches, realizations from the SIS are sampled, possibly a multiple number of times, and these realizations are used in the next SIS step. Three sampling procedures that have been proposed are Multinomial Sampling [13], Residual Sampling [16, 23], and Minimum Variance Sampling [20, 7].

In this paper, the Minimum Variance Sampling procedure will be used. To implement it, one samples N points $\{U_1, \ldots, U_N\}$ in the interval $[0, 1]$, with each of the points a distance N^{-1} apart. (That is, $U_1 \sim Uniform[0, N^{-1}]$, and $U_i \equiv U_1+(i-1)N^{-1}; i = 2, \ldots, N$.) Then the i-th *resampled* realization, $\tilde{\mathbf{X}}_{1:t}^{(i)}$, is $\mathbf{X}_{1:t}^{(l)}$, where $\sum_{j=1}^{l-1} \tilde{w}_t(\mathbf{X}_{1:t}^{(j)}) \leq U_i < \sum_{j=1}^{l} \tilde{w}_t(\mathbf{X}_{1:t}^{(j)})$. Thus, N_l, the number of times $\mathbf{X}_{1:t}^{(l)}$ appears in the resample, is the number of points from $\{U_i\}$ that are between $\sum_{j=1}^{l-1} \tilde{w}_t(\mathbf{X}_{1:t}^{(j)})$ and $\sum_{j=1}^{l} \tilde{w}_t(\mathbf{X}_{1:t}^{(j)})$, and N_l satisfies $\lfloor N\tilde{w}_t(\mathbf{X}_{1:t}^{(l)})\rfloor \leq N_l \leq \lceil N\tilde{w}_t(\mathbf{X}_{1:t}^{(l)})\rceil$. This procedure can be implemented efficiently since only a single uniform random number on $[0, N^{-1}]$ needs to be generated; the other two procedures may require the generation of up to N random numbers. In addition, this resampling procedure induces a smaller variance on the $\{N_l\}$.

After a resampling step, the resulting sample $\tilde{\mathbf{X}}_{1:t}^{(1)}, \ldots, \tilde{\mathbf{X}}_{1:t}^{(N)}$ is an equally weighted sample from $p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})$. Therefore, the unnormalized weights $w_t(\mathbf{X}_{1:t}^{(i)})$, after resampling, must be reset to N^{-1}. After making this adjustment, the weights at future times, $t+1, t+2, \ldots$ (until the next resampling), are determined by equation (11).

How often to resample will depend on the problem of interest and the choice of proposal distribution. One approach is to set a fixed resampling schedule. That is, resample at times $t = m, 2m, 3m, \ldots$, where m is a prespecified resampling rate. A second approach is to monitor the weights $\{\tilde{w}_t(\mathbf{X}_{1:t}^{(i)})\}$ and resample when they start to become badly behaved. Liu [22] recommends monitoring the weights by their coefficent of variation and resampling when this exceeds a certain level. In the examples that follow, a fixed resampling schedule is used with $m = 1$.

2.3 Using a single sample to answer multiple questions

As exhibited by equations (8) and (9), the expectation of any integrable function of the state variables can be easily estimated with realizations from the importance sampling distribution. For example, the posterior mean of the state $\mathbf{X}_t$ can be estimated by

$$\hat{E}_p[\mathbf{X}_t|\mathbf{Z}_{1:t}] = \sum_{i=1}^{N} \mathbf{X}_t^{(i)} \tilde{w}_t(\mathbf{X}_{1:t}^{(i)});$$

and the danger field at location $\mathbf{s}$ and time t, as described in Section 1, can be estimated by

$$\hat{E}_p[g(\mathbf{s}, t; \mathbf{X}_t)|\mathbf{Z}_{1:t}] = \sum_{i=1}^{N} g(\mathbf{s}, t; \mathbf{X}_t^{(i)}) \tilde{w}_t(\mathbf{X}_{1:t}^{(i)}),$$

where the importance samples are obtained from SIS. Thus, in theory, a single SIS run can be used to answer multiple questions. However, the number of realizations N needed for the estimator to reach a specified level of precision will depend on the variance of the function of interest. So, in choosing N, one should have a good idea of the quantities of interest and it should be chosen so that the estimation variance of each quantity be within a pre-specified precision.

3 Description of the Battlespace

To examine the properties of SIS and compare its associated UPF to the UKF and the EKF, data generated by an object-oriented combat simulation program [6] was used. The program simulates the movement of five tanks in a 100 km by 100 km battlespace, $[0, 100] \times [0, 100]$. The battlespace is assumed to consist of flat terrain, with a town located in the southern part centered at (57.5, 18.5); see Figure 1.

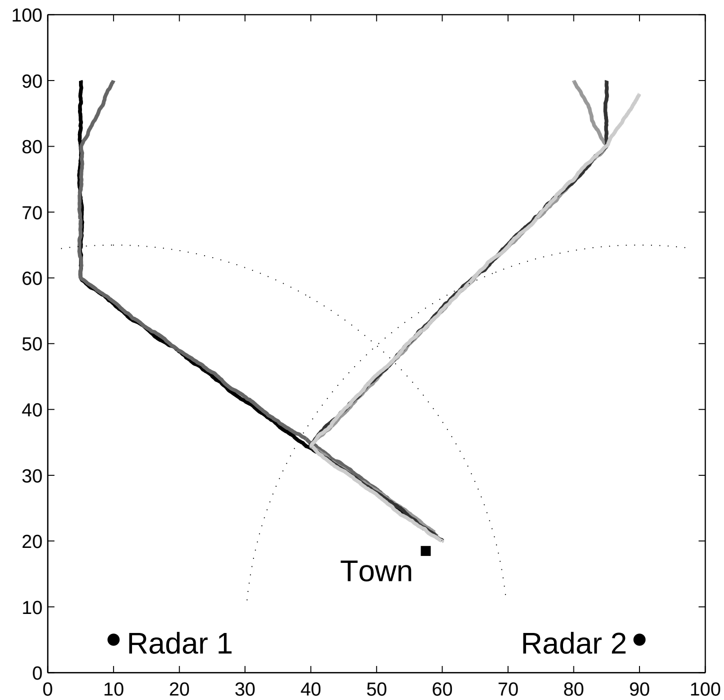

Figure 1: Battlespace features. The solid lines, in different grey scales, show the five true paths of the tanks in the 100 km x 100 km battlespace. The dotted lines show the range of the two radar stations. Also displayed are the locations of the two radar stations and the town.

Each of the five tanks has the same damage-potential parameters. The damage-potential function $\delta(\cdot)$ is given by (1) with maximum damage potential $\alpha = 1$, explosive radius $R = 0.05$ km (50 meters), and powers $p_1 = p_2 = 3$. The maximum range of the ordinance is 4 km.

The aiming-accuracy distribution f is described by a lognormal/normal cone shooting distribution: For a weapon at $\mathbf{Y}$ and a target at $\mathbf{s}$, denote the true distance between the two locations by $r_0 \equiv \|\mathbf{Y} - \mathbf{s}\|$ and the angle (in radians) from $\mathbf{Y}$ to $\mathbf{s}$ by θ_0. Then r, the distance from the weapon at $\mathbf{Y}$ to the impact location $\mathbf{w}$, is distributed as

$$r = r_0 \times \epsilon_r,$$

where ϵ_r has a lognormal distribution with mean 1 and standard deviation σ_r. The angle from $\mathbf{Y}$ to $\mathbf{w}$, θ, is distributed as

$$\theta = \theta_0 + \epsilon_\theta,$$

where $\epsilon_\theta \sim N(0, \sigma_\theta^2)$ and ϵ_r and ϵ_θ are independent. In the example, $\sigma_r = 0.01$ and $\sigma_\theta = \pi/180$ radians (or $1°$).

4 Battlespace Data

The five tanks start out in two groups, with two in the northwest (at (5,90) and (10,90)) and three in the northeast (at (80,90), (85,90), and (90,88)). Each tank has a sequence of three spatial waypoints to reach along its path, with the last waypoint for each near the town. A waypoint $\mathbf{W}$ of an object is, in general, a point in space and time that the object attempts to reach; see [29]. Note that for this example, the tanks have proximate waypoints and they are traveling at comparable velocities; the true paths of the tanks are shown in Figure 1.

4.1 Movement algorithm

The movement of the five tanks are independent given each of their waypoints, and the path of each tank is updated every 15 seconds for a period of 5 hours. Except for the waypoints, the movement algorithm is the same for each tank, so for the purpose of explanation, we concentrate on a single object moving through the two-dimensional battlespace. At time t, assume that the tank is at location $\mathbf{Y}_t = (y_{1t}, y_{2t})$, moving at speed v_t in direction θ_t (v_t and θ_t are the speed and angle that resulted in the object's moving from position $\mathbf{Y}_{t-1}$ to $\mathbf{Y}_t$). The speed and angle used to derive the object's new position at time $(t+1)$ are given by the random processes,

$$\begin{aligned} v_{t+1} &= \rho_v v_t + (1 - \rho_v) V_{t+1}, \\ \theta_{t+1} &= \rho_\theta \theta_t + (1 - \rho_\theta) \Theta_{t+1}, \end{aligned} \tag{16}$$

where ρ_v and ρ_θ are autocorrelation parameters (both specified to be 0.1 in the simulation), and V_{t+1} and Θ_{t+1} are random speed and angle changes. The random speed and angle changes are specified to fluctuate around the targeting speed and the direction that the object is trying to maintain, respectively. The targeting speed was fixed at 20 km/h for the whole simulation, and V_{t+1} was generated according to,

$$V_{t+1} = 20 \times \epsilon_{V,t+1},$$

where $\epsilon_{V,t+1}$ has a lognormal distribution with mean 1 and standard deviation 0.2. However, the targeting angle changes at each time point.

The targeting angle at time t, A_t, is specified to be the angle from the current location of the object, $\mathbf{Y}_t$, to its next waypoint, $\mathbf{W}$. The random angle deviation was then generated according to,

$$\Theta_{t+1} = A_t + \epsilon_{\Theta,t+1},$$

where $\epsilon_{\Theta,t+1}$ has a normal distribution with mean 0 and standard deviation $\pi/9$ radians (or 20°). This results in a (wrapped) normal distribution for Θ_{t+1} centered at A_t. The position of the object at time $(t+1)$, $\mathbf{Y}_{t+1}$, is then given by,

$$\begin{aligned} y_{1,t+1} &= y_{1t} + (v_{t+1}\Delta t)\cos(\theta_{t+1}), \\ y_{2,t+1} &= y_{2t} + (v_{t+1}\Delta t)\sin(\theta_{t+1}), \end{aligned} \tag{17}$$

where the time increment Δt is specified to be 15 seconds. That is, $\Delta t = (0.25/60)$ hours.

4.2 Observations

To observe the movement of the tanks, two radar stations were specified to be at (10,5) and (90,5). Each radar has an observation radius of 60 km, as shown in Figure 1. Notice that each radar can observe each tank for only part of its path, so that at any time a tank may be observed by 0, 1, or 2 radar stations.

Observations on the tank locations within range are taken every 15 seconds during the battlespace simulation, and these observations are specified to have a distribution given by the radius-angle distribution [6]. Specifically, if the true angle between a radar station and a target at $\mathbf{Y}$ is θ_1, then the observed angle is $\theta = \theta_1 + \epsilon_\theta$, where ϵ_θ is distributed with mean 0 and variance σ_θ^2; and if the true distance is r_1, then the observed distance r is randomly distributed with mean $r_1/(1-\frac{1}{2}\sigma_\theta^2)$ and variance σ_r^2. Note that although r is biased for r_1, this choice allows the observed locations, $\mathbf{Z}$, to be approximately unbiased estimators of the true locations, after converting back to standard Cartesian coordinates. More precisely, by applying Taylor series approximations, the observed location $\mathbf{Z}$ has,

$$\begin{gathered} E[\mathbf{Z}|\mathbf{Y}] \simeq \mathbf{Y} \\ var[\mathbf{Z}|\mathbf{Y}] \simeq [\sigma_r^2 - r_1^2(1-\gamma) - \sigma_\theta^2(\sigma_r^2 - r_1^2\gamma)]\mathbf{pp'} + \sigma_\theta^2[\sigma_r^2 - r_1^2\gamma]\mathbf{qq'}, \end{gathered} \tag{18}$$

where $\mathbf{p} = (\cos\theta_1, \sin\theta_1)'$, $\mathbf{q} = (\sin\theta_1, -\cos\theta_1)'$, and $\gamma = (1-\frac{1}{2}\sigma_\theta^2)^{-2}$. Both radar stations have the same measurement-error parameters with $\sigma_r = 0.005$ and $\sigma_\theta = \pi/360$ radians (or 0.5°). The observed weapon locations were assumed to be normally distributed with mean vector and covariance matrix as given above. For the purpose of illustration, Figure 2 shows tank 3's true path and its observed locations for the last 20 minutes.

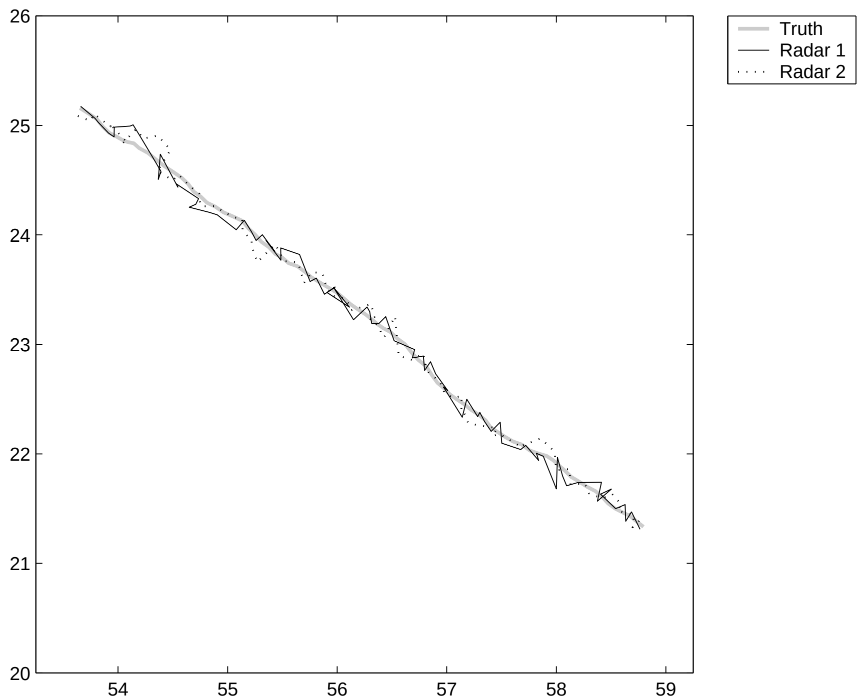

Figure 2: True path for tank 3 along with the two sets of radar observations for the last 20 minutes.

5 Analysis of Battlespace Data

5.1 Battlespace features of interest

To investigate the properties of the three estimation procedures of interest, the Unscented particle filter (UPF), the Unscented Kalman filter (UKF), and the Extended Kalman filter (EKF), a number of features of the battlespace will be investigated. First, estimates of the paths of the five tanks will be obtained. Next, estimates of the danger field for the whole battlespace at two times (3 and 5 hours) during the simulation will be calculated. Finally, three features of the danger at (57.5, 18.5), the center of the town, for the last 20 minutes of the simulation will be investigated. Specifically, the danger due to each tank, the probability that the danger due to each tank exceeds 0.25 damage units, and the combined danger of all five tanks will be examined.

A secondary factor of interest is the effect of data-collection frequency on the danger-field estimates. What happens when data are collected every 15 seconds is compared to what happens when data are collected every 60 seconds during the simulation. The latter dataset is generated from the original dataset by keeping every fourth observation.

5.2 Movement model used for analysis

Without loss of generality, consider a single object moving through the battlespace. The state of the object at time t is defined as $\mathbf{X}_t \equiv (\mathbf{Y}_t, v_t, \theta_t)$, where $\mathbf{Y}_t = (y_{1t}, y_{2t})$ is its position, v_t is its speed, and θ_t is its direction of travel at time t. We assume that $\{\mathbf{X}_t\}$ is obtained from the following movement equations:

$$\begin{aligned} \log v_{t+1} &= \log v_t + \epsilon_{v,t+1}, \\ \theta_{t+1} &= \theta_t + \epsilon_{\theta,t+1}, \end{aligned} \tag{19}$$

where $\epsilon_{v,t+1}$ and $\epsilon_{\theta,t+1}$ are independently distributed with $\epsilon_{v,t+1} \sim N(0, \sigma_v^2)$ and $\epsilon_{\theta,t+1} \sim N(0, \sigma_\theta^2)$. Then, given the object is in state $\mathbf{X}_t$ at time t, its position $\mathbf{Y}_{t+1}$ at time $(t+1)$ is given by,

$$\begin{aligned} y_{1,t+1} &= y_{1t} + \frac{1}{2}(v_{t+1}\cos\theta_{t+1} + v_t \cos\theta_t) \\ y_{2,t+1} &= y_{2t} + \frac{1}{2}(v_{t+1}\sin\theta_{t+1} + v_t \sin\theta_t), \end{aligned} \tag{20}$$

and its state at time $(t+1)$ is $\mathbf{X}_{t+1} = (\mathbf{Y}_{t+1}, v_{t+1}, \theta_{t+1})$. Clearly, $\{\mathbf{X}_t\}$ is a first-order Markov process; see (12).

This is a generic movement model that assumes constant acceleration and change in angle between time t and time $(t+1)$. Notice that the movement model (19), (20) does *not* match the algorithm that describes the manner in which the data were simulated; see (16), (17). It was deliberately chosen this way, since in practice one cannot expect to know the true movement process of an object under consideration.

The measurement process $\mathbf{Z}_t$ given $\mathbf{X}_t$ is described in Section 4.2. However, since the variance matrix (18) depends on the unknown parameters r_1 and θ_1, they need to be estimated. Let $\hat{r}_1$ be the distance between $\mathbf{Z}_t$ and the radar station, and let $\hat{\theta}_1$ be the angle from the radar station to $\mathbf{Z}_t$. Then substituting $\hat{r}_1$ and $\hat{\theta}_1$ into (18), an estimate of the variance matrix used by each of the filters is obtained.

To implement the three filters, the two variances, σ_v^2 and σ_θ^2, need to be specified. In this instance, they were obtained from the true simulated paths by matching moments, yielding the values $\sigma_v = 0.1$ and $\sigma_\theta = 0.2$ radians. In addition, a starting state is needed for each weapon. For each tank, the true state of the weapon just prior to being detected by the first radar was used. For the UPF and UKF, the Scaled-Unscented-Transformation parameters were set to $\alpha = 1$, $\beta = 2$, and $\kappa = 0$; see Appendix A. For the UPF, analyses are based on $N = 1000$ imputations for each tank with resampling done at each step. This choice for the number of imputations was based on timing and accuracy considerations and test runs.

5.3 Results

Estimates for the various features of interest described in Section 5.1 were obtained for the three filters (UPF, UKF, and EKF). Implementation was in `MATLAB` and analyses were performed on Red Hat Linux 7.3 servers with dual AMD Athlon 1800+ MP processors running at 1.533 GHz with 3 GB RAM.

Path estimates

The three filters tend to give similar path estimates during intervals between waypoints when the direction and speed of the tanks do not vary greatly. However, after the tank passes a waypoint where there is an abrupt change of direction, the UKF and EKF deviate further from the true path. Clearly, the movement of the tank during the change is not consistent with the movement model (19) and (20) used for analysis, but the UPF can adapt to this more quickly than the UKF or the EKF. This is apparent from Figures 3 and 4.

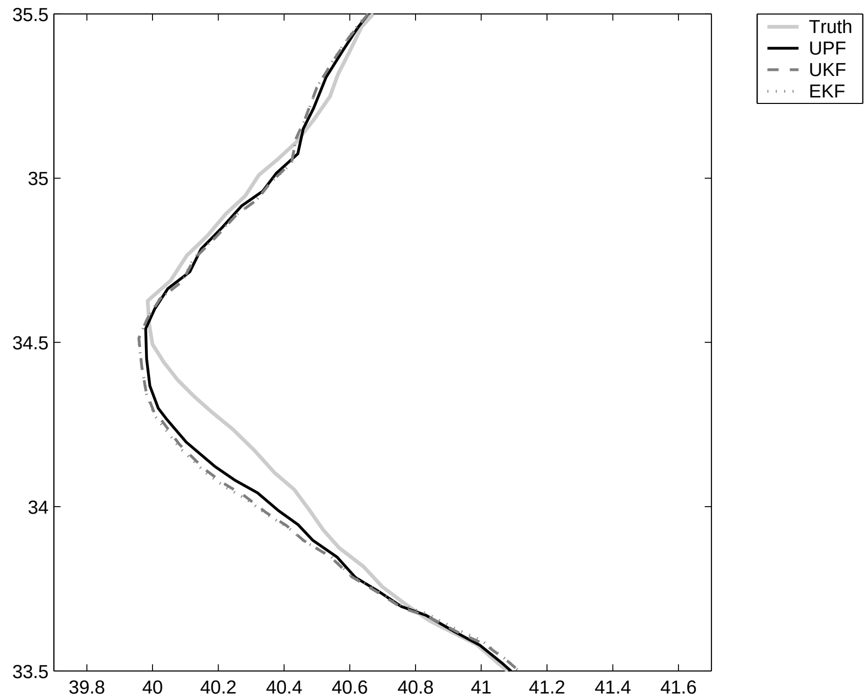

Figure 3: Estimated path for tank 3 for the three estimators, during the abrupt change of direction shown on the true path.

Danger for the whole battlespace

In Figures 5 and 6, estimates of the danger field at 3 and 5 hours are shown for the estimation procedure based on the UPF. In comparison, UPF- and EKF-based estimates of the danger field are too concentrated. This can be seen clearly in Table 1, which shows the areas of the regions with positive danger. As the UKF and EKF

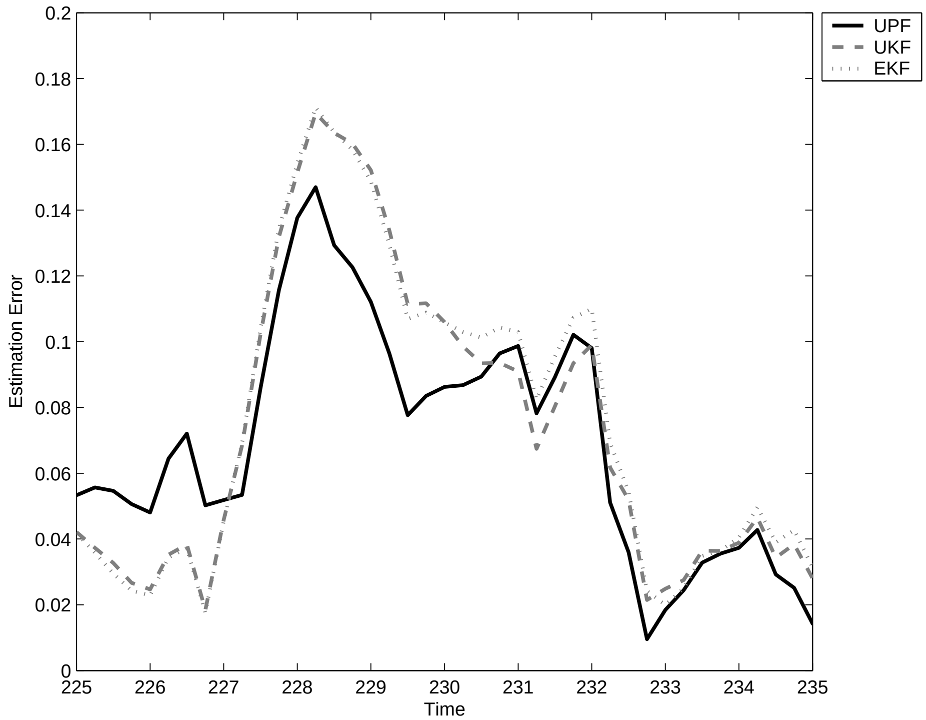

Figure 4: Error in estimated position for tank 3 for the three estimators, during the abrupt change of direction shown in Figure 3.

estimators are both plug-in estimators based on (6), and the estimates of the tank locations are similar for the two methods, the similar areas of positive danger are to be expected. The area of positive danger based on the UPF accounts for uncertainty of the danger field caused by uncertainty in the locations, and hence the area is not overoptimistic.

Table 1: Areas of regions (in km^2) with positive danger at 3 and 5 hours into the simulation.

Time	UPF	UKF	EKF
3 hours	158.69	127.43	127.43
5 hours	81.11	67.59	67.62

Danger at the center of the town

Estimates of the danger at the center of town due to tank 5, during the last 20 minutes of the simulation, is shown in Figure 7. The estimates based on the UKF and EKF are similar for both tanks, which is to be expected, since the estimated tank locations

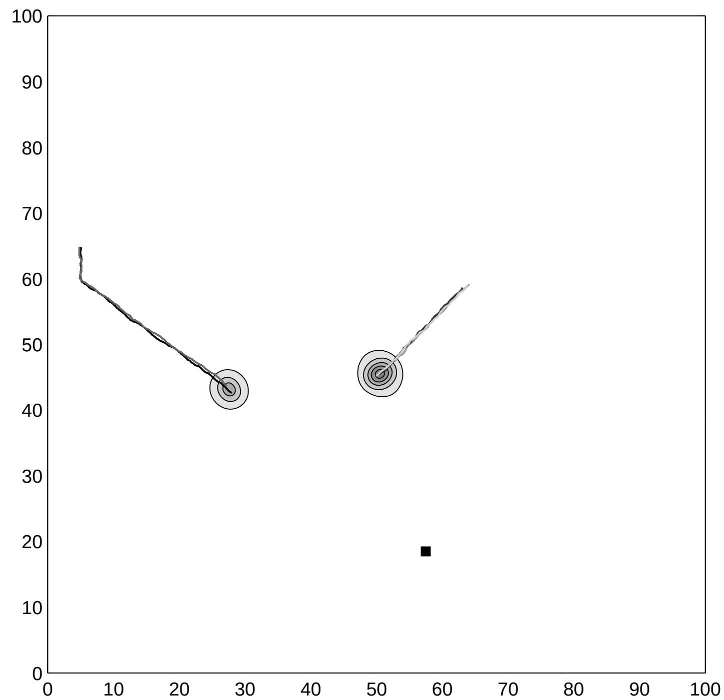

Figure 5: Estimated danger field at 3 hours based on the UPF estimator. The darker the gray colour, the higher the danger. The estimated paths for the 5 tanks are included.

are similar. These two estimates are also similar to the estimate based on the UPF, once the tank has clearly moved within range of the town. However, at times before that, the UKF and the EKF yield poor estimates, rather different from the town center's posterior mean danger (4). The posterior mean danger, given by an estimate obtained from the UPF, shows evidence of danger earlier, and this represents a clear advantage.

The total danger to the center of the town is shown in Figure 8. The sharp changes in danger occur when a new tank moves within range of the town. As with the individual tank estimates, the UPF-based estimate of danger does a better job of accounting for the uncertainties in the tank locations.

Data-collection frequency

The effects of data-collection frequency can be seen in Figure 9. The danger estimates based on the UPF due to tank 1 are shown for two sampling frequencies. While the width of the 90% credibility interval around the danger estimate is fairly constant for the 4-samples-per-minute sampling frequency, this doesn't hold for the 1-sample-per-minute sampling frequency. Instead, there is a sawtooth pattern in the width that

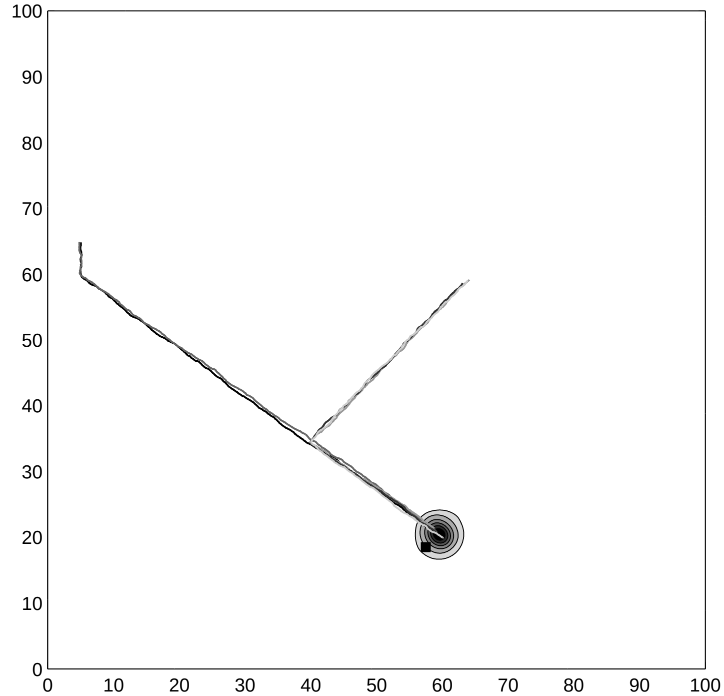

Figure 6: Estimated danger field at 5 hours based on the UPF estimator. The darker the gray colour, the higher the danger. The estimated paths for the 5 tanks are included.

increases as one gets further from the last observed data point, until a new radar observation is taken. Also, as to be expected, the credibility interval for the higher sampling frequency is tighter, since it is based on more data. While there is greater precision with the higher sampling frequency, the two UPF-based estimates of danger during the last 20 minutes are similar.

Direct estimation of danger field

Instead of using (6), the plug-in estimate of danger for the UKF or EKF, one could potentially estimate danger using these filters directly on the danger field. Unfortunately, this is not feasible for the EKF as it would involve a complicated calculation of partial derivatives of the danger field. However, it is feasible for the UKF, since no derivatives are required. While not as good in accounting for uncertainty as the UPF-based estimate, the danger-field, UKF-based estimate does account for some of the uncertainty of the tank position. The latter has one main drawback, namely if there is interest in the danger for many locations simultaneously, the size of the matrices involved in the UKF may become too large for efficient calculation.

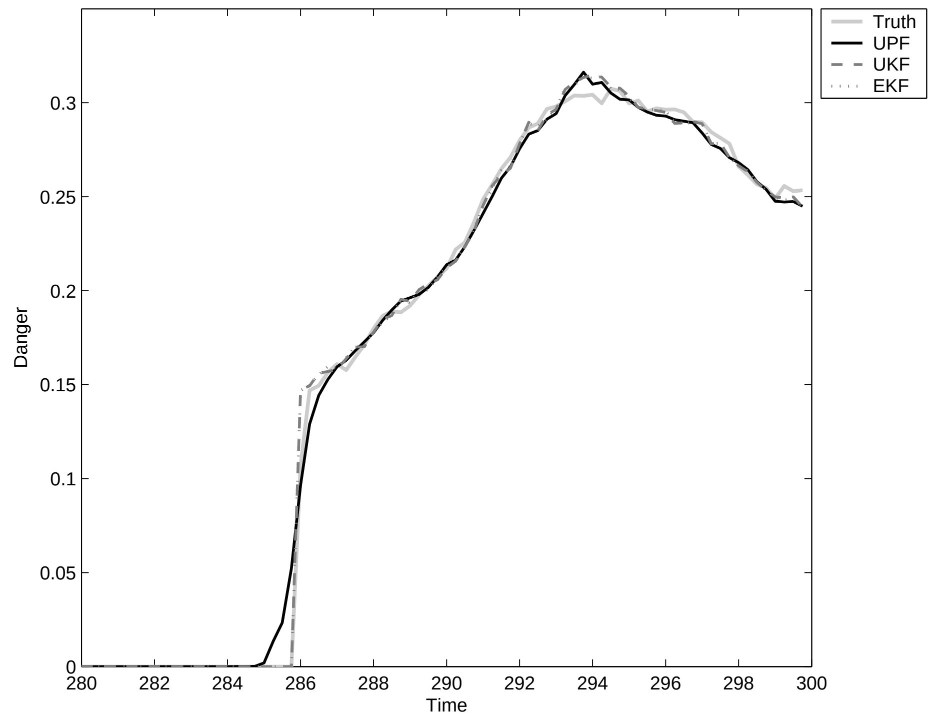

Figure 7: Estimated danger at the center of town due to tank 5 for the last 20 minutes.

6 Discussion and Conclusions

For the danger field to be a useful tool to the battle commander, quick and accurate estimation of the field and its features are needed. This can be performed by SIS procedures, in particular the UPF. The utility of the UPF has been demonstrated with a small example of tank movement in a battlespace. This example was also analyzed using two Kalman-filter approaches, the UKF and EKF. While these two approaches meet the fast-computation requirements, due to their approximative nature they can break down with respect to accuracy; see below.

As seen in Section 5, the UKF- and EKF-based estimates can be very similar to the UPF-based estimates, but on occasions they can miss important features of the danger field that the UPF-based estimates can detect. In the cases where the three filters give similar results, such as the estimation of the tanks' locations, the functionals being investigated are approximately linear. However, the Kalman filters tend to give poorer answers when the functional of interest is non linear, particularly for the case of extrema. For example, if we wish to estimate the danger at a location near the maximum range of a tank, the damage-potential function (1) is highly non linear in this situation because small changes in distance from the target may lead to large changes in damage.

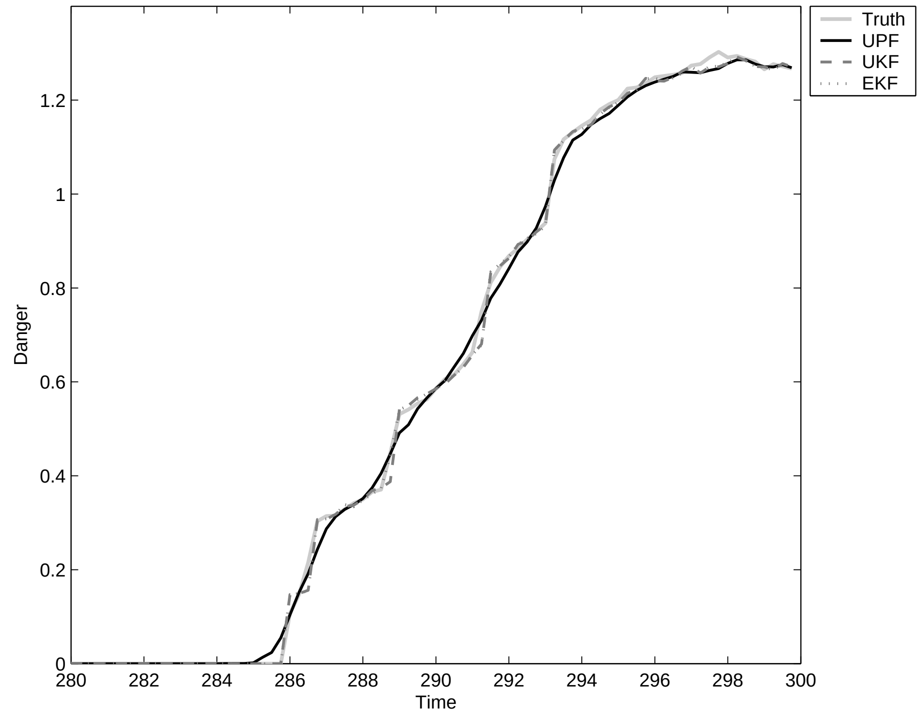

Figure 8: Estimated danger at the center of town due to all 5 tanks for the last 20 minutes.

While the UKF can miss important features, it can be useful when run simultaneously with the the UPF. The UKF can be used to give instaneous, though possibly rough, answers. If alerted by the UKF results, the UPF running in the background could then be used to derive more accurate results if needed. This approach should have the advantange of lowering the overall computation burden.

Processing time for the UPF on the state of the tanks, $\mathbf{X}_t$, averages approximately 3 seconds per time point t, about 20% of the data collection time. Processing time for the danger-field summaries vary, depending on the complexity of the summary. As `MATLAB` is an interpreted language, recoding the procedures in a compiled language such as C will likely lead to significant speed increases, which would allow real-time processing of the danger field by the UPF. Computing time for the two Kalman-filter-based approaches is not a concern as updates are virtually instanteous. However, the time advantage of the Kalman filters is offset by their potential to miss important non-linear features of the processes of interest.

One potential improvement in computation time for SIS-based methods, such as the UPF, is the use of parallel processing. SIS, as described in this paper, is based on performing N independent imputations at each time point t, which can be implemented naturally in a parallel fashion. Moreover, if the tanks are processed

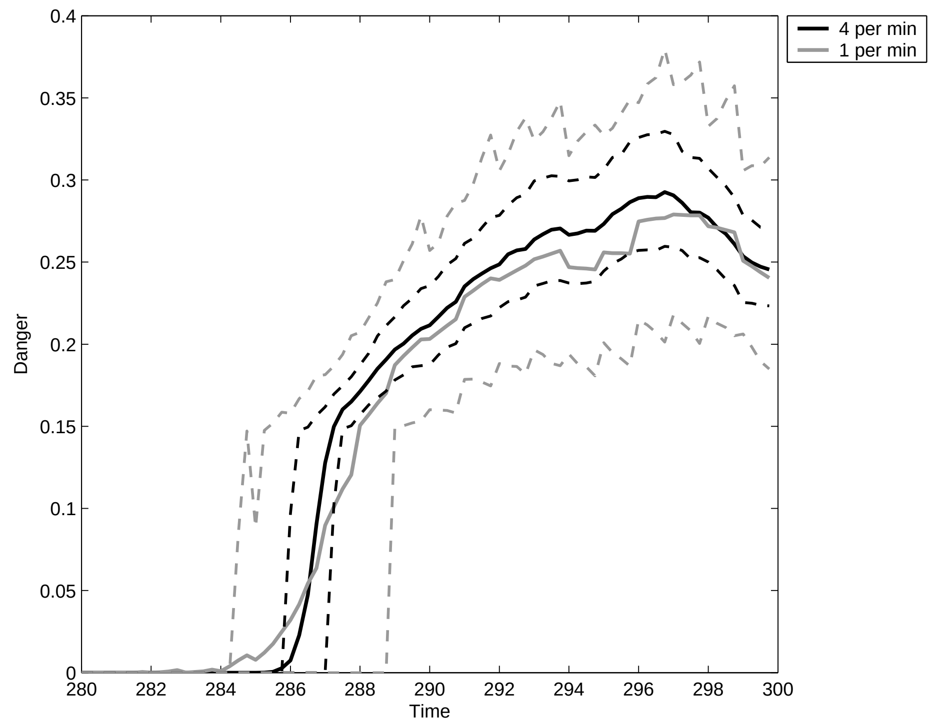

Figure 9: Estimated danger (solid lines) with 90% credibility intervals (dashed lines) due to tank 1 for two sampling frequencies (4 per minute and 1 per minute) for the last 20 minutes.

independently (as they are in the example in Section 5), there is a further opportunity for parallel processing.

The statistical model described in Section 5 is hierarchical. Consequently, another simulation-based approach called Markov Chain Monte Carlo (MCMC) could be used to examine the danger field. However, with MCMC samplers, it can be difficult to exploit the sequential nature of the process and the data. To filter the process at time $(t+1)$ by MCMC, the direct approach requires rerunning the chain given all previous data and the latest data at time $(t+1)$. This leads to an increasing computational burden as more data are collected. With SIS approaches, past states do not need to be reprocessed, since the necessary adjustment in the posterior distribution due to the new data is accomplished through the importance sampling weights and the resampling procedure. Although less efficient than SIS, there are occasions when an MCMC/SIS hybrid approach is useful. For a certain class of dynamic models, Berzuini et al. [3] propose a Metropolis-Hastings importance resampling approach that avoids rerunning the chain when new data are acquired.

Acknowledgements

This research was supported the Office of Naval Research under grants N00014-99-1-0214 and N00014-02-1-0052. The authors would like to expess their appreciation to John Kornak for a number of helpful suggestions.

Appendix A

In the description of the UKF that follows, we rely heavily on the exposition of the method given by van der Merwe et al. [27]. To implement the UKF, the distributions describing the evolution of the state process $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ and the measurement model $p(\mathbf{Z}_t|\mathbf{X}_t)$ need to be given in the following form:

- State-process evolution

$$\mathbf{X}_t = u(\mathbf{X}_{t-1}, \boldsymbol{\delta}_t)\,.$$

- Measurement model

$$\mathbf{Z}_t = v(\mathbf{X}_t, \boldsymbol{\epsilon}_t)\,,$$

where $\boldsymbol{\delta}_t$ and $\boldsymbol{\epsilon}_t$ are independent with means, $E[\boldsymbol{\delta}_t] = \mathbf{d}_t$ and $E[\boldsymbol{\epsilon}_t] = \mathbf{e}_t$, and variances, $var[\boldsymbol{\delta}_t] = \mathbf{D}_t$ and $var[\boldsymbol{\epsilon}_t] = \mathbf{E}_t$.

The essential ingredient of the UKF is the Scaled Unscented Tranformation (SUT). The state-process variable $\mathbf{X}_{t-1}$ is augmented with the random evolutions of the state process and the measurement model, yielding the augmented random variable $\mathbf{A}_t = [\mathbf{X}_{t-1}^T\ \boldsymbol{\delta}_t^T\ \boldsymbol{\epsilon}_t^T]^T$. Following van der Merwe et al. [27], the UKF that updates the posterior mean $\boldsymbol{\mu}_t$, and posterior variance $\boldsymbol{\Sigma}_t$, of the state variable $\mathbf{X}_t$, is obtained through the following sequence of steps:

1. Set $t = 1$ and define

$$\begin{aligned} \boldsymbol{\mu}_0 &= E[\mathbf{X}_0] \\ \boldsymbol{\Sigma}_0 &= var[\mathbf{X}_0]\,. \end{aligned}$$

2. For $t > 0$,

   (a) Calculate the mean vector and variance matrix of the augmented state variable $\mathbf{A}_t$:

$$\begin{aligned} \mathbf{m}_t &= E[\mathbf{A}_t] = [\boldsymbol{\mu}_{t-1}^T\ \mathbf{d}_t^T\ \mathbf{e}_t^T]^T \\ \mathbf{C}_t &= var[\mathbf{A}_t] = \begin{bmatrix} \boldsymbol{\Sigma}_{t-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_t \end{bmatrix}. \end{aligned}$$

Let n_a be the dimension of $\mathbf{A}_t$. Note that n_a may vary over time as it depends on the dimension of $\boldsymbol{\epsilon}_t$, which depends on the number of observations at time t.

(b) Calculate the $(2n_a + 1)$ SUT sigma points and weights:

$$\begin{aligned}
\boldsymbol{\mathcal{A}}_{0,t} &= \mathbf{m}_t \\
\boldsymbol{\mathcal{A}}_{j,t} &= \mathbf{m}_t + (\sqrt{(n_a + \lambda)\mathbf{C}_t})_j;\ j = 1, \dots, n_a \\
\boldsymbol{\mathcal{A}}_{-j,t} &= \mathbf{m}_t - (\sqrt{(n_a + \lambda)\mathbf{C}_t})_j;\ j = 1, \dots, n_a \\
W_0^{(m)} &= \lambda/(n_a + \lambda) \\
W_0^{(c)} &= \{\lambda/(n_a + \lambda)\} + (1 - \alpha^2 + \beta) \\
W_j^{(m)} &= W_{-j}^{(m)} = W_j^{(c)} = W_{-j}^{(c)} = 1/(2(n_a + \lambda));\ j = 1, \dots, n_a,
\end{aligned}$$

where $\lambda = \alpha^2(n_a + \kappa) - n_a$ is a scaling parameter, and $(\sqrt{(n_a + \lambda)\mathbf{C}_t})_j$ is the jth column of a matrix square root of the matrix $(n_a + \lambda)\mathbf{C}_t$. The paramters α, β, and κ are tuning parameters of the SUT. To guarantee positive-semidefiniteness of the covariance matrix, set $\kappa \geq 0$. It is also necessary to have $0 \leq \alpha \leq 1$ and $\beta \geq 0$.

(c) Time update:
Write $\boldsymbol{\mathcal{A}}_{j,t} = [(\boldsymbol{\mathcal{A}}_{j,t}^x)^T\ (\boldsymbol{\mathcal{A}}_{j,t}^\delta)^T\ (\boldsymbol{\mathcal{A}}_{j,t}^\epsilon)^T]^T$. Then define

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_{j,t|t-1} &= u(\boldsymbol{\mathcal{A}}_{j,t}^x, \boldsymbol{\mathcal{A}}_{j,t}^\delta);\ j = -n_a, \dots, n_a \\
\boldsymbol{\mu}_{t|t-1} &= \sum_{j=-n_a}^{n_a} W_j^{(m)} \boldsymbol{\mathcal{X}}_{j,t|t-1} \\
\boldsymbol{\Sigma}_{t|t-1} &= \sum_{j=-n_a}^{n_a} W_j^{(c)} [\boldsymbol{\mathcal{X}}_{j,t|t-1} - \boldsymbol{\mu}_{t|t-1}][\boldsymbol{\mathcal{X}}_{j,t|t-1} - \boldsymbol{\mu}_{t|t-1}]^T \\
\boldsymbol{\mathcal{Z}}_{j,t|t-1} &= v(\boldsymbol{\mathcal{X}}_{j,t|t-1}, \boldsymbol{\mathcal{A}}_{j,t}^\epsilon);\ j = -n_a, \dots, n_a \\
\boldsymbol{\nu}_{t|t-1} &= \sum_{j=-n_a}^{n_a} W_j^{(m)} \boldsymbol{\mathcal{Z}}_{j,t|t-1}\,.
\end{aligned}$$

(d) Measurement update:

Define

$$\mathbf{C}_{\mathbf{Z}_t\mathbf{Z}_t} = \sum_{j=-n_a}^{n_a} W_j^{(c)}[\boldsymbol{\mathcal{Z}}_{j,t|t-1} - \boldsymbol{\nu}_{t|t-1}][\boldsymbol{\mathcal{Z}}_{j,t|t-1} - \boldsymbol{\nu}_{t|t-1}]^T$$

$$\mathbf{C}_{\mathbf{X}_t\mathbf{Z}_t} = \sum_{j=-n_a}^{n_a} W_j^{(c)}[\boldsymbol{\mathcal{X}}^x_{j,t|t-1} - \boldsymbol{\mu}_{t|t-1}][\boldsymbol{\mathcal{Z}}_{j,t|t-1} - \boldsymbol{\nu}_{t|t-1}]^T$$

$$\mathbf{K}_t = \mathbf{C}_{\mathbf{X}_t\mathbf{Z}_t}\mathbf{C}^{-1}_{\mathbf{Z}_t\mathbf{Z}_t}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{Z}_t - \boldsymbol{\nu}_{t|t-1})$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{C}_{\mathbf{Z}_t\mathbf{Z}_t}\mathbf{K}_t^T .$$

The state at time t is estimated using $\boldsymbol{\mu}_t$, which is the UKF-estimate of the posterior mean, with its uncertainty given by $\boldsymbol{\Sigma}_t$, which is the UKF-estimate of the posterior variance. Also, the danger at a location $\mathbf{s}$ at time t is given by the "plug-in" estimate (6), obtained by "plugging in" $\boldsymbol{\mu}_t$ into (6).

3. Increment t by 1 time unit and return to step 2.

Appendix B

Implementation of the Unscented particle filter (UPF) when the state process $\{\mathbf{X}_t : t = 0, 1, 2, \dots\}$ is described by a first-order Markov process, has a similar structure to the UKF. However, one major difference is that the Scaled Unscented Transformation (SUT) is performed on a smaller-dimensional random vector, since the state variable $\mathbf{X}_{t-1}$ is assumed fixed during each implementation of the transformation. Let $\mathbf{A}_t = [\boldsymbol{\delta}_t^T \ \boldsymbol{\epsilon}_t^T]^T$, where $\boldsymbol{\delta}_t$ and $\boldsymbol{\epsilon}_t$ are as defined in Appendix A. Then the UPF is obtained through the following sequence of steps.

1. Set $t = 1$ and initialize the filter:
   (a) Sample $\mathbf{X}_0^{(1)}, \dots, \mathbf{X}_0^{(N)}$ from $p(\mathbf{X}_0)$
   (b) Initialize the importance sampling weights $w_0(\mathbf{X}_0^{(i)}) = \frac{1}{N}; i = 1, \dots, N$.
2. For $t > 0$,
   (a) Calculate the mean vector and variance matrix of $\mathbf{A}_t$:

$$\mathbf{m}_t = E[\mathbf{A}_t] = [\mathbf{d}_t^T \ \mathbf{e}_t^T]^T$$

$$\mathbf{C}_t = var[\mathbf{A}_t] = \begin{bmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_t \end{bmatrix}$$

   Let n_a be the dimension of $\mathbf{A}_t$. As in the UKF case, n_a may vary over time.

(b) Calculate the $(2n_a+1)$ SUT sigma points and weights:

$$\begin{aligned}
\boldsymbol{\mathcal{A}}_{0,t} &= \mathbf{m}_t \\
\boldsymbol{\mathcal{A}}_{j,t} &= \mathbf{m}_t + (\sqrt{(n_a+\lambda)\mathbf{C}_t})_j;\ \ j=1,\dots,n_a \\
\boldsymbol{\mathcal{A}}_{-j,t} &= \mathbf{m}_t - (\sqrt{(n_a+\lambda)\mathbf{C}_t})_j;\ \ j=1,\dots,n_a \\
W_0^{(m)} &= \lambda/(n_a+\lambda) \\
W_0^{(c)} &= \{\lambda/(n_a+\lambda)\} + (1-\alpha^2+\beta) \\
W_j^{(m)} &= W_{-j}^{(m)} = W_j^{(c)} = W_{-j}^{(c)} = 1/(2(n_a+\lambda));\ \ j=1,\dots,n_a,
\end{aligned}$$

where the tuning parameters are as specified in Appendix A.

(c) Time update:
Write $\boldsymbol{\mathcal{A}}_{j,t} = [(\boldsymbol{\mathcal{A}}_{j,t}^{\delta})^T\ (\boldsymbol{\mathcal{A}}_{j,t}^{\epsilon})^T]^T$. Then for each $i=1,\dots,N$, define:

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)} &= u(\mathbf{X}_{t-1}^{(i)}, \boldsymbol{\mathcal{A}}_{j,t}^{\delta});\ \ j=-n_a,\dots,n_a \\
\boldsymbol{\mu}_{t|t-1}^{(i)} &= \sum_{j=-n_a}^{n_a} W_j^{(m)} \boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)} \\
\boldsymbol{\Sigma}_{t|t-1}^{(i)} &= \sum_{j=-n_a}^{n_a} W_j^{(c)} [\boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)} - \boldsymbol{\mu}_{t|t-1}^{(i)}][\boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)} - \boldsymbol{\mu}_{t|t-1}^{(i)}]^T \\
\boldsymbol{\mathcal{Z}}_{j,t|t-1}^{(i)} &= v(\boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)}, \boldsymbol{\mathcal{A}}_{j,t}^{\epsilon});\ \ j=-n_a,\dots,n_a \\
\boldsymbol{\nu}_{t|t-1}^{(i)} &= \sum_{j=-n_a}^{n_a} W_j^{(m)} \boldsymbol{\mathcal{Z}}_{j,t|t-1}^{(i)}\,.
\end{aligned}$$

(d) Measurement update:
Define

$$\begin{aligned}
\mathbf{C}_{\mathbf{Z}_t\mathbf{Z}_t}^{(i)} &= \sum_{j=-n_a}^{n_a} W_j^{(c)} [\boldsymbol{\mathcal{Z}}_{j,t|t-1}^{(i)} - \boldsymbol{\nu}_{t|t-1}^{(i)}][\boldsymbol{\mathcal{Z}}_{j,t|t-1}^{(i)} - \boldsymbol{\nu}_{t|t-1}^{(i)}]^T \\
\mathbf{C}_{\mathbf{X}_t\mathbf{Z}_t}^{(i)} &= \sum_{j=-n_a}^{n_a} W_j^{(c)} [\boldsymbol{\mathcal{X}}_{j,t|t-1}^{(i)} - \boldsymbol{\mu}_{t|t-1}^{(i)}][\boldsymbol{\mathcal{Z}}_{j,t|t-1}^{(i)} - \boldsymbol{\nu}_{t|t-1}^{(i)}]^T \\
\mathbf{K}_t^{(i)} &= \mathbf{C}_{\mathbf{X}_t\mathbf{Z}_t}^{(i)} (\mathbf{C}_{\mathbf{Z}_t\mathbf{Z}_t}^{(i)})^{-1} \\
\boldsymbol{\mu}_t^{(i)} &= \boldsymbol{\mu}_{t|t-1}^{(i)} + \mathbf{K}_t^{(i)}(\mathbf{Z}_t - \boldsymbol{\nu}_{t|t-1}^{(i)}) \\
\boldsymbol{\Sigma}_t^{(i)} &= \boldsymbol{\Sigma}_{t|t-1}^{(i)} - \mathbf{K}_t^{(i)} \mathbf{C}_{\mathbf{Z}_t\mathbf{Z}_t}^{(i)} (\mathbf{K}_t^{(i)})^T\,.
\end{aligned}$$

(e) Importance sampling:

For $i = 1, \ldots, N$, sample $\mathbf{X}_t^{(i)}$ from $q(\mathbf{X}_t | \mathbf{X}_{t-1}^{(i)}, \mathbf{Z}_t) = N(\boldsymbol{\mu}_t^{(i)}, \boldsymbol{\Sigma}_t^{(i)})$, and update the importance-sampling weight (up to a normalizing constant)

$$w_t(\mathbf{X}_{1:t}^{(i)}) = w_{t-1}(\mathbf{X}_{1:t-1}^{(i)}) \frac{p(\mathbf{X}_t^{(i)} | \mathbf{X}_{t-1}^{(i)}) p(\mathbf{Z}_t | \mathbf{X}_t^{(i)})}{q(\mathbf{X}_t^{(i)} | \mathbf{X}_{t-1}^{(i)}, \mathbf{Z}_t)}.$$

Normalize the weights.

(f) Resample (if desired) as described in Section 2.2.

The state at time t and its contribution to the danger-field at a location $\mathbf{s}$ are estimated using the methods described in Section 2.3.

3. Increment t by 1 time unit and return to step 2.

References

[1] Anderson, B. D., and Moore, J. B. *Optimal Filtering.* Prentice-Hall, New Jersey, 1979.

[2] Bergman, N. Posterior Cramer-Rao bounds for sequential estimation. In *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and H. J. Gordon, Eds. Springer Verlag, New York, 2001, pp. 321–338.

[3] Berzuini, C., Best, N. G., Gilks, W. R., and Larizza, C. Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association 92* (1997), 1403–1412.

[4] Blake, A., Isard, M., and MacCormick, J. Statistical models of visual shape and motion. In *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and H. J. Gordon, Eds. Springer Verlag, New York, 2001, pp. 339–357.

[5] Cohn, S. E., and Todling, R. Approximate data assimilation schemes for stable and unstable dynamics. *Journal of the Meteorological Society of Japan 74* (1996), 63–75.

[6] Cressie, N., Wendt, D., Johannesson, G., Mugglin, A., and Hrafnkelsson, B. A spatial-temporal statistical approach to problems in command and control. In *Proceedings of U.S. Army Conference on Applied Statistics 2000* (2002), B. Bodt, Ed. in press.

[7] Crisan, D. Particle filters - a theoretical perspective. In *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and H. J. Gordon, Eds. Springer Verlag, New York, 2001, pp. 17–41.

[8] Doucet, A. On sequential simulation-based methods for Bayesian filtering. Technical report CUED/F-INFENG/TR 310, Department of Engineering, Cambridge University, 1998.

[9] Doucet, A., Gordon, N. J., and Krishnamurthy, V. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing 49* (2002), 613–624.

[10] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research 99* (1994), 10143–10162.

[11] Ferguson, T. S. *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, San Diego, 1967.

[12] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. *Bayesian Data Analysis.* Chapman and Hall, London, 1995.

[13] Gordon, N. *Bayesian Methods for Tracking.* PhD thesis, Imperial College, University of London, 1994.

[14] Gordon, N., Marrs, A., and Salmond, D. Sequential analysis of nonlinear dynamic systems using particles and mixtures. In *Nonlinear and Nonstationary Signal Processing*, W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. Young, Eds. Cambridge University Press, Cambridge, 2001.

[15] Heemink, A. W. Modeling and prediction of environmental data in space and time using Kalman filtering. In *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (Delft, Netherlands, 2000), G. B. M. Heuvelink and M. J. P. M. Lemmens, Eds., Delft University Press, pp. 283–291.

[16] Higuchi, T. Monte Carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation 59* (1997), 1–23.

[17] Irwin, M. E., Cox, N., and Kong, A. Sequential imputation for multilocus linkage analysis. *Proceedings of the National Academy of Science, USA 91* (1994), 11684–11688.

[18] Julier, S. J. The scaled unscented transformation. Manuscript - http://citeseer.nj.nec.com/julier99scaled.html, 1999.

[19] Julier, S. J., and Uhlmann, J. K. A new extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI, SPIE Proceedings* (Bellingham, WA, 1997), I. Kadar, Ed., vol. 3068, SPIE, pp. 182–193.

[20] Kitagawa, G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics 5* (1996), 1–25.

[21] Kong, A., Liu, J. S., and Wong, W. H. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association 89* (1994), 278–288.

[22] Liu, J. *Monte Carlo Strategies in Scientific Computing.* Springer Verlag, New York, 2001.

[23] Liu, J., and Chen, R. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association 93* (1998), 1032–1044.

[24] Liu, J., Chen, R., and Logvinenko, T. A theoretical framework for sequential importance sampling with resampling. In *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and H. J. Gordon, Eds. Springer Verlag, New York, 2001, pp. 17–41.

[25] Pitt, M. K., and Shephard, N. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94* (1999), 590–599.

[26] Smith, A. F. M., and Roberts, G. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B 55* (1993), 3–102.

[27] van der Merwe, R., Doucet, A., de Freitas, N., and Wan, E. The unscented particle filter. Technical report CUED/F-INFENG/TR 380, Department of Engineering, Cambridge University, 2000.

[28] van der Merwe, R., Doucet, A., de Freitas, N., and Wan, E. The unscented particle filter. In *Advances in Neural Information Processing Systems (NIPS13)* (Cambridge, USA, 2001), T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., MIT Press.

[29] Wendt, D., Cressie, N., and Irwin, M. E. Waypoint analysis for command and control. Technical Report 691, Department of Statistics, The Ohio State University, 2002.