HAL
open science

Approches bayésiennes pour le pistage radar de cibles de surface potentiellement manoeuvrantes

Clément Magnant

To cite this version:

Clément Magnant. Approches bayésiennes pour le pistage radar de cibles de surface potentiellement manoeuvrantes. Autre. Université de Bordeaux, 2016. Français. NNT : 2016BORD0136 . tel-01398519

HAL Id: tel-01398519

https://theses.hal.science/tel-01398519v1

Submitted on 17 Nov 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.

THÈSE

présentée
pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

Spécialité : Automatique, Productique, Signal et Image, Ingénierie Cognitique

par

Clément MAGNANT

Approches bayésiennes pour le pistage radar de cibles de surface potentiellement manœuvrantes

Soutenue le 21 Septembre 2016 devant le jury composé de :

M. Noyer Jean-Charles	Professeur à l'Université du Littoral Côte d'Opale, Calais	Rapporteur
M. Chabert Marie	Professeur à l'INP de Toulouse	Rapporteur
M. Giovannelli Jean-François	Professeur à l'Universite de Bordeaux, Talence	Examinateur
M. Ratton Laurent	Responsable du service Algorithmie Amont à Thales Systèmes Aéroportés, Elancourt	Référent scientifique
M. Giremus Audrey	Maître de Conférences à l'Université de Bordeaux, Talence	Co-encadrante de thèse
M. Grivel Eric	Professeur à l'ENSEIRB-MATMECA, Bordeaux INP, Talence	Directeur de thèse
M. Montigny Richard	Responsable du service Département Algorithmie et Traitement, Pessac	Membre invité
M. Joseph Bernard	Ingénieur Etudes Radar, Expert pistage à Thales Systèmes Aéroportés, Pessac	Membre invité

Préparée à Thales Systèmes Aéroportés
25, avenue Gustave Eiffel - 33608 Pessac Cedex
et au Laboratoire IMS, UMR CNRS 5218, Université de Bordeaux, Bordeaux INP
351, avenue de la Libération - 33405 Talence Cedex

Remerciements

Comme pour tous les doctorants, les premières lignes d'un manuscrit sont souvent celles écrites en dernier. N'échappant pas à la règle, je vais veiller à n'oublier personne.

Tout d'abord, je souhaite remercier mes rapporteurs, Marie Chabert et Jean-Charles Noyer, non seulement pour le temps qu'ils ont consacré à la relecture de mon manuscrit mais également pour leurs remarques pertinentes. Mes remerciements s'adressent aussi à Jean-François Giovannelli pour avoir accepté de présider mon jury de thèse.

Tout naturellement, je remercie à présent mes encadrants. Tout d'abord (honneur aux dames), merci à Audrey Giremus pour son encadrement, ses précieux conseils et pour avoir été à mon écoute tout au long de ces trois années. Ensuite, merci à Eric Grivel, pour sa confiance, son encadrement, son dévouement et pour m'avoir encouragé à toujours donner le meilleur de moi-même.

Côté entreprise, mes remerciements les plus chaleureux s'adressent à mes encadrants : Bernard Joseph et Laurent Ratton qui n'ont eu cesse de me soutenir et m'accompagner durant cette thèse. Je remercie également Richard Montigny pour m'avoir accueilli au sein de son service et pour m'avoir permis d'évoluer dans les meilleures conditions.
Un grand merci également à mes collègues et amis avec qui j'ai partagé de très bons moments. Ils auront rendu ces trois années autant agréables que mémorables : Alexandre, Benoit, Jérémy, Julien Po., Julien Pe., Léo, Lionel, Mickaël, Philippe, Sylvain, Timothée, Vincent C., Vincent M. puis tous les autres !

Côté laboratoire, merci aux différentes personnes du laboratoire IMS avec qui j'ai pu interagir. Beaucoup de bons souvenirs, avec notamment Ahmed, Brendan, Emilie, Guillaume B., Guillaume F., Guillaume Z., Jacques, Lionel, Marc, Mouna, Olivier, Pascal, Romain, Roxana, Salma, Vincent P., Yannick et bien d'autres !

Je viens à bout de cet exercice périlleux en adressant les plus profonds remerciements à ma famille sans qui je n'aurais pas pu accomplir tout ça. Ils m'ont toujours soutenu et cru en moi bien qu'ils n'aient jamais réellement compris l'objet de mes travaux ! Un merci tout particulier à Jenny. Enfin, une pensée tout particulière à mes grand-parents qui ont toujours été fiers que leur petit-fils puisse devenir docteur !

Table des matières

Remerciements **i**

Liste des abréviations **xv**

Liste des notations **xvii**

Introduction **1**

1 Etat de l'art sur le pistage de cibles **5**

1.1 Introduction 6

1.2 Estimation bayésienne pour le pistage monocible 6

1.2.1 Représentation dans l'espace d'état 7

1.2.2 Filtre de Bayes 8

1.2.3 Filtre de Kalman 8

1.2.4 Filtre particulaire 11

1.3 Pistage de cibles manœuvrantes : algorithmes à modèles multiples 18

1.3.1 Estimateur à modèles multiples dynamique optimal 19

1.3.2 Algorithme à modèles multiples interactifs (IMM) 22

1.3.3 Problématiques liées à l'utilisation des modèles multiples 25

1.4 Ensembles finis aléatoires pour le pistage multi-cibles 26

1.4.1 Définitions, notations et abréviations 26

1.4.2 Distributions de probabilités multi-cibles 28

1.4.3 Représentation multi-cibles du système 30

1.5 Filtrage et estimation bayésienne multi-cibles 32

1.5.1 Filtre de Bayes multi-cibles 33

1.5.2 Filtre *probability hypothesis density* (PHD) 33

1.5.3 Filtre *cardinalized probability hypothesis density* (CPHD) 35

1.6 Evaluation de performances pour les algorithmes de pistage multi-cibles 37

1.6.1 Estimateurs d'états multi-cibles 37

1.6.2 Distance OSPA 38

1.7 Conclusions 39

2 Comparaison de modèles d'état fondée sur la divergence de Jeffrey **41**

2.1 Introduction 42

2.2 Rappels 43

2.2.1 Modèles de mouvement 43
2.2.2 Divergences de Kullback-Leibler (KL) et de Jeffrey (JD) 46
2.3 JD entre deux modèles d'état 47
2.3.1 Calcul récursif 47
2.3.2 Analyse des termes additifs dans l'expression récursive de la JD 49
2.3.3 JD dans différentes bases de l'espace d'état 50
2.4 Classification d'un ensemble de M modèles d'état fondée sur la JD 51
2.4.1 Structuration en matrice des JDs entre modèles d'état 52
2.4.2 Analyse fondée sur une décomposition en valeurs propres 54
2.4.3 Application à la conception d'algorithmes à modèles multiples 56
2.5 Simulations et résultats 58
2.5.1 Application à l'estimation de paramètres : comparaison de $M = 2$ marches aléatoires 58
2.5.2 Application à la poursuite de cibles : comparaison de $M = 2$ modèles d'état 59
2.5.3 Application à la poursuite de cibles : classification de $M = 6$ modèles d'état 63
2.6 Conclusions et perspectives 68

3 Approches bayésiennes non-paramétriques pour le pistage de cibles 71
3.1 Introduction 72
3.2 Processus de Dirichlet 73
3.2.1 Introduction aux méthodes bayésiennes non-paramétriques 73
3.2.2 Définition des processus de Dirichlet 75
3.2.3 Représentation dite en *stick breaking* 76
3.2.4 Urne de Polya 78
3.3 Estimation non-paramétrique de la matrice de covariance du bruit de modèle . . 80
3.3.1 Description du problème 80
3.3.2 Représentation dans l'espace d'état non-paramétrique 82
3.4 Cas des matrices de covariance avec une structure prédéfinie 84
3.4.1 Modèle hiérarchique considéré 85
3.4.2 Calcul de la distribution d'importance optimale 86
3.5 Application au pistage d'une cible 92
3.5.1 Protocoles de simulation 92
3.5.2 Résultats de simulation 97
3.6 Conclusions et perspectives 101

4 Pistage et classification conjoints de cibles 103
4.1 Introduction 104
4.2 Position du problème 105
4.2.1 Modélisation du système fondée uniquement sur les paramètres cinématiques de cible 105
4.2.2 Modélisation du système incluant la longueur de cible 106
4.3 Pistage et classification d'une cible à partir de mesures cinématiques et de mesures d'étalement distance 108
4.3.1 Filtre IMM multi-classes (MC-IMM) 109

4.3.2 Filtre particulaire multi-modèles multi-classes (MC-MMPF) 113
4.3.3 Extension au cas où les paramètres du modèle d'état sont inconnus 116
4.3.4 Application au pistage d'une cible marine 118
4.4 Pistage et classification de plusieurs cibles à partir de mesures cinématiques et de mesures d'étalement distance 127
4.4.1 Modélisation multi-cibles 127
4.4.2 Algorithme PHD multi-modèles (MM-PHD) multi-classes 128
4.4.3 Equations du filtre MM-PHD et implémentation par mélange de gaussiennes (MM-GMPHD) 129
4.4.4 Equations théoriques 129
4.4.5 Application au pistage de cibles marines 135
4.5 Conclusions et perspectives 143

Conclusions générales et perspectives **145**

Annexes **151**

A Pistage de cibles et algorithmes d'association de données **151**
A.1 Position du problème 151
A.2 Algorithmes *nearest neighbour* (NN) et *global nearest neighbour* (GNN) 153
A.3 Algorithmes *probabilistic data association* (PDA) et *joint probabilistic data association* (JPDA) 154
A.4 Algorithme *multiple hypothesis tracking* (MHT) 156

B Approximation de la loi optimale pour l'estimation de la matrice de précision du bruit de modèle **157**
B.1 Modélisation non-paramétrique du problème 158
B.2 Echantillonnage de la matrice de précision du bruit d'état 159
B.3 Echantillonnage du vecteur d'état 160
B.4 Echantillonnage de l'indice du cluster 161

C Filtre particulaire auxiliaire **163**

D Compléments mathématiques **165**
D.1 Conjugaison de deux lois normales 165
D.2 Distribution t de Student 167
D.3 Distribution t multivariée 168
D.4 Distribution t matricielle 169
D.5 Distribution de Wishart 169

Bibliographie **170**

Résumé **181**

Table des figures

1.1 Cycle du filtre de Kalman 10
1.2 Conditions d'utilisation des différents filtres bayésiens présentés 18
1.3 Evolution des modèles dans le cas optimal et selon l'algorithme IMM. Dans le premier cas, l'exposant au niveau de l'état estimé $\hat{\boldsymbol{x}}_k$ désigne la séquence de modèles. 21
1.4 Schéma bloc de l'algorithme IMM 25

2.1 Exemple de trois réalisations générées selon une REE donnée, dans le cadre de la modélisation du mouvement 44
2.2 Représentation des JDs entre modèles : exemple avec $M = 5$ modèles 53
2.3 Evolution de la matrice de dissimilarité transformée $E(k)$ au cours du temps . . 53
2.4 Evolution des valeurs propres pour la comparaison de M modèles de mouvement de type MUA 55
2.5 Schéma bloc de l'approche proposée pour la conception d'algorithmes à modèles multiples 57
2.6 Evolution de la JD selon le nombre d'échantillons et le rapport q des variances des processus générateurs des deux modèles 59
2.7 Comparaison de deux marches aléatoires selon le rapport q des variances des processus générateurs des deux modèles 59
2.8 Trajectoires générées à partir de modèles MUA et de Singer 60
2.9 Evolution de la JD en fonction du rapport des variances du jerk des modèles considérés 61
2.10 Influence du rapport des variances du jerk des modèles considérés sur la JD pour des mouvements de type MUA 61
2.11 Evolution de la JD en fonction de la constante du modèle de Singer τ_{Sin} et du nombre d'échantillons 62
2.12 Comparaison des modèles de Singer et MUA selon la valeur de τ_{Sin} 63
2.13 Evolution des valeurs propres au cours du temps où le paramètre $\lambda = 0.15$ est défini selon (2.52) 64
2.14 Evolution des valeurs propres au cours du temps avec le paramètre d'échelle $\lambda = 0.01$ 64
2.15 Classification d'un ensemble de six modèles en trois sous-ensembles de deux modèles 65

3.1 Réalisation du processus représentant le bruit de modèle associé à la trajectoire d'une cible manœuvrante et histrogramme associé. 74

3.2 Influence du paramètre d'échelle α_0 et du vecteur de probabilités initiales p_0 sur la distribution du processus de Dirichlet. A gauche : courbes de niveau de la densité de Dirichlet projetées sur le simplexe. A droite : nuages de points correspondant à des réalisations d'un processus de Dirichlet. 76
3.3 Représentation graphique du *stick breaking* 77
3.4 Réalisations d'un processus de Dirichlet de distribution de base $G_0 = \mathcal{N}(0,1)$ pour différentes valeurs du paramètre d'échelle α_0. 78
3.5 Séquences d'échantillons obtenues à partir de la représentation en urne de Polya pour $\alpha_0 = 1$ ou $\alpha_0 = 100$ et $G_0 \sim \mathcal{N}(0,1)$. 79
3.6 Représentation graphique du modèle hiérarchique 83
3.7 Réalisation de la trajectoire de la cible et de l'évolution de l'indice de la famille de modèles 94
3.8 Distributions *a priori* de type Gamma sur le paramètre de précision γ_k pour les différentes familles de modèles 95
3.9 Représentation au cours du temps des indices de famille de modèles estimés . . . 99
3.10 Influence du nombre de particules sur l'EQM 100
3.11 Influence du paramètre d'échelle sur l'EQM 100

4.1 Modèle elliptique de la cible 107
4.2 Schéma bloc des algorithmes proposés 110
4.3 Représentation graphique du modèle hiérarchique 117
4.4 Loi initiale sur la longueur de la cible 119
4.5 Trajectoire de la cible 2 et mesures associées 121
4.6 Probabilités des classes de la cible moyennées sur 100 simulations de Monte Carlo 122
4.7 Probabilités des classes pour une cible de 60 mètres manœuvrante (cas 1) ou non-manœuvrante (cas 2) 123
4.8 Temps de convergence des probabilités de classes selon la forme de la loi *a priori* sur la longueur de cible 124
4.9 Estimation de la taille de la cible, moyennée sur 100 simulations de Monte Carlo 126
4.10 Comparaison des performances du *MC-IMM* avec et sans les mesures de l'étalement distance de la cible 126
4.11 Trajectoires de cibles. Les points de départ sont désignés par des triangles alors que les points d'arrivée sont marqués par des cercles. 136
4.12 Trajectoires de cibles et états associés 137
4.13 Estimation de la cardinalité 138
4.14 Distances OSPA 138
4.15 Trajectoires de quatre cibles en croisement. Les points de départ sont désignés par des triangles alors que les points d'arrivée sont marqués par des cercles. . . . 139
4.16 Trajectoires de quatre cibles en croisement et états associés. Résultat d'une première réalisation. 140
4.17 Trajectoires de quatre cibles en croisement et états associés. Résultat d'une seconde réalisation. 140
4.18 Trajectoires de deux cibles en croisement et états associés 141

4.19 Estimation du nombre de cibles pour un scénario de croisement de deux cibles . 142
4.20 Estimation de la distance OSPA pour un scénario de croisement de deux cibles . 142

A.1 Fenêtrage statistique 152
A.2 Association de données dans un contexte de cibles proches 153

B.1 Représentation graphique de l'interdépendance des variables 158

D.1 Distribution de probabilité t de Student pour différentes valeurs du nombre de degrés de liberté ν 168

Liste des tableaux

2.1 EQM entre le vecteur d'état et son estimation pour des trajectoires de 100 échantillons, moyennée sur 200 simulations de Monte Carlo. 66
2.2 EQM entre le vecteur d'état et son estimation 67

3.1 Ordre de grandeur des coûts calculatoires des différentes approches testées 97
3.2 EQM moyennée sur 100 simulations de Monte Carlo 98
3.3 EQM moyennée sur 100 simulations de Monte Carlo pour différents réglages de la TPM. 99

4.1 Caractéristiques des différentes cibles marines 119
4.2 EQM pour la cible 1 (non-manœuvrante) 125
4.3 EQM pour la cible 2 (potentiellement manœuvrante) 125

Liste des algorithmes

1.1 Etapes du filtre particulaire générique 17
3.1 Estimation du paramètre de précision : une itération de l'algorithme (*Algo1*) proposé (tenant compte d'une structure prédéfinie de la matrice de covariance du bruit de modèle et fondé sur la loi optimale) 91
4.1 Etapes de l'algorithme IMM multi-classes proposé (MC-IMM) 112
4.2 Etapes de l'algorithme particulaire multi-classes proposé (MC-MMPF) 115
4.3 Etapes de *pruning* et de *merging* 134
B.1 Estimation de la matrice de précision du bruit de modèle : une itération de l'algorithme (*Algo2*) 162

Liste des abréviations

APF	*Auxiliary particle filter*, filtre particulaire auxiliaire
AR	Autorégressif
BNP	Bayésien non-paramétrique
CPHD	*Cardinalized probability hypothesis density*, densité d'hypothèse de probabilité de cardinal
EAP	*Expected a posteriori*, espérance *a posteriori*
EFA	Ensemble fini aléatoire
EM	*Expectation maximization*, espérance-maximisation
EQM	Erreur quadratique moyenne
FISST	*Finite set statistics*, statistiques des ensembles finis
FK	Filtre de Kalman
FKE	Filtre de Kalman étendu
FKSP	Filtre de Kalman sans parfum
FSG	Filtre par somme de gaussiennes
GLMB	*Generalized labeled multi-Bernoulli*, multi-Bernoulli labélisé généralisé
GM	*Gaussian mixture*, mélange de gaussiennes
GNN	*Global nearest neighbour*, algorithme du plus proche voisin global
GPS	*Global positioning system*, géopositionnement par satellite
HRR	*High-range resolution*, résolution à grande distance
i.i.d	Indépendants et identiquement distribués
IMM	*Interacting multiple models*, modèles multiples intéractifs
JD	*Jeffrey's divergence*, divergence de Jeffrey
JPDA	*Joint probability data association*, association de données aux probabilités jointes
JTC	*Joint tracking and classification*, pistage et classification conjoints
JVC	Jonker-Volgenant-Castanon
KL	Kullback-Leibler
LMB	*Labeled multi-Bernoulli*, multi-Bernoulli labélisé

LOS	*Line of sight*, ligne de vue
MA	Moyenne ajustée
MAP	Maximum *a posteriori*
MC	Monte Carlo
MemBer	*Multi-Bernoulli*
MHT	*Multiple hypothesis tracking*, poursuite par hypothèses multiples
MM	Modèles multiples
MRU	Mouvement rectiligne uniforme
MUA	Mouvement uniformément accéléré
NN	*Nearest neighbour*, plus proche voisin
OMAT	*Optimal mass transfer*, transfert de masse optimal
OSPA	*Optimal sub-pattern assignment*, affectation optimale
PDA	*Probability data association*, association de données probabiliste
PHD	*Probability hypothesis density*, densité d'hypothèse de probabilité
RBPF	*Rao-Blackwellized particle filter*, filtre particulaire Rao-Blackwellizé
REE	Représentation dans l'espace d'état
RFS	*Random finite set*, ensemble fini aléatoire
SIR	*Sequential importance resampling*, ré-échantillonage d'importance séquentiel
SIS	*Sequential importance sampling*, échantillonnage d'importance séquentiel
SMC	*Sequential Monte Carlo*, Monte-Carlo séquentiel
TPM	*Transition probability matrix*, matrice des probabilités de transition
TVAR	*Time-varying autoregressive*, autorégressif variant dans le temps
UKF	*Unscented Kalman filter*, filtre de Kalman sans parfum

Liste des notations

Chapitre 1

$\boldsymbol{x}_k$	Vecteur d'état à l'instant k
$\boldsymbol{y}_k$	Vecteur d'observation à l'instant k
$x_{k_1:k_2}$	Séquence des valeurs du processus aléatoire x entre les instants k_1 et k_2
$\mathbb{X}$	Espace d'état
$\mathbb{Y}$	Espace des observations
n_x	Dimension du vecteur d'état
n_y	Dimension du vecteur d'observation
f_k	Fonction d'évolution, aussi appelée fonction de transition, du vecteur d'état
$\boldsymbol{v}_k$	Bruit de modèle, aussi appelé bruit d'état
$\boldsymbol{w}_k$	Bruit de mesure
h_k	Fonction de mesure reliant le vecteur d'état au vecteur d'observation
F_k	Matrice d'évolution, aussi appelée matrice de transition
H_k	Matrice d'observation
Q_k	Matrice de covariance du bruit de modèle
R_k	Matrice de covariance du bruit de mesure
$\hat{\boldsymbol{x}}_{k\|k-1}$	Estimation *a priori* du vecteur d'état à l'instant k
$\hat{\boldsymbol{x}}_{k\|k}$	Estimation *a posteriori* du vecteur d'état à l'instant k
$P_{k\|k-1}$	Matrice de covariance de l'erreur d'estimation *a priori* à l'instant k
$P_{k\|k}$	Matrice de covariance de l'erreur d'estimation *a posteriori* à l'instant k
$\mathcal{N}(x;\mu,\Sigma)$	Densité de probabilité gaussienne pour la variable x de moyenne μ et de matrice de covariance Σ.
K_k	Gain de Kalman à l'instant k
S_k	Matrice de covariance de l'innovation $\tilde{y}_{k\|k-1}$ à l'instant k
$\hat{\boldsymbol{y}}_{k\|k-1}$	Mesure prédite à l'instant k étant donné les $k-1$ dernières observations
$\tilde{\boldsymbol{y}}_{k\|k-1}$	Innovation du filtre de Kalman à l'instant k
$\delta(\cdot)$	Mesure de Dirac
$\boldsymbol{x}_k^{(i)}$	$i^{\text{ème}}$ particule
$w_k^{(i)}$	Poids de la $i^{\text{ème}}$ particule
N_p	Nombre de particules
N_{eff}	Nombre efficace (mesure de dégénérescence)
M_i	$i^{\text{ème}}$ modèle représenté au sein d'une structure à modèles multiples

$m(k)$	Modèle suivi par le système à l'instant k
$m_i(k)$	Notation représentant l'évènement $m(k) = M_i$
$F_k^{m(k)}$	Matrice de transition du modèle $m(k)$
$Q_k^{m(k)}$	Matrice de covariance du bruit de modèle associée au modèle $m(k)$
$H_k^{m(k)}$	Matrice d'observation du modèle $m(k)$
$R_k^{m(k)}$	Matrice de covariance du bruit de mesure associée au modèle $m(k)$
$\Gamma(\cdot)$	Fonction Gamma
μ_k^i	Poids du $i^{\text{ème}}$ modèle à l'instant k
$\mu_{k\|k-1}^i$	Poids prédit du $i^{\text{ème}}$ modèle à l'instant k
$\mu_{k-1\|k}^{ij}$	Poids de mélange, probabilité de passer du $i^{\text{ème}}$ modèle au $j^{\text{ème}}$ de l'instant $k-1$ à k
$\hat{\boldsymbol{x}}_{k-1\|k-1}^i$	Estimation *a posteriori* de l'état associée au modèle i à l'instant $k-1$
$\hat{\boldsymbol{x}}_{k\|k-1}^i$	Etat estimé *a priori* à l'instant k par le $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples
$\hat{\boldsymbol{x}}_{k\|k}^i$	Etat estimé *a posteriori* à l'instant k par le $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples
$P_{k\|k-1}^i$	Matrice de covariance de l'erreur d'estimation *a priori* à l'instant k pour le $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples
$P_{k\|k}^i$	Matrice de covariance de l'erreur d'estimation *a posteriori* à l'instant k pour le $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples
$\boldsymbol{x}_{k-1\|k-1}^{0i}$	Etat à l'entrée du $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples après l'étape de mélange
$P_{k-1\|k-1}^{0i}$	Matrice de covariance de l'erreur d'estimation *a posteriori* du $i^{\text{ème}}$ filtre d'un algorithme à modèles multiples après l'étape de mélange
Π	Matrice de transition stockant les probabilités de transition entre les différents modèles
π_{ij}	Elément de la matrice de transition Π représentant la probabilité de passer de l'état i à l'état j entre deux instants consécutifs.
$\boldsymbol{X}$	Ensemble fini aléatoire
$\|\boldsymbol{X}\|$	Cardinal de l'ensemble fini aléatoire $\boldsymbol{X}$
$\mathcal{F}(\mathbb{X})$	Espace des sous-ensembles de l'espace $\mathbb{X}$
$\mathcal{F}_n(\mathbb{X})$	Espace des sous-ensembles de $\mathbb{X}$ contenant n éléments
P_D	Probabilité de détection
P_S	Probabilité de survie
$v(\cdot)$	Fonction d'intensité, aussi appelée densité d'hypothèse de probabilité (PHD)
$N(k)$	Nombre de cibles à l'instant k
$M(k)$	Nombre de mesures à l'instant k
$\mathbb{E}[\cdot]$	Espérance mathématique
$\mathbb{N}$	Ensemble des entiers naturels
$\mathbb{R}$	Ensemble des nombres réels
$\pi(\cdot)$	Densité de probabilité multi-cibles
$\rho(\cdot)$	Densité du nombre de cibles

X_k	Etat multi-cibles
Y_k	Etat multi-observations
$\mathcal{U}(\cdot)$	Densité de probabilité uniforme
$S_{k\|k-1}$	Ensemble fini aléatoire des cibles survivantes
Γ_k	Ensemble fini aléatoire associé aux cibles naissantes
$\pi_B(\cdot)$	Distribution des cibles naissantes
$\pi_c(\cdot)$	Distribution du nombre de fausses alarmes
$\hat{N}$	Estimation du nombre de cibles
C_n^k	Coefficient binomial, combinaison de k éléments parmi n
A_n^k	Nombre d'arrangements de k éléments parmi n

Chapitre 2

$KL(p,q)$	Divergence de Kullback-Leibler entre les densités p et q
$JD(p,q)$	Divergence de Jeffrey entre les densités p et q
$JD^{ij}(k)$	Divergence de Jeffrey entre les densités associées aux modèles i et j à l'instant k
$D(k)$	Matrice de dissimilarité à l'instant k composée des JDs entre les modèles considérés
M	Nombre de modèles d'état
T	Période d'échantillonnage / temps de rafraichissement
F^i	Matrice de transition du modèle i
H^i	Matrice d'observation du modèle i
Q^i	Matrice de covariance du bruit de modèle du modèle i
R^i	Matrice de covariance du bruit de mesure du modèle i
$\otimes$	Produit de Kronecker
$W(k)$	Matrice des JDs après application de l'algorithme de Dijkstra
$E(k)$	Matrice des JDs après application de l'algorithme de Dijkstra et d'une transformation non-linéaire
$\sigma^2_{ma_i}$	Variance du processus générateur de la $i^{\text{ème}}$ marche aléatoire

Chapitre 3

G_0	Distribution de base du processus de Dirichlet
α_0	Facteur d'échelle du processus de Dirichlet
$\mathcal{D}(\cdot)$	Distribution de Dirichlet
c_k	Indice de cluster
$\mathcal{W}(\cdot)$	Distribution de Wishart
$\mathcal{B}(\cdot)$	Distribution Beta
$\mathcal{G}(\cdot)$	Distribution Gamma
$\boldsymbol{U}_j$	Atome ou cluster associé aux réalisations d'un processus de Dirichlet
$\boldsymbol{\theta}_k$	Vecteur des variables latentes
$J(k)$	Nombre de cluster à l'instant k

$\boldsymbol{\theta}_i^*$	$i^{\text{ème}}$ valeur distincte parmi la séquence des variables $\boldsymbol{\theta}_{1:k}$
Γ_k	Matrice de précision (inverse de la matrice de covariance du bruit de modèle k)
γ_k	Paramètre de précision
I_{n_x}	Matrice identité de taille $n_x \times n_x$
z_k	Indice de famille de modèles à l'instant k

Chapitre 4

C	Nombre de classes
c	Indice de la classe de la cible
$r(c)$	Nombre de modèles considérés pour la $c^{\text{ème}}$ classe
Π^c	Matrice des probabilités de transition pour la classe c
l	Longueur de la cible
L_k	Mesure d'étalement distance à l'instant k
h_{taille}	Fonction de mesure de la taille d'une cible
ϕ	Angle entre la direction de la cible et la LOS radar-cible
σ_d^2	Variance du bruit de mesure sur l'axe distance
σ_{az}^2	Variance du bruit de mesure sur l'axe azimut
σ_{ed}^2	Variance du bruit de mesure de l'étalement distance
Δ_x	Distance radar-cible sur l'axe x
Δ_y	Distance radar-cible sur l'axe y
$\boldsymbol{X}_k$	Vecteur d'état étendu incluant la longueur de la cible
$\boldsymbol{Y}_k$	Vecteur d'observation étendu incluant la mesure d'étalement distance
m_k^c	Modèle d'évolution pour la classe c à l'instant k
M_i^c	$i^{\text{ème}}$ modèle parmi les modèles considérés dans la $c^{\text{ème}}$ classe
Π^c	Matrice des probabilités de transition pour la classe c
π_{ij}^c	Elément (i,j) de Π^c
$\boldsymbol{x}_k$	Vecteur d'état composé des paramètres cinématiques de la cible
$\boldsymbol{X}_k$	Vecteur d'état incluant la longueur de la cible
$\boldsymbol{v}_k^{m_k^c}$	Bruit de modèle associé au modèle m_k^c
$\boldsymbol{V}_k^{m_k^c}$	Bruit de modèle incluant la composante de longueur de cible associée au modèle m_k^c
$Q^{m_k^c}$	Matrice de covariance du bruit de modèle $\boldsymbol{v}_k^{m_k^c}$
$F^{m_k^c}$	Matrice de transition associée au modèle m_k^c
$\boldsymbol{Q}^{m_k^c}$	Matrice de covariance du bruit de modèle $\boldsymbol{V}_k^{m_k^c}$
$\boldsymbol{F}^{m_k^c}$	Matrice de transition incluant la composante sur la longueur de la cible associée au modèle m_k^c
$J_{k\|k-1}^c$	Nombre de gaussiennes à l'instant k pour la classe c
$X_{k\|k}^{(i,c)}$	Etat étendu estimé *a posteriori* à partir du $i^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe
$P_{k\|k}^{(i,c)}$	Matrice de covariance de l'erreur d'estimation *a posteriori* pour le $i^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe
$\mu_{k\|k}^{(i,)}$	Poids associé au $i^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe à l'instant k

$\boldsymbol{H}^{(i,c)}$	Matrice jacobienne de la fonction de mesure $\boldsymbol{h}$ évaluée en $\boldsymbol{X}_{k\|k-1}^{(i,c)}$
$\boldsymbol{S}^{(i,c)}$	Matrice de covariance de l'innovation associée au $i^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe
$\hat{l}_{k\|k}^{(i,c)}$	Estimation de la longueur de la cible par le $i^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe
$var(\cdot)$	Variance

Introduction

Le radar, de l'anglais *RAdio Detection And Ranging*, est un système électronique utilisé dans de nombreux contextes : en météorologie, pour la surveillance du trafic routier, pour le contrôle du trafic aérien civil ou militaire, etc. Dans le cadre des activités militaires, la fonction principale d'un radar aéroporté équipant un avion de combat ou de surveillance est d'assurer la détection à distance de cibles aériennes, terrestres ou maritimes et de suivre leurs évolutions au cours du temps. L'architecture des traitements associés peut se décomposer en deux phases spécifiques.
Dans un premier temps, une forme d'onde est émise par le radar. Elle est rétrodiffusée par les cibles, puis captée par l'antenne en réception. Les signaux reçus sont alors traités afin de maximiser la probabilité de détection des échos utiles et minimiser le taux de fausses alarmes. Ce traitement délivre alors des informations comme la distance séparant la cible du radar, l'azimut, la surface équivalente radar (SER) ou la fréquence Doppler.
Dans un second temps, ces informations sont exploitées. Selon les cas, elles permettent par exemple de déduire un profil distance/Doppler en vue d'une classification ou d'une identification. Pour le pistage, elles sont utilisées pour estimer les trajectoires des objets d'intéret, également appelées pistes. Il s'agit d'estimer à chaque instant les paramètres cinématiques de chaque cible (position, vitesse et si besoin accélération) à partir de mesures bruitées fournies par le radar. Ces paramètres constituent ce que l'on appelle l'état des cibles. Son estimation repose classiquement sur une approche récursive visant à exprimer la densité de probabilité *a posteriori* de l'état et ses moments d'ordre un et deux. Elle est tout d'abord prédite en tirant profit du modèle d'état qui décrit l'évolution de la cible au cours du temps ; puis, elle est mise à jour lorsqu'une nouvelle mesure est disponible. En pistage monocible, cette formulation récursive peut être par exemple mise en œuvre par filtrage de Kalman [Kal60] ou encore par filtrage particulaire [Dou01a]. Dans le cadre du pistage de plusieurs cibles, l'état d'une cible peut être estimé indépendamment des autres. Néanmoins comme les mesures disponibles peuvent être dues à différentes cibles voire du clutter, des algorithmes dits d'association de données sont utilisés pour définir des probabilités d'association des mesures aux cibles. Récemment, des algorithmes alternatifs de pistage multi-cibles ont été introduits par Mahler [Mah07]. Ils traitent les cibles comme un ensemble non ordonné et permettent de s'affranchir de certains problèmes inhérents aux algorithmes d'association. Mahler propose une extension du filtre de Bayes au cas multi-cibles en s'appuyant sur la théorie statistique des ensembles finis. Dans ce contexte, un état du filtre de Bayes multi-cibles est modélisé par un ensemble fini aléatoire (EFA) pour représenter la totalité des cibles présentes sur la zone d'intérêt. Par ailleurs, le filtre de Bayes multi-cibles ne pouvant pas être calculé explicitement, plusieurs approximations comme le filtre *probability hypothesis density* (PHD) [Mah03] et le filtre *cardinalized probability hypothesis density* (CPHD)

[Mah07] ont été proposées.

De manière générale, quel que soit le cadre d'étude, c'est-à-dire le pistage monocible ou multicibles, les techniques d'estimation récursives utilisées sont fondées sur la définition *a priori* de la représentation dans l'espace d'état (REE) du système. Elle est constituée du modèle d'état traduisant l'évolution de l'état au fil du temps et du modèle de mesure caractérisant la relation entre la mesure et l'état. La sélection de ces modèles n'est pas nécessairement triviale. D'une part, la REE pour laquelle l'utilisateur a opté peut être ou devenir inadaptée au comportement du système, ce qui a alors une influence sur les performances d'estimation de l'algorithme utilisé. D'autre part, cette REE du système dépend de paramètres qu'il est difficile de définir *a priori*. Pour pallier ces difficultés, les algorithmes à modèles multiples ont été introduits et trois générations ont été développées [Bar95, Li96]. Leur principe est d'associer en parallèle plusieurs filtres correspondant à différentes hypothèses d'évolution du système et de combiner les estimations de l'état à chaque instant en fonction de leur pertinence. Cependant, quelle que soit la génération d'algorithmes, il a été montré qu'utiliser un trop grand nombre de modèles en parallèle ne permet pas nécessairement d'améliorer les performances d'estimation [Bar95]. En général, deux à trois modèles sont associés. En poursuite de cibles, il s'agit de choisir des modèles de mouvement suffisamment dissimilaires les uns des autres pour couvrir un large panel de trajectoires [Bar93]. Pourtant, jusqu'à présent, peu d'études ont été menées sur la manière de réaliser ce choix *a priori*. Aussi, bien que les approches par modèles multiples soient très utilisées depuis le début des années 2000, plusieurs questions ouvertes subsistent : comment sélectionner les modèles ? Comment mesurer les similarités entre modèles ?
Pour contribuer aux réflexions menées sur ces sujets, nous traitons dans cette thèse de l'analyse des dissimilarités entre modèles d'état par une approche statistique. Nous étudions la pertinence d'une méthode fondée sur la divergence de Jeffrey (JD) qui est une version symétrique de la divergence de Kullback-Leibler (KL) [Kul51, Jef48]. Afin de mesurer la dissimilarité entre deux modèles d'état, nous proposons une expression récursive de la JD entre les distributions jointes des vecteurs d'état successifs associés aux deux modèles d'état considérés [Mag13]. Nous évaluons ensuite l'intérêt de cette étude en l'appliquant à un ensemble de plus de deux modèles d'état afin de définir une méthode de sélection de modèles applicable à la conception d'algorithmes à modèles multiples [Mag14, Mag15d].

Cependant, l'une des limites des approches d'estimation par modèles multiples est qu'elles restent conditionnées par la valeur des paramètres des modèles d'état et de mesure, notamment les matrices de covariance des bruits de modèle et de mesure. En pratique, ces paramètres sont fixés *a priori* lors de la conception de l'algorithme de pistage. Pourtant, ils peuvent évoluer et prendre un nombre indéfini de valeurs différentes au cours du temps.
Une alternative fondée sur l'utilisation de méthodes bayésiennes non-paramétriques a été récemment introduite [Car08]. Cette approche permet de relâcher toute hypothèse vis-à-vis du nombre de modèles et des distributions de leurs paramètres. Elle a l'avantage de pouvoir représenter des bruits multi-modaux sans avoir à spécifier *a priori* le nombre de modes et leurs caractéristiques. Il est alors possible d'estimer les paramètres caractérisant la REE à mesure que des observations sont disponibles. L'originalité de ces méthodes est de relâcher l'hypothèse que la distribution du bruit est connue et de la considérer comme aléatoire. Ainsi, elles associent des

distributions de probabilités à des distributions de probabilités [Car08, Fox10, Fox11b]. Cela revient à considérer que les paramètres du système peuvent permuter entre différentes valeurs qui réapparaissent au cours du temps sans limitation sur le nombre de valeurs possibles. Dans le cadre de cette thèse, nous traitons de ces techniques et proposons un modèle hiérarchique fondé sur l'utilisation des processus de Dirichlet. Notre contribution par rapport à l'état de l'art est de tirer profit des formes structurelles des matrices de covariance des bruits de modèles rencontrées en pistage de cibles pour n'estimer qu'un nombre réduit d'hyperparamètres [Mag15b]. Le problème d'estimation étant non-linéaire non-gaussien, l'inférence conjointe du vecteur d'état et des paramètres inconnus est réalisée par filtrage particulaire [Mag16a].

Enfin, afin de définir des modèles pertinents à utiliser au sein d'une structure à modèles multiples vis-à-vis du type de cibles pistées, il peut être avantageux de prendre en compte des informations sur chacune d'elles [Mei07, Cao13, Lan13, Sun14, Cao14]. C'est l'une des raisons pour lesquelles des approches effectuant le pistage conjointement à la classification ont été introduites. En anglais, on parle de *joint tracking and classification* (JTC). Dans cette thèse, nous définissons plusieurs classes différenciables selon la manœuvrabilité des cibles. Outre ses caractéristiques dynamiques, différents attributs de la cible comme sa forme ou sa fréquence Doppler peuvent être utilisés pour aider à la classification. En partant de l'*a priori* que la dynamique d'une cible est étroitement liée à sa taille, nous proposons d'exploiter la mesure d'étalement distance qui correspond à la projection de sa longueur sur l'axe radar-cible. Dans ce cas, comme la longueur de cible n'est pas directement mesurée, nous l'intégrons au vecteur d'état pour pouvoir l'estimer. Au sein de chaque classe de cibles, la trajectoire de la cible est représentée selon plusieurs modèles de mouvement adaptés à sa dynamique, ce qui nous amène à considérer des algorithmes multi-classes multi-modèles.
Dans un premier temps, nous proposons dans un contexte monocible de mener l'estimation de l'état au travers de deux approches qui diffèrent selon la manière de traiter le caractère non-linéaire du problème d'estimation. La première est fondée sur l'utilisation d'un algorithme IMM multi-classes alors que la seconde met en œuvre un filtre particulaire multi-classes multi-modèles. Dans un second temps, nous abordons l'extension de cette approche JTC au contexte de pistage multi-cibles en tirant profit du formalisme des EFAs. Un algorithme multi-cibles multi-classes fondé sur l'utilisation de la version multi-modèles du filtres PHD est alors étudié.

Ce manuscrit de thèse s'organise comme suit.

Les concepts fondamentaux pour le pistage de cibles et le formalisme des approches bayésiennes dans un contexte monocible sont introduits dans le chapitre 1. En particulier, nous détaillons le filtre de Bayes, le filtre de Kalman et le filtre particulaire. Puis, nous présentons les approches à modèles multiples. Enfin, le formalisme des EFAs est abordé et les filtres multi-cibles sont présentés.

Dans le chapitre 2, nous établissons une mesure de dissimilarité fondée sur la divergence de Jeffrey entre deux densités de probabilités associées à deux modèles d'état différents. Celle-ci est appliquée à la comparaison de modèles de mouvement. Elle est ensuite utilisée pour décomposer un ensemble de plusieurs modèles d'état en différents sous-ensembles. Cette étude est enfin mise à profit pour proposer une méthode de sélection *a priori* des modèles constituant des algorithmes à modèles multiples.

Le chapitre 3 est consacré à l'utilisation de modèles bayésiens non-paramétriques pour estimer les statistiques du bruit de modèle. Dans un premier temps, nous rappelons la théorie des processus de Dirichlet. Puis, nous introduisons la modélisation BNP pour l'estimation de toute la matrice de covariance du bruit de modèle. Dans un second temps, nous proposons un modèle hiérarchique tirant profit des formes structurelles des matrices de covariance associées aux modèles de mouvement pour n'estimer qu'un nombre réduit d'hyperparamètres. Nous détaillons ensuite les différentes étapes de calcul de la distribution d'importance optimale utilisée pour réaliser par filtrage particulaire l'inférence conjointe des paramètres cinématiques de la cible et des hyperparamètres. Enfin, une étude comparative avec différents algorithmes de pistage, notamment fondés sur des filtrages particulaires auxiliaires, est menée.

Enfin, la mise en œuvre de nouvelles méthodes JTC fondées sur l'utilisation de la longueur des cibles est détaillée dans le chapitre 4. Nous présentons tout d'abord dans un contexte monocible un algorithme IMM multi-classes puis un algorithme particulaire multi-classes multi-modèles. Nous étudions ensuite une extension de l'approche JTC au cas où les paramètres du modèle d'état sont inconnus en mettant à profit les modèles BNP. Finalement, nous détaillons l'extension au cas multi-cibles. Dans tous les cas, nous illustrons les avantages et inconvénients des approches proposées au travers de résultats de simulation.

En outre, ce mémoire est complété par quatre annexes.
L'annexe A expose succinctement les méthodes d'association de données qui sont généralement combinées à un algorithme de filtrage bayésien afin de déterminer à chaque instant la mesure utilisée pour mettre à jour l'estimateur associé à la cible.
L'annexe B porte sur le développement de l'expression de la distribution d'importance optimale lorsque la distribution de toute la matrice de covariance est modélisée par un processus de Dirichlet et l'estimation réalisée par filtrage particulaire.
L'annexe C présente le filtre particulaire auxiliaire utilisé à titre de comparaison dans le chapitre 3.
Enfin, l'annexe D introduit des compléments mathématiques utilisés dans le manuscrit.

Chapitre 1

Etat de l'art sur le pistage de cibles

Sommaire

1.1 Introduction **6**

1.2 Estimation bayésienne pour le pistage monocible **6**

1.2.1 Représentation dans l'espace d'état 7

1.2.2 Filtre de Bayes 8

1.2.3 Filtre de Kalman 8

1.2.4 Filtre particulaire 11

1.3 Pistage de cibles manœuvrantes : algorithmes à modèles multiples **18**

1.3.1 Estimateur à modèles multiples dynamique optimal 19

1.3.2 Algorithme à modèles multiples interactifs (IMM) 22

1.3.3 Problématiques liées à l'utilisation des modèles multiples 25

1.4 Ensembles finis aléatoires pour le pistage multi-cibles **26**

1.4.1 Définitions, notations et abréviations 26

1.4.2 Distributions de probabilités multi-cibles 28

1.4.3 Représentation multi-cibles du système 30

1.5 Filtrage et estimation bayésienne multi-cibles **32**

1.5.1 Filtre de Bayes multi-cibles 33

1.5.2 Filtre *probability hypothesis density* (PHD) 33

1.5.3 Filtre *cardinalized probability hypothesis density* (CPHD) 35

1.6 Evaluation de performances pour les algorithmes de pistage multi-cibles **37**

1.6.1 Estimateurs d'états multi-cibles 37

1.6.2 Distance OSPA 38

1.7 Conclusions **39**

1.1 Introduction

Le pistage d'une ou de plusieurs cibles s'effectue généralement en utilisant des approches d'estimation bayésiennes. Dans un contexte monocible, le filtre de Kalman et le filtre particulaire sont principalement appliqués. Dans un contexte industriel, le filtre de Kalman étant moins coûteux calculatoirement que le filtre particulaire, il est plus souvent utilisé. Ces différentes méthodes sont fondées sur la REE du système qui est caractérisée par une équation dite d'état modélisant l'évolution du système au cours du temps et une équation de mesure caractérisant le lien entre la mesure et l'état. Dans le cas d'une cible manœuvrante, un modèle englobant caractérisant différents types de trajectoires peut être utilisé. Cependant, une modélisation unique ne permet pas nécessairement d'obtenir des performances d'estimation satisfaisantes. C'est pour cette raison que les algorithmes à modèles multiples, qui permettent à chaque instant de selectionner le modèle de mouvement le plus pertinent tout en fournissant une estimation de l'état, ont été introduits.

Dans un contexte multi-cibles, deux philosophies peuvent être envisagées. D'une part, les approches précédentes peuvent s'appliquer en menant le pistage de chacune des cibles indépendamment les unes des autres. Cependant, différents problèmes se posent notamment lorsque les cibles sont proches et/ou les mesures sont trop imprécises. D'autre part, une alternative aux filtres de Kalman et particulaire fondée sur la théorie des EFAs a été récemment introduite par Mahler [Mah01]. Elle consiste à estimer conjointement le nombre de cibles et leurs états au cours du temps. Ainsi, une extension du filtre de Bayes, qui décrit l'évolution de la loi *a posteriori*, a été développée dans le cadre du pistage multi-cibles. Comme ce filtre n'est pas directement implémentable, des approximations de ce dernier ont été développées. Il s'agit notamment des filtres *probability hypothesis density* (PHD) [Mah03] et *cardinalized probability hypothesis density* (CPHD) [Mah07].

L'objet de ce premier chapitre est de préciser les concepts cités précédemment. Aussi, nous introduisons les méthodes d'estimation bayésienne et les modèles sur lesquels elles reposent. En particulier, les filtres de Bayes, de Kalman et particulaire sont présentés. Ensuite, les approches à modèles multiples et les problématiques inhérentes à leur utilisation sont exposées. Enfin, le formalisme relatif au pistage de plusieurs cibles selon la théorie des EFAs ainsi que les différents filtres bayésiens associés sont détaillés.

1.2 Estimation bayésienne pour le pistage monocible

D'un point de vue général, il s'agit de définir des estimateurs des variables d'un système dynamique sujet à des perturbations à partir d'observations partielles et bruitées. Dans la suite de ce mémoire, nous considérons que les observations sont disponibles jusqu'à l'instant courant ; on parle dans ce cas de filtrage.
Dans un premier temps, nous présentons la REE, puis, le formalisme du filtrage bayésien est décrit. Il consiste à estimer récursivement la densité de probabilité *a posteriori* de l'état d'un système, c'est-à-dire la distribution de l'état conditionnellement aux observations jusqu'à l'instant courant. Enfin, nous décrivons le filtre de Bayes ainsi que le filtre de Kalman et le filtre particulaire.

1.2.1 Représentation dans l'espace d'état

Dans la suite, nous considérons la REE à temps discret. Elle est caractérisée à l'instant k par un vecteur d'état noté $\boldsymbol{x}_k$, de dimension n_x, contenant les variables d'état à estimer et qui prend ses valeurs dans l'espace d'état $\mathbb{X} \subseteq \mathbb{R}^{n_x}$. L'état de la cible est indirectement observé à partir d'une mesure bruitée et possiblement incomplète. Cette mesure $\boldsymbol{y}_k$, de dimension n_y, prend ses valeurs dans l'espace des observations $\mathbb{Y} \subseteq \mathbb{R}^{n_y}$.

L'objectif est alors d'estimer récursivement l'état de la cible $\boldsymbol{x}_k$ à partir de la séquence de mesures reçues jusqu'à l'instant courant, notée $\boldsymbol{y}_{1:k}$.

Modélisation de l'état du système

L'évolution du vecteur d'état au cours du temps est décrite à partir d'un modèle de transition markovien comme suit :

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}, \boldsymbol{v}_k) \tag{1.1}$$

avec f_k la fonction de transition et $\boldsymbol{v}_k$ le bruit de modèle, aussi appelé bruit d'état.

On peut alors de façon équivalente représenter l'évolution du vecteur d'état selon une distribution de transition de Markov notée $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$.

Modélisation de la relation entre l'état et la mesure

On modélise les mesures à partir de l'équation d'observation suivante :

$$\boldsymbol{y}_k = h_k(\boldsymbol{x}_k, \boldsymbol{w}_k) \tag{1.2}$$

avec h_k la fonction de mesure et $\boldsymbol{w}_k$ le bruit de mesure.

On peut alors représenter la relation entre le vecteur d'état $\boldsymbol{x}_k$ et le vecteur de mesure $\boldsymbol{y}_k$ à partir de la fonction de vraisemblance $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ qui représente la densité de probabilité qu'à l'instant k, l'état $\boldsymbol{x}_k$ génère la mesure $\boldsymbol{y}_k$. De plus, on suppose l'indépendance temporelle entre les mesures reçues conditionnellement à l'état. Ainsi, la vraisemblance peut se décomposer comme suit :

$$p(\boldsymbol{y}_{1:k}|\boldsymbol{x}_{0:k}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{y}_{k-1}|\boldsymbol{x}_{k-1})...p(\boldsymbol{y}_1|\boldsymbol{x}_1) \tag{1.3}$$

Estimation bayésienne

Dans le cadre bayésien, à partir de la REE définie par (1.1) et (1.2), il s'agit de calculer la densité de probabilité *a posteriori* $p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$ ou l'une des marginales possibles telle que $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. Ces densités de probabilité contiennent toute l'information disponible sur le vecteur d'état à l'instant courant.

Par ailleurs, afin d'extraire une estimation *a posteriori* de l'état $\boldsymbol{x}_k$ à partir de sa densité de probabilité, différents estimateurs optimaux peuvent être calculés selon le critère d'erreur considéré. Les plus fréquents sont l'estimateur *expected a posteriori* (EAP) qui minimise l'erreur quadratique moyenne (EQM) et l'estimateur *maximum a posteriori* (MAP) qui minimise le coût binaire moyen. Ils sont définis respectivement comme suit :

$$\hat{\boldsymbol{x}}_k^{EAP} = \int_{\mathbb{X}} \boldsymbol{x}_k p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) d\boldsymbol{x}_k \tag{1.4}$$

$$\hat{\boldsymbol{x}}_k^{MAP} = argsup_{\boldsymbol{x}_k}\{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})\} \tag{1.5}$$

Théoriquement, la densité *a posteriori* $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ peut être calculée récursivement à partir de deux équations formant le filtre de Bayes.

1.2.2 Filtre de Bayes

La résolution du problème de filtrage et plus particulièrement le calcul de la densité de probabilité *a posteriori* $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ peut se faire de manière récursive à partir du filtre de Bayes. Ce dernier opère en deux étapes : une étape de prédiction fondée sur le modèle d'état et une étape de mise à jour qui tire profit de la mesure reçue à l'instant courant.
A chaque instant, la densité de probabilité prédite à l'instant $k-1$, notée $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$, est calculée à partir de la densité de probabilité *a posteriori* précédente en utilisant l'équation de Chapman-Kolmogorov :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \int_{\mathbb{X}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})d\boldsymbol{x}_{k-1} \tag{1.6}$$

L'étape de mise à jour utilise alors la règle de Bayes et les indépendances conditionnelles pour tirer profit de l'information apportée par l'observation à l'instant k comme suit :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) &= p(\boldsymbol{x}_k|\boldsymbol{y}_k, \boldsymbol{y}_{1:k-1}) \\ &= \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{y}_{1:k-1})p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})} \\ &= \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})} \end{aligned} \tag{1.7}$$

Ainsi, si la densité de probabilité de l'état à l'instant initial $p(\boldsymbol{x}_0)$ est connue, le filtre de Bayes permet de calculer récursivement la densité de probabilité de l'état du système à chaque instant à partir de (1.6) et (1.7). Sa mise en œuvre nécessite néanmoins de calculer des intégrales multidimensionnelles. Or, cela n'est pas toujours réalisable ou alors sous certaines conditions. Par exemple, dans le cas où la REE est linéaire et gaussienne et où $p(\boldsymbol{x}_0)$ est gaussienne, la densité de probabilité *a posteriori* est gaussienne. On dispose alors d'expressions exactes au travers de l'algorithme du filtre de Kalman. Dans les autres cas, il est nécessaire de recourir à des approximations soit par des méthodes analytiques faisant intervenir une simplification du modèle, soit par des méthodes numériques fondées sur une discrétisation de l'espace d'état ou enfin par des méthodes de Monte Carlo.
Dans la suite, nous présentons le filtre de Kalman et ses variantes ainsi que le filtre particulaire et analysons leurs avantages et inconvénients.

1.2.3 Filtre de Kalman

Le filtre de Kalman propage à chaque instant les moments statistiques d'ordre un et deux de la distribution de probabilité *a posteriori* $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$.
La démonstration des expressions du filtre de Kalman, telles qu'introduites dans la version originale [Kal60], reposent sur la minimisation de l'EQM associée à l'estimation du vecteur d'état. Dans la suite, nous proposons de revenir sur les expressions des équations du filtre de

Kalman en partant d'une formulation bayésienne [Ho64].
Le filtre de Kalman constitue l'approche exacte de la récursion bayésienne dans le cas linéaire et gaussien. En effet, lorsque les fonctions f_k et h_k définies en (1.1) et (1.2) respectivement sont linéaires et que les bruits de modèle et de mesure $\boldsymbol{v}_k$ et $\boldsymbol{w}_k$ sont additifs et gaussiens, la distribution *a posteriori* est gaussienne et sa moyenne et sa covariance peuvent être calculées récursivement. La REE à temps discret du système est alors caractérisée par les équations suivantes :

$$\boldsymbol{x}_k = F_k \boldsymbol{x}_{k-1} + \boldsymbol{v}_k \tag{1.8}$$

$$\boldsymbol{y}_k = H_k \boldsymbol{x}_k + \boldsymbol{w}_k \tag{1.9}$$

où F_k désigne la matrice de transition de l'instant $k-1$ à l'instant k et H_k la matrice d'observation. Elles sont de taille $n_x \times n_x$ et $n_y \times n_x$ respectivement.
De plus, le bruit de modèle $\boldsymbol{v}_k$ et le bruit de mesure $\boldsymbol{w}_k$ sont blancs, gaussiens et indépendants. Ils satisfaisont :

$$\boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{v}_k; 0, Q_k) \tag{1.10}$$

$$\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{w}_k; 0, R_k) \tag{1.11}$$

Ainsi, on définit la densité de transition et la vraisemblance de la mesure comme suit :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k; F_k \boldsymbol{x}_{k-1}, Q_k) \tag{1.12}$$

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k) \tag{1.13}$$

Soient $\hat{\boldsymbol{x}}_{k|k-1} = \mathbb{E}[\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}]$ la prédiction de l'état $\boldsymbol{x}_k$ et $\hat{\boldsymbol{x}}_{k|k} = \mathbb{E}[\boldsymbol{x}_k|\boldsymbol{y}_{1:k}]$ l'estimation de l'état *a posteriori*, c'est-à-dire celle qui prend en compte l'information apportée par la mesure. Alors, étant donné (1.12) et (1.13), les étapes du filtre de Kalman sont les suivantes :

Etape de prédiction

Supposons qu'à l'instant $k-1$, la densité *a posteriori* est définie comme suit :

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}) = \mathcal{N}(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1|k-1}, P_{k-1|k-1}) \tag{1.14}$$

avec $P_{k-1|k-1}$ la matrice de covariance de la densité de probabilité *a posteriori* de $\boldsymbol{x}_{k-1}$ qui est également la matrice de covariance de l'erreur d'estimation *a posteriori* à l'instant $k-1$.

La densité de probabilité *prédite* à l'instant k s'exprime alors comme suit :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}, P_{k|k-1}) \tag{1.15}$$

avec :

$$\hat{\boldsymbol{x}}_{k|k-1} = F_k \hat{\boldsymbol{x}}_{k-1|k-1} \tag{1.16}$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{1.17}$$

Etape de mise à jour

La mise à jour de la densité de probabilité de l'état du système peut être effectuée selon la règle de Bayes énoncée en (1.7).

En considérant tout d'abord la densité de probabilité de la mesure conditionnellement à l'état :

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) \sim \mathcal{N}(\boldsymbol{y}_k; H_k\boldsymbol{x}_k, R_k) \tag{1.18}$$

Puis, en intégrant (1.15) et (1.18) dans (1.7), il vient :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \propto \mathcal{N}(\boldsymbol{y}_k; H_k\boldsymbol{x}_k, R_k)\mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}, P_{k|k-1}) \tag{1.19}$$

En remplaçant les gaussiennes par leurs expressions analytiques puis en développant [1] les formes quadratiques à l'intérieur des exponentielles, on peut montrer en appliquant le lemme d'inversion matricielle que la densité de probabilité de l'état *a posteriori* à l'instant k est gaussienne et satisfait :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k}, P_{k|k}) \tag{1.20}$$

avec :

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + K_k(\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1}) \tag{1.21}$$

$$P_{k|k} = [I_{n_x} - K_kH_k]P_{k|k-1} \tag{1.22}$$

$$K_k = P_{k|k-1}H_k^TS_k^{-1} \tag{1.23}$$

La matrice K_k est par définition le gain de Kalman et la grandeur $\tilde{\boldsymbol{y}}_{k|k-1} = (\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1})$ est appelée innovation et a pour matrice de covariance S_k. Elle est dénommée ainsi car elle représente l'information nouvelle apportée par la mesure sur $\boldsymbol{x}_k$ par rapport à la prédiction.

La figure 1.1 résume les étapes d'un cycle du filtre de Kalman.

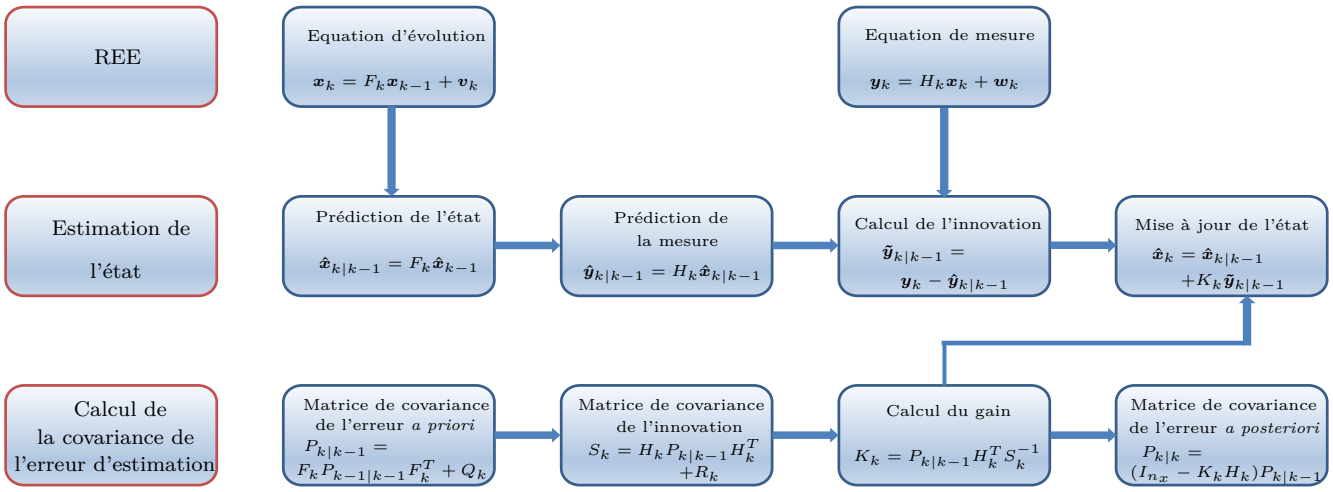

FIGURE 1.1 – Cycle du filtre de Kalman

1. Pour plus de détails sur les étapes de calcul à suivre, le lecteur peut se reporter à l'annexe D.1.

Finalement, notons que sous les hypothèses de gaussianité et de linéarité des équations de modèle et de mesure, le filtre de Kalman est le filtre de Bayes optimal. Dans ce cas, quel que soit l'instant considéré, les densités de probabilité *prédite* et *a posteriori* sont gaussiennes.

Dans la suite, nous discutons de certaines variantes du filtre de Kalman couramment utilisées lorsque les REEs ne sont pas linéaires et gaussiennes.

Si les systèmes sont faiblement non-linéaires, les équations d'état et de mesure peuvent être linéarisées à partir d'un développement de Taylor des fonctions d'évolution et de mesure définies en (1.1) et (1.2) respectivement. Cela conduit au filtre de Kalman étendu (FKE) pour un développement au premier ordre [Ath68]. Sa simplicité de mise en œuvre et sa faible complexité calculatoire le rendent très populaire. Pour des non-linéarités plus importantes, comme l'approximation au premier ordre peut ne pas suffire, le FKE est susceptible de diverger. Il est alors possible d'utiliser un FKE du second ordre [Rot11]. Une alternative à la linéarisation consiste à utiliser un filtre par sigma point du type Kalman sans parfum (FKSP) (unscented Kalman filter -UKF- en anglais) [Wan01], un filtre de Kalman par différence centrale (CDKF) ou encore les versions quadrature ou cubature [Ara09, Qui00].

Si les systèmes sont linéaires mais non gaussiens, un filtre du type somme de gaussiennes (FSG) est préféré [Als72]. Ce filtre est une extension de l'approximation gaussienne du filtre de Kalman étendu à une somme pondérée de gaussiennes visant à approcher la densité de probabilité *a posteriori* de l'état. Notons qu'il nécessite d'ajuster le nombre de gaussiennes.

Dans le cas d'hypothèses fortement non-linéaires et/ou non gaussiennes, on opte généralement pour des approches différentes telles que le filtre particulaire présenté ci-après.

1.2.4 Filtre particulaire

1.2.4.1 Intégration de Monte-Carlo

Avant de présenter les principes fondateurs du filtrage particulaire, nous fournissons dans un premier temps les connaissances théoriques nécessaires à la compréhension de cet algorithme.
Le principe des méthodes de Monte Carlo est d'utiliser des échantillons obtenus par simulation pour approcher une distribution de probabilité qu'il n'est pas possible de calculer analytiquement. Cet échantillonnage permet par exemple d'approcher les intégrales de grande dimension qui interviennent dans le filtre de Bayes.
De façon plus générale, si $p(\boldsymbol{x})$ désigne la densité de probabilité de $\boldsymbol{x}$, l'objectif d'une méthode de Monte Carlo est d'approcher une intégrale du type :

$$I(f) = E_p[f(\boldsymbol{x})] = \int_{\mathbb{X}} f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \tag{1.24}$$

pour toute fonction $f(\cdot)$ continue et bornée sur $\mathbb{X}$. Selon les formes de $p(\boldsymbol{x})$ et de la fonction $f(\cdot)$, le calcul analytique de cette intégrale n'est pas toujours envisageable. Les méthodes de Monte Carlo permettent alors de mettre en œuvre une méthode d'approximation qui est indépendante de la dimension de l'intégrale. Le principe est d'utiliser des réalisations indépendantes d'un vecteur aléatoire distribué selon la distribution $p(\boldsymbol{x})$. Ces réalisations, notées $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_p)}$, permettent

d'écrire :

$$I_{N_p}(f) = \frac{1}{N_p} \sum_{n=1}^{N_p} f(\boldsymbol{x}^{(n)}) \tag{1.25}$$

La loi des grands nombres, qui énonce que la moyenne de différentes variables aléatoires d'espérance commune et de variance finie converge quasi sûrement vers leur espérance commune, assure la convergence de $I_{N_p}(f)$ vers $I(f)$, soit :

$$Pr\left[\lim_{N_p \to +\infty} I_{N_p}(f) = I(f)\right] = 1 \tag{1.26}$$

Etant donné (1.24) et (1.25), l'estimateur obtenu est sans biais. De plus, si la variance de $f(\boldsymbol{x})$ par rapport à $p(\boldsymbol{x})$ notée σ_f^2 est finie, alors le théorème de la limite centrale permet de garantir que l'erreur d'estimation converge en distribution, soit :

$$\sqrt{N_p}(I_{N_p}(f) - I(f)) \underset{N_p \to +\infty}{\Longrightarrow} \mathcal{N}(0, \sigma_f^2) \tag{1.27}$$

où $\Longrightarrow$ représente la convergence en distribution.

A la différence des méthodes d'intégration numérique, la convergence de l'estimation de Monte-Carlo est indépendante de la dimension de l'espace d'état sur lequel on intègre. Cette propriété est l'atout principal des méthodes de Monte-Carlo qui a contribué à les rendre attractives.
Par ailleurs, l'approximation de Monte-Carlo se révèle pertinente dès lors que l'on sait générer des échantillons selon la loi $p(\cdot)$. Malheureusement, cela n'est pas toujours possible. Il existe néanmoins plusieurs approches permettant de passer outre cette difficulté. Certaines utilisent une densité arbitraire $q(\cdot)$, appelée densité instrumentale ou densité d'importance, dont le support contient au minimum celui de $p(\cdot)$ et selon laquelle il est facile de simuler des échantillons. Parmi ces méthodes, on compte entre autres l'échantillonnage d'acceptation-rejet [Neu51]. Ces méthodes étant trop coûteuses calculatoirement, la méthode d'échantillonnage d'importance que nous présentons dans la suite est souvent préférée.

1.2.4.2 Echantillonnage d'importance

Soient $p(\cdot)$ et $q(\cdot)$ une densité de probabilité dite cible qu'il n'est pas possible d'échantillonner et une densité d'importance respectivement. Les échantillons tirés selon la densité d'importance sont associés à des poids afin de corriger la différence entre la distribution d'importance $q(\cdot)$ et la distribution cible $p(\cdot)$. Cette pondération de l'ensemble des échantillons rend l'estimation de Monte-Carlo possible du moment que la densité d'importance a un support contenant celui de $p(\cdot)$. En effet, dans ce cas, l'intégrale (1.24) peut s'exprimer comme suit :

$$I(f) = \int_{\mathbb{X}} f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{X}} f(\boldsymbol{x})\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}q(\boldsymbol{x})d\boldsymbol{x} = E_q[f(\boldsymbol{x})\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}] \tag{1.28}$$

où $E_q[\cdot]$ désigne l'espérance mathématique sur la base de $q(\cdot)$.
Une estimation de Monte-Carlo consiste alors à générer N_p échantillons indépendants $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_p)}\}$

selon la densité d'importance $q(\cdot)$ et à former la somme pondérée suivante :

$$I_{N_p}(f) = \frac{1}{N_p} \sum_{i=1}^{N_p} f(\boldsymbol{x}^{(i)}) \tilde{w}^{(i)} \tag{1.29}$$

où $\{\tilde{w}^{(i)}\}_{i=1,\ldots,N_p}$ sont les poids d'importance. Ils satisfont :

$$\tilde{w}^{(i)} = \frac{p(\boldsymbol{x}^{(i)})}{q(\boldsymbol{x}^{(i)})} \tag{1.30}$$

Dans les cas où le facteur de normalisation de la densité cible est inconnu, une solution est de normaliser les poids. Il vient alors :

$$I_{N_p}(f) = \sum_{i=1}^{N_p} f(\boldsymbol{x}^{(i)}) w^{(i)} \tag{1.31}$$

où les poids normalisés $\{w^{(i)}\}_{i=1,\ldots,N_p}$ satisfont :

$$w^{(i)} = \frac{\tilde{w}^{(i)}}{\sum_{j=1}^{N_p} \tilde{w}^{(j)}} \tag{1.32}$$

En tirant profit des propriétés de la mesure de Dirac, (1.31) peut être ré-écrit comme suit :

$$I_{N_p}(f) = \int_{\mathbb{X}} f(\boldsymbol{x}) \sum_{i=1}^{N_p} w^{(i)} \delta(\boldsymbol{x} - \boldsymbol{x}^{(i)}) d\boldsymbol{x} \tag{1.33}$$

La densité cible $p(\cdot)$ peut alors être approchée à partir des échantillons tirées selon la loi d'importance $q(\cdot)$ comme suit :

$$p(\boldsymbol{x}) \approx \sum_{i=1}^{N_p} w^{(i)} \delta(\boldsymbol{x} - \boldsymbol{x}^{(i)}) \tag{1.34}$$

On aboutit donc à une approximation pondérée discrète de la densité $p(\cdot)$.

La mise en œuvre de l'échantillonnage d'importance à chaque instant pour calculer les intégrales du filtre de Bayes constitue la méthode d'échantillonnage d'importance séquentiel, *sequential importance sampling* (SIS) en anglais.
Dans la suite, nous présentons les étapes de fonctionnement du filtrage particulaire qui repose en particulier sur l'approche SIS.

1.2.4.3 Mise en œuvre du filtrage particulaire

Nommé *bootstrap filter* ou encore *condensation filter*, le filtre particulaire est fondé sur l'utilisation de méthodes de Monte-Carlo séquentielles (SMC) permettant d'approcher le filtre de Bayes. L'intérêt de telles méthodes est l'absence de contrainte quant aux conditions de gaussianité et de linéarité sur les modèles dynamique et de mesure.

Le principe du filtrage particulaire est d'estimer la densité de probabilité *a posteriori* notée $p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$ par un ensemble de particules afin de former une approximation Monte Carlo discrète de la distribution *a posteriori*.
On forme ainsi des couples du type $\{\boldsymbol{x}_{0:k}^{(i)}, w_k^{(i)}\}_{i=1}^{N_p}$ de telle manière que les poids soient normalisés, i.e. $\sum_{i=1}^{N_p} w_k^{(i)} = 1$. La distribution de probabilité *a posteriori* à l'instant k prend alors la forme suivante :

$$p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k}) \approx \sum_{i=1}^{N_p} w_k^{(i)} \delta(\boldsymbol{x}_{0:k} - \boldsymbol{x}_{0:k}^{(i)}) \tag{1.35}$$

Si les points support $\boldsymbol{x}_{0:k}^{(i)}$ sont générés selon une distribution d'importance $q(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$, les poids définis en (1.35) s'expriment comme suit :

$$w_k^{(i)} \propto \frac{p(\boldsymbol{x}_{0:k}^{(i)}|\boldsymbol{y}_{1:k})}{q(\boldsymbol{x}_{0:k}^{(i)}|\boldsymbol{y}_{1:k})} \tag{1.36}$$

Il est alors souhaitable de ne pas ré-échantillonner les particules entre l'instant initial et l'instant courant mais simplement de simuler les particules courantes $\boldsymbol{x}_k^{(i)}$ selon $q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k})$. Pour calculer les poids récursivement, il s'agit d'exprimer récursivement $p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$ à partir de $p(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1})$, en considérant la factorisation suivante de la densité d'importance :

$$q(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k}) = q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{1:k}) q(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1}) \tag{1.37}$$

Sans la factorisation (1.37), il serait nécessaire d'échantillonner tout l'état $\boldsymbol{x}_{0:k}$ à chaque instant afin d'avoir des échantillons distribués selon la loi $q(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$. Or, cela deviendrait problématique au regard des contraintes liées aux applications en temps réel. La factorisation (1.37) permet donc de générer séquentiellement N_p vecteurs d'état à chaque instant puis de les concaténer avec l'ensemble existant, soit $\boldsymbol{x}_{0:k}^{(i)} = \{\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{x}_k^{(i)}\}$. Cette factorisation a aussi pour avantage de simplifier la forme des poids d'importance. On peut montrer que l'équation de mise à jour satisfait :

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k}^{(i)}, \boldsymbol{y}_{1:k-1}) p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k})} \tag{1.38}$$

Dans la mesure où les observations sont supposées indépendantes conditionnellement au processus d'état et que leurs distributions ne dépendent que de l'état courant, il vient :

$$p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k}^{(i)}, \boldsymbol{y}_{1:k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) \tag{1.39}$$

Dans ce cas, on peut écrire :

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{q(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k})} \tag{1.40}$$

Une mise en œuvre de l'algorithme SIS, qui approche la densité marginale *a posteriori*, peut alors être réalisée de la manière suivante :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{i=1}^{N_p} w_k^{(i)} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}) \tag{1.41}$$

où les poids d'importance $\{w_k^{(i)}\}_{i=1,\ldots,N_p}$ sont normalisés.
Selon la loi des grands nombres, lorsque $N_p \to +\infty$, la densité de probabilité $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ est approchée avec exactitude.
Finalement, le filtrage à partir de l'algorithme SIS, qui vise à propager récursivement les poids d'importance et les points supports à chaque instant, constitue la brique de base du filtre particulaire. Cependant, il est sujet au phénomène de dégénérescence, qui se traduit par une augmentation importante de la variance des poids au cours du temps.

1.2.4.4 Problème de dégénéresence

Dans [Kon94], Kong *et al.* ont notamment montré que lorsque la densité d'importance $q(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$ est différente de la densité cible $p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$, la variance des poids ne peut qu'augmenter. En effet, après un certain nombre d'itérations, toutes les particules sauf une ont des poids négligeables. Ainsi, l'algorithme met à jour un grand nombre de particules dont la contribution à l'approximation de $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ est quasi nulle. Il n'est pas possible d'empêcher la dégénérescence d'un filtre SIS, mais il existe une mesure permettant de la caractériser. C'est le nombre effectif qui représente le nombre équivalent de particules distribuées directement selon la cible et conduisant à la même variance de l'estimateur construit par échantillonnage d'importance. Il est défini de la manière suivante :

$$N_{eff} = \frac{N_p}{1 + var[w_k^{(*,i)}]} \tag{1.42}$$

où $w_k^{(*,i)}$ correspondent aux poids définis en (1.38) après normalisation et $var[w_k^{(*,i)}]$ désigne la variance des poids calculée par rapport à la distribution $q(\cdot|\boldsymbol{y}_{1:k})$.
Comme on ne peut pas calculer explicitement cette valeur, elle est approchée classiquement par :

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_p}(w_k^{(i)})^2} \tag{1.43}$$

Par définition, $1 \leq N_{eff} \leq N_p$. Une valeur faible de N_{eff} traduit une forte dégénérescence.
Afin de limiter cette dégénérescence, il existe deux facteurs sur lesquels l'utilisateur peut agir. D'une part, une étape de ré-échantillonnage, permettant de conserver un ensemble non dégénéré de particules, peut être considérée. D'autre part, il s'agit de réaliser un choix pertinent de la densité d'importance $q(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{0:k})$ de façon à ce qu'elle soit aussi proche que possible de la densité cible $p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{0:k})$.

1.2.4.5 Ré-échantillonnage

L'étape de ré-échantillonnage consiste, lorsque N_{eff} est plus petit qu'un seuil prédéfini, à éliminer les particules dont les poids sont les plus faibles et à concentrer un nouvel ensemble de particules $\{\boldsymbol{x}_k^{(*,i)}\}_{i=1}^{N_p}$ là où se situent les particules de poids les plus élevés. Ainsi, il vient :

$$\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1,\ldots,N_p} \to \{\boldsymbol{x}_k^{(*,i)}, \frac{1}{N_p}\}_{i=1,\ldots,N_p} \tag{1.44}$$

avec :

$$Pr\left[\boldsymbol{x}_k^{(*,i)} = \boldsymbol{x}_k^{(j)}\right] = w_k^{(j)} \tag{1.45}$$

Cela revient à effectuer un tirage selon l'approximation discrète de la loi *a posteriori* qui est une loi catégorielle.

Notons qu'il existe un autre critère que la mesure de dégénérescence pour appliquer l'étape de ré-échantillonnage. Ce dernier, proposé par Pham [PHA01], est fondé sur le calcul de l'entropie des poids d'importance.

Bien que le ré-échantillonnage limite le phénomène de dégénérescence, il introduit un effet indésirable. La duplication des particules de poids fort a tendance à appauvrir le système de particules, ce qui peut être problématique lorsque la dynamique du bruit d'état est faible. Les particules ont dans ce cas tendance à se concentrer dans certaines régions provoquant un perte de diversité. Cela augmente le risque qu'aucune des particules ne soit cohérente avec les mesures suivantes, provoquant ainsi la divergence du filtre. Afin de limiter ce phénomène, le choix de la densité d'importance est crucial.

1.2.4.6 Choix de la densité d'importance

Le choix de la distribution d'importance $q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_k)$ est l'un des problèmes les plus critiques associé au filtrage particulaire. En effet, pour un ensemble fini de particules, les performances de l'algorithme sont étroitement liées à ce choix.
En pratique, la distribution d'importance est souvent choisie comme étant la distribution de transition *a priori*, soit :

$$q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_k) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \tag{1.46}$$

L'intérêt est que les poids sont mis à jour simplement :

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}) \tag{1.47}$$

Cependant, un grand nombre de particules est alors nécessaire pour approcher convenablement la densité de probabilité *a posteriori* car les particules sont propagées à l'aveugle et peuvent se retrouver localisées en majorité dans des régions de faible vraisemblance de l'espace d'état. Une alternative est de choisir la distribution d'importance comme suit :

$$q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_k)_{opt} = p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{1:k}) \tag{1.48}$$

Ce choix est optimal au sens où il minimise la variance des poids et donc maximise le nombre effectif N_{eff} [DOU00b]. Dans ce cas, les poids d'importance sont mis à jour comme suit :

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_{k-1}^{(i)}) \tag{1.49}$$

Le poids $w_k^{(i)}$ est alors indépendant du tirage de l'état courant de la particule et peut être calculé avant que la particule correspondante ne se propage.
Finalement, pour utiliser la densité d'importance optimale, il est nécessaire d'échantillonner

selon $p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_k)$ puis d'évaluer $p(\boldsymbol{y}_k|\boldsymbol{x}_{k-1}^{(i)})$. Ces deux étapes ne sont pas nécessairement réalisables facilement. En effet, obtenir une forme analytique de la distribution $p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_k)$ est difficile pour de nombreuses REEs. Ainsi, afin d'exploiter l'information apportée par la mesure $\boldsymbol{y}_k$, il est souvent souhaitable de recourir à une approximation de la densité d'importance optimale. En réalité, il existe deux cas où l'utilisation de la densité optimale est possible : lorsque l'espace d'état est discret ou lorsque la REE est linéaire gaussienne. Dans ce dernier cas, il est préférable de réaliser l'estimation par filtrage de Kalman.

Finalement, en considérant le nombre effectif $\hat{N}_{eff}$ comme mesure de dégénérescence puis en combinant le ré-échantillonnage avec le filtre SIS, il est possible de définir une extension de ce filtre. L'**Algorithme 1.1** présenté ci-dessous en décrit les principales étapes.

Algorithme 1.1 Etapes du filtre particulaire générique

- Filtrer selon l'algorithme SIS : Pour $i = 1, ..., N_p$
  - Echantillonner $\boldsymbol{x}_k^{(i)} \sim q(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_k)$
  - Calculer les poids d'importance selon (1.40)
- Normaliser les poids d'importance : Pour $i = 1, ..., N_p$

$$w_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N_p} w_k^{(j)}} \tag{1.50}$$

- Calculer le nombre efficace $\hat{N}_{eff}$ selon (1.42)
- Si $\hat{N}_{eff} < C$ où C est une constante à définir, alors ré-échantillonner.
  Le nouvel ensemble de particules est alors noté $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)} = \frac{1}{N_p}\}$
- Estimer l'état :

$$\hat{\boldsymbol{x}}_k = \sum_{i=1}^{N_p} w_k^{(i)} \boldsymbol{x}_k^{(i)} \tag{1.51}$$

Le filtre particulaire constitue donc une alternative au filtre de Kalman notamment dans un contexte de non-linéarité et/ou non-gaussianité. Néanmoins, sa mise œuvre n'est pas toujours évidente notamment du fait du problème de dégénérescence auquel il est sujet. C'est par exemple le cas lors de l'estimation de paramètres constants.
De nombreuses variantes du filtre particulaire ont été proposées à ce jour :

- l'algorithme *sequential importance resampling* (SIR), obtenu à partir de l'algorithme SIS en choisissant la densité de transition comme densité d'importance et en ré-échantillonnant systématiquement,
- le filtre particulaire Rao-Blackwellisé (RBPF) aussi appelé filtre particulaire marginalisé. Il s'applique aux modèles qui deviennent linéaires gaussiens conditionnellement à certaines variables,
- le filtre particulaire auxiliaire (APF) [Pit99], permettant de guider les particules dans des régions prometteuses en incluant une variable auxiliaire et cela sans considérer la distribution d'importance optimale,
- etc.

Pour plus de détails, le lecteur peut se référer aux ouvrages suivants [Ris04a, Dou01a, Dou00a, Dou01b].

Finalement, nous concluons cette partie en résumant sur la figure 1.2 les conditions d'utilisation des différents filtres introduits précédemment.

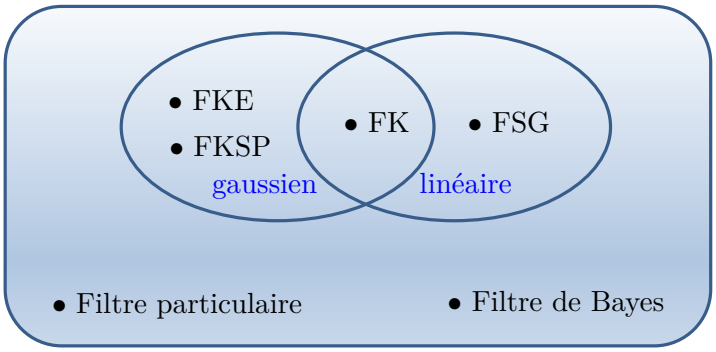

Figure 1.2 – Conditions d'utilisation des différents filtres bayésiens présentés

Les différentes approches bayésiennes de filtrage introduites précédemment sont toutes fondées sur l'utilisation d'un modèle d'évolution du vecteur d'état correspondant à un modèle de mouvement dans le cadre du pistage de cibles. Lorsque la cible est manœuvrante au cours du temps, le modèle utilisé peut ne pas convenir à certains instants. Il s'ensuit une erreur d'estimation sur l'état pouvant mener à un décrochage de la piste. Pour remédier à ce phénomène, il est possible d'utiliser les algorithmes à modèles multiples qui permettent de considérer plusieurs modèles d'évolution simultanément.

1.3 Pistage de cibles manœuvrantes : algorithmes à modèles multiples

Bien définir le modèle de mouvement sur lequel repose le filtre réalisant l'estimation de la trajectoire est primordial. En effet, les performances de l'algorithme en dépendent fortement. Cependant, un choix pertinent suppose de disposer *a priori* d'informations sur la dynamique de la cible. De plus, il se peut que le type de mouvement de la cible varie au cours de la trajectoire. Un modèle qui convenait pour le début de la trajectoire peut ne plus être adapté pour la suite. Dans ce cas, la trajectoire ne peut plus être représentée par un modèle unique mais doit être décomposée en plusieurs phases, chacune correspondant à un modèle de mouvement différent. Les algorithmes à modèles multiples ont été introduits dans le but de gérer ce type de problème. Le principe général est de disposer d'un banc de filtres fondés sur des modèles de mouvement différents. Ainsi, l'incertitude liée au modèle est contournée en déterminant à chaque instant le modèle le plus adapté pour chaque partie de la trajectoire.

Il existe trois générations d'approches par modèles multiples. Dans les algorithmes de première génération, on tire avantage de plusieurs estimateurs indépendants les uns des autres fondés sur des hypothèses différentes de modèle. A chaque instant, le modèle le plus vraisemblable est sélectionné. Ces approches sont dites statiques. Par opposition, dans la deuxième

génération, les estimateurs sont interactifs. Le système est supposé pouvoir changer de modèle au cours du temps. Dans ce cas, on parle d'estimateurs dynamiques.
Les algorithmes à modèles multiples de première et deuxième génération ont une structure dite fixe puisqu'ils utilisent le même jeu de modèles à chaque instant. Pour obtenir de "bonnes" performances avec des structures fixes, on pourrait *a priori* penser que plus on associe de modèles, plus l'estimation est précise. Pourtant, considérer un trop grand nombre de modèles de mouvement ne s'avère pas nécessairement être un choix judicieux pour plusieurs raisons. D'une part, la complexité calculatoire augmente avec le nombre de modèles. D'autre part, comme les matrices de covariance peuvent augmenter significativement selon les modèles considérés, cela ne permet pas nécessairement pas d'améliorer des performances de l'algorithme [Li96]. Les modèles multiples à structure variable ont été étudiés comme alternative. Les plus connus sont les algorithmes *model-group switching* et *likely model set* [Li00]. Ils constituent la famille des algorithmes de 3ème génération pour lesquels le nombre et la nature des estimateurs considérés peuvent évoluer à chaque instant.

Dans la suite, nous présentons succinctement l'estimateur à modèles multiples dynamique optimal puis détaillons le principe de l'algorithme *interacting multiple model* (IMM) qui est l'algorithme à modèles multiples de 2ème génération le plus couramment utilisé à ce jour.

1.3.1 Estimateur à modèles multiples dynamique optimal

On considère un système composé de r modèles qui s'appuient sur des modèles de mouvement différents notés $\{M_1, ..., M_r\}$. Si $m(k)$ désigne le modèle suivi par le système à l'instant k, la séquence $\{m(k)\}_{k>0}$ est supposée être une chaîne de Markov de matrice de probabilités de transition (TPM) $\Pi = \{\pi_{ij}\}_{i=1,...,r}^{j=1,...,r}$. Par souci de simplicité, on notera $m_i(k)$ l'évènement $m(k) = M_i$ pour $i = 1, ..., r$. Chaque élément π_{ij} de la matrice Π définit la probabilité de passer du $i^{\text{ème}}$ modèle au $j^{\text{ème}}$ modèle entre deux instants consécutifs, soit :

$$\pi_{ij} = Pr[m_j(k)|m_i(k-1)] \tag{1.52}$$

La chaîne de Markov étant homogène, les probabilités de transition sont supposées invariantes au cours du temps et indépendantes du modèle d'état.
La REE du système caractérisée par (1.8) et (1.9) devient alors :

$$\begin{aligned} m(k)|m(k-1) &\sim \Pi && (1.53)\\ \boldsymbol{x}_k &= F_k^{m(k)}\boldsymbol{x}_{k-1} + \boldsymbol{v}_k^{m(k)} && (1.54)\\ \boldsymbol{y}_k &= H_k^{m(k)}\boldsymbol{x}_k + \boldsymbol{w}_k^{m(k)} && (1.55) \end{aligned}$$

où (1.53) caractérise l'évolution du modèle suivi par la cible à chaque instant selon les probabilités de transition contenues dans la matrice Π. De plus, le bruit de modèle $\boldsymbol{v}_k^{m(k)}$ et le bruit de mesure $\boldsymbol{w}_k^{m(k)}$ ont pour matrice de covariance $Q_k^{m(k)}$ et $R_k^{m(k)}$ respectivement.
Notons que le système décrit par (1.53), (1.54) et (1.55) est une version généralisée d'un modèle de Markov caché.

Dans la suite, il s'agit de montrer comment estimer l'état de la cible selon l'estimateur

dynamique optimal.
A l'instant k, il y a r^k séquences de modèles différentes qu'a pu suivre le système depuis l'instant initial. Le nombre de séquences possibles augmente donc de façon exponentielle au cours du temps. La $l^{\text{ème}}$ séquence à l'instant k, notée $M^{k,l}$, satisfait :

$$M^{k,l} = \{M_{i_{1,l}}, ..., M_{i_{k,l}}\} \qquad l = 1, ..., r^k \tag{1.56}$$

avec $i_{\kappa,l}$ l'indice du modèle à l'instant κ issu de la séquence l, soit :

$$1 \leq i_{\kappa,l} \leq r \qquad \kappa = 1, ..., k \tag{1.57}$$

La probabilité conditionnelle *a posteriori* de la $l^{\text{ème}}$ séquence à l'instant k, notée $\mu^{k,l}$, satisfait :

$$\mu^{k,l} = Pr[M^{k,l}|\boldsymbol{y}_{1:k}] \tag{1.58}$$

De plus, la $l^{\text{ème}}$ séquence à l'instant k peut se décomposer à partir de la séquence mère, indexée par la lettre s, à partir de laquelle elle a été construite à l'instant précédent et en lui ajoutant le modèle à l'instant courant. Si M_j est le modèle associé à la séquence $M^{k,l}$ à l'instant k, alors :

$$M^{k,l} = \{M^{k-1,s}, m_j(k)\} \tag{1.59}$$

Ainsi, si i est l'indice du modèle de la séquence s à l'instant $k-1$, il vient par définition :

$$Pr[m_j(k)|M^{k-1,s}] = Pr[m_j(k)|m_i(k-1)] = \pi_{ij} \tag{1.60}$$

En prenant en considération les notations précédentes, la densité de probabilité de l'état à l'instant k est obtenue selon le théorème des probabilités totales comme suit :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \sum_{l=1}^{r^k} p(\boldsymbol{x}_k|M^{k,l}, \boldsymbol{y}_{1:k}) \underbrace{Pr[M^{k,l}|\boldsymbol{y}_{1:k}]}_{\mu^{k,l}\text{d'après (1.58)}} \tag{1.61}$$

avec :

$$\begin{aligned} \mu^{k,l} &= Pr(M^{k,l}|\boldsymbol{y}_{1:k}) \\ &\propto p(\boldsymbol{y}_k|M^{k,l}, \boldsymbol{y}_{1:k-1}) Pr[m_j(k)|M^{k-1,s}]\mu^{k-1,s} \\ &\propto p(\boldsymbol{y}_k|M^{k,l}, \boldsymbol{y}_{1:k-1}) Pr[m_j(k)|m_i(k-1)]\mu^{k-1,s} \\ &\propto p(\boldsymbol{y}_k|M^{k,l}, \boldsymbol{y}_{1:k-1}) \pi_{ij}\mu^{k-1,s} \end{aligned} \tag{1.62}$$

Compte tenu de (1.62), il apparaît que malgré le caractère Markovien du système, le conditionnement se fait par rapport à la séquence complète des modèles et non pas selon seulement le dernier modèle. De plus, r^k séquences doivent être évaluées à chaque instant, ce qui implique de considérer un nombre croissant de filtres. Cela rend la mise en œuvre de l'approche optimale impossible.
La seule solution permettant de réduire le nombre croissant d'hypothèses à évaluer au cours du temps est de considérer des approches sous-optimales. Par exemple, seule une certaine profon-

deur de l'historique des modèles est conservée dans certaines approches. Il s'agit des algorithmes GPB pour *generalized pseudo Bayesian* [Bar95]. L'algorithme GPB du premier ordre, appelé GPB1, ne conserve que les r hypothèses de modèles correspondant à l'instant précédent. L'algorithme GPB du deuxième ordre, appelé GPB2, conserve les r^2 hypothèses correspondant aux combinaisons de modèles possibles sur les deux derniers instants. Enfin, notons que de façon similaire au filtrage particulaire, les séquences de modèles conservées peuvent aussi être tirées aléatoirement.

Dans la suite, nous présentons une autre alternative qui nécessite seulement r filtres. Elle a globalement le même coût calculatoire qu'un GPB1 mais a des performances d'estimation similaires au GPB2. C'est l'algorithme IMM. Comme illustré par la figure 1.3, qui présente la structure de l'arbre des modèles de l'IMM à deux modèles comparé au cas optimal, l'intérêt d'une telle approche est de réduire de façon significative le coût calculatoire tout en conservant de bonnes performances d'estimation.

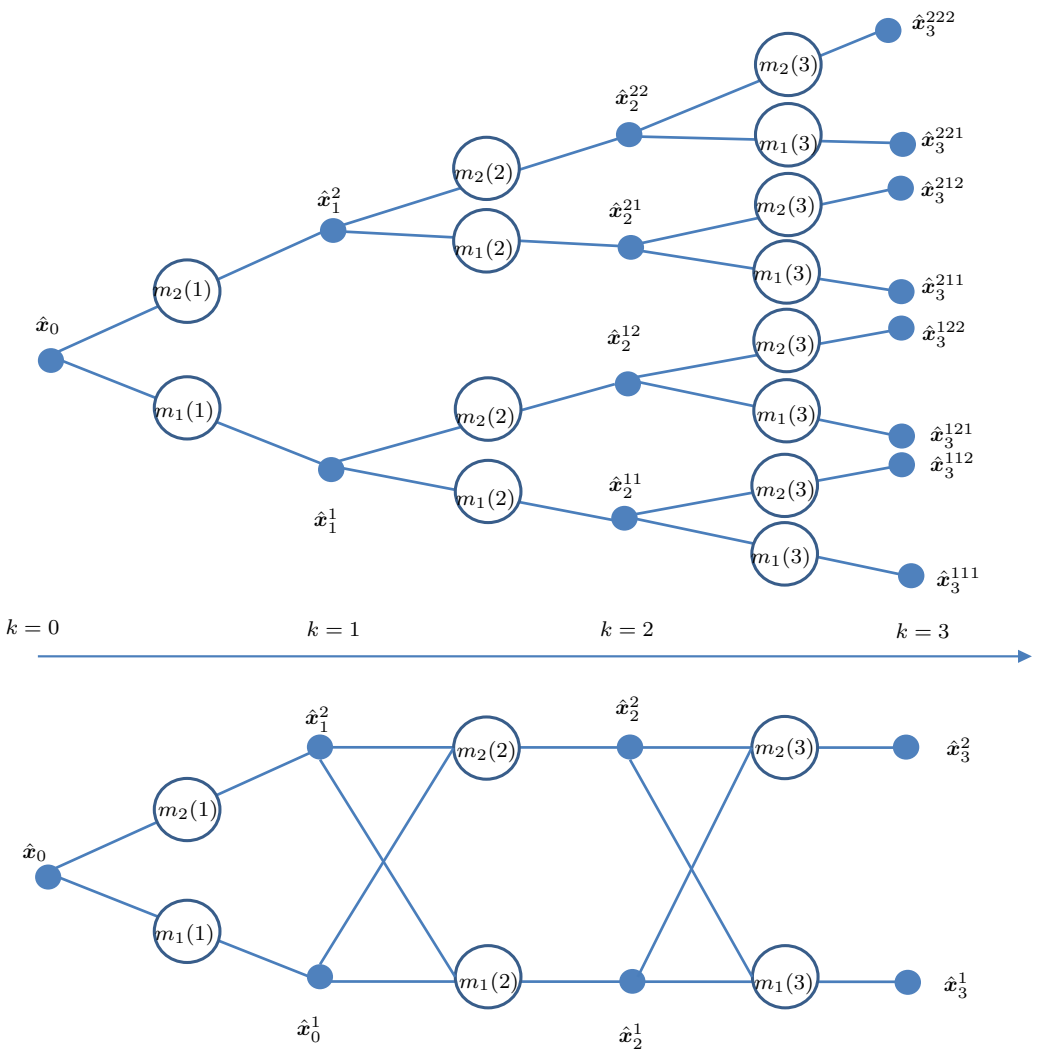

Figure 1.3 – Evolution des modèles dans le cas optimal et selon l'algorithme IMM. Dans le premier cas, l'exposant au niveau de l'état estimé $\hat{\boldsymbol{x}}_k$ désigne la séquence de modèles.

1.3.2 Algorithme à modèles multiples interactifs (IMM)

Le principe de l'algorithme IMM est d'estimer l'état de la cible à l'instant courant en combinant les estimés de r filtres associés à r modèles différents. L'originalité de cet algorithme est de considérer en entrée de chaque filtre une combinaison linéaire des états estimés à l'instant précédent à partir de chacun des r filtres.

Dans la suite, nous détaillons les étapes de prédiction et de mise à jour de l'algorithme IMM telles qu'introduites par Bar-Shalom [Bar95].

Etape de prédiction :
Considérons la densité de probabilité de l'état *a posteriori* à l'instant $k-1$ notée $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$. Dans l'approche originale de l'algorithme IMM [Bar95], Bar-Shalom propose de considérer qu'à chaque instant cette densité de probabilité peut s'écrire sous la forme d'un mélange de gaussiennes. Selon le théorème de probabilités totales, il vient :

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}) = \sum_{i=1}^{r} \mu_{k-1}^{i} \mathcal{N}(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1|k-1}^{i}, P_{k-1|k-1}^{i}) \tag{1.63}$$

où $\mu_{k-1}^{i} = Pr(m_i(k-1)|\boldsymbol{y}_{1:k-1})$ désigne la probabilité *a posteriori* que le système suive le modèle i à l'instant $k-1$. De plus, $\hat{\boldsymbol{x}}_{k-1|k-1}^{i}$ et $P_{k-1|k-1}^{i}$ désignent respectivement la moyenne et la matrice de covariance de chaque composante gaussienne.

La densité de probabilité prédite, notée $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$, s'exprime alors comme suit :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) &= \sum_{i=1}^{r} p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1}) Pr(m_i(k)|\boldsymbol{y}_{1:k-1}) \\ &= \sum_{i=1}^{r} p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1}) \mu_{k|k-1}^{i} \end{aligned} \tag{1.64}$$

où les poids prédits $\mu_{k|k-1}^{i}$ sont obtenus en tirant profit de la structure en chaîne de Markov considérée. Ils satisfont :

$$\mu_{k|k-1}^{i} = \sum_{j=1}^{r} \pi_{ji} \mu_{k-1}^{j} \tag{1.65}$$

Dans la suite, il s'agit de montrer que la densité de probabilité $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ définie en (1.64) peut s'écrire sous la forme d'une gaussienne. En effet, cette dernière peut se décomposer comme suit :

$$\begin{aligned} & p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1}) \\ &= \int_{\boldsymbol{x}_{k-1}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, m_i(k)) p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}, m_i(k)) d\boldsymbol{x}_{k-1} \\ &= \int_{\boldsymbol{x}_{k-1}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, m_i(k)) \times \sum_{j=1}^{r} Pr[m_j(k-1)|m_i(k), \boldsymbol{y}_{1:k-1}] p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}, m_j(k-1)) d\boldsymbol{x}_{k-1} \\ &\underset{(1.63)}{=} \int_{\boldsymbol{x}_{k-1}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, m_i(k)) \times \sum_{j=1}^{r} \mu_{k-1|k}^{ji} \mathcal{N}(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1|k-1}^{j}, P_{k-1|k-1}^{j}) d\boldsymbol{x}_{k-1} \end{aligned} \tag{1.66}$$

où $\mu_{k-1|k}^{ji} = Pr[m_j(k-1)|m_i(k), \boldsymbol{y}_{1:k-1}]$ représente la probabilité d'avoir été dans le modèle j à l'instant $k-1$ sachant que l'on est dans le modèle i à l'instant k. Ces probabilités de mélange sont calculées comme suit :

$$\begin{aligned}\mu_{k-1|k}^{ji} &\triangleq Pr[m_j(k-1)|m_i(k), \boldsymbol{y}_{1:k-1}] \\ &= \frac{Pr[m_i(k)|m_j(k-1), \boldsymbol{y}_{1:k-1}]Pr[m_j(k-1)|\boldsymbol{y}_{1:k-1}]}{\sum_{l=1}^{r} Pr[m_i(k)|m_l(k-1), \boldsymbol{y}_{1:k-1}]Pr[m_l(k-1)|\boldsymbol{y}_{1:k-1}]} \\ &= \frac{\pi_{ji}\mu_{k-1}^{j}}{\sum_{l=1}^{r} \pi_{li}\mu_{k-1}^{l}}\end{aligned} \tag{1.67}$$

Cette relation met en évidence l'importance du réglage de la matrice des probabilités de transition Π. Ce dernier influe sur l'évolution des probabilités associées aux modèles de l'IMM et donc sur les performances d'estimation de l'algorithme.

Ainsi, en associant (1.64) et (1.66), il apparaît que la densité prédite $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$ définie en (1.64) peut se décomposer comme un mélange de r^2 gaussiennes.
A ce stade, afin de limiter la complexité calculatoire, Bar Shalom [Bar01] propose d'approcher la densité de probabilité $p(\boldsymbol{x}_{k-1}|m_i(k), \boldsymbol{y}_{1:k-1})$ par une seule gaussienne. Ainsi, il vient :

$$p(\boldsymbol{x}_{k-1}|m_i(k), \boldsymbol{y}_{1:k-1}) \approx \mathcal{N}(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1|k-1}^{0i}, P_{k-1|k-1}^{0i}) \tag{1.68}$$

où les moyennes $\{\hat{\boldsymbol{x}}_{k-1|k-1}^{0i}\}_{i=1,\dots,r}$ et matrices de covariance $\{P_{k-1|k-1}^{0i}\}_{i=1,\dots,r}$, utilisées en entrée du $i^{\text{ème}}$ filtre, sont obtenues selon une étape de mélange combinant les sorties de tous les filtres :

$$\begin{aligned}\hat{\boldsymbol{x}}_{k-1|k-1}^{0i} &= \sum_{j=1}^{r} \mu_{k-1|k}^{ji} \hat{\boldsymbol{x}}_{k-1|k-1}^{j} \\ P_{k-1|k-1}^{0i} &= \sum_{j=1}^{r} \mu_{k-1|k}^{ji} \left(P_{k-1|k-1}^{j} + (\hat{\boldsymbol{x}}_{k-1|k-1}^{j} - \hat{\boldsymbol{x}}_{k-1|k-1}^{0i})(\hat{\boldsymbol{x}}_{k-1|k-1}^{j} - \hat{\boldsymbol{x}}_{k-1|k-1}^{0i})^T\right)\end{aligned} \tag{1.69}$$

Comme mis en avant par la figure 1.3 et (1.61), par rapport au cas optimal, l'approximation (1.68) permet de ne considérer à chaque instant que r filtres plutôt que r^k dans le cas optimal. En substituant (1.68) dans (1.66), il vient :

$$\begin{aligned}p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1}) &= \int_{\boldsymbol{x}_{k-1}} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, m_i(k)) \times \mathcal{N}(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1|k-1}^{0i}, P_{k-1|k-1}^{0i}) d\boldsymbol{x}_{k-1} \\ &= \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}^{i}, P_{k|k-1}^{i})\end{aligned} \tag{1.70}$$

avec :

$$\hat{\boldsymbol{x}}_{k|k-1}^{i} = F_k^{m_i(k)} \hat{\boldsymbol{x}}_{k-1|k-1}^{0i} \tag{1.71}$$

$$P_{k|k-1}^{i} = F_k^{m_i(k)} P_{k-1|k-1}^{0i} (F_k^{m_i(k)})^T + Q_k^{m_i(k)} \tag{1.72}$$

En combinant (1.64) et (1.70), on montre que la densité de probabilité prédite peut donc se mettre sous la forme d'un mélange de gaussiennes tel que :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \sum_{i=1}^{r} \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}^{i}, P_{k|k-1}^{i}) \mu_{k|k-1}^{i} \tag{1.73}$$

Etape de mise à jour :
Supposons que pour l'ensemble des estimateurs considérés, les densités de probabilités prédites $p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1})$ sont connues. D'après la règle de Bayes, il vient :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{i=1}^{r} \mu_k^i \frac{p(\boldsymbol{y}_k|m_i(k), \boldsymbol{x}_k)}{p(\boldsymbol{y}_k|m_i(k), \boldsymbol{y}_{1:k-1})} p(\boldsymbol{x}_k|m_i(k), \boldsymbol{y}_{1:k-1}) \tag{1.74}$$

où si la fonction de vraisemblance associée au $i^{\text{ème}}$ filtre $p(\boldsymbol{y}_k|m_i(k), \boldsymbol{y}_{1:k-1})$ est notée Λ_k^i, les probabilités associées aux modèles sont mises à jour de la manière suivante :

$$\mu_k^i = \frac{\Lambda_k^i \mu_{k|k-1}^i}{\sum_{l=1}^{r} \Lambda_k^l \mu_{k|k-1}^l} \tag{1.75}$$

De plus, d'après (1.70) et considérant que la fonction de vraisemblance satisfait :

$$\Lambda_k^i = \mathcal{N}(\boldsymbol{y}_k; \hat{\boldsymbol{y}}_{k|k-1}^i, S_k^i) \tag{1.76}$$

avec $\hat{\boldsymbol{y}}_{k|k-1}^i = H_k^{m_i(k)} \hat{\boldsymbol{x}}_{k|k-1}^i$ la mesure prédite et S_k^i la matrice de covariance de l'innovation associée au $i^{\text{ème}}$ modèle. L'équation (1.74) peut alors se réécrire comme un mélange de r gaussiennes :

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{i=1}^{r} \mu_k^i \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k}^i, P_{k|k}^i) \tag{1.77}$$

où les moyennes $\hat{\boldsymbol{x}}_{k|k}^i$ et les matrices de covariances $P_{k|k}^i$ peuvent être calculées selon l'étape de mise à jour du filtre de Kalman comme suit :

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k}^i &= \hat{\boldsymbol{x}}_{k|k-1}^i + K_k(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1}^i) \\ P_{k|k}^i &= (I_{n_x} - K_k H_k^{m_i(k)}) P_{k|k-1}^i \\ K_k &= P_{k|k-1}^i (H_k^{m_i(k)})^T (S_k^i)^{-1} \end{aligned} \tag{1.78}$$

Finalement, l'estimation globale peut être obtenue en recombinant les estimations de chaque filtre à partir des probabilités *a posteriori*, soit :

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k} &= \sum_{i=1}^{r} \hat{\boldsymbol{x}}_{k|k}^i \mu_k^i \\ P_{k|k} &= \sum_{i=1}^{r} \left(P_{k|k}^i + (\hat{\boldsymbol{x}}_k^i - \hat{\boldsymbol{x}}_{k|k})(\hat{\boldsymbol{x}}_{k|k}^i - \hat{\boldsymbol{x}}_{k|k})^T \right) \mu_k^i \end{aligned} \tag{1.79}$$

Les principales étapes de l'algorithme IMM sont représentées sur la figure 1.4. L'algorithme IMM permet donc de prendre en compte plusieurs modèles d'évolution simultanément en les faisant intéragir. Cependant, en pratique, il existe certaines problématiques notamment liées au réglage de l'algorithme.

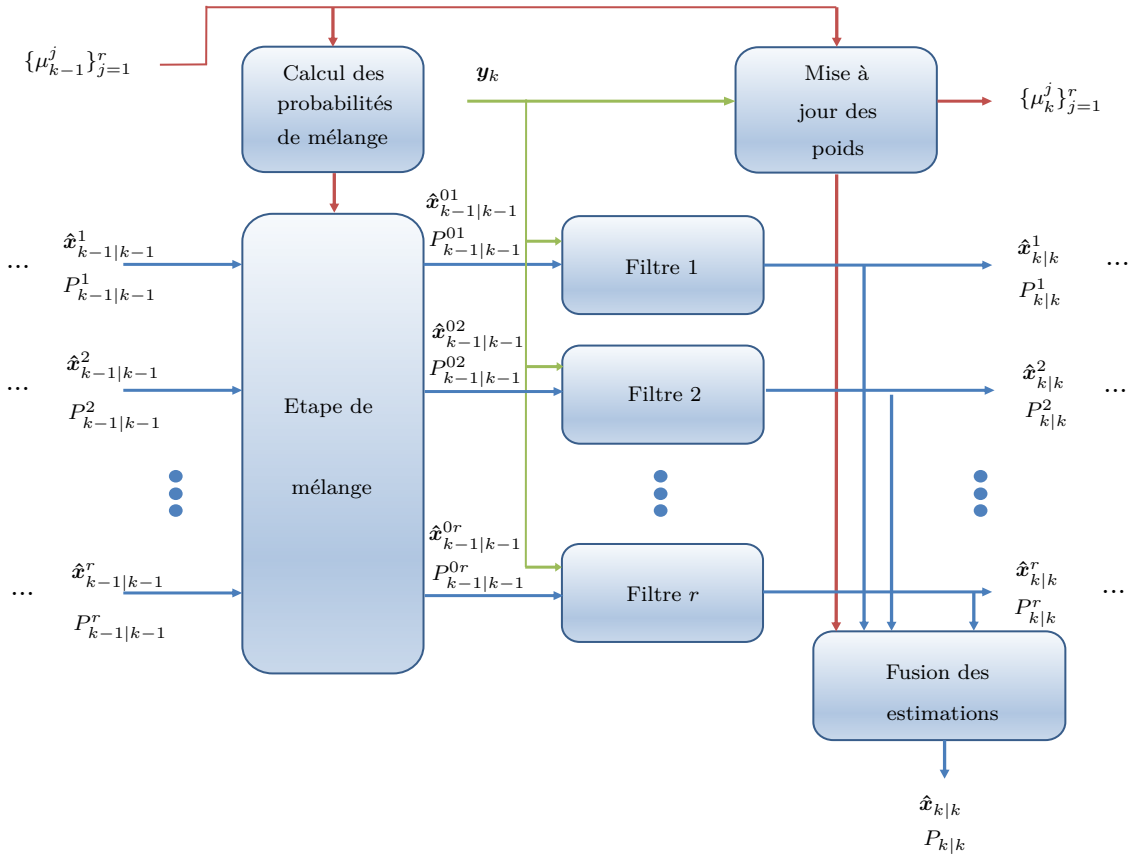

FIGURE 1.4 – Schéma bloc de l'algorithme IMM

1.3.3 Problématiques liées à l'utilisation des modèles multiples

L'utilisation simultanée de plusieurs modèles d'évolution permet de relâcher les hypothèses sur le mouvement suivi par la cible. D'une part, plusieurs familles de mouvement peuvent être envisagées pour représenter l'évolution du système au cours du temps. D'autre part, pour éviter un réglage *a priori* des paramètres d'un modèle de mouvement, plusieurs modèles, associés à une même famille de mouvement mais avec différentes valeurs de leurs paramètres, peuvent être associés.
Cependant, il susbsite une problématique importante inhérente à la phase de conception de l'algorithme : quels modèles associer ensemble et comment faire le choix de ces modèles ?

Les algorithmes à modèles multiples de troisième génération ont en partie permis de répondre à ces questions en utilisant à chaque instant un sous-ensemble de modèles parmi un ensemble initial prédéfini. Néanmoins, ces méthodes nécessitent une connaissance fine des modèles considérés puisqu'il faut être en mesure de définir les liens existants (que l'on nomme "adjacence") entre ces modèles et qui permettent de réactiver certains modèles au cours du temps. Or, il n'existe pas à notre connaissance de méthode pour définir ces relations. Pourtant, le choix du type de modèles et de leurs paramètres incombe à l'utilisateur. Cette selection est subjective et fondée sur son expérience et la connaissance *a priori* du système.
De plus, l'une des difficultés liée à l'utilisation de l'IMM est le choix de la matrice des pro-

babilités de transition. En effet, selon le choix effectué *a priori* par l'utilisateur, les performances d'estimation peuvent varier de façon plus ou moins importante. Pour pallier cette difficulté, certaines méthodes consistent à estimer les éléments de cette matrice au cours du temps [Tug82, Jil04, Ben15]. Par exemple, Jilkov *et al.* proposent d'exprimer la matrice des probabilités de transition comme une somme pondérée de matrices prédéfinies puis d'estimer le poids de chacune d'elles de manière récursive.

Le formalisme introduit précédemment est valable lorsque la mesure associée à l'estimateur à chaque instant est toujours celle issue de la cible. Or, dans la pratique, une cible peut ne pas être détectée à chaque instant. De plus, la présence de clutter de mer ou de terre est à l'origine de fausses alarmes. Ainsi, le nombre de cibles et le nombre de mesures reçues varient au cours du temps et sans traitement préalable il n'est pas possible de définir quelle mesure est issue de quelle cible. Afin de traiter cette problématique, les approches de filtrage présentées précédemment sont généralement associées à des algorithmes d'association de données pour pouvoir lever l'ambiguité sur l'origine des mesures reçues. Une présentation de ces approches est proposée en annexe A. Leur limite principale est que lorsque les cibles sont proches les unes des autres et/ou que le bruit de mesure est important, les erreurs d'association conduisent généralement à une mauvaise estimation de l'état de chaque piste.
Plus récemment, une alternative fondée sur la théorie des ensembles finis aléatoires (EFAs) a été introduite par Mahler [Mah01] dans le contexte multi-cibles. Dans la partie suivante, sans être exhaustif, nous revenons sur ces concepts et sur certains des filtres bayésiens utilisés dans cette thèse qui s'inscrivent dans ce cadre théorique.

1.4 Ensembles finis aléatoires pour le pistage multi-cibles

Une alternative à la combinaison des traitements de filtrage avec les méthodes d'association a récemment été introduite dans un contexte multi-cibles par Mahler. Elle est fondée sur la théorie des statistiques des ensembles finis. Ce formalisme a permis de généraliser le filtre de Bayes monocible au cas multi-cibles. Le principe est de modéliser les états et les observations par des EFAs qui sont des concepts associés à la théorie de la mesure. Les EFAs offrent un cadre unifié pour modéliser l'incertitude sur le cardinal de ces ensembles et la valeur de chacun des vecteurs aléatoires les constituant. Les cibles ne sont pas gérées indépendamment les unes des autres mais pistées dans leur ensemble de sorte que l'étape d'association n'est pas nécessaire.
Dans la suite de cette partie, nous exposons les principes fondamentaux de la théorie des EFAs. Puis nous introduisons le filtre de Bayes multi-cibles ainsi que deux de ses approximations : les filtres PHD et CPHD. Pour plus de détails sur ces concepts, le lecteur peut se référer aux tutoriels écrits par Mahler intitulés *Statistics 101* [Mah04] et *Statistics 102* [Mah13] qui présentent les principes fondamentaux permettant de se familiariser avec les statistiques des ensembles finis.

1.4.1 Définitions, notations et abréviations

Un ensemble fini aléatoire (EFA ou *random finite set* -RFS- en anglais) est un ensemble fini non ordonné de variables aléatoires dont la cardinalité est elle-même aléatoire. Contrairement à un vecteur aléatoire dont les composantes ne sont pas interchangeables, les variables aléatoires

qui constituent l'EFA ne sont pas ordonnées. Cet ensemble est entièrement caractérisé par la distribution de probabilité discrète du cardinal et une famille de densités de probabilités jointes pour décrire les éléments de l'EFA.
Dans le contexte du pistage multi-cibles, la cardinalité d'un EFA X, notée $|X|$, peut réprésenter le nombre de cibles ou le nombre d'observations reçues à chaque instant alors que les valeurs des éléments de X peuvent représenter les états des cibles ou les observations. Ce formalisme permet de modéliser naturellement la situation de pistage multi-cibles.

Dans la suite, afin de distinguer les EFAs des vecteurs aléatoires, on note les états multi-cibles avec une lettre majuscule contrairement à l'état monocible qui est noté à l'aide d'une lettre minuscule. De plus, pour simplifier les notations, aucune différence n'est faite entre un EFA et sa réalisation. L'espace des sous-ensembles finis de $\mathbb{X}$ est noté $\mathcal{F}(\mathbb{X})$ et l'espace des sous-ensembles finis contenant exactement n éléments est noté $\mathcal{F}_n(\mathbb{X})$.

Introduisons dans un premier temps d'autres notations.
Le produit scalaire de deux fonctions continues $f(\boldsymbol{x})$ et $g(\boldsymbol{x})$ à energie finie à valeurs dans $\mathbb{R}^{n_x}$ satisfait :

$$\langle f, g \rangle \triangleq \int f(\boldsymbol{x}) g(\boldsymbol{x}) d\boldsymbol{x} \tag{1.80}$$

Pour des séquences discrètes, l'intégrale dans (1.80) est remplacée par une somme.

La notation d'exponentielle multi-cibles [Vo13], définie comme suit :

$$h^X = \prod_{\boldsymbol{x} \in X} h(\boldsymbol{x}) \tag{1.81}$$

est utilisée pour simplifier les notations pour les fonctions à valeur réelles $h(x)$ qui doivent être évaluées pour tout vecteur d'état d'un EFA X. De plus, notons que par définition $h^X = 1$ si $X = \emptyset$.

La fonction Kronecker généralisée aux EFAs est définie comme suit :

$$\delta_Y(X) \triangleq \begin{cases} 1, & \text{si } X = Y, \\ 0, & \text{sinon} \end{cases} \tag{1.82}$$

La fonction indicatrice ou fonction d'inclusion définie par :

$$1_Y(X) \triangleq \begin{cases} 1, & \text{si } X \subseteq Y, \\ 0, & \text{sinon} \end{cases} \tag{1.83}$$

est utilisée pour savoir si un ensemble X est un sous-ensemble ou non d'un ensemble Y.

Dans la sous-partie suivante, nous introduisons plusieurs EFAs communément utilisés dans le contexte du pistage multi-cibles.

1.4.2 Distributions de probabilités multi-cibles

Tout comme dans le cas monocible où un vecteur aléatoire $\boldsymbol{x}$ est caractérisé par sa densité de probabilité $p(\boldsymbol{x})$, un état multi-cibles est caractérisé par sa densité de probabilité multi-cibles notée $\pi(X)$. Cette densité de probabilité incorpore l'incertitude sur l'état même des cibles mais également sur le cardinal de l'ensemble. Elle est définie comme suit :

$$\pi(X) \triangleq \begin{cases} \pi_0(\emptyset), & \text{si } X = \emptyset, \\ \pi_1(\{\boldsymbol{x}^{(1)}\}), & \text{si } X = \{\boldsymbol{x}^{(1)}\}, \\ \pi_2(\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\}), & \text{si } X = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\}, \\ \vdots & \vdots \end{cases} \tag{1.84}$$

L'intégrale de la densité de probabilité multi-cibles est définie par :

$$\int \pi(X)\delta X = \sum_{i=0}^{+\infty} \frac{1}{i!} \int \pi_i(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(i)}\}) d\boldsymbol{x}^{(1)} ... \boldsymbol{x}^{(i)} = 1 \tag{1.85}$$

où $\pi_i(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(i)}\}) = 0$ si $|\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(i)}\}| \neq i$.

Enfin, la distribution de la cardinalité d'un EFA X qui correspond à la probabilité de contenir exactement n vecteurs est :

$$\rho(n) = Pr(|X| = n) = \frac{1}{n!} \int \pi_n(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}\}) d\boldsymbol{x}^{(1)} ... \boldsymbol{x}^{(n)} \tag{1.86}$$

Ainsi, à partir de la distribution de probabilité de la cardinalité définie par (1.86), il est possible d'obtenir une estimation du nombre de cibles représentées par l'EFA X.

Dans la suite, nous décrivons différents EFAs couramment utilisés dans le cadre du pistage multi-cibles.

<u>EFA caractérisé par des clusters i.i.d :</u>
Un EFA noté X modélisé par un cluster de variables indépendantes et identiquement distribuées (i.i.d) a sa densité de probabilité multi-cibles définie comme suit :

$$\pi_n(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}\}) = n!\rho(n) \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)}) \tag{1.87}$$

où ρ est la distribution de probabilité de la cardinalité et $p(\boldsymbol{x}^{(i)})$ correspond à la distribution spatiale de la $i^{\text{ème}}$ cible.

<u>EFA de Poisson :</u>
Un EFA de Poisson X sur $\mathbb{X}$ est entièrement caractérisé par sa fonction d'intensité $v(\cdot)$ qui correspond au moment du premier ordre de la densité multi-cibles. C'est le seul EFA pour lequel la distribution de la cardinalité de X est de type Poisson de moyenne :

$$N = \int v(\boldsymbol{x}) d\boldsymbol{x} \tag{1.88}$$

et pour une cardinalité donnée, les éléments de X sont i.i.d selon $v(.)/N$.

La densité de probabilité d'un EFA de Poisson peut être directement obtenue à partir de (1.87) en remplacant la distribution de cardinalité $\rho(n)$ par :

$$\rho(n) = e^{-\lambda}\frac{\lambda^n}{n!} \tag{1.89}$$

Ainsi, il vient :

$$\pi(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}\}) = e^{-\lambda}\lambda^n \prod_{i=1}^{n} p(\boldsymbol{x}^{(i)}) = e^{-\lambda}\lambda^n \prod_{i=1}^{n} \frac{v(\boldsymbol{x}^{(i)})}{N} \tag{1.90}$$

Notons que dans de nombreuses applications de pistage, il est usuel de modéliser le processus de fausses alarmes par un EFA de type Poisson dont la densité d'intensité est $\kappa(\cdot) = \lambda_c \times \mathcal{U}(\cdot)$, avec λ_c le nombre moyen de fausses alarmes et $\mathcal{U}(\cdot)$ une densité de probabilité supposée uniforme sur l'espace des mesures.

<u>EFA de Bernoulli :</u>
Un EFA de Bernouli X sur $\mathbb{X}$ a la probabilité $1-r$ d'être vide et la probabilité complémentaire r d'être un singleton distribué selon la densité de probabilité $p(\cdot)$ définie sur $\mathbb{X}$. La distribution de cardinalité d'un EFA de Bernoulli est alors une distribution de Bernoulli de paramètre r et sa densité de probabilité est définie comme suit :

$$\pi(X) = \begin{cases} 1-r, & \text{si } \ X = \{\emptyset\} \\ rp(\boldsymbol{x}), & \text{si } \ X = \{\boldsymbol{x}\} \\ 0, & \text{sinon} \end{cases} \tag{1.91}$$

<u>EFA multi-Bernoulli :</u>
Un EFA dit multi-Bernoulli noté X sur $\mathbb{X}$ est l'union d'un nombre fini de M EFAs de Bernoulli indépendants notés $X^{(i)}$ avec la probabilité $r^{(i)} \in [0,1]$ et la densité de probabilité $p^{(i)}(\cdot)$ définie sur $\mathbb{X}$ pour $i = 1, ..., M$, soit :

$$X = \bigcup_{i=1}^{M} X^{(i)} \tag{1.92}$$

La cardinalité moyenne de l'EFA multi-Bernoulli est alors $\sum_{i=1}^{M} r^{(i)}$. Un EFA de ce type est entièrement caractérisé par l'ensemble des paramètres de multi-Bernoulli associés, soit $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$. Sa densité de probabilité est alors :

$$\pi(\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}\}) = \begin{cases} \prod_{i=1}^{M}(1-r^{(i)}), & \text{si } \ n = 0 \\ \prod_{i=1}^{M}(1-r^{(i)}) \sum_{1 \leq i_1 \neq ... \neq i_n \leq M} \prod_{j=1}^{n} \frac{r^{(i_j)} p^{(i_j)}(\boldsymbol{x}^{(j)})}{1-r^{(i_j)}}, & \text{si } \ n \leq M \\ 0, & \text{sinon} \end{cases} \tag{1.93}$$

Dans la suite, par souci de simplicité, la densité définie en (1.93) est notée :

$$\pi(\{(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)\}) = \{(r^{(i)}, p^{(i)})\}_{i=1}^{M} \tag{1.94}$$

La distribution de cardinalité s'écrit :

$$\rho(n) = \prod_{i=1}^{M} (1 - r^{(j)}) \sum_{1 \leq i_1 \neq ... \neq i_n \leq M} \prod_{j=1}^{n} \frac{r^{(i_j)}}{1 - r^{(i_j)}} \tag{1.95}$$

Dans [Mah07], Mahler montre que la fonction d'intensité d'un EFA de type multi-Bernoulli a l'expression suivante :

$$v(\boldsymbol{x}) = \sum_{j=1}^{M} r^{(j)} p^{(j)}(\boldsymbol{x}) \tag{1.96}$$

et a pour cardinalité moyenne $\hat{N} = \sum_{j=1}^{M} r^{(j)}$.

1.4.3 Représentation multi-cibles du système

A chaque instant, parmi les cibles qui survivent et celles qui apparaissent, chacune d'entre elles peut donner lieu ou non une mesure. Celles-ci seront élaborées par le senseur avec d'autres mesures issues de fausses alarmes. De ce fait, l'évolution des cibles au cours du temps ainsi que l'origine de chaque mesure est inconnue. Une telle incertitude et une telle variabilité peuvent se caractériser naturellement par l'utilisation d'EFAs afin de construire des modèles stochastiques pour la dynamique des états et la génération des observations.

On suppose qu'à l'instant k, il y a $N(k)$ cibles, associées à leurs vecteurs d'états $\boldsymbol{x}_k^{(1)}, ..., \boldsymbol{x}_k^{(N(k))}$, prenant leurs valeurs dans l'espace d'état $\mathbb{X} \subseteq \mathbb{R}^{n_x}$, et $M(k)$ mesures $\boldsymbol{y}_k^{(1)}, ..., \boldsymbol{y}_k^{(M(k))}$, prenant leurs valeurs dans l'espace d'observations $\mathbb{Y} \subseteq \mathbb{R}^{n_y}$. Puisque le nombre d'états et d'observations est variable et que l'ordre est indifférent, ces ensembles peuvent être représentés à chaque instant par des EFAs comme suit :

$$\begin{aligned} X_k &= \{\boldsymbol{x}_k^{(1)}, ..., \boldsymbol{x}_k^{(N(k))}\} \in \mathcal{F}(\mathbb{X}) \\ Y_k &= \{\boldsymbol{y}_k^{(1)}, ..., \boldsymbol{y}_k^{(M(k))}\} \in \mathcal{F}(\mathbb{Y}) \end{aligned} \tag{1.97}$$

L'ensemble fini X_k est appelé ensemble d'état multi-cibles et l'ensemble fini Y_k est appelé ensemble d'observation multi-cibles.

1.4.3.1 Modèle d'état multi-cibles

Dans cette sous-partie, nous décrivons un EFA permettant de modéliser l'état multi-cibles. Celui-ci permet de prendre en compte la dynamique des cibles ainsi que les processus de naissance et de disparition.

Une cible avec l'état $\boldsymbol{x}_{k-1}$ survit à l'instant suivant avec la probabilité $P_S(\boldsymbol{x}_{k-1})$. Elle prend alors l'état $\boldsymbol{x}_k$ à l'instant k selon la densité de probabilité $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$. Sinon, elle meurt avec la probabilité $1 - P_S(\boldsymbol{x}_{k-1})$ et prend la valeur $\emptyset$.

Considérant la réalisation X_{k-1} de l'EFA multi-cibles à l'instant $k-1$, le comportement de chaque cible d'état $\boldsymbol{x}_k \in X_k$ peut être modélisé par l'EFA de Bernoulli suivant :

$$S_{k|k-1}(\boldsymbol{x}_{k-1}) = \{P_S(\boldsymbol{x}_{k-1}), p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\} \tag{1.98}$$

De plus, de nouvelles cibles peuvent apparaître à chaque instant. Ainsi, à l'instant k, l'état multi-cibles est représenté par l'EFA X_k qui peut se décomposer de la manière suivante [2] :

$$X_k = \left(\bigcup_{x_{k-1} \in X_{k-1}} S_{k|k-1}(\boldsymbol{x}_{k-1}) \right) \cup \Gamma_k \tag{1.99}$$

où Γ_k désigne l'EFA modélisant les cibles naissantes à l'instant k. Selon le type de filtre, il peut être caractérisé par un EFA de Poisson, un EFA de cluster i.i.d ou encore un EFA multi-Bernoulli. Enfin, on suppose généralement que les EFAs $S_{k|k-1}$ et Γ_k sont indépendants [Vo08].

Densité de transition markovienne :
L'équation (1.99) peut également se mettre sous la forme d'une densité de transition multi-cibles qui évalue la probabilité de passer de l'état X_{k-1} à l'instant $k-1$ à l'état X_k à l'instant k. Contrairement à la densité de transition monocible, la densité de transition multi-cibles doit représenter l'apparition de nouvelles cibles et la survie des cibles d'un instant à l'autre. Si les $N(k)$ cibles considérées à chaque instant évoluent indépendamment les unes des autres et que les nouvelles cibles apparaissent selon une distribution de naissance du type Poisson $\pi_B(\boldsymbol{X})$ de paramètre λ_B, alors la densité de transition markovienne multi-cibles s'écrit comme suit :

$$p(\boldsymbol{X}_k|\boldsymbol{X}_{k-1}) = \pi_B(\boldsymbol{X}_k) \prod_{i=1}^{N(k)} \left(1 - P_S(\boldsymbol{x}_{k-1}^{(i)})\right) \sum_{\theta} \prod_{j \setminus \theta(j)>0} \frac{P_S(\boldsymbol{x}_{k-1}^{\theta(j)}) p(\boldsymbol{x}_k^{(j)}|\boldsymbol{x}_{k-1}^{\theta(j)})}{(1 - P_S(\boldsymbol{x}_{k-1}^{\theta(j)}))\lambda_B p_B(\boldsymbol{x}_k^{(j)})} \tag{1.100}$$

où λ_B représente le nombre moyen de cibles naissantes distribuées selon la densité de probabilité $p_B(\cdot)$. Notons que la somme dans (1.100) permet de parcourir toutes les hypothèses d'association $\theta : \{1, ..., N(k)\} \to \{0, ..., N(k)\}$, avec $\theta(i) > 0$ impliquant que la cible a survécu entre les instants $k-1$ et k alors que $\theta(i) = 0$ correspond à la disparition de la cible ou au fait qu'elle n'était pas présente à l'instant précédent.
Ajoutons également que la probabilité que toutes les cibles de l'EFA prédit correspondent à des cibles naissantes est :

$$\pi_B(\boldsymbol{X}_k) = e^{-\lambda_B} \prod_{i=1}^{N(k)} \lambda_B p_B(\boldsymbol{x}_k^{(i)}) \tag{1.101}$$

1.4.3.2 Modèle d'observation multi-cibles

Considérons qu'une cible d'état $\boldsymbol{x}_k \in X_k$ à l'instant k soit détectée avec la probabilité $P_D(\boldsymbol{x}_k)$ et génère une observation $\boldsymbol{y}_k$ de vraisemblance $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ ou alors n'est pas détectée avec la probabilité $1 - P_D(\boldsymbol{x}_k)$ et génère la valeur $\emptyset$. Etant donné la mesure $\boldsymbol{y}_k$ associée à l'état $\boldsymbol{x}_k \in X_k$ à l'instant k, son comportement est modélisé par l'EFA de Bernoulli suivant :

$$\Theta_k(\boldsymbol{x}_k) = \{P_D(\boldsymbol{x}_k), p(\boldsymbol{y}_k|\boldsymbol{x}_k)\} \tag{1.102}$$

2. Par souci de simplicité, on ne considère pas dans ce manuscrit l'EFA des cibles non-résolues.

On peut alors modéliser la détection et la génération des mesures à l'instant courant par l'EFA Y_k qui satisfait :

$$Y_k = \left(\bigcup_{\boldsymbol{x}_k \in X_k} \Theta_k(\boldsymbol{x}_k) \right) \cup K_k \tag{1.103}$$

où K_k est l'EFA des fausses alarmes à l'instant k. Selon le type de filtre, il est caractérisé par un EFA de Poisson, un EFA de cluster i.i.d ou encore un EFA multi-Bernoulli. Enfin, on suppose que les EFAs Θ_k et K_k sont indépendants [Vo08].

Fonction de vraisemblance :
L'équation (1.103) peut également se mettre sous la forme d'une vraisemblance multi-cibles qui évalue la densité de probabilité $p(Y_k|X_k)$ traduisant qu'à l'instant k l'état multi-cibles X_k génère l'EFA d'observations multi-cibles Y_k. Contrairement à la fonction de vraisemblance monocible, la fonction de vraisemblance multi-cibles doit représenter l'ensemble des mesures issues de cibles et les fausses alarmes.
Notons que du fait des fausses alarmes, de l'incertitude de détection ainsi que du bruit de mesure sur la position des mesures provenant de cibles, l'association des mesures aux pistes est ambigüe. Le formalisme des EFAs permet de tester toutes les hypothèses d'associations.
Ainsi la fonction de vraisemblance multi-cibles est calculée pour toutes les hypothèses θ : $\{1, ..., N(k)\} \rightarrow \{0, ..., M(k)\}$ pour $N(k)$ cibles et $M(k)$ mesures où 0 représente l'hypothèse de non-détection.
Si le nombre de fausses alarmes suit une distribution de type Poisson notée $\pi_c(\cdot)$ de paramètre λ_c et les fausses alarmes sont distribuées selon une distribution $\mathcal{U}(\cdot)$ qui est généralement choisie comme uniforme, la fonction de vraisemblance multi-cibles a l'expression suivante :

$$p(\boldsymbol{Y}_k|\boldsymbol{X}_k) = \pi_c(\boldsymbol{Y}_k) \prod_{i=1}^{N(k)} \left(1 - P_D(\boldsymbol{x}_k^{(i)})\right) \sum_{\theta} \prod_{j:\theta(j)>0} \frac{P_D(\boldsymbol{x}_k^{(j)}) p(\boldsymbol{y}_k^{(\theta(j))}|\boldsymbol{x}_k^{(j)})}{(1 - P_D(\boldsymbol{x}_k^{(j)}))\lambda_c \mathcal{U}(\boldsymbol{y}_k^{(\theta(j))})} \tag{1.104}$$

avec

$$\pi_c(\boldsymbol{Y}_k) = e^{-\lambda_c} \prod_{\boldsymbol{y}_k \in \boldsymbol{Y}_k} \lambda_c \mathcal{U}(\boldsymbol{y}_k) \tag{1.105}$$

Notons que pour une hypothèse où la probabilité de détection ne dépend pas de l'état, soit $P_D(\boldsymbol{x}_k^{(i)}) = P_D$, alors la fonction de vraisemblance multi-cibles est égale, à une constante près, à la vraisemblance utilisée au sein de l'algorithme d'association MHT, Cf. annexe A.

Dans la sous-partie suivante, nous détaillons le filtre de Bayes multi-cibles ainsi que les filtres PHD et CPHD qui s'appuient sur le formalisme précédemment introduit.

1.5 Filtrage et estimation bayésienne multi-cibles

Le fondement mathématique des méthodes de filtrage développées dans la suite repose sur la théorie des ensembles aléatoires introduite initialement par Matheron. A partir de cette approche mais également de la théorie des mesures, Mahler [Mah01] développa la théorie de la statistique des ensembles finis afin d'étendre les approches bayésiennes du cas monocible aux cas multi-

cibles et multi-senseurs. Les observations ainsi que les états des cibles sont représentés par des EFAs à partir desquels il est possible de formaliser un filtre bayésien multi-cibles multi-senseurs optimal : le filtre de Bayes multi-cibles. Ce dernier ne pouvant pas directement être mis en œuvre, les filtres PHD et CPHD ont entre autres été développés. Dans la suite, nous présentons ces filtres en nous appuyant sur les travaux suivants [Goo97, Mah01, Mah07, Vo05b, Pan07].

1.5.1 Filtre de Bayes multi-cibles

On peut montrer que la problématique de filtrage multi-cibles peut être traitée à partir d'une approche bayésienne, d'espace d'état multi-cibles $\mathcal{F}(\mathbb{X})$ et d'espace d'observation multi-cibles $\mathcal{F}(\mathbb{Y})$.
Rappelons qu'à chaque instant, toute l'information liée à l'état du système est contenue dans la densité de probabilité *a posteriori* multi-cibles notée $p(X_k|Y_{1:k})$, conditionnée par l'ensemble des observations reçues jusqu'à l'instant courant.
Considérons les densités de probabilités multi-cibles prédite et *a posteriori* à l'instant k notées $p(X_k|Y_{1:k-1})$ et $p(X_k|Y_{1:k})$ respectivement. Dans ce cas, la récursion du filtre de Bayes multi-cibles est constituée des équations de prédiction et de mise à jour suivantes :

$$p(X_k|Y_{1:k-1}) = \int p(X_k|X_{k-1})p(X_{k-1}|Y_{1:k-1})\delta X_{k-1} \tag{1.106}$$

et

$$p(X_k|Y_{1:k}) = \frac{p(Y_k|X_k)p(X_k|Y_{1:k-1})}{\int p(Y_k|X)p(X|Y_{1:k-1})\delta X} \tag{1.107}$$

où (1.106) est une intégrale d'ensemble de la forme (1.85).
L'interêt de cette approche est qu'elle évite l'étape d'association explicite des mesures aux pistes. Cependant, le calcul des densités est difficilement implémentable. Il a donc été nécessaire de développer des alternatives comme par exemple le filtre PHD.

1.5.2 Filtre *probability hypothesis density* (PHD)

Le filtre *Probability Hypothesis Density* [Mah03, Tob06, Vo05a] est une alternative sous-optimale du filtre de Bayes multi-cibles. Il approche la densité de probabilité multi-cibles *a posteriori* par son moment statistique d'ordre un, appelé fonction d'intensité [Mah03, Mah07], et le propage récursivement.
Dans le contexte du pistage de cibles, la fonction d'intensité est plus communément appelée densité d'hypothèse de probabilité ou *probability hypothesis density* (PHD) en anglais. Le PHD, qui est défini sur l'espace d'état monocible $\mathbb{X}$, n'est pas une distribution de probabilité puisque son intégrale sur cet espace ne vaut pas l'unité mais représente le nombre moyen de cibles à l'instant courant, soit :

$$\hat{N} = \int_{\mathbb{X}} v(\boldsymbol{x})d\boldsymbol{x} \tag{1.108}$$

où $\hat{N}$ est l'estimation du nombre moyen de cibles.
De façon plus générale, $\int_\Omega v(\boldsymbol{x})d\boldsymbol{x}$ représente le nombre moyen de cibles dans la région Ω.
Les maxima locaux de la fonction d'intensité $v(\cdot)$ représentent donc les régions de $\mathbb{X}$ avec une forte concentration de cibles.

Dans la suite, nous détaillons les étapes du filtre PHD.

Considérons que $v_{k|k-1}$ et $v_{k|k}$ représentent les fonctions d'intensité associées aux états multi-cibles *a priori* et *a posteriori* respectivement. Si l'on formule les hypothèses suivantes :

- 1/ les cibles évoluent indépendamment les unes des autres et génèrent des mesures indépendantes.
- 2/ les EFAs multi-cibles prédit et *a posteriori* sont approchés par des EFAs du type Poisson.

Alors les équations du filtre PHD sont les suivantes :

$$v_{k|k-1}(\boldsymbol{x}_k) = \int P_S(\boldsymbol{x}_{k-1})p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})v_{k-1|k-1}(\boldsymbol{x}_{k-1})d\boldsymbol{x}_{k-1} + \gamma_k(\boldsymbol{x}_k) \tag{1.109}$$

$$v_{k|k}(\boldsymbol{x}_k) = (1 - P_D(\boldsymbol{x}_k))v_{k|k-1} + \sum_{\boldsymbol{y}\in Y_k} \frac{P_D(\boldsymbol{x}_k)p(\boldsymbol{y}|\boldsymbol{x}_k)v_{k|k-1}(\boldsymbol{x}_k)}{\kappa_k(\boldsymbol{y}) + \langle P_D p(\boldsymbol{y}|\cdot), v_{k|k-1}\rangle} \tag{1.110}$$

avec $\kappa_k(\cdot)$ la fonction d'intensité de l'EFA représentant le clutter et $\gamma_k(\cdot)$ la fonction d'intensité de l'EFA représentant les naissances de cibles.
Le nombre de cibles présentes dans l'espace de surveillance peut donc être déduit à chaque instant comme suit :

$$\hat{N}(k) = \int v_{k|k}(\boldsymbol{x})d\boldsymbol{x} \tag{1.111}$$

Notons que la première hypothèse est commune dans la plupart des applications multi-cibles, mais la seconde est une hypothèse simplificatrice utilisée afin de faciliter le dévelopement des équations dans la démonstration du filtre PHD. De plus, comme un EFA de Poisson a pour cardinal moyen $N = \int v(\boldsymbol{x})d\boldsymbol{x}$ et ses éléments sont i.i.d selon v/N, il est entièrement caractérisé par sa fonction d'intensité $v(\cdot)$.

Grâce à l'approximation à l'ordre un de la densité multi-cibles, le filtre PHD ne nécessite pas le calcul fastidieux d'intégrales d'ensemble. Les équations du filtre PHD sont donc plus simples à calculer que celles du filtre de Bayes multi-cibles. Néanmoins, elles reposent sur l'évaluation d'intégrales multidimensionnelles. Afin d'implémenter cet algorithme à moindre coût, des solutions approchées ont été développées. Il s'agit des filtres *gaussian-mixture* PHD (GM-PHD) [Vo03, Vo06, Cla06, Pac10] et *sequential Monte Carlo* PHD (SMC-PHD) [Vo05b, Sid03, Pun05, Pan05].

Dans le cas de l'implémentation SMC-PHD, l'intensité *a posteriori* est approchée par :

$$v_{k|k}(\boldsymbol{x}) = \sum_i w^{(i)}\delta_{\boldsymbol{x}^{(i)}}(\boldsymbol{x}) \tag{1.112}$$

avec $\sum_i w^{(i)} = \hat{N}$, le nombre estimé de cibles. Contrairement à l'implémentation SMC dans le cas monocible, les poids des particules $w^{(i)}$ ne sont pas normalisés puisque leur somme permet d'évaluer le nombre de cibles présentes.

Concernant l'implémentation GM-PHD, de façon similaire au filtre somme de gaussiennes dans le cas monocible, la fonction d'intensité est approchée par un mélange de gaussiennes

comme suit :

$$v_{k|k}(\boldsymbol{x}) = \sum_i w^{(i)} \mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{x}}^{(i)}, P^{(i)}) \tag{1.113}$$

avec $\hat{\boldsymbol{x}}^{(i)}$ la moyenne de la $i^{\text{ème}}$ distribution gaussienne et $P^{(i)}$ la matrice de covariance associée. Comme précédemment les poids ne sont pas normalisés mais permettent de calculer le nombre de cibles présentes, i.e. $\sum_i w^{(i)} = \hat{N}$. Notons que comme pour le filtre de Kalman, le filtre GM-PHD suppose que les modèles d'évolution et de mesure sont gaussiens et linéaires. De plus, les probabilité de survie et de détection sont supposées ne pas dépendre de l'état de la cible, soit $P_S(x) = P_S$ et $P_D(x) = P_D$.

Comme introduit dans la partie suivante, à partir du nombre de cibles estimé selon (1.111) et de la fonction d'intensité *a posteriori* $v_{k|k}$, il est possible d'extraire à chaque instant l'état multi-cibles qui se compose des $\hat{N}(k)$ maxima locaux de $v_{k|k}$. Néanmoins, l'un des inconvénients du filtre PHD est qu'il ne permet pas d'associer les états d'un instant à un autre afin d'extraire des trajectoires de cibles. En effet, d'un point de vue pratique, le filtre PHD ne permet pas à l'utilisateur d'identifier quels états proviennent de quelles cibles et donc de créer des pistes. Cependant, des extensions de ces filtres ont été proposées pour pallier cette faiblesse [Cla05, Pan05, Pan06b].
De plus, le filtre PHD est instable quant à l'estimation du nombre de cibles dans certains contextes [Erd05] notamment lorsque la probabilité de détection est faible. C'est la conséquence du fait que le PHD ne propage que le nombre moyen de cibles. Cela a conduit au dévelopement du filtre CPHD qui propage le moment statistique d'ordre supérieur.

1.5.3 Filtre *cardinalized probability hypothesis density* (CPHD)

Dans [Erd05], Erdinc met en avant que le filtre PHD présente des faiblesses d'estimation du nombre de cibles en raison de l'utilisation du premier ordre de la densité jointe uniquement. Le filtre CPHD [Mah07] permet de pallier ce problème en propageant conjointement à la fonction d'intensité $v(x)$, une distribution de probabilité $\rho(n)$ modélisant l'incertitude sur l'estimation du nombre de cibles présentes [Mah07, Vo07].
Dans cette sous-partie, le système considéré présente les mêmes caractéristiques que pour le filtre PHD, à savoir que nous considérons la modélisation du système multi-cibles définie par les équations (1.99) et (1.103).
Les hypothèses d'utilisation de ce filtre sont identiques à celles utilisées pour le filtre PHD excepté que les EFAs sont approximés par des clusters i.i.d, plutôt que des EFAs de Poisson pour le PHD.

Notons $v_{k|k-1}$ et $\rho_{k|k-1}(n) = Pr[N = n|\boldsymbol{y}_{1:k-1}]$ la fonction d'intensité *a priori* et la distribution de cardinalité *a priori* respectivement, associées à la densité d'état multi-cibles. De façon similaire, $v_{k|k}$ et $\rho_{k|k}(n) = Pr[N = n|\boldsymbol{y}_{1:k}]$ désignent la fonction d'intensité *a posteriori* et la distribution de cardinalité *a posteriori* respectivement, associées à la densité d'état multi-cibles.
La fonction d'intensité prédite du filtre CPHD est équivalente à celle du filtre PHD, soit :

$$v_{k|k-1}(\boldsymbol{x}_k) = \int_{\xi \in \mathbb{X}} P_S(\xi) p(\boldsymbol{x}_k|\xi) v_{k-1|k-1}(\xi) d\xi + \gamma_k(\boldsymbol{x}_k) \tag{1.114}$$

et la distribution de cardinalité *a priori* $\rho_{k|k-1}$ s'exprime comme suit :

$$\rho_{k|k-1}(n) = \sum_{j=0}^{n} \rho_{\Gamma,k}(n-j) \sum_{l=0}^{+\infty} C_j^l \frac{\langle P_S, v\rangle^j \langle 1-P_S, v\rangle^{l-j}}{\langle 1, v\rangle^l} \rho_{k-1|k-1}(l) \tag{1.115}$$

avec $\rho_{\Gamma,k}(.)$, la distribution de cardinalité des cibles naissantes et $C_j^l = \frac{l!}{j!(l-j)!}$ le coefficient binomiale.

On peut exprimer la fonction d'intensité $v_{k|k}$ et la distribution de cardinalité $\rho_{k|k}$ *a posteriori* en fonction de $v_{k-1|k}$ et $\rho_{k-1|k}$ comme suit :

$$\rho_{k|k}(n) = \frac{\Upsilon_k^0[v_{k|k-1}; Y_k](n)\rho_{k|k-1}(n)}{\langle \Upsilon_k^0[v_{k|k-1}; Y_k], \rho_{k|k-1}\rangle} \tag{1.116}$$

$$\begin{aligned} v_{k|k}(\boldsymbol{x}) = (1-P_D(\boldsymbol{x}))&\frac{\langle \Upsilon_k^1[v_{k|k-1}; Y_k], \rho_{k|k-1}\rangle}{\langle \Upsilon_k^0[v_{k|k-1}; Y_k], \rho_{k|k-1}\rangle} v_{k|k-1}(\boldsymbol{x}) \\ &+ \sum_{\boldsymbol{y}\in Y_k} \psi_{k,\boldsymbol{y}}(\boldsymbol{x}) \frac{\langle \Upsilon_k^1[v_{k|k-1}; Y_k - \{\boldsymbol{y}\}], \rho_{k|k-1}\rangle}{\langle \Upsilon_k^0[v_{k|k-1}; Y_k], \rho_{k|k-1}\rangle} v_{k|k-1}(\boldsymbol{x}) \end{aligned} \tag{1.117}$$

avec :

$$\Upsilon_k^u[v, Y](n) = \sum_{j=0}^{min(|Y|,n)} (|Z|-j)!\rho_c(|Y|-j)P_{j+u}^n \frac{\langle 1-P_D, v\rangle^{n-(j+u)}}{\langle 1, v\rangle^n} e_j\left(\Xi(v, Y)\right) \tag{1.118}$$

la fonction de vraisemblance pour un ensemble de mesure Y. ρ_c représente la distribution de cardinalité des mesures issues du clutter et $P_j^n = \frac{n!}{(n-j)!}$ est le coefficient de permutation.
La vraisemblance de la mesure peut s'exprimer à partir de la fonction d'intensité du clutter $\kappa_k(\cdot)$ et de la probabilité de détection $P_D(\cdot)$ comme suit :

$$\psi_{\boldsymbol{y}}(\boldsymbol{x}) = \frac{\langle 1, \kappa_k\rangle}{\kappa_k(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{x}) P_D(\boldsymbol{x}) \tag{1.119}$$

De plus, par définition on considère :

$$\Xi(v, Y) = \{\langle v, \psi_{k,\boldsymbol{y}}\rangle : \boldsymbol{y} \in Y\} \tag{1.120}$$

et $e_j(Y)$ représente la fonction symmétrique élémentaire d'ordre j et est définie comme suit :

$$e_j(Y) = \sum_{S\subseteq Y, |S|=j} \left(\prod_{\zeta\in S} \zeta\right) \tag{1.121}$$

Notons que la fonction d'intensité de prédiction du filtre CPHD (1.114) est la même que pour le filtre PHD (1.109). La cardinalité (1.115) et la fonction d'intensité de prédiction (1.114) ne sont pas couplées contrairement à la cardinalité (1.116) et la fonction d'intensité de mise à jour (1.117). De plus, (1.116) peut s'interpréter comme une mise à jour bayésienne avec $\Upsilon_k^0[v_{k|k-1}; Y_k](n)$ la vraisemblance des observations multi-cibles Y_k sachant qu'il y a n cibles, et $\langle \Upsilon_k^0[v_{k|k-1}; Y_k](n), \rho_{k|k-1}\rangle$ une constante de normalisation.

Par ailleurs, le nombre de cibles présentes dans la zone d'intérêt à chaque instant peut être calculée soit comme pour le filtre PHD à partir de (1.111) soit de la manière suivante :

$$\hat{N}(k) = \underset{n}{argsup}\ \rho_{k|k}(n)d\boldsymbol{x} \tag{1.122}$$

Enfin, de façon similaire au filtre PHD, le filtre CPHD peut être implémenté en tirant profit de méthodes SMC [Mah07] ou de mélanges de gaussiennes [Vo07]. Comparativement au filtre PHD, le filtre CPHD fournit une estimation plus précise mais également plus stable de la cardinalité grâce à la propagation de la distribution de cardinalité.

Quel que soit le filtre considéré, PHD ou CPHD, les critéres d'évaluation de performances définis dans un contexte monocible tels que l'EQM ne sont pas applicables dans un contexte d'estimation multi-cibles. En effet, d'autres mesures de performances doivent être considérées pour les algorithmes multi-cibles estimant un ensemble d'état et le nombre de cibles.
Dans la suite, nous présentons les solutions existantes.

1.6 Evaluation de performances pour les algorithmes de pistage multi-cibles

1.6.1 Estimateurs d'états multi-cibles

Une fois la densité de probabilité multi-cibles obtenue, il faut être en mesure d'extraire l'état des cibles à chaque instant. Pour cela, un estimateur d'état multi-cibles est nécessaire.
Dans la suite, considérons que l'on souhaite estimer l'état multi-cibles X_k à partir de sa densité de probabilité jointe *a posteriori* $p(X_k|Y_{1:k})$ ou à partir de la densité de probabilité *a posteriori* de sa cardinalité $\rho_{k|k}(\cdot)$ et de la fonction d'intensité associée $v_{k|k}(\cdot)$. Il existe alors deux possibilités permettant d'extraire l'état multi-cibles selon que l'on utilise $p(X_k|Y_{1:k})$ ou $v_{k|k}(\cdot)$ et $\rho_{k|k}(\cdot)$.

Visualisation à partir du moment d'ordre 1 :
La fonction d'intensité d'un EFA représentant en un point donné le nombre moyen de cibles, les maxima de cette fonction indiquent les lieux de forte concentration du nombre moyen de cibles. Le nombre de cibles peut être calculé à partir de (1.111) ou (1.122) selon que l'on utilise un algorithme PHD ou CPHD. A partir du nombre estimé de cibles, il s'agit ensuite d'extraire les $\hat{N}(k)$ vecteurs d'état $\hat{\boldsymbol{x}}^{(1)}, ..., \hat{\boldsymbol{x}}^{(\hat{N}(k))}$ tels que $v_k(\hat{\boldsymbol{x}}^{(1)}), ..., v_k(\hat{\boldsymbol{x}}^{(\hat{N}(k))})$ soient les $\hat{N}$ plus grandes valeurs de $v(\cdot)$.
L'état multi-cibles satisfait alors $X_k = \{\hat{\boldsymbol{x}}^{(1)}, ..., \hat{\boldsymbol{x}}^{(\hat{N}(k))}\}$.

Estimateur multi-cibles marginal :
Cet estimateur permet d'estimer l'état multi-cibles en deux étapes. Dans un premier temps, comme précédemment le nombre de cibles est estimé à partir de (1.111) ou (1.122). Puis, les états sont extraits en utilisant la densité de probabilité jointe *a posteriori* de l'état des cibles :

$$X_k = \underset{X:|X|=\hat{N}(k)}{argsup}\ p(X_k|Y_{1:k}) \tag{1.123}$$

Dans la suite, nous présentons les moyens disponibles permettant d'évaluer la précision d'un algorithme multi-cibles.
En pistage monocible, afin d'évaluer l'erreur commise lors de l'estimation de l'état du système, on utilise classiquement un calcul de distance entre l'état estimé et l'état théorique. Pour cela, la mesure communément utilisée est l'erreur quadratique moyenne. Néanmoins, les filtres multi-cibles étant fondés sur les EFAs, ils ne prennent pas en compte l'ordre des cibles. De ce fait, il a été nécessaire de redéfinir des critères d'évaluation des estimations fournies.
Pour évaluer les performances des algorithmes de filtrage multi-cibles, le concept de distance multi-cibles a été introduit. Il existe plusieurs distances utilisées dans la littérature scientifique pour mesurer la différence entre deux ensembles. Historiquement, la première introduite a été la distance de Hausdorff dont les perfomances sont jugées insatisfaisantes du fait qu'elle ne soit pas sensible aux différences de cardinalités des ensembles. Hoffman et Mahler ont alors proposé une métrique appelée *optimal mass transfer* (OMAT) fondée sur la distance de Wasserstein qui permet en partie de résoudre le comportement non désiré de la métrique de Hausdorff [Hof04]. Enfin, Schumacher *et al.* [Sch08] ont montré que la métrique proposée par Mahler souffre de quelques inconvénients et ont alors suggéré une nouvelle métrique appelée distance *optimal sub-pattern assignment* (OSPA) que nous présentons dans la suite.

1.6.2 Distance OSPA

Afin de définir la distance OSPA, on introduit la distance notée $d^{(c)}$ entre deux vecteurs $\boldsymbol{x} \in W$ et $\boldsymbol{y} \in W$ telle que :

$$d^{(c)}(\boldsymbol{x}, y) = min(c, d(\boldsymbol{x}, y)) \tag{1.124}$$

avec $c \geq 0$ le paramètre de coupure.
Considérons deux ensembles finis $X = \{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(n)}\}$ et $Y = \{\boldsymbol{y}^{(1)}, ..., \boldsymbol{y}^{(m)}\}$ tels que $m \leq n$. Si $\mathcal{P}_k$ dénote l'ensemble des permutations sur $\{1, 2, ..., k\}$ pour tout $k \in \mathbb{N}$, alors pour $1 \leq p \leq +\infty$, la distance OSPA d'ordre p et de paramètre de coupure c est définie comme suit :

$$d_p^{(c)}(X, Y) = \left(\frac{1}{n} \left(\underset{t \in \mathcal{P}_n}{min} \sum_{i=1}^{m} d^c(\boldsymbol{x}^{(i)}, y^{t(i)})^p + c^p(n-m) \right) \right)^{\frac{1}{p}} \tag{1.125}$$

Notons que si $m \geq n$, il suffit d'échanger m et n dans (1.125).
De plus, pour $p = +\infty$, on a :

$$d_\infty^{(c)}(X, Y) = \begin{cases} \underset{t \in \mathcal{P}_n}{min} \underset{1 \leq i \leq n}{max} d^{(c)}(\boldsymbol{x}^{(i)}, y^{t(i)}) \text{ si } m = n \\ = c \text{ autrement} \end{cases} \tag{1.126}$$

Il est démontré dans [Sch08] que $d_p^{(c)}(X, Y)$ vérifie bien les propriétés d'une distance.
Notons que le paramètre c détermine selon sa valeur l'importance que l'on donne aux erreurs d'estimation de cardinal. Des valeurs faibles de c mettent en avant des erreurs d'estimation sur la position tandis que la métrique est peu sensible aux erreurs sur le cardinal. En revanche, pour des valeurs de c plus importantes, la métrique porte plus intérêt à l'estimation du cardinal de

l'ensemble fini qu'aux positions. Une valeur de c qui serait de l'ordre de grandeur de l'écart-type du bruit de mesure est considérée comme faible, alors qu'une valeur de c qui serait de l'ordre de la distance maximale qui sépare deux cibles est considérée comme grande. L'influence de la valeur de ce paramètre sur l'évaluation des performances a été étudiée dans [Sch08].
Enfin, une valeur élevée pour l'ordre p donne plus de poids aux grandes distances, ce qui a tendance a plus pénaliser les *outliers*.

1.7 Conclusions

Dans ce chapitre traitant des généralités sur le pistage de cibles, nous avons dans un premier temps décrit les principes de base de l'estimation bayésienne pour le pistage monocible. Nous traitons plus précisément de la représentation du système dans l'espace d'état puis des filtres bayésiens tels que le filtre de Bayes, le filtre de Kalman et le filtre particulaire en rappelant pour ce dernier les difficultés de mise en œuvre, les problèmes de dégénérescence du nuage de particules et les choix possibles sur la densité d'importance. Les algorithmes à modèles multiples, utilisés pour la poursuite de cibles manœuvrantes, sont ensuite introduits. En particulier, nous mettons l'accent sur l'algorithme IMM, sa mise en œuvre et les différentes problématiques associées.
Dans un second temps, nous abordons les principes du pistage multi-cibles. Les filtres PHD et CPHD, fondés sur la théorie des EFAs, sont détaillés. Enfin, les différentes mesures de performance utilisées en pistage multi-cibles sont rappelées.

Les performances d'estimation des différentes approches de filtrage présentées dans ce chapitres sont étroitement liées au choix du ou des modèles sélectionnés pour effectuer l'estimation. Ce choix est subjectif puisqu'il est essentiellement guidé par la connaissance *a priori* de l'utilisateur sur le système. Pour un algorithme d'estimation à modèles multiples combinant plusieurs modèles d'état, il est préférable que ces derniers soient les plus dissimilaires possible afin de représenter une large diversité des types de trajectoires de la cible.
Aussi, dans le chapitre 2, nous évaluons la pertinence d'une approche statistique fondée sur une mesure de dissimilarité entre deux modèles d'état. Cette étude est notamment mise à profit pour comparer entre eux plusieurs modèles d'état afin d'en sélectionner un nombre réduit à associer au sein d'un algorithme à modèles multiples.

Chapitre 2

Comparaison de modèles d'état fondée sur la divergence de Jeffrey

Sommaire

2.1 Introduction **42**

2.2 Rappels **43**

2.2.1 Modèles de mouvement 43

2.2.2 Divergences de Kullback-Leibler (KL) et de Jeffrey (JD) 46

2.3 JD entre deux modèles d'état **47**

2.3.1 Calcul récursif 47

2.3.2 Analyse des termes additifs dans l'expression récursive de la JD 49

2.3.3 JD dans différentes bases de l'espace d'état 50

2.4 Classification d'un ensemble de M modèles d'état fondée sur la JD **51**

2.4.1 Structuration en matrice des JDs entre modèles d'état 52

2.4.2 Analyse fondée sur une décomposition en valeurs propres 54

2.4.3 Application à la conception d'algorithmes à modèles multiples 56

2.5 Simulations et résultats **58**

2.5.1 Application à l'estimation de paramètres : comparaison de $M = 2$ marches aléatoires 58

2.5.2 Application à la poursuite de cibles : comparaison de $M = 2$ modèles d'état 59

2.5.3 Application à la poursuite de cibles : classification de $M = 6$ modèles d'état 63

2.6 Conclusions et perspectives **68**

2.1 Introduction

Afin de traiter les problématiques d'estimation séquentielle, des approches bayésiennes comme le filtre de Kalman et ses variantes [Kal60] ou le filtre particulaire [Ver02] sont communément utilisées. Leurs performances en termes d'estimation sont étroitement liées au choix du modèle d'état. Par exemple, pour le pistage de cibles manœuvrantes, le mouvement de la cible est en général inconnu et peut changer au cours du temps. La cible peut adopter un mouvement rectiligne uniforme (MRU) [Li03, Bar93], puis un mouvement uniformément accéléré (MUA) [Li03, Bar93], puis un mouvement de Singer [Sin70], etc. De plus, quel que soit le mouvement de la cible, les paramètres dynamiques tels que la variance du jerk ou de l'accélération peuvent varier au cours du temps. Dans ce cas, utiliser un estimateur fondé sur une unique REE peut ne pas suffire à caractériser pleinement le système.
Afin de relâcher cette hypothèse sur le modèle *a priori*, les modèles multiples ont été développés. Or, comme nous l'avons indiqué dans le chapitre précédent, il n'existe pas de méthodes permettant de guider l'utilisateur dans le choix des modèles à associer au sein d'une structure à modèles multiples, même si l'un des principes est de considérer des modèles dissimilaires afin d'être le plus représentatif possible de la diversité des mouvements des cibles pistées.

Dans ce chapitre, nous proposons et évaluons la pertinence d'une approche statistique fondée sur la divergence de Jeffrey (JD) [Jef48], qui est la version symétrique de la divergence de KL [Kul51], afin de mesurer les dissimilarités entre modèles d'état. En particulier, nous proposons de calculer la JD entre les distributions jointes des vecteurs d'état successifs associés aux deux modèles d'état considérés [Mag13]. Dans un cas général, il est difficilement envisageable de calculer directement l'expression de la JD. Cependant, nous tirons profit des propriétés markoviennes et de gaussianité des modèles à comparer pour déduire une expression récursive de la JD. Ainsi, au lieu d'un calcul direct qui dépendrait de matrices de taille croissante, la formulation récursive a l'avantage de n'inverser que des matrices de taille fixe et de même dimension que le vecteur d'état.
Nous étendons ensuite notre étude à un groupe de modèles afin d'en identifier des sous-ensembles à partir des JDs calculées entre les modèles deux à deux [Mag14].
Enfin, nous appliquons cette approche dans le cadre de la sélection de modèles *a priori* et montrons sa pertinence pour la poursuite de cibles par modèles multiples [Mag15d].

Avant d'aborder la JD, nous rappelons les différents modèles de mouvement considérés en pistage et que l'on utilise dans la suite de ce mémoire.

2.2 Rappels

Dans les domaines du contrôle ou du traitement du signal, les approches récursives bayésiennes sont couramment utilisées. Elles sont fondées sur la REE du système qui est un modèle mathématique du système dynamique [Bar11, Hau12]. Comme rappelé dans le chapitre 1, la REE se compose classiquement de deux équations : la première traduit l'évolution du système au cours du temps tandis que la seconde décrit la relation entre le vecteur d'état et le vecteur d'observation à l'instant courant. Dans la suite, nous nous intéressons au cas où le système est linéaire gaussien et invariant par morceaux. Plus précisément, il peut commuter entre M modèles différents dont le $i^{\text{ème}}$ est supposé défini comme suit :

$$\boldsymbol{x}_k = F^i \boldsymbol{x}_{k-1} + \boldsymbol{v}_k^i \tag{2.1}$$

$$\boldsymbol{y}_k = H^i \boldsymbol{x}_k + \boldsymbol{w}_k^i \tag{2.2}$$

où $\boldsymbol{x}_k$ désigne le vecteur d'état à l'instant k, de taille n_x, et F^i la matrice de transition. Le bruit de modèle $\boldsymbol{v}_k^i$ est supposé blanc gaussien à moyenne nulle et de matrice de covariance Q^i. Dans la suite de ce chapitre, Q^i est supposée inversible ce qui équivaut raisonnablement à supposer que la densité de transition est non dégénérée. Le vecteur $\boldsymbol{y}_k$ désigne l'observation et H^i la matrice d'observation. De plus, $\boldsymbol{w}_k^i$ est un bruit blanc gaussien centré non corrélé avec $\boldsymbol{v}_k^i$ et de matrice de covariance R^i. Notons que la formulation probabiliste des équations (2.1) et (2.2) satisfait :

$$p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k; F^i\boldsymbol{x}_{k-1}, Q^i) \tag{2.3}$$

$$p^i(\boldsymbol{y}_k|\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{y}_k; H^i\boldsymbol{x}_k, R^i) \tag{2.4}$$

Dans le contexte de la poursuite d'une cible, la REE permet de modéliser le mouvement suivi par la cible, c'est-à-dire non seulement sa position mais aussi sa dynamique. Ce point est illustré en figure 2.1, où le vecteur d'état contient la position et la vitesse de la cible à chaque instant dans un plan 2D. Les observations collectées par le système radar peuvent correspondre à des versions bruitées des positions en coordonnées polaires ou cartésiennes.

En pratique, différents modèles de mouvement existent. Dans la suite, nous détaillons la REE des modèles de mouvement considérés dans le cadre de cette thèse.

2.2.1 Modèles de mouvement

Trois modèles de mouvement sont envisagés dans cette thèse [1] : le mouvement rectiligne uniforme (MRU), le mouvement uniformément accéléré (MUA) et le mouvement de Singer.

Modèle rectiligne uniforme :
Dans ce cas, le vecteur d'état est défini comme suit :

$$\boldsymbol{x}_k = [x_k, \dot{x}_k]^T \tag{2.5}$$

où x_k correspond à la position et $\dot{x}_k$ la vitesse. Notons que dans le cas général, le vecteur d'état

1. Notons qu'il en existe d'autres comme le mouvement curviligne ou le mouvement rotationnel par exemple [Bar95].

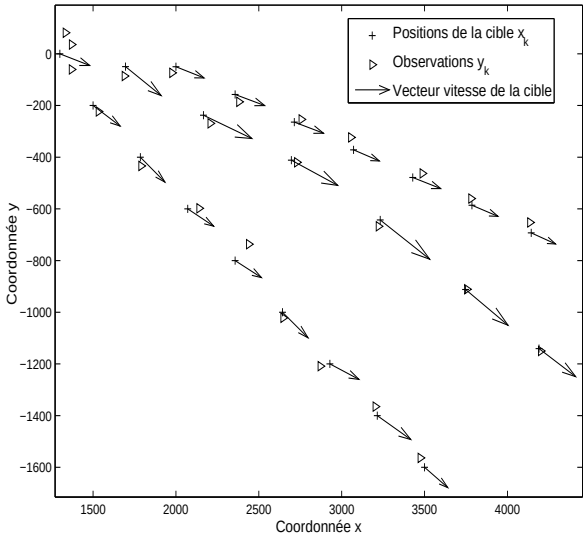

FIGURE 2.1 – Exemple de trois réalisations générées selon une REE donnée, dans le cadre de la modélisation du mouvement

peut également contenir une ou deux dimensions supplémentaires selon les axes y et z. Cependant, dans la suite, pour faciliter la lecture, nous nous concentrons sur le cas monodimensionnel.

La REE du MRU est caractérisée par sa matrice de transition et la matrice de covariance de son bruit de modèle. Elles sont définies comme suit :

$$F^{MRU} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \tag{2.6}$$

$$Q^{MRU} = \sigma^2_{MRU} \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \tag{2.7}$$

où T est la période d'échantillonnage et σ^2_{MRU} la variance de l'accélération.

Le MRU est classiquement choisi pour modéliser le comportement d'une cible se déplaçant en ligne droite à vitesse quasi-constante.

<u>Modèle uniformément accéléré :</u>
Le vecteur d'état est défini dans ce cas comme suit :

$$\boldsymbol{x}_k = [x_k, \dot{x}_k, \ddot{x}_k]^T \tag{2.8}$$

où $\ddot{x}_k$ est la composante accélération.

La matrice de transition et la matrice de covariance du bruit de modèle pour le MUA sont

les suivantes :

$$F^{MUA} = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

$$Q^{MUA} = \sigma^2_{MUA} \begin{bmatrix} T^5/20 & T^4/8 & T^3/6 \\ T^4/8 & T^3/3 & T^2/2 \\ T^3/6 & T^2/2 & T \end{bmatrix} \tag{2.10}$$

où σ^2_{MUA} est la variance sur le jerk qui correspond à la dérivée de l'accélération.

Le MUA est classiquement utilisé pour modéliser le comportement d'une cible en phase d'accélération ou de décélération.

Modèle de Singer

Le modèle de Singer suppose que l'accélération correspond à un processus de Markov d'ordre un tel que sa fonction d'autocorrélation puisse s'écrire comme suit [Sin70] :

$$R_a(\tau) = E[\ddot{x}_{k+\tau}\ddot{x}_k] = \sigma^2_{Sin} e^{-\alpha_{Sin}|\tau|} \tag{2.11}$$

avec σ^2_{Sin} la variance sur l'accélération et $\alpha = \frac{1}{\tau_{Sin}}$ où τ_{Sin} désigne la constante de temps du modèle de Singer.
On peut alors montrer que :

$$F^{Sin} = \begin{bmatrix} 1 & T & \frac{1}{\alpha^2}(\alpha T - 1 + \rho) \\ 0 & 1 & \frac{1}{\alpha}(1-\rho) \\ 0 & 0 & \rho \end{bmatrix} \tag{2.12}$$

où $\rho = e^{(-\alpha T)}$. De plus, la matrice de covariance du bruit de modèle satisfait :

$$Q^{Sin} = 2\alpha\sigma^2_{Sin} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \tag{2.13}$$

où les éléments $\{q_{ij}\}_{i=1,\dots,3}^{j=1,\dots,3}$ de (2.13) sont définis comme suit :

$$\begin{aligned}
q_{11} &= \frac{1}{2\alpha^5}(2\alpha T - 2\alpha^2 T^2 + 2\alpha^3 T^3/3 - 4\alpha T\rho - \rho^2 + 1) \\
q_{12} &= q_{21} = \frac{1}{2\alpha^4}(\alpha^2 T^2 + 1 + \rho^2 + \rho(2\alpha T - 2) - 2\alpha T) \\
q_{22} &= \frac{1}{2\alpha^3}(2\alpha T - 3 + 4\rho - \rho^2) \\
q_{13} &= q_{31} = \frac{1}{2\alpha^3}(1 - 2\alpha T\rho - \rho^2) \\
q_{23} &= q_{32} = \frac{1}{2\alpha^2}(1-\rho)^2 \text{ and } q_{33} = -\frac{1}{2\alpha}(\rho^2 - 1)
\end{aligned}$$

Remarquons que les matrices de covariance définies en (2.7), (2.10) et (2.13) sont inversibles.

Le modèle de Singer est classiquement utilisé pour modéliser le comportement d'une cible manœuvrante. Le réglage de la constante de temps τ_{Sin} combiné au réglage de la variance sur l'accélération permet de représenter différents types de mouvement. En particulier, on peut montrer que lorsque la valeur de τ_{Sin} est faible, le modèle de Singer tend vers un modèle MRU, alors que lorsqu'elle augmente, le modèle Singer tend vers un modèle de type MUA.
Cette flexibilité fait du modèle de Singer l'un des modèles les plus couramment utilisés en pistage de cibles puisqu'il permet de pister des cibles aux caractéristiques dynamiques variées.

Dans la suite de ce manuscrit, la notion de modèle fait référence à une famille de mouvements (MRU, MUA, Singer, etc.) et un jeu de paramètres (variance sur l'accélération, constante de temps, etc.). Plusieurs peuvent être associés au sein d'une structure à modèles multiples. Néanmoins, lors de la conception d'un tel algorithme d'estimation, l'utilisateur doit faire un choix. Celui-ci rend la conception de l'algorithme délicate d'autant plus qu'en général, il est déconseillé d'associer plus de deux ou trois modèles pour des raisons de précision [Bar95]. Cependant, il est difficile de connaître *a priori* la pertinence d'un modèle plutôt qu'un autre pour modéliser au mieux l'évolution de la cible. En réalité, il est préférable d'associer des modèles suffisamment dissimilaires afin de représenter la plus grande diversité possible de trajectoires. Pour cela, il faut être en mesure de comparer des modèles entre eux. Or, il n'existe pas à notre connaissance de méthode permettant de mener cette comparaison et ainsi de guider la conception d'un algorithme à modèles multiples.

Dans la suite, nous étudions la dissimilarité entre deux modèles d'état à partir d'un outil statistique fondé sur la divergence de Jeffrey.

2.2.2 Divergences de Kullback-Leibler (KL) et de Jeffrey (JD)

Classiquement, lorque l'on compare deux densités de probabilité, l'une des mesures de dissimilarité les plus communes est la divergence de KL [Kul51, Ull96]. C'est par définition une mesure non négative de la dissimilarité entre deux densités de probabilité. La divergence de KL entre deux densités de probabilité $p(\cdot)$ et $q(\cdot)$ satisfait :

$$KL(p,q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{2.14}$$

La divergence de KL, qui est une divergence de Bregman [Nie09], n'est pas une métrique puisque elle ne respecte ni l'inégalité triangulaire, ni la propriété de symétrie. Cependant, sa version symétrique, plus connue sous le nom de divergence de Jeffrey, a été introduite et définie comme suit [Jef48] :

$$JD(p,q) = KL(p,q) + KL(q,p) \tag{2.15}$$

Cette mesure de dissimilarité est utilisée dans de nombreux domaines : en statistique, en apprentissage automatique (*machine learning* en anglais), etc. [Coe05, Sch13]. Elle a également été appliquée dans la littérature pour comparer des processus gaussiens [Ras06], mais dans ce cas seulement des expressions asymptotiques ont été proposées. Des travaux plus récents présentent

également une expression de cette divergence pour les cas des modèles à moyennes ajustées (MA) de premier ordre [Mak10, Leg16] et des sources de Markov [Rac04]. Dans cette thèse, nous étudions le cas plus général des modèles d'état gaussiens et linéaires.

Dans la partie suivante, nous montrons comment tirer profit de ces divergences pour analyser les dissimilarités entre deux modèles indexés par les lettres i et j.

2.3 JD entre deux modèles d'état

Afin de mesurer la dissimilarité entre deux modèles d'état, nous calculons la JD entre les distributions jointes $p(\cdot)$ et $q(\cdot)$ des $k+1$ valeurs successives du vecteur d'état notées $p^i(\boldsymbol{x}_{0:k})$ et $p^j(\boldsymbol{x}_{0:k})$ respectivement.
Dans un premier temps, le calcul de la divergence de KL entre ces deux distributions est proposé. Puis, la JD est exprimée en symétrisant le résultat obtenu. Le calcul analytique direct de ces divergences en passant par le développement des intégrales est relativement fastidieux. En effet, il s'agirait de concaténer les vecteurs d'état successifs sous forme d'un unique vecteur gaussien dont il faudrait inverser la covariance, de dimension $(k+1) \times n_x$. Sauf structure particulière de cette matrice, le coût calculatoire serait alors rapidement prohibitif avec l'augmentation de k et des difficultés numériques pourraient survenir. Ainsi, tirant profit des propriétés des processus de Markov, nous proposons en alternative une forme récursive pour le calcul de la JD.

2.3.1 Calcul récursif

Selon (2.14), l'expression de la divergence de KL entre les deux modèles [2] est définie comme suit :

$$KL\left(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k})\right) = \int_{\boldsymbol{x}_{0:k}} p^i(\boldsymbol{x}_{0:k}) \log \frac{p^i(\boldsymbol{x}_{0:k})}{p^j(\boldsymbol{x}_{0:k})} d\boldsymbol{x}_{0:k} \tag{2.16}$$

En tirant avantage des propriétés associées aux processus de Markov d'ordre un, il vient :

$$\forall\ m \in \{i, j\},\ p^m(\boldsymbol{x}_{0:k}) = p^m(\boldsymbol{x}_{0:k-1}) p^m(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}) \tag{2.17}$$

En insérant (2.17) dans (2.16), on a :

$$\begin{aligned} KL\left(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k})\right) = & \int_{\boldsymbol{x}_{0:k-1}} p^i(\boldsymbol{x}_{0:k-1}) \log \frac{p^i(\boldsymbol{x}_{0:k-1})}{p^j(\boldsymbol{x}_{0:k-1})} \left(\int_{\boldsymbol{x}_k} p^i(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}) d\boldsymbol{x}_k \right) d\boldsymbol{x}_{0:k-1} \\ & + \int_{\boldsymbol{x}_{0:k}} p^i(\boldsymbol{x}_{0:k}) \log \frac{p^i(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})}{p^j(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})} d\boldsymbol{x}_{0:k} \end{aligned} \tag{2.18}$$

Or, comme on a $\int_{\boldsymbol{x}_k} p^i(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}) d\boldsymbol{x}_k = 1$, (2.18) peut se réécrire comme suit :

$$\begin{aligned} KL(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k})) &= \int_{\boldsymbol{x}_{0:k-1}} p^i(\boldsymbol{x}_{0:k-1}) \log \frac{p^i(\boldsymbol{x}_{0:k-1})}{p^j(\boldsymbol{x}_{0:k-1})} d\boldsymbol{x}_{0:k-1} + G \\ &= KL(p^i(\boldsymbol{x}_{0:k-1}), p^j(\boldsymbol{x}_{0:k-1})) \ + G \end{aligned} \tag{2.19}$$

2. Dans la suite, nous parlerons par abus de langage de "divergence entre deux modèles" pour parler de la divergence entre les distributions jointes associées au vecteurs d'état successifs des deux modèles considérés.

où G est le second terme de l'addition dans (2.18) ; il satisfait :

$$G = \int_{\boldsymbol{x}_{k-1:k}} p^i(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}) \log \frac{p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{p^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})} d\boldsymbol{x}_{k-1:k} \tag{2.20}$$

Utilisant la formule de Bayes appliquée à $p^i(\boldsymbol{x}_k, \boldsymbol{x}_{k-1})$, l'équation (2.20) peut alors être reformulée de la manière suivante [3]:

$$\begin{aligned} G &= \int_{\boldsymbol{x}_{k-1}} p^i(\boldsymbol{x}_{k-1}) \int_{\boldsymbol{x}_k} p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \log \frac{p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{p^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})} d\boldsymbol{x}_k d\boldsymbol{x}_{k-1} \\ &= \mathrm{E}_{p^i(x_{k-1})}[KL(p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}),\ p^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}))] \end{aligned} \tag{2.21}$$

où les densités $p^i(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ et $p^j(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ peuvent, en considérant (2.1) et (2.3), s'écrire comme suit :

$$\forall\ m \in \{i, j\},\quad p^m(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \sim \mathcal{N}(\boldsymbol{x}_k; F^m \boldsymbol{x}_{k-1},\ Q^m) \tag{2.22}$$

Ensuite, en tirant profit de la formule de la divergence de KL entre deux densités de probabilité normales multivariées [4] $\mathcal{N}_i$ et $\mathcal{N}_j$, (2.21) peut être développée comme suit :

$$G = \frac{1}{2}\left[\mathrm{Tr}((Q^j)^{-1}Q^i) - n_x - \log\frac{\det Q^i}{\det Q^j} + \int_{\boldsymbol{x}_{k-1}} [\Delta F \boldsymbol{x}_{k-1}]^T (Q^j)^{-1} [\Delta F \boldsymbol{x}_{k-1}] p^i(\boldsymbol{x}_{k-1}) d\boldsymbol{x}_{k-1}\right] \tag{2.24}$$

avec :

$$\Delta F = F^j - F^i \tag{2.25}$$

En utilisant les propriétés de la trace et en intégrant, on a :

$$G = \frac{1}{2}\left[\mathrm{Tr}((Q^j)^{-1}Q^i) - n_x - \log\frac{\det Q^i}{\det Q^j} + \mathrm{E}_{p^i(\boldsymbol{x_{k-1}})}\ \left[\mathrm{Tr}\left(\boldsymbol{x}_{k-1}\boldsymbol{x}_{k-1}^T(\Delta F^T (Q^j)^{-1}\Delta F)\right)\right]\right] \tag{2.26}$$

Puis, par linéarité de la trace, il vient :

$$G = \frac{1}{2}\left[\mathrm{Tr}((Q^j)^{-1}Q^i) - n_x - \log\frac{\det Q^i}{\det Q^j} + \mathrm{Tr}\ [P_{k-1}^i(\Delta F^T (Q^j)^{-1}\Delta F)]\right] \tag{2.27}$$

où la matrice $P_{k-1}^i = \mathrm{E}_{p_i(\boldsymbol{x_{k-1}})}\left[\boldsymbol{x}_{k-1}\boldsymbol{x}_{k-1}^T\right]$ peut être calculée récursivement en s'appuyant sur (2.1) comme suit :

$$P_{k-1}^i = F^i P_{k-2}^i (F^i)^T + Q^i \tag{2.28}$$

puisque $\boldsymbol{x}_{k-2}$ et $\boldsymbol{v}_{k-2}$ sont décorrélés.
En combinant (2.19), (2.27) et (2.28), on obtient une expression récursive de $KL(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k}))$. Puis, en opérant de façon similaire pour $KL(p^j(\boldsymbol{x}_{0:k}), p^i(\boldsymbol{x}_{0:k}))$, on peut en déduire l'expression

3. $\mathrm{E}_{p^i(x_{k-1})}$ correspond à l'espérance mathématique par rapport à $p^i(x_{k-1})$.
4. La divergence de KL entre deux densités normales multivariées est :

$$KL(\mathcal{N}_i, \mathcal{N}_j) = \frac{1}{2}\ [\mathrm{Tr}((Q^j)^{-1}Q^i) - n_x - \log\frac{\det Q^i}{\det Q^j} + (\mu_j - \mu_i)^T (Q^j)^{-1}(\mu_j - \mu_i)\ ] \tag{2.23}$$

où μ_i, μ_j et Q^i, Q^j sont les moyennes et les matrices de covariance des deux densités, n_x est la dimension du vecteur d'état et Tr est la trace de la matrice.

récursive de la JD entre les modèles i et j à l'instant k. Elle s'exprime comme suit :

$$\begin{aligned} JD^{ij}(k) &= JD\left(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k})\right) \\ &= KL\left(p^i(\boldsymbol{x}_{0:k}), p^j(\boldsymbol{x}_{0:k})\right) + KL\left(p^j(\boldsymbol{x}_{0:k}), p^i(\boldsymbol{x}_{0:k})\right) \\ &= JD\left(p^i(\boldsymbol{x}_{0:k-1}), p^j(\boldsymbol{x}_{0:k-1})\right) + A + B \\ &= JD^{ij}(k-1) + A + B \end{aligned} \tag{2.29}$$

où :

$$\begin{aligned} A &= -n_x + \frac{1}{2}\left[\mathrm{Tr}((Q^i)^{-1}Q^j) + \mathrm{Tr}((Q^j)^{-1}Q^i)\right] \\ B &= \frac{1}{2}\left[\mathrm{Tr}\ [P^i_{k-1}(\Delta F^T(Q^j)^{-1}\Delta F)] + \mathrm{Tr}\ [P^j_{k-1}(\Delta F^T(Q^i)^{-1}\Delta F)]\right] \end{aligned} \tag{2.30}$$

En remarquant que les bruits de modèles satisfont :

$$\begin{aligned} \boldsymbol{v}^i_{\boldsymbol{k}} &\sim \mathcal{N}(\boldsymbol{v}^i_k; 0, Q^i) \\ \boldsymbol{v}^j_{\boldsymbol{k}} &\sim \mathcal{N}(\boldsymbol{v}^j_k; 0, Q^j) \end{aligned} \tag{2.31}$$

et compte tenu de (2.23), le terme A de (2.30) peut s'interpréter comme la JD entre les distributions (2.31) des bruits de modèles.

Dans la suite, nous analysons le comportement des termes additifs A et B au cours du temps ainsi que leur influence respective sur l'évolution de la JD.

2.3.2 Analyse des termes additifs dans l'expression récursive de la JD

Analyse de A :
Notons tout d'abord que A est constant.
Considérons la décomposition en valeurs propres de $(Q^i)^{-1}Q^j$ qui est une matrice inversible et n'admet donc pas de valeurs propres nulles. Si l'ensemble $\{\eta_i\}_{i=1,\ldots,n_x}$ représente les valeurs propres de $(Q^i)^{-1}Q^j$, alors $\{\frac{1}{\eta_i}\}_{i=1,\ldots,n_x}$ sont les valeurs propres de $(Q^j)^{-1}Q^i$. Dans ce cas, (2.30) devient :

$$\begin{aligned} A &= -n_x + \frac{1}{2}\left(\mathrm{Tr}((Q^i)^{-1}Q^j) + \mathrm{Tr}((Q^j)^{-1}Q^i)\right) \\ &= -n_x + \frac{1}{2}\sum_{i=1}^{n_x}(\eta_i + \frac{1}{\eta_i}) \end{aligned} \tag{2.32}$$

Dans la suite, nous analysons le signe de ces valeurs propres. Pour cela, on considère le préambule suivant :

- *Si C et D sont deux matrices réelles définies positives et symétriques, alors il existe M et N telles que $C = M^TM$ et $D = N^TN$. Ainsi, $CD = M^TMN^TN$. Son polynôme caractéristique est alors le même que celui de NM^TMN^T, soit $(MN^T)^TMN^T$. Comme cette dernière est positive et symétrique, ses valeurs propres sont réelles et positives. Finalement, les valeurs propres de CD sont également réelles et positives.*

Comme les matrices Q^i et Q^j utilisées dans (2.32) sont deux matrices réelles définies positives et symétriques, $(Q^j)^{-1}Q^i$ et $(Q^i)^{-1}Q^j$ admettent des valeurs propres réelles et positives $\{\eta_i\}_{i=1,\ldots,l}$.

De plus, on peut montrer aisément que si $\eta_i \geq 0$, alors :

$$\eta_i + \frac{1}{\eta_i} \geq 2 \tag{2.33}$$

Par conséquent, le terme A est nécessairement positif ou nul.
Intéressons-nous maintenant au cas où A est égal à zéro :

$$A = 0 \Leftrightarrow \sum_{i=1}^{n_x} (\eta_i + \frac{1}{\eta_i}) = 2n_x \tag{2.34}$$

En considérant (2.33), il vient :

$$\eta_i + \frac{1}{\eta_i} = 2 \Leftrightarrow \eta_i = 1 \tag{2.35}$$

Ainsi, comme la diagonalisation se fait dans une base orthonormée de vecteurs propres puisque la matrice est symétrique, A est égal à zéro si et seulement si $Q^i = Q^j$.

Analyse de B :
Le terme B est nécessairement positif et évolue au cours du temps. Par conséquent, il est à l'origine du caractère non-linéaire de la JD par rapport à k s'il est différent de zéro.
Le cas particulier où $B = 0$ se produit seulement si $\Delta F = 0$.
Finalement, dans (2.30), le terme A est positif et constant, alors que B est également positif mais peut évoluer au cours du temps. Par conséquent, la JD est une fonction croissante du temps.
Par ailleurs, lorsque l'on compare deux modèles d'état, la JD associée est la même quelle que soit la base de l'espace d'état considérée. Cette propriété est démontrée dans la partie suivante.

2.3.3 JD dans différentes bases de l'espace d'état

Un système est caractérisé par une infinité de REEs, on peut passer de l'une à l'autre par une matrice de changement de base. La REE canonique est communément utilisée, mais d'autres peuvent être déduites à partir de celle-ci. L'objectif dans la suite est de montrer que la JD entre deux modèles d'état ne dépend pas de la base de l'espace d'état dans laquelle ils sont comparés.
Si T correspond à une matrice inversible, alors (2.1) et (2.2) deviennent :

$$\boldsymbol{x}_{k+1,T} = F_T^i \boldsymbol{x}_{k,T} + T^{-1} \boldsymbol{v}_k^i \tag{2.36}$$

$$\boldsymbol{y}_k = H_T^i \boldsymbol{x}_{k,T} + \boldsymbol{w}_k^i \tag{2.37}$$

avec $\boldsymbol{x}_{k+1,T} = T^{-1}\boldsymbol{x}_{k+1}$ et où $F_T^i = T^{-1}F_iT$ et $H_T^i = H^iT$ sont les matrices de transition et d'observation respectivement dans la nouvelle base de l'espace d'état. De plus, la matrice de covariance du bruit de modèle est égale à $Q_T^i = T^{-1}Q^i(T^{-1})^T$, alors que celle du bruit de mesure ne change pas.
Dans ce cas, on a :

$$\begin{aligned} JD_T^{ij}(k) &= JD(p^i(\boldsymbol{x}_{0:k,T}), p^j(\boldsymbol{x}_{0:k,T})) \\ &= JD_T^{ij}(k-1) + A_T + B_T \end{aligned} \tag{2.38}$$

avec A_T et B_T définies comme suit :

$$\begin{aligned} A_T &= -n_x + \frac{1}{2}[\mathrm{Tr}\left((T^{-1}(Q^i)^{-1}(T^{-1})^T)^{-1}T^{-1}Q^j(T^{-1})^T\right) \\ &\quad + \mathrm{Tr}((T^{-1}(Q^j)^{-1}(T^{-1})^T)^{-1}T^{-1}Q^i(T^{-1})^T)] \\ B_T &= \frac{1}{2}[\mathrm{Tr}\ [P^i_{k-1,T}\Delta F_T^T(Q^j_T)^{-1}\Delta F_T] + \mathrm{Tr}\ [P^j_{k-1,T}\Delta F_T^T(Q^i_T)^{-1}\Delta F_T]] \end{aligned} \tag{2.39}$$

avec $\Delta F^T = F^j_T - F^i_T$ et $P^i_{k-1,T}$ qui satisfait :

$$\begin{aligned} P^i_{k-1,T} &= \mathrm{E}_{p_i(\boldsymbol{x_{k-1,T}})}\left[\boldsymbol{x}_{k-1,T}\boldsymbol{x}^T_{k-1,T}\right] \\ &= T^{-1}P^i_{k-1}(T^{-1})^T \end{aligned} \tag{2.40}$$

De plus, le terme $\Delta F_T^T(Q^i_T)^{-1}\Delta F_T$ peut être développé de la manière suivante :

$$\begin{aligned} \Delta F_T^T(Q^i_T)^{-1}\Delta F_T &= (T^{-1}\Delta FT)^T T^T(Q^i)^{-1}TT^{-1}\Delta FT \\ &= T^T\Delta F(Q^i)^{-1}\Delta F\ T \end{aligned} \tag{2.41}$$

Finalement, lorsque l'on développe (2.39) en utilisant (2.40), (2.41) et les propriétés de la trace, il vient :

$$\begin{aligned} A_T &= A \\ B_T &= B \end{aligned} \tag{2.42}$$

Ainsi, lorsque l'on compare des modèles d'état, la JD associée est la même quelle que soit la base considérée de l'espace d'état.

Nous proposons dans la partie qui suit d'analyser le pertinence de cette expression récursive pour la classification de modèles d'état.

2.4 Classification d'un ensemble de M modèles d'état fondée sur la JD

Etant donné un ensemble de M modèles satisfaisant (2.1) et (2.2), nous supposons qu'il existe des modèles avec suffisamment de similarité les uns avec les autres pour former des sous-ensembles de modèles représentant ce que l'on appelle des modes [5]. L'objectif est de tirer profit des JDs calculées deux à deux entre les différents modèles pour décomposer l'ensemble des modèles d'état en plusieurs sous-ensembles selon leurs dégrés de similarité.
Nous présentons tout d'abord la façon de structurer les JD calculées deux à deux sous forme matricielle. Puis, nous proposons une méthode d'analyse fondée sur la décomposition en valeurs propres de la matrice résultante. Elle permet de déterminer le nombre de sous-ensembles et leur cardinal. Enfin, dans la dernière partie, nous abordons le processus de classification des modèles dans les sous-ensembles.

5. Un mode fait référence à la caractéristique physique du système, alors qu'un modèle décrit la représentation mathématique du système considéré sur laquelle est fondé l'estimateur [Li00].

2.4.1 Structuration en matrice des JDs entre modèles d'état

Etant donné les JDs entre les paires de modèles, nous les structurons sous forme matricielle. Ainsi, à chaque instant k, il est possible de construire une matrice de dissimilarité, notée $D(k)$, définie comme suit :

$$D(k) = \begin{bmatrix} JD^{11}(k) & JD^{12}(k) & ... & JD^{1M}(k) \\ JD^{21}(k) & JD^{22}(k) & ... & JD^{2M}(k) \\ \vdots & \vdots & \ddots & \vdots \\ JD^{M1}(k) & JD^{M2}(k) & ... & JD^{MM}(k) \end{bmatrix} \tag{2.43}$$

où $k = 0, ..., N$. Elle satisfait :

$$\forall\ i,\ j \in [\![1,\ M]\!] \begin{cases} JD^{ij}(k) = JD^{ji}(k) & \text{si } i \neq j \\ JD^{ij}(k) = 0 & \text{si } i = j \end{cases} \tag{2.44}$$

Dans la suite, il s'agit d'exploiter les informations apportées par la matrice $D(k)$ pour identifier des sous-groupes de modèles parmi les différents modèles considérés. Or, l'analyse de $D(k)$ peut être difficile du fait du nombre d'éléments potentiellement important. Pour pallier cette difficulté, nous proposons de tirer avantage des méthodes de réduction de dimension fondées sur les matrices de corrélation. Pour cela, nous proposons d'appliquer à la matrice $D(k)$ plusieurs transformations successives visant à obtenir une matrice ayant les mêmes caractéristiques qu'une matrice de corrélation.

Il s'agit tout d'abord de s'assurer que les éléments composant la matrice $D(k)$ satisfont les propriétés d'une distance. Or, comme toute autre divergence de Bregman [Nie09], la JD ne satisfait pas l'inégalité triangulaire ; nous modifions donc $D(k)$ à chaque instant afin que cette propriété soit respectée. A cet effet, un algorithme du plus court chemin permet de trouver la plus faible distance entre deux modèles d'état. En particulier, nous proposons d'utiliser l'algorithme de Dijkstra [6] [Dij59]. On obtient donc une nouvelle matrice, notée $W(k)$, composée d'éléments présentant les propriétés d'une distance. Ils sont définis comme suit :

$$\begin{aligned} &\forall\ i,\ j \text{ et } m \in [\![1,\ M]\!] \text{ tels que } \ m \neq i, j \\ &\begin{cases} W^{ij}(k) = JD^{im}(k) + JD^{mj}(k) \ \text{ si } \ JD^{ij}(k) > JD^{im}(k) + JD^{mj}(k) \\ W^{ij}(k) = JD^{ij}(k) \ \text{ sinon} \end{cases} \end{aligned} \tag{2.45}$$

Nous suggérons alors de transformer $W(k)$ en une matrice semblable à une matrice de corrélation notée $E(k)$. Plus particulièrement et par souci de simplicité, les valeurs $E^{ij}(k)$ devront être comprises dans l'intervalle [0 1]. De plus, si $JD^{ij}(k) = 0$, $E^{ij}(k)$ doit être égale à la valeur maximale qu'il est possible d'obtenir, soit 1. Par ailleurs, plus $JD^{ij}(k)$ est élevé, plus $E^{ij}(k)$ doit être petit. Autrement dit, la transformation doit être une fonction monotone décroissante. Comme suggéré dans [Sch13], on peut introduire une fonction non-linéaire telle que :

$$E^{ij}(k) = \exp\left(-\lambda W^{ij}(k)\right) \tag{2.46}$$

6. L'algorithme de Dijkstra est un algorithme de recherche de graphe permettant de résoudre le problème du plus court chemin pour des graphes dont les composantes sont non-négatives. L'application la plus connue est la recherche du plus court chemin pour aller d'une ville à une autre connaissant le réseau routier d'une région.

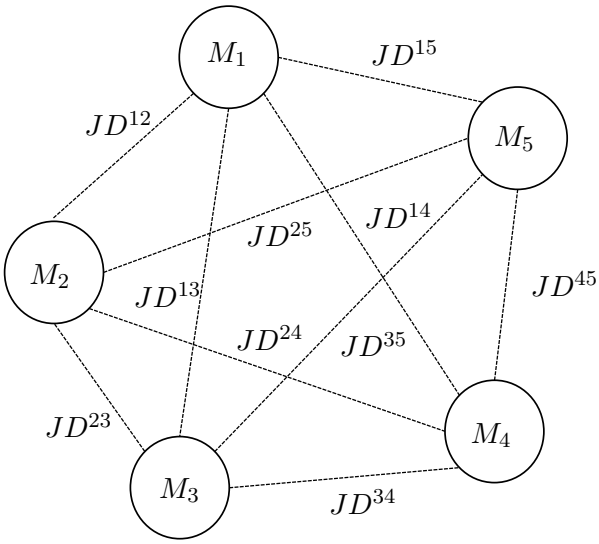

FIGURE 2.2 – Représentation des JDs entre modèles : exemple avec $M = 5$ modèles

où λ est un paramètre d'échelle à régler par l'utilisateur.
Cette transformation non-linéaire permet de renforcer les différences entre les petites valeurs de JD et les plus grandes valeurs. Notons que ce phénomène peut être plus ou moins accentué selon la valeur de λ. En effet, lorsque λ a une valeur suffisamment élevée, les éléments $\{E^{ij}(k)\}_{i\neq j}$ sont proches de 0 et du même ordre de grandeur alors que les éléments diagonaux $\{E^{ii}(N)\}_{i=1,\ldots,\ M}$ sont égaux à 1.
Une fois que les matrices $\{E(k)\}_{k=0,\ldots,N}$ ont été definies et organisées comme illustré par la figure 2.3, nous proposons d'analyser leurs décompositions en valeurs propres. Nous montrons en particulier que l'évolution des valeurs propres permet de déduire le nombre de sous-ensembles et le nombre d'éléments de chacun d'entre eux.

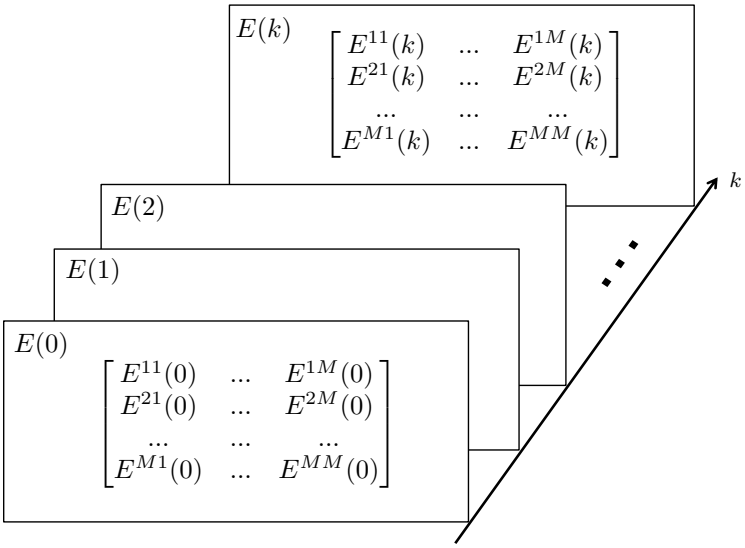

FIGURE 2.3 – Evolution de la matrice de dissimilarité transformée $E(k)$ au cours du temps

2.4.2 Analyse fondée sur une décomposition en valeurs propres

Dans un premier temps, rappelons que $E(k)$ est une matrice symétrique. Elle est semi-définie positive donc ses valeurs propres, notées $\{\beta_n\}_{n=1,...,M}$, sont nécessairement positives ou nulles [THE92]. De plus, notons que les éléments diagonaux de $E(k)$ sont toujours égaux à un. Sa trace est donc égale au nombre de modèles considérés, soit M. Par conséquent

$$\sum_{n=1}^{M} \beta_n = M \text{ et } \beta_n \geq 0 \ \forall \ n \tag{2.47}$$

Dans la suite, nous interprétons la décomposition en valeurs propres de la matrice $E(k)$.

A l'instant initial, comme les conditions sont supposées les mêmes pour tous les modèles, les JDs entre modèles sont toutes égales à zéro et les éléments de $E(k)$ sont tous égaux à un. De ce fait, une valeur propre de $E(0)$ est égale à M tandis que les autres sont toutes égales à 0. Cette valeur propre prédominante caractérise un ensemble de M modèles similaires.

Selon le théorème du cercle de Gershgorin [VAR04], les valeurs propres de $E(k)$ appartiennent au plan complexe et sont incluses dans au moins l'un des cercles de Gershgorin centrés en $\{E^{ii}(k)\}_{i=1,...,M} = 1$ et de rayon :

$$R_i(k) = \sum_{j=1,...,M; i \neq j} |E^{ij}(k)| \tag{2.48}$$

Comme $E(k)$ est une matrice réelle symétrique, les valeurs propres de $E(k)$ sont au moins dans l'un des intervalles $[1 - R_i(k),\ 1 + R_i(k)]$ pour $i = 1, ..., M$.
De plus, puisque $E(k)$ est aussi semi-définie positive, ces intervalles se réduisent à $[max(0, 1 - R_i(k)),\ 1 + R_i(k)]$. Lorsque $k = 0$, cela signifie que les valeurs propres sont dans l'intervalle $[0\ M]$. Puis, lorsque k augmente, les JDs entre les différents modèles augmentent également ou restent constantes [7]. Ainsi, il vient :

$$JD^{ij}(k+1) \geq JD^{ij}(k) \geq 0 \quad \text{pour } i \neq j \tag{2.49}$$

De ce fait, selon (2.45), pour $i \neq j$ on a :

$$\begin{aligned} W^{ij}(k+1) &\geq W^{ij}(k) \\ E^{ij}(k+1) &\leq E^{ij}(k) \end{aligned} \tag{2.50}$$

De plus, $E^{ii}(k+1) = E^{ii}(k) = 1 \ \forall \ i = 1, ..., M$. En combinant les résultats précédents, les valeurs propres de $E(k)$ sont dans des intervalles qui sont de plus en plus petits lorsque k augmente. Cela explique pouquoi la valeur propre la plus grande est dans un premier temps égale à M puis tend vers un au cours du temps. Finalement, le nombre de valeurs propres prédominantes, c'est-à-dire celles qui sont plus grandes que un à un instant donné, peut varier au cours du temps.

7. Si les modèles sont identiques, i.e. ils ont la même REE, une valeur propre est égale au nombre de modèles et les autres sont toutes égales à zéro quel que soit l'instant.

Les valeurs propres peuvent donc évoluer selon l'une des configurations suivantes :
1/ La valeur propre prédominante à l'instant initial qui correspond au nombre de modèles M peut uniquement diminuer ou rester contante.
2/ Concernant les valeurs propres nulles à l'instant initial, certaines d'entre elles augmentent de 0 à 1 au cours du temps.
3/ Les autres augmentent jusqu'à une valeur maximale supérieure à 1 puis diminuent.

A titre d'exemple, considérons la classification d'un ensemble de $M = 4$ modèles de mouvement du type MUA dont les variances sur le jerk sont respectivement réglées comme suit : $\{0.1, 0.2, 0.8, 0.9\}m.s^{-\frac{5}{2}}$. Ces valeurs sont choisies de manière à ce que deux ensembles de deux modèles puissent *a priori* être identifiés. L'objectif est de vérifier ce résultat avec notre approche en analysant le comportement des valeurs propres.
La figure 2.4 représente l'évolution des valeurs propres calculées avec l'approche proposée ainsi que les intervalles $[max(0, 1 - R_i(k)),\ 1 + R_i(k)]$ pour chaque valeur de k.

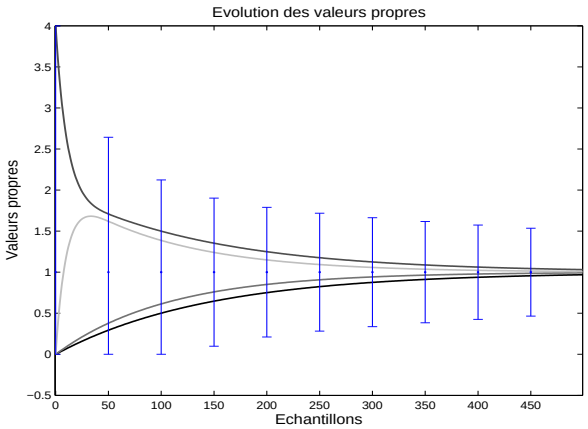

FIGURE 2.4 – Evolution des valeurs propres pour la comparaison de M modèles de mouvement de type MUA

Dans cet exemple, une des quatre valeurs propres augmente de 0 à 2 alors que la valeur propre prédominante à l'instant initial diminue de 4 à 2 sur le même intervalle d'étude. Toutes les valeurs propres convergent ensuite vers 1. Cette évolution traduit, comme attendu, le fait que l'ensemble initial de modèles d'état peut se décomposer en deux sous-ensembles de deux modèles.
Finalement, cette évolution des valeurs propres permet de décomposer l'ensemble initial de modèles en plusieurs sous-ensembles de modèles. En effet, le nombre de valeurs propres prédominantes peut être utilisé pour connaître le nombre de sous-ensembles de modèles. Bien que les valeurs des valeurs propres ne soient pas nécessairement des entiers, elles donnent une information sur le cardinal de chaque sous-ensemble.

Enfin, remarquons que comme les modèles sont associés à des REEs différentes, ils deviennent nécessairement dissimilaires au cours du temps. Ainsi, nous suggérons arbitrairement de choisir un instant N auquel tous les modèles soient considérés comme "également dissimilaires". De

cette façon, les éléments $\{E^{ij}(N)\}_{i\neq j}$ doivent être proches de 0, avec le même ordre de grandeur. Comme les éléments diagonaux $\{E^{ii}(N)\}_{i=1,\ldots,\ M}$ restent égaux à un, $E(k)$ peut s'exprimer comme la somme d'une matrice identité et d'une petite matrice de perturbation. En utilisant le théorème de Wielandt-Hoffman ou encore celui du cercle de Gershgorin, la décomposition en valeurs propres de $E(N)$ donne lieu à des valeurs propres qui sont nécessairement proches de l'unité [WIL70] [MOL04]. Par conséquent, à l'instant N, à condition que les modèles soient considérés comme dissimilaires, il y a M valeurs propres proches de l'unité, correspondant à M sous-ensembles comprenant chacun un modèle.

Intéressons-nous à présent à la valeur du paramètre d'échelle λ défini par la transformation (2.46). Etant donné les considérations précédentes, nous définissons une valeur possible pour ce paramètre. Pour cela, considérons la matrice $E(N)$. Afin de s'assurer que les valeurs propres soient proches de un à l'instant N, les éléments non-diagonaux de $E(N)$ doivent être proches de zéro et satisfaire selon (2.46) :

$$\forall\ i,\ j \text{ tels que } i \neq j,\ E^{ij}(N) = \exp\left(-\lambda W^{ij}(N)\right) \leq \mathcal{T} \tag{2.51}$$

avec $\mathcal{T}$ avec un seuil à définir.
Ainsi, si $W^{min}(N)$ correspond à la valeur minimale des éléments $\{W^{ij}(N)\}_{i\neq j}$, nous suggérons de régler λ de manière à ce que $\exp(-\lambda W^{min}(N))$ soit plus petit que $\mathcal{T}$. Il vient :

$$\lambda \geq -\frac{\ln(\mathcal{T})}{W^{min}(N)} \tag{2.52}$$

Dans cette cette sous-partie, nous avons proposé une façon d'organiser les JDs sous forme matricielle de manière à ce que cette matrice ait les mêmes propriétés qu'une matrice de corrélation après deux transformations explicitées en (2.45) et (2.46). Dans la sous-partie suivante, nous traitons d'une approche permettant de classer les modèles considérés en plusieurs sous-ensembles de modèles.

2.4.3 Application à la conception d'algorithmes à modèles multiples

Classiquement, un enjeu important en classification non-supervisée est de déterminer le nombre de sous-ensembles du jeu de données considéré. En se fondant sur l'analyse proposée dans la sous-partie précédente, nous détaillons à présent la procédure pour déterminer le nombre de sous-ensembles de modèles et leurs cardinaux.

Nous suggérons de procéder à une classification supervisée en considérant que le nombre de classes correspond au nombre de valeurs propres prédominantes. Nous proposons alors d'utiliser l'algorithme K-means avec K réprésentant le nombre de valeurs propres supérieures à un. Ainsi, certains modèles sont suffisamment similaires entre eux pour être distingués des autres et constituer un sous-ensemble de modèles. Finalement, cette approche permet de définir des sous-ensembles de modèles à partir d'un ensemble initial de modèles.

Cette classification est alors d'intérêt pour la phase de conception d'algorithmes à modèles

multiples. En effet, elle peut être mise à profit pour identifier K modèles, chacun appartenant à une classe distincte, à associer afin de maximiser la diversité entre modèles.

Un schéma bloc résumant les différentes étapes de notre approche est fourni en figure 2.5.

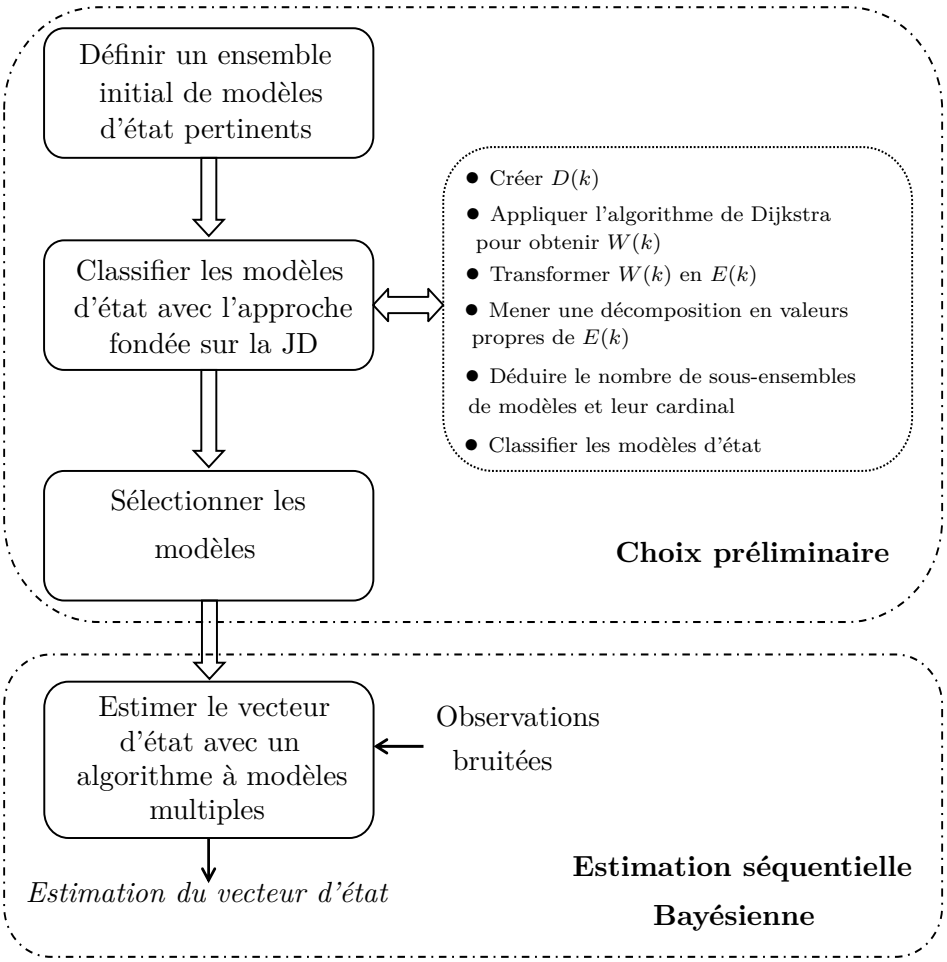

FIGURE 2.5 – Schéma bloc de l'approche proposée pour la conception d'algorithmes à modèles multiples

Dans la partie suivante, nous appliquons la classification de modèles d'état à partir de la JD aux modèles de mouvements.

2.5 Simulations et résultats

Dans cette partie, nous illustrons tout d'abord notre approche en l'appliquant à un exemple simple visant à comparer $M = 2$ marches aléatoires. Nous nous intéressons ensuite à la comparaison de modèles de mouvement dans le contexte du pistage de cibles.

2.5.1 Application à l'estimation de paramètres : comparaison de $M = 2$ marches aléatoires

Dans de nombreux problèmes d'estimation comme la poursuite de fréquence ou l'estimation de paramètres TVAR, la variation temporelle des paramètres peut être décrite par une marche aléatoire. La poursuite s'effectue à partir d'un filtre de Kalman ou d'une de ses variantes. Nous proposons dans cette sous-partie d'appliquer notre approche afin de comparer deux hypothèses d'évolution du vecteur d'état. Plus particulièrement, nous comparons deux marches aléatoires notées ma_1 et ma_2 et définies comme suit :

$$x_{k+1} = x_k + u_k^i \quad i = 1,\ 2 \tag{2.53}$$

où x_k est la valeur du paramètre à l'instant k et u_k^i est un échantillon d'un bruit blanc gaussien centré de variance $\sigma^2_{ma_i}$ où $i \in \{1, 2\}$ correspond au $i^{\text{ème}}$ modèle. De plus, le rapport des variances des processus générateurs des deux modèles est $q = \frac{\sigma^2_{ma_2}}{\sigma^2_{ma_1}}$.
Dans la suite, nous analysons les dissimilarités entre ces deux modèles. Dans ce cas particulier, les matrices de transition F^i sont égales à un pour les deux modèles. Ainsi, comme ΔF vaut zéro, le terme B défini en (2.30) est également nul. Par conséquent, si la JD à l'instant initial est nulle, l'évolution de la JD est régie par la relation suivante :

$$JD^{12}(k) = JD^{12}(k-1) + A = kA \tag{2.54}$$

avec

$$A = -1 + \frac{1}{2}(q + \frac{1}{q}) \geq 0 \tag{2.55}$$

Afin de comparer les deux processus, nous posons $\sigma^2_{ma_1}$ égale à 1 et q varie de 1 à 20. Nous suggérons dans un premier temps de régler λ en utilisant (2.52) avec $\mathcal{T} = 0.02$ et $N = 30$, soit $\lambda = 0.05$. La figure 2.6 représente l'évolution de la JD entre les deux modèles considérés pour différentes valeurs du rapport des variances.

La figure 2.7 représente l'évolution des valeurs propres associées à la matrice de dissimilarité qui caractérise les deux modèles.

Lorsque $q = 1$, il y a une valeur propre prédominante qui est égale à deux quel que soit le nombre d'échantillons tandis que l'autre reste nulle. En effet, comme le terme A défini en (2.55) est égal à zéro, la JD entre les deux modèles est nulle. Cependant, quand q augmente, les deux valeurs propres convergent vers un au cours du temps. Comme l'on pouvait s'y attendre *a priori*, la convergence de plus en plus rapide des valeurs propres lorsque la valeur de q augmente traduit le fait que les deux modèles sont de plus en plus dissimilaires.

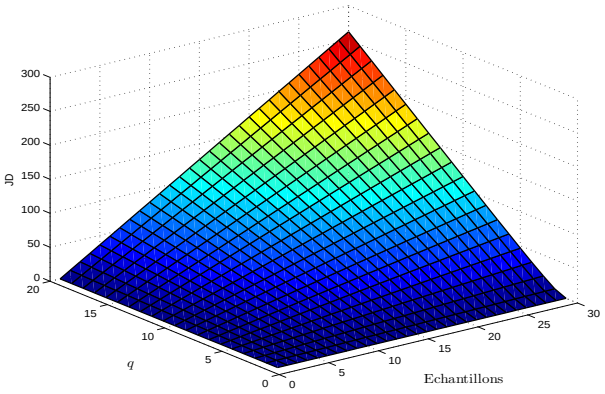

FIGURE 2.6 – Evolution de la JD selon le nombre d'échantillons et le rapport q des variances des processus générateurs des deux modèles

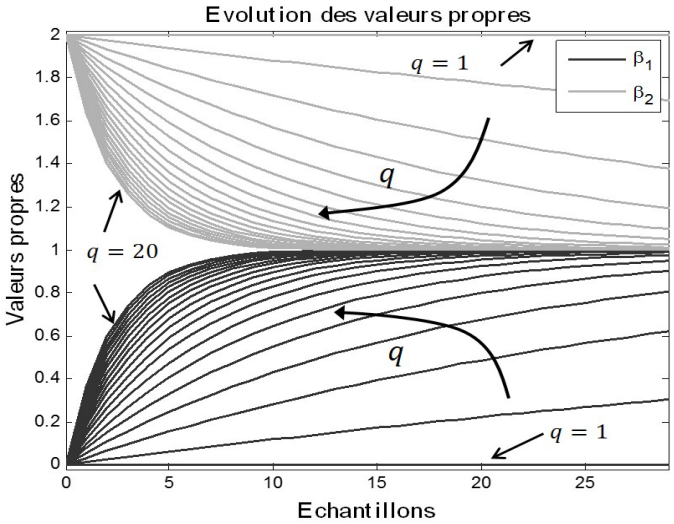

FIGURE 2.7 – Comparaison de deux marches aléatoires selon le rapport q des variances des processus générateurs des deux modèles

2.5.2 Application à la poursuite de cibles : comparaison de $M = 2$ modèles d'état

Dans le paragraphe suivant, nous analysons la JD dans les deux cas suivants :

- pour comparer deux MUAs selon la valeur du rapport des variances de leur jerk,
- pour comparer un MUA et un modèle de Singer selon sa constante de temps.

2.5.2.1 Comparaison de $M = 2$ modèles du type MUA

Si la période d'échantillonnage T est la même pour les deux modèles considérés, $\Delta F = 0$. Ainsi, la JD entre les deux modèles notés MUA_1 et MUA_2, dépend du rapport de variances des jerks noté $q = \frac{\sigma^2_{MUA_2}}{\sigma^2_{MUA_1}}$. Dans ce cas, (2.29) et (2.30) deviennent respectivement :

$$JD^{12}(k) = JD^{12}(k-1) + A = kA \tag{2.56}$$

avec

$$A = n_x \left(\frac{1+q^2}{2q} - 1 \right) = n_x \frac{(1-q)^2}{2q} \tag{2.57}$$

La JD est donc une fonction linéaire au cours du temps et la pente dépend uniquement du rapport des variances q et de la taille du vecteur d'état n_x.

Sur la figure 2.8, on représente 50 trajectoires obtenues selon chacun des deux modèles MUAs considérés pour lesquels le rapport de variance des jerks est égal à $q = 3$. On considère que les deux modèles sont initialisés de la même manière, soit $E[\boldsymbol{x}_0] = [0\ 0\ 0]^T$. Dans ce cas, quel que soit l'instant k, on a :

$$\mathrm{E}[x_k] = 0,\ \mathrm{E}[\dot{x}_k] = 0,\ \mathrm{E}[\ddot{x}_k] = 0 \tag{2.58}$$

De plus, la matrice de covariance à l'instant k du vecteur d'état $\boldsymbol{x}_k$, notée $P(k) = \{P_{ij}(k)\}_{i=1,\ldots,3}^{j=1,\ldots,3}$, est définie comme suit :

$$P(k) = \mathrm{E}[\boldsymbol{x}_k \boldsymbol{x}_k^T] = F^{MUA} P(k-1) (F^{MUA})^T + Q^{MUA} \tag{2.59}$$

avec $P(0)$ une matrice nulle.
A noter que les barres verticales sur la figure 2.8 correspondent à l'intervalle $[-3\sqrt{P_{11}(k)},\ +3\sqrt{P_{11}(k)}]$ aux différents instants.
En analysant les trajectoires résultantes, il apparaît logiquement que pour le modèle MUA avec la plus forte variance sur l'accélération, la plage des positions définie par toutes les trajectoires à chaque instant est plus grande que celle associée au modèle de plus faible variance.

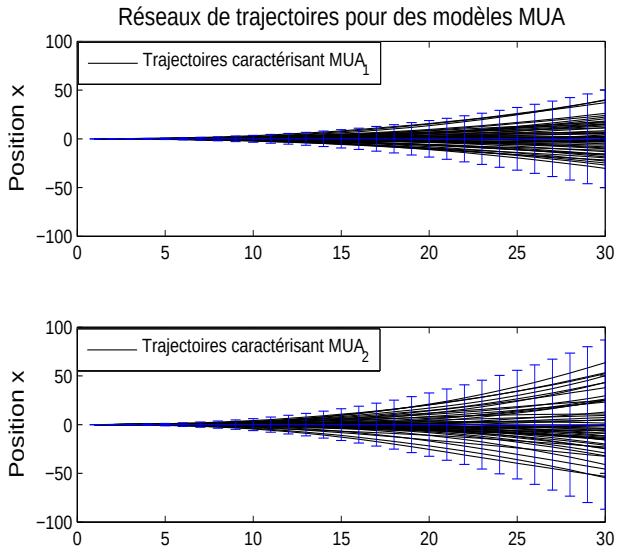

FIGURE 2.8 – Trajectoires générées à partir de modèles MUA et de Singer

Nous illustrons sur la figure 2.9 l'évolution de la JD entre deux modèles MUA selon le nombre d'échantillons et le rapport des variances de jerk entre les deux modèles considérés.

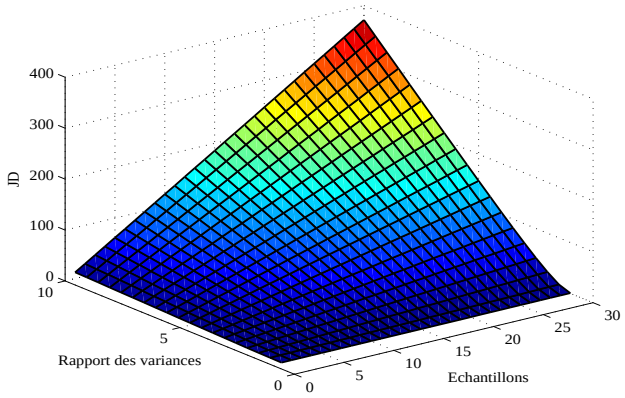

FIGURE 2.9 – Evolution de la JD en fonction du rapport des variances du jerk des modèles considérés

Alternativement, nous représentons sur la figure 2.10 l'évolution des valeurs propres de la matrice de dissimilarité définie à partir de deux modèles MUA selon le rapport des variances de jerk entre les deux modèles considérés.

Pour cet ensemble de simulations, q évolue de 1 à 10 par pas de 0.5. De plus, en s'appuyant sur figure 2.8, si nous admettons qu'avec $q = 3$ les deux modèles peuvent être considérés comme dissimilaires après $N = 20$ échantillons, nous suggérons de régler λ en utilisant (2.52) avec $\mathcal{T} = 0.02$, soit $\lambda = 0.02$. Nous gardons également la même valeur pour les autres valeurs de q.

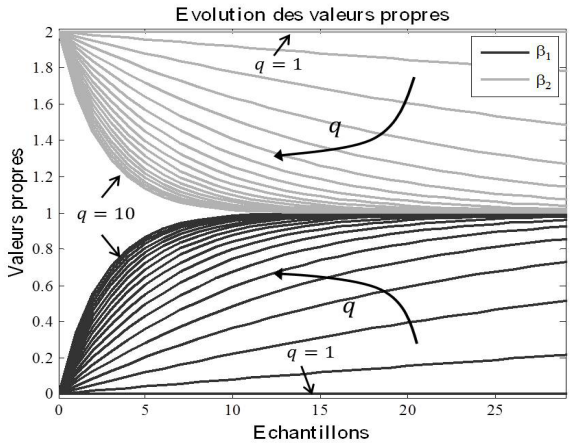

FIGURE 2.10 – Influence du rapport des variances du jerk des modèles considérés sur la JD pour des mouvements de type MUA

Comme attendu, on peut remarquer sur la figure 2.10 que lorsque $q = 1$, c'est-à-dire quand les deux modèles sont identiques, une valeur propre est égale à deux tandis que l'autre est égale à zéro. Les deux modèles étant les mêmes, cela caractérise un sous-ensemble de deux modèles. Puis, lorsque q augmente, la valeur propre prédominante n'est plus égale à deux mais décroit vers un au cours du temps alors que la seconde valeur propre augmente de zéro à un. Ce phénomène de convergence des valeurs propres est de plus en plus rapide lorsque la valeur de q est grande. Cette évolution progressive traduit le passage d'un sous-ensemble de deux modèles à

deux sous-ensembles constitués chacun d'un modèle.

Dans le paragraphe suivant, nous tirons profit de notre approche pour vérifier la propriété que le modèle de Singer est un modèle générique permettant d'approcher un MUA ou un MRU selon la valeur de sa constante de temps τ_{Sin}. En effet, si τ_{Sin} est suffisamment faible, le modèle de Singer peut être assimilé à un MRU alors que si τ_{Sin} est suffisamment grand, le modèle de Singer se rapproche d'un MUA [Bar95].

2.5.2.2 Comparaison de $M = 2$ modèles du type MUA et Singer

Notre approche est appliquée à un ensemble de deux modèles pour différentes valeurs de τ_{Sin} variant de 0.1s à 30s. En théorie, comme les modèles de Singer et MUA sont considérés comme dissimilaires pour de faibles valeurs de τ_{Sin}, nous suggérons de régler λ en utilisant (2.52) pour le cas τ_{Sin} =0.1s avec $\mathcal{T} = 0.02$ et $N = 30$, soit $\lambda = 0.03$. Puis, pour cette valeur de λ, on fait varier τ_{Sin}.
De plus, $T = 1s$ et les variances σ^2_{MUA} et σ^2_{Sin} définies en (2.10) et (2.13) vérifient la relation suivante :

$$\sigma^2_{MUA} = 2\alpha\sigma^2_{Sin} \tag{2.60}$$

Nous représentons sur la figure 2.11 l'évolution de la JD en fonction du nombre d'échantillons et de τ_{Sin}.

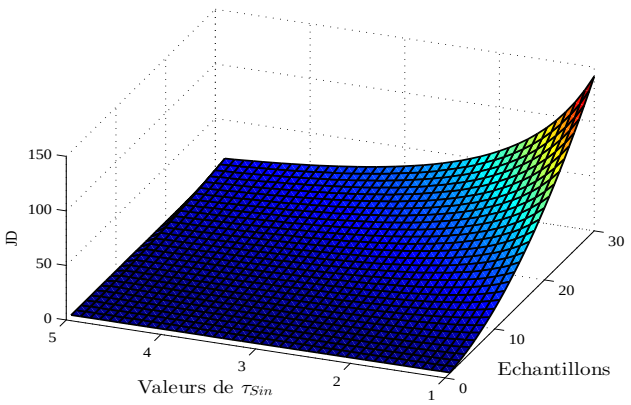

Figure 2.11 – Evolution de la JD en fonction de la constante du modèle de Singer τ_{Sin} et du nombre d'échantillons

Comme illustré par les figures 2.11 et 2.12, on vérifie que plus la constante de temps du modèle de Singer est faible, plus les valeurs propres convergent rapidement vers un. Par conséquent, les modèles sont de plus en plus dissimilaires lorsque τ_{Sin} diminue. Cette remarque confirme que le modèle de Singer est similaire au MUA pour des valeurs suffisamment grandes de τ_{Sin} alors que ces modèles sont de plus en plus dissimilaires pour des faibles valeurs de la constante de temps de Singer [Bar93].

Dans la partie suivante, on considère un ensemble initial de modèles d'état composé de

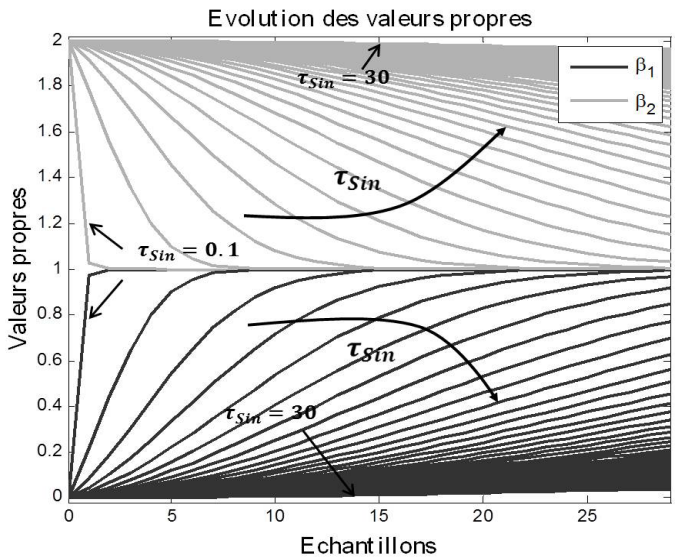

FIGURE 2.12 – Comparaison des modèles de Singer et MUA selon la valeur de τ_{Sin}

MUAs et de modèles de Singer. L'objectif est d'identifier des sous-ensembles de modèles afin de concevoir des algorithmes à modèles multiples pour la poursuite de cibles.

2.5.3 Application à la poursuite de cibles : classification de $M = 6$ modèles d'état

Dans cette sous-partie, nous proposons de déterminer des sous-ensembles de modèles d'état à partir d'un ensemble initial de $M = 6$ modèles composé de modèles de Singer et de MUAs.

2.5.3.1 Description de l'ensemble de modèles considéré

Comparons $M = 6$ modèles d'état, sur $N = 100$ échantillons avec $T = 1s$. Cet ensemble de modèles contient quatre MUAs et deux modèles de Singer respectivement notés $\{MUA_i\}_{i=1}^{4}$ et $\{Sin_i\}_{i=1}^{2}$. Les écarts-types sur le jerk pour les MUAs et les modèles de Singer sont les suivants : $\sigma_{MUA,i} = \sqrt{2\alpha}g\{0.27, 0.30, 0.60, 0.68\}$ et $\sigma_{Sin,i} = g\{0.17, 0.19\}$, où τ_{Sin} vaut $15s$ pour les deux modèles. Notons que ces paramètres sont délibérément choisis de façon à ce que trois sous-ensembles de modèles puissent être *a priori* identifiés : $\{MUA_1, MUA_2\}$, $\{MUA_3, MUA_4\}$ et $\{Sin_1, Sin_2\}$.

2.5.3.2 Analyse et interprétation de l'évolution des valeurs propres

Dans ce paragraphe, nous analysons la pertinence de notre approche sur l'ensemble considéré de modèles. Nous suggérons de régler λ à partir de (2.52), soit $\lambda = 0.15$, puis, si nécessaire, nous affinons notre choix selon la zone d'évolution des valeurs propres que nous souhaitons observer.

A partir de la figure 2.13, nous pouvons noter qu'il y a trois valeurs propres prédominantes. Deux d'entre elles partent de zéro puis augmentent jusqu'à une valeur maximale proche de deux, mais avec une vitesse différente. L'autre valeur propre vaut initialement six puis diminue

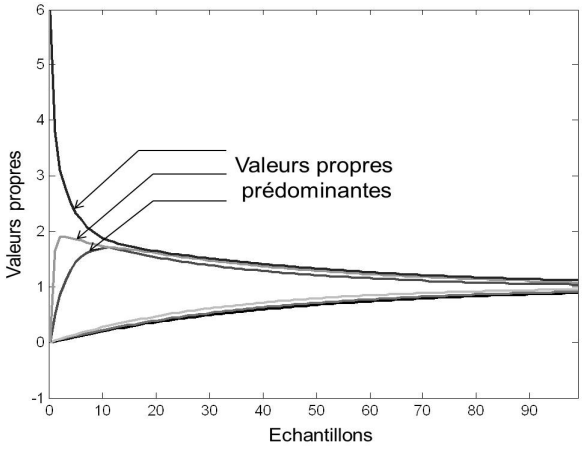

FIGURE 2.13 – Evolution des valeurs propres au cours du temps où le paramètre $\lambda = 0.15$ est défini selon (2.52)

au cours du temps. On observe qu'après cinq à dix échantillons, trois des six valeurs propres sont proches de deux. Cela traduit le fait que l'ensemble initial de modèles d'état peut se décomposer en trois sous-ensembles de deux modèles.
Afin d'analyser de façon plus précise ce phénomène et de mettre en avant l'influence du paramètre d'échelle λ, nous proposons d'affiner la valeur de λ. Ainsi, lorsque λ vaut 0.01, l'évolution des valeurs propres est illustrée par la figure 2.14.

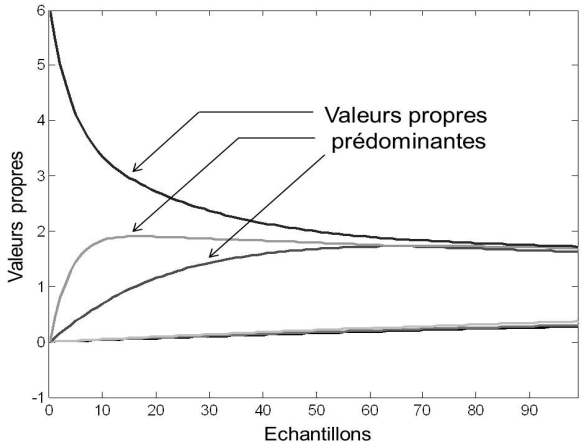

FIGURE 2.14 – Evolution des valeurs propres au cours du temps avec le paramètre d'échelle $\lambda = 0.01$

Finalement, nous remarquons que plus le paramètre λ est grand, plus les valeurs propres convergent rapidement. Régler λ à une valeur suffisamment élevée permet de s'assurer de la détection des maxima des valeurs propres et également de mettre en évidence le phénomène de convergence de l'ensemble des valeurs propres vers un. La valeur de λ n'a pas d'influence sur l'interprétation associée au nombre de sous-ensembles. Elle modifie la vitesse de convergence des valeurs propres mais le résultat de la décomposition de l'ensemble initial en plusieurs sous-ensembles ne varie pas.
Par conséquent, on suggère à l'utilisateur de choisir une valeur suffisamment grande de λ dans un premier temps, pour laquelle toutes les valeurs propres ont convergé vers un sur la fenêtre

d'analyse.

2.5.3.3 Classification de l'ensemble de modèles d'état

A partir des remarques précédentes, on propose d'utiliser un algorithme de classification pour identifier les sous-ensembles de modèles. Plus précisément, l'algorithme des K-means [MAC67] est appliqué à $W(k)$, dont les éléments sont définis par (2.45). La valeur de l'instant k auquel la classification est menée dépend de la valeur K retenue pour l'algorithme des K-means. Par exemple, comme illustré par la figure 2.14, où on observe trois valeurs propres prédominantes, on peut appliquer la classification avec $K = 3$ sous-ensembles à l'instant $k = 60$. Le résultat obtenu est alors :

$$\begin{aligned} &\text{Sous-ensemble 1 (SE 1)} \ : MUA_1, MUA_2 \\ &\text{Sous-ensemble 2 (SE 2)} \ : MUA_3, MUA_4 \\ &\text{Sous-ensemble 3 (SE 3)} \ : Sin_1, Sin_2 \end{aligned} \tag{2.61}$$

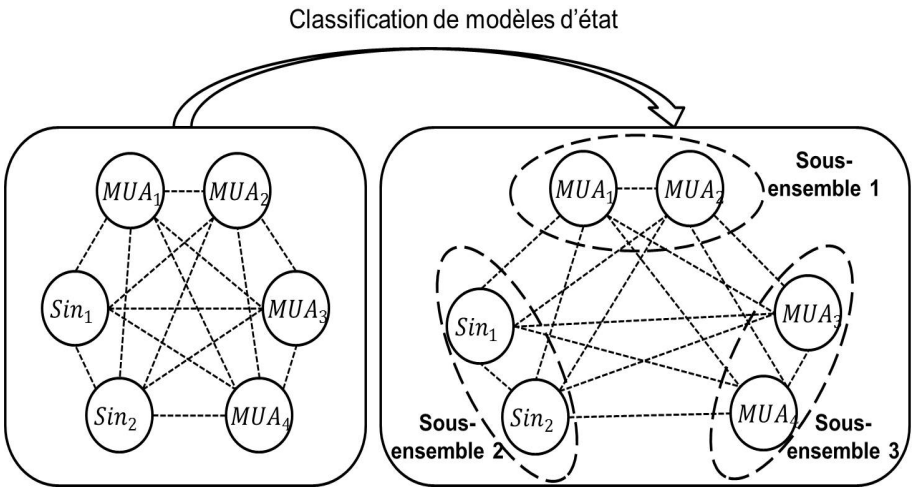

FIGURE 2.15 – Classification d'un ensemble de six modèles en trois sous-ensembles de deux modèles

Finalement, les résultats obtenus et illustrés sur la figure 2.15 sont cohérents avec ce qui était attendu initialement.

Dans la sous-partie suivante, nous proposons de confirmer les résultats précédents en comparant les modèles d'état avec une approche par simulation.

2.5.3.4 Pertinence de la classification

Afin d'analyser la pertinence de la classification indiquée par (2.61) pour la poursuite d'objets mobiles, nous générons une trajectoire de cible à partir de chaque modèle d'état qui a été considéré dans la sous-partie précédente. Puis, cette trajectoire de cible est estimée à partir d'observations bruitées en utilisant des filtres de Kalman fondés sur chacun de ces modèles de mouvement. Les différents algorithmes obtenus sont alors comparés en termes d'EQM sur la position, en moyennant sur 200 réalisations de la trajectoire de référence et du bruit de mesure.

De plus, chaque trajectoire est composée de $N = 100$ échantillons, la période d'échantillon vaut $T = 1s$ et la variance du bruit de mesure est égale à $R = 100m^2$.

			Modèles de trajectoire					
			Sous-ensemble 1		Sous-ensemble 2		Sous-ensemble 3	
			MUA_1	MUA_2	MUA_3	MUA_4	Sin_1	Sin_2
Modèles de poursuite	SE 1	MUA_1	**9.56**	10.06	11.26	11.87	9.61	9.73
		MUA_2	9.58	**10.05**	11.12	11.69	9.64	9.77
	SE 2	MUA_3	9.82	10.12	**10.53**	10.93	9.88	10.01
		MUA_4	9.87	10.14	10.58	**10.85**	9.93	10.05
	SE 3	Sin_1	10.37	10.63	11.91	12.24	**9.24**	9.36
		Sin_2	10.36	10.61	11.92	12.29	9.29	**9.33**

TABLE 2.1 – EQM entre le vecteur d'état et son estimation pour des trajectoires de 100 échantillons, moyennée sur 200 simulations de Monte Carlo.

En analysant le tableau 1, on remarque tout d'abord que pour chacun des modèles de trajectoires, l'EQM minimale est celle associée au modèle de poursuite qui correspond au modèle de trajectoire considéré. Ainsi, les valeurs minimales sont toutes sur la diagonale principale du tableau 2.1 et surlignées en gras. De plus, pour chaque modèle de trajectoire, le modèle de poursuite qui a la deuxième plus faible EQM est le modèle appartenant au même sous-ensemble que le modèle de trajectoire considéré. Ainsi, ces résultats de simulations sont cohérents avec la classification obtenue en (2.61) et permettent de vérifier la pertinence de la méthode fondée sur la JD.

2.5.3.5 Classification de modèles d'état pour la conception d'algorithmes à modèles multiples

Dans cette sous-partie, nous utilisons la classification de modèles d'état introduite précédemment pour concevoir un algorithme à modèles multiples. Supposons qu'une cible puisse suivre plusieurs modèles de mouvement connus *a priori*. Dans ce cas, un algorithme à modèles multiples est communément utilisé pour estimer le vecteur d'état de la cible. Comme évoqué dans l'introduction de ce chapitre, l'utilisation d'un trop grand nombre de modèles en parallèle en même temps au sein d'une structure à modèles multiples ne permet pas nécessairement d'améliorer les performances de l'algorithme en terme d'estimation par rapport à un filtre de Kalman fondé sur une hypothèse de mouvement unique [Li96]. Ainsi, il est préférable d'utiliser un nombre réduit de modèles de mouvement.

Dans la suite, l'objectif est de montrer que notre approche peut être utile pour guider l'utilisateur de modèles multiples dans son choix de modèles. Considérons alors une trajectoire d'une cible générée à partir des modèles de mouvement étudiés dans (2.61) et composée de $N = 500$ échantillons. A chaque instant, le système peut changer de modèle aléatoirement parmi les six modèles considérés. Ces transitions se font selon une chaîne de Markov qui est caractérisée par une matrice de transition dont les éléments diagonaux sont égaux à 0.75 (cela correspond à un temps de séjour de $4T$) tandis que les autres éléments sont égaux à 0.05. De plus, la période

d'échantilonnage est fixée à $T = 1s$ et la variance du bruit de mesure à $R = 100m^2$.

L'état de la cible est alors estimé en utilisant plusieurs algorithmes IMM caractérisés par différentes combinaisons de modèles d'état. Finalement, les estimations du vecteur d'état sont comparées en termes d'EQM sur la position. Les résultats de dix IMM, notés $\{IMM_n\}_{n=1,\ldots,10}$ sont présentés dans la Table 2.2. Parmi ceux-ci, six algorithmes IMM à deux modèles, quatre algorithmes IMM à trois modèles et un algorithme à 6 modèles sont considérés. Les résultats présentés dans la suite sont moyennés sur 200 simulations de Monte Carlo. Notons également que l'EQM calculée lorsque l'estimation correspond à l'observation bruitée, c'est-à-dire lorsqu'aucun filtrage n'est appliqué, est égale à 8.96.

La table 1.2 montre que les algorithmes IMM qui sont caractérisés par la plus petite EQM sont ceux dont les modèles d'état appartiennent à différents sous-ensembles de modèles selon (2.61).

Algorithmes IMM	Modèles d'état utilisés	EQM
IMM_1	$\{UAM_1,\ UAM_2\}$	8.95
IMM_2	$\{UAM_1,\ UAM_3\}$	8.35
IMM_3	$\{UAM_3,\ UAM_4\}$	8.47
IMM_4	$\{UAM_3,\ Sin_1\}$	8.04
IMM_5	$\{UAM_4,\ Sin_2\}$	7.95
IMM_6	$\{Sin_1,\ Sin_2\}$	8.71
IMM_7	$\{UAM_1,\ Sin_1,\ Sin_2\}$	8.79
IMM_8	$\{UAM_3,\ UAM_4,\ Sin_2\}$	8.94
IMM_9	$\{UAM_1,\ UAM_3,\ Sin_1\}$	7.93
IMM_{10}	$\{UAM_2,\ UAM_4,\ Sin_2\}$	7.86

TABLE 2.2 – EQM entre le vecteur d'état et son estimation

Parmi les IMM à deux modèles, IMM_1 , IMM_3 et IMM_6 dont les modèles d'état appartiennent au même sous-ensemble de modèles présentent les plus grandes EQMs. Plus précisément, contrairement à IMM_3, les algorithmes IMM_1 et IMM_6 ne prennent pas en compte les périodes manœuvrantes de la cible. Ainsi, leurs EQMs sont parmi les plus grandes car les phases non-manœuvrantes de la cible ne sont pas correctement modélisées et donc filtrées. En effet, dans ce cas, le gain de Kalman est surestimé puisque la variance du bruit des modèles d'état est elle-même surestimée. Les modèles d'état utilisés pour IMM_2 , IMM_4 et IMM_5 appartiennent dans chacun des trois cas à des sous-ensembles différents. Ces algorithmes présentent des EQMs relativement faibles puique toutes les manœuvres de la cible sont représentées par la combinaison de modèles d'état utilisée. Plus particulièrement, parmi les trois derniers algorithmes cités, IMM_5 présente la plus faible RMSE puisque tous les modèles MUAs et de Singer peuvent être représentés par les modèles englobants UAM_4 et Sin_2.

Par ailleurs, parmi les algorithmes IMM à trois modèles, IMM_7 et IMM_8 sont moins performants puisqu'ils ne permettent pas de filtrer correctement toutes les manœuvres suivies par la cible. En effet, dans le cas de IMM_7, les phases de mouvement modélisées par UAM_3 ou UAM_4 ne sont pas représentées. Pour IMM_8, ce sont les phases modélisées par MUA_1 et MUA_2 qui

ne sont pas représentées. Cependant, IMM_9 et IMM_{10} sont les algorithmes avec la plus petite EQM. Selon (2.61), les trois modèles d'état de ces deux algorithmes appartiennent à différents sous-ensembles de modèles. Par conséquent, les manœuvres de la cible peuvent toutes être représentées au fil du temps par au moins l'un des modèles d'état qui est utilisé par ces algorithmes. De plus, nous pouvons remarquer que l'EQM de IMM_{10} est légérement plus faible que celle de IMM_9 puisque les modèles d'état de IMM_{10} sont ceux parmi chaque sous-ensemble avec les plus grandes variances de bruit.

Finalement, étant donné qu'il est préférable d'associer des modèles dissimilaires pour concevoir un algorithme modèles multiples, on constate que notre méthode de classification de modèles d'état est d'intéret afin de choisir judicieusement quels modèles peuvent être associés au sein d'un algorithme à modèles multiples.

2.6 Conclusions et perspectives

L'un des nombreux enjeux en approche paramétrique pour l'estimation est de définir *a priori* un modèle qui est utilisé pour représenter le système considéré. Pour relâcher les contraintes sur ces choix *a priori*, les approches à modèles multiples ont été développées. Lorsque l'on applique ces méthodes, il n'existe pas de règle de conception concernant la façon dont les modèles doivent être choisis. Néanmoins, il est préférable qu'ils soient suffisamment dissimilaires entre eux et représentatifs du panel de mouvements possibles effectués par la cible pour assurer une certaine diversité entre les modèles. Cependant, il n'y a pas à notre connaissance de méthode permettant de comparer les modèles d'état entre eux.

Dans ce chapitre, nous avons étudié la JD entre deux densités de probabilités jointes des vecteurs d'état successifs correspondant à deux modèles distincts.
Nous montrons tout d'abord que l'expression récursive de la JD dépend de deux termes : le premier peut s'interpréter comme la JD entre les distributions des bruits de modèles alors que le second dépend notamment des matrices de transition. Ces deux termes sont en outre positifs ou nuls. Enfin, quelle que soit la base choisie pour représenter l'espace d'état, la JD est la même. Puis, nous avons développé une approche statistique visant à organiser un groupe de plusieurs modèles d'état en différents sous-ensembles de modèles similaires. Pour ce faire, nous structurons les JDs entre paires de modèles d'état sous forme matricielle. Nous disposons ainsi d'un ensemble de matrices de JD reflétant les différences entre modèles en tenant compte d'un nombre d'échantillons de plus en plus élevé. Afin de tirer avantage des méthodes de réduction de dimension fondées sur des matrices de corrélation, plusieurs transformations sont ensuite menées sur ces matrices. L'analyse des valeurs propres prédominantes de la matrice résultante permet alors de déduire le nombre de sous-ensemble de modèles. Enfin, un algorithme de classification tel que l'algorithme k-means permet d'identifier quels modèles appartiennent à chaque sous-ensemble. Cette approche a d'abord été testée avec succès pour comparer deux modèles MUA ou de Singer puis pour extraire des sous-ensembles de modèles à partir d'un groupe de six modèles. Nous avons alors montré comment tirer profit de cette classification pour guider un utilisateur dans

la phase de conception d'un algorithme à modèles multiples.
Nous avons ensuite appliqué notre approche au pistage d'une cible et avons montré que cette analyse est pertinente pour la classification de modèles d'état.

Notons que nous avons étendu ce travail à des processus de Markov d'ordre p quelconque et montré sa pertinence pour la comparaison de processus AR ou encore AR variant dans le temps (TVAR) [Mag15a, Mag15e].

Comme perspectives d'étude, plutôt que d'utiliser une divergence pour comparer les distributions jointes, une alternative serait d'étudier la distance de Rao entre les matrices de covariance de ces distributions. De plus, plutôt que de considérer des décompositions en valeurs propres successives et d'analyser leurs évolutions au cours du temps, nous pourrions investiguer l'analyse tensorielle et plus particulièrement exploiter une décomposition en valeurs singulières d'ordre supérieur.

Les approches développées dans ce chapitre reposent sur des modéles paramétriques qu'il est difficile de régler *a priori*. Afin de pallier ce problème, des modèles bayésiens non-paramétriques peuvent être utilisés. Dans le chapitre suivant, nous traitons de l'utilisation de ces modèles pour estimer les propriétés statistiques du bruit de modèle.

Chapitre 3

Approches bayésiennes non-paramétriques pour le pistage de cibles

Sommaire

3.1 Introduction **72**

3.2 Processus de Dirichlet **73**

3.2.1 Introduction aux méthodes bayésiennes non-paramétriques 73

3.2.2 Définition des processus de Dirichlet 75

3.2.3 Représentation dite en *stick breaking* 76

3.2.4 Urne de Polya 78

3.3 Estimation non-paramétrique de la matrice de covariance du bruit de modèle **80**

3.3.1 Description du problème 80

3.3.2 Représentation dans l'espace d'état non-paramétrique 82

3.4 Cas des matrices de covariance avec une structure prédéfinie **84**

3.4.1 Modèle hiérarchique considéré 85

3.4.2 Calcul de la distribution d'importance optimale 86

3.5 Application au pistage d'une cible **92**

3.5.1 Protocoles de simulation 92

3.5.2 Résultats de simulation 97

3.6 Conclusions et perspectives **101**

3.1 Introduction

La REE du système joue un rôle important vis-à-vis des performances d'estimation de l'algorithme récursif utilisé. En particulier, un réglage inapproprié des matrices de covariance des bruits de modèle et de mesure dégrade la précision des estimations et peut même conduire à des divergences.
Afin de relâcher cette contrainte, différentes techniques visant à estimer les matrices de covariance de ces bruits ont été proposées dans la continuité des travaux fondateurs de Mehra [Meh70, Meh71]. Elles peuvent être classées en trois catégories : les algorithmes mettant à profit les corrélations et covariances [Meh72, Ode06, Raj09], ceux reposant sur l'estimation au sens du maximum de vraisemblance [Axe11] et enfin ceux fondés sur l'estimation bayésienne [Fre12, Ge14]. Ces méthodes se sont avérées pertinentes dans différents domaines d'application comme le rehaussement de la parole [Gab99] ou le pistage de cibles non-manœuvrantes [Bar11] pour lesquels les matrices de covariance à estimer varient lentement ou sont quasiment constantes par morceaux. Cependant, cette hypothèse n'est pas nécessairement valide dans d'autres contextes. Par exemple, en navigation GPS [Spa10], les multitrajets ou les interférences induisent des sauts fréquents de variance du bruit de mesure. De façon similaire, en pistage de cibles manœuvrantes, le type de mouvement suivi par une cible peut évoluer rapidement au cours du temps ; pour chaque type de mouvement, la dynamique du bruit de modèle peut également changer. Ces changements de familles de modèles et de dynamiques sont classiquement pris en compte à partir de plusieurs REEs caractérisées par différentes valeurs de la matrice de covariance du bruit de modèle. L'estimation est alors réalisée par l'intermédiaire d'algorithmes à modèles multiples tels que l'IMM [Bar95, Bar11], le *variable structure*-IMM (VS-IMM) [Li05] ou un filtre particulaire pour les systèmes de Markov à sauts [Dou01b]. Néanmoins, dans ces cas, il est nécessaire de définir le nombre de modèles *a priori* et leurs paramètres.

Plus récemment, les méthodes bayésiennes non-paramétriques (BNP), introduites par Fergusson [Fer73] et bien connues en statistiques ou *machine learning*, ont été exploitées en traitement du signal. Elles ont été mises à profit par exemple pour les communications numériques [Val15], en radio cognitive [Per14], en traitement vidéo [Elg13], en traitement de la parole [Fox11a], ou encore en pistage de cibles [Cha14, Fox06, Fox07]. Ces approches permettent de relâcher toute hypothèse sur le nombre de modèles et la distribution de leurs paramètres. Le principe est de considérer que cette distribution est inconnue et aléatoire. Elle se voit donc attribuer une loi *a priori* telle qu'un processus de Dirichlet [Car08, Fox10, Fox11b]. Cela revient à considérer que le système peut changer de modèle parmi une collection illimitée. Ces modèles peuvent réapparaître au cours du temps et leurs caractéristiques statistiques sont inconnues.
Plus précisément, dans [Car08], la matrice de transition est définie *a priori* mais les distributions des bruits de modèle et de mesure sont supposées inconnues. Plus particulièremet, elles sont estimées à partir de processus de Dirichlet à mélange (DPM), qui peuvent être vus comme des mélanges infinis de densités de probabilité. Caron propose la mise en œuvre de son approche dans un cas scalaire et réalise l'inférence du paramètre inconnu à partir d'un filtre particulaire Rao-Blackwellizé. Dans [Fox11b], Fox *et al.* traitent du cas où les matrices d'état, c'est-à-dire la matrice de transition et la matrice de covariance du bruit d'état, sont supposées inconnues et les estiment conjointement à la matrice des probabilités de transition entre modèles en utilisant des

processus de Dirichlet hiérarchiques. Cette représentation est plus complète que celle proposée par Caron mais aussi plus complexe à mettre en œuvre. L'estimation est réalisée *offline* selon des approches du type *markov chain Monte-Carlo* (MCMC).

Dans ce chapitre, nous étudions l'intérêt d'un modèle BNP fondé sur les processus de Dirichlet pour traiter l'estimation conjointe du vecteur d'état et de la matrice de covariance du bruit d'état [1] d'un système linéaire à sauts.
Une approche possible est celle initialement proposée par Caron [Car08] et décrite précédemment qui consiste à modéliser directement la distribution de toute la matrice de covariance par un processus de Dirichlet. Néanmoins, selon la dimension de la matrice, le nombre de paramètres à estimer conjointement peut être élevé. Par conséquent, il s'agit de trouver un compromis entre la dimensionnalité du problème d'estimation et la flexibilité de l'algorithme. Aussi, nous proposons une approche fondée sur l'hypothèse que les modèles d'évolution appartiennent à des familles connues. Au sein de chacune d'entre elles, les matrices de covariance des bruits de modèle partagent la même forme fonctionnelle mais diffèrent selon un nombre réduit d'hyperparamètres que l'on cherche à estimer. C'est la distribution de ces paramètres qui est alors supposée suivre un processus de Dirichlet.
Finalement, l'inférence du vecteur d'état et de la matrice de covariance du bruit d'état est réalisée à partir d'un filtre particulaire dont le calcul de la distribution d'importance est développé dans ce chapitre.

Dans la suite du chapitre, nous présentons dans un premier temps la théorie des processus de Dirichlet et nous introduisons la modélisation bayésienne non-paramétrique du problème d'estimation de la matrice de covariance du bruit d'état. Ensuite, nous développons l'expression de la distribution d'importance optimale utilisée au sein du filtre particulaire considéré. Enfin, nous menons une étude comparative, fondée sur des résultats de simulation, avec d'autres approches.

3.2 Processus de Dirichlet

3.2.1 Introduction aux méthodes bayésiennes non-paramétriques

L'estimation bayésienne est une approche pour l'inférence statistique reposant sur l'utilisation de modèles *a priori* sur les variables à estimer. Les méthodes paramétriques s'appuient généralement sur l'hypothèse que le système peut être caractérisé par un vecteur de paramètres de dimension finie. Par exemple, en classification non supervisée, chaque classe est caractérisée par une forme de loi et un vecteur de paramètres. Ainsi, les données suivent un mélange fini de densités de probabilités de vecteurs de paramètres inconnus que l'on souhaite estimer. Chaque mode du mélange correspond alors à un nuage de points. L'inconvénient de ce type de méthodes est que seules quelques directions dans l'espace des paramètres possibles sont explorées et que l'on ne s'adapte pas à l'apparition de nouveaux modes. Afin de rendre cette technique plus robuste, il peut être judicieux de considérer une distribution inconnue dont le support est plus

1. Comme les statistiques sur le bruit de mesure peuvent être obtenues dans de nombreuses applications et en particulier en traitement radar, seule la matrice de covariance du bruit d'état est supposée inconnue dans la suite.

large que celui fourni par un modèle paramétrique.

Ces deux dernières décennies ont vu le développement des méthodes BNP qui apportent un cadre flexible pour la spécification de lois *a priori* en permettant de relâcher un certain nombre d'hypothèses *a priori*. Au lieu de considérer une distribution *a priori* sur un espace de dimension finie, les modèles BNP définissent une distribution de probabilité sur des espaces de dimension infinie. Ainsi, un modèle non paramétrique peut être simplement vu tel un modèle statistique avec un nombre infini de paramètres.

La figure 3.1 représente l'évolution temporelle du bruit de modèle associé à la trajectoire d'une cible manœuvrante. Elle met en évidence plusieurs modes qui apparaissent à des instants différents, ce qui se traduit sur l'histogramme des valeurs du bruit par un mélange de lois. Par conséquent, la distribution du bruit de modèle pourrait être représentée à partir d'un mélange de distributions. Cependant, se pose le problème du choix du nombre de modes pour que le mélange soit suffisamment représentatif. La particularité des méthodes BNP est d'apprendre le comportement stochastique des sauts de paramètres entre différents modes et d'autoriser de nouveaux modes à mesure que le nombre de données augmente.

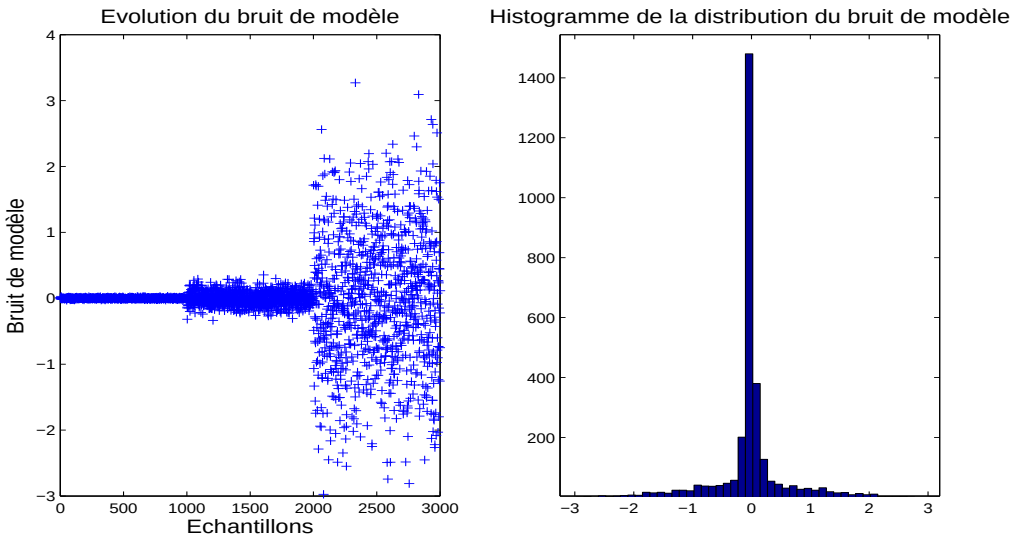

FIGURE 3.1 – Réalisation du processus représentant le bruit de modèle associé à la trajectoire d'une cible manœuvrante et histrogramme associé.

Les modèles bayésiens non paramétriques les plus populaires sont les processus gaussiens et les processus de Dirichlet. Bien que les processus de Dirichlet soient des objets statistiques connus depuis le début des années 1970, ces modèles ne sont devenus populaires que plus récemment grâce au développement des méthodes de Monte Carlo permettant d'estimer leurs paramètres. Ils ont depuis connu un grand essor pour l'estimation de densités et la classification non supervisée dans divers domaines d'application.

3.2.2 Définition des processus de Dirichlet

Le processus de Dirichlet a été introduit par Ferguson [Fer73] comme étant une mesure de probabilité sur l'espace des mesures de probabilité.

Soit $(\mathcal{X}, \mathcal{A})$ un espace mesurable, on introduit G_0 et α_0 une mesure de probabilité sur $(\mathcal{X}, \mathcal{A})$ et un réel strictement positif respectivement. Une distribution de probabilité G est distribuée selon un processus de Dirichlet de distribution de base G_0 et de facteur d'échelle α_0 si quelle que soit la partition $A_1, ..., A_K$ de $\mathcal{X}$, le vecteur de probabilités aléatoires $[G(A_1)...G(A_K)]$ suit une distribution de Dirichlet, c'est-à-dire :

$$[G(A_1)...G(A_K)] \sim \mathcal{D}(\alpha_0 G_0(A_1), ..., \alpha_0 G_0(A_K)) \tag{3.1}$$

où $\mathcal{D}(\cdot)$ représente la distribution de Dirichlet qui est une distribution sur un jeu de probabilités discrètes de dimension K.
Par définition, si S_k désigne le simplexe de dimension $K-1$ et $\Gamma(\cdot)$ est la fonction gamma, la distribution de Dirichlet de paramètre d'échelle α_0 et de vecteur de probabilités p_0 de dimension K dont le $j^{\text{ème}}$ élément est noté $p_0(j)$ satisfait :

$$\mathcal{D}(p, \alpha_0 p_0) = \begin{cases} \frac{\Gamma(\sum_{j=1}^{K} \alpha_0 p_0(j))}{\prod_{j=1}^{K} \Gamma(\alpha_0 p_0(j))} (p(j))^{\alpha_0 p_0(j)-1} & \text{si } p \in S_k \\ 0 \text{ sinon} \end{cases} \tag{3.2}$$

Le simplexe correspond à l'ensemble des vecteurs de probabilité de dimension K tels que $\sum_j p(j) = 1$ et $p(j) \geq 0 \ \forall \ j$.
Notons que la moyenne et la variance de la distribution sont définies comme suit :

$$E[p(j)] = p_0(j) \tag{3.3}$$

$$Var[p(j)] = \frac{(1-p_0(j))p_0(j)}{1+\alpha_0} \tag{3.4}$$

L'influence du paramètre d'échelle α_0 est illustrée par la figure 3.2 où sont représentées plusieurs distributions de Dirichlet caractérisées par différentes valeurs du paramètres d'échelle α_0 et du vecteur de probabilité moyen p_0. Conformément à la formule (3.4), on remarque que plus α_0 est grand, plus la variance est faible et donc plus les réalisations sont concentrées autour du vecteur de probabilité moyen p_0. Au contraire, plus α_0 est petit, plus les valeurs sont dispersées.

De plus, concernant le processus de Dirichlet, considérant (3.1), il peut être montré que pour tout $A \in \mathcal{X}$ son espérance et sa variance sont définies comme suit :

$$E_{DP(G_0,\alpha_0)}[G(A)] = G_0(A) \tag{3.5}$$

$$Var_{DP(G_0,\alpha_0)}[G(A)] = \frac{(1-G_0(A))G_0(A)}{1+\alpha_0} \tag{3.6}$$

Par conséquent, la distribution de base G_0 représente la valeur moyenne du processus de Dirichlet. Par ailleurs, le paramètre d'échelle α_0 est homogène à un paramètre d'inverse variance. Plus α_0 est grand, plus la variance du processus de Dirichlet est petite et le processus de Dirichlet est concentré autour de la distribution de base G_0.

Il existe plusieurs manières de représenter les réalisations d'un processus de Dirichlet. L'une

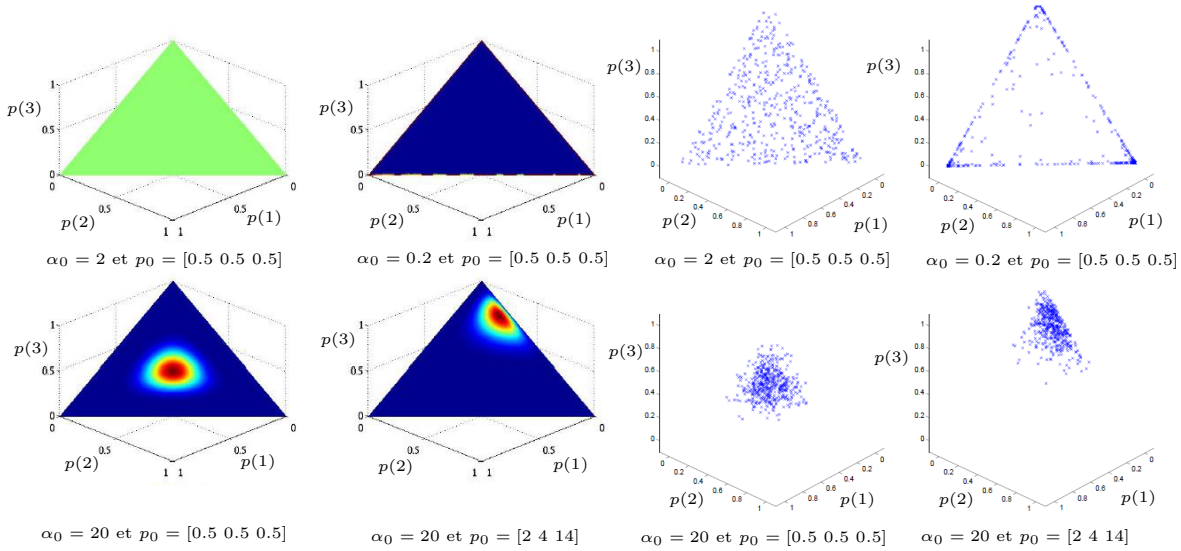

FIGURE 3.2 – Influence du paramètre d'échelle α_0 et du vecteur de probabilités initiales p_0 sur la distribution du processus de Dirichlet. A gauche : courbes de niveau de la densité de Dirichlet projetées sur le simplexe. A droite : nuages de points correspondant à des réalisations d'un processus de Dirichlet.

d'entre elles est introduite ci-après. Il s'agit de la représentation dite en *stick-breaking*.

3.2.3 Représentation dite en *stick breaking*

Avec une probabilité égale à un, les réalisations G d'un processus de Dirichlet sont des distributions discrètes à support infini [BLA73]. Elles peuvent s'exprimer à partir de la représentation dite en *stick-breaking* [SET94] :

$$G(\boldsymbol{\theta}_k) = \sum_{j=1}^{+\infty} \pi_j \delta_{\boldsymbol{U}_j}(\boldsymbol{\theta}_k) \tag{3.7}$$

où $\delta_{\boldsymbol{U}_j}(\cdot)$ est la mesure de Dirac centrée en $\boldsymbol{U}_j$. Les positions $\boldsymbol{U}_j$, aussi appelées atomes, sont indépendantes et identiquement distribuées selon la distribution de base G_0, i.e. :

$$\boldsymbol{U}_j \overset{i.i.d}{\sim} G_0 \tag{3.8}$$

Comme illustré par la figure 3.3, les probabilités π_j sont obtenues à partir d'une procédure inspirée d'un cassé de bâton (*stick-breaking* en anglais). Cette procédure consiste à casser en morceaux successifs un bâton de longueur unité. Le bâton est initialement cassé en deux. Il en résulte un morceau de longueur $\pi_1 = \beta_1$ et un autre de longueur $L_1 = 1 - \beta_1$. Le morceau de longueur L_1 est ensuite cassé en deux. Un des deux morceaux restants est de taille $\pi_2 = \beta_2(1-\beta_1)$ alors que le second est de taille $L_2 = (1-\beta_1)(1-\beta_2)$. Cette étape peut être réitérée une infinité de fois. Ainsi, à la $j^{\text{ème}}$ itération, pour $j \geq 2$, il vient :

$$\pi_j = \beta_j \prod_{l=1}^{j-1} (1 - \beta_l) \tag{3.9}$$

Pour les processus de Dirichlet, les β_j sont des variables aléatoires indépendantes suivant une distribution Beta, i.e. :

$$\beta_j \sim \mathcal{B}(1,\ \alpha_0) \tag{3.10}$$

avec α_0 le paramètre d'échelle du processus de Dirichlet. Par ailleurs, notons que par construction, l'équation (3.7) définit une mesure de probabilité puisque $\sum_{j=1}^{+\infty} \pi_j = 1$ et $0 \leq \pi_j \leq 1$.

Ainsi, les réalisations d'un processus de Dirichlet correspondent à des mélanges infinis de mesures de Dirac. Comme illustré par la figure 3.4, la distribution de base G_0 détermine les positions des atomes ou points supports de G, tandis que le facteur d'échelle α_0 pilote la variance des poids.

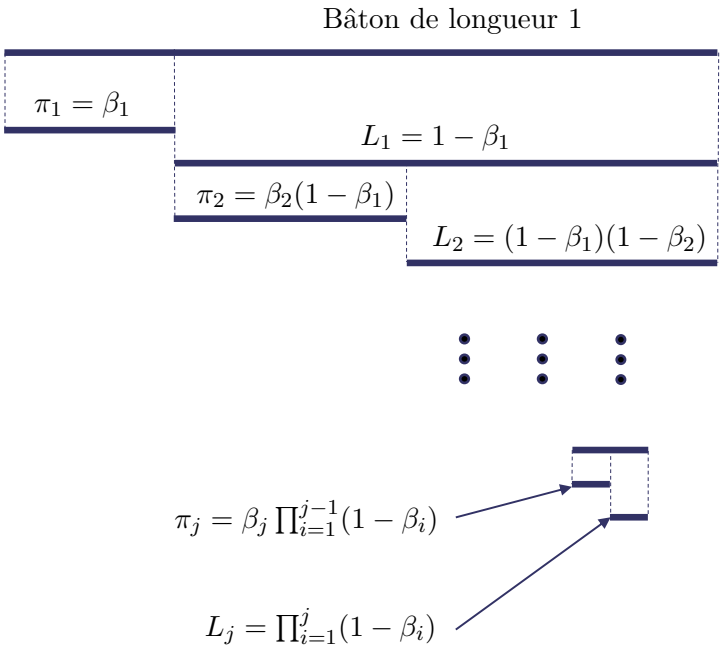

FIGURE 3.3 – Représentation graphique du *stick breaking*

Selon la distribution à partir de laquelle sont générées les variables β_j, il existe d'autres types de construction en *stick-breaking* correspondant à des processus différents du processus de Dirichlet. Par exemple, si $\mathcal{B}(\cdot)$ désigne la distribution beta, on peut considérer le processus Beta à deux paramètres [ISH00] tel que $\beta_j \sim \mathcal{B}(a, b)$ ou encore le processus de Pitman-Yor également appelé processus Poisson-Dirichlet [PIT97] tel que $\beta_j \sim \mathcal{B}(1 - a, b + ja)$.

Un propriété intéressante des processus de Dirichlet est que G peut être marginalisée pour obtenir directement la distribution des vecteurs de paramètres $\boldsymbol{\theta}_k$ en utilisant la représentation en urne de Polya.

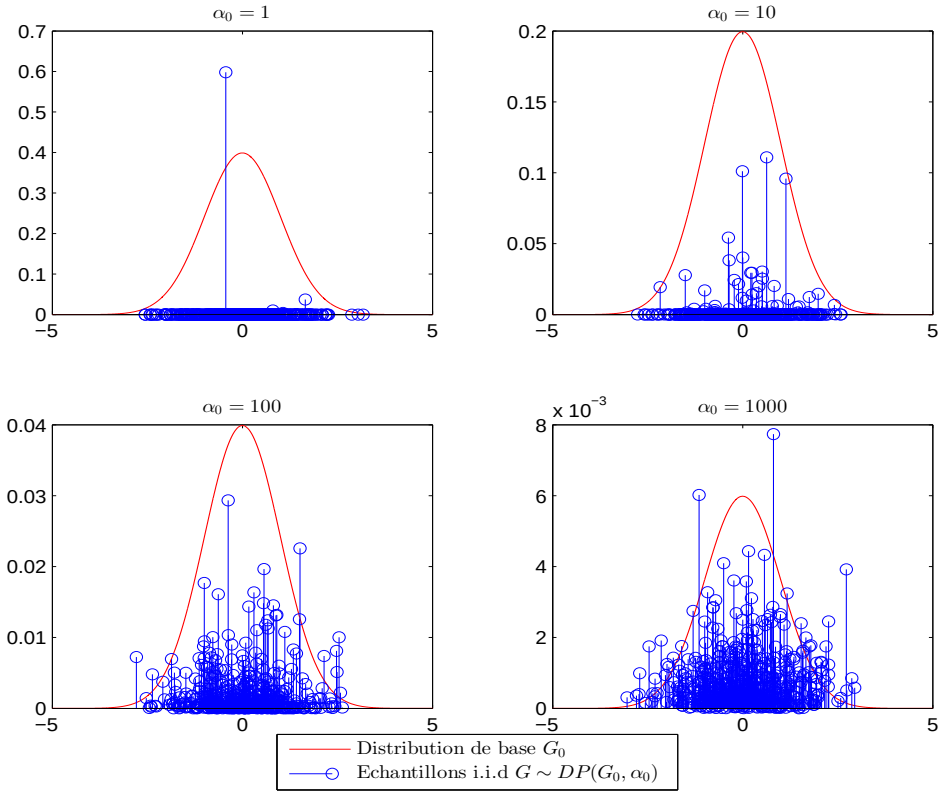

FIGURE 3.4 – Réalisations d'un processus de Dirichlet de distribution de base $G_0 = \mathcal{N}(0,1)$ pour différentes valeurs du paramètre d'échelle α_0.

3.2.4 Urne de Polya

Considérons la séquence $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k$ *i.i.d* selon une distribution inconnue G distribuée selon un processus de Dirichlet, soit :

$$\begin{aligned} \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k | G \sim G \\ G \sim \mathcal{DP}(\alpha_0, G_0) \end{aligned} \tag{3.11}$$

Comme les mesures de probabilités échantillonnées à partir d'un processus de Dirichlet sont discrètes, il y a une probabilité non-nulle d'observer des valeurs identiques parmi les échantillons $\{\boldsymbol{\theta}_k\}_{k=1,...,+\infty}$.
Blackwell et MacQueen [BLA73] ont montré que la distribution prédictive des vecteurs de paramètres $\boldsymbol{\theta}_k$ peut être obtenue en marginalisant G et s'exprime à partir de la représentation en urne de Polya comme suit :

$$\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_{1:k} \sim \frac{\alpha_0}{\alpha_0 + k} G_0 + \frac{1}{\alpha_0 + k} \sum_{i=1}^{k} \delta_{\boldsymbol{\theta}_i} \tag{3.12}$$

Ainsi, conditionnellement aux valeurs précédentes des variables $\boldsymbol{\theta}_{1:k}$, le nouvel échantillon $\boldsymbol{\theta}_{k+1}$ est égal à un échantillon déjà tiré avec la probabilité $\frac{k}{\alpha_0+k}$ ou est distribué selon G_0 avec la probabilité $\frac{\alpha_0}{\alpha_0+k}$. Par conséquent, plus une réalisation a été tirée dans le passé, plus elle a de chance d'être tirée de nouveau.

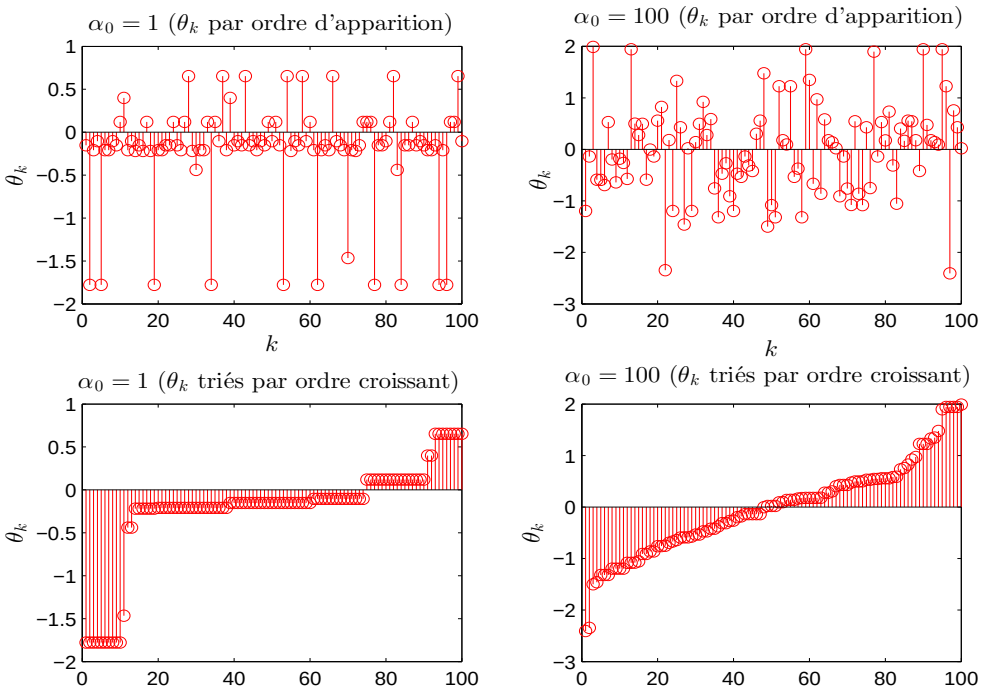

FIGURE 3.5 – Séquences d'échantillons obtenues à partir de la représentation en urne de Polya pour $\alpha_0 = 1$ ou $\alpha_0 = 100$ et $G_0 \sim \mathcal{N}(0, 1)$.

Comme illustré par la figure 3.5, plusieurs vecteurs peuvent prendre des valeurs identiques et ainsi être associés au même atome $\boldsymbol{U}_j$. Cet effet dit de renforcement est plus ou moins accentué selon la valeur du paramètre d'échelle α_0 du processus de Dirichlet. On peut montrer que le nombre moyen de valeurs distinctes à l'instant k tend vers $\alpha_0 \log(1 + \frac{k}{\alpha_0})$ lorsque k tend vers l'infini [ANT74].
Par conséquent, le paramètre d'échelle α_0 joue un rôle clé. Si α_0 tend vers zéro, la majorité des échantillons prennent la même valeur alors que si α_0 tend vers l'infini, les échantillons deviennent indépendants et identiquement distribués selon la distribution de base G_0. La distribution de base joue également un rôle important puisqu'elle détermine les valeurs possibles des points supports.

L'effet de renforcement, encore appelé effet de *clustering*, peut être illustré par la métaphore dite du restaurant chinois.
Dans la suite, on appelera cluster, un ensemble de vecteurs de paramètres qui prennent la même valeur. A l'instant $k-1$, on note $J(k-1)$ le nombre de valeurs distinctes de $\{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{k-1}\}$, notées $\{\boldsymbol{\theta}^*_1, ..., \boldsymbol{\theta}^*_{J(k-1)}\}$. De plus, on introduit $c_{k-1} \in \{1, ..., J(k-1)\}$ comme étant une variable discrète d'assignation à l'instant $k-1$ telle que $\boldsymbol{\theta}_{k-1} = \boldsymbol{\theta}^*_{c_{k-1}}$. Alors, la distribution prédictive de c_k sachant $c_{1:k-1}$ peut être directement calculée en utilisant le processus du restaurant chinois [ALD83] comme suit :

$$Pr(c_k|c_{1:k-1}, \alpha_0) = \frac{1}{\alpha_0 + k - 1} \sum_{j=1}^{k-1} \delta_{c_j}(c_k) + \frac{\alpha_0}{\alpha_0 + k - 1} \delta_{J(k-1)+1}(c_k) \tag{3.13}$$

En d'autres termes, l'indice du cluster à l'instant k est :

$$c_k = \begin{cases} J(k-1)+1 & \text{avec la probabilité } \frac{\alpha_0}{\alpha_0+k-1} \\ i & \text{avec la probabilité } \frac{\sum_{j=1}^{k-1} \delta_{c_j}(i)}{\alpha_0+k-1} \text{ pour } 1 \leq i \leq J(k-1) \end{cases} \tag{3.14}$$

L'analogie au restaurant chinois s'interprète comme suit : considérons $\boldsymbol{\theta}_k$ comme le plat servi au $k^{\text{ème}}$ client dans un restaurant chinois où il y a une infinité de tables et où à chaque table un plat unique est servi. Alors, chaque nouveau client choisit soit une table déjà occupée avec une probabilité proportionnelle au nombre de personnes présentes à cette table, soit une nouvelle table non occupée avec une probabilité proportionnelle à α_0.

Les propriétés de l'urne de Polya permettent une formulation simple des mécanismes d'inférence bayésienne. Ainsi, en combinant (3.12) et (3.14), on peut écrire :

$$\begin{aligned} \boldsymbol{\theta}_k | \{c_k = J(k-1)+1\} &\sim G_0 \\ \boldsymbol{\theta}_k | \{c_k = i\} &\sim \delta_{\boldsymbol{\theta}_i^*}(\boldsymbol{\theta}_k) \text{ pour } 1 \leq i \leq J(k-1) \end{aligned} \tag{3.15}$$

Dans la partie suivante, nous utilisons le formalisme précédemment introduit afin de présenter les modèles BNP pour l'estimation des statistiques du bruit de modèle.

3.3 Estimation non-paramétrique de la matrice de covariance du bruit de modèle

3.3.1 Description du problème

Nous considérons un système dont la REE est la suivante :

$$\begin{aligned} \boldsymbol{x}_k = F_k \boldsymbol{x}_{k-1} + \boldsymbol{v}_k \quad &\rightarrow \quad \boldsymbol{x}_k | \boldsymbol{x}_{k-1} \sim p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k; F_k \boldsymbol{x}_{k-1}, Q_k) \\ \boldsymbol{y}_k = H_k \boldsymbol{x}_k + \boldsymbol{w}_k \quad &\rightarrow \quad \boldsymbol{y}_k | \boldsymbol{x}_k \sim p(\boldsymbol{y}_k | \boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k) \end{aligned} \tag{3.16}$$

où $\boldsymbol{x}_k$ désigne le vecteur d'état à l'instant k, F_k la matrice de transition, $\boldsymbol{y}_k$ le vecteur d'observation et H_k la matrice d'observation. De plus, notons que les matrices des bruits de modèle $\boldsymbol{v}_k$ et de mesure $\boldsymbol{w}_k$, qui sont supposés non-corrélés, sont notées Q_k et R_k respectivement.

Dans un cas général, l'objectif est d'estimer récursivement le vecteur d'état $\boldsymbol{x}_k$ à partir de la séquence des observations bruitées $\boldsymbol{y}_{1:k}$. Pour cela, des informations *a priori* sur le système et plus particulièrement sur les statistiques du bruit doivent être disponibles. Comme un réglage inapproprié des matrices de covariance des bruits de modèle et de mesure peut conduire à une dégradation de l'estimation, une attention particulière a été apportée ces dernières décennies à la manière de les estimer. Plusieurs méthodes ont été proposées. Ce sont soit des méthodes dites *off-line* soit des méthodes *on-line*.
D'une part, les méthodes d'estimation *off-line* incluent les méthodes de sous-espaces [Van94] ou les méthodes fondées sur les erreurs de prédiction [Dun08]. D'autre part, parmi les méthodes dites *on-line*, on distingue les approches bayésiennes et non-bayésiennes. Dans un cadre non-bayésien, aucun modèle *a priori* n'est supposé pour les matrices de covariance. En général,

les estimations du vecteur d'état et des matrices de covariance sont réalisées à partir de deux algorithmes intéractifs de telle façon que les sorties de l'un soient utilisées comme entrées de l'autre. Parmi eux, certains algorithmes exploitent les calculs de corrélation alors que d'autres sont fondés sur le maximum de vraisemblance [Meh70, Sol14]. Dans la suite, nous traitons exclusivement de méthodes bayésiennes. Celles-ci estiment un vecteur d'état étendu $\boldsymbol{X}_k$ qui contient à la fois le vecteur d'état $\boldsymbol{x}_k$ et les paramètres d'intérêt tels que les élements de Q_k par exemple. Dans ce cas, le problème d'inférence se ramène au calcul récursif de la densité jointe *a posteriori* $p(\boldsymbol{X}_k|\boldsymbol{y}_{1:k})$.
En supposant que $\boldsymbol{x}_k$ et Q_k sont indépendants et en considérant par exemple un modèle Markovien pour Q_k, la probabilité de transition du vecteur d'état étendu peut s'exprimer comme suit :

$$p(\boldsymbol{X}_k|\boldsymbol{X}_{k-1}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, Q_k)p(Q_k|Q_{k-1}) \tag{3.17}$$

Un modèle d'évolution *a priori* pour Q_k est alors nécessaire.
Dans le cas scalaire où $Q_k = \sigma_k^2$, un choix standard et par exemple considéré dans [Ver02] et [Car07], consiste en la mise à jour suivante :

$$\log(\sigma_k) = \log(\sigma_{k-1}) + \lambda_k, \tag{3.18}$$

où λ_k est un bruit blanc gaussien. Dans un cas plus général, si la matrice Q_k est supposée diagonale, chaque composante peut suivre ce modèle là. Il convient particulièrement bien pour représenter des variations lentes pour des paramètres à valeurs positives. Cependant, quelle que soit la densité de probabilité *a priori* sélectionnée, l'inférence jointe du vecteur d'état et de la matrice de covariance est un problème non linéaire non-gaussien. Dans ce cas, la distribution *a posteriori* n'admet pas d'expression analytique. Un filtre particulaire est alors généralement utilisé [Gor93].

La plupart des approches bayésiennes pour l'estimation de matrices de covariance s'appliquent seulement à des variations lentes ou à des matrices constantes par morceaux comme dans [Gir10]. Cependant, cette hypothèse est assez restrictive dans bon nombre d'applications comme le pistage de cibles manœuvrantes où les matrices de covariance peuvent changer rapidement d'une valeur à une autre au cours du temps.
Nous proposons d'aborder ce problème à partir d'un nouveau modèle hiérarchique permettant de traiter les sauts fréquents de paramètres qui constituent la REE du système. On considère alors que le nombre *a priori* de modes, c'est-à-dire de valeurs différentes prises par les paramètres, n'est pas limité. Plus précisément, le modèle hiérarchique proposé inclut des distributions bayésiennes non-paramétriques fondées sur les processus de Dirichlet.
Dans la suite, nous traitons tout d'abord des modèles BNP pour l'estimation de matrice de covariance dans le cas général, puis nous développons le modèle BNP dans le cas où l'on suppose que la matrice a une structure prédéfinie et qu'elle se caractérise uniquement par un nombre réduit d'hyperparamètres.

3.3.2 Représentation dans l'espace d'état non-paramétrique

Considérons le cas où la matrice de transition est connue et fixée contrairement à la matrice de covariance. Cette dernière est supposée pouvoir prendre différentes valeurs qui se reproduisent au cours du temps. Dans la suite, on appellera mode du système une REE correspondant à une réalisation donnée de la matrice de covariance du bruit de modèle.

Soit $\boldsymbol{\theta}_k$ un vecteur contenant des variables latentes qui représentent tous les paramètres inconnus de la REE (3.16), et plus particulièrement dans notre cas, les paramètres de la matrice de covariance du bruit de modèle Q_k.
Afin de prendre en compte les changements de mode et l'apparition de nouveaux modes à mesure que de nouvelles observations sont disponibles, une modélisation non-paramétrique est utilisée pour estimer Q_k. Ainsi, nous considérons que $\boldsymbol{\theta}_k$ suit une distribution inconnue G qui est supposée aléatoire et vérifie $G \sim \mathcal{DP}(G_0, \alpha_0)$ avec G_0 la distribution de base et α_0 le paramètre d'échelle du processus de Dirichlet. Ainsi, la séquence des vecteurs $\boldsymbol{\theta}_k$ suit le modèle hiérarchique suivant :

$$G \sim \mathcal{DP}(G_0, \alpha_0) \tag{3.19}$$

$$\boldsymbol{\theta}_k | G \overset{i.i.d}{\sim} G(\boldsymbol{\theta}_k) \tag{3.20}$$

Les distributions G sont infinies et discrètes et peuvent s'exprimer à partir de la représentation en stick-breaking (3.7)-(3.9). De plus, comme montré par Blackwell *et al.* dans [Bla73], l'inférence de $\boldsymbol{\theta}_k$ peut s'exprimer directement à partir de la représentation en urne de Polya (3.13)-(3.15) présentée dans la partie 3.3 sans faire explicitement intervenir la distribution G. Dans ce cas, les différentes valeurs de $\boldsymbol{\theta}_k$ peuvent être regroupées en clusters indexés par la variable c_k dont la distribution suit (3.13).
Le modèle d'estimation (3.16) peut donc être étendu en intégrant la modélisation BNP de $\boldsymbol{\theta}_k$:

$$Pr[c_k = J(k-1)+1 | c_{0:k-1}] = \frac{\alpha_0}{\alpha_0 + k - 1} \tag{3.21}$$

$$Pr[c_k = i | c_{0:k-1}] = \frac{\sum_{j=1}^{k-1} \delta_{c_j}(i)}{\alpha_0 + k - 1} \text{ pour } 1 \leq i \leq J(k-1) \tag{3.22}$$

$$\boldsymbol{\theta}_k | \{c_k = J(k-1)+1\} \sim G_0 \tag{3.23}$$

$$\boldsymbol{\theta}_k | \{c_k = i\} \sim \delta_{\boldsymbol{\theta}_i^*}(\boldsymbol{\theta}_k) \text{ pour } i \leq J(k-1) \tag{3.24}$$

$$\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \boldsymbol{\theta}_k \sim \mathcal{N}(\boldsymbol{x}_k; F_k \boldsymbol{x}_{k-1}, Q_k(\boldsymbol{\theta}_k)) \tag{3.25}$$

$$\boldsymbol{y}_k | \boldsymbol{x}_k \sim \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k) \tag{3.26}$$

Si l'on définit le vecteur d'état étendu $\boldsymbol{X}_k = [c_k \ \boldsymbol{\theta}_k \ \boldsymbol{x}_k]$, l'objectif est alors de calculer séquentiellement la distribution de probabilité *a posteriori* $p(\boldsymbol{X}_k | \boldsymbol{y}_{1:k})$ à partir des observations $\boldsymbol{y}_{1:k}$.

Par ailleurs, afin de faciliter les développements calculatoires, plutôt que d'estimer Q_k, il est équivalent d'estimer son inverse plus communément appelée matrice de précision et notée Γ_k. Aussi, dans la suite de ce chapitre, nous détaillerons les calculs en estimant Γ_k. Comme la relation entre Q_k et Γ_k est un simple changement de variable, ce choix n'a pas d'influence sur le résultat final. Néanmoins, le choix de la distribution de base G_0 du processus de Dirichlet en

dépend. Plusieurs distributions peuvent être envisagées mais dans le cadre d'une estimation par filtrage particulaire, il s'avère plus judicieux d'opter pour des distributions conjuguées. Elles ont l'avantage de rendre possibles les calculs de la loi de proposition optimale.
Comme le bruit d'état est gaussien, G_0 est généralement choisie [2] comme étant une distribution de Wishart. En statistique, la distribution de Wishart, dont l'expression est rappelée ci-après, est la conjuguée de l'inverse de la matrice de covariance d'un vecteur aléatoire normal multivarié. Pour toute matrice symétrique definie positive X de taille $n \times p$, la densité de probabilité de la distribution de Wishart est définie comme suit :

$$\mathcal{W}(X; W_0, v_0) = B(W_0, v_0)|X|^{\frac{v_0-p-1}{2}} \exp\left(-\frac{1}{2}Tr[W_0^{-1}X]\right) \tag{3.27}$$

où W_0 et v_0 sont la matrice d'échelle et le nombre de degrés de liberté respectivement avec $B(W_0, v_0)$ la constante de normalisation définie comme suit :

$$B(W_0, v_0) = \frac{1}{2^{\frac{v_0 p}{2}}|W_0|^{\frac{v_0}{2}}\Gamma_p(\frac{v_0}{2})} \tag{3.28}$$

où $\Gamma_p(\cdot)$ est la fonction Gamma multivariée.
Les équations (3.23)-(3.25) du modèle hiérarchique initial deviennent :

$$\Gamma_k|\{c_k = J(k-1)+1\} \sim G_0 = \mathcal{W}(\Gamma_k; W_0, v_0) \tag{3.29}$$

$$\Gamma_k|\{c_k = i\} \sim \delta_{\Gamma_i^*}(\Gamma_k) \text{ pour } i \leq J(k-1) \tag{3.30}$$

$$\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k \sim \mathcal{N}(\boldsymbol{x}_k; F_k\boldsymbol{x}_{k-1}, \Gamma_k^{-1}) \tag{3.31}$$

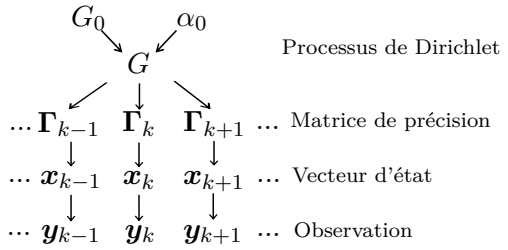

FIGURE 3.6 – Représentation graphique du modèle hiérarchique

Il est alors possible d'estimer conjointement c_k, Γ_k et $\boldsymbol{x}_k$ par filtrage particulaire en tirant profit des relations entre ces différentes variables, cf. figure 3.6. Contrairement à Caron [CAR08] qui propose la mise en œuvre de son approche à partir d'un filtre particulaire Rao-Blackwellizé, nous proposons d'estimer toutes les variables inconnues par filtrage particulaire en utilisant la distribution d'importance optimale. L'algorithme auquel nous aboutissons est ainsi une variante qui se distingue du travail originel de Caron. Cependant, nous avons fait le choix de présenter les détails de cette démarche en annexe B pour ne pas alourdir le chapitre 3 et pour mettre en avant un nouvel algorithme dans la suite de ce chapitre de thèse. Ce dernier se fonde sur deux caractéristiques :
La première consiste en un nouveau modèle hiérarchique tirant profit que dans de nombreuses

2. Si c'était la matrice de covariance Q_k qui était estimée, alors G_0 serait généralement choisie comme étant une distribution inverse Wishart.

applications et notamment en pistage de cibles, la structure de la matrice de covariance du bruit de modèle a des formes fonctionnelles prédéfinies. Ainsi, seulement un nombre réduit d'hyperparamètres caractérisant Γ_k constituent le vecteur $\boldsymbol{\theta}_k$.
La seconde est liée au choix de la distribution d'importance optimale des particules. Dans la mesure où ce problème non-linéaire et non-gaussien est traité par filtrage particulaire, nous détaillons le calcul de la distribution d'importance optimale des particules.

3.4 Cas des matrices de covariance avec une structure prédéfinie

Lorsque la dimension de la matrice Γ_k (égale à la dimension de Q_k) augmente, il est nécessaire d'augmenter le nombre de particules pour conserver la même précision d'estimation. Pour éviter que le coût calculatoire dépende de la dimension du problème d'estimation considéré, nous avons proposé dans [Mag15b] une alternative tirant profit de la structure de la matrice. En effet, dans de nombreuses applications comme la poursuite d'objets, la matrice de covariance ne peut prendre que des formes fonctionnelles connues et traduisant les relations physiques entre les différents paramètres constituant le vecteur d'état. Ainsi, seulement un nombre réduit d'hyperparamètres doivent être estimés. Dans la suite, nous considérons la décomposition suivante :

$$Q_k = \frac{1}{\gamma_k}\tilde{Q} \text{ soit } \Gamma_k = \gamma_k\tilde{\Gamma} \tag{3.32}$$

où $\tilde{Q}$ et $\tilde{\Gamma}$ sont des matrices dont les valeurs sont constantes et connues de l'utilisateur alors que γ_k est une variable scalaire auquel il est fait référence par la suite sous le terme de paramètre de précision.
Ces décompositions permettent de réduire la dimension de l'espace de paramètres à explorer. Pour illustrer notre démarche, prenons le cas du MRU. Sa matrice de covariance du bruit de modèle satisfait :

$$Q^{MRU} = \sigma^2_{MRU}\tilde{Q}^{MRU} = \sigma^2_{MRU}\begin{bmatrix} \frac{T^3}{2} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \tag{3.33}$$

soit encore :

$$\Gamma^{MRU} = \gamma_{MRU}\tilde{\Gamma}^{MRU} = \gamma_{MRU}\begin{bmatrix} \frac{T^3}{2} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}^{-1} \tag{3.34}$$

où T est la période d'échantillonnage qi est *a priori* connue et le paramètre de précision $\gamma_{MRU} = \frac{1}{\sigma^2_{MRU}}$ est un scalaire et correspond à l'inverse de la variance de l'accélération.

Par ailleurs, plutôt que de ne considérer qu'une seule famille de mouvements, ce qui pourrait s'avérer être un choix trop restrictif, nous supposons que le système peut changer de famille de modèles au cours du temps. Dans ce cas, chacune de ces familles de modèles est caractérisée par des expressions différentes des matrices de covariance qui vérifient toutes (3.32).
Par exemple, on peut considérer les modèles du type MRU, MUA ou Singer explicités au chapitre 2. Au sein de chaque famille, toutes les REEs partagent la même forme fonctionnelle $\tilde{\Gamma}$ qui est *a priori* connue mais elles diffèrent les unes des autres par le paramètre de précision γ_k.
Les distributions de ces hyperparamètres sont modélisées selon des processus de Dirichlet et estimées par filtrage particulaire. Dans un premier temps, nous avons proposé dans [Mag15b]

d'utiliser la distribution *a priori* pour propager les particules. Puis, afin d'améliorer l'estimation pour le même nombre de particules, nous avons utilisé dans [MAG16a] une approximation de la distribution d'importance optimale.
Dans la sous-partie suivante, nous détaillons le calcul de la distribution d'importance optimale du filtre particulaire qui permet d'estimer conjointement $\boldsymbol{X}_k$ et γ_k.

3.4.1 Modèle hiérarchique considéré

Dans la suite, nous modélisons l'évolution de la cible à partir de plusieurs familles de modèles dont la matrice de covariance est connue à un paramètre près.

Soit z_k l'indice de la famille de modèles à l'instant k tel que $z_k \in \{1, ..., M\}$. La séquence $\{z_k\}_{k>0}$ est supposée être une chaîne de Markov de matrice de probabilités de transition $\Pi = \{\pi_{ij}\}_{i=1,...,M}^{j=1,...,M}$. Contrairement aux approches standards à modèles multiples, z_k ne fait pas référence à un modèle avec une matrice de précision du bruit d'état prédéfinie mais à une famille qui comprend une infinité de modèles, chacun d'entre eux correspondant à une valeur différente du paramètre de précision.

Conditionnellement à z_k, la distribution du paramètre de précision γ_k est modélisée par un processus de Dirichlet. On considère alors autant de processus de Dirichlet que de familles de modèles. Pour chaque famille m, avec $m = 1, ..., M$, les réalisations $\{G^m\}_{m=1,...,M}$ sont caractérisées par une distribution de base G_0^m et un facteur d'échelle α_0^m :

$$G^m \sim \mathcal{DP}(G_0^m, \alpha_0^m) \text{ pour } m = 1, ..., M \tag{3.35}$$

$$\gamma_k | \{z_k, \{G^m\}_{m=1,...,M}\} \sim G^{z_k}(\gamma_k) \tag{3.36}$$

À chaque instant, conditionnellement à l'indice de la famille de modèles z_k, la modélisation BNP de γ_k est pilotée par l'indice de cluster $c_k^{z_k}$, la distribution de base $G_0^{z_k}$ et le facteur d'échelle $\alpha_0^{z_k}$ du processus de Dirichet.
Les relations entre z_k, γ_k, $\boldsymbol{x}_k$ et $\boldsymbol{y}_k$ peuvent être décrites par le modèle hiérarchique suivant :

$$z_k | z_{k-1} \sim \pi_{z_k z_{k-1}} \text{ pour } z_k = 1, ..., M \tag{3.37}$$

$$Pr[c_k^{z_k} = J_{z_{k-1}}(k-1) + 1 | c_{0:k-1}^{z_k}] = \frac{\alpha_0^{z_k}}{\alpha_0^{z_k} + k - 1} \tag{3.38}$$

$$Pr[c_k^{z_k} = i | c_{0:k-1}^{z_k}] = \frac{\sum_{j=1}^{k-1} \delta_{c_j^{z_k}}(i)}{\alpha_0^{z_k} + k - 1} \text{ pour } 1 \leq i \leq J_{z_{k-1}}(k-1) \tag{3.39}$$

$$\gamma_k | \{c_k^{z_k} = J_{z_{k-1}}(k-1) + 1, z_k\} \sim G_0^{z_k} = \mathcal{G}(\gamma_k; a_{z_k}, b_{z_k}) \tag{3.40}$$

$$\gamma_k | \{c_k^{z_k} = i, z_k\} \sim \delta_{\gamma_i^{*,z_k}}(\gamma_k) \tag{3.41}$$

$$\boldsymbol{x}_k | \{\boldsymbol{x}_{k-1},\ \gamma_k,\ z_k\} \sim \mathcal{N}(\boldsymbol{x}_k; F_k^{z_k} \boldsymbol{x}_{k-1}, (\gamma_k \tilde{\Gamma}^{z_k})^{-1}) \tag{3.42}$$

$$\boldsymbol{y}_k | \boldsymbol{x}_k \sim \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k) \tag{3.43}$$

où γ_i^{*,z_k} et $J_{z_k}(k)$ désignent le $i^{\text{ème}}$ cluster et le nombre de clusters à l'instant k pour la famille de modèles z_k respectivement et $\mathcal{G}$ désigne une densité de probabilité Gamma.
Notons que le choix d'une loi *a priori* gamma pour le paramètre de précision est classique. De plus, il facilite le calcul de la loi de proposition optimale. En effet, la distribution Gamma est

utilisée comme distribution conjuguée pour différents paramètres tels que les paramètres d'inverse variance [Bis06].

Selon la représentation en urne de Polya, les distributions inconnues $\{G^m\}_{m=1,\dots,M}$ peuvent être marginalisées du modèle hiérarchique. Cependant, le changement entre les différentes familles de modèles doit être pris en compte. La distribution de prédiction de γ_k devient alors :

$$\gamma_k|\gamma_{1:k-1}, z_k \sim \frac{1}{\alpha_0^{z_k} + n_{z_k}} \sum_{\substack{j=1 \\ s.t.\ z_j=z_k}}^{k-1} \delta_{\gamma_j}(\gamma_k) + \frac{\alpha_0^{z_k}}{\alpha_0^{z_k} + n_{z_k}} G_0^{z_k}(\gamma_k) \text{ pour } 1 \leq i \leq J_{z_{k-1}}(k-1) \quad (3.44)$$

où n_{z_k} est le nombre de fois où la famille de modèles z_k est apparue auparavant.
Le modèle hiérarchique défini par (3.38)-(3.43) se réduit ainsi à (3.37), (3.44), (3.42) et (3.43). Sur la base de ce modèle hiéarchique, il s'agit ensuite d'estimer par filtrage particulaire la distribution jointe *a posteriori* de tous les paramètres inconnus $p(\boldsymbol{x}_{0:k}, z_{0:k}, \gamma_{0:k}|\boldsymbol{y}_{1:k})$.

Comme expliqué dans le chapitre 1, le choix de la loi de proposition est crucial pour garantir une bonne précision du filtre particulaire utilisé avec un nombre limité de particules. Dans la suite, nous optons pour la loi dite optimale au sens défini par A. Doucet [Dou00b].

3.4.2 Calcul de la distribution d'importance optimale

Dans un premier temps, par souci de simplicité, l'indice de la famille de modèles z_k peut être directement échantillonné selon les probabilités de transition contenues dans la TPM notée $\Pi = \{\pi_{ij}\}_{i=1,\dots,M}^{i=1,\dots,M}$ comme suit :

$$z_k|z_{k-1} \sim \pi_{z_k z_{k-1}} \quad (3.45)$$

Il s'agit alors de montrer comment effectuer l'échantillonnage conjoint du vecteur d'état $\boldsymbol{x}_k$ et du paramètre de précision γ_k en prenant en compte l'information apportée par la mesure courante. En pratique, cela peut être fait en trois étapes de la manière suivante :

$$c_k^{z_k} \sim Pr(c_k^{z_k}|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, c_{0:k-1}^{z_k}, \boldsymbol{y}_k) \quad (3.46)$$

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, c_k^{z_k}, \boldsymbol{y}_k) \quad (3.47)$$

$$\gamma_k \sim p(\gamma_k|\boldsymbol{x}_{0:k}, z_k, c_k^{z_k}, \gamma_{0:k-1}, \boldsymbol{y}_k) \quad (3.48)$$

Dans la suite, les expressions (3.46)-(3.48) sont détaillées. Pour chacune de ces distributions, on considère deux cas : $c_k^{z_k} = J_{z_{k-1}}(k-1)+1$ qui fait référence à un nouveau cluster et $c_k^{z_k} = i$ qui dénote le $i^{\text{ème}}$ cluster existant.

3.4.2.1 Echantillonnage du paramètre de précision

Il s'agit de calculer $p(\gamma_k|\boldsymbol{x}_{0:k}, z_k, c_k^{z_k}, \gamma_{0:k-1}, \boldsymbol{y}_k)$. Or, notons que γ_k ne dépend pas de l'observation $\boldsymbol{y}_k$ conditionnellement à $\boldsymbol{x}_{0:k}$, z_k, $c_k^{z_k}$ et $\gamma_{0:k-1}$. De plus, tirant profit des propriétés markoviennes de l'état, on développe en réalité l'expression suivante $p(\gamma_k|\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, z_k, c_k^{z_k}, \gamma_{0:k-1})$.

D'une part, si $c_k^{z_k} = J_{z_{k-1}}(k-1)+1$, c'est-à-dire si un nouveau cluster est créé, alors :

$$p(\gamma_k|\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \gamma_{0:k-1}) \propto p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, \gamma_k)p(\gamma_k|c_k^{z_k} = J_{z_{k-1}}(k-1)+1, z_k) \tag{3.49}$$

avec :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, \gamma_k) &\sim \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}^{z_k}, (\gamma_k\tilde{\Gamma}^{z_k})^{-1}) && (3.50)\\ p(\gamma_k|c_k^{z_k} = J_{z_{k-1}}(k-1)+1, z_k) &\sim \mathcal{G}(\gamma_k; a_{z_k}, b_{z_k}) && (3.51) \end{aligned}$$

où $\hat{\boldsymbol{x}}_{k|k-1}^{z_k} = F_k^{z_k}\boldsymbol{x}_{k-1}$ est le vecteur d'état prédit à l'instant k pour la famille z_k.

En combinant (3.50) et (3.51) puis en tirant profit du fait que la loi gamma est conjuguée de la loi gaussienne pour le paramètre de précision, $p(\gamma_k|\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \gamma_{0:k-1})$ peut s'exprimer comme une nouvelle distribution Gamma :

$$\begin{aligned} &p(\gamma_k|\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \gamma_{0:k-1})\\ &\quad\propto p(\gamma_k\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \gamma_{0:k-1})\\ &\quad= (2\pi)^{-\frac{n_x}{2}}|\gamma_k\tilde{\Gamma}^{z_k}|^{\frac{1}{2}}exp\left(-\frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})^T(\gamma_k\tilde{\Gamma}^{z_k})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})\right) \times \frac{b_{z_k}^{a_{z_k}}exp(-b_{z_k}\gamma_k)}{\Gamma(a_{z_k})}\gamma_k^{a_{z_k}-1}\\ &\quad= \mathcal{G}(\gamma_k; a_{z_k}^{post}, b_{z_k}^{post}) \end{aligned} \tag{3.52}$$

avec :

$$\begin{aligned} a_{z_k}^{post} &= a_{z_k} + \frac{n_x}{2}\\ b_{z_k}^{post} &= b_{z_k} + \frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})^T\tilde{\Gamma}^{z_k}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k}) \end{aligned} \tag{3.53}$$

D'autre part, si $c_k^{z_k} = i$ pour $i \leq J_{z_{k-1}}(k-1)$, c'est-à-dire si un cluster déjà existant est sélectionné, comme on a :

$$\gamma_k|\{c_k^{z_k} = i, \gamma_{0:k-1}\} \sim \delta_{\gamma_i^*, z_k}(\gamma_k) \tag{3.54}$$

alors :

$$\begin{aligned} &p(\gamma_k|\boldsymbol{x}_k, z_k, c_k^{z_k} = i, \gamma_{0:k-1}) \propto p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, \gamma_k)p(\gamma_k|\gamma_{0:k-1}, z_k, c_k^{z_k} = i)\\ &p(\gamma_k|\boldsymbol{x}_k, z_k, c_k^{z_k} = i, \gamma_{0:k-1}) = \delta_{\gamma_i^*, z_k}(\gamma_k) \end{aligned} \tag{3.55}$$

3.4.2.2 Echantillonnage du vecteur d'état

Développons à présent (3.47), i.e. $p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, c_k^{z_k}, \boldsymbol{y}_k)$. Comme précédemment, étudions deux cas : celui où $c_k^{z_k} = J_{z_{k-1}}(k-1)+1$ et celui où $c_k^{z_k} = i$ pour $i \leq J_{z_{k-1}}(k-1)$.
D'une part, si $c_k^{z_k} = J_{z_{k-1}}(k-1)+1$, alors on a :

$$\begin{aligned} &p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \boldsymbol{y}_k)\\ &\qquad\propto p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) \end{aligned} \tag{3.56}$$

avec

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) \sim \mathcal{N}(\boldsymbol{y}_k; H_k\boldsymbol{x}_k, R_k) \tag{3.57}$$

et le second terme du produit de (3.56) peut s'exprimer comme suit :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J(k-1)+1) &= \int_{\gamma_k} p(\boldsymbol{x}_k, \gamma_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1)d\gamma_k \\ &= \int_{\gamma_k} p(\boldsymbol{x}_k|\gamma_k, \boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1)p(\gamma_k|z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1)d\gamma_k \end{aligned} \tag{3.58}$$

Or, en utilisant (3.52) et (3.50), il vient :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) = \int_{\gamma_k} \frac{(2\pi)^{-n_x/2}|\tilde{\Gamma}^{z_k}|^{1/2}\Gamma(a_{z_k}^{post})b_{z_k}^{a_{z_k}}}{(b_{z_k}^{post})^{a_{z_k}^{post}}\Gamma(a_{z_k})}\mathcal{G}(\gamma_k; a_{z_k}^{post}, b_{z_k}^{post})d\gamma_k \tag{3.59}$$

Puis, en intégrant par rapport à γ_k, on a alors :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) = \frac{(2\pi)^{-n_x/2}|\tilde{\Gamma}_{z_k}|^{1/2}\Gamma(a_{z_k}^{post})b_{z_k}^{a_{z_k}}}{(b_{z_k}^{post})^{a_{z_k}^{post}}\Gamma(a_{z_k})} \tag{3.60}$$

A ce stade, le seul terme qui dépend de $\boldsymbol{x}_k$ est $(b_{z_k}^{post})^{a_{z_k}^{post}}$.
En remplaçant $a_{z_k}^{post}$ et $b_{z_k}^{post}$ par leurs valeurs, d'après (3.53), on a la relation suivante :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) &\propto \left(b_{z_k} + \frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})^T\tilde{\Gamma}^{z_k}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})\right)^{-(a_{z_k}+\frac{n_x}{2})} \\ &\propto b_{z_k}\left(1 + \frac{b_{z_k}^{-1}}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})^T\tilde{\Gamma}^{z_k}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}^{z_k})\right)^{-(a_{z_k}+\frac{n_x}{2})} \end{aligned} \tag{3.61}$$

On identifie alors une distribution t matricielle dont l'expression est donnée en annexe D. Ainsi, on note :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) \propto T_{n_x,1}(\boldsymbol{x}_k; a_{z_k}, \hat{\boldsymbol{x}}_{k|k-1}^{z_k}, 2b_{z_k}, (\tilde{\Gamma}^{z_k})^{-1}) \tag{3.62}$$

Si le nombre de degrés de liberté est suffisamment élevé, (3.62) peut être approchée par une distribution gaussienne :

$$p(\boldsymbol{x}_k|\hat{\boldsymbol{x}}_{k|k-1}^{z_k}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1) \propto \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}^{z_k}, \frac{b_{z_k}}{a_{z_k}}(\tilde{\Gamma}^{z_k})^{-1}) \tag{3.63}$$

Finalement, en combinant (3.57) avec (3.63), (3.56) peut s'exprimer comme suit :

$$p(\boldsymbol{x}_k|c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \boldsymbol{x}_{0:k-1}, z_{0:k}, \gamma_{0:k-1}, \boldsymbol{y}_k) \approx \mathcal{N}(\boldsymbol{x}_k; m_x^0, \Sigma_x^0) \tag{3.64}$$

avec

$$\Sigma_x^0 = \left(H_k^T R_k^{-1} H_k + \frac{a_{z_k}}{b_{z_k}}\tilde{\Gamma}^{z_k}\right)^{-1} \tag{3.65}$$

$$m_x^0 = \Sigma_x^0\left(H_k^T R_k^{-1}\boldsymbol{y}_k + \frac{a_{z_k}}{b_{z_k}}\tilde{\Gamma}^{z_k}\hat{\boldsymbol{x}}_{k|k-1}^{z_k}\right) \tag{3.66}$$

Notons que du fait de l'approximation de la distribution t par une gaussienne, ce n'est pas exac-

tement la distribution d'importance optimale qui est calculée.

D'autre part, si $c_k^{z_k} = i$ pour $i \leq J_{z_{k-1}}(k-1)$:

$$p(\boldsymbol{x}_k|c_k^{z_k} = i, \boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, \boldsymbol{y}_k) \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k, z_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = i) \tag{3.67}$$

avec :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = i) &\sim \mathcal{N}(\boldsymbol{x}_k; F_k^{z_k}\boldsymbol{x}_{k-1}, (\gamma_i^{*,z_k}\tilde{\Gamma}^{z_k})^{-1}) & (3.68)\\ p(\boldsymbol{y}_k|\boldsymbol{x}_k) &\sim \mathcal{N}(\boldsymbol{y}_k; H_k\boldsymbol{x}_k, R_k) & (3.69)\end{aligned}$$

En combinant les deux gaussiennes, il vient :

$$p(\boldsymbol{x}_k|c_k^{z_k} = i, \boldsymbol{x}_{0:k-1}, z_{0:k-1}, \gamma_{0:k-1}, \boldsymbol{y}_k) \approx \mathcal{N}(\boldsymbol{x}_k; m_x^i, \Sigma_x^i) \tag{3.70}$$

avec

$$\begin{aligned} \Sigma_x^i &= \left(H_k^T R_k^{-1} H_k + \gamma_i^{*,z_k}\tilde{\Gamma}^{z_k}\right)^{-1} & (3.71)\\ m_x^i &= \Sigma_x^i \left(H_k^T R_k^{-1}\boldsymbol{y}_k + \gamma_i^{*,z_k}\tilde{\Gamma}^{z_k}\hat{\boldsymbol{x}}_{k|k-1}^{z_k}\right) & (3.72)\end{aligned}$$

3.4.2.3 Echantillonnage de l'indice du cluster

Intéressons-nous à présent à l'expression du terme (3.46), i.e. $Pr(c_k^{z_k}|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, \boldsymbol{y}_k)$.
Nous opérons une nouvelle fois en distinguant deux cas.
D'une part, si $c_k^{z_k} = J_{z_{k-1}}(k-1)+1$, alors :

$$\begin{aligned} &Pr(c_k^{z_k} = J_{z_{k-1}}(k-1)+1|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, \boldsymbol{y}_k)\\ &\quad\propto p(\boldsymbol{y}_k|c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}) \times Pr[c_k^{z_k} = J_{z_{k-1}}(k-1)+1|c_{0:k-1}^{z_k}]\end{aligned} \tag{3.73}$$

En utilisant la représentation en urne de Polya, on a :

$$Pr[c_k^{z_k} = J_{z_{k-1}}(k-1)+1|c_{0:k-1}^{z_k}] = \frac{\alpha_0^{z_k}}{\alpha_0^{z_k}+k-1} \tag{3.74}$$

Il vient alors :

$$\begin{aligned} &Pr(c_k^{z_k} = J_{z_{k-1}}(k-1)+1|\boldsymbol{x}_{0:k-1}, z_{0:k}, \gamma_{0:k-1}, \boldsymbol{y}_k)\\ &\qquad\propto p(\boldsymbol{y}_k|c_k^{z_k} = J_{z_{k-1}}(k-1)+1, \boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1})\frac{\alpha_0^{z_k}}{\alpha_0^{z_k}+k-1}\\ &\qquad\propto \frac{\alpha_0^{z_k}}{\alpha_0^{z_k}+k-1}\int_{\boldsymbol{x}_k} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, c_k^{z_k} = J_{z_{k-1}}(k-1)+1)d\boldsymbol{x}_k\end{aligned} \tag{3.75}$$

Finalement, en combinant les deux distributions gaussiennes définies par (3.57) et (3.63), on aboutit à :

$$\begin{aligned} Pr(c_k^{z_k} = J_{z_{k-1}}(k-1)+1|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, \boldsymbol{y}_k) \\ \propto \frac{\alpha_0^{z_k}}{\alpha_0^{z_k}+k-1}\mathcal{N}(\boldsymbol{y}_k; H_k\hat{\boldsymbol{x}}_{k|k-1}^{z_k}, R_k + \frac{b_{z_k}}{a_{z_k}}H_k(\tilde{\Gamma}^{z_k})^{-1}H_k^T) \end{aligned} \tag{3.76}$$

D'autre part, si $c_k^{z_k} = i$ pour $i \leq J_{z_{k-1}}(k-1)$, en procédant comme précédemment, il vient :

$$\begin{aligned} p(c_k^{z_k} = i|\boldsymbol{x}_{0:k-1}, z_k, \gamma_{0:k-1}, \boldsymbol{y}_k) &\propto p(\boldsymbol{y}_k|c_k^{z_k} = i, \boldsymbol{x}_{k-1}, z_{0:k}, \gamma_{0:k-1})Pr[c_k^{z_k} = i|c_{0:k-1}^{z_k}] \\ &= p(\boldsymbol{y}_k|c_k^{z_k} = i, \boldsymbol{x}_{k-1}, z_k, \gamma_{0:k-1})\frac{\#i}{\alpha_0^{z_k}+k-1} \\ &= \frac{\#i}{\alpha_0^{z_k}+k-1}\int_{\boldsymbol{x}_k}\int_{\gamma_k} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, \gamma_k)\delta_{\gamma_i^{*,z_k}}(\gamma_k)d\gamma_k d\boldsymbol{x}_k \\ &= \frac{\#i}{\alpha_0^{z_k}+k-1}\int_{\boldsymbol{x}_k} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, z_k, \gamma_i^{*,z_k})d\boldsymbol{x}_k \\ &= \frac{\#i}{\alpha_0^{z_k}+k-1}\mathcal{N}(\boldsymbol{y}_k; H_k\hat{\boldsymbol{x}}_{k|k-1}^{z_k}, R_k + H_k(\gamma_i^{*,z_k}\tilde{\Gamma}^{z_k})^{-1}H_k^T) \end{aligned} \tag{3.77}$$

où $\#i$ désigne le nombre de fois où $c_j = i$ pour $j \leq k$.

Etant donné les développements que nous avons effectués dans les parties 3.4.2.1, 3.4.2. et 3.4.2.3, on aboutit à un algorithme dont les étapes principales sont présentées dans **Algorithme 3.1** où N_p désigne le nombre de particules.

Algorithme 3.1 Estimation du paramètre de précision : une itération de l'algorithme (*Algo1*) proposé (tenant compte d'une structure prédéfinie de la matrice de covariance du bruit de modèle et fondé sur la loi optimale)

- Pour $i = 1, .., N_p$: Echantillonner la famille du modèle selon (3.45)
- Pour $i = 1, .., N_p$:
  - Calculer $P(i, 0) \underset{(3.76)}{=} Pr(c_k^{z_k} = J_{z_{k-1}}^i(k-1) + 1 | \boldsymbol{x}_{0:k-1}^i, \gamma_{0:k-1}^i, \boldsymbol{y}_k)$
  - pour $j = 1, ..., J_{z_{k-1}}^i(k-1)$
    Calculer $P(i, j) \underset{(3.77)}{=} p(c_k^{z_k} = j | \boldsymbol{x}_{0:k-1}^i, \gamma_{0:k-1}^i, \boldsymbol{y}_k)$
- Pour $i = 1, .., N_p$: Mettre à jour les poids d'importance

$$w_k^i = w_{k-1}^i \sum_{j=0}^{J_{z_{k-1}}^i(k-1)} P(i, j)$$

- Pour $i = 1, .., N_p$:
  Si $rand = \mathcal{U}(0, 1) \leq P(0)$:
  - Echantillonner $\boldsymbol{x}_k$ selon (3.64)
  - Echantillonner γ_k selon (3.52)
  - Mettre à jour le nombre de clusters $J_{z_k}^i(k) = J_{z_{k-1}}^i(k-1) + 1$

  sinon
  - Echantillonner $\boldsymbol{x}_k$ selon (3.70)
  - Echantillonner γ_k selon (3.55)
  - Mettre à jour le nombre de clusters $J_{z_k}^i(k) = J_{z_{k-1}}^i(k-1)$
- Pour $i = 1, .., N_p$: Normaliser les poids d'importance

$$w_k^i = \frac{w_k^i}{\sum_{j=1}^{N_p} w_k^j}$$

- Calculer $N_{eff} = \frac{1}{\left(\sum_{i=1}^{N} (w_k^i)^2\right)}$
- Si $N_{eff} \leq C$, où C est une constante à définir, alors ré-échantillonner.
  Le nouvel ensemble de particules est donc $\{\boldsymbol{x}_k^i, w_k^i = \frac{1}{N_p}\}$.

Dans la sous-partie suivante, notre approche est appliquée au pistage monocible. La loi déduite précédemment est utilisée à chaque itération de l'algorithme de filtrage particulaire pour propager les particules. Les autres étapes de ce filtre sont rappelées dans le chapitre 1.

3.5 Application au pistage d'une cible

Dans cette partie, nous présentons les résultats obtenus à partir de l'approche développée précédemment dans un contexte de pistage de cibles. Nous précisons dans un premier temps les deux protocoles de simulation considérés et donnons des éléments sur la manière dont sont réglés les algorithmes à tester. Puis, les résultats de simulation sont fournis en termes de complexité calculatoire et d'EQM. L'influence de certains paramètres comme la TPM ou le nombre de particules est ensuite étudiée.

3.5.1 Protocoles de simulation

La pertinence de notre approche est analysée en l'appliquant à des ensembles de trajectoires générés selon deux protocoles de simulation.

3.5.1.1 Génération de trajectoires selon le protocole 1

L'ensemble de trajectoires est dans le premier cas généré à partir d'un modèle de mouvement du second ordre. Plus particulièrement, le vecteur d'état est défini comme suit :

$$\boldsymbol{x}_k = [x_k, \dot{x}_k, \ddot{x}_k]^T \tag{3.78}$$

où x_k dénote la position, $\dot{x}_k$ la vitesse et $\ddot{x}_k$ l'accélération. De plus, la matrice de transition est de la forme :

$$F = \begin{bmatrix} 1 & T & T^2/2 \\ 1 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{3.79}$$

où T est la période d'échantillonnage. La matrice d'observation satisfait :

$$H = [1\ 0\ 0] \tag{3.80}$$

Finalement, la matrice de précision du bruit d'état est supposée pouvoir changer de valeurs au cours du temps avec une probabilité de 0.05 et chaque valeur est distribuée selon une distribution de Wishart de paramètres $W_0 = 10 I_{n_x}$ et $v_0 = 4$. De plus, la variance du bruit de mesure R est fixée à 25 m^2.

3.5.1.2 Génération de trajectoires selon le protocole 2

Dans la second cas, l'ensemble de trajectoires est généré à partir de $M = 3$ familles de modèles de mouvement à chaque instant : MRU, MUA et Singer. Les vecteurs d'état associés satisfont respectivement :

$$\boldsymbol{x}_k^{MRU} = [x_k, \dot{x}_k]^T \quad ; \quad \boldsymbol{x}_k^{MUA} = [x_k, \dot{x}_k, \ddot{x}_k]^T \quad ; \quad \boldsymbol{x}_k^{Sin} = [x_k, \dot{x}_k, \ddot{x}_k]^T \tag{3.81}$$

De plus, les matrices de transition et d'observation de ces familles de modèles sont définies comme suit :

$$\begin{aligned} F^{MRU} &= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \;;\; F^{MUA} = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \;;\; F^{Sin} = \begin{bmatrix} 1 & T & \frac{(\alpha T+\rho)}{\alpha^2} \\ 0 & 1 & \frac{(2-\rho)}{\alpha} \\ 0 & 0 & \rho+1 \end{bmatrix} \\ H^{MRU} &= [1\ 0] \;;\; H^{MUA} = [1\ 0\ 0] \;;\; H^{Sin} = [1\ 0\ 0] \end{aligned} \tag{3.82}$$

où $\rho = e^{(-\alpha T)} - 1$ et $\alpha = \frac{1}{\tau_{Sin}}$ avec τ_{Sin} la constante de temps du modèle de Singer.
Les matrices de précision du bruit d'état pour ces modèles de mouvement sont définies respectivement comme suit :

$$\begin{aligned} &\Gamma^{MRU} = \gamma_{MRU}\tilde{\Gamma}^{MRU} = \gamma_{MRU}\begin{bmatrix} \frac{T^3}{2} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}^{-1} \;;\quad \Gamma^{MUA} = \gamma_{MUA}\tilde{\Gamma}^{MUA} = \gamma_{MUA}\begin{bmatrix} \frac{T^5}{20} & \frac{T^4}{8} & \frac{T^3}{6} \\ \frac{T^4}{8} & \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^3}{6} & \frac{T^2}{2} & T \end{bmatrix}^{-1} \\ &\Gamma^{Sin} = \gamma_{Sin}\tilde{\Gamma}^{Sin} = \gamma_{Sin}\left(\tilde{Q}^{Sin}\right)^{-1} \end{aligned} \tag{3.83}$$

où les éléments $\{q_{ij}\}_{i=1,\dots,3}^{j=1,\dots,3}$ de $\tilde{Q}^{Sin}$ dépendent de α, ρ et T selon (2.14).
Les paramètres de précision associés peuvent alors s'exprimer comme suit :

$$\begin{aligned} \gamma_{MRU} &= \frac{1}{\sigma^2_{MRU}} \\ \gamma_{MUA} &= \frac{1}{\sigma^2_{MUA}} \\ \gamma_{Sin} &= \frac{1}{\sigma^2_{Sin}} \end{aligned} \tag{3.84}$$

À chaque instant, le système peut changer de famille de modèles de mouvement selon la TPM suivante :

$$\Pi = \begin{bmatrix} p_s & \frac{1-p_s}{2} & \frac{1-p_s}{2} \\ \frac{1-p_s}{2} & p_s & \frac{1-p_s}{2} \\ \frac{1-p_s}{2} & \frac{1-p_s}{2} & p_s \end{bmatrix} \tag{3.85}$$

Dans un premier temps, $p_s = 0.96$. Ce réglage correspond à un temps moyen de séjour dans chaque famille de mouvement de $\frac{1}{1-p_s}T = 25s$, puisque la période d'échantillonnage est $T = 1s$.

De plus, on définit une probabilité $p_r = 0.9$ correspondant à la probabilité que le paramètre de précision reste le même à l'instant suivant. Ainsi, ce réglage permet aux paramètres γ_{MRU}, γ_{MUA} et γ_{Sin} d'évoluer de façon non continue pendant une période au cours de laquelle la famille de modèles reste inchangée.

Finalement, les paramètres de précision d'état des modèles de mouvement sont uniformément distribués dans un intervalle prédéfini tel que :

$$\begin{aligned}
&\sigma_{MRU} \in [5.10^{-2} \quad 10^{-1}]m.s^{-\frac{3}{2}} \rightarrow \gamma_{MRU} \in [100\ 400] \\
&\sigma_{MUA} \in [2.5\ 4.5]m.s^{-\frac{5}{2}} \rightarrow \gamma_{MUA} \in [0.05\ 0.16] \\
&\sigma_{Sin} \in [2.5\ 4.5]m.s^{-2} \rightarrow \gamma_{Sin} \in [0.05\ 0.16]
\end{aligned} \tag{3.86}$$

Notons que les paramètres de précision ne représentent pas la même quantité physiquement. Pour cette raison, ils ne sont pas nécessairement du même ordre de grandeur.
De plus, la constante du modèle de Singer est $\tau_{Sin} = 15s$.
Enfin, la variance du bruit de mesure sur la position R est de $25m^2$.

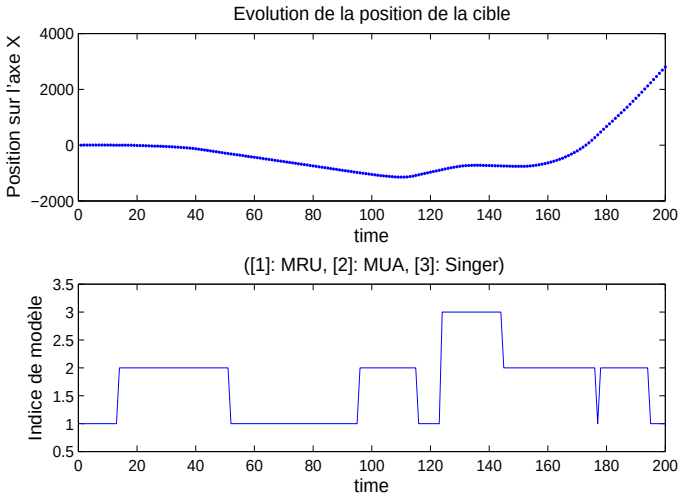

FIGURE 3.7 – Réalisation de la trajectoire de la cible et de l'évolution de l'indice de la famille de modèles

3.5.1.3 Réglage de l'algorithme proposé

Dans l'approche que nous proposons et que l'on notera par la suite *Algo1*, trois modèles de mouvement sont considérés : un MRU, un MUA et un Singer dont la constante de temps est supposée connue *a priori*. Etant donné (3.83), les formes fonctionnelles des matrices de précision associées sont connues et seul le paramètre de précision γ_k peut évoluer et alterner entre différentes valeurs au cours du temps. La variable z_k désignant la famille de modèles considérée à chaque instant suit une chaîne de Markov de matrice de probabilités de transition $\Pi^* = \begin{bmatrix} p_s^* & \frac{1-p_s^*}{2} & \frac{1-p_s^*}{2} \\ \frac{1-p_s^*}{2} & p_s^* & \frac{1-p_s^*}{2} \\ \frac{1-p_s^*}{2} & \frac{1-p_s^*}{2} & p_s^* \end{bmatrix}$ où dans un premier temps compte tenu de (3.85), $p_s^* = p_s$ soit $\Pi^* = \Pi$.

Le processus de Dirichlet associé à la famille de modèles z_k satisfait :

$$\begin{aligned}
&G^{z_k} \sim \mathcal{DP}(G_0^{z_k}, \alpha_0^{z_k}) \\
&\alpha_0^{z_k} = 10 \\
&G_0^{z_k} \sim \mathcal{G}(a_{z_k}, b_{z_k})
\end{aligned} \tag{3.87}$$

où selon la famille de modèles de mouvement considérée, les paramètres de la distribution Gamma prennent les valeurs suivantes :

$$a_{MRU} = 4 \; ; b_{MRU} = 40 \tag{3.88}$$

$$a_{MUA} = 40 \, ; b_{MUA} = 0.01 \tag{3.89}$$

$$a_{Sin} = 40 \, ; b_{Sin} = 0.01 \tag{3.90}$$

Les distributions associées sont représentées sur la figure 3.8.

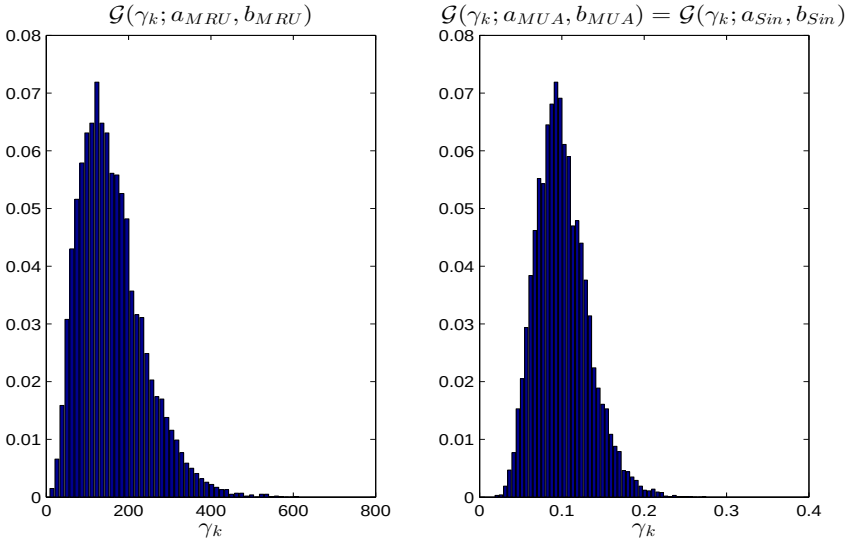

FIGURE 3.8 – Distributions *a priori* de type Gamma sur le paramètre de précision γ_k pour les différentes familles de modèles

Notons que le support des valeurs possibles prises par $\{\gamma_{z_k}\}_{z_k=MRU,MUA,Sin}$ est pertinent compte tenu des valeurs définies en (3.86) utilisées pour la génération de trajectoires. Quelle que soit la méthode de filtrage considérée, le choix des paramètres du processus de Dirichlet a une influence sur les performances de l'algorithme. Dans le cas présent, les valeurs de α_0, a_{z_k} et b_{z_k} ont été réglées selon la connaissance *a priori* du système mais il est également possible de les estimer en ligne [RAB12, PER13].

3.5.1.4 Approches de filtrage et modélisations alternatives

Dans un premier temps, nous considérons à titre de comparaison une approche BNP, notée *Algo2* fondée sur un unique modèle d'évolution et qui permet d'estimer à chaque instant l'intégralité de la matrice de précision du bruit de modèle tel qu'introduit dans la section 3.3.2. La modélisation BNP de cette approche est fondée sur les travaux de Caron [CAR08], mais l'inférence conjointe du vecteur d'état et de la matrice de précision est réalisée à partir d'un filtre particulaire mis en œuvre avec la loi d'importance optimale dont le calcul est développé dans l'annexe B.

Comme cette approche est fondée sur un unique modèle d'évolution, le choix du modèle doit permettre de pister des cibles pouvant évoluer selon les trois familles de modèles de mouvement

considérées. On suggère alors d'utiliser un modèle englobant du second ordre dont la matrice de transition est de la forme générale suivante :

$$F = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{3.91}$$

et dont la matrice de précision du bruit d'état Γ_k est supposée inconnue. L'inférence du vecteur d'état $\boldsymbol{x}_k$ et de la matrice de précision Γ_k est menée par filtrage particulaire et les particules sont propagées selon la distribution d'importance calculée en annexe B.
La matrice de précision Γ_k suit une distribution G inconnue telle que :

$$G \sim \mathcal{DP}(G_0, \alpha_0) \tag{3.92}$$

avec $\alpha_0 = 15$ alors que G_0 est une distribution de Wishart satisfaisant $G_0 \sim \mathcal{W}(\Gamma_k; W_0, v_0)$ avec $W_0 = 10 I_{n_x}$ et $v_0 = 4$.

Outre *Algo1* qui estime conjointement le vecteur d'état, la famille de modèles et le paramètre de précision et *Algo2*, présenté en annexe B et qui estime conjointement le vecteur d'état et toute la matrice de covariance du bruit de modèle par filtrage particulaire, on considère les alternatives de filtrage suivantes :

1/ Un filtre particulaire similaire à *Algo1*, noté *tpd-PF1*, dont la distribution d'importance est la distribution *a priori*.

2/ Un filtre particulaire similaire à *Algo2*, noté *tpd-PF2*, dont la distribution d'importance est la distribution *a priori*.

3/ Un filtre particulaire auxiliaire qui estime les hyperparamètres de la matrice de précision du bruit d'état conditionnellement à la famille de modèles. On le note *A-PF1*.

4/ Un filtre particulaire auxiliaire [Pit99] qui estime la matrice de précision du bruit d'état. On le note *A-PF2*. Le filtre particulaire auxiliaire est une alternative à l'algorithme *sequential importance sampling and resampling* permettant de guider les particules dans des régions prometteuses en incluant une variable auxiliaire et ainsi sans considérer la distribution d'importance optimale. Pour plus de détails, le lecteurs peut se référer à l'annexe C.

5/ Un filtre de Kalman fondé à chaque instant sur le modèle de mouvement utilisé pour la génération de la trajectoire et avec la valeur exacte de la matrice de précision. Cet algorithme de référence est noté *Ref*.

6/ Un IMM, noté *IMM*, combinant trois filtres de Kalman, chacun étant associé à l'une des trois familles de modèles de mouvement considérées, i.e. MRU, MUA ou Singer. De plus, les paramètres de variances sont réglés de manière à représenter la valeur maximale possible de la

dynamique de la cible pour cette famille de modèles. Plus précisément, on a $\sigma_{MRU} = 10^{-1} m.s^{-\frac{3}{2}}$, $\sigma_{MUA} = 4.5 m.s^{-\frac{5}{2}}$ et $\sigma_{Sin} = 4.5 m.s^{-2}$, soit $\gamma_{MRU} = 100$, $\gamma_{MUA} = 0.05$ et $\gamma_{Sin} = 0.05$. Notons que ce choix utilise seulement trois modèles contrairement aux six utilisés pour la génération de la trajectoire puisque comme cela fut montré dans [Bar95], utiliser plus de deux ou trois modèles au sein d'une structure à modèles multiples peut conduire à la dégradation des performances d'estimation.
Remarquons que de façon similaire à *Algo1* et *Algo2*, les approches 1 à 4 sont toutes fondées sur l'utilisation de DPs. De plus, les TPMs utilisées pour *tdp-PF1*, *A-PF1* et *IMM* sont égales à Π^*. Enfin, les filtres particulaires sont lancés avec $N_p = 2000$ particules.

3.5.2 Résultats de simulation

Dans un premier temps, une étude comparative en termes de complexité calculatoire est menée. Puis, l'EQM sur la position est analysée pour les protocoles 1 et 2. Finalement, l'influence de plusieurs paramètres de simulation est illustrée pour le protocole 2. Les résultats présentés ci-dessous sont moyennés sur 100 simulations de Monte Carlo pour des trajectoires de 200 échantillons.

3.5.2.1 Analyse du coût calculatoire

Au regard du tableau 3.1, comme l'on pourrait s'y attendre, les approches particulaires sont calculatoirement plus coûteuses que les méthodes fondées sur le filtrage de Kalman. De plus, n'estimer que le paramètre de précision dans *Algo1* contrairement à *Algo2* qui estime toute la matrice réduit légèrement le coût calculatoire. En effet, *Algo1* permet d'éviter un calcul d'inversion matricielle pour chaque particule à chaque itération. Par conséquent, le gain en temps de calcul est d'autant plus important que les matrices de précision sont de grandes dimensions. Finalement, l'utilisation de la distribution d'importance optimale plutôt que la distribution de transition *a priori* pour la propagation des particules est calculatoirement plus coûteux. La raison est que les paramètres de la distribution simulée diffèrent pour chaque particule et doivent être mis à jour à chaque itération.
Enfin, la complexité calculatoire des variantes avec variables auxiliaires est moins importante que celles utilisant la distribution d'importance optimale mais légèrement plus élevée que celles utilisant la distribution de transition *a priori* du fait de l'étape intermédiaire mise en jeu dans les filtres particulaires auxiliaires.

Algorithme	*Ref*	*IMM*	*tpd-PF1*	*tpd-PF2*	*A-PF1*	*A-PF2*	*Algo1*	*Algo2*
Coûts calculatoires	1	$\times 3$	$\times 25$	$\times 30$	$\times 40$	$\times 50$	$\times 200$	$\times 250$

Table 3.1 – Ordre de grandeur des coûts calculatoires des différentes approches testées

3.5.2.2 Analyse de la précision d'estimation

Nous présentons les EQMs entre le vecteur d'état et son estimation sur 200 échantillons. Notons que l'EQM moyenne liée au bruit de mesure est de 4.11.

Algorithme	*Ref*	*IMM*	*tpd-PF1*	*tpd-PF2*	*A-PF1*	*A-PF2*	*Algo1*	*Algo2*
EQM protocole 1	3.03	3.78	3.51	3.43	3.42	3.34	3.32	**3.21**
EQM protocole 2	2.91	3.42	3.11	3.23	3.06	3.15	**2.96**	3.08

TABLE 3.2 – EQM moyennée sur 100 simulations de Monte Carlo

Selon le tableau 3.2, l'EQM associée à *Ref* est la plus faible puisqu'elle considère à chaque instant le bon modèle de mouvement et la bonne matrice de précision. Les EQMs de *Algo2* et *Algo1* sont plus faibles que celles associées à *IMM* car la matrice de précision est estimée au cours du temps alors qu'elle est réglée à une valeur prédéfinie pour un algorithme du type IMM. Finalement, l'utilisation de la distribution optimale est pertinente puisque *Algo2* et *Algo1* ont de meilleures performances que *tpd-PF1* et *tpd-PF2* respectivement. Concernant les variantes auxiliaires, elles ont des performances d'estimation intermédiaires en termes d'EQM.

Concernant les résultats obtenus à partir du second protocole de simulation, même si *Algo1* a de meilleures performances d'estimation que *Algo2*, la comparaison de ces deux approches n'est pas tellement justifiée. En effet, *Algo2* est fondé sur l'utilisation d'un seul type de modèle d'évolution alors que *Algo1* prend en compte plusieurs familles de modèles de mouvement. De manière générale, un modèle du second ordre comme *Algo2* convenablement réglé, c'est-à-dire caractérisé par une variance sur l'accélération suffisamment importante, peut être utilisé pour le pistage de cibles manœuvrantes. Cependant, l'estimation est moins précise que pour un algorithme considérant plusieurs modèles d'évolution, comme *Algo1* qui permet de prendre en compte différentes phases de mouvement de la cible. C'est l'une des raisons pour lesquelles *Algo1* a de meilleures performances que *Algo2*. L'autre raison est que *Algo1*, prenant en compte les formes fonctionnelles des matrices de précision du bruit d'état, permet d'estimer un seul paramètre au lieu de toute la matrice de précision pour *Algo2*. Comme illustré dans la partie 3.6.2.4, pour une nombre fixe de particules, l'inférence est alors plus précise pour *Algo1* que pour *Algo2*.

Notons que la précision obtenue avec *Algo1* est aussi liée au choix des familles de modèles. Or, l'information concernant les familles de modèles et leurs formes structurelles associées peut ne pas être disponible dans toutes les applications. *Algo2* serait alors dans ce cas plus approprié. Par exemple, pour le protocole 1, *Algo1* est légèrement moins performant que *Algo2* puisque le modèle utilisé dans ce cas ne correspond pas nécessairement aux modèles de mouvement suivis par la cible. De ce fait, selon la connaissance *a priori* du système, le choix entre *Algo2* et *Algo1* est un compromis entre la complexité calculatoire et la précision d'estimation souhaitées. Finalement, les deux approches proposées sont d'intérêt puisqu'elles ont globalement de meilleures performances que les autres approches existantes, et cela grâce à l'utilisation de modèles non-paramétriques et de la distribution d'importance optimale.

Enfin, notons que l'indice de la famille de modèles est mieux estimé avec *Algo1* qu'avec l'IMM pour lequel l'indice de la famille de modèles correspond à chaque instant au modèle avec le poids *a posteriori* le plus élevé [BAR95]. Ce phénomène est illustré par la figure 3.9 où l'estimation de z_k obtenu avec l'algorithme IMM et *Algo1* est donnée pour une réalisation de la trajectoire de cible. En moyennant cette estimation sur 500 simulations de Monte Carlo, on remarque que

l'approche proposée estime la bonne famille de modèles dans 93% des cas pour *Algo1* et 88% pour IMM. Cela peut s'expliquer du fait de l'étape de mélange au sein de l'algorithme IMM qui ne permet pas de rendre le poids en faveur de la bonne famille de modèles prépondérant.

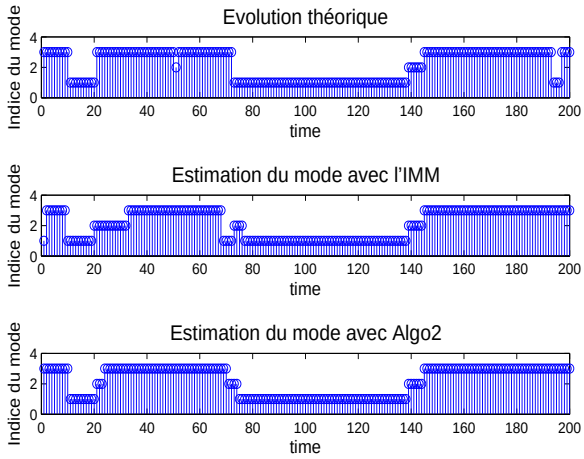

FIGURE 3.9 – Représentation au cours du temps des indices de famille de modèles estimés

Dans la suite, les influences de la TPM, du nombre de particules et du paramètre d'échelle du DP sur les performances d'estimation sont étudiés. Par souci de simplicité, nous focalisons notre analyse sur *Algo1*, *tpd-PF1* et *A-PF1* mais les remarques qui suivent sont également valables pour *Algo2*, *tpd-PF2* et *A-PF2*.

3.5.2.3 Influence du réglage de la TPM

Parmi les approches de filtrage qui ont été proposées précédemment, certaines d'entre elles sont fondées sur l'utilisation d'une TPM. Dans la suite, nous analysons la sensibilité de ces approches par rapport à ce choix *a priori*.

Les résultats de cette analyse sont présentés au sein du tableau 3.3 où les éléments diagonaux sont choisis égaux à $p_s^* = 0.99, 0.96, 0.9, 0.8$ correspondant à des temps de séjour de 100s, 25s, 10s et 5s respectivement.

Algorithme d'estimation	*IMM*	*Algo1*	*tpd-PF1*	*A-PF1*
$p_s^* = 0.99$	3.46	2.99	3.15	3.10
$p_s^* = 0.96$ (valeur vraie)	**3.42**	**2.96**	3.11	3.06
$p_s^* = 0.9$	3.60	3.13	3.30	3.34
$p_s^* = 0.8$	3.80	3.32	3.53	3.54

TABLE 3.3 – EQM moyennée sur 100 simulations de Monte Carlo pour différents réglages de la TPM.

Si la TPM n'est pas réglée à la même valeur que celle utilisée pour la génération des trajectoires, les performances sont légèrement dégradées. Cependant, la méthode non paramétrique, qui présente globalement de meilleurs résultats en termes d'EQM, est l'approche pour laquelle la dégradation de performance liée au mauvais réglage de la TPM est la moins importante. En effet, les approches BNP sont plus flexibles selon la valeur de α_0.

3.5.2.4 Influence du nombre de particules

Les EQMs obtenues pour les trois approches particulaires sont représentées par la Fig. 3.10 en fonction du nombre de particules.

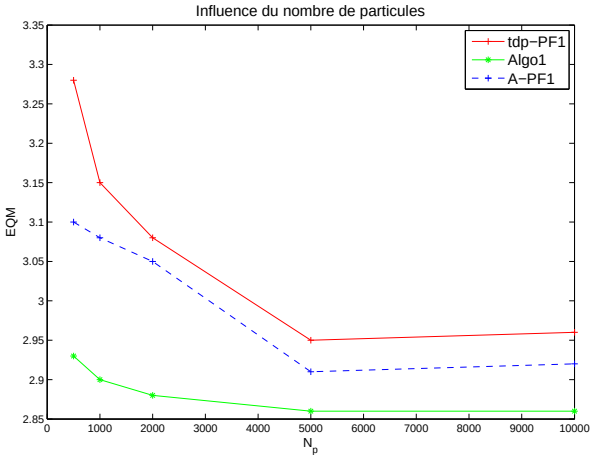

FIGURE 3.10 – Influence du nombre de particules sur l'EQM

Notons que si le nombre de particules est inférieur à environ 5000, *Algo1* est bien meilleur que les autres approches. De plus, du fait de l'étape de pré-échantillonnage, *A-PF1* est légèrement meilleur que *tpd-PF1*. Finalement, pour un nombre suffisamment élevé de particules, la différence entre ces trois algorithmes est de moins en moins importante.

3.5.2.5 Influence du paramètre d'échelle

Dans ce cas, on fait varier le paramètre d'échelle de 0.1 à 50. L'EQM obtenue est représentée en Fig. 3.11. Nous remarquons que si α_0 est trop petit ou au contraire trop grand, l'EQM associée se dégrade. En effet, le premier cas induit un faible nombre de clusters ne permettant pas à la matrice de précision d'explorer suffisamment de valeurs pour converger vers la bonne valeur. Dans le second cas, une nouvelle valeur est tirée quasiment à chaque instant, l'estimation de Q_k ne peut donc pas converger correctement. Cette dernière remarque met en avant la difficulté pour l'utilisateur de paramétrer ce type de méthodes.

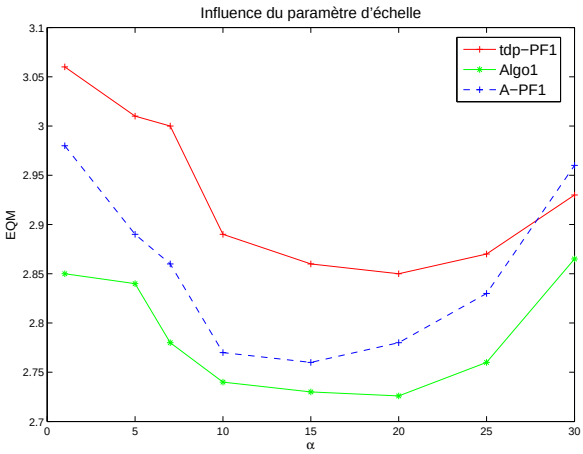

FIGURE 3.11 – Influence du paramètre d'échelle sur l'EQM

3.6 Conclusions et perspectives

Dans ce chapitre, nous avons décrit une modélisation non-paramétrique du problème d'estimation jointe de la matrice de covariance du bruit d'état et du vecteur d'état. En particulier, nous insistons sur l'intérêt des processus de Dirichlet et des méthodes BNP en général pour traiter l'estimation de ces paramètres qui évoluent de façon discontinue et dont les valeurs peuvent réapparaître plus ou moins fréquemment au cours du temps.
Dans ce contexte, l'estimation non-paramétrique de la matrice de covariance du bruit de modèle étant un problème non-linéaire non-gaussien, il est traité par filtrage particulaire. Dans la continuité des travaux initiés par Caron pour estimer toute la matrice de covariance du bruit de modèle et vis-à-vis de l'état de l'art, nous développons deux contributions.

Dans un premier temps, nous proposons un modèle hiérarchique BNP fondé sur l'utilisation de différentes familles de modèles de mouvement. Pour chacune d'elles, nous tirons profit qu'en pistage de cibles, les matrices de covariance des bruits d'état ont des formes fonctionnelles connues. Ainsi, conditionnellement à la famille de modèles considérée, seulement un paramètre de précision est estimé. Dans un second temps, nous détaillons les différentes étapes de calcul de la loi d'importance optimale utilisée pour mener l'inférence conjointe du vecteur d'état et du paramètre de précision par filtrage particulaire.

Nous montrons ensuite l'intérêt de ces travaux pour le pistage monocible au travers d'une étude comparative. Outre notre algorithme, nous testons une approche particulaire inspirée des travaux de Caron [Car08] qui estime conjointement au vecteur d'état toute la matrice de covariance du bruit de modèle et qui utilise la distribution d'importance optimale développée en annexe B. Ces deux approches sont également comparées à des algorithmes particulaires similaires fondés sur la distribution d'importance *a priori* ou sur des filtres particulaires auxiliaires. Nous montrons que ces approches BNP ont de meilleures performances d'estimation qu'un algorithme IMM dont les matrices de covariance des bruits de modèles sont fixées *a priori*. C'est notamment le cas pour des scénarios où les mouvements de la cible varient beaucoup et donc pour lesquels un algorithme d'estimation, fondé sur un réglage *a priori* qui n'est pas nécessairement adapté, est mis en difficulté. De manière plus globale, les résultats de simulation obtenus mettent en avant l'intérêt de notre modèle hiérarchique et la pertinence de la distribution d'importance optimale. Néanmoins, les approches BNP testées sont caractérisées par un coût calculatoire important, notamment vis-à-vis des approches de pistage fondées sur du filtrage de Kalman.

Enfin, notons que le réglage *a priori* des paramètres des processus de Dirichlet a une influence non-négligeable sur les performances d'estimation de l'état de la cible. Afin de s'affranchir de cette contrainte, nous envisageons de les estimer en ligne comme proposé dans [Rab12, Per13].

En général, le choix de la famille de modèles est important au regard des performances de l'algorithme d'estimation utilisé. Il peut alors être pertinent de l'adapter au type de cible rencontré. En effet, en pratique, que ce soit en pistage de cibles marines, terrestres ou aériennes, une grande diversité de cibles est souvent rencontrée. Pour cela, il est possible d'appliquer des méthodes du type JTC effectuant conjointement la classification et le pistage de la cible afin d'améliorer la qualité de l'estimation. Nous traitons de ce type d'approches dans le prochain

chapitre et nous les combinons à des approches BNP.

Chapitre 4

Pistage et classification conjoints de cibles

Sommaire

4.1 Introduction 104

4.2 Position du problème 105

4.2.1 Modélisation du système fondée uniquement sur les paramètres cinématiques de cible 105

4.2.2 Modélisation du système incluant la longueur de cible 106

4.3 Pistage et classification d'une cible à partir de mesures cinématiques et de mesures d'étalement distance 108

4.3.1 Filtre IMM multi-classes (MC-IMM) 109

4.3.2 Filtre particulaire multi-modèles multi-classes (MC-MMPF) 113

4.3.3 Extension au cas où les paramètres du modèle d'état sont inconnus . . . 116

4.3.4 Application au pistage d'une cible marine 118

4.4 Pistage et classification de plusieurs cibles à partir de mesures cinématiques et de mesures d'étalement distance 127

4.4.1 Modélisation multi-cibles 127

4.4.2 Algorithme PHD multi-modèles (MM-PHD) multi-classes 128

4.4.3 Equations du filtre MM-PHD et implémentation par mélange de gaussiennes (MM-GMPHD) 129

4.4.4 Equations théoriques 129

4.4.5 Application au pistage de cibles marines 135

4.5 Conclusions et perspectives 143

4.1 Introduction

En pistage de cibles, la prise en compte d'informations sur le type de cibles pistées peut aider à définir des modèles d'évolution pertinents [Mei07, Cao13, Lan13, Sun14, Cao14]. Cette constatation a motivé l'introduction d'approches proposant d'effectuer conjointement le pistage et la classification de cibles [Cha01, Mei04, Ris04b]. On parle en anglais de *joint tracking and classification* (JTC). Dans ce chapitre, nous définissons plusieurs classes qui se distinguent les unes des autres selon la manœuvrabilité des cibles. Outre ses caractéristiques dynamiques, différents attributs de la cible comme sa forme ou sa fréquence Doppler peuvent être utilisés pour aider à la classification. Partant de l'*a priori* que la longueur d'une cible apporte une information sur sa manœuvrabilité, nous proposons de tirer avantage de cette information pour aider à la classification de cible. De plus, comme la longueur de cible n'est pas directement mesurée, nous l'intégrons au vecteur d'état pour pouvoir l'estimer. Pour les approches classiques de pistage, une hypothèse de cible ponctuelle est généralement formulée [Cha01, Ang04, Li07]. Cependant, les mesures d'étalement distance, qui sont des mesures de la longueur de la cible selon l'axe radar-cible [Sal03] [Ris04c], peuvent être fournies par des radars à haute résolution distance, *high-range resolution* en anglais (HRR). Leur exploitation requiert néanmoins une modélisation géométrique de la cible.

Dans ce chapitre, nous présentons tout d'abord dans un contexte monocible deux algorithmes bayésiens JTC multi-modèles. Ils reposent tous deux sur l'utilisation conjointe des mesures cinématiques et des mesures de l'étalement distance de la cible mais diffèrent par leur traitement des mesures non-linéaires [Mag15c]. Le premier utilise au sein de chaque classe un algorithme IMM combinant des FKEs, alors que le second repose sur un filtre particulaire à modèles multiples. Nous proposons ensuite une extension au cas où les paramètres du bruit de modèle d'état sont supposés inconnus. Dans ce cas, nous combinons une approche JTC avec un modèle BNP permettant d'estimer la distribution du bruit de modèle.
Enfin, afin de traiter non pas une mais plusieurs cibles conjointement, nous étendons ces approches JTC au pistage multi-cibles [Mag16b]. Nous présentons un algorithme multi-classes dont la mise en œuvre est fondée sur l'utilisation de filtres PHD à modèle multiples noté MM-PHD.

Dans la suite de ce chapitre, la partie 4.2 présente la problématique étudiée. En particulier, la modélisation du système reposant sur les paramètres cinématiques de la cible ainsi que sur sa longueur est introduite. Les méthodes bayésiennes multi-classes multi-modèles respectivement fondées sur des algorithmes IMM et particulaire sont présentées dans la partie 4.3. Nous traitons également dans cette partie de l'extension au cas où les paramètres du bruit de modèle sont supposés inconnus et estimés à partir d'un modèle bayésien non-paramétrique. Ensuite, la partie 4.4 présente la généralisation au cas multi-cibles. Enfin, les conclusions et perspectives sont exposées en partie 4.5.

4.2 Position du problème

En pratique, que ce soit en pistage de cibles aériennes, terrestres ou marines, les différentes cibles pistées n'ont pas toutes les mêmes caractéristiques dynamiques. Par exemple, en poursuite de cibles aériennes, on peut distinguer les avions militaires des avions civils. Dans un contexte de surveillance maritime, on peut distinguer les petites embarcations des gros navires. Cette classification peut permettre à l'utilisateur d'adapter la modélisation de la dynamique de l'objet au type de cibles pistées.
Dans ce chapitre, nous suggérons de tirer profit de la longueur des cibles pour aider à déterminer leur type et ainsi régler plus finement le ou les modèles d'évolution utilisés au sein de l'algorithme de pistage.
Dans la suite, nous présentons la REE d'une cible conditionnée par sa classe, puis nous introduisons une extension de cette REE incluant la longueur de cible dans l'état du système et la mesure d'étalement distance dans les observations.

4.2.1 Modélisation du système fondée uniquement sur les paramètres cinématiques de cible

Lors du pistage d'une cible manœuvrante, il y a une forte incertitude quant au modèle de dynamique de la cible. Cependant, connaître le type de cible pistée permet en partie de lever cette indétermination. Dans ce travail, nous supposons qu'une cible appartient à une classe c parmi C classes prédéfinies. Cet attribut ne varie pas au cours du temps et caractérise un type de cible. De plus, nous faisons l'hypothèse que pour chaque classe, la trajectoire de cible peut être décrite par un nombre fini $r(c)$ de modèles de mouvement notés $\{M_1^c, ..., M_{r(c)}^c\}$.

Au sein de chaque classe, la transition entre les différents modèles de mouvement est classiquement décrite par une chaîne de Markov dont la matrice de transition est notée Π^c. Si m_k^c dénote le modèle de mouvement de la cible à l'instant k, les probabilités de transition entre les modèles d'une même classe, notées $\{\Pi_{ji}^c\}_{i=1,...,r(c)}^{j=1,...,r(c)}$ et éléments de Π^c, satisfont :

$$\Pi_{ji}^c = Pr(m_{k+1}^c = M_i^c | m_k^c = M_j^c) \tag{4.1}$$

Pour le modèle de mouvement m_k^c à l'instant k, l'évolution du système est décrite par :

$$\boldsymbol{x}_k = F^{m_k^c} \boldsymbol{x}_{k-1} + \boldsymbol{v}_k^{m_k^c} \tag{4.2}$$

où

- le vecteur d'état $\boldsymbol{x}_k$ est défini pour les modèles du second et du troisième ordre respectivement comme suit :

$$\boldsymbol{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k]^T \; ; \; \boldsymbol{x}_k = [x_k, \dot{x}_k, \ddot{x}_k, y_k, \dot{y}_k, \ddot{y}_k]^T \tag{4.3}$$

  avec x_k et y_k les positions, $\dot{x}_k$ et $\dot{y}_k$ les vitesses et $\ddot{x}_k$ et $\ddot{y}_k$ les accélérations selon les dimensions x et y. Dans ce travail, nous considérons la poursuite de cibles de surface dans un repère 2D. Néanmoins notre étude pourrait être généralisée sans difficulté au pistage

de cibles aériennes dans un repère 3D.

- $F^{m_k^c}$ correspond à la matrice de transition associée au modèle m_k^c de la classe c.
- $\boldsymbol{v}_k^{m_k^c}$ est un bruit blanc gaussien centré et de matrice de covariance $Q^{m_k^c}$.

Comme un radar aéroporté fournit généralement la distance du radar à la cible notée d_k et une mesure angulaire, notée θ_k, la relation entre la mesure $\boldsymbol{y}_k$ et le vecteur d'état $\boldsymbol{x}_k$ satisfait :

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{4.4}$$

où le bruit de mesure $\boldsymbol{w}_k$, non-corrélé avec $\boldsymbol{v}_k^{m_k^c}$ quel que soit m_k^c, est un bruit blanc gaussien centré de matrice de covariance $R = diag(\sigma_d^2, \sigma_\theta^2)$ avec σ_d^2 et σ_θ^2 désignant les variances des bruits de mesure de distance et d'angle respectivement. De plus, la fonction de mesure satisfait :

$$h(\boldsymbol{x}_k) = \begin{bmatrix} d_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} \sqrt{(x_k - x_r)^2 + (y_k - y_r)^2} \\ tan^{-1}(\frac{y_k - y_r}{x_k - x_r}) \end{bmatrix} \tag{4.5}$$

où $(x_r,\ y_r)$ désignent les coordonnées du radar projetées sur la surface où évolue la cible. Elles sont supposées connues à chaque instant. La relation entre les mesures et les coordonnées cartésiennes de la cible est représentée à la figure 4.1.

Cette modélisation du système, fondée uniquement sur l'utilisation de mesures dynamiques de la cible, est largement utilisée en pistage. Néanmoins, l'apport d'informations complémentaires portant sur les caractéristiques de la cible d'intérêt peut être d'un réel intérêt. Dans la suite, nous présentons une REE qui inclut la longueur de la cible dans le vecteur d'état et proposons de mettre à profit cette information pour développer une approche JTC.

4.2.2 Modélisation du système incluant la longueur de cible

En pratique, les dimensions d'une cible ne sont pas directement disponibles. Cependant, l'étalement distance de la cible, qui est une fonction de la longueur de la cible et de la géométrie relative entre le radar et la cible, peut être estimé en utilisant un radar à haute résolution distance. Sous certaines hypothèses, cette mesure est explicitement reliée à la longueur de la cible par une formule analytique introduite ci-après. Afin d'exploiter les mesures d'étalement distance de la cible quelle que soit son orientation vis-à-vis du radar, Salmond *et al.* [Sal03] ont introduit un modèle elliptique et l'utilisent pour améliorer les performances des algorithmes d'association de données. Dans ce modèle, la position (x_k, y_k) est supposée correspondre au barycentre de l'ellipse.

Comme illustré par la figure 4.1, si une cible est supposée ellipsoïdale, on peut montrer que son étalement distance, noté $\mathcal{E}$, s'exprime comme suit :

$$\mathcal{E}(\phi_k, l) = l\sqrt{cos^2(\phi_k) + \left(\frac{a}{b}\right)^2 sin^2(\phi_k)} \tag{4.6}$$

où ϕ_k représente l'angle entre la direction de la cible et la LOS radar-cible à l'instant k. De plus, l est la longueur de la cible et $\frac{a}{b}$ est le rapport entre le petit axe et le grand axe de l'ellipse qui est supposé connu.

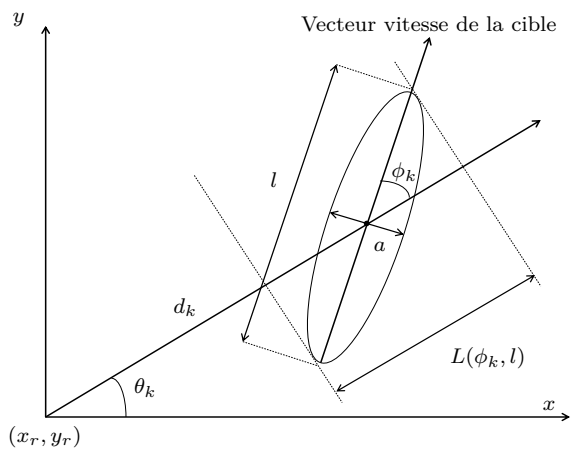

FIGURE 4.1 – Modèle elliptique de la cible

Dans la suite de ce chapitre, nous nous placons dans le contexte de la poursuite de cibles marines. Dans ce cas, on peut considérer que toute cible est orientée selon la même direction que son vecteur vitesse. Il vient alors :

$$\begin{aligned}\mathcal{E}(\boldsymbol{x}_k, l) &= \frac{l\sqrt{(\dot{y}_k\Delta_y + \dot{x}_k\Delta_x)^2 + \left(\frac{a}{b}\right)^2(\dot{y}_k\Delta_x - \dot{x}_k\Delta_y)^2}}{\sqrt{{\Delta_x}^2 + {\Delta_y}^2}\sqrt{\dot{x}_k^2 + \dot{y}_k^2}} \\ &= h_{taille}(\boldsymbol{x}_k) \times l\end{aligned} \tag{4.7}$$

où $\Delta_x = x_k - x_r$, $\Delta_y = y_k - y_r$.
L'équation (4.7) met en évidence que la mesure d'étalement distance présente un double intérêt. Premièrement, elle permet d'apporter une information complémentaire visant à dissocier différents types de cibles selon leur taille et ainsi d'affiner les modèles d'évolution en conséquence. De plus, étant donné (4.7), comme l'étalement distance peut s'exprimer à partir des paramètres cinématiques de la cible, la prise en compte de cette information devrait contribuer à améliorer les performances du pistage.

Etant donné (4.2) et (4.7), un vecteur d'état étendu qui inclut la longueur de la cible est introduit. Il satisfait :

$$\boldsymbol{X}_k = [\boldsymbol{x}_k^T \ l]^T \tag{4.8}$$

où l ne varie pas au cours du temps contrairement à $\boldsymbol{x}_k$.
Dans ce cas, combinant (4.2) et (4.8), le vecteur d'état suit la dynamique suivante :

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}^{m_k^c}\boldsymbol{X}_k + \boldsymbol{V}_k^{m_k^c} \tag{4.9}$$

où $\boldsymbol{V}_k^{m_k^c} = [(\boldsymbol{v}_k^{m_k^c})^T \ 0]^T$ est le bruit de modèle étendu avec pour matrice de covariance $\boldsymbol{Q}^{m_k^c}$.
De plus, les matrices de transition et de covariance sont respectivement définies comme suit :

$$\begin{aligned}\boldsymbol{F}^{m_k^c} &= blkdiag([F^{m_k^c} \ 1]) \\ \boldsymbol{Q}^{m_k^c} &= blkdiag([Q^{m_k^c} \ 0])\end{aligned} \tag{4.10}$$

Concernant l'équation de mesure, si L_k désigne la mesure bruitée de l'étalement distance $\mathcal{E}$, (4.4) devient :

$$\boldsymbol{Y}_k = [\boldsymbol{y}_k^T \ L_k]^T = \boldsymbol{h}(\boldsymbol{X}_k) + \boldsymbol{W}_k \tag{4.11}$$

avec :

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{X}_k) &= [h(\boldsymbol{x}_k)^T \ \mathcal{E}(\boldsymbol{x}_k, l)]^T \\ &= [h(\boldsymbol{x}_k)^T \ h_{taille}(\boldsymbol{x}_k) \times l]^T \end{aligned} \tag{4.12}$$

De plus, le bruit sur la mesure d'étalement distance, noté $\boldsymbol{u}_k$, est supposé à moyenne nulle et de variance σ_{ed}^2. Il n'est ni corrélé avec le bruit de mesure sur les paramètres cinématiques $\boldsymbol{w}_k$ ni avec $\boldsymbol{v}_k^{m_k^c}$. Ainsi, la matrice de covariance de $\boldsymbol{W}_k = [\boldsymbol{w}_k^T \ \boldsymbol{u}_k]^T$ satisfait :

$$\boldsymbol{R} = blkdiag(R, \sigma_{ed}^2) \tag{4.13}$$

Comme nous l'avons mis en évidence dans l'équation (4.9) de remise à jour du vecteur d'état étendu, le modèle d'évolution du système peut changer d'un instant à un autre. Pour prendre en compte cette propriété lors de l'estimation, les algorithmes à modèles multiples peuvent être utilisés [Bar95, Li05]. Pour un modèle de mouvement donné, comme l'équation de mesure (4.11) est non-linéaire, un filtre de Kalman étendu, un filtre de Kalman du second ordre ou encore un filtre de Kalman à sigma-points tels que le filtre de Kalman sans parfum [Wan01, Ito00, Ara09] peuvent être envisagés. Le filtre de Kalman étendu, qui présente généralement un bon compromis entre complexité calculatoire et précision d'estimation est souvent préféré. Alternativement, selon la non-linéarité de la fonction de mesure définie par (4.11), une mise en œuvre par filtrage particulaire peut s'avérer plus efficace.

Nous détaillons dans la partie suivante deux approches JTC à modèles multiples. La première combine plusieurs FKEs alors que la seconde est fondée sur un filtre particulaire à modèles multiples. Nous évaluons ensuite l'intérêt de chaque méthode selon les types de cibles et leurs dynamiques.

4.3 Pistage et classification d'une cible à partir de mesures cinématiques et de mesures d'étalement distance

Dans un contexte bayésien, l'estimation de l'état et de la classe est fondée sur la distribution jointe *a posteriori* $Pr(\boldsymbol{X}_k, c|\boldsymbol{Y}_{1:k})$. Ce problème peut être décomposé en deux étapes. Premièrement, la distribution *a posteriori* de la $c^{\text{ème}}$ classe de cible conditionnellement aux observations peut être séquentiellement calculée comme suit :

$$\begin{aligned} Pr(c|\boldsymbol{Y}_{1:k}) &= \frac{p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c)Pr(c|\boldsymbol{Y}_{1:k-1})}{p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1})} \\ &= \frac{p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c)Pr(c|\boldsymbol{Y}_{1:k-1})}{\sum_{c'=1}^{C} p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c')Pr(c'|\boldsymbol{Y}_{1:k-1})} \end{aligned} \tag{4.14}$$

où à l'instant initial, si aucune information *a priori* n'est disponible sur la cible, on considère :

$$\forall\ c\ \in \{1, ..., C\},\ Pr(c) = \frac{1}{C} \tag{4.15}$$

Puis, en utilisant la règle de Bayes, la densité de probabilité *a posteriori* de l'état étendu, à savoir $p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k})$, s'exprime comme suit :

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}) = \sum_{c=1}^{C} p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}, c) Pr(c|\boldsymbol{Y}_{1:k}) \tag{4.16}$$

A l'instant initial, on considère les hypothèses suivantes :

$$p(\boldsymbol{X}_0|c) = [p(\boldsymbol{x}_0)\ p(l|c)] \tag{4.17}$$

où $p(\boldsymbol{x}_0)$ est gaussien et

$$p(l|c) = \mathcal{N}(l; l_0^c, P_0^c) \tag{4.18}$$

avec l_0^c et P_0^c fixées *a priori* pour chacune des classes.

Dans (4.17), $p(l|c)$ permet d'introduire de l'information *a priori* sur le type de cible selon sa longueur. Ce sont en effet les valeurs des paramètres $\{l_0^c\}_{c=1,...,C}$ et $\{P_0^c\}_{c=1,...,C}$ qui influent sur la capacité de l'algorithme à déterminer rapidement la classe d'une cible.

Afin de traiter le problème décrit par (4.14) et (4.16), deux méthodes du type JTC sont proposées. Elles ont toutes la même architecture illustrée par la figure 4.2, mais diffèrent par la manière dont l'estimation est réalisée au sein de chaque classe.
Les deux méthodes permettent de prendre en compte la non-linéarité de l'équation de mesure (4.11). La première, notée MC-IMM, est fondée sur un algorithme IMM utilisant des FKEs. La seconde, notée MC-MMPF, utilise des filtres particulaires à modèles multiples.

4.3.1 Filtre IMM multi-classes (MC-IMM)

Il s'agit d'estimer $\boldsymbol{X}_k$ au sein de chaque classe en utilisant un algorithme IMM fondé sur des FKEs. Comme nous l'avons vu dans la partie 1.2.1 du chapitre 1, le principe de l'IMM est d'approcher à chaque instant les distributions de prédiction et *a posteriori* par un mélange de distributions gaussiennes. En adaptant les notations au contexte d'étude et en se référant aux équations (1.73) et (1.77), il vient :

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k-1}, c) \simeq \sum_{i=1}^{r(c)} \mu_{k|k-1}^{(i,c)} \mathcal{N}(\boldsymbol{X}_k; \boldsymbol{X}_{k|k-1}^{(i,c)}, P_{k|k-1}^{(i,c)}) \tag{4.19}$$

et

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}, c) \simeq \sum_{i=1}^{r(c)} \mu_k^{(i,c)} \mathcal{N}(\boldsymbol{X}_k; \boldsymbol{X}_{k|k}^{(i,c)}, P_{k|k}^{(i,c)}) \tag{4.20}$$

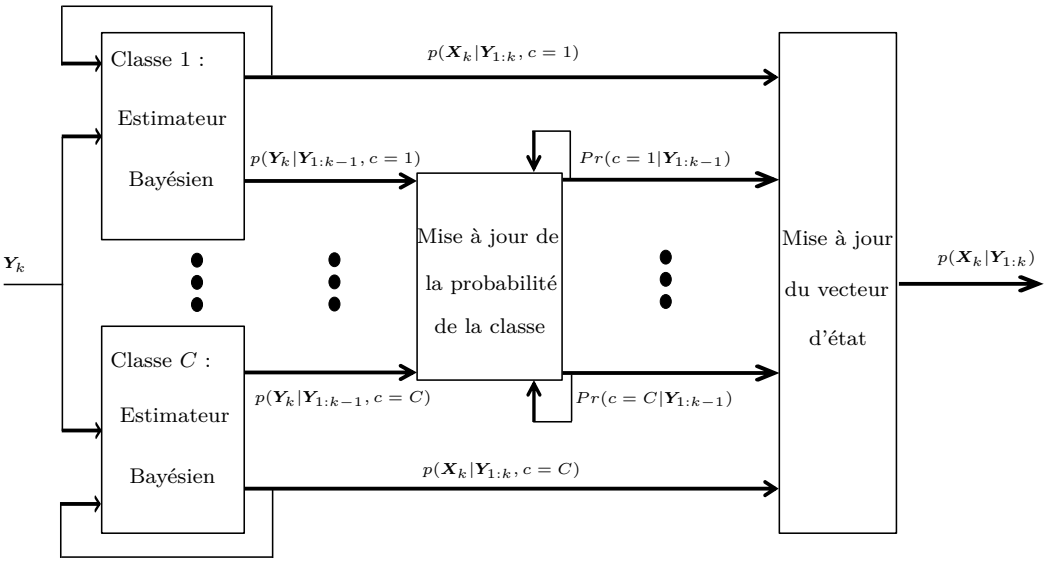

FIGURE 4.2 – Schéma bloc des algorithmes proposés

où $\boldsymbol{X}_{k|k-1}^{(i,c)}$ et $\boldsymbol{X}_{k|k}^{(i,c)}$ désignent la prédiction et l'estimation du vecteur d'état $\boldsymbol{X}_k$ calculées par le $i^{\text{ème}}$ FKE de la $c^{\text{ème}}$ classe à l'instant k. $\boldsymbol{P}_{k|k-1}^{(i,c)}$ et $\boldsymbol{P}_{k|k}^{(i,c)}$ sont les matrices de covariance de leurs erreurs d'estimation respectives.
De plus, pour chaque classe, les poids de chaque filtre notés $\mu_k^{(i,c)}$ à l'instant k satisfont les équations récursives suivantes :

$$\mu_{k|k-1}^{(i,c)} = \sum_{j=1}^{r(c)} \Pi_{ji}^{c} \mu_{k-1}^{(j,c)} \tag{4.21}$$

$$\mu_k^{(i,c)} \propto p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, m_k^c = i)\mu_{k|k-1}^{(i,c)} \tag{4.22}$$

La constante de proportionalité utilisée en (4.22) est telle que :

$$\sum_{i=1}^{r(c)} \mu_k^{(i,c)} = 1 \tag{4.23}$$

Finalement, (4.20) est substituée dans (4.16) pour obtenir une approximation de la distribution *a posteriori* du vecteur d'état, soit :

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}) \approx \sum_{c=1}^{C} Pr(c|\boldsymbol{Y}_{1:k}) \sum_{i=1}^{r(c)} \mu_k^{(i,c)} \mathcal{N}(\boldsymbol{X}_k; \boldsymbol{X}_{k|k}^{(i,c)}, P_{k|k}^{(i,c)}) \tag{4.24}$$

Dans (4.24), il reste à évaluer la distribution *a posteriori* de la classe notée $Pr(c|\boldsymbol{Y}_{1:k})$. Elle peut être calculée récursivement en utilisant (4.14) où $p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c)$ doit être explicitée. Pour cela, plusieurs étapes sont nécessaires.
Dans un premier temps, on peut réécrire $p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c)$ en faisant intervenir le vecteur d'état

comme suit :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c) = \int_{\boldsymbol{X}_k} p(\boldsymbol{Y}_k|\boldsymbol{X}_k,c)p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k-1},c)d\boldsymbol{X}_k \tag{4.25}$$

En insérant (4.19) dans (4.25), on peut envisager l'approximation suivante pour $p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c)$:

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c) \simeq \sum_{i=1}^{r(c)} \mu_{k|k-1}^{(i,c)} \underbrace{\int_{\boldsymbol{X}_k} p(\boldsymbol{Y}_k|\boldsymbol{X}_k,c)\mathcal{N}(\boldsymbol{X}_k;\boldsymbol{X}_{k|k-1}^{(i,c)},P_{k|k-1}^{(i,c)})d\boldsymbol{X}_k}_{p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c,i)} \tag{4.26}$$

Afin de calculer explicitement l'intégrale définie en (4.26), le filtre de Kalman étendu procède en linéarisant la fonction de mesure $\boldsymbol{h}(\cdot)$ autour de l'état prédit $\boldsymbol{X}_{k|k-1}^{(i,c)}$ de telle manière que :

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_{k|k-1}^{(i,c)}) + \boldsymbol{H}^{(i,c)}(\boldsymbol{X}_k - \boldsymbol{X}_{k|k-1}^{(i,c)}) + \boldsymbol{W}_k \tag{4.27}$$

avec :

$$\begin{aligned} \boldsymbol{W}_k|\boldsymbol{Y}_{1:k-1},c,i &\sim \mathcal{N}(\boldsymbol{W}_k;0,\boldsymbol{R}) \\ \boldsymbol{X}_k|\boldsymbol{Y}_{1:k-1},c,i &\sim \mathcal{N}(\boldsymbol{X}_k;\boldsymbol{X}_{k|k-1}^{(i,c)},P_{k|k-1}^{(i,c)}) \end{aligned} \tag{4.28}$$

et où $\boldsymbol{H}^{(i,c)}$ est la matrice Jacobienne de la fonction de mesure $\boldsymbol{h}$ évaluée en $\boldsymbol{X}_{k|k-1}^{(i,c)}$. Ses éléments sont explicités dans [Sal03] et rappelés ci-après.
Si $r = \sqrt{{\Delta_x}^2 + {\Delta_y}^2}$ et $u = \sqrt{\dot{x}_k^2 + \dot{y}_k^2}$, la matrice Jacobienne de taille 3×7 de la fonction de mesure $\boldsymbol{h}(\cdot)$ est définie comme suit :

$$\begin{bmatrix} \frac{\Delta_x}{r} & 0 & 0 & \frac{\Delta_y}{r} & 0 & 0 & 0 \\ \frac{-\Delta_y}{r^2} & 0 & 0 & \frac{\Delta_x}{r^2} & 0 & 0 & 0 \\ \frac{\Delta_y}{r^2}\frac{d\mathcal{E}}{d\phi} & \frac{-\dot{y}_k}{u^2}\frac{d\mathcal{E}}{d\phi} & 0 & \frac{-\Delta_x}{r^2}\frac{d\mathcal{E}}{d\phi} & \frac{\dot{x}_k}{u^2}\frac{d\mathcal{E}}{d\phi} & 0 & \frac{\mathcal{E}}{l} \end{bmatrix} \tag{4.29}$$

avec

$$\frac{d\mathcal{E}}{d\phi} = \frac{-l^2}{\mathcal{E}}\left(1 - (\frac{b}{a})^2\right)\frac{(\dot{y}_k\Delta_y + \dot{x}_k\Delta_x)(\dot{x}_k\Delta_y - \dot{y}_k\Delta_x)}{r^2u^2} \tag{4.30}$$

et

$$\phi = tan^{-1}(\frac{\dot{x}_k\Delta_y - \dot{y}_k\Delta_x}{\dot{x}_k\Delta_x + \dot{y}_k\Delta_y}) \tag{4.31}$$

Notons que l'expression (4.29) est définie pour des modèles du troisième ordre. Pour des modèles du second ordre, au regard de (4.3), la matrice Jacobienne est en réalité de taille 3×5 et est définie à partir de la Jacobienne (4.29) en enlevant les colonnes 3 et 6.

A ce stade du développement, pour simplifier (4.26) à partir de (4.27) et (4.28), on utilise le lemme suivant.

- *Soit Z un vecteur aléatoire s'exprimant comme suit : $Z = AX + BY + C$, où C est une constante. Si X et Y sont deux vecteurs aléatoires gaussiens indépendants qui satisfont $X \sim \mathcal{N}(X;m_X,P_X)$ et $Y \sim \mathcal{N}(Y;m_Y,P_Y)$ respectivement, alors Z est gaussien et satisfait $Z \sim \mathcal{N}(Z;m_Z,P_Z)$ avec :*

$$\begin{aligned} m_Z &= Am_X + Bm_Y + C \\ P_Z &= AP_XA^T + BP_YB^T \end{aligned} \tag{4.32}$$

Finalement, en appliquant ce lemme à (4.26), on obtient :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c) \simeq \sum_{i=1}^{r(c)} \mu_{k|k-1}^{(i,c)} \mathcal{N}(\boldsymbol{Y}_k; \boldsymbol{Y}_{k|k-1}^{(i,c)}, S_k^{(i,c)}) \tag{4.33}$$

où $\mathcal{N}(\boldsymbol{Y}_k; \boldsymbol{Y}_{k|k-1}^{(i,c)}, S_k^{(i,c)})$ désigne la densité de probabilité de l'innovation associée au $i^{\text{ème}}$ FKE pour l'IMM de la classe c à l'instant k et où $S_k^{(i,c)}$ et $\boldsymbol{Y}_{k|k-1}^{(i,c)}$ correspondent à la matrice de covariance de l'innovation et la mesure prédite respectivement définies comme suit :

$$\begin{aligned} \boldsymbol{Y}_{k|k-1}^{(i,c)} &= \boldsymbol{h}(\boldsymbol{X}_{k|k-1}^{(i,c)}) \\ S_k^{(i,c)} &= \boldsymbol{H}^{(i,c)} P_{k|k-1}^{(i,c)} (\boldsymbol{H}^{(i,c)})^T + \boldsymbol{R} \end{aligned} \tag{4.34}$$

L'**Algorithme 4.1** ci-dessous décrit les principales étapes du filtre IMM multi-classes présenté.

Algorithme 4.1 Etapes de l'algorithme IMM multi-classes proposé (MC-IMM)

Etape de prédiction :
- Etat du système : $p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k-1},c) \underset{(4.19)}{\simeq} \sum_{i=1}^{r(c)} \mu_{k|k-1}^{(i,c)} \mathcal{N}(\boldsymbol{X}_k; \boldsymbol{X}_{k|k-1}^{(i,c)}, P_{k|k-1}^{(i,c)})$
  avec :
  - $\boldsymbol{X}_{k|k-1}^{(i,c)} = \boldsymbol{F}^{m_k^c} \boldsymbol{X}_{k-1|k-1}^{(i,c)}$
  - $P_{k|k-1}^{(i,c)} = \boldsymbol{F}^{m_k^c} P_{k-1|k-1}^{(i,c)} (\boldsymbol{F}^{m_k^c})^T + \boldsymbol{Q}^{m_k^c}$
- Poids des modèles : $\mu_{k|k-1}^{(i,c)} \underset{(4.21)}{=} \sum_{j=1}^{r(c)} \Pi_{ji}^c \mu_{k-1}^{(j,c)}$

Etape de mise à jour :
- Etat du système : $p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k},c) \underset{(4.20)}{\simeq} \sum_{i=1}^{r(c)} \mu_k^{(i,c)} \mathcal{N}(\boldsymbol{X}_k; \boldsymbol{X}_{k|k}^{(i,c)}, P_{k|k}^{(i,c)})$
  avec :
  - $\boldsymbol{X}_{k|k}^{(i,c)} = \boldsymbol{X}_{k|k-1}^{(i,c)} + K_k^{(i,c)} (\boldsymbol{Y}_k - \boldsymbol{h}(\boldsymbol{X}_{k|k-1}^{(i,c)}))$
  - $P_{k|k}^{(i,c)} = (Id - K_k^{(i,c)} \boldsymbol{H}^{(i,c)}) P_{k|k-1}^{(i,c)}$
- Poids des modèles : $\mu_k^{(i,c)} \underset{(4.22)}{\propto} p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, m_k^c = i) \mu_{k|k-1}^{(i,c)}$
- Probabilités des classes : $Pr(c|\boldsymbol{Y}_{1:k}) \propto Pr(c|\boldsymbol{Y}_{1:k-1}) \sum_{i=1}^{r(c)} \mu_{k|k-1}^{(i,c)} \mathcal{N}(\boldsymbol{Y}_k; \boldsymbol{Y}_{k|k-1}^{(i,c)}, S_k^{(i,c)})$

L'avantage de cette IMM multi-classes est principalement sa facilité de mise en œuvre et son faible coût calculatoire notamment grâce à l'utilisation de FKEs. Néanmoins, cette approche peut être limitée étant donné l'approximation de la non-linéarité par un devéloppement de Taylor au premier ordre.

Dans la suite, nous étudions comme alternative, la pertinence d'une approche JTC reposant sur un filtrage particulaire multi-modèles.

4.3.2 Filtre particulaire multi-modèles multi-classes (MC-MMPF)

Nous proposons désormais d'utiliser un filtre MC-MMPF afin d'estimer conjointement au sein de chaque classe l'état $\{\boldsymbol{x}_k, l\}$ et le modèle de mouvement m_k^c en utilisant le filtrage particulaire. Introduisons dans un premier temps la notation suivante :

$$\bar{\boldsymbol{x}}_k = [\boldsymbol{x}_k^T \ m_k^c]^T \tag{4.35}$$

Selon (4.7), on peut observer que conditionnellement à $\bar{\boldsymbol{x}}_k$, le modèle dans l'espace d'état est linéaire gaussien vis-à-vis de la taille de cible l. Dans ce cas, on peut opter pour des filtres particulaires marginalisés car ils permettent de réduire la variance de l'erreur d'estimation pour un nombre donné de particules en résolvant analytiquement une partie du problème. Ils sont fondés sur la décomposition suivante de la densité de probabilité jointe *a posteriori* :

$$p(\bar{\boldsymbol{x}}_{0:k}, l|\boldsymbol{Y}_{1:k}, c) = p(l|\bar{\boldsymbol{x}}_{0:k}, \boldsymbol{Y}_{1:k}, c)p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k}, c) \tag{4.36}$$

Or, comme la densité de probabilité de longueur de la cible dépend seulement des mesures d'étalement distance $L_{1:k}$ et non des mesures $\boldsymbol{y}_{1:k}$, on peut ré-écrire (4.36) comme suit :

$$p(\bar{\boldsymbol{x}}_{0:k}, l|\boldsymbol{Y}_{1:k}, c) = p(l|\bar{\boldsymbol{x}}_{0:k}, L_{1:k}, c)p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k}, c) \tag{4.37}$$

où $p(l|\bar{\boldsymbol{x}}_{0:k}, L_{1:k}, c)$ est gaussien et ses moments du premier ordre et du second ordre peuvent être calculés récursivement par filtrage de Kalman.
Ainsi, pour chaque classe, seulement $p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k}, c)$ est estimé par filtrage particulaire comme suit :

$$\hat{p}(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k}, c) = \sum_{i=1}^{N_p} w_k^{(i,c)} \delta_{\bar{\boldsymbol{x}}_{0:k}^{(i,c)}}(\bar{\boldsymbol{x}}_{0:k}) \tag{4.38}$$

où comme introduit au chapitre 1, les points supports $\bar{\boldsymbol{x}}_{0:k}^{(i,c)}$ sont générés séquentiellement à partir d'une distribution d'importance notée $q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c)$ et les poids $\{w_k^{(i,c)}\}_{i=1,\ldots,N_p}$ sont calculés récursivement pour corriger la dégénérescence entre la vraie distribution *a posteriori* et la distribution d'importance. En exploitant les indépendances conditionnelles, on peut écrire :

$$\begin{aligned} \bar{w}_k^{(i,c)} &= \frac{p(\boldsymbol{Y}_k|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, \boldsymbol{Y}_{1:k-1})p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)}{q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c)} \\ &= w_{k-1}^{(i,c)} \frac{p(\boldsymbol{y}_k|\bar{\boldsymbol{x}}_k^{(i,c)})p(L_k|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1})p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)}{q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c)} \end{aligned} \tag{4.39}$$

et

$$w_k^{(i,c)} = \frac{\bar{w}_k^{(i,c)}}{\sum_i \bar{w}_k^{(i,c)}} \tag{4.40}$$

Puis, pour chaque particule $\bar{\boldsymbol{x}}_{0:k}^{(i,c)}$, il s'agit de calculer la forme exacte de la densité de probabilité conditionnelle de la longueur de la cible par filtrage de Kalman. Dans ce cas, il vient :

$$p(l|\bar{\boldsymbol{x}}_{0:k}^{i,c}, L_{1:k}, c) = \mathcal{N}(l; l_{k|k}^{(i,c)}, P_{k|k}^{(i,c)}) \tag{4.41}$$

où $l_{k|k}^{(i,c)}$ et $P_{k|k}^{(i,c)}$ sont les estimations de la longueur de la cible et de la matrice de covariance associées au $i^{\text{ème}}$ filtre de Kalman. L'algorithme consiste donc en un banc de filtres de Kalman, chacun étant associé à une des particules du filtre particulaire.
En insérant (4.38) et (4.41) dans (4.37) puis en intégrant par rapport à $\bar{\boldsymbol{x}}_{0:k}$, il vient :

$$\hat{p}(l|\boldsymbol{Y}_{1:k}, c) = \sum_{i=1}^{N_p} w_k^{(i,c)} \mathcal{N}(l; l_{k|k}^{(i,c)}, P_{k|k}^{(i,c)}) \tag{4.42}$$

Etant donné la distribution (4.42), la longueur de la cible peut alors être estimée au sens de l'espérance *a posteriori* comme suit :

$$l_k^c = \sum_{i=1}^{N_p} w_k^{(i,c)} l_{k|k}^{(i,c)} \tag{4.43}$$

Concernant la probabilité de la classe, elle peut être estimée récursivement à partir de (4.14) en calculant premièrement :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c) = \int_l \int_{\bar{\boldsymbol{x}}_{0:k}} p(\boldsymbol{Y}_k|l, \bar{\boldsymbol{x}}_{0:k}, \boldsymbol{Y}_{1:k-1}, c) p(l, \bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k-1}, c) d\bar{\boldsymbol{x}}_{0:k} dl \tag{4.44}$$

où en prenant en compte les indépendances conditionnelles, on a :

$$p(\boldsymbol{Y}_k|l, \bar{\boldsymbol{x}}_{0:k}, \boldsymbol{Y}_{1:k-1}, c) = p(L_k|l, \bar{\boldsymbol{x}}_{0:k}, L_{1:k-1}) p(\boldsymbol{y}_k|\bar{\boldsymbol{x}}_k) \tag{4.45}$$

et d'après Bayes :

$$p(l, \bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k-1}, c) = p(l|\bar{\boldsymbol{x}}_{0:k}, \boldsymbol{Y}_{1:k-1}, c) p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k-1}, c) \tag{4.46}$$

où au sein du filtre particulaire, la distribution *a priori* $p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k-1}, c)$ est approchée par :

$$p(\bar{\boldsymbol{x}}_{0:k}|\boldsymbol{Y}_{1:k-1}, c) \simeq \sum_{i=1}^{N_p} w_{k|k-1}^{(i,c)} \delta_{\bar{\boldsymbol{x}}_{0:k}^{(i,c)}}(\bar{\boldsymbol{x}}_{0:k}) \tag{4.47}$$

avec

$$w_{k|k-1}^{(i,c)} \propto w_{k-1}^{(i,c)} \frac{p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)}{q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, c, \boldsymbol{Y}_{1:k})} \tag{4.48}$$

En remplaçant (4.45), (4.46) et (4.47) dans (4.44) et en intégrant par rapport à $\bar{\boldsymbol{x}}_{0:k}$, on obtient :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1}, c) \approx \int_l \sum_{i=1}^{N_p} w_{k|k-1}^{(i,c)} p(L_k|l, \bar{\boldsymbol{x}}_{0:k}^{(i,c)}, c) p(\boldsymbol{y}_k|\bar{\boldsymbol{x}}_k^{(i,c)}) p(l|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1}, c) dl \tag{4.49}$$

avec :

$$p(L_k|l, \bar{\boldsymbol{x}}_{0:k}^{(i,c)}, c) = \mathcal{N}(L_k; h_{taille}(\bar{\boldsymbol{x}}_k^{(i,c)}) l, \sigma_{ed}^2) \tag{4.50}$$

$$p(l|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1}, c) = \mathcal{N}(l; l_{k|k-1}^{(i,c)}, P_{k|k-1}^{(i,c)}) \tag{4.51}$$

En regroupant les densités de probabilités qui font intervenir l, (4.49) devient :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c) \approx \sum_{i=1}^{N_p} w_{k|k-1}^{(i,c)} p(y_k|\bar{\boldsymbol{x}}_k^{(i,c)}) \underbrace{\int_l p(L_k|l,\bar{\boldsymbol{x}}_{0:k}^{(i,c)}) p(l|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1}) dl}_{p(L_k|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1})} \tag{4.52}$$

Finalement, (4.44) peut se réécrire comme suit :

$$p(\boldsymbol{Y}_k|\boldsymbol{Y}_{1:k-1},c) \approx \sum_{i=1}^{N_p} w_{k|k-1}^{(i,c)} p(y_k|\bar{\boldsymbol{x}}_k^{(i,c)}) \mathcal{N}(L_k; L_{k|k-1}^{(i,c)}, S_k^{(i,c)}) \tag{4.53}$$

où $L_{k|k-1}^{(i,c)}$ est la mesure d'étalement distance prédite associée à la $i^{\text{ème}}$ particule à l'instant k et $S_k^{(i,c)}$ est sa variance telle que :

$$L_{k|k-1}^{(i,c)} = h_{taille}(\bar{\boldsymbol{x}}_k^{(i,c)}) l_{k|k-1}^{(i,c)} \tag{4.54}$$

$$S_k^{(i,c)} = \left(h_{taille}(\bar{\boldsymbol{x}}_k^{(i,c)})\right)^2 P_{k|k-1}^{(i,c)} + \sigma_{ed}^2 \tag{4.55}$$

Notons qu'en pratique, nous utilisons comme distribution d'importance la distribution de transition *a priori* soit : $q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c) = p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)$.
L'**Algorithme 4.2** ci-après décrit les principales étapes du filtre particulaire multi-modèles multi-classes présenté.

Algorithme 4.2 Etapes de l'algorithme particulaire multi-classes proposé (MC-MMPF)

Pour chaque classe :

Générer les particules :

$$\bar{\boldsymbol{x}}_{0:k}^{(i,c)} \sim q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c) = p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)$$

Calculer les poids d'importance :

$$\bar{w}_k^{(i,c)} \underset{(4.39)}{=} w_{k-1}^{(i,c)} \frac{p(\boldsymbol{y}_k|\bar{\boldsymbol{x}}_k^{(i,c)}, \boldsymbol{y}_{1:k-1}) p(L_k|\bar{\boldsymbol{x}}_{0:k}^{(i,c)}, L_{1:k-1}) p(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{k-1}^{(i,c)}, c)}{q(\bar{\boldsymbol{x}}_k^{(i,c)}|\bar{\boldsymbol{x}}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c)} \tag{4.56}$$

Calculer la longueur de cible par filtrage de Kalman :

$$p(l|\bar{\boldsymbol{x}}_{0:k}^{i,c}, L_{1:k}, c) \underset{(4.41)}{=} \mathcal{N}(l; l_{k|k}^{(i,c)}, P_{k|k}^{(i,c)}) \tag{4.57}$$

Mettre à jour des probabilités de classes :

$$Pr(c|\boldsymbol{Y}_{1:k}) \propto Pr(c|\boldsymbol{Y}_{1:k-1}) \sum_{i=1}^{N_p} w_{k|k-1}^{(i,c)} p(y_k|\bar{\boldsymbol{x}}_k^{(i,c)}) \mathcal{N}(L_k; L_{k|k-1}^{(i,c)}, S_k^{(i,c)}) \tag{4.58}$$

Dans la suite, nous tirons profit du modèle BNP proposé dans le chapitre précédent pour concevoir une approche JTC qui estime, conjointement à la classe de la cible et au vecteur d'état, les paramètres de la matrice de précision du bruit de modèle.

4.3.3 Extension au cas où les paramètres du modèle d'état sont inconnus

Les algorithmes présentés précédemment reposent sur l'hypothèse qu'une cible appartenant à une classe donnée peut seulement alterner entre un nombre fini de modèles de mouvement dont les variances des bruits de modèles sont connues et réglées *a priori*. Afin de tirer profit des approches développées dans le chapitre précédent, nous proposons de mettre en œuvre une approche JTC en levant les *a priori* sur la matrice de covariance des bruits de modèles pour les différentes familles de mouvement considérées. Pour cela, nous introduisons un modèle BNP, similaire à celui développé dans le chapitre 3, afin d'estimer la distribution des paramètres de précision des différentes familles de mouvement considérées.

4.3.3.1 Modèle hiérarchique proposé

Parmi les C classes considérées, chacune est cette fois associée à $f(c)$ familles de modèles. On note $z_k^c \in \{1, ..., f(c)\}$ l'indice désignant la famille de modèles de mouvement de la cible à l'instant k pour la classe c. La séquence $\{z_k^c\}_{k>0}$ est alors supposée être une chaîne de Markov de matrice de transition $\pi^c = \{\pi_{ij}^c\}_{i=1,...,f(c)}^{j=1,...,f(c)}$. Notons que contrairement aux approches introduites jusqu'à présent dans ce chapitre, l'indice z_k^c ne fait pas référence à un modèle bien défini mais à une famille composée d'une infinité de modèles caractérisés par différentes valeurs de la matrice de précision $\Gamma^{z_k^c}$. Plus précisément, comme suggéré dans le chapitre 3, la matrice de covariance du bruit de modèle $\Gamma^{z_k^c}$ est supposée avoir une forme fonctionnelle connue *a priori* et qui ne varie que selon le paramètre de précision γ_k.
Dans la suite, il s'agit d'estimer pour chaque famille de modèles z_k^c le paramètre de précision afin d'en déduire la matrice de covariance du bruit de modèle.

Conditionellement à z_k^c, la distribution de γ_k est représentée par un processus de Dirichlet. On considère donc autant de processus de Dirichlet que de familles de modèles pour une classe de cible donnée. Leurs réalisations sont notées $\{G^{c,m}\}_{m=1,...,f(c)}^{c=1,...,C}$.
Chaque processus de Dirichlet est caractérisé par sa propre distribution de base $G_0^{c,m}$ ainsi que son propre facteur d'échelle $\alpha_0^{c,m}$. A partir de ces notations, les relations entre z_k^c, $G^{c,m}$, γ_k, $\boldsymbol{X}_k$ et $\boldsymbol{Y}_k$ peuvent être décrites selon le modèle hiérarchique suivant :

$$z_k^c | z_{k-1}^c \sim \pi_{z_k^c z_{k-1}^c}^c \tag{4.59}$$

$$G^{c,m} \sim \mathcal{DP}(G_0^{c,m}, \alpha_0^{c,m}) \text{ pour } \; m = 1, ..., f(c) \tag{4.60}$$

$$\gamma_k | \{z_k^c, \{G^{c,m}\}_{m=1,...,f(c)}\} \sim G^{c,z_k^c}(\gamma_k) \tag{4.61}$$

$$\boldsymbol{X}_k | \{\boldsymbol{X}_{k-1}, \gamma_k, z_k^c\} \sim p(\boldsymbol{X}_k | \boldsymbol{X}_{k-1}, \gamma_k, z_k^c) \tag{4.62}$$

$$\boldsymbol{Y}_k | \boldsymbol{X}_k \sim p(\boldsymbol{Y}_k | \boldsymbol{X}_k) \tag{4.63}$$

En s'appuyant sur la représentation en Urne de Polya, la distribution de prédiction de γ_k est

définie comme suit :

$$\gamma_k|\{\gamma_{1:k-1}, z^c_{1:k}, c\} \sim \frac{1}{\alpha_0^{z^c_k} + n_{z^c_k}} \sum_{\substack{j=1 \\ t.q.\ z^c_j = z^c_k}}^{k-1} \delta_{\gamma_j}(\gamma_k) + \frac{\alpha_0^{z^c_k}}{\alpha_0^{z^c_k} + n_{z^c_k}} G_0^{z^c_k}(\gamma_k) \tag{4.64}$$

où $n_{z^c_k}$ désigne le nombre de fois où la famille de modèles z^c_k est apparue précédemment. Il satisfait :

$$n_{z^c_k} = \sum_{j=1}^{k-1} \delta_{z^c_k}(z^c_j) \tag{4.65}$$

Le modèle hiérarchique défini par (4.59)-(4.63) se réduit ainsi à (4.59), (4.64), (4.62) et (4.63). Les relations existant entre les différentes variables sont représentées par la figure 4.3.

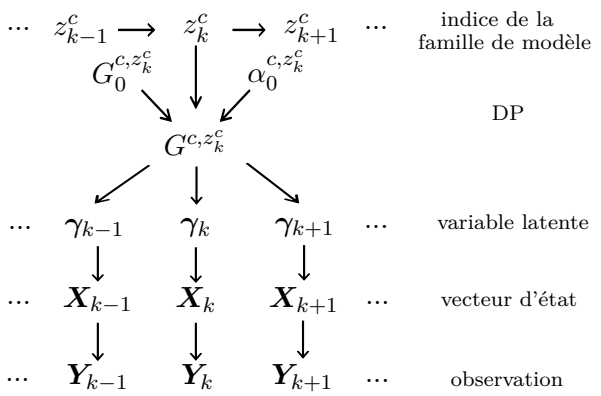

FIGURE 4.3 – Représentation graphique du modèle hiérarchique

L'objectif est alors d'estimer pour chaque classe la distribution jointe *a posteriori* notée $p(l, \boldsymbol{x}_{0:k}, z^c_{0:k}, \gamma_{0:k}|\boldsymbol{Y}_{1:k}, c)$ de tous les paramètres inconnus. Cette densité n'admet pas d'expression analytique du fait de la non-linéarité et la non-gaussianité du modèle proposé. Par conséquent, des techniques de filtrage particulaire sont utilisées.

Comme le modèle hiérarchique est linéaire gaussien pour la longueur de cible l conditionnellement à $\boldsymbol{x}_{0:k}$, $z^c_{0:k}$ et $\gamma_{0:k}$, un filtre particulaire marginalisé est utilisé au sein de chaque classe. Il estime les états $\boldsymbol{x}_{0:k}$, les indices de la famille de modèle $z^c_{0:k}$ et les paramètres de précision $\gamma_{0:k}$ alors que la taille de la cible est estimée par filtrage de Kalman. Le principe de l'algorithme étant le même que dans la partie 4.3.2, la mise en œuvre n'est pas redétaillée.

Concernant les particules, on pourrait les simuler selon le modèle *a priori* (4.59), (4.62), (4.64). Dans ce cas, les poids des particules seraient simplement proportionnels à la vraisemblance suivante :

$$w_k^{(i,c)} \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i,c)}) p(L_k|\boldsymbol{x}_{0:k}^{(i,c)}, \gamma_{0:k}^{(i,c)}, z_{0:k}^{(i,c)}, L_{1:k-1}) \tag{4.66}$$

Cependant, un nombre important de particules serait nécessaire pour approcher finement la densité de probabilité *a posteriori* qui est multi-dimensionnelle.

Nous considérons donc, comme au chapitre 3, une approximation de la distribution d'importance optimale [DOU01a], soit :

$$q(\gamma_k^{(i,c)}, \boldsymbol{x}_k^{(i,c)}, z_k^{(i,c)} | \gamma_{0:k-1}^{(i,c)}, \boldsymbol{x}_{0:k-1}^{(i,c)}, z_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}^{(i,c)}) = q(\gamma_k^{(i,c)} | \gamma_{0:k-1}^{(i,c)}, \boldsymbol{x}_{0:k}^{(i,c)}, \boldsymbol{Y}_{1:k}, c, z_k^{(i,c)}) \times q(\boldsymbol{x}_k^{(i,c)} | \boldsymbol{x}_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}, c, z_k^{(i,c)}) Pr(z_k^{(i,c)} | z_{k-1}^{(i,c)}) \tag{4.67}$$

Finalement, concernant l'estimation de la classe, la même expression que dans (4.39) est utilisée mais l'expression des poids de prédiction est modifiée comme suit :

$$w_{k|k-1}^{(i,c)} \propto w_{k-1}^{(i,c)} \frac{p(\boldsymbol{x}_k^{(i,c)} | \boldsymbol{x}_{k-1}^{(i,c)}, \gamma_k^{(i,c)}, c) p(\gamma_k^{(i,c)} | \gamma_{0:k-1}^{(i,c)}, c)}{q(\gamma_k^{(i,c)}, \boldsymbol{x}_k^{(i,c)} | \gamma_{0:k-1}^{(i,c)}, \boldsymbol{x}_{0:k-1}^{(i,c)}, z_{0:k-1}^{(i,c)}, \boldsymbol{Y}_{1:k}^{(i,c)})} \tag{4.68}$$

Dans la suite, nous appliquons les trois approches présentées précédemment au problème de pistage de cibles marines.

4.3.4 Application au pistage d'une cible marine

4.3.4.1 Etape préliminaire : classe d'une cible et paramétrage

Avant d'analyser la qualité de nos approches, nous définissons les classes de cibles considérées. Selon le tableau 4.1 où les longueurs de cible et les accélérations maximales sont données pour différents types de cibles, on peut en déduire qu'une cible est potentiellement manœuvrante si sa longueur est suffisamment petite. Notons que le terme "potentiellement" est utilisé car les cibles considérées comme manœuvrantes peuvent ne pas manœuvrer.

Dans la suite, une cible marine est supposée appartenir à l'une des $C = 2$ classes suivantes :
Classe 1 : cibles non-manœuvrantes,
Classe 2 : cibles manœuvrantes.

Notons néanmoins que l'on aurait pu sans restriction considérer plus de classes.

Décrivons à présent le nombre et le type de modèles de mouvement dans chaque classe.
Pour la classe de cible non-manœuvrante, $r(1) = 1$: le modèle de mouvement M_1^1 est un MRU avec $\sigma_{MRU,1}$ l'écart-type sur le bruit de modèle.
Pour la classe des cibles manœuvrantes, nous suggérons d'utiliser $r(2) = 2$ modèles : M_1^2 est un modèle de mouvement dont l'écart-type de l'accélération est $\sigma_{MRU,2}$ et M_2^2 est un modèle de Singer avec pour écart-type sur l'accélération et constante de temps $\sigma_{Sin,2}$ et $\tau_{Sin,2}$ respectivement. Le passage d'un modèle à un autre est piloté par la chaîne de Markov (4.1).
Enfin, notons que la loi initiale sur la longueur en fonction de la classe représentée en figure 4.4 est définie comme suit :

$$\begin{aligned} p(l) &= \frac{1}{C} \sum_{c'=1}^{C} p(l|c') \\ &= \frac{1}{C} \sum_{c'=1}^{C} \mathcal{N}(l; l_0^{c'}, P_0^{c'}) \end{aligned} \tag{4.69}$$

Type de cible		Longeur de cible (m)	accélération maximale	Classe de cible
	Cargos	≥ 100	$0.02g$	Cibles
	Pétroliers	≥ 75	$0.02g$	non-
	Remorqueurs	≥ 50	$0.02g$	manœuvrantes
	Bâteaux de pêche	≈ 20	$0.1g$	
	Bâteaux de plaisance	≈ 10	$0.2g$	Cibles
	Bâteaux pneumatiques	≈ 10	$0.4g$	manœuvrantes
	Jetski	≈ 3	$0.5g$	

TABLE 4.1 – Caractéristiques des différentes cibles marines

avec $l_0^1 = 120\ m$, $P_0^1 = 400\ m^2$, $l_0^2 = 10\ m$ et $P_0^2 = 400\ m^2$.

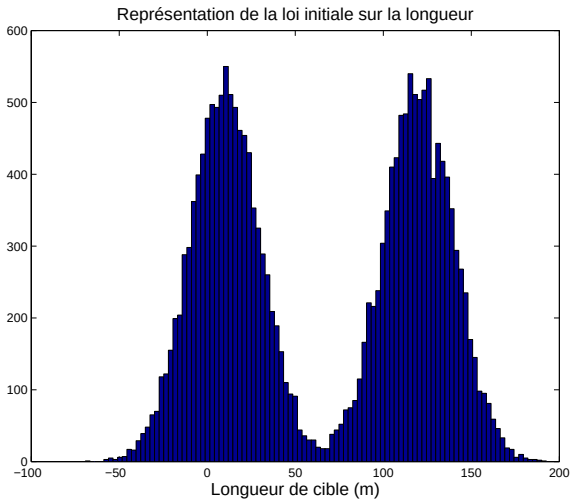

FIGURE 4.4 – Loi initiale sur la longueur de la cible

4.3.4.2 Protocole de simulation

Un cargo de 120 m (cible 1) et un bâteau pneumatique de 10 m (cible 2) sont poursuivis séparément.
Nous décrivons dans la suite le scénario adopté pour générer les trajectoires de cibles ainsi que le paramétrage retenu pour les différentes approches JTC proposées dans ce chapitre. Bien que le scénario choisi et présenté ci-après soit spécifique, il se veut représentatif du contexte de pistage de cibles marines. Nous testerons néanmoins quelques variations dans la suite de cette partie.

Génération des trajectoires :
A chaque instant, la trajectoire de la cible 1 est générée à partir d'un modèle de mouvement MRU dont l'écart-type de l'accélération $\sigma_{MRU,1}$ peut prendre deux valeurs, à savoir 5×10^{-2} $m.s^{-3/2}$ et 0.2 $m.s^{-3/2}$. La probabilité de changer d'une valeur à une autre est fixée à 0.02.
Concernant la cible 2, sa trajectoire est générée à partir d'une chaîne de Markov d'ordre deux dont la matrice de transition est $\begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$. Le premier état caractérise un MRU alors que le second est un modèle de Singer. Tant que le mouvement de la cible est décrit par la même famille de modèles, le bruit de modèle associé reste constant et peut prendre l'une des valeurs suivantes de façon équiprobable : pour le modèle MRU, $\sigma_{MRU} \in \{10^{-1}; 2 \times 10^{-1}\}$ $m.s^{-3/2}$ et pour le modèle de Singer $\sigma_{Sin} \in \{1; 2; 3; 4; 5\} m.s^{-2}$.

Estimation des trajectoires des cibles :
Les trajectoires des cibles sont estimées à partir des mesures bruitées qui incluent la distance et l'angle à laquelle elles se situent ainsi que leur étalement distance sur l'axe radar-cible. Concernant les bruits de mesure sur la distance d_k et l'angle θ_k, trois configurations sont testées :

- *Configuration 1* : $\sigma_d = 10$ m, $\sigma_\theta = 0.001$ $rad.$
- *Configuration 2* : $\sigma_d = 10$ m, $\sigma_\theta = 0.005$ $rad.$
- *Configuration 3* : $\sigma_d = 20$ m, $\sigma_\theta = 0.001$ $rad.$

De plus, l'écart-type du bruit de mesure sur la mesure d'étalement distance de la cible est fixé à $\sigma_{ed} = 5$ m et la période d'échantillonnage vaut $T = 1$ s.
A titre illustratif, une réalisation de trajectoire pour la cible 2 et les mesures radar associées sont présentées sur la figure 4.5.

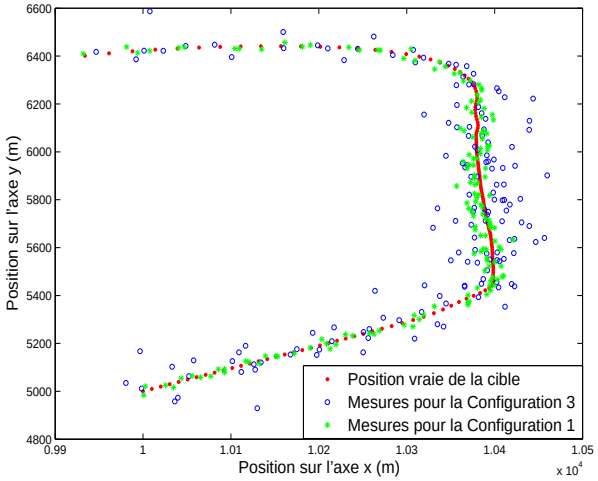

FIGURE 4.5 – Trajectoire de la cible 2 et mesures associées

Les paramètres pour les trois approches proposées sont les suivants :

Algorithme IMM multi-classes (*MC-IMM*) :
Les paramètres du modèle sont : $\sigma_{MRU,1} = \sigma_{MRU,2} = 0.2\ m.s^{-3/2}$, $\sigma_{Sin,2} = 5\ m.s^{-2}$ et $\tau_{Sin,2} = 15\ s$. De plus, la TPM est : $\Pi^{c=2} = \begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$. Cela correspond à un temps de séjour moyen de $50T$ et $10T$ dans le premier et le second modèle respectivement.

Algorithme MMPF multi-classes (*MC-MMPF*) :
Pour les deux classes, les mêmes paramètres que pour l'algorithme MC-IMM sont considérés. De plus, l'algorithme est mis en œuvre avec $N_p = 3000$ particules.

Algorithme non-paramétrique multi-classes (*MC-BNP*) :
Pour chaque classe, la variance du bruit de modèle γ_k correspondant à la famille de modèles z_k^c est estimée.
Pour la première classe, $f(1) = 1$. $z_k^c = 1$ fait ainsi référence au MRU, alors que pour la seconde classe $f(2) = 2$ avec $z_k^c \in \{1, 2\}$ qui fait référence soit à un MRU soit à un modèle de Singer. Dans ce dernier cas, la chaîne de Markov associée à cette famille de modèles est caractérisée par la TPM: $\Pi^{c=2} = \begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$.
Comme suggéré dans [CAR08], la distribution de base G_0 est définie comme étant une loi Gamma conjuguée *a priori* $\Gamma(a_{z_k^c}, b_{z_k^c})$ sur le paramètre de précision. Pour la première classe : $a_1 = 4$, $b_1 = 40$. Pour la seconde : $a_1 = 4$, $b_1 = 40$ et $a_2 = 10$, $b_2 = 0.01$. Enfin, l'algorithme est également mis en œuvre avec $N_p = 3000$ particules.

Les résultats présentés dans la suite sont obtenus avec des trajectoires de 200 échantillons moyennés sur 100 simulations de Monte Carlo.

4.3.4.3 Résultats de simulation

Dans la suite, nous donnons quelques informations sur le coût calculatoire et la pertinence de ces algorithmes concernant l'estimation de la classe ainsi que l'estimation de la position et de la longueur de la cible.

Coût calculatoire :
Au regard des tests qui ont été menés, par rapport à un FKE fondé sur une unique hypothèse de mouvement et un algorithme IMM qui n'incluent pas la classification, la complexité calculatoire de *MC-IMM* est seulement légèrement plus élevée. Cependant, concernant l'algorithme *MC-MMPF*, sa complexité est plus importante (environ 50 fois plus élévée que *MC-IMM* lorsque $N_p = 3000$) et est étroitement liée au nombre de particules utilisées. Enfin, la complexité calculatoire de *MC-BNP* est la plus importante parmi toutes les approches proposées (environ 5 fois plus élevée que *MC-MMPF*) mais c'est essentiellement lié à l'utilisation de la loi de proposition optimale.
Enfin, bien que ce ne soit pas le choix que nous ayons retenu, notons que quelle que soit l'approche JTC considérée, les calculs pourraient être menés en parallèle pour chacune des classes considérées. Dans ce cas, la complexité calculatoire ne serait pas liée au nombre de classes.

Estimation de la classe :
Selon la figure 4.6, où sont représentées les probabilités des classes pour les trois approches proposées et les 40 premiers échantillons, utiliser les mesures cinématiques et les mesures de l'étendue de la cible sur l'axe distance permet d'estimer la classe de la cible correctement et rapidement.

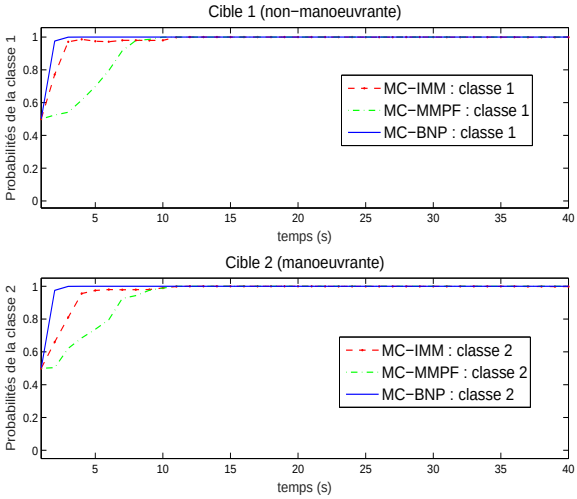

FIGURE 4.6 – Probabilités des classes de la cible moyennées sur 100 simulations de Monte Carlo

Quelle que soit la méthode considérée, la mise à jour de la probabilité de classe se fait selon un terme de vraisemblance qui caractérise la pertinence de la classe selon la position estimée et la longueur estimée de la cible. Ainsi, la discrimination d'une classe par rapport à une autre se fait à deux niveaux : par le choix des modèles de mouvement associés à chaque classe et par l'estimation de la longueur de cible vis-à-vis de la loi *a priori* définie en (4.69).

Pour les cas représentés sur la figure 4.6, les probabilités de classes convergent très rapidement et de façon similaire pour les trois approches proposées. Plus particulièrement, du fait de l'estimation de la variance du bruit, la convergence de *MC-BNP* est légèrement plus rapide.
De plus, bien qu'étant liée au choix des modèles de mouvement pour chacune des classes, pour les cas analysés, le vitesse de convergence dépend essentiellement du choix de la loi *a priori* de la longueur de cible définie en (4.69).

A titre d'exemple et comme cas limite, nous représentons sur la figure 4.7, les probabilités de classes pour une cible de 60 mètres de long. Compte tenu de (4.69), il est dans ce cas plus difficile de déterminer la classe à laquelle appartient la cible. Cela dépend alors essentiellement du fait que la cible manœuvre ou non.
Nous considérons alors deux cas. Dans le premier cas, la cible se déplace en MRU tout le long de la simulation. Dans le second cas, la cible se déplace en MRU pendant 100 secondes puis effectue une manœuvre qui dure 30 secondes puis se redéplace en MRU.

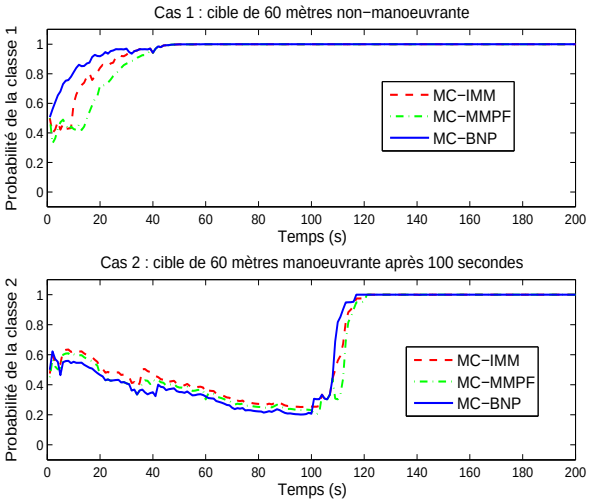

FIGURE 4.7 – Probabilités des classes pour une cible de 60 mètres manœuvrante (cas 1) ou non-manœuvrante (cas 2)

En analysant la figure 4.7, on s'aperçoit au travers de ce cas limite que la classification est plus délicate. Contrairement aux cas présentés précédemment où l'estimation de la longueur de la cible contribuait de façon importante à sa classification, la convergence des probabilités des classes est dans ce cas plus lente quelle que soit l'approche considérée. Néanmoins, pour le cas 2, lorsque la cible effectue alors une manœuvre après 100 secondes, l'estimation de l'état de la cible est plus précise pour la classe 2 que la classe 1. Cela permet d'identifier très rapidement que la cible appartient à la classe des cibles potentiellement manœuvrantes.

Notons que pour une longueur de cible donnée, la vitesse de convergence dépend également de l'écart entre les moyennes l_0^1 et l_0^2 définies en (4.69). Afin de vérifier ce comportement, nous représentons en figure 4.8, le temps de convergence à 95% de la probabilité de classe pour les

deux cibles pistées en fonction de δl qui modifie la loi *a priori* définie en (4.69) comme suit :

$$\begin{aligned} l_0^1 &= 60 - \delta l \\ l_0^2 &= 60 + \delta l \end{aligned} \tag{4.70}$$

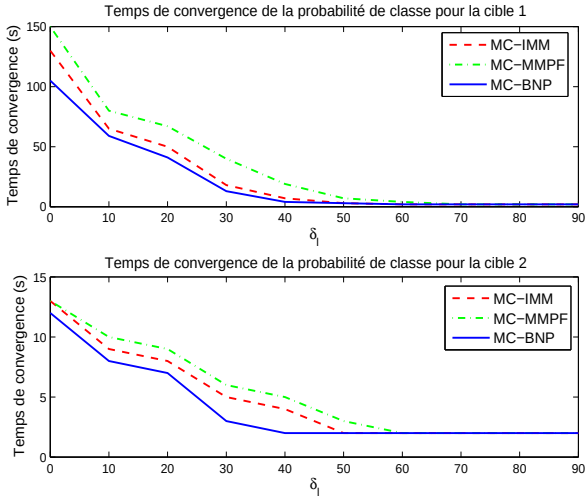

FIGURE 4.8 – Temps de convergence des probabilités de classes selon la forme de la loi *a priori* sur la longueur de cible

Pour les deux cibles, les probabilités de classes convergent de plus en plus vite avec δ_l. Pour la cible non-manœuvrante, lorsque δ_l varie de 0 à 90, on passe d'un temps de convergence de 150 secondes à 2 secondes. Lorsque δ_l est nul, c'est exclusivement la précision de l'estimation liée au choix des modèles de chaque classe qui permet d'identifier la classe de la cible. Il faut alors un certain temps pour que les probabilités des classes convergent. Lorsque δ_l augmente, la mesure de longueur est de plus en plus discriminante, ce qui permet d'identifier la classe de la cible plus rapidement.

Pour la cible manœuvrante, la cible suit à l'instant initial un MRU puis peut à chaque instant changer de mouvement pour effectuer une manœuvre. La cible peut alors manœuvrer à tout instant mais avec une probabilité de plus en plus élevée avec le temps. Comme lorsqu'une cible manœuvre la probabilité de classe converge rapidement, les temps de convergence sont plus faibles pour la cible manœuvrante que la cible non-manœuvrante.

Enfin, notons que le MC-MMPF converge dans ce cas un peu moins rapidement que le MC-IMM. Néanmoins, cette observation est étroitement liée à la non-linéarité du problème et donc plus précisément à la période d'échantillonnage T et à la dynamique des cibles.

Estimation de la position :

L'EQM sur les positions pour les trois approches proposées est comparée avec :

1/ l'EQM associée au bruit de mesure, i.e. sans appliquer d'algorithme de filtrage. Elle est notée *Mes.*

2/ l'EQM lorsqu'un FKE est utilisé, noté *Ref.* et qui se fonde sur le modèle de mouvement qui correspond au modèle de génération de la trajectoire à chaque instant.

3/ l'EQM lorsqu'un *IMM* sans classification combinant deux FKEs est mis en œuvre. Le premier

est fondé sur un MRU avec $\sigma_{MRU} = 2.10^{-1}\ m.s^{-\frac{3}{2}}$. Le second repose sur un modèle de Singer avec $\sigma_{Sin} = 5\ m.s^{-2}$.

Configuration	*Mes.*	*Ref.*	*IMM*	*MC-IMM*	*MC-MMPF*	*MC-BNP*
Configuration 1	15.33	7.88	8.62	8.16	8.27	**8.04**
Configuration 2	53.91	26.44	27.59	27.01	26.95	**26.62**
Configuration 3	59.25	33.43	34.91	34.37	34.23	**33.67**

Table 4.2 – EQM pour la cible 1 (non-manœuvrante)

Configuration	*Mes.*	*Ref.*	*IMM*	*MC-IMM*	*MC-MMPF*	*MC-BNP*
Configuration 1	14.75	8.08	8.92	8.91	8.78	**8.43**
Configuration 2	54.23	27.57	28.46	28.43	28.35	**27.85**
Configuration 3	60.68	34.32	35.68	35.66	35.32	**34.77**

Table 4.3 – EQM pour la cible 2 (potentiellement manœuvrante)

Selon les tableaux 1.2 et 1.3, comparativement à l'*IMM*, les trois approches proposées ont de meilleures performances pour l'estimation de la position. En particulier, comme la variance du bruit de modèle est estimée à chaque instant avec *MC-BNP*, cette approche présente de meilleurs résultats en termes d'estimation que les approches pour lesquelles la variance est fixée *a priori*. De plus, *MC-MMPF* présente une EQM plus faible que *MC-IMM* du fait de la non-linéarité de la fonction de mesure.
Finalement, on peut remarquer que *MC-IMM* et *IMM* ont des performances proches lorsqu'il s'agit de la cible manœuvrante mais différentes pour la cible non-manœuvrante. En effet, pour la cible 2, si la classification est correcte, les algorithmes *IMM* et *MC-IMM* sont équivalents, alors que pour la cible 1 *MC-IMM* se réduit à un FKE fondé sur un seul MRU. Ce dernier est alors plus performant que *MC-IMM* en raison du mélange des estimations dans la stratégie de coopération de l'IMM qui a pour conséquence le fait que le poids en faveur du modèle suivi par la cible n'est pas toujours prépondérant.

<u>Influence de l'estimation de la taille de la cible :</u>
Finalement, nous analysons l'estimation de la taille de la cible pour une réalisation de trajectoire de la cible 1 (non-manœuvrante) et de la cible 2 (manœuvrante) obtenues avec la *Configuration* 2.

Comme illustré par la figure 4.9, l'estimation de la longueur de la cible converge rapidement vers la vraie valeur. Cette caractéristique est en particulier à l'origine de la convergence rapide des probabilités des classes. Pour mettre en avant l'intérêt d'estimer la longueur de cible, nous représentons sur la figure 4.10 l'évolution de la probabilité de classe pour l'approche *MC-IMM* avec et sans estimation de la longueur. On remarque que la convergence est plus lente lorsque l'estimation de la longueur n'est pas considérée. Par conséquent, l'EQM obtenue sans intégrer la taille de la cible au vecteur d'état est légèrement plus importante que celle obtenue avec les mesures d'étalement distance.

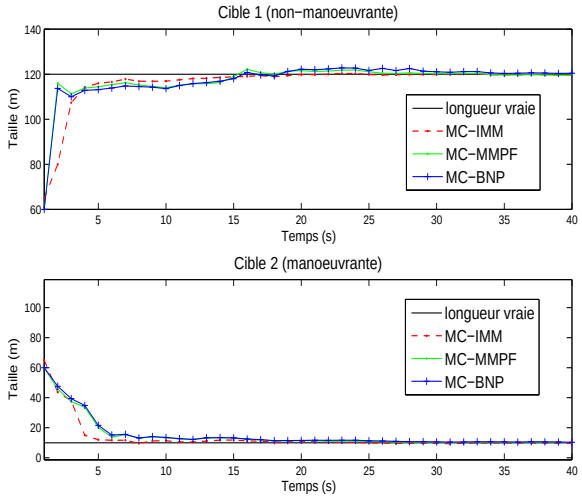

FIGURE 4.9 – Estimation de la taille de la cible, moyennée sur 100 simulations de Monte Carlo

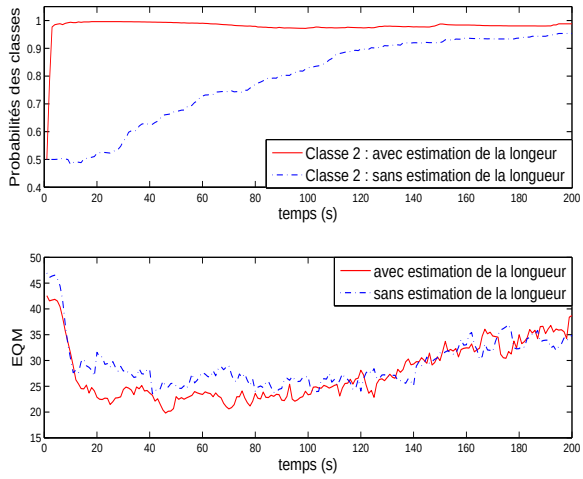

FIGURE 4.10 – Comparaison des performances du *MC-IMM* avec et sans les mesures de l'étalement distance de la cible

L'utilisation de mesures de l'étalement distance d'une cible au cours du temps s'avère donc être un choix algorithmique judicieux afin d'améliorer les performances du pistage dans un contexte monocible. De plus, le résultat de la classification obtenue pourrait être utilisé à d'autres fins que le pistage lui-même et pourrait apporter une information complémentaire dans le cadre d'une mission opérationnelle. Par exemple, en surveillance maritime, en cas de situation anormale comme par exemple la détection d'une petite cible en pleine mer, l'opérateur pourrait demander un changement de mode du radar pour caractériser la menace potentielle par une image radar afin de confirmer ou non la menace.

Dans la partie suivante, nous analysons la pertinence de ces approches JTC vis-à-vis des algorithmes de pistage multi-cibles. La classification aide alors à discriminer les différentes cibles les unes des autres.

4.4 Pistage et classification de plusieurs cibles à partir de mesures cinématiques et de mesures d'étalement distance

Ces dernières années, le pistage multi-cibles a été un sujet largement traité. Deux approches ont été principalement introduites : celles fondées sur les algorithmes d'association telles que l'algorithme JPDA [For83] ou MHT [Bla04] et les approches fondées sur la théorie des EFAs [Mah07]. Dans ce dernier cas, l'état multi-cibles ainsi que l'ensemble des observations sont modélisés à partir d'EFAs. Ainsi, les cibles sont pistées dans leur ensemble sans être ordonnées ou explicitement reliées aux mesures. Cette représentation du système a donné lieu dans un premier temps au filtre de Bayes multi-cibles qui propage récursivement la densité de probabilité de l'état multi-cibles. Néanmoins, son implémentation n'étant pas directement réalisable, le filtre PHD qui correspond à son approximation au premier ordre a été introduit. Plus particulièrement, le filtre GM-PHD [Vo06] et le filtre SMC-PHD [Pun08] ont été développés.

Dans la partie précédente, nous avons présenté un algorithme JTC fondé sur l'utilisation de mesures cinématiques et des mesures d'étalement distance dans un contexte de pistage monocible. Dans la suite, nous proposons une extension au cas multi-cibles. Dans le cadre de la thèse, nous proposons d'utiliser des filtres du type PHD qui sont des algorithmes fondés sur les EFAs qui ont été le plus largement étudiés. De façon similaire au cas monocible, chaque classe de cibles est caractérisée par ses propres modèles de mouvement et est donc associée à un filtre GM-PHD multi-modèles noté MM-GMPHD. Les pistes sont alors extraites pour chaque classe puis recombinées *a posteriori* en prenant en compte les probabilités de classes.

4.4.1 Modélisation multi-cibles

Dans la suite de cette sous-partie, nous présentons tout d'abord la représentation multi-cibles adoptée puis nous détaillons les équations théoriques de l'algorithme MM-PHD multiclasses. Enfin, nous présentons son implémentation par mélange de gaussiennes.

4.4.1.1 Représentation multi-cibles

Dans un contexte multi-cibles, le nombre de cibles peut évoluer au cours du temps. De plus, le nombre de mesures reçues à chaque instant peut être différent du nombre de cibles du fait de la probabilité de détection P_D mais aussi à cause des fausses alarmes essentiellement dues à la présence de clutter. Dans ces conditions, comme introduit dans le chapitre 1, le pistage multi-cibles peut être traité à partir de la théorie des EFAs. Ainsi, si l'on considère un ensemble de $N(k)$ cibles et un ensemble de $M(k)$ mesures à l'instant k, l'état multi-cibles $\boldsymbol{\mathcal{X}}_k$ et l'observation multi-cibles $\boldsymbol{\mathcal{Y}}_k$ peuvent être représentés par les ensembles finis suivants :

$$\begin{aligned} \boldsymbol{\mathcal{X}}_k &= \{\boldsymbol{X}_k^{(1)}, ..., \boldsymbol{X}_k^{(N(k))}\} \in \mathcal{F}(\mathbb{X}) \\ \boldsymbol{\mathcal{Y}}_k &= \{\boldsymbol{Y}_k^{(1)}, ..., \boldsymbol{Y}_k^{(M(k))}\} \in \mathcal{F}(\mathbb{Y}) \end{aligned} \tag{4.71}$$

avec $\{\boldsymbol{X}_k^{(i)}\}_{i=1}^{N(k)}$ et $\{\boldsymbol{Y}_k^{(i)}\}_{i=1}^{M(k)}$ définis selon (4.8) et (4.11) respectivement. Rappelons que $N(k)$ et $M(k)$ sont eux-mêmes considérés comme des variables aléatoires. De plus, $\mathcal{F}(\mathbb{X})$ et $\mathcal{F}(\mathbb{Y})$ sont les collections de tous les sous-ensembles finis de l'espace d'état monocible $\mathbb{X}$ et de l'espace

d'observation $\mathbb{Y}$ respectivement.

4.4.2 Algorithme PHD multi-modèles (MM-PHD) multi-classes

Considérons que chaque classe de cibles est caractérisée par un ensemble de modèles cinématiques et est associée à un filtre PHD multi-modèles (MM-PHD). Alors, à chaque filtre correspond un ensemble d'états extraits. Ceux-ci sont ensuite recombinés selon la probabilité *a posteriori* de chaque classe.

Dans le cas multi-cibles, l'objectif est de déterminer pour chaque cible la classe à laquelle elle appartient et d'estimer l'état multi-cibles $\boldsymbol{\mathcal{X}}_k$ étant donné les ensembles de mesures entre les instants 1 et k notés $\boldsymbol{\mathcal{Y}}_{1:k}$.
Dans le cas monocible, pour estimer l'état *a posteriori*, nous avons vu qu'il s'agissait dans un premier temps de calculer les probabilités de chacune des classes selon (4.14), puis de calculer l'estimation finale selon (4.16) en pondérant l'estimation de chaque classe par la probabilité associée.
Dans le cas multi-cibles, il n'est pas possible de pondérer un état multi-cibles $\boldsymbol{\mathcal{X}}_k^c$ par une probabilité de classe puisque celle-ci est nécessairement différente pour chacun des éléments de $\boldsymbol{\mathcal{X}}_k^c$. Par conséquent, nous proposons d'attribuer une probabilité de classe à chacun de ses éléments, et l'état multi-cibles estimé est obtenu comme suit :

$$\hat{\boldsymbol{\mathcal{X}}}_k = \{\hat{\boldsymbol{X}}_k^{(1)}, ..., \hat{\boldsymbol{X}}_k^{(\hat{N}(k))}\} \tag{4.72}$$

où $\hat{N}(k)$ est le nombre estimé de cibles et chaque état $\{\hat{\boldsymbol{X}}_k^{(i)}\}_{i=1,...,\hat{N}(k)}$ vérifie :

$$\hat{\boldsymbol{X}}_k^{(i)} = \sum_{c=1}^{C} Pr(c, i|\boldsymbol{Y}_{1:k})\hat{\boldsymbol{X}}_k^{(i,c)} \tag{4.73}$$

Il est alors nécessaire d'identifier les états correspondant à la même cible pour chacun des filtres PHD considérés. Nous verrons dans la partie suivante comment cela peut être réalisé simplement selon une étape d'association inter-classes.

Dans la suite, nous présentons les équations récursives du filtre MM-PHD pour chaque classe et son implémentation par mélange de gaussiennes pour un modèle d'état non linéaire. Notre contribution consiste en l'application de ce filtre au contexte JTC.

4.4.3 Equations du filtre MM-PHD et implémentation par mélange de gaussiennes (MM-GMPHD)

4.4.4 Equations théoriques

Comme introduit dans le chapitre 1, le filtre PHD est fondé sur deux étapes : prédiction et mise à jour. Pour la classe c, notons v_k^c et $v_{k|k-1}^c$ les intensités de la densité multi-cibles *a posteriori* p_k^c et de la densité multi-cibles prédite $p_{k|k-1}^c$ respectivement. Les hypothèses suivantes sont formulées :

- toutes les cibles évoluent et génèrent des mesures indépendamment les unes des autres,
- le processus de naissance de cibles est caractérisé par un EFA de type Poisson,
- le processus de fausses alarmes est modélisé par un EFA de type Poisson dont la densité d'intensité est $\kappa_k(\cdot) = \lambda_c \times \mathcal{U}(\cdot)$, avec λ_c le nombre moyen de fausses alarmes et $\mathcal{U}(\cdot)$ une densité de probabilité supposée uniforme sur l'espace des mesures.
- la densité prédite multi-cibles $p_{k|k-1}^c$ et la densité *a posteriori* $p_{k|k}^c$ sont du type Poisson,
- les probabilités de détection et de survie sont considérées comme invariantes et indépendantes du temps, soit $p_{S,k}(\boldsymbol{x}) = P_S$ et $p_{D,k}(\boldsymbol{x}) = P_D$.

Définissions à présent les notations employées dans la suite de cette partie. Soient :

- $\mu_k^{c,q}$, le poids définissant la probabilité *a posteriori* du $q^{\text{ème}}$ modèle de la classe c à l'instant k.
- $\mu_{k|k-1}^{c,qr}$, la probabilité de mélange que dans la classe c, le système suive le modèle r à l'instant k sachant qu'il suit le modèle q à l'instant précédent.
- $\Pi^c = \{\Pi_{qr}^c\}$, la matrice de probabilités de transition pour la classe c.
- $f_k(\boldsymbol{x}|\boldsymbol{\xi}, c, m_k^c = M_r^c)$, la fonction d'évolution de l'état $\boldsymbol{\xi}$ vers l'état $\boldsymbol{x}$ à l'instant k.
- $g_k(\boldsymbol{y}|\boldsymbol{x})$, la fonction de vraisemblance de la mesure $\boldsymbol{y}$.

On peut montrer que l'intensité *a posteriori* peut être propagée au cours du temps selon les équations récursives présentées ci-après [Wei12b]. Considérons $\boldsymbol{\xi} \in \mathbb{X}$ comme l'état précédent, les intensités sont alors mélangées de la manière suivante :

$$v_{k-1|k}^c(\boldsymbol{\xi}, m_k^c = M_r^c) = \sum_{q=1}^{r(c)} \mu_{k-1|k}^{c,qr} v_{k-1}^c(\boldsymbol{\xi}, m_{k-1}^c = M_q^c) \tag{4.74}$$

où les probabilités de mélange $\mu_{k-1|k}^{c,qr}$ à l'instant k s'expriment comme suit :

$$\mu_{k-1|k}^{c,qr} = \frac{\Pi_{qr}^c \mu_{k-1}^{c,q}}{\sum_{q'=1}^{r(c)} \Pi_{q'r}^c \mu_{k-1}^{c,q'}} \tag{4.75}$$

avec $\mu_{k-1}^{c,q}$ la probabilité que le système évolue selon le $q^{\text{ème}}$ modèle de la $c^{\text{ème}}$ classe à l'instant $k-1$. Puis, pour tout modèle $m_k^c \in \{M_1^c, ..., M_{r(c)}^c\}$ et considérant les cibles naissantes, l'intensité prédite est :

$$v_{k|k-1}^c(\boldsymbol{x}, m_k^c) = \gamma_k^c(\boldsymbol{x}) + \int P_S f_k(\boldsymbol{x}|\boldsymbol{\xi}, m_k^c) v_{k-1|k}^c(\boldsymbol{\xi}, m_k^c) d\boldsymbol{\xi} \qquad \boldsymbol{x}, \boldsymbol{\xi} \in \mathbb{X} \tag{4.76}$$

avec $\gamma_k^c(.)$ la fonction d'intensité de l'EFA des cibles naissantes pour la $c^{\text{ème}}$ classe et $f_k(\cdot|\boldsymbol{\xi})$ la fonction de transition.

Finalement, l'intensité jointe *a posteriori* à l'instant k est mise à jour comme suit :

$$v_k^c(\boldsymbol{x}, m_k^c) = (1 - P_D) v_{k|k-1}^c(\boldsymbol{x}, m_k^c) + \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}_k} \frac{P_D g_k(\boldsymbol{y}|\boldsymbol{x}, m_k^c) v_{k|k-1}^c(\boldsymbol{x}, m_k^c)}{\kappa_k(\boldsymbol{y}) + \int P_D g_k(\boldsymbol{y}|\boldsymbol{x}) v_{k|k-1}^c(\boldsymbol{x}, m_k^c) d\boldsymbol{x}} \tag{4.77}$$

où $\kappa_k(.)$ est la fonction d'intensité de l'EFA du clutter et $g_k(\boldsymbol{y}|\boldsymbol{x})$ est la vraisemblance que la mesure $\boldsymbol{y}$ soit générée par une cible dont l'état est $\boldsymbol{x}$.
Ainsi, l'intensité marginale mise à jour en considérant toutes les classes possibles est :

$$v_k^c(\boldsymbol{x}) = \sum_{r=1}^{r(c)} v_k^c(\boldsymbol{x}, m_k^c = r) \tag{4.78}$$

Notons que pour ne pas alourdir les notations, l'intensité jointe *a posteriori* définie en (4.77) et sa marginale définie en (4.78) sont notées de façon identique mais le nombre d'arguments suffit à les différencier.

Nous introduisons par la suite une implémentation par mélange de gaussiennes de l'algorithme MM-PHD, notée MM-GMPHD.

4.4.4.1 Implémentation du filtre MM-GMPHD multi-classes

Dans la suite, nous présentons les équations caractérisant la mise en œuvre du filtre MM-PHD consistant à approcher à chaque instant les fonctions d'intensité de chaque classe par des mélanges de gaussiennes.
Considérons les hypothèses suivantes :

- L'intensité de naissance s'exprime à partir d'un mélange de gaussiennes comme suit :

$$\gamma_k^c(\boldsymbol{x}) = \sum_{i=1}^{J_k^{\gamma,c}} w_k^{\gamma,c,i} \mathcal{N}(\boldsymbol{x}; \alpha_k^{\gamma,c,i}, P_k^{\gamma,c,i}) \tag{4.79}$$

  avec $J_k^{\gamma,c}$, $w_k^{\gamma,c,i}$, $\alpha_k^{\gamma,c,i}$ et $P_k^{\gamma,c,i}$ des paramètres du modèle permettant de déterminer la forme de la fonction d'intensité. Notons que contrairement aux mélanges de gaussiennes classiques, $\sum_{i=1}^{J_k^{\gamma,c}} w_k^{\gamma,c,i} \neq 1$ mais correspond à l'espérance du nombre de cibles naissantes.
- La modèle d'évolution de l'état est linéaire et gaussien. De plus, le modèle de mesure considéré est non-linéaire et indépendant de la classe. Dans ce cas, pour le modèle $m_k^c = M_r^c$ de la classe c, il s'ensuit :

$$\begin{aligned} f_k(\boldsymbol{x}|\boldsymbol{\xi}, c, m_k^c = M_r^c) &= \mathcal{N}(\boldsymbol{x}; \boldsymbol{F}^{c,r}\boldsymbol{\xi}, \boldsymbol{Q}^{c,r}) \\ g_k(\boldsymbol{y}|\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{y}; \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{R}) \end{aligned} \tag{4.80}$$

  où $g(\cdot)$ désigne la vraisemblance, $\boldsymbol{F}^{c,r}$ la matrice de transition définie en (4.10) et $\boldsymbol{h}(\cdot)$ la fonction de mesure définie en (4.12). De plus, $\boldsymbol{Q}^{c,r}$ et $\boldsymbol{R}$ sont les matrices de covariance des bruits de modèle et de mesure définies en (4.10) et (4.13) respectivement.

Dans la suite, dans la mesure où les hypothèses précédentes sont supposées vérifiées, la fonction

d'intensité *a posteriori* à l'instant $k-1$ peut s'écrire comme suit :

$$\hat{v}_{k-1}^{c}(\boldsymbol{\xi}) = \sum_{i=1}^{J_{k-1}^{c}} \sum_{r=1}^{r(c)} w_{k-1}^{(i),c,r} \mu_{k-1}^{(i),c,r} \mathcal{N}(\boldsymbol{\xi}; \alpha_{k-1}^{(i),c,r}, P_{k-1}^{(i),c,r}) \tag{4.81}$$

où J_{k-1}^{c} est le nombre de composantes gaussiennes à l'instant $k-1$. Alors, les principales étapes du filtre MM-GMPHD multi-classes proposé sont les suivantes, où l'exposant (i) est introduit pour faire référence à chaque composante gaussienne :

1/ Tout d'abord, supposons que l'intensité prédite à l'instant k a pour expression :

$$\hat{v}_{k|k-1}^{c}(\boldsymbol{x}) = v_{s,k|k-1}^{c}(\boldsymbol{x}) + \gamma_{k}^{c}(\boldsymbol{x}) \tag{4.82}$$

où $\gamma_{k}^{c}(\boldsymbol{x})$ est l'intensité de naissance décrite en (4.79) et l'intensité prédite associée aux cibles survivantes, notée $v_{s,k|k-1}^{c}(\boldsymbol{x})$, est définie comme suit :

$$v_{s,k|k-1}^{c}(\boldsymbol{x}) = \sum_{i=1}^{J_{k-1}^{c}} \sum_{r=1}^{r(c)} w_{k|k-1}^{(i),c,r} \mu_{k|k-1}^{(i),c,r} \mathcal{N}(\boldsymbol{x}; \alpha_{k|k-1}^{(i),c,r}, P_{k|k-1}^{(i),c,r}) \tag{4.83}$$

avec :

$$\begin{aligned}
w_{k|k-1}^{(i),c,r} &= P_S w_{k-1}^{(i),c,r} && (4.84)\\
\mu_{k|k-1}^{(i),c,r} &= \sum_{q=1}^{r(c)} \Pi_{qr}^{c} \mu_{k-1}^{(i),c,q} && (4.85)\\
\alpha_{k|k-1}^{(i),c,r} &= \boldsymbol{F}^{c,r} \alpha_{m,k-1}^{(i),c,r} && (4.86)\\
P_{k|k-1}^{(i),c,r} &= \boldsymbol{Q}^{c,r} + \boldsymbol{F}^{c,r} P_{m,k-1}^{(i),c,r} (\boldsymbol{F}^{c,r})^{T} && (4.87)
\end{aligned}$$

où comme introduit dans la partie dédiée à l'IMM au chapitre 1, $\alpha_{m,k-1}^{(i),c,r}$ et $P_{m,k-1}^{(i),c,r}$ sont calculés à partir d'une étape de mélange préliminaire et satisfont :

$$\begin{aligned}
\alpha_{m,k-1}^{(i),c,r} &\propto \sum_{q=1}^{r(c)} \Pi_{qr}^{c} \mu_{k-1}^{(i),c,r} \alpha_{k-1}^{(i),c,r} \\
P_{m,k-1}^{(i),c,r} &\propto \sum_{q=1}^{r(c)} \Pi_{qr}^{c} \left(P_{k-1}^{(i),c,r} + (\alpha_{k-1}^{(i),c,r} - \alpha_{m,k-1}^{(i),c,r})(\alpha_{k-1}^{(i),c,r} - \alpha_{m,k-1}^{(i),c,r})^{T} \right)
\end{aligned} \tag{4.88}$$

Remarquons que compte tenu de (4.79) et (4.83), le nombre de gaussiennes *a priori* à l'instant k est :

$$J_{k|k-1}^{c} = r(c) \times J_{k-1}^{c} + J_{k}^{\gamma,c} \tag{4.89}$$

De plus, on peut alors montrer que le nombre de cibles prédit associé à la fonction d'intensité $v_{k|k-1}^{c}$ s'exprime alors comme suit [Vo06] :

$$\hat{N}_{k|k-1}^{c} = \hat{N}_{k-1}^{c} P_S + \sum_{i=1}^{J_{k}^{\gamma,c}} w_{k}^{\gamma,c,i} \tag{4.90}$$

2/ L'équation (4.82) est supposée s'écrire de manière unifiée comme suit en représentant les cibles survivantes et naissantes avec le même formalisme.

$$v^c_{k|k-1}(\boldsymbol{x}) = \sum_{i=1}^{J^c_{k|k-1}} \sum_{r=1}^{r(c)} w^{(i),c,r}_{k|k-1} \mu^{(i),c,r}_{k|k-1} \mathcal{N}(\boldsymbol{x}; \alpha^{(i),c,r}_{k|k-1}, P^{(i),c,r}_{k|k-1}) \tag{4.91}$$

En prenant en compte les mesures courantes, chaque gaussienne peut être mise à jour selon les équations d'un filtre de Kalman étendu. Dans un premier temps, la mesure prédite, la matrice de covariance de l'innovation, le gain de Kalman et la matrice de covariance de l'erreur d'estimation *a posteriori* sont calculés comme suit :

$$\eta^{(i),c,r}_{k|k-1} = \boldsymbol{h}(\alpha^{(i),c,r}_{k|k-1}) \tag{4.92}$$

$$S^{(i),c,r}_k = \boldsymbol{R} + H^{(i),c,r}_k P^{(i),c,r}_{k|k-1} (H^{(i),c,r}_k)^T \tag{4.93}$$

$$K^{(i),c,r}_k = P^{(i),c,r}_{k|k-1} (H^{(i),c,r}_k)^T (S^{(i),c,r}_k)^{-1} \tag{4.94}$$

$$P^{(i),c,r}_k = (I - K^{(i),c,r}_k H^{(i),c,r}_k) P^{(i),c,r}_{k|k-1} \tag{4.95}$$

avec $\eta^{(i),c,r}_{k|k-1}$ et $S^{(i),c,r}_k$ la mesure prédite et la matrice de covariance de l'innovation. De plus, $H^{(i),c,r}_k$ est la matrice jacobienne, définie en (4.29), de la fonction de mesure $\boldsymbol{h}(\cdot)$ évaluée en $\alpha^{(i),c,r}_{k|k-1}$.

3/ Ensuite, la fonction d'intensité *a posteriori* est calculée en considérant que chaque gaussienne peut être mise à jour soit à partir de chacune des mesures reçues à l'instant courant en se fondant sur les équations (4.92)-(4.95) soit en formulant une hypothèse de non détection de la cible. Dans ce cas, la mise à jour se décompose de la manière suivante :

$$\begin{aligned} \hat{v}^c_k(\boldsymbol{x}) = & \sum_{i=1}^{J^c_{k|k-1}} \sum_{r=1}^{r(c)} w^{(i,0),c,r}_k \mu^{(i,0),c,r}_k \mathcal{N}(\boldsymbol{x}; \alpha^{(i,0),c,r}_k, P^{(i,0),c,r}_k) \\ & + \sum_{l=1}^{M(k)} \sum_{i=1}^{J^c_{k|k-1}} \sum_{r=1}^{r(c)} w^{(i,l),c,r}_k \mu^{(i,l),c,r}_k \mathcal{N}(\boldsymbol{x}; \alpha^{(i,l),c,r}_k, P^{(i,l),c,r}_k) \end{aligned} \tag{4.96}$$

où les termes correspondant à l'hypothèse de non détection sont indiqués par l'exposant (0) et calculés ainsi :

$$w^{(i,0),c,r}_k = (1 - P_D) w^{(i),c,r}_{k|k-1} \tag{4.97}$$

$$\mu^{(i,0),c,r}_k = \mu^{(i),c,r}_{k|k-1} \tag{4.98}$$

$$\alpha^{(i,0),c,r}_k = \alpha^{(i),c,r}_{k|k-1} \tag{4.99}$$

$$P^{(i,0),c,r}_k = P^{(i),c,r}_{k|k-1} \tag{4.100}$$

et les termes correspondant à une mise à jour selon l'une des mesures reçues, ici la $l^{\text{ème}}$, sont

définis comme suit :

$$\bar{w}_k^{(i,l),c,r} = P_D w_{k|k-1}^{(i),c,r} \mathcal{L}_k^{(i,l),c,r} \tag{4.101}$$

$$\mu_k^{(i,l),c,r} = \mu_{k|k-1}^{(i),c,r} \tag{4.102}$$

$$\alpha_k^{(i,l),c,r} = \alpha_{k|k-1}^{(i),c,r} + K_k^{(i),c,r}(\boldsymbol{y}_k^{(l)} - \eta_{k|k-1}^{(i),c,r}) \tag{4.103}$$

$$P_k^{(i,l),c,r} = P_k^{(i),c,r} \tag{4.104}$$

avec $\mathcal{L}_k^{(i,l),c,r} = \mathcal{N}(\boldsymbol{y}_k^{(l)}; \eta_{k|k-1}^{(i),c,r}, S_k^{(i),c,r})$ la vraisemblance de la mesure $\boldsymbol{y}_k^{(l)}$ pour le filtre associé au modèle r de la classe c.
Puis, pour chaque mesure l, les poids des composantes gaussiennes sont ensuite normalisés de la manière suivante :

$$w_k^{(i,l),c,r} = \frac{\bar{w}_k^{(i,l),c,r}}{\kappa_k(\boldsymbol{y}_k^{(l)}) + \sum_{i=0}^{J_{k|k-1}^c} \bar{w}_k^{(i,l),c,r}} \tag{4.105}$$

On peut remarquer que (4.96) a la même forme que (4.81) avec $J_k^c = J_{k|k-1}^c \times (M(k)+1)$.
Notons également que le nombre de cibles estimé *a posteriori* associé à la fonction d'intensité v_k^c s'exprime comme suit :

$$\hat{N}_k^c = \hat{N}_{k|k-1}^c(1-P_D) + \sum_{l=1}^{M(k)} \sum_{i=1}^{J_{k|k-1}^c} \sum_{r=1}^{r(c)} w_k^{(i,l),c,r} \tag{4.106}$$

Avec les hypothèses formulées, on a bien demontré que si l'intensité *a posteriori* à l'instant $k-1$ est un mélange de gaussiennes alors l'intensité *a posteriori* à l'instant k conserve la même expression. Les équations de mise à jour des paramètres du mélange ont été développées.

4/ A ce stade, les composantes gaussiennes de chaque classe qui ont un poids *a posteriori* suffisamment important sont extraites. Cette extraction est effectuée en mélangeant dans un premier temps pour chaque point de l'espace d'état les composantes gaussiennes des différents modèles de mouvement considérés telle que :

$$v_k^c(\boldsymbol{x}) \simeq \sum_{l=0}^{M(k)} \sum_{i=1}^{J_{k|k-1}^c} w_k^{(i,l),c} \mathcal{N}(\boldsymbol{x}; \alpha_k^{(i,l),c}, P_k^{(i,l),c}) \tag{4.107}$$

où $w_k^{(i,l),c} = \sum_{r=1}^{r(c)} w_k^{(i,l),c,r}$.
De façon plus générique, on peut réécrire (4.107) comme suit :

$$v_k^c(\boldsymbol{x}) \simeq \sum_{j=1}^{J_k^c} w_k^{(j),c} \mathcal{N}(\boldsymbol{x}; \alpha_k^{(j),c}, P_k^{(j),c}) \tag{4.108}$$

où chaque valeur de l'indice j représente un couple unique (i,l) pour $i = 1, ..., J_{k|k-1}^c$ et

$l = 0, ..., \ M(k)$. Dans ce cas, les paramètres $w_k^{(j),c}$, $\alpha_k^{(j),c}$ et $P_k^{(j),c}$ vérifient :

$$\begin{aligned} w_k^{(j),c} &= w_k^{(i,l),c} = \sum_{r=1}^{r(c)} w_k^{(i,l),c,r} \mu_k^{(i,l),c,r} \\ \alpha_k^{(j),c} &= \alpha_k^{(i,l),c} = \sum_{r=1}^{r(c)} \alpha_k^{(i,l),c,r} \mu_k^{(i,l),c,r} \\ P_k^{(j),c} &= P_k^{(i,l),c} = \sum_{r=1}^{r(c)} \mu_k^{(i,l),c,r} [P_k^{(i,l),c,r} + (\alpha_k^{(i,l),c} - \alpha_k^{(i,l),c,r})(\alpha_k^{(i,l),c} - \alpha_k^{(i,l),c,r})^T] \end{aligned} \tag{4.109}$$

De plus, par souci de simplicité par la suite, notons également la vraisemblance comme suit :

$$\mathcal{L}_k^{(j),c,r} = \mathcal{L}_k^{(i,l),c,r} \tag{4.110}$$

On constate alors que le nombre de composantes gaussiennes peut augmenter de façon importante pour chaque filtre MM-GMPHD. En effet, à chaque instant le mélange de gaussiennes nécessite $(J_{k-1}^c + J_k^{\gamma,c})(1 + M(k)) = O(J_{k-1}^c M(k))$ composantes pour représenter v_k^c.
Ainsi, afin d'éviter que le nombre de gaussiennes n'augmente sans limite, une étape de *pruning*, qui réduit le nombre de gaussiennes à propager à chaque instant, est effectuée pour chaque classe de cible. Il s'agit de ne propager que les gaussiennes dont le poids $w_k^{(j),c}$ est supérieur à une valeur prédéfinie notée $TH_{pruning}$.
De plus, certaines composantes gaussiennes sont tellement proches les unes des autres, qu'elle peuvent être approchées par une seule gaussienne en les fusionnant. On parle d'étape de *merging*.
Ces deux étapes sont effectuées comme présenté dans l'**Algorithme 4.3**.

Algorithme 4.3 Etapes de *pruning* et de *merging*

Etape de *pruning*
Former l'ensemble I de composantes gaussiennes tel que :

$$I = \{i = 1, ..., J_k^c | w_k^{(i),c} > TH_{pruning}\}$$

où $w_k^{(i),c}$ sont les poids définis en (4.109).

Etape de *merging*
Soit $l = 0$.
Tant que I est non vide, répéter :
- $l = l + 1$
- $j = \underset{i \in I}{argmax} \ \ w_k^{(i),c}$
- $L = \{i \in I | (\alpha_k^{(i),c} - \alpha_k^{(j),c})^T P_k^{(i),c} (\alpha_k^{(i),c} - \alpha_k^{(j),c}) \leq TH_{merging}\}$
- $\tilde{w}_k^{(l),c} = \sum_{i \in L} w_k^{(i),c}$
- $\tilde{\alpha}_k^{(l),c} = \frac{1}{\tilde{w}_k^{(l),c}} \sum_{i \in L} w_k^{(i),c} \alpha_k^{(i),c}$
- $\tilde{P}_k^{(l),c} = \frac{1}{\tilde{w}_k^{(l),c}} \sum_{i \in L} w_k^{(i),c} (P_k^{(i),c} + (\tilde{\alpha}_k^{(l),c} - \alpha_k^{(i),c})(\tilde{\alpha}_k^{(l),c} - \alpha_k^{(i),c})^T)$
- $I = I \backslash L$

Sélectionner les gaussiennes caractérisées par $\{\tilde{w}_k^{(l),c}, \tilde{\alpha}_k^{(l),c}, \tilde{P}_k^{(l),c}\}$

A l'issue de cette étape, il s'agit d'extraire l'état multi-cibles associé à chaque classe. Pour cela, on pourrait choisir d'extraire les $\hat{N}_k$ maxima locaux de la fonction d'intensité. Néanmoins, la hauteur de chaque pic dépend du poids et de la covariance de la gaussienne associée. Cette démarche peut alors conduire à l'extraction de gaussiennes de poids faibles. Une alternative est d'extraire les moyennes de chaque composante gaussienne dont les poids *a posteriori* sont plus grands qu'un seuil prédéterminé, habituellement égal à 0.5 [Vo06]. Ainsi, l'état multi-cibles pour chaque classe vérifie :

$$\begin{aligned}\hat{\boldsymbol{\mathcal{X}}}_k^c &= \{\boldsymbol{X}_k^{(1),c}, ..., \boldsymbol{X}_k^{(\mathcal{N}_k^c),c}\} \\ &= \{\alpha_k^{(j),c} | w_k^{(j),c} > 0.5\}_{j=1}^{J_k^c - \mathcal{N}_{Pru,k}^c - \mathcal{N}_{Mer,k}^c}\end{aligned} \tag{4.111}$$

avec $\mathcal{N}_k^c$ le nombre de composantes gaussiennes extraites au sein de la classe c après l'étape de *pruning* et de *merging*. De plus, $\mathcal{N}_{Pru,k}^c$ et $\mathcal{N}_{Mer,k}^c$ désignent respectivement le nombre de gaussiennes ayant été supprimées ou fusionnées.

5/ Finalement, les composantes gaussiennes extraites des différents filtres MM-GMPHD associés aux différentes classes sont regroupés. Plus précisément, les composantes qui sont proches spatialement c'est-à-dire qui sont susceptibles de représenter la même cible, sont fusionnées selon une étape de *merging* interclasses. Ainsi, chaque état monocible $\hat{\boldsymbol{X}}_k^{(j)}$ de $\hat{\boldsymbol{\mathcal{X}}}_k$ s'obtient comme suit :

$$\begin{aligned}\hat{\boldsymbol{X}}_k^{(j)} &\simeq \sum_{c=1}^{C_p(j)} \hat{\boldsymbol{X}}_k^{(j),c} Pr(c, j | \boldsymbol{\mathcal{Y}}_{1:k-1}) \\ &\sum_{c=1}^{C_p(j)} \hat{\boldsymbol{X}}_k^{(j),c} \frac{Pr(c, j | \boldsymbol{\mathcal{Y}}_{1:k-1}) p(\boldsymbol{\mathcal{Y}}_k | c, j, \boldsymbol{\mathcal{Y}}_{1:k-1})}{\sum_{c'=1}^{C} p(\boldsymbol{\mathcal{Y}}_{1:k} | c, j, \boldsymbol{\mathcal{Y}}_{1:k-1})} \\ &\simeq \sum_{c=1}^{C_p(j)} \hat{\boldsymbol{X}}_k^{(j),c} \frac{Pr(c, j | \boldsymbol{\mathcal{Y}}_{1:k-1}) \sum_{r'=1}^{r(c)} \mu_{k|k-1}^{(j),c,r'} \mathcal{L}_k^{(j),c,r'}}{\sum_{c'=1}^{C_p(j)} \sum_{r'=1}^{r(c)} \mu_{k|k-1}^{(j),c',r'} \mathcal{L}_k^{(j),c',r'}}\end{aligned} \tag{4.112}$$

où $C_p(j)$ représente le nombre de classes où l'état $\hat{\boldsymbol{X}}_k^{(j),c}$ est présent. En effet, en pratique, il se peut que le nombre de gaussiennes extraites ne soit pas le même pour toutes les classes. Autrement dit, chaque estimation $\hat{\boldsymbol{X}}_k^{(j)}$ calculée selon (4.112) est en réalité obtenue à partir d'au plus C termes $\hat{\boldsymbol{X}}_k^{(j),c}$.

Dans la sous-partie suivante, nous appliquons l'algorithme MM-GMPHD au pistage de plusieurs cibles dans un contexte de surveillance maritime.

4.4.5 Application au pistage de cibles marines

4.4.5.1 Protocole de simulation

Pour analyser la pertinence de l'approche proposée, quatre cibles sont considérées : un cargo (cible 1), un tanker (cible 2), un bâteau de plaisance (cible 3) et un jetski (cible 4). Les cibles apparaissent sur la zone de surveillance à $t = 1s$, $t = 50s$, $t = 20s$ et $t = 80s$ respectivement.

A chaque instant, les trajectoires des cibles 1 et 2 sont générées à partir d'un modèle de mouvement de type MRU dont l'écart-type sur l'accélération σ_{MRU} vaut 0.2 $m.s^{-3/2}$. Quant

aux cibles 3 et 4, leurs trajectoires sont pilotées par une chaîne de Markov d'ordre dont la matrice de transition de probabilité est $\begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$. Le premier état est caractérisé par un modèle de mouvement de type MRU dont l'écart-type sur l'accélération est $\sigma_{MRU} = 0.2\ m.s^{-3/2}$ alors que le second état est un modèle de Singer avec $\sigma_{Sin} = 5m.s^{-2}$ et $\tau_{Sin} = 15\ s$. La matrice de covariance $\boldsymbol{R} = diag(\sigma_d^2, \sigma_\theta^2, \sigma_{ed}^2)$ est une matrice diagonale composée des variances des bruits sur la distance d_k, l'angle θ_k et l'étalement distance $\mathcal{E}$ tels que : $\sigma_d = 10\ m$, $\sigma_\theta = 0.003\ rad$ et $\sigma_{ed} = 5\ m$. De plus, la période d'échantillonnage vaut $T = 1s$.

Enfin, la détection de la cible est modélisée selon une probabilité égale à 0.98 et le processus de clutter est un EFA de type Poisson dont la fonction d'intensité est $\kappa(\boldsymbol{y}) = \lambda_c S\mathcal{U}(\boldsymbol{y})$, où $\mathcal{U}(\cdot)$ désigne la densité uniforme sur la zone de surveillance de surface S et $\lambda_c = 10^{-6}m^{-2}$ représente la densité spatiale de clutter.

La figure 4.11 illustre une réalisation des trajectoires des quatre cibles dans le clutter.

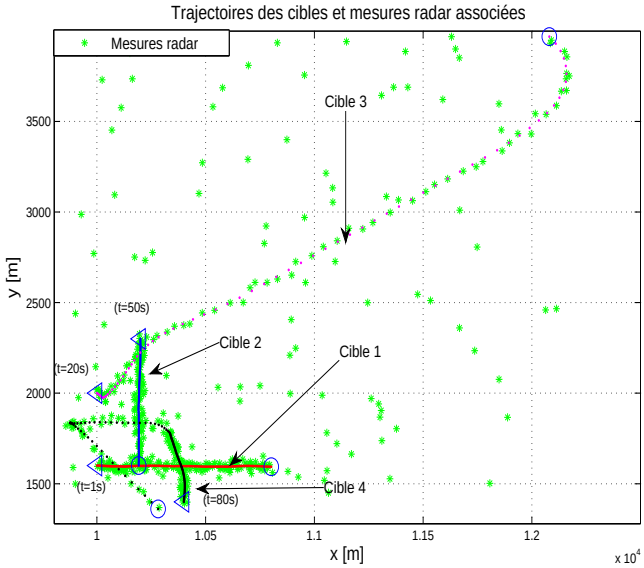

FIGURE 4.11 – Trajectoires de cibles. Les points de départ sont désignés par des triangles alors que les points d'arrivée sont marqués par des cercles.

Définissons à présent le paramétrage de l'algorithme MM-GMPHD multi-classes proposé. Dans ce cas également, $C = 2$ classes de cibles sont considérées. Pour la classe des cibles non-manœuvrantes, on considère $r(1) = 1$ modèle de mouvement noté M_1^1 qui est un MRU dont l'écart-type sur le bruit de modèle vaut $\sigma_{MRU,1} = 0.2\ m.s^{-3/2}$. Pour le pistage des cibles manœuvrantes, $r(2) = 2$ modèles sont combinés : M_1^2 représente un MRU dont l'écart-type sur l'accélération est réglé à $\sigma_{MRU,2} = 0.2\ m.s^{-3/2}$ et M_2^2 désigne un modèle de Singer de paramètres $\sigma_{Sin,2} = 5\ m.s^{-2}$ et $\tau_{Sin,2} = 15s$. Le changement d'un modèle à l'autre est piloté par (4.1). De plus, la TPM associée est $\Pi^{c=2} = \begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$. A noter que c'est la même matrice que celle utilisée pour générer les données. Enfin, la fonction d'intensité du processus de Poisson

pour les cibles naissantes $\gamma_k(\cdot)$ introduit en (4.79) est défini par : $J_k^{\gamma,c} = 4$ et

$$\begin{aligned}
\alpha_k^{\gamma,c,1} &= [10000, 0, 0, 1600, 0, 0, l^c] \\
\alpha_k^{\gamma,c,2} &= [10000, 0, 0, 2000, 0, 0, l^c] \\
\alpha_k^{\gamma,c,3} &= [10200, 0, 0, 2300, 0, 0, l^c] \\
\alpha_k^{\gamma,c,4} &= [10400, 0, 0, 1400, 0, 0, l^c]
\end{aligned} \tag{4.113}$$

avec $l^{c=1} = 100m$ et $l^{c=2} = 15m$.
De plus, $P_k^{\gamma,c,i} = diag([400, 100, 0, 400, 100, 0, 400])$, $i = 1, 2, 3, 4$.

Dans la suite, l'algorithme multi-classes MM-GMPHD est comparé avec un filtre GMPHD fondé sur un modèle MRU avec $\sigma_{MRU} = 0.2m.s^{-3/2}$. Il est également comparé à un filtre MM-GMPHD fondé sur un modèle MRU et un modèle de Singer avec $\sigma_{MRU} = 0.2m.s^{-3/2}$, $\sigma_{Sin} = 5m.s^{-2}$, $\tau_{Sin} = 15s$ et $\Pi = \begin{bmatrix} 0.98 & 0.02 \\ 0.1 & 0.9 \end{bmatrix}$. Ces deux filtres incluent la longueur de cible dans le vecteur d'état mais qu'aucune classification n'est mise en œuvre.

4.4.5.2 Résultats de simulation

La figure 4.12 représente les trajectoires de cible estimées à partir de l'algorithme multi-classes MM-GMPHD étudié, pour la réalisation représentée en figure 4.11.

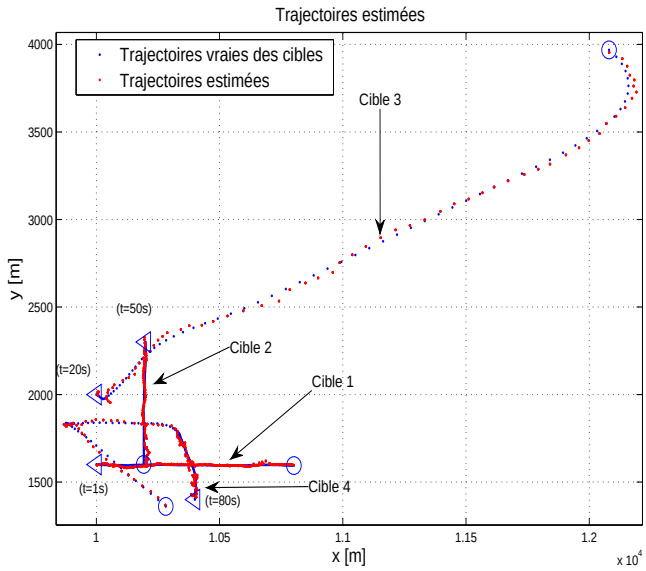

FIGURE 4.12 – Trajectoires de cibles et états associés

D'après la figure 4.12 mais également des figures 4.13 et 4.14 qui représentent respectivement l'estimation du nombre de cibles et la distance OSPA pour les différents filtres considérés, la formulation JTC permet au pistage d'être plus précis.

Les filtres MM-GMPHD multi-classes et MM-GMPHD sont logiquement plus performants que le filtre GMPHD puisque ce dernier ne permet pas de pister correctement les cibles ma-

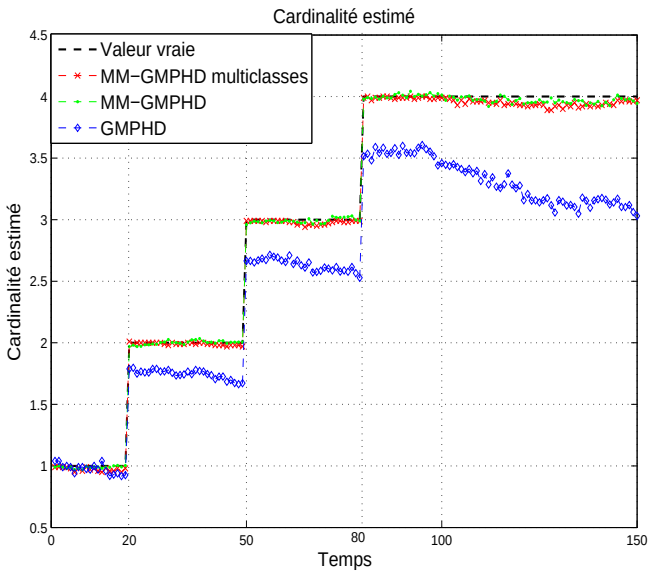

FIGURE 4.13 – Estimation de la cardinalité

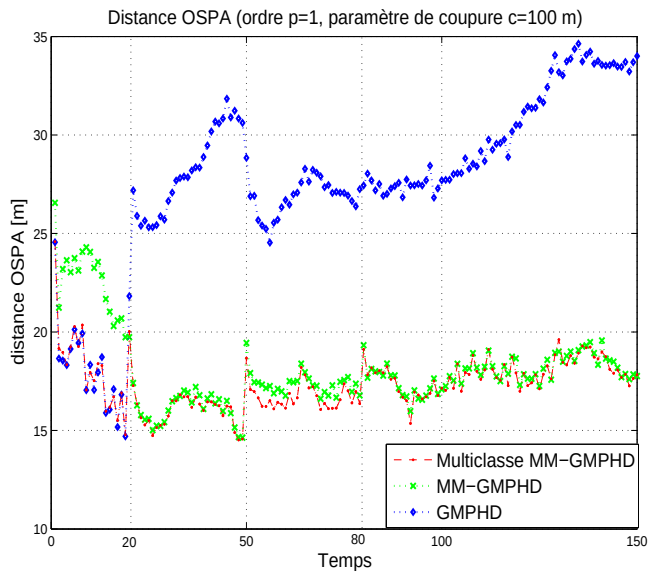

FIGURE 4.14 – Distances OSPA

nœuvrantes. Cela peut notamment être mis en évidence à partir de $k = 20$, c'est-à-dire lorsque la première cible manœuvrante apparaît. De plus, bien que le nombre estimé de cibles soit sensiblement le même pour les filtres MM-GMPHD multi-classes et MM-GMPHD, la distance OSPA est moins élevée pour le filtre MM-GMPHD multiclasses. En effet, comme observé dans la partie précédente dans le cas monocible, l'estimation de la longueur des cibles permettant d'identifier rapidement les cibles 1 et 2 comme étant non-manœuvrantes, le poids associé à la classe non-manœuvrante pour ces cibles est prépondérant et donc l'algorithme GMPHD associé à cette classe contribue à estimer plus précisément les trajectoires. C'est ce que l'on constate sur la figure 4.14 entre $k = 1$ et $k = 20$ où à $k = 50$ lors de l'apparition des cibles 1 et 2. Dans ce cas, le filtre multi-classes MM-GMPHD a des performances très similaires au filtre GMPHD.

Les résultats précédents illustrent la capacité de notre algorithme à pister plusieurs cibles de longueurs et dynamiques différentes à partir d'une approche multi-classes multi-modèles.
En général, quelle que soit l'approche de pistage considérée, l'estimation est de plus en plus difficile lorsque :

- la probabilité de détection diminue,
- la densité de fausses alarmes augmente,
- les cibles évoluent très proches les unes des autres.

Le filtre PHD a pour intérêt de prendre en compte la probabilité de détection dans la phase de filtrage. Ainsi, si celle-ci est convenablement paramétrée, le filtre PHD et ses variantes peuvent extraire les états d'une cible même si cette dernière n'est pas détectée à chaque instant. Pour cette raison, il est généralement recommandé pour des applications où la probabilité de détection n'est pas nécessairement élevée. La densité de fausses alarmes est également prise en compte dans la phase de filtrage. Néanmoins, en présence d'un trop grand nombre de fausses alarmes, le filtre PHD n'estime généralement pas précisément le nombre de cibles présentes.

Dans la suite, nous proposons d'étudier le comportement de notre approche dans le cas où les cibles évoluent proches les unes des autres. Pour cela, nous considérons le scénario décrit par la figure 4.15 où les quatre cibles considérées précédemment apparaissent dès l'instant initial et se déplacent toutes en MRU l'une vers l'autre. Elles se croisent alors à l'instant $t = 50s$.

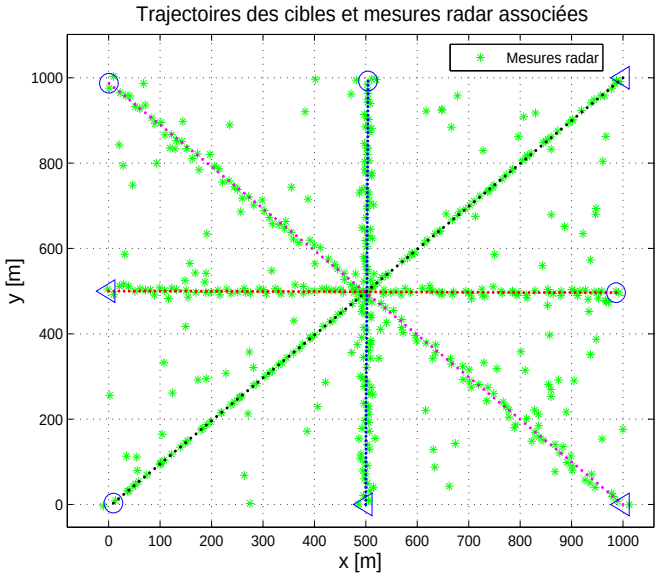

FIGURE 4.15 – Trajectoires de quatre cibles en croisement. Les points de départ sont désignés par des triangles alors que les points d'arrivée sont marqués par des cercles.

Nous représentons sur la figure 4.16 les états estimés pour une réalisation.

Dans le cas présenté par la figure 4.16, il s'avère que les positions des quatre cibles sont correctement estimées malgré leur croisement. Néanmoins, en rejouant plusieurs fois le scénario, on s'aperçoit qu'une voire deux des quatre cibles considérées peuvent parfois ne plus être suivies après l'instant de croisement. C'est ce qui est illustré en figure 4.17.

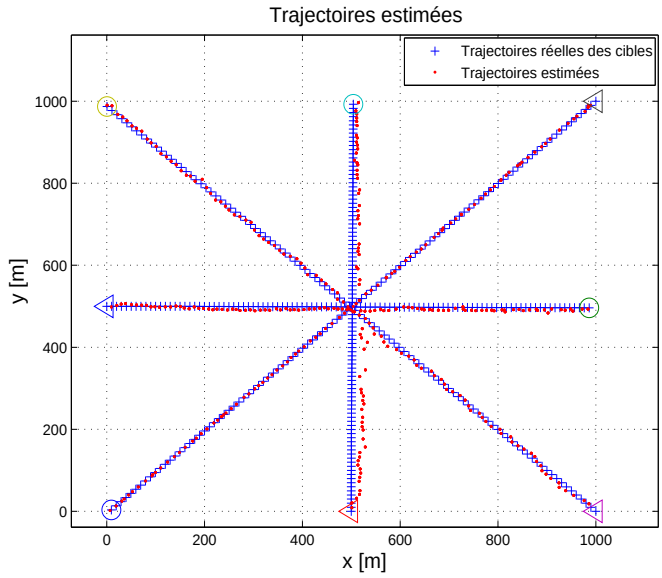

FIGURE 4.16 – Trajectoires de quatre cibles en croisement et états associés. Résultat d'une première réalisation.

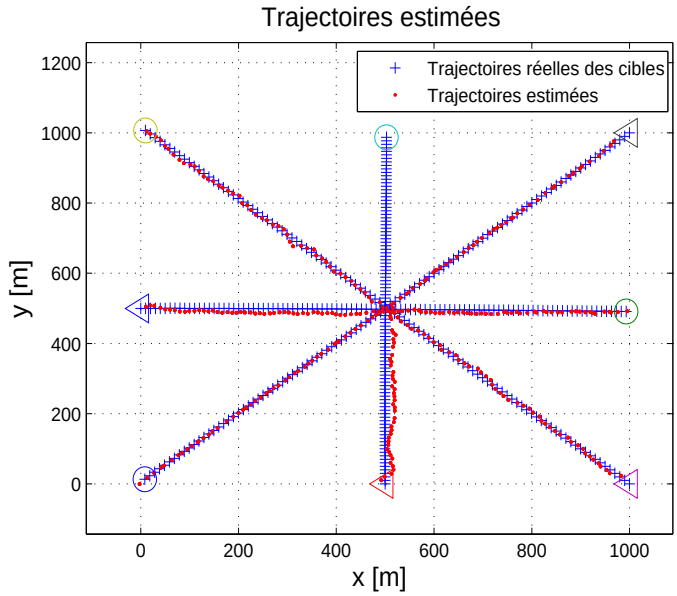

FIGURE 4.17 – Trajectoires de quatre cibles en croisement et états associés. Résultat d'une seconde réalisation.

Ce phénomène, qui est une limite de notre algorithme, est la conséquence des étapes de *merging* intra et interclasses mises en jeu dans les étapes 4/ et 5/ du filtre MM-GMPHD multi-classes proposé. En effet, ces étapes de *merging* ont pour effet de fusionner des gaussiennes associées à des cibles proches à la fois en distance, vitesse et accélération mais également en longueur. Plus particulièrement, concernant le scénario étudié et la mise en œuvre du filtre PHD par mélange de gaussiennes, celles associées aux cibles 1 et 2 ainsi que celles associées aux cibles 3 et 4 ont tendance à être fusionnées.

Cependant, cette configuration reste un cas difficile à traiter pour lequel la plupart des algorithmes de pistage sont mis en difficulté. Dans un cas général où deux cibles de classes différentes se croisent, l'intérêt des algorithmes JTC est qu'ils contribuent à les discriminer plus facilement. Par exemple, considérons le croisement de la cible 1 et de la cible 3 appartenant aux classes 1 et 2 respectivement. La figure 4.18 illustre le résultat de la simulation avec l'algorithme MM-GMPHD multi-classes proposé.

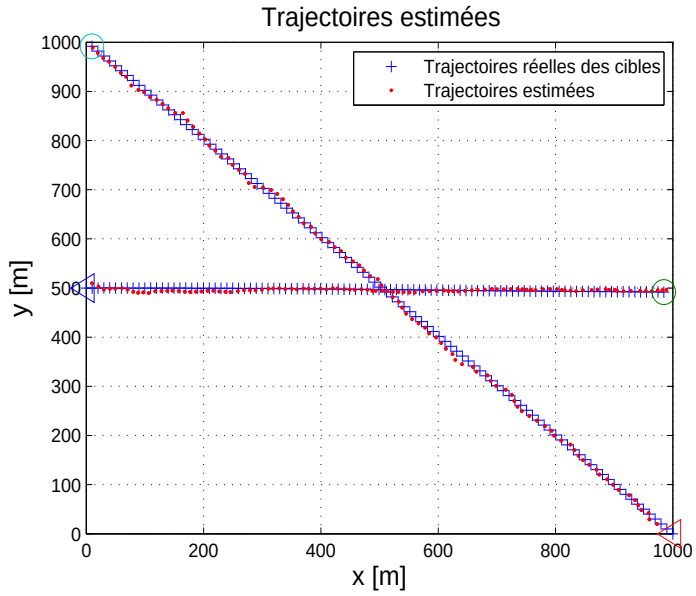

FIGURE 4.18 – Trajectoires de deux cibles en croisement et états associés

Dans la suite, nous comparons alors les résultats obtenus avec un algorithme GMPHD n'intégrant pas la longueur dans le vecteur d'état et représentons sur les figures 4.19 et 4.20 le nombre estimé de cibles et la distance OSPA pour les deux approches. Les résultats sont moyennés sur 100 simulations de Monte Carlo.

Comme dans le cas monocible, on peut voir à partir de la figure 4.19 que la formulation JTC permet de mieux discriminer les cibles entre elles comparé à un filtre GMPHD standard qui n'intègre pas la mesure de longueur à l'état des cibles. De plus, cette erreur sur l'estimation du cardinal a une influence directe sur la distance OSPA.

Finalement, il apparaît que bien que notre approche soit mise en difficulté dans la première configuration étudiée, elle se comporte convenablement lorsque les deux cibles peuvent être discriminées selon leurs mesures de longueur.

De manière plus générale, plusieurs traitements ou modifications pourraient être apportées à notre approche selon son utilisation.
Dans un premier temps, si l'utilisateur souhaite traiter du pistage de cibles proches, pour améliorer les performances de notre approche, d'autres attributs caractérisant les cibles comme la mesure de SER ou la fréquence Doppler pourraient être exploités en les intégrant à l'état des cibles.

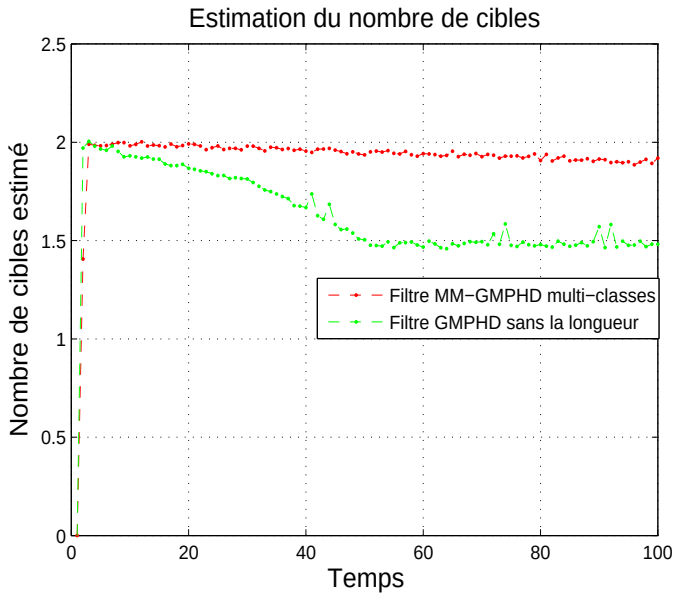

FIGURE 4.19 – Estimation du nombre de cibles pour un scénario de croisement de deux cibles

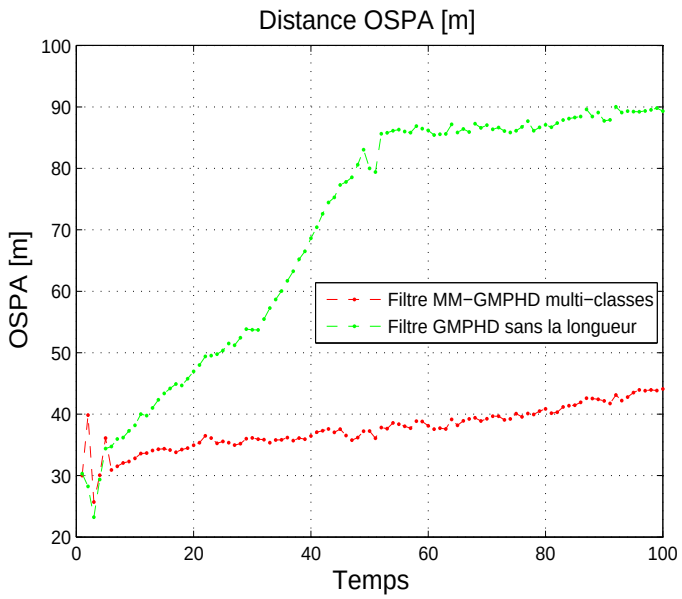

FIGURE 4.20 – Estimation de la distance OSPA pour un scénario de croisement de deux cibles

Dans un deuxième temps, nous faisons l'hypothèse dans nos simulations que le processus de naissance est spatialement localisé. En pratique, cela peut être le cas dans des configurations très spécifiques comme le pistage de cibles autour d'aéroports ou de zones portuaires. Néanmoins, en général, la zone d'apparition de nouvelles cibles n'est pas nécessairement connue. Par exemple, en pistage de cibles marines, un sous-marin peut sortir son périscope n'importe où et n'importe quand. Ainsi, un processus de naissance étalé sur toute la zone de surveillance serait plus approprié. Cependant, si la densité de fausses alarmes est importante, le nombre de cibles estimé risque de ne pas être représentatif de la réalité. Une solution consisterait à combiner une approche de filtrage multi-cibles avec par exemple un algorithme *track before detect* [PUN05, PAP13] permettant d'éliminer en partie les fausses alarmes.

Enfin, l'inconvénient majeur du filtre PHD est qu'il ne permet pas d'identifier quels états sont associés à quelles cibles, autrement dit il n'est pas possible d'extraire à chaque instant des pistes plutôt que seulement des états. Pour pallier cela, il est néanmoins possible de considérer des approches alternatives associant les états extraits aux différents instants [Cla05, Pan09]. En particulier, Panta *et al.* discutent dans [Pan09] des différents problèmes à gérer, notamment : la création, la propagation et la suppression de pistes.

4.5 Conclusions et perspectives

Dans ce chapitre, nous traitons de l'utilisation des mesures d'étalement distance des cibles pour définir de nouvelles approches JTC.

Dans un premier temps, nous proposons trois approches dans le cas monocible. Les deux premières approches diffèrent selon la manière de réaliser l'inférence au sein de chaque classe. Dans le premier cas, le vecteur d'état est estimé à partir d'un IMM fondé sur des FKEs alors que dans le second cas il est estimé à partir d'un filtre particulaire à modèles multiples. Dans les simulations étudiées, les deux approches ont des performances d'estimation similaires. Néanmoins, la pertinence de l'approche particulaire vis-à-vis de celle fondée sur des FKEs est surtout liée à la non-linéarité du problème étudié et s'avère plus adaptée pour des cibles à forte dynamique. La troisième approche combine le principe JTC avec les modèles BNP introduits au chapitre précédent. Ainsi, il s'agit d'estimer au sein de chaque classe et pour chaque famille de modèles considérée le paramètre de précision associé. En termes d'estimation, cette dernière approche est légèrement plus performante que les deux premières mais l'utilisation de la loi d'importance optimale la rend calculatoirement plus coûteuse. Nous montrons également que pour les trois algorithmes proposés, le choix de la loi *a priori* sur la longueur de cible joue un rôle important vis-à-vis des performances d'estimation. En particulier, pour des cibles de taille intermédiaire, la classification est plus délicate et le gain sur l'estimation de la trajectoire est moindre que pour des cibles clairement identifiées comme appartenant à l'une des deux classes. On pourrait dans ce cas relâcher les hypothèses pour rendre la loi *a priori* sur la longueur de cible moins discriminante et considérer d'autres attributs de cible pour aider un peu plus la classification.

Dans un second temps, nous développons une extension au cas multi-cibles fondée sur des filtres MM-PHD et mise en œuvre par mélange de gaussiennes. Au travers des différentes simulations réalisées, nous mettons en évidence l'intérêt de cet l'algorithme résultant, appelé MM-GMPHD multi-classes, lorsque les cibles pistées ont des caractéristiques dynamiques différentes. Il s'avère en effet plus performant que les algorithmes GMPHD et MM-GMPHD existants. Nous relevons également les limites de notre approche. En particulier, dans un contexte de cibles proches, l'implémentation par mélange de gaussiennes impliquant une étape de *merging*, le nombre de cibles pistées peut ne pas toujours être correctement estimé. De plus, l'utilisation de filtres PHD ne permet pas d'identifier des trajectoires de cibles mais seulement d'extraire un ensemble d'états à chaque instant.

Concernant ces travaux, nous envisageons les perspectives suivantes.
Premièrement, comme dans le cas monocible, selon la non-linéarité de la fonction de mesure par rapport à l'état, une implémentation par filtrage particulaire de l'algorithme MM-PHD pourrait être plus performante. Nous pourrions donc développer une mise en œuvre SMC puis la comparer à celle proposée dans ce manuscrit.
Deuxièmement, des paramètres supplémentaires pourraient être intégrés à l'état afin d'enrichir le modèle et d'améliorer l'estimation. Par exemple, on pourrait supposer que le rapport $\frac{a}{b}$ du modèle elliptique défini en (4.6) est dépendant de la classe de cibles et estimé conjointement aux paramètres cinématiques et à la longueur de cible.
Troisièmement, dans la continuité de nos travaux sur le multi-cibles et compte tenu de l'incapacité du filtre PHD à identifier des pistes, nous pourrions envisager deux approches. La première exploiterait les travaux de Panta *et al.* [Pan06a] pour labéliser les gaussiennes dans notre approche et ainsi être en mesure d'identifier des trajectoires de cibles. La seconde consisterait à tirer avantage de la théorie des EFAs labélisés pour étendre notre approche aux filtres de Bernoulli [Reu13, Reu15].
Enfin, afin de valider la pertinence de nos approches, il serait pertinent de les appliquer à des données réelles issues d'enregistrements en vol et mener une étude comparative avec les approches existantes actuellement exploitées.

Conclusions générales et perspectives

Cette thèse porte sur l'amélioration des techniques de pistage de cibles de surface dans un contexte monocible ou multi-cibles, notamment dans les situations critiques où elles manœuvrent et sont potentiellement proches entre elles.

Dans le premier chapitre de ce manuscrit, nous avons décrit les principes de base de l'estimation bayésienne pour le pistage monocible. Nous traitons plus précisément de la représentation dans l'espace d'état puis des filtres bayésiens tels que le filtre de Bayes, le filtre de Kalman et le filtre particulaire qui visent à estimer les paramètres cinématiques de la cible. Nous présentons également les algorithmes à modèles multiples, utilisés pour la poursuite de cibles manœuvrantes. En particulier, nous mettons l'accent sur l'algorithme IMM, sa mise en œuvre et les différentes problématiques associées.
Dans un second temps, nous abordons les principes du pistage multi-cibles. Les filtres PHD et CPHD, qui sont les plus connus et fondés sur la théorie des EFAs, sont détaillés. Enfin, les différentes mesures de performance utilisées en pistage multi-cibles sont introduites.

Plusieurs problématiques se dégagent à l'issue de ce chapitre.
Premièrement, pour un algorithme d'estimation à modèles multiples combinant plusieurs modèles d'état, il est préférable que ces derniers soient les plus dissimilaires possible afin de représenter une large diversité des mouvements de la cible. Or, le choix des modèles se fait selon la connaissance *a priori* du système et l'expérience de l'utilisateur. Il n'existe pas de règles permettant de définir les types de mouvement à associer et leurs paramètres.
Deuxièmement, lorsqu'un type de mouvement est utilisé, ses paramètres sont généralement fixés *a priori* mais les valeurs considérées ne sont pas nécessairement adaptées à toutes les phases du mouvement. L'une des difficultés majeures réside alors dans la manière de définir et de faire évoluer la matrice de covariance du bruit de modèle.
Enfin, on sait que le choix de la famille de modèles est important au regard des performances de l'algorithme d'estimation utilisé. Or, dans un contexte de pistage multi-cibles, les paramètres du modèle de mouvement diffèrent selon le type de cibles. Le prendre en compte permet de mieux les discriminer notamment dans les situations où elles sont proches les unes des autres.

L'étude menée dans cette thèse se décline alors en plusieurs points.

Dans un premier temps, nous nous sommes intéréssés à la comparaison de modèles d'état afin de définir des modèles d'évolution pertinents à associer au sein d'une structure à modèles multiples. Notre contribution a porté sur la définition d'une mesure de dissimilarité fondée sur la divergence de Jeffrey entre les lois jointes des vecteurs d'état successifs associés à deux modèles d'état différents. Celle-ci est appliquée à la comparaison de modèles de mouvement. Elle est ensuite utilisée pour classifier un ensemble de plusieurs modèles d'état en sous-groupes afin de proposer une méthode non supervisée de sélection *a priori* des modèles constituant des algorithmes à modèles multiples. Nous montrons par simulation l'intérêt de cet outil statistique pour aider l'utilisateur à la conception d'algorithmes à modèles multiples.
Ces travaux ont été présentés dans [Mag13, Mag14, Mag15a, Mag15d].
Nous envisageons, comme perspective à cette étude, d'analyser comme alternative à la JD certaines distances définies dans le contexte de la géométrie de l'information [Ama00]. De plus, plutôt que de considérer des décompositions en valeurs propres successives de la matrice de dissimilarité puis d'analyser leurs évolutions au cours du temps, il pourrait être intéressant de voir l'ensemble de ces JD entre modèles au cours du temps comme un tenseur et d'étudier si des approches fondées sur la décomposition en valeurs singulières d'ordre supérieur pourraient être adaptées pour déduire le nombre de sous-familles de modèles. Différents travaux ont été initiés en ce sens.

Dans un deuxième temps, notre contribution a porté sur l'utilisation de modèles BNP pour estimer les statistiques du bruit de modèle. L'utilisation de processus de Dirichlet étant avantageuse pour représenter des bruits multimodaux sans avoir à spécifier *a priori* le nombre de modes et leurs caractéristiques, nous présentons un algorithme dont l'originalité réside à deux niveaux : d'une part, nous proposons un modèle hiérarchique tirant profit des formes structurelles des matrices de covariance associées aux modèles de mouvement pour n'estimer qu'un nombre réduit d'hyperparamètres. D'autre part, l'estimation conjointe des paramètres cinématiques de la cible et de la matrice de précision du bruit de modèle étant réalisée à partir d'un filtre particulaire, nous développons le calcul de la distribution d'importance optimale.
Il est à noter que que nous avons également développé un second algorithme qui s'inscrit dans la continuité des travaux de Caron [Car08] où toute la matrice de covariance du bruit de modèle est supposée inconnue. Notre contribution, présentée en annexe pour ne pas alourdir le chapitre 3, porte sur l'expression de la distribution d'importance.
Nous avons montré à partir de simulations de Monte Carlo que les approches BNP présentent des performances d'estimation supérieures aux algorithmes paramétriques standards de pistage de cibles fondés sur du filtrage de Kalman ou particulaire par exemple. De plus, comparativement à l'approche inspirée des travaux de Caron qui estime toute la matrice de covariance du bruit de modèle, notre approche est légèrement plus performante et est calculatoirement moins coûteuse. Au travers de différentes déclinaisons de ces deux algorithmes, nous montrons également la pertinence du calcul de la distribution d'importance optimale dans chacun des cas.
Ces travaux ont été présentés dans [Mag15b, Mag16a].

Dans la continuité de cette étude, afin de s'affranchir du réglage des paramètres des processus de Dirichlet, nous envisageons de les estimer en ligne comme proposé dans [Rab12, Per13]. De plus, dans le cadre de cette thèse, la distribution des paramètres inconnus est supposée multimodale et stationnaire. On peut envisager des approches considérant une distribution dont les modes évoluent au cours du temps. Il s'agit d'une question délicate et des premiers modèles bayésiens non-paramétriques ont déjà été étudiés en ce sens [Car07, Zhu05]. Nous pourrions également proposer une extension fondée sur les processus de Dirichlet hiérarchiques consistant à introduire des dépendances entre les différents modes et à estimer la matrice de transition des probabilités associée. Enfin, comme l'utilisation de processus de Dirichlet favorise certains modes qui ont tendance à réapparaître très fréquemment, une solution alternative serait de considérer les processus de Pitman-Yor [Pit97]. Ces derniers sont une généralisation des processus de Dirichlet qui selon le réglage de leurs paramètres admettent des lois de puissance à queue plus lourde.

Enfin, dans la troisième partie de la thèse, nous tirons profit des méthodes JTC pour mener simultanément la classification de la cible et l'inférence de ses paramètres. Dans ce cas, à chaque classe de cible est associé un ensemble de modèles d'évolution qui lui est propre. Afin de réaliser la classification, nous utilisons les mesures d'étalement distance conjointement aux mesures de position de la cible. Nous appliquons ces approches dans un contexte de pistage monocible puis dans un contexte multi-cibles.

Dans le premier cas, deux algorithmes bayésiens multi-classes multi-modèles sont tout d'abord présentés : un algorithme multi-classes IMM combinant des FKEs et un second fondé sur un filtre particulaire à modèles multiples. Dans les simulations étudiées, les deux approches ont des performances d'estimation similaires. Toutefois, l'approche particulaire est plus adaptée dans le cas de cibles présentant de fortes dynamiques qui entraînent une non-linéarité du problème d'estimation plus importante. Par ailleurs, un algorithme BNP est également détaillé dans ce contexte monocible. Il est légèrement plus performant que les deux premiers mais l'utilisation de la loi d'importance optimale le rend calculatoirement plus coûteux.

Dans le second cas, un algorithme multi-classes MM-PHD implémenté par mélange de gaussiennes est présenté. Les différentes simulations menées contribuent à mettre en évidence l'intérêt cette approche notamment lorsque plusieurs cibles, caractérisées par des dynamiques variées, sont pistées.

Ces travaux ont été présentés dans [Mag15c, Mag16b].

Les perspectives sont multiples. Tout d'abord, nous pourrions améliorer la modélisation elliptique de la forme de chaque cible en estimant pour chacune d'entre elles le rapport entre le petit axe et le grand axe de l'ellipse. De plus, compte tenu des limites du filtre PHD qui ne permet pas d'extraire des pistes, mais seulement des états, nous prévoyons de poursuivre dans un contexte multi-cibles sur l'utilisation d'EFAs labélisés et en particulier de mettre en œuvre nos approches en utilisant les filtres de type multi-Bernoulli. Nous pourrions également envisager de faire évoluer nos approches en intégrant d'autres informations que celle apportée par la mesure d'étalement distance d'une cible. Ainsi, nous pourrions être en mesure de définir d'autres classes de cibles que celles proposées dans ce manuscrit. Par exemple, on pourrait considérer les classes de cibles suivantes : cibles manœuvrantes, cibles non-manœuvrantes, cibles naissantes, cibles furtives, cibles au comportement suspect, détections correspondant à des fausses alarmes, etc.

Finalement, de manière plus générale, nous envisageons de tester les différentes approches développées dans le cadre de cette thèse sur des données réelles récemment acquises lors d'enregistrements en vol à partir d'un radar de surveillance afin d'évaluer leurs valeurs ajoutées sur des scénarios opérationnels.

Annexes

Annexe A

Pistage de cibles et algorithmes d'association de données

Lorsque l'on considère le pistage d'une ou plusieurs cibles dans un environnement où la probabilité de détection n'est pas égale à un et que d'autres mesures induites par le clutter sont détectées, les filtres bayésiens tels que les filtres de Kalman ou particulaire ne peuvent pas s'appliquer directement. Il existe une incertitude quant à la mesure à utiliser pour mettre à jour l'état de la cible. Il est alors nécessaire d'identifier la mesure issue de la cible et de l'associer à la piste considérée avant de filtrer son état selon l'une des approches bayésiennes introduites dans le chapitre 1. Ce traitement constitue ce que l'on appelle l'étape d'association des mesures aux pistes ou plus communément "association de données".

Parmi les algorithmes d'association de données dans un contexte monocible, l'algorithme *nearest neighbour* (NN) et l'algorithme *probabilistic data association* (PDA) sont principalement utilisés. Dans un contexte multi-cibles, d'autres approches telles que l'algorithme *global nearest neighbour* (GNN), l'algorithme *joint probabilistic data association* (JPDA) et l'algorithme *multiple hypothesis tracking* (MHT) ont été introduites.
Dans cette annexe, il s'agit de détailler ces différentes approches et de rappeler les avantages et les inconvénients de chacun des algorithmes énoncés précédemment.

A.1 Position du problème

L'association de données consiste à affecter à chaque cible poursuivie une ou plusieurs des mesures radar reçues à l'instant courant, afin de mettre à jour l'état du système. A chaque balayage d'antenne du radar, aussi appelé *scan*, deux types de mesures sont potentiellement reçues : des mesures dues aux cibles et des fausses alarmes essentiellement générées par le clutter de terre ou de mer environnant. Cette étape d'association est essentielle par rapport aux performances d'estimation des algorithmes utilisés puisqu'une mauvaise association ou une non-association peut conduire à une erreur d'estimation sur le positionnement de la cible.

Généralement, avant de procéder à l'étape d'association des mesures aux pistes, un traitement préliminaire appelé fenêtrage statistique ou *gating* est nécessaire afin de sélectionner parmi l'ensemble des mesures reçues à l'instant courant les mesures les plus probables autour de la

prédiction de la piste [Bar95]. Ces mesures sont dites candidates à l'association.
Si l'estimation est réalisée par filtrage de Kalman, l'ensemble des mesures appartenant à la fenêtre statistique, également appelée fenêtre de validation, vérifient :

$$d_{Mah} = (\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1})^T S_k^{-1}(\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1}) \leq \gamma \tag{A.1}$$

où d_{Mah} représente la distance de Mahalanobis entre la mesure $\boldsymbol{y}_k$ et la mesure prédite $H_k\hat{\boldsymbol{x}}_{k|k-1}$. De plus, $(\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1})$ désigne l'innovation, S_k est la matrice de covariance de l'innovation associée à la piste considérée et γ définit un seuil dont la valeur dépend de la probabilité de *gating* P_G que la fenêtre contienne bien la mesure associée à la cible. Comme l'innovation est gaussienne, pour une valeur fixe et pré-établie de P_G, on déduit la valeur du seuil associé en utilisant la fonction inverse du χ^2.

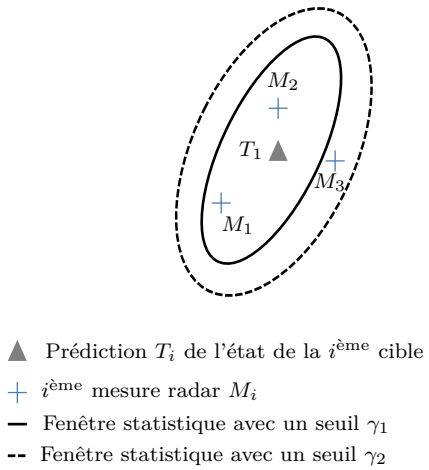

Figure A.1 – Fenêtrage statistique

Comme illustré par la figure A.1, le choix de γ peut avoir une influence sur le résultat de l'association. En effet, pour deux seuils différents notés γ_1 et γ_2 et respectivement liés aux probabilités de *gating* P_{G_1} et P_{G_2} telles que $P_{G_2} > P_{G_1}$, le résultat de l'étape de fenêtrage statistique n'est pas identique. Avec le seuil γ_1, seules deux mesures sont retenues pour l'étape d'association des mesures aux pistes, alors qu'avec le seuil γ_2, trois sont sélectionnées.

Un algorithme d'association doit être en mesure de déterminer les mesures à utiliser pour la mise à jour de l'état des cibles et celles qui doivent être considérées comme des fausses alarmes ou comme de nouvelles cibles et ainsi initialiser de nouvelles pistes. La figure A.2 illustre la complexité du problème d'association des mesures aux pistes dans un cas multi-cibles.
A ce jour, il existe différents types d'algorithmes permettant de prendre une décision quant à la « meilleure » association possible. Certains d'entre eux sont des algorithmes dits *monoscan* tels que le NN ou GNN pour lesquels la prise de décision est effectuée à l'instant courant, alors que d'autres sont des algorithmes *multiscans* comme le MHT pour lesquels la prise de décision n'est pas effectuée à l'instant où l'on reçoit la mesure, mais est différée de quelques instants en attendant que plusieurs mesures consécutives soient disponibles pour résoudre le problème d'association de données.
Parmi les algorithmes *monoscan*, il existe également des algorithmes probabilistes pour lesquels

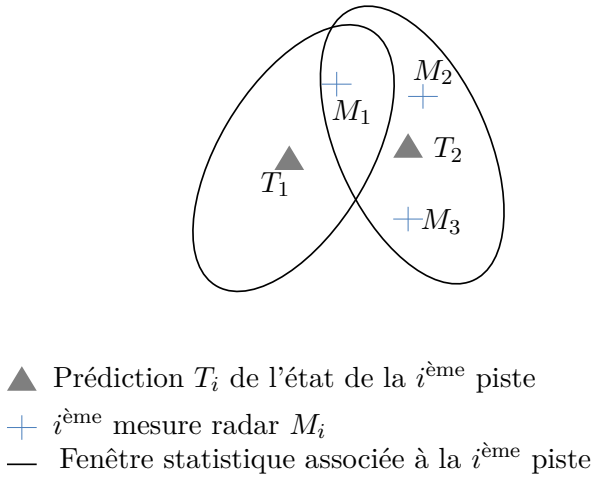

FIGURE A.2 – Association de données dans un contexte de cibles proches

la mise à jour de l'état de la cible se fait selon une somme pondérée des innovations calculées à partir de toutes les mesures dans la fenêtre statistique de la piste. C'est notamment le cas des algorithmes PDA et JPDA qui s'appliquent dans un contexte monocible et multi-cibles respectivement.
Dans la suite, nous présentons ces différents algorithmes d'association.

A.2 Algorithmes *nearest neighbour* (NN) et *global nearest neighbour* (GNN)

L'algorithme NN est une méthode d'association point à point, c'est-à-dire qu'une piste n'est mise à jour qu'à partir d'une seule mesure, et une mesure ne peut être associée qu'à une seule piste. Le principe de cet algorithme est d'associer à chaque piste la mesure la plus proche de la prédiction. Afin de tenir compte de l'imprécision sur les mesures et sur les estimations respectivement liées aux bruits de mesure et de modèle, la distance de Mahalanobis est généralement préférée à la distance euclidienne et est calculée selon (A.1), entre la prédiction de la piste associée et chaque mesure présente à l'intérieur de la fenêtre de validation.

Le principe de l'algorithme NN est de construire un vecteur de coût C dont les éléments c_{ij} correspondent à la distance de Mahalanobis entre la $i^{\text{ème}}$ mesure et la $j^{\text{ème}}$ piste. Il s'agit ensuite de sélectionner la mesure dont la distance de Mahalanobis est minimale, puis de l'associer à la piste.

Intéressons-nous à présent à l'algorithme GNN, qui est la généralisation de l'algorithme NN au cas multi-cibles.

L'algorithme GNN est également un algorithme d'association point à point. Cet algorithme résout le problème d'association de données de façon optimale au sens où il maximise le nombre d'associations et minimise la somme totale des distances correspondantes.

Le principe de l'algorithme est de sélectionner un élément de la matrice coût C par ligne et par colonne de façon à ce que la somme des éléments soit la plus petite possible. Cette dé-

marche revient à trouver la combinaison d'associations mesures-pistes qui minimise la distance totale et qui maximise le nombre d'associations. La résolution d'un tel problème peut se faire à partir d'algorithmes d'optimisation combinatoire comme l'algorithme de Munkres [Mun57] ou les algorithmes Jonker-Volgenant-Castanon (JVC) et Auction [Ber91]. L'algorithme le plus couramment utilisé est l'algorithme de Munkres qui correpond à la généralisation de l'algorithme Hongrois pour des matrices rectangulaires et permet de résoudre le problème d'affectation en temps polynomial plutôt qu'en temps exponentiel où l'ensemble des combinaisons doit être testé [Bla99].

Dans un contexte où le taux de fausses alarmes est important, il est probable que la situation où plusieurs mesures se trouvent dans la fenêtre de validation se produise fréquemment. Cela peut alors conduire à une mauvaise association et donc une estimation erronée des trajectoires des cibles et des paramètres associés. Afin de pallier ce problème, d'autres approches fondées sur la mise à jour du vecteur d'état à partir de toutes les mesures présentes dans la fenêtre de validation ont été développées. C'est le cas de l'algorithme PDA applicable au contexte monocible et de l'algorithme JPDA qui constitue sa généralisation au cas multi-cibles.

A.3 Algorithmes *probabilistic data association* (PDA) et *joint probabilistic data association* (JPDA)

L'algorithme PDA consiste à calculer pour chaque mesure présente dans la fenêtre de validation la probabilité *a posteriori* qu'elle corresponde à la cible. Pour cela, dans chaque cas on suppose qu'au plus une mesure est issue de la cible alors que toutes les autres mesures sont des fausses alarmes.

Contrairement aux approches précédentes, celle-ci n'est pas une approche point à point, puisqu'il ne s'agit pas d'associer à une piste une seule mesure, mais de considérer que toutes les mesures sélectionnées après l'étape de fenêtrage statistique peuvent être associées.
La mise à jour de l'état du système est donc réalisée à partir des innovations de toutes ces mesures, pondérées par les probabilités *a posteriori* que chacune d'elles soit la vraie mesure issue de la cible. Ainsi, si $\boldsymbol{Y}_k = \{\boldsymbol{y}_k^{(1)}, ..., \boldsymbol{y}_k^{(M(k))}\}$ représente l'ensemble des $M(k)$ mesures reçues à l'instant k, il vient :

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \sum_{j=0}^{M(k)} \beta^{(j)} p(\boldsymbol{x}_k|\boldsymbol{y}_k^{(j)}) \tag{A.2}$$

où $j = 0$ représente l'hypothèse de non détection de la piste et où les poids $\beta^{(j)}$ définissent la probabilité que la $j^{\text{ème}}$ mesure soit celle issue de la cible. Ils satisfont :

$$\beta^{(j)} \approx Pr[\boldsymbol{y}_k = \boldsymbol{y}_k^{(j)}|\boldsymbol{Y}_{1:k}] \tag{A.3}$$

De plus, ils vérifient :

$$\sum_{j=0}^{M(k)} \beta^{(j)} = 1 \tag{A.4}$$

Comme (A.2) représente une distribution multimodale, elle n'est pas gaussienne. Afin de

faciliter l'applicabilité du filtre de Kalman, elle est approchée par une seule gaussienne. Ainsi, l'estimation *a posteriori* de l'état est obtenue comme suit :

$$\hat{\boldsymbol{x}}_k = \sum_{j=0}^{M(k)} \beta^{(j)} \hat{\boldsymbol{x}}_k^{(j)} \tag{A.5}$$

avec :

$$\hat{\boldsymbol{x}}_k^{(0)} = \hat{\boldsymbol{x}}_{k|k-1} \tag{A.6}$$

$$\hat{\boldsymbol{x}}_k^{(j)} = \hat{\boldsymbol{x}}_{k|k-1} + K_k^{(j)}(\boldsymbol{y}_k^{(j)} - H_k \hat{\boldsymbol{x}}_{k|k-1}) \qquad j = 1, ..., M(k) \tag{A.7}$$

De plus, la matrice de covariance de l'erreur d'estimation est mise à jour de la manière suivante :

$$P_k = \sum_{j=0}^{M(k)} \beta^{(j)} \left(P_{k|k-1} - K_k^{(j)} S_k^{(j)} (K_k^{(j)})^T + (\hat{\boldsymbol{x}}_k^{(j)} - \hat{\boldsymbol{x}}_k)(\hat{\boldsymbol{x}}_k^{(j)} - \hat{\boldsymbol{x}}_k)^T \right) \tag{A.8}$$

Pour plus de détails sur ce filtre et en particulier le calcul et l'expression des poids $\{\beta_j\}_{j=1,...,M(k)}$, le lecteur peut se référer à l'article de référence [Bar09].

Notons que cet algorithme s'applique à des pistes existantes mais ne permet pas de créer de nouvelles pistes. Cependant, il existe une autre approche ne nécessitant pas cet *a priori* et permettant d'initialiser de nouvelles pistes en présence de clutter. Il s'agit de l'algorithme *integrated* PDA (IPDA) [Mus94].

Dans un contexte multi-cibles, l'algorithme PDA n'est plus adapté. En effet, une mesure peut appartenir à la fenêtre de validation d'une ou de plusieurs pistes. Dans ce cas, afin d'améliorer les performances de l'algorithme de pistage, l'algorithme JPDA est utilisé. Les probabilités d'association des mesures aux cibles sont alors calculées conjointement pour toutes les cibles partageant des mesures dans leurs fenêtres de validation respectives.

Si les cibles n'interfèrent pas entre elles, c'est-à-dire si elles n'ont pas de mesures en commun dans leurs fenêtres de validation respectives, les deux algorithmes sont équivalents. Alors que l'algorithme PDA suppose l'existence d'une cible unique dont la piste a été initialisée, l'algorithme JPDA fonctionne pour un nombre connu et fixe de cibles. La différence réside dans la manière de calculer les probabilités d'association. Alors que le PDA vise à calculer l'ensemble des probabilités d'association séparément pour toutes les cibles, en supposant que toutes les mesures qui ne sont pas dans la fenêtre de validation de la piste considérée sont des fausses alarmes, l'algorithme JPDA calcule les probabilités jointes pour l'ensemble des pistes considérées. Par conséquent, l'une des limites de cet algorithme est qu'il nécessite une connaissance *a priori* du nombre de cibles. De plus, dans un contexte où les cibles sont proches, deux cibles peuvent ne former qu'une seule et même piste. On parle dans ce cas d'un phénomène de coalescence [Bar95].

Dans la suite, nous discutons de l'algorithme MHT fondé sur la propagation d'hypothèses d'association au cours du temps.

A.4 Algorithme *multiple hypothesis tracking* (MHT)

Contrairement aux approches introduites précédemment, l'algorithme MHT est *multiscans*. Ainsi, le principe de l'algorithme MHT est de prendre une décision sur le problème d'association des mesures aux pistes de façon différée. Pour cela, différentes hypothèses d'association des mesures aux pistes sont formulées à chaque instant et propagées sur plusieurs instants en supposant que les observations futures permettent de résoudre l'incertitude sur l'origine des mesures.
L'avantage de cet algorithme est qu'il ne nécessite pas la connaissance *a priori* du nombre de cibles puisque cette méthode *multiscans* est en mesure de gérer l'apparition de nouvelles pistes sur la zone de surveillance.
Le principe de cet algorithme consiste à rechercher de façon exhaustive toutes les associations mesures-pistes possibles dans un intervalle de temps donné.
Dès qu'un nouvel ensemble de mesures est disponible, un nouvel ensemble d'hypothèses est formé sur la base des hypothèses existantes. L'hypothèse caractérisée par la vraisemblance maximale est ensuite sélectionnée.
Notons qu'afin de limiter la complexité calculatoire, d'un instant à l'autre, une étape de *pruning* peut être effectuée afin de ne retenir que les hypothèses de forte probabilité.
Cet algorithme peut être implémenté selon deux variantes :

- le MHT orienté hypothèse ou *Hypothesis-Oriented MHT* (HOMHT). Cette approche a été développée par Reid [Rei79]. C'est une méthode exhaustive énumérant toutes les hypothèses d'association possibles. Dès qu'un nouvel ensemble de mesures est disponible, il s'agit d'étendre les hypothèses existantes en considérant toutes les associations mesures-pistes contenues dans les hypothèses de l'instant courant.
  On peut alors représenter l'ensemble des hypothèses sous forme d'un arbre d'hypothèses.
- le MHT orienté piste ou *Track-Oriented MHT* (TOMHT). Cette approche a été introduite par Kurien [Kur88, Kur91], et contrairement à l'approche HOMHT, elle ne conserve pas les hypothèses d'un instant à un autre. Dès lors qu'un nouvel ensemble de mesures est reçu, un nouvel ensemble de pistes est formé en considérant toutes les associations mesures-pistes possibles. Cette méthode ne maintient donc pas un ensemble d'hypothèses d'un instant à l'autre, mais un ensemble de pistes dites non-compatibles c'est-à-dire n'ayant pas de mesures en commun dans leurs fenêtres de validation respectives.

L'algorithme *multiscans* MHT est plus performant que les algorithmes monoscans introduits précédemment. Notamment dans des contextes de cibles proches en présence de fausses alarmes, l'algorithme MHT est plus fiable. Il permet alors d'estimer plus précisément les trajectoires des cibles. Cependant, la génération puis l'évaluation de toutes les hypothèses à chaque instant rendent la complexité calculatoire de l'algorithme très importante. Il existe certaines alternatives visant à diminuer le coût calculatoire. Par exemple, une méthode permettant de générer seulement les meilleures hypothèses a été développée par Cox et *al* [Cox96]. Cette approche propose une manière efficace d'implémenter l'algorithme MHT en utilisant l'algorithme de Murty [Mur68, Mil97] afin de définir les meilleures hypothèses sans générer les hypothèses de faible probabilité.

Annexe B

Approximation de la loi optimale pour l'estimation de la matrice de précision du bruit de modèle

En pistage de cibles, plutôt que d'utiliser des approches à modèles multiples pour prendre en compte les changements de manœuvre d'une cible au cours du temps, il est possible de considérer un modèle d'évolution unique caractérisé par une fonction de transition fixée *a priori* et dont la matrice de précision du bruit de modèle est supposée pouvoir évoluer au cours du temps. Il s'agit alors d'estimer conjointement au cours du temps le vecteur d'état de la cible et les paramètres de la matrice de précision du bruit de modèle.
Afin de prendre en compte les changements de mode et l'apparition de nouveaux modes au cours du temps, une modélisation non-paramétrique peut être utilisée pour réaliser cette estimation.

Dans la suite, nous considérons que la matrice de précision notée Γ_k suit une distribution inconnue G, supposée aléatoire, qui vérifie :

$$G \sim DP(G_0, \alpha_0) \tag{B.1}$$

avec G_0 la distribution de base et α_0 le paramètre d'échelle du processus de Dirichlet.
Cette modélisation BNP permet d'estimer une matrice de précision Γ_k qui évolue au cours du temps mais dont les instanciations à différents instants peuvent être identiques. Afin de prendre en compte cet effet de *clustering*, nous introduisons une variable discrète d'assignation $c_k \in \{1, ..., J(k)\}$ où $J(k)$ représente le nombre de matrices distinctes parmi l'ensemble $\{\Gamma_1, ..., \Gamma_k\}$. Ces matrices distinctes sont notées $\{\Gamma_1^*, ..., \Gamma_k^*\}$. Ainsi, à chaque instant, la matrice de précision du bruit de modèle satisfait :

$$\Gamma_k = \Gamma_{c_k}^* \tag{B.2}$$

Nous considérons alors le modèle hiérarchique suivant qui définit les relations entre le vecteur d'état $\boldsymbol{x}_k$, la matrice de précision Γ_k et la variable discrète d'assignation $c_k \in \{1, ..., J(k)\}$:

$$\Gamma_k | \{c_k = J(k-1)+1\} \sim G_0 = \mathcal{W}(\Gamma_k; W_0, v_0) \tag{B.3}$$

$$\Gamma_k | \{c_k = i\} \sim \delta_{\Gamma_i^*}(\Gamma_k) \text{ pour } i \leq J(k-1) \tag{B.4}$$

$$\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \Gamma_k \sim \mathcal{N}(\boldsymbol{x}_k; F_k \boldsymbol{x}_{k-1}, \Gamma_k^{-1}) \tag{B.5}$$

Comme le bruit d'état est gaussien, G_0 est généralement choisie comme étant une distribution de Wishart de paramètre d'échelle W_0 et avec v_0 degrés de liberté.

Il est alors possible d'estimer conjointement c_k, Γ_k et $\boldsymbol{x}_k$ par filtrage particulaire en tirant profit des relations existantes entre ces différentes variables. Dans les travaux existants [Car07, Fox07, Dor09], la loi d'importance *a priori* est généralement utilisée. Dans cette thèse, l'une de nos contributions a été de développer le calcul de la distribution d'importance optimale. L'objectif de cette annexe est de présenter les différentes étapes de ce calcul.

B.1 Modélisation non-paramétrique du problème

Soit $\boldsymbol{X}_k = [c_k \ \Gamma_k \ \boldsymbol{x}_k]$ le vecteur d'état étendu où la dimension de $\boldsymbol{x}_k$ est notée n_x. L'objectif est de montrer que $\boldsymbol{X}_k$ peut être obtenu par simulation selon une approximation de la loi optimale comme suit :

$$\boldsymbol{X}_k \sim p(\boldsymbol{X}_k | \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_{1:k}) \tag{B.6}$$

En tirant profit des indépendances respectives entre les variables constituant le vecteur d'état et représentées en figure B.1, on montre que cela peut se faire en trois étapes :

$$c_k \sim Pr(c_k | \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \tag{B.7}$$

$$\boldsymbol{x_k} \sim p(\boldsymbol{x}_k | c_k, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \tag{B.8}$$

$$\Gamma_k \sim p(\Gamma_k | c_k, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \tag{B.9}$$

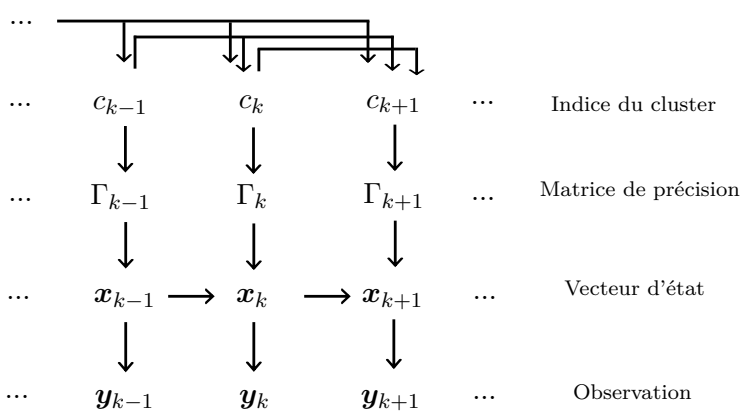

Figure B.1 – Représentation graphique de l'interdépendance des variables

Dans la suite, nous détaillons les expressions analytiques des distributions définies par (B.7)-(B.9). En particulier, deux cas sont développés : $c_k = J(k-1)+1$ et $c_k = i$ pour $i \leq J(k-1)$.

B.2 Echantillonnage de la matrice de précision du bruit d'état

Dans un premier temps, focalisons-nous sur $p(\Gamma_k|c_k, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k)$. En particulier, comme Γ_k ne dépend pas de l'observation $\boldsymbol{y}_k$ conditionnellement à c_k, $\boldsymbol{x}_k$ et $\boldsymbol{X}_{0:k-1}$, nous développons l'expression de $p(\Gamma_k|c_k, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1})$.
D'une part, si l'on traite d'un nouveau cluster, i.e. si $c_k = J(k-1)+1$, il vient :

$$\begin{aligned} p(\Gamma_k|c_k = J(k-1)+1, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1}) &\propto p(\Gamma_k|c_k = J(k-1)+1, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1}) p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k) \\ &\propto p(\Gamma_k|c_k = J(k-1)+1) p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k) \end{aligned} \tag{B.10}$$

où en utilisant (3.31), on a :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k) = (2\pi)^{-n_x/2} |\Gamma_k|^{1/2} \times exp\left(-\frac{1}{2} Tr\left[(\boldsymbol{x}_k - F_k \boldsymbol{x}_{k-1})(\boldsymbol{x}_k - F_k \boldsymbol{x}_{k-1})^T \Gamma_k\right]\right) \tag{B.11}$$

et étant donné (3.29) :

$$\begin{aligned} p(\Gamma_k|c_k = J(k-1)+1) &= \mathcal{W}(\Gamma_k; W_0, v_0) \\ &= B(W_0, v_0) |\Gamma_k|^{\frac{v_0 - n_x - 1}{2}} exp\left(-\frac{1}{2} Tr[W_0^{-1} \Gamma_k]\right) \end{aligned} \tag{B.12}$$

où la constante de normalisation de la distribution de Wishart $B(W_0, v_0)$ est définie comme suit :

$$B(W_0, v_0) = \frac{1}{2^{\frac{v_0 p}{2}} |W_0|^{\frac{v_0}{2}} \Gamma_p(\frac{v_0}{2})} \tag{B.13}$$

avec p la dimension de la matrice de précision Γ_k, i.e $p = n_x$.
En combinant (B.11) et (B.12), et tirant profit des propriétés de la fonction exponentielle et de la linéarité de la trace, on peut montrer que (B.10) devient :

$$p(\Gamma_k|c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}, \boldsymbol{x}_k) \propto \mathcal{W}(\Gamma_k; v_{post}, W_{post}) \tag{B.14}$$

avec

$$v_{post} = v_0 + 1 \tag{B.15}$$

$$W_{post} = \left(W_0^{-1} + (\boldsymbol{x}_k - F_k \boldsymbol{x}_{k-1})(\boldsymbol{x}_k - F_k \boldsymbol{x}_{k-1})^T\right)^{-1} \tag{B.16}$$

$$\tag{B.17}$$

D'autre part, si $c_k = i$ pour $1 \leq i \leq J(k-1)$, selon (B.4), on a :

$$p(\Gamma_k|c_k = i, \boldsymbol{x}_k, \boldsymbol{X}_{0:k-1}) = \delta_{\Gamma_i^*}(\Gamma_k) \tag{B.18}$$

B.3 Echantillonnage du vecteur d'état

Traitons à présent (B.8), c'est-à-dire l'expression de $p(\boldsymbol{x}_k|c_k, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k)$.

D'une part, si $c_k = J(k-1)+1$, alors on a :

$$p(\boldsymbol{x}_k|c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \propto p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) \tag{B.19}$$

où selon (3.16), $p(\boldsymbol{y}_k|\boldsymbol{x}_k) \sim \mathcal{N}(\boldsymbol{y}_k; H_k\boldsymbol{x}_k, R_k)$ et $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1)$ peut être obtenue à partir de $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k)$ en intégrant par rapport à Γ_k :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) = \int_{\Gamma_k} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_k)p(\Gamma_k|c_k = J(k-1)+1)d\Gamma_k \tag{B.20}$$

où les termes de l'intégrale sont (B.11) et (B.12) respectivement. Il vient :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) = \int_{\Gamma_k} (2\pi)^{-n_x/2} B(W_0, v_0) \frac{\mathcal{W}(\Gamma_k; v_{post}, W_{post})}{B(W_{post}, v_{post})} d\Gamma_k \tag{B.21}$$

Ensuite, en intégrant par rapport à Γ_k, (B.21) peut se réécrire comme suit :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) = (2\pi)^{-n_x/2} \frac{B(W_0, v_0)}{B(W_{post}, v_{post})} \tag{B.22}$$

A ce stade, étant donné (B.13) et (B.14), le seul terme de (B.22) qui dépend de $\boldsymbol{x}_k$ est W_{post}. Ainsi, on a la relation suivante :

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) &\propto |W^{post}|^{\frac{v_{post}}{2}} \\ &\propto |W_0^{-1} + (\boldsymbol{x}_k - F_k\boldsymbol{x}_{k-1})(\boldsymbol{x}_k - F_k\boldsymbol{x}_{k-1})^T|^{\frac{-v_{post}}{2}} \end{aligned} \tag{B.23}$$

Comme le vecteur d'état $\boldsymbol{x}_k$ est de dimension n_x, en utilisant l'annexe D, on peut identifier la distribution de student multivariée (cf. annexe D) comme suit :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) \propto T_{n_x,1}(v_0 - n_x + 1, F_k\boldsymbol{x}_{k-1}, W_0^{-1}, 1) \tag{B.24}$$

On peut alors montrer en utilisant la propriété décrite dans l'annexe D que, si v_0 est suffisamment grand, (B.24) peut être approchée par la distribution gaussienne suivante :

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) \approx \mathcal{N}(\boldsymbol{x}_k; F_k\boldsymbol{x}_{k-1}, \frac{W_0^{-1}}{v_0 - n_x + 1}) \tag{B.25}$$

Finalement, en combinant (B.25) et (3.16) puis en utilisant la propriété décrite dans l'annexe D, (B.19) devient :

$$\begin{aligned} p(\boldsymbol{x}_k|c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) &\propto p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, c_k = J(k-1)+1) \\ &= \mathcal{N}(\boldsymbol{x}_k; m_x^0, Q_x^0) \end{aligned} \tag{B.26}$$

avec :

$$Q_x^0 = \left(H_k^T R_k^{-1} H_k + W_0(v_0 - n_x + 1)\right)^{-1} \tag{B.27}$$

$$m_x^0 = Q_x^0 \left(H_k^T R_k^{-1} \boldsymbol{y}_k + (v_0 - n_x + 1) W_0 F_k \boldsymbol{x}_{k-1}\right) \tag{B.28}$$

D'autre part, si $c_k = i$ pour $1 \leq i \leq J(k-1)$:

$$p(\boldsymbol{x}_k | c_k = i, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \propto p(\boldsymbol{y}_k | \boldsymbol{x}_k) p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, c_k = i) \tag{B.29}$$

avec $p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, c_k = i) \sim \mathcal{N}(\boldsymbol{x}_k; F_k \boldsymbol{x}_{k-1}, (\Gamma_i^*)^{-1})$ et $p(\boldsymbol{y}_k | \boldsymbol{x}_k) \sim \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k)$. Une fois de plus, en combinant les deux distributions gaussiennes, il vient :

$$p(\boldsymbol{x}_k | c_k = i, \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \propto \mathcal{N}(\boldsymbol{x}_k; m_x^i, Q_x^i) \tag{B.30}$$

avec :

$$Q_x^i = (H_k^T R_k^{-1} H_k + \Gamma_i^*)^{-1} \tag{B.31}$$

$$m_x^i = Q_x^i (H_k^T R_k^{-1} \boldsymbol{y}_k + \Gamma_i^* F_k \boldsymbol{x}_{k-1}) \tag{B.32}$$

B.4 Echantillonnage de l'indice du cluster

Nous nous intéressons enfin à (B.7), c'est-à-dire à $p(c_k | \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k)$.

D'une part, si $c_k = J(k-1) + 1$, alors on a :

$$\begin{aligned} Pr(c_k = J(k-1)+1 | \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) &\propto p(\boldsymbol{y}_k | c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}) Pr[c_k = J(k-1)+1 | \boldsymbol{X}_{0:k-1}] \\ &\propto p(\boldsymbol{y}_k | c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}) \frac{\alpha_0}{\alpha_0 + k - 1} \end{aligned} \tag{B.33}$$

En prenant en compte les dépendances sur $\boldsymbol{x}_k$:

$$p(\boldsymbol{y}_k | c_k = J(k-1)+1, \boldsymbol{X}_{0:k-1}) = \int_{\boldsymbol{x}_k} p(\boldsymbol{y}_k | \boldsymbol{x}_k) p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, c_k = J(k-1)+1) d\boldsymbol{x}_k \tag{B.34}$$

où $p(\boldsymbol{y}_k | \boldsymbol{x}_k) \sim \mathcal{N}(\boldsymbol{y}_k; H_k \boldsymbol{x}_k, R_k)$ et $p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, c_k = J(k-1)+1)$ est définie par (B.25).
Finalement, en combinant ces deux distributions gaussiennes, il vient :

$$Pr(c_k = J(k-1)+1 | \boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) \propto \frac{\alpha_0}{\alpha_0 + k - 1} \mathcal{N}\left(\boldsymbol{y}_k; H_k F_k \boldsymbol{x}_{k-1}, R_k + H_k \frac{W_0^{-1}}{v_0 + n_x - 1} H_k^T\right) \tag{B.35}$$

D'autre part, si $c_k = i$ pour $1 \leq i \leq J(k-1)$, en procédant comme précédemment, il vient :

$$\begin{aligned} Pr(c_k = i|\boldsymbol{X}_{0:k-1}, \boldsymbol{y}_k) &\propto p(\boldsymbol{y}_k|c_k = i, \boldsymbol{X}_{0:k-1}) Pr[c_k = i|\boldsymbol{X}_{0:k-1}] \\ &= p(\boldsymbol{y}_k|c_k = i, \boldsymbol{X}_{0:k-1}) \frac{\#i}{\alpha_0 + k - 1} \\ &= \frac{\#i}{\alpha_0 + k - 1} \int_{\boldsymbol{x}_k} p(\boldsymbol{y}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \Gamma_i^*) d\boldsymbol{x}_k \\ &= \frac{\#i}{\alpha_0 + k - 1} \mathcal{N}(\boldsymbol{y}_k; H_k F_k \boldsymbol{x}_{k-1}, R_k + H_k (\Gamma_i^*)^{-1} H_k^T) \end{aligned} \tag{B.36}$$

où $\#i$ désigne le nombre de fois que $c_j = i$.
Les principales étapes du filtre particulaire fondé sur la distribution d'importance optimale sont détaillées dans **Algorithme B.1** où N_p désigne le nombre de particules.

Algorithme B.1 Estimation de la matrice de précision du bruit de modèle : une itération de l'algorithme (*Algo2*)

- Pour $i = 1, .., N_p$:
  - Calculer $P(i, 0) \underset{(B.35)}{=} p(c_k = J^i(k-1) + 1|\boldsymbol{X}_{0:k-1}^i, \boldsymbol{y}_k)$
  - Pour $j = 1, ..., J(k-1)$
    Calculer $P(i, j) \underset{(B.36)}{=} p(c_k = j|\boldsymbol{X}_{0:k-1}^i, \boldsymbol{y}_k)$
- Pour $i = 1, .., N_p$: Mise à jour des poids d'importance

$$w_k^i = w_{k-1}^i \sum_{j=0}^{J_{z_{k-1}}^i(k-1)} P(i, j)$$

- Pour $i = 1, .., N_p$:
  Si $rand = \mathcal{U}(0, 1) \leq P(0)$:
  - Echantillonner $\boldsymbol{x}_k$ selon (B.26)
  - Echantillonner Γ_k selon (B.14)
  - Mise à jour du nombre de clusters $J^i(k) = J_{z_{k-1}}^i(k-1) + 1$

  sinon
  - Echantillonner $\boldsymbol{x}_k$ selon (B.30)
  - Echantillonner Γ_k selon (B.18)
  - Mise à jour du nombre de clusters $J^i(k) = J_{z_{k-1}}^i(k-1)$
- Pour $i = 1, .., N_p$: Normaliser les poids d'importance

$$w_k^i = \frac{w_k^i}{\sum_{j=1}^{N_p} w_k^j}$$

- Calculer $N_{eff} = \frac{1}{\left(\sum_{i=1}^{N_p} (w_k^i)^2\right)}$
- Si $N_{eff} \leq C$, où C est une constante à définir, alors ré-échantillonner les particules.
  Cela donne lieu à un ensemble de particules $\{\boldsymbol{x}_k^i, w_k^i = \frac{1}{N_p}\}$.

Annexe C

Filtre particulaire auxiliaire

Comme introduit dans le chapitre 1, le filtre particulaire repose sur la méthode d'échantillonnage d'importance séquentiel (SIS). Une alternative est l'algorithme *sequential importance resampling* (SIR) obtenu à partir de l'algorithme SIS en choisissant la densité de transition comme densité d'importance et en ré-échantillonnant systématiquement.

Une variante algorithmique du filtre SIR est le filtre particulaire auxiliaire (APF) aussi appelé *auxiliary SIR filter* en anglais [Pit99]. Il permet de prendre en compte les mesures lors de l'étape de simulation des particules sans avoir recours à la distribution d'importance optimale. Le principe est de ne propager que les particules les plus vraisemblables en introduisant une variable auxiliaire.

Comme introduit dans le chapitre 1, dans l'algorithme "bootstrap", les nouvelles particules sont simulées à l'instant k selon la loi de mélange suivante :

$$\hat{p}(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \sum_{i=1}^{N_p} w_{k-1}^{(i)} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k-1}) \tag{C.1}$$

Pitt et Shephard proposent de modifier les probabilités de mélange et ainsi de remplacer $w_{k-1}^{(i)}$ de manière à prendre en compte l'information apportée par la mesure à l'instant courant. Ainsi, (C.1) devient :

$$\hat{p}(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \sum_{i=1}^{N_p} \tilde{w}_k^{(i)} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k-1}) \tag{C.2}$$

avec :

$$\tilde{w}_k^{(i)} = w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k-1}) \tag{C.3}$$

où $p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k-1})$ correspond à la vraisemblance prédite. Comme son expression analytique n'est généralement pas calculable, elle est approchée par une distribution du type $p(\boldsymbol{y}_k|\eta_k^{(i)})$ où $\eta_k^{(i)}$ est un paramètre caractérisant la distribution *a priori* $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k-1})$ tel que sa moyenne.

Le terme "auxiliaire" associé à ce filtre provient du tirage des particules qui s'effectue dans un espace de plus grande dimension que l'espace d'état. En effet, plutôt que d'échantillonner des ensembles de particules $\{x_k^{(i)}\}_{i=1,\ldots,N_p}$, ce sont des couples de la forme $\{x_k^{(i)}, j^i\}_{i=1,\ldots,N_p}$ qui

sont simulés où $\{j^i\}_{i=1,...,N_p}$ désignent les composantes de la loi de mélange ou encore l'indice des particules propagées, soit :

$$j^i \sim \sum_{l=1}^{N_p} \tilde{w}_{k-1}^{(i)} \delta(l - j^i),\ i = 1, ..., N_p \tag{C.4}$$

$$\boldsymbol{x}_k^{(i)} \sim p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}^{(j^i)}, \boldsymbol{y}_{1:k-1}),\ i = 1, ..., N_p \tag{C.5}$$

Les poids sont alors mis à jour comme suit :

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k | \boldsymbol{x}_k^{(j^i)}, \boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k | \eta_k^{(j^i)})} \tag{C.6}$$

La principale limitation de cette approche est le fait de caractériser la distribution *a priori* seulement par $\eta_k^{(i)}$ qui est souvent insuffisant pour pleinement la décrire.
Notons que cette variante algorithmique est surtout pertinente lorsque le bruit de modèle est faible.

Annexe D

Compléments mathématiques

D.1 Conjugaison de deux lois normales

Il s'agit de démontrer que sous l'hypothèse que la vraisemblance d'un vecteur aléatoire X_k connaissant les observations Y_k est gaussienne et que la loi *a priori* est également gaussienne, alors la loi *a posteriori* est gaussienne.
Cette démonstration permet en particulier de justifier le choix d'hypothèses gaussiennes lors de l'utilisation du filtrage de Kalman.

Supposons que la représentation dans l'espace d'état du système se met sous la forme linéaire suivante :

$$X_k = F_{k-1}X_{k-1} + V_k \tag{D.1}$$

$$Y_k = H_kX_k + W_k \tag{D.2}$$

avec V_k et W_k les bruits dits de modèle et de mesure respectivement. On suppose dans la suite que ces bruits sont gaussiens à moyenne nulle indépendants et identiquement distribués, soit :

$$V_k \sim \mathcal{N}(V_k; 0, Q_k) \tag{D.3}$$

$$W_k \sim \mathcal{N}(W_k; 0, R_k) \tag{D.4}$$

On suppose également que le vecteur d'état à l'instant $k-1$ suit la loi *a priori* suivante :

$$p(X_k|X_{k-1}) = \mathcal{N}(X_k; F_{k-1}X_{k-1}, P_{k|k-1}) \tag{D.5}$$

avec $P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_k$ où $P_{k-1|k-1}$ désigne la matrice de covariance sur l'erreur d'estimation à l'instant $k-1$.
De plus, la vraisemblance de l'état X_k par rapport aux observations Y_k est également une loi normale multivariées de paramètre Y_k, de moyenne H_kX_k et de matrice de covariance R_k :

$$p(Y_k|X_k) = \mathcal{N}(Y_k; H_kX_k, R_k) \tag{D.6}$$

Dans la suite, il s'agit de démontrer que la loi *a posteriori* $p(X_k|Y_k)$ peut également prendre la forme d'une loi normale multivariée.

Pour cela, rappelons que selon la formule de Bayes, la distribution *a posteriori* peut s'écrire sous la forme suivante :

$$p(X_k|Y_k) = \frac{p(Y_k|X_k)P(X_k)}{p(Y_k)} \tag{D.7}$$

où $p(Y_k)$ peut s'interpréter comme une constante de normalisation puisqu'elle satisfait :

$$p(Y_k) = \int p(Y_k|X_k)p(X_k)dX_k \tag{D.8}$$

Finalement, à la constante de normalisation près, on peut écrire selon (D.5) et (D.6) :

$$p(X_k|Y_k) \sim \mathcal{N}(X_k; F_{k-1}X_{k-1}, P_{k|k-1})\mathcal{N}(Y_k; H_kX_k, R_k) \tag{D.9}$$

$$\sim (2\pi)^{-N_x/2}(2\pi)^{-N_y/2}|R_k|^{-1/2}|P_{k|k-1}|^{-1/2}e^{-\frac{1}{2}q(X_k)} \tag{D.10}$$

avec $q(X_k)$ défini comme suit :

$$q(X_k) = (X_k - F_{k-1}X_{k-1})^T P_{k|k-1}^{-1}(X_k - F_{k-1}X_{k-1}) + (Y_k - H_kX_k)^T R_k^{-1}(Y_k - H_kX_k) \tag{D.11}$$

Il s'agit ensuite de faire apparaître une seule forme quadratique en développant comme suit :

$$\begin{aligned} q(X_k) = {} & X_k^T P_{k|k-1}^{-1}X_k + X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} - 2X_k^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} \\ & + Y_kR_k^{-1}Y_k + X_k^T H_k^T R_k^{-1}H_kX_k - 2X_k^T H_k^T R_k^{-1}Y_k \end{aligned} \tag{D.12}$$

En regroupant les termes de même degré, on obtient :

$$\begin{aligned} q(X_k) = {} & X_k^T(P_{k|k-1}^{-1} + H_k^T R_k^{-1}H_k)X_k - 2X_k^T(H_k^T R_k^{-1}Y_k + P_{k|k-1}^{-1}F_{k-1}X_{k-1}) \\ & + Y_k^T R_k^{-1}Y_k + X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} \end{aligned} \tag{D.13}$$

On introduit à présent les notations suivantes :

$$C_x^{-1} = P_{k|k-1}^{-1} + H_k^T R_k^{-1}H_k \tag{D.14}$$

$$\mu_x = C_x(H_k^T R_k^{-1}Y_k + P_{k|k-1}^{-1}F_{k-1}X_{k-1}) \tag{D.15}$$

Etant donné (D.13), (D.14) et (D.15), on a donc :

$$q(X_k) = X_k^T C_x^{-1}X_k - 2X_k^T C_x^{-1}\mu_x + Y_k^T R_k^{-1}Y_k + X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} \tag{D.16}$$

A partir de l'équation (D.16) et en faisant apparaître le développement de la forme quadratique $(X_k - \mu_x)^T C_x^{-1}(X_k - \mu_x)$, on peut identifier la forme quadratique d'une loi normale. En effet, il vient :

$$\begin{aligned} q(X_k) = {} & (X_k^T C_x^{-1}X_k - 2X_k^T C_x^{-1}\mu_x + \mu_x^T C_x^{-1}\mu_x) - \mu_x^T C_x^{-1}\mu_x + Y_k^T R_k^{-1}Y_k \\ & + X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} \end{aligned} \tag{D.17}$$

Soit :

$$q(X_k) = (X_k - \mu_x)^T C_x^{-1}(X_k - \mu_x) - \mu_x^T C_x^{-1}\mu_x + Y_k^T R_k^{-1}Y_k + X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1}F_{k-1}X_{k-1} \tag{D.18}$$

Ensuite, en combinant (D.18) avec (D.10), on obtient une expression pour la loi *a posteriori* de X_k à la constante de normalisation $p(Y_k)$ près :

$$\begin{aligned} p(X_k|Y_k) \sim & \frac{(2\pi)^{-N_x/2}(2\pi)^{-N_y/2}|R_k|^{-1/2}|P_{k|k-1}|^{-1/2}}{(2\pi)^{-N_x/2}|C_x|^{-1/2}} \\ & \times exp(1/2\mu_x^T C_x^{-1}\mu_x - 1/2Y_k^T R_k^{-1} Y_k - 1/2X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1} F_{k-1} X_{k-1}) \\ & \times \mathcal{N}(X_k; \mu_x, C_x) \end{aligned} \tag{D.19}$$

On peut ensuite obtenir la constante de normalisation en intégrant comme suit :

$$\begin{aligned} p(Y_k) = & \int \frac{(2\pi)^{-N_y/2}|R_k|^{-1/2}|P_{k|k-1}|^{-1/2}}{|C_x|^{-1/2}} \\ & \times exp(1/2\mu_x^T C_x^{-1}\mu_x - 1/2Y_k^T R_k^{-1} Y_k - 1/2X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1} F_{k-1} X_{k-1}) \\ & \times \mathcal{N}(X_k; \mu_x, C_x)\ dX_k \end{aligned} \tag{D.20}$$

$$\begin{aligned} p(Y_k) = & \frac{(2\pi)^{-N_y/2}|R_k|^{-1/2}|P_{k|k-1}|^{-1/2}}{|C_x|^{-1/2}} \\ & \times exp(1/2\mu_x^T C_x^{-1}\mu_x - 1/2Y_k^T R_k^{-1} Y_k - 1/2X_{k-1}^T F_{k-1}^T P_{k|k-1}^{-1} F_{k-1} X_{k-1}) \end{aligned} \tag{D.21}$$

Finalement, en effectuant la normalisation de (D.19) à partir de (D.21), on montre que la densité *a posteriori* de X_k décrite en (D.10) suit une loi normale de moyenne μ_x et de matrice de covariance C_x :

$$p(X_k|Y_k) = \mathcal{N}(X_k; \mu_x, C_x) \tag{D.22}$$

D.2 Distribution t de Student

Si y est une variable aléatoire distribuée selon une loi normale centrée réduite et z est une variable aléatoire qui suit une loi du χ^2 à ν degrés de liberté, alors la variable aléatoire x suit une loi de Student à ν degrés de liberté si :

$$x = \frac{y}{\sqrt{\frac{z}{\nu}}} \tag{D.23}$$

La densité de x, notée $p_x(\cdot)$ satisfait pour $\nu > 1$:

$$p_x(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{D.24}$$

où $\Gamma(\cdot)$ est la fonction Gamma.

L'espérance mathématique et la variance de x vérifient :

$$E[x] = 0 \text{ si } \nu > 1 \tag{D.25}$$

$$Var[x] = \frac{\nu}{\nu - 2} \text{ si } \nu > 2 \tag{D.26}$$

Propriété :
La distribution de Student s'apparente à la distribution de la loi normale. Elle est symétrique et centrée en zéro. Néanmoins, comme illustré par la figure D.1 elle varie selon le degré de liberté ν.
Si $\nu \to +\infty$, la distribution t de Student tend vers une distribution normale centrée réduite. En pratique, la distribution de Student peut être approchée par une distribution normale centrée réduite pour des valeurs de ν supérieures ou égales à 30.

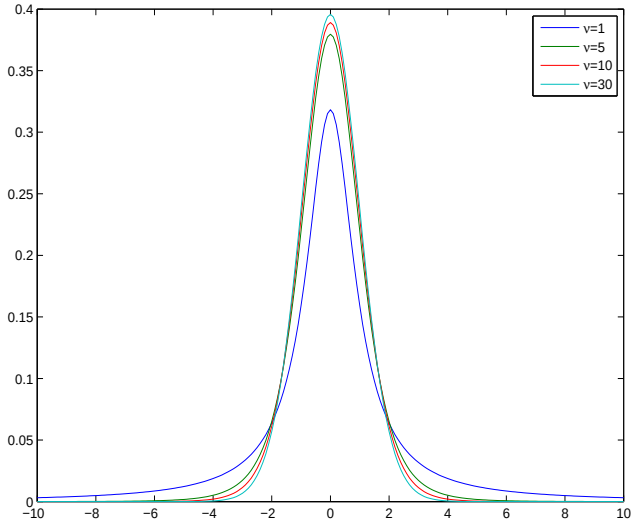

FIGURE D.1 – Distribution de probabilité t de Student pour différentes valeurs du nombre de degrés de liberté ν

D.3 Distribution t multivariée

La distribution t multivariée constitue la généralisation de la distribution t de Student aux vecteurs aléatoires.
Notons p la dimension du vecteur aléatoire $\boldsymbol{x}$. Si $\boldsymbol{y}$ est un vecteur aléatoire distribué selon une loi normale multivariée centrée et de matrice de covariance Σ et que u est une variable aléatoire distribuée selon une loi du χ^2 à ν degrés de liberté, alors en posant :

$$\boldsymbol{x} = \boldsymbol{y}\sqrt{\frac{\nu}{u}} + \boldsymbol{\mu} \tag{D.27}$$

le vecteur aléatoire $\boldsymbol{x}$ est supposé suivre une distribution t multivariée de paramètres Σ, $\boldsymbol{\mu}$ et ν satisfaisant :

$$t_\nu(\boldsymbol{\mu}, \Sigma) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\nu^{p/2}\pi^{p/2}|\Sigma|^{1/2}(1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^{\frac{\nu+p}{2}}} \tag{D.28}$$

La distribution est de moyenne μ pour $\nu > 1$.
De plus, elle a pour variance $\frac{\nu}{\nu-2}\Sigma$ si $\nu > 2$.

Propriété :
La distribution t multivariée peut être approchée par une distribution normale multivariée pour des valeurs de ν suffisamment grandes.

D.4 Distribution t matricielle

La distribution t matricielle constitue la généralisation de la distribution t multivariée aux matrices.

Une matrice aléatoire T de taille $p \times m$ suit une distribution t matricielle de paramètres M, Σ, Ω et $\nu > 0$ si sa densité de probabilité satisfait [WEI12a] :

$$T_{p,m}(\nu, M, \Sigma, \Omega) = \frac{\Gamma_p(\frac{\nu+m+p-1}{2})}{\Gamma_p(\frac{\nu+p-1}{2})\pi^{mp/2}} |\Sigma|^{-m/2} |\Omega|^{-p/2} |I_p + \Sigma^{-1}(T-M)\Omega^{-1}(T-M)'|^{-\frac{\nu+m+p-1}{2}} \tag{D.29}$$

où $\Gamma_p(\cdot)$ est la fonction Gamma multivariée de paramètre p, $T \in \mathbb{R}^{p\times m}$ et $M \in \mathbb{R}^{p\times m}$. De plus, Ω et Σ sont des matrices semi-défines positives de dimension $m \times m$ et $p \times p$ respectivement.
La distribution est de moyenne M si $\nu + m - p > 1$.
De plus, elle a pour variance $\frac{M\otimes\Omega}{\nu-2}$ si $\nu > 2$, où $\otimes$ désigne le produit de Kronecker.
Propriété :
Selon la valeur de ν, la distribution t matricielle peut être approchée par une distribution normale matricielle. En effet si $T \sim T_{p,m}(\nu, M, \nu\Sigma, \Omega)$ alors quand $\nu \to \infty$, T converge vers X avec $X \sim \mathcal{N}(X; M, \Sigma \otimes \Omega)$.

D.5 Distribution de Wishart

En statistique, la distribution de Wishart est la conjuguée *a priori* de l'inverse de la matrice de covariance d'un vecteur aléatoire normal multivarié. Pour toute matrice symétrique définie positive X de taille $n \times p$, la densité de probabilité de la distribution de Wishart satisfait :

$$\mathcal{W}(X; W_0, \nu) = B(W_0, \nu)|X|^{\frac{\nu-p-1}{2}} \exp\left(-\frac{1}{2}Tr[W_0^{-1}X]\right) \tag{D.30}$$

où W_0 et ν désignent la matrice d'échelle et le nombre de degrés de liberté respectivement avec $B(W_0, \nu)$ la constante de normalisation qui vaut :

$$B(W_0, \nu) = \frac{1}{2^{\frac{\nu p}{2}} |W_0|^{\frac{\nu}{2}} \Gamma_p(\frac{\nu}{2})} \tag{D.31}$$

où $\Gamma_p(\cdot)$ est la fonction Gamma multivariée.

Bibliographie

[Ald83] D. Aldous, *Exchangeability and Related Topics*, Ecole d'été de Probabilités de Saint-Flour XIII, 1983.

[Als72] D. Alspach and H. Sorenson, "A New Method for the Nonlinear Transformation of Means ans Covariances in Filters and Estimators," *IEEE Trans. on Automatic Control*, Vol. 17, pp. 439–448, 1972.

[Ama00] S. Amari and H. Nagaoka, *Methods on Information Geometry, Translations of Mathematical Monographs*, American Mathematical Society, 2000.

[Ang04] D. Angelova and L. Mihaylova, "Sequential Monte Carlo Algorithms for Joint Target Tracking and Classification Using Kinematic Radar Information," In: *Int. Conf. on Information Fusion*, pp. 709–716, 2004.

[Ant74] C. Antoniak, "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, Vol. 2, pp. 1152–1174, 1974.

[Ara09] I. Arasaratnam and S. Haykin, "Cubature Kalman Filters," *IEEE Trans. on Automatic Control*, Vol. 54, pp. 1254–1259, 2009.

[Ath68] M. Athans, R. Wishner and A. Bertolini, "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *IEEE Trans. on Automatic Control*, Vol. 13, pp. 504–514, 1968.

[Axe11] P. Axelsson, U. Orguner, F. Gustafsson and M. Norrlof, "ML Estimation of Process Noise Variance in Dynamic System," In: *International Federation of Automatic Control*, pp. 5609–5614, 2011.

[Bar01] Y. Bar-Shalom, X. R. Li and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*, Wiley, 2001.

[Bar09] Y. Bar-Shalom, F. Daum and J. Huang, "The Probabilistic Data Association Filter," *IEEE Control Systems*, Vol. 29, pp. 82–100, 2009.

[Bar11] Y. Bar-Shalom, P. Willett and X. Tian, *Tracking and Data Fusion: a Handbook of Algorithms*, YBS Publishing, 2011.

[Bar93] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques and Software*, Artech House, 1993.

[Bar95] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, 1995.

[Ben15] M. Ben Mabrouk, E. Grivel, C. Magnant, G. Ferré and N. Deltimple, "Compensating Power Amplifier Distortion in Cognitive Radio Systems with Adaptive Interacting Multiple Model," In: *European Signal Processing Conference, EUSIPCO 2015*, 2015.

[Ber91] D. P. Bertsekas, *Linear Network Optimization*, MIT Press, 1991.

[Bis06] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[Bla04] S. S. Blackman, "Multiple Hypothesis Tracking for Multiple Target Tracking," *IEEE Trans. on Aerospace and Elec. Syst.*, Vol. 19, pp. 5–18, Jan. 2004.

[Bla73] D. Blackwell and B. MacQueen, "Ferguson Distributions via Polya Urn Schemes," *The annals of statistics*, Vol. 1, pp. 353–355, 1973.

[Bla99] S. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Norwood, MA : Artech House, 1999.

[Cao13] W. Cao, J. Lan and X. R. Li, "Extended Object Tracking and Classification Based on Recursive Joint Decision and Estimation," In: *Int. Conf. on Information Fusion*, pp. 1670–1677, 2013.

[Cao14] W. Cao, J. Lan and X. R. Li, "Joint Tracking and Classification of Extended Object Based on Recursive Joint Decision and Estimation Using Multi-Sensor Data," In: *Int. Conf. on Information Fusion*, pp. 1–8, 2014.

[Car07] F. Caron, M. Davy and A. Doucet, "Generalized Polya Urn for Time-Varying Dirichlet Process Mixtures," In: *Conf. on Uncertainty Artificial Intelligence*, 2007.

[Car08] F. Caron, M. Davy, A. Doucet, E. Duflos and P. Vanheeghe, "Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures," *IEEE Trans. on Signal Processing*, Vol. 56, pp. 71–84, 2008.

[Cha01] S. Challa and G. W. Pulford, "Joint Target Tracking and Classification Using Radar and ESM Sensors," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 37, pp. 1039–1055, 2001.

[Cha14] P. Chavali and A. Nehorai, "Hierarchical Particle Filtering for Multi-Modal Data Fusion with Application to Multiple-Target Tracking," *Signal Processing*, Vol. 97, pp. 207–220, 2014.

[Cla05] D. Clark and J. Bell, "Data Association for the PHD Filter," In: *International Conf. on the Intelligent Sensor, Sensor Network and Information Processing*, pp. 217–222, 2005.

[Cla06] D. Clark, B. Vo and J. Bell, "The GM-PHD Filter Multiple Target Tracker," In: *International Conf. on Information Fusion, Florence, Italy*, pp. 1–8, Jul. 2006.

[Coe05] F. M. Coetzee, "Correcting the Kullback-Leibler Distance for Feature Selection," *Pattern Recognition Letters*, Vol. 26, pp. 1675–1683, 2005.

[Cox96] I. Cox and S. Hingorani, "An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and its Evaluation for the Purpose of Visual Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 138–150, 1996.

[Dij59] E. Dijkstra, "A Note on Two Problems in Connexion with Graphs," In: *Numerishe Mathematik 1*, pp. 269–271, 1959.

[Dor09] R. M. Dorazio, "On Selecting a Prior for the Precision Parameter of Dirichlet Process Mixture Models," *Journal of Statistical Planning*, pp. 3384–3390, 2009.

[Dou00a] A. Doucet, N. de Freitas, K. Murphy and S. Russell, "Rao-Blackwellized Particle Filtering for Dynamic Bayesian Networks," In: *Conf. on Uncertainty Artificial Intelligence*, pp. 176–183, 2000.

[Dou00b] A. Doucet, S. Godsill and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, Vol. 10, pp. 197–208, 2000.

[Dou01a] A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.

[Dou01b] A. Doucet, N. Gordon and V. Krisnamurthy, "Particle Filtering State Estimation of Jump Markov Linear Systems," *IEEE Trans. on Signal Processing*, Vol. 49, pp. 613–624, 2001.

[Dun08] M. Dunik, J. ad Simandl, "Estimation of State and Measurement Noise Covariance Matrices by Multi-Step Prediction," In: *International Federation of Automatic Control (IFAC) World Congress*, 2008.

[Elg13] T. Elguebaly and N. Bouguila, "Simultaneous Bayesian Clustering and Feature Selection Using RJMCMC-Based Learning of Finite Generalized Dirichlet Mixture Models," *Signal Processing*, Vol. 93, pp. 1531–1546, 2013.

[Erd05] O. Erdinc, P. Willett and Y. Bar-Shalom, "Probability Hypothesis Density Filter for Multitarget Multisensor Tracking," In: *International Conf. on Information Fusion*, pp. 146–153, 2005.

[Fer73] T. Fergusson, "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, Vol. 1, pp. 209–230, 1973.

[For83] T.-E. Fortmann, Y. Bar-Shalom and M. Scheffe, "Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association," *IEEE Journal of Oceanic Engineering*, Vol. 8, pp. 173–184, July 1983.

[Fox06] E. Fox, D. S. Choi and A. S. Willsky, "Nonparametric Bayesian Methods for Large Scale Multi-Target Tracking," In: *Conf. on Signals, Systems and Computers*, pp. 2009–2013, 2006.

[Fox07] E. Fox, E. Sudderth and A. Willsky, "Hierarchical Dirichlet Processes for Tracking Maneuvering Targets," In: *Int. Conf. on Information Fusion*, pp. 1–8, 2007.

[Fox10] E. Fox, E. Sudderth, M. I. Jordan and A. Willsky, "Bayesian Nonparametric Methods for Learning Markov Switching Processes," *IEEE Signal Processing Magazine*, Vol. 27, pp. 43–54, 2010.

[Fox11a] E. Fox, E. Sudderth, M. I. Jordan and A. Willsky, "A Sticky HDP-HMM with Application to Speaker Diarization," *Annals of Applied Statistics*, Vol. 5, pp. 1020–1056, 2011.

[Fox11b] E. Fox, E. Sudderth, M. I. Jordan and A. Willsky, "Bayesian Nonparametric Inference of Switching Dynamic Linear Models," *IEEE Trans. on Signal Processing*, Vol. 59, pp. 1569–1585, 2011.

[Fre12] M. Frei and H. Kunsch, "Sequential State and Observation Noise Covariance Estimation Using Combined Ensemble Kalman and Particle Filters," *Monthly Weather Review*, Vol. 140, pp. 1476–1495, 2012.

[Gab99] M. Gabrea, E. Grivel and M. Najim, "A Single Microphone Kalman Filter-Based Noise Canceller," *IEEE Signal Processing Letters*, Vol. 6, pp. 55–57, 1999.

[Ge14] M. Ge and E. C. Kerrigan, "Noise Covariance Estimation for Time-Varying and Nonlinear Systems," In: *International Federation of Automatic Control (IFAC) world congress*, pp. 9545–9550, 2014.

[Gir10] A. Giremus, J.-Y. Tourneret and A. Doucet, "Fixed Lag Particle Filter for the Joint Detection/Compensation of Interference Effects in GPS Navigation," *IEEE Trans. on Signal Processing*, Vol. 58, pp. 6066–6079, 2010.

[Goo97] I. Goodman, R. Mahler and H. Nguyen, *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997.

[Gor93] N. J. Gordon, D. J. Salmond and A. F. M. Smith, "A Novel Approach to Non-Linear and Non-Gaussian Bayesian State Estimation," In: *IEE Proceedings of Radar and Signal Processing*, pp. 107–113, 1993.

[Hau12] A. J. Haug, *Bayesian Estimation and Tracking: a Practical Guide*, Wiley, 2012.

[Ho64] Y. Ho and R. Lee, "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Trans. on Control Systems Society*, Vol. 9, pp. 333–339, 1964.

[Hof04] J. Hoffman and R. Mahler, "Multi-target Miss Distance via Optimal Assignments," *IEEE Trans. on Systems, Man and Cybernatics*, Vol. 34, pp. 327–336, 2004.

[Ish00] H. Ishwaran and M. Zarepour, "Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models," *Biometrika*, Vol. 87, pp. 371–390, 2000.

[Ito00] K. Ito and K. Xiong, "Gaussian Filters for Non-Linear Filtering Problems," *IEEE Trans. on Automatic Control*, Vol. 45, pp. 910–927, 2000.

[Jef48] H. Jeffreys, *Theory of Probability*, 2nd edition, Oxford, 1948.

[Jil04] V. P. Jilkov and X. R. Li, "Online Bayesian Estimation of Transition Probabilities for Markovian Jump Systems," *IEEE Trans. on Signal Processing*, Vol. 52, pp. 1620–1630, 2004.

[Kal60] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. of ASME-Journal of Basic Engineering*, Vol. 82, pp. 35–45, 1960.

[Kon94] A. Kong, J. S. Liu and W. H. Wong, "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, Vol. 89, pp. 278–288, 1994.

[Kul51] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, pp. 79–86, 1951.

[Kur88] T. Kurien and M. Liggins, "Report-to-Target Assignment in Multisensor Multitarget Tracking," In: *IEEE Conf. on Decision and Control*, pp. 2484–2488, Dec. 1988.

[Kur91] T. Kurien, "Framework for Integrated Tracking and Identification of Multiple Targets," In: *IEEE/AIAAA Conf. on Digital Avionics Systems*, pp. 362–366, Oct. 1991.

[Lan13] J. Lan and X. R. Li, "Joint Tracking and Classification of Extended Object Using Random Matrix," In: *Int. Conf. on Information Fusion*, pp. 1550–1557, 2013.

[Leg16] L. Legrand and E. Grivel, "Evaluating Dissimilarities Between Two Moving-Average Models: Comparative Study Between Jeffrey's Divergence and Rao Distance," In: *European Signal Processing Conference, EUSIPCO 2016*, 2016.

[Li00] X. R. Li, *Engineer's Guide to Variable-Structure Multiple-Model Estimation for Tracking*, in Multitarget-Multisensor Tracking: Applications and Advances, Ch. 10, pp. 499-567, Artech House, 2000.

[Li03] X. R. Li and V. P. Jilkov, "Survey of Maneuvering Target Tracking. Part I: Dynamic Models," *IEEE Trans. on Aerospace and Elec. Syst.*, Vol. 39, pp. 1333–1364, 2003.

[Li05] X. R. Li and V. P. Jilkov, "Survey of Maneuvering Target Tracking. Part V: Multiple Model Methods," *IEEE Trans. on Aerospace and Elec. Syst.*, Vol. 41, pp. 1255–1321, Oct. 2005.

[Li07] X. R. Li, M. Yang and J. Ru, "Joint Tracking and Classification Based on Bayes Joint Decision and Estimation," In: *Int. Conf. on Information Fusion*, 2007.

[Li96] X. R. Li and Y. Bar-Shalom, "Multiple Model Estimation with Variable Structure-Part IV," *IEEE Trans. on Automatic Control*, Vol. 41, pp. 478–493, 1996.

[Mac67] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Obervations," In: *5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[Mag13] C. Magnant, A. Giremus and E. Grivel, "Jeffreys Divergence Between State Models: Application to Target Tracking using Multiple Models," In: *European Signal Processing Conference, EUSIPCO 2013*, pp. 1–5, 2013.

[Mag14] C. Magnant, E. Grivel, A. Giremus, B. Joseph and L. Ratton, "Classifying Several State Models Using Jeffrey's Divergence: Application to Target Tracking," In: *IEEE Radar Conf., RADAR'14*, pp. 1–6, 2014.

[Mag15a] C. Magnant, A. Giremus and E. Grivel, "On Computing Jeffrey's Divergence Between Time-Varying Autoregressive Models," *IEEE Signal Processing Letters*, Vol. 22, pp. 915–919, 2015.

[Mag15b] C. Magnant, A. Giremus, E. Grivel, L. Ratton and B. Joseph, "Dirichlet-Process-Mixture-Based Bayesian Nonparametric Method for Markov Switching Process Estimation," In: *European Signal Processing Conference, EUSIPCO 2015*, 2015.

[Mag15c] C. Magnant, A. Giremus, E. Grivel, L. Ratton and B. Joseph, "Joint Tracking and Classification Based on Kinematic and Target Extent Measurements," In: *Int. Conf. on Information Fusion*, pp. 1748–1755, 2015.

[Mag15d] C. Magnant, E. Grivel, A. Giremus, B. Joseph. and L. Ratton, "Jeffrey's Divergence for State-Space Model Comparison," *Signal Processing*, Vol. 114, pp. 61–74, 2015.

[Mag15e] C. Magnant, E. Grivel, A. Giremus, L. Ratton and B. Joseph, "Classifying Autoregressive Models Using Dissimilarity Measures: a Comparative Study," In: *European Signal Processing Conference, EUSIPCO 2015*, 2015.

[Mag16a] C. Magnant, A. Giremus, E. Grivel, L. Ratton and B. Joseph, "Bayesian Non-Parametric Methods for Dynamic State-Noise Covariance Matrix Estima-

tion: Application to Target Tracking," *Signal Processing, DOI information: 10.1016/j.sigpro.2016.02.013*, 2016.

[MAG16b] C. Magnant, A. Giremus, E. Grivel, L. Ratton and B. Joseph, "Multi-target Tracking Using a PHD-Based Joint Tracking and Classification Algorithm," In: *IEEE Radar Conf., RADAR'16*, 2016.

[MAH01] R. Mahler, *Random Set Theory for Target Tracking and Identification*, Handbook of Multisensor Data Fusion, 2001.

[MAH03] R. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 39, pp. 1152–1178, 2003.

[MAH04] R. Mahler, "Statistics 101 for Multisensor, Multitarget Data Fusion," *IEEE Trans. on Aerospace and Electronics Systems Magazine*, Vol. 19, pp. 53–64, 2004.

[MAH07] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Artech House, 2007.

[MAH13] R. Mahler, "Statistics 102 for Multisource, Multitarget Detection and Tracking," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 7, pp. 376–389, 2013.

[MAK10] E. Makalic and F. Schmidt, "Fast Computation of the Kullback-Leibler Divergence and Exact Fisher Information for the First-Order Moving Average Model," *IEEE Signal Processing Letters*, Vol. 17, pp. 391–393, 2010.

[MEH70] R. K. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE. Trans. on Automatic Control*, Vol. 15, pp. 175–184, 1970.

[MEH71] R. K. Mehra, "On-line Identification of Linear Dynamic Systems with Applications to Kalman Filtering," *IEEE. Trans. on Automatic Control*, Vol. 16, pp. 12–21, 1971.

[MEH72] R. K. Mehra, "Approaches to Adaptive Filtering," *IEEE Trans. on Automatic Control*, Vol. 15, pp. 903–908, 1972.

[MEI04] W. Mei, G.-L. Shan and X. R. Li, "An Efficient Bayesian Algorithm for Joint Target Tracking and Classification," In: *International Conference on Signal Processing*, pp. 2090–2098, 2004.

[MEI07] W. Mei, G.-L. Shan and X. R. Li, "Simultaneous Tracking and Classification: a Modularized Scheme," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 43, pp. 581–599, 2007.

[MIL97] M. Miller, H. Stone and I. Cox, "Optimizing Murty's Ranked Assignment Method," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 33, pp. 851–862, July 1997.

[MOL04] C. B. Moler, *Numerical Computing with Matlab, Chapter 10: Eigenvalues and Singular Values*, Society for Industrial and Applied Mathematics, 2004.

[MUN57] J. Munkres, "Algorithm for the Assignment and Transportation Problems," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 5, pp. 32–38, 1957.

[MUR68] K. Murty, "An Algorithm for Ranking all the Assignments in Order of Increasing Cost," *Operation Research*, Vol. 16, pp. 682–687, May 1968.

[Mus94] D. Musicki and B. L. Scala, "Integrated Probability Data Association," *IEEE Trans. on Automatic Control*, Vol. 39, pp. 1237–1241, 1994.

[Neu51] J. V. Neumann, "Various Techniques Used in Connection with Random Digits," *Nat. Bureau Standards*, Vol. 12, pp. 36–38, 1951.

[Nie09] F. Nielsen and R. Nock, "Sided and Symmetrized Bregman Centroids," *IEEE Trans. on Information Theory*, Vol. 55, pp. 2882–2904, 2009.

[Ode06] B. J. Odelson, M. R. Rajamani and J. B. Rawlings, "A New Autocovariance Least Squares Method for Estimating Noise Covariance," *Automatica*, Vol. 42, pp. 532–540, 2006.

[Pac10] M. Pace and F. Caron, "Comparison of Implementation of Gaussian Mixture PHD Filters," In: *International Conf. on Information Fusion*, 2010.

[Pan05] K. Panta, B. Vo and S. Singh, "Improved Probability Hypothesis Density (PHD)," In: *International Conf. on Intelligent Sensing and Information Processing*, 2005.

[Pan06a] K. Panta, B. Vo and D. Clark, "An Efficient Track Management Systems for the Gaussian Mixture Probability Hypothesis Density (GM-PHD) Tracker," In: *International Conf. on Intelligent Sensing and Information Processing*, Dec. 2006.

[Pan06b] K. Panta, B. Vo and S. Singh, "Novel Data Association Technique for the Probability Hypothesis Density Filter," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 43, pp. 556–570, 2006.

[Pan07] K. Panta. *Multi-Target Tracking Using 1st Moment of Random Finite.* PhD thesis, University of Melbourne, supervised by Ba-Ngu Vo, 2007.

[Pan09] K. Panta, D. Clark and B. Vo, "Data Association and Track Management for the Gaussian-Mixture Probability Hypothesis Density Filter," *IEEE Trans. on Aerospace and Electronics Systems*, pp. 1003–1016, 2009.

[Pap13] F. Papi, B. Vo, M. Bocquel and B. Vo, "Multi-target Track-Before-Detect Using Labeled Random Finite Sets," In: *Conf. on Control, Automation and Information Sciences (ICCAIS)*, pp. 116–121, 2013.

[Per13] V. Pereira, A. Giremus, A. Rabaoui and E. Grivel, "Online EM Estimation of the Dirichlet Process Mixture Scale Parameter to Model the GPS Multipath Error," In: *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, pp. 6625–6629, 2013.

[Per14] V. Pereira, G. Ferre, A. Giremus and E. Grivel, "Relevance of Dirichlet Processes Mixtures for Modeling Interferences in Underlay Cognitive Radio," In: *European Signal Processing Conference, EUSIPCO 2014*, pp. 176–180, 2014.

[Pha01] D. T. Pham, "Stochastic Methods for Sequential Data Assimilation in Strongy Nonlinear Systems," *Monthly Weather Review*, Vol. 129, pp. 1194–1207, 2001.

[Pit97] J. Pitman, "The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator," *The Annals of Probability*, Vol. 25, pp. 855–900, 1997.

[Pit99] M. Pitt and N. Shephard, "Filtering via Simulation: Auxiliary Particle Filtering," *Journal of the American Statistical Association*, Vol. 94, pp. 590–599, 1999.

[Pun05] K. Punithakumar and T. Kirubarajan, "A Sequential Monte Carlo Probability Hypothesis Density Algorithm for Multitarget Track-Before-Detect," In: *SPIE Signal and Data Processing of Small Targets*, pp. 505–512, 2005.

[Pun08] K. Punithakumar, T. Kirubarajan and A. Sinha, "Multiple-Model Probability Hypothesis Density Filter for Tracking Maneuvering Targets," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 44, pp. 87–98, 2008.

[Qui00] B. M. Quine, "A Derivative-Free Implementation of the Extended Kalman filter," *Automatica*, Vol. 42, pp. 1927–1934, 2000.

[Rab12] A. Rabaoui, N. Viandier, J. Marais, E. Duflos and P. Vanheeghe, "Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons," *IEEE Trans. on Signal Processing*, Vol. 60, pp. 1638–1655, 2012.

[Rac04] Z. Rached, F. Alajaji and L. L. Campbell, "The Kullback Leibler Divergence Rate Between Markov Sources," *IEEE Trans. on Information Theory*, Vol. 50, pp. 917–921, 2004.

[Raj09] M. R. Rajamani and J. B. Rawlings, "Estimation of the Disturbance Structure from Data Using Semidefinite Programming and Optimal Weighting," *Automatica*, Vol. 45, pp. 142–148, 2009.

[Ras06] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning, Chap. 7*, MIT Press, 2006.

[Rei79] D. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Trans. on Automatic Control*, Vol. 24, pp. 843–854, 1979.

[Reu13] S. Reuter. *Multi-Object Tracking Using Random Finite Sets.* PhD thesis, Ulm University, supervised by K. Dietmayer, 2013.

[Reu15] S. Reuter, A. Scheel and K. Dietmayer, "The Multiple Model Labeled Multi-Bernoulli Filter," In: *Int. Conf. on Information Fusion*, pp. 1574–1580, 2015.

[Ris04a] B. Ristic, S. Arulampalam and N. Gordon, *Beyond the Kalman Filter*, Artech House Pulishers, 2004.

[Ris04b] B. Ristic, N. Gordon and A. Bessell, "On Target Classification Using Kinematic Data," *Journal on Information Fusion*, Vol. 5, pp. 15–21, 2004.

[Ris04c] B. Ristic and D. J. Salmond, "A Study of Nonlinear Filtering Problem for Tracking an Extended Target," In: *Int. Conf. on Information Fusion*, 2004.

[Rot11] M. Roth and F. Gustafsson, "An Efficient Implementation of the Second Order Extended Kalman Filter," In: *Int. Conf. on Information Fusion*, pp. 1–6, 2011.

[Sal03] D. J. Salmond and M. C. Parr, "Track Maintenance Using Measurements of Target Extent," *IEE Proceedings of Radar, Sonar and Navigation*, Vol. 150, pp. 389–395, 2003.

[Sch08] D. Schuhmacher, B.-T. Vo and B.-N. Vo, "A Consistent Metric for Performance Evaluation of Multi-Object Filters," *IEEE Trans. on Signal Processing*, Vol. 56, pp. 3447–3457, 2008.

[Sch13] A. Schutz, L. Bombrun and Y. Berthoumieu, "Centroid-Based Texture Classification Using the SIRV Representation," In: *IEEE Int. Conf. Image Processing, ICIP 2013*, pp. 3810–3814, 2013.

[Set94] J. Sethuraman, "A Contructive Definition of Dirichlet Priors," *Statistica Sinica*, Vol. 4, pp. 639–650, 1994.

[Sid03] H. Sidenbladh, "Multi-Target Particle Filtering for the Probability Hypothesis Density," In: *International Conf. of Fusion*, pp. 1110–1117, 2003.

[Sin70] R. A. Singer, "Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets," *IEEE Trans. on Aerospace and Elec. Syst.*, Vol. 6, pp. 473–483, 1970.

[Sol14] A. Solonen, J. Hakkarainen, A. Ilin, M. Abbas and A. Bibov, "Estimating Model Error Covariance Matrix Parameters in Extended Kalman Filtering," *Nonlinear Processes in Geophysics*, Vol. 21, pp. 919–927, 2014.

[Spa10] M. Spangenberg, V. Calmettes, O. Julien, J.-Y. Tourneret and G. Duchateau, "Detection of Variance Changes and Mean Value Jumps in Measurement Noise for Multipath Mitigation in Urban Navigation," *Journal of the Institute of Navigation*, Vol. 57, pp. 35–52, 2010.

[Sun14] L. Sun, J. Lan and X. R. Li, "Joint Tracking and Classification of Extended Object Based on Support Functions," In: *Int. Conf. on Information Fusion*, pp. 1–8, 2014.

[The92] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, 1992.

[Tob06] M. Tobias. *Probability Hypothesis Densities for Multitarget, Multisensor Tracking with Application to Passive Radar.* PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, supervised by A. Lanterman, 2006.

[Tug82] J. K. Tugnait, "Adaptive Estimation and Identification for Discrete Systems with Markov Jump Parameters," *IEEE Trans. on Automatic Control*, Vol. 27, pp. 1054–1065, 1982.

[Ull96] A. Ullah, "Entropy, Divergence and Distance Measures with Econometric Applications," *Statistical Planning and Inference*, Vol. 49, pp. 137–162, 1996.

[Val15] I. Valera, F. Ruiz, L. Svensson and F. Perez-Cruz, "A Bayesian Nonparametric Approach for Blind Multiuser Channel Estimation," In: *Invited talk at the special session on Bayesian non-parametrics for Signal and Image Processing at European Signal Processing Conference*, 2015.

[Van94] P. Van Overschee and B. De Moor, "N4SID: Subspace Algorithm for the Identification of Combined Deterministic and Stochastic Systems," *Automatica*, Vol. 30, pp. 75–93, 1994.

[Var04] R. S. Varga, *Gershgorin and his Circles*, Springer Science, 2004.

[Ver02] J. Vermaak, C. Andrieu, A. Doucet and S. J. Godsill, "Particle Methods for Bayesian Modeling and Enhancement of Speech Signals," *IEEE Trans. on Speech and Audio Processing*, Vol. 10, pp. 173–185, 2002.

[Vo03] B. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo Implementation of the PHD Filter for Multitarget Tracking," In: *International Conf. of Fusion*, pp. 792–799, 2003.

[Vo05a] B.-N. Vo and S. Singh "Technical Aspects of the Probability Hypothesis Density Recursion,", Tech. Rep., Dept. of Electrical Engineering, University of Melbourne, 2005.

[Vo05b] B.-N. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo Methods for Multi-target Filtering with Random Finite Sets," *IEEE Trans. on Aerospace and Elec. Syst.*, Vol. 41, pp. 1224–1245, 2005.

[Vo06] B.-T. Vo and W.-K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Trans. on Signal Processing*, pp. 4091–4104, 2006.

[Vo07] B.-T. Vo, B.-N. Vo and A. Cantoni, "Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter," *IEEE Trans. on Signal Proc.*, Vol. 55, pp. 3553–3567, 2007.

[Vo08] B.-T. Vo. *Random Finite Sets in Multi-Object Filtering.* PhD thesis, University of Western Australian, supervised by A. Cantoni, 2008.

[Vo13] B.-T. Vo and B.-N. Vo, "Labeled Random Finite Sets and Multi-Object Conjugate Priors," *IEEE Trans. on Signal Processing*, Vol. 61, pp. 3460–3475, 2013.

[Wan01] E. A. Wan and R. V. de Merwe, *Chapter 7: The Unscented Kalman Filter, in Kalman Filtering and Neural Networks*, Wiley, 2001.

[Wei12a] X. Wei and C. Li, "The Infinite Student's t-Mixture for Robust Modeling," *Signal Processing*, Vol. 92, pp. 224–234, 2012.

[Wei12b] Y. Wei, F. Yaowen, L. Jianqian and L. Xiang, "Joint Detection, Tracking, and Classification of Multiple Targets in Clutter Using the PHD Filter," *IEEE Trans. on Aerospace and Eletronic Systems*, Vol. 48, pp. 3594–3608, 2012.

[Wil70] J. H. Wilkinson "Elementary Proof of the Wielandt-Hoffman Theorem and of its Generalization,", Tech. Rep., School of Humanities and Sciences, Stanford University, 1970.

[Zhu05] X. Zhu, Z. Ghahramani and J. Lafferty "Time-Sensitive Dirichlet Process Mixture Models,", Tech. Rep., Carnegie Mellon University, Pittsburgh, PA, 2005.

Résumé

Dans le cadre de la surveillance maritime ou terrestre par radar aéroporté, l'un des principaux objectifs est de détecter et de poursuivre une grande diversité de cibles au cours du temps. Ces traitements s'appuient généralement sur l'utilisation d'un filtre Bayésien pour estimer récursivement les paramètres cinématiques (position, vitesse et accélération) des cibles. Il repose sur la représentation dans l'espace d'état du système et plus particulièrement sur la modélisation *a priori* de l'évolution des cibles à partir d'un modèle de mouvement (mouvement rectiligne uniforme, mouvement uniformément accéléré, mouvement rotationnel, etc.). Si les cibles pistées sont manœuvrantes, plusieurs modèles de mouvement, chacun avec une dynamique prédéfinie, sont classiquement combinés au sein d'une structure à modèles multiples. Même si ces approches s'avèrent pertinentes, des améliorations peuvent être apportées à plusieurs niveaux, notamment sur la manière de sélectionner et définir *a priori* les modèles utilisés.

Dans ce contexte d'étude, plusieurs problématiques doivent être traitées.
1/ Lors de l'utilisation d'une structure à modèles multiples, on considère en général deux à trois modèles. Ce choix est fait lors de la phase de conception de l'algorithme selon la connaissance du système et l'expertise de l'utilisateur. Cependant, il n'existe pas à notre connaissance d'outils ou de règles permettant de définir les types de mouvement à associer et leurs paramètres.
2/ Il est préférable que le choix du ou des modèles de mouvement soit cohérent avec le type de cible pisté.
3/ Lorsqu'un type de mouvement est utilisé, ses paramètres sont fixés *a priori* mais ces valeurs ne sont pas nécessairement adaptées à toutes les phases du mouvement. L'une des difficultés majeures réside dans la manière de définir et de faire évoluer la matrice de covariance du bruit de modèle.

Le travail présenté dans ce mémoire vise à proposer des solutions algorithmiques aux problématiques précédentes afin d'améliorer l'estimation des trajectoires des cibles d'intérêt.
Dans un premier temps, nous établissons une mesure de dissimilarité fondée sur la divergence de Jeffrey entre deux densités de probabilité associés à deux modèles d'état différents. Celle-ci est appliquée à la comparaison de modèles de mouvement. Elle est ensuite utilisée pour comparer un ensemble de plusieurs modèles d'état. Cette étude est alors mise à profit pour proposer une méthode de sélection *a priori* des modèles constituant des algorithmes à modèles multiples.
Puis, nous présentons des modèles Bayésiens non-paramétriques (BNP) utilisant les processus de Dirichlet pour estimer les statistiques du bruit de modèle. Cette modélisation a l'avantage de pouvoir représenter des bruits multimodaux sans avoir à spécifier *a priori* le nombre de modes et

leurs caractéristiques. Deux cas sont traités. Dans le premier, on estime la matrice de précision du bruit de modèle d'un unique modèle de mouvement sans émettre d'*a priori* sur sa structure. Dans le second, nous tirons profit des formes structurelles des matrices de précision associées aux modèles de mouvement pour n'estimer qu'un nombre réduit d'hyperparamètres. Pour les deux approches, l'estimation conjointe des paramètres cinématiques de la cible et de la matrice de précision du bruit de modèle est réalisée par filtrage particulaire. Les contributions appportées sont notamment le calcul de la distribution d'importance optimale dans chacun des cas.
Enfin, nous tirons profit des méthodes dites de classification et pistage conjoints (joint tracking and classification -JTC-) pour mener simultanément la classification de la cible et l'inférence de ses paramètres. Dans ce cas, à chaque classe de cible est associé un ensemble de modèles d'évolution qui lui est propre. Afin de réaliser la classification, nous utilisons les mesures de position de la cible et les mesures d'étalement distance correspondant à la projection de la longueur de cible sur la ligne de vue radar-cible. Notons que cette démarche est appliquée dans un contexte de pistage monocible puis dans un contexte multi-cibles.

Mots-clés : pistage de cibles, approche à modèles multiples, divergence de Jeffrey, approches Bayésiennes non-paramétriques, classification et pistage conjoints (joint tracking and classification).

Bayesian approaches for surface potentially-maneuvering target tracking

Abstract

As part of the ground or maritime surveillance by using airborne radars, one of the main objectives is to detect and track a wide variety of targets over time. These treatments are generally based on Bayesian filtering to estimate recursively the kinematic parameters (position, velocity and acceleration) of the targets. It is based on the state-space representation and more particularly on the prior modeling of the target evolutions (uniform motion, uniformly accelerated motion, movement rotational, etc.). If maneuvering targets are tracked, several motion models, each with a predefined dynamic, are typically combined in a multiple-model structure. Although these approaches are relevant, improvements can be made at several levels, including how to select and define *a priori* the models to be used.

In this framework, several issues must be addressed.
1 / When using a multiple-model structure, it is generally considered two to three models. This choice is made in the algorithm design stage according to the system knowledge and the user expertise. However, it does not exist in our knowledge tools or/and rules to define the types of motions and their associated parameters.
2 / It is preferable that the choice of the motion model(s) is consistent with the type of target to be tracked.
3 / When a type of motion model is used, its parameters are fixed *a priori* but these values ??are not necessarily appropriate in all phases of the movement. One of the major challenges is the way to define the covariance matrix of the model noise and to model its evolution.

The work presented in this thesis consists of algorithmic solutions to the previous problems in order to improve the estimation of target trajectories.
First, we establish a dissimilarity measure based on Jeffrey divergence between probability densities associated with two different state models. It is applied to the comparison of motion models. It is then used to compare a set of several state models. This study is then harnessed to provide a method for selecting *a priori* models constituting multiple-model algorithms.
Then we present non-parametric Bayesian models (BNP) using the Dirichlet process to estimate model noise statistics. This model has the advantage of representing multimodal noises without specifying *a priori* the number of modes and their features. Two cases are treated. In the first one, the model noise precision matrix is estimated for a single motion model without issue of any *a priori* on its structure. In the second one, we take advantage of the structural forms of precision matrices associated to motion models to estimate only a small number of hyperparameters. For both approaches, the joint estimation of the kinematic parameters of the target and the precision matrix of the model noise is led by particle filtering. The contributions include calculating the distribution optimal importance in each case.
Finally, we take advantage of methods known as joint tracking and classification (JTC) for simultaneously leading the classification of the target and the inference of its parameters. In this case, each target class is associated with a set of evolution models. In order to achieve the classification, we use the target position measurements and the target extent measurements corresponding to the projection of the target length on the line of sight radar-target. Note that

this approach is applied in a single target tracking context and a multiple-target environment.

Keywords: target tracking, multiple-model approaches, Jeffrey divergence, non-parametric approaches, joint tracking and classification.