

# Opérateurs d'agrégation pour les données multiples. L'exemple du triple $\Pi$ et de ses variantes

Sébastien Régis\*, Richard Emilion\*\*  
Andrei Doncescu\*\*\*

\*LAMIA, Université des Antilles et de la Guyane, 97159 Pointe-à-Pitre, Guadeloupe France  
sregis@univ-ag.fr,

<http://www.lamia.univ-ag.fr>

\*\*MAPMO Université d'Orléans

autre-adresse@email

<http://www.une-autre-page.html>

\*\*\*LAAS-CNRS Av. Colonel Roche Toulouse

adocnesc@univ-ag.fr

<http://www.laas.fr>

**Résumé.** Cet article traite du sujet des opérateurs d'agrégation issus de la logique floue. En particulier, un exemple d'opérateur d'agrégation nommé triple Pi, ainsi que certaines de ses variantes et de ses généralisations, sont présentés ; le triple Pi présente certaines potentialités intéressantes pour la fusion d'informations. Quelques applications de cet opérateur sont également présentées. De façon plus général, à travers cet exemple, on cherche à illustrer quelques-unes des nombreuses possibilités offertes par l'utilisation des opérateurs d'agrégations dans le domaine de la fouille de données complexes et massives.

## 1 Introduction

Les avancées technologiques au niveau des capteurs et du stockage ont permis une véritable explosion des données numériques, augmentant sans cesse la quantité, la qualité et la diversité de ces informations. Ainsi, au début du siècle la question était "Peut-on récupérer et stocker ces données ?" et aujourd'hui la problématique est plutôt la suivante : "comment exploiter toutes ces données ?". En effet l'adage qui stipule que trop d'information tue l'information n'a jamais été aussi vrai. Les masses de données, leur volume et leur complexité sont telles aujourd'hui, qu'elles nécessitent des outils informatiques de plus en plus sophistiqués. Diverses approches issues de l'informatique et des mathématiques appliquées ont été proposées pour gérer ces informations. Par exemple, les opérateurs d'agrégation issus de la théorie des sous-ensembles flous permettent une synthèse de ces informations diverses et variées. Il existe de nombreux opérateurs d'agrégation et le choix d'un de ces opérateurs dépend de nombreux paramètres. Mais ce choix est avant tout lié à la *fusion* elle-même. Avant d'aller plus loin il nous semble important de rappeler la définition de la fusion. Nous reprenons la définition proposée par (Bloch et Hunter, 2001) qui fut donnée lors des travaux du Groupe Européen de

Travail sur la Fusion (FUSION) :

*La fusion consiste à réunir ou agréger des informations provenant de différentes sources, et à exploiter cette information réunie ou agrégée, dans diverses applications comme la réponse à une question, la prise de décision, une estimation numérique, etc.*

Cette définition met l'accent sur deux éléments principaux. D'abord, elle met l'emphase sur la combinaison de l'information. Puis l'accent est mis sur l'objectif de la fusion. Pour définir la nature de cette combinaison et l'objectif de fusion sous-jacent, il convient donc de connaître le type de données que l'on cherche à fusionner. Nous reprenons succinctement les types de données proposés dans Bloch et Hunter (2001) et repris dans Dubois et Prade (2004) :

- *les observations*. Elles décrivent le monde d'un point de vue plus ou moins particulier. Il s'agit le plus souvent de données numériques fournies par des capteurs.
- *la connaissance*. Elle décrit la façon dont le monde est *en général*. Il s'agit de données plus subjectives que les observations, et qui sont souvent issues de personnes plutôt que de capteurs.
- *la préférence*. Ce sont des informations subjectives qui décrivent comment *on aimerait* que le monde soit. Il s'agit là aussi de données issues de personnes.
- *les régulations*. Il s'agit d'informations génériques qui sont la plupart du temps énoncées sous forme de lois.

Les opérateurs d'agrégation permettent de fusionner la plupart des informations et sont bien adaptés pour synthétiser des éléments issus des 3 premiers types d'informations. Par ailleurs, ils sont parfaitement adaptés pour agréger des informations ponctuelles issues de plusieurs variables. Ainsi, si à un instant  $t$  on dispose pour chaque variable d'une information ponctuelle, on pourra fusionner toutes ces informations ponctuelles à cet instant  $t$  via un opérateur d'agrégation.

Dans cet article, nous nous proposons, à travers un exemple d'opérateur particulier, le triple  $\Pi$ , d'illustrer quelques-unes des potentialités des opérateurs d'agrégation dans le domaine des données multiples. Nous cherchons donc à faire découvrir ces outils et leurs possibilités pour les utilisateurs confrontés à la complexité de données multiples et massives. Dans la section 2, nous faisons un survol des différents opérateurs d'agrégation. La section 3 est consacrée à l'opérateur triple  $\Pi$  et à ses variantes dans différents cas : fini, discret et continu. La section 4 présente deux exemples d'application de cet opérateur et de ses variantes. Quelques perspectives sont présentées dans la section 6 avant de conclure dans la section 7.

## 2 Fusion d'informations et opérateurs d'agrégation

Les opérateurs d'agrégation, eux sont issus de la théorie des sous-ensembles flous. Un opérateur d'agrégation peut être définie comme une fonction  $f : [0, 1]^n \rightarrow [0, 1]$ ,  $n \in \mathbb{N}$ , qui vérifient les deux conditions suivantes (Dubois et Prade, 2004; Detyniecki, 2000; Campanella et Ribeiro, 2011) :

$$\begin{aligned} f(\underbrace{0, 0, \dots, 0}_{n \text{ fois}}) &= 0 \\ f(\underbrace{1, 1, \dots, 1}_{n \text{ fois}}) &= 1 \end{aligned} \tag{1}$$

et pour tout  $x, y \in [0, 1]^n$  :

$$x \leq y \Rightarrow f(x) \leq f(y) \quad (2)$$

L'équation 1 est une condition dite de préservation des bords tandis que l'équation 2 est appelée condition de monotonie.

Il est à noter que les opérateurs d'agrégation ont donc un domaine de définition compris entre 0 et 1. Cela ne pose pas vraiment problème car, *a priori*, quelque soient les données numériques traitées, il est en général aisé de se ramener à cet intervalle. On aura aussi noté que les opérateurs mathématiques classiques (moyenne arithmétique, maximum, minimum, médian...) peuvent être définis comme des opérateurs d'agrégation.

Parmi les opérateurs les plus connus, on peut citer (Dubois et Prade, 2004; Martin, 2005; Pradera et Trillas, 2006; Beliakov et al., 2007) :

- les normes triangulaires (ou t-normes). Il s'agit d'opérateurs conjonctifs. Ces opérateurs sont généralement utilisés quand toutes les sources sont supposées fiables. Les plus connus sont le minimum et le produit. Citons également les copulas (voir Nelsen (1999); Dubois et Prade (2004)) qui sont des généralisations d'opérateurs conjonctifs. On notera enfin qu'il existe aussi des normes triangulaires paramétriques qui possèdent des paramètres modifiables (Yager, 1980; Weber, 1983; Zimmermann, 2001).
- les conormes triangulaires (ou t-conormes). Ce sont des opérateurs disjonctifs. Ils sont utilisés quand il existe au moins une source fiable, les autres sources pouvant être incertaines. La plus connue des t-conormes est le maximum. Il existe également des conormes triangulaires paramétriques (Yager, 1980; Weber, 1983; Zimmermann, 2001).
- les connectifs mixtes ou connectifs mixtes de compensation (on parle aussi d'opérateurs hybrides). Il s'agit de combinaisons linéaires ou non linéaires de t-normes et de t-conormes. L'objectif de ces combinaisons est d'allier les caractéristiques des t-normes et des t-conormes en faisant varier le ou les paramètres de la combinaison. En utilisant ces connectifs mixtes, il est possible de faire varier le comportement de l'opérateur entre les situations extrêmes, allant de la conjonction à la disjonction : on espère ainsi que la t-norme contre-balancera la t-conorme et inversement, d'où la notion de compensation (Zimmermann et Zynso, 1980; Piera-Carreté et al., 1988). Il faut remarquer qu'une grande partie des uninormes et des normes nulles (présentées juste après) sont aussi des connectifs mixtes.

Voici deux exemples de connectifs mixtes. Notons  $T$  une t-norme et  $C$  sa t-conorme duale. Parmi les connectifs mixtes les plus connus on trouve :

- le connectif mixte linéaire :

$$\alpha T(x_1, \dots, x_n) + (1 - \alpha)C(x_1, \dots, x_n) \quad (3)$$

- le connectif mixte géométrique :

$$T(x_1, \dots, x_n)^\alpha + C(x_1, \dots, x_n)^{(1-\alpha)} \quad (4)$$

où  $0 \leq \alpha \leq 1$ .

Piera-Carreté et al. (1988) ont montré que le nombre de classes fournies par le connectif mixte de compensation linéaire varie en fonction de la valeur du paramètre  $\alpha$  : plus  $\alpha$  est proche de 1, plus le nombre de classes augmente. On notera cependant qu'il a été

## Opérateurs d'agrégation et données multiples

montré (Yager et Rybalov, 1998) qu'un connectif mixte ne permet pas toujours d'allier les caractéristiques des t-normes et des t-conormes.

- les moyennes. Comme leur nom l'indique, il s'agit d'opérateurs qui fournissent une valeur comprise entre le minimum et le maximum possible. Citons par exemple les moyennes arithmétiques et géométriques ou encore la médiane. Les moyennes pondérées ordonnées (Yager, 1988, 2004), appelées *OWA* (*Ordered Weighted Averaging*) sont aussi des opérateurs moyennes dans lesquelles on introduit une pondération dépendant de l'importance et de la fiabilité des sources.
- les normes zéro (*nullnorms*) définies par Calvo et al. (2001). Ce sont des opérateurs commutatifs, associatifs et possédant un élément absorbant  $a \in [0, 1]$  que l'opérateur peut en général fixer a priori.
- les uninormes proposées par Yager et Rybalov (Yager, 1994; Yager et Rybalov, 1996, 1998). Il s'agit en fait d'opérateurs commutatifs, associatifs et possédant un élément neutre  $e \in [0, 1]$  que l'utilisateur peut, dans la plupart des cas, fixer a priori. Dans la pratique, une uninorme est souvent définie par une t-norme sur l'intervalle  $[0, e]$  et par une t-conorme sur l'intervalle  $[e, 1]$ . Il faut noter que le triple  $\Pi$  développé par Yager et Rybalov (1998) est un cas particulier d'uninorme. Notons aussi que ce triple  $\Pi$  est également une somme symétrique (Waissman-Vilanova, 2000). On rappelle que les sommes symétriques, introduites par Silvert (1979), sont des opérateurs dont la particularité est d'être symétrique par rapport à un sous-ensemble et à son complément.

Ces différents opérateurs possèdent des propriétés intrinsèques qui peuvent être utiles en fonctions des objectifs visés et du domaine d'application. Une des propriétés les plus remarquables est celle du renforcement total. Un opérateur totalement renforcé est un opérateur renforcé positivement et renforcé négativement. Le renforcement total est une propriété intéressante car elle modélise le comportement humain (Elkan, 1994), (Yager et Rybalov, 1998).

Un opérateur d'agrégation  $L$  dont les arguments sont dans l'intervalle  $[0, 1]$ , a la propriété de renforcement positif, si lorsque tous ses attributs sont affirmatifs (i.e. supérieur ou égaux à 0,5), il vérifie :

$$L(x_1, \dots, x_n) \geq \max_i [L(x_i)] \quad (5)$$

Les t-normes ont la particularité d'être positivement renforcés.

De façon similaire, un opérateur d'agrégation  $L$  dont les arguments sont dans l'intervalle  $[0, 1]$ , a la propriété de renforcement négatif si lorsque tous ses attributs sont non-affirmatifs (i.e. inférieur ou égaux à 0,5), il vérifie :

$$L(x_1, \dots, x_n) \leq \min_i [L(x_i)] \quad (6)$$

Les t-conormes ont la particularité d'être négativement renforcés.

Un opérateur qui possède les deux propriétés est défini comme étant *totalement renforcé* (*full reinforced*). Les uninormes sont des opérateurs totalement renforcés (Yager et Rybalov, 1996). Le triple  $\Pi$ , qui est une uninorme particulière, est totalement renforcé (Yager et Rybalov, 1998). Cependant, comme nous l'avons déjà signalé plus haut, un connectif mixte (combinaison d'une t-norme et d'une t-conorme) n'est pas forcément totalement renforcé.

### 3 Les opérateurs $3\Pi$ et $M3\Pi$

#### 3.1 L'opérateur $3\Pi$

Avant de fournir la définition de l'opérateur  $3\Pi$ , rappelons deux propriétés importantes de cet opérateur :

- l'opérateur  $3\Pi$  est une uninorme (Yager et Rybalov, 1998).
- l'opérateur  $3\Pi$  est une généralisation des sommes symétriques (Waissman-Vilanova, 2000).

Les uninormes ont été proposées par (Yager et Rybalov, 1996) tandis que les sommes symétriques ont été proposées par (Silvert, 1979).

**Définition** (Yager et Rybalov, 1996) Une uninorme est une fonction  $U : J \rightarrow I$  telle que pour tout  $x, y, z, s, t \in I^n, n = 1, 2, \dots$ , les propriétés suivantes sont vérifiées :

1. Commutativité :  $U(x, y) = U(y, x)$
2. Monotonie :  $U(x, y) \geq U(s, t)$  si  $x \geq s, y \geq t$
3. Associativité :  $U(x, U(y, z)) = U(U(x, y), z)$
4. Élément neutre : il existe un élément fixe  $g \in [0, 1]$  tel que  $U(x, g) = U(x)$ .

**Définition** (Silvert, 1979) Une somme symétrique  $S$  est un opérateur d'agrégation continu auto-dual, où l'auto-dualité signifie que :

$$S(x_1, x_2, \dots, x_n) = 1 - S(1 - x_1, 1 - x_2, \dots, 1 - x_n) \quad (7)$$

Pour  $x_i \in [0, 1], i = 1, \dots, n$ .

Comme indiqué dans Detyniecki (2000), Silvert a montré qu'une somme symétrique avec deux arguments peut être écrite sous la forme :

$$S(x, y) = \frac{G(x, y)}{G(x, y) + G(1 - x, 1 - y)} \quad (8)$$

où  $G$  est une fonction positive, continue, croissante satisfaisant  $G(0, 0) = 0$ .

**Définition** (Yager et Rybalov, 1998) Le triple  $\Pi$  est défini comme suit :

$$PI(x_1, \dots, x_n) = \frac{\prod_{j=1}^n x_j}{\prod_{j=1}^n x_j + \prod_{j=1}^n (1 - x_j)} \quad (9)$$

avec  $x_j \in [0, 1]$ .

On notera que dans le cas problématique où le dénominateur serait nul, Yager propose de définir arbitrairement le triple  $\Pi$  comme étant égal à 0 ou à 1 (Yager, 2002). Cet opérateur  $3\Pi$  a comme particularité d'être totalement renforcé c'est-à-dire qu'il est renforcé positivement et négativement (Yager et Rybalov, 1998). Cet opérateur triple  $\Pi$  et ses variantes ont été utilisés pour des applications dans le domaine des systèmes multi-agents (Yager, 2002), dans les bio-procédés (Doncescu et al., 2007), les systèmes automatiques de navigations (Faurie, 2011), ou le traitement d'images (Nagau et al., 2010).

### 3.2 La moyenne $M3\Pi$

**Définition** (Emilion et al., 2004) L'opérateur  $M3\Pi$ , qui dérive du  $3\Pi$ , se définit comme suit :

$$M3\Pi(x_1, \dots, x_n) = \frac{\prod_{j=1}^n (x_j)^{(1/n)}}{\prod_{j=1}^n (x_j)^{(1/n)} + \prod_{j=1}^n (1 - x_j)^{(1/n)}} \quad (10)$$

avec  $x_j \in [0, 1]$ . Cet opérateur ne peut être totalement renforcé du fait de sa nature (c'est une moyenne) mais il possède une propriété ayant une certaine similitude avec celle de renforcement total du  $3\Pi$ . Cette propriété est présentée ci-dessous.

Soit  $M3\Pi$  la moyenne triple  $\Pi$ . On considère la moyenne arithmétique classique  $\frac{1}{n} \sum_{j=1}^n x_j$ . Alors :

Si  $\forall j \in 1, \dots, n$ , on a  $x_j \geq 0,5$ , on a :

$$M3\Pi(x_1, \dots, x_n) \geq \frac{1}{n} \sum_{j=1}^n x_j \quad (11)$$

Si  $\forall j \in 1, \dots, n$ , on a  $x_j \leq 0,5$ , on a :

$$M3\Pi(x_1, \dots, x_n) \leq \frac{1}{n} \sum_{j=1}^n x_j \quad (12)$$

Cette propriété a été baptisée *renforcement moyen* (Doncescu et al., 2007) en référence au renforcement total du triple  $\Pi$ .

### 3.3 Vers une généralisation du triple $\Pi$ pour agréger d'autres types de données

En suivant la même démarche, on peut généraliser cet opérateur afin d'appliquer ces opérateurs dérivés sur de nouveaux types de données. En effet, comme nous l'avons déjà signalé, les opérateurs d'agrégation sont parfaitement adaptés pour synthétiser des données ponctuelles issues de sources diverses et mesurées à un instant  $t$ . Or les données sont telles aujourd'hui qu'elles sont souvent organisées en séries temporelles, voire en flux de données, on doit donc chercher à traiter des plages de données sur des intervalles de temps plutôt qu'une information ponctuelle. Or les opérateurs d'agrégation sont peu usités pour ces données arrivant en continu. On notera cependant qu'ils sont utilisés sur des séries temporelles afin de :

- lisser des données sur ces séries temporelles (Yager, 2008)
- faire de la prévision sur ces celles-ci (Komornik et al., 2006; Jilani et Burney, 2008)

En d'autres termes, on peut dire que les opérateurs d'agrégation sont surtout utilisés d'un point de vue local et non pour résumer l'information sur un intervalle temporel. La question qui en découle est la suivante : Peut-on trouver ou définir des opérateurs issus du triple  $\Pi$  qui synthétisent des plages de données issues de séries temporelles ?

En fait, le paragraphe suivant propose diverses généralisations du triple  $\Pi$  qui peuvent être utilisées sur des séries temporelles. Ces généralisations ne sont pas au sens propre du terme des opérateurs d'agrégation mais elles peuvent permettre de fusionner des données issues de séries temporelles.

## 4 Généralisation de l'opérateur $3\Pi$ pour des séries temporelles

### 4.1 Liminaire

Définissons la généralisation d'une somme symétrique :

$$PI(x_1, \dots, x_n) = \frac{\prod_{j=1}^n G(x_j)}{\prod_{j=1}^n G(x_j) + \prod_{j=1}^n G(1-x_j)} \quad (13)$$

où  $G(x)$  est une fonction appelée *fonction génératrice* et qui est non négative et croissante (Silver, 1979; Waissman-Vilanova, 2000). Il est clair que pour définir l'opérateur triple  $\Pi$  on prend la fonction  $G(x_j) = x_j$  et pour la moyenne triple  $\Pi$  on prend la fonction  $G(x_j) = x_j^{1/n}$ . A partir de l'équation 13, on cherche cette fois une généralisation de la fonction  $G$  pour différents cas : ensembles finis, cas discret et cas continu. On va donc définir une généralisation de l'opérateur que l'on pourra appliquer à des fonctions.

### 4.2 Définition et inégalité

#### Définition

Soit  $f : \Omega \rightarrow [0, 1]$  une fonction mesurable définie sur un espace de probabilité  $(\Omega, \mathcal{F}, \mu)$  et prenant ses valeurs dans  $[0, 1]$ . Alors,

$$G3\Pi = \frac{e^{\int_{\Omega} \log(f) d\mu}}{e^{\int_{\Omega} \log(f) d\mu} + e^{\int_{\Omega} \log(1-f) d\mu}} \quad (14)$$

est appelée *généralisation de  $3\Pi$*  et est écrit  $G3\Pi$ .

Evidemment, le  $G3\Pi$  n'est pas un opérateur d'agrégation mais il possède des propriétés similaires à celles du  $3\Pi$  comme nous le voyons ci-dessous.

#### Proposition 1

Soit  $f : \Omega \rightarrow (0, \frac{1}{2}]$  une fonction mesurable définie sur un espace de probabilités  $(\Omega, \mathcal{F}, \mu)$  et prenant ses valeurs dans  $(0, \frac{1}{2}]$ . Alors

$$\frac{e^{\int_{\Omega} \log(f) d\mu}}{e^{\int_{\Omega} \log(f) d\mu} + e^{\int_{\Omega} \log(1-f) d\mu}} \leq \int_{\Omega} f d\mu \quad (15)$$

Si  $f$  prend ses valeurs dans  $[\frac{1}{2}, 1)$  alors l'inégalité inverse est vérifiée.

Notons que l'intégrale ci-dessus est bien définie puisque  $\log(f) < 0$   $\mu$  presque partout et  $e^{-\infty} = 0$ .

Cette inégalité théorique et générale est la plus importante car toutes les autres inégalités en découlent. Elle a été démontrée dans (Emilion et al., 2013b,a). Le cas fini correspond à un opérateur d'agrégation ; c'est en fait la définition de la moyenne  $M3\Pi$  et de ses propriétés. Le cas discret et le cas continu permettent de considérer de nouvelles applications pour les séries temporelles (voire des flots de données), et montrent qu'on peut trouver une généralisation du  $3\Pi$  pour des séries temporelles. Nous détaillerons les 3 cas (fini, discret et continu) que nous venons de mentionner.

### 4.3 Le cas fini

Soit  $\Omega = \{1, \dots, n\}$  et soit  $\mu$  une mesure de probabilité uniforme  $\mu\{i\} = \frac{1}{n}$  pour  $i = 1, \dots, n$ . Si  $f(i)$  est noté  $x_i$ , alors, avec la convention que  $\log(0) = -\infty$  et  $e^{-\infty} = 0$ , on a :

$$\int_{\Omega} f d\mu = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\int_{\Omega} \log(f) d\mu = \frac{1}{n} \sum_{i=1}^n \log(x_i) = \log((x_1 \dots x_n)^{\frac{1}{n}})$$

et

$$e^{\int_{\Omega} \log(f) d\mu} = (x_1 \dots x_n)^{\frac{1}{n}}$$

ainsi dans le cas fini l'inégalité (15) peut être écrite :

$$\frac{(x_1 \dots x_n)^{\frac{1}{n}}}{(x_1 \dots x_n)^{\frac{1}{n}} + ((1-x_1) \dots (1-x_n))^{\frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n x_i \quad (16)$$

quand  $0 \leq x_i \leq \frac{1}{2}$  pour tout  $i$  ; l'inégalité inverse est vérifiée quand  $\frac{1}{2} \leq x_i \leq 1$  pour tout  $i$ .

Comme nous l'avons indiqué plus haut, l'inégalité (16) a déjà été prouvée et elle représente la principale propriété de la Moyenne 3Π (Emilion et al., 2004). L'intérêt majeur de M3Π est de fournir une meilleure discrimination que la moyenne arithmétique classique.

### 4.4 Le cas discret

Plus généralement, soit  $\Omega = \{0, 1, \dots, n, \dots\}$  et soit  $\mu$  une distribution de probabilité sur  $\Omega$ . Si  $f(i)$  est notée  $x_i$ , alors, avec les mêmes conventions que ci-dessus, on a :

$$\int_{\Omega} f d\mu = \sum_{i=0}^{+\infty} x_i \mu\{i\},$$

$$\int_{\Omega} \log(f) d\mu = \sum_{i=0}^{+\infty} \log(x_i) \mu\{i\} = \log(\prod_{i=0}^{+\infty} x_i^{\mu\{i\}})$$

et par conséquent

$$e^{\int_{\Omega} \log(f) d\mu} = \prod_{i=0}^{+\infty} x_i^{\mu\{i\}}$$

de sorte que dans le cas discret, l'inégalité (15) peut être réécrite comme suit :

$$\frac{\prod_{i=0}^{+\infty} x_i^{\mu\{i\}}}{\prod_{i=0}^{+\infty} x_i^{\mu\{i\}} + \prod_{i=0}^{+\infty} (1-x_i)^{\mu\{i\}}} \leq \sum_{i=0}^{+\infty} x_i \mu\{i\}. \quad (17)$$

Par exemple si  $\mu$  est une distribution géométrique ( $p$ ) avec le paramètre  $0 \leq p \leq 1$ , telle que  $\mu\{i\} = (1-p)p^i$ , on a :

$$\frac{\prod_{i=0}^{+\infty} x_i^{(1-p)p^i}}{\prod_{i=0}^{+\infty} x_i^{(1-p)p^i} + \prod_{i=0}^{+\infty} (1-x_i)^{(1-p)p^i}} \leq (1-p) \sum_{i=0}^{+\infty} x_i p^i \quad (18)$$

D'autre part si  $\mu$  est une distribution de Poisson( $\lambda$ ) avec le paramètre  $\lambda > 0$ , telle que  $\mu\{i\} = e^{-\lambda} \frac{\lambda^i}{i!}$ , on a :

$$\frac{\prod_{i=0}^{+\infty} x_i^{e^{-\lambda} \frac{\lambda^i}{i!}}}{\prod_{i=0}^{+\infty} x_i^{e^{-\lambda} \frac{\lambda^i}{i!}} + \prod_{i=0}^{+\infty} (1-x_i)^{e^{-\lambda} \frac{\lambda^i}{i!}}} \leq e^{-\lambda} \sum_{i=0}^{+\infty} x_i \frac{\lambda^i}{i!}. \quad (19)$$

Bien sûr, (19) est un cas particulier de (17) quand on prend  $\mu\{i\} = \frac{1}{i}$  si  $i = 1, \dots, n$  et  $\mu\{i\} = 0$  sinon.

Les figures 1 et 2 donnent deux exemples simulés pour illustrer l'inégalité de la généralisation du  $3\Pi$ .

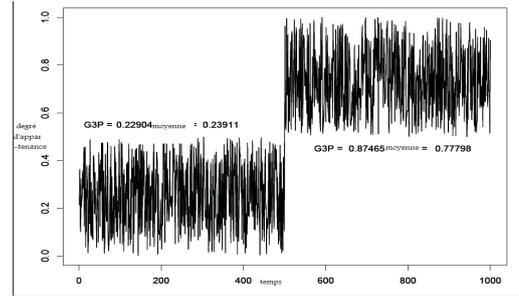


FIG. 1 – Exemple de l'utilisation de la propriété du  $G3\Pi$ . Les valeurs de la courbe représentent des degrés d'appartenance à une classe donnée, allant de 0 (non appartenance à cette classe) à 1 (appartenance à cette classe). Cette série temporelle a été simulée uniformément dans l'intervalle de valeurs  $[0; 0.5]$  sur l'intervalle temporel  $[0; 500]$  et uniformément dans l'intervalle de valeurs  $[0.5; 1]$  sur l'intervalle temporel  $[501; 1000]$ . Comme attendu, l'opérateur  $G3\Pi$  donne une valeur plus petite (respectivement plus grande) que la moyenne sur l'intervalle temporel  $[0; 500]$  (respectivement sur l'intervalle temporel  $[501; 1000]$ ).

Cette inégalité est certainement la plus prometteuse, car elle correspond à une fusion d'informations sur une série temporelle, afin d'avoir un point de vue global sur un intervalle de temps donné. Ainsi, cette inégalité qui peut s'appliquer à des cas discrets, permettra une plus grande discrimination sur un intervalle temporel donné. Comme on l'a vu, on peut utiliser cette généralisation d'opérateur sur des phénomènes connus comme par exemple les arrivées de type poissonnien.

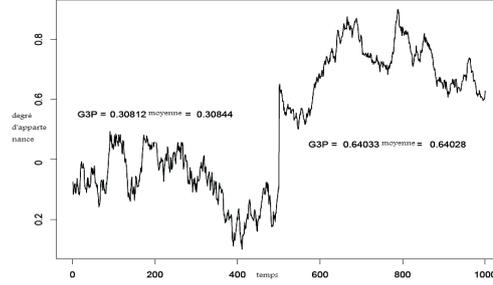


FIG. 2 – Autre exemple d'utilisation de la propriété du G3Π. Ici il s'agit d'une série temporelle simulée par un processus stochastique de type brownien dans un intervalle de valeurs de  $[0; 0.5]$  (respectivement dans un intervalle de valeurs  $[0.5; 1]$ ) sur un intervalle de temps  $[0; 500]$  (respectivement sur un intervalle de temps  $[501; 1000]$ ). Comme attendu, l'opérateur G3Π donne une valeur plus petite (respectivement plus grande) que la moyenne sur l'intervalle temporel  $[0; 500]$  (respectivement sur l'intervalle temporel  $[501; 1000]$ ). Dans ce cas la différence entre le G3Π et la moyenne est très faible.

## 4.5 Le cas continu

### 4.5.1 Cas général

Soit  $\Omega = \mathbb{R}$  l'ensemble des réels et soit  $\mu = g.\lambda$  une distribution de probabilité sur  $\Omega$  avec une densité  $g$  pour la mesure de Lebesgue  $\lambda$ . Alors l'inégalité (15) peut être écrite comme suit :

$$\frac{e^{\int_{\mathbb{R}} \log(f(t))g(t)dt}}{e^{\int_{\mathbb{R}} \log(f(t))g(t)dt} + e^{\int_{\mathbb{R}} \log(1-f(t))g(t)dt}} \leq \int_{\mathbb{R}} f(t)g(t)dt. \quad (20)$$

Par exemple si  $\mu$  est la distribution uniforme sur un intervalle fini  $[a, b]$ ,  $a < b$  dans  $\mathbb{R}$ , alors l'inégalité (20) donne :

$$\frac{e^{\frac{1}{b-a} \int_a^b \log(f(t))dt}}{e^{\frac{1}{b-a} \int_a^b \log(f(t))dt} + e^{\frac{1}{b-a} \int_a^b \log(1-f(t))dt}} \leq \frac{1}{b-a} \int_a^b f(t)dt \quad (21)$$

D'autre part, si  $\mu$  est une distribution exponentielle sur  $[0, +\infty)$  avec le paramètre  $\theta > 0$ , on a :

$$\frac{e^{\theta \int_0^{+\infty} \log(f(t))e^{-\theta t} dt}}{e^{\theta \int_0^{+\infty} \log(f(t))e^{-\theta t} dt} + e^{\theta \int_0^{+\infty} \log(1-f(t))e^{-\theta t} dt}} \leq \theta \int_0^{+\infty} f(t)e^{-\theta t} dt. \quad (22)$$

On trouve une inégalité semblable avec une distribution standard comme la distribution de type normale  $(\mu, \sigma^2)$  ou une distribution de type exponentielle.

### 4.5.2 Fonctions positives

Maintenant, on suppose que  $f > 0$ . Alors, comme on a  $0 < \frac{f}{2(1+f)} \leq \frac{1}{2}$ , l'inégalité (15) implique ce qui suit.

Soit  $f : \Omega \rightarrow (0, +\infty)$  une fonction positive mesurable sur un espace de probabilité alors,

$$\frac{e^{\int_{\Omega} \log\left(\frac{f}{2(1+f)}\right) d\mu}}{e^{\int_{\Omega} \log\left(\frac{f}{2(1+f)}\right) d\mu} + e^{\int_{\Omega} \log\left(\frac{2+f}{2(1+f)}\right) d\mu}} \leq \int_{\Omega} \frac{f}{2(1+f)} d\mu \quad (23)$$

De la même façon, dans l'équation ci-dessus,  $\frac{f}{2(1+f)}$  peut être remplacée par  $\phi(f)$ ; où  $\phi$  est une fonction mesurable projetant  $\mathbb{R}_+$  sur l'intervalle  $(0, \frac{1}{2}]$ .

Ce dernier cas (cas continu) montre l'intérêt de cet opérateur et de l'inégalité sous-jacente pour la fusion de plages de données d'une série temporelle, quand bien même celle-ci serait le résultat d'une modélisation mathématique d'un phénomène expérimental.

## 5 Exemples d'applications

Dans cette partie nous présentons deux exemples d'applications possibles : une pour l'opérateur d'agrégation triple  $\Pi$  dans le domaine des bioprocédés (cas fini), l'autre pour sa généralisation (cas continu) dans le cas d'analyse de données issues de station d'énergie solaire. Par ces deux exemples, nous cherchons à illustrer les possibilités offertes par ces opérateurs, l'accent sera donc mis sur les propriétés utilisées plus que sur les résultats expérimentaux.

### 5.1 Cas fini : classification d'états physiologiques dans un bioprocédé

Les procédés actuels qu'ils soient industriels ou expérimentaux sont aujourd'hui dans leur grande majorité assistés ou pilotés par des outils logiciels de plus en plus sophistiqués. Ainsi dans le domaine des bioprocédés (utilisant le plus souvent des micro-organismes), les capteurs permettent de mesurer en temps réel un grand nombre de variables biochimiques qui représentent l'activité biochimique du système analysé. Dans le cas que nous présentons (Doncescu et al., 2007), il s'agit d'un bioprocédé dont l'objectif est la production de biomasse et la fermentation. Les variables biochimiques (toutes mesurées au même instant et à une fréquence régulière) sont considérées comme les sources d'informations. A partir de ces sources d'information, on cherche à déterminer les différents états physiologiques des micro-organismes. Les experts en microbiologie s'appuie sur des mesures biologiques complémentaires et réalisées *a posteriori* pour définir les différents états (voir figure 3).

Le bioprocédé de type batch que nous étudions est un bioprocédé de fermentation utilisant les micro-organismes appelés *Saccharomyces Cerevisiae*. L'expérience dure environ 20 heures et correspond à 1012 points de mesures pour chacun des 22 paramètres biochimiques analysés (le début de l'expérience est considéré comme étant égal à  $t=0h$ ). Trois états physiologiques principaux sont ici présents :

1. la fermentation (production d'éthanol) : elle va de 0h jusqu'à environ 9h ce qui représente un total de 590 points mesurés.
2. la diauxie : cet état commence à environ 9h et se termine à 9h46 ce qui représente environ 33 points de mesure. C'est le plus petit état physiologique parmi les 3 et le plus difficile à caractériser.

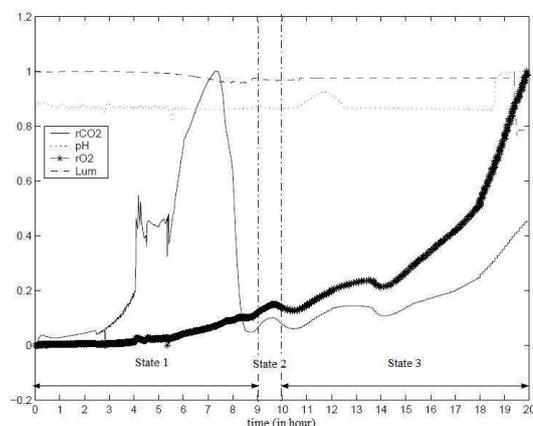


FIG. 3 – *Quelques-uns des paramètres biochimiques. 3 états physiologiques sont présents dans ce bioprocédé.*

3. l'oxydation (production de biomasse) : elle commence à 9h46 et se termine en même temps que la fin de l'expérience à 20h ce qui représente 389 points.

Le problème est que la classification fournie par les experts est faite après la fin du bioprocédé et qu'une analyse en temps réel et une caractérisation des différents états pendant ce procédé est nécessaire pour optimiser le système. Pour pallier cela chaque source est analysée par une méthode de clustering baptisée LAMDA (Piera-Carreté et Aguilar-Martin, 1991; Doncescu et al., 2007). LAMDA est une méthode de clustering et de classification basée sur l'utilisation des sous-ensembles flous qui a fourni des résultats intéressants dans le domaine des bioprocédés (Régis et al., 2003). Un état physiologique doit être représenté par une classe ou un ensemble de classes fournies par LAMDA. Pour chaque variable biochimique (c'est-à-dire chaque source d'information) et pour chaque classe, LAMDA fournit un degré d'appartenance pour chaque élément mesuré à un instant  $t$ . A chaque instant  $t$  et pour chaque classe, on peut donc fusionner des informations issues des différentes sources, et ce, via un opérateur d'agrégation. Cependant l'une des difficultés rencontrées survient au niveau des transitions entre les classes et par conséquent au niveau des transitions entre les états physiologiques. En effet, il s'agit de phénomène biologique et la délimitation des frontières entre différents états n'est pas chose aisée. Même pour des experts en microbiologie, qui pourtant restent la référence en matière de détection et de classification d'états, la délimitation des états physiologique pose parfois problème (voir figure 4).

Pour résoudre ce problème on se propose d'utiliser le triple  $\Pi$  en s'appuyant sur sa propriété de renforcement total. Ainsi le triple  $\Pi$  est l'opérateur d'agrégation choisi pour fusionner les informations issues de LAMDA et ce pour chaque classe et pour instant  $t$ . Pour diminuer les incertitudes au niveau des transitions entre états, on va rajouter des informations pertinentes sous forme de fonction par paliers. Ces fonctions par paliers changent de palier chaque fois qu'une singularité est détectée sur une variable biochimique (Nakkabi et al., 2002; Régis et al., 2002). En introduisant ces fonctions par paliers comme nouvelle source d'information en plus

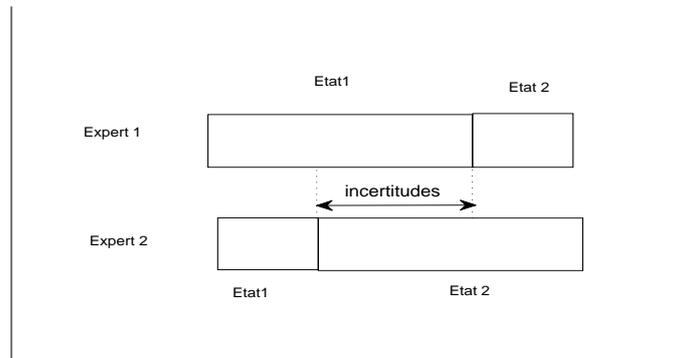


FIG. 4 – Deux experts en microbiologie peuvent trouver des transitions légèrement différentes créant ainsi une incertitude. La définition d'une frontière entre des états physiologiques reste donc une étape délicate.

des 22 variables biochimiques, et en s'appuyant sur les propriétés de renforcement total du triple II (en particulier sur le renforcement positif) on tient compte implicitement des transitions entre états (voir figure 5).

On pourrait se dire que l'utilisation d'un simple maximum comme opérateur d'agrégation donnerait les mêmes résultats mais ce n'est pas le cas. En effet, du fait que le triple II soit totalement renforcé, il permet de faire un tri entre les différents paliers. Seuls les paliers qui sont en adéquation avec le reste des variables biochimiques sont considérés comme pertinents, leurs degrés d'appartenance aux diverses classes influent donc le résultat uniquement dans le cas où ces paliers vont dans le même sens que les variables biochimiques. Si l'on utilise le maximum, ces paliers influenceront fortement le résultat, qu'ils soient considérés comme biologiquement pertinents ou non.

Ainsi, cet exemple permet de voir un des avantages de la fusion via un opérateur comme le triple II qui améliore la classification au niveau des transitions entre classes.

Nous donnons, par ailleurs, un exemple succinct d'utilisation de la moyenne  $M3II$  toujours sur les données issues de bioprocédé. De par sa propriété de consensus, la moyenne  $M3II$  ne peut améliorer les transitions entre les classes. Par contre il a été montré (Doncescu et al., 2007) qu'elle est plus robuste que le triple II. Ainsi si les données sont bruitées (voir figure 6), il est préférable d'utiliser la moyenne  $M3II$ ; en effet, la moyenne  $M3II$  continue à donner des résultats cohérents alors que le triple II ne fournit plus que des résultats abhérents (voir Doncescu et al. (2007)).

Ainsi l'utilisation du triple II et de la moyenne  $M3II$  peut fournir une meilleure prise en main des données multiples en fonction du type, de la qualité des données et des objectifs de

## Opérateurs d'agrégation et données multiples

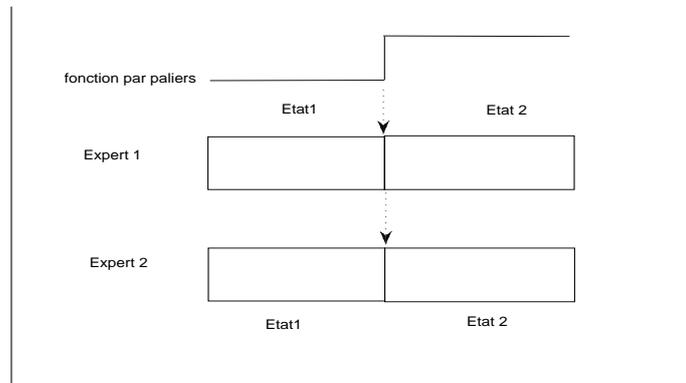


FIG. 5 – L'introduction de nouvelles informations, à savoir des fonctions par paliers et l'utilisation du triple  $\Pi$  permettent de mieux définir les transitions.

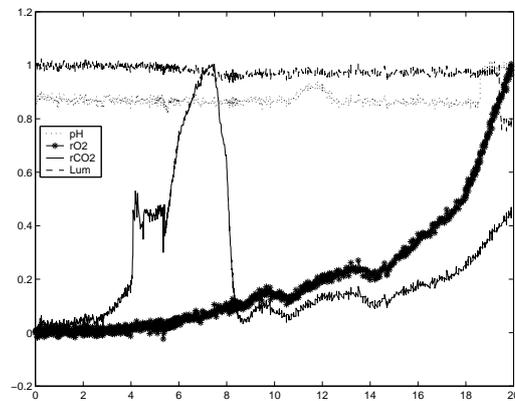


FIG. 6 – 4 des 22 variables biochimiques sont artificiellement bruités (Rapport Signal Bruit=40dB). Dans ce cas le triple  $\Pi$  fournit une fusion inexploitable tandis que la moyenne  $M3\Pi$  fournit encore des résultats cohérents (près de 12,5% de classification correcte, voir Doncescu et al. (2007)).

la fusion. Rappelons que la majorité des applications est faite pour la version finie c'est-à-dire le triple  $\Pi$  (Yager, 2002; Nagau et al., 2010; Faurie, 2011).

## 5.2 Cas continu

Voici un exemple de degré d'appartenance à temps continu.

Dans le domaine de l'énergie solaire, la radiation solaire dite extraterrestrielle ou hors atmosphère ou encore directe ( $W/m^2$ ) est la radiation maximale qui pourrait être reçue directement du soleil si le ciel était d'une clarté idéale sans nuages. Elle est connue de façon quasi déterministe en tout point précis du globe  $p$  et à tout instant  $t$ , nous la notons  $e_p(t)$ . La radiation globale mesurée, elle, est aléatoire, car elle dépend des conditions météorologiques et des objets environnants qui réfléchissent la radiation, nous la notons  $g_p(t)$ . Le quotient  $k_p(t) = \frac{g_p(t)}{e_p(t)}$ , appelé indice de clarté, est un nombre entre 0 et 1 que nous interprétons donc comme un degré d'appartenance instantané à une classe théorique de clarté.

Le graphique 7 représente une simulation de  $k_p(t)$  d'après un modèle d'EDS (Equation Différentielle Stochastique) en milieu aléatoire. Le milieu est modélisé par une chaîne de Markov cachée (Hidden Markov Model), c'est-à-dire dont on n'observe pas directement la dynamique, représentant les effets de l'environnement (Tran, 2013).

En comparant sur le graphique la moyenne mobile de triple  $\Pi$  à un opérateur classique d'agrégation temporelle comme par exemple la moyenne mobile KAMA (Kaufman Adaptive Moving Average) définie par Kaufman (Kaufman, 1995), une variante de AMA (Adaptive Moving average), on s'aperçoit que sur certains intervalles de temps (entre 0 et 5500 secondes, entre 8500 et 10000 secondes), la moyenne mobile de triple  $\Pi$  est plus proche que KAMA de la courbe de  $k_p$ . On peut en conclure, que triple  $\Pi$  améliore la prédiction sur certains intervalles, notamment quand le signal observé est bas. Une étude plus spécifique et plus approfondie serait nécessaire pour déterminer d'avance les instants et le choix de l'opérateur d'agrégation de moyenne afin d'avoir la meilleure prédiction. C'est une problématique qui nous semble nouvelle.

## 6 Perspectives : applications possibles

Nous avons présenté plus haut une généralisation de l'opérateur  $3\Pi$ , nommée  $G3\Pi$  et nous avons donné certaines propriétés et inégalités. L'exemple suivant est certainement la plus simple application pour le  $G3\Pi$ . Supposons que l'on ait une série temporelle représentant un phénomène quelconque, et supposons que l'on ait deux classes possibles pour chaque élément de la série : la classe numéro 1 représente l'état "danger" et la classe numéro 2 représente l'état "sans danger". On suppose que la figure 8 représente le degré d'appartenance pour la classe numéro 1 (sur un intervalle de temps donné) des éléments de la série temporelle. Comme les degrés d'appartenance à la classe 1 sont supérieurs à 0.5 sur l'intervalle de temps [26;50] ; il y a danger dans cet intervalle. L'utilisation du  $G3\Pi$  va renforcer le résultat "danger" comparativement à une moyenne classique (voir aussi les exemples des figures 1 et 2). Cet exemple permet d'effleurer certaines des potentialités du  $G3\Pi$ . En particulier, son utilisation dans le domaine de la biclassification est une voie à explorer.

Cependant, les applications potentielles pour cette généralisation restent à découvrir. Nous donnons deux options possibles pour l'utilisation de la généralisation de  $3\Pi$ , en fonction de

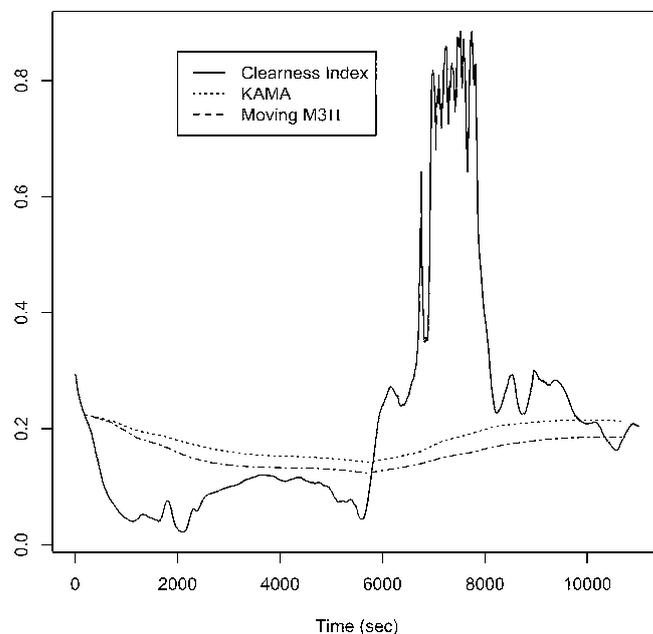


FIG. 7 – Courbe simulée de  $k_p$  suivant un modèle d'EDS en milieu aléatoire et moyennes mobiles KAMA et G3II

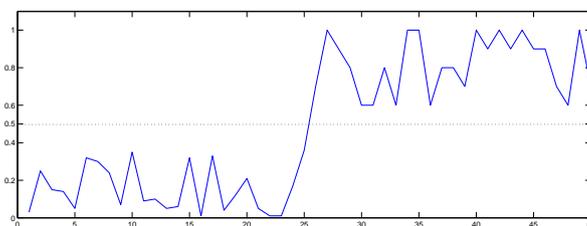


FIG. 8 – Courbe des degrés d'appartenance à la classe 1 (classe "danger") pour les éléments de la série temporelle sur un petit intervalle de temps [1 ; 50]. Les valeurs sont comprises entre 0 (non appartenance à la classe "danger") et 1 (appartenance à la classe "danger"). Sur l'intervalle [1 ; 25] les degrés d'appartenance ont des valeurs faibles (inférieures à 0.5). Cependant sur l'intervalle [26 ; 50], les degrés d'appartenance ont des valeurs supérieures à 0.5 montrant une appartenance forte à la classe "danger". Sur cet intervalle [26 ; 50], l'utilisation du G3II va renforcer l'appartenance à la classe "danger", comparativement à la moyenne géométrique par exemple.

l'objectif prioritaire que l'on a. Si l'analyse local des signaux est primordiale, alors l'approche suivante (illustrée par la figure 9) peut être appliquée :

1. pour chaque élément de temps mesuré et pour chaque variable, on calcule le  $3\Pi$  : le résultat est un signal (à partir de l'agrégation des différentes variables à un instant donné) défini pour chaque instant donné (étape 1).
2. la valeur de la généralisation du  $3\Pi$ , c'est-à-dire le  $G3\Pi$  est alors calculé à partir de ce signal "résultat" (étape 2).

Si l'analyse global d'un signal est primordiale alors la méthode suivante peut être appliquée (voir figure 10) :

1. Pour chaque variable, on calcule la généralisation, le  $G3\Pi$  : le résultat est une seule valeur par variable. (étape 1).
2. agréger toutes les valeurs avec le  $3\Pi$  (étape 2)

Le choix d'une approche parmi les deux applications dépend de l'application et de son cadre. Quelque soit la méthode choisie, les deux approches sont configurées pour agréger des séries temporelles. Evidemment, pour des applications expérimentales, le cas discret semble être le plus adapté; cependant le choix adéquate et l'utilisation d'une fonction  $g$  (voir équation 20) permettrait également l'utilisation du cas continu pour des applications expérimentales. Une fonction de pondération serait l'option la plus naturelle et semble être la perspective la plus évidente à creuser. Quoiqu'il en soit de nombreuses applications semblent possibles : analyse médicale de signaux physiologiques, analyse fonctionnelle, etc. Les perspectives d'applications du  $G3\Pi$  sont donc prometteuses, même s'il est vrai qu'à l'heure actuelle, les tests et les études en sont à leurs balbutiements. Comme on l'a déjà indiqué, des problématiques issues de la bi-classification ou du bi-clustering utilisant des séries temporelles pourraient donner des résultats probants via l'utilisation du  $G3\Pi$ . A moyen terme, une comparaison avec d'autres opérateurs d'agrégation (totalement renforcés ou non) permettra de mieux positionner le  $G3\Pi$ . Quoiqu'il en soit les perspectives semblent donc variées et intéressantes.

## 7 Conclusion

Nous avons proposé, à travers un exemple d'opérateur particulier (le triple  $\Pi$ ), d'illustrer quelques-unes des potentialités des opérateurs d'agrégation dans le domaine des données multiples. Nous avons donc présenté un opérateur d'agrégation totalement renforcé, le  $3\Pi$  ainsi que quelques-unes de ces variantes : le  $M3\Pi$  moyenne issue de cet opérateur, et la généralisation de l'opérateur  $3\Pi$  appelée  $G3\Pi$ . Le triple  $\Pi$  et la moyenne  $M3\Pi$  peuvent être utilisés sur des données multiples qu'il faut agréger ponctuellement. La généralisation  $G3\Pi$  permet notamment d'agrégation des informations à partir d'une série temporelle sur un intervalle temporel donné. Agréger des informations d'une série temporelle sur un intervalle donné, permettra d'avoir une vision plus globale plutôt que d'avoir une vision locale du problème. Cette généralisation permettra le management d'un plus grand volume de données. Une meilleure compréhension, et donc, une meilleure prise de décision pourrait résulter de l'utilisation de la généralisation  $G3\Pi$ . Quoiqu'il en soit l'utilisation d'opérateur d'agrégation, et en particulier le triple  $\Pi$  et ses variantes présentées dans cet article, peuvent permettre de bien synthétiser l'information issue de données multiples et faciliter la compréhension et l'interprétation de données de plus en

Opérateurs d'agrégation et données multiples

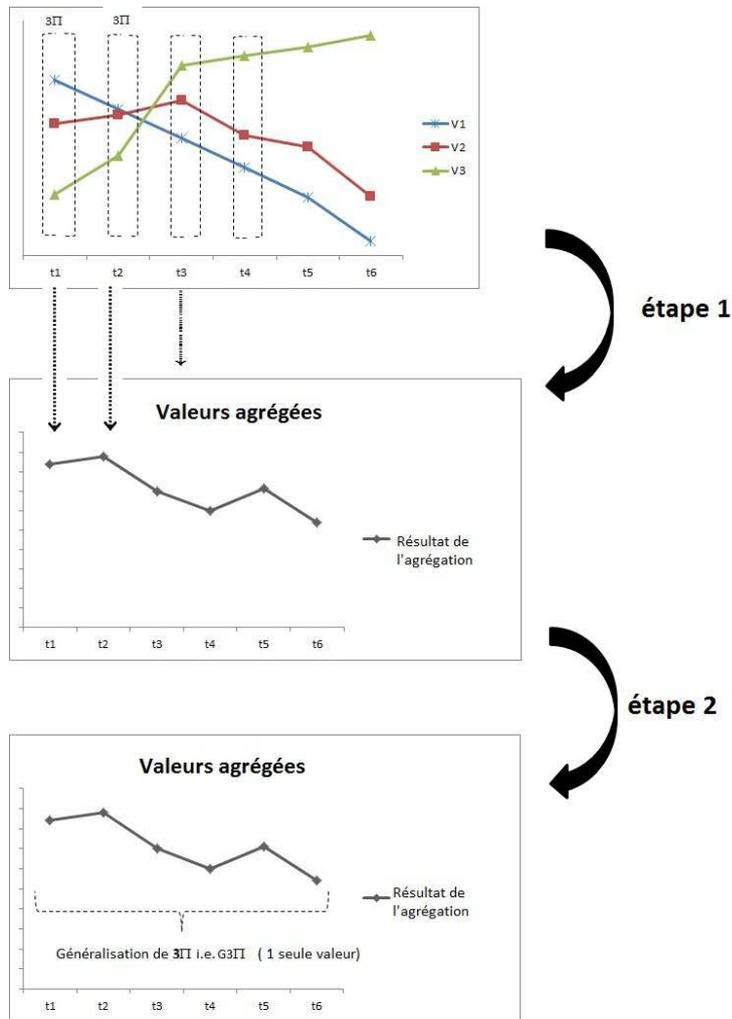


FIG. 9 – Cette approche favorise l'analyse locale : d'abord on applique le 3II, puis la généralisation du 3II, le G3II.

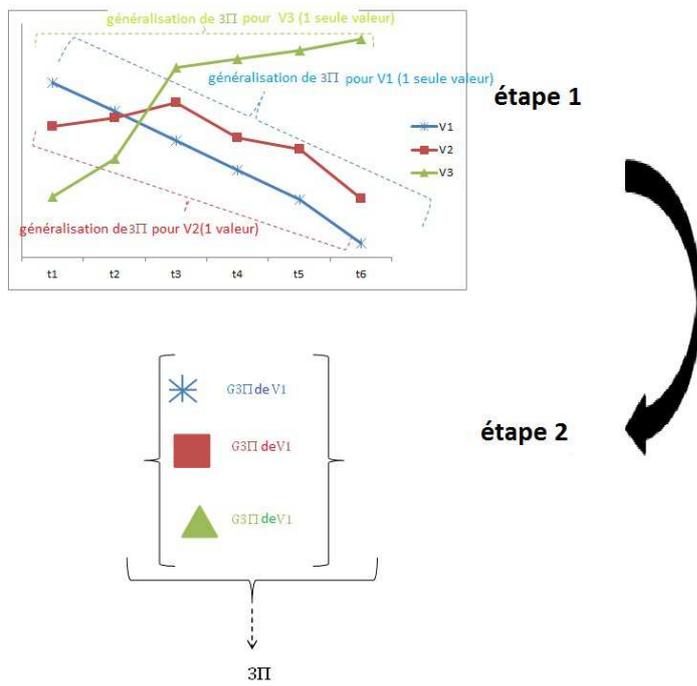


FIG. 10 – Cette approche favorise l'analyse globale : premièrement on applique le G3II puis le 3II.

plus complexes et volumineuses. Cet article ne présente que partiellement les potentialités du  $3\Pi$  et du  $G3\Pi$ , et de nombreuses applications sont encore à trouver pour mettre en évidence les atouts de ces opérateurs d'agrégation.

## Remerciements

Les auteurs tiennent à remercier les équipes de recherche du Laboratoire de Biotechnologie-Bioprocédés de l'INSA de Toulouse pour leur aide et leurs conseils.

## Références

- Beliakov, G., T. Calvo, et A. Pradera (2007). Handling of neutral information by aggregation information. *Fuzzy sets and systems* 158, 861–880.
- Bloch, I. et A. Hunter (Eds 2001). Fusion : General concepts and characteristics. *International Journal of Intelligent Systems* 16, 1107–1134.
- Calvo, T., B. D. Baets, et J. Fodor (2001). The functional equations of franck and alsina for uninorms and nullnorms. *Fuzzy Sets and Systems* 120, 385–394.
- Campanella, G. et R. Ribeiro (2011). A framework for dynamic multiple-criteria decision making. *Decision Support Systems* 52, 52–60.
- Detyniecki, M. (2000). *Mathematical Aggregation Operators and their Application to Video Querying*. Phd Thesis, Pierre et Marie Curie University (Paris VI).
- Doncescu, A., S. Régis, K. Inoue, et R. Emilion (2007). Analysis of new aggregation operators : Mean  $3\pi$ . *J. of Advanced Computational Intelligence and Intelligent Informatics* 11(6), 561–569.
- Dubois, D. et H. Prade (2004). On the use of aggregation operations in information fusion process. *Fuzzy Sets and Systems* 142, 143–161.
- Elkan, C. (1994). The paradoxical success of fuzzy logic. *IEEE Expert*, 3–8.
- Emilion, R., S. Régis, et A. Doncescu (2013a). A general version of the triple  $\pi$  operator. *International Journal of Intelligent Systems*. accepted, to appear.
- Emilion, R., S. Régis, et A. Doncescu (2013b). Généralisation d'un opérateur d'agrégation totalement renforcé pour des séries temporelles. In *EGC-AFDC 2013*, Toulouse, France. Actes de l'Atelier de Fouille de données complexes : complexité liée aux données multiples et massives.
- Emilion, R., S. Régis, A. MÉRIL, et A. Doncescu (2004). Un nouvel opérateur de type moyenne pour la fusion de données. In *Logique Floue et Application (LFA)*, Nantes.
- Faurie, F. (2011). *Algorithmes de contrôle d'intégrité pour la navigation hybride GNSS et systèmes de navigation inertielle en présence de multiples mesures satellitaires défailantes*. Thèse de Doctorat (Phd Thesis), Université de Bordeaux.
- Jilani, T. A. et S. M. A. Burney (2008). Multivariate stochastic fuzzy forecasting models. *Expert Systems with Applications* 35, 691–700.
- Kaufman, P. (1995). *Smarter Trading*. McGraw-Hill.

- Komornik, J., M. Komornikova, R. Mesiar, D. Szokeova, et J. Szolgay (2006). Comparison of forecasting performance of nonlinear models of hydrological time series. *Physics and Chemistry of the Earth* 31, 1127–1145.
- Martin, A. (2005). La fusion d'informations. Polycopié de cours ENSIETA 1484.
- Nagau, J., S. Régis, et J.-L. Henry (2010). Etude d'opérateurs d'agrégation pour l'ordonnement de clusters dans des images numériques de plantes. In *Atelier "Fouille de Données Complexes" (EGC'2010)*, Hammamet, Tunisia.
- Nakkabi, Y., S. Régis, J. Desachy, A. Doncescu, et G. Roux (2002). Apport de la transformée en ondelettes pour affiner les résultats de classifications. In *IXe Rencontre de la Société Francophone de Classification*, Toulouse, France, pp. 287–291.
- Nelsen, R. (1999). *An introduction to Copulas*, Volume 139 of *Lecture Notes in Statistics*. New York : Springer.
- Piera-Carreté, N. et J. Aguilar-Martin (1991). Controlling selectivity in non-standard pattern recognition algorithm. *Trans. in Syst. Man, Cybernetics* 21, 71–82.
- Piera-Carreté, N., J. Aguilar-Martin, et M. Sanchez (1988). Mixed connectives between min and max. In *8th Inter. Symp. on Multiple Valued Logic*.
- Pradera, A. et E. Trillas (2006). Aggregation, non-contradiction and excluded-middle. *Mathware and Soft Computing* 13, 189–201.
- Régis, S., A. Doncescu, J. Aguilar-Martin, et G. Goma (2002). Classification of bioprocesses using wavelets coefficient. In *International Conference on Trends in Monitoring and Control of Life Science Applications*, Lynby, Denmark.
- Régis, S., A. Doncescu, J. Desachy, et J. Aguilar-Marti (2003). Comparaison de classifications non supervisées de données biotechnologiques. In *10ième Rencontre de la SFC*, Neuchâtel, Suisse.
- Silvert, W. (1979). Symmetric summation : A class of operations on fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics* 9(10), 657–659.
- Tran, V. L. (2013). *Modèles stochastiques pour des processus de radiation solaire*. Thèse de Doctorat (Phd Thesis), Université d'Orléans.
- Waissman-Vilanova, J. (2000). *Construction d'un modèle comportemental pour la supervision de procédés : application à une station de traitement des eaux*. Thèse de Doctorat (Phd Thesis), LAAS - CNRS.
- Weber, S. (1983). A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy sets and systems* 11, 103–113.
- Yager, R. (1980). On a general class of fuzzy connectives. *Fuzzy sets and Systems* 4, 235–242.
- Yager, R. (1988). On ordered weighted averaging operators in multi-criteria decision making. *IEEE Transaction on Systems, Man, and Cybernetics*, 183–190.
- Yager, R. (1994). Misrepresentations and challenges : A response to elkan. *IEEE Expert*, 41–42.
- Yager, R. (2002). Defending against strategic manipulation in uninorm-based multi-agent decision making. *European Journal of Operational Research* 141, 217–232.
- Yager, R. (2004). Owa aggregation over a continuous interval argument with applications to

## Opérateurs d'agrégation et données multiples

- decision making. *IEEE Transactions on Systems, Man and Cybernetics-Part B : Cybernetics* 34(5), 1952–1963.
- Yager, R. (2008). Time series smoothing and owa aggregation. *IEEE Transactions on Fuzzy Systems* 16(4), 994–1007.
- Yager, R. et A. Rybalov (1996). Uninorm aggregation operators. *Fuzzy Sets and Systems* 80, 111–120.
- Yager, R. et A. Rybalov (1998). Full reinforcement operators in aggregation techniques. *IEEE Transactions on Systems, Man and Cybernetics-Part B : Cybernetics* 28(6), 757–769.
- Zimmermann, H. (2001). *Fuzzy set theory - and its applications* (Springer ed.).
- Zimmermann, H. et P. Zynso (1980). Latent connectives in human decision making. *Fuzzy Sets and Systems* 4, 37–51.

## Summary

This article deals with the aggregation operators from fuzzy logic. Particularly, an example of aggregation operator the 3II, and others operators derived from it, are presented. Triple II presents some interested potentialities for the multiple data fusion. Some applications are also presented. More generally, through this instance of 3II, the aim is to show the possibilities provided by the aggregation operator in the field of multiple and massive data mining.