Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME 2023)
19-21 July 2023, Tenerife, Canary Islands, Spain

A long simulation scenario for evaluation of multi-target tracking methods

Edmund Førland Brekke
Dept. of Engineering Cybernetics
Norwegian University of Science and Technology
Trondheim, Norway
edmund.brekke@ntnu.no

Audun Gullikstad Hem
Dept. of Engineering Cybernetics
Norwegian University of Science and Technology
Trondheim, Norway
audun.g.hem@ntnu.no

***Abstract*—Algorithms for multi-target tracking are typically evaluated by means of simulations. Typically, the simulated scenarios are fairly short and of very simple geometry. In contrast, this paper describes a scenario which was deliberately and explicitly designed to exhibit a large challenging behavior for several targets moving in close formation for about 23 minutes.The scenario is inspired by experiments with autonomous surface vehicles (ASVs). We use the scenario to study the performance differences between tracking methods in the Poisson Multi-Bernoulli Mixture (PMBM) family. A multiple models approach outperforms the single model approach, but is to a large extent outperformed by the multiple hypothesis approach.**

***Index Terms*—simulation, trajectory tracking, multi-target tracking, multiple models, multiple hypotheses**

I. Introduction

Algorithms for multi-target tracking play a fundamental role in many safety-critical systems, including vehicle autonomy, surveillance and defense. To verify that these algorithms perform as intended, they must be tested on relevant data. The ultimate test is to test the algorithm as part of the complete real-world system, e.g., as part of a closed-loop collision avoidance system for vehicle autonomy. Before this stage is reached, the algorithm is typically tested on previously recorded data sets, and before that on simulated data sets. Simulations play an important role because they enable testing a much larger number of scenarios, because of reproducability, because one can explore the impact of small changes in the test data, and because simulations can easily be designed to cover edge cases, that may not even be safe to explore in the real world.

Multi-target tracking methods typically utilize models of target kinematics, sensor resolution, clutter intensity, target birth intensity etc. Thus, one approach to simulation is simply to generate data directly from these models. If this philosophy is strictly adhered to, one should let target trajectories be born according to the birth model, as done in e.g. [1] [2]. The simulated targets will then for most of the time be far away from each other. This is not unreasonable if the goal is to reproduce typical behavior in the real world, but it will fail to provide significant focus on the bona-fide multi-target challenge of closely spaced targets.

Another approach is something we can call the *common midpoint technique*. In this approach, several targets are simulated so that they come towards the same midpoint from different directions [3]. The latter parts of their trajectories are simulated by means of the kinematics model, while the former parts are simulated by means of the corresponding reverse kinematics model. By playing with the sampling time and the process noise covariance, the common midpoint technique can provide a fairly rich ensemble of scenarios. However, the realism of all interacting targets approaching the same point can be questioned. A possible adaptation of the common midpoint technique can be to simulate several subsequent obstacle targets that attempt at intercepting an "ownship" target [4].

In this paper, we simulate several targets which move in close formation for more than a thousand time steps, where each time step is assumed to be in the order of one second. This is inspired by the experiment and data set recorded in [5], where 3 high-speed rigid inflatable bodies (RIBs) unintentionally took part in an autonomous collision avoidance experiment. The scenario described here can be seen as a scaled up version of this real scenario. We call the simulated scenario "The 9 ravens".

The main contribution of the paper is to present a simulated benchmark scenario, consisting of measurements from 1397 simulated radar scans in 2 dimensions. The scenario has already been used to assess novel techniques for multiple hypothesis data association in [6] and [7]. We also provide the underlying trajectory data, from which one can generate variations of the same scenario, with different time resolution, measurement noise, missed detections and clutter intensity. In contrast to the aforementioned approaches, the scenario is to a large extent manually designed. The explanation of this design process is also a contribution of the paper. A final contribution of the paper is to provide a comparison of an Interacting

This work was supported in part by the Research Council of Norway through Projects 223254, 295033 and 309230.

979-8-3503-2297-2/23/$31.00

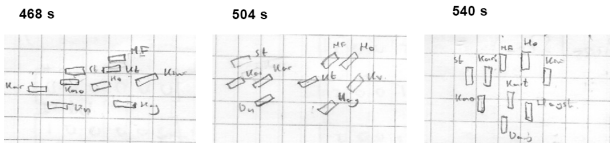


Fig. 1: Original drawings.

Multiple Model Joint Probabilistic Data Association (IMM-JIPDA) filter and a Poisson Multi-Bernoulli Mixture (PMBM) filter for maneuvering target tracking. The paper also presents a technique for trajectory tracking not published before.

II. The design process

The intention behind the scenario was to simulate a formation of several highly agile vessels or drones, with interactions and non-trivial dynamics. The construction of the scenario can be divided into a design phase, a simulation phase and a data phase. The design phase of the scenario consisted of the following steps:

1) Draw the instantaneous configurations at 36s intervals.
2) Make an average path for the formation.
3) Make an average velocity for the formation.

The simulation phase involved the following elements:

1) A curvilinear trajectory tracker for following the nominal trajectories from the design phase.
2) A collision avoidance method to ensure that the vessels do not collide with each other or with other vessels.
3) Environmental noises.

In the data phase the measurements used for multi-target tracking were generated from the trajectories, according to a generalized measurement model encompassing:

1) a choice of radar-carrying ownship.
2) a choice of sensor resolution.
3) detection probability.
4) clutter intensity.

A. The design phase

61 scenes involving the 9 vessels were drawn on paper. In each scene, the vessels were placed so that the sequence of configurations could be translated to meaningful maneuvres and interactions. Some of these drawings are shown in Figure 1.

Based on these drawings, the scenes were translated into numbers in an Excel sheet. In addition to these lateral displacements, height displacements were also made, so that the scenario can be used in 3-dimensional mode if desirable. The next step was to draw the average path of the formation. Here Photoshop was used, and the path was extracted using a Hough-based curve tracking tool implemented in Matlab. For the entire path, a velocity assignment was made by spline interpolation around 11 control points. This profile is shown in Figure 2. The profile begins with velocities typical for surface craft, and then increases towards velocities more typical for cars and small aircraft. Based on the velocity profile, the path was changed into a time-parameterized trajectory. Individual trajectories for each vessel were generated by adding the formation displacements from the Excel table to this average trajectory.

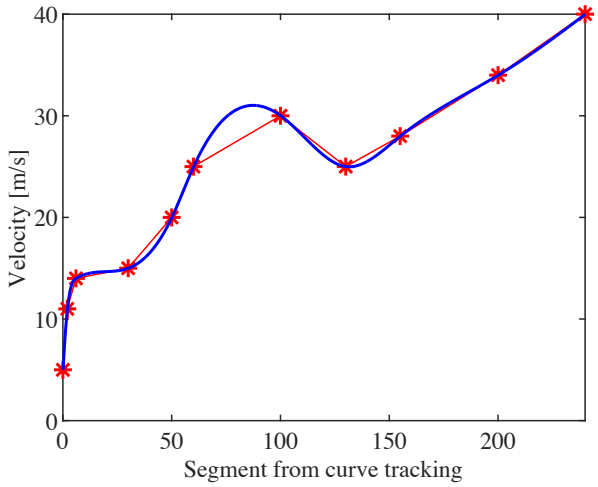


Fig. 2: Velocity profile used to generate trajectory from path.

B. The simulation phase

In order to simulate the individual vessels, the two possibilities of *path following* and *trajectory tracking* were considered. To ensure that the formation would stay together, without the need for more complicated formation control schemes, trajectory tracking along the individual trajectories was preferred over path following.

Because the average path was curvilinear, and because we wanted to simulate underactuated vessels, a curvilinear non-holonomic trajectory tracker suitable for horizontal motion was needed. For the sake of simplicity, a novel line-of-sight approach inspired by [8] was invented for these simulations. A first iteration of this technique was presented in [9]. For each vessel, we are given the desired trajectory $\boldsymbol{\rho}_d(t)$, a positive time lag L, the vehicle position $\boldsymbol{\rho}(t)$, its body-relative velocities u and v, and its heading ψ. The desired heading and velocity are calculated as

$$\begin{aligned}\psi_d &= \angle(\boldsymbol{\rho}_d(t) - \boldsymbol{\rho}(t)) \\ u_d &= -v\xi + \sqrt{1+\xi}(U - k_x \begin{bmatrix}1 & 0\end{bmatrix} \mathbf{x}_e)\end{aligned}$$

where

$$\begin{aligned}\xi &= \tan\left(\beta + \arctan\left(\begin{bmatrix}0 & 1\end{bmatrix} \mathbf{x}_e/\Delta\right)\right) \\ \Delta &= \|\boldsymbol{\rho}_d(t+L) - \boldsymbol{\rho}_d(t)\| \\ \mathbf{x}_e &= \mathsf{R}(\gamma)^{\mathsf{T}}(\boldsymbol{\rho}(t) - \boldsymbol{\rho}_d(t)) \\ \gamma &= \angle(\boldsymbol{\rho}_d(t+L) - \boldsymbol{\rho}_d(t)) \\ U &= \|\dot{\boldsymbol{\rho}}_d(t)\|\end{aligned}$$

and k_x is a positive gain constant. Feedback linearization along the lines of [8] was used to obtain the control inputs affecting $\dot{u}$ and $\dot{\psi}$.

A problem with this scheme is that u_d can potentially grow very large in cases where the vehicle by accident is moving in the wrong direction. As a solution, we saturate the desired velocity towards an upper limit by means of a smooth, concave function. The vessel will then try to fix its orientation first, and then try to catch up with the trajectory. Let $u_b = 2$, $\epsilon = 0.4$ and $c = \pi/3$. We define

$$\begin{aligned} u_0 =& u_b \left(-U/u_b - \|\mathbf{x}_e\|^2\right) \\ \gamma_b =& (\psi - \psi_d)^2 - c^2 \\ \chi_b =& 0.5 \left(1 + \frac{\gamma_b}{\sqrt{\epsilon^2 + \gamma_b^2}}\right) \end{aligned}$$

and find the new desired velocity as $\chi_b u_0 + (1 - \chi_B) u_d$.

Environmental noises were included in the surge and sway accelerations. These were drawn from correlated K-distributions according to the recipe in [10]. The rationale behind this was that the K-distribution gives a more realistic variability than purely Gaussian noise would give.

The nominal values of u_d and ψ_d were overridden in cases where they would lead to collisions. In addition to the 9 vessels of the formation, some additional vessels were simulated with straight-line kinematics in order to force sudden deviations from the nominal formation pattern. For this a velocity obstacle method was used [11], with 24 u_d-ψ_d-pairs distributed by varying either u_d or ψ_d away from the guidance values.

C. The data phase

The input to the data phase was a collection of the 9 time-parameterized trajectories the formation vessels, including position, orientation and velocity, stored as a file. In the subsequent simulations, vessel number 1 was used as ownship. That is, the sensor data were simulated to come from a radar mounted on vessel number 1. A sensor model of the form

$$\mathbf{z}_p^t = \mathrm{c2p}(\mathsf{R}(-\psi^o)(\boldsymbol{\rho}^t - \boldsymbol{\rho}^o) + \mathbf{w} + \mathbf{e}$$

where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_p)$ was used to generate measurements in polar coordinates. The superscripts t and o stand for target and ownship, respectively. $\mathsf{R}(\cdot)$ is a 2-dimensional rotation matrix, and $\mathrm{p2c}(\cdot)$ is the polar-to-Cartesian mapping, with inverse $\mathrm{c2p}(\cdot)$. The polar covariance matrix $\mathbf{R}_p = \mathrm{diag}(\sigma_r^2, \sigma_\theta^2)$ was designed to mimic the resolution of a typical recreational boat radar, by setting $\sigma_r = 2$ and $\sigma_\theta = 2^\circ$. The polar measurements were converted to Cartesian measurements in the world frame according to

$$\begin{aligned} \mathbf{z}_c^t =& \mathsf{R}(\psi^o)\mathrm{p2c}(\mathbf{z}_p^t) + \mathbf{e} \\ \mathbf{e} \sim& \mathcal{N}(\mathbf{0}, \mathsf{R}(\psi^o)\mathbf{R}_e\mathsf{R}(\psi^o)^\mathsf{T}) \end{aligned}$$

where $\mathbf{e}$ was a noise component representing the impact of target extent. The matrix $\mathbf{R}_e = \mathrm{diag}(1^2, 0.5^2)$ could represent the centroid uncertainty of a small recreational vessel, or possibly a fixed wing unmanned aerial vehicle (UAV). Ownship navigational uncertainty was not included in the simulation.

The detection probability was set to $P_\mathrm{D} = 0.9$. This means that for each vessel, at any time step, there was a 10% probability that its measurement would be discarded. It can easily be reduced by omitting more measurements. We simulated the clutter measurements in the world frame as well. These were generated according to a Poisson point process with constant intensity 0.0013 defined over a disk with radius $100\mathrm{m}$ around the ownship.

III. The scenario

The scenario consists of 1397 frames of measurements, together with the underlying ground truth trajectories, and the correct target-to-measurement associations. It is available on codeocean.com with digital object identifier 10.24433/CO.7713078.v1.

An overview picture of the trajectories is shown in Figure 4. One can see how the motion varies between fairly straight segments and turns of varying sharpness. The average velocity is overall increasing, from about $5\mathrm{m/second}$ to $40\mathrm{m/second}$, thus covering values suitable for recreational boats, RIBs, cars and drones, see Figure 2. While the the overall picture of the velocities was designed by hand, the accelerations result from the guidance, disturbances and collision avoidance of the simulation phase. Figure 5 shows how the accelerations suddenly may change from a calm regime to a more chaotic regime. In this case it was triggered by a collision avoidance maneuver of the dark blue vessel around time step 1110.

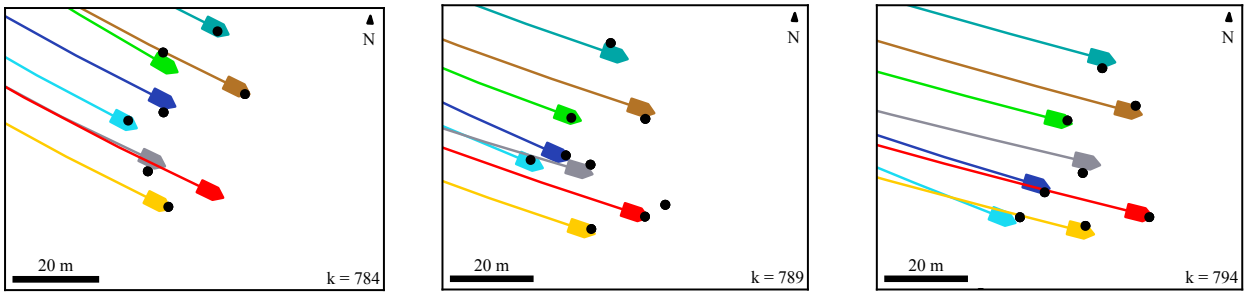
Some snapshots from the scenario are shown in Figure 3. Let us make the following three observations. First, we notice that both misdetections (Red target at $784\mathrm{second}$) and clutter (near Red target at $789\mathrm{second}$) are present. Second, significant changes in the formation may happen during a $5\mathrm{second}$ interval (follow the relative placement of Cyan target). Third, the high velocity means that there is a little to no overlap between the convex hulls of the formation at consecutive time steps. As a consequence, if a track needs to be re-initialized, several possible track-to-measurement assignments must be considered.

In Figure 6 we compare the minimal and average distances between targets in the formation with the average measurement noise strength, defined as the square root of the determinant of the measurement noise covariance matrices. We see that almost all the time, at least two of the targets in the formation are so close to each other that it will not be obvious which measurement came from which target.

IV. Comparison of VIMMJIPDA and PMBM

We investigated the performance of three open source multi-target tracking methods on the simulated data set: the implementation of the Poisson Multi-Bernoulli Mixture filter reported in [3], and the Visibility Interacting Multiple Model Joint Integrated Probabilistic Data Association (VIMMJIPDA) of [5], in triple-mode and in single-mode versions. Parameters for the tracking methods are provided in Table I.

The PMBM filter is in principle an optimal multi-hypothesis tracking method. However, as with any model-based technique, it can only be expected to deliver good results when the models are adequate. The implementation uses standard



(a) 784s (b) 789s (c) 794s

Fig. 3: Snapshots of the simulated targets (colored) and measurements (black dots) at three time steps.

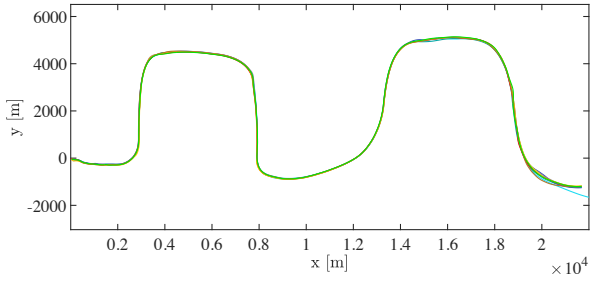


Fig. 4: Overview of the scenario

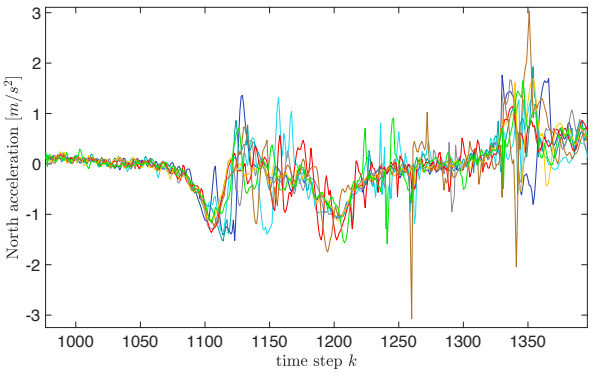


Fig. 5: Accelerations of the 8 targets during the last 7 minutes.

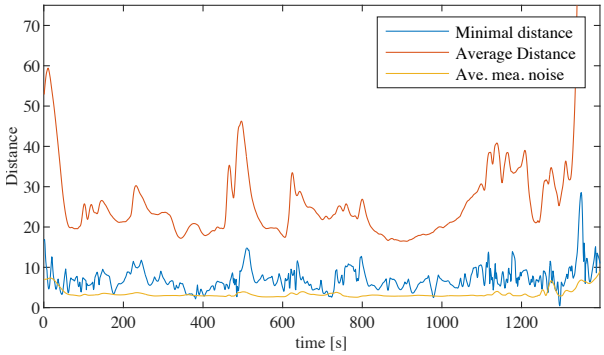


Fig. 6: Minimal and average target distances

TABLE I: Tracking system parameters for PMBM, VIMMJIPDA and JIPDA (which uses only JIPDA CV model 2).

Quantity	Symbol	Unit	Value
Common parameters			
Radar sample interval	T	[s]	1
Cartesian noise std.	σ_{c_R}	[m]	1.0
Polar range std.	σ_r	[m]	2.0
Polar bearing std.	σ_θ	[°]	0.0035
Detection probability	P_{D}	[−]	0.9
Survival probability	P_{S}	[−]	1
Gate size	γ^2	[−]	20
Max. number of hypotheses	$N_{\max}$	[−]	400
Area radius	R	[m]	100
Clutter intensity	λ	$[\mathrm{m}^{-2}]$	$4/(\pi R^2)$
Initial velocity std.	σ_v	$[\mathrm{m\,s}^{-1}]$	10
PMBM parameters			
Poisson pruning threshold	T_{Pp}	[−]	1.0×10^{-5}
Bernoulli pruning threshold	T_{Bp}	[−]	0
Process noise intensity	$q_{a,1}$	$[\mathrm{m}^2\,\mathrm{s}^{-3}]$	0.8^2
Birth weight		$[\mathrm{m}^{-2}]$	1×10^{-3}
JIPDA parameters			
Existence confirmation threshold		[−]	0.99
Existence termination threshold		[−]	1.0×10^{-5}
CV model 1 process noise intensity	$q_{a,1}$	$[\mathrm{m}^2\,\mathrm{s}^{-3}]$	0.1^2
CV model 2 process noise intensity	$q_{a,2}$	$[\mathrm{m}^2\,\mathrm{s}^{-3}]$	1.5^2
Turn rate process noise intensity		$[\mathrm{rad}^2\,\mathrm{s}^{-3}]$	0.15^2
Initial existence probability	r_v	[-]	0.18
Initial visibility probability	η_v	[-]	0.9
Visibility transition probability	$\pi^{\tilde{s}s}$	[%]	$\begin{bmatrix} 48 & 52 \\ 10 & 90 \end{bmatrix}$
Initial model probability	μ_v^s	[%]	$\begin{bmatrix} 80 & 10 & 10 \end{bmatrix}$
IMM transition probability	$\pi^{\tilde{s}s}$	[%]	$\begin{bmatrix} 99 & .5 & .5 \\ .5 & 99 & .5 \\ .5 & .5 & 99 \end{bmatrix}$

constant velocity (CV) model. In the implementation, the PMBM filter accounts for the top 400 hypotheses.

The VIMMJIPDA does not maintain multiple hypotheses. Instead, it maintains multiple motion models. For each track and motion model, it uses mixture reduction over the different track-to-measurement hypotheses so that only a single Gaussian results. In its triple-mode version the VIMMJIPDA uses a low noise CV model, a high noise CV model and a constant turn-rate (CT) model. In the single-mode version it only uses the high noise CV model.

To analyze performance, we first look at the generalized

	GOSPA^2	TLE	TFR	TFAR	TPD	TSR	Runtime
PMBM	**51.47**	**1.79**	2.87×10^{-3}	6.36×10^{-3}	0.994	**0.054**	760.34 s
VIMMJIPDA	69.06	2.08	$\mathbf{0.72 \times 10^{-3}}$	**0.00**	**0.998**	0.165	**375.09** s
JIPDA	140.73	3.23	0.89×10^{-3}	8.95×10^{-5}	**0.998**	0.150	381.80 s

TABLE II: Performance indicators

optimal sub-pattern assignment (GOSPA) [12] metric, which provides a distance measure between the multi-target state estimate of a tracking method and the ground truth. It can be decomposed into a localization part, a misdetection part and a false alarm part, see Figure 7.

In Table II we also look at additional performance indicators: Track localization error (TLE), track fragmentation rate (TFR), track false-alarm rate (TFAR), track probability of detection (TPD) and track swap rate (TSR). All of these except the TSR are defined in [13]. We define the TSR as the number of track swaps per second. A track swap occurs if a track, after following a target for a while, starts following a different target. For each track we note from which targets the measurement in the most likely measurement association come from. A track follows a target if, after N associations with target-originated measurements, it has associated with said target over $N/2$ times. If, after having followed a target for N time steps, a track starts following another for N time steps, we consider this a track swap.

Both versions of VIMMJIPDA have about 3 times as many track swaps as the PMBM filter. Their higher GOSPA values are probably related to this, and also to issues with track coalescence. On the other hand, the VIMMJIPDA comes out slightly better with regards to track management measures, especially TPD. The reason for this could be that the PMBM filter is unable to decompose the scenario into independent clusters. Each hypothesis must account for everything that happens in the entire surveillance region, and this makes it inevitable, even with 400 hypotheses, that for some tracks only the most likely association will be considered. In some cases, this may be a faulty association, ruling out the true association. See also [6] for further discussion of this.

The reduction in GOSPA achieved by going from JIPDA to VIMMJIPDA is much larger than the additional reduction achieved by going from VIMMJIPDA to PMBM. Nevertheless, PMBM is able to beat VIMMJIPDA, even though it does not use multiple models. The run-time of PMBM is 2-3 times that of the JIPDAs, while there is no significant difference between JIPDA and VIMMJIPDA. The runtime of PMBM is close to linear in the number of allowed hypotheses.

V. Conclusion

This paper has shown an example of how simulations with an intermediate level of fidelity can be designed. While the data set itself is limited in scope, there exist several possibilities for tweaking it into more challenging scenarios, or easier scenarios. We hope that this work can serve as inspiration for other advances in benchmarking simulations for robotic perception solutions.

A core idea in this work was to arrange the different phases of the simulation as a cascade, where the intermediate knowledge could be stored as a file to be passed on to the next phase. In contrast, it is also of interest to design simulations of a similar level of fidelity where target tracking and sensor fusion is part of a closed loop system where guidance and collision avoidance depends on the perception methods [4]. In any case, if this work was to be repeated, more sophisticated approaches to guidance would be worth considering [14]. Ideas similar to those proposed here could be used to generate trajectories for high-fidelity simulators such as [15].

References

[1] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, Jul. 2013.

[2] E. F. Brekke and M. A. Chitre, "Success rates and posterior probabilities in multiple hypothesis tracking," in *Proc. Fusion*, Cambridge, UK, July 2018, pp. 949–956.

[3] A. F. Garcia-Fernandez, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1883–1901, 2018.

[4] E. S. Henriksen, "Automatic testing of maritime collision avoidance methods with sensor fusion," Master's thesis, NTNU, Jun. 2018.

[5] E. F. Brekke, A. G. Hem, and L.-C. N. Tokle, "Multitarget tracking with multiple models and visibility: Derivation and verification on maritime radar data," *IEEE Journal of Oceanic Engineering*, vol. 46, no. 4, pp. 1272–1287, 2021.

[6] E. F. Brekke and L.-C. N. Tokle, "Hypothesis exploration in multiple hypothesis tracking with multiple clusters," in *Proc. Fusion*, Linköping, Sweden, 2022, pp. 1–8.

[7] O. A. Severinsen, L.-C. N. Tokle, and E. F. Brekke, "Belief propagation for marginal probabilities in multiple hypothesis tracking," 2023, submitted to Fusion 2023, Charleston, SC, USA.

[8] A. Lekkas and T. Fossen, "Minimization of cross-track and along-track errors for path tracking of marine underactuated vehicles," in *Proc. ECC*, June 2014, pp. 3004–3010.

[9] A. L. Flåten and E. F. Brekke, "Stability of line-of-sight based trajectory-tracking in two dimensions," in *Proc. CCTA*, Aug 2017, pp. 760–765.

[10] K. D. Ward, R. J. A. Tough, and S. Watts, *Sea Clutter: Scattering, the K Distribution and Radar Performance*, ser. IET Radar, Sonar and Navigation. London, UK: The Institution of Engineering and Technology, 2006, vol. 20.

[11] P. Fiorini and Z. Shiller, "Motion planning in dynamic environments using velocity obstacles," *International Journal of Robotics Research*, vol. 17, no. 7, pp. 760 – 72, 1998.

[12] A. S. Rahmathullah, A. F. Garcia-Fernandez, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proc. Fusion*, Xi'An, China, 7 2017.

[13] S. Coraluppi, D. Grimmett, and P. D. Theije, "Benchmark evaluation of multistatic trackers," in *Proc. Fusion*, Florence, Italy, July 2006.

[14] C. Paliotta, E. Lefeber, K. Y. Pettersen, J. Pinto, M. Costa, and J. T. de Sousa, "Trajectory tracking and path following for underactuated marine vehicles," *IEEE Transactions on Control Systems Technology*, pp. 1–15, 2018.

[15] K. Vasstein, E. Brekke, R. Mester, and E. Eide, "Autoferry Gemini: A real-time simulation platform for electromagnetic radiation sensors on autonomous ships," in *Proc. ICMASS*, Virtual conference, 2020.

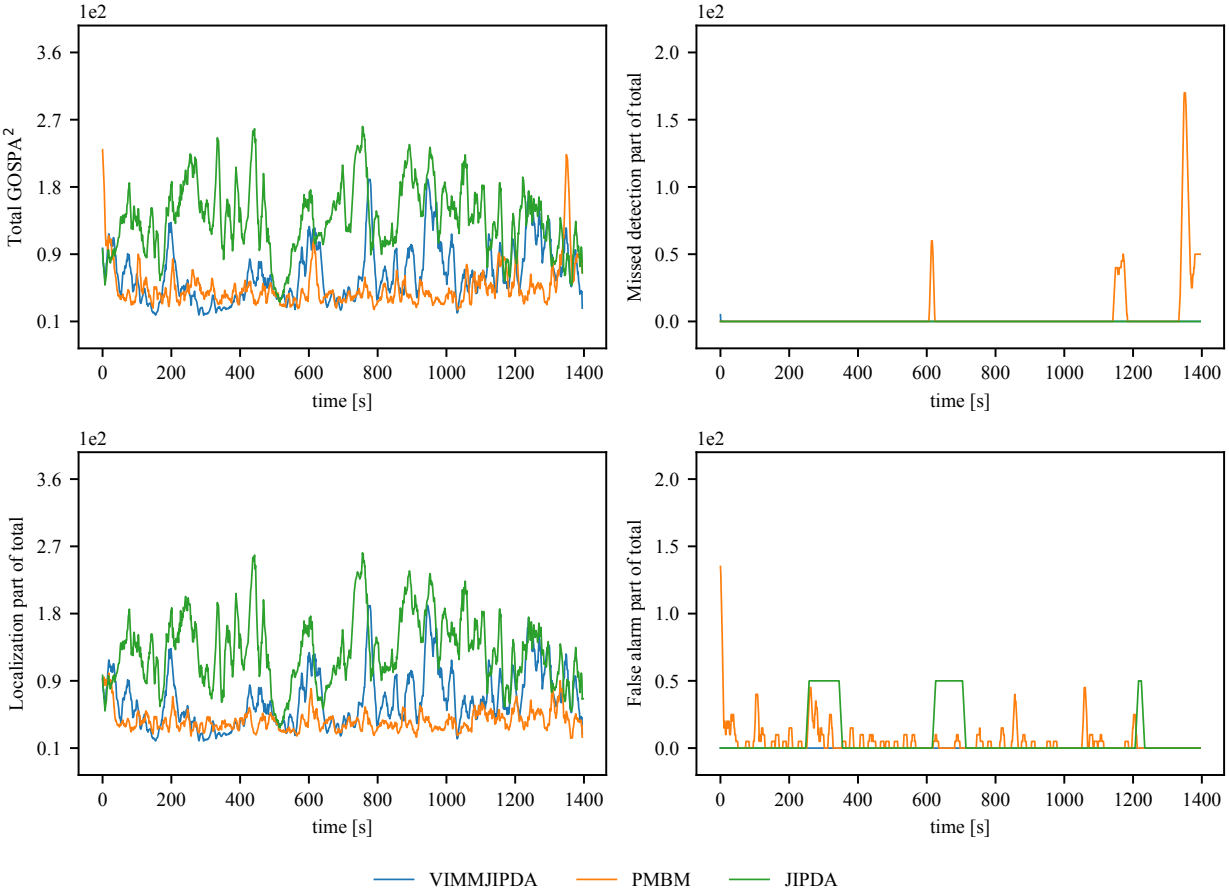
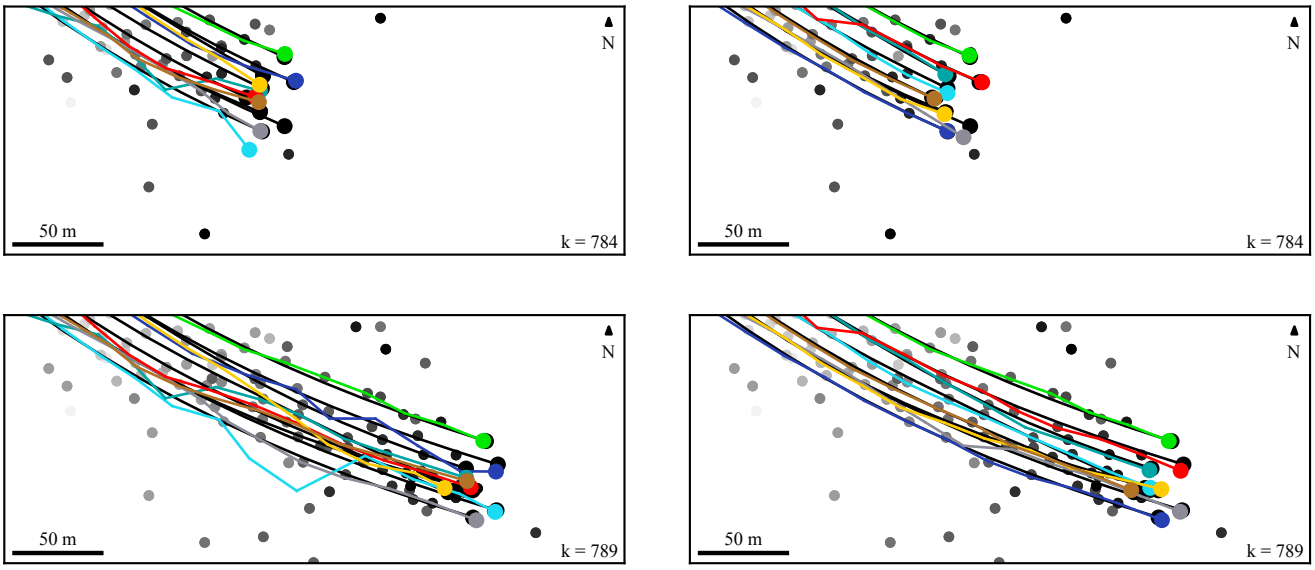


Fig. 7: Moving average of GOSPA^2 error for the PMBM, VIMMJIPDA and JIPDA filters, over 10 time steps.



(a) VIMMJIPDA

(b) PMBM

Fig. 8: Scenario between time steps 779 and 789. The left column shows the targets until time step 784, and the right column shows the targets until time step 789.