Sociedad Española de Estadística e Investigación Operativa

Test

Volume 11, Number 2. December 2002

Spatial-Temporal Nonlinear Filtering Based on Hierarchical Statistical Models

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

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 249–302

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 249–302

Spatial-Temporal Nonlinear Filtering Based on Hierarchical Statistical Models

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

Abstract

A hierarchical statistical model is made up generically of a data model, a process model, and occasionally a prior model for all the unknown parameters. The process model, known as the state equations in the filtering literature, is where most of the scientist's physical/chemical/biological knowledge about the problem is used. In the case of a dynamically changing configuration of objects moving through a spatial domain of interest, that knowledge is summarized through equations of motion with random perturbations. In this paper, our interest is in dynamically filtering noisy observations on these objects, where the state equations are nonlinear. Two recent methods of filtering, the Unscented Particle filter (UPF) and the Unscented Kalman filter, are presented and compared to the better known Extended Kalman filter. Other sources of nonlinearity arise when we wish to estimate nonlinear functions of the objects positions; it is here where the UPF shows its superiority, since optimal estimates and associated variances are straightforward to obtain. The longer computing time needed for the UPF is often not a big issue, with the ever faster processors that are available. This paper is a review of spatial-temporal nonlinear filtering, and we illustrate it in a Command and Control setting where the objects are highly mobile weapons, and the nonlinear function of object locations is a two-dimensional surface known as the danger-potential field.

Key Words: battlespace, danger-potential field, Kalman filter, particle filter, resampling, scaled unscented transformation, sequential importance sampler, unscented particle filter.

AMS subject classification: 62M20, 62M30

1 Introduction

The Kalman filter (e.g. Anderson and Moore (1979)) was instrumental in allowing humans to land on the moon and return safely with their samples

*Correspondence to: Noel Cressie, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus OH 43210, USA. Phone: 614-292-5194. Fax: 614-292-2096. E-mail: ncressie@stat.ohio-state.edu.

Received: September 2002; Accepted: October 2002

to Earth. Vehicles move through a space-time continuum according to the laws of motion, but what is often not clear is the intent of the commanders at the controls of the vehicles. Based on current and past (noisy) measurements on the position of a vehicle, a model for measurement, and a model for vehicle movement, one can estimate the current position of the vehicle. The algorithm that yields the estimate is usually called a filter; the ubiquitous Kalman filter is based on both models being linear.

In this article, we review some of the recent developments in spatial-temporal nonlinear filtering, by putting the filtering problem in a hierarchical-statistical-model setting. In this context, our approach is empirical Bayesian, since we do not impose a prior distribution on unknown parameters. We feature a filter known as the Unscented Particle filter (UPF), which was recently proposed in van der Merwe et al. (2001, 2000) and is based on sequential importance sampling. While computationally demanding, the ever faster processors available often make this filter preferable to the better known Extended Kalman filter.

A natural application of this nonlinear-filtering methodology is to Command and Control (C2) and the monitoring of objects in a battlespace. 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. In developing tools for C2, 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.

One tool for describing the state of the battlespace is the mapping of the *danger-potential field* (Cressie et al. (2002)). 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, it is assumed 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:

$$\delta(r) = \begin{cases} \alpha(1-(r/R)^{p_1})^{p_2}, & \text{for } 0 \le r \le R \\ 0, & \text{otherwise,} \end{cases} \tag{1.1}$$

where α, R, p_1, and p_2 are all 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 $\mathbf{s}$:

$$g(\mathbf{s}, t; \mathbf{X}_t) = \int \delta(r_{\mathbf{s},\mathbf{w}}) f(\mathbf{w}|\mathbf{s}, \mathbf{X}_t) d\mathbf{w}; \quad \mathbf{s} \in D, \tag{1.2}$$

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 Cressie et al. (2002). 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:

$$g(\mathbf{s}, t; \{\mathbf{X}_{kt}\}) = \sum_k \int \delta(r_{\mathbf{s},\mathbf{w}_k}) f(\mathbf{w}_k|\mathbf{s}, \mathbf{X}_{kt}) d\mathbf{w}_k, \tag{1.3}$$

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 5 we obtain the total danger field (1.3) for five tanks. The only unknown in (1.2) 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

$$\hat{g}(\mathbf{s}, t; \mathbf{Z}_{1:t}) \equiv E[g(\mathbf{s}, t; \mathbf{X}_t)|\mathbf{Z}_{1:t}] = \int g(\mathbf{s}, t; \mathbf{X}_t) p(\mathbf{X}_t|\mathbf{Z}_{1:t}) d\mathbf{X}_t, \tag{1.4}$$

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 (1.4) minimizes the Bayes risk based on the squared-error-loss function (Ferguson (1967)).

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

$$\hat{\mathbf{X}}_t = E[\mathbf{X}_t|\mathbf{Z}_{1:t}]. \tag{1.5}$$

Hence, an alternative estimator of the danger field would be to use the "plug-in" method mentioned earlier and substitute (1.5) into (1.2) to obtain,

$$\tilde{g}(\mathbf{s}, t; \mathbf{Z}_{1:t}) \equiv g(\mathbf{s}, t; \hat{\mathbf{X}}_t). \tag{1.6}$$

Note that this estimator $\tilde{g}$ is different from $\hat{g}$ given by (1.4). While (1.4) is an unbiased estimator of the danger field in the sense that $E[\hat{g}] = E[g]$, the estimator (1.6) will generally be biased. The size of the bias depends on the degree of non-linearity of the damage-potential function $\delta(\cdot)$; see (1.1) for an example of a non-linear $\delta(\cdot)$.

While (1.4) 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 (1.4),

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) (Anderson and Moore (1979)). 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 means, variances, and covariances of the unknown state variables in $\mathbf{X}_t$.

Another approach is the Unscented Kalman filter (UKF), due to Julier and Uhlmann (1997); see also van der Merwe et al. (2000). This is based on the Unscented Transformation and the Scaled Unscented Transformation (Julier (1999)). 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) (Evensen (1994); Heemink (2000)). 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) (Cohn and Todling (1996); Heemink (2000)). 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 (Heemink (2000)). 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 (see van der Merwe et al. (2000); Liu et al. (2001); Gordon et al. (2001)). SIS has been used successfully to analyze data from terrain navigation (Bergman (2001)), genetics (Irwin et al. (1994)), and image analysis (Blake et al. (2001)), for example. Another Monte Carlo approach, known as Markov Chain Monte Carlo, or MCMC (Smith and Roberts (1993); Gelman et al. (1995)), 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 $\{\mathbf{X}_t^{(i)} : i = 1, \ldots, N\}$, 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 (1.4), by approximating it with the (weighted) av-

erage of the associated danger fields $\left\{g(\mathbf{s}, t; \mathbf{X}_t^{(i)}) : i = 1, \ldots, N\right\}$. 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 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 danger could be of great use to a 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 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 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 non-linear 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 (Doucet

(1998); Pitt and Shephard (1999)). A more promising suggestion is the Unscented particle filter (UPF), where the Unscented Kalman filter is used to generate the proposal distribution (van der Merwe et al. (2000, 2001)); 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. Analysis 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

$$p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t}) = \frac{p(\mathbf{X}_{1:t})p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t})}{p(\mathbf{Z}_{1:t})} \propto p(\mathbf{X}_{1:t})p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t}), \tag{2.1}$$

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 (Gelman et al. (1995)).

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 tractable 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:

$$\begin{aligned} E_p[h(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}] &= \int h(\mathbf{X}_{1:t})p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= \int h(\mathbf{X}_{1:t})\frac{p(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})}q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= \int h(\mathbf{X}_{1:t})w_t(\mathbf{X}_{1:t})q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t})d\mathbf{X}_{1:t} \\ &= E_q[h(\mathbf{X}_{1:t})w_t(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}], \end{aligned} \tag{2.2}$$

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

$$\begin{aligned} \hat{E}_p[h(\mathbf{X}_{1:t})|\mathbf{Z}_{1:t}] &\equiv \frac{\sum_{i=1}^N h(\mathbf{X}_{1:t}^{(i)})w_t(\mathbf{X}_{1:t}^{(i)})}{\sum_{i=1}^N w_t(\mathbf{X}_{1:t}^{(i)})} \\ &= \sum_{i=1}^N h(\mathbf{X}_{1:t}^{(i)})\tilde{w}_t(\mathbf{X}_{1:t}^{(i)}), \end{aligned} \tag{2.3}$$

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

$$q(\mathbf{X}_{1:t}|\mathbf{Z}_{1:t}) = q(\mathbf{X}_1|\mathbf{Z}_1)\prod_{k=2}^{t} q(\mathbf{X}_k|\mathbf{X}_{1:k-1}, \mathbf{Z}_{1:k}). \tag{2.4}$$

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 et al. (2002)), 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 (Kong et al. (1994); Irwin et al. (1994); Liu and Chen (1998)). 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 $\left\{w_t(\mathbf{X}_{1:t}^{(i)})\right\}$ can be decomposed as,

$$w_t(\mathbf{X}_{1:t}^{(i)}) = w_{t-1}(\mathbf{X}_{1:t-1}^{(i)})\frac{p(\mathbf{X}_{1:t}^{(i)}|\mathbf{Z}_{1:t})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{1:t-1}^{(i)}, \mathbf{Z}_{1:t})p(\mathbf{X}_{1:t-1}^{(i)}|\mathbf{Z}_{1:t-1})}. \tag{2.5}$$

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

$$p(\mathbf{X}_{1:t}) = p(\mathbf{X}_1)\prod_{k=2}^{t} p(\mathbf{X}_k|\mathbf{X}_{k-1}); \tag{2.6}$$

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

$$p(\mathbf{Z}_{1:t}|\mathbf{X}_{1:t}) = \prod_{k=1}^{t} p(\mathbf{Z}_k|\mathbf{X}_k). \tag{2.7}$$

Then recall that the optimal choice for q is,

$$q^*(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) \propto p(\mathbf{X}_t|\mathbf{X}_{t-1})p(\mathbf{Z}_t|\mathbf{X}_t). \quad (2.8)$$

However, even though the structure is relatively simple, (2.8) may still not be tractable for easy generation of importance samples; the model to be presented in Section 5 is one such example. Often, the importance sampling distribution (2.8) 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 (2.8) is, up to a normalizing 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 (van der Merwe et al. (2000)) 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 (2.4) 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 (2.7), the update formula (2.5) for the unnormalized importance sampling weights satisfies,

$$w_t(\mathbf{X}_{1:t}^{(i)}) \propto w_{t-1}(\mathbf{X}_{1:t-1}^{(i)}) \frac{p(\mathbf{X}_t^{(i)}|\mathbf{X}_{t-1}^{(i)})p(\mathbf{Z}_t|\mathbf{X}_t^{(i)})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{t-1}^{(i)}, \mathbf{Z}_t)}. \quad (2.9)$$

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 (Kong et al. (1994)). 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 (see Liu et al. (2001); van der Merwe et al. (2000)). 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 (Gordon (1994)), Residual Sampling (Higuchi (1997); Liu and Chen (1998)), and Minimum Variance Sampling (Kitagawa (1996); Crisan (2001)).

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 (2.5).

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 (2001) recommends monitoring the weights by their coefficient 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 (2.2) and (2.3), 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 (Cressie et al. (2002)) 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.

Each of the five tanks has the same damage-potential parameters. The damage-potential function $\delta(\cdot)$ is given by (1.1) 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.

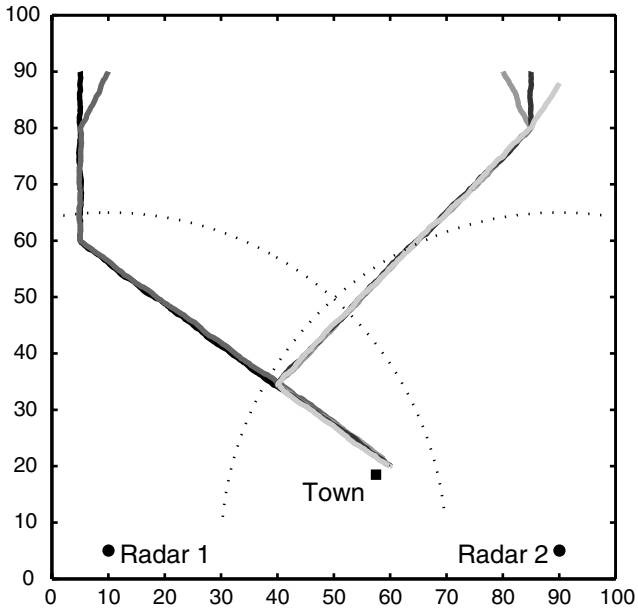

Figure 1: *Battlespace features. The solid lines, in different grey scales, show the five true paths of the tanks in the 100 km x 100 km battlespace. The dotted lines show the range of the two radar stations. Also displayed are the locations of the two radar stations and the town.*

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 Wendt et al. (2002)). Note that for this example, the tanks have proximate waypoints and they are travelling 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,

$$\begin{aligned} v_{t+1} &= \rho_v v_t + (1-\rho_v)V_{t+1}, \\ \theta_{t+1} &= \rho_\theta \theta_t + (1-\rho_\theta)\Theta_{t+1}, \end{aligned} \tag{4.1}$$

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,

$$\begin{aligned} y_{1,t+1} &= y_{1t} + (v_{t+1}\Delta t)\cos(\theta_{t+1}), \\ y_{2,t+1} &= y_{2t} + (v_{t+1}\Delta t)\sin(\theta_{t+1}), \end{aligned} \tag{4.2}$$

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 (Cressie et al. (2002)). 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,

$$\begin{aligned} E[\mathbf{Z}|\mathbf{Y}] &\simeq \mathbf{Y} \\ var[\mathbf{Z}|\mathbf{Y}] &\simeq [\sigma_r^2 - r_1^2(1-\gamma) - \sigma_\theta^2(\sigma_r^2 - r_1^2\gamma)]\mathbf{p}\mathbf{p}' + \sigma_\theta^2[\sigma_r^2 - r_1^2\gamma]\mathbf{q}\mathbf{q}', \end{aligned} \tag{4.3}$$

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.

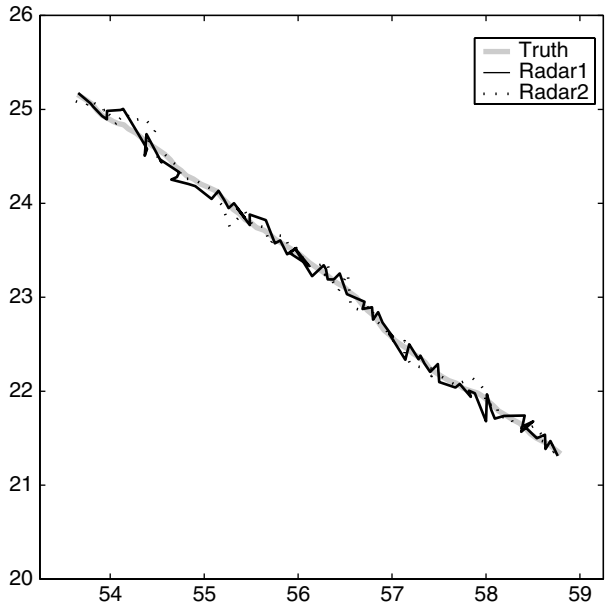

Figure 2: *True path for tank 3 along with the two sets of radar observations for the last 20 minutes.*

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:

$$\begin{aligned} \log v_{t+1} &= \log v_t + \epsilon_{v,t+1}, \\ \theta_{t+1} &= \theta_t + \epsilon_{\theta,t+1}, \end{aligned} \tag{5.1}$$

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,

$$\begin{aligned} y_{1,t+1} &= y_{1t} + \frac{1}{2}(v_{t+1}\cos\theta_{t+1} + v_t \cos\theta_t) \\ y_{2,t+1} &= y_{2t} + \frac{1}{2}(v_{t+1}\sin\theta_{t+1} + v_t \sin\theta_t), \end{aligned} \tag{5.2}$$

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 (2.6).

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 (5.1), (5.2) does *not* match the algorithm that describes the manner in which the data were simulated; see (4.1), (4.2). 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 (4.3) 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 (4.3), 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 analysis 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 (5.1) and (5.2) 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.

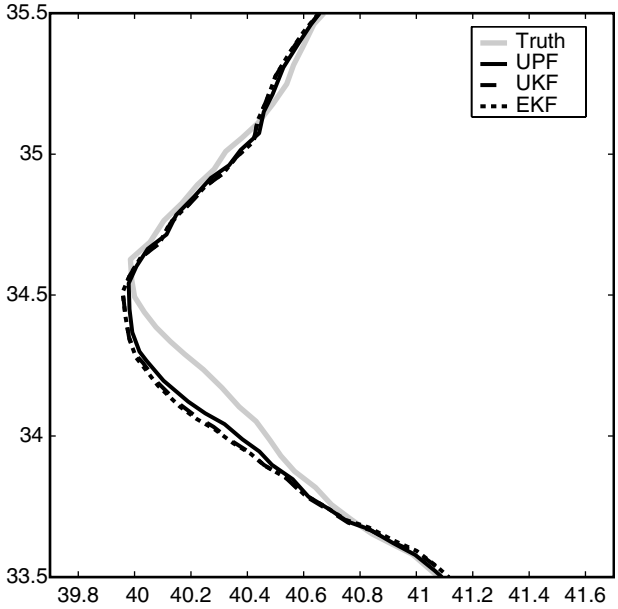

Figure 3: *Estimated path for tank 3 for the three estimators, during the abrupt change of direction shown on the true path.*

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 estimators are both plug-in estimators based on (1.6), 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.

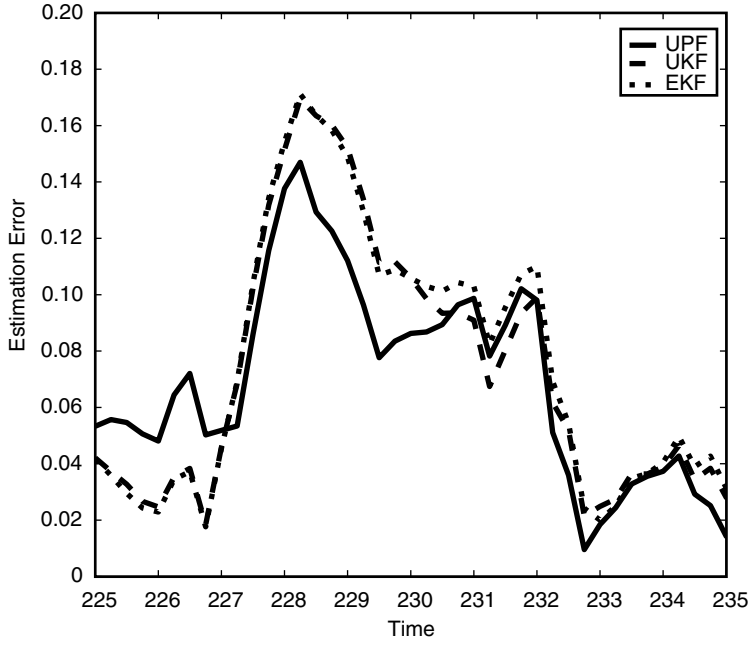

Figure 4: *Error in estimated position for tank 3 for the three estimators, during the abrupt change of direction shown in Figure 3.*

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

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

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 methods, which is to be expected, since the estimated tank locations 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

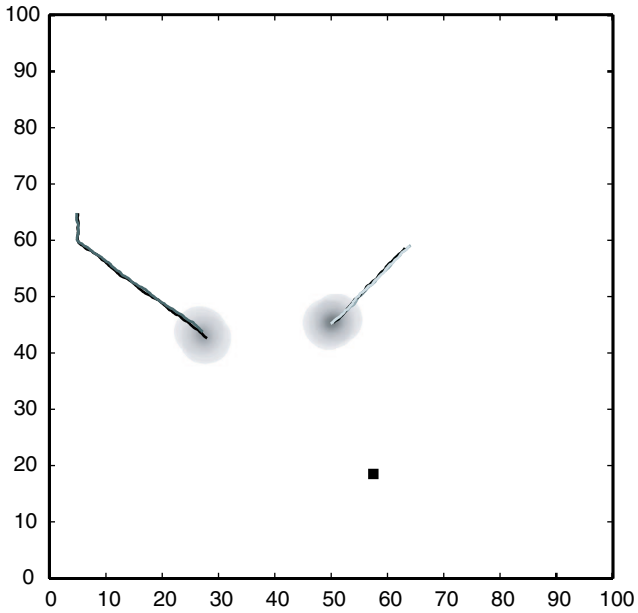

Figure 5: *Estimated danger field at 3 hours based on the UPF estimator. The darker the gray color, the higher the danger. The estimated paths for the 5 tanks are included.*

center's posterior mean danger (1.4). 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,

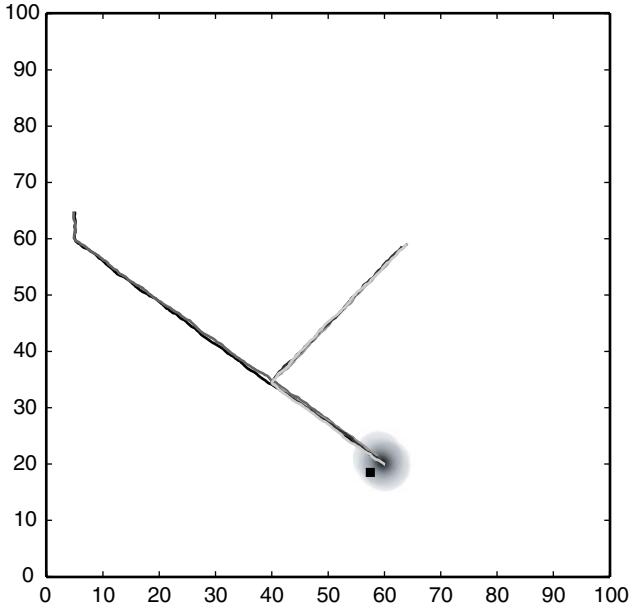

Figure 6: *Estimated danger field at 5 hours based on the UPF estimator. The darker the gray color, the higher the danger. The estimated paths for the 5 tanks are included.*

this doesn't hold for the 1-sample-per-minute sampling frequency. Instead, there is a sawtooth pattern in the width that 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 (1.6), 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

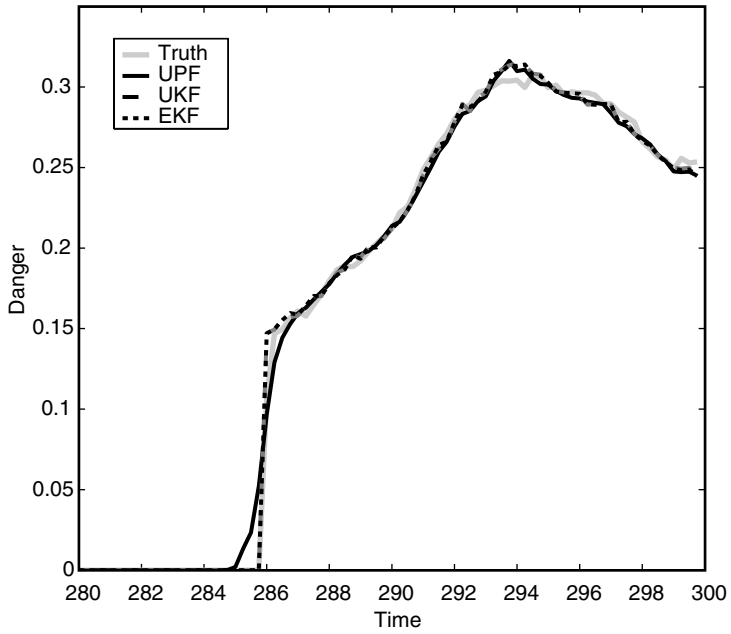

Figure 7: *Estimated danger at the center of town due to tank 5 for the last 20 minutes.*

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.

6 Discussion and conclusions

For the danger field to be a useful tool to a 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

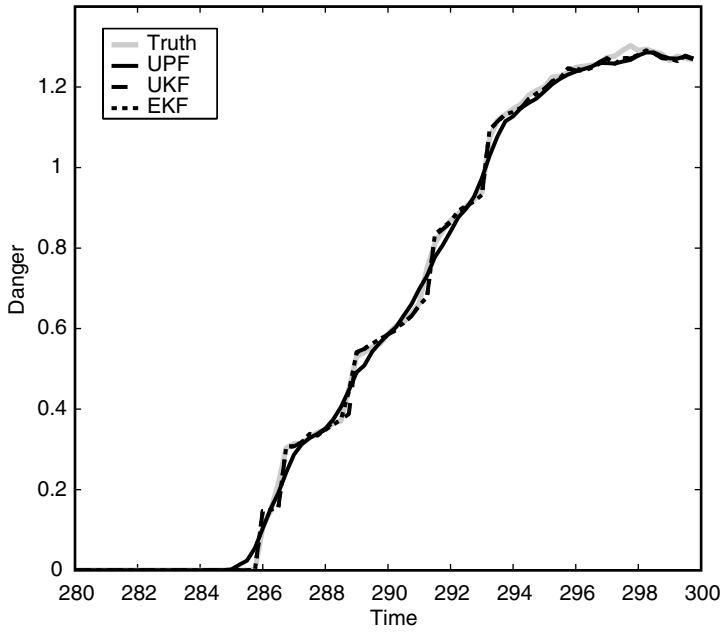

Figure 8: *Estimated danger at the center of town due to all 5 tanks for the last 20 minutes.*

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 (1.1) is highly non linear in this situation because small changes in distance from the target may lead to large changes in damage.

While the UKF can miss important features, it can be useful when run simultaneously with the UPF. The UKF can be used to give instantaneous, though possibly rough, answers. If alerted by the UKF results, the UPF

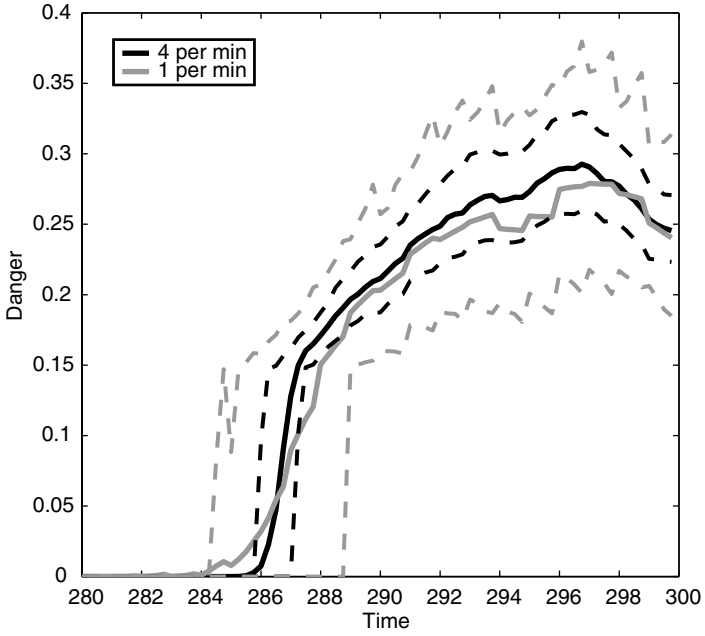

Figure 9: *Estimated danger (solid lines) with 90% credibility intervals (dashed lines) due to tank 1 for two sampling frequencies (4 per minute and 1 per minute) for last 20 minutes.*

running in the background could then be used to derive more accurate results if needed. This approach should have the advantage 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 instantaneous. 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 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. (1997) 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 express 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. (2000). 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 Transformation (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. (2000), 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:

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

where $\left(\sqrt{(n_a+\lambda)\mathbf{C}_t}\right)_j$ is the jth column of a matrix square root of the matrix $(n_a+\lambda)\mathbf{C}_t$, and $\lambda = \alpha^2(n_a+\kappa) - n_a$ is a scaling parameter. The parameters 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 $\boldsymbol{\mathcal{A}}_{j,t} = \left[\left(\boldsymbol{\mathcal{A}}_{j,t}^x\right)^T \ \left(\boldsymbol{\mathcal{A}}_{j,t}^\delta\right)^T \ \left(\boldsymbol{\mathcal{A}}_{j,t}^\epsilon\right)^T\right]^T$. Then define

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

(d) Measurement update:

Define

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

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 (1.6), obtained by "plugging in" $\boldsymbol{\mu}_t$ into (1.6).

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 $\{\mathbf{X}_t : t = 0, 1, 2, \ldots\}$ 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 (van der Merwe et al. (2001)).

1. Set $t = 1$ and initialize the filter:
   (a) Sample $\mathbf{X}_0^{(1)}, \ldots, \mathbf{X}_0^{(N)}$ from $p(\mathbf{X}_0)$.
   (b) Initialize the importance sampling weights $w_0\left(\mathbf{X}_0^{(i)}\right) = \frac{1}{N}; i = 1, \ldots, N$.
2. For $t > 0$,

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

$$\mathbf{m}_t = E[\mathbf{A}_t] = \left[\mathbf{d}_t^T \ \mathbf{e}_t^T\right]^T$$
$$\mathbf{C}_t = var[\mathbf{A}_t] = \begin{bmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_t \end{bmatrix}.$$

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:

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

where the tuning parameters are as specified in Appendix A.

(c) Time update:

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

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

(d) Measurement update:

Define

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

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

$$\mathbf{K}^{(i)}_t = \mathbf{C}^{(i)}_{\mathbf{X}_t\mathbf{Z}_t}\left(\mathbf{C}^{(i)}_{\mathbf{Z}_t\mathbf{Z}_t}\right)^{-1}$$

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

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

(e) Importance sampling:

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

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

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.

DISCUSSION

José M. Angulo and María D. Ruiz-Medina
Universidad de Granada, Spain

In this interesting review paper the authors present a detailed discussion, well connected with the current literature, on spatial-temporal nonlinear filtering, with a clear exposition of sequential Monte Carlo approaches (known as *Sequential Importance Samplers* (SIS) or *particle filters*) recently proposed. They illustrate and compare different methods with a simulation study developed in a Command and Control setting. Specifically, they compare the Unscented Particle filter (UPF) and the Unscented Kalman filter (UKF) to the Extended Kalman filter (EKF), concluding that UPF-based estimates outperform Kalman-filter-based estimates in the approximation of nonlinear features of the functional of interest.

A key aspect in the design of an effective SIS is the definition of the conditional sampling distribution or importance sampling distribution. In this respect, the authors comment that SIS can be improved using Kalman-filter-based methods (EKF and UKF) in the calculation of such a distribution. However, the use of the extended Kalman particle filter leads to a linearized calculation of the importance sampling distribution. Furthermore, the calculation of the importance sampling distribution by using UKF involves the second-order moments of the conditional distribution of the state process at time t given the state at time $t-1$. The use of a Gaussian approximation to the importance sampling distribution can be interpreted as an implicit linearization of the system, since the functional of interest is defined as a nonlinear function of a Gaussian process. We think that the use of maximum-entropy-based methods (see, for example, Christakos (2000)) to obtain the importance sampling distribution would improve SIS-based estimates in the description of nonlinear features.

The model considered in the application to Command and Control (C2) is essentially temporal, since the movement equations, describing the weapon's state in terms of the vectorial process $\mathbf{X}_t = (Y_{1t}, Y_{2t}, V_t, \theta_t)$, reflect the evolution of the components of $\mathbf{X}$, and since, in the definition of the danger-potential field, the possible spatial-temporal interaction between different weapon state processes is not taken into account. Thus, the

application considered basically provides an illustration of how the above-mentioned nonlinear filtering techniques work in the temporal case. It would be very interesting to illustrate their performance with a more complex model involving spatial-temporal interaction. For example, in the context (C2) considered, the incorporation of the spatial-temporal interaction between weapons to the state model for the definition of the damage potential and the danger field would be of interest. In the definition of the observation model, an interesting extension would be to consider the optimum design of the placement of observation devices, possibly moving over time, jointly with the time sampling frequency.

In our opinion, some interpretations derived from the application, which are used to establish the conclusion in the paper on the outperformance of UPF, are not strongly supported by the results obtained. For instance, one can see that UPF-based estimates of position (see Figure 4) and danger (see Figures 7 and 8) are sometimes better, sometimes worse than EKF-based or UKF-based estimates. Furthermore, we cannot agree with the appreciation on the UPF-based estimates being advantageous for 'showing evidence of danger earlier', since by the same way of thinking one would be led to say that 1 minute sampling gives better results than 1/4 minute sampling (see Figure 9), which would be paradoxical. Further clarifications on these aspects would be appreciated.

Rong Chen
University of Illinois at Chicago, USA
Jun S. Liu
Harvard University, USA

We congratulate the authors for an excellent review of some Sequential Monte Carlo (SMC) algorithms for nonlinear filtering and a very interesting application using SMC. It has been shown that SMC is a powerful tool in handling nonlinear and non-Gaussian dynamic systems. This paper provides another evidence in this regard. Here we would like to make one general comment and discuss two issues, one general and one specific.

SIS versus MCMC: It is commonly believed that SMC and its variations are just cheap (and inferior) alternatives to the more computationally demanding MCMC procedures. This is true to a certain extent, especially

for a class of the state-space models. However, the extreme flexibility of SMC methods sometimes make them the primary choices for certain problems, with efficiencies far exceeding that of the standard MCMC procedure. Two dramatic examples in this direction are the counting and inference of zero-one tables with fixed margins (Liu (2001, §3.4.2 and §4.3)) and the simulation of long-chain polymers (Zhang and Liu (2002)). These examples demonstrate that innovative SIS designs can outperform most MCMC schemes in certain difficult problems where the involved variables (or part of which) are highly correlated (or interlocked). It has also been shown that using MCMC steps in SMC and, conversely, using SMC in MCMC can be beneficial (Liu (2001)).

The use of the current observation Z_t: The authors advocate the use of information in the current observation Z_t to construct the sampling distribution, first proposed by Liu and Chen (1998). In UPF, significant amount of computational resources are devoted to achieve this, using the scaled unscented transformation. It should be noted that the efficiency of SMC is not solely measured by the variance of the weights. The amount of computation is also an important issue. A good measure should be the accuracy of the final estimate given the same amount of computational time. For example, if UPF uses 10 times of computation as a simple particle filter (SPF) that uses only the state equation, then one should compare the accuracy of the final estimate between a UPF using m samples and a SPF using $10m$ samples. The key here is actually how much information the current observation Z_t brings in. Let us examine two extreme situations: (1) If the observation noise is very large, then there is virtually no benefit to include Z_t in the sampling. In this case, a SPF with 10 times more samples would probably work better. (2) If the observation noise is very small, or the state equation is not adequate (e.g. one sample per 3 minutes for the battle field example), then Z_t becomes very important. In fact, in this case one can do the complete opposite of the SPF – using only the observation equation for sampling and the state equation for updating the weight. One way of doing this is to generate $Z_t^{(j)} = Z_t + e_t^{(j)}$ and construct $X_t^{(j)}$ by inverting the observation equation using $Z_t^{(j)}$, if it is not too difficult. Procedures such as UPF may show significant benefit when the information from both the observation equation and the state equation are comparable.

The Movement Model: The movement model used in the example does not have a maneuvering component. A very simple multilevel model (e.g. Bar-Shalom and Fortmann (1988, p. 125–127), Chen and Liu (2000)) can be very useful in cases like this. Figure 3 in the paper shows that the current system has some difficulties tracking the target after a maneuvering at the pathway. It may be partially due to the fact that the filter "trusts" the movement model more than the observation model (due to the relatively large observation measurement error), even at the time the movement model is not true. With a multilevel model which assumes several levels of uncertainty in the movement model, the filter may be able to detect persistent one-sided deviations between the estimated states and the observations, hence engage a maneuvering mode to increase the uncertainty in the movement model. Then the filter will weight more on the observations and allows to track the maneuvering better.

Montserrat Fuentes
North Carolina State University, USA

Irwin, Cressie and Johannesson (ICJ) have provided an interesting review of several generalizations of the Kalman-filter algorithm for nonlinear problems. ICJ compare the performance of two recent methods of filtering, Unscented Particle filter (UPF) and Unscented Kalman filter (UKF) to the traditional Extended Kalman filter (EKF) in a Command and Control (C2) setting. This article is Bayesian in the sense that the main focus is on currently updated posterior distributions.

I certainly agree that the Bayesian perspective on the Kalman filter is the most natural way of viewing this sequential estimation procedure that predicts the dynamically changing configuration of objects in a C2 setting. My specific comments are of two types: First, regarding the movement model in the C2 setting presented by the authors, and secondly regarding the estimation of relevant parameters. I present here an alternative model of the the movement and suggest a different danger potential. In the illustration to the C2 setting that ICJ present, the state parameters are fixed and treated as known. I suggest here a procedure to estimate parameters and take into account the uncertainty about these parameters in the subsequent prediction.

There are both deterministic and stochastic approaches to describe the trajectory of moving objects. Motion in Newtonian dynamics is generally described by a potential function, $P(y,t)$ (see Nelson (1967)). Here y is location and t is time. The equation of motion takes the form

$$\begin{aligned} dy(t) &= v(t)dt \\ dv(t) &= -\beta v(t)dt - \beta \nabla P(y(t),t)dt \end{aligned} \tag{1}$$

with $y(t)$ the object's location at time t, $v(t)$ is the object's velocity, ∇P the external force field acting on the object, β is the coefficient of friction, and ∇ is the gradient operator. The function P can be interpreted as controlling the object's direction and velocity. For example

$$P(y) = 1/|y-a|^2$$

corresponds to a point of repulsion at a.

ICJ incorporate a stochastic component in the movement model equation for their analysis by simply adding some white noise to the speed and direction. I suggest instead a stochastic version of the equation of motion (1). A probabilistic concept for dynamic situations is a stochastic differential equation (SDE), e.g. Nelson (1967), Bhattacharya and Waymire (1990). Such equations lead to Markov processes and take the form

$$dy(t) = \mu(y(t),t)dt + \Sigma(y,t)dB(t), \tag{2}$$

where B is a bivariate Brownian motion, μ is the drift parameter and Σ is the variance matrix or diffusion parameter. Many properties are known concerning solutions of SDEs, and this model has been successfully used to estimate trajectories of different objects (for instance by Brillinger et al. (2000)). I recommend movement model (2) for the analysis of C2. Thus, we can model the movement of the objects when they circle around a city, the repulsion between objects (so the tanks do not crush), the terrain effects or many other sources of external forces that might be acting on the object.

It seems a little bit arbitrary how ICJ define the function representing the potential damage. Also they do not provide much insight regarding the interpretation and estimation of the potential damage parameters. In my opinion the selection of the damage potential deserves more attention, because this function is one of the most important and critical factors to model the damage of mobile weapons in a C2 setting. In the simulation

presented by ICJ, the parameters are fixed. In a real application, the parameters need to be estimated, and most likely they will be different for different objects. Thus, I suggest to reduce the number of parameters to three and use as damage potential

$$\alpha e^{r/R^p},$$

where α is the maximum damage potential, R the explosive radius, and p the parameter that controls how smooth (differentiable) this function is. The function proposed by ICJ can lead to problems when using likelihood based methods for estimating the unknown parameters, especially for some values of the power parameters. If the computational advantage (and the easier interpretation) of having the damage potential identically 0 beyond some distance is needed in a particular application, we could use then a truncated version of the function proposed here. I personally prefer to avoid functions that are identically 0 beyond some value, because they are always problematic when using a likelihood method for estimation.

ICJ treat the parameters of interest, i.e. the parameters in the movement model and the damage potential, as known. In a real C2 setting we need to estimate these parameters and quantify the uncertainty about the estimation. A plug-in algorithm would not capture the uncertainty in the trajectory prediction of the estimation of the parameters. Also, when the number of parameters is large, it is common to have problems with lack of identifiability. If we use proper priors for all parameters, then the solution (the prediction) will most likely be completely determined by the priors. If we put uniform priors to all parameters, we might get improper posteriors, especially once we add some more structure in the equation for the movement, and we allow for some parameters to explain the spatial dependency. So, there is not an easy answer.

A Bayesian attempt to solve the problem of unknown parameters in the movement model is the use of discount factors (West and Harrison (1997)). Suppose that we define the state evolution by

$$X_t = G_t X_{t-1} + u_t,$$

where $u_t \sim N(0, W_t)$. If $\delta \in (0, 1]$ is the known discount factor and C_{t-1} is the variance of X_{t-1} given all the information of the sample at time $t-1$, then it is easy to show that $W_t = G_t C_{t-1} G_t'(1-\delta)/\delta$, thus solving the problem of specification of the unknown value of W_t. This approach would

be particularly easy to implement for prediction of trajectories because the state model is almost linear. Discount factors can be also used with nonlinear models (West et al. (1985)). A procedure to estimate the unknown parameters in the state equation and the observation model, is to integrate expressions with respect to X_t and to obtain predictive likelihood depending only on the unknown hyperparameters (Harvey (1989)). This likelihood could be maximized to estimate the hyperparameters. The estimation of unknown quantities and the uncertainty in the prediction about the estimated parameters should be carefully addressed in hierarchical Bayesian models. This is particularly important in C2 settings, where is critical to understand the degree of confidence in the prediction of trajectories for mobile weapons, because of costly economic and social consequences.

John Kent and Kanti Mardia
University of Leeds, UK

One aspect of this paper involves time series models for angular data. The topic has not received much serious attention in the literature, especially for tracking problems, mainly due to a lack of analytically tractable models. For simplicity we limit attention here to AR(1)-type models, or equivalently, Markov chain models on the circle, though many of the ideas generalize easily to higher-order processes on higher-dimensional spheres.

After removing a mean effect from equation (4.1) in the paper, the model for the angular component takes the linear AR(1) form,

$$\theta_{t+1} = \rho\theta_t + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0, \sigma_\epsilon^2), \tag{1}$$

with $\theta_t \in (-\pi, \pi)$. This model is fine for concentrated data, but is less satisfactory for dispersed data due to the lack of continuity of the drift term at $\theta_t = \pi$ $(= -\pi \mod 2\pi)$. Thus it is worth exploring models which better accommodate the periodic nature of θ_t.

A natural starting point is to note that any conditional bivariate angular density $f(\theta_{t+1}|\theta_t)$ can be used to generate a Markov chain. For a current review of various possibilities see, e.g. Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001). Some more recent contributions are given by Downs and Mardia (2002) and Singh et al. (2002).

Let $x = (\cos\theta, \sin\theta)^T$ be a unit vector on $\mathbb{R}^2$. The von Mises distribution for x, written $M(\mu)$ in terms of a two-dimensional parameter vector

$\mu = \kappa(\cos\alpha, \sin\alpha)^T$, has "mean direction" α and "concentration parameter" κ. One of the simplest bivariate conditional models takes the form

$$x_{t+1}|x_t \sim M(a\ x_t + b\ e_1) \tag{2}$$

(Breckling, 1989, p. 175), where a and $b > 0$ are scalars and $e_1^T = (1, 0)$. Under large concentration (with $a + b \gg 0$), this model reduces to the linear AR(1) model above, with first order auto-correlation $\rho = a/(a + b)$ and innovation variance $\sigma_\epsilon^2 = 1/(a + b)$. It turns out that the equilibrium distribution of this process has density

$$I_0(\parallel ax + b\ e_1 \parallel) \exp(b\ e_1^T x), \tag{3}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order 0. Note that in this approach the conditional distributions are tractable, but the marginal bivariate distribution is awkward.

An alternative AR(1)-type process on the circle in continuous time (Kent (1978)) can be generated by the SDE

$$d\theta_{t+1} = -\lambda \sin\theta_t\ dt + dw_t,$$

where w_t is a real-valued Brownian motion. This process is analogous to the Ornstein-Uhlenbeck process on the line. In this case the equilibrium distribution is tractable, $M(2\lambda e_1)$, but the conditional distributions are intractable. Intractability of at least one or other of the marginals or conditionals in directional models seems to be a universal "law".

Reverting to the model in equation (1), the next step is to treat the $\{x_t\}$ process as an unobserved signal which is combined with an observation process

$$y_t|x_t \sim M(\kappa x_t), \quad \kappa > 0. \tag{4}$$

Then Kalman filter ideas can be investigated to recover the signal from the observations. Unfortunately, although the conditionals (2) and (3) are tractable, the conditional distribution of $x_t|(y_1, \ldots y_t)$ is not. This lack of tractability has hampered the development of these models. Recent advances in particle filters and related ideas should give a boost to the area. Indeed, a key example of this new approach is Pitt and Shephard (1999) who implemented a particle filter approach for a tracking problem involving a wrapped Cauchy observational process.

Jorge Mateu
Universitat Jaume I de Castellón, Spain.
Francisco Montes
Universitat de Valencia, Spain.

We enjoyed reading this stimulating paper, which is both interesting and useful. It is a pleasure for us to congratulate the authors for an able and comprehensive survey of an area of statistics which is only now assuming the importance which its many applications demand. Particularly welcome in this account is the variety of model proposals, based on different postulates about underlying mechanisms. It is surely crucial that models should be chosen not only for their mathematical convenience, but because they reflect the scientist's insight into the nature of the phenomena observed.

This paper presents a review of spatial-temporal nonlinear filtering combining both sound theory and constructive practice.

Developing appropriate stochastic models and statistical methods for space-time processes is one of the great open challenges in statistics. In the last decade, a rapidly expanding literature has been experienced. In particular, when the state equations of the space-time process are nonlinear and the aim is focused on dynamically filtering noisy observations, the available research references are not so extensive, and in this sense this paper is a welcomed addition to this area of the scientific research based on spatial-temporal modelling. As illustrated below we find this a promising direction of research.

It is our belief that modelling this and similar datasets combining statistical and physical models can only bring good scientific reports. Indeed, the authors state that the scientist's physical/chemical/biological knowledge is to be considered in the process model. We would expect that such models would produce better predictions than models using either just a physical approach or just a statistical approach. Furthermore, the differences between the observations and a physical model prediction may have a simpler structure than the observations themselves, for example, by being more nearly stationary and isotropic in space. To carry out such a combined modelling approach would undoubtedly take the active collaboration of statisticians and atmospheric scientists and we can only see good things coming out of such an endeavor.

A first appreciation of this paper reminds us of several problems within the computational intelligence (CI) research area. And here are some lines to remind the reader of these kind of problems.

The 1990s were witnessing the beginnings of a new quantitative revolution based upon computational intelligence technologies very relevant to spatial analysis. The *raison d'ĉtre* of CI-based spatial analysis is to exploit the tolerance for imprecision and uncertainty in large-scale spatial analysis problems in an analysis process which is increasingly driven by the availability of huge quantities of spatial data, and to achieve tractability, robustness, computational adaptivity, real-time analysis and low cost.

There are two principal areas of CI that are particularly relevant to spatial analysis: *evolutionary computation* which includes genetic algorithms and artificial life; and *neural networks* which are also known as neurocomputing. Biologically inspired evolutionary computation (i.e. genetic algorithms, genetic programming, non-Darwinian evolutionary algorithms etc.) provide the basis for developing new solutions to complex spatial optimization problems as well as building blocks for new kinds of spatial analysis techniques and models; for example, the artificial life based pattern hunting creatures or the automated modelling system developed by Openshaw and associates (Openshaw (1994)).

Much of the recent interest in neural network modelling in spatial analysis stems from the growing realization of the limitations of conventional tools and models as vehicles for exploring patterns and relationships in GIS (geographical information systems) and RS (remote sensing) environments and from the consequent hope that these limitations may be overcome by judicious use of neural net approaches. The attractiveness of these approaches extends far beyond the high computation rates provided by massive parallelism and essentially stems from the following features: (a) the greater representational flexibility and freedom from linear model design constraints; (b) the built-in capability to incorporate rather than ignore the special nature of spatial data; (c) the greater degree of robustness or fault tolerance to deal with *noisy data*, *missing and fuzzy information*; (d) the ability to deal efficiently with very large data sets and thus to provide the prospect to obtain better results by being able to process finer resolution data or real-time analysis.

Neural networks which may be viewed as non-linear extensions of conventional spatial statistical models are applicable to two major domains:

first, as universal function approximators to areas such as spatial regression, spatial interaction, spatial choice and space-time series analysis; and second, as pattern recognizers and classifiers to intelligently allow the user to sift through the data, reduce dimensionality, and find patterns of interest in data-rich environments. In this sense, and the analysis presented in this paper actually brings light on this, neural pattern classifiers have an important role to play in high dimensional problems of pattern recognition and classification in remote sensing environments.

Another aspect it is worth noting here, is that the methodological proposals presented (and reviewed) in this paper can be clearly adapted to solve many other problems coming in a variety of scientific fields. The authors specify a body of applications within the area of armed forces. But the field of possible real applications goes further than that. For example, biological or ecological problems related to migratory movements of several species would benefit from this kind of methodology. Controlling, chasing and estimation of movement paths through large regions is an important problem in this field. Also, hydrogeological problems could be considered. With the necessary adaptation of the proposed methodology, in terms of movement and transport equations, we could use spatial-temporal nonlinear filtering methods to propose statistical modelling of chemical (or other substances) flows within certain classes of aquifers.

Our methodological concern relates to the empirical Bayesian versus a fully Bayesian approaches. It is shown in the paper that UPF does a better job of accounting for the uncertainties in the tank locations and danger field compared to the performance shown by UKF and EKF estimators. The latter estimators are both plug-in estimators and do not take into account these kind of uncertainties. Thus, our query is the following. Using a fully Bayesian methodology, could the uncertainty be taken into account in a more methodological fashion through proposing priors to the model parameters, and under this case, could UKF and EKF be more efficient at detecting these kind of uncertainties?

Finally, though in a different context, in Gregori et al. (2002) we have developed a class of spatial point processes, called *generalized area-interaction point processes*, which are particularly devoted to solve problems in certain communication applications, for example detection of cell-phones in continuous movement and its relation to the corresponding fixed location of the reception antennas. These processes can be generalized to incorpore

spatio-temporal information, and in this sense could be related to the aim of the Irwin, Cressie and Johannesson paper.

Michael L. Stein
University of Chicago, USA.

Irwin, Cressie and Johannesson provide a nice review of various recent innovations in nonlinear filtering and demonstrate how methods based on as the scaled unscented transformation (SUT) can be successfully applied to a problem of assessing the danger of highly mobile weapons. The basic idea behind the SUT, whether as part of the UKF or the UPF, is to use a deterministic sample from the joint conditional distribution of the state, the process error and the measurement error given a current estimate of the state of the system to determine an updated mean and covariance matrix for the state of the system and for the possibly nonlinear functions of the state that are of practical concern. In comparison to the extended Kalman filter, one is replacing derivatives by finite differences, with the potential advantage of better capturing the nonlinear behavior of the system. In comparison to simple random sampling, deterministic sampling should generally provide much more precise estimates of conditional means and covariance matrices for the quantities of interest. However, as with any systematic sample, there is always the potential for serious bias. Furthermore, there is considerable ground between a simple random sample and a purely deterministic sample. I would like to explore the possibility of adding some randomness to the SUT as a way of getting the good properties of both deterministic and random sampling.

Neither Irwin, Cressie and Johannesson nor the few other references I have consulted make any mention of how to choose the matrix square root $\mathbf{C}_t^{1/2}$ in step 2(b) of the UKF and UPF algorithms. Two obvious choices are the square root obtained from the Cholesky decomposition and the symmetric positive definite square root. If the system is linear and the quantities of interest are linear in the state, then the SUT is constructed so that it does not matter which square root is used. To see how the choice of square root can matter for a badly nonlinear system, suppose $\mathbf{C}_t$ is diagonal, in which case, both the Cholesky decomposition and the symmetric positive definite square root give the elementwise square root of $\mathbf{C}_t$. Now suppose $u(\mathbf{X}_{t-1}, \boldsymbol{\delta}_t) = \mathbf{X}_{t-1} + \delta_{t1}^4 \mathbf{v}$, where δ_{tj} is the jth component

of $\boldsymbol{\delta}_t$ and $\mathbf{v}$ is some fixed vector. Then $E(\mathbf{X}_t \mid \mathbf{X}_{t-1}) = \mathbf{X}_{t-1} + 3\sigma_{\delta 1}^4 \mathbf{v}$, whereas $\boldsymbol{\mu}_{t|t-1} = \mathbf{X}_{t-1} + (n_a + \lambda)\sigma_{\delta 1}^4 \mathbf{v}$, which is increasingly in error as n_a increases. A general principle of statistical design is that if there is a choice to be made between differing designs for which one does not have any substantive basis for preferring one to another, the choice should be made randomly as a way of guarding against unanticipated sources of bias. This principle suggests randomly rotating $\mathbf{C}_t^{1/2}$: replacing $\mathbf{C}_t^{1/2}$ by $\mathbf{C}_t^{1/2}\mathbf{H}$, where $\mathbf{H}$ is an orthogonal matrix that rotates vectors by an angle chosen from the uniform distribution on the unit sphere. For large n_a, using a random rotation will approximately remove the bias in the estimate of $E(\mathbf{X}_t \mid \mathbf{X}_{t-1})$ for this example.

There could in principle be some further bias reduction by multiplying $\mathbf{C}_t$ by a random scalar with distribution $\chi^2_{n_a}/n_a$, where χ^2_n is a chi-squared distribution with n degrees of freedom. This choice gives the correct variability to each component of $\boldsymbol{\mathcal{A}}_{j,t}$ for $j \neq 0$. If one does this random rescaling, it is possible (I have not checked this) that one may need to modify the $W_j^{(m)}$s and $W_j^{(c)}$s in 2(b) to ensure the procedure works correctly when the process is linear. By combining the random rotation with the random scaling, each $\boldsymbol{\mathcal{A}}_{j,t}$ for $j \neq 0$ now follows a multivariate normal distribution. However, the sample is far different from a simple random sample from this multivariate normal. For example, if $\mathbf{C}_t$ is proportional to the identity matrix, for $j \neq 0$, the $\boldsymbol{\mathcal{A}}_{j,t}$s will all be on a sphere with center $\mathbf{m}_t$ and the vectors $\boldsymbol{\mathcal{A}}_{j,t} - \mathbf{m}_t$ will be orthogonal. This highly but not completely systematic sampling scheme should have similar precision as the purely deterministic SUT, but with better bias characteristics.

When using these random rotations and rescalings in practice, one does need to decide whether the same rotations and rescalings can ever be reused. For the UKF, I would suggest picking a new rotation and rescaling at each time step. For the UPF, the conservative choice would be to have independently chosen rotations and rescalings for each time step and each particle. However, if choosing a separate rotation for each particle added substantially to the computational burden, one could consider using the same rotation for every particle within a given time step.

Whether adding these random elements to the scheme will actually help in any particular circumstance depends on the nature of the nonlinearities. In many cases, there will be little or no gain for the extra effort of computing the random rotations and rescalings, but it may not be easy to know *a priori*

when, so that if the added computations do not slow down the algorithm much, it would be prudent to include these random elements.

Rejoinder by M. E. Irwin, N. Cressie, and G. Johannesson

We would like to thank each of the discussants for their insightful comments on the paper. All agree that modern particle filtering is a technology that can address important nonlinear spatial-temporal problems.

There were a number of common themes in their comments. The three themes involved the choice of movement model, selection of the proposal distribution and approximations, and the estimation of model parameters.

Movement models

The most common theme brought up by the discussants was the choice of movement model. The movement model, as described in equations (5.1) and (5.2), was chosen to display the properties of the procedures described in the paper, and not proposed as a truly realistic description of tank movement. This model was chosen to be similar in structure to many of the linear movement models used for target tracking as discussed by Ramachandra (2000). However, to take the example past the exploratory stage, careful consultation with military experts would be needed to determine appropriate movement models associated with various kinds of vehicles.

Chen and Liu's suggestion of multi-level movement models seems particularly appealing given the hierarchical modelling approach taken in our paper. While the simple movement model we use can easily be expanded by allowing different types of movement (e.g. go straight, turn, accelerate, decelerate), their suggestion has a potentially greater benefit. The multi-level approach would allow environmental effects such as weather, battle conditions (e.g. attack versus retreat), and tank-fitness information to be included in the state equations for modelling the tanks' movements. Some terrain effects could also be handled in this way.

Additional terrain effects could potentially be handled by the suggestion of Fuentes to base movement on a potential function $P(\mathbf{s}, t)$. While the

continuous-time stochastic-differential-equation approach she suggests may not be feasible due to a need for the filter to provide quick updates, discrete-time approximations should be. In fact, the model described by equations (5.1) and (5.2) is a special case of continuous-time movement approximated using a sequence of equally spaced time steps.

In general, the UPF approach we recommend will be feasible when the movement model can be described by a function of the form,

$$\mathbf{X}_t = \mathbf{u}_t(\mathbf{X}_{1:t-1}, \delta_t),$$

where δ_t is a random variable whose moments could depend on t and $\mathbf{X}_{1:t-1}$. Notice that usually $\mathbf{u}_t$ depends on $\mathbf{X}_{1:t-1}$ through $\mathbf{X}_{t-1}$, which yields state equations of Markovian form. The movement models for simulating the data (Section 4.1) and filtering the data (equations (5.1) and (5.2)) both fit into this structure. The excellent suggestions given by the discussants can be put into this framework. For example, one could replace the direction component given in (5.1) with Kent and Mardia's suggestion of modelling it by the von Mises distribution and its extensions.

Angulo and Ruiz-Medina comment that the model used in the example treats each of the tanks as independent. We agree with them that adding spatial-temporal interactions between the different elements is important in modelling a less-benign battlespace. In the example, the tanks are travelling in convoy, but separated so they will not interact. However, if the local movement of the tanks within a convoy required a tight formation, and the global movement of various convoys was determined by a sequence of spatial-temporal waypoints (Wendt et al. (2002)), the interaction of the various objects should not be ignored. Another example where such interactions may be extremely important is the case of minefield detection, where it may be desirable to maximize the area examined while minimizing the overlap of coverage of the various detectors as they pass through the region of interest. This is an example of Angulo and Ruiz-Medina's proposal for mobile monitoring devices. The incorporation of interactions would also be necessary to apply filtering in the applications suggested by Mateu and Montes.

One approach to deal with spatial-temporal interactions between objects would be to modify the process model, given by the equation above, to be a function of the states of all objects in the domain of interest. For

example, with two objects in the space, the equation could be modified to

$$(\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}) = \mathbf{u}_t(\mathbf{X}_{1:t-1}^{(1)}, \mathbf{X}_{1:t-1}^{(2)}, \delta_t^{(1)}, \delta_t^{(2)}),$$

where the superscripts correspond to the two objects of interest. Also, in this formulation, it will often be necessary for the random variables $\delta_t^{(1)}$ and $\delta_t^{(2)}$ to be dependent.

Proposal distribution and approximations

A second theme in the discussants' comments deals with choice of the sequential components of the proposal distribution $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$ in equation (2.4) of the SIS. Chen and Liu make the important point that the interaction between the computational burden of the sampler and its sampling variance should be the driving force in choosing the proposal distribution. This is particularly important in the danger-field example, as rapid, accurate updates of the danger field are required. For this example, Stein's interesting suggestion of selecting a random rotation of the matrix square root probably wouldn't be computationally feasible, a potential drawback that was noted by Stein in his discussion. Stein's suggestion is more important for the UKF than for the UPF. The biases inherent in approximation could accumulate in the Kalman filter, and the random rotations should average out the biases as Stein suggests. This bias accumulation should not be a problem in the UPF since the sampling procedure and adjustment by the importance sampling weights should deal with the bias accumulation. Furthermore, the tuning parameters of the Scaled Unscented Transformation or SUT (α, β, and κ) can be adjusted to account for potential skewness and kurtosis, and this should also ameliorate bias.

The bottom line in selecting a proposal distribution is to find something "close" to the optimal choice of $q^*(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t}) = p(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Z}_{1:t})$, under a reasonable computational burden. For many problems, such as the one examined in the paper, using the SUT achieves this goal, since much of the computing can be done with rapid matrix calculations. While a Gaussian proposal was suggested for computational convenience, it can break down. One potential problem is that the tails of the Gaussian proposal may be too light relative to the posterior distribution, a common worry in importance sampling. A potential solution to this problem would

be to choose from a class of distributions with heavier tails, such as a t-distribution with low or moderate degrees of freedom. This approach, of deviating from a Gaussian proposal distribution, can be generalized. The underlying idea of the UPF is to use the SUT to approximate, numerically, parameters of transformed random variables. In our paper we use the SUT to approximate the mean and variance, but other parameters such as higher-order moments or coefficients of variation could be approximated as well. For example, with a gamma proposal distribution, the scale and shape parameters would need to be approximated. This idea may be an approach to implement, in a computationally efficient manner, Angulo and Ruiz-Medina's suggestion of using maximum entropy to choose the proposal. (The notion of maximum entropy is important in some areas of physics and astronomy; see Jaynes (1957) and Grandy and Schick (1991). It can be justified in a decision-theoretic setting as a maximin criterion; see Bernardo (1979).)

Parameter estimation - Bayes versus empirical Bayes

The last major theme in the discussions deals with the estimation of model parameters. The approach taken in the paper was Empirical Bayes (EB), where the data were used to estimate the unknown parameters, and these estimates were plugged into the filtering equations. While this approach will lead to under-estimating the uncertainties of the filter, much of the variability is already captured in the hierarchical model, and in the example it will have little effect. Also, in response to Mateu and Montes' question, we could consider the fully Bayesian approach as giving a more complicated (and non-linear) state-space model, and hence we expect the fully Bayesian UPF to maintain or perhaps increase its advantage over the other two filters.

When possible, a fully Bayesian solution for dealing with the unknown model parameters would be preferable. Fuentes' suggestion to use discount factors (West and Harrison (1997)) is one approach. Another approach can be used in the situation where conjugate priors are available. Kong et al. (1994) base their sequential sampler on the predictive distribution of the state variables. A nice property of their approach is that it is easy to reweight their sampler under different choices for the prior, which allows for sensitivity analysis of the prior distribution. Modifications to their

approach should be possible when the prior is non-conjugate.

In response to Angulo and Ruiz-Medina's last comment, the anticipatory nature of the UPF for the danger field is a function of two factors, the form of the estimator and the amount of data. The UPF gives an estimate of the posterior mean of the danger field, which comes from minimizing the Bayes risk for squared-error loss. The anticipatory effect comes from averaging the danger field over the possible tank locations. Since the Kalman-filter estimates for danger are based on plugging in a single estimate of location, they cannot be anticipatory for danger since they ignore the uncertainty in the tank locations. Because of the averaging in the UPF, a lower sampling frequency must lead to earlier anticipation of danger, since there is more uncertainty in the true tank locations. Note that early anticipation of danger can have costs that may argue for increasing the sampling frequency; however, choosing an estimator that ignores the uncertainty of the tank positions may have more severe costs. Notice that a fully Bayesian solution with the UPF will tend to increase the anticipation time slightly, as the tank locations will also need to be averaged over the prior distributions for the model parameters. In conclusion, there are occasions when the UPF-based estimates are comparable with the UKF-based and EKF-based estimates, but we have shown that the UPF can have distinct advantages when the true movement process changes drastically (Figures 3 and 4), and when the questions we ask are highly non-linear (Figures 7 and 8).

References

Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, New Jersey.

Bar-Shalom, Y. and Fortmann, T. E. (1988). *Tracking and Data Association*. Academic Press, Boston.

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

Bernardo, J. J. (1979). Expected information as expected utility. *Annals of Statistics*, 7:686–690.

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

Bhattacharya, R. N. and Waymire, E. (1990). *Stochastic Processes with Applications*. Wiley, New York.

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

Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Springer-Verlag, Berlin.

Brillinger, D. R., Preisler, H. K., Ager, A. A., and Kie, J. G. (2000). *The Use of Potential Functions in Modelling Animal Movement*. Data Analysis from Statistical Foundations. Nova Science, New York.

Chen, R. and Liu, J. S. (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society, Series B*, 62:493–508.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*. Oxford University Press, Oxford.

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

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

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

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

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

Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika*, 89:683–697.

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

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, San Diego.

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

Gordon, N. (1994). *Bayesian Methods for Tracking*. Ph.D. thesis, Imperial College, University of London.

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

Grandy, W. T. and Schick, L. H., eds. (1991). *Maximum Entropy and Bayesian Methods*. Kluwer, Dordrecht.

Gregori, P., van Lieshout, M. N. M., and Mateu, J. (2002). Generalized area-interaction point processes. In preparation.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

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

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

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

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.

Jaynes, E. T. (1957). Information theory and statistical mechanics, II. *Physics Review*, 108:171–190.

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

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

Kent, J. T. (1978). Time reversible diffusions. *Advances in Applied Probability*, 10:819–835.

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

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

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer Verlag, New York.

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

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

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, New York.

Nelson, E. (1967). *Dynamical Theories of Brownian Motion*. Princeton University Press, Princeton.

Openshaw, S. (1994). *Two Exploratory Space-Time-Attribute Pattern Analysers Relevant to GIS*. Ed. Stewart Fotheringham and Peter Rogerson. Bristol:Taylor & Francis Inc., pp. 83–104.

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

Ramachandra, K. V. (2000). *Kalman Filtering Techniques for Radar Tracking*. Marcel Dekker, New York.

Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, 89:719–723.

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

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

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

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

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamics Models*. Springer-Verlag, New York, 2nd ed.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, 80:73–96.

Zhang, J. L. and Liu, J. S. (2002). A new sequential importance sampling method and its application to the 2D HP model. *Journal of Chemical Physics*, 93:1032–1043.