

## CHAPITRE 4

### STRATÉGIES D'APPRENTISSAGE SPÉCIALISÉES BASÉES SUR L'OPTIMISATION DES PARAMÈTRES DU RÉSEAU FUZZY ARTMAP ÉVALUÉES SUR LES BASES SYNTHÉTIQUES

Lors de notre première réflexion sur la performance des réseaux fuzzy ARTMAP, nous avons identifié plusieurs caractéristiques qui, selon nous, pourraient entraîner la dégradation des performances. Ces caractéristiques sont :

- a. le nombre d'époques d'entraînement;
- b. la taille de la base d'entraînement;
- c. la technique de normalisation;
- d. la structure (dispersion et chevauchement) des données générées;
- e. l'ordre de présentation des patrons d'entraînement;
- f. le type de MatchTracking.

L'analyse des résultats présentés au chapitre 3 a démontré que le nombre d'époques d'entraînement et la taille de la base de données d'apprentissage pouvaient engendrer une dégradation des performances due au phénomène de sur-apprentissage pour les bases avec chevauchement ( $DB_{\mu}$  et  $DB_{\sigma}$ ). Par contre, la technique de normalisation ainsi que la structure des bases de données n'ont aucune incidence sur la dégradation des performances dans les réseaux fuzzy ARTMAP. L'impact de l'ordre de présentation des patrons d'entraînement n'est pas directement abordé, car ce sujet est très vaste. Par contre, l'optimisation de la taille de la base d'entraînement semble atténuer la dégradation des performances causée par cette caractéristique.

La polarité du MatchTracking utilisée fait varier l'erreur de sur-apprentissage de manière significative pour les bases de données avec chevauchement. En analysant ces résultats, il faut s'interroger sur l'effet combiné des quatre paramètres internes des réseaux fuzzy ARTMAP, soit: le paramètre de choix ( $\alpha$ ), de vitesse d'apprentissage ( $\beta$ ), de MatchTracking ( $\epsilon$ ) et de vigilance de base ( $\bar{\rho}$ ). En effet, si l'erreur en généralisation

peut varier en changeant la polarité d'épsilon (MT- vs MT+), il doit exister une combinaison de paramètres permettant d'optimiser la performance en généralisation. Dans le même ordre d'idées, l'erreur nette restante après l'optimisation de la taille de la base d'apprentissage (voir figure 38) peut-elle être réduite avec une meilleure sélection des paramètres? La dégradation des performances due à la taille de la base d'apprentissage et au nombre d'époques d'entraînement sont-elles influencées par les valeurs des paramètres utilisées? Pour répondre à ces questions et ainsi analyser l'impact des valeurs des paramètres internes sur la performance en généralisation du FAM, nous utilisons les stratégies d'apprentissage spécialisées pour FAM, avec l'algorithme PSO, que nous avons développées.

Ces stratégies permettent de trouver des valeurs pour chacun des paramètres dans le but d'optimiser les performances en généralisation des réseaux fuzzy ARTMAP, et ce, peu importe la base de données utilisée. Ainsi, ce chapitre analyse les effets des paramètres internes du FAM lorsque ceux-ci sont optimisés avec l'algorithme PSO par nos stratégies d'apprentissage spécialisées, pour les bases de données synthétiques. Les résultats obtenus dans ce chapitre ont contribué à la publication d'un article [35].

À noter que les temps de convergence incluent toutes les époques effectuées par toutes les particules lors de l'optimisation PSO. Le temps de convergence varie entre 660 époques (15 particules PSO, 11 itérations PSO, 4 répétitions PSO, et 1 époque d'entraînement FAM) et  $6 \times 10^6$  époques (15 particules PSO, 100 itérations PSO, 4 répétitions PSO, et 1000 époques d'entraînement FAM).

#### **4.1 Bases de données avec chevauchement**

Cette section traite des effets de l'optimisation des paramètres internes du réseau fuzzy ARTMAP avec les bases de données  $DB_{\mu}$  et  $DB_{\sigma}$ . Étant donné qu'aucune différence n'a

été remarquée entre les deux techniques de normalisation (voir section 3.3), la normalisation MinMax est utilisée.

#### 4.1.1 Résultats

Les figures suivantes présentent les erreurs en généralisation, les temps de convergence ainsi que les taux de compression obtenus pour les quatre stratégies d'apprentissage et les bases de données  $DB_{\mu}(1\%)$ ,  $DB_{\mu}(9\%)$ ,  $DB_{\mu}(25\%)$  et  $DB_{\sigma}(9\%)$ . La stratégie d'apprentissage  $HV(MT+)$ , le classificateur quadratique Bayésien (CQB) ainsi que le  $k$ NN sont également présentés comme référence. De plus, les valeurs optimisées des quatre paramètres internes des réseaux fuzzy ARTMAP sont également analysées.

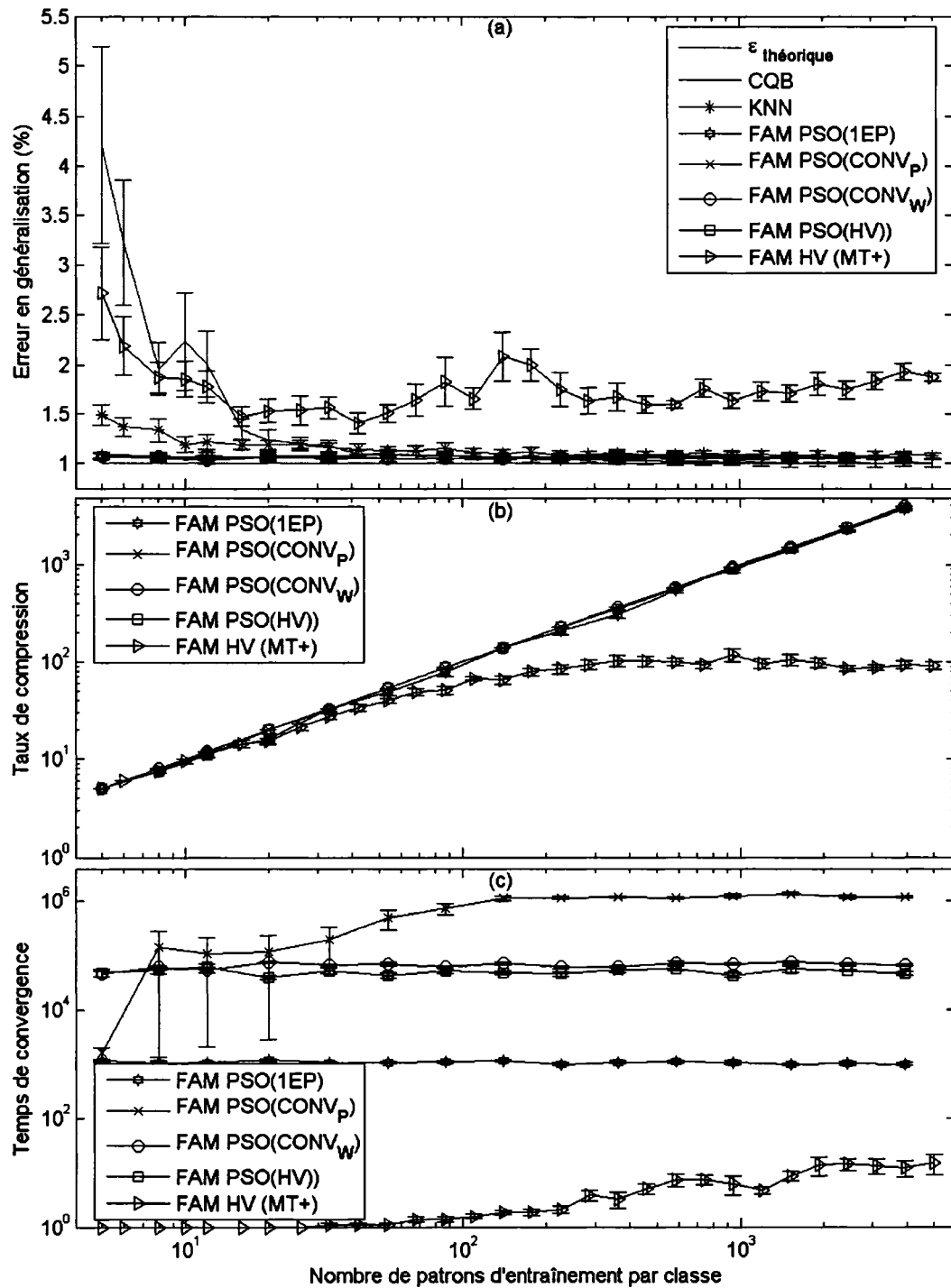


Figure 40 Performance du FAM avec les stratégies PSO sur la base  $DB_{\mu}(1\%)$   
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression

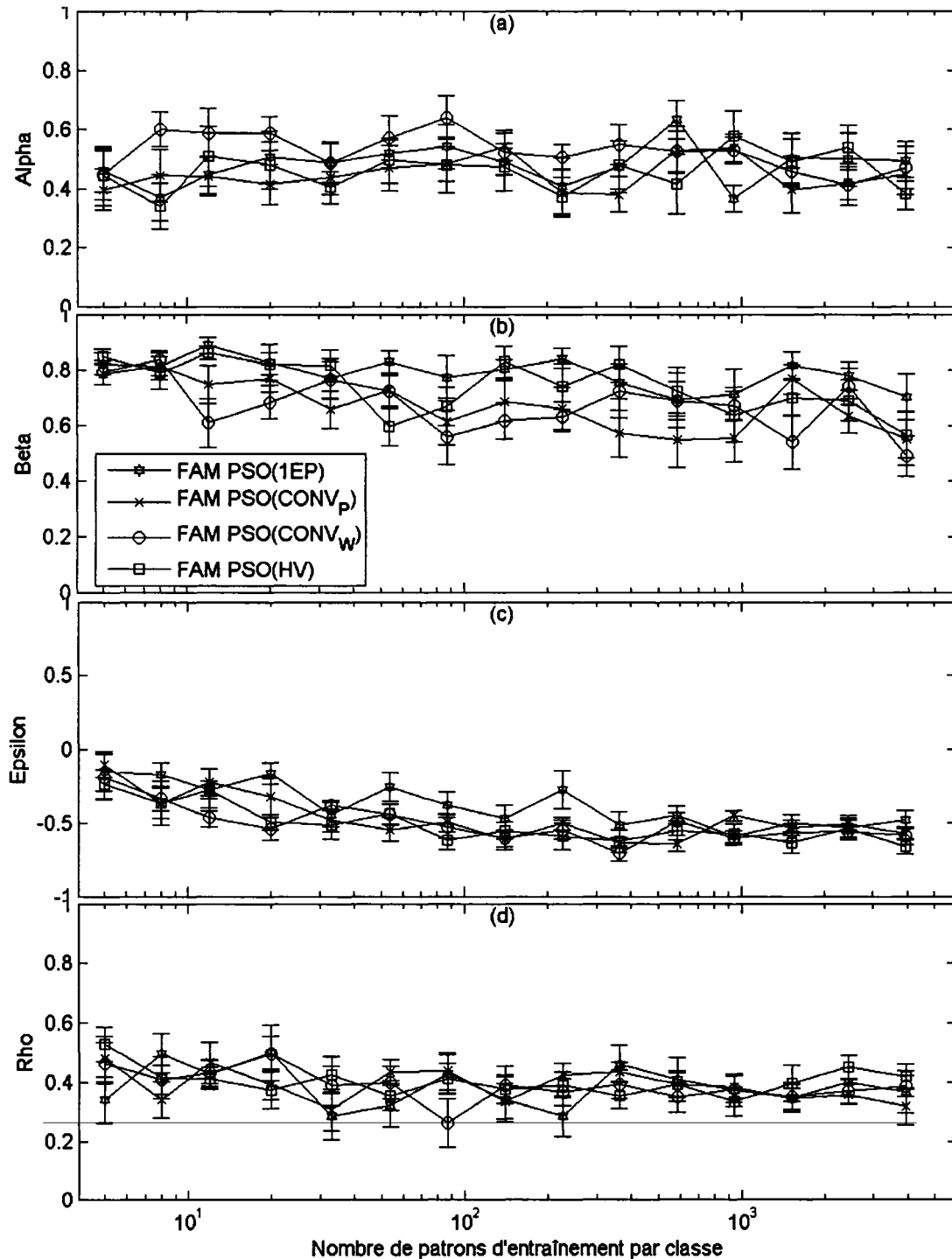


Figure 41 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base  $DB_{\mu}(1\%)$

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base

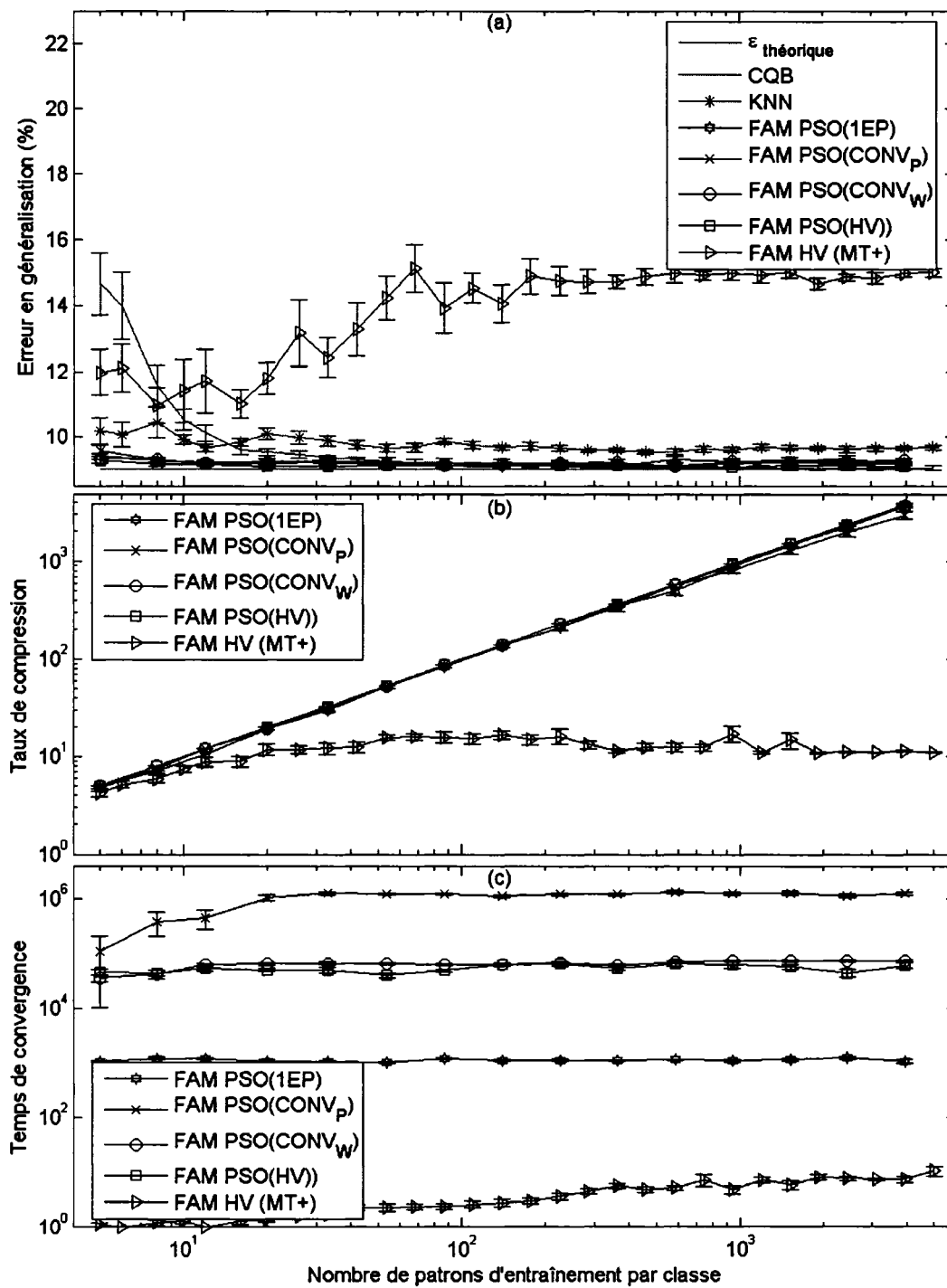


Figure 42 Performance du FAM avec les stratégies PSO sur la base  $DB_{\mu}(9\%)$   
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression

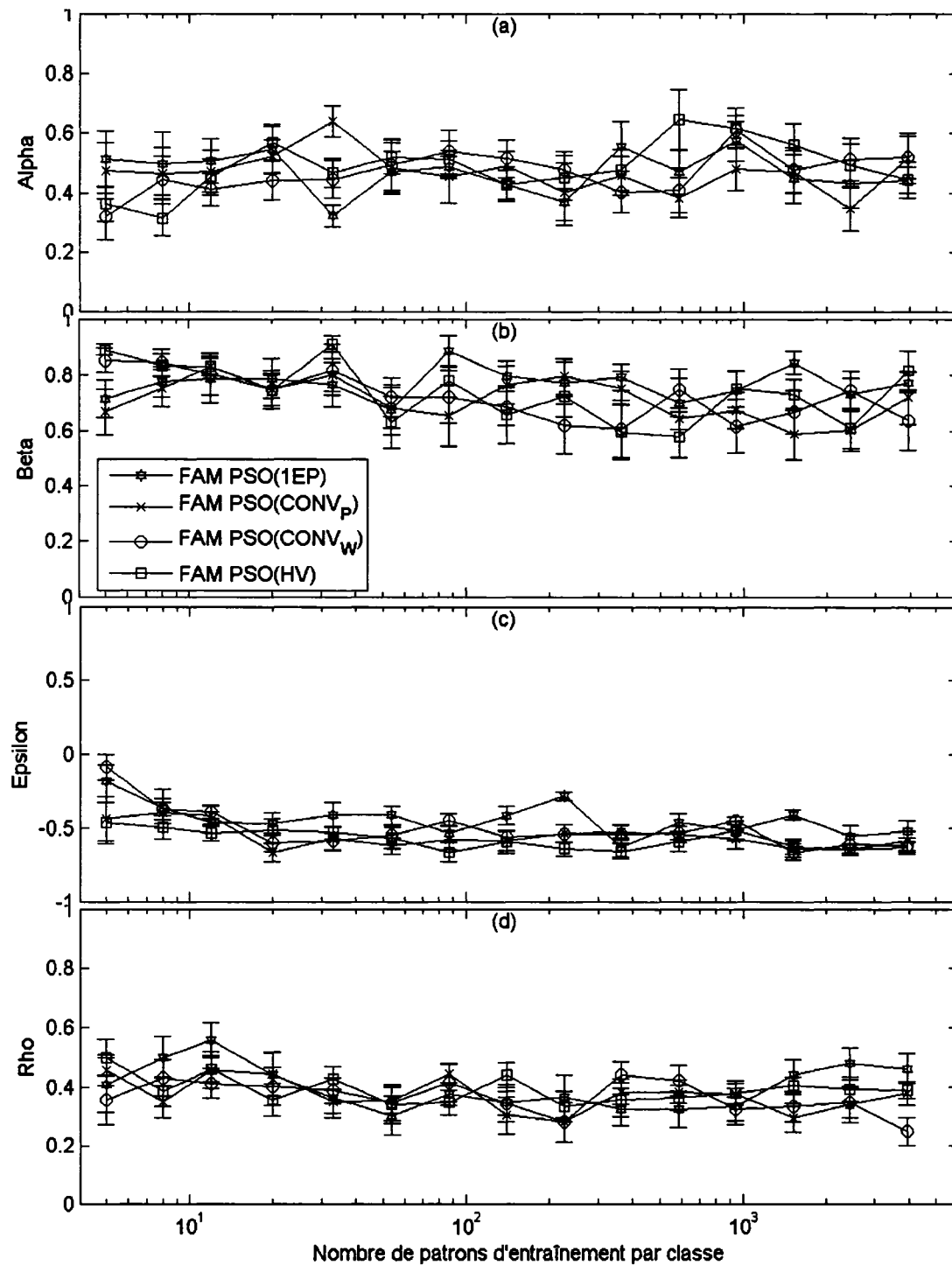


Figure 43 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base  $DB_{\mu}(9\%)$

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base

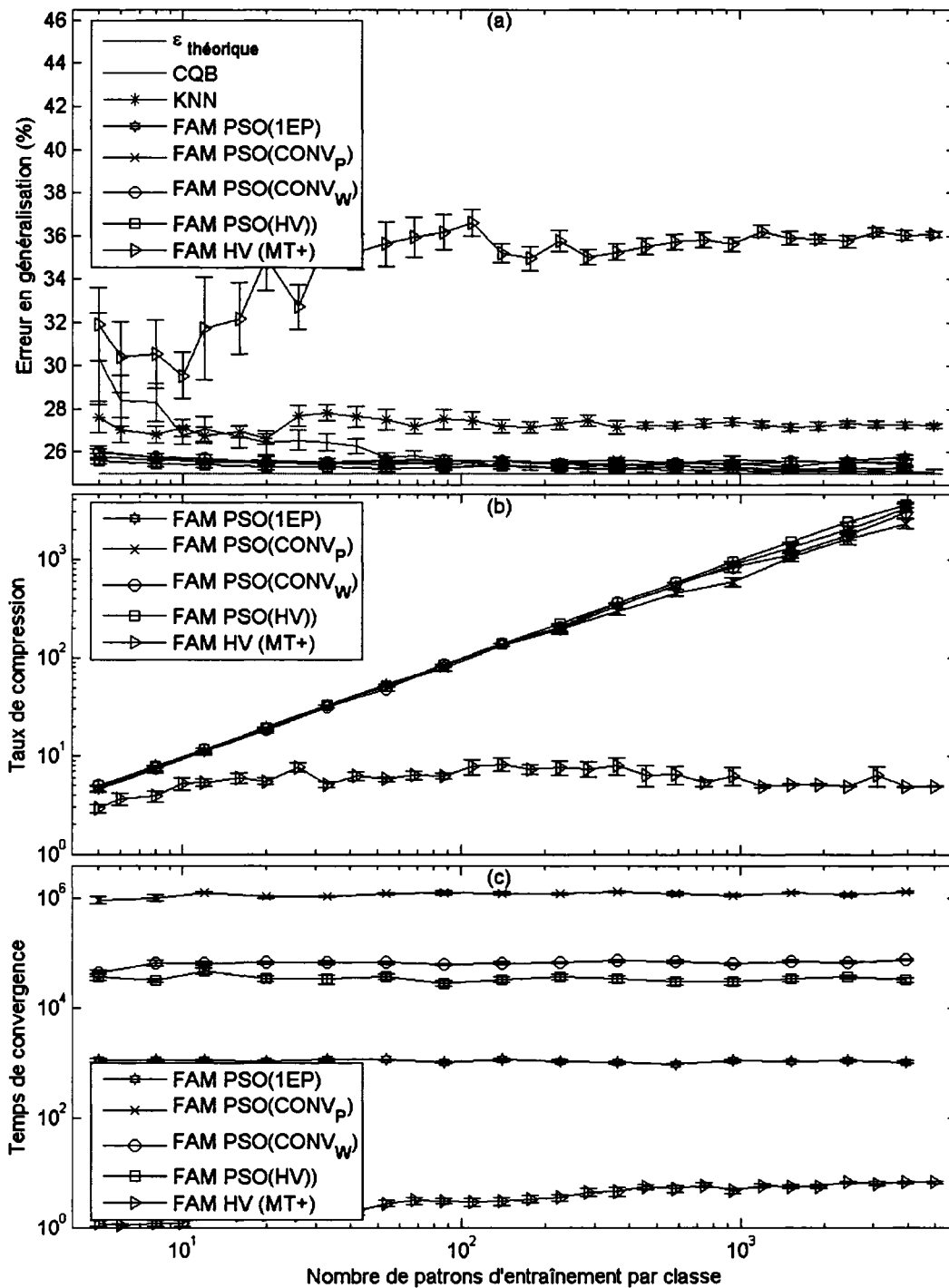


Figure 44 Performance du FAM avec les stratégies PSO sur la base  $DB_{\mu}(25\%)$   
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression



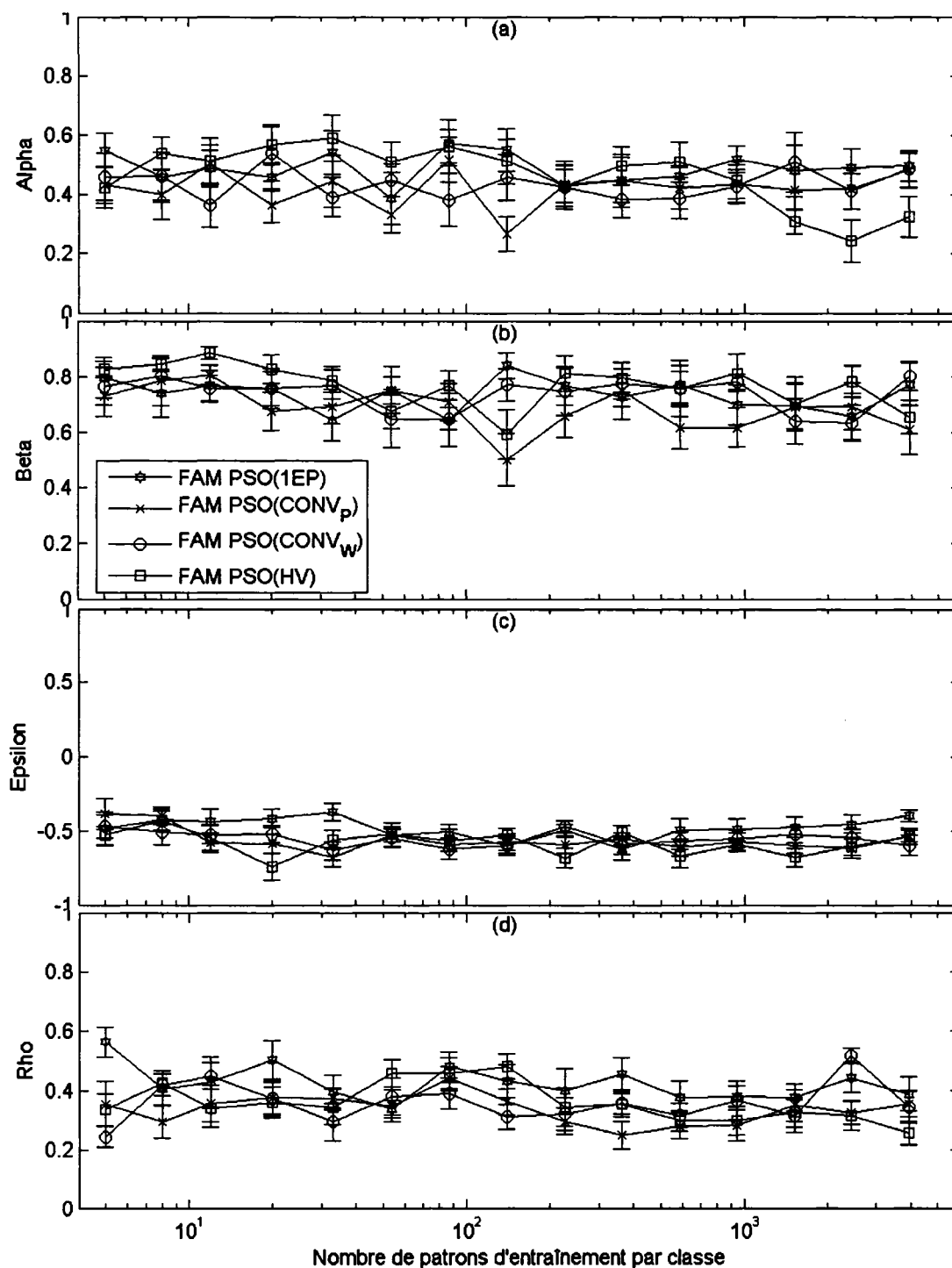


Figure 45 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base  $DB_{\mu}(25\%)$

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base

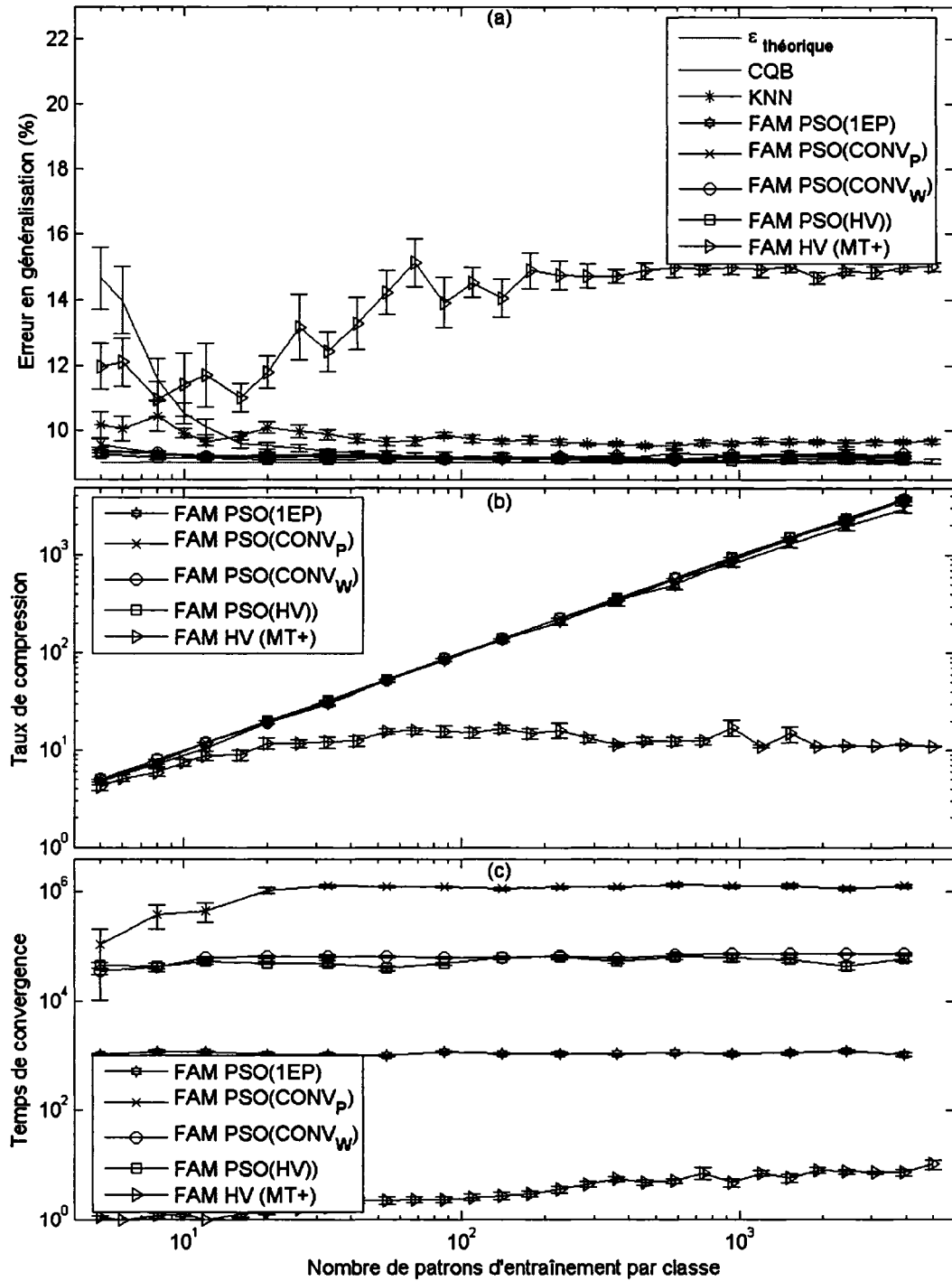


Figure 46 Performance du FAM avec les stratégies PSO sur la base DB $\sigma$ (9%)  
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression

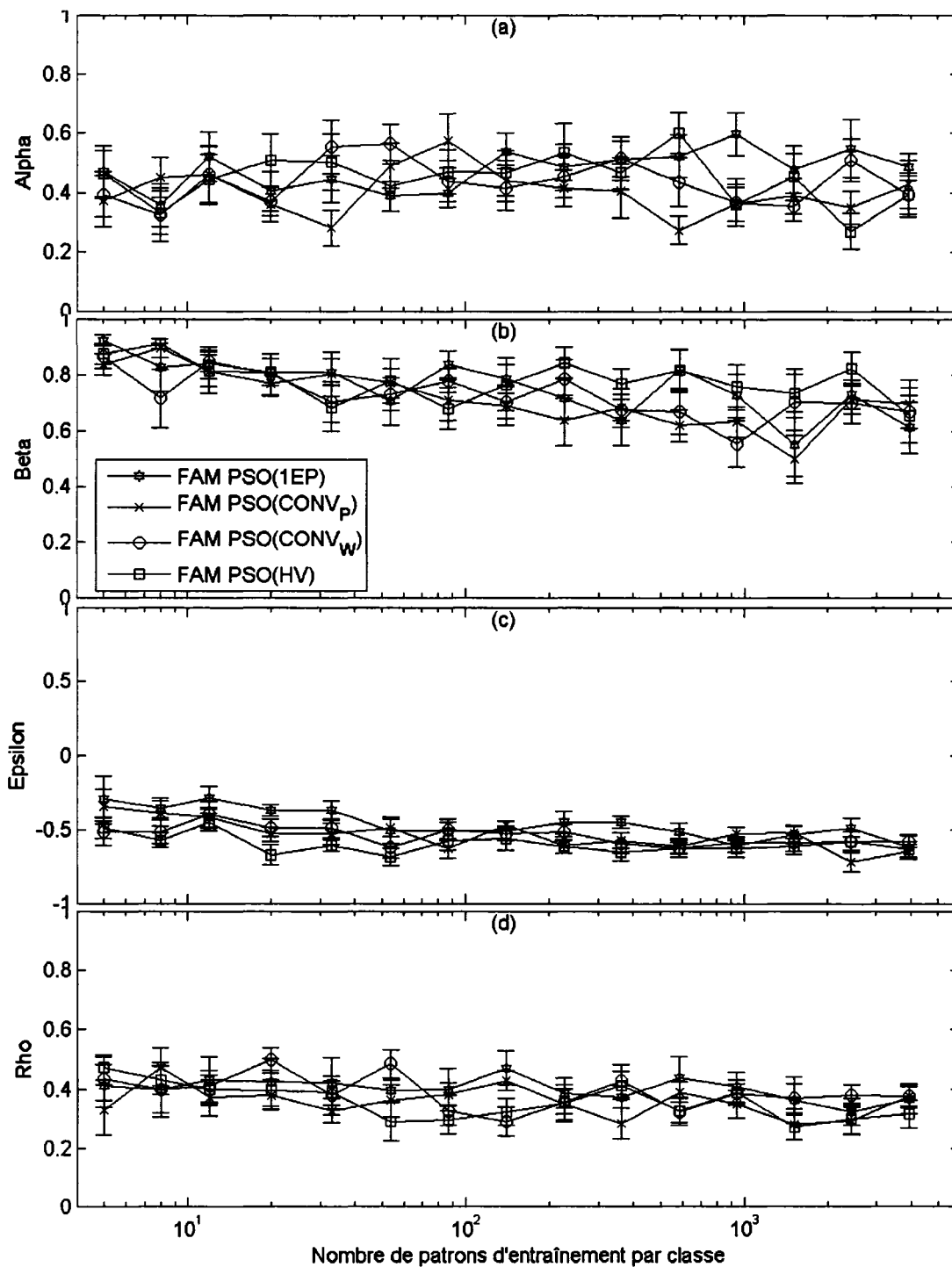


Figure 47 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base  $DB_{\sigma}(9\%)$

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base

### 4.1.2 Analyse

Tel que le montrent les figures 39 à 46, il est clair que l'optimisation des paramètres internes des réseaux fuzzy ARTMAP est très avantageux au point de vue des erreurs en généralisation et des taux de compression.

L'optimisation des paramètres permet d'obtenir des erreurs en généralisation très proches des limites théoriques et ce, pour toutes les tailles des bases d'apprentissage ainsi que pour toutes les stratégies d'apprentissage testées avec les bases de données possédant du chevauchement. Ainsi, même avec la plus petite taille de base d'apprentissage, les réseaux créés à l'aide de nos stratégies d'apprentissage spécialisées obtiennent des performances proches de l'erreur théorique et ce, pour tous les degrés de chevauchement testés. De plus, les erreurs en généralisation réalisées par les réseaux fuzzy ARTMAP sont généralement plus petites que celles obtenues avec le classificateur  $k$ NN et sont comparables à celles obtenues avec CQB.

Au chapitre 3, lors de l'optimisation de la taille de la base d'apprentissage sur un des tests de la base  $DB_{\mu}(9\%)$ , nous avons réussi à diminuer l'erreur en généralisation à 9.87% (MT+ et HV) alors qu'elle était de 14.85% avec la taille maximale de la base d'entraînement. En optimisant les paramètres avec cette même base de données nous obtenons une erreur en généralisation de 9.07% avec la taille maximale et lors de l'optimisation de la taille de la base d'entraînement nous obtenons une erreur de 8.99% avec 1519 patrons par classe<sup>2</sup>. Afin de mieux comprendre l'impact de l'optimisation des paramètres, la Figure 48 présente les bornes de décision de ces tests pour  $DB_{\mu}(9\%)$ , avec

---

<sup>2</sup>Certaines erreurs en généralisation obtenues sont juste inférieures à l'erreur théorique. Ce phénomène est provoqué par le manque de patrons dans les bases de données. Ainsi, le nombre de patrons contenus dans la base de test n'est pas suffisant pour représenter parfaitement la distribution calculée par les coefficients du Tableau III et du Tableau IV. Les erreurs réelles de ces bases de données sont donc très proches des erreurs théoriques, mais elles ne sont pas identiques.

la taille de la base d'entraînement optimisée et maximale, lors de l'optimisation des paramètres ainsi que lors de l'utilisation des paramètres standard (MT+).

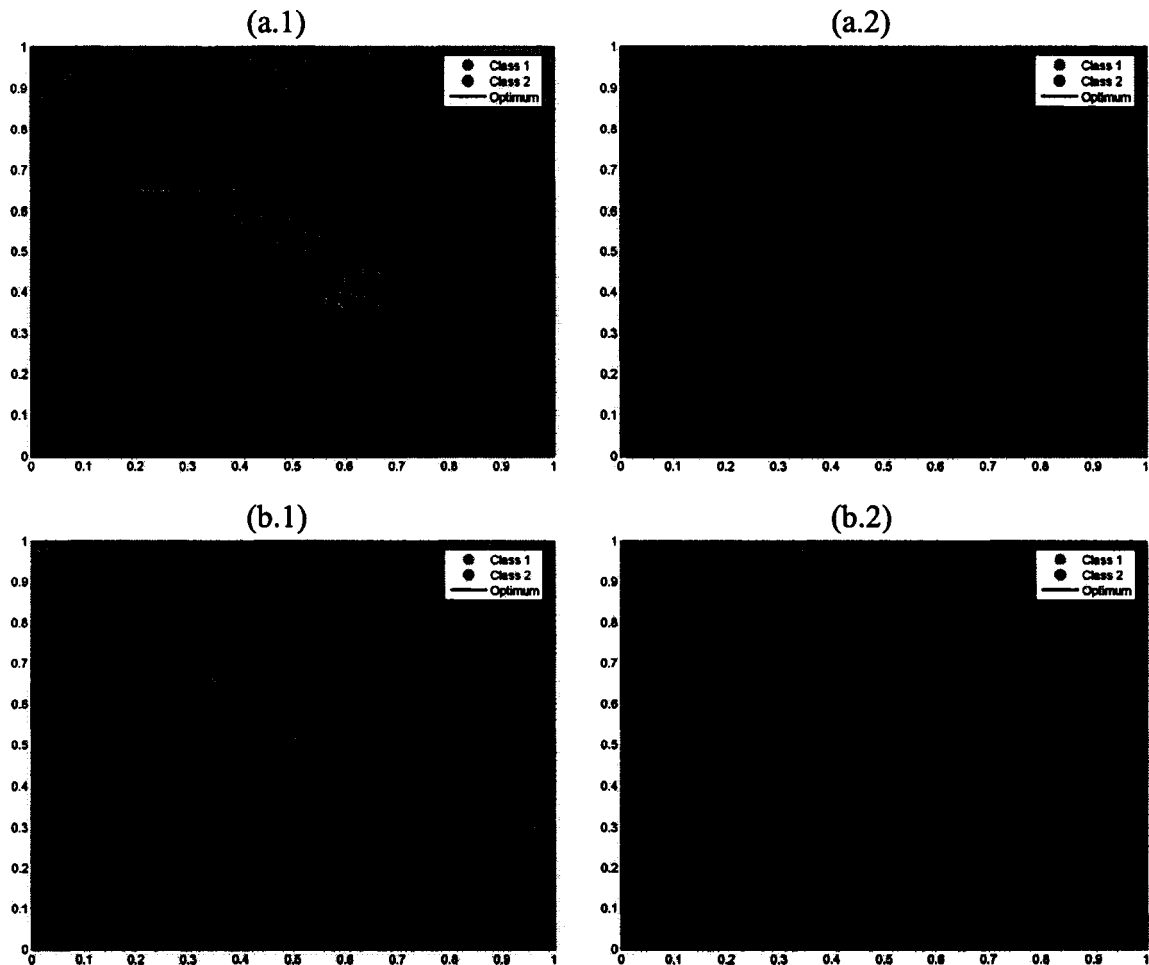


Figure 48 Bornes obtenues avec et sans optimisation des paramètres ainsi qu'avec et sans optimisation de la taille de la base d'apprentissage pour  $DB_{\mu}(9\%)$

Soit (a) sans optimisation des paramètres et (b) avec optimisation des paramètres, (1) avec la taille maximale de la base d'entraînement (5000 patrons par classe) et (2) avec la taille de la base d'apprentissage optimisée (a.2 = 178 patrons par classe et b.2 = 1519 patrons par classe); le tout avec la stratégie d'apprentissage HV.

Le tableau X présente le sommaire des résultats des quatre réseaux fuzzy ARTMAP présentés à la figure 48.

Tableau X

Résultats avec et sans optimisation des paramètres avec HV pour une série de  $DB_{\mu}(9\%)$

<b>HV</b>	<b>Nombre de patrons d'entraînement par classe</b>	<b>Erreur en généralisation</b>	<b>Nombre de catégories</b>	<b>Taux de compression</b>
<b>Sans PSO</b>	5000	14.85%	916	5.46
<b>Sans PSO</b>	178	9.87%	4	44.5
<b>Avec PSO</b>	5000	9.07%	2	2500
<b>Avec PSO</b>	1519	8.99%	2	759.5

Les réseaux obtenus avec l'optimisation des paramètres possèdent, en moyenne, une catégorie par classe. En sachant que les classes des bases de données  $DB_{\mu}$  et  $DB_{\sigma}$  sont des distributions gaussiennes et que les performances en généralisation obtenues sont très proches des erreurs théoriques, les catégories créées par ces réseaux représentent donc les centres de masse des distributions gaussiennes.

On remarque également que les quatre stratégies d'apprentissage offrent des performances équivalentes au niveau des erreurs en généralisation ainsi qu'au niveau des taux de compression. Par contre, le temps de convergence de la stratégie PSO(1EP) est inférieur aux trois autres stratégies d'apprentissage spécialisées. Ainsi, bien que l'optimisation des paramètres requière toujours des temps de convergence beaucoup plus élevés que sans optimisation, nous sommes en mesure de réduire ce coût en utilisant la stratégie d'apprentissage PSO(1EP) sans dégrader les performances en généralisation ainsi que les taux de compression.

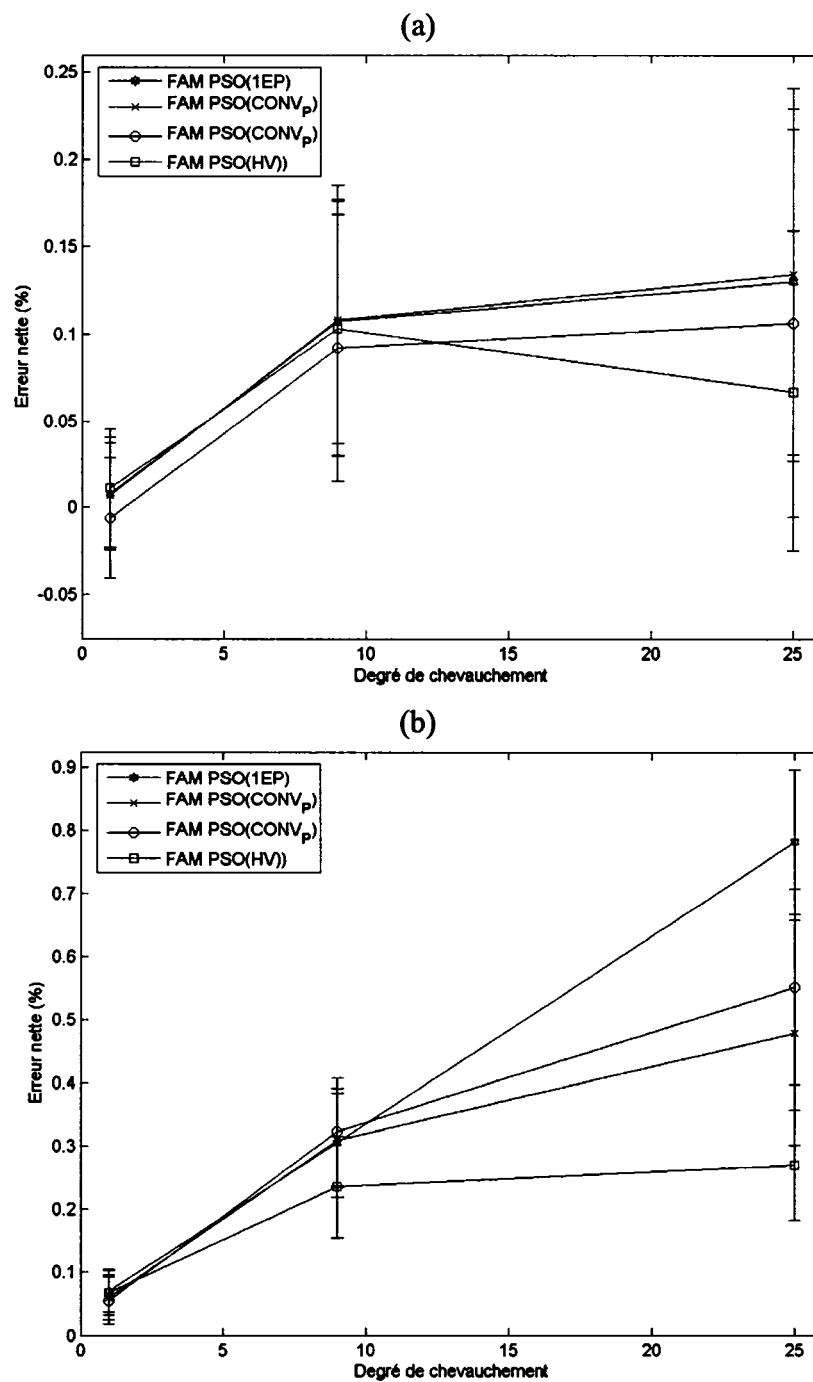


Figure 49 Erreur nette avec PSO en fonction du chevauchement avec  $DB_{\mu}$   
 (a) avec la taille de la base d'apprentissage optimisée et (b) avec la taille maximale de la base d'apprentissage

Pour appuyer cette analyse, la figure 49 présente l'erreur nette obtenue lors de l'utilisation de la taille maximale de la base d'apprentissage et lors de l'optimisation du nombre de patrons d'entraînement. Tel qu'on peut le voir, avec la taille maximale de la base d'entraînement (figure 49(b)) la stratégie d'apprentissage PSO(1EP) présente des erreurs en généralisation supérieures aux autres stratégies avec 25% de chevauchement. Malgré ce fait, une fois l'optimisation du nombre de patrons d'entraînement effectuée (Figure 49(a)), toutes les stratégies obtiennent des erreurs en généralisation similaires.

Pour montrer que la stratégie PSO(1EP) obtient des performances en généralisation similaires à PSO(HV), les figures 49, 50 et 51 présentent les différences des résultats obtenus pour ces deux stratégies, et ce, pour l'erreur en généralisation, le taux de compression, le temps de convergence et le nombre de catégories créées avec, respectivement, les bases  $DB_{\mu}(1\%)$ ,  $DB_{\mu}(9\%)$  et  $DB_{\mu}(25\%)$ . Ainsi, lorsque la courbe est positive, la valeur obtenue avec la stratégie PSO(HV) est plus grande que celle obtenue avec la stratégie PSO(1EP), et vice-versa.

On remarque avec ces deux figures que les erreurs en généralisation avec la stratégie PSO(1EP) sont similaires de celles obtenues avec PSO(HV). Les taux de compression obtenus avec la stratégie PSO(1EP) sont similaires ou meilleurs que ceux obtenus avec PSO(HV). De ce fait, le nombre de catégories créées par PSO(1EP) est généralement égal ou plus petit que celui obtenu avec PSO(HV). Le temps de convergence est nettement plus petit avec PSO(1EP). En utilisant PSO(1EP) plutôt que PSO(HV), on épargne, en moyenne, aux alentours de  $3 \times 10^4$  époques d'entraînement.



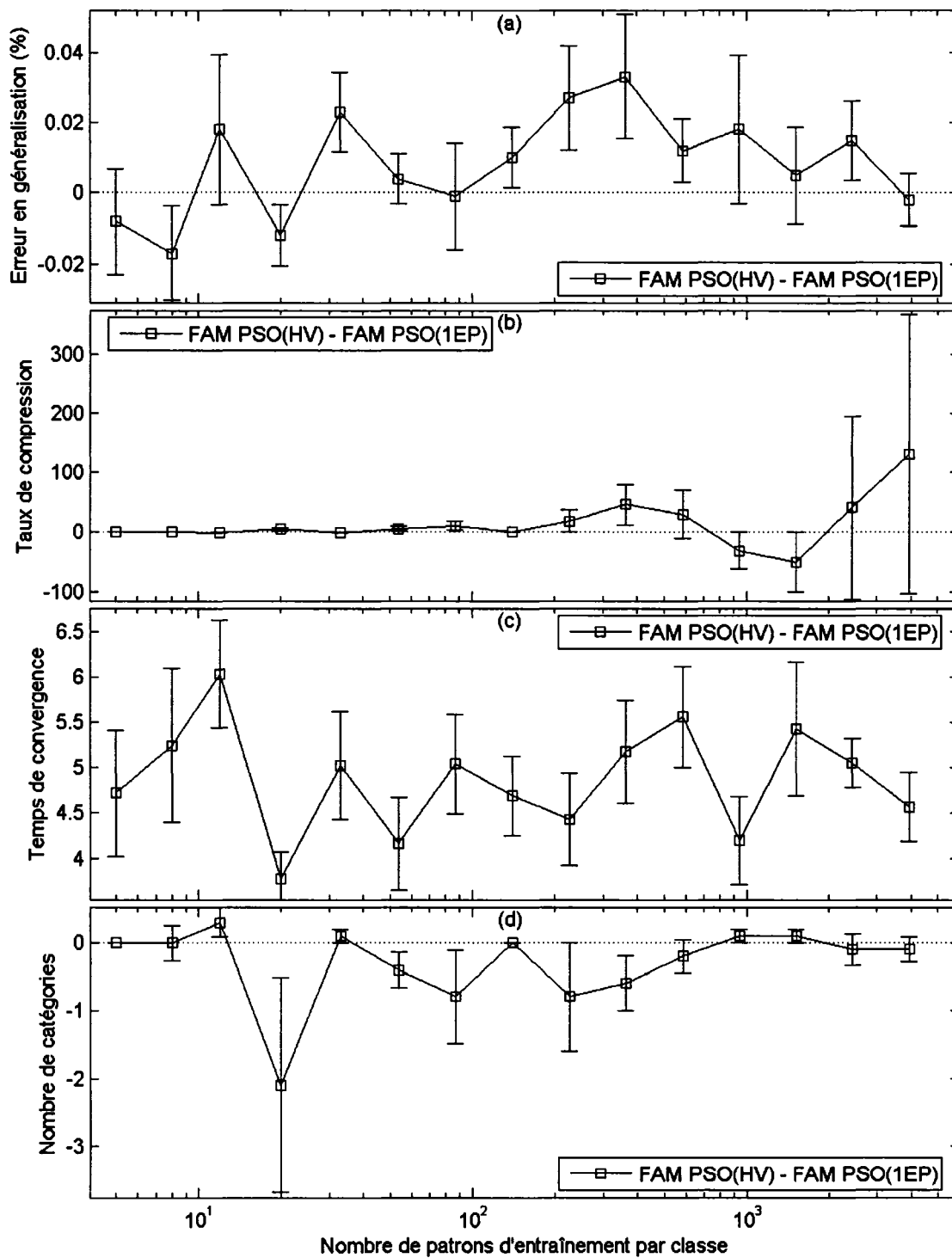


Figure 50 Différence entre PSO(HV) et PSO(1EP) avec la base  $DB_{\mu}(1\%)$

(a) Erreur en généralisation, (b) Taux de compression, (c) Temps de convergence et (d) Nombre de catégories.

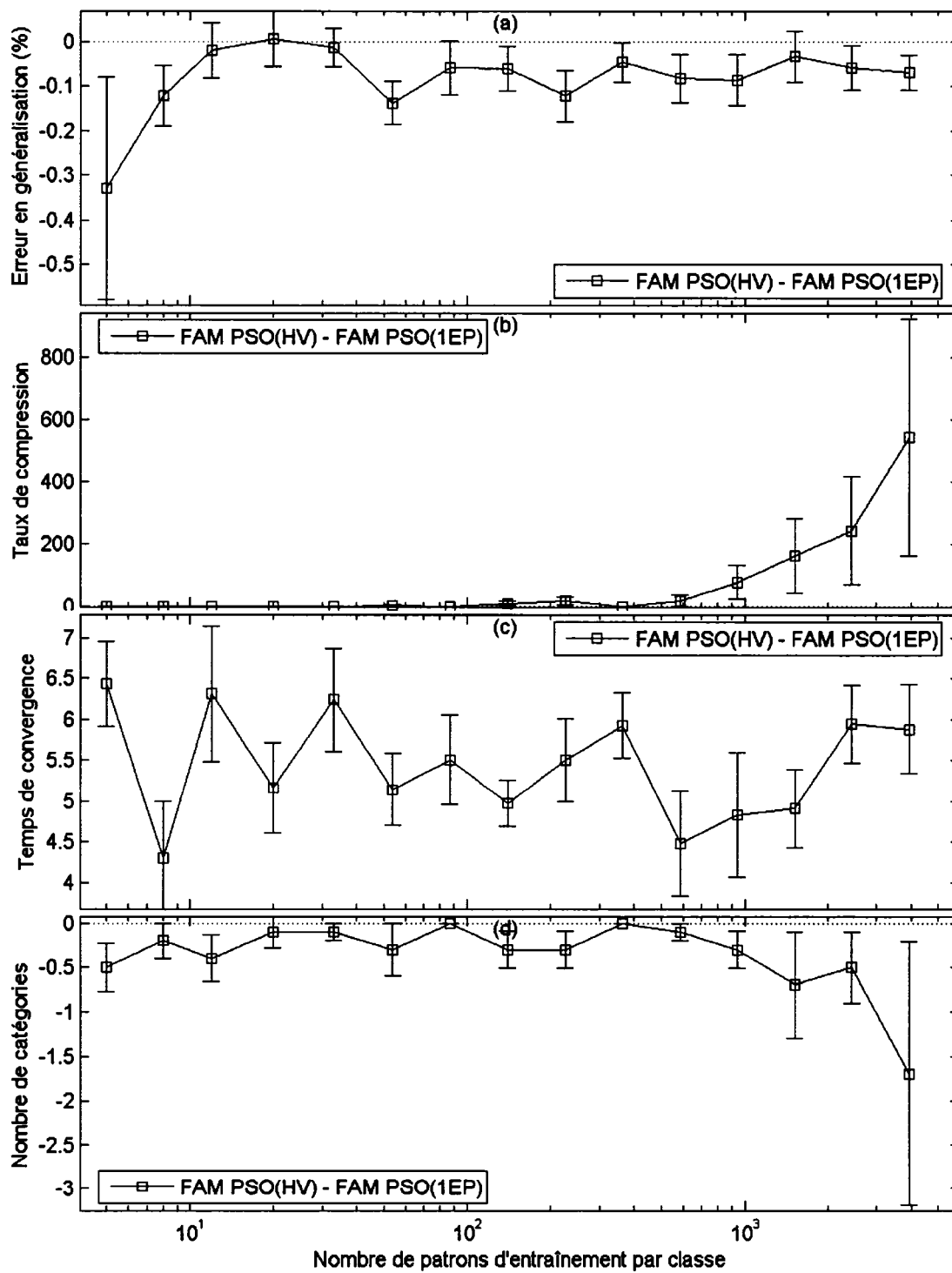


Figure 51 Différence entre PSO(HV) et PSO(1EP) avec la base  $DB_{\mu}(9\%)$

(a) Erreur de généralisation, (b) Taux de compression, (c) Temps de convergence et (d) Nombre de catégories.

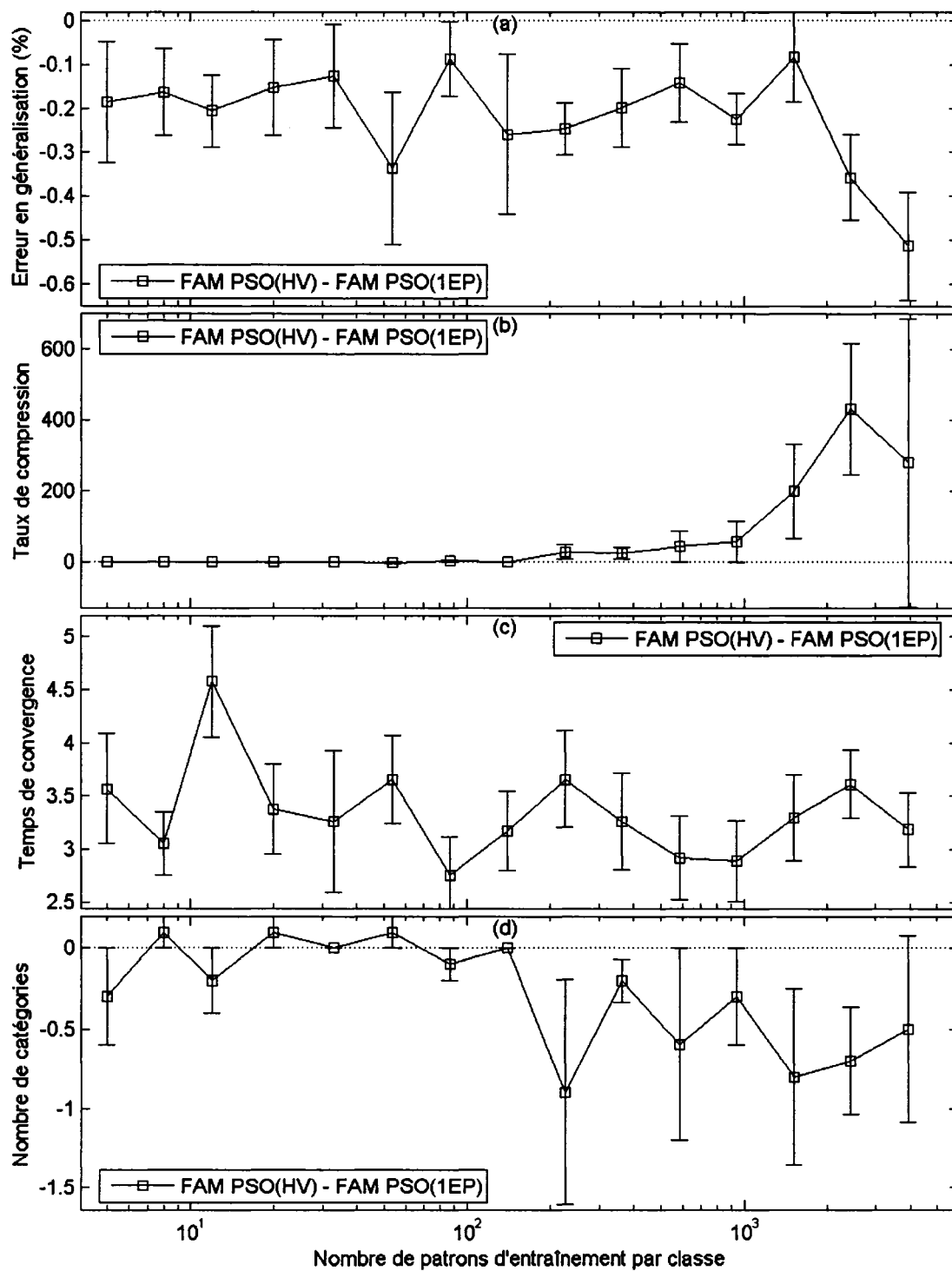


Figure 52 Différence entre PSO(HV) et PSO(1EP) avec la base  $DB_{\mu}(25\%)$

(a) Erreur en généralisation, (b) Taux de compression, (c) Temps de convergence et (d) Nombre de catégories.

La Figure 49(a) démontre également qu'une fois les paramètres optimisés, il reste encore une légère dégradation des performances en généralisation engendrée par la taille de la base d'entraînement. Pour mieux voir ce phénomène, la Figure 53 présente l'erreur de sur-apprentissage due à la taille de la base d'entraînement lors de l'utilisation des stratégies d'apprentissage spécialisées pour FAM avec les bases  $DB_{\mu}$ .

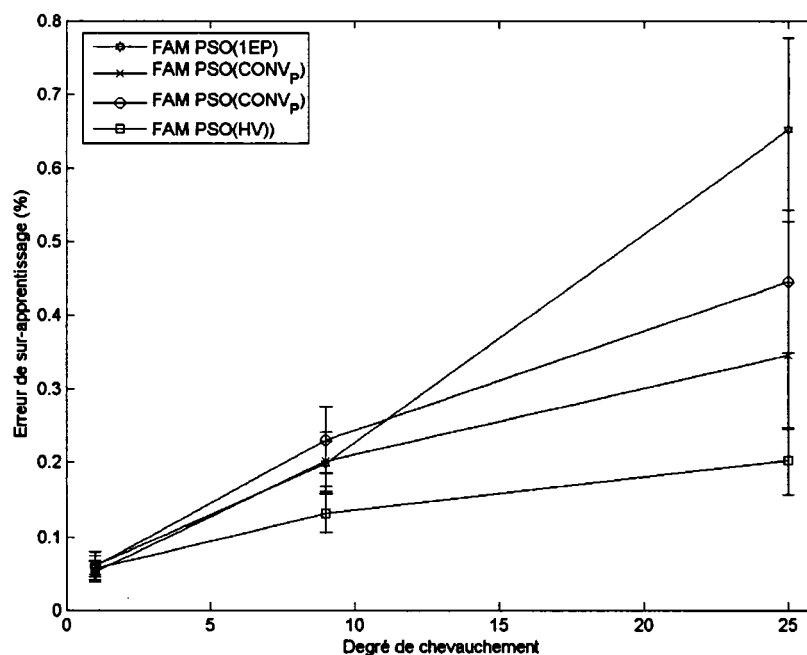


Figure 53 Erreur de sur-apprentissage due à la taille de la base d'entraînement avec les stratégies d'apprentissage spécialisées pour  $DB_{\mu}$

En tenant compte de ces résultats, nous venons de démontrer que, contrairement à ce qu'il a été énoncé dans la conclusion du chapitre 3, le facteur causant la plus grande dégradation des performances à l'intérieur des réseaux FAM n'est pas la taille de la base d'apprentissage. Il faut plutôt dire que, lors de l'utilisation des paramètres standard MT+, la taille de la base d'apprentissage est le facteur le plus influent dans la création d'erreur de sur-apprentissage. Par contre, lors de l'optimisation des paramètres des réseaux FAM avec PSO, nous avons remarqué, qu'en fait, ce sont les paramètres internes de ces réseaux qui ont le plus grand impact sur la dégradation des performances.

Ainsi, la cause première des mauvaises performances des réseaux fuzzy ARTMAP est la sélection des valeurs des paramètres utilisés lors de la phase d'entraînement.

On remarque également sur les figures 39 à 46 que le taux de compression augmente de façon linéaire. Ceci indique que le nombre de catégories créées avec les stratégies PSO reste constant pour toutes les trilles des bases d'apprentissage, soit autour de 2.2 en moyenne. Ce faible nombre de catégories est dû aux valeurs attribuées aux paramètres lors de l'optimisation. Ces valeurs permettent ainsi d'éviter la création de catégories qui n'amélioreront pas les performances en généralisation et ainsi obtenir un réseau compact, simple, dont les performances sont très proches de l'erreur théorique.

On remarque également que les valeurs utilisées pour le paramètre epsilon sont toutes négatives. Ceci montre, tel qu'énoncé au chapitre 3, que le MatchTracking négatif était la bonne direction à suivre pour les bases de données possédant du chevauchement, mais que la valeur absolue de ce paramètre n'était pas assez grande. Les valeurs d'epsilon obtenues oscillent généralement entre -0,2 et -0,7 comparativement à -0,001 lors de l'utilisation des paramètres par défaut MT-. Une grande valeur positive d'epsilon favorise la création de nouvelles catégories, lors de la phase d'apprentissage, lorsque la première catégorie sélectionnée pour classier un patron d'entraînement n'appartient pas à la bonne classe. À l'opposé, une grande valeur négative de ce paramètre favorise la réutilisation des catégories déjà existantes et ainsi évite la création de nouvelles catégories.

La vigilance de base  $\bar{\rho}$  oscille en moyenne entre 0.5 et 0.9, et ce pour tous les degrés de chevauchement testés. Ce paramètre influence directement le nombre de catégories créées lors de la phase d'apprentissage. Une grande valeur de ce paramètre réduit la probabilité qu'une catégorie soit associée au patron d'entraînement présenté et ainsi favorise l'assignation d'une nouvelle catégorie. À l'opposé, une faible valeur permet à plusieurs catégories d'être sélectionnées comme étant la bonne catégorie pour classier

le patron d'entraînement et ainsi évite de créer un trop grand nombre de catégories. De ce fait, ce paramètre influence directement la taille maximale des catégories, soit:

$$|w_j| \leq 2(1 - \rho).$$

Le paramètre de choix  $\alpha$  oscille en moyenne entre 0.25 et 0.75 avec les bases de données possédant un degré de chevauchement. Ce paramètre influence le nombre d'itérations de recherche des catégories pendant la phase d'apprentissage avant de créer une nouvelle catégorie. Ainsi, une grande valeur de ce paramètre entraîne une recherche rapide à travers les catégories existantes lors de l'entraînement, favorisant la création de nouvelles catégories. Une faible valeur favorise les catégories existantes face à l'assignation d'une nouvelle catégorie, soit une recherche plus approfondie à l'intérieur des catégories existantes.

La vitesse d'apprentissage  $\beta$  oscille en moyenne entre 0.5 et 0.9. Ce paramètre détermine la vitesse à laquelle les catégories s'adaptent aux patrons d'entraînement, chaque fois qu'une catégorie est modifiée. Une grande valeur permet aux catégories de s'adapter le plus rapidement possible alors qu'une faible valeur diminue la vitesse de changement de la taille de la catégorie.

Les valeurs des paramètres obtenues avec les stratégies d'apprentissage spécialisées utilisant l'algorithme d'optimisation PSO favorisent la création d'un petit nombre de catégories (epsilon négatif) possédant une grande surface (vigilance de base élevée).

#### **4.2 Effets de la structure des bases de données**

Cette section présente une comparaison entre les deux méthodes de création du degré de chevauchement lors de l'optimisation des paramètres des réseaux fuzzy ARTMAP.

### 4.2.1 Résultats

Dans la section 3.2, nous avons conclu que les deux structures utilisées pour la création du chevauchement ( $DB_{\mu}$  et  $DB_{\sigma}$ ) engendrent des bases caractérisées par un degré de difficulté semblable pour les réseaux fuzzy ARTMAP. Ainsi, les réseaux obtenus avec ces deux méthodes ont des performances analogues sur le plan des erreurs en généralisation, des temps de convergence ainsi que des taux de compression. Cette section montre que ce phénomène est encore présent lors de l'apprentissage avec l'optimisation des paramètres internes des réseaux fuzzy ARTMAP.

Les figures 53, 54 et 55 présentent une comparaison entre la base  $DB_{\mu}(9\%)$  et  $DB_{\sigma}(9\%)$  au niveau des erreurs en généralisation, des temps de convergence ainsi que des taux de compression, pour les quatre stratégies d'apprentissage spécialisées pour FAM. Ainsi, lorsque la courbe est positive, la valeur obtenue avec la base  $DB_{\mu}(9\%)$  est plus grande que celle obtenue avec la base  $DB_{\sigma}(9\%)$ , et vice-versa.

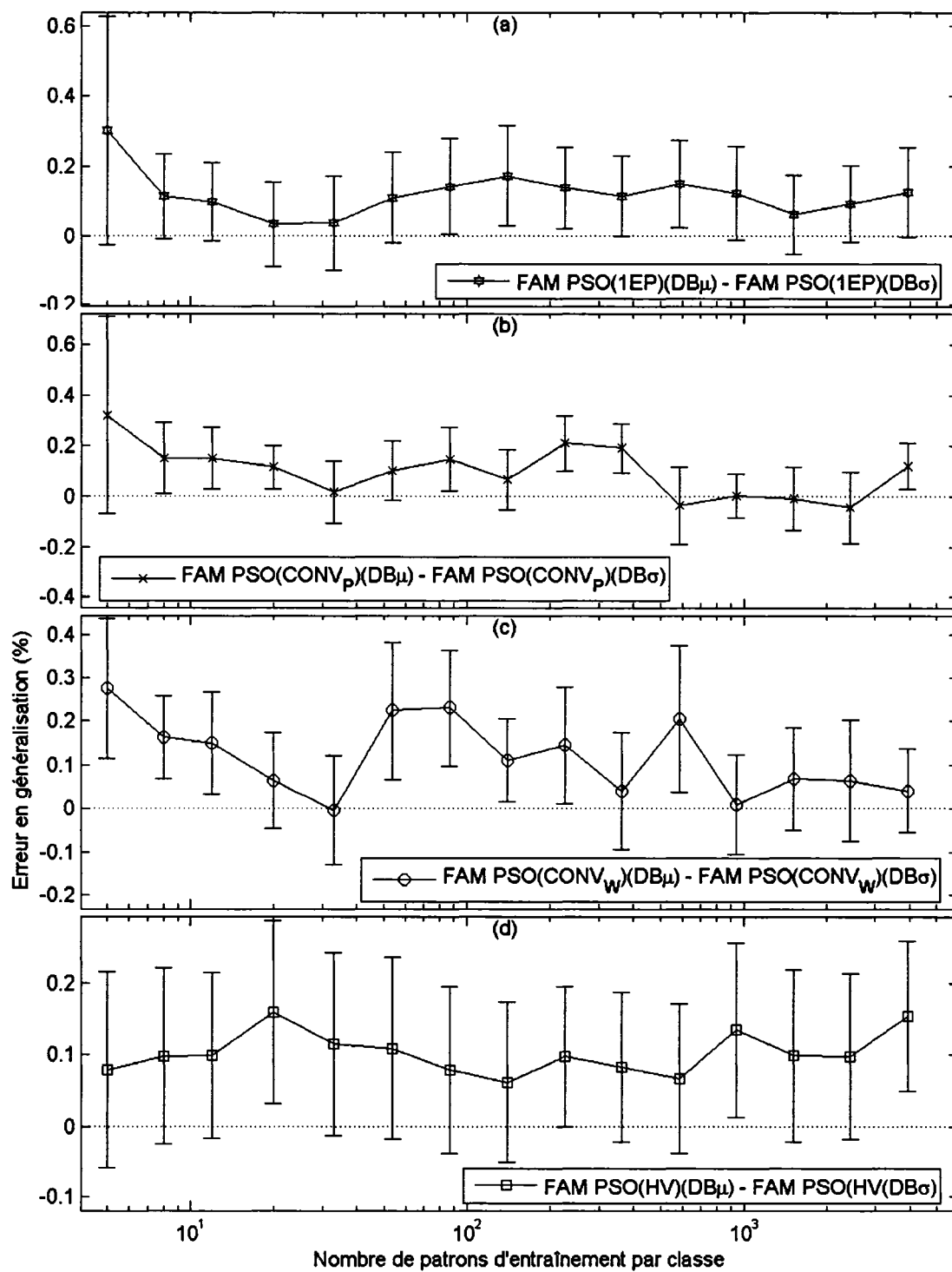


Figure 54  $DB_\mu(9\%)$  versus  $DB_\sigma(9\%)$  sur l'erreur en généralisation avec PSO  
 (a) Une époque, (b) Convergence des patrons, (c) Convergence des poids synaptiques et  
 (d) Validation hold-out.



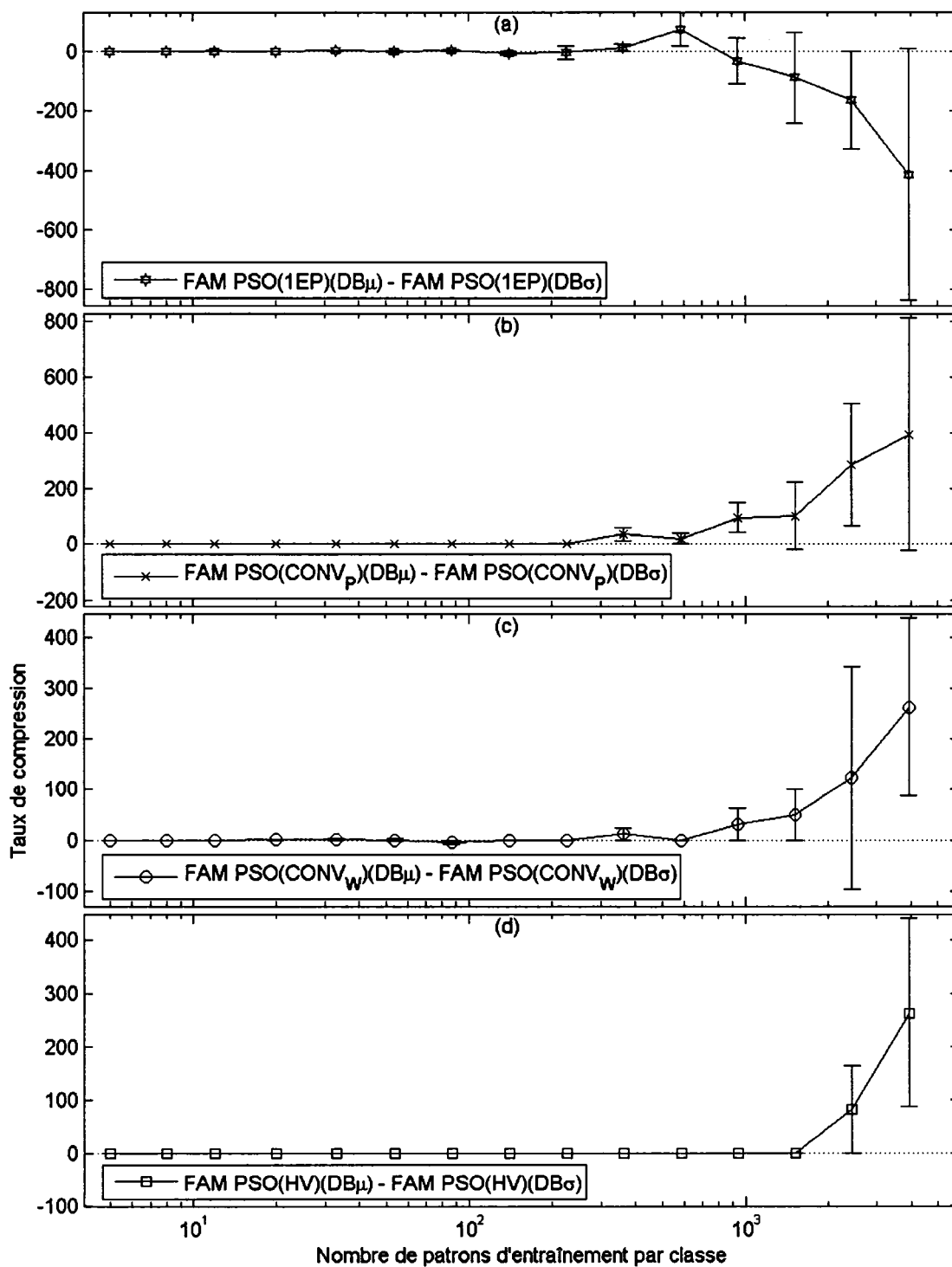


Figure 55  $DB_\mu(9\%)$  versus  $DB_\sigma(9\%)$  sur le taux de compression avec PSO  
 (a) Une époque, (b) Convergence des patrons, (c) Convergence des poids synaptiques et  
 (d) Validation hold-out.

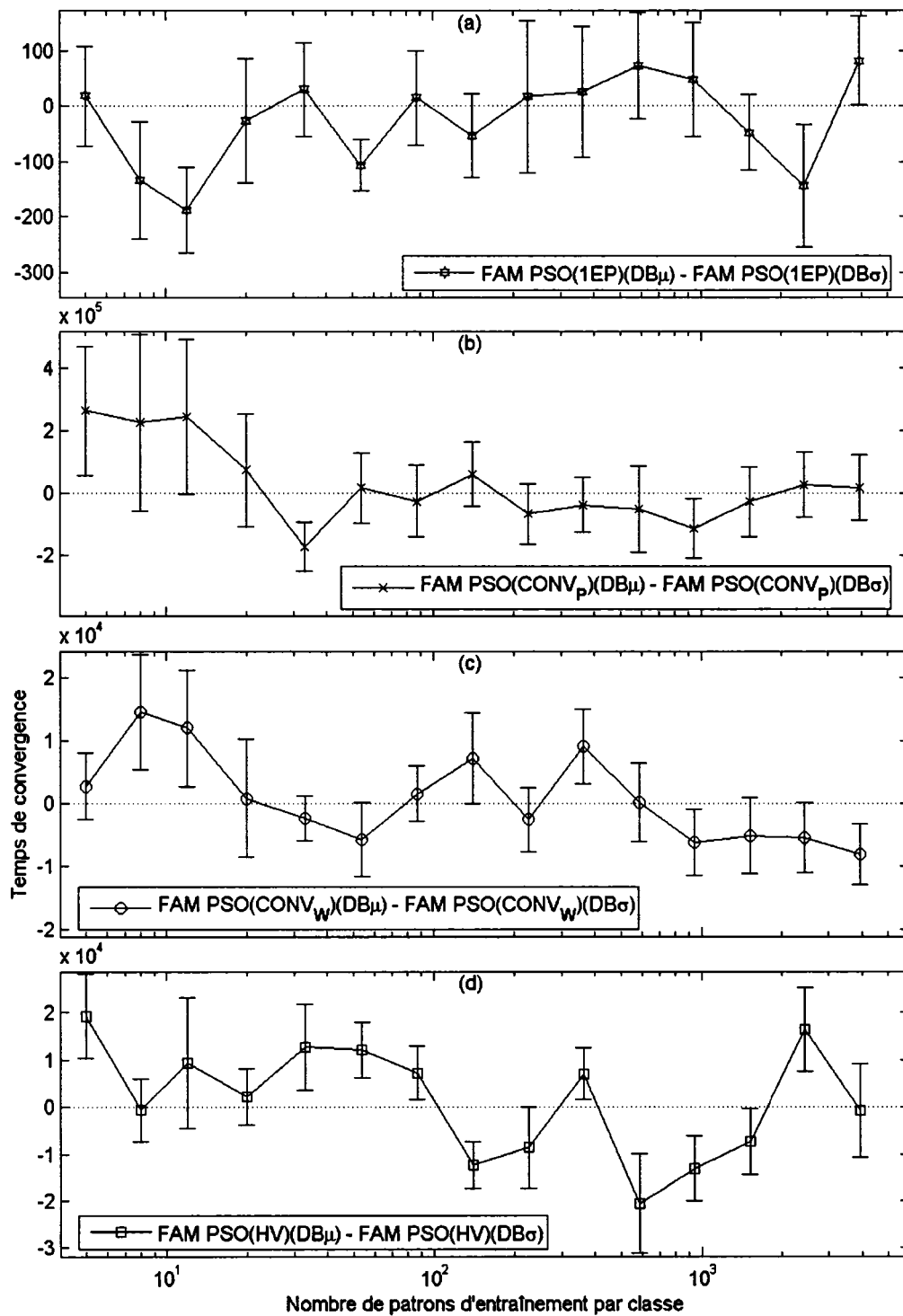


Figure 56 DB $\mu$ (9%) versus DB $\sigma$ (9%) sur le temps de convergence avec PSO  
 (a) Une époque, (b) Convergence des patrons, (c) Convergence des poids synaptiques et  
 (d) Validation hold-out.

#### 4.2.2 Analyse

La Figure 54 montre que les performances en généralisation obtenues avec les bases de données  $DB_{\mu}(9\%)$  et  $DB_{\sigma}(9\%)$  sont similaires. La dispersion des erreurs en généralisation est calculée avec l'erreur standard (voir équation (2.8)) et puisque les erreurs en généralisation s'étendent sur les deux régions, positives et négatives, ces résultats ne sont pas significativement différents. Ainsi, statistiquement, les deux méthodes de création de chevauchement génèrent des bases caractérisées par un degré de difficulté semblable pour les réseaux fuzzy ARTMAP.

Au niveau des taux de compression, la Figure 55 illustre que, généralement, les résultats obtenus sont similaires. Par contre, lors de l'utilisation de la stratégie d'apprentissage  $PSO(CONV_w)$ , nous obtenons des taux de compression supérieurs avec la base  $DB_{\mu}(9\%)$  lors de l'emploi des grandes tailles de la base d'entraînement.

La Figure 56 présente la comparaison entre la base  $DB_{\mu}(9\%)$  et la base  $DB_{\sigma}(9\%)$  pour les temps de convergence. Ces résultats sont majoritairement similaires à l'exception de ceux obtenus avec la stratégie d'apprentissage  $PSO(CONV_w)$ . Cette stratégie demande un plus grand nombre d'époques d'entraînement avec la base  $DB_{\sigma}(9\%)$  qu'avec  $DB_{\mu}(9\%)$ , lors de l'utilisation des grandes tailles de la base d'apprentissage.

Malgré les temps de convergence, la majorité des résultats obtenus par les deux structures des bases de données avec chevauchement sont semblables. Ainsi, nous en tirons la même conclusion qu'au chapitre 3, soit qu'il n'y a pas de différence entre ces deux structures, même lors de l'optimisation des paramètres internes des réseaux fuzzy ARTMAP.

### **4.3 Bases de données sans chevauchement**

Cette sous-section traite des effets de l'optimisation des paramètres internes des réseaux fuzzy ARTMAP avec les stratégies d'apprentissage spécialisées pour les bases de données sans chevauchement ( $DB_{P2}$  et  $DB_{CIS}$ ). Étant donné qu'aucun effet dû à la normalisation n'est présent avec les paramètres standard FAM (voir section 3.3), la technique de normalisation MinMax est utilisée.

#### **4.3.1 Résultats**

Les figures suivantes présentent les erreurs en généralisation, les temps de convergence et les taux de compression obtenus pour les quatre stratégies d'apprentissage spécialisées avec les bases de données  $DB_{CIS}$  et  $DB_{P2}$ . De plus, les valeurs optimisées des quatre paramètres internes des réseaux fuzzy ARTMAP sont présentées. À noter que le temps de convergence inclut toutes les époques effectuées par toutes les particules lors de l'optimisation PSO.

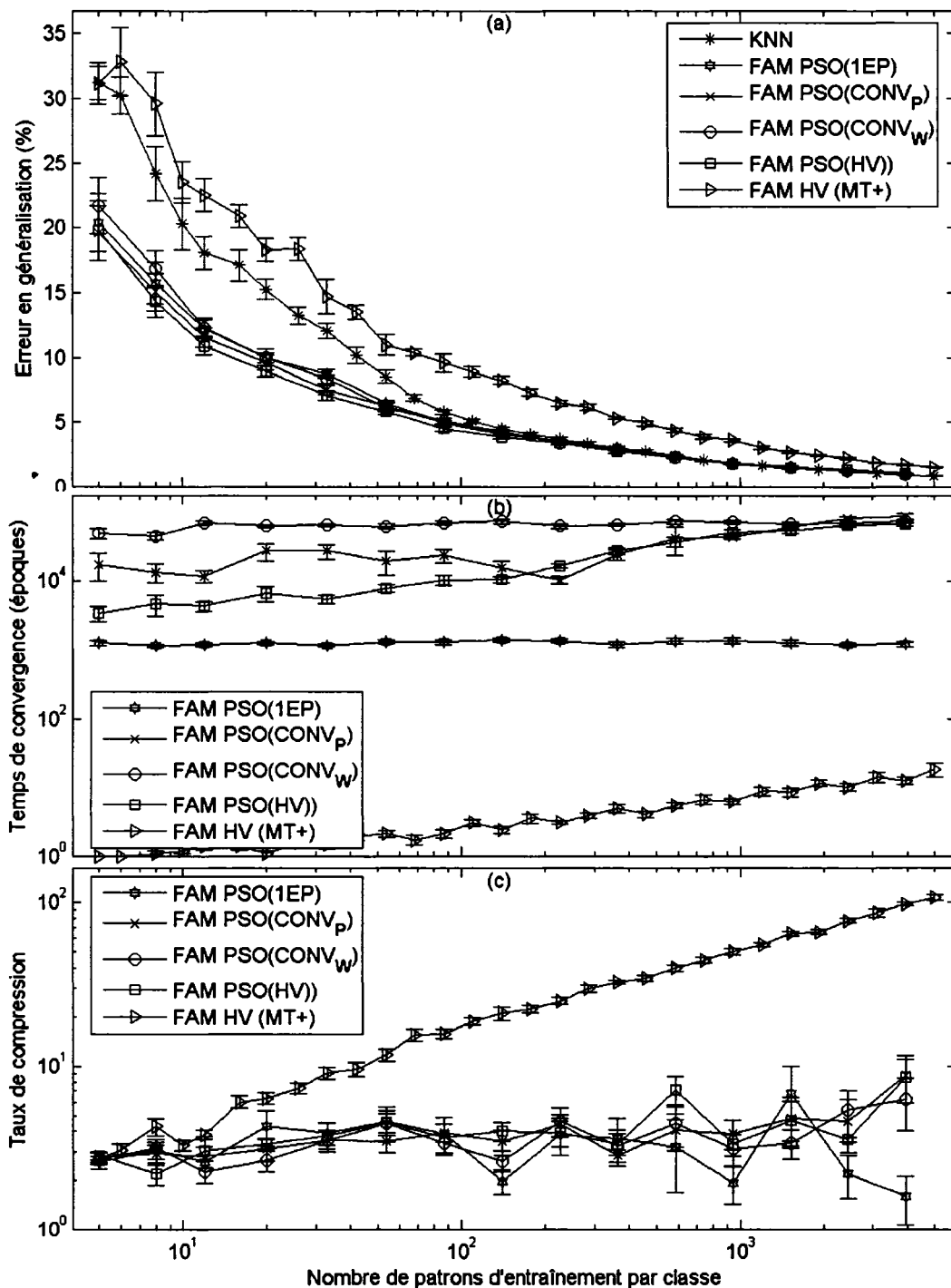


Figure 57 Performance du FAM avec les stratégies PSO sur la base DB<sub>CIS</sub>  
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression

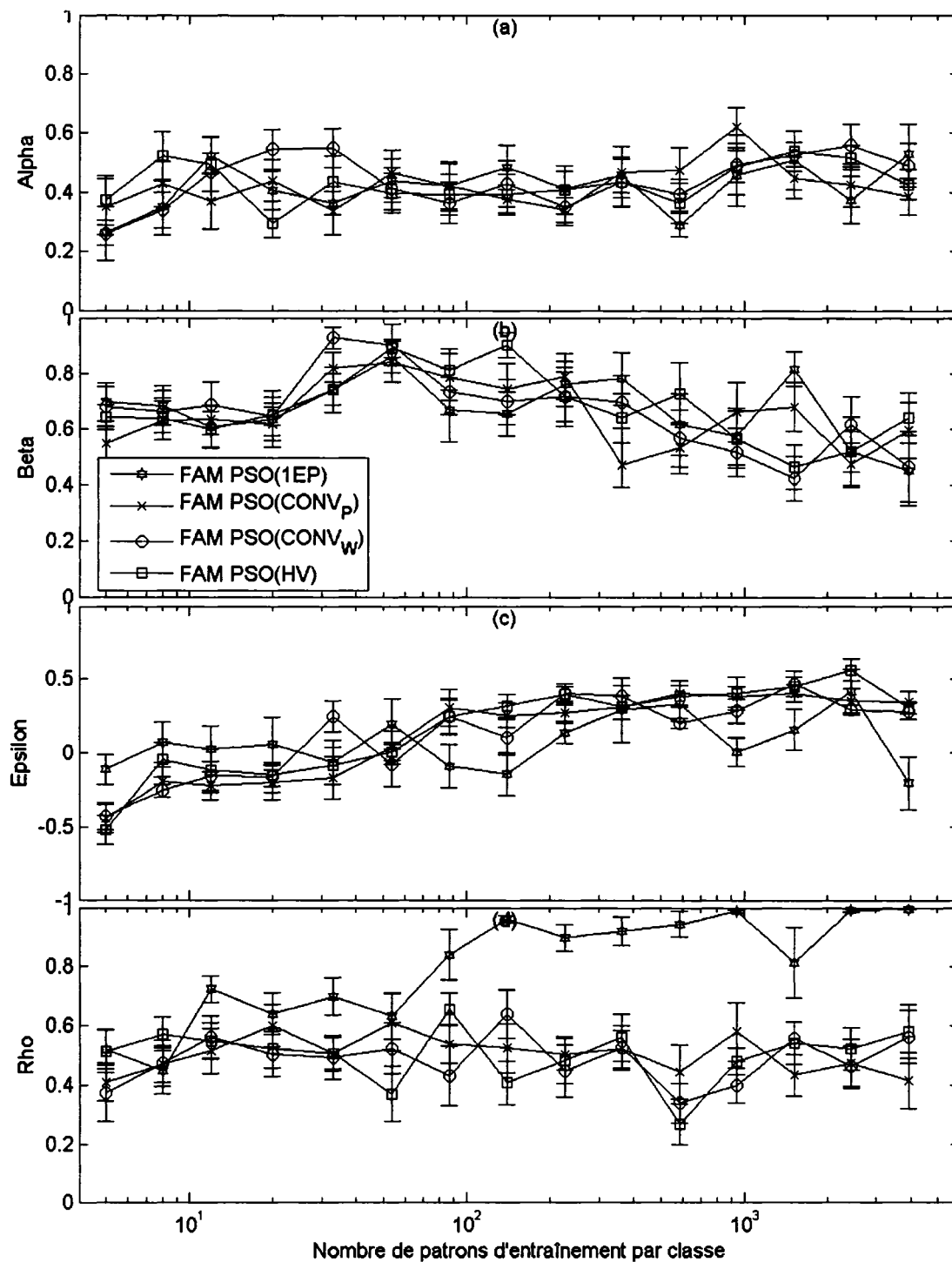


Figure 58 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base DB<sub>CIS</sub>

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base

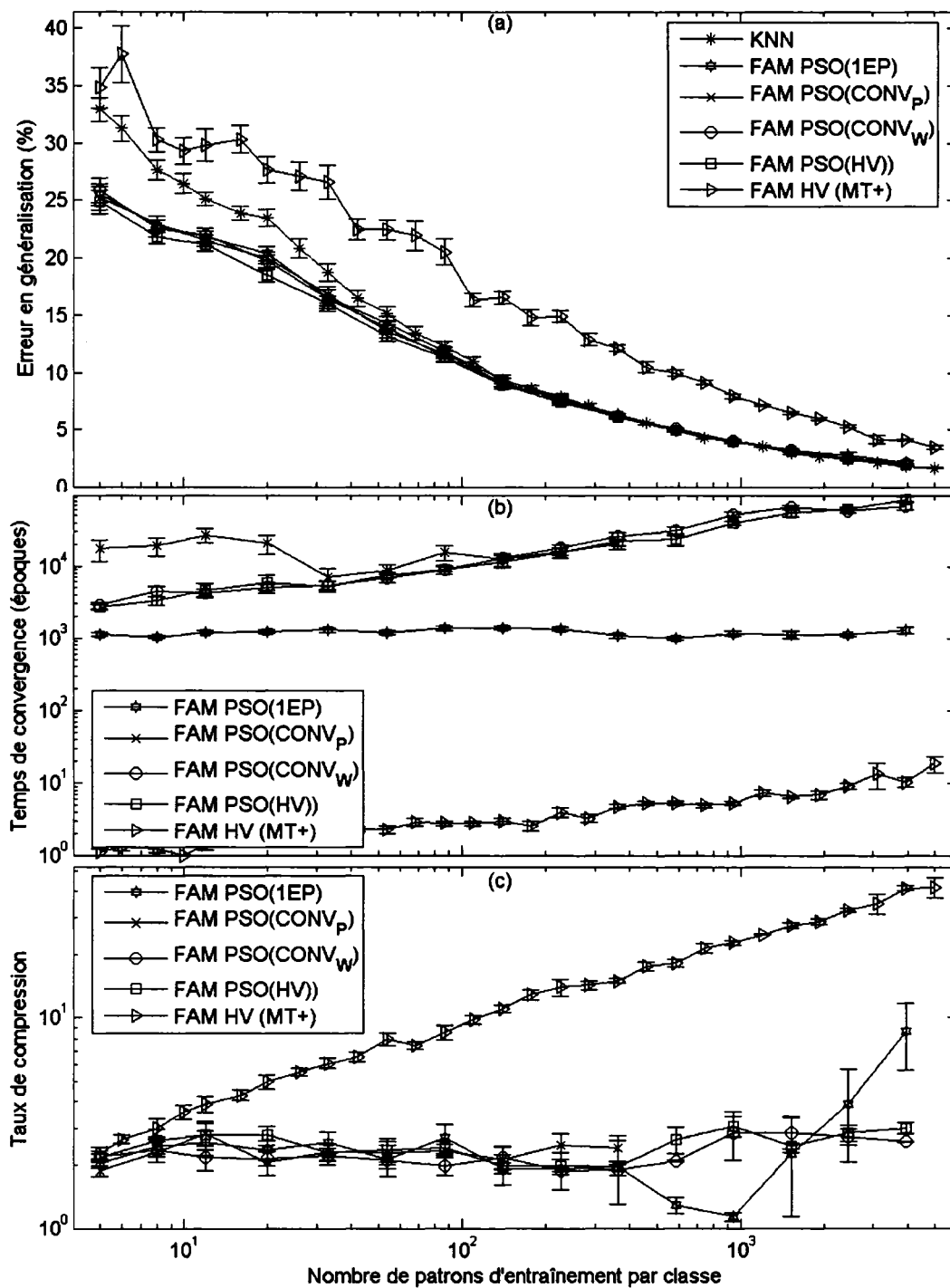


Figure 59 Performance du FAM avec les stratégies PSO sur la base  $DB_{P2}$   
 (a) Erreur en généralisation, (b) temps de convergence et (c) taux de compression

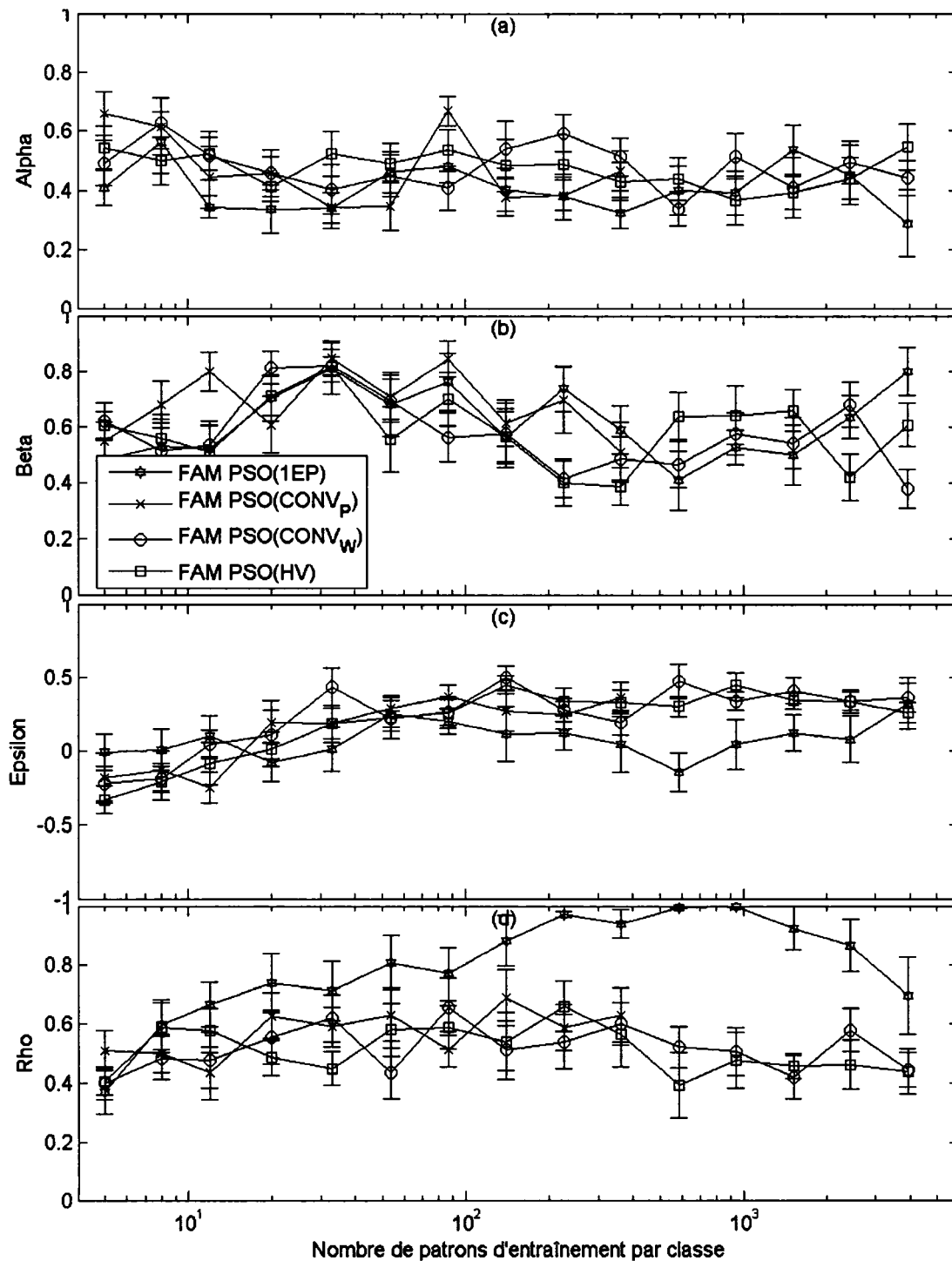


Figure 60 Valeurs des paramètres internes du FAM lors de l'utilisation des stratégies spécialisées PSO avec la base  $DB_{P_2}$

(a)  $\alpha$  : paramètre de choix, (b)  $\beta$  : vitesse d'apprentissage, (c)  $\varepsilon$  : paramètre de MatchTracking et (d)  $\bar{\rho}$  : vigilance de base



### 4.3.2 Analyse

Les figures précédentes montrent que les stratégies d'entraînement spécialisées, effectuant l'optimisation des paramètres internes des réseaux fuzzy ARTMAP, permettent de meilleures performances en généralisation, particulièrement avec les bases d'apprentissage de petites tailles.

L'optimisation de ces paramètres permet également de générer des performances en généralisation comparables à celles obtenues avec le  $k$ NN et ce, pour les quatre stratégies d'apprentissage utilisées. À noter que lors de l'utilisation de  $k$ NN avec les bases de données sans chevauchement, les valeurs de  $k$  sélectionnées sont presque toujours égales à 1. Avec un faible nombre de patrons d'apprentissage, les erreurs en généralisation relevées sont moindres que les celles obtenues avec le  $k$ NN. Par contre, plus la taille de la base d'entraînement augmente, plus cet écart diminue, pour finalement disparaître.

On remarque également que les performances en généralisation des quatre stratégies d'apprentissage sont similaires. Puisque la stratégie d'entraînement PSO(1EP) obtient les temps de convergence les plus rapides, il serait préférable de l'utiliser lors de l'optimisation des paramètres, afin de minimiser le temps de convergence avec les bases de données sans chevauchement et ce, sans dégradation des performances en généralisation. Au niveau des taux de compression, on remarque des résultats similaires entre les quatre stratégies d'apprentissage, sauf lorsque la taille de la base d'entraînement est très grande. Ainsi, avec le nombre maximal de patrons d'apprentissage, les taux de compression obtenus avec la stratégie PSO(1EP) sont différents de ceux obtenus avec les trois autres stratégies. Avec la base DB<sub>CIS</sub>, PSO(1EP) obtient des taux plus faibles que les trois autres stratégies, alors qu'avec DB<sub>P2</sub> elle obtient les meilleurs taux. Malgré ce phénomène, nous recommandons quand même d'utiliser la stratégie d'apprentissage PSO(1EP) avec l'optimisation des paramètres.

Pour appuyer cette analyse, les figures 60 et 61 présentent la comparaison entre les résultats obtenus pour les stratégies PSO(HV) et PSO(1EP), et ce pour l'erreur en généralisation, le taux de compression, le temps de convergence et le nombre de catégories créées avec les bases  $DB_{CIS}$  et  $DB_{P2}$ . Ainsi, lorsque la courbe est positive, la valeur obtenue avec la stratégie PSO(HV) est plus grande que celle obtenue avec la stratégie PSO(1EP), et vice-versa.

On observe sur ces deux figures que, lors de l'utilisation de la taille maximale de la base d'apprentissage, les différences entre les stratégies PSO(HV) et PSO(1EP) sont pratiquement nulles. Les taux de compression de la base  $DB_{CIS}$  favorisent la stratégie PSO(HV) alors que ceux obtenus avec  $DB_{P2}$  favorisent PSO(1EP). Le grand avantage d'utiliser la stratégie d'apprentissage PSO(1EP) se situe au niveau des temps de convergence. Avec la taille maximale de la base d'apprentissage, les temps de convergence obtenus avec la stratégie PSO(1EP) sont de  $7 \times 10^4$  à  $9 \times 10^4$  époques d'entraînement plus rapides qu'avec la stratégie PSO(HV).

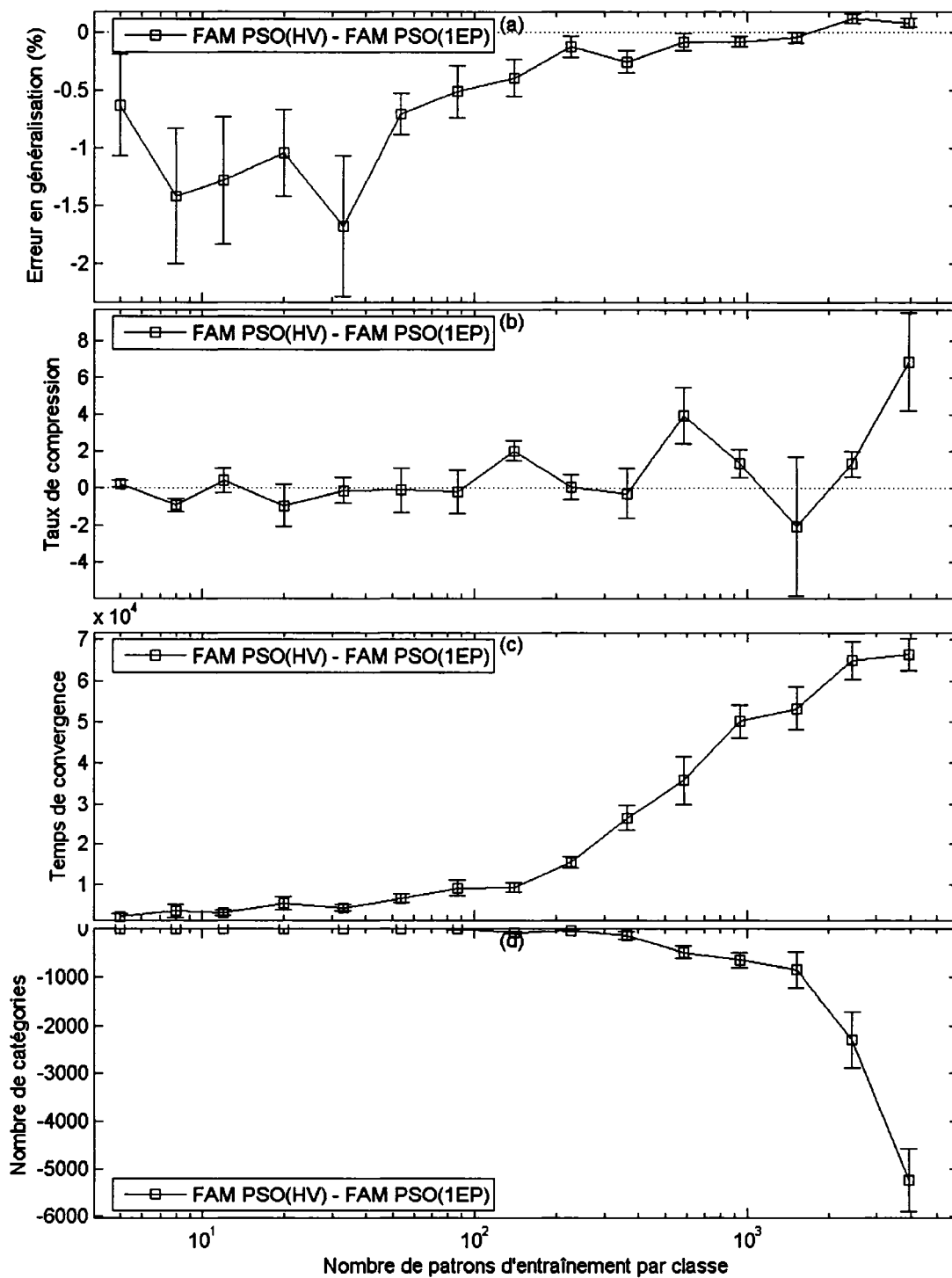


Figure 61 Différence entre PSO(HV) et PSO(1EP) avec la base  $DB_{CIS}$

(a) Erreur en généralisation, (b) Taux de compression, (c) Temps de convergence et (d) Nombre de catégories.

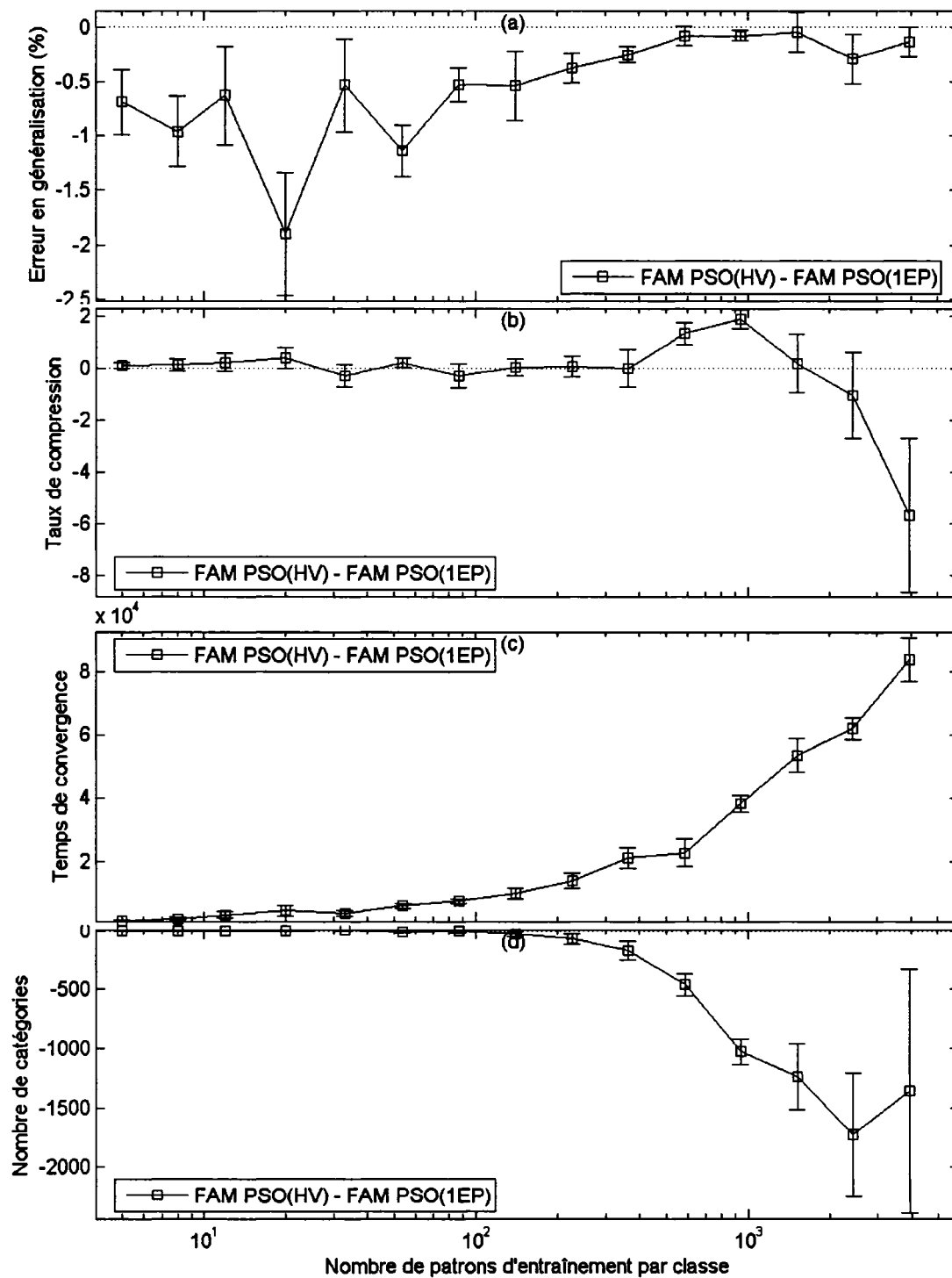


Figure 62 Différence entre PSO(HV) et PSO(1EP) avec la base DB<sub>P2</sub>

(a) Erreur en généralisation, (b) Taux de compression, (c) Temps de convergence et (d) Nombre de catégories.

Sans l'optimisation des paramètres, il n'y a pas de dégradation des performances en généralisation causée par la taille de la base d'entraînement avec les bases de données sans chevauchement. Ceci est encore le cas lorsque les paramètres sont optimisés par les stratégies d'apprentissage spécialisées. Au chapitre 3, la Figure 15 présente les bornes de décision ainsi que les catégories créées lors de l'augmentation progressive de la taille de la base de données d'entraînement pour une des répétitions avec la base  $DB_{CIS}$ . Cette augmentation fait passer l'erreur en généralisation de 19.6% (12 patrons par classe), à 5.59% (226 patrons par classe) pour atteindre 1.66% (5000 patrons par classe), et ce pour la stratégie d'apprentissage HV. Lors de l'optimisation des paramètres, nous obtenons, pour cette même expérience, les erreurs en généralisation suivantes : 15.58% (12 patrons par classe), 3.51% (226 patrons par classe) et 1.36% (5000 patrons par classe) avec la stratégie PSO(HV). Pour mieux comprendre le gain des performances en généralisation obtenu lors de l'utilisation de la stratégie PSO(HV) avec la base  $BD_{CIS}$ , la Figure 63 présente les bornes de décision pour ce cas.

La Figure 63 montre que les bornes de décision obtenues avec l'optimisation des paramètres sont plus nettes et suivent mieux la borne optimale qu'avec l'utilisation de paramètres standard MT+.

La Figure 58 nous donne également un indicatif sur les valeurs optimisées des paramètres lors de l'application de l'algorithme PSO avec la base  $DB_{CIS}$ . Ainsi, on remarque que la valeur d'epsilon a tendance à être positive lors de l'utilisation d'un grand nombre de patrons d'apprentissage. Le paramètre  $\bar{\rho}$  tend vers des valeurs positives aux alentours de 0,5. Pour mieux voir l'effet de ces paramètres et comprendre comment ceux-ci peuvent obtenir une meilleure borne de décision, la Figure 64 présente une comparaison entre les catégories obtenues avec et sans optimisation lors de l'utilisation de la taille maximale de la base d'apprentissage (5000 patrons par classe) sur une des 10 répétitions avec la base  $DB_{CIS}$ .

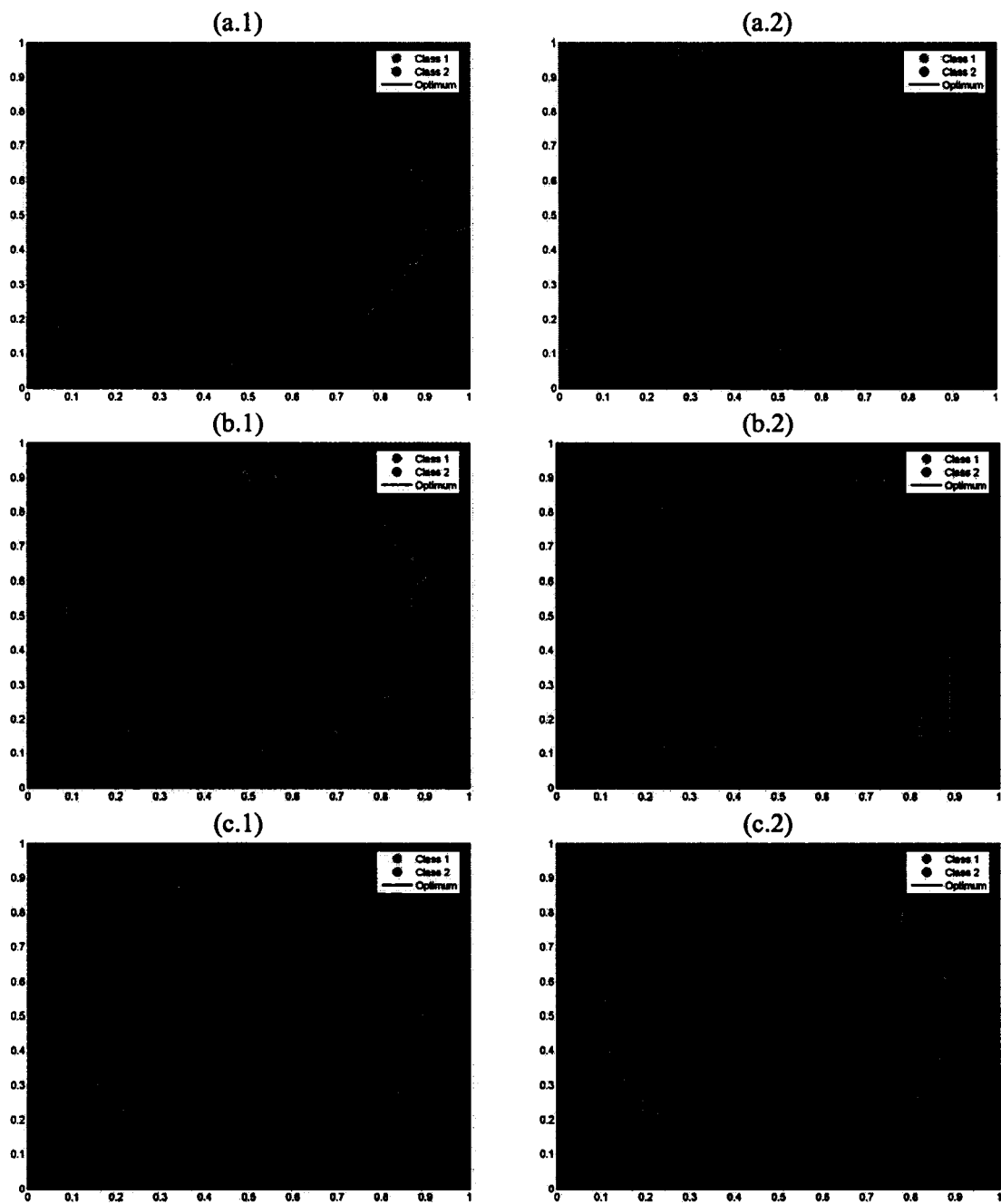


Figure 63 Bornes obtenues lors de l'accroissement de la taille de la base d'apprentissage avec la base  $DB_{CIS}$ , avec (PSO(HV)) et sans optimisation (HV MT+)

Soit : (a) 12 patrons, (b) 226 patrons et (c) 5000 patrons d'entraînement par classe, et (.1) avec optimisation des paramètres et (.2) sans optimisation avec MT+.

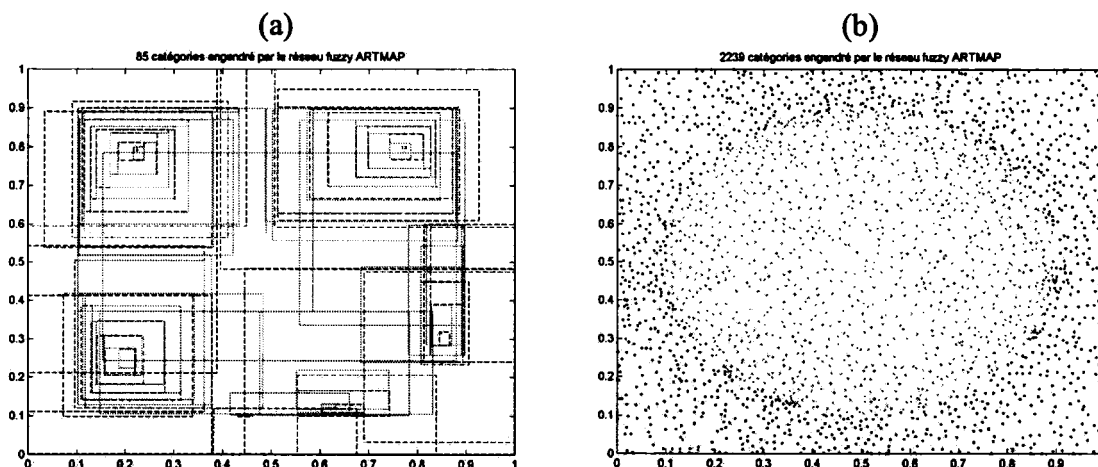


Figure 64 Catégories obtenues avec 5000  $p/\omega$  de la base  $DB_{CIS}$

(a) sans optimisation des paramètres HV MT+ et (b) avec optimisation des paramètres PSO(HV)

Ainsi, lors de l'optimisation des paramètres internes des réseaux fuzzy ARTMAP avec PSO(HV) sur la base  $DB_{CIS}$ , les paramètres obtenus tendent à créer un très grand nombre de petites catégories, soit 2239 catégories avec l'optimisation (PSO(HV)) comparativement à 85 sans optimisation (HV MT+). On remarque que, sur ces 2239 catégories, seulement 134 représentent plus d'un patron. Ainsi, 94.20% des catégories obtenues avec l'optimisation des paramètres représentent un seul patron. Ce phénomène soulève un fait intéressant, soit que les réseaux fuzzy ARTMAP peuvent tendre vers le comportement d'un classificateur 1NN. Cependant, la fonction de calcul de la distance est différente. En outre, puisque tous les patrons de la base d'entraînement ne deviennent pas des catégories, on peut parler d'un réseau semblable au 1NN mais possédant un meilleur taux de compression. En fait, on peut établir qu'un classificateur 1NN a une compression de 1, soit une catégorie pour chaque patron d'apprentissage. Les réseaux fuzzy ARTMAP obtenant des performances en généralisation semblables à celles de  $kNN$  ( $k$  est majoritairement égal à 1 avec les bases de données sans chevauchement) possèdent des compressions allant de 2 à 9 suivant l'augmentation graduelle de la base d'apprentissage avec  $DB_{CIS}$  (voir figure 57.c).

#### 4.4 Conclusion

Ce chapitre a clairement montré l'importance d'une sélection adéquate des paramètres internes des réseaux fuzzy ARTMAP en fonction du type de bases de données à classifier. Cependant, puisqu'il n'est pas aisé d'établir des valeurs précises pour chaque cas, les stratégies d'apprentissage spécialisées pour les réseaux FAM que nous avons développées résolvent ce problème. L'algorithme PSO optimise les valeurs des paramètres, obtenant ainsi de meilleures performances en généralisation et ce, pour tous les problèmes que nous avons testés. Il faut également noter que d'autres algorithmes d'optimisation pourraient être utilisés tels que les algorithmes génétiques. D'autre part, les taux de compression obtenus avec les bases possédant un degré de chevauchement sont nettement meilleurs qu'avec l'utilisation des paramètres standard MT- ou MT+. En optimisant ces paramètres, le réseau FAM tend soit vers un estimateur des centres de masses de chaque classe (comme avec les bases synthétiques avec chevauchement), soit vers un réseau semblable au INN (comme avec les bases synthétiques sans chevauchement).

La stratégie PSO(1EP) obtient des performances en généralisation et des taux de compression majoritairement similaires aux trois autres stratégies d'apprentissage tout en obtenant les temps de convergence les plus courts et ce, pour toutes les bases de données testées. Il est donc recommandé d'employer uniquement la stratégie PSO(1EP) afin d'accélérer les temps de convergence de l'optimisation.

Finalement, la caractéristique du fuzzy ARTMAP qui engendre le plus de dégradation des performances est la valeur des paramètres. On ne peut donc se fier aux valeurs standard (MT- et MT+) lors de l'utilisation des réseaux fuzzy ARTMAP et l'optimisation des ces paramètres pour ces réseaux devient une nécessité. Cependant, même avec les paramètres optimisés, il reste encore une légère erreur de sur-apprentissage due à la taille de la base d'apprentissage.



Le tableau XI présente un sommaire des résultats obtenus lors de l'utilisation de la taille maximale de la base d'apprentissage avec les différentes stratégies d'entraînement pour les diverses bases de données. Ces résultats sont l'erreur en généralisation moyenne et la dispersion des résultats obtenues sur les 10 réplifications. L'annexe 7 présente une synthèse de tous les degrés de chevauchement testés.

Tableau XI

Résultats sommaires avec 5k patrons par classe

Stratégies d'apprentissage	Erreur en généralisation moyenne (dispersion des résultats) %				
	DB $\mu$ (1%)	DB $\mu$ (9%)	DB $\mu$ (25%)	DB $\sigma$	DB $\rho$
Erreur théorique	1,00	9,00	25,00	0,00	0,00
CQB	1,00 (0,04)	9,12 (0,08)	25,11 (0,10)	ND	ND
kNN	1,08 (0,03)	9,88 (0,08)	27,23 (0,12)	0,86 (0,03)	1,65 (0,04)
1NN	1,54 (0,03)	13,35 (0,10)	33,49 (0,16)	0,84 (0,02)	1,61 (0,04)
FAM 1EP MT-	2,75 (0,20)	22,49 (0,87)	40,58 (0,47)	4,20 (0,25)	8,89 (0,44)
FAM HV MT-	2,17 (0,08)	20,80 (0,56)	39,83 (0,31)	1,69 (0,07)	4,26 (0,16)
FAM CONV <sub>w</sub> MT-	2,74 (0,18)	20,45 (0,46)	40,31 (0,27)	1,77 (0,09)	4,48 (0,38)
FAM CONV <sub>p</sub> MT-	2,56 (0,09)	22,00 (0,66)	40,42 (0,31)	1,59 (0,04)	4,51 (0,19)
FAM 1EP MT+	2,51 (0,14)	18,78 (0,38)	38,81 (0,36)	3,98 (0,21)	7,33 (0,33)
FAM HV MT+	1,88 (0,05)	15,17 (0,13)	36,10 (0,20)	1,58 (0,05)	3,68 (0,07)
FAM CONV <sub>w</sub> MT+	1,97 (0,09)	15,30 (0,16)	35,94 (0,15)	1,64 (0,05)	3,66 (0,08)
FAM CONV <sub>p</sub> MT+	1,90 (0,07)	15,44 (0,15)	36,14 (0,20)	1,47 (0,05)	3,61 (0,08)
FAM PSO(1EP)	1,04 (0,03)	9,35 (0,08)	25,50 (0,09)	1,13 (0,12)	2,05 (0,10)
FAM PSO(HV) <sup>3</sup>	1,07 (0,04)	9,24 (0,08)	25,27 (0,09)	1,06 (0,05)	1,99 (0,04)
FAM PSO(CONV <sub>w</sub> ) <sup>3</sup>	1,06 (0,03)	9,32 (0,08)	25,56 (0,16)	1,02 (0,05)	2,02 (0,04)
FAM PSO(CONV <sub>p</sub> ) <sup>3</sup>	1,06 (0,04)	9,31 (0,07)	25,48 (0,18)	1,00 (0,03)	2,01 (0,03)

MCours.com

<sup>3</sup> Résultats obtenus avec 3940 patrons par classe au lieu de 5000 patrons par classe.