

THALES

THALES Air Defence
7-9, rue des Mathurins
92221 Bagneux Cedex
France



INTEGRATION TEMPORELLE POUR DES METHODES DE CLASSIFICATION RADAR

Francisco SÁNCHEZ VEGA

Responsable à *Thales* : **Marie-Christine SOUTY**
Service JRS/RBPEF

Responsable à *l'ENST* : **Florence TUPIN**
Département TSI

RAPPORT DE STAGE INGENIEUR

Paris, le 1 octobre 2004

Sommaire

1 REMERCIEMENTS	0
2 RESUME	2
3 INTRODUCTION	4
3.1 PRESENTATION DU GROUPE THALES	4
3.1.1 <i>Les origines</i>	4
3.1.2 <i>THALES aujourd'hui</i>	7
3.1.3 <i>THALES Air Systems Division et le site de Bagneux</i>	9
3.2 OBJECTIFS DU STAGE ET MISSION PROPOSEE	12
4 LA RECONNAISSANCE DE CIBLES NON-COOPERATIVE (NCTR)	14
4.1 PROBLEMATIQUE GENERALE	14
4.1.1 <i>Les techniques coopératives</i>	14
4.1.2 <i>Les techniques non coopératives : origines et justification</i>	14
4.1.3 <i>Le NCTR radar</i>	15
4.2 LA FUSION DE DONNEES POUR L'IDENTIFICATION NCTR	17
4.2.1 <i>Les systèmes multicapteurs et multiparamètres</i>	17
4.2.2 <i>L'approche bayésienne</i>	18
4.2.3 <i>La théorie de l'évidence de Dempster – Shafer</i>	19
4.2.4 <i>La théorie des ensembles flous et des possibilités</i>	21
4.2.5 <i>Un cadre commun pour des probabilités, croyances et possibilités</i>	24
4.2.6 <i>D'autres techniques pour l'intégration</i>	25
4.3 UN MODELE D'IMPLEMENTATION CONCRET BASE SUR LA LOGIQUE FLOUE	26
4.3.1 <i>Schéma général</i>	26
4.3.2 <i>Les fonctions d'appartenance et de densité de possibilité</i>	28
4.3.3 <i>Calcul de la possibilité et la nécessité</i>	29
4.3.4 <i>Intégration spatiale et intégration temporelle</i>	29
4.3.5 <i>Prise de décision</i>	30
5 DES OUTILS POUR L'EVALUATION DES PERFORMANCES	32
5.1 FONCTIONS D'EVALUATION	32
5.1.1 <i>Fonction « A »</i>	33
5.1.2 <i>Fonction « B »</i>	33
5.1.3 <i>Fonction « C »</i>	34
5.1.4 <i>Fonction taux d'erreur et taux de réussite</i>	35
5.2 MESURES SUR LES MATRICES DE CONFUSION	35
5.2.1 <i>Erreur quadratique</i>	37
5.2.2 <i>Indicateurs de précision</i>	37
5.2.3 <i>Coefficient kappa</i>	38
6 ETUDE SUR L'INTEGRATION TEMPORELLE POUR LA VARIABLE POSSIBILITE	40
6.1 DESCRIPTION DU CONTEXTE DE SIMULATION ET DES DONNEES UTILISEES	40
6.2 LES RESULTATS EN ABSENCE D'INTEGRATION TEMPORELLE	41
6.3 LES OPERATEURS DE MOYENNE	43
6.3.1 <i>La méthode « k sur M »</i>	43
6.3.2 <i>Moyennes généralisées</i>	45
6.3.3 <i>Ordered Weighted Average (OWA)</i>	48
6.3.4 <i>Conclusions sur les opérateurs de moyenne</i>	59
6.4 NORMES TRIANGULAIRES	59
6.4.1 <i>t-normes de Hamacher</i>	61
6.4.2 <i>t-norme de Frank</i>	63
6.4.3 <i>t-normes de Dombi</i>	64
6.4.4 <i>t-normes de Yager</i>	66
6.4.5 <i>Conclusions sur les t-normes triangulaires</i>	68

6.5	ESSAIS DE DOUBLE INTEGRATION	70
6.5.1	<i>Moyenne exponentielle</i>	70
6.5.2	« Double norme ».....	73
6.5.3	<i>Intégration hybride : « k sur M » + « double norme drastique »</i>	74
6.5.4	<i>Conclusions sur la double intégration</i>	77
6.6	L'APPROCHE « LEAKY BUCKET »	78
7	QUELQUES CONSIDERATIONS ET PERSPECTIVES AUTOUR DE LA VARIABLE NECESSITE	82
8	CONCLUSION	84
9	REFERENCES BIBLIOGRAPHIQUES.....	86

1 Remerciements

A Thales, je tiens à remercier ma responsable de stage madame Marie-Christine Souty de sa confiance, ses conseils, sa disponibilité et sa gentillesse pendant toute la durée de mon stage.

Je tiens aussi à remercier monsieur Roch Settineri de m'avoir proposé ce sujet de stage, de m'avoir fait confiance depuis notre première rencontre et de m'avoir bien accueilli au sein de son équipe.

Je remercie monsieur Michel Moruzzis pour ses conseils techniques et son aide avec la relecture critique de mon rapport.

Je tiens également à remercier monsieur Jean-Pierre Larvor de m'avoir accueilli au sein du service dont il assume la direction.

Merci enfin à l'ensemble des personnes du service pour leur disponibilité pour répondre à mes questions ainsi que de leur convivialité qui a contribué à rendre mon stage très agréable.
Muchas gracias a todos.

A l'ENST, je tiens à remercier ma correspondante de stage Florence Tupin de ses conseils et de sa disponibilité.

Je tiens aussi à remercier madame Isabelle Bloch de ses orientations et son amabilité lors de notre rencontre.

2 Résumé

Mon stage s'est déroulé entre les mois de juillet et septembre 2004 à *Thales Air Defence*, sur le site de Bagneux, et plus concrètement au sein de l'équipe qui travaille sur la reconnaissance non-coopérative de cibles aériennes (NCTR) appartenant au service RBPEF de *Joint Radar Sensors*. J'ai commencé par faire de la recherche bibliographique sur les systèmes NCTR et j'ai étudié un modèle concret utilisant les outils mathématiques du domaine de la logique floue et la théorie des possibilités. Dans le cadre de ce modèle, j'ai travaillé sur des différentes approches d'intérêt concernant l'intégration temporelle des données floues, l'objectif de mon stage étant l'étude et comparaison des différentes stratégies envisageables à ce propos. Des données issues de mesures radar réelles ont été mises à ma disposition. Tout d'abord, j'ai décidé d'établir une métrique variée, réaliste et cohérente pour l'évaluation des performances des divers algorithmes, à travers la définition de quelques fonctions d'évaluation, de critères concrets pour la définition des taux d'erreur et de réussite et de quelques mesures sur les matrices de confusion associées à chaque processus de classification. Par la suite, je suis parti de l'approche « k sur M » utilisée initialement à Thales et j'ai évalué ses performances pour toutes les valeurs possibles du paramètre k (pour M fixé). J'ai étudié aussi d'autres opérateurs de moyenne, notamment des moyennes généralisées et des OWA (avec plusieurs approches pour ceux-ci: distributions de poids naïves, MEOWA's et apprentissage du vecteur de poids). Ces opérateurs m'ont permis de constater l'existence d'un compromis entre l'adoption d'approches plus risquées (taux de réussite et taux d'erreur plus élevés) quand l'on se place près de l'opérateur « min » et le choix de postures plus conservatrices (moins discriminantes mais aussi avec un taux d'erreur plus petit) du côté du « max ». Les normes triangulaires m'ont permis d'aller au delà de l'opérateur « min » jusqu'à la limite imposée par la t -norme drastique, en poursuivant la tendance observée pour les opérateurs de moyenne. Après l'étude initiale de ces opérateurs, j'ai évalué l'approche consistant à effectuer une double intégration à l'aide d'une moyenne exponentielle et aussi à l'aide des normes triangulaires. Les performances observées m'ont permis de constater des différences importantes dans la nature des paramètres à intégrer, qui ont du être prises en compte afin d'améliorer le processus d'identification jusqu'à arriver à une stratégie « hybride » visant une double intégration par t -norme drastique précédée par un filtrage « k sur M » pour les paramètres que l'on estime moins fiables. Finalement, j'ai proposé un algorithme original, inspiré des méthodes « *leaky bucket* » propres au domaine de la gestion du trafic réseau, qui repose sur les principes de l'intégration « hybride » et qui nous permet de profiter de ses avantages, en ajoutant une plus grande simplicité, une meilleure versatilité et une optimisation des ressources en termes de temps de calcul et utilisation de mémoire.

Non-cooperative air target recognition (NCTR) represents an emerging field of research in which fuzzy logic and possibility theory are likely to play a major role in the near future. During my internship at the Joint Radar Sensors unit from Thales Air Defence in Bagneux (near Paris), a concrete model for a NCTR system circumscribed to this theoretical framework was taken into account following an initial stage of bibliographic research, and performances of several strategies for fuzzy data time integration were studied and compared. Data from real radar measures were used for all our integration simulations. Only possibility measures were considered for the analysis, mainly due to the limitations imposed by a short-period internship. First of all, a set of tools for measuring algorithm performances was proposed and defined with the aim of providing a varied, realistic and coherent approach for the evaluation of the different fusion methods later to be considered. These tools included several evaluation functions and some specific criteria for the definition of the main success and error (misclassification) rates, as well as the selection of a certain number of concrete measures intended to evaluate the quality of the recognition process through the use of confusion matrixes. " k th out of M " approach for time integration was the first to be evaluated, and the performances obtained for all possible values of k were measured. After that, other averaging operators were studied, namely generalised means and OWA (including several different implementations of this last concept, such as intuitive weights distributions, Maximum Entropy OWAs and weights vector learning). The results obtained showed a compromise between the choice of higher risk operators (with higher values for both misidentification and success rates), the level of risk apparently increasing towards the "min" operator, and the adoption of other operators closer to "max", which resulted in more conservative strategies (their discrimination power being lower, they actually presented reduced values for both misclassification and success rates). Triangular norms allowed me to go beyond the "min" operator frontier and reach the boundary imposed by the so-called "drastic t -norm", all the while observing a continuity in the behaviour and properties of the operators under study, as well as a correlation in their performances. After that, I decided to consider the possibility of performing a double integration through the use of an exponential time average first, and a t -norm afterwards, for the fusion of the newly integrated values with those obtained in the past and in order to add some sort of recursive "memory" or "persistence" to our system. This new approach pointed out the intrinsic differences between the different parameters considered and suggested the convenience of taking these particularities into account so as to improve the general identification performances. This led up to the proposal of a hybrid integration strategy consisting of an double drastic t -norm preceded by a " k th out of M " error filter and so that the latter should be adjusted for each parameter as a function of the level of robustness and reliability associated to the parameter itself. Finally, an original algorithm inspired of the "leaky bucket" techniques, which are commonly used in other domains such as that of network traffic management, was proposed. This approach followed the principles of hybrid integration and allows us to enjoy its advantages, adding a greater simplicity, more versatility and a most welcomed optimisation in terms of computational resources.

3 Introduction

Le contenu de ce rapport de stage a été structuré en quatre sections principales, au delà de celles correspondant au résumé, les conclusions, les remerciements et les références bibliographiques :

- On peut trouver une première section d'introduction, dans laquelle on fera une brève description du groupe Thales et puis de Thales Air Défense jusqu'à arriver au niveau de la section dans laquelle j'ai travaillé pendant ces trois derniers mois. On retrouvera ici aussi une description des objectifs initiaux et la mission proposée définissant mon sujet de stage.
- Ensuite, on présentera le cadre théorique dans lequel se place le stage, notamment celui de la fusion de données appliquée aux techniques NCTR. On présentera ici le modèle concret pour le système de reconnaissance avec lequel on travaillera dans la suite. Ce chapitre constitue la synthèse du travail de recherche bibliographique réalisé pendant les premières semaines de mon stage et a comme but primordial de fournir une perspective générale sur le domaine de référence considéré.
- Une troisième partie présentera les outils mathématiques choisis pour l'évaluation et la comparaison des performances des différents algorithmes que l'on considérera pour notre étude sur l'intégration temporelle. On combinera des mesures traditionnelles sur les matrices de confusion avec des fonctions d'évaluation dont j'ai développé le concept, de nature intuitive ou suivant les observations faites par les professionnels de Thales, et toujours considérant les besoins spécifiques relatifs au problème à traiter.
- Finalement, on trouvera la partie la plus importante du rapport, au cours de laquelle on utilise les outils dont on vient de parler pour réaliser une étude sur les différentes approches possibles pour l'intégration temporelle de la variable possibilité. On considérera quelques opérateurs plutôt « classiques » et on arrivera à proposer un algorithme d'intégration original qui semble apporter de nombreux avantages par rapport aux autres.

3.1 PRESENTATION DU GROUPE THALES¹

3.1.1 Les origines

3.1.1.1 La création de Thomson-CSF

Thomson-CSF est issue de la fusion, en **1968**, des activités d'électronique professionnelle de Thomson-Brandt avec la Compagnie Générale de Télégraphie Sans Fil (C.S.F.).

Les origines de **Thomson-Brandt** remontent à 1893, avec la création de la Compagnie Française Thomson-Houston, chargée d'exploiter en France les brevets de la Thomson-Houston Electric Corporation, dans le domaine, alors émergent, de la production et du transport de l'électricité.

Créée en 1918, la **C.S.F.** a été, avec sa filiale Société Française Radioélectrique (la S.F.R. absorbée en 1957), l'un des pionniers dans le domaine des transmissions hertziennes. Les deux sociétés ont joué un rôle particulièrement important, avant la Seconde Guerre Mondiale, dans le développement de la radiodiffusion, des radiocommunications sur ondes courtes, de l'électro-acoustique et déjà du radar et de la télévision.

3.1.1.2 Les années 70

Après les deux chocs pétroliers de 1973 et 1979, Thomson-CSF conclut ses premiers **grands contrats à l'exportation** avec des pays du Moyen-Orient, notamment pour la vente de ses systèmes de défense anti-aériens Shahine puis Crotale.

¹ sources : [23], [24]

Cette décennie sera également marquée par la **diversification** des activités, avec principalement le développement accéléré de la commutation téléphonique, tout d'abord par croissance interne puis, en 1976, avec l'acquisition des filiales françaises de l'Américain ITT (Le Matériel Téléphonique - LMT - et la société Lignes Télégraphiques et Téléphoniques - LTT) et du Suédois Ericsson (Société Française des Téléphones Ericsson). La société participe, également sous l'égide des pouvoirs publics, au développement en France de l'industrie des semi-conducteurs silicium. Et en 1979, elle reprend à sa société mère Thomson-Brandt la société d'imagerie médicale CGR (Compagnie Générale de Radiologie).

3.1.1.3 La nationalisation et le recentrage stratégique

A la **nationalisation, en février 1982**, de Thomson-Brandt, la situation financière de Thomson-CSF est fortement dégradée : à l'issue de la croissance accélérée des années 70, le portefeuille d'activités, très diversifié, inclut de nombreux domaines où la taille et les parts de marché, et donc la rentabilité, de Thomson-CSF sont insuffisantes. Et, malgré les ressources perçues au titre des premiers grands contrats avec des pays du Golfe, l'endettement s'est fortement accru. Les exercices 1982 et 1983 se soldent par de lourdes pertes et à fin 1983, les fonds propres sont négatifs de 337 MF et l'endettement financier net approche les 7 milliards de francs.

Pour assurer le redressement de la situation financière, la nouvelle direction engage rapidement la **restructuration du portefeuille d'activités** et décide de le focaliser autour du métier cœur, l'électronique de défense. Outre les sorties d'activités périphériques, les désinvestissements majeurs réalisés à partir de 1982 concernent les télécommunications civiles, au travers d'un important accord conclu en 1983 avec la C.G.E., et le médical, cédé à General Electric en 1987 dans le cadre de l'accord organisant également la reprise, par Thomson S.A., des activités d'électronique grand public de la firme américaine.

Les **activités de semi-conducteurs**, elles aussi très déficitaires au début de la décennie 80 en raison de leur taille "sous-critique", feront, dans un premier temps, l'objet d'un **développement interne vigoureux**, complété en 1985 par l'acquisition, à United Technologies, de sa filiale Mostek. Mais la montée en puissance des concurrents japonais rendant de plus en plus difficile l'accession par croissance interne seule à une part de marché suffisante, Thomson-CSF et le groupe public italien IRI-Finmeccanica décidaient en 1987 de fusionner leurs activités. La joint venture SGS-THOMSON- issue de cette fusion se plaçait dès sa création au deuxième rang des fabricants européens de semi-conducteurs, derrière Philips.

Outre l'impact, très favorable sur la rentabilité opérationnelle, de la politique de concentration du portefeuille d'activités, le redressement des résultats a également bénéficié de l'apport très positif des **activités financières** développées en interne à partir de 1984 pour gérer la trésorerie générée par d'importants contrats conclus à l'exportation. Ces activités financières, regroupées en 1987 au sein de Thomson-CSF Finance, feront l'objet d'un accord conclu à la fin de 1989 avec le Crédit Lyonnais et organisant dès 1990, leur reprise progressive par la Banque, en échange d'une participation de Thomson-CSF au capital de celle-ci. La sortie définitive de Thomson-CSF Finance (rebaptisé Altus) sera effective en juin 1993.

3.1.1.4 La politique de croissance externe

A partir de la **fin des années 80** et de la chute du Mur de Berlin, l'environnement des industries de la défense est profondément et durablement altéré. Au déclin des budgets dans la quasi-totalité des pays occidentaux - la France faisant alors figure d'exception - s'ajoute la contraction de la demande dans les pays du Moyen-Orient. La demande solvable émergente, identifiée dans la région du Sud-Est Asiatique, est encore insuffisante pour compenser cette double contraction.

Dès 1987, Thomson-CSF, anticipant la baisse inéluctable des budgets et en prévision de l'achèvement des grands contrats-export alors en cours de facturation, entame la **restructuration** en profondeur de ses activités : allègement et adaptation des structures, diminution des effectifs, concentration des sites industriels et des moyens de recherche et développement, réorganisation des réseaux commerciaux internationaux, actions pour réduire le coût des approvisionnements,... Les frais de restructuration cumulés sur la période 1987-1995 dépassent 6,5 milliards de francs, un montant qui représente presque 2% du chiffre d'affaires de la période.

Parallèlement, une politique active de **croissance externe** (acquisitions et alliances au sein de joint-ventures) est menée, principalement en Europe, qui permet d'atténuer le déclin des ventes. L'addition des chiffres d'affaires des sociétés entrées dans le périmètre de consolidation depuis 1990 s'élève à **5 milliards de francs**. L'acquisition la plus importante de la période a été la reprise, finalisée en décembre 1989, des activités d'électronique de défense du groupe Philips, à savoir les sociétés hollandaise Signaal, belge MBLE et française TRT (dénominations sociales modifiées depuis, pour les deux dernières), qui a apporté un chiffre d'affaires de 4,5 milliards de francs en 1990.

Au travers de ces acquisitions, Thomson-CSF a tout d'abord acquis ou consolidé sa présence sur l'ensemble des segments de l'électronique de défense et étoffé son très **large spectre**, désormais comparable à celui de ses grands concurrents américains. Mais la société a également accru la part de ses activités d'**électronique professionnelle civile** : sur les 15 milliards acquis, environ 11 milliards concernent l'électronique de défense et 4 milliards les applications civiles.

Enfin, elle a, en quelques années, développé sa présence industrielle hors des frontières nationales : en 1989, 95 % du chiffre d'affaires consolidé provenait des implantations françaises ; six ans plus tard, en 1995, les **filiales étrangères** ont contribué pour plus de 20 % aux ventes consolidées.

En 1995, pour renforcer la performance de son organisation opérationnelle, Thomson-CSF achève le processus de **filialisation** initié pour partie quelques années plus tôt.

Entre 1996 et 1997, la cession des participations dans le **Crédit Lyonnais** et dans **SGS-Thomson** (devenu **ST Microelectronics**) génèrent des ressources qui permettent de financer la poursuite du développement international à l'ensemble des domaines de la sécurité civile.

3.1.1.5 La privatisation et l'accord de coopération

L'**année 1996** marque le début du processus de **privatisation**. Après avoir annoncé en février 1996 sa volonté de procéder à la privatisation du Groupe, le gouvernement décide en décembre de privatiser Thomson-CSF selon une procédure distincte de celle de Thomson multimédia. En février 1997, il opte pour une privatisation selon une procédure de gré à gré avec cahiers des charges, procédure qui sera abandonnée suite à la dissolution du gouvernement.

C'est finalement en **octobre 1997** que le nouveau gouvernement décide de regrouper autour de Thomson-CSF, dans le cadre d'un **partenariat stratégique avec Alcatel**, les activités d'électronique spatiale et de défense et les activités de communications militaires d'Alcatel, les activités d'électronique professionnelle et de défense de **Dassault Electronique**, ainsi que les activités satellites d'**Aérospatiale**.

Le 14 avril 1998 les sociétés Aérospatiale, Alcatel, Dassault Industries, Thomson SA et Thomson-CSF **signent l'Accord de Coopération** qui définit les modalités de mise en œuvre des orientations décidées par le gouvernement. Enfin, les 22 et 23 juin 1998, l'Assemblée générale des actionnaires et le Conseil d'administration respectivement approuvent les opérations d'apports. Désormais, la majorité du capital de Thomson-CSF, soit 53,06 %, est détenue par des **actionnaires du secteur privé**.

3.1.1.6 Dualité et Multidomesticité (Années 1999-2004).

La dynamique d'implantation "multidomestique" des activités de défense, poursuivie tout au long de la décennie 90, dépasse les frontières du Continent européen : Afrique du Sud, Australie, Corée, Singapour...

En juin 2000, l'acquisition par OPA de la société britannique **Racal Electronics** fait du Royaume-Uni le deuxième pays d'implantation; les activités de défense et les technologies de l'information et services (IT&S) en sont renforcées.

Le développement du groupe, par croissance interne et par acquisitions, a aussi profondément modifié le spectre de ses activités : une réflexion stratégique met en évidence la part croissante des applications civiles et la "dualité" des technologies qui font la force du groupe. En juillet 2000, une nouvelle organisation en trois pôles est mise en place, en cohérence avec la nouvelle approche stratégique. En

décembre 2000 Thomson-CSF, devenu **Thales**, annonce la création avec l'Américain **Raytheon** de la première joint-venture transatlantique entre industriels de la défense.

En 2001, il y a une poursuite du recentrage des activités : Thales prend le contrôle total de plusieurs de ses joint-ventures en défense et en aéronautique, et se retire d'Alcatel Space; en IT&S, priorité est donnée aux activités synergiques et en forte croissance, principalement le positionnement par satellite et les opérations sécurisées. La structure de l'actionariat évolue : le flottant frôle les 40%, tandis qu'Alcatel se désengage partiellement et que l'Etat français passe sous la barre des 33%, au profit de l'actionariat salarié.

De 2001 à 2003, dans un contexte géopolitique et économique profondément bouleversé à la suite des attentats du 11 septembre 2001, poursuite du recentrage des activités civiles et confirmation de la stratégie multi-domestique : Thales prend le contrôle total de plusieurs de ses joint-ventures en défense et en aéronautique, et se retire d'Alcatel Space ; Thales Defence Ltd, devenue la deuxième société de défense britannique, est retenue dans plusieurs programmes ou appels d'offres prioritaires du ministère de la Défense du Royaume. En IT & S, les désinvestissements sont poursuivis dans les secteurs non stratégiques, dont certains sont fortement déficitaires, tandis que les activités liées au domaine de la sécurité civile prennent de plus en plus d'importance. Fin 2003, Thales lance une initiative majeure, Thales SHIELD, marquant l'engagement du groupe dans les systèmes et technologies qui s'appliquent à l'ensemble des domaines de la sécurité civile.

3.1.2 THALES aujourd'hui

Aujourd'hui, les activités de Thales relèvent de l'électronique et s'adressent essentiellement à des marchés professionnels, qu'il s'agisse d'équipements, de systèmes ou des services associés. Elles requièrent la mise en œuvre de technologies complexes, dont une partie importante est développée en interne par les ingénieurs de Thales.

Les marchés de **Défense** représentent leur principale destination, avec environ 70% du chiffre d'affaires consolidé réalisé en 2003. L'offre de Thales pour ses clients militaires s'adresse à tous les segments, l'Air, la Terre, la Mer, ainsi qu'à la défense aérienne, et aux forces inter-armées. Elle englobe une très large gamme d'équipements et de systèmes et s'oriente de plus en plus vers des "solutions globales" de "systèmes de systèmes", de maîtrise d'œuvre et de services, sur toute la durée de vie des équipements ou des plates-formes. Cette évolution reflète le mouvement, observé dans de nombreux pays, de partage croissant de la responsabilité des grands programmes entre les industriels et leurs prescripteurs, voire de transfert de l'ensemble de la responsabilité à un groupe maître d'œuvre de la plate-forme, dans son ensemble.

Les activités du groupe trouvent aussi de plus en plus de débouchés dans le domaine de la **sécurité collective et civile**, un domaine dont la croissance est largement liée au nouveau contexte géopolitique, marqué par la montée de l'insécurité. La sécurité collective et civile recouvre de nombreux champs d'application pour les compétences de Thales : si la sécurité du transport aérien est l'un de ses métiers traditionnels en Aéronautique (avionique, contrôle du trafic aérien et entraînement de pilotes), le groupe est aussi présent sur l'ensemble des marchés de la sécurité civile, au travers de ses activités IT&S(2) : Thales est un acteur reconnu notamment dans les domaines de la sécurité des transports terrestres, des réseaux informatiques et des transactions électroniques ainsi que dans la protection des sites et des personnes.

Les activités de Thales partagent un socle technologique commun essentiel, dont la maîtrise de l'information, et sa diffusion en temps réel, sont l'un des enjeux majeurs : il s'agit de l'ensemble des technologies de collecte des informations (capteurs de toutes natures), de transmission et de traitement sécurisés de toutes les formes d'informations ainsi collectées (phonie, messagerie, images), avec pour principale évolution technologique de ces dernières années la numérisation.

La deuxième caractéristique des activités du groupe est leur caractère très international : en termes de marchés - elles s'adressent pour la plupart à des marchés globaux - comme en termes d'implantation géographique – les effectifs en France font désormais à peine plus de la moitié (55%) des effectifs du

groupe contre plus de 95 % à la fin des années 1980. Cette évolution résulte de la politique de croissance externe hors du territoire national, conduite depuis plus de dix ans. Dans le domaine de la défense en particulier, où les clients sont des Etats nationaux, l'approche "multidomestique" s'affirme comme un axe stratégique majeur du groupe : disposer d'une implantation industrielle locale permet en effet d'apporter au client une réponse adaptée à ses attentes, et de réaliser sur place une partie importante des développements spécifiques qu'il demande, tout en lui assurant le niveau technologique de premier plan d'un leader mondial.

3.1.2.1 Chiffres clés

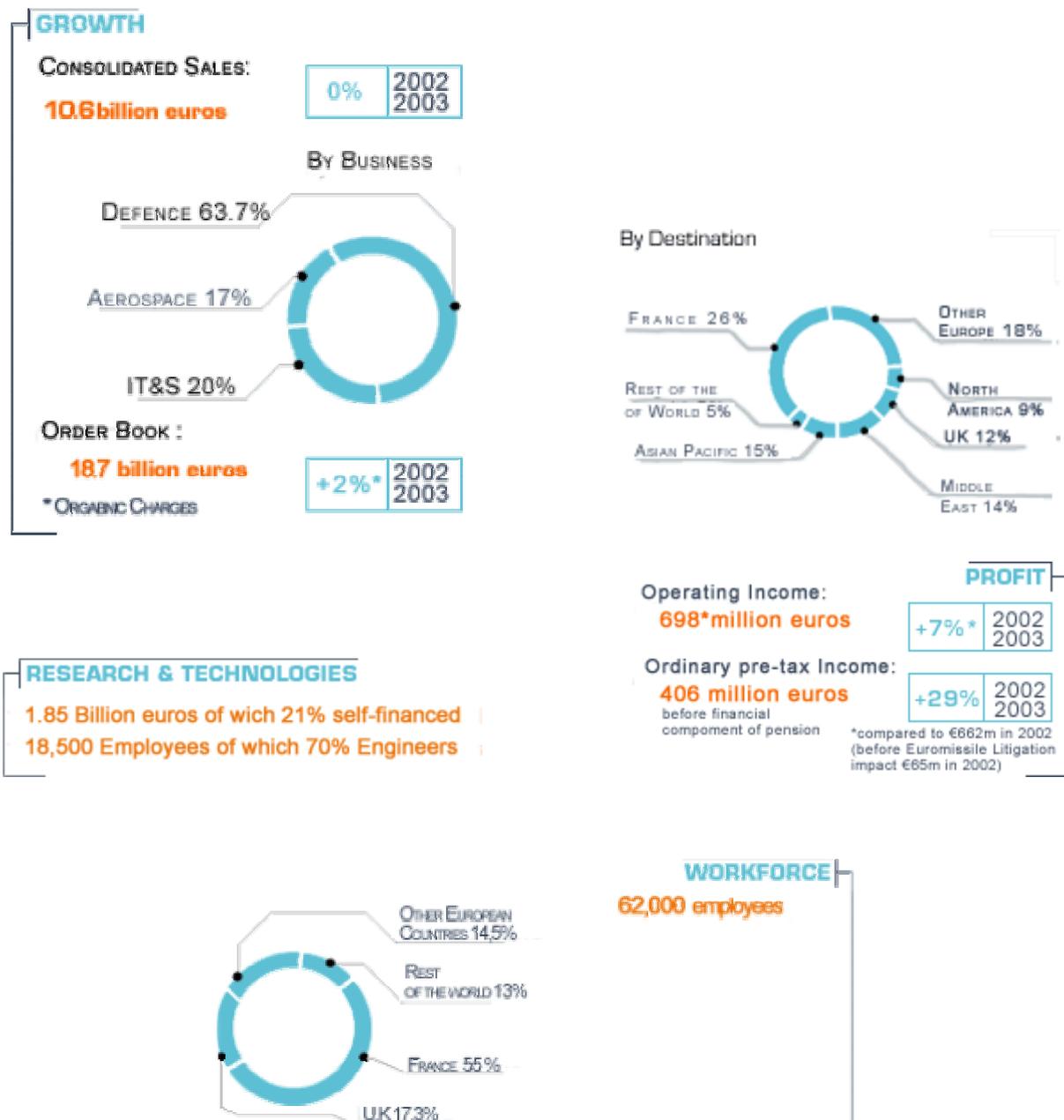


Figure 1. Chiffres clés pour le groupe Thales [23]

3.1.2.2 L'organisation interne du groupe Thales

Thales a été restructuré en Juillet 2004. L'organisation est passée d'une structure en pôles (Aéronautique, Défense, IT&Services) et « *Business groups* » à une structure en six divisions.

La nouvelle structure se construit selon le schéma suivant :

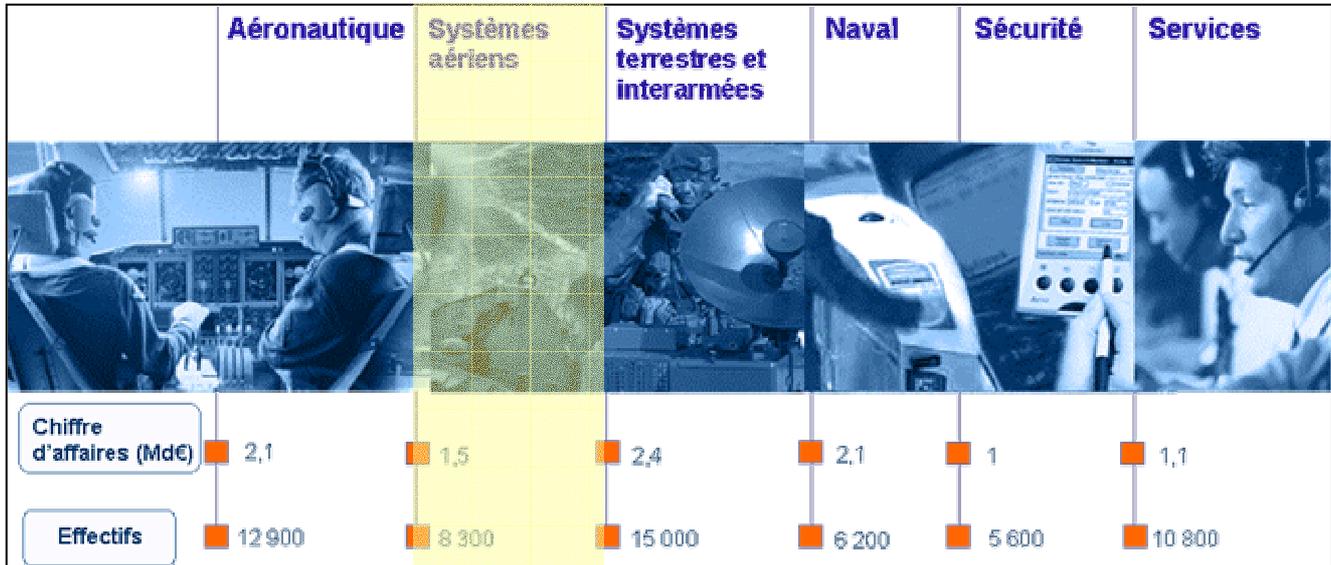


Figure 2. Organisation générale du groupe Thales

3.1.3 THALES Air Systems Division et le site de Bagnoux

L'ancienne entité appelée *Thales Air Defence* est récemment disparue du point de vue organisationnel et ses activités *Joint Radar Sensors* ont été regroupées dans la division Systèmes Aériens.

Thales est l'un des trois acteurs mondiaux capables d'assurer la maîtrise d'œuvre complète et l'architecture globale des systèmes de défense aérienne à l'échelle d'un pays. Son offre s'étend des systèmes de détection, d'identification et de conduite des opérations aériennes (défensives et offensives), à la protection du champ de bataille et des sites sensibles, par des systèmes d'armes sol/air très courte, courte et moyenne portées, ainsi qu'aux domaines de l'armement et de la propulsion.

Les deux leaders mondiaux sont les industriels américains Raytheon et Lockheed-Martin. Les autres intervenants Européens sont MBDA, pour les systèmes de missiles sol-air, Finmeccanica et BAE Systems pour les systèmes de défense aérienne et de détection sur le champ de bataille.

Le marché de la défense aérienne comprend les systèmes de commandement et de contrôle des opérations aériennes ainsi que les systèmes de défense aérienne du champ de bataille. Ces systèmes permettent l'engagement des forces défensives et préparent les missions offensives sur un territoire donné. Les récents conflits militaires ont mis en évidence que les menaces aériennes ne venaient plus seulement des avions et des missiles de croisière mais aussi des drones et des missiles balistiques tactiques. Ils ont aussi montré que la mobilité et l'aéro-transportabilité des systèmes sont devenus des impératifs pour les forces armées qui doivent intervenir "en projection" sur les théâtres d'opérations.

En outre, pour mener à bien des missions de maintien de la paix ou de résolution de conflits dans un contexte multinational, les forces alliées doivent désormais disposer de systèmes interopérables, de connexions entre les senseurs, les systèmes d'armes et les centres de commande et de contrôle, dans un environnement où la maîtrise des systèmes d'information et de communication est fondamentale.

Depuis dix ans, Thales a pris en compte ce nouvel environnement et conçoit et réalise des systèmes de défense interopérables. Dans le domaine des systèmes de commandement et de contrôle des opérations

aériennes, Thales est notamment impliqué dans les besoins de renouvellement de l'ensemble des infrastructures de l'OTAN (interopérabilité des systèmes alliés à l'échelle continentale et des forces projetées) via sa filiale commune avec Raytheon, la société Thales Raytheon Systems (TRS).

Après son implication dans des grands projets à niveau européen et américain, dont ceux associées aux technologies M3R (Radar Mobile, Multifonctions et Modulaire) font un bon exemple, Thales a poursuivi en 2003 le développement de nouveaux programmes comme Rapsodie (futur radar de surveillance au sol de l'armée française), Cobra (radar de contrebatterie, réalisé dans le cadre du GIE EuroArt) et Martha, premier système au monde capable de piloter simultanément des systèmes d'armes très courte, courte et moyenne portées, développé pour l'armée de terre française depuis 2000.

3.1.3.1 Organigramme général

L'organigramme du groupe Thales, spécifié jusqu'au niveau du service et de la section au sein desquels j'ai effectué mon stage, est le suivant :

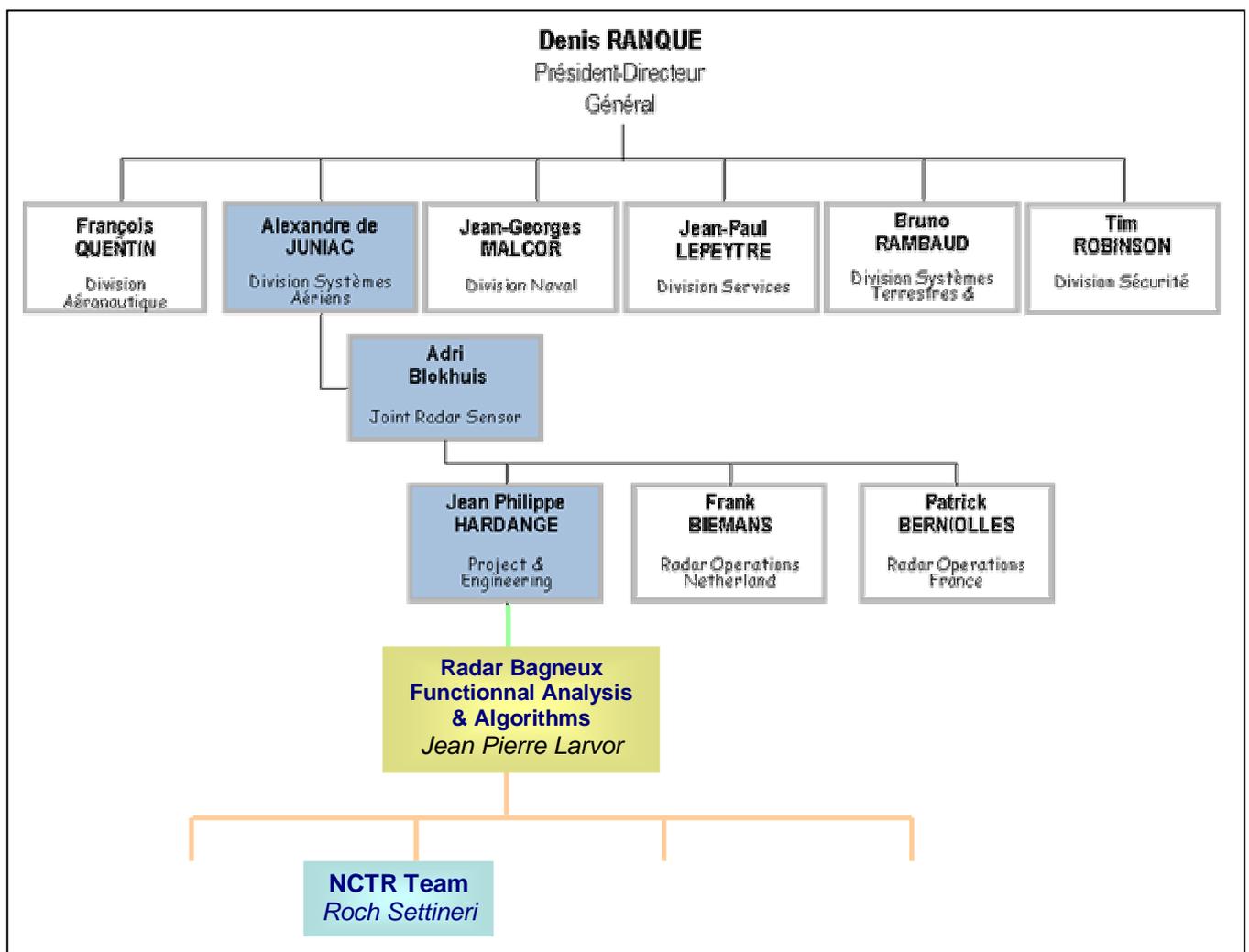


Figure 3. Organigramme de Thales – Division Systèmes Aériens

La **Division de Systèmes Aériens** du groupe Thales a en charge les domaines de la gestion du trafic aérien, la défense aérienne et d'autres projets comme le *Galileo* européen. Elle regroupe plusieurs directions opérationnelles dont *Joint Radar Sensors*.

Au sein du département Défense Aérienne, la direction opérationnelle **Joint Radar Sensor (JRS)** a la particularité d'être un regroupement entre plusieurs services de l'ancien *Thales Air Defence* (TAD) et de *Thales Naval Nederland* (TNN). Ses missions incluent la conception et fabrication de systèmes radar pour les marchés civils et militaires, et notamment pour la défense aérienne, le domaine naval, la gestion du trafic aérien, la surveillance du sol et la localisation d'armes .

Project & Engineering (P&E) est un département de *Joint Radar Sensors* (JRS). Son rôle est de définir les nouveaux produits et de réaliser des études amont. P&E est également responsable de l'architecture radar, et traite un projet jusqu'à la signature du contrat. C'est ensuite le département *Radar Operations* qui réalise le programme.

3.1.3.2 Le site de Bagneux

Près de 4000 personnes travaillent sur le site de Bagneux, avec environ 1300 personnes chez l'ancien TAD. Les autres branches de Thales présentes sur le site sont Thales ATM (*Air Traffic Management*) et TNF (*Thales Naval France*).

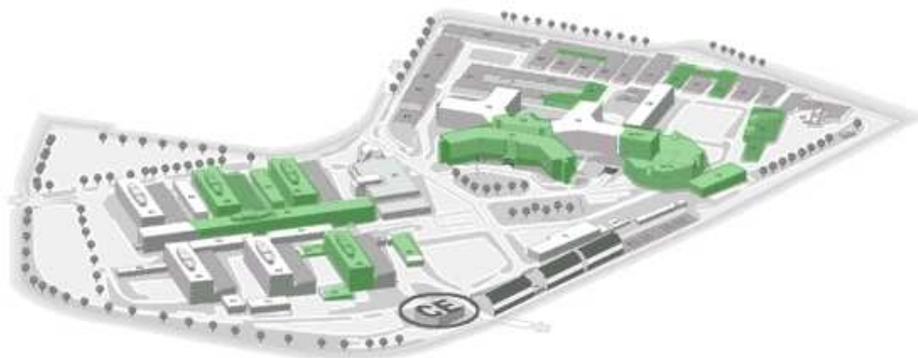


Figure 4. TAD (en vert) sur le site de Thales à Bagneux

La partie du service P&E sur le site de Bagneux ajoute les lettres RB (Radar Bagneux) à sa dénomination (RBPE). **Functional Analysis & Algorithms (RBPEF)**, est donc le service de P&E à Bagneux qui participe à la définition des traitements radar, à la réalisation de simulations fonctionnelles et à la conduite d'études amonts. Il compte une trentaine de personnes.

La section Traitements **Non-Cooperative Target Recognition (NCTR)**, est une section du service RBPEF qui travaille sur le sujet de l'identification de cibles aériennes non coopérative . Ils travaillent sur des études portant sur des radars futurs qui sont en cours de développement à Thales, mais aussi sur des radars existants pour lesquels le client demande une fonction NCTR.

3.2 OBJECTIFS DU STAGE ET MISSION PROPOSEE

Mon stage, d'une durée de trois mois, s'est passé au sein de l'équipe Traitements NCTR chez Thales à Bagnaux.

Comme on verra plus en détail dans les sections suivantes, l'équipe Traitements NCTR travaille sur un système d'identification radar pour des cibles aériennes qui s'appuie sur les principes mathématiques de la logique floue. Il s'agit d'un système de reconnaissance monocapteur radar et multiparamètres qui évolue en temps réel et cela implique la nécessité de fusionner des données issues de sources multiples (les différents paramètres) à des instants variés, de sorte que ces données puissent contribuer à déterminer l'identité de la cible en question d'une manière efficace.

L'intégration spatiale restant hors le cadre de mon stage, mon objectif principal était celui de travailler sur l'intégration temporelle des données floues utilisées pour le processus d'identification, et plus concrètement :

- Me renseigner à propos des possibles approches dans le domaine de la fusion de données susceptibles d'être adoptées pour le cas considéré.
- Etudier l'adéquation de ces approches aux caractéristiques du problème traité et leur influence sur les performances du système utilisé chez Thales.
- Réaliser des simulations Matlab sur des données procédant de mesures radar réelles afin d'évaluer « sur le terrain » le comportement des différents opérateurs d'intégration.
- Faire une comparaison, sur des critères rigoureux et objectifs, des performances des différents algorithmes envisageables pour l'intégration temporelle à partir des résultats de ces simulations et en déduire des conclusions et des recommandations pour le futur.

4 La reconnaissance de cibles non-cooperative (NCTR)

Reconnaissance, classification et identification sont des concepts assez proches dans leur signification, qu'il faudra tout d'abord bien distinguer afin d'éviter des possibles malentendus dans la suite.

En effet, on appelle classification la répartition systématique en classes, en catégories d'êtres, de choses ou de notions ayant des caractères communs. L'identification, par contre, est une discrimination faite entre plusieurs objets appartenant à une même classe déterminée. Le terme reconnaissance désigne tout simplement l'action de poser un certain élément comme déjà connu.

Les systèmes de reconnaissance dont on parlera dans ce rapport sont, pour le moment, des systèmes de classification capables de faire une distinction entre des types, des modèles différents de cibles aériennes.

4.1 PROBLEMATIQUE GENERALE

La reconnaissance de cibles aériennes trouve des applications très importantes dans les domaines civil et militaire. Du point de vue de la sécurité du trafic aérien, les contrôleurs ont besoin de connaître à chaque instant le statut et l'identité des avions présents dans l'espace aérien local. Dans le domaine militaire, l'identification fiable des cibles aériennes a une relevance évidente puisqu'elle peut permettre de faire la distinction entre un avion ami et un objectif ennemi.

Traditionnellement, on trouve deux grandes approches pour la classification de cibles aériennes : celle des cibles amies faite à travers de méthodes coopératives et celle des cibles inconnues ou hostiles dont on n'envisage pas d'obtenir aucun type de coopération active.

4.1.1 Les techniques coopératives

Dans le premier groupe, on trouve les techniques d'« **Identification Friend or Foe** » (**IFF**), qui ont été utilisées depuis longtemps dans le domaine militaire. Le processus d'identification IFF commence avec l'envoi d'un signal d'interrogation vers la cible à identifier. Les avions considérés « amis » sont censés avoir un transpondeur qui répond avec un code d'identification particulier, connu par l'interrogateur. La réponse peut être plus ou moins complexe, selon le mode d'opération IFF, de sorte que le transpondeur peut inclure des informations telles que l'altitude actuelle de l'avion ou être cryptée, par exemple. En général, une cible hostile ne sera pas capable d'émettre une réponse valide par manque d'un transpondeur approprié et sera donc considéré « hostile » ou « non amie ».

Dans le domaine de l'aviation civile, on utilise une technique similaire à l'IFF ; la **Surveillance Secondaire Radar (SSR)**. Bien que la mission principale de la SSR soit celle de fournir des informations permettant la localisation des avions civils, elle apporte aussi des informations concernant l'identité de l'avion en question. En principe, tous les avions civils portent des transpondeurs SSR et donc on pourrait supposer qu'ils devraient être facilement identifiables, mais il faut faire attention à ce type de suppositions. Les attentats du 11 Septembre 2001 constituent un exemple d'une situation dans laquelle il semble que les terroristes ont désactivé les transpondeurs SSR pour empêcher – ou au moins essayer de retarder - la détection de la modification des trajectoires des avions kidnappés [9].

4.1.2 Les techniques non coopératives : origines et justification

Cet exemple montre l'inconvénient principal des techniques de reconnaissance coopérative: en effet, la coopération n'est pas toujours garantie, que ce soit par des problèmes techniques, par des erreurs humaines ou par des intentions hostiles de la part de la cible à identifier. Les **techniques de reconnaissance non coopérative (NCTR)** ont pour but de surmonter ces obstacles en s'appuyant sur des informations obtenues par des capteurs propres pour découvrir la vraie identité de l'objectif.

D'après quelques sources non – officielles [27], les origines des techniques NCTR pourraient se trouver au milieu des années 1970, avec le programme militaire américain 'Musketeer', où l'on a essayé d'utiliser le radar pour obtenir une signature unique de chaque modèle d'avion ennemi à partir du nombre et des caractéristiques techniques de ses moteurs. La nouvelle technologie a dû attendre une dizaine d'années pour que les moyens informatiques nécessaires se soient suffisamment développés. Depuis 1985, les F-15C de l'USAF incorporent déjà des modes NCTR pour le combat. Pendant la première Guerre du Golfe, entre 1990 et 1991, les ROE (*Rules of Engagement* ou « Règles d'engagement ») de l'USAF spécifiaient la nécessité d'une double identification des cibles de la part des chasseurs américains avant de lancer une attaque. Une de ces identifications pouvait être fournie par les AWACS, l'autre étant de type NCTR. En plus, ces techniques rendaient possible l'utilisation de missiles *air-to-air* de type BVR (*Beyond Visual Range*, ou « au delà du champ visuel »), qui pouvaient être utilisés à des longues distances avec la sûreté de ne pas détruire des cibles amies. Pour beaucoup d'analystes, c'est grâce aux techniques NCTR que les F-15C furent les seuls chasseurs de la coalition à avoir pu opérer sur l'Irak, au détriment d'autres modèles comme les F-14 de l'époque. Au cours des dix années suivant le premier conflit du Golfe, le NCTR a été incorporé à d'autres chasseurs modernes, notamment aux F-14, F-16 et F-18 américains, quelques F-15 vendus à des gouvernements étrangers par les USA, aux Tornado britanniques et aux Mirage français.

Parmi les atouts principaux de l'approche NCTR on trouve :

- Dans le domaine civil, une protection face à des possibles problèmes techniques ou humains traditionnellement associés aux techniques coopératives.
- Dans le domaine militaire,
 - Une meilleure compréhension de la situation tactique et une meilleure estimation des possibles menaces en combat, ce qui peut rendre possible, par exemple, une réponse stratégique hiérarchisée face aux cibles identifiées.
 - Une réduction du risque de dommage fratricide.
 - En termes défensifs, une réaction plus efficace face à une menace réelle (réduction du temps de réaction, possibilité de choisir les systèmes d'armement en fonction de la cible détectée, etc.).
 - En termes d'attaque, l'utilisation des techniques de type BVR comme on a déjà commenté, ce qui implique une diminution considérable du risque pris par rapport à une situation où l'on est obligé d'établir un contact visuel.
 - En termes de surveillance, ces techniques peuvent être utilisées pour optimiser le processus général de détection, identification, localisation et pistage, en adaptant les moyens utilisés (par exemple la forme d'onde) au type de cible identifiée en première instance.

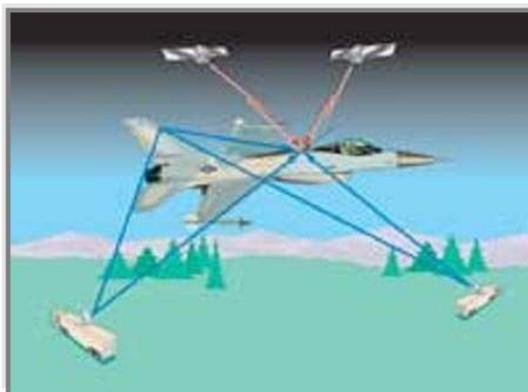


Figure 5. Scénario NCTR multicapteur [26]

4.1.3 Le NCTR radar

D'un point de vue technique, la reconnaissance non coopérative peut être abordée par différentes approches : capteurs radar, ESM (*Electronic Support Measures*), acoustiques, infra-rouges, optroniques, etc. L'approche radar est très attractive parce qu'elle peut être utilisée pour des distances de plusieurs centaines de kilomètres, indifféremment pendant le jour ou la nuit, et elle est relativement insensible aux variations des conditions climatologiques.

Le radar est capable de fournir plusieurs paramètres physiques associés à la cible considérée, qui peuvent être classifiés, par exemple, dans trois domaines :

- **Cinématique.** Paramètres tels que l'altitude, la vitesse, l'accélération, la vitesse ascensionnelle, l'orientation, etc. Elles peuvent être utiles pour faire la distinction entre des grands groupes de cibles (missiles, hélicoptères, chasseurs, etc.).
- **Haute Résolution Distance (HRD).** Paramètres telles que la longueur, la Surface Equivalente Radar (SER ou RCS en anglais), le profil Distance, etc. Elles peuvent apporter des informations plus spécifiques permettant de distinguer entre des cibles grandes et petites, par exemple.
- **Doppler.** L'analyse des données procédant des parties mobiles, qui seront la cause de singularités sur le spectre mesuré, peut nous fournir des informations très utiles pour l'identification. Les niveaux à la sortie de chacun des filtres Doppler utilisés permettent de caractériser le spectre de la cible.

Dans leur article de 1998, Delhotte et Moruzzis [5] proposent un modèle schématique pour les systèmes de reconnaissance radar NCTR qui se structure sur 4 points :

1. Mesure des caractéristiques physiques de la cible à identifier en utilisant des formes d'onde appropriées et des techniques typiques dans le domaine des mesures radar (Doppler, HRD, 1D et 2D ISAR, etc.).
2. Extraction des paramètres pertinents pour l'identification.
3. Fusion spatiale et temporelle des données obtenues et comparaison avec une base de données auxiliaire (construite « off-line » à partir de connaissances préalables) jusqu'à l'obtention d'une valeur de confiance pour chaque classe considérée.
4. Prise d'une décision finale en ce qui concerne la possible identité de la cible.

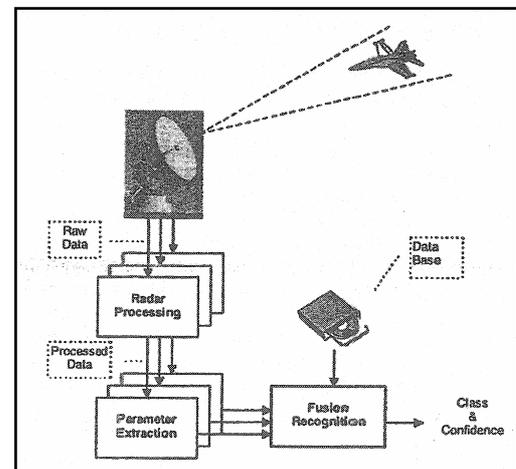


Figure 6. Modèle NCTR haut-niveau [5]

Les performances du processus d'identification dépendent de plusieurs facteurs qu'il faut prendre en compte :

- la qualité des mesures radar (résolution, SNR, erreur de mesure, etc.)
- le degré de complémentarité et redondance entre les capteurs
- le choix des variables et paramètres à considérer
- l'optimalité du processus de fusion de données
- la représentativité et l'exhaustivité du contenu de la base de données utilisée
- le critère utilisé pour la prise de décision finale

4.2 LA FUSION DE DONNEES POUR L'IDENTIFICATION NCTR

4.2.1 Les systèmes multicateurs et multiparamètres

Les niveaux de détail, précision et fiabilité souhaités aujourd'hui pour les systèmes de reconnaissance automatique rendent parfois indispensable l'utilisation conjointe de nombreux capteurs et/ou de nombreux paramètres, qui peuvent être de différente nature, et dont on devra profiter dans un double esprit de complémentarité et redondance.

En effet, la redondance existante entre les différents capteurs peut nous permettre d'augmenter notre degré de confiance pour la mesure obtenue, tandis que la complémentarité entre capteurs peut nous permettre de raffiner le niveau d'identification, de le rendre plus discriminant, en améliorant parfois aussi les performances générales et en réduisant parfois le temps nécessaire pour arriver aux conclusions définitives.

En général, on peut travailler avec un ou plusieurs paramètres provenant d'un ou plusieurs capteurs, et ces paramètres peuvent évoluer temporellement ou pas, ce qui rend possible l'existence de plusieurs stratégies envisageables pour l'intégration spatiale (la fusion des différents paramètres ou du même paramètre procédent de différents capteurs) et temporelle (la fusion des différents valeurs temporelles pour le même paramètre).

La complexité de quelques systèmes avec lesquels l'on travaille est telle que l'on peut être obligé de considérer aussi leurs interactions avec d'autres parties du système de défense qu'ils intègrent (on peut penser, par exemple, à un traitement conjoint identification pistage pour des cibles aériennes). C'est pourquoi la conception du système et le choix d'une architecture adéquate deviennent des aspects pas du tout négligeables.

En principe, on peut faire une distinction entre deux grandes familles de systèmes multicateurs et multiparamètres, selon s'ils possèdent une structure centralisée ou décentralisée :

- Dans les **systèmes décentralisés**, une partie du traitement est associée directement à chaque capteur et s'effectue à niveau local, de sorte que les données envoyées au système central possèdent un caractère symbolique au delà des simples mesures numériques. Dans certains cas, la nécessité de considérer des connaissances contextuelles peut conseiller l'utilisation d'une approche de ce type : on peut penser, par exemple, au traitement des profils Haute Résolution Distance produits par un radar cohérent, pour lesquels il faut prendre en compte des informations relatives au clutter, des paramètres associés à la forme d'onde et les filtres utilisés, etc. Au lieu d'envoyer toutes ces données au système central, il peut être plus avantageux de les traiter au niveau local et envoyer seulement le résultat d'intérêt pour le paramètre souhaité. On peut aussi considérer des systèmes dans lesquels ce que l'on envoie est déjà une mesure du degré d'appartenance à une classe pour une cible déterminée, obtenue à partir de la fusion de plusieurs paramètres.

Du point de vue de la fusion de données, cette architecture correspond bien au modèle de la « **fusion de décisions** », selon lequel on prend d'abord des décisions locales au niveau de chaque source séparément pour les fusionner après dans une décision globale. Du côté positif, on peut dire qu'il est très approprié quand les sources ne sont pas disponibles simultanément, qu'il facilite le développement de procédures spécifiques pour chaque source (et donc l'obtention de réponses plus rapides ou mieux adaptées, ce qui peut être très intéressant pour des systèmes temps réel) et qu'il facilite l'introduction de sources supplémentaires. Du côté négatif, on peut citer la perte de certaines informations qui restent à niveau local, une plus grande difficulté pour la prise en compte des possibles corrélations ou dépendances entre cibles et capteurs, ou encore une plus grande difficulté pour la gestion des possibles conflits causés par des décisions locales contradictoires.

- Dans les **systèmes centralisés**, par contre, on envoie les données numériques tels qu'elles sont acquises par le capteur, sans traitement préalable, vers un centre de fusion sur lequel aura lieu l'intégration. Cette approche permet de diminuer la complexité des capteurs périphériques, mais implique une surcharge en terme de capacité de calcul centrale et en terme de flux d'informations entre les différents composants du système.

Intuitivement, on peut ici établir une correspondance avec le modèle de la « **fusion globale** » dans le domaine de la fusion de données, puisque cette architecture permet de prendre une décision directe à partir de toutes les informations venant de toutes les sources, sans perte d'éléments de jugement à cause de sous-traitements locaux ; dans ce modèle globale aucune information n'est négligée au prix d'une complexité plus grande.

Dans la pratique, on peut considérer des modèles hybrides dans lesquels on fait quelques combinaisons pour certains paramètres à niveau local conduisant à des paramètres « dérivés » (imaginons, par exemple, la fusion locale de tous les paramètres cinématiques pour une cible donnée), que l'on enverra au sous-système central et que l'on considérera de manière conjointe avec d'autres paramètres élémentaires ou dérivés pour la prise finale de décisions globales. Il faudra, en tout cas, essayer d'arriver à avoir un compromis entre les avantages et les inconvénients associés à chacune des approches exposées.

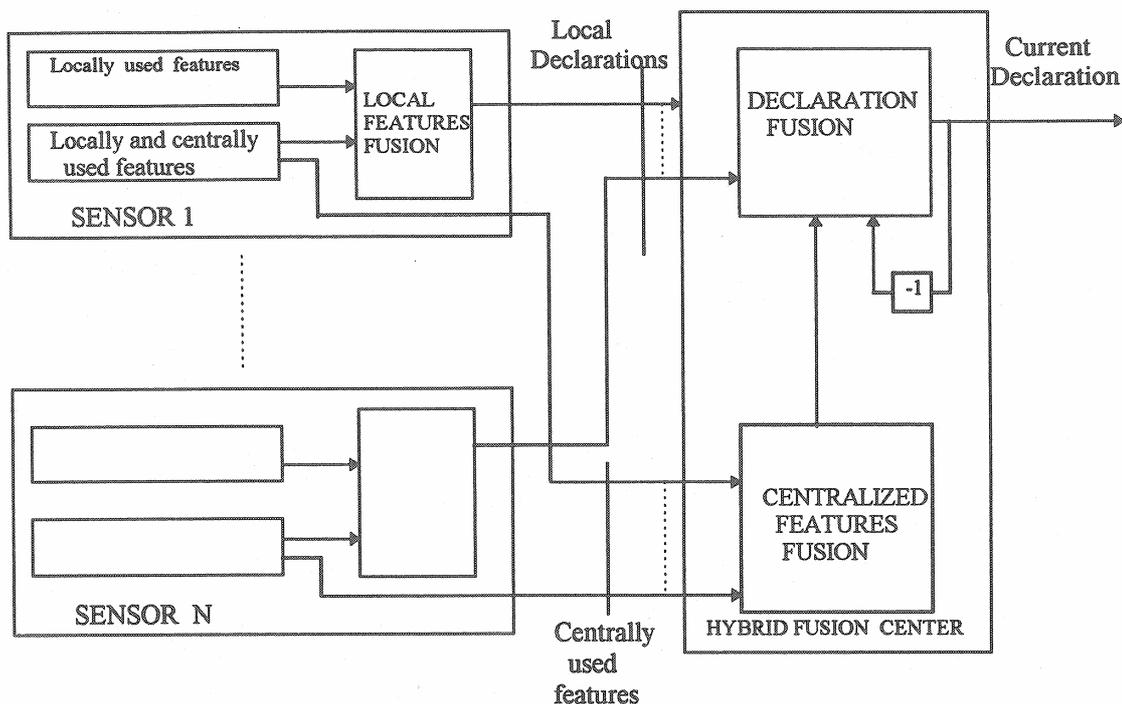


Figure 7. Modèle d'architecture hybride pour un système NCTR multicapteur [6]

Au delà de l'architecture considérée, le processus de fusion de données comportera toujours quatre étapes bien définies : modélisation, estimation, combinaison et décision, pour lesquelles on devra choisir un formalisme mathématique concret.

Dans les sous-sections suivantes on fera une petite description des formalismes les plus utilisés dans le cadre de la reconnaissance de cibles NCTR.

4.2.2 L'approche bayésienne

A fin de gérer les imperfections et incertitudes associées aux mesures radar, on peut essayer de les modéliser par des distributions de probabilité, ce qui nous permettra de nous appuyer sur des théories mathématiques solides et bien maîtrisées depuis longtemps.

On rappelle que, étant donné un ensemble de référence X , une **mesure de probabilité** \Pr est une fonction définie sur l'ensemble des parties $P(X)$, prenant ses valeurs dans $[0,1]$, telle que :

- a) $\Pr(X) = 1$
- b) $\forall A \in P(X), \forall B \in P(X)$ tels que $A \cap B = \emptyset$, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Ceci dit, une **mesure de probabilité** vérifie, quels que soient les événements A et B :

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

$$A \supseteq B \Rightarrow \Pr(A) \geq \Pr(B)$$

$$\Pr(A) + \Pr(\bar{A}) = 1$$

Si Pr est une mesure de probabilité définie sur X, une distribution de probabilité p est une fonction qui associe la valeur Pr(A) à tout événement A d'un système {A1, A2, ...} qui est complet, c'est à dire, tel que les événements soient disjoints et vérifient $\Pr(\cup_{i=1,2,\dots} A_i) = 1$.

Dans le cas de l'identification de cibles, on supposera que chaque objet appartient à une et à une seule classe « I » entre N classes possibles et on appellera « x » à la mesure fournie par le capteur considéré à un instant donné.

Afin de calculer la probabilité d'appartenance de la cible à chaque classe, on a besoin de certaines connaissances préalables à propos des classes susceptibles d'être identifiées, notamment :

- les probabilités « a priori » pour chaque classe P(J)
- la densité de probabilité conditionnelle « a priori » pour chaque classe P(x/J).

A partir de ces informations, il suffit d'utiliser la traditionnelle règle de Bayes :

$$P(I/x) = \frac{P(x/I) \cdot P(I)}{\sum_J P(x/J) \cdot P(J)}$$

Et finalement, on peut retenir la classe qui maximise la probabilité « a posteriori ».

Naturellement, cette approche plutôt simple peut être enrichie avec des concepts plus complexes (en introduisant, par exemple, la notion de coûts de décision), mais l'idée sous-jacente reste toujours la même.

Puisque la mesure x peut-être un vecteur, l'approche bayésienne offre une méthode assez intuitive et directe pour la fusion spatiale et temporelle de données, et elle peut être utilisée indifféremment pour travailler avec des données discrètes ou continues.

En tout cas, le lecteur intéressé trouvera plus d'informations dans la riche littérature disponible sur le sujet. Notre intention n'étant pas celle d'effectuer une description approfondie de cette approche, on s'arrête ici et on continue avec la présentation des principales techniques de fusion dans le domaine NCTR.

4.2.3 La théorie de l'évidence de Dempster – Shafer

La théorie de l'évidence ou théorie des croyances de Dempster – Shafer est née au cours des années 1970, inspirée des notions de probabilités supérieures et inférieures, mais son utilisation dans le domaine de la fusion de données est relativement récente.

La théorie peut être vue comme un modèle formel quantitatif de degrés de confiance, dans lequel on parle d'une fonction de croyance et d'une fonction de plausibilité par opposition à la fonction de probabilité classique de l'approche bayésienne.

La théorie de l'évidence considère un **univers de référence** ou **cadre de discernement X** contenant un certain nombre d'hypothèses (ex. $X = \{H_1, H_2, H_3\}$), supposé fini pour simplifier, sur lequel sont déterminés des coefficients de croyance obtenus en répartissant une masse globale de croyance égale à 1 entre tous les événements possibles et en attribuant à chacun le degré m(A) avec lequel un groupe d'observateurs croit en sa réalisation.

Voici quelques définitions utiles dans le cadre de la théorie de Dempster – Shafer :

- On appelle **cadre de discernement** l'ensemble d'hypothèses considérées.
- Une **masse de croyance** est une fonction m qui attribue un coefficient compris entre 0 et 1 aux parties de X de telle sorte que :

$$\sum_{A \in P(X)} m(A) = 1 \quad \text{et} \quad m(\emptyset) = 0$$

- On appelle **élément focal** toute partie non vide E de X telle que $m(E) \neq 0$ et on note \mathbf{E} leur ensemble. On appelle **corps d'évidence** le couple (\mathbf{E}, m) .

Dès lors qu'un jeu de masses a été défini sur l'ensemble des parties de X , il devient possible de calculer la plausibilité d'un élément A de $P(X)$. La **fonction de plausibilité** est donc définie grâce à la relation :

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$$

Lorsqu'une masse est affectée à une partie B de E , nul ne sait a priori sur quel élément particulier de B cette masse peut être attribuée. Lors du calcul de la plausibilité de A , on considère que pour chaque élément B dont l'intersection avec A est non nulle, cette masse est attribuée uniquement sur les éléments communs à A et à B . Cela revient, en définitive, à attribuer toute l'incertitude en faveur de A . Par conséquent la plausibilité représente la probabilité maximale qu'il serait possible d'estimer si tous les événements s'exprimaient en faveur de A .

Dans le même esprit, on définit la **fonction de croyance** d'un sous-ensemble quelconque A de $P(X)$ par :

$$Bel(A) = \sum_{A \supseteq B} m(B)$$

Inversement au calcul précédent, qui définissait une probabilité maximale, la crédibilité représente en fait une probabilité minimale. Comme précédemment pour toutes les parties B telles que $A \cap B \neq \emptyset$ on considère que l'incertitude va en défaveur de l'élément A . Par conséquent toute la masse attribuée à B est en définitive concentrée sur les éléments de B qui n'appartiennent pas à l'intersection $A \cap B$. Seules les parties B telles que $B \subseteq A$ contribuent donc au calcul de cette probabilité inférieure.

Intuitivement, on observe que dans le cas d'une ignorance totale sur un événement A , aucun élément focal n'intervient en sa faveur ni en sa défaveur, c'est à dire qu'aucun élément focal n'est inclus dans A ni dans \bar{A} et par conséquent $Bel(A) = Bel(\bar{A}) = 0$. De même, tous les éléments focaux ont une intersection non vide avec A et \bar{A} , ce qui implique que $Pl(A) = Pl(\bar{A}) = 1$.

Parmi les propriétés fondamentales des fonctions de croyance et plausibilité on trouve :

$$Bel(A) \leq Pl(A)$$

$$Pl(A) + Bel(\bar{A}) = 1$$

$$Bel(A) + Bel(\bar{A}) \leq 1$$

$$Pl(A) + Pl(\bar{A}) \geq 1$$

Et l'union et intersection de sous ensembles vérifient :

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

$$Pl(A \cup B) \leq Pl(A) + Pl(B) - Pl(A \cap B)$$

D'une façon générale, les masses de croyance nous permettent de calculer la probabilité des hypothèses du cadre de discernement considéré en attribuant à chaque hypothèse incluse dans une partie A une masse égale à la masse totale de A divisée par le nombre d'hypothèses contenues dans A (distribution équi-répartie) et en additionnant l'ensemble des masses qui lui sont attribuées. Si $|A|$ représente le cardinal de A on a alors la relation suivante :

$$\bar{P}(H_i) = \sum_{H_j \in A} \frac{m(H_j)}{|A|}$$

On peut vérifier que la consistance entre la plausibilité, la crédibilité et la probabilité d'une hypothèse exprimée par la relation suivante est bien vérifiée :

$$Bel(H_i) \leq \bar{P}(H_i) \leq Pl(H_i)$$

4.2.3.1 Règle de combinaison de Dempster

En présence de plusieurs sources d'informations fournissant chacune un jeu de masses sur un même cadre de discernement, la règle de Dempster permet de construire un jeu de masses unique par sommation orthogonale. Une particularité de cette règle résulte de ce que les éléments focaux de la masse issue de la combinaison sont généralement différents des éléments focaux des masses élémentaires, si bien que la première étape de cette règle doit être consacrée à la définition de l'ensemble des éléments focaux de la masse résultant de la combinaison.

Soient $m_1()$ et $m_2()$ les masses élémentaires fournies par deux sources d'informations indépendantes sur un même cadre de discernement E , soit $m()$ la masse obtenue après combinaison de ces deux masses élémentaires, nous noterons :

$$m() = m_1() (+) m_2().$$

Chaque élément focal de $m()$ est le résultat de l'intersection d'un élément focal de $m_1()$ avec un élément focal de $m_2()$. Par conséquent l'ensemble des éléments focaux de $m()$ est constitué par toutes les intersections possibles.

Il convient d'attribuer une masse à chacun de ces éléments. La première étape de cette affectation se fait en attribuant pour chaque intersection d'un élément du support de $m_1()$ avec un élément du support de $m_2()$ le produit des deux masses correspondantes. Puis pour chaque élément du support de $m()$ on recense l'ensemble des intersections qui aboutissent à cette élément et on additionne toutes les masses. La seconde étape, de normalisation, consiste à diviser l'ensemble des masses par $1-K$ où K est le facteur de conflit qui se calcule en additionnant l'ensemble des masses qui sont attribuées à l'élément \emptyset .

La règle de Dempster s'exprime grâce à la relation suivante :

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K} \quad \forall A \text{ et } B \in P(X)$$

avec

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C) \quad \forall A \text{ et } B \in P(X)$$

4.2.4 La théorie des ensembles flous et des possibilités

La théorie des ensembles flous offre un nouveau paradigme pour la modélisation de l'incertitude. Malgré quelques essais précédents dans le domaine de la science et la philosophie, notamment dans le cadre de la logique multivaluée et les concepts vagues, le professeur Lofti A. Zadeh, de l'université de California à Berkeley a été le premier à proposer une théorie mathématique pour ce type d'ensembles et leur logique associée dans son article « *Fuzzy sets* » publié dans la revue *Information and controls* en 1965 [17].

Basiquement, un ensemble flou est un ensemble dont les membres peuvent avoir des degrés d'appartenance entre 0 et 1, par opposition aux ensembles classiques pour lesquels chaque élément doit avoir un degré d'appartenance égal à 1 ou 0, « oui » ou « non ».

Pour une description très complète des concepts utilisés dans le cadre de cette théorie, on recommande la lecture du chapitre 8 dans [1].

La théorie des possibilités, dérive de la théorie des ensembles flous et a été aussi introduite par L.Zadeh dans son article « *Fuzzy Sets as a Basis for a Theory of Possibility* » en 1978 [18]. Elle fournit une méthode pour formaliser des incertitudes subjectives sur des événements, c'est à dire, dans quelle mesure la réalisation d'un événement est possible et dans quelle mesure on en est certain, sans toutefois avoir à sa disposition l'évaluation de la probabilité de cette réalisation, par exemple parce qu'on ne connaît pas d'événement analogue auquel se référer, ou parce que l'incertitude est la conséquence d'une absence de fiabilité des instruments d'observation ou d'un doute de l'observateur lui-même. Ces deux évaluations subjectives se formalisent à travers une mesure de possibilité et une mesure de nécessité.

Etant donné un ensemble de référence X , une **mesure de possibilité** Π est une fonction définie sur l'ensemble $P(X)$, prenant des valeurs dans $[0,1]$, telle que :

$$i) \quad \Pi(\emptyset) = 0, \Pi(X) = 1$$

$$ii) \quad \forall \text{ famille d'ensembles } A_i \in P(X)$$

$$\Pi\left(\bigcup_{i=1,2,\dots} A_i\right) = \sup_{i=1,2,\dots} \Pi(A_i)$$

ou sup indique le supremum des valeurs concernées, c'est à dire la plus grande d'entre elles dans le cas fini.

Une mesure de possibilité est totalement définie si l'on attribue un coefficient de possibilité à toute partie de l'ensemble de référence X , mais l'on peut y arriver plus facilement si l'on indique les coefficients attribués aux seules parties élémentaires de X , les singletons, un sous-ensemble quelconque de X pouvant être regardé comme l'union des singletons qu'il contient.

Une **distribution de possibilité** π est une fonction définie sur X , prenant ses valeurs dans $[0,1]$, satisfaisant la condition de normalisation suivante :

$$\sup_{x \in X} \pi(x) = 1$$

L'ensemble des éléments de X pour lesquels la distribution de possibilité π est égale à 1 est son **noyau** et celui des éléments pour lesquels la distribution est différente de 0 est son **support**. Mesure et distribution de possibilité peuvent être associées bijectivement.

Quelques propriétés d'intérêt concernant la possibilité sont :

$$\forall (A, B) \in P(X)^2 \quad \Pi(A \cap B) \leq \min(\Pi(A), \Pi(B))$$

$$\max(\Pi(A), \Pi(\bar{A})) = 1$$

$$\Pi(A) + \Pi(\bar{A}) \geq 1$$

Pour compléter l'information décrivant l'incertitude d'un événement, la théorie des possibilités évalue aussi le degré avec lequel la réalisation de cet événement est certaine, par l'intermédiaire d'une mesure de nécessité.

Une **mesure de nécessité** N est une fonction définie sur l'ensemble $P(X)$ des parties de l'ensemble de référence fini X , prenant des valeurs dans $[0,1]$, telle que :

$$i) \quad \Pi(\emptyset) = 0, \Pi(X) = 1$$

$$ii) \quad \forall \text{ famille d'ensembles } A_i \in P(X)$$

$$\Pi\left(\bigcap_{i=1,2,\dots} A_i\right) = \inf_{i=1,2,\dots} \Pi(A_i)$$

ou inf indique l'infimum des valeurs concernées, c'est à dire la plus grande d'entre elles dans le cas fini.

Les axiomes définissant mesure de possibilité et mesure de nécessité sont duaux et donc on peut définir le concept de distribution de nécessité à partir d'un raisonnement similaire à celui utilisé pour le cas antérieur.

En fait, étant donné un ensemble de référence X , et soit \bar{A} le complémentaire de toute partie A de X , une mesure de nécessité N peut être obtenue à partir de la donnée d'une mesure de possibilité Π , par :

$$\forall A \in P(X) \quad N(A) = 1 - \Pi(\bar{A})$$

Ceci nous permet de définir la mesure de nécessité N duale de Π par :

$$\forall A \in P(X) \quad N(A) = \inf_{x \notin A} (1 - \pi(x))$$

Quelques propriétés d'intérêt concernant la nécessité sont :

$$\forall (A, B) \in P(X)^2 \quad N(A \cup B) \geq \max(\Pi(A), \Pi(B))$$

$$\min(\Pi(A), \Pi(\bar{A})) = 0$$

$$N(A) + N(\bar{A}) \leq 1$$

La dualité entre possibilité et nécessité est mise en évidence grâce aux propriétés suivantes pour une partie A de X :

$$\Pi(A) \geq N(A)$$

$$\max(\Pi(A), 1 - N(A)) = 1$$

$$N(A) \neq 0 \Rightarrow \Pi(A) = 1$$

$$\Pi(A) \neq 1 \Rightarrow N(A) = 0$$

Les deux dernières relations indiquent que tout événement dont on est, au moins un peu, certain est tout à fait possible, et qu'on ne peut avoir la moindre certitude sur un événement qui n'est que relativement possible.

4.2.4.1 L'agrégation de composantes floues

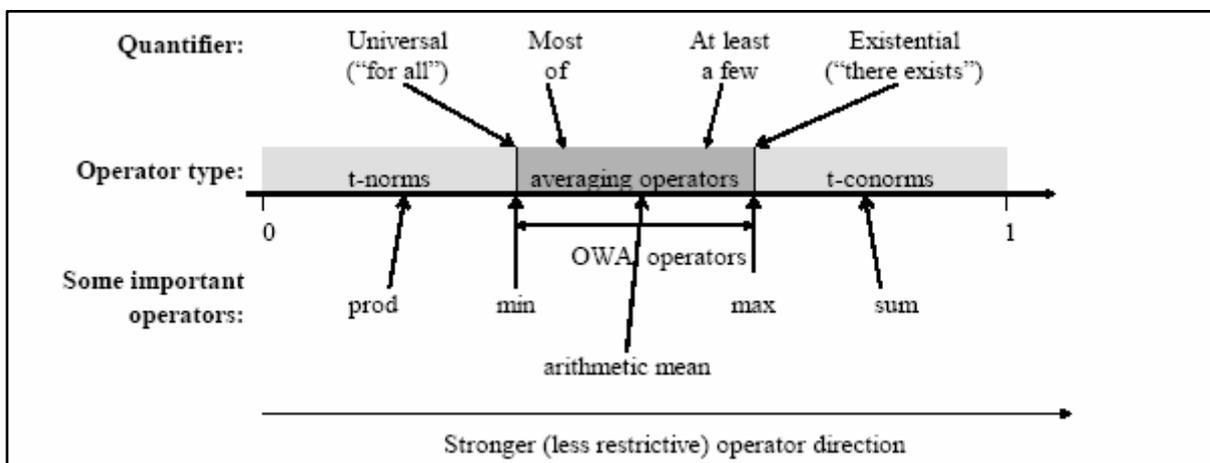
Dans le domaine de la logique floue, la fusion des avis différents sur la caractérisation d'un même critère ou des préférences obtenues par diverses éventualités à partir de l'utilisation de plusieurs critères est accompli par le choix et utilisation d'un opérateur d'agrégation, qui déterminera la façon dont interviennent chacun des éléments dans le résultat agrégé cherché.

Soit donc un ensemble X sur lequel sont définis des sous-ensembles flous A_1, \dots, A_n . On cherche un opérateur h défini sur $[0, 1]^n$, à valeurs dans $[0, 1]$, qui permette de construire un résultat agrégé A à partir des sous-ensembles considérés.

Les opérateurs envisageables sont très nombreux et peuvent être groupés dans des différentes familles :

- Opérateurs de moyenne (moyennes classiques, OWA's, intégrales floues ...)
- Normes et conormes triangulaires
- Sommes symétriques
- Opérateurs adaptatifs

On parlera plus en détail de certains opérateurs de moyenne (classiques, OWA), ainsi que des normes et conormes triangulaires, dans le cadre de l'étude décrite dans la section dédiée à la variable possibilité.



En ce qui concerne le reste, on peut faire ici une description très rapide, afin d'en avoir une idée intuitive :

- Les **intégrales floues** de Choquet et Sugeno sont des opérateurs idempotents, continus, croissants et compris entre le minimum et le maximum.

Soit un espace mesuré flou (Ω, B, μ) , où B est une σ -algèbre de parties de Ω et μ une mesure floue sur B et soit une fonction mesurable f définie sur Ω et à valeurs dans $[0, 1]$.

On note $f^\alpha = \{x \in \Omega / f(x) > \alpha\}$.

L'**intégrale de Sugeno** de f est définie par :

$$\int_S f \circ \mu = \sup_{\alpha \in [0,1]} \min(\alpha, \mu(f^\alpha))$$

et l'**intégrale de Choquet** de f est définie par :

$$\int_C f \cdot d\mu = \int_0^1 \mu(f^\alpha) \cdot d\alpha$$

On y trouve comme cas particulier les statistiques d'ordre et donc le minimum, le maximum et la médiane. Les intégrales de Choquet définies par rapport à une mesure additive μ sont équivalentes à une moyenne arithmétique pondérée, dans laquelle les poids w_i affectés aux valeurs x_i sont égaux à $\mu(\{x_i\})$.

- Les **sommes symétriques** sont des opérateurs qui se caractérisent par avoir une propriété d'autodualité, qui correspond à l'invariance du résultat de l'opération par inversion de l'échelle des valeurs à combiner. Cette autodualité s'oppose à la dualité qu'on trouvera quand on étudiera les t-normes et t-conormes.

Leur forme générale est donnée par :

$$\sigma(x, y) = \frac{g(x, y)}{g(x, y) + g(1-x, 1-y)}$$

où g est une fonction de $[0, 1] \times [0, 1]$ dans $[0, 1]$, croissante, positive et continue telle que $g(0, 0) = 0$. Typiquement, on peut prendre pour g une t-norme ou une t-conorme continue.

- Les **opérateurs adaptatifs** prennent en compte le conflit entre les différentes distributions de possibilité à fusionner et peuvent se comporter comme un « min » si les distributions sont consonantes et comme un max si elles présentent un fort conflit. Au delà des distributions de possibilité considérées (par exemple π_1 et π_2 définies sur D), ces opérateurs utiliseront une mesure h de conflit qui peut être, par exemple, de la forme :

$$h(\pi_1, \pi_2) = 1 - \max_{c \in D} \min(\pi_1(c), \pi_2(c))$$

Pour une description plus en détail de ces opérateurs, avec des références à des articles permettant d'approfondir sur chacun d'entre eux, on renvoie le lecteur au chapitre 8 de [1].

4.2.5 Un cadre commun pour des probabilités, croyances et possibilités

Dans son ouvrage « *La logique floue et ses applications* » [2], Bernadette Bouchon-Meunier explique comment les théories des possibilités et des probabilités peuvent être considérées comme des cas particuliers de la théorie de l'évidence.

4.2.5.1 Cas des probabilités

Imaginons des observateurs qui émettent des avis très nets et que les croyances émises concernent des éléments de X pris individuellement, les éléments focaux seront des singletons de X . L'indétermination sur l'occurrence des divers événements est alors moins importante que précédemment puisque les événements élémentaires reçoivent une part de la masse de croyance.

Si les éléments focaux sont les singletons de X , la fonction de plausibilité et la fonction de croyance sont identiques et ont les propriétés d'une mesure de probabilité :

$$Bel(A) = Pl(A) = Pr(A)$$

4.2.5.2 Cas des possibilités

Si, par contre des observateurs n'attribuent pas de croyance à des événements incompatibles des événements ne seront affectés de masses de croyance non nulles que s'ils sont concordants. Le corps d'évidence est cohérent et donc les éléments focaux $E_1, E_2 \dots E_k$ sont emboîtés, c'est à dire tels que $E_1 \supseteq E_2 \supseteq \dots \supseteq E_k$.

Dans ce cas là, on peut démontrer que la fonction de plausibilité a les propriétés d'une mesure de possibilité et la fonction de croyance celles d'une mesure de nécessité.

La figure ci-jointe, extraite de l'ouvrage de Bouchon-Meunier cité au-dessus, montre une récapitulation des conclusions qu'on vient d'expliquer :

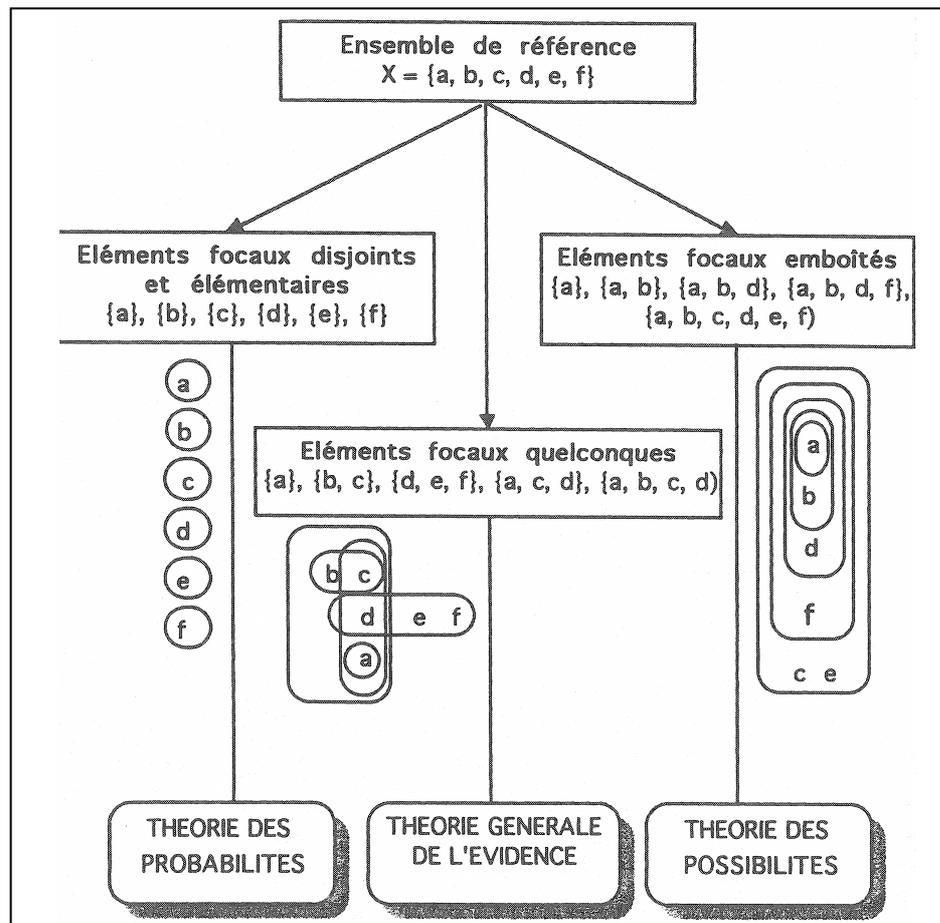


Figure 9. Position de la théorie des possibilités et de la théorie des probabilités dans le cadre de la théorie de l'évidence [2]

4.2.6 D'autres techniques pour l'intégration

Les techniques et les modèles formels présentés dans cette section pour la modélisation et la fusion d'informations dans le cadre de l'identification NCTR constituent ce qu'on pourrait considérer comme le « noyau traditionnel » pour ce type de méthodes (s'il est possible de parler de traditions dans un domaine tellement récent).

En tout cas, il y a d'autres approches possibles telles que celles basées sur l'utilisation de systèmes experts, de réseaux de neurones ou même des approches hybrides comme celle des systèmes « *neuro-fuzzy* » qui se trouvent hors le cadre de ce rapport mais qu'il serait peut-être intéressant de considérer pour des études futures.

4.3 UN MODELE D'IMPLEMENTATION CONCRET BASE SUR LA LOGIQUE FLOUE

Dans cette section on fera la description d'un exemple de système d'identification NCTR multiparamétrique et temps réel dont le cadre formel utilisé est celui de la logique floue et la théorie des possibilités.

L'architecture et le fonctionnement général du système étaient proposés par M. Moruzzis et N. Colin dans leur article « *Radar Target Recognition by Fuzzy Logic* » [4].

C'est à partir de ce modèle qu'on effectuera notre travail de simulation et recherche autour des possibles algorithmes envisageables pour l'intégration temporelle des données floues dans les sections suivantes.

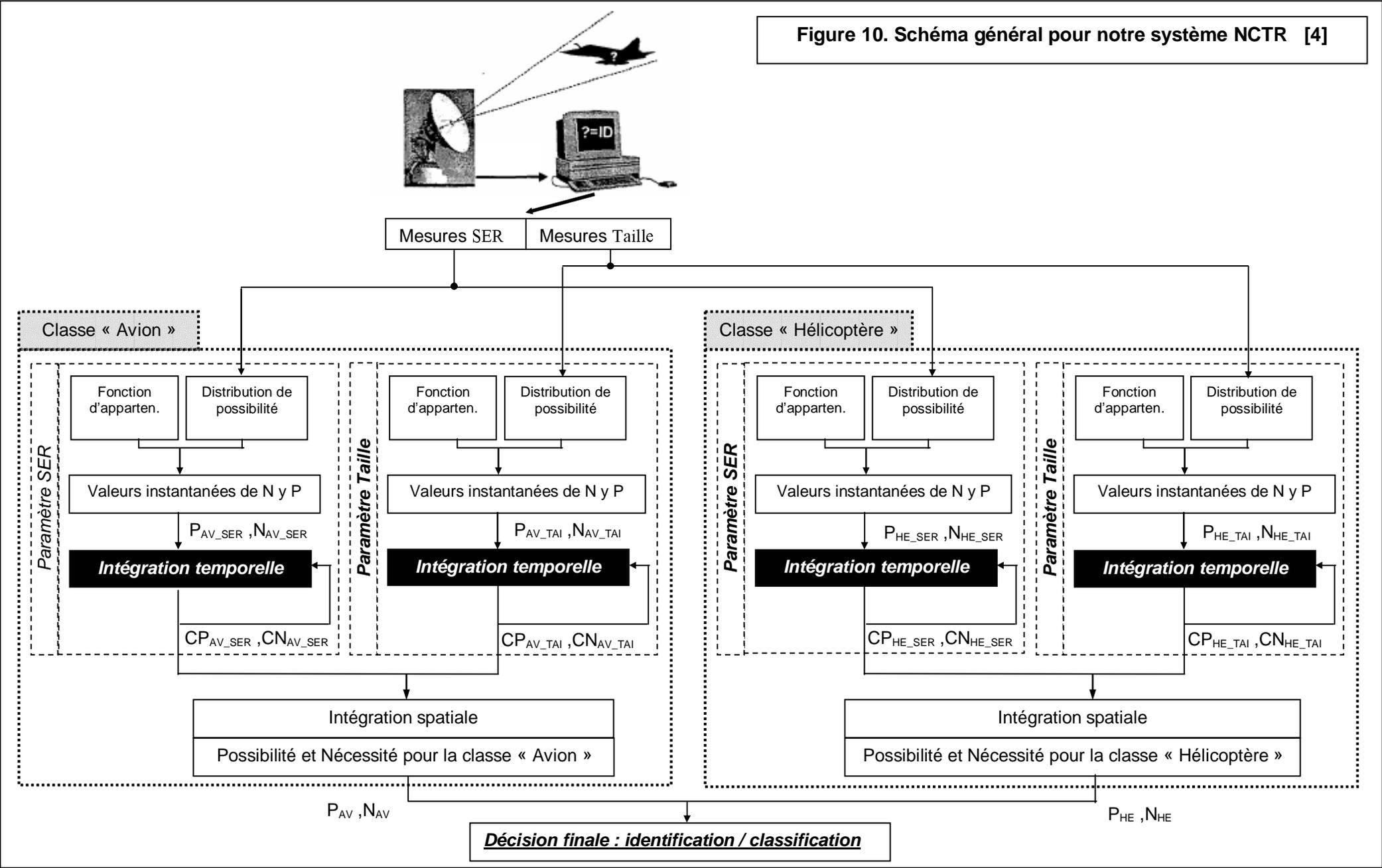
4.3.1 Schéma général

On va considérer une situation dans laquelle un ou plusieurs capteurs radar vont fournir certains paramètres pour chacune des cibles à identifier. Comme on a déjà expliqué, ces paramètres peuvent être mesurés d'une manière plutôt « directe » par le radar (c'est le cas, par exemple, des paramètres de type cinématique) ou extraits des mesures radar grâce à des techniques plus complexes (analyse Doppler, techniques HRD, ISAR 1D et 2D, etc...).

Pour chaque paramètre et chaque classe considérée on calculera des valeurs de possibilité et de nécessité instantanées, qui feront l'objet d'un processus d'intégration temporelle afin d'obtenir les valeurs pertinentes pour la possibilité et la nécessité « cumulées » à chaque instant. Ensuite, on fera une intégration de type « spatiale » entre les différents paramètres pour chaque classe qui nous permettra d'obtenir une valeur finale de possibilité et nécessité d'appartenance à la classe en question pour la cible considérée. En regardant ces valeurs, obtenues pour chaque classe, on devra prendre une décision en ce qui concerne la classification de la cible.

Dans la figure ci-jointe, on présente un schéma correspondant à un petit exemple dans lequel on a considéré deux classes (« avion » et « hélicoptère ») et deux paramètres (« Surface Equivalent Radar (SER) » et « taille »). Le modèle peut être facilement étendu à d'autres systèmes plus complexes, qui prennent en compte plus de paramètres et plus de classes possibles pour l'identification.

Figure 10. Schéma général pour notre système NCTR [4]



4.3.2 Les fonctions d'appartenance et de densité de possibilité

Les fondements théoriques sous-jacents à ce modèle sont extraites de l'article « *Weighted Fuzzy Pattern Matching* » écrit par Dubois, Prade et Testemale [11].

L'idée initiale étant celle de construire un système de reconnaissance de formes sur la base de la théorie des ensembles flous et des possibilités, on associera une distribution de possibilité et un ensemble flou à chacun des paramètres venant des capteurs radar.

Plus concrètement, on travaillera avec deux fonctions, qui prendront des valeurs dans le domaine X de mesure de chaque paramètre considéré et donneront des résultats dans l'intervalle $[0,1]$:

- La **fonction d'appartenance $M_P(x)$** , qu'on utilise pour montrer et caractériser le degré de compatibilité entre une classe donnée et un certain paramètre. Pour chaque élément x de X , $M_P(x) = 1$ implique une totale sûreté à propos du fait que la valeur mesurée x est compatible avec la classe P tandis que $M_P(x) = 0$ implique une totale sûreté du fait qu'elle ne l'est pas. Cette fonction sera fixe dans le temps pour chaque classe et chaque paramètre.

Du point de vue de la représentation des connaissances, cette fonction permet d'introduire une incertitude dans la caractérisation a priori de chaque classe, caractérisation qui doit souvent être faite à partir de l'expérience d'identifications passées ou des avis d'experts.

L'approche la plus habituelle est celle de considérer une fonction trapézoïdale. Dans des conditions normales, un expert sera capable de terminer les limites dans lesquels il est sûr que les valeurs d'un certain paramètre correspondent ou ne correspondent pas à la classe considérée. Entre ces limites, il est difficile, même pour un expert, de décider si l'objet appartient à la classe ou pas, et donc on peut considérer une variation monotone que l'on peut modéliser par une ligne droite.

La figure ci-jointe montre un exemple d'une fonction de ce type :

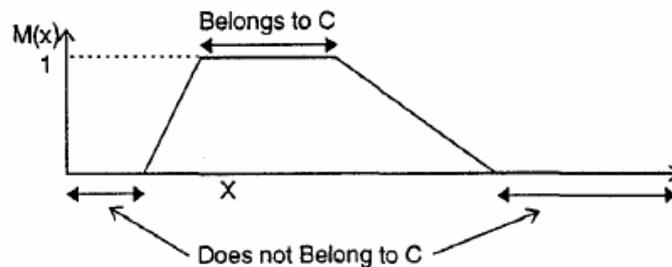


Figure 11. Exemple de fonction d'appartenance

- La densité de possibilité $D(x)$ est directement relié à la notion d'erreur ou imprécision dans la mesure radar, est devra être obtenue à chaque instant à partir des valeurs instantanées fournies par le capteur impliqué ($x_C(t)$) et sa précision (σ_C). On parle, donc, d'une fonction avec une forte dépendance des caractéristiques techniques du radar utilisé et qui varie à chaque instant.

De même que pour la fonction d'appartenance, on peut considérer des fonctions trapézoïdales pour la densité de possibilité. Un critère habituel est celui de considérer un modèle gaussien de variance σ_C et utiliser cette valeur pour définir les limites du trapèze en question :

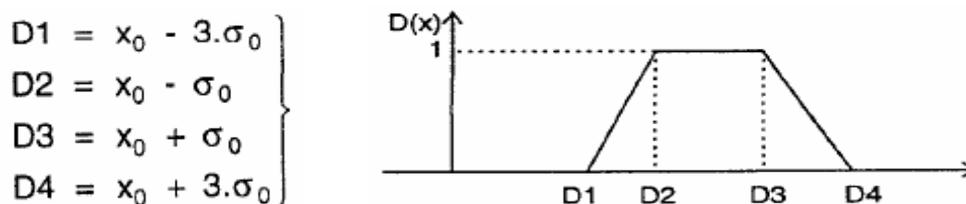


Figure 12. Exemple pour la densité de possibilité

4.3.3 Calcul de la possibilité et la nécessité

Après l'introduction plutôt formelle aux principes de la théorie des possibilités qu'on a fait dans la section précédente, on peut vouloir avoir une vision plus intuitive des concepts de possibilité et nécessité pour notre cas concret d'application.

C'est ainsi que la possibilité peut être regardée comme une mesure du degré d'intersection entre la fonction d'appartenance et la distribution de possibilité pour chaque classe et chaque paramètre mesuré, tandis que la nécessité peut être considéré une mesure d'inclusion.

A partir de cette idée, on peut calculer facilement la possibilité et la nécessité instantanées, selon les relations suivantes :

$$P(t) = \underset{x}{\text{Max}}\{\text{Min}[M(x), D(x, t)]\}$$

$$N(t) = \underset{x}{\text{Min}}\{\text{Max}[M(x), 1 - D(x, t)]\}$$

Graphiquement, on trouve une situation de la forme :

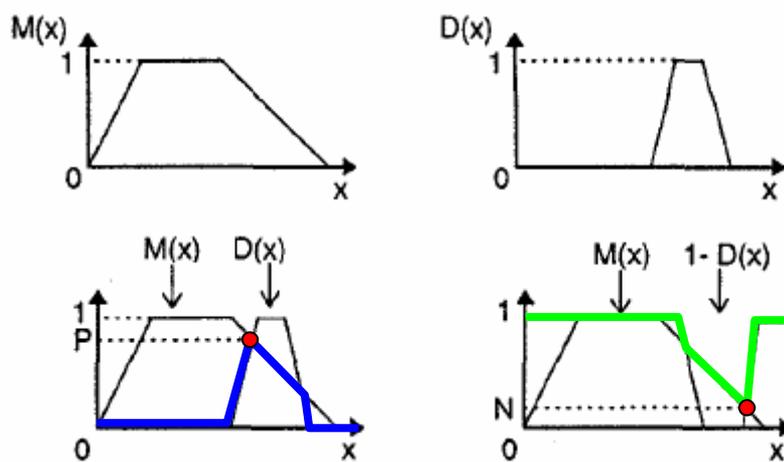


Figure 13. Calcul de la possibilité et la nécessité instantanées

4.3.4 Intégration spatiale et intégration temporelle

Une fois qu'on a calculé les valeurs instantanées de possibilité et nécessité pour chaque classe et chaque paramètre, il faudra les intégrer en une seule valeur par classe, afin de faciliter le processus final de prise de décision. De plus, il peut être souhaitable de prendre en compte l'évolution temporelle des valeurs considérées pour améliorer les performances générales du système d'identification. Les techniques de fusion de données dont on a déjà parlé joueront ici un rôle essentiel.

Le premier type d'intégration mentionné, l'intégration dite « spatiale » peut être accomplie par une approche plutôt simple à caractère conjonctif ou disjonctif. Puisque ce type d'intégration ne constitue pas l'objet d'étude spécifique de l'actuel stage, on considérera que, dans le cadre de nos simulations et pour le modèle d'architecture proposé, elle s'effectue toujours avec un opérateur de type conjonctif (« min » ou « AND »).

Cela veut dire que, après l'intégration temporelle, les différentes valeurs cumulées de P et N pour les paramètres utilisés pour chaque classe sont fusionnées dans une valeur finale par classe en calculant leur minimum. On utilisera toujours ce critère afin de faciliter l'étude indépendamment de l'influence du choix de l'algorithme d'intégration temporelle sur les performances générales d'identification.

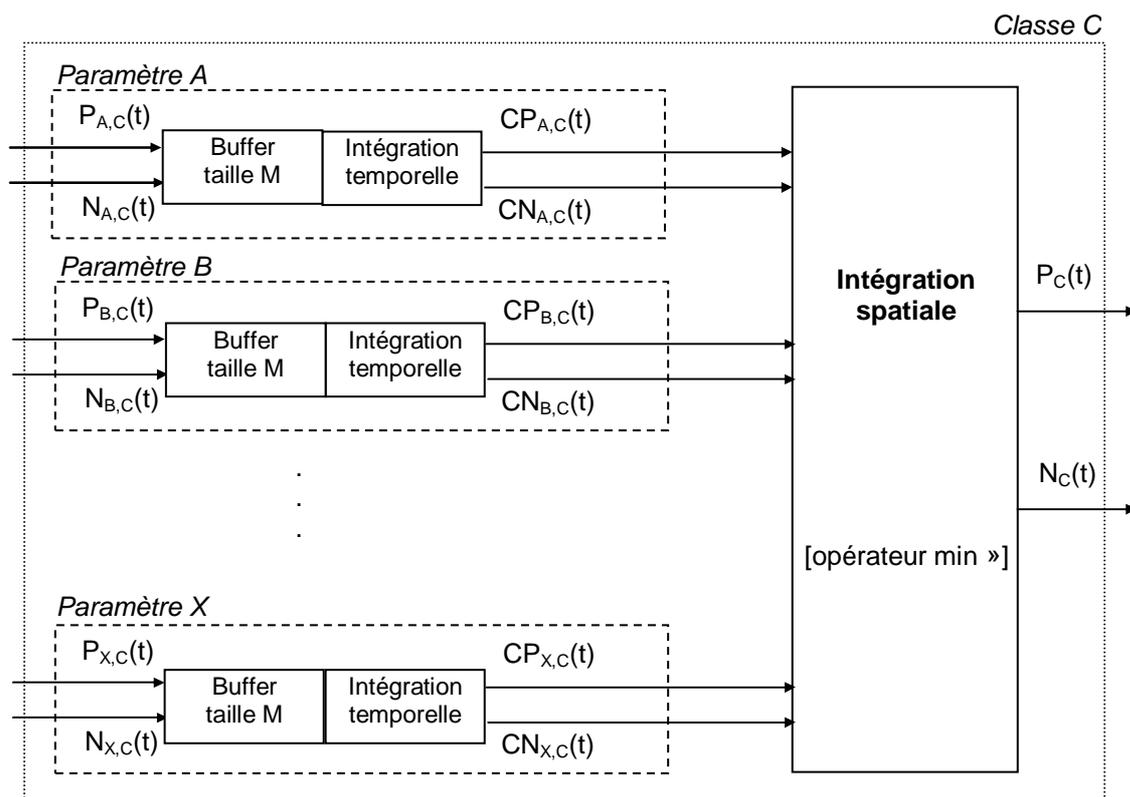


Figure 14. Intégration temporelle vs. intégration spatiale

Le processus d'intégration temporelle sera le sujet principale d'une section future et donc, on se limitera ici à signaler que, en principe, cette intégration sera faite sur les M dernières valeurs instantanées considérées (typiquement $M=10$), qui seront stockées à cet effet dans un buffer en mémoire. Comme on le verra plus tard, parfois il peut être néanmoins intéressant de considérer aussi d'autres stratégies visant faire l'intégration avec le résultat cumulé correspondant à l'instant précédent.

4.3.5 Prise de décision

Pour la prise finale de la décision concernant la classification de chaque cible, on partira d'un système de double seuil conduisant à trois résultats possibles par cible et classe : acceptation, rejet ou doute.

Idéalement, on pourrait considérer un « seuil de rejet » sur la possibilité de manière que l'on puisse rejeter directement la classe (on affirme que la cible n'appartient pas à cette classe) si après les processus d'intégration temporelle et spatiale une cible présente pour celle-ci une valeur de possibilité inférieure au seuil. Si, par contre, cette valeur est supérieure au seuil et que la valeur correspondante de nécessité est supérieure à un autre seuil dit « d'acceptation », on pourra affirmer l'appartenance de la cible à la classe considérée. Si l'on se trouve au-dessus du seuil de rejet pour la possibilité et au-dessous du seuil d'acceptation pour la nécessité, on ne pourra pas rejeter ni accepter la classe et on parlera de doute.

D'après cette approche, on travaillera avec un vecteur de décision pour chaque identification, de dimension 1, dont la longueur sera égale au nombre de classes considérées et dont les éléments pourront prendre trois valeurs seulement. Il sera donc possible d'accepter plusieurs identités pour une cible, c'est-à-dire, une cible pourra être identifiée comme appartenant à plusieurs classes si les informations fournies par les capteurs ne nous permettent pas de réduire le nombre de candidates jusqu'à l'unité.

5 Des outils pour l'évaluation des performances

Afin de bien conduire notre étude sur les différentes approches envisageables pour l'intégration temporelle des données floues dans le cadre du système d'identification NCTR que l'on vient de décrire, on devra tout d'abord choisir quelques critères pour l'évaluation des performances des différents algorithmes et opérateurs.

De cette façon, on disposera de quelques indicateurs quantitatifs, précis et objectifs capables de mesurer les résultats obtenus au cours de nos futures simulations.

Les indicateurs que l'on a décidé d'utiliser peuvent être classés tout d'abord dans deux grands groupes : des fonctions d'évaluation et des mesures sur les matrices de confusion.

5.1 FONCTIONS D'ÉVALUATION

La première idée que l'on a eu pour comparer les différentes méthodes d'intégration est de noter les résultats obtenus par moyen de quelques fonctions que l'on a développées en tenant compte de nos besoins spécifiques et avec des critères très particuliers pour chaque cas.

Bien que ces fonctions puissent avoir des comportements ou des résultats « discordants » au niveau ponctuel, on s'attend à observer une certaine cohérence globale qui nous permettra d'obtenir quelques conclusions d'intérêt.

Les trois premières fonctions A, B et C répondent à des approches assez intuitives, tandis que la dernière fonction se limite à calculer un taux d'erreur et un taux de réussite objectifs selon les critères habituels adoptés à Thales pour les produits d'identification radar.

On travaillera avec une base de données constituée par des mesures réelles correspondant aux différentes classes utiles pour l'identification de quelques cibles connues. Cela veut dire que pour une cible d'un type X connu, on devra considérer des différents paramètres et pour chacun d'entre eux on disposera d'un nombre de mesures égal à chaque instant à la quantité totale de classes susceptibles d'être identifiées (plus concrètement, on considérera 6 classes C1,C2..C6 donnés sur 6 cibles T1, T2...T6 appartenant aux 6 classes en question).

Exemple. Si l'on considère un contexte d'identification avec 6 classes possibles, pour une cible du type 3 et un paramètre P donnés, on devra considérer 6 ensembles de mesures de possibilité et nécessité fournies par les capteurs radars tout au long d'un même intervalle temporel, de sorte que l'on sera obligé de travailler avec un ensemble de données (pre et post – intégration temporelle) de la forme :

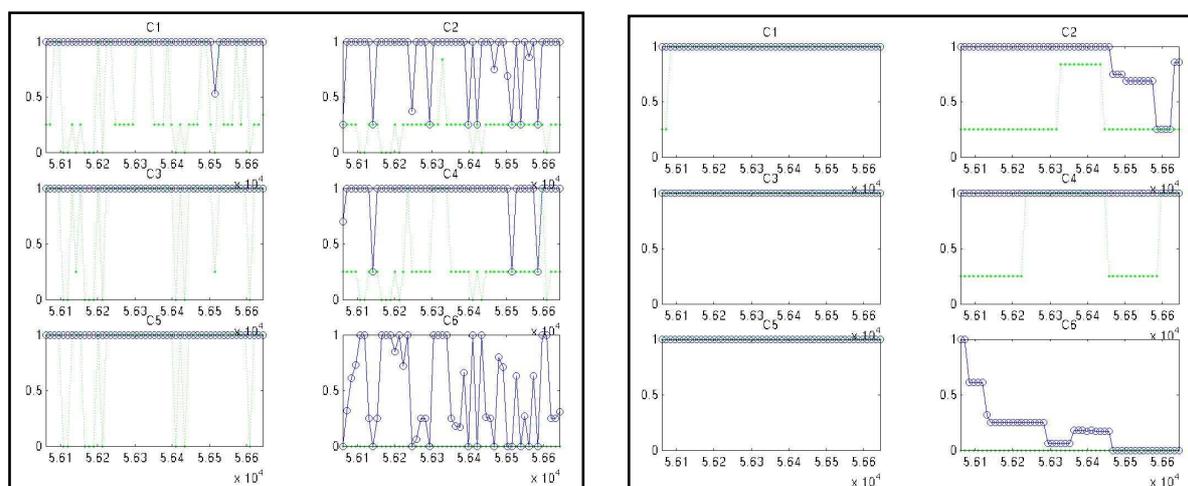


Figure 15. Exemple des données réelles utilisées – possibilité (bleu) et nécessité (vert) avant (gauche) et après (droite) intégration. La figure montre les mesures correspondant à six classes pour un certain paramètre.

5.1.1 Fonction « A »

Cette première fonction est la seule qui travaille directement sur les résultats de l'intégration temporelle et spatiale, sans intervention d'aucun type de processus de prise de décision.

Il s'agit tout simplement de mesurer de manière objective le degré de similarité entre les résultats obtenus pour une méthode d'intégration temporelle déterminé et les résultats de classification pour un cas idéal.

On va donner une bonne note quand on obtient des valeurs de possibilité et nécessité élevées pour la bonne classe associée à chaque cible et aussi quand on obtient des valeurs de possibilité et nécessité faibles pour les classes qui ne correspondent pas à la cible en question. Cela traduit l'idée intuitive qu'un algorithme sera bon dès qu'il sera capable d'affirmer l'appartenance à la bonne classe, mais aussi de rejeter l'appartenance de la cible aux classes erronées.

D'un point de vue mathématique, cela revient à considérer comme des notes les valeurs de la possibilité et la nécessité finales elles-mêmes pour la bonne classe et leurs complémentaires pour les classes erronées. Puis, on fera une normalisation et une pondération afin d'avoir une valeur finale entre 0 et 1 dans laquelle la note apportée par la bonne classe ait le même poids que la note apportée par l'ensemble des classes erronées.

Si l'on considère Q classes possibles (et donc Q valeurs finales de P_i et N_i , avec i qui varie entre 1 et Q) et une cible de type C ($1 \leq C \leq Q$), la fonction d'évaluation pour chaque instant sera de la forme :

$$f_A(\text{données_intégrées}) = \frac{1}{2} \left(\frac{P_C + N_C}{2} + \frac{1}{Q-1} * \sum_{i=1, i \neq C}^Q \frac{(1-P_i) + (1-N_i)}{2} \right)$$

$$0 \leq f \leq 1$$

Une note égale à 1 veut dire que, après les processus d'intégration, on a obtenu des valeurs de possibilité et nécessité égales à 1 pour la bonne classe et la bonne cible et égales à 0 pour tout le reste de combinaisons erronées classe - cible. Une note égale à 0 veut dire que l'on est arrivé à avoir des valeurs égales à 0 pour la bonne paire classe - cible et égales à 1 pour le reste.

5.1.2 Fonction « B »

Pour la deuxième fonction considérée, on effectuera un processus de décision par comparaison avec un double seuil au niveau de chaque classe et chaque paramètre, avec trois résultats possibles : acceptation, rejet ou doute. Il y aura, donc, une dépendance des résultats par rapport aux seuils utilisés.

Dans le cadre de cette fonction, on considérera qu'il y a eu une réussite :

- Si l'on accepte un paramètre pour la bonne paire classe - cible
- Si l'on rejette un paramètre pour une paire classe - cible erronée

On considérera qu'il y a eu une erreur :

- Si l'on rejette un paramètre pour la bonne paire classe - cible
- Si l'on accepte un paramètre pour une paire classe - cible erronée

Après l'évaluation de tous les paramètres mesurés pour toutes les classes possibles pour une cible donnée, on comptabilise le nombre de réussites et d'erreurs et on calcule leurs pourcentages respectifs par rapport à la quantité totale de décisions prises. Le pourcentage de classifications « neutres » peut être obtenu comme le complémentaire de l'ensemble des pourcentages d'erreur et de réussite.

5.1.3 Fonction « C »

La troisième fonction d'évaluation développée dans le cadre de ce stage est un peu plus complexe que les précédentes. Elle travaille sur les valeurs finales de possibilité et nécessité pour chaque classe (après intégration temporelle et spatiale).

Elle s'appuie sur la notion d'erreur employée à Thales pour ses applications d'identification NCTR radar, selon laquelle il y a une erreur :

1. Si l'on rejette la bonne classe
2. Si l'on n'accepte pas la bonne classe et on accepte une autre classe (erronée).

L'objectif de la fonction C est de fournir une note globale qui prenne en compte, avant tout, le fait de ne pas avoir d'erreurs, mais aussi qui soit capable de mettre en valeur la puissance discriminante de l'algorithme d'intégration évalué : si l'on accepte toujours toutes les cibles, ou que l'on les déclare toujours « douteuses » (ni acceptés, ni rejetés), on aura un algorithme avec un taux d'erreur nul, mais qui ne semble pas trop performant quand même. C'est pourquoi la fonction C va noter les résultats de l'intégration en tenant compte, tout d'abord du statut affecté à la bonne classe mais aussi du nombre de classes acceptées et rejetées dans le cadre de chaque identification.

Voici un modèle schématique pour le calcul des résultats :

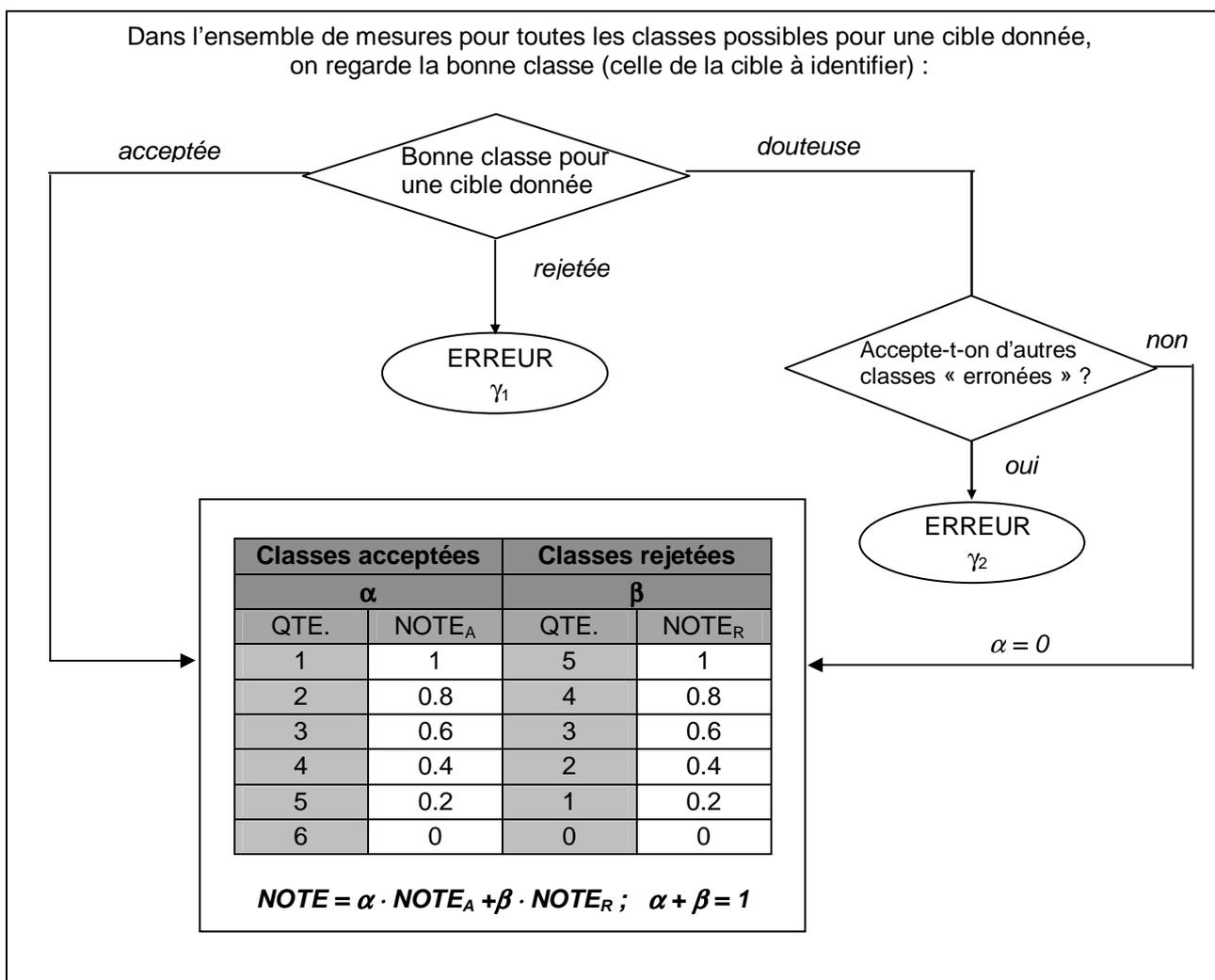


Figure 16. Schéma pour la fonction C

L'algorithme d'évaluation est le suivant :

1. On regarde la classification faite pour la bonne classe pour une cible donnée.
2. Si la bonne classe est rejetée, on donne une note γ_1 (négative) et on finit.
3. Si la bonne classe est acceptée, on donne une note qui dépend de la quantité de classes acceptées et rejetées en total (grâce au système de double seuil, une classe ne peut pas être acceptée et rejetée au même temps).
4. Si la bonne classe est douteuse, on regarde le reste de classes et ...
 - a) ... si une autre classe est acceptée, on donne une note γ_2 (négative) et on finit.
 - b) ... si aucune classe n'est acceptée, on donne une note qui dépend de la quantité de classes rejetées (et qui sera pondéré par le paramètre β de telle sorte que cela sera la note maximale).

Certaines propriétés de cette fonction peuvent être contrôlées grâce aux paramètres qui interviennent dans le processus d'évaluation :

- Les paramètres γ nous permettent de contrôler la pénalisation associée aux erreurs d'identification. On considérera ici des valeurs négatives (ex. $\gamma = -1$).
- Les paramètres α et β nous permettent d'établir l'importance relative d'accepter ou pas la bonne classe face à l'importance de rejeter les classes erronées.

5.1.4 Fonction taux d'erreur et taux de réussite

La dernière fonction utilisée est la plus conventionnelle de toutes, puisqu'elle fournit une valeur pour le taux d'erreur et un autre pour le taux de réussite selon le critère utilisé traditionnellement à Thales pour les systèmes d'identification NCTR :

Comme on a déjà expliqué, on considère qu'il y a une erreur :

1. Si l'on rejette la bonne classe
2. Si l'on n'accepte pas la bonne classe et on accepte une autre classe (erronée).

Par contre, on considère qu'il y a une réussite si l'on accepte la bonne classe et seulement la bonne classe pour une cible donnée.

A nouveau, on considérera ici les valeurs de possibilité et nécessité utilisées pour la prise de décision finale pour chaque classe, après intégration temporelle et spatiale.

Il s'agit d'une fonction qui nous apporte une vision très nette, très facile à comprendre, de la performance d'un algorithme d'intégration, puisqu'elle fournit une mesure du pourcentage de cibles bien et mal classées « au sens strict », ce qui la rend très appropriée en tant qu'indicateur de qualité globale.

5.2 MESURES SUR LES MATRICES DE CONFUSION

Les matrices de confusion ou tableaux de contingence sont des outils qui servent à évaluer la qualité d'une classification ou identification.

Les matrices de confusion que l'on utilisera dans le cadre de notre étude seront construites en mettant les données (cibles testées) sur les lignes et les résultats de la classification faite par notre système sur les colonnes. C'est ainsi que chaque valeur occupant la position ij donne une idée de la fréquence avec laquelle une cible de type i a été classifié comme appartenant à la classe j .

On va considérer des types différents de matrices de confusion et deux manières différentes de les construire pour chacun de ces types.

D'abord, on fera une distinction entre les matrices d'acceptation et les matrices de rejet :

- Les matrices d'acceptation expriment la probabilité avec laquelle on accepte les différentes classes pour une cible donnée. Dans le cas idéal, on acceptera seulement la bonne classe pour chaque cible considérée et on obtiendra, donc, une matrice dans laquelle tous les éléments qui ne se trouvent pas sur la diagonale sont nuls (matrice identité).

Exemple :

	C1	C2	C3	C4	C5	C6
T1	0.9637	0.7289	0.9912	0.9556	0.8928	0.0015
T2	0.9969	0.9969	0.6114	0.0748	0.0305	0
T3	0.0573	0.1293	0.9863	0.7093	1.0000	0.0020
T4	0.0789	0.2341	0.7875	0.9924	1.0000	0.2935
T5	0	0	0	0.2647	1.0000	0.2114
T6	0	0	0	0.8776	0.6327	1.0000

- Les matrices de rejet expriment la probabilité avec laquelle une classe donnée est rejetée au cours du processus d'identification pour une cible d'un certain type. Dans le cas idéal, tous les éléments de la diagonale seront nuls et les autres égaux à 1 puisque on ne rejettera jamais la bonne classe pour la cible à identifier et on rejettera toujours les autres.

Exemple :

0.0140	0.1255	0.0074	0.0282	0.0280	0.9985
0.0031	0.0031	0.0031	0.3844	0.9038	1.0000
0.8751	0.6029	0.0137	0.2674	0	0.9980
0.8932	0.5343	0.0381	0	0	0.6816
1.0000	1.0000	0.9846	0.7353	0	0.7853
1.0000	1.0000	0.9796	0.0612	0.1837	0

On fera aussi une deuxième distinction par rapport à la manière dont les matrices sont construites :

- On travaillera avec des matrices normalisées, avec des valeurs entre 0 et 1, qui expriment le pourcentage de temps (~probabilité) pendant lequel chaque classe a été acceptée ou rejetée pour chaque type de cible. (voir exemple ci-dessus).
- On travaillera aussi avec des matrices non normalisées, qui montreront la quantité totale de classifications effectuées pour chaque paire classe - cible, chaque unité correspondant à un instant dans un axe temporel discrétisé.

Exemple :

685	596	707	680	665	1
805	805	487	52	20	0
50	89	600	433	612	1
31	92	276	326	329	59
0	0	0	161	652	128
0	0	0	43	31	49

Bien que la simple observation des matrices de confusion offre déjà une idée générale de la qualité du processus de classification évalué, on travaillera avec des mesures quantitatives plus rigoureuses, objectives et précises.

5.2.1 Erreur quadratique

On définira cette mesure pour les matrices à valeurs normalisées. Elle nous apporte une indication du degré de dissimilitude par rapport aux matrices idéales d'acceptation (matrice identité) et de rejet (matrice de uns moins matrice identité) :

$$E_{Q,accept} = \sum_i^n \sum_j^n \left((Id - M_{A,n})_{ij} \right)^2$$

$$E_{Q,rejet} = \sum_i^n \sum_j^n \left(((ones - Id) - M_{R,n})_{ij} \right)^2$$

5.2.2 Indicateurs de précision

On calculera ces indicateurs sur les matrices non normalisées d'acceptation.

On parlera de trois types de précision :

- La **précision pour l'utilisateur** est le pourcentage de fois qu'une certaine classe a été bien affectée à son type cible correspondant (une valeur par colonne).

$$P_{util,j} = \frac{M_{ij}}{\sum_i M_{ij}}$$

- La **précision pour le réalisateur** est le pourcentage de fois que, pour une cible d'un certain type, on a affecté la bonne classe (une valeur par ligne).

$$P_{real,i} = \frac{M_{ii}}{\sum_j M_{ij}}$$

- La **précision totale** est la relation entre la quantité de bonnes identifications et la quantité totale d'identifications effectuées (une valeur par table).

$$P = \frac{\sum_k M_{kk}}{\sum_i \sum_j M_{ij}}$$

Exemple.

685	596	707	680	665	1	0.2055
805	805	487	52	20	0	0.3711
50	89	600	433	612	1	0.3361
31	92	276	326	329	59	0.2929
0	0	0	161	652	128	0.6929
0	0	0	43	31	49	0.3984
0.4360	0.5088	0.2899	0.1923	0.2824	0.2059	0.3293

- Précisions pour l'utilisateur
- Précisions pour le réalisateur
- Précision générale

5.2.3 Coefficient kappa

On travaillera ici, à nouveau, avec les matrices de confusion d'acceptation avec des valeurs non normalisées.

Le coefficient Kappa est une mesure de concordance sur un processus de classification ou identification faite par deux observateurs. Il a été utilisé initialement dans le domaine de la médecine et la psychiatrie, afin de mesurer le degré de concordance entre les diagnostics fournis par plusieurs experts. D'autres domaines d'application habituelle incluent aujourd'hui, par exemple, celui de la classification de pixels pour le traitement d'images satellitaires.

La définition originelle de l'indice kappa se fait à partir l'expression suivante, où l'on travaille avec une matrice de confusion contenant des valeurs de probabilité :

$$\kappa = \frac{\sum \pi_{ii} - \sum \pi_{i+} \cdot \pi_{+i}}{1 - \sum \pi_{i+} \cdot \pi_{+i}}$$

ou π_{ii} sont les éléments de la diagonale, π_{i+} est la somme de tous les éléments de la ligne i et π_{+i} es la somme de tous les éléments de la colonne i .

Cette expression est très utile pour avoir une vision intuitive de la sémantique du coefficient. En effet, on peut supposer des variables aléatoires indépendantes qui représentent la classification d'un même objet (ou cible) par deux observateurs. La probabilité pour qu'un objet soit classifié par les deux dans la même catégorie C est donnée par $\pi_{C+} \cdot \pi_{+C}$. Si on étend ce raisonnement à toutes les catégories possibles, $\sum_C \pi_{C+} \cdot \pi_{+C}$ est la probabilité de ce que les observateurs coïncident par des causes attribuables au hasard exclusivement. Cet ainsi que la valeur kappa κ montre la relation entre l'excès de concordance au delà de celui attribuable au hasard $\left(\sum_C \pi_{CC} - \sum_C \pi_{C+} \cdot \pi_{+C} \right)$ et l'excès maximum possible pour le cas de classification idéale $\left(1 - \sum_C \pi_{C+} \cdot \pi_{+C} \right)$.

Au delà de cette approche basée sur l'utilisation de probabilités, on peut obtenir une estimation de l'indice kappa en remplaçant celles-ci par des fréquences d'échantillonnage :

$$\kappa = \frac{\sum \pi_{ii} - \sum \pi_{i+} \cdot \pi_{+i}}{1 - \sum \pi_{i+} \cdot \pi_{+i}} \cong \frac{\sum \left(\frac{X_{ii}}{N} \right) - \sum \left(\frac{X_{i+}}{N} \right) \cdot \left(\frac{X_{+i}}{N} \right)}{1 - \sum \left(\frac{X_{i+}}{N} \right) \cdot \left(\frac{X_{+i}}{N} \right)} = \frac{N \sum X_{ii} - \sum X_{i+} \cdot X_{+i}}{N^2 - \sum X_{i+} \cdot X_{+i}}$$

ou N est la somme de toutes les valeurs de la matrice de confusion non normalisée (X).

Dans le cadre de notre étude, on utilisera toujours cette dernière expression.

En ce qui concerne les possibles valeurs de kappa, on trouve que la concordance maximale possible correspond à une valeur $\kappa=1$. On obtient $\kappa=0$ pour un degré de concordance qui coïncide avec celui que l'on pourrait attendre à cause du hasard seul. Si la concordance est plus grande que celle attribuable au hasard, kappa sera positif et vice-versa. La valeur minimum de kappa dépend des distributions marginales X_{i+} et X_{+i} .

On peut aussi considérer l'échelle numérique proposée dans [12], résumée dans ce tableau :

Valeur de κ	Force de la concordance
< 0.20	Pauvre
0.21 – 0.40	Faible
0.41 – 0.60	Modérée
0.61 – 0.80	Bonne
0.81 – 1.00	Très bonne

Un traitement plus rigoureux de ce coefficient impliquerait un travail avec des intervalles de confiance en fonction du nombre d'échantillons considérés, mais au cours de nos simulations nous nous limiterons à travailler avec des valeurs ponctuelles qui resteront valides, en tout cas, pour nous donner une idée générale de l'évolution des performances pour la classification associées à chacun des algorithmes étudiés.

6 Etude sur l'intégration temporelle pour la variable possibilité

On arrive à la section la plus importante de ce rapport, qui constitue le noyau du travail effectué au cours de mon stage.

Après avoir décrit un modèle concret d'architecture pour un système d'identification NCTR radar, on étudiera des approches envisageables pour l'intégration temporelle des données floues et on comparera leurs performances à l'aide des outils introduits dans la section précédente.

Il est important de noter que, dans cette section, on travaillera seulement avec des valeurs de possibilité, suivant des indications de la part de Thales. Les mesures de nécessité étant encore en phase de perfectionnement, il semble plus intéressant pour notre étude de les laisser un peu de côté afin de nous centrer sur le traitement de la possibilité. On se limitera donc à réaliser quelques brèves observations dans la section suivante à propos de la nécessité et son utilisation conjointe avec la possibilité.

Le fait de travailler exclusivement avec la possibilité implique le besoin d'adopter un nouveau schéma pour la prise de décision par rapport à celui décrit à la fin de la section dédiée au sujet de la NCTR. On maintiendra les trois états possibles pour chaque cible et chaque classe : acceptation, rejet et doute, mais on utilisera une stratégie à double seuil plus simple, selon laquelle on rejettera une classe pour une cible donnée si la valeur finale de possibilité est inférieure à un seuil de rejet, on l'acceptera si cette valeur est supérieure à un seuil dit d'acceptation et on la déclarera « douteuse » (ou « inconnue ») dans tous les autres cas.

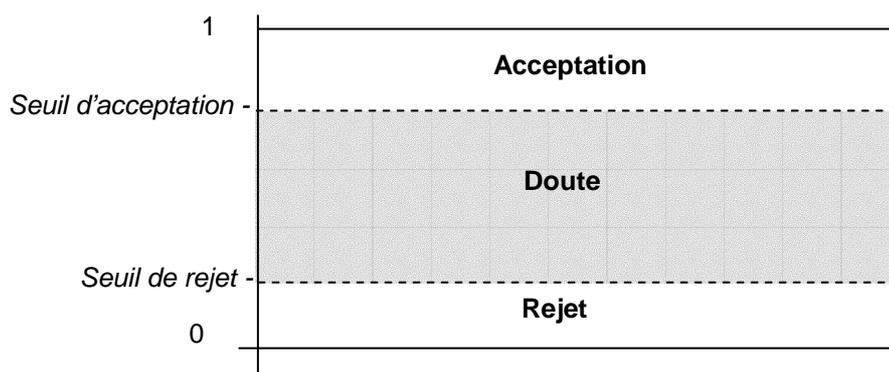


Figure 17. Modèle à double seuil

6.1 DESCRIPTION DU CONTEXTE DE SIMULATION ET DES DONNEES UTILISEES

On travaillera avec des cibles aériennes de six types ou classes différentes, que l'on désignera avec des chiffres (C1, C2, C3, C4, C5, C6) et dont on ne connaît pas l'identité réelle pour des raisons de confidentialité.

Pour les cinq premiers types de cibles (T1 à T5), on travaillera avec trois jeux de données réelles, correspondant à trois mesures différentes fournies par des capteurs radar réels, obtenues dans des conditions diverses et couvrant des intervalles temporelles différents (mais égaux au niveau de chaque ensemble, c'est à dire, les conditions et les intervalles considérés pour chacun des trois ensembles de données sont les mêmes pour toutes les cibles). Pour la sixième classe de cible on disposera uniquement d'un seul jeu de données simulées. Naturellement, on connaît en chaque cas la cible réelle à laquelle appartient chacune des mesures considérées et on utilisera cette connaissance pour évaluer les performances des différents algorithmes.

En ce qui concerne les paramètres utilisés pour la classification, ils seront de deux types que l'on dénotera par A et B, à nouveau à cause de critères de confidentialité. En tout cas, il s'agit de paramètres utilisés habituellement dans le cadre des systèmes NCTR, similaires à ceux dont on a

parle dans la section de la NCTR radar . On considérera trois paramètres de type A (que l'on appellera A1, A2 et A3) et deux paramètres de type B (que l'on appellera B1 et B2).

Toutes les fonctions d'évaluation utilisées ont été conçues pour donner comme résultat les valeurs correspondantes aux paramètres type A et type B isolés (après l'intégration spatiale partielle A1+A2+A3 et B1+B2, mais avant l'intégration spatiale globale A+B) et aussi la valeur finale après intégration spatiale entre les ensembles A et B (en total, trois valeurs par évaluation).

Dans la suite de ce rapport, on utilisera le critère suivant pour toutes les figures représentant des résultats des ces trois types :

- **Couleur rouge, symbole étoile (*) et trait pointé (.) pour les paramètres type A**
- **Couleur vert, symbole triangle (Δ) et trait tirets (--) pour les paramètres type B**
- **Couleur bleu, symbole croix (x), et trait continu (-) pour les résultats de la fusion des paramètres A et B.**

Pour les figures présentant des résultats pour les mesures sur des matrices de confusion, on respectera toujours les critères suivants :

- **Couleur bleu, symbole croix (x), et trait continu (-) pour la précision et l'erreur quadratique d'acceptation.**
- **Couleur rouge, symbole étoile (*), trait pointé (.) pour le coefficient kappa et l'erreur quadratique de rejet.**

6.2 LES RESULTATS EN ABSENCE D'INTEGRATION TEMPORELLE

Il semble intéressant de regarder la structure des données considérés et les résultats obtenus pour les fonctions d'évaluation et les matrices de confusion dans un cas hypothétique où l'on n'effectue aucun type d'intégration temporelle, c'est à dire, les valeurs pour la possibilité cumulée sont égales à celles de la possibilité instantanée.

Si l'on représente les histogrammes de ces données avec à une division de l'intervalle [0,1] en dix sous - intervalles, on obtient ces résultats pour la bonne paire cible - classe et pour les paires erronées :

Bonne classe (T_i, C_i):

Valeur	≥ 0	≥ 0.1	≥ 0.2	≥ 0.3	≥ 0.4	≥ 0.5	≥ 0.6	≥ 0.7	≥ 0.8	≥ 0.9	=1
Occurrences	9	0	2	1	1	1	2	4	1	1	4150

Classes erronées ($T_i, C_j \ i \neq j$):

Valeur	≥ 0	≥ 0.1	≥ 0.2	≥ 0.3	≥ 0.4	≥ 0.5	≥ 0.6	≥ 0.7	≥ 0.8	≥ 0.9	=1
Occurrences	6225	374	435	498	372	361	409	451	320	331	11054

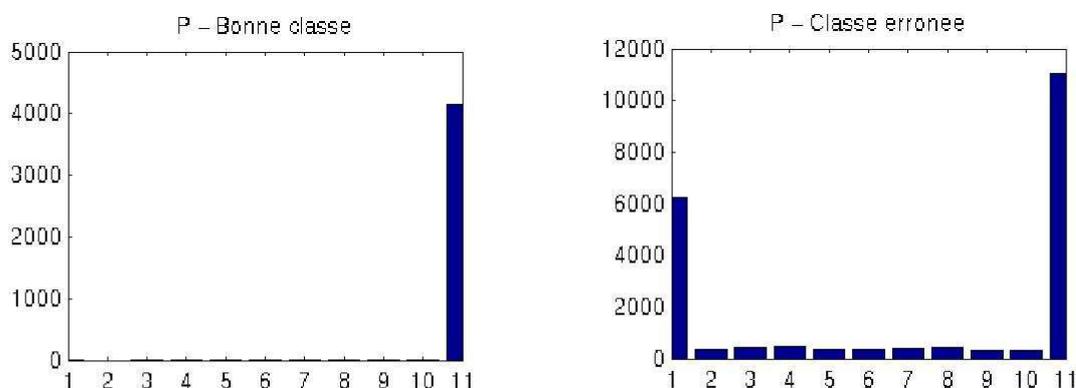


Figure 18. Distribution des valeurs de possibilité pour les données sans intégration temporelle

On observe que pour la bonne paire classe - cible les valeurs sont presque toujours égales à 1. Par contre, pour les paires erronées il y a de nombreuses valeurs au-dessous de 1 et une grande quantité de valeurs proches de 0.

Les fonctions d'évaluation donnent ces résultats :

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.5809	0.3483	0.2326

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.7633	0.7187	0.6195	0.1949	0.2536	0.3494

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.1547	0.1260	0.1260

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.1430	0	0.0937
Erreur	0.0101	0.0016	0.0089

Voici la matrice de confusion normalisée pour l'acceptation :

0.9637	0.7289	0.9912	0.9556	0.8928	0.0015
0.9969	0.9969	0.6114	0.0748	0.0305	0
0.0573	0.1293	0.9863	0.7093	1.0000	0.0020
0.0789	0.2341	0.7875	0.9924	1.0000	0.2935
0	0	0	0.2647	1.0000	0.2114
0	0	0	0.8776	0.6327	1.0000

Erreur quadratique : 9.1751

Voici la matrice non normalisée avec les mesures de précision :

685	596	707	680	665	1	0.2055
805	805	487	52	20	0	0.3711
50	89	600	433	612	1	0.3361
31	92	276	326	329	59	0.2929
0	0	0	161	652	128	0.6929
0	0	0	43	31	49	0.3984
0.4360	0.5088	0.2899	0.1923	0.2824	0.2059	<u>0.3293</u>

kappa = 0.3016

Et voici finalement la matrice de confusion pour le rejet:

0.0140	0.1255	0.0074	0.0282	0.0280	0.9985
0.0031	0.0031	0.0031	0.3844	0.9038	1.0000
0.8751	0.6029	0.0137	0.2674	0	0.9980
0.8932	0.5343	0.0381	0	0	0.6816
1.0000	1.0000	0.9846	0.7353	0	0.7853
1.0000	1.0000	0.9796	0.0612	0.1837	0

Erreur quadratique : 11.6448

Ces résultats que l'on vient d'obtenir seront très intéressants dans la suite puisqu'ils nous serviront de référence, de point de départ, pour avoir une idée plus précise de la qualité des résultats après intégration temporelle.

6.3 LES OPERATEURS DE MOYENNE

Un opérateur de moyenne est une fonction $m ; [0,1] \times [0,1] \rightarrow [0,1]$ telle que :

- le résultat de la combinaison est toujours compris entre le min et le max.
 $\forall (x, y) \in [0,1]^2, \min(x,y) \leq m(x,y) \leq \max(x,y)$, mais m n'est pas l'opérateur min ni max
- m est commutative
- m est croissante par rapport aux deux variables :
 $\forall (x, x', y, y') \in [0,1]^4, (x \leq x' \text{ et } y \leq y') \Rightarrow m(x,y) \leq m(x',y')$ [1]

6.3.1 La méthode « k sur M »

La première approche que l'on a considérée est celle de la méthode dite « k sur M », proposée comme méthode de départ par mes responsables à Thales.

Il s'agit d'une méthode assez intuitive qui consiste à considérer un vecteur contenant les M dernières valeurs mesurées pour la possibilité instantanée, les ranger en ordre croissant et prendre celle de position k comme nouvelle valeur pour la possibilité cumulée à chaque instant. Une valeur de $k=1$ équivaut à prendre le minimum des dix derniers échantillons, tandis que une valeur $k=10$ est l'équivalent de prendre le maximum.

Dans la documentation interne que j'ai eu l'opportunité de regarder à mon arrivée, on proposait d'utiliser une valeur de $k=3$, afin d'éviter quelques erreurs dues à des possibles valeurs anormales, et contrôler toujours le taux d'erreur.

Une première idée intuitive a été celle d'évaluer les performances de cet algorithme d'intégration en faisant varier le paramètre k entre 1 et 10.

Les résultats, présentés dans les figures ci-dessous, montrent une tendance monotone dans tous les cas. Les mesures de qualité globale obtenues pour les fonctions A, B et C s'améliorent quand la valeur de k diminue, de telle sorte que les meilleurs résultats pour ces fonctions s'obtiennent pour $k=1$, c'est à dire, pour le minimum. Par contre, le taux d'erreur global augmente dans la même ligne.

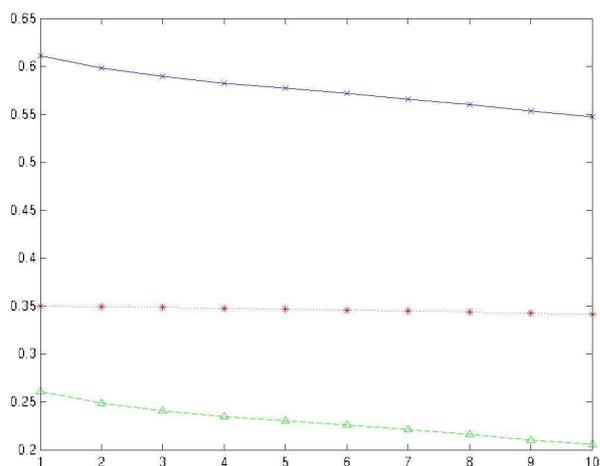


Figure 20. Fonction A - "k sur M"

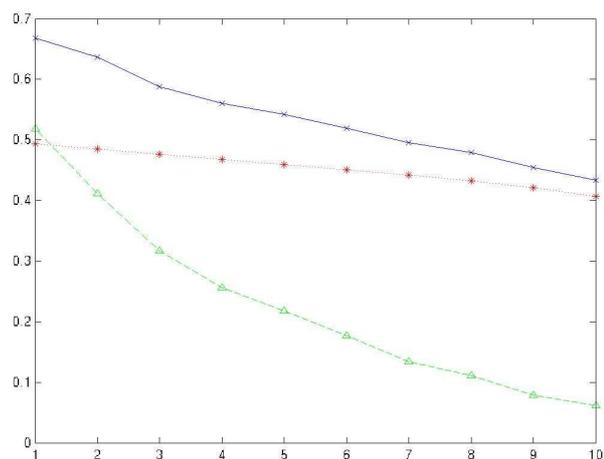


Figure 19. Fonction C - "k sur M"

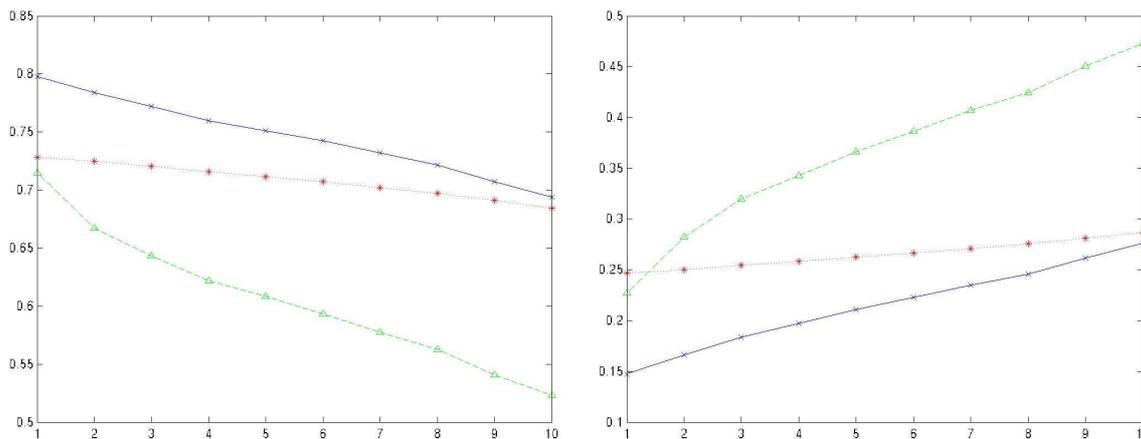


Figure 21. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "K sur M"

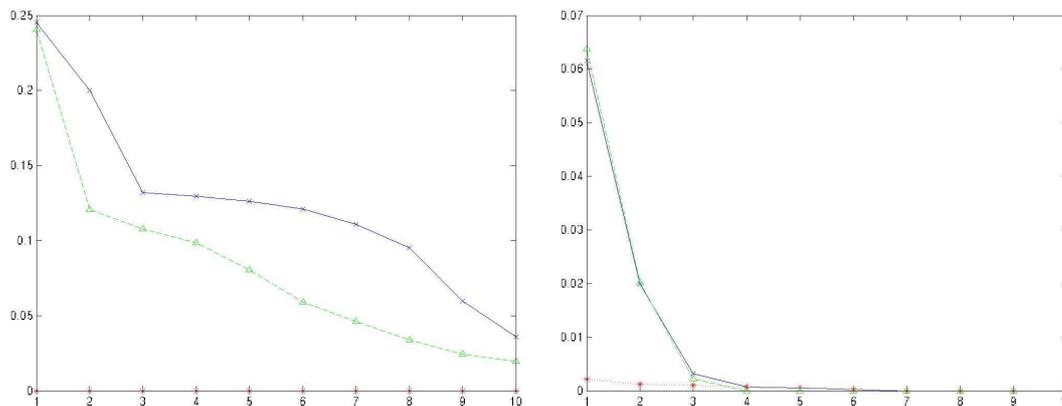


Figure 22. Taux de réussite (gauche) et d'erreur (droite) - "k sur M"

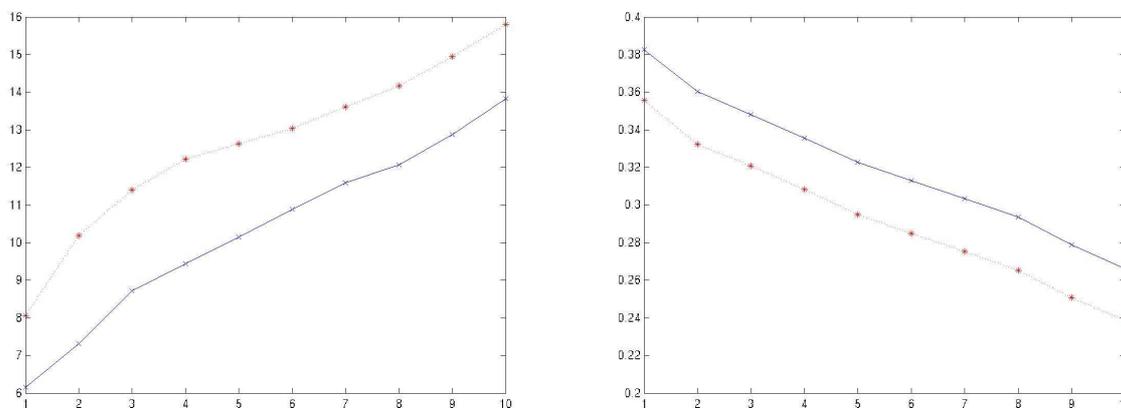


Figure 23. Erreurs quadratiques (gauche) et précision-kappa (droite) – « k sur M »

Voici un résumé des résultats quantitatifs le plus importants qui accompagnent les graphiques d'au-dessus :

Fonction A

k	FUSION FINALE	PARAMETRES A	PARAMETRES B
1	0.6113	0.3501	0.2612
3	0.5893	0.3485	0.2408
10	0.5473	0.3416	0.2057

Fonction B

k	TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
	FUSION	A	B	FUSION	A	B
1	0.7979	0.7283	0.7150	0.1474	0.2467	0.2271
3	0.7718	0.7205	0.6431	0.1838	0.2546	0.3202
10	0.6936	0.6844	0.5232	0.2764	0.2867	0.4731

Fonction C

k	FUSION FINALE	PARAMETRES A	PARAMETRES B
1	0.1772	0.1318	0.1434
3	0.1617	0.1274	0.1024
10	0.1181	0.1090	0.0206

k	Taux	FUSION FINALE	PARAMETRES A	PARAMETRES B
1	Réussite	0.2454	0	0.2405
	Erreur	0.0615	0.0021	0.0637
3	Réussite	0.1321	0	0.1078
	Erreur	0.0032	0.0010	0.0023
10	Réussite	0.0360	0	0.0197
	Erreur	0	0	0

Matrices de confusion:

	k = 1	k = 3	K = 10
Erreur quadratique acceptation	6.1555	8.7214	13.8264
Erreur quadratique rejet	8.0644	11.4002	15.8039
Précision générale	0.3826	0.3482	0.2660
Coefficient kappa	0.3557	0.3208	0.2385

Intuitivement, ce comportement peut être expliqué par le fait que prendre une valeur petite de k implique une stratégie plus « risquée » pour le processus de classification, c'est à dire, plus discriminante, mais à la fois plus susceptible de provoquer des erreurs d'identification pour la bonne classe. Le fait de prendre une valeur plus grande, par contre, suppose une attitude plus conservatrice : on acceptera plus de classes pour chaque identification, on deviendra moins discriminant, mais on commettra moins d'erreurs.

6.3.2 Moyennes généralisées

Une fois qu'on a étudié l'approche « k sur M », et sans sortir du cadre des opérateurs de moyenne, on peut considérer un autre type d'opérateurs qui peuvent être utilisés pour la fusion d'informations floues : les moyennes généralisées.

L'idée mathématique derrière cette approche est celle d'utiliser une expression générale dépendant d'un paramètre α , telle que en faisant varier sa valeur on puisse parcourir les différents types de

moyennes classiques (arithmétique, géométrique, harmonique, etc.) sur un chemin allant de l'opérateur *min* jusqu'à l'opérateur *max* :

$$h_{\alpha}(a_1, a_2, \dots, a_n) = \left(\frac{1}{n} \sum_i a_i^{\alpha} \right)^{\frac{1}{\alpha}}$$

$\alpha \rightarrow -\infty : h_{-\infty}(A) \rightarrow \min_i(a_i)$	Opérateur <i>min</i>
$\alpha = -1 : h_{-1}(A) = \frac{n}{\sum_{i=1}^n \frac{1}{a_i}}$	M. harmonique
$\alpha = 0 : h_0(A) = \left(\prod_{i=1}^n (a_i) \right)^{\frac{1}{n}}$	M. géométrique
$\alpha = 1 : h_1(A) = \frac{\sum_{i=1}^n a_i}{n}$	M. arithmétique
$\alpha \rightarrow +\infty : h_{+\infty}(A) \rightarrow \max_i(a_i)$	Opérateur <i>max</i>

Dans les figures suivantes, on observe les variations des performances pour une variation logarithmique de α entre -100 et 100 (selon une expression du type $\alpha = \pm 10^{\frac{1}{2}x-2}$, avec x entre 0 et 8):

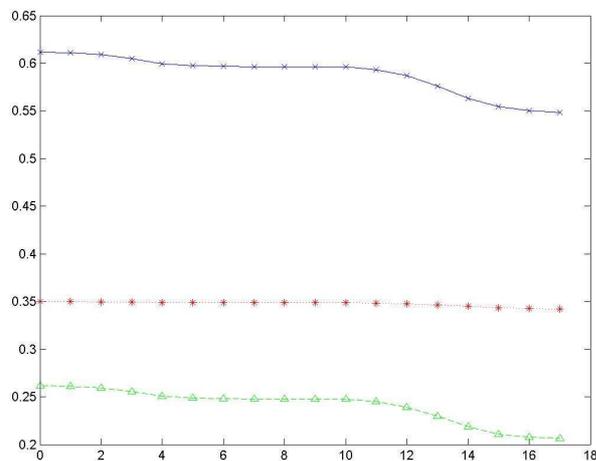


Figure 24. Fonction A - "Moyennes gen."

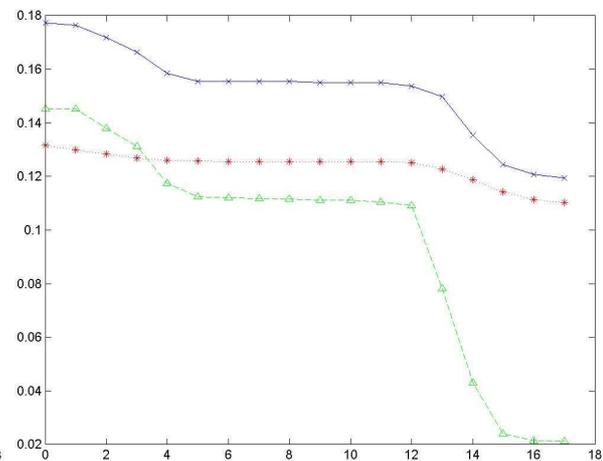


Figure 25. Fonction C - "Moyennes gen."

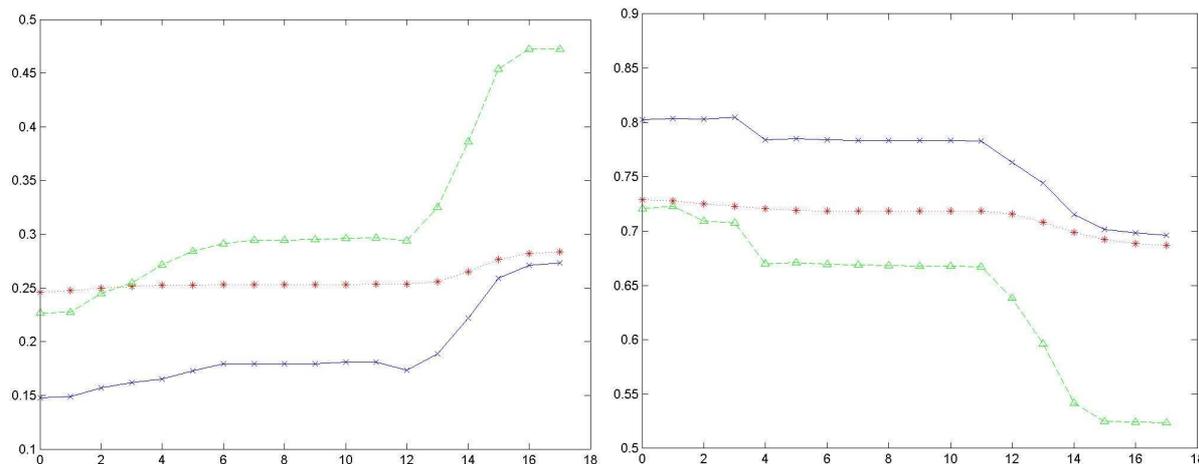


Figure 26. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Moyennes gen."

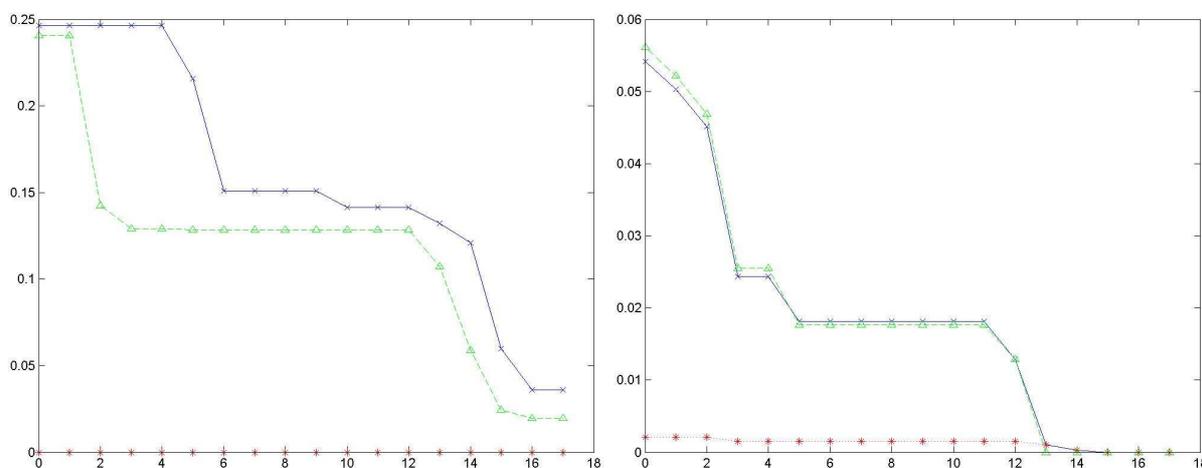


Figure 27. Taux de réussite et d'erreur - "Moyennes gen."

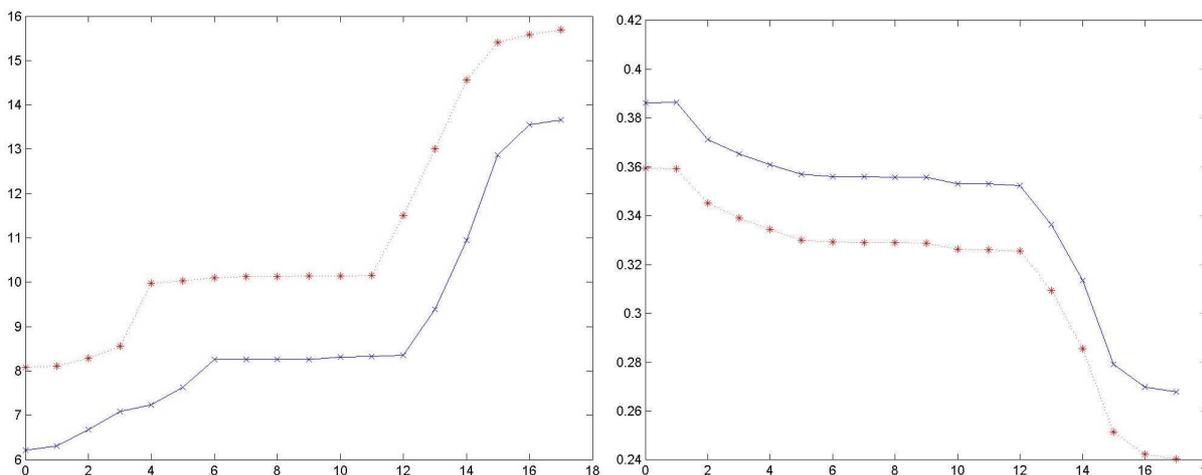


Figure 28. Erreurs quadratiques et précision-kappa – « Moyennes gen. »

A cause du comportement monotone observé, il semble intéressant de regarder les valeurs numériques correspondant aux valeurs extrêmes du paramètre, mais puisque pour celles-ci on obtient un opérateur qui va vers le *min* et le *max*, on renvoie le lecteur aux résultats obtenus dans la section précédente pour les valeurs $k = 1$ (*min*) et $k = 10$ (*max*).

En effet, les moyennes généralisées nous apportent une alternative à l'approche « k sur M » qui parcourt le même intervalle de performances traversant ce que l'on pourrait considérer comme « un chemin différent ». Parmi les avantages introduits par les moyennes on peut nommer une précision

plus grande grâce à la possibilité de faire varier le paramètre α de manière continue et aussi la méthode de calcul employée (selon les circonstances et l'environnement de calcul à notre disposition, il peut être ou non plus intéressant de calculer une moyenne arithmétique, par exemple, que de ranger un vecteur par ordre croissant et choisir un élément dans une position déterminée).

6.3.3 Ordered Weighted Average (OWA)

Les opérateurs OWA (*Ordered Weighted Average*) sont un type particulier d'opérateurs de moyenne pondérée qui ont été introduits par Yager dans son article de 1988 [14].

Si a_1, a_2, \dots, a_n sont les valeurs à intégrer, elles doivent être ordonnées dans une suite telle que :

$$a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(n)}$$

L'opérateur OWA est paramétré par une n-tuple de poids, $W=(w_1, \dots, w_n)$ tels que :

$$\sum_{i=1}^n w_i = 1 \quad \forall i, 1 \leq i \leq n, \text{ et } w_i \in [0,1]$$

Ceci dit, la fusion de données par un opérateur OWA est définie par la relation :

$$h_{OWA}(a_1, a_2, \dots, a_n) = \sum_{i=1}^n w_i a_{(i)}$$

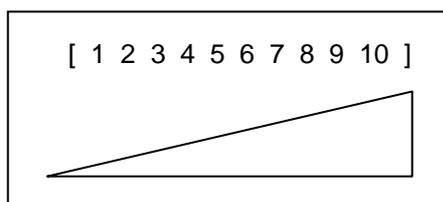
6.3.3.1 Approche « naïve »

Une première option pour travailler avec ce type d'opérateurs est celle de choisir quelques formes géométriques plutôt simples et intuitives pour la distribution de poids, comme par exemple une distribution triangulaire croissante, triangulaire décroissante ou pyramidale.

Il s'agit d'une approche assez naïve, à laquelle il manque la rigueur formelle de celles que l'on considérera plus tard, mais qui peut apporter des informations d'intérêt à propos du comportement des opérateurs OWA.

6.3.3.2 Distribution de poids triangulaire croissante

Ceci dit, on commence par considérer une distribution de poids de forme triangulaire croissante, selon la pondération (normalisée) :



On peut considérer des permutations circulaires successives afin d'étudier les performances pour des positions différentes du poids maximal dans la distribution, toujours avec la forme relative ci-dessus.

Les figures ci-jointes montrent les performances obtenues. L'axe des x indique la position du poids maximal :

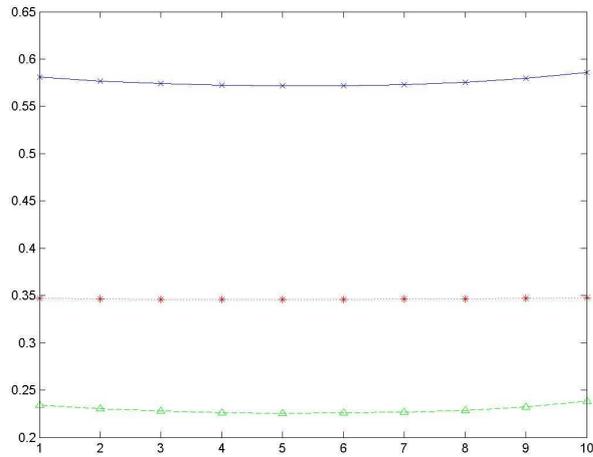


Figure 29. Fonction A - "OWA- T.croissante"

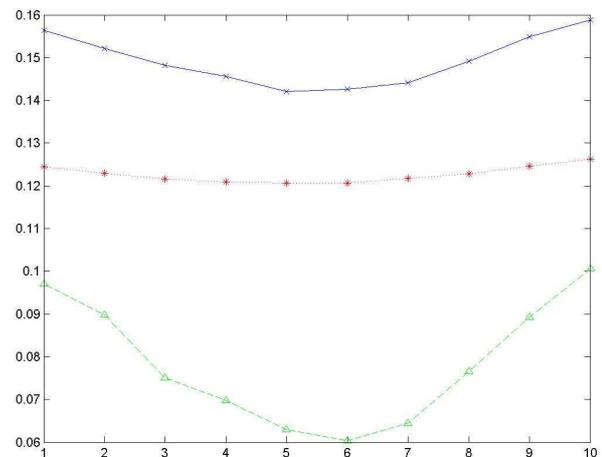


Figure 30. Fonction C - "OWA- T.croissante"

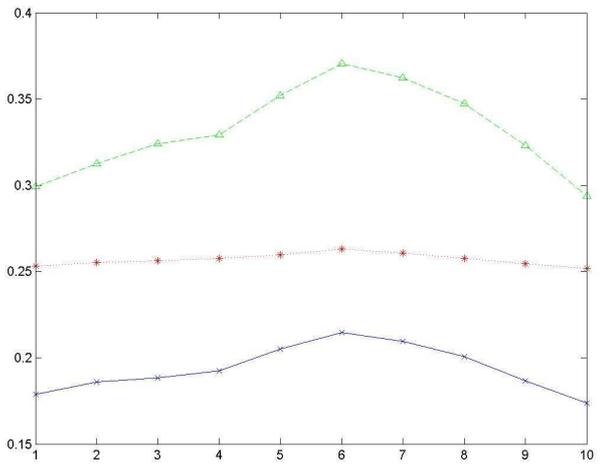
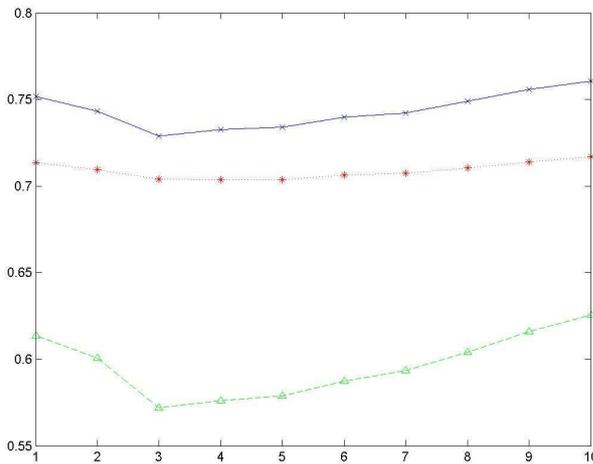


Figure 31. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "OWA - T. croissante"

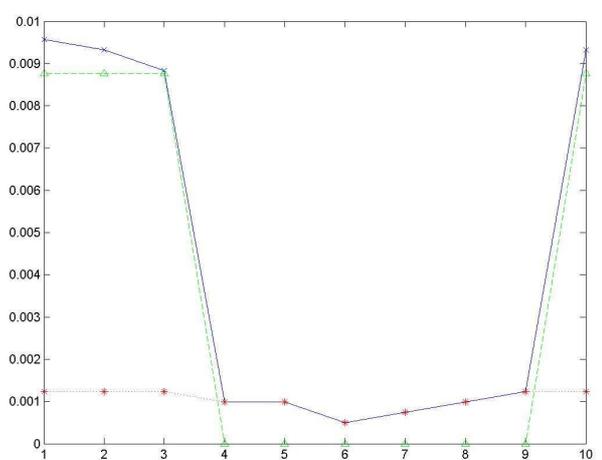
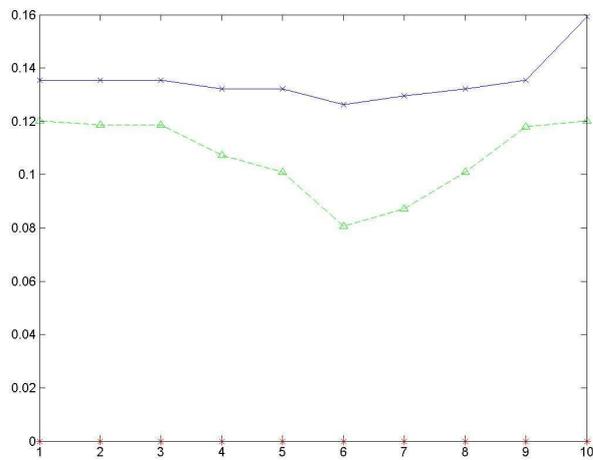


Figure 32. Taux de réussite et d'erreur - "OWA - T. croissante"

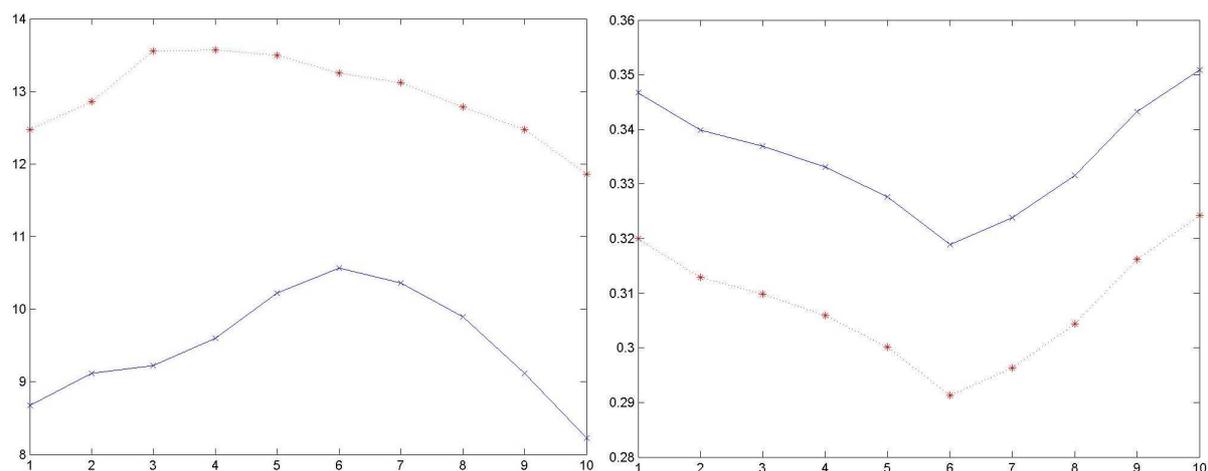
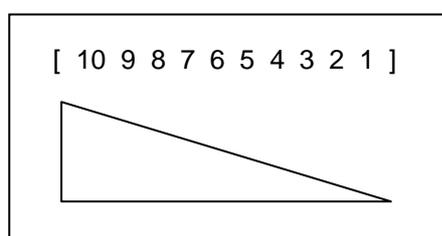


Figure 33. Erreurs quadratiques et précision-kappa – «OWA - T. croissante»

Les performances observées pour les fonctions d'évaluation (et à exception de le taux d'erreur globale), toujours au-dessous de celles obtenues pour l'opérateur min, sont meilleures quand le poids maximal se trouve en dernière position (distribution triangulaire « croissante pure »). Cela est logique, puisque cette distribution est la plus proche à l'opérateur min, et les résultats obtenus sont donc cohérents avec ceux de la section précédente. En général, on observe que quand la distribution de poids est telle que les valeurs maximales obtiennent la plus part du poids, les performances générales deviennent pires (ce qui arrive pour des valeurs centrales pour le poids maximum) et vice-versa.

6.3.3.3 Triangulaire décroissante

Si l'on continue avec l'esprit de l'approche naïve, on peut considérer une distribution de poids complémentaire à celle que l'on vient d'utiliser, une distribution triangulaire décroissante :



Si l'on applique un critère analogue au cas antérieur pour les rotations, on obtient des résultats complémentaires à ceux obtenus auparavant :

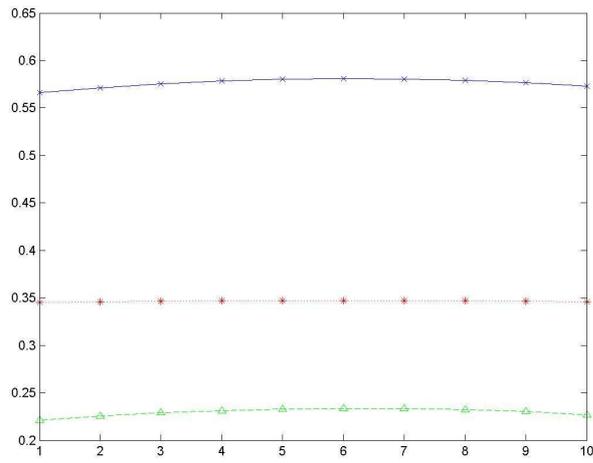


Figure 34. Fonction A - "OWA- T. décroissante"

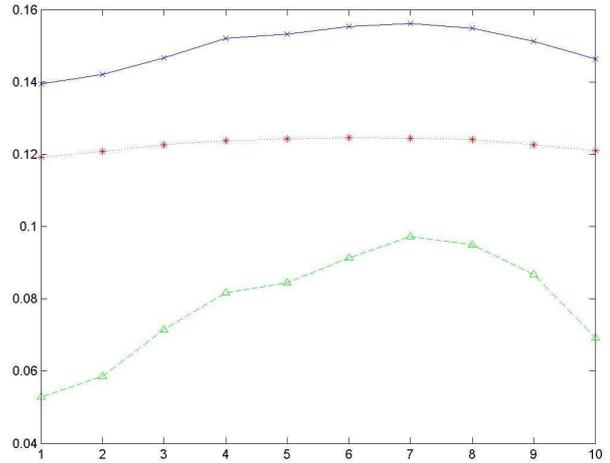


Figure 35. Fonction C - "OWA- T. décroissante"

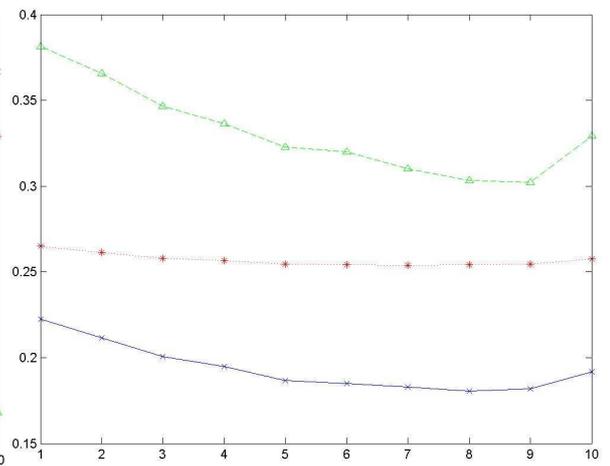
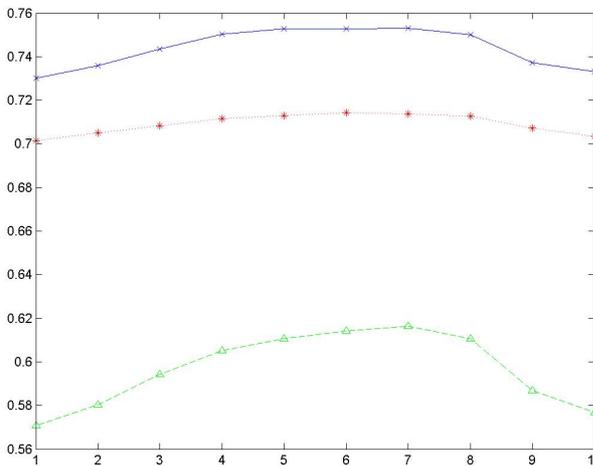


Figure 36. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "OWA - T. décroissante"

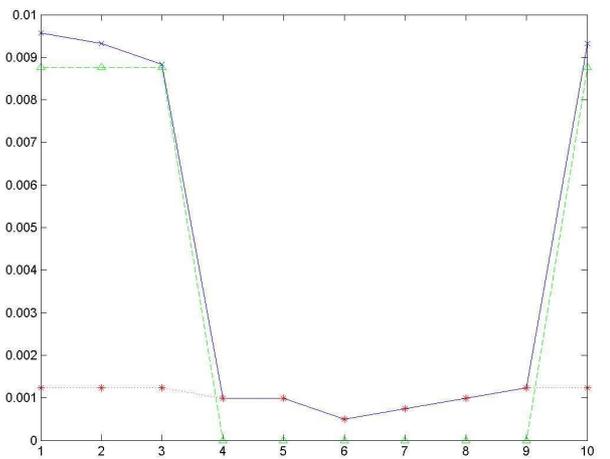
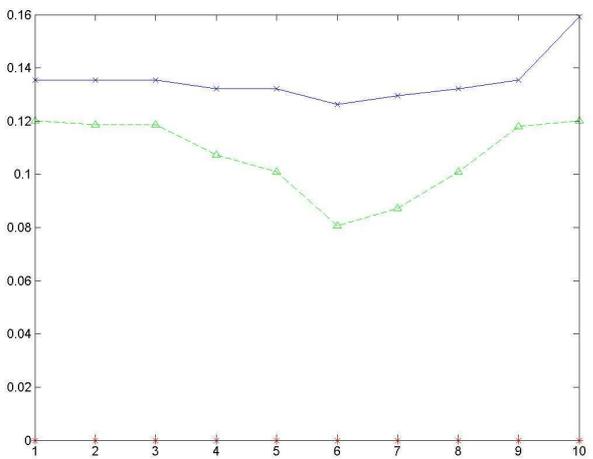


Figure 37. Taux de réussite et d'erreur - "OWA - T. décroissante"

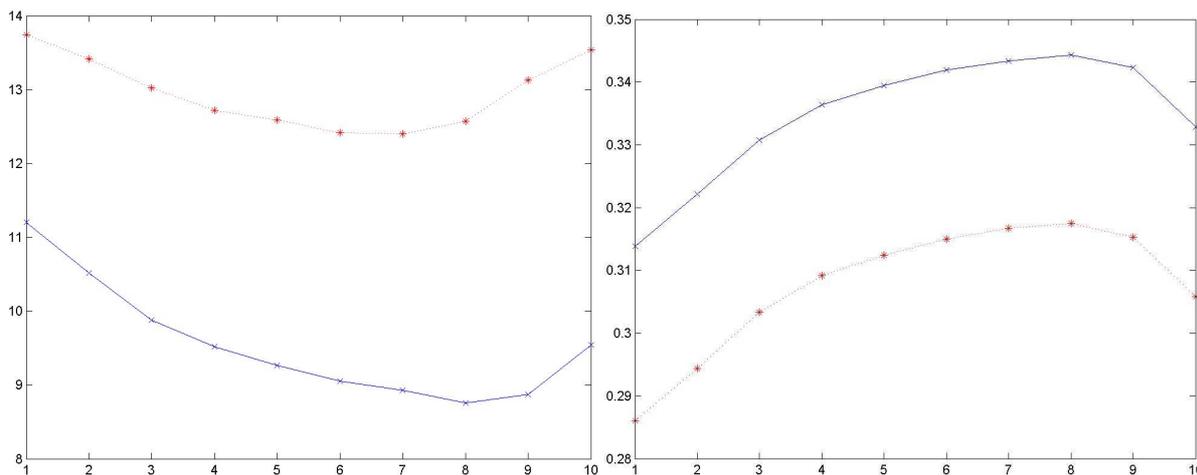
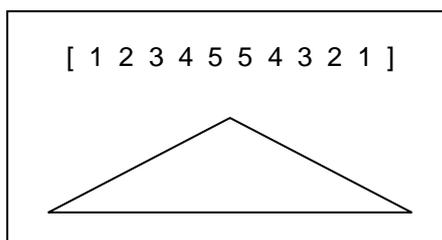


Figure 38. Erreurs quadratiques et précision-kappa – «OWA - T. décroissante»

Maintenant les meilleurs résultats, qui restent toujours assez au-dessous de ceux obtenus pour l'opérateur min « pur », s'obtiennent pour les positions de poids maximal qui offrent une proportion totale du poids plus grande aux valeurs les plus petites du vecteur ordonné, c'est-à-dire, aux valeurs minimales de l'ensemble de valeurs à intégrer.

6.3.3.4 Pyramidale

Une troisième idée dans le cadre de l'approche OWA avec des distributions « naïves » est d'utiliser une distribution de structure pyramidale avec des poids relatifs (normalisés) de la forme :



Si, à nouveau, on fait varier la position du maximum, on trouve les performances suivantes :

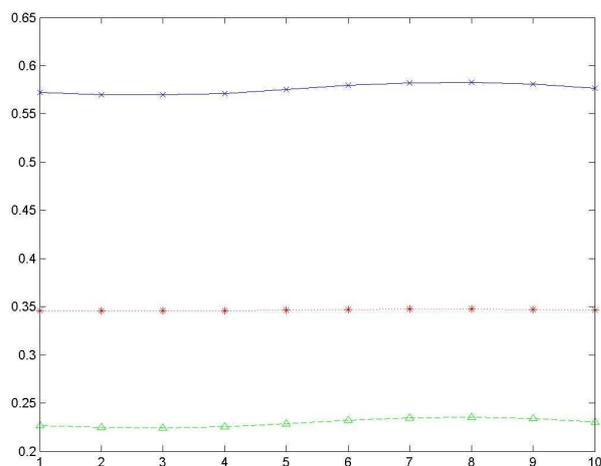


Figure 39. Fonction A - "OWA- T. pyramidale"

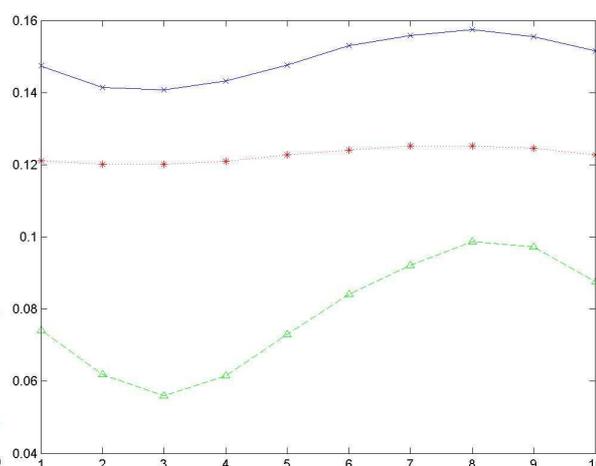


Figure 40. Fonction C - "OWA- pyramidale"

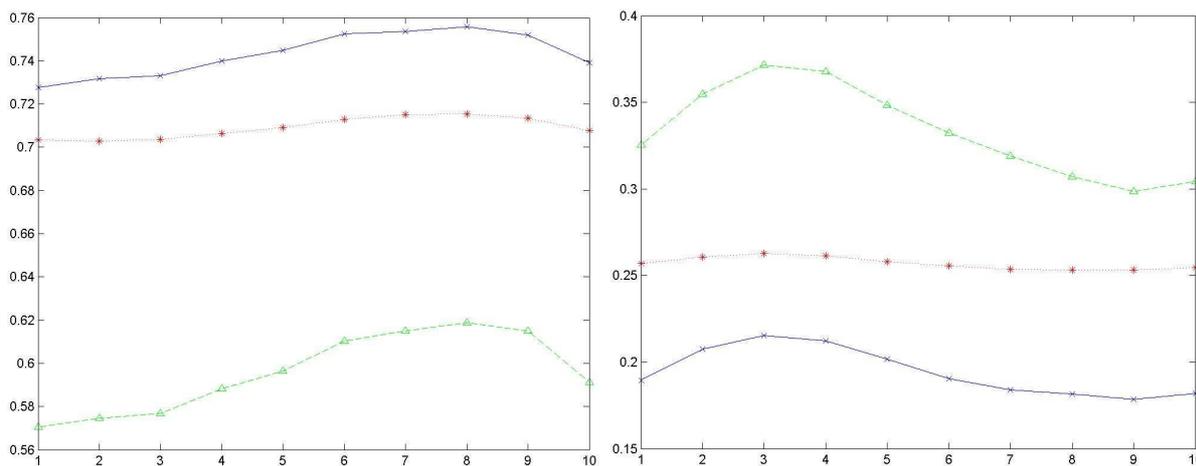


Figure 41. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "OWA - pyramidale"

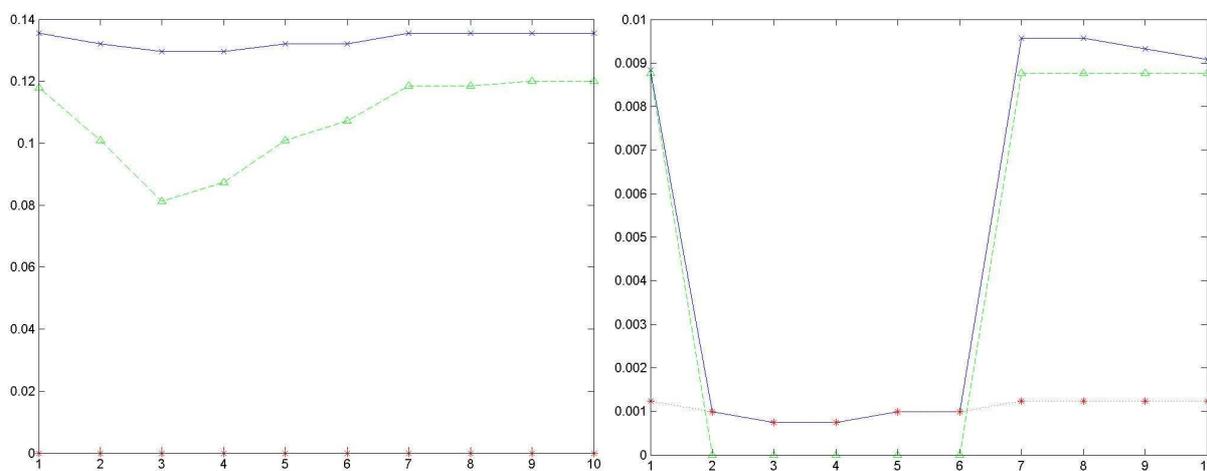


Figure 42. Taux de réussite et d'erreur - "OWA - pyramidale"

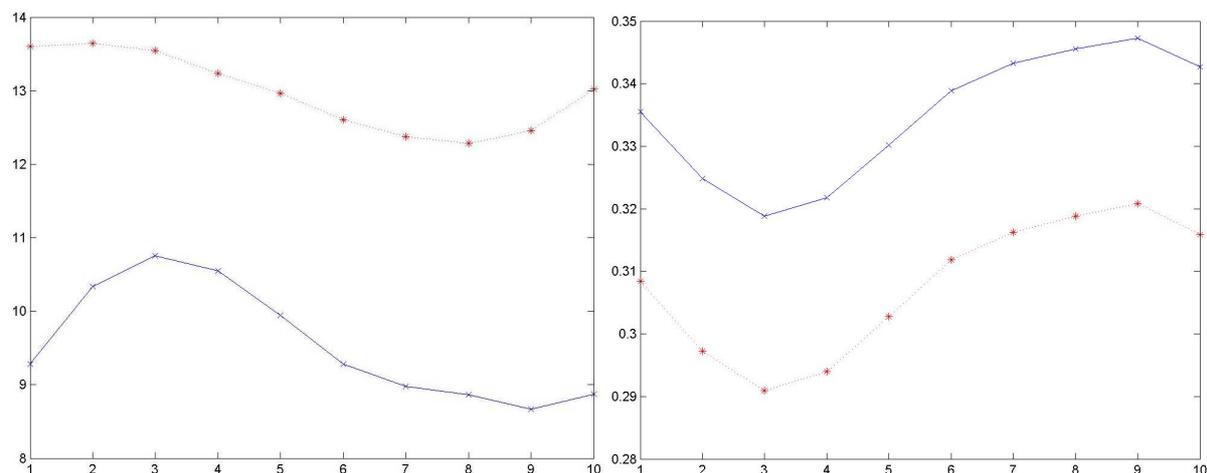


Figure 43. Erreurs quadratiques et précision-kappa – «OWA - pyramidale»

A nouveau, on trouve les performances les plus attirantes (sauf taux d'erreur) pour les distributions qui montrent une tendance à concentrer la plupart du poids dans la partie du vecteur des valeurs à intégrer correspondant aux valeurs minimales. Les premières observations dérivées de l'utilisation des opérateurs OWA semblent recommander donc de s'approcher des caractéristiques de l'opérateur *min* pour améliorer les performances.

En tout cas, les résultats obtenus correspondent à un comportement assez « conservateur », peu « risquée », avec un taux de réussite (toujours inférieur au 16%) et un taux d'erreur (toujours inférieur au 1%), valeurs globales très basses pour tous les distributions et tous les rotations considérées.

6.3.3.5 MEOWA

Au delà de ces premières tentatives à caractère plutôt « naïf », les opérateurs OWA offrent une variété presque illimitée de stratégies envisageables pour le choix du vecteur de poids et donc pour la fusion des données floues.

Ici, on essaiera d'établir un critère plus rigoureux, plus systématique, pour l'évaluation de ce type d'opérateurs. On commencera, tout d'abord, par définir quelques concepts d'intérêt dans le domaine des OWA:

- La mesure de « **orness** » est un indicateur de la similarité entre un opérateur OWA donnée et l'opérateur OR (max). On aura toujours $orness(W) \in [0,1]$.

$$orness(w) = \frac{1}{n-1} \sum_{i=1}^n ((n-i)w_i)$$

- La mesure de « **andness** » est une mesure complémentaire de l'antérieur et indique la similarité entre l'opérateur OWA considéré et l'opérateur AND (min). A nouveau, on aura toujours $andness(W) \in [0,1]$.

$$andness(w) = 1 - orness(w) = \frac{1}{n-1} \sum_{i=1}^n ((i-1)w_i)$$

- Finalement, on peut définir un troisième indicateur, celui de la **dispersion**, qui montre le degré jusqu'au quel un opérateur OWA prend en compte toute l'information disponible dans le vecteur de données à intégrer pour le processus de fusion. La notion de dispersion est proche de celle d'entropie. Malgré l'existence de nombreux types d'entropie (entropies de Rényi, entropies d'ordre β introduites par Daróczy, entropies R-normes, etc.) on se placera ici près de l'approche originelle proposée par Shannon en 1948 et on considèrera une mesure de dispersion normalisée de la forme :

$$dispersion(w) = -\frac{1}{\ln n} \sum_{i=1}^n (w_i \ln w_i)$$

Dans la suite, on utilisera les résultats présentés par Péter Majlender dans son article « *OWA operators with maximal entropy* » [15]. Cet article décrit l'approche originellement proposée par O'Hagan [16] pour obtenir un vecteur de poids OWA avec l'entropie maximale Shannon pour une certaine valeur d'orness.

C'est ainsi que l'on essaiera d'évaluer les performances des opérateurs OWA en faisant varier la valeur d'orness considérée et en calculant en chaque cas le vecteur de poids associé (qui aura toujours l'entropie maximale à cause du procédé choisi pour le calcul).

Mathématiquement, on devra résoudre un problème de maximisation de la forme :

$$\begin{aligned} &\text{maximiser} && -\sum_{i=1}^n w_i \ln w_i \\ &\text{étant donné} && \frac{1}{n-1} \sum_{i=1}^n ((n-i)w_i) = \alpha, && 0 \leq \alpha \leq 1 \\ &&& \sum_{i=1}^n w_i = 1 && 0 \leq w_i \leq 1, i = 1, \dots, n \end{aligned}$$

La solution générique pour un opérateur OWA de n composants proposée par O'Hagan (dont la démonstration et l'explication en détail se trouvent dans l'article de Majlender) est basée sur l'utilisation de ces formules :

$$\begin{aligned} \text{Obtention de } w_1 : w_1 [(n-1)\alpha + 1 - mw_1]^n &= ((n-1)\alpha)^{n-1} [((n-1)\alpha - n)w_1 + 1] \\ \text{Reste de valeurs : } w_n &= \frac{((n-1)\alpha - n)w_1 + 1}{(n-1)\alpha + 1 - mw_1} \\ w_j &= \sqrt[n-1]{w_1^{n-j} w_n^{j-1}} \end{aligned}$$

Si l'on utilise ces expressions pour évaluer les performances des opérateurs MEOWA en fonction des différentes valeurs de orness dans le cadre de notre étude sur l'intégration temporelle de la variable possibilité, on obtient des résultats de la forme :

[L'axe x des figures montre le degré d'orness, pour 101 valeurs dans l'intervalle [0,1] telles que orness=0.01·x, 0≤x≤100]

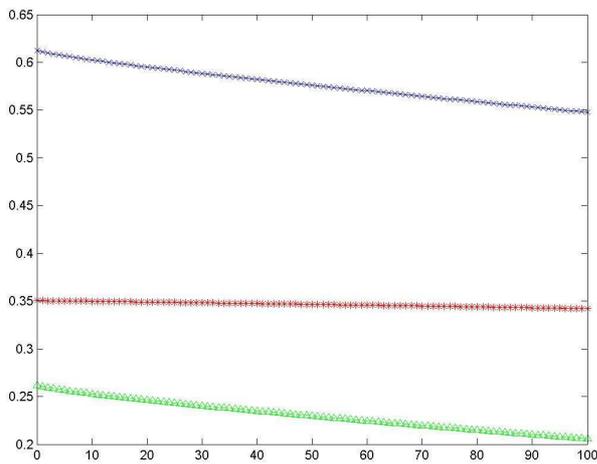


Figure 44. Fonction A - "MEOWA"

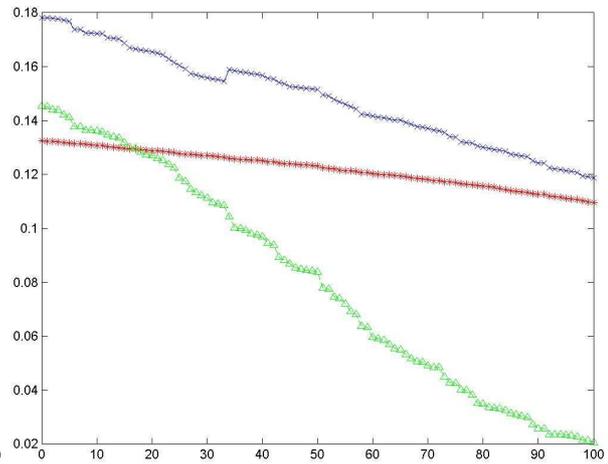


Figure 45. Fonction C - "MEOWA"

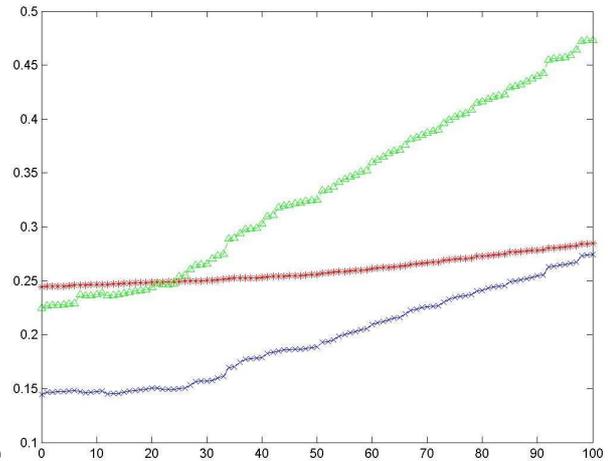
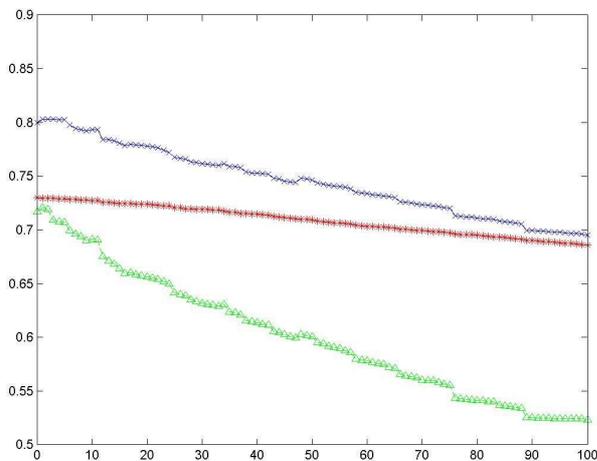


Figure 46. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "MEOWA"

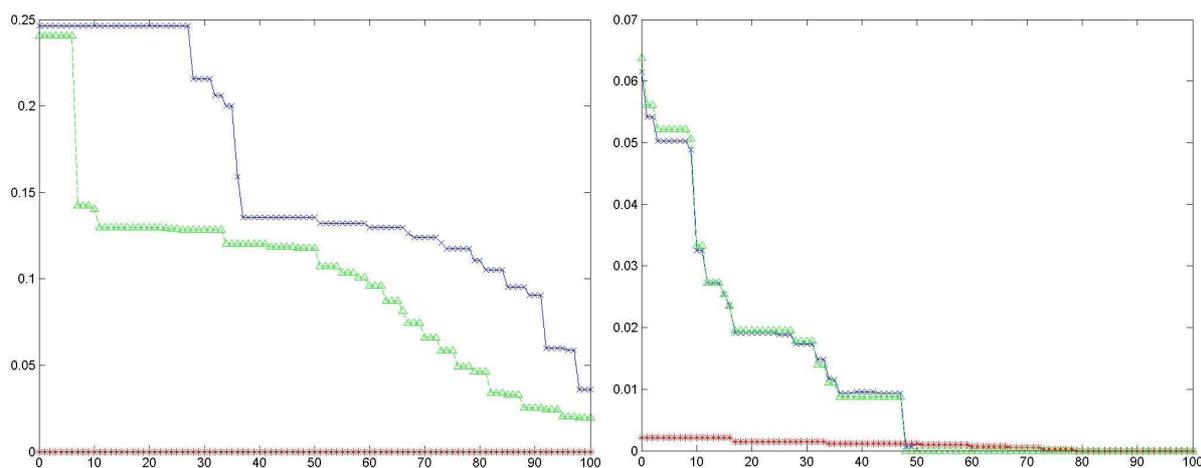


Figure 47. Taux de réussite et d'erreur - "MEOWA"

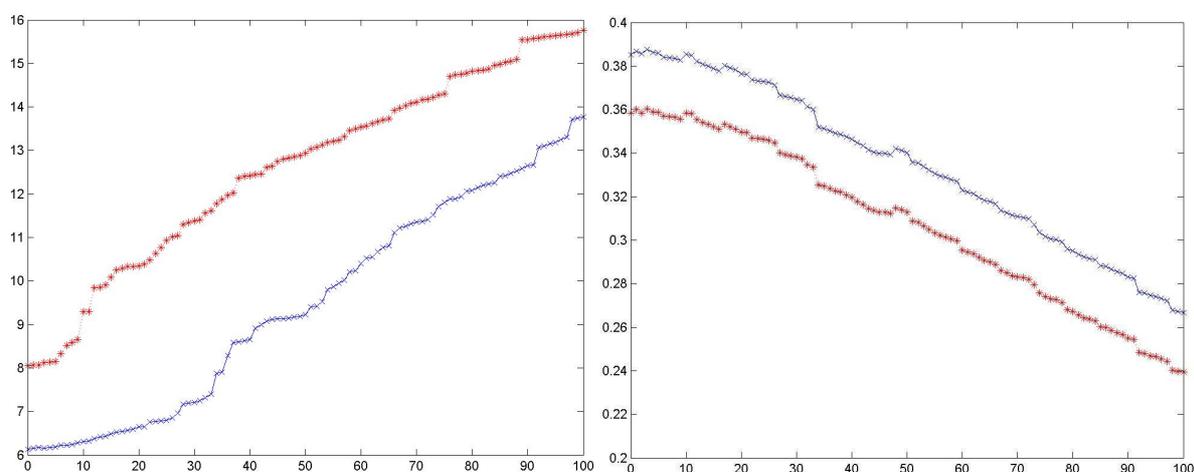


Figure 48. Erreurs quadratiques et précision-kappa – «MEOWA»

L'observation des résultats obtenus nous permet d'obtenir des conclusions similaires à celles obtenues pour l'approche « k sur M », où l'on se déplaçait aussi entre le min et le max, et les meilleures performances en termes de discrimination s'obtenaient pour le min avec une pénalisation importante en terme de taux d'erreur.

A nouveau, on trouve que les performances générales sauf le taux d'erreur améliorent quand l'*orness* diminue, jusqu'à un maximum (pour lequel le taux général de réussite se trouve proche de 25%) obtenu pour *orness*=0 (opérateur AND ou min). Par contre, le fait d'adopter une approche plus risquée comme celle de l'opérateur min cause une augmentation très importante du taux d'erreur, qui monte au-dessus de 6%, valeur non tolérable dans le cadre d'applications militaires dans lequel l'on travaille.

En tout cas, l'utilisation des ME-OWA nous apporte un outil très intéressant pour le futur, puisque l'on obtient une variation avec un caractère beaucoup plus continu que celle associée aux opérateurs « k sur M », parcourant le même intervalle de performances, ainsi qu'une précision théoriquement infinie pour le choix du point de l'espace métaphorique entre les opérateurs min et max où l'on veut se placer (par opposition aux dix possibilités discrets de « k sur M »).

6.3.3.6 Apprentissage des poids

Cette troisième approche pour le travail avec des opérateurs OWA utilise l'idée décrite par Ronald Yager et Dimitar Filev dans leur article d'avril de 1999 « *Induced Ordered Weighted Averaging Operators* ».

Les chercheurs Filev et Yager ont proposé dans leur article [13] une méthode pour l'apprentissage du vecteur de poids associé à un opérateur OWA à partir des observations préalables que l'on explique par la suite. Il s'agit d'une technique qui se rapproche de la théorie des réseaux de neurones et qui est aussi basée sur la méthode de la descente du gradient.

On considérera que l'on a une collection de m ensembles de données $(a_{k1}, a_{k2}, \dots, a_{kn})$ que l'on ordonnera dans les ensembles $(b_{k1}, b_{k2}, \dots, b_{kn})$ et pour chacun desquels on connaît la valeur agrégée idéale d_k associé. On peut essayer de trouver une méthode itérative qui nous permette « d'apprendre » la distribution de poids idéale $W^T = [w_1, w_2, \dots, w_n]$ telle que :

$$b_{k1}w_1 + b_{k2}w_2 + \dots + b_{kn}w_n = d_k \text{ pour } k \text{ entre } 1 \text{ et } m$$

En fait, on cherchera le vecteur de poids W^T qui minimise l'erreur quadratique « instantanée » :

$$e_k = (b_{k1}w_1 + b_{k2}w_2 + \dots + b_{kn}w_n - d_k)^2$$

A nouveau, on trouve un problème d'optimisation avec les contraintes imposées par la définition des OWA:

1. $\sum_{i=1}^n w_i = 1$
2. $\forall i, 1 \leq i \leq n$, et $w_i \in [0,1]$

Ce problème peut être transformé en un problème non linéaire sans contraintes si l'on représente chacun des poids OWA selon l'expression :

$$w_i = \frac{e^{\lambda_i}}{\sum_{j=1}^n e^{\lambda_j}}$$

On remplace cette valeur dans l'expression de l'erreur e_k , et puis on applique une technique de descente de gradient, similaire à celles utilisées pour la rétro-propagation dans le domaine des réseaux de neurones, de telle sorte que l'on obtienne une règle pour l'actualisation des valeurs de λ de la forme :

$$\lambda_i(l+1) = \lambda_i(l) - \beta \left. \frac{\partial e_k}{\partial \lambda_i} \right|_{\lambda_i = \lambda_i(l)}$$

où β est le « taux d'apprentissage » ($0 < \beta < 1$) et $\lambda_i(l)$ est l'estimation de λ_i après la l -ième itération.

On dénotera par \hat{d}_k l'estimation de la valeur agrégée d_k , et donc, on a :

$$\hat{d}_k = b_{k1}w_1 + b_{k2}w_2 + \dots + b_{kn}w_n$$

Et la dérivée partielle à calculer sera de la forme :

$$\frac{\partial e_k}{\partial \lambda_i} = w_i (b_{ki} - \hat{d}_k) (\hat{d}_k - d_k) \quad i=[1,n]$$

A partir des expressions précédentes on obtient la règle finale pour l'actualisation de lambda :

$$\lambda_i(l+1) = \lambda_i(l) - \beta w_i (b_{ki} - \hat{d}_k) (\hat{d}_k - d_k)$$

Ceci dit, l'algorithme proposé par Yager et Filev pour l'apprentissage du vecteur de poids est le suivant :

1. On commence avec une estimation arbitraire des valeurs $\lambda_i(l)$ (par exemple, $\lambda_i(l)=0$, qui correspond à un vecteur de poids avec tous ses éléments égaux à $1/n$) et une nouvelle observation de valeurs ordonnées $(b_{k1}, b_{k2}, \dots, b_{kn})$ avec leur valeur agrégée correspondante d_k .

2. On calcule l'estimation actuelle des poids à partir de $\lambda_i(l)$:

$$w_i(l) = \frac{e^{\lambda_i(l)}}{\sum_{j=1}^n e^{\lambda_j(l)}}$$

3. On utilise les poids estimés pour calculer une estimation de la valeur agrégée à partir du vecteur de valeurs à intégrer :

$$\hat{d}_k = b_{k1}w_1(l) + b_{k2}w_2(l) + \dots + b_{kn}w_n(l)$$

4. Finalement, on actualise les valeurs des λ_i estimés :

$$\lambda_i(l+1) = \lambda_i(l) - \beta w_i(l) (b_{ki} - \hat{d}_k) (\hat{d}_k - d_k)$$

5. On répète le processus un certain nombre de fois ou jusqu'à ce que l'on arrive à avoir une erreur d'approximation inférieure à un certain seuil.

Dans notre cas, les valeurs à intégrer sont des ensembles de valeurs de possibilités instantanées et la valeur agrégée de référence pour chacun de ces ensembles sera 1 pour les bonnes paires classe - cible et 0 pour les paires erronées.

Si l'on choisit un taux d'apprentissage $\beta=0.35$, il suffit de très peu d'itérations pour observer une tendance évidente vers le vecteur de poids $W^T = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]$, c'est-à-dire, le vecteur de poids qui minimise l'erreur quadratique et qui semble être celui qui correspond à l'opérateur min. On constate avec plaisir que ce résultat est cohérent avec ceux obtenus pour tous les autres opérateurs jusqu'à présent :

Itération	w ₁	w ₂	w ₃	w ₄	w ₅	w ₆	w ₇	w ₈	w ₉	w ₁₀
0	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
1	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	0.0008	0.0009	0.0012	0.9933
2	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006	0.9966
3	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003	0.0004	0.9977
4	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.9983
5	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.9986
6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.9988
7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.9990
8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.9991
9	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.9992
10	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.9993

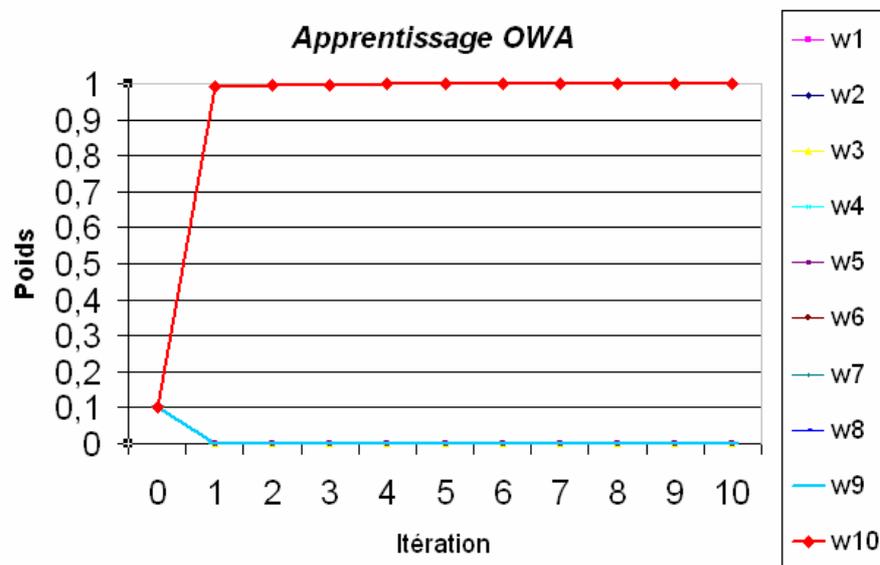


Figure 49. Résultats de l'apprentissage pour le vecteur de poids OWA

6.3.4 Conclusions sur les opérateurs de moyenne

A partir de l'approche « k sur M », on a commencé notre étude des différents opérateurs de fusion pour l'intégration temporelle avec les opérateurs de moyenne. On a fait des simulations pour des moyennes généralisées et pour des opérateurs OWA.

Les résultats obtenus nous permettent de tirer quelques conclusions :

- Le pouvoir discriminant de l'algorithme augmente quand on s'approche de l'opérateur min. Cela signifie que l'on adopte une approche plus « risquée », qui implique une augmentation des taux de réussite et erreur globales.
- C'est ainsi que l'opérateur min est celui qui nous permet d'optimiser les résultats obtenus à travers les différentes fonctions d'évaluation, ainsi que les mesures sur les matrices de confusion et le taux général de réussite. Par contre, le taux d'erreur obtenu avec cet opérateur est trop élevé pour les requis des applications dans le cadre desquelles on travaille.
- Ceci dit, il semble souhaitable d'approfondir dans la direction de l'opérateur min et, éventuellement, d'essayer de trouver une solution pour contrôler et réduire le taux d'erreur associé.
- En tout cas, les opérateurs étudiés dans cette section restent des outils avec des caractéristiques très intéressantes dont la possible utilisation dans le cadre de la fusion floue NCTR ne doit pas être rejetée face à l'avenir.

6.4 NORMES TRIANGULAIRES

Tout d'abord, on fera une description rapide des normes triangulaires en tant qu'opérateurs utilisés pour la fusion de données floues. Cette description a été extraite de l'ouvrage [1].

En général, on peut considérer que les t-normes généralisent la notion d'intersection aux ensembles flous ainsi que le « et » logique.

Le concept de norme triangulaire provient du domaine de la géométrie stochastique et fait référence à une fonction $t : [0,1] \times [0,1] \rightarrow [0,1]$ telle que :

- t est commutative : $\forall (x, y) \in [0,1]^2, t(x, y) = t(y, x)$
- t est associative : $\forall (x, y, z) \in [0,1]^3, t[t(x, y), z] = t[x, t(y, z)]$
- 1 est élément neutre : $\forall x \in [0,1], t(x, 1) = t(1, x) = x$
- t est croissante par rapport aux deux variables :
 $\forall (x, x', y, y') \in [0,1]^4, (x \leq x', y \leq y') \Rightarrow t(x, y) \leq t(x', y')$

En plus, on trouve que $t(0,1) = t(0,0) = t(1,0) = 0$, $t(1,1) = 1$, et 0 est élément absorbant : $(\forall x \in [0,1], t(x, 0) = 0)$.

Le « min » est la plus grande t-norme et donc on peut considérer que toute t-norme a un comportement conjonctif. En effet, on vérifie :

$$\forall (x, y) \in [0,1]^2, t(x, y) \leq \min(x, y)$$

A l'opposé, on peut considérer la t-norme dite « drastique », qui est toujours plus petite que toutes les autres, et qui est définie par :

$$\forall (x, y) \in [0,1]^2, t_0(x, y) = \begin{cases} x, & \text{si } y = 1 \\ y, & \text{si } x = 1 \\ 0 & \text{autrement} \end{cases}$$

Par conséquent, on peut affirmer que toutes les t-normes seront comprises entre le min et la t-norme drastique mais, en tout cas, il faut noter qu'il n'y a pas d'ordre total sur l'ensemble des t-normes.

A partir d'une t-norme et d'une complémentation c , un autre opérateur T appelé t-conorme peut être défini par dualité :

$$\forall (x, y) \in [0,1]^2, T(x, y) \Rightarrow c[t(c(x), c(y))]$$

Les t-conormes généralisent aux ensembles flous la notion de réunion ou de « ou » logique, par opposition au caractère conjonctif des t-normes.

Selon un raisonnement parallèle à celui utilisé pour les t-normes, on peut dire que les t-conormes sont des fonctions $T : [0,1] \times [0,1] \rightarrow [0,1]$ telles que :

- T est commutative, associative et croissante par rapport aux deux variables.
- 0 est élément neutre : $\forall x \in [0,1], t(x, 0) = t(0, x) = x$

De plus, on trouve que $T(0,1) = T(1,1) = T(1,0) = 1$, $T(0,0) = 0$, et 1 est élément absorbant : ($\forall x \in [0,1], t(x, 1) = 1$).

L'opérateur « max » est la t-conorme la plus petite, c'est-à-dire, tout t-conorme vérifie :

$$\forall (x, y) \in [0,1]^2, T(x, y) \geq \max(x, y)$$

Et, à nouveau, on peut définir une t-conorme plus grande que le reste, dite drastique, selon l'expression :

$$\forall (x, y) \in [0,1]^2, T_0(x, y) = \begin{cases} x, & \text{si } y = 0 \\ y, & \text{si } x = 0 \\ 1 & \text{autrement} \end{cases}$$

Quelques propriétés d'intérêt pour les t-normes et t-conormes cités par I. Bloch dans son introduction [1] sont :

- Toute t-norme ou t-conorme est distributive par rapport au « min » et au « max » et l'on a donc des égalités du type :

$$\forall (x, y, z) \in [0,1]^3, t[x, \min(y, z)] = \min[t(x, y), t(x, z)]$$

- Les seules t-normes et t-conormes qui sont mutuellement distributives sont le « min » et le « max ».
- Les seules t-normes et t-conormes idempotentes sont le « min » et le « max ».
- A partir de toute t-norme t et de toute fonction continue strictement croissante h de $[0,1]$ dans $[0,1]$ telle que $h(0) = 0$ et $h(1) = 1$, on peut construire une autre t-norme t' par la formule :

$$\forall (x, y) \in [0,1]^2, t'(x, y) = h^{-1}[t(h(x), h(y))]$$

ce qui donne un moyen de générer des familles de t-normes à partir d'un exemple concret.

Dans le cadre de notre étude pour la variable possibilité, après avoir constaté que les résultats les plus intéressants pour les opérateurs de moyenne étaient ceux proches du min, on utilisera exclusivement des t-normes dans le reste de cette section pour aller au delà de cette frontière et essayer d'améliorer les performances d'intégration.

6.4.1 t-normes de Hamacher

La première famille de t-normes considérée dans le cadre de notre étude est celle des t-normes de Hamacher.

Il s'agit des seules t-normes archimédiennes, strictement monotones et d'expression rationnelle, et elles ont été définies par Hamacher dans [19].

Une t-norme archimédienne strictement monotone vérifie, parmi d'autres, les propriétés suivantes :

- $\forall x \in [0,1], t'(x, x) < x$
- $\forall (x, y, y') \in [0,1]^3, y < y' \Rightarrow t(x, y) < t(x, y')$

Pour plus d'informations sur les t-normes et co-normes archimédiennes, on renvoie le lecteur au chapitre 8 de [1].

L'expression concrète de l'opérateur est de la forme:

$$\forall (x, y) \in [0,1]^2, t_H(x, y) = \frac{xy}{\gamma + (1-\gamma)(x + y - xy)}$$

où γ est un paramètre positif.

Parmi les cas particuliers pour cette t-norme on doit signaler que :

- Pour $\gamma = 1$ on retrouve le produit (opérateur dit « probabiliste »).
- Pour $\gamma \rightarrow \infty$ on retrouve la t-norme drastique décrite ci-dessus.

Les résultats obtenus par les simulations d'évaluation de performances pour les t-normes de Hamacher sont de la forme :

[Dans toutes les figures de cette section, l'axe des x montre la variation du paramètre caractéristique associé à chacune des différentes familles de t-normes considérées, selon une relation de type $param\grave{e}tre = 10^{\pm x}$ ($-10 \leq x \leq 10$). On a choisit cette échelle pour essayer d'évaluer les différents comportements pour des valeurs des paramètres entre 0 et ∞ .]

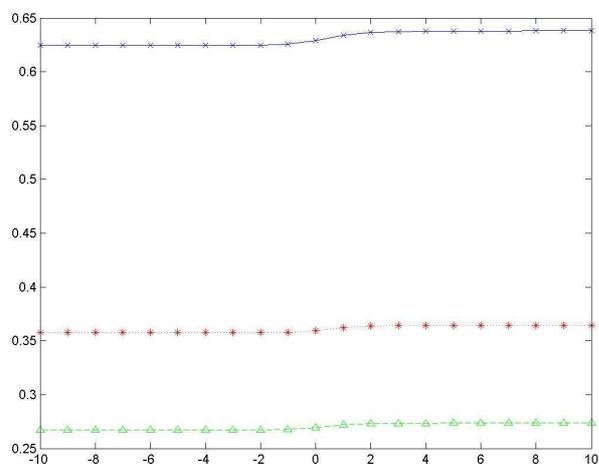


Figure 50. Fonction A - "Hamacher"

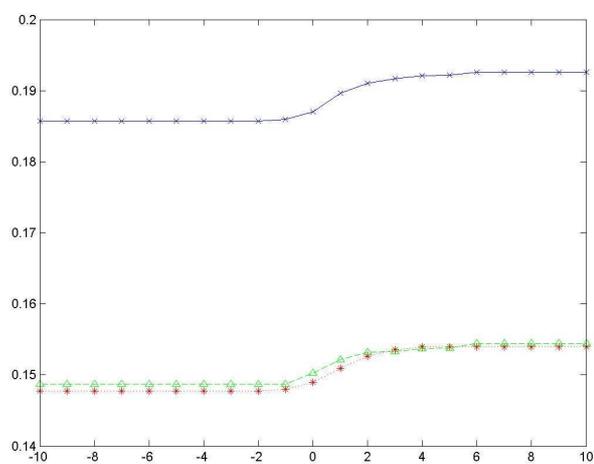


Figure 51. Fonction C - "Hamacher"

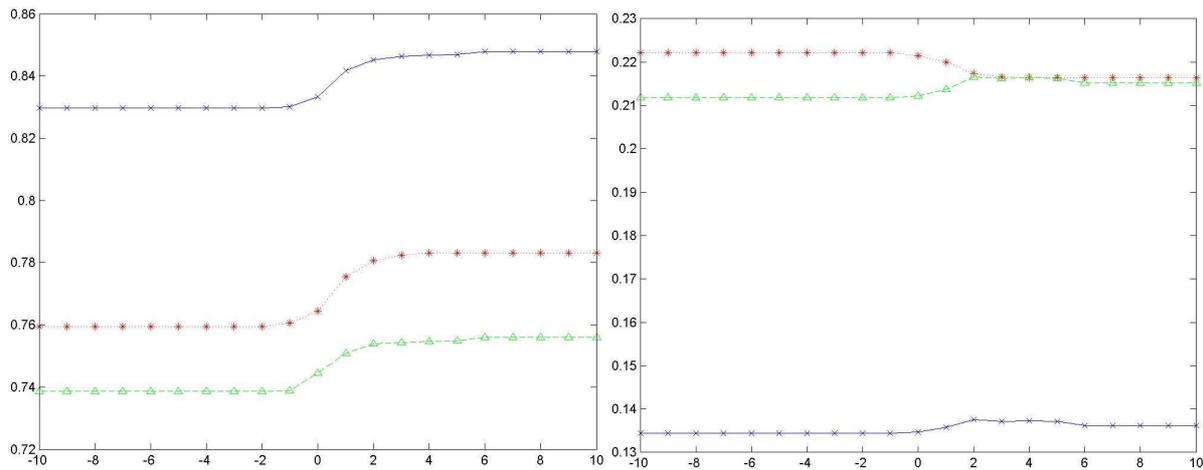


Figure 52. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Hamacher"

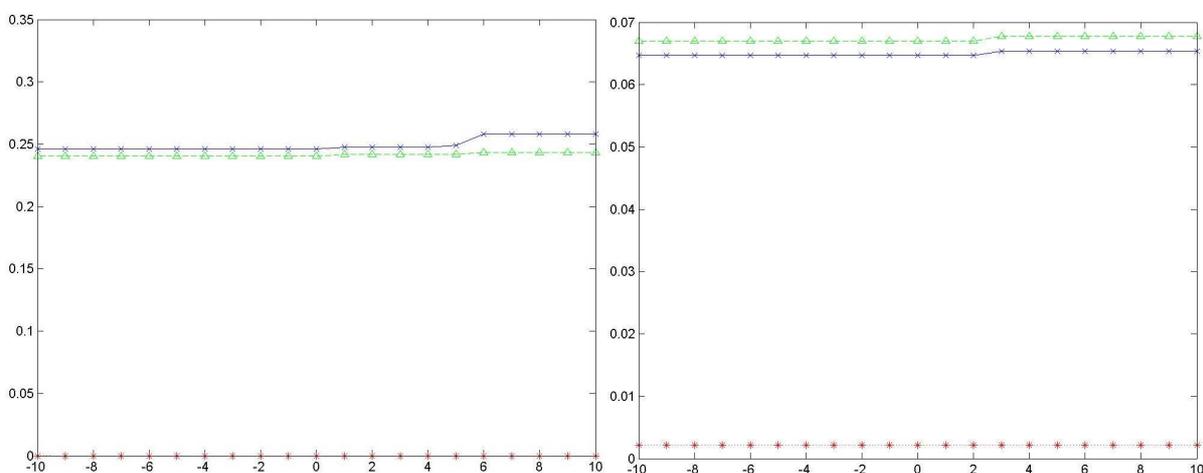


Figure 53. Taux de réussite et d'erreur - "Hamacher"

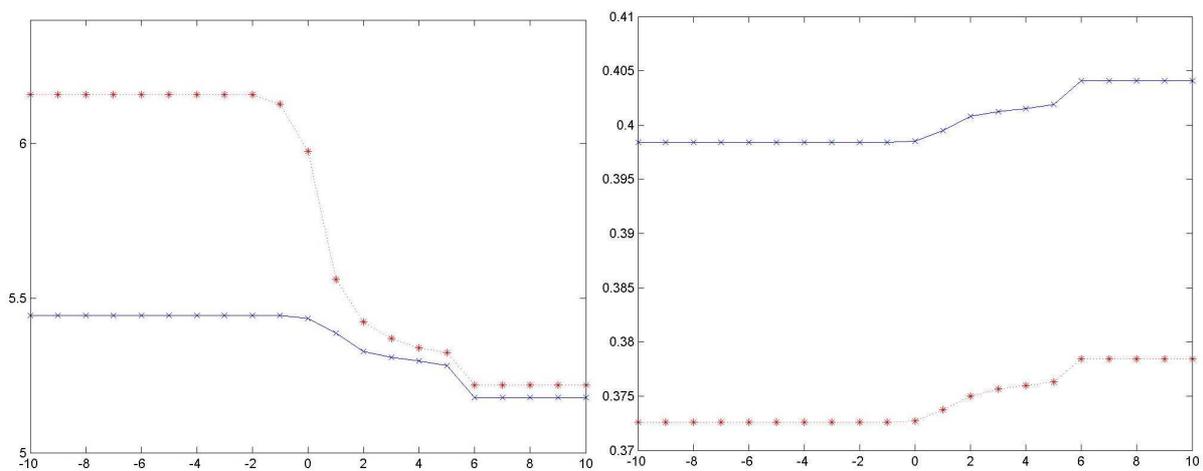


Figure 54. Erreurs quadratiques et précision-kappa – «Hamacher»

Malgré une montée importante de l'erreur, on trouve une certaine amélioration du reste des performances générales par rapport à l'opérateur min. Ce fait est plus évident pour des valeurs de γ qui tendent vers l'infini.

6.4.2 t-norme de Frank

Les t-normes de M. Frank [20] constituent une autre famille paramétrique de la même classe que les précédents.

$$\forall (x, y) \in [0,1]^2, t_F(x, y) = \log_s \left[1 + \frac{(s^x - 1)(s^y - 1)}{s - 1} \right]$$

où s est un paramètre strictement positif.

A nouveau, on trouve quelques cas particuliers :

- Si $s \rightarrow 0$, la t-norme tend vers le minimum.
- Si $s \rightarrow \infty$, la t-norme tend vers la t-norme dite « de Luckasiewicz ».
- Si $s = 1$ on retrouve le produit nouvellement.

Les t-normes et t-conormes de Frank présentent la propriété unique de vérifier l'égalité suivante:

$$\forall (x, y) \in [0,1]^2, t(x, y) + T(x, y) = x + y$$

Les résultats obtenus pour les t-normes de Frank sont de la forme :

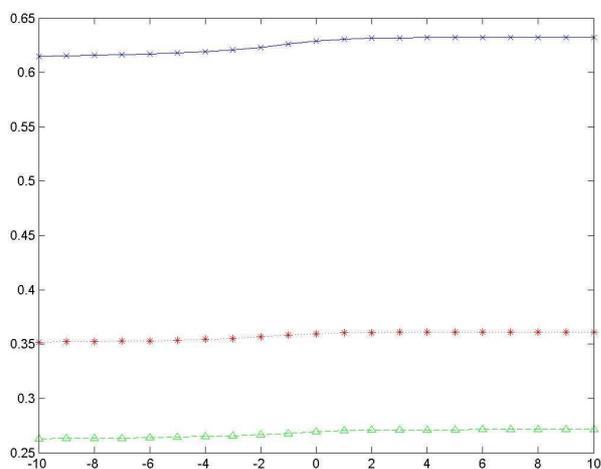


Figure 55. Fonction A - "Frank"

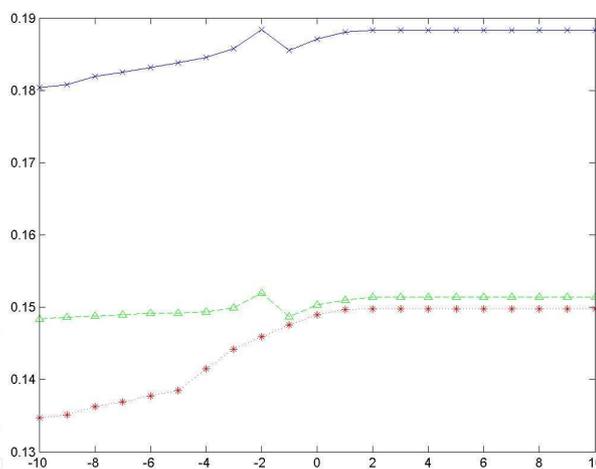


Figure 56. Fonction C - "Frank"

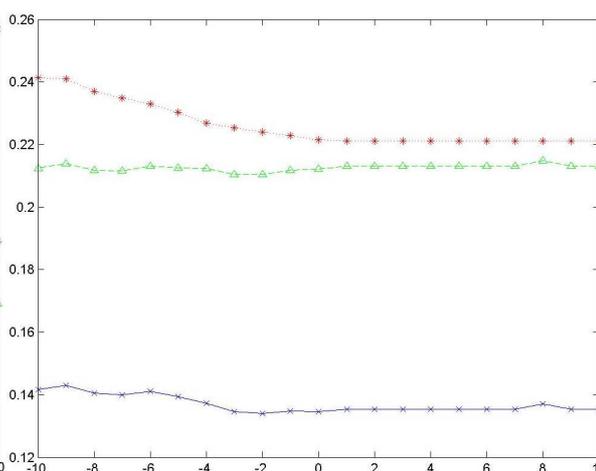
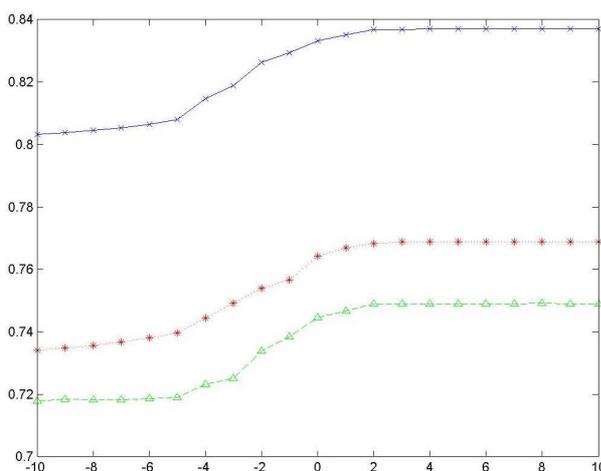


Figure 57. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Frank"

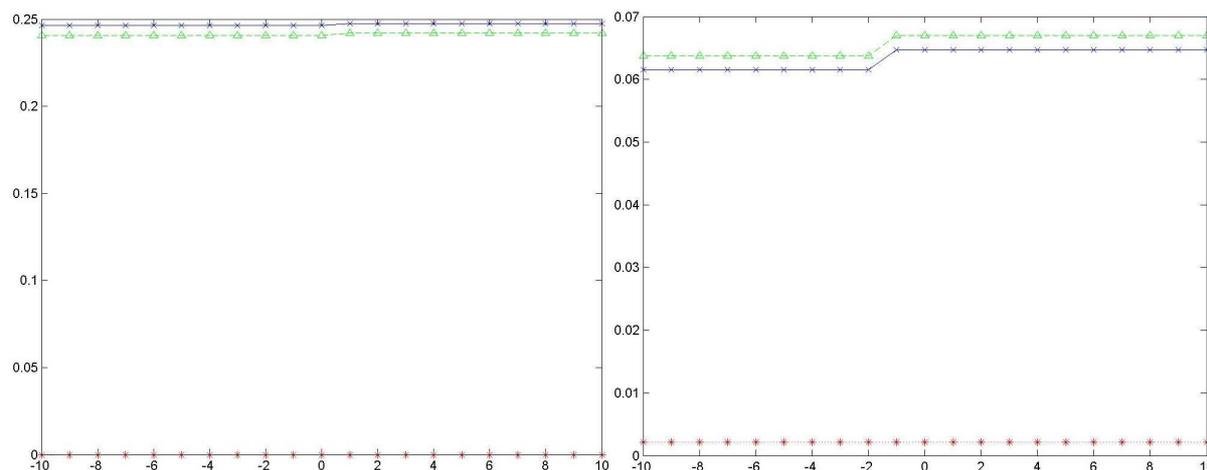


Figure 58. Taux de réussite et d'erreur - "Frank"

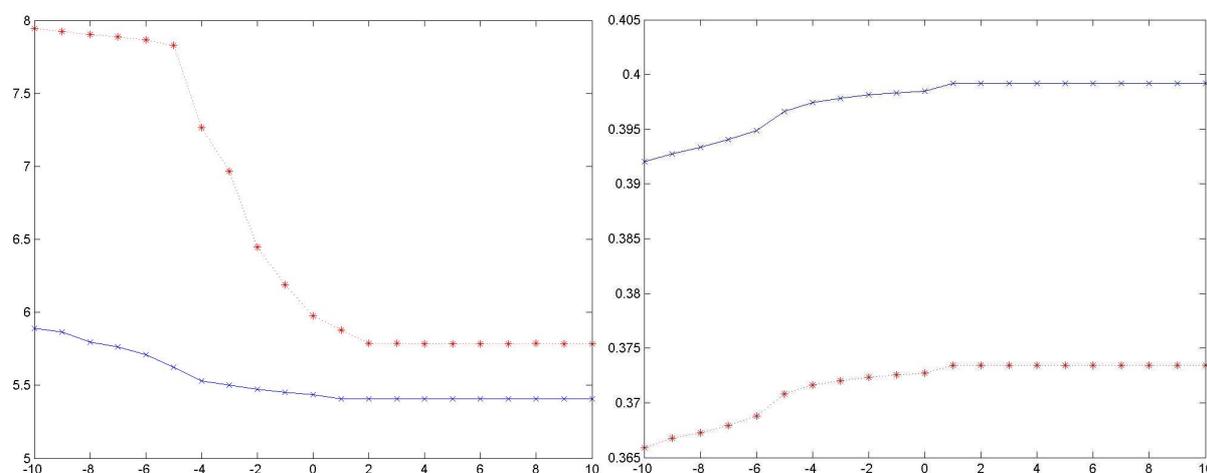


Figure 59. Erreurs quadratiques et précision-kappa – «Frank»

A nouveau, on observe des performances générales légèrement meilleures pour des valeurs élevées du paramètre s, bien qu'ici l'augmentation relative de l'erreur soit plus accentuée et plus évidente à partir d'un certain seuil.

6.4.3 t-normes de Dombi

Les t-normes de Dombi constituent une autre famille de t-normes archimédiennes paramétriques d'intérêt, et elles ont été introduites par Dombi en 1982 [21].

Elles sont définies par une expression de la forme :

$$t_D(x, y) = \begin{cases} 0 & \text{si } x = 0 \text{ ou } y = 0 \\ \frac{1}{\left(\left(\frac{1}{x} - 1 \right)^\alpha + \left(\frac{1}{y} - 1 \right)^\alpha \right)^{\frac{1}{\alpha}} + 1} & \text{sinon} \end{cases}$$

ou α est strictement positif

La classe Dombi de t-normes nous permet de parcourir tout le espace de t-normes entre le min et la t-norme drastique, de telle sorte que :

- Pour $\alpha \rightarrow 0$, la t-norme tend vers la t-norme drastique.
- Pour $\alpha \rightarrow \infty$, la t-norme tend vers l'opérateur « min »

- Pour $\alpha = 1$ on obtient un opérateur de la forme :

$$t_{D1} = \begin{cases} 0 & \text{si } x = 0 \text{ ou } y = 0 \\ \frac{xy}{x + y - xy} & \text{sinon} \end{cases}$$

Les résultats pour l'intégration de la variable possibilité dans le cadre de notre système NCTR sont de la forme :

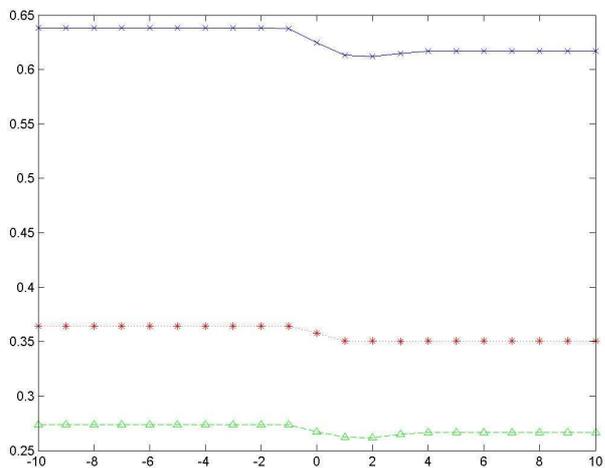


Figure 60. Fonction A - "Dombi"

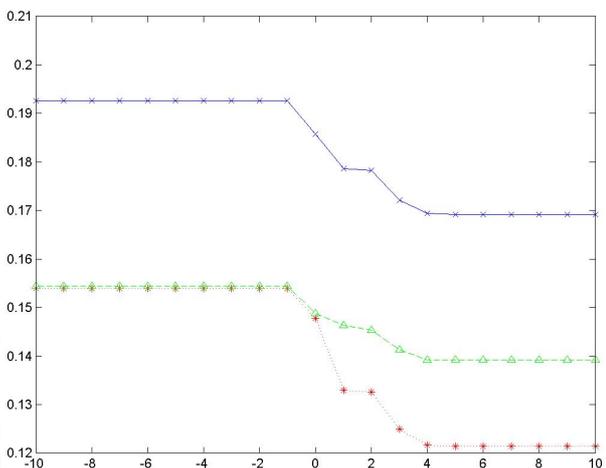


Figure 61. Fonction C - "Dombi"

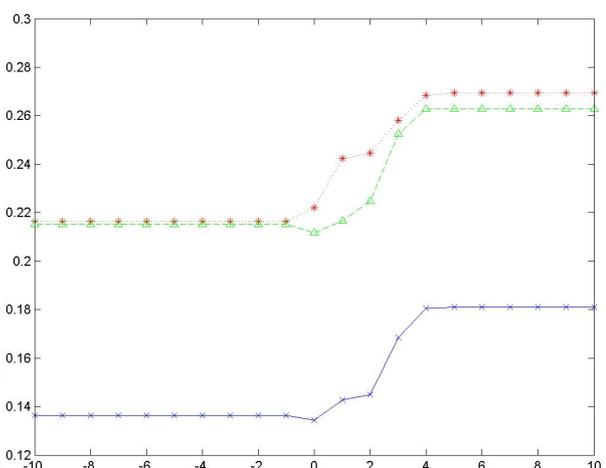
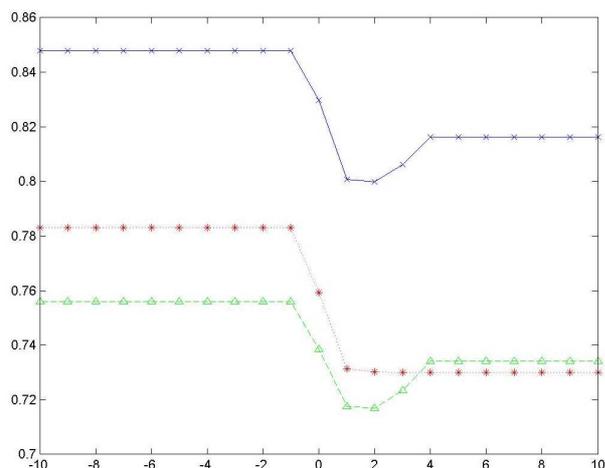


Figure 62. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Dombi"

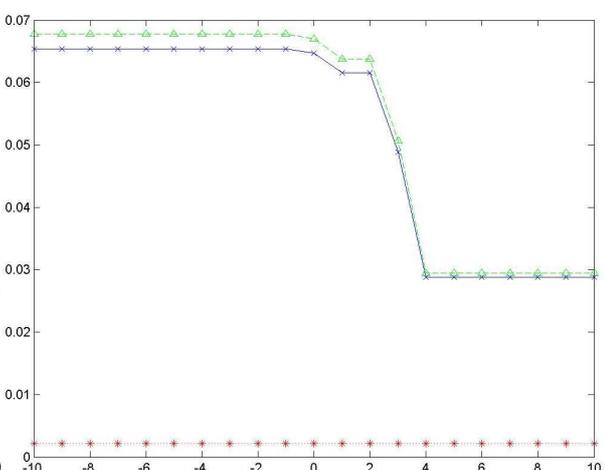
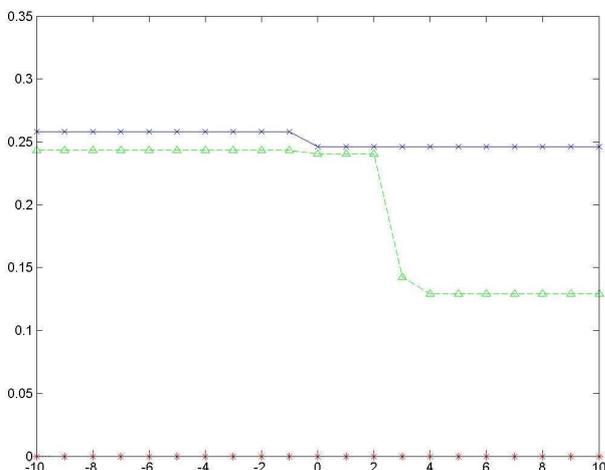


Figure 63. Taux de réussite et d'erreur - "Dombi"

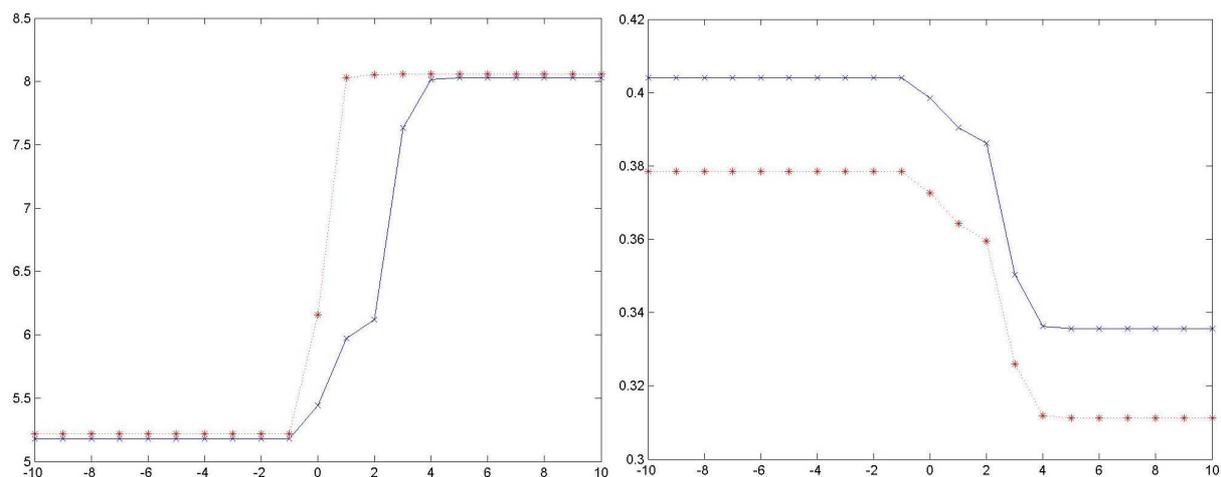


Figure 64. Erreurs quadratiques et précision-kappa – «Dombi»

On constate à nouveau la même tendance d'auparavant en ce qui concerne le compromis entre les performances générales et taux d'erreur associé à une approche plus ou moins « risquée ». C'est ainsi que l'on obtient les meilleures performances générales (avec un taux de réussite supérieur à 25%), mais aussi le taux d'erreur le pire (au-dessus du 6.5%) pour des valeurs de p qui tendent vers 0, ce qui correspond à la t -norme drastique.

6.4.4 t -normes de Yager

Finalement, le dernier groupe de t -normes que l'on considérera, la famille des t -normes de Yager [22], possèdent la particularité d'être des opérateurs nilpotents :

$$\forall (x, y) \in [0, 1]^2, t(x, y) = f * [f(x) + f(y)]$$

où f est une bijection décroissante de $[0, 1]$ dans $[0, 1]$, telle que $f(0)=1$, $f(1)=0$ et $f^*(x)=f^{-1}(x)$ si $x \in [0, 1]$, avec $f^*(x)=0$ si $x \geq 1$.

Puisque ces considérations commencent à sortir du cadre de notre étude, on se limitera à dire que ces opérateurs satisfont les lois du tiers exclu et de non-contradiction.

La forme concrète des t -normes de Yager est donnée par l'expression :

$$t_Y(x, y) = \max\left(1 - \left((1-x)^p + (1-y)^p\right)^{\frac{1}{p}}, 0\right)$$

Et parmi les cas particuliers d'intérêt, on peut mentionner :

- Pour $p \rightarrow \infty$, la t -norme tend vers l'opérateur « min »
- Pour $p=1$, on obtient à nouveau la t -norme de Lukasiewicz

Les résultats obtenus pour les t -normes de Yager sont de la forme :

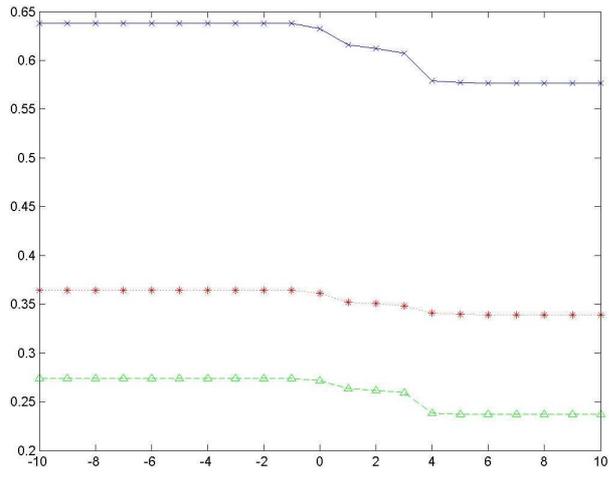


Figure 65. Fonction A - "Yager"

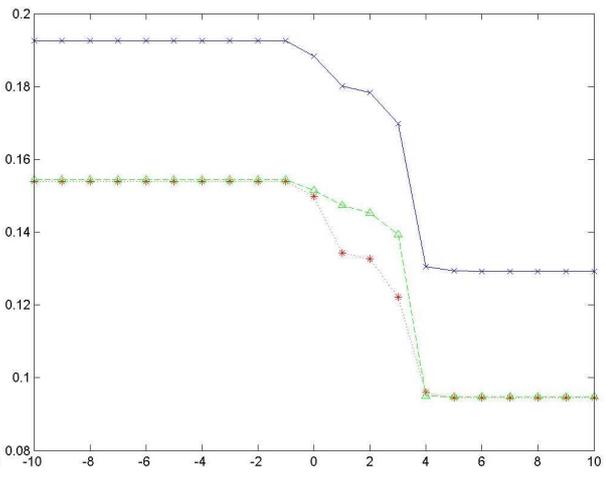


Figure 66. Fonction C - "Yager"

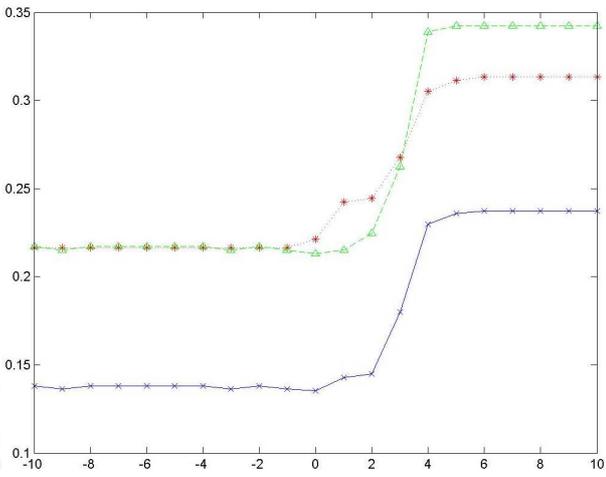
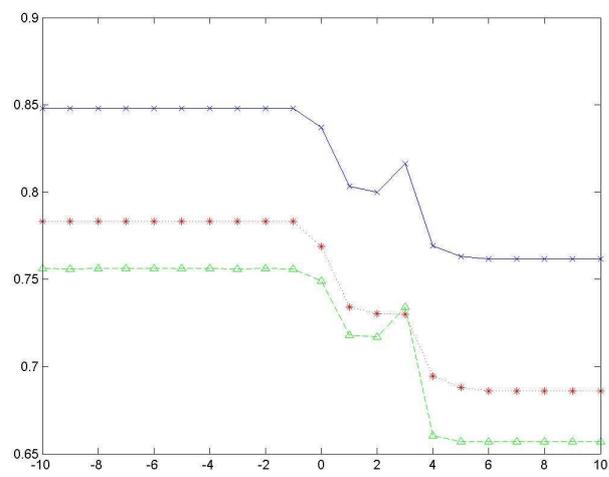


Figure 67. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Yager"

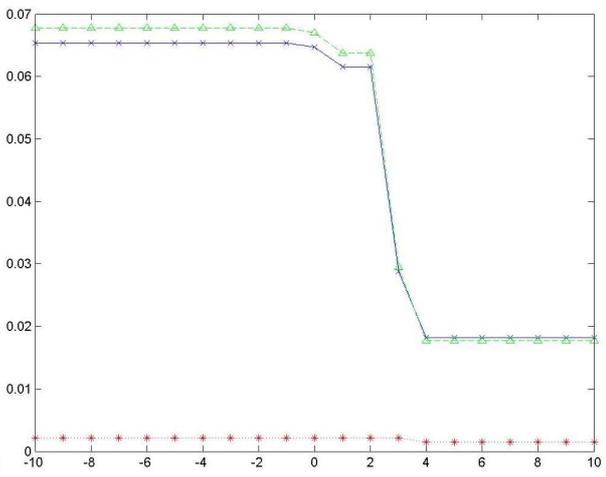
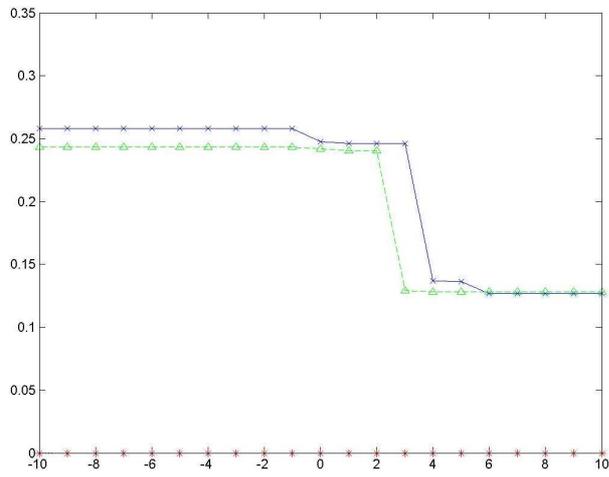


Figure 68. Taux de réussite et d'erreur - "Yager"

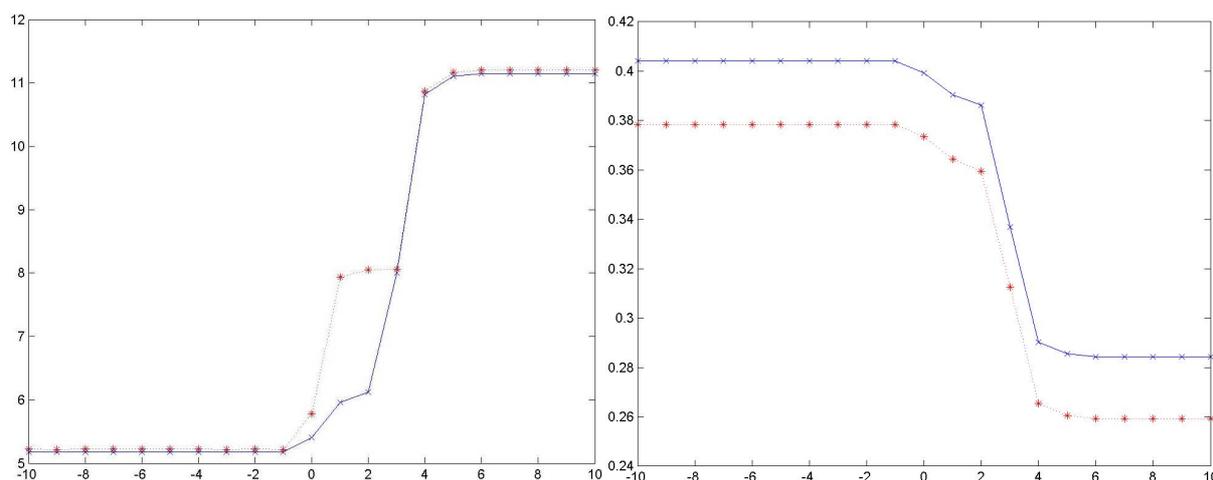


Figure 69. Erreurs quadratiques et précision-kappa – «Yager»

On observe un comportement assez similaire à celui des t-normes de Dombi. En fait, on retrouve les résultats les plus risqués, c'est-à-dire, les plus discriminants mais aussi les plus susceptibles de commettre des erreurs, à nouveau, pour des valeurs du paramètre α qui tendent vers 0 et qui correspondent donc à la t-norme drastique. En général, il semble que le comportement des familles de Dombi et de Yager offre des performances similaires pour les données considérées.

6.4.5 Conclusions sur les t-normes triangulaires

L'utilisation des t-normes triangulaires nous permet d'obtenir des résultats très intéressants qui complètent ceux fournis auparavant par les opérateurs de moyenne.

En général, les t-normes constituent une continuation que l'on pourrait considérer « naturelle » par rapport aux opérateurs étudiés dans la section précédente, du fait qu'elles apportent une continuité aux résultats obtenus: au delà de l'opérateur "min", les notes fournies par les fonctions d'évaluation, le taux de réussite globale et les mesures sur les matrices de confusion s'améliorent pour les t-normes quand elles deviennent petites, jusqu'à ce que l'on arrive à la t-norme drastique, pour laquelle on obtient les résultats les meilleurs. Le taux d'erreur, par contre, continue d'augmenter poursuivant également la tendance observée pour les opérateurs de moyenne et l'on arrive donc à une situation de compromis analogue à celle mentionnée pour les opérateurs de moyenne, mais encore plus accentuée.

Dans cet esprit, les résultats les plus extrêmes sont ceux qui correspondent à la t-norme drastique (on les a obtenus pour Hamacher quand $\gamma \rightarrow \infty$ et pour Yager quand $p \rightarrow 0$) et que sont résumées dans les tableaux suivants, que l'on peut comparer avec leurs équivalents pour les données sans intégration et pour l'approche « k sur M »:

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.6380	0.3643	0.2737

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.8479	0.7831	0.7562	0.1380	0.2164	0.2170

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.1926	0.1540	0.1544

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.2582	0	0.2433
Erreur	0.0654	0.0021	0.0678

Voici la matrice de confusion normalisée pour l'acceptation :

0.7926	0.1789	0.9912	0.5905	0.6808	0
0.9047	0.9687	0.1772	0.0038	0.0027	0
0	0.0086	0.9221	0.2154	1.0000	0.0020
0	0	0.4426	0.9242	0.9924	0.0647
0	0	0	0.0936	1.0000	0.0820
0	0	0	0.4286	0.1224	1.0000

Erreur quadratique : 5.1779

Voici la matrice non normalisée avec les mesures de précision :

551.0000	172.0000	707.0000	498.0000	581.0000	0	0.2196
733.0000	778.0000	143.0000	3.0000	2.0000	0	0.4690
0	6.0000	544.0000	134.0000	612.0000	1.0000	0.4194
0	0	173.0000	299.0000	326.0000	13.0000	0.3687
0	0	0	56.0000	652.0000	50.0000	0.8602
0	0	0	21.0000	6.0000	49.0000	0.6447
0.4291	0.8138	0.3472	0.2957	0.2992	0.4336	0.4041

kappa = 0.3785

Et voici finalement la matrice de confusion pour le rejet:

0.1424	0.7889	0.0074	0.3874	0.3148	1.0000
0.0953	0.0313	0.7984	0.9962	0.9962	1.0000
1.0000	0.9893	0.0779	0.7716	0	0.9980
1.0000	0.9975	0.5574	0.0076	0.0076	0.8756
1.0000	1.0000	1.0000	0.9064	0	0.9180
1.0000	1.0000	0.9796	0.5714	0.8776	0

Erreur quadratique : 5.2243

La conclusion semble évidente: les t-normes nous permettent d'augmenter le niveau de risques pris pour l'intégration, nous permettent d'adopter des approches plus « risquées » (avec des taux d'erreur et de réussite plus élevées au même temps) que celles correspondants aux opérateurs de moyenne. Il s'agit donc d'un type d'outils qui peut apporter des grands avantages, mais qu'il faudra utiliser avec précaution.

6.5 ESSAIS DE DOUBLE INTEGRATION

Après avoir étudié les performances de plusieurs types d'opérateurs de fusion, on a pensé à la possibilité de faire évoluer la stratégie d'intégration elle-même, en s'inspirant des techniques utilisées dans le domaine des télécoms.

6.5.1 Moyenne exponentielle

En effet, c'est ainsi que l'on a pensé tout d'abord aux techniques de moyennage exponentiel utilisées pour des applications telles que celle du calcul de la congestion et la gestion du trafic dans les réseaux IP comme l'Internet. Il semble que dans ce domaine, on utilise une moyenne pondérée simple pour fusionner l'estimation actuelle de congestion (estimation fournie par le réseau) avec la valeur finale calculée pour l'instant précédent (processus récursif), de telle sorte que cette moyenne devienne exponentielle avec le temps, et qu'elle permette de régler l'importance relative que l'on donne aux nouvelles informations par rapport aux connaissances préalables (pour plus d'info voir [28]).

Tous les algorithmes considérés jusqu'à présent pour notre étude utilisaient en tout instant les M (M=10) derniers échantillons fournis par le radar. Par contre, l'échantillon fourni M+1 instants en avant ne jouait plus aucun rôle dans le calcul de la possibilité cumulée pour un instant donné. Intuitivement, on a pensé que, d'un côté, il pourrait être intéressant de prendre en compte les données passés au-delà des derniers M pour gagner de la cohérence et de la consistance dans le processus d'intégration temporelle. D'un autre côté, l'expérience des radaristes montre que, souvent, c'est mieux d'oublier le passé le plus vite possible car les données les plus récentes sont plus précises ou plus cohérentes avec la réalité. Ceci dit, pourquoi ne pas essayer d'évaluer et comparer les performances des deux approches par nous-mêmes ?

C'est ainsi que l'on a décidé de faire quelques simulations dans lesquelles on fait l'intégration temporelle avec une t-norme triangulaire, mais aussi une sorte de « double intégration » puisque l'on fusionne l'apport de chaque nouvelle ensemble de valeurs instantanées avec la valeur calculée pour la possibilité cumulée correspondante à l'instant antérieur, selon une expression de la forme :

$$P_{cumulee}(t+1) = w \cdot f_{P,t-norme}(t+1) + (1-w) \cdot P_{cumulee}(t)$$

ou w est un paramètre entre 0 et 1.

Evidemment, pour $w = 1$, la contribution « du passé » est nulle et l'on retrouve la stratégie que l'on avait utilisée dans tous les cas précédents.

Voici quelques exemples pour un paramètre quelconque et une intégration de base de type « k sur M » sur laquelle on a réalisé une moyenne avec différentes valeurs de w :

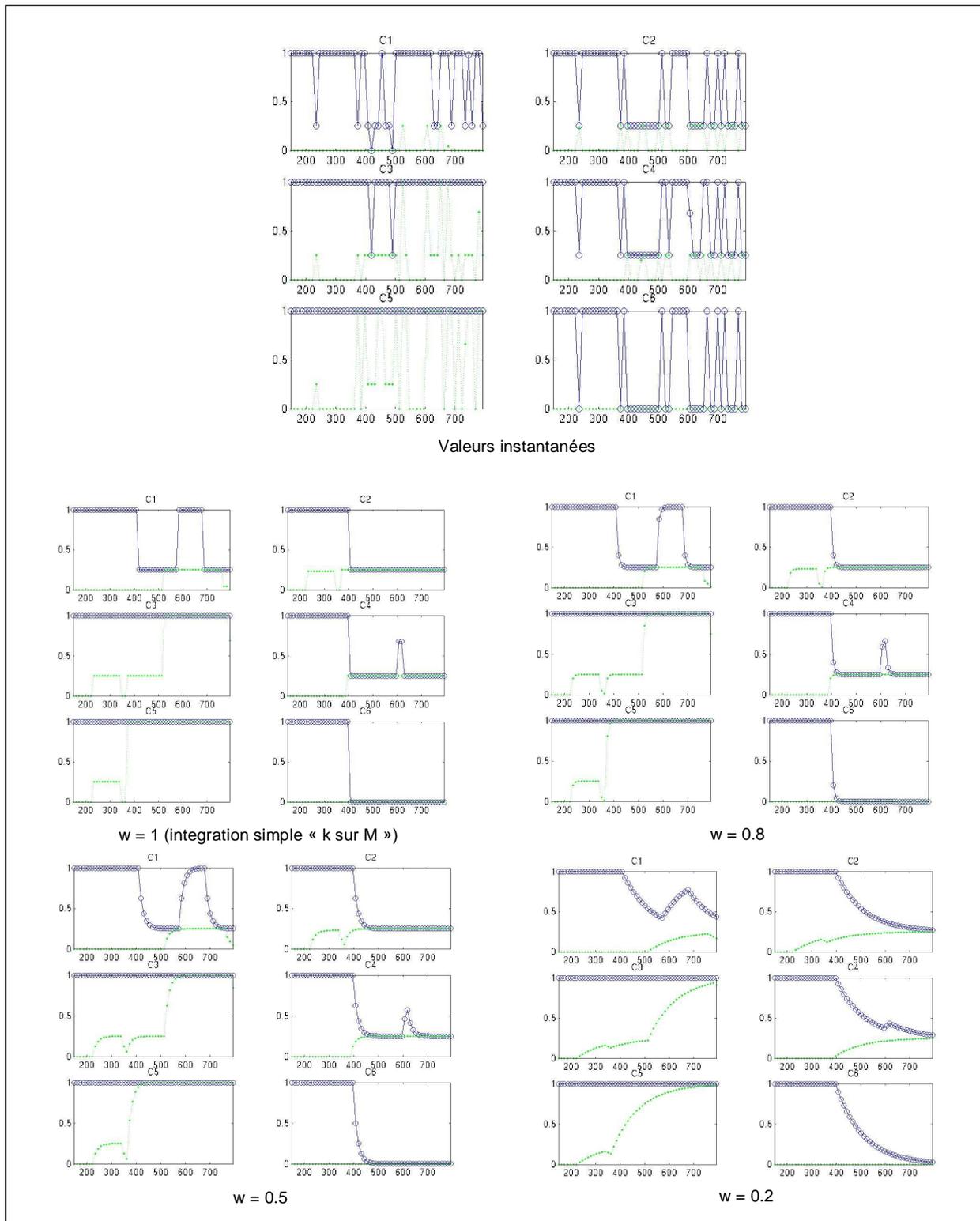


Figure 70. Exemples de moyenne exponentiel pour la double intégration temporelle (cas générique, avec la possibilité en bleu et la nécessité en vert)

Les résultats obtenus pour cette nouvelle approche, appliquée à la t-norme drastique (Yager, avec $\alpha \rightarrow 0$) sont présentés dans les figures ci-jointes, dans lesquelles l'axe des x correspond aux différentes valeurs de w dans l'expression de la moyenne exponentielle ci-dessus, telles que $w=0.1x$:

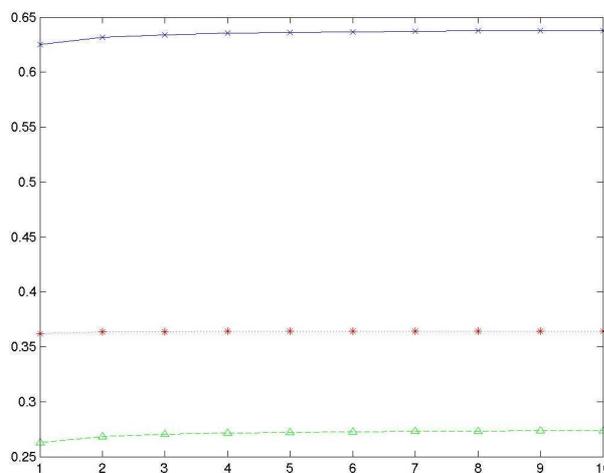


Figure 71. Fonction A - "Yager+m.exp."

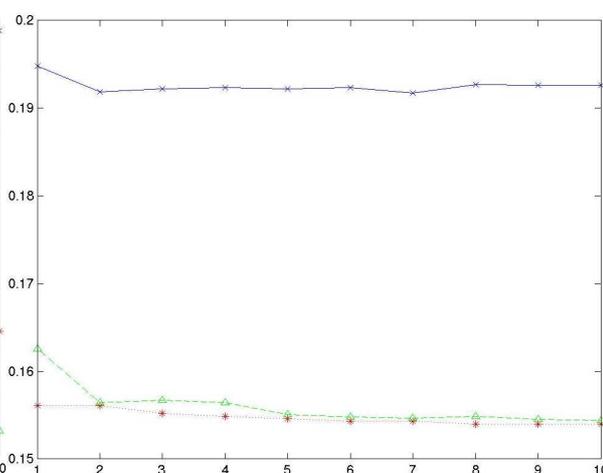


Figure 72. Fonction A - "Yager+m.exp."

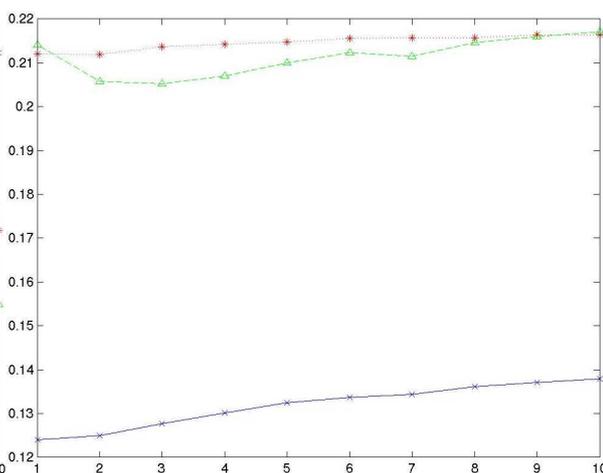
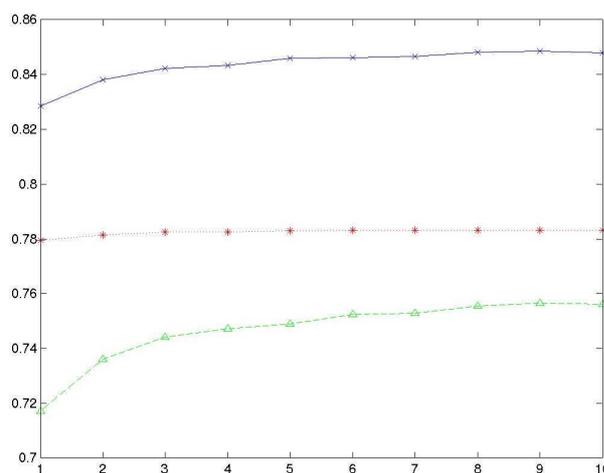


Figure 73. Fonction B – réussite (gauche) et erreur (droite) élémentaires - "Yager+m.exp."

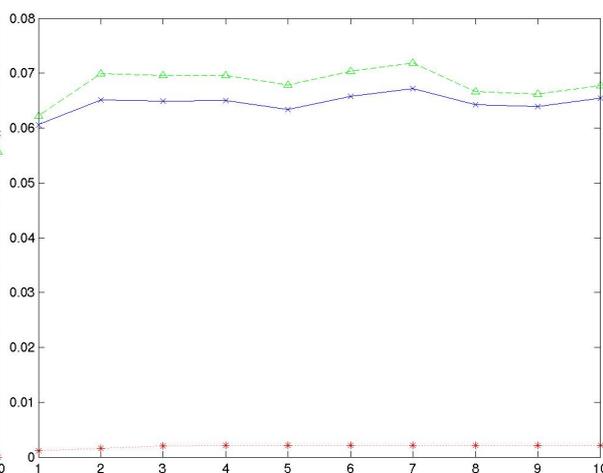
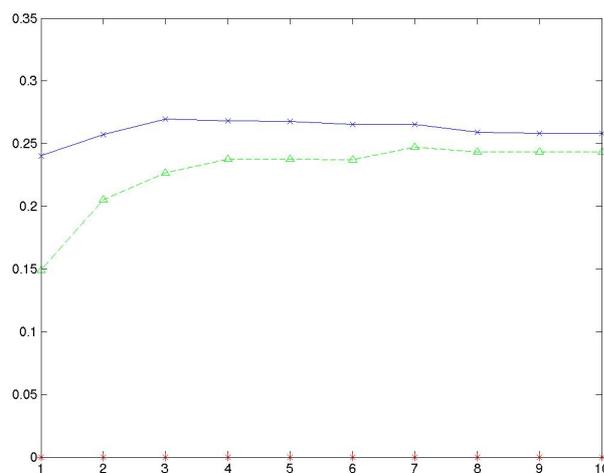


Figure 74. Taux de réussite et d' erreur - "Yager+m.exp."

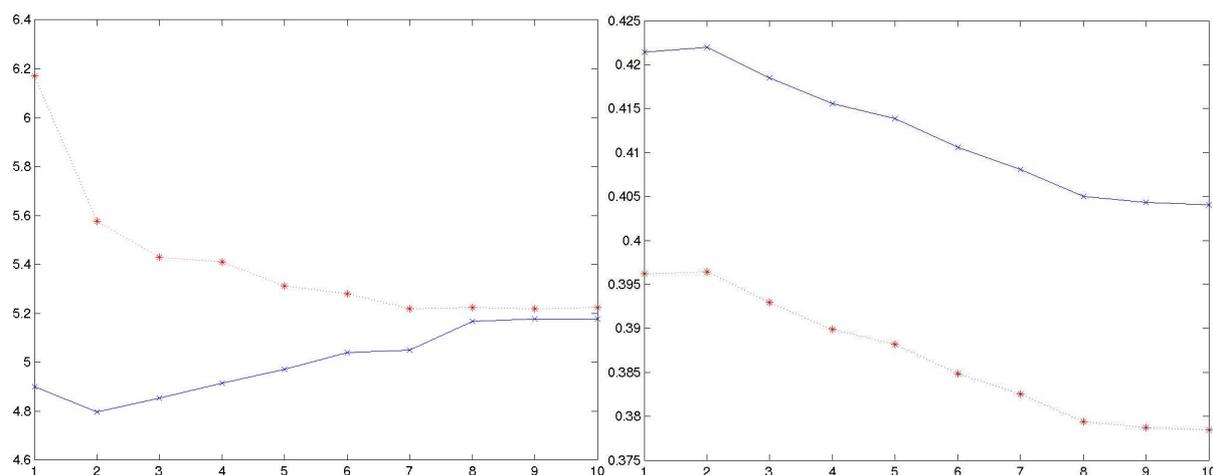


Figure 75. Erreurs quadratiques et précision-kappa – «Yager+m.exp.»

Ces résultats montrent que les notes fournies par les fonctions d'évaluation sont, en général, meilleures pour des valeurs proches de $w=1$, c'est-à-dire, pour les cas où il n'y a pas de double intégration. On observe, par contre, que les résultats correspondant aux mesures sur les matrices de confusion et aussi au taux de réussite général s'améliorent parfois grâce à la technique de la moyenne. Le taux d'erreur, par contre, augmente en général et dépasse le 7% pour quelques valeurs de w .

On observe donc des performances qui ont expérimenté des modifications plutôt complexes dans leur tendances (elles sont devenues parfois monotones avec le paramètre considéré et parfois pas), mais en tout cas ce sont des modifications d'une magnitude pas trop importante. La moyenne exponentielle n'apporte pas d'amélioration évidente sur les performances, et pour certaines valeurs du paramètre w celles-ci deviennent pires que leurs équivalents pour l'intégration simple. Il semble intéressant, quand même, de poursuivre notre enquête sur la double intégration à la recherche de comportements plus adaptés à nos besoins en terme de taux d'erreur et de taux global de réussite.

6.5.2 « Double norme »

C'est ainsi que, puisque l'idée d'utiliser une « double intégration » temporelle avec moyenne exponentielle ne semblait pas apporter des avantages significatifs, on a considéré la possibilité d'effectuer le « moyennage » (ou plutôt la fusion/intégration, puisque l'on se retrouve hors le cadre des opérateurs de moyenne) avec une norme triangulaire.

En effet, on a déjà vu que les opérateurs de moyenne ne sont pas ceux pour lesquels on obtient normalement les comportements les plus discriminants et, puisque les t-normes nous permettent d'améliorer toutes les performances générales hors taux d'erreur pour l'intégration simple, on a décidé d'essayer une intégration double de type « double t-norme drastique », par exemple.

L'algorithme utilisé est le même que dans le cas précédent, mais on remplace la moyenne pondérée par la t-norme drastique, ce qui revient à considérer que la possibilité cumulée vaut 1 si toutes les valeurs de possibilité passées ont été égales à 1, et 0 dans le cas contraire.

Les résultats obtenus sont résumés dans les tableaux suivants :

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.6450	0.3678	0.2772

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.8069	0.7928	0.7345	0.1906	0.2068	0.2618

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.0804	0.1597	0.0610

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.3014	0	0.2567
Erreur	0.2088	0.0021	0.2129

Matrices de confusion:

Erreur quadratique acceptation	3.9949
Erreur quadratique rejet	3.9934
Précision générale	0.4246
Coefficient kappa	0.3980

On observe une augmentation simultanée très importante des taux de réussite et d'erreur. En réalité, on trouve ici une approche notamment plus drastique, plus radicale, plus « risquée » que dans les cas précédents. Le comportement devient binaire : la possibilité cumulée est égale à 1 jusqu'à ce que l'on trouve une seule valeur instantanée différente de 1 et à partir de cet instant là elle restera toujours égale à 0.

Apparemment, cette stratégie est tellement « risquée » que les avantages obtenus en terme d'une plus grande puissance de discrimination sont annulés par les nouvelles erreurs provoquées par les valeurs non idéales pour la possibilité instantanée fournies sporadiquement par les capteurs pour la bonne paire cible-classe.

Idéalement, on voudrait profiter de l'amélioration des performances auprès du taux de réussite, mais en contrôlant le taux d'erreur. Cela sera notre objectif dans la suite.

6.5.3 Intégration hybride : « k sur M » + « double norme drastique »

Après tout ce que l'on vient d'expliquer, on pourrait penser que l'approche de la double intégration avec double t-norme pourrait être très utile grâce à son pouvoir discriminant dans le cas où toutes les valeurs de possibilité pour la bonne paire classe-cible sont toujours égales à 1.

Si l'on observe les données avec lesquelles on travaille, on trouve que cela est déjà le cas pour tous les paramètres sauf celui du paramètre B1.

En effet, tous les paramètres sauf le B1 fournissent toujours des valeurs de possibilité égales à 1 pour les bonnes paires classe-cible et ceci pour tous les ensembles de données réelles avec lesquelles on travaille. On pourrait penser à des précisions de mesure très fiables (maîtrise de la fonction de densité de possibilité) ou à des cibles pour lesquelles on maîtrise la définition des fonctions d'appartenance jusqu'au point d'avoir un niveau très robuste de fiabilité pour les mesures correspondant à la bonne classe pour chaque cible. Le paramètre B1 présente, par contre, quelques fluctuations dans la valeur calculée pour la possibilité instantanée pour la bonne paire classe-cible qui peuvent être considérées inhérentes au processus de détection.

Il semble intéressant, donc, d'étudier les résultats pour le même processus d'intégration « double t-norme drastique » si l'on ignore le paramètre B1 :

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.6495	0.4597	0.1898

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.8882	0.7928	0.7595	0.1110	0.2068	0.2382

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.2086	0.1597	0.1270

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.3291	0	0.1815
Erreur	0.0021	0.0021	0

Matrices de confusion:

Erreur quadratique acceptation	4.4949
Erreur quadratique rejet	4.5019
Précision générale	0.4446
Coefficient kappa	0.4157

Voilà un taux de réussite de 32,91% avec un taux d'erreur associée de 0,21% - ce qui constitue les meilleurs résultats que l'on a obtenu jusqu'à présent.

Les résultats confirment par ailleurs nos suppositions initiales en ce qui concerne le paramètre B1 : on obtient des meilleures performances sans considérer le paramètre B1, avec la double intégration sur t-norme drastique, qu'en l'utilisant.

Quelques simulations faites avec et sans le paramètre B1 (que l'on ne reproduira pas ici pour des raisons de place) montrent que, en général, le paramètre B1 améliore les performances d'identification pour tout les opérateurs d'intégration considérés dans les autres sections. C'est seulement quand on utilise une approche tellement drastique, tellement « radicale », comme celle de la double t-norme drastique que la prise en compte de B1 apporte une contribution négative.

Comme on l'a déjà dit, B1 présente des valeurs de possibilité égales à 1 la plupart du temps, mais il présente aussi quelques fluctuations qui empêchent la bonne identification selon l'approche « double drastique ». Comment pourrait-on donc profiter de l'information contenue dans B1 pour améliorer le processus d'identification ?

Une solution possible est de faire ce que l'on pourrait considérer comme « un prétraitement » du paramètre B1 avec une intégration additionnelle utilisant la méthode « k sur M », ce qui nous permet de réaliser une espèce de « filtrage », de transformer en 1 les valeurs plus ou moins isolées (selon le choix de k) de possibilité au-dessous de 1, afin de corriger les fluctuations non désirées qui perturbent nos données floues pour la bonne paire classe-cible. Si on considère, par exemple, une valeur de k=4, on pourra tolérer 3 valeurs différentes de 1 sur chaque ensemble de M(M=10) échantillons correspondants à la possibilité instantanée pour le paramètre B1, valeurs qui seront mises à 1 au cours de ce prétraitement. Sans doute, on augmentera aussi parfois les valeurs de possibilité pour des paires classe-cible erronées, mais le résultat global semble apporter des résultats satisfaisants.

Les figures ci-dessous montrent un exemple de ce « filtrage » avec la méthode « k sur M » sur une mesure du paramètre B1 et ses répercussions sur le processus d'intégration.

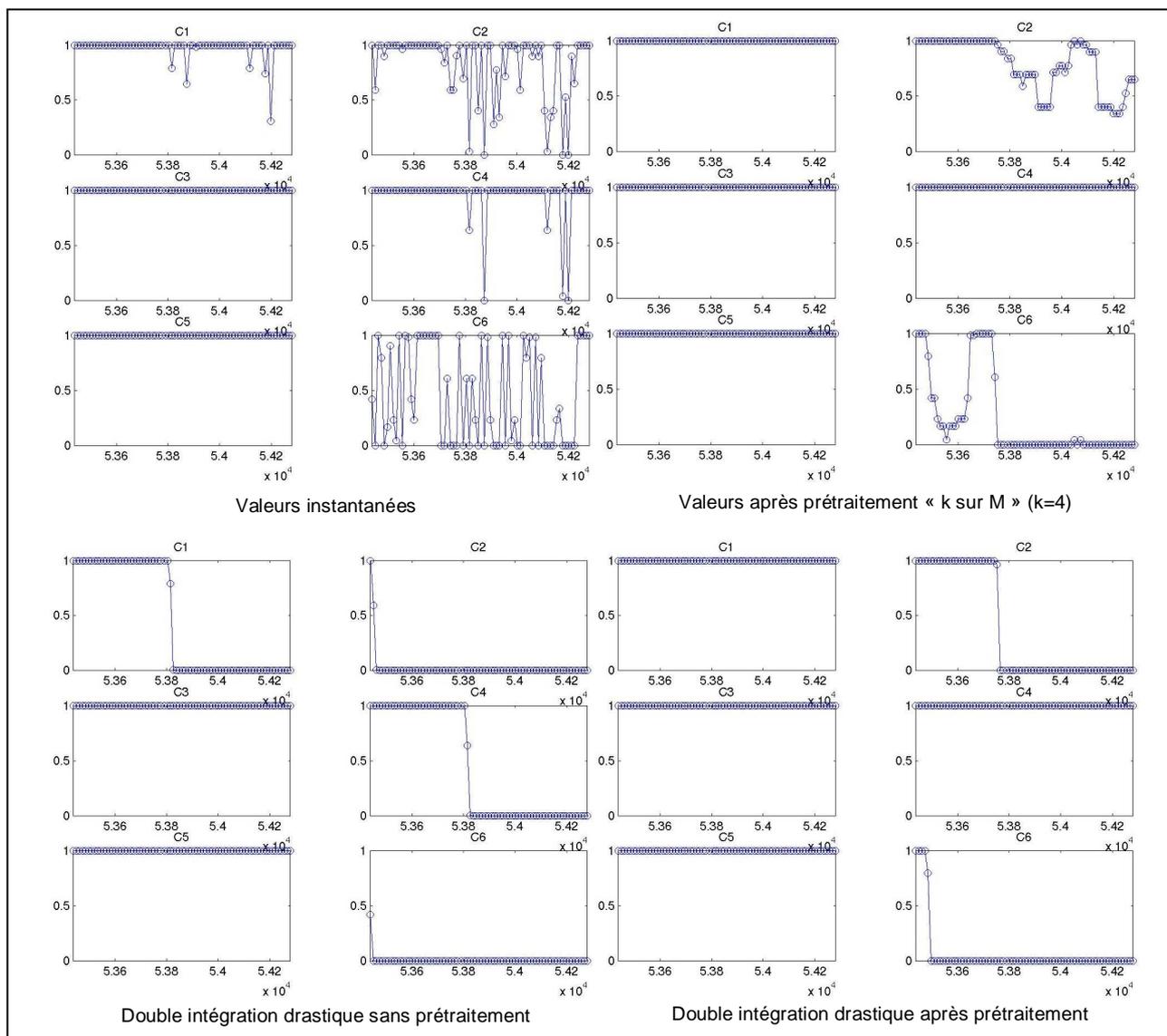


Figure 76. Exemples de double intégration avec et sans prétraitement

Les résultats numériques obtenus pour cette stratégie que l'on pourrait considérer comme «hybride», consistant en un prétraitement « k sur M » avec $k = 4$ exclusivement pour le paramètre B1 et une double intégration avec t-norme drastique postérieure pour tous les paramètres sont :

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.6511	0.3678	0.2833

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.8958	0.7928	0.7981	0.1029	0.2068	0.1992

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.2148	0.1597	0.1565

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.3540	0	0.2547
Erreur	0.0021	0.0021	0

Matrices de confusion:

Erreur quadratique acceptation	4.0098
Erreur quadratique rejet	4.0371
Précision générale	0.4620
Coefficient kappa	0.4338

On observe donc une amélioration relativement significative des performances (taux de réussite à 35,4% pour le même taux d'erreur) par rapport à la non utilisation du paramètre B1.

En tout cas, on retrouve une nouvelle situation de compromis en ce qui concerne le choix de la valeur de k pour le paramètre B1 :

K (B1)	Réussite	Erreur	Précision	Kappa
1	0.3014	0.2088	0.4241	0.3975
2	0.3612	0.0467	0.4600	0.4333
3	0.3540	0.0245	0.4627	0.4356
4	0.3540	0.0021	0.4620	0.4338
5	0.3527	0.0021	0.4612	0.4330
6	0.3291	0.0021	0.4567	0.4287
7	0.3291	0.0021	0.4546	0.4265
8	0.3291	0.0021	0.4490	0.4204
9	0.3291	0.0021	0.4446	0.4157
10	0.3291	0.0021	0.4446	0.4157

Si l'on considère une valeur très petite de k pour B1, on aura un meilleur taux de réussite mais aussi un taux d'erreur plus grand (approche plus risquée, plus discriminante), taux d'erreur qui diminuera avec des valeurs plus grandes de k (approche plus conservatrice).

6.5.4 Conclusions sur la double intégration

Après toutes les simulations réalisées pour les opérateurs et les diverses approches d'intégration considérées, on peut proposer une stratégie d'intégration avec des résultats assez satisfaisants (au moins en ce qui concerne les données réelles avec lesquelles on a travaillé).

Cette stratégie consiste à faire une distinction entre les paramètres considérés pour l'intégration :

- On pourra considérer, d'un côté, les paramètres dont les capteurs radar associés fournissent des valeurs fiables pour la bonne paire classe-cible. Pour ces paramètres, l'option la plus intéressante pour l'intégration semble être celle de la double t-norme drastique. En effet, si l'on accepte que l'on va avoir toujours une valeur de possibilité égale à 1 pour la bonne paire classe-cible, l'option qui nous permettra d'obtenir les performances les meilleures sera la plus drastique, la plus risquée, la plus radicale, puisqu'elle affectera seulement les paires classe-cible erronées.
- Si l'on travaille avec des paramètres pour lesquels on peut avoir des valeurs différentes de 1 pour la possibilité instantanée de la bonne classe, on propose de faire un prétraitement avec une méthode « k sur M » (ou un autre opérateur de moyenne équivalent), afin de filtrer les fluctuations indésirables et transformer, en définitive, le paramètre en question en un autre paramètre similaire à ceux mentionnés dans le paragraphe précédent. Le choix de la valeur concrète de k (ou de la valeur du paramètre pour les moyennes généralisées, ou de l'orness pour les OWA) dépendra de la fiabilité que l'on associe au paramètre à traiter. Plus le paramètre est fiable, plus notre approche pour le prétraitement pourra (et devra, si l'on veut optimiser les performances d'classification) être risquée.

On arrive donc, à proposer un modèle de traitement différencié pour chaque paramètre, en fonction de sa fiabilité, qui comporte la réalisation d'une double intégration temporelle avec t-norme drastique précédée d'un prétraitement « de filtrage », avec un niveau de risque adapté au paramètre en question, dans les cas où celui-ci est nécessaire.

6.6 L'APPROCHE « LEAKY BUCKET »

OBJECTIF : Profiter des avantages de l'intégration hybride [prétraitement « k sur M » + double drastique] avec une implémentation simple, versatile et efficace.

IDEE : Utiliser un algorithme de type « *leaky bucket* » (« panier percé »), similaire à ceux utilisés fréquemment dans le domaine de la gestion du trafic réseau.

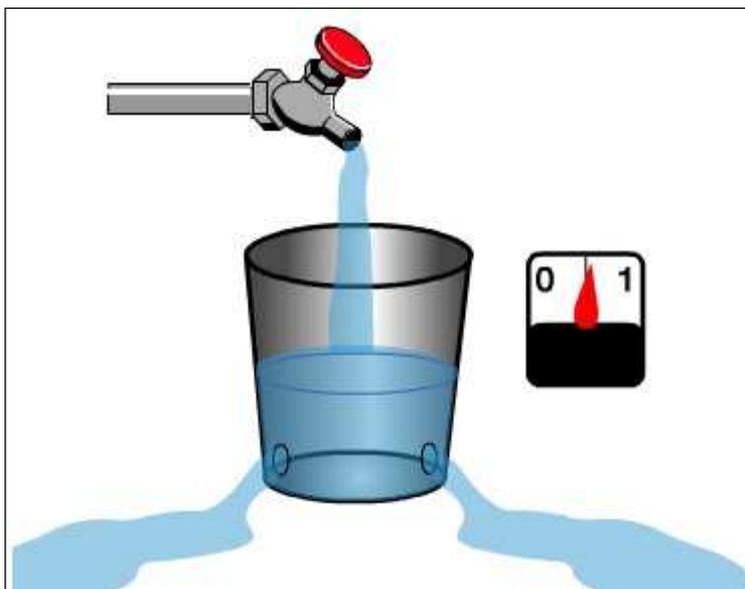


Figure 77. Concept de "leaky bucket" ou "panier percé"

Dans la section précédente, on a parlé du prétraitement « k sur M » qui présentait un grand intérêt car il permettait d'implémenter une certaine « tolérance » face à des « suites » de valeurs différentes de 1 pour la bonne paire classe-cible.

Cette idée de tolérance face à des « suites » ou « rafales » (dans le sens intuitif du mot, et appliqué aux données floues, rien à avoir avec les rafales radar) de valeurs différentes de 1 (il faut pas oublier que l'on considère toujours une « fenêtre » temporelle de M échantillons) nous rappelle le concept de tolérance aux retards relatifs entre paquets de données propre du domaine de la gestion et modélisation du trafic réseau. Dans ce domaine d'étude des Telecoms, l'approche « *leaky bucket* » ou « panier percé » a été utilisée depuis longtemps et il n'est pas donc difficile de trouver une vaste littérature qui la décrit en rigueur et profondeur. Pour une définition plutôt basique et facile à comprendre, on recommande le chapitre X de l'œuvre *Computer Networks* [29]. Ici, on se limitera à présenter cette approche en tant qu'idée originelle pour l'intégration temporelle des données floues née dans le cadre de l'étude qui nous occupe.

On peut donc essayer d'appliquer cette approche pour l'intégration temporelle des données floues dans le cadre de notre système NCTR, afin d'obtenir des performances similaires à celles de la double intégration drastique avec prétraitement par opérateur de moyenne, mais avec une plus grande simplicité et en optimisant l'utilisation de ressources.

Il s'agit tout simplement de considérer un « seau » ou un « panier » avec une certaine capacité et un trou au fond. Chaque fois que l'on trouve une valeur de possibilité différente de 1, on ajoute une certaine quantité d'eau dans le seau. Si l'eau déborde, on active « l'alarme » et considère que la possibilité sera toujours égale à 0 à partir de ce moment-là.

C'est ainsi que, pour chaque paramètre et pour chaque classe possible, on va considérer une variable « bucket_level » initialisée à zéro.

- D'un côté, chaque fois que l'on détecte une valeur de possibilité différente de 1, on ajoute une quantité « possibility_deficit » à cette variable.
- D'un autre côté, on diminue la valeur de la variable « bucket_level » une quantité « empty_rate » après l'arrivée de chaque nouvelle valeur de possibilité. Si « bucket_level » devient négatif, on le remet à zéro (seau vide).
- Initialement, la valeur de la possibilité cumulée est toujours 1. Si le seau déborde, la valeur de la possibilité cumulée devient 0.

6.6.1.1 Avantages de cette approche

L'approche ainsi définie permet d'obtenir les mêmes performances que la double intégration avec norme drastique et prétraitement « k sur M » et présente quelques avantages additionnels d'intérêt, notamment :

- **Optimisation de ressources :**
 - On utilise **une seule variable « bucket_level » en mémoire**. On n'a plus besoin de stocker les dix dernières valeurs de possibilité pour chaque classe.
 - Le **Traitement très simple du point de vue calcul**: une somme, une reste et une comparaison avec un seuil. On n'a plus besoin d'ordonner un vecteur de M échantillons pour calculer chaque nouvelle valeur cumulée ni d'effectuer d'autres types d'opérations arithmétiques (formules des t-normes et moyennes généralisées, produit pour les OWA, etc ...). Cela peut être un aspect critique dans certaines applications face au traitement « temps réel ».
 - Les **performances sont indépendantes du nombre d'échantillons considérés**. Le paramètre M se modélise à travers les relations entre les nouveaux paramètres « possibility_deficit », « bucket_capacity » et « empty_rate », et donc les performances en termes du temps de calcul et utilisation de mémoire devraient être les mêmes pour M=10 et pour M=1.000.000, par exemple.
- Plus de **versatilité** :
 - Si l'on considère une capacité de seau nulle, on obtient un simple détecteur de valeurs différentes de 1 (similaire à l'approche « double t-norme drastique » considérée toute seule). En réglant les différents paramètres, on peut définir **différentes « sensibilités » qui pourraient être fixées en regardant la nature des différents paramètres** et capteurs utilisés pour l'identification. Tout cela peut-être fait avec un seul algorithme, ce qui nous garantit un traitement homogène et plus rapide des divers paramètres (pas plus besoin de faire un prétraitement et puis une double intégration).
 - Nous pouvons imaginer, par exemple, une **version « continue » de l'algorithme** par opposition à cette première proposition « discrète » qui nous permette de considérer un seuil pour le déficit de possibilité cumulé pendant une fenêtre temporelle d'une longueur fixe. Cela revient à transformer la quantité fixe « possibility_deficit » en une quantité variable et égale en chaque instant à la différence : $1 - P_{\text{INSTANTANÉE}}$, au lieu de considérer une « pénalisation » discrète comme dans l'exemple propose. De cette manière, l'on pourrait imaginer une situation où le seau ne déborde pas pour une possibilité moyenne au-dessus de 0.95 dans le cadre d'une certaine fenêtre temporelle, mais il déborde pour une possibilité moyenne de 0.94.

6.6.1.2 Algorithme en pseudocode :

```

for class =1:Number_of_classes

    flag = 1;
    bucket_level = 0;

    for i=1:total_of_instants

        if (flag~=0)

            if (instant_P_matrix(i,class) < 1 )
                bucket_level = bucket_level + possibility_deficit;
            end if

            bucket_level = bucket_level - empty_rate;

            if (bucket_level > bucket_capacity)
                flag = 0;
            elseif (bucket_level < 0)
                bucket_level = 0;
            end_if

        endif

        integrated_matrix(i,class) = flag;

    endfor
endfor

```

6.6.1.3 Résultats obtenus

Avec l'approche « *leaky bucket* », on obtient des résultats similaires à ceux de la section antérieure, et on les améliore même légèrement, de telle sorte que cette méthode d'intégration apporte les meilleurs résultats obtenus dans le cadre de notre étude :

Fonction A

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.6542	0.3681	0.2862

Fonction B

TAUX DE REUSSIT ELEMENTAIRE			TAUX D'ERREUR ELEMENTAIRE		
FUSION	A	B	FUSION	A	B
0.8998	0.7941	0.8085	0.1002	0.2059	0.1915

Fonction C

FUSION FINALE	PARAMETRES A	PARAMETRES B
0.2180	0.1602	0.1689

Fonction D

	FUSION FINALE	PARAMETRES A	PARAMETRES B
Réussite	0.3566	0	0.2559
Erreur	0.0021	0.0021	0

Et donc, on obtient un taux de réussite de 35,66 % pour un taux d'erreur du 0.21 % .

Voici la matrice de confusion normalisée pour l'acceptation :

0.9873	0.2561	0.9912	0.5298	0.4605	0
0.5066	1.0000	0.0859	0	0	0
0	0.0135	1.0000	0.0640	1.0000	0.0020
0	0	0.5101	1.0000	0.7922	0.0647
0	0	0	0.0820	1.0000	0.0820
0	0	0	0.4286	0.0816	1.0000

Erreur quadratique = 3.9049

Voici la matrice non normalisée avec les mesures de précision :

702.0000	194.0000	707.0000	515.0000	440.0000	0	0.2744
430.0000	808.0000	74.0000	0	0	0	0.6159
0	11.0000	612.0000	45.0000	612.0000	1.0000	0.4778
0	0	200.0000	329.0000	247.0000	13.0000	0.4170
0	0	0	50.0000	652.0000	50.0000	0.8670
0	0	0	21.0000	4.0000	49.0000	0.6622
0.6201	0.7976	0.3842	0.3427	0.3335	0.4336	0.4659

kappa = 0.4376

Et voici finalement la matrice normalisée de confusion pour le rejet :

0.0127	0.7439	0.0088	0.4702	0.5395	1.0000
0.4934	0	0.9141	1.0000	1.0000	1.0000
1.0000	0.9865	0	0.9360	0	0.9980
1.0000	1.0000	0.4899	0	0.2078	0.9353
1.0000	1.0000	1.0000	0.9180	0	0.9180
1.0000	1.0000	1.0000	0.5714	0.9184	0

Erreur quadratique = 3.9049

Les résultats ci-dessous ont été obtenus pour des paramètres d'intégration de la forme :

	Paramètre B1	Autres paramètres
<i>bucket_capacity</i>	1.5	0.5
<i>error_penalization</i>	1	1
<i>empty_rate</i>	0.3	0

7 Quelques considérations et perspectives autour de la variable nécessité

Face à l'avenir, il faut rappeler que le travail réalisé jusqu'au présent concerne seulement la variable possibilité, et qu'il serait très souhaitable de poursuivre cette étude avec premièrement la prise en compte de la nécessité seule et avec l'étude du comportement conjoint possibilité – nécessité dans une étape postérieure. Les bases théoriques pour l'étude ont été déjà proposées, de même que des outils pertinents pour l'évaluation des performances et le code Matlab utilisé lui-même, qui pourront tous être réutilisés à ce propos.

En fait, on pourrait inclure ici les résultats pour la variable nécessité des simulations équivalentes à celles que l'on a réalisées pour la possibilité, puisque tout le code nécessaire a été déjà développé et il suffirait de remplacer les données de possibilité instantanée à l'entrée par les valeurs correspondantes de nécessité. Quelque chose de similaire peut être dite à propos des résultats correspondant à la considération conjointe des variables possibilité et nécessité. Par contre, je crois que pour l'interprétation rigoureuse et raisonnée de ces résultats on aurait besoin du temps équivalent, au moins, à la durée d'un autre stage comme le mien et donc j'ai décidé de ne pas entrer à fond dans le sujet.

On peut esquisser néanmoins quelques résultats plutôt intuitifs si l'on applique, par exemple, la fonction A (qui ne dépend d'aucun seuil et qui, par conséquent, nous évite d'en faire le choix) à l'approche « MEOVA » de manière simultanée pour les valeurs de possibilité et nécessité :

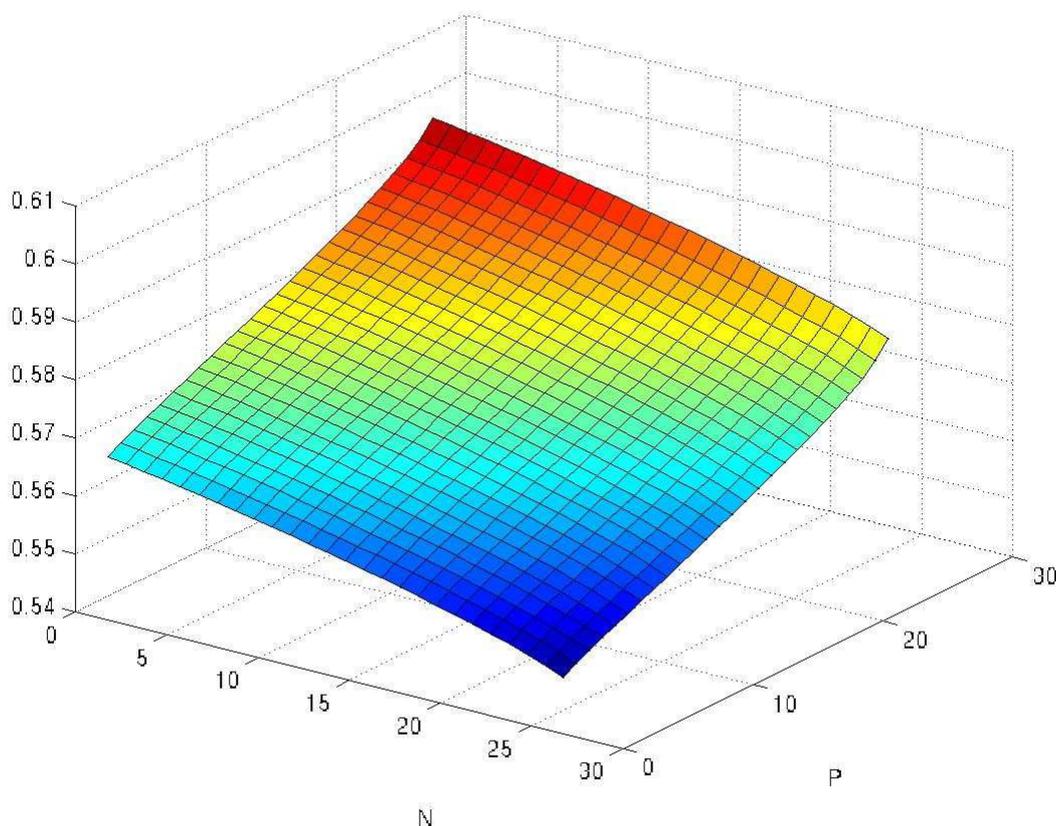


Figure 78. Fonction A - "k sur M" - utilisation conjointe de la possibilité et la nécessité

Sur la figure, les axes représentent des valeurs d'andness pour les opérateurs OWA considérés. On trouve que, apparemment, les résultats les plus proches du cas idéal correspondent à des valeurs grandes d'andness pour la possibilité (opérateurs conjonctifs du type AND ou « min », comme on avait déjà vu) et à des valeurs petites pour la nécessité (opérateurs disjonctifs du type OR ou « max »).

Cela, veut-il dire que l'on devra répéter notre parcours vers le « max » et au-delà pour les valeurs de nécessité, de manière complémentaire à ce que l'on a fait pour la possibilité ? Peut-être oui, mais pas nécessairement, puisque un simple coup d'œil aux résultats de la fonction A appliqué à la nécessité seule montre un comportement pas toujours monotone pour les paramètres de type A :

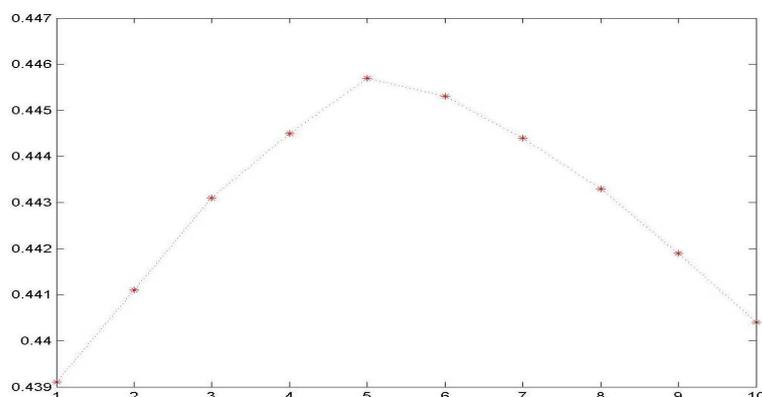


Figure 79. Fonction A - "k sur M" - Intégration des paramètres type A pour la nécessité isolée

On trouve, donc, une complexité apparemment supérieure inhérente au travail avec la nécessité qu'il ne faudra pas ignorer. En plus, les valeurs que l'on a vu jusqu'au présent pour la nécessité semblent parfois être contradictoires (on trouve des cas où la bonne paire classe - cible reste toujours à zéro, tandis que des paires erronées prennent des valeurs plus grandes) et donc il faudrait s'assurer de la bonne qualité des valeurs de nécessité fournies par des niveaux plus bas avant de se lancer dans un étude rigoureuse.

Finalement, on doit signaler que, en ce qui concerne l'utilisation conjointe de possibilité et nécessité, il faudra être prudent pour ne pas perdre le contenu sémantique associé à ces concepts dans le cadre de la théorie des possibilités. En effet, comme Isabelle Bloch l'a bien remarqué au cours d'une conversation sur le sujet, le fait de réaliser l'intégration sur des valeurs de possibilité et nécessité calculées (et pas sur les mesures venant du radar ou sur les densités de possibilité associées) peut nous mener à obtenir parfois des valeurs cumulées de possibilité plus petites que les valeurs correspondantes de nécessité pour un même instant, ce qui constitue une contradiction à éviter et que l'on doit prendre en compte pendant la phase de conception du traitement.

8 Conclusion

Le cadre mathématique de la logique floue et de la théorie des possibilités offre un grand potentiel pour le développement d'applications pratiques complexes dont aujourd'hui on commence seulement à avoir les premiers exemples. Les techniques floues pour la reconnaissance de cibles aériennes non coopérative sont des bons exemples et montrent jusqu'à quel point ces outils formels de base sont puissants en termes de performances dans des situations réelles.

Tout au long de ces pages, on a montré les résultats obtenus pour l'étude de quelques approches diverses concernant l'intégration temporelle des données floues pour la possibilité dans le cadre d'un système d'identification NCTR réel.

Ces conclusions peuvent être résumés par les figures ci-dessous :

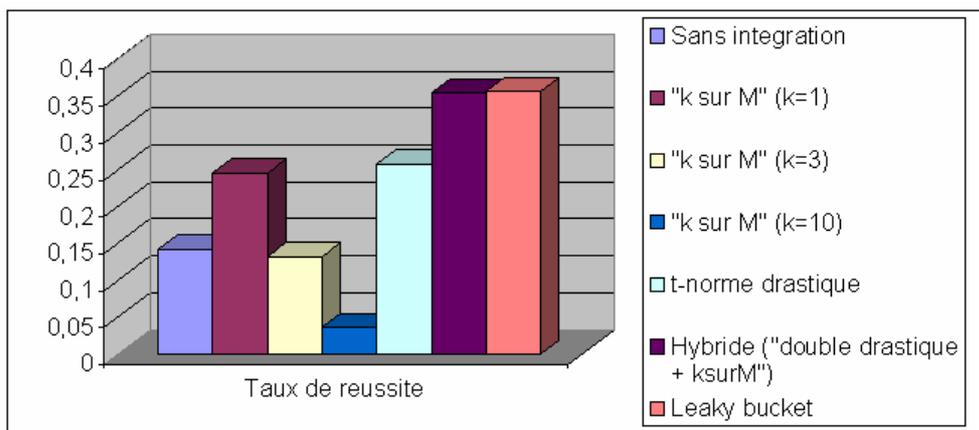


Figure 80.
Comparaison de performances:
taux de réussite

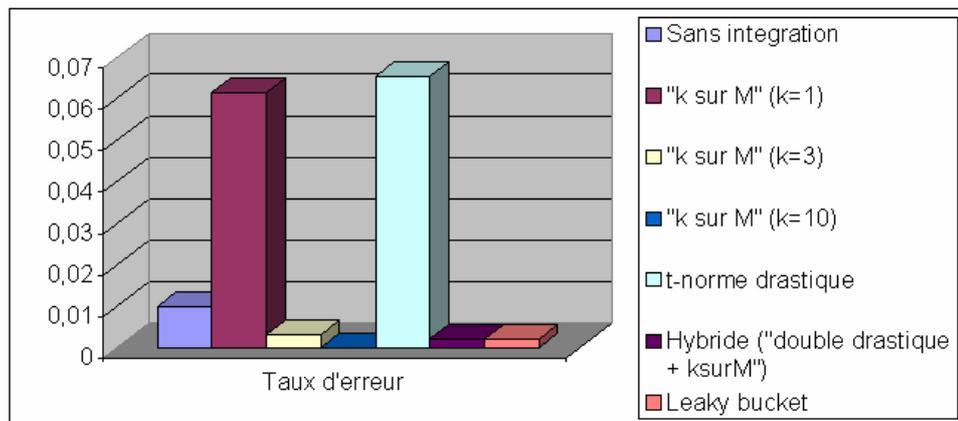


Figure 81
Comparaison de performances:
taux d'erreur

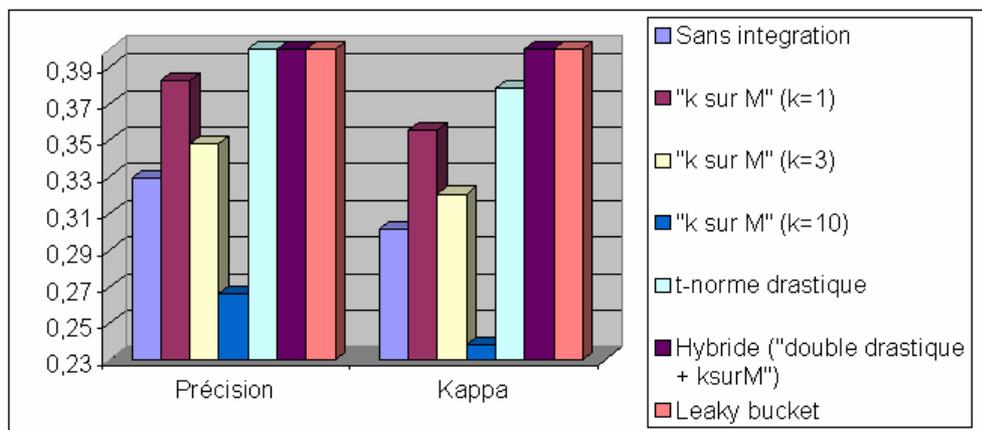


Figure 82
Comparaison de performances:
précision et kappa

	Sans intégration	"k sur M" (k=1)	"k sur M" (k=3)	"k sur M" (k=10)	t-norme drastique	Hybride	"Leaky bucket"
Taux de réussite	0,143	0,2454	0,1321	0,036	0,2582	0,354	0,3566
Taux d'erreur	0,0101	0,0615	0,0032	0	0,0654	0,0021	0,0021
Précision	0,3293	0,3826	0,3482	0,266	0,4041	0,462	0,4659
Kappa	0,3016	0,3557	0,3208	0,2385	0,3785	0,4338	0,4376

Ces résultats nous permettent d'arriver à quelques conclusions pratiques concernant les différents opérateurs étudiés:

- L'approche « k sur M » nous impose un compromis qui ne convient pas toujours à nos intérêts : on doit choisir entre une bonne discrimination avec un taux d'erreur élevé, ou un taux d'erreur satisfaisant au détriment du reste des performances mesurées.
- L'utilisation de la t-norme drastique par elle-même offre des taux de réussite et d'erreur assez similaires à ceux de l'opérateur « min », bien qu'elle apporte une certaine amélioration en termes de précision et coefficient kappa. En tout cas, le taux d'erreur associé est trop élevé pour le type d'application dans le cadre duquel on travaille.
- On observe que les meilleures performances correspondent à l'approche hybride « double t-norme drastique » avec prétraitement « k sur M » et à l'approche « *leaky bucket* », puisqu'elles présentent les valeurs les plus élevées pour le taux de réussite, pour la précision et pour le coefficient kappa, avec le taux d'erreur le plus petit.

Les opérateurs de moyenne (moyennes généralisées et OWA) n'apparaissent pas dans la figure de manière explicite, mais leurs performances peuvent être comparées à celles de l'approche « k sur M », du fait qu'ils permettent de parcourir par des chemins divers, et avec des pas plus ou moins longs (en utilisant la même métaphore, rappelons la « précision infinie » pour le choix de la valeur d'orness associée aux OWA), l'intervalle des performances comprises entre celles associées à l'opérateur « min » et l'opérateur « max ».

Finalement, la plus grande simplicité de l'algorithme « *leaky bucket* », ainsi que ses avantages déjà expliqués en termes de versatilité et optimisation de ressources nous mènent à le proposer comme méthode préférable pour l'intégration temporelle des données floues dans notre système.

En tout cas, il ne faut pas oublier que les simulations que l'on a effectuées ont été faites sur une base d'exemples réels limitée, et il serait donc souhaitable d'étendre cette base avec des nouvelles mesures radar dans des conditions diverses en termes de distance de la cible au radar, présentation de la cible (en rapprochement, en éloignement, en manœuvres, etc.), caractéristiques techniques du radar et d'autres.

C'est pourquoi, bien qu'il faille rester toujours prudent face aux fortes améliorations en termes de performances apportées par les deux dernières approches par rapport à la première (on passe d'un taux de réussite de 13,21% et un taux d'erreur associé de 0,32% pour l'approche « k sur M » avec k=3 à un taux de réussite de 35,66% et un taux d'erreur de 0,21% pour l'algorithme « *leaky bucket* »), on peut considérer que les traitements d'intégration proposés au cours du stage offrent des perspectives intéressantes.

Sur un plan personnel, j'espère que le travail ici présenté sera d'utilité pour d'autres études futures et qu'il contribuera à enrichir notre compréhension d'un domaine passionnant dans le cadre de la fusion de données qui se trouve actuellement en plein développement.

9 Références bibliographiques

Livres

- [1] « *Fusion d'informations en traitement du signal et des images* » sous la direction d'Isabelle Bloch.
Série IC2 – Traitement du Signal et de l'Image. Hermes science – Lavoisier. 2003
- [2] Bernadette Bouchon-Meunier. « *La logique floue et ses applications* ».
Collection vie artificielle. Addison-Wesley, France, 1990.

Articles

- [3] I. Bloch, « *Information Combination Operators for Data Fusion : A Comparative Review with Classification* ».
IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, vol. 26, no. 1.
Janvier 1996
- [4] N. Colin, M. Moruzzis. « Radar target recognition by fuzzy logic »
- [5] C. Delhotte, M. Moruzzis. « *MILORD : a Technical Demonstrator for Long Range Radar Identification* »
NATO – RTO Meeting Proceedings 6, April 1998
- [6] M. Moruzzis, N. Colin, G. Milhem. « *Applications of Multisensor Data Fusion to Target Recognition* »
AGARD – Mission Systems Panels 3rd Symposium, May 1995
- [7] Z. Chair, P.K. Varshney. « *Optimal Data Fusion in Multiple Sensor Detection Systems* ».
IEEE Transactions on Aerospace and Electronic Systems. Vol. AES-22, no 1. Janvier 1986
- [8] V. Nimier. « *Application de la théorie de Dempster Shafer a la fusion de classifieurs* »
ONERA – Rapport technique 6 janvier 1995
- [9] J. Portegies. « *Aircraft Recognition from Features Extracted from Measured and Simulated Radar Range Profiles* »
ASCI dissertation series number 92, Amsterdam, october 2003
- [10] L. Hong, A. Lynch. « *Recursive Temporal-Spatial Information Fusion with Applications to Target Identification* »
IEEE Transactions on Aerospace and Electronic Systems, vol. 29, no. 2. Avril 1993.
- [11] D. Dubois, H. Prade, C. Testemale. « *Weighted Fuzzy Pattern Matching* »
Fuzzy Sets and Systems 28 (1998) , 313-331
- [12] I. Ullibarrí, S. Pita. « *Medidas de concordancia : el indice de Kappa* »
Unidad de Epidemiología Clínica y Bioestadística. Hospital Universitario Juan Canalejo
Cad Aten Primaria 1999 ; 6 : 169-171
- [13] R. Yager, D. Filev. « *Induced Ordered Weighted Averaging Operators* »
IEEE Transactions on systems, man and cybernetics – Part B: Cybernetics, vol. 29, no.2, Avril 1999

- [14] R. Yager, «*On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making*»
IEEE Transactions on systems, man and cybernetics – Part B: Cybernetics, vol. 18, no.1, p.183-190, 1988
- [15] Peter Majlender. « *OWA operators with maximal entropy* ».
- [16] M. O'Hagan. « *Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic* ».
Proc. 22nd Annu. IEEE Asilomar Conf. Signals, Systems, Computers, Pacific Grove CA 1988, pp. 681-689
- [17] L. Zadeh. « *Fuzzy Sets* »
Information and Control, vol. 8, p. 338-353, 1965.
- [18] L. Zadeh. « *Fuzzy Sets as a Basis for a Theory of Possibility* »
Fuzzy Sets and Systems, vol. 1, p. 3-28, 1978.
- [19] H. Hamacher. « *Über logische Verknüpfungen Aussagen und deren zugehörige Bewertungsfunktionen*»
Working Paper, 75/14, 1975, RWTH, Aachen.
- [20] M. Frank. «*On the Simultaneous Associativity of $F(x,y)$ and $x+y-F(x,y)$* »
Aequationes Mathematicae, vol. 19, p. 194-226, 1979.
- [21] J. Dombi. « *A general class of fuzzy operators* »
Fuzzy Sets and Systems, 8(2): 149-163.
- [22] Yager, R. R., «*On a general class of fuzzy connectives*»
Fuzzy Sets and Systems 4, 235-242, 1980.

Autres

- [23] *Corporate Brochure* du groupe Thales 2004
- [24] Rapport Annuel des résultats du groupe Thales pour l'année 2003
- [25] Notes du cours de Henrik Legind Larsen sur la logique floue et la théorie des ensembles flous.
Aalborg University Esbjerg
- [26] Site web sur la recherche NCTR du laboratoire de Traitement du signal (TSD) de POSTECH et Yeungnam University (Korea). <http://yu.ac.kr/~kkt/>
- [27] Plusieurs sources non officielles :
- « *Situation Awareness in Air Combat* » article en ligne de Peter Grining, 2000 (source non contrastée).
- « *The Bombrats's impending war on terrorism* », article en ligne de Tony Holmes chez www.global-defence.com
- Air Combat Information Group Forum (www.acig.org)
- [28] J. Kurose et K. Ross. « *Computer Networking: A Top-Down Approach Featuring the Internet*»
Pearson - Addison Wesley. Second Edition (juillet 2002)
- [29] A. Tanenbaum. « *Computer Networks* »
Pearson – Prentice hall (octobre 2002)

Finalement, voici quelques ouvrages consultés ponctuellement afin d'avoir une certaine culture de base sur des généralités radar:

- *Les RADARS : Principes de base*
C. Delhotte
Formation interne THALES - 24 Juin 2003, Bagneux
- *Bases physiques du radar*
F . Le Chevalier
Revue technique THOMSON-CSF vol. 25 N°4 Decembre 1993
- *Cibles et environnement et chaines fonctionnelles TS/TD*
Guy Desodt
Suite communication AIRSYS / RD / RAN – 250/98
- *Que sais-je ? Le radar*
Michel-Henri Carpentier
Presses Universitaires de France
- *Le Radar : 1904-2004*
Yves BLANCHARD
Ellipses – THALES
- *Introduction to Radar Systems 3rd Ed*
M.I. Skolnik
McGraw-Hill, 2002
- *Traitement des images de radar à synthèse d'ouverture*
Sous la direction de Henri Maître
Série IC2 – Traitement du Signal et de l'Image. Hermes science – Lavoisier