



# A Mixture Regularized Rao-Blackwellized Particle Filter for Terrain Positioning

Achille Murangira, Christian Musso, Karim Dahia

## ► To cite this version:

Achille Murangira, Christian Musso, Karim Dahia. A Mixture Regularized Rao-Blackwellized Particle Filter for Terrain Positioning. IEEE Transactions on Aerospace and Electronic Systems, 2016, 52 (4), 10.1109/TAES.2016.150089 . hal-01384329

**HAL Id: hal-01384329**

**<https://hal.science/hal-01384329>**

Submitted on 19 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Mixture Regularized Rao-Blackwellized Particle Filter for Terrain Positioning

Achille Murangira, Christian Musso, and Karim Dahia

**Abstract**—This study is concerned with the development of a robust particle filtering algorithm tailored to the problem of terrain aided positioning (TAP) via radar-altimeter measurements. The Rao-Blackwellized Particle Filter (RBPF) is a popular particle filtering algorithm for TAP that takes advantage of the nature of the state-space model by sampling particles in a subspace of the state space, yielding more efficient estimators than the standard particle filter.

Like most Monte Carlo filters, the standard RBPF uses the transition kernel as the proposal distribution during the particle update step. However, in contexts where the likelihood function is peaky, this may be highly inefficient since samples may fall in regions of low posterior probability. To address this issue, it is often advocated to use an importance sampling density which takes into account the latest observation. In a sequential importance sampling context, an optimal importance density is available but can only be easily sampled for specific state-space models, which raises the question of how to design a proposal density that is efficient, yet easy to sample from.

In this article, we propose a particle filtering importance sampling method adapted to multimodal distributions. It hinges on the use of a robust proposal density as well as a cluster-based representation of the multimodal posterior. This leads to a novel marginalized particle filter, the regularized Rao-Blackwellized Particle Filter, that is evaluated on a challenging terrain positioning application.

## I. INTRODUCTION

**I**NERTIAL NAVIGATION SYSTEMS (INS) are often the main source of navigation data in aircraft since they are autonomous and reliable. The aircraft position, velocity, attitude and heading components are computed by dead-reckoning. However, it is well known that these components drift due to alignment error and the accumulation of sensor errors over time. For certain classes of INS, this drift can be inconsistent with the navigation performance requirements. To ensure accurate navigation, one or more aiding source can be used in conjunction with the INS. A widespread aiding source is the GPS (Global Positioning System) and in this case, the GPS/INS sensor fusion can be performed with an Extended Kalman Filter (EKF) or an Unscented

Kalman Filter (UKF). One drawback of this approach is that the GPS can experience intentional jamming. This is especially true of military grade aircraft. These vehicles would use alternate aiding sources such as radar-altimeters to ensure autonomous navigation [1]. A Radar-altimeter provides ground clearance measurements along the flight path enabling navigation by comparing accumulated relative heights with a terrain elevation database. This form of navigation is commonly termed Terrain Aided Positioning (TAP).

In order to perform TAP/INS integration to estimate the drift of all kinematic components, it's necessary to resort to non-linear filtering algorithms. The goal is to approximate the probability density function (pdf) of a state vector (INS error) given the sequence of ground-clearance measurements accumulated up until the current time step. Due to the multimodality of this posterior (at least in the early stages of the INS update), methods such as the EKF or the UKF are inefficient since the Gaussian approximations are no longer relevant. Grid based methods such as the Point Mass Filter (PMF, [2]) perform better than the EKF but are mostly suited to estimating the position components only.

Particle filters (PF) [3] are an efficient solution to the non-linear filtering problem since they can, in theory, approximate any posterior density without any hypothesis regarding the linearity of the process or observation model or the Gaussian nature of process and observation noise. Moreover particle filters remain somewhat tractable when the state dimension increases. In TAP, the state vector may include attitude sensor bias components additionally to position and velocity yielding a dimension greater than 9.

The problem of TAP/INS integration using particle filters can be addressed by a specific algorithm, the Rao-Blackwellized Particle Filter (RBPF) otherwise known as the Marginalized Particle Filter [4], [5]. This filter takes advantage of the state-space model structure where, conditionally to part of the state vector (the horizontal position error component), the model is linear Gaussian. As reported in [5], [4], compared to a standard particle filter, less particles are needed to achieve similar performance.

In this paper, issues encountered when using particle filtering algorithms such as the Regularized Particle Filter (RPF) [6] or the Rao-Blackwellized Particle Filter in the context of terrain aided positioning are addressed. It is well known that standard particle filtering methods struggle when the standard deviation of the measurement noise is small w.r.t. the typical range of the observations. The influence of the intensity of the observation noise statistics on PF performance is not specific to TAP/INS integration and has been reported in various applications ranging from robot positioning [7] to target tracking [8]–[10]. This is caused by a mismatch between the proposal distribution used at the prediction step, and the likelihood function. A small observation noise can thus cause the likelihood function to be peaky and have little overlap with the proposal density. From a filtering standpoint, very few particles are significantly weighted, which can lead to filter divergence despite the resampling step. While the RBPF improves upon the standard particle filter by sampling particles in a lower dimensional state space, it can still suffer from the previous limitations if the transition density is used as the proposal in the prediction step.

Different approaches have been proposed in the particle filtering literature in order to design filters that are more robust in the context of low observation noise and/or peaky observation function. Most enhanced particle filtering algorithms proceed by replacing the usual transition kernel density with an approximation of the optimal importance density (OID). For instance, the particle filter with EKF proposal [11] proceeds by local linearisation of the state-space model yielding. The EKF proposal tends to perform well when the optimal importance density is unimodal and the non-linearities are moderate, conditions which are rarely met in terrain positioning and many other applications. The contribution of this paper is a novel particle filtering methodology dealing with highly multimodal posterior. To this end, we reconsider mixture particle filtering and its efficient implementation through particle clustering. This framework allows us to derive an importance sampling method aimed at generating particles around the modes of the posterior distribution in the event of a mismatch between the transition density and the likelihood function typically caused by low sensor noise. This proposal is then used during the particle update step of a Rao-Blackwellized Particle Filter. The paper's outline is as follows: Section II recalls the basics of particle filtering and Rao-Blackwellization. We also describe mixture modelling for PF via clustering for the sake of a practical implementation of a particle filtering algorithm with a mixture proposal; section III introduces the importance

sampling methodology and offers an analysis of the influence of the choice of the covariance matrix on the stability of the algorithm in case of a Gaussian proposal. The resulting algorithm, the mixture Rao-Blackwellized Particle Filter, is detailed at the end of this section. Finally, in section IV, the resulting filter is compared to standard PF algorithms in terms of mean squared error and robustness to divergence, in typically challenging settings.

## II. THE RAO-BLACKWELLIZED PARTICLE FILTER AND MIXTURE MODELS IN NON-LINEAR FILTERING

### A. Bayesian estimation and particle filtering

Consider the following general state-space model:

$$\begin{cases} x_k &= f_k(x_{k-1}) + w_k \\ y_k &= h_k(x_k) + v_k \end{cases} \quad (1)$$

where  $k$  is the time index and  $\{x_k, k \geq 0\}$  is the unobserved  $\mathbb{R}^d$  valued discrete time process,  $d \geq 1$  and  $\{y_k, k \geq 0\}$  is the sequence of observations.  $\{w_k, k \geq 0\}$  and  $\{v_k, k \geq 0\}$  are respectively the process and measurement noise sequences. The sequences  $\{w_k, k \geq 0\}$  and  $\{v_k, k \geq 0\}$  are assumed i.i.d., mutually independent and independent of  $x_0$ .

The aim of sequential filtering is the recursive computation of the posterior density  $p(x_k|y_{0:k})$  where  $y_{0:k} = \{y_0, \dots, y_k\}$ . This can be done in two steps:

- prediction:

$$p(x_k|y_{0:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{0:k-1})dx_{k-1} \quad (2)$$

- correction:

$$p(x_k|y_{0:k}) = \frac{p(y_k|x_k)p(x_k|y_{0:k-1})}{\int p(y_k|x_k)p(x_k|y_{0:k-1})dx_k} \quad (3)$$

It is generally assumed that  $w$  and  $v$  have known probability density functions, which yields closed form the transition  $p(x_k|x_{k-1})$  and the likelihood  $p(y_k|x_k)$ . In the remainder, we will sometimes use the short hand notation  $g_k(x_k) = p(y_k|x_k)$ .

The idea of particle filtering is to approximate the posterior distribution with a weighted sum of Dirac measures:

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^N \omega_k^i \delta_{x_k^i} \quad (4)$$

where  $\delta_x$  denotes the Dirac measure centered at  $x$  and the  $x_k^i, i = 1, \dots, N$  are the particles. The importance weights  $\{\omega_k^i\}_{i=1}^N$  sum up to one. Particle filtering algorithms stem from the Sequential Importance Resampling (SIR) algorithm that recursively updates the particle/weights systems [3]. Assuming a weighted

particle set  $\{x_{k-1}^i, \omega_{k-1}^i\}_{i=1}^N$  is available at time  $k-1$ , the SIR computes the particle approximation at time  $k$  through the following steps :

- 1) Particle update: sample  $x_k^i \sim \tilde{q}(x_k^i | x_{k-1}, y_k)$ ,  $i = 1, \dots, N$  where  $\tilde{q}$  is suitably chosen proposal density
- 2) Weight update: compute  $\tilde{\omega}_k^i = \frac{p(x_k^i | x_{k-1}) p(y_k | x_k^i)}{\tilde{q}(x_k^i | x_{k-1}, y_k)}$  and  $\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{j=1}^N \tilde{\omega}_k^j}$
- 3) Resampling: sample with replacement  $N$  particles  $\tilde{x}_k^i$  so that  $\forall j = 1, \dots, N$ ,  $\mathbb{P}(\tilde{x}_k^j = x_k^j) = \omega_k^j$ . Set  $x_k^i = \tilde{x}_k^i$  and  $\omega_k^i = \frac{1}{N}$

The resampling step avoids in principle the weight degeneracy where after a few iterations, all but one particle have negligible weight. It essentially discards particles with low weight and duplicates those with high weight. In practice, resampling is triggered by monitoring a criterion such as the effective sample size:

$$N_{eff} \approx \frac{1}{\sum_{i=1}^N (\omega_k^i)^2} \quad (5)$$

Resampling is triggered whenever  $N_{eff} \geq N_{th}$  where  $N_{th}$  is a threshold between 0 and  $N$ .

### B. The Rao-Blackwellized Particle Filter

The Rao-Blackwellized Particle Filter, or marginalized particle filter, is an efficient implementation for conditionally linear Gaussian models. It has been widely used in several engineering applications such as aircraft navigation [5], target tracking [12] and robot positioning [13]. For a general discussion regarding the algorithm we refer the reader to [14], [15], [16] and [4].

Assume the state-space model takes the following form :

$$\begin{cases} x_{k,1} &= f_{k-1,1}(x_{k-1,1}) + F_{k-1,1}(x_{k-1,1})x_{k-1,2} \\ &+ G_{k-1,1}(x_{k-1,1})w_{k-1,1} \\ x_{k,2} &= f_{k-1,2}(x_{k-1,1}) + F_{k-1,2}(x_{k-1,1})x_{k-1,2} \\ &+ G_{k-1,2}(x_{k-1,1})w_{k-1,2} \\ y_k &= h_k(x_{k,1}) + H_k(x_{k,1})x_{k,2} + v_k \end{cases} \quad (6)$$

where  $x_k^T = [x_{k,1}^T, x_{k,2}^T]$ .  $x_{k,1}$  is the non-linear part of  $x_k$  and  $x_{k,2}$  the linear part. Assume that  $w_k$  and  $v_k$  are white mutually independent Gaussian noises with respective covariance  $Q_k$  and  $R_k$  and such that

$$Q_k = \begin{pmatrix} Q_k^{11} & Q_k^{12} \\ (Q_k^{12})^T & Q_k^{22} \end{pmatrix}$$

We also suppose that  $x_{0,2} \sim \mathcal{N}(\bar{x}_{0,2}, P_0)$ . The initial density  $p(x_{0,1})$  for the non-linear substate is also known. The RBPF proceeds by interleaving a Kalman filter that updates the linear part and a particle filter that targets

the non-linear part. For a detailed derivation of the algorithm, we refer the reader to [4] and only describe the RBPF algorithm in appendix A.

### C. Mixture particle filters

Mixture filtering [17], [18] is based on the use of a mixture model for the posterior density, i.e.  $p(x_k | y_{0:k})$  is expressed as

$$p(x_k | y_{0:k}) = \sum_{j=1}^M \alpha_{j,k} p_j(x_k | y_{0:k}) \quad (7)$$

where  $M$  is the number of mixture components and the mixture weights  $\alpha_{j,k}$  are positive and satisfy  $\sum_{j=1}^M \alpha_{j,k} = 1$ .

The filter prediction step is performed according to

$$p(x_k | y_{0:k-1}) = \sum_{j=1}^M \alpha_{j,k-1} p_j(x_k | y_{0:k-1}) \quad (8)$$

where

$$p_j(x_k | y_{0:k-1}) = \int p(x_k | x_{k-1}) p_j(x_{k-1} | y_{0:k-1}) dx_{k-1}$$

and the filter correction step involves both the update of the mixture weights and the measurement update of the mixture densities:

$$p(x_k | y_{0:k}) = \sum_{j=1}^M \alpha_{j,k} p_j(x_k | y_{0:k})$$

where

$$p_j(x_k | y_{0:k}) = \frac{p(y_k | x_k) p_j(x_k | y_{0:k-1})}{\int p(y_k | x_k) p_j(x_k | y_{0:k-1}) dx_k} \quad (9)$$

and

$$\alpha_{j,k} = \frac{\alpha_{j,k-1} p_j(y_k | y_{0:k-1})}{\sum_{l=1}^M \alpha_{l,k-1} p_l(y_k | y_{0:k-1})} \quad (10)$$

Mixture particle filters are obtained by considering a set of  $N$  particles  $\{x_k^i\}_{i=1}^N$  and associated weights  $\{\omega_k^i\}_{i=1}^N$  so that each mixture component is targeted by a *subset* of weights and particles. The particle approximation of the posterior density is expressed as:

$$p(x_k | y_{0:k}) \approx \sum_{j=1}^M \alpha_{j,k} \sum_{i \in I_j} \omega_k^i \delta_{x_k^i} \quad (11)$$

where  $I_j \subseteq \{1, \dots, N\}$  is the subset of particles targeting the  $j$ -th mixture component  $p_j(x_k | y_{0:k})$  and  $\sum_{i \in I_j} \omega_k^i = 1$ . The  $\{\omega_k^i, i \in I_j\}$  are the intra-component importance weights. The particle prediction step at time  $k$  is identical to that of the SIR and consists in sampling

$N$  particles according to some proposal  $\tilde{q}(x_k|x_{k-1}, y_k)$ . The measurement update steps involves the update of the importance weights and the update of the mixture weights :

- compute  $\tilde{\omega}_k^i = \omega_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{\tilde{q}(x_k^i|x_{k-1}^i, y_k)}$
- for  $i = 1 \in I_j$ ,  $\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{l \in I_j} \tilde{\omega}_k^l}$
- for  $j = 1, \dots, M$ ,

$$\alpha_{j,k} \approx \frac{\alpha_{j,k-1} w_{j,k}^+}{\sum_{p=1}^M \alpha_{p,k-1} w_{p,k}^+} \quad (12)$$

where  $w_{j,k}^+ = \sum_{i \in I_j} \tilde{\omega}_k^i$

The main advantage of mixture particle filters is that they are able to maintain multimodality longer than standard particle filtering algorithms. Moreover, we have demonstrated in a previous paper [19] that they allow a more robust implementation of the regularization procedure. Indeed, the standard regularized particle filter [6] assumes a unimodal posterior which is not the case when dealing with ambiguous measurements. Another advantage which will be highlighted in the next section is the possibility to use a different proposal for each mixture component  $p_j(x_k|y_{0:k})$ . This makes it easier to adapt the sampling distribution to the particular shape of  $p_j(x_k|y_{0:k})$ .

Note that in practice, we wish to have one mixture component per mode of the posterior density. Since the number of modes may change with time as ambiguity is progressively resolved, a clustering procedure may be necessary to group together particles associated to the same modes. In [17], the authors use the k-means algorithm. In terrain navigation since the number of modes of the posterior is usually unknown, we advocate the use of the mean-shift clustering algorithm [19]. Denoting  $M_k$  the number of modes at time  $k$ , the posterior is then expressed as

$$p(x_k|y_{0:k}) = \sum_{j=1}^{M_k} \alpha_{j,k} p_j(x_k|y_{0:k}) \quad (13)$$

$$\approx \sum_{j=1}^{M_k} \alpha_{j,k} \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i} \quad (14)$$

where  $I_{j,k} \subseteq \{1, \dots, N\}$  is the set of particle index targeting the component  $p_j(x_k|y_{0:k})$  at time  $k$ . At time  $k+1$ , assuming that the number of components remains

identical,

$$p(x_{k+1}|y_{0:k+1}) \approx \hat{p}_{k+1|k+1} \quad (15)$$

$$= \sum_{j=1}^{M_k} \alpha_{j,k+1} \sum_{i \in I_{j,k}} \omega_{k+1}^i \delta_{x_{k+1}^i} \quad (16)$$

$$= \sum_{i=1}^N \alpha_{c_1(i),k+1} \omega_{k+1}^i \delta_{x_{k+1}^i} \quad (17)$$

where  $c_1(i) = j$  if  $i \in I_{j,k}$ . In reality, each empirical mixture component  $\sum_{i \in I_{j,k}} \omega_{k+1}^i \delta_{x_{k+1}^i}$ ,  $j = 1, \dots, M_k$  may contain multiple modes or be merged with another spatially close component (in particular  $M_{k+1} > M_k$  is possible). To account for this variation in the number of modes, the posterior is expressed as a mixture of  $M_{k+1}$  components as

$$\hat{p}_{k+1|k+1} = \sum_{l=1}^{M_{k+1}} \beta_{l,k+1} \sum_{i \in I_{l,k+1}} \nu_{k+1}^i \delta_{x_{k+1}^i} \quad (18)$$

where  $\sum_{l=1}^{M_{k+1}} \beta_{l,k+1} = 1$ ,  $\sum_{i \in I_{j,k+1}} \nu_{k+1}^i = 1$ . Equation (18) can be rewritten as

$$\hat{p}_{k+1|k+1} = \sum_{i=1}^N \beta_{c_2(i),k+1} \nu_{k+1}^i \delta_{x_{k+1}^i} \quad (19)$$

where  $c_2(i) = l$  if  $i \in I_{l,k+1}$ . It follows that (see [17] for instance),

$$\beta_{l,k+1} = \sum_{i \in I_{l,k+1}} \alpha_{c_1(i),k+1} \omega_{k+1}^i \quad (20)$$

and

$$\nu_{k+1}^i = \frac{\alpha_{c_1(i),k+1} \omega_{k+1}^i}{\beta_{c_2(i),k+1}} \quad (21)$$

### III. AN IMPORTANCE SAMPLING METHOD FOR MULTIMODAL DISTRIBUTIONS BASED ON POSTERIOR MODES

As discussed in the introduction, choosing an adequate proposal density is crucial for the robustness of the filter, especially in challenging situations of interest, namely when the observation noise is small or when the observation function exhibits peakiness, more generally when the Fisher information is high. Many alternate importance sampling schemes involve some approximation of the optimal importance density (OID) at time  $k$ ,  $p(x_k|x_{k-1}, y_k)$  such as the particle filter with EKF proposal [11], the unscented particle filter [20] or the progressive proposal particle filter [21]. In this work, we take another perspective by using mixture modelling and proposing a different importance distribution for each mixture component  $p_j(x_k|y_{0:k})$ . Each proposal is centered on the maximum a posteriori of  $p_j(x_k|y_{0:k})$ .

It is well known that the proposal density should be as close as possible in shape to the target density. Since we assume each component  $p_j(x_k|y_{0:k})$  to be unimodal, proximity can be achieved by centering the proposal at the mode of  $p_j$ . If the proposal is a Gaussian density, the covariance can be chosen so as to match the orientation and dispersion of  $p_j(x_k|y_{0:k})$ . Choices of the corresponding covariance will also be discussed in this section.

In order to implement this importance sampling strategy, it is necessary to determine, without excessive computation overhead, the modes of the posterior density. We have previously introduced [22] a method suitable for partially linear models which will be detailed here for the sake of clarity. We will also discuss some issues related to the stability of the importance sampling procedure, which will lead in particular, to robust choices of the covariance matrix for Gaussian proposals. Finally, a mixture Rao-Blackwellized Particle Filter using this importance sampling strategy will be presented.

#### A. Standard importance sampling in Bayesian inference

Let  $X$  be a hidden state distributed according to a density  $q$  and partially observed through a measurement  $Y = h(X) + V$  where  $h$  is a non-linear observation function and  $V$  is a measurement noise independent from  $X$ .

Importance sampling targets posterior expectations of the form

$$\mathbb{E}_p(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x)p(x|y)dx \quad (22)$$

where  $p(x|y) = \frac{g(x)q(x)}{\int g(x)q(x)dx}$  is the conditional density of  $X$  given  $Y$ ,  $\varphi$  is a real-valued measurable test function of  $\mathbb{R}^d$ ,  $g$  is the likelihood function and  $\mathbb{E}_p$  refers to the expectation w.r.t. density  $p$ . In general, it is not straightforward to sample from the posterior which is one of the reasons for using importance sampling.

Let  $\tilde{q}$  be a density whose support included that of  $\varphi q$ . The importance sampling (IS) density estimator of (22) is given by

$$\mathbb{E}_p(\varphi(X)) \approx \sum_{i=1}^N \omega^i \varphi(x^i) \quad (23)$$

where  $x^i \sim \tilde{q}$ ,  $i = 1, \dots, N$  are i.i.d. samples from  $\tilde{q}$ , a so-called importance density. The importance weights  $\omega^i$  are computed according to  $\omega^i = \frac{\tilde{\omega}^i}{\sum_{j=1}^N \tilde{\omega}^j}$  where  $\tilde{\omega}^i = \frac{g(x^i)q(x^i)}{\tilde{q}(x^i)}$ .

If additionally,

- $\mathbb{E}(\varphi^2(X)|Y) < \infty$

- $\mathbb{E}(\varphi^2(X)\frac{q(X)}{\tilde{q}(X)}|Y) < \infty$

then the estimator (23) is unbiased and strongly consistent [23]. Another indicator of importance sampling performance which is of interest in this paper is the asymptotic variance of the importance weights  $V_{\tilde{q}}$  defined by

$$V_{\tilde{q}} = \frac{1}{N} \left( \frac{\int_{\mathbb{R}^d} \frac{g^2(x)q^2(x)}{\tilde{q}(x)}dx}{\left( \int_{\mathbb{R}^d} g(x)q(x)dx \right)^2} - 1 \right) \quad (24)$$

(24) is a measure of IS performance [24] and will be used as a criterion to design the importance density  $\tilde{q}$ . Indeed, it is known (see [24] for example) that  $V_{\tilde{q}} = 0$  for  $\tilde{q} = p(x|y)$  which would indicate that a good choice of  $\tilde{q}$  is one that is similar to  $p(x|y)$ .

For a general mixture model  $p(x|y) = \sum_{j=1}^M \alpha_j p_j(x|y)$  where  $p_j(x|y) = \frac{g(x)q_j(x)}{\int g(x)q_j(x)dx}$  and  $q_j$  is the prior density corresponding the  $j$ -th component, we may use a different proposal for each component. In this case,

$$\begin{aligned} \mathbb{E}_p(\varphi(X)|Y) &= \int \varphi(x) \sum_{j=1}^M \alpha_j p_j(x|y) dx \\ &= \sum_{j=1}^M \alpha_j \int \varphi(x) \frac{g(x)q_j(x)}{\int g(z)q_j(z)dz} dx \\ &= \sum_{j=1}^M \alpha_j \frac{\mathbb{E}_{\tilde{q}_j} \left( \frac{g(X)q_j(X)\varphi(X)}{\tilde{q}_j(X)} \right)}{\mathbb{E}_{\tilde{q}_j} \left( \frac{g(X)q_j(X)}{\tilde{q}_j(X)} \right)} \\ &\approx \sum_{j=1}^M \alpha_j \frac{\frac{1}{N_j} \sum_{i \in I_j} \tilde{\omega}^i \varphi(x^i)}{\frac{1}{N_j} \sum_{l \in I_j} \tilde{\omega}^l} \\ &\approx \sum_{j=1}^M \alpha_j \sum_{i \in I_j} \omega^i \varphi(x^i) \end{aligned}$$

where  $x^i, i \in I_j$  are i.i.d. samples from  $\tilde{q}_j$ ,  $\omega^i = \frac{\tilde{\omega}^i}{\sum_{l \in I_j} \tilde{\omega}^l}$  and  $I_j \subseteq \{1, \dots, N\}$  is the subset of samples targeting the  $j$ -th mixture component as in section II-C and such that  $|I_j| = N_j$ .

#### B. Designing the importance density

Given that we have indicated how to implement importance sampling when targeting a mixture density, we shall assume that the posterior is unimodal in this subsection: if this is not the case then it suffices to find a decomposition of the posterior into a mixture of unimodal densities as outlined in III-A.

As argued previously, a general rule of thumb to design a suitable importance density is to choose one that is close

to the posterior. For instance, we can use a proposal  $\tilde{q}$  centered on an approximation of the conditional expectation  $\mathbb{E}(X|Y)$  and with covariance close to the posterior covariance  $\text{Cov}(X|Y)$ . However, for this choice of  $\tilde{q}$  we must also ensure  $V_{\tilde{q}} < \infty$ . In the sequel we study several proposal densities, designed to match the posterior while guaranteeing a finite estimation variance, as measured by the asymptotic variance of the importance weights  $V_{\tilde{q}}$ . These covariances will be candidates for the proposals used in the mixture Rao-Blackwellized particle filter introduced in III-D.

In order to approximate  $\text{Cov}(X|Y)$  we use Laplace's method: this approach has been used in the context of importance sampling in [25] and [9]. Laplace's method approaches the posterior covariance by the inverse of the observed Fisher information matrix  $J$  evaluated at the mode of  $p(x|y)$ , while a first order approximation of the posterior expectation is given by the mode of  $p(x|y)$ .

$$\mathbb{E}(X|Y) = \hat{x}^* - F_0(J(\hat{x}^*), J'(\hat{x}^*)) \quad (25)$$

$$\text{Cov}(X|Y) = J^{-1}(\hat{x}^*) - G_0(J(\hat{x}^*), J'(\hat{x}^*), J''(\hat{x}^*)) \quad (26)$$

where  $\hat{x}^* = \arg \max_{x \in \mathbb{R}^d} p(x|y)$  is the maximum a posteriori (MAP) and

$$J(x) = -\frac{\partial^2 \log g}{\partial x^2}(x) - \frac{\partial^2 \log q}{\partial x^2}(x) \quad (27)$$

and  $F_0$  and  $G_0$  are given functions (see [9] for instance). In the sequel, we only keep the first terms of expansions (25) and (26), ie.

$$\begin{aligned} \mathbb{E}(X|Y) &\approx \hat{x}^* \\ \text{Cov}(X|Y) &\approx J^{-1}(\hat{x}^*) \end{aligned}$$

Assuming a gaussian prior  $q$  with covariance  $P$  and zero mean gaussian noise with covariance  $R$ , as will be done subsequently, one can compute  $J$  as

$$J(x) = 2 \left[ \frac{\partial^2 h}{\partial x^2} R^{-1} (y - h(x)) - R^{-1} \left( \frac{\partial h}{\partial x} \right) \left( \frac{\partial h}{\partial x} \right)^T \right] + P^{-1} \quad (28)$$

The assumption of a gaussian prior in a filtering setting is somewhat reasonable since  $q$  is supposed unimodal here.

Therefore, assuming we have an estimate for  $\hat{x}^*$ , we can use a Gaussian proposal density  $\tilde{q}$  with mean  $\hat{x}^*$  and covariance  $J^{-1}(\hat{x}^*)$ . However, as pointed out in [22], [23], this can lead to non-consistent importance sampling estimates since the tails of the proposal may be lighter than the tails of the posterior.

It is straightforward that whenever  $\sup_{x \in \mathbb{R}^d} \frac{q(x)}{\tilde{q}(x)} < \infty$ ,  $V_{\tilde{q}}$  is automatically finite. Hence, assuming that the prior  $q$  is Gaussian, one can choose a t-distribution centered on the mode  $\hat{x}^*$ , with a suitable scale matrix  $\Sigma$  with

$\nu > 2$  degrees of freedom (d.o.f.). For  $\nu > 2$ , the covariance matrix of a variable distributed according to the  $t$ -distribution with scale matrix  $\Sigma$  is equal to  $\frac{\nu}{\nu-2} \Sigma$ . Therefore, to achieve a covariance equal to Laplace's approximation of  $\text{Cov}(X|Y) \approx J^{-1}(\hat{x}^*)$ , one must set the scale matrix to

$$\Sigma = \frac{\nu-2}{\nu} J^{-1}(\hat{x}^*) \quad (29)$$

The d.o.f. parameter  $\nu$  tunes the fatness of the  $t$ -distribution's tail : lower values correspond to a fat tail while higher ones lead to a lighter tail.

Let us now examine the choice of a Gaussian proposal density  $\tilde{q}(x) = \mathcal{N}(x; \hat{x}^*, \Sigma)$  whose covariance  $\Sigma$  is chosen under the constraint that  $V_{\tilde{q}} < \infty$ .

Since  $\sup_{x \in \mathbb{R}^d} \frac{q(x)}{\tilde{q}(x)} < \infty \Rightarrow V_{\tilde{q}} < \infty$ , one only needs to find a matrix  $\Sigma$  such that  $\sup \frac{q}{\tilde{q}} < \infty$ . The following proposition provides such a matrix.

**Proposition 1.** *Let  $P^{-1} = D^T D$  be the Choleski decomposition of  $P^{-1}$ . If the largest eigenvalue of  $D^{-T} \Sigma^{-1} D^{-1}$  is smaller than 1, then  $\frac{q}{\tilde{q}}$  is bounded from above and  $V_{\tilde{q}}$  is finite.*

*Proof.* See appendix B.  $\square$

This proposition gives us a practical way of designing Gaussian proposals with covariance  $\Sigma$  that yield consistent importance sampling estimates.

We have previously proposed in [22] a covariance  $\Sigma$ . While giving good results in simulations, there was no guarantee that the asymptotic variance would be finite. This choice is obtained by rotating the ellipsoid of the prior covariance  $P$  along the principal axes of the ellipsoid of the Laplace approximation of the posterior covariance. More precisely, this covariance is defined by

$$P_{rot} \triangleq E_J \Lambda_P E_J^T \quad (30)$$

where  $E_J$  is the matrix of eigenvectors of  $\hat{J}^{-1} = J^{-1}(\hat{x}^*)$  and  $\Lambda_P$  the matrix of eigenvalues of  $P$ . In [22], we have found that setting  $\Sigma = P_{rot}$  in the proposal  $\tilde{q}(x) = \mathcal{N}(x; \hat{x}^*, \Sigma)$ , could, in some cases, yield a lower estimation error compared to the use of the prior covariance  $\Sigma = P$ .

Still, to ensure that the variance of the asymptotic weights  $V_{\tilde{q}}$  remains finite, we may set  $\Sigma = s P_{rot}$  where  $s > 0$  and find the smallest  $s$  such that  $V_{\tilde{q}} < \infty$ . According to the previous proposition, we only need to determine

$$\inf \{s > 0 \mid \lambda_{\max}(\frac{1}{s} D^{-T} P_{rot}^{-1} D^{-1}) < 1\}$$

where  $\lambda_{\max}(B)$  is the largest eigenvalue of matrix  $B$  when it exists. Hence, for all  $s > s^*$ ,  $V_{\tilde{q}} < \infty$  where  $s^*$

is defined as

$$s^* = \lambda_{\max}(D^{-T}P_{rot}^{-1}D^{-1}) \quad (31)$$

Finally, we may also seek a value  $\Sigma$  that is close to the Laplace approximation of the posterior covariance, while also ensuring that  $V_{\hat{q}}$  is finite. Since the latter constraint is met whenever  $\Sigma - P > 0$  (cf. lemma 1), we can solve the following optimization problem :

$$\text{minimise } \|\Sigma - \hat{J}^{-1}\|_F^2 \text{ such that } \Sigma - P > 0$$

where  $\|\cdot\|_F$  is the Frobenius norm. The minimizer may be obtained via (see proof in appendix C)

$$\Sigma_F = \frac{1}{2}(\hat{J}^{-1} + P + H) + \kappa I_d \quad (32)$$

where  $\kappa$  is a small positive constant,  $U$  and  $H$  are such that  $\hat{J}^{-1} - P = UH$ ,  $U^T U = I$  and  $I_d$  is the  $d$  dimensional identity matrix.

Up until now, we have two different choices for the covariance of the Gaussian proposal density. Each one of them tries to be close to the posterior covariance while guaranteeing stable importance sampling estimates, in the sense that the asymptotic variance of the importance weights remains finite :

- $\Sigma = s^*P_{rot}$  as defined by (30) and (31). This essentially consists in rotating the prior covariance  $P$  along the principal axes of the Laplace approximation of the posterior covariance, denoted  $\hat{J}^{-1}$ .
- $\Sigma = \Sigma_F$  as defined by (62).

Of course, as outlined in section III-A, it remains possible to use a Student  $t$ -distribution density centered on the mode of the posterior and with scale matrix proportional to  $\hat{J}^{-1}$  (or  $P_{rot}$ ). As a summary, we have plotted in figure 1 the 99% confidence ellipsoids of each covariance/scale matrix associated to the different Gaussian or t-density proposals centered on the MAP:

- 1) Laplace approximation of the posterior covariance:  $\hat{J}^{-1}$
- 2) prior covariance with rotation along posterior covariance main direction:  $P_{rot}$
- 3) scaled rotated prior covariance:  $s^*P_{rot}$  (achieves finite variance)
- 4) closest covariance matrix to  $\hat{J}^{-1}$  for the Frobenius norm:  $\Sigma_F$  (achieves finite variance)

These ellipsoids have equation

$$(x - \hat{x}^*)^T \Sigma^{-1} (x - \hat{x}^*) = \chi_{2,.99}^2$$

where  $\chi_{2,.99}^2$  is the 0.99 quantile of the Chi-squared distribution with 2 degrees of freedom. This illustration assumes a Gaussian prior with covariance  $P$ . The first choice is not new and has been widely used in the

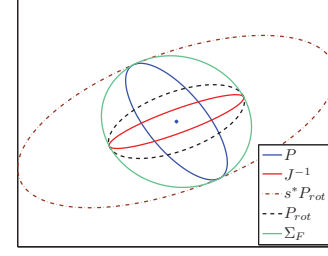


Fig. 1. 99% confidence ellipsoids associated to covariance matrix  $P$ ,  $J^{-1}$ ,  $P_{rot}$ ,  $s^*P_{rot}$  and  $\Sigma_F$

particle filtering literature, while the second one has been introduced in a previous paper [22] and the next two have been introduced in this paper. The main take away from this figure is that all ellipsoids are aligned with the ellipsoid corresponding to the Gaussian approximation of the posterior which is given by  $\hat{J}^{-1}$ . However, as argued previously, the size of the covariance should be sufficient to guarantee consistent estimation as attested by the greater areas covered by the proposal ellipsoids.

### C. Efficient MAP approximation for partially linear models

In accordance to the previous, section we assume a static setting where the posterior density is unimodal. We wish to compute the mode of the posterior density defined as

$$\hat{x}^* = \arg \max_{x \in \mathbb{R}^d} p(x|y) = \arg \max_{x \in \mathbb{R}^d} g(x)q(x) \quad (33)$$

When dealing with a multimodal density  $p(x|y)$ , we will assume that a finite mixture representation  $p(x|y) = \sum_{j=1}^M \alpha_j p_j(x|y)$  is available where each  $p_j(x|y)$  is assumed unimodal with unknown modes. Hence the problem of localizing all the modes of the full posterior consists in finding the mode of each individual mixture component.

Most filtering applications (target tracking, data assimilation, etc.) deal with high-dimensional state-vectors therefore deriving an exact or an approximate expression (33) is a non-trivial task. We have previously proposed [22] and we detail further a methodology suited for partially linear and Gaussian observation models, i.e. models where the observation equation takes the following form :

$$y = Ax_2 - h(x_1) + v \quad (34)$$

where  $x_1$  and  $x_2$  are sub-vectors of  $x$ , i.e.  $x = [x_1^T, x_2^T]^T$ ,  $h$  a non-linear function of  $x_1$  and  $A$  a matrix independent of  $x$ . Radar-altimeter observations can be modelled by

such an equation, and more generally, any observation model where only a substate  $x_1$  of the state-vector affects the measurement function.

We suggested in [22] a method to reduce the optimization dimension by assuming that the prior density  $q$  is Gaussian with mean  $\mu$  and covariance matrix  $P$ . The impact of this hypothesis will be discussed in the following section. We also assume the likelihood  $g$  to be a zero mean Gaussian density with covariance  $R$ .

Let  $\mu = (\mu_1^T, \mu_2^T)^T$  and  $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$ . We showed in [22] that the maximization can be performed in two steps :

- find  $\hat{x}_1^* = \arg \min_{x_1} (x_1 - \mu_1)^T P_{11}^{-1} (x_1 - \mu_1) + \Gamma(x_1)$
- set  $\hat{x}_2^* = \gamma(\hat{x}_1^*)$  and  $(\hat{x}^*)^T = [x_1^*, x_2^*]$

where  $\Gamma$  and  $\gamma$  are non-linear functions of  $x_1$  defined by

$$\begin{aligned} \Gamma(x_1) &= [y + h(x_1) - A\gamma(x_1)] R^{-1} [y + h(x_1) - A\gamma(x_1)]^T \\ &\quad + [\gamma(x_1) - \mathbb{E}_1(x_2)] P_{21}^{-1} [\gamma(x_1) - \mathbb{E}_1(x_2)]^T \\ \gamma(x_1) &= \Omega \left[ A^T R^{-1} (y + h(x_1)) + P_{21}^{-1} \mathbb{E}_1(x_2) \right] \\ \Omega &= \left( A^T R^{-1} A + P_{21}^{-1} \right)^{-1} \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}_1(x_2) &= \mathbb{E}(x_2|x_1) = \mathbb{E}(x_2) + P_{21}P_{11}^{-1}(x_1 - \mathbb{E}(x_1)) \\ &= \mu_2 + P_{21}P_{11}^{-1}(x_1 - \mu_1) \end{aligned}$$

Likewise,  $P_{2|1} = P_{22} - P_{21}P_{11}^{-1}P_{12}$  is the covariance of  $x_2$  given  $x_1$ .

#### D. Adaptive importance sampling in a multimodal filtering setting

In the previous subsections, we have reviewed and suggested possible methods for designing proposal densities based on local modes of the posterior in a static setting. We also recalled a method for approximate MAP computation developed in a previous paper. These methods can be extended to a filtering context where we deal with multimodal posterior.

In accordance to the previous notations, the observation model is assumed partially linear and is expressed as

$$y_k = A_k x_{k,2} - h_k^n(x_{k,1}) + v_k \quad (35)$$

where  $v_k \sim \mathcal{N}(0, R)$  and  $A_k$  is matrix independent of  $x_k$ . The state vector is decomposed into its linear part  $x_{k,2}$  and its non-linear part  $x_{k,1}$ ,  $x_k = \begin{bmatrix} x_{k,1}^T & x_{k,2}^T \end{bmatrix}^T$  while  $h_k^n$  is a non-linear function of  $x_{k,1}^T$ . To deal with multimodal posterior densities, we resort

to a mixture representation of the posterior as in section II-C, i.e.:

$$p(x_k|y_{0:k}) = \sum_{j=1}^M \alpha_{j,k} p_j(x_k|y_{0:k})$$

where  $\sum_{j=1}^M \alpha_{j,k} = 1$  and

$$p(x_k|y_{0:k}) = \frac{g_k(x_k) p(x_k|y_{0:k-1})}{\int g_k(x_k) p(x_k|y_{0:k-1}) dx_k} \quad (36)$$

where we recall that  $g_k$  is the likelihood function at time  $k$ . Now, assume a particle approximation of the predictive density  $p(x_k|y_{0:k-1})$  is available.

$$p(x_k|y_{0:k-1}) \approx \sum_{j=1}^M \alpha_{j,k-1} \sum_{i \in I_j} \omega_{k-1}^i \delta_{x_{k|k-1}^i} \quad (37)$$

where  $x_{k|k-1}^i \sim p(x_k|x_{k-1}^i)$  are i.i.d. samples. Recall

$$p(x_k|y_{0:k}) = \frac{\sum_{j=1}^M \alpha_{j,k-1} g_k(x_k) p_j(x_k|y_{0:k-1})}{\sum_{l=1}^M \alpha_{l,k-1} \int g_k(x_k) p_l(x_k|y_{0:k-1}) dx_k}$$

Obtaining an approximation of  $p(x_k|y_{0:k})$  through importance sampling requires sampling particles  $x_k^i$  according to some suitable proposal as well as computing the importance weights  $\{\omega_k^i\}_{i=1}^N$  and the mixture weights  $\{\alpha_{j,k}\}_{j=1}^M$ . As suggested previously, a different importance density  $\tilde{q}_j$  may be used for each mixture component  $p_j(x_k|y_{0:k})$ . Note that, in the context of filtering, the prior  $q$  is the predictive density  $p_j(x_k|y_{0:k-1})$ .

Let us introduce  $M_k$  importance densities  $\tilde{q}_j$ , each centered on the mode of  $p_j(x_k|y_{0:k})$ . When using Gaussian densities,  $\tilde{q}_j(x_k) = \varphi(x_k|\hat{x}_{j,k}^*, \Sigma_{j,k})$  where  $\varphi(\cdot|m, \Theta)$  denotes the multivariate normal density with mean  $m$  and covariance matrix  $\Theta$ . The covariance  $\Sigma_{j,k}$  of the proposal can be chosen as described in subsection III-B. This scheme requires the computation of  $M_k$  modes  $\{\hat{x}_{j,k}^*\}_{j=1}^{M_k}$  defined by

$$\hat{x}_{j,k}^* = \arg \max_{x_k \in \mathbb{R}^d} g_k(x_k) p_j(x_k|y_{0:k-1}) \quad (38)$$

We obtain these modes by using the method detailed in subsection III-C. It is then necessary to work out a Gaussian approximation for the predictive density  $p_j(x_k|y_{0:k-1})$ , which is assumed unimodal, by using for instance the sample mean and sample covariance defined respectively as

$$\bar{x}_{j,k|k-1} = \sum_{i \in I_{j,k-1}} \omega_{k-1}^i x_k^i \quad (39)$$

and

$$\hat{P}_{j,k|k-1} = \sum_{i \in I_{j,k-1}} \omega_{k-1}^i (x_k^i - \bar{x}_{j,k|k-1})(x_k^i - \bar{x}_{j,k|k-1})^T \quad (40)$$

This Gaussian approximation is also used to work out the covariance of the proposal  $\Sigma_{j,k}$ .

Once parameters for the current proposal  $\tilde{q}_j$  are derived, we first obtain particles  $\{x_k^i\}_{i=1}^N$  by sampling for each mixture component of index  $j$ ,  $N_j$  samples  $x_k^i \sim \tilde{q}_j$ ,  $i \in I_{j,k}$ . Hence, the particle weights read :

- $\tilde{\omega}_k^i = \omega_{k-1}^i \frac{g_k(x_k^i) p_j(x_k^i | y_{0:k-1})}{\tilde{q}(x_k^i)}$
- $\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{l \in I_j} \tilde{\omega}_k^l}$

Notice that in order to compute the importance weights it is necessary to evaluate  $p_j(x_k^i | y_{0:k-1})$ , the predictive density for each sampled particle. A possibility is to use the approximation  $p_j(x_k^i | y_{0:k-1}) \approx \sum_{l \in I_j} \omega_{k-1}^l p(x_k^i | x_{k-1}^l)$ . However, this requires  $\mathcal{O}(N^2)$  operations, which can be costly for online applications. As an alternative, we use a Gaussian approximation for the predictive density  $p_j(x_k | y_{0:k-1})$  which is readily available since it is also used for the approximate mode computation. In this case the particle weight update becomes

$$\tilde{\omega}_k^i = \omega_{k-1}^i \frac{g_k(x_k^i) \varphi(x_k^i | \hat{x}_{j,k}^*, \Sigma_{j,k})}{\tilde{q}(x_k^i)} \quad (41)$$

$$\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{l \in I_j} \tilde{\omega}_k^l} \quad \forall i \in I_{j,k} \quad (42)$$

This approach has the disadvantage that the particle estimate is no longer consistent. However, simulations on a TAP/INS integration and on a range/bearings based target tracking scenario have suggested good filter behaviour.

It remains to compute the mixture weights  $\{\alpha_{j,k}\}_{j=1}^M$ . According to (10),

$$\alpha_{j,k} = \frac{\alpha_{j,k-1} \int g_k(x_k) p_j(x_k | y_{0:k-1}) dx_k}{\sum_{l=1}^M \alpha_{l,k-1} \int g_k(x_k) p_l(x_k | y_{0:k-1}) dx_k} \quad (43)$$

Importance sampling based on the proposal  $\tilde{q}_j$  can be used to work out the integrals in the numerator and the denominator which yields the following mixture weight estimate:

$$\alpha_{j,k} \approx \frac{\frac{\alpha_{j,k-1}}{N_j} \sum_{i \in I_{j,k-1}} \tilde{\omega}_k^i}{\sum_{l=1}^M \frac{\alpha_{l,k-1}}{N_l} \sum_{i \in I_{l,k-1}} \tilde{\omega}_k^i} \quad (44)$$

The general IS approach taken here is similar to the Gaussian approximation of the optimal importance density [11], which is, in a sequential importance sampling framework, equal to  $p(x_k | x_{k-1}, y_k)$ . However, the main difference is that the mode of the OID  $p(x_k | x_{k-1}, y_k)$  has to be computed for *each* particle which makes this method more computationally intensive, unless a straightforward expression for these modes is available. It is also necessary to evaluate the inverse of Hessian of the log-likelihood  $N$  times ; this can be time consuming whenever the evaluation of the observation function per particle is more costly than other standard particle filter operations (random variable generation, resampling).

To limit the algorithm's computing cost, we can choose to use the importance sampling only when detecting excessive variance in the particle weights. We monitor the effective sample size (ESS) for each mixture component represented by a cluster of weighted particles and trigger importance sampling only when the ESS falls significantly below the resampling threshold. The rationale is that when the mismatch between the predictive density  $p_j(x_k | y_{0:k-1})$  and the likelihood  $g_k(x_k)$  isn't too high, resampling/regularization can suffice to avoid filter divergence. Therefore, we only trigger the MAP computation and importance sampling only in the event of very low ESS. More specifically, let  $N_{eff}^j$  be the ESS of the particle approximation of  $\hat{p}_j = \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$  of  $p_j(x_k | y_{0:k})$  where the particles  $x_k^i$  and associated weights are computed by using the transition density as the proposal.  $N_{eff}^j$  is computed similarly to (5):

$$N_{eff}^j \approx \frac{1}{\sum_{i=1}^{N_j} (\omega_k^i)^2} \quad (45)$$

The threshold  $N_{th,MAP}$  for triggering importance sampling is defined as a fraction  $\zeta \in [0, 1]$  of the resampling threshold  $N_{th}$ , i.e.

$$N_{th,MAP} = \zeta N_{th} \quad (46)$$

Low values of  $\zeta$  imply a moderate use of importance sampling (that is, only in cases of severe weight degeneracy) while higher ones lead to a more frequent use. Let us observe that the formula for the mixture weight update when using MAP based importance sampling for every mixture component measurement update (see equation (44)) is different from the case where IS is not systematically triggered. In the latter case, the following formula (47) should be used:

$$\alpha_{j,k} = \frac{\alpha_{j,k-1} \beta_{j,k}}{\sum_{l=1}^M \alpha_{l,k-1} \beta_{l,k}} \quad (47)$$

where

$$\beta_{j,k} = \begin{cases} \sum_{i \in I_{j,k-1}} \omega_{k-1}^i g_k(x_k^i) & \text{if } \frac{N_{eff}^j}{N_j} \leq N_{th,MAP} \\ \frac{1}{N_j} \sum_{i \in I_{j,k-1}} \frac{g_k(x_k^i) p_j(x_k^i | y_{0:k-1})}{\tilde{q}_j(x_k^i)} & \text{otherwise} \end{cases} \quad (48)$$

Let us now detail the mixture Rao-Blackwellized particle filter algorithm using MAP based importance sampling. In the standard Rao-Blackwellized particle filter (see Algorithm 1), the Kalman filter measurement update is done after the particle measurement update and the resampling step. However, in our importance sampling strategy, the mixture component posterior density of the *entire* state  $x_k = [x_{k,1}^T, x_{k,2}^T]^T$ ,  $p_j(x_k | y_{0:k})$  is obtained. This entails that the Kalman measurement update and the particle measurement update cannot be split into different stages separated by the resampling step if we mix the adaptive importance sampling scheme in a standard RBPF framework. The solution is then to have a variable  $\text{indexMAP}_j$  that equals 1 if importance sampling is triggered for the  $j$ -th mixture component and equals 0 otherwise. In the event where  $\text{indexMAP}_j = 1$ , we only need to compute the covariance matrix  $P_k$  during the Kalman measurement update step while the computation of the updated means  $\hat{x}_{k,2}^i$ ,  $i = 1, \dots, N$  is not performed since it is already carried out at the importance sampling stage. The resulting algorithm, called MRBPF-MAP, is outlined in appendix D.

#### IV. PERFORMANCE EVALUATION

In this section, we study the performance of the MRBPF-MAP algorithm (Mixture Rao-Blackwellized Particle Filter with MAP aide proposal), in a context of low process and observation noise. As mentioned in the introduction, this can impact negatively standard particle filters as the prediction step using the transition kernel may sample particles in low probability regions of the state-space. We consider the problem of TAP/INS integration via particle filtering methods and compare our algorithm with standard filters used in terrain navigation applications. Performance indicators such as RMSE (Root Mean Squared Error) and percentage of non-divergent tracks are presented.

##### A. Motion and observation model

We consider an aircraft equipped with an Inertial Navigation System (INS). The accumulation of sensor errors in the rate gyros and in the accelerometers is responsible for the drift in the position, velocity and attitude component. letting  $k$  be the time index, we

denote  $[\tilde{\lambda} \ \tilde{\phi} \ \tilde{z} \ \tilde{v}_n \ \tilde{v}_e \ \tilde{v}_d]^T$  the vector comprising the inertial latitude, longitude and altitude as well as the north (n), east (e) and down (d) components of the velocity components in the navigation frame.

We wish to estimate the drift vector  $x_k$  defined as the metric error between the true position and velocity coordinates and their inertial equivalent and expressed in the navigation frame. It is shown in [26] that the drift can be accurately modelled by a linear model according to

$$x_{k+1} = Fx_k + G_k w_k$$

where

$$F = \begin{pmatrix} I_3 & \Delta I_3 \\ 0_3 & I_3 \end{pmatrix} \quad G_k = \begin{pmatrix} \frac{\Delta^2}{2} I_3 \\ \Delta I_3 \end{pmatrix} \quad (49)$$

$\Delta$  is the sampling period and  $w_k \sim \mathcal{N}(0, Q)$  is white Gaussian noise.

The aircraft is equipped with a DTED (Digital Terrain Elevation Data) with 3 arc seconds accuracy ( $\approx 100$  m), which enables a matching with the reconstructed elevation profile from altimeter measurements. Let  $y_k$  be the observation delivered by the radar-altimeter and let  $h_{DTED}(\lambda, \phi)$  be the digital terrain elevation function which maps a latitude  $\lambda$  and longitude  $\phi$  to the corresponding terrain elevation  $h_{DTED}(\lambda, \phi)$ . Then the ground clearance  $y_k$  is modelled according to

$$y_k = \tilde{z}_k - x_{k,3} - h_{DTED} \left( \tilde{\lambda} + \frac{x_{k,1}}{R_{\tilde{\lambda}} + \tilde{z}_k}, \tilde{\phi}_k + \frac{x_{k,2}}{(R_{\tilde{\phi}} + \tilde{z}_k) \cos \tilde{\lambda}_k} \right) + v_k \quad (50)$$

where  $v_k$  is an additive white Gaussian noise with standard deviation  $\sigma_v$ . This simple model does not take into account possible biases or altitude-dependent errors and is used to generate altimeter measurements at a frequency of 10 Hz.  $R_{\tilde{\lambda}}$  and  $R_{\tilde{\phi}}$  are respectively the north and east radius and are defined as

$$R_{\tilde{\lambda}} = a \frac{1 - e^2}{(1 - e^2 \sin^2(\tilde{\lambda}))^{3/2}} \quad (51)$$

$$R_{\tilde{\phi}} = \frac{a}{(1 - e^2 \sin^2(\tilde{\lambda}))^{1/2}} \quad (52)$$

$a = 6378137$  m,  $b = 6356752.3$  m are the earth's ellipsoid semi major axis and semi minor axis as defined by the World Geodetic System (WGS 84), while  $e = \sqrt{\frac{b(a-b)}{a}}$  is the eccentricity. The true aircraft trajectory is a straight line with near constant velocity as shown in figure 2 and an altitude of 2923 m.

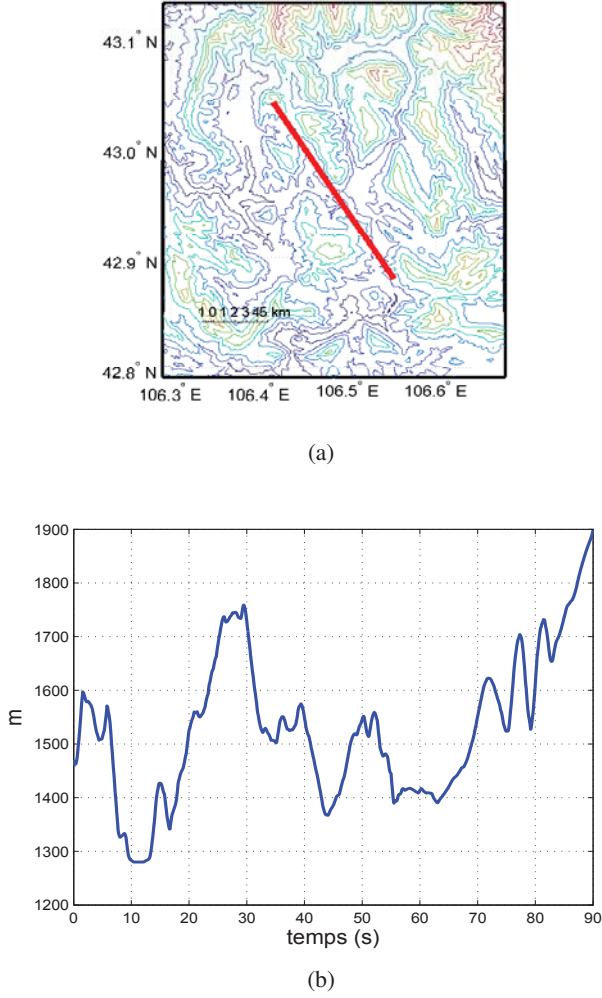


Fig. 2. Aircraft trajectory (a) and terrain elevation profile (b)

### B. Simulation settings and numerical results

We compare hereafter, the performance of the RBPF, the MRBPF and MRBPF-MAP in a challenging setting. The initial drift  $x_0$  is assumed to follow a zero-mean Gaussian distribution with covariance

$$P_0^{INS} = \text{diag}(1000^2, 1000^2, 100^2, 3^2, 3^2, 1^2)$$

all units being in  $m$  or  $m.s^{-1}$  where appropriate. This initial uncertainty is simply obtained by integration of the INS sensor errors from an initially known position, and during a given time interval, after which INS/TAP integration starts. The INS sensor errors are modelled by a Gauss-Markov type bias term plus a white Gaussian noise as in [5]:

- accelerometer bias standard deviation/correlation period:  $\sigma_a = 3 \times 10^{-5} m.s^{-2}$ ,  $\tau_a = 60 s$
- accelerometer wide band Gaussian noise noise st.dev.  $\sigma_{wa} = 5 \times 10^{-4} m.s^{-2}$
- gyrometer bias st.dev/correlation period :  $\sigma_g = 10^{-6} rad.s^{-1}$ ,  $\tau_g = 60 s$

- gyrometer wide band Gaussian noise noise st.dev.  $\sigma_{wg} = 1.7 \times 10^{-4} \frac{\pi}{180} rad.s^{-1}$

The radar-altimeter noise standard deviation is set to two different values  $\sigma_v = 5m$  and  $\sigma_v = 15m$ . Given the initial uncertainty and the nature of the elevation data,  $\sigma_v = 5m$  corresponds to very informative measurements and proves in practical simulations to be a challenging noise setting. The baseline value  $\sigma_v = 15m$  corresponds to medium level noise similar to the work [5]. Additionally, for the state dynamics, we considered  $Q = \text{diag}(1^2, 1^2, 0.01^2)$  which gives rather low process noise, which is another source of difficulty for particle filters as particles tend to be stranded in the same regions limiting their ability to explore the state-space. The filter parameters were set as follows:

- $N_{RBPF} = 4000$  particles for the RBPF
- $N_{MRBPF} = 4000$  particles for the MRBPF
- $N_{MRBPF-MAP} = 3000$  particles for the MRBPF-MAP
- resampling threshold :  $N_{th} = \frac{N}{3}$  where  $N$  is the number of particles
- threshold for cluster elimination  $\alpha_{min} = 10^{-20}$

The different number of particles between the MRBPF-MAP and the other algorithms ensures a similar computational load between all algorithms. To obtain a mixture representation for the particle cloud, a mean-shift algorithm was used as in [22]. Moreover, the modes were located as described in III-C by using a non-linear conjugate gradient optimization algorithm. To reduce the algorithmic complexity associated to this mode search, MAP based estimation is only carried out when 20 or less modes are detected by the clustering routine. For this choice of parameters, the computing cost of the MRBPF-MAP is approximately twice that of the RBPF.

*a) Influence of the modified proposal:* First, we focus on the impact of a MAP based proposal compared with the sole use of a mixture representation (MRBPF algorithm) and with the standard RBPF algorithm. In this analysis, the chosen proposal is a Gaussian density, with covariance matrix equal to  $P_{rot}$  (see (30)). For each algorithm, we have computed the RMSE for the non-divergent tracks by carrying out 200 filter runs as well as the percentage of non-divergent tracks. A filter run is said to be non-divergent if the final estimate  $\hat{x}_n$  is inside the 99% confidence ellipsoid associated to the Gaussian distribution centered on the true state with covariance the empirical covariance  $\hat{P}_n$  given by the particle cloud. The posterior Cramer-Rao lower bound (PCRB) [27] is also worked out for this model to assess filtering efficiency.

TABLE I  
INFLUENCE OF THE MODIFIED PROPOSAL: PERCENTAGE OF NON DIVERGENT TRACKS

	$\sigma_v = 5$	$\sigma_v = 15$
RBPF	69	80
MRBPF	71	81
MRBPF-MAP	79	90

Note that when the posterior is multimodal, this bound is not tight hence the large gap between the filter RMSE and the PCRB usually observed in the first few iterations. As can be seen in figures 3 and 4 which correspond to a very informative measurements setting, the RMSE given by the MRBPF-MAP, which uses a proposal centered around the posterior modes, is significantly lower for a good portion of the trajectory, especially for the position components the state vector. In the baseline setting (increased measurement noise), it appears that the improvement in mean squared error yielded by the MRBPF-MAP is less significant (see fig. 5 and 6). The percentage of non-divergent tracks for both sensor noise levels is displayed in table I. The results indicate both a gain in robustness (fewer divergent tracks) and accuracy for the MRBPF-MAP compared to the standard Rao-Blackwellized Particle filter (RBPF) and its adaptation (MRBPF) in a multimodal setting.

*b) Influence of the covariance matrix in the MAP based proposal:* Second, let us examine the choice of the covariance of the proposal density  $\tilde{q}$  that is used locally in each mixture distribution. To this end, we have used either a Gaussian density or a Student t-density proposal with scale matrix proportional to the target posterior covariance  $\Sigma$ , as explained in section III. Here, three different implementations of the MRBPF-MAP using different target posterior covariance  $\Sigma$  are compared :

- MRBPF-MAP 1: this algorithm uses for each mixture component, a Gaussian proposal centered on the MAP and with covariance  $\Sigma = P_{rot}$
- MRBPF-MAP 2: this algorithm uses a Gaussian proposal with covariance  $\Sigma = \Sigma_F$
- MRBPF-MAP 3: this version uses a t-density proposal with  $\nu = 8$  degrees of freedom. For each component, the proposal covariance is  $\Sigma = \hat{J}^{-1}$  computed via the Laplace approximation given by (26) and (28), where  $R = \sigma_v^2$ . As discussed in previous sections, the use of a Gaussian proposal with covariance  $\Sigma = \hat{J}^{-1}$  can potentially lead to infinite variance for the importance sampling estimator, whereas the t-Student density theoretically avoids this pitfall.

TABLE II  
INFLUENCE OF PROPOSAL'S COVARIANCE MATRIX: PERCENTAGE OF NON DIVERGENT TRACKS

	$\sigma_v = 5$	$\sigma_v = 15$
MRBPF-MAP 1	79	90
MRBPF-MAP 2	81	93
MRBPF-MAP 3	82	92
RBPF	69	80

As in the previous case study the number of particles is set to 3000 for the three versions of the algorithm. There is no change in other particle filtering parameters. The percentages of non-divergent tracks for each version of the MRBPF-MAP is shown in table II. The use of the student density in conjunction with covariances  $\Sigma_F$  or  $\Sigma = \hat{J}^{-1}$  improves the percentage of non-divergent tracks in both sensor noise situations. However, as can be seen in figures 7 and 8 which correspond to  $\sigma_v = 5m$ , there is no significant difference in terms of RMSE. For  $\sigma_v = 15m$  (not shown here), the same observation holds. We therefore conclude that the covariance matrix used is of moderate importance in this particular case.

## V. CONCLUSION

This paper introduces and discusses a practical importance sampling procedure in case of state-space models with informative and ambiguous measurements. The motivation is essentially terrain navigation but the algorithm we have presented can be used in other applications with similar characteristics (multimodality and/or low observation noise) such as target tracking or gravity gradiometer navigation. We have proposed a mixture proposal density whose components are centered on the modes of the posterior distribution. This is different from the use of a proxy to the optimal importance density in the sequential importance sampling/resampling framework, since in the latter, the proposal density depends on the locations of the particles at the previous time step and its parameters have to be computed as many times as there are particles. In contrast, the proposed methodology only requires the approximation of the modes of the posterior which is a priori less time consuming than computing the mode of the OID for each particle. We have recalled an efficient method for maximum a posteriori approximation in cases of interest, i.e. when the likelihood function is a function of a smaller part of the state vector. The ensuing mixture proposal lends itself to the mixture particle filtering framework which is convenient when dealing with multimodality. We have then combined this importance sampling procedure with

the Mixture Rao-Blackwellized Particle Filter. Moreover, we have proposed several choices for the covariance matrix as an additional parameter of the importance density. These are based on an approximation of the mixture component posterior covariance which ensures that the importance sampling asymptotic variance remains finite. We have compared the mixture RBPF with mode based importance sampling, with the simple mixture RBPF and the standard RBPF in a terrain aided positioning scenario. We considered both a challenging setting consisting in very accurate measurements as well as more usual situation with a higher variance of the additive noise. In both situations, it was found that the use of the mode based mixture in the mixture RBPF yielded less divergent tracks. In the case of a low sensor noise, the filter accuracy in terms of mean squared error was also greatly improved. The effect of different covariance matrices used in the proposal density was also investigated and it was found to have moderate effect in the positioning application. However, we believe it can have a higher impact in other applications where the measurement is not as ambiguous since the difference in shape and orientation between the posterior and the predictive distribution at a given time step may be important since, in that case, one single measurement carries noisy but near-complete information about the hidden state. Overall, this implementation of a mixture particle filter is an alternative to the auxiliary particle filter in cases where it is difficult to give an appropriate approximation of the predictive likelihood, a step necessary to compute the auxiliary weights : for instance, strategies that fit a Gaussian distribution may fail when dealing with multi-modal distributions. In the same way, the MRBPF avoids linearizations such as in the particle filter with EKF proposal that can yield an unstable filter whenever the measurement function is highly non-linear. Nevertheless, care must be taken when the number of modes in the posterior is unknown and high as the clustering step can be time consuming since non-parametric techniques (e.g. the mean-shift algorithm) have to be used.

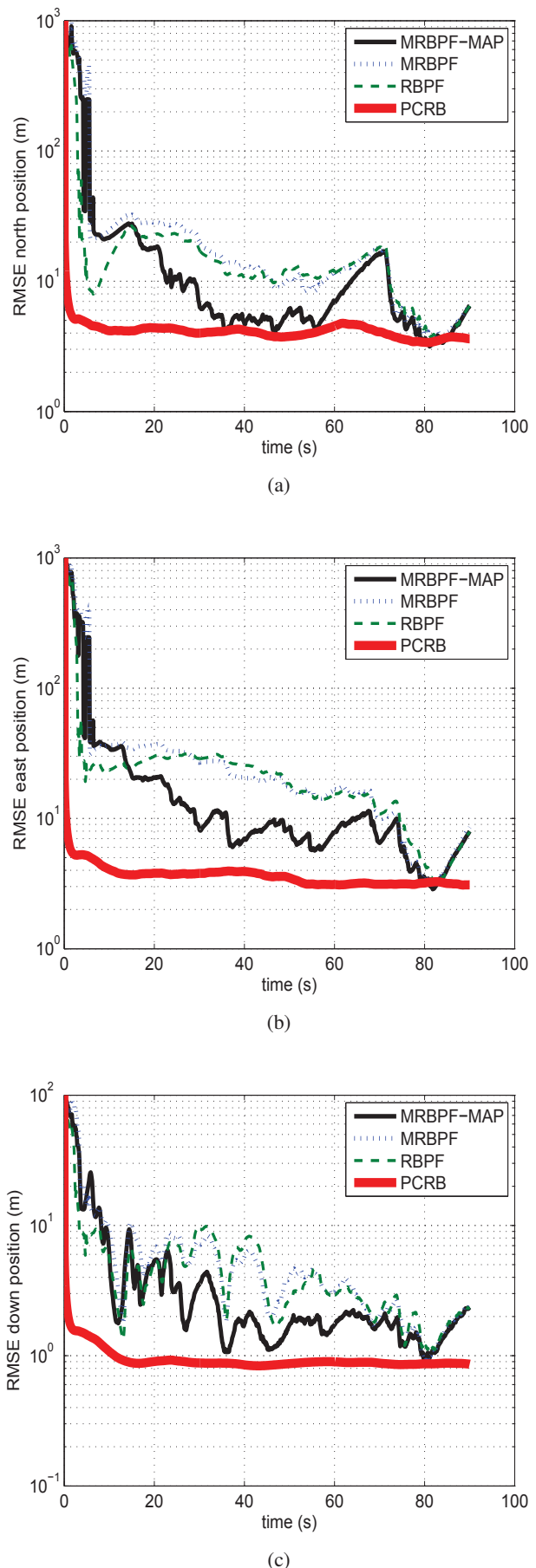


Fig. 3.  $\sigma_v = 5m$ : RMSE of the position drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)

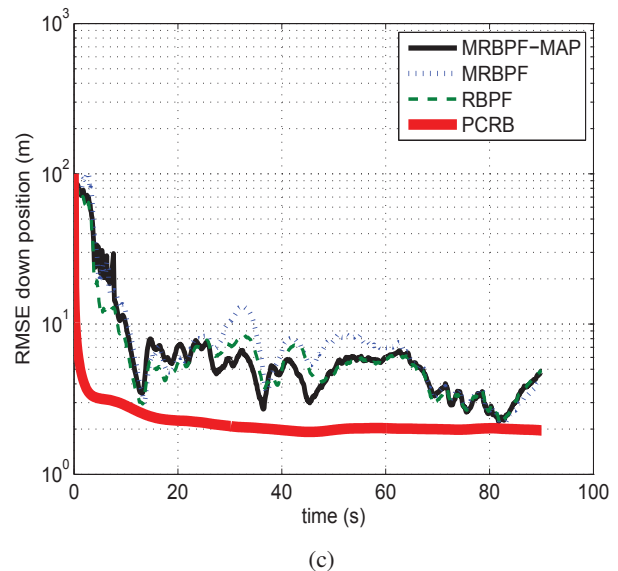
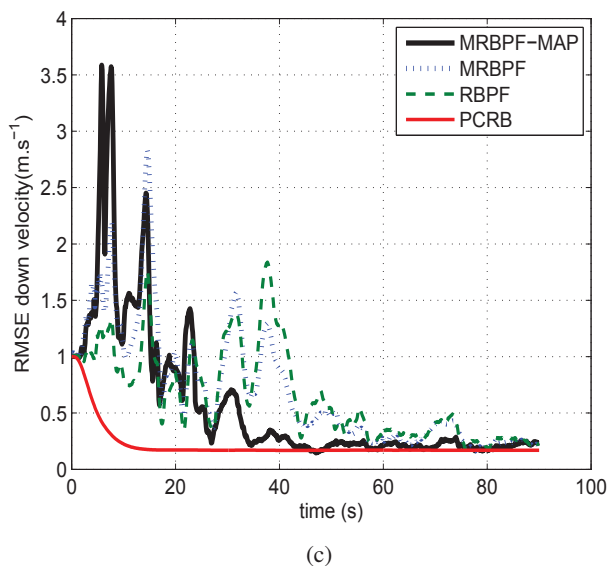
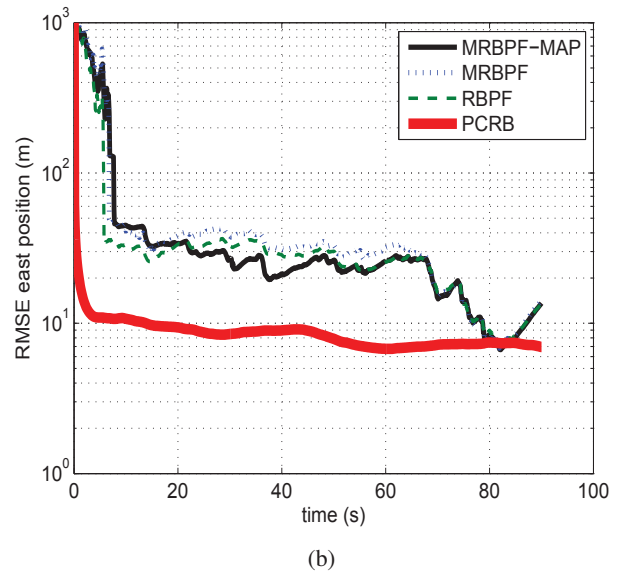
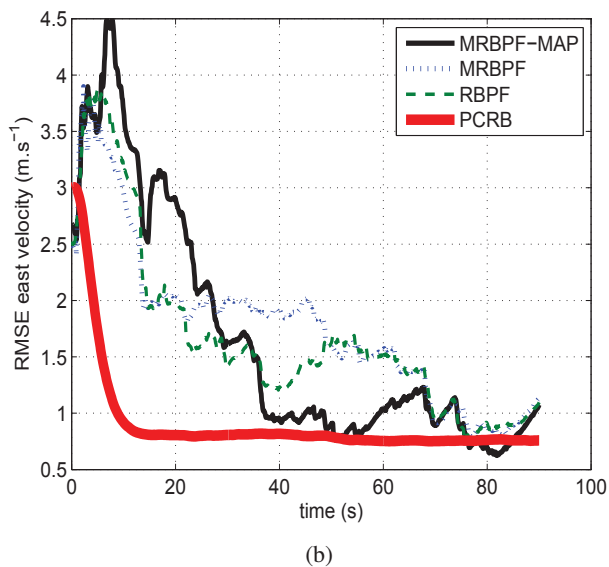
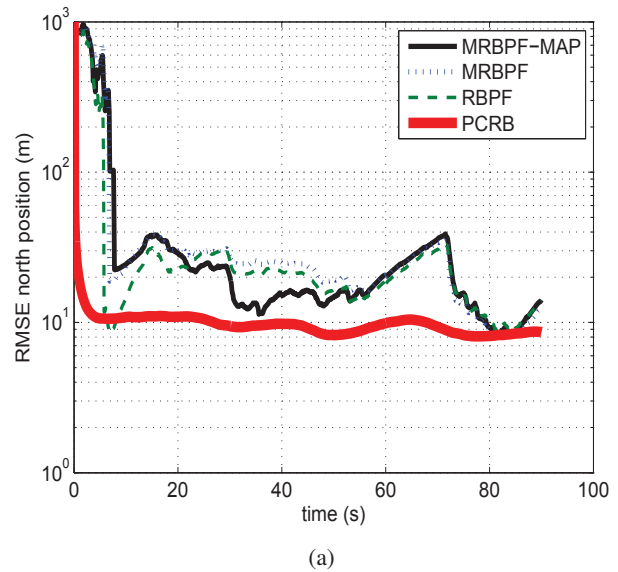
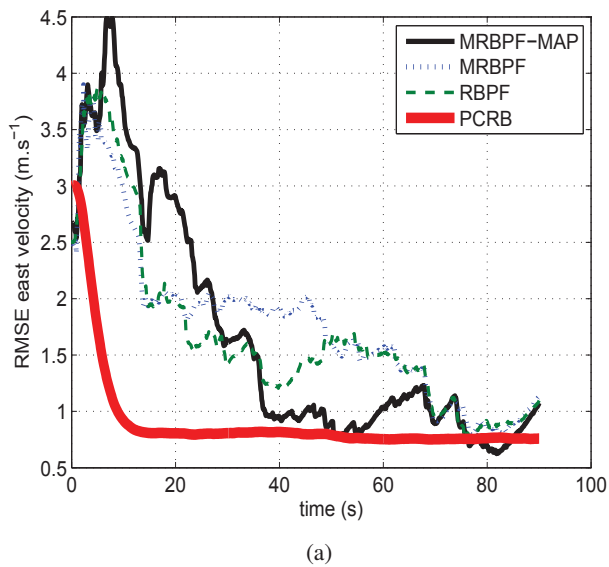


Fig. 4.  $\sigma_v = 5m$ : RMSE of the velocity drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)

Fig. 5.  $\sigma_v = 15m$ : RMSE of the position drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)

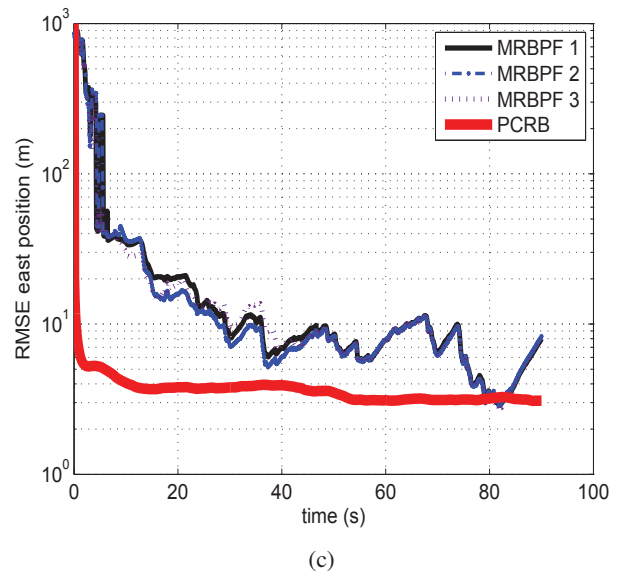
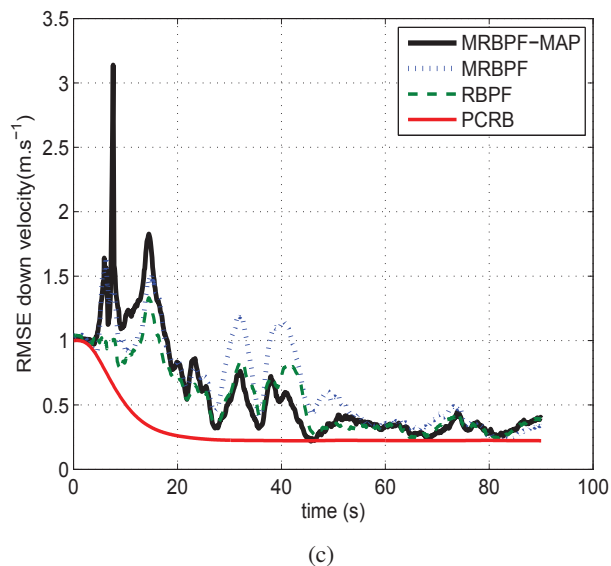
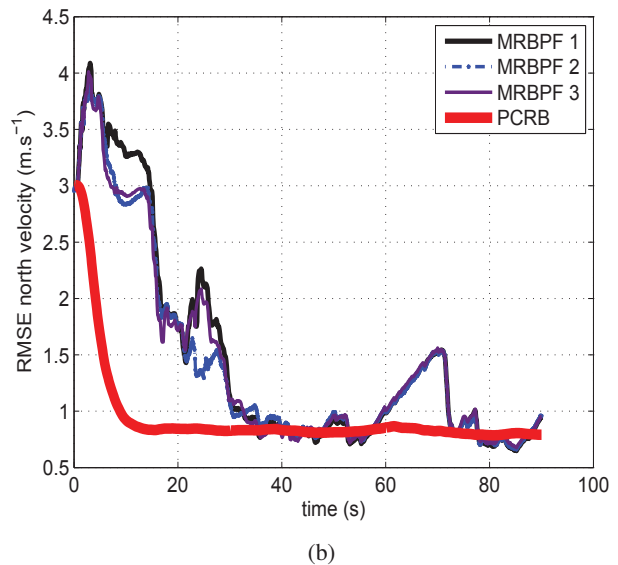
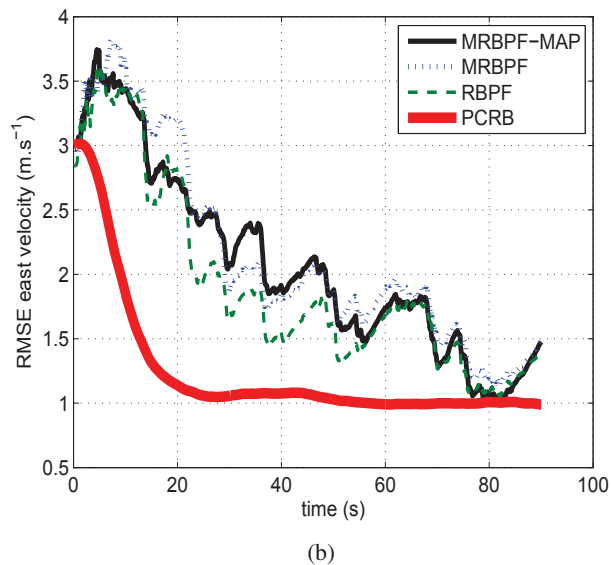
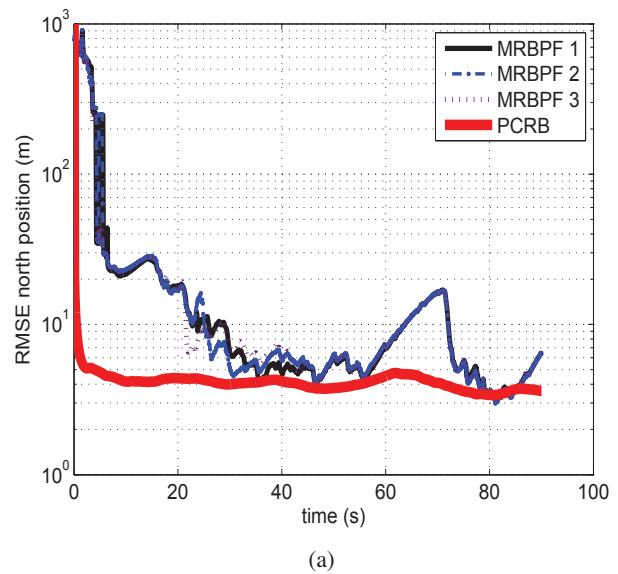
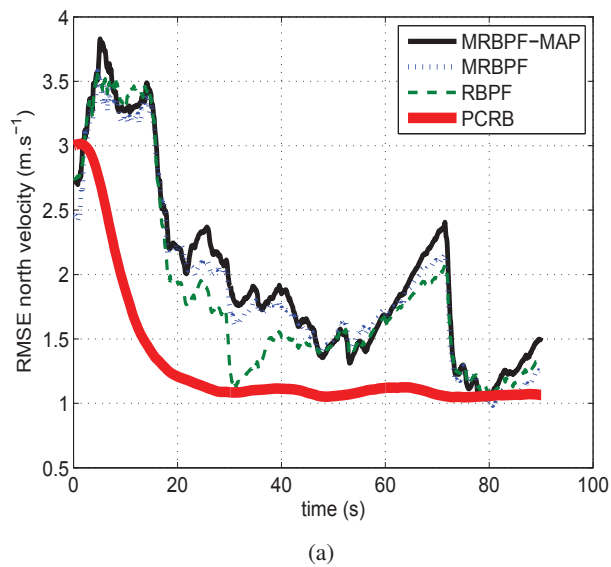
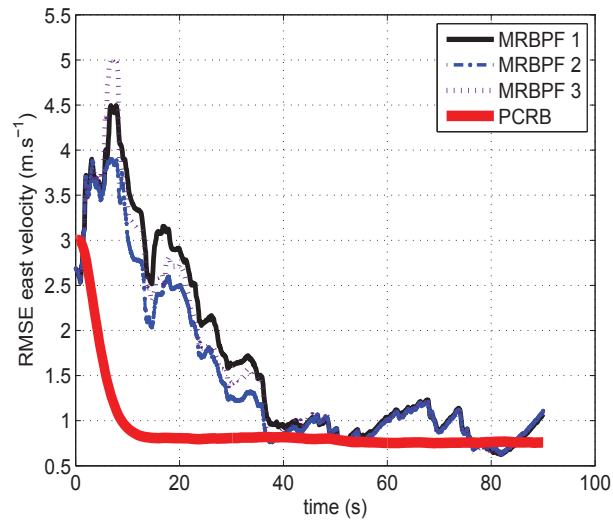
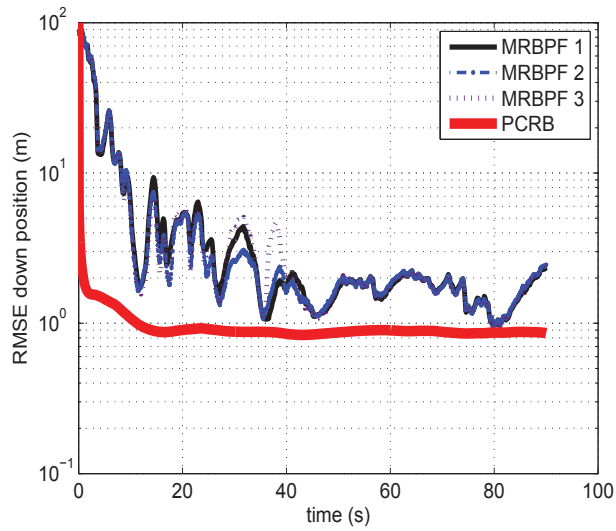


Fig. 6.  $\sigma_v = 15m$ : RMSE of the velocity drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)

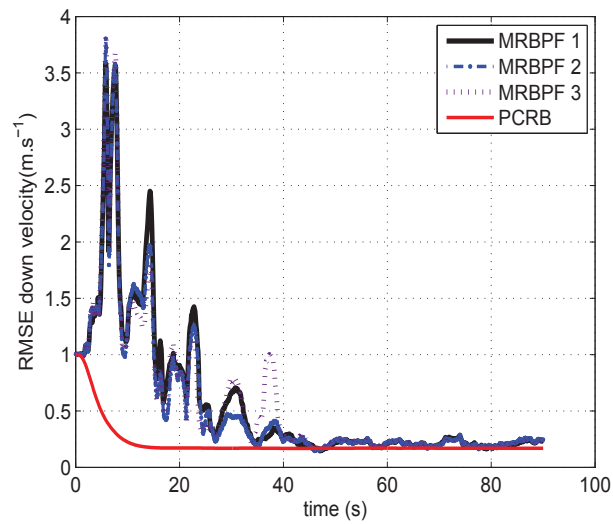
Fig. 7. (c) RMSE of the position drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)



(a)



(b)



(c)

Fig. 8. (c) RMSE of the velocity drift estimates ; red line : Posterior Cramer-Rao bound (PCRB)

## APPENDIX A

### RBPF ALGORITHM

**Algorithm 1 : Rao-Blackwellized Particle Filter**

**for**  $i = 1, \dots, N$

- sample  $x_{0,1}^i \sim p(x_{0,1})$
- set  $x_{0|-1,2}^i = \bar{x}_{0,2}$  and  $P_{0|-1}^i = P_0$

**end for**

**for**  $k = 1, 2, \dots$

- *[Particle measurement update]* compute  $\tilde{\omega}_k^i = \omega_{k-1}^i p(y_k | x_{k,1}^i, y_{0:k-1})$   
where

$$p(y_k | x_{k,1}^i, y_{0:k-1}) = \mathcal{N} \left( h_k(x_{k,1}^i) + H_k^T(x_{k,1}^i) \hat{x}_{k|k-1,2}^i, H_k(x_{k,1}^i) P_{k|k-1}^i H_k^T(x_{k,1}^i) + R_k \right) \quad (53)$$

- let  $\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{j=1}^N \tilde{\omega}_k^j}$ ,  $i = 1, \dots, N$
- compute  $N_{eff}$  according to (5)
- *[Resampling]* if  $N_{eff} < N_{th}$  resample the particle system
- *[Kalman filter update]* for  $i = 1, \dots, N$ , compute  $\hat{x}_{k,2}^i$  and  $P_k^i$  according to:

$$\hat{x}_{k,2}^i = \hat{x}_{k|k-1,2}^i + K_k^i (y_k - h_k(x_{k,1}^i) - H_k^T(x_{k,1}^i) \hat{x}_{k|k-1,2}^i) \quad (54)$$

$$P_k^i = P_{k|k-1}^i - K_k^i M_k^i (K_k^i)^T \quad (55)$$

where

$$\begin{aligned} M_k^i &= H_k(x_{k,1}^i) P_{k|k-1}^i H_k^T(x_{k,1}^i) + R_k \\ K_k^i &= P_{k|k-1}^i H_k^T(x_{k,1}^i) (M_k^i)^{-1} \end{aligned}$$

- *[Particle filter time update]* for  $i = 1, \dots, N$ , sample  $x_{k+1,1}^i \sim p(x_{k+1,1} | x_{0:k,1}^i, y_{0:k})$   
where  $p(x_{k+1,1} | x_{0:k,1}^i, y_{0:k})$  is the Gaussian density

$$\mathcal{N} \left( f_{k+1,1}(x_{k,1}) + F_{k+1,1}(x_{k,1}) \hat{x}_{k,2}^i, F_{k+1,1} P_k(F_{k+1,1})^T + G_{k+1,1} Q_{k+1}^{11} (G_{k+1,1})^T \right) \quad (56)$$

- *[Kalman filter time update]* for  $i = 1, \dots, N$ , compute  $\hat{x}_{k+1|k,2}^i$  and  $P_{k+1|k}^i$

$$\left\{ \begin{array}{lcl} \hat{x}_{k+1|k,2}^i & = & \bar{F}_{k-1,2} \hat{x}_{k,2}^i + G_{k,2} (Q_k^{21})^T (G_{k,1} Q_k^{11})^{-1} z_{k+1,1}^i + f_{k,2}(x_{k,1}^i) \\ & + & L_k^i (z_{k,1}^i - A_{k,1} \hat{x}_{k,2}^i) \\ P_{k|k-1}^i & = & \bar{F}_{k,2} P_k^i (\bar{F}_{k,2})^T + G_{k,2} \bar{Q}_k^{22} (G_{k,2})^T - L_k^i N_k^i L_k^{i,T} \\ N_k^i & = & F_{k,1} P_k^i (F_{k,1})^T + G_{k,1} Q_k^{11} (G_{k,1})^T \\ L_k^i & = & \bar{F}_{k,2} P_k^i (F_{k,1})^T (N_k^i)^{-1} \end{array} \right. \quad (57)$$

where

$$\left\{ \begin{array}{lcl} z_{k,1} & = & x_{k+1,1} - f_{k,1}(x_{k,1}) \\ z_{k,2} & = & y_k - h_k(x_{k,1}) \end{array} \right.$$

and

$$\left\{ \begin{array}{lcl} \bar{F}_{k,2} & = & F_{k,2} - G_{k,2} (Q_k^{21})^T (G_{k,1} Q_k^{11})^{-1} F_{k,1} \\ \bar{Q}_k^{22} & = & Q_k^{22} - (Q_k^{21})^T (Q_k^{11})^{-1} Q_k^{21} \end{array} \right.$$

**end for**

## APPENDIX B

### PROOF OF PROPOSITION 1

**Lemma 1.** *Assume the prior  $q$  is a Gaussian density with mean  $\mu$  and covariance  $P$ . Then if  $\Sigma - P$  is positive definite then  $\sup_{x \in \mathbb{R}^d} \frac{q(x)}{\hat{q}(x)} < \infty$  and  $V_{\hat{q}}$  is finite.*

The proof is readily obtained by working out the log ratio  $\log \frac{q(x)}{\hat{q}(x)}$  and is omitted here for brevity. Since a sufficient condition for finite asymptotic variance of the importance weights is given by  $\Sigma - P > 0$ , let us characterize this inequality. First, this is equivalent to  $P^{-1} - \Sigma^{-1} > 0$ .

$$\Sigma^{-1} - P^{-1} < 0 \Leftrightarrow x^T(\Sigma^{-1} - P^{-1})x < 0 \quad \forall x \in \mathbb{R}^d \setminus \{0\} \Leftrightarrow \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} < 1 \quad (58)$$

Let  $P^{-1} = D^T D$  be the Choleski decomposition of  $P^{-1}$  where  $D$  is an upper triangular matrix. Let  $C = D^{-T} \Sigma^{-1} D^{-1}$  where  $D^{-T} = (D^{-1})^T$ . Then,

$$\frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} = \frac{x^T D^T C D x}{x^T D^T D x}$$

Setting  $y = Dx$ ,

$$\frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} = \frac{y^T C y}{y^T y}$$

yields

$$\sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} < 1 \Leftrightarrow \sup_{y \in \mathbb{R}^d \setminus \{0\}} \frac{y^T C y}{y^T y} < 1 \quad (59)$$

It is a well known result ([28, p. 176]) that the supremum of the Rayleigh quotient  $\frac{y^T C y}{y^T y}$  equals the largest eigenvalue of  $C$ .

## APPENDIX C

### PROOF: NEAREST SYMMETRIC POSITIVE DEFINITE MATRIX SATISFYING FINITE IMPORTANCE WEIGHTS ASYMPTOTIC VARIANCE

Recall the optimization problem

$$\text{minimise } \|\Sigma - \hat{J}^{-1}\|_F^2 \text{ such that } \Sigma - P > 0$$

The minimizer may be obtained thanks to Higham's theorem.

**Theorem 1** (N. Higham [29]). *Let  $A$  a real symmetric matrix. The following optimization problem:*

$$\text{minimise } \|X - A\|_F^2 \text{ s.t. } X \geq 0$$

has a unique solution  $X^*$  defined by

$$X^* = \frac{1}{2}(A + H) \quad (60)$$

where  $H$  is a symmetric positive semidefinite matrix such that  $A = UH$  where  $U^T U = I$ .  $A = UH$  is the polar decomposition of  $A$ .

Noticing that  $\|\Sigma - \hat{J}^{-1}\|_F^2 = \|(\Sigma - P) - (\hat{J}^{-1} - P)\|_F^2$ , we have

$$\min_{\Sigma - P \geq 0} \|\Sigma - \hat{J}^{-1}\|_F^2 = \min_{X \geq 0} \|X - (\hat{J}^{-1} - P)\|_F^2$$

Let  $X_F = \arg \min_{X \geq 0} \|X - (\hat{J}^{-1} - P)\|_F$ . Then according to Higham's theorem,

$$X_F = \frac{1}{2}(\hat{J}^{-1} - P + H)$$

where  $U$  and  $H$  are such that  $\hat{J}^{-1} - P = UH$  and  $U^T U = I$ . It follows that

$$\Sigma_F = \frac{1}{2}(\hat{J}^{-1} + P + H) \quad (61)$$

Note that the initial goal was to minimise  $\|\Sigma - \hat{J}^{-1}\|_F^2$  subject to  $\Sigma - P > 0$  while the above minimizer satisfies the constraint  $\Sigma - P \geq 0$ . Therefore, to satisfy the positive definite constraint, one may consider the matrix

$$\Sigma_F = \frac{1}{2}(\hat{J}^{-1} + P + H) + \kappa I_d$$

where  $\kappa$  is a small positive constant and  $I_d$  is the  $d$  dimensional identity matrix. Formula (61) requires the derivation of  $H$  which can be obtained as outlined in [29] as follows: Let  $\hat{J}^{-1} - P = Z\tilde{\Lambda}Z^T$  be the orthogonal diagonalization of the symmetric real matrix  $\hat{J}^{-1} - P$ , where  $\tilde{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d)$  is the diagonal eigenvalue matrix and  $ZZ^T = I$ . Then setting

$$\tilde{\Lambda}_+ = \text{diag}(\max(0, \tilde{\lambda}_1), \dots, \max(0, \tilde{\lambda}_d))$$

entails

$$\Sigma_F = Z\tilde{\Lambda}_+Z^T + P \quad (62)$$

## APPENDIX D MRBPF-MAP ALGORITHM

---

### Algorithm 2: MRBPF-MAP

---

**if**  $k = 0$

- sample  $x_{0,1}^i \sim p(x_{0,1})$ ,  $i = 1, \dots, N$
- set  $x_{0,2}^i = \bar{x}_{0,2}$  and  $P_0^i = \bar{P}_0$

**end if**

**if**  $k = 1, \dots, n$ ,

- *[Particle time update]* draw  $x_{k|k-1,1}^i \sim p(x_k^n | x_{0:k-1}^n, y_{0:k-1})$  (cf. (56)),  $i = 1, \dots, N$ . Set  $\omega_{k|k-1}^i = \omega_{k-1}^i$ ,  $i = 1, \dots, N$ .
- *[Kalman filter time update]* for  $i = 1, \dots, N$ , compute  $\hat{x}_{k|k-1,2}^i$  and  $P_{k|k-1}^i$  according to (57).
- *[Clustering]* Cluster the particle set into  $M_k$  clusters  $C_{j,k}$  using meanshift clustering for instance with bandwidth parameter  $h_{MS}$ .  
 $[C'_{j,k}, I'_{j,k}] = \text{meanshiftClustering}(\{x_k^i\}_{i=1}^N, \{\omega_k^i\}_{i=1}^N, K, h^{MS})$
- *[Particle measurement update]* Compute the unnormalized weights  $\tilde{\omega}_k^i = \omega_{k-1}^i p(y_k | x_{k,1}^i, y_{0:k-1})$ ,  $i = 1, \dots, N$  where the likelihood  $p(y_k | x_{k,1}^i, y_{0:k-1})$  is given by (53).
- *[Importance sampling]* for  $j = 1, \dots, M_k$ 
  - $\beta_j = \sum_{i \in I_{j,k}} \tilde{\omega}_k^i$
  - $\omega_k^i = \frac{\tilde{\omega}_k^i}{\beta_j}$ ,  $i \in I_{j,k}$
  - compute  $N_{eff}^j$  according to (45)
  - $\text{indexMAP}_j = 0$
  - if  $N_{eff}^j / N_j \leq N_{th,MAP}$ 
    - \* set  $\bar{x}_{k|k-1,1}^j = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i x_{k|k-1,1}^i$
    - \* set  $\bar{x}_{k|k-1,2}^j = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i x_{k|k-1,2}^i$
    - \* set  $\hat{P}_{j,k|k-1}^{11} = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i (x_{k|k-1,1}^i - \bar{x}_{k|k-1,1}^j)(x_{k|k-1,1}^i - \bar{x}_{k|k-1,1}^j)^T$
    - \* set  $\hat{P}_{j,k|k-1}^{22} = \sum_{i \in I_{j,k}} \left( P_{k|k-1}^i + \omega_{k|k-1}^i (x_{k|k-1,2}^i - \bar{x}_{k|k-1,2}^j)(x_{k|k-1,2}^i - \bar{x}_{k|k-1,2}^j)^T \right)$
    - \* set  $\hat{P}_{j,k|k-1}^{12} = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i (x_{k|k-1,1}^i - \bar{x}_{k|k-1,1}^j)(x_{k|k-1,2}^i - \bar{x}_{k|k-1,2}^j)^T$
    - \* set
$$\hat{P}_{j,k|k-1} = \begin{pmatrix} \hat{P}_{j,k|k-1}^{11} & \hat{P}_{j,k|k-1}^{12} \\ \left( \hat{P}_{j,k|k-1}^{12} \right)^T & \hat{P}_{j,k|k-1}^{22} \end{pmatrix}$$
  - and  $\bar{x}_{j,k|k-1} = \begin{bmatrix} \left( \bar{x}_{k|k-1,1}^j \right)^T & \left( \bar{x}_{k|k-1,2}^j \right)^T \end{bmatrix}^T$ .
  - \* let  $q_j(x) = \varphi(x | \bar{x}_{j,k|k-1}, \hat{P}_{j,k|k-1})$  be the Gaussian approximation of the prior of the  $j$ -th component.
  - \* compute  $\hat{x}_{j,k}^* = \arg \max_{x_k \in \mathbb{R}^d} g_k(x_k) q_j(x_k)$  as outlined in section III-C.
  - \* set  $\hat{q}_j(x) = \varphi(x | \hat{x}_{j,k}^*, \Sigma_j^*)$  where  $\Sigma_j^* = \hat{P}_{j,k|k-1}$ ,  $\Sigma_j^* = P_{rot}$  or  $\Sigma_j^* = \Sigma_F$  (see (32))
  - \* draw  $N_j$  i.i.d. samples  $\xi^i \sim \hat{q}_j$ ,  $i \in I_{j,k}$  and  $N_j = \text{Card}(I_{j,k})$
  - \* set  $x_{k,1}^i = \xi_1^i$  and  $x_{k,2}^i = \xi_2^i$  where  $\xi^i = \begin{bmatrix} \xi_1^i \\ \xi_2^i \end{bmatrix}$ .
  - \* compute  $\tilde{\omega}_k^i$  and  $\omega_k^i$  according to (42)
  - \* set  $\beta_{j,k} = \frac{1}{N_j} \sum_{i \in I_{j,k}} \tilde{\omega}_k^i$
  - \*  $\text{indexMAP}_j = 1$
- for  $j = 1, \dots, M_k$ , compute  $\alpha_{j,k}$  as a function of  $\{\alpha_{j,k-1}\}_{1 \leq j \leq M_{k-1}}$  and  $\{\beta_{j,k-1}\}_{1 \leq j \leq M_{k-1}}$  using equation (47)

- Compute the weighted average of the non-linear substate  $\hat{x}_{k,1} = \sum_{j=1}^{M_k} \sum_{i \in I_{j,k}} \omega_k^i x_{k,1}^i$ .
  - *[Resampling]* let  $j = 1$ .
    - while  $j \leq M_k$ 
      - if  $\alpha_{j,k} < \alpha_{min}$  remove the cluster  $C_{j,k}$  and sample  $N_j$  new particles according to the empirical mixture distribution with weights  $\frac{\alpha_{l,k}}{\sum_{r \neq j} \alpha_{r,k}}$ ,  $l \neq j$
      - else
        - \* compute  $N_{eff}^j$  according to (45)
        - \* if  $\frac{N_{eff}^j}{N_j} < N_{th}$  resample the particle cloud of the current cluster. Additional local regularization may be performed as described in [19]
        - \*  $j = j + 1$
  - *[Kalman measurement update]*
  - compute  $P_k^i$  according to (55).
    - for  $j = 1, \dots, M_k$ 
      - \* if  $\text{indexMAP}_j = 0$ , compute  $\hat{x}_k^{l,i}$ ,  $i \in I_{j,k}$  using (54)
  - work out the posterior mean of the linear substate  $\hat{x}_{k,2} = \sum_{j=1}^{M_k} \sum_{i \in I_{j,k}} \omega_k^i x_{k,2}^i$
- end for**
-

## REFERENCES

- [1] R. Runnalls and P. Groves, "Terrain-referenced navigation using the igmap data fusion algorithm," in *ION 61st Annual Meeting The MITRE Corporation & Draper Laboratory*, 2005.
- [2] N. Bergman, "Recursive Bayesian Estimation: Navigation and Tracking Applications," Doctorate Thesis, Linkping University, 1999.
- [3] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, apr 1993.
- [4] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.
- [5] P.-J. Nordlund and F. Gustafsson, "Marginalized particle filter for accurate and reliable terrain-aided navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 4, pp. 1385–1399, oct. 2009.
- [6] C. Musso and N. Oudjane, "Regularization schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem," in *Proceedings of the Irish Signals Systems Conference*, 1998.
- [7] D. Fox, S. Thrun, F. Dellaert, and W. Burgard, "Particle filters for mobile robot localization," in *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.
- [8] C. Musso, N. Oudjane, and F. Le Gland, "Improving regularized particle filters," in *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001, pp. 247–271.
- [9] C. Musso, P. Quang, and F. Le Gland, "Introducing the laplace approximation in particle filtering," in *Proceedings of the 14th International Conference on Information Fusion*, july 2011, pp. 1–8.
- [10] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter : particle filters for tracking applications*. Artech House, 2004.
- [11] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [12] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.
- [13] G. Grisetti, G. D. Tipaldi, C. Stachniss, W. Burgard, and D. Nardi, "Fast and accurate slam with rao-blackwellized particle filters," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 30–38, 2007.
- [14] C. Andrieu and A. Doucet, "Particle filtering for partially observed gaussian state space models," *Journal of the Royal Statistical Society*, vol. 64, no. 4, pp. 827–836, 2002.
- [15] G. Casella and C. P. Robert, "Rao-Blackwellisation of Sampling Schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.
- [16] R. Chen and J. Liu, "Mixture Kalman filters," *J. R. Statist. Soc. B*, vol. 62, pp. 493–508, 2000.
- [17] J. Vermaak, A. Doucet, and P. Perez, "Maintaining multimodality through mixture tracking," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, oct. 2003, pp. 1110–1116 vol.2.
- [18] K. Okuma, A. Taleghani, N. Freitas, J. Little, and D. Lowe, "A boosted particle filter: Multitarget detection and tracking," *European Conference on Computer Vision*, vol. 3021, pp. 28–39, 2004.
- [19] A. Murangira, C. Musso, K. Dahia, and J. Allard, "Robust regularized particle filter for terrain navigation," in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, july 2011, pp. 1–8.
- [20] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan, "The unscented particle filter," in *Advances in Neural Information Processing Systems 13*, Nov. 2001.
- [21] P. Bunch and S. Godsill, "Particle filtering with progressive gaussian approximations to the optimal importance density," in *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2013, pp. 360–363.
- [22] A. Murangira and C. Musso, "Proposal distribution for particle filtering applied to terrain navigation," in *Proceedings of the 21st European Signal Processing Conference*, 2013.
- [23] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–39, November 1989.
- [24] A. Kong, J. S. Liu, and W. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, pp. 278–288, 1994.
- [25] B. Jungbacker and S. J. Koopman, "Monte Carlo estimation for nonlinear non-Gaussian state space models," *Biometrika*, vol. 94, no. 4, pp. 827–839, 2007.
- [26] K. Dahia, "Nouvelles méthodes en filtrage particulaire. Application au recalage de navigation inertielle par mesures altimétriques," Doctoral Thesis, Université Joseph Fourier, 2005.
- [27] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, 1998.
- [28] R. A. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [29] N. J. Higham, "Computing a nearest symmetric positive semidefinite matrix," *Linear Algebra and Its Applications*, vol. 103, pp. 103–118, 1988.



Achille Murangira Achille Murangira received his M.Sc. in Transmission Technologies from Telecom ParisTech in 2008 and a Ph.D. in systems optimization and dependability from the Université de Technologie de Troyes under the supervision of Igor Nikiforov in 2014. He carried his research in particle filtering and INS technology at ONERA from 2014 to 2013. From 2014 to october 2015, he was a postdoctoral researcher at IFPEN, a French research institute in energy, transport and environment.



Christian Musso Christian Musso received a Ph.D. in applied mathematics from the Université Joseph Fourier in Grenoble (France) in 1993. He worked for the French naval division (DCAN) in 1986 where he implemented optimal observer maneuvers algorithms and quick estimators in Target Motion Analysis. Since 1991, he's been a research engineer at ONERA, the French Aerospace Lab, in the field of signal processing and information processing. His main research interests are estimation, tracking, detection and Monte Carlo methods for nonlinear filtering and for optimization.



**Karim Dahia** Karim Dahia is a senior navigation engineer with the ONERA French Aerospace Lab in Palaiseau, France. His Ph.D. from the Université Joseph Fourier - Grenoble focused on the application of particle filtering to aircraft motion estimation. Dr. Dahia's research interests include robust and optimal navigation as well as filtering for aerospace systems.