

# Contribution au choix d'un modèle optimal par la variance non reconstruite

## Sommaire

---

<b>3.1</b>	<b>Introduction</b>	<b>63</b>
<b>3.2</b>	<b>Détection et détectabilité de défauts</b>	<b>64</b>
3.2.1	Détectabilité généralisée de défauts	64
3.2.1.1	Indice $T^2$ de Hotelling	66
3.2.1.2	Indice $SPE$	67
3.2.1.3	Indice $SWE$	68
3.2.1.4	Indice combiné	69
3.2.1.5	Indice de Mahalanobis	70
3.2.2	Influence de la modélisation sur la détectabilité de défauts	71
3.2.2.1	Effet d'une sous-estimation du modèle	71
3.2.2.2	Effet d'une surestimation du modèle	72
<b>3.3</b>	<b>Différentes variances non reconstruites</b>	<b>73</b>
3.3.1	Principe de la reconstruction unidimensionnelle	73
3.3.2	Variance non reconstruite généralisée	74
3.3.3	Comportements des différents critères VNR	76
3.3.3.1	VNR utilisant l'indice $SPE$	76
3.3.3.2	VNR utilisant l'indice $SWE$	76
3.3.3.3	VNR utilisant l'indice $T^2$ de Hotelling	77
3.3.3.4	VNR utilisant l'indice de Mahalanobis	78
3.3.3.5	VNR utilisant un indice exprimé dans le sous-espace principal	79
3.3.3.6	VNR utilisant l'indice combiné	80
<b>3.4</b>	<b>Nouveaux critères VNR</b>	<b>81</b>
3.4.1	VNR utilisant un nouvel indice combiné	81
3.4.2	Changement de représentation des données	83

3.4.2.1	Motivation et théorie d'une nouvelle approche ACPVI	83
3.4.2.2	<b>VNR</b> basée sur l'ACPVI . . . . .	86
3.4.2.3	Consistance théorique du critère <b>VNRVI</b> . . . . .	87
<b>3.5</b>	<b>Résultats de simulation</b> . . . . .	<b>89</b>
<b>3.6</b>	<b>Conclusion</b> . . . . .	<b>93</b>

---

### 3.1 Introduction

Les méthodes de détection et de localisation de défauts en s'appuyant sur l'ACP ont été largement utilisées pour la surveillance de processus. Le principe de la surveillance basée sur l'approche de l'ACP repose principalement sur une modélisation du comportement de processus en fonctionnement normal. Les défauts sont alors détectés en comparant le comportement observé par rapport à celui donné par le modèle ACP. En effet, la phase de détection de défauts est liée à une étape de génération de résidus ou d'indices de détection qui a pour but de générer à partir des mesures observées et d'un modèle ACP, des signaux révélateurs de la présence de défauts. A partir de l'analyse de ces indices, l'étape de détection doit alors indiquer l'existence ou non de défauts.

Dans ce cadre, quelques indices typiques pour la détection des fonctionnements anormaux ont été proposés dans la littérature (Qin, 2003). En revanche, la plupart des méthodes de diagnostic utilisent plus particulièrement l'erreur quadratique de prédiction (*squared prediction error* : *SPE*) et la statistique  $T^2$  de Hotelling qui sont souvent connues par les statistiques Q et D respectivement (Kresta et al., 1991; Kourti et MacGregor, 1995; Dunia et al., 1996; Dunia et Qin, 1998c; Qin, 2003). On note que ces deux indices de détection jouent des rôles différents dans la stratégie de surveillance par ACP. La statistique  $T^2$  décrit le comportement des variables du processus qui sont corrélées avec les composantes principales, tandis que la statistique *SPE* dépend de toutes les variables à surveiller. En outre, celle-ci représente un test global qui cumule les erreurs de modélisation présentes sur chaque résidu (Harkat et al., 2006). L'indice *SPE* est utilisé dans le sous-espace résiduel. Tandis que l'indice  $T^2$  de Hotelling est utilisé dans le sous-espace principal. L'indice  $T_H^2$  de Hawkins (Hawkins, 1974) aussi appelé *SWE* (*squared weighted error*) représente aussi les variations des données dans le sous-espace résiduel. Sa particularité par rapport à l'indice *SPE* se manifeste par une pondération des résidus par les inverses des valeurs propres résiduelles. Néanmoins, d'autres indices sensibles à l'ensemble de l'espace de représentation des données ont été également utilisés comme la distance combinée (Yue et Qin, 2001) et la distance de Mahalanobis. Dans le but d'améliorer les capacités de détection en utilisant la méthode d'ACP, un test basé sur les dernières composantes principales a été proposé par Harkat et al. (2002, 2005, 2006).

Une telle description des indices de détection nous permettra de définir dans ce chapitre de nouveaux critères de sélection du nombre optimal des CPs en se basant sur la variance non reconstruite afin de remédier aux limitations des critères comparés dans le chapitre précédent. En effet, toute procédure d'un diagnostic de défauts repose d'une manière cruciale sur la précision et l'efficacité du critère considéré. Pour cela, nous allons prouver théoriquement l'influence de la modélisation par ACP sur la détectabilité de défauts. En s'appuyant sur le principe de la variance de l'erreur de reconstruction, il s'est avéré possible d'établir une variance non reconstruite associée à chacun de ces indices de détection. Ce résultat nous a permis de proposer un critère empirique relatif à la distance combinée (Mnassri et al., 2010a). Ensuite, un nouveau critère de même type, basé sur une nouvelle statistique combinée représente notre deuxième contribution (Mnassri et al., 2010b). Ces contributions ont été également enrichies par un critère plus performant. Ce dernier est basé sur un changement de représentation de données en envisageant que d'autres données sont beaucoup mieux révélatrices d'informations que les données observées réellement.

Cela nous a permis de proposer une nouvelle théorie reposant sur une ACP de variances inversées (ACPVI) (Mnassri et al., 2010c, 2011b). Ainsi, ce troisième critère de sélection, appelé **VNRVI**, montrera une efficacité importante en remédiant au problème souvent rencontré qui est relatif à la présence des variables indépendantes et quasi-indépendantes. Les résultats de simulation valideront cette nouvelle approche.

## 3.2 Détection et détectabilité de défauts

Dans le cadre de l'ACP, tous les indices de détection disponibles dans la littérature se caractérisent par une forme quadratique (Yue et Qin, 2001; Qin, 2003; Alcalá et Qin, 2009, 2011). Par conséquent, les procédures de détection, isolation et diagnostic de défauts peuvent être généralisées en considérant un indicateur généralisé ou unifié.

Le succès de l'utilisation de l'ACP pour la surveillance de processus a été enrichi par le développement de certains concepts fondamentaux de performance comme la détectabilité de défauts. En effet, la détectabilité représente la capacité d'un indice donné, à détecter la présence de défauts. Dans la littérature, ce concept a été développé plus particulièrement pour l'indice *SPE* et la distance *T2* de Hotelling (Dunia et Qin, 1998b,c,a; Yue et Qin, 2001; Qin, 2003). Pour cela, nous proposons une étude généralisée de détectabilité valable pour tout indice de détection ayant une forme quadratique.

### 3.2.1 Détectabilité généralisée de défauts

Considérons  $\gamma$  et  $\Gamma^2$  respectivement un indice quadratique de détection et sa limite de contrôle.  $\gamma$  peut être n'importe quel indice de détection parmi ceux qui existent dans la littérature (voir tableau 3.1). Mathématiquement,  $\gamma$  représente une distance quadratique qui est égale au carré de la norme euclidienne du vecteur  $\mathbf{x}(k)$  projeté dans un sous-espace vectoriel  $\mathcal{S}_\gamma = \text{span}\{\mathcal{M}^{\frac{1}{2}}\}$  :

$$\gamma(k) = \|\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k)\|^2 = \mathbf{x}^T(k) \mathcal{M} \mathbf{x}(k) \quad (3.1)$$

où  $\mathcal{M}^{\frac{1}{2}}$  est une matrice semi-définie positive.  $k$  est le numéro de l'observation considérée.

En s'appuyant sur les travaux de Box (1954), le seuil de détection de la distance quadratique  $\gamma$  pour un nombre d'observations  $N$  important peut être approximé par une distribution de la forme  $g_\gamma \chi_{(h_\gamma, \alpha)}^2$ , où  $\chi_{(h_\gamma, \alpha)}^2$  est la distribution du  $\chi^2$  avec  $h_\gamma$  degrés de liberté et un seuil de signification  $\alpha$ . On note que le niveau de confiance est égale à  $(1 - \alpha)$ . En se basant sur l'indice  $\gamma$ , le processus est considéré en fonctionnement normal à la  $k$ ème observation si :

$$\gamma(k) \leq \Gamma^2 = g_\gamma \chi_{(h_\gamma, \alpha)}^2 \quad (3.2)$$

Les paramètres  $g_\gamma$  et  $h_\gamma$  peuvent être déterminés comme suit :

$$g_\gamma = \frac{\text{tr}[(\Sigma \mathcal{M})^2]}{\text{tr}[\Sigma \mathcal{M}]} \quad (3.3)$$

Indice de détection	Matrice caractéristique	Limite de contrôle
$\gamma$	$\mathcal{M}^{\frac{1}{2}}$	$\Gamma^2$
$T2$	$\hat{\mathbf{P}}\hat{\Lambda}^{-\frac{1}{2}}\hat{\mathbf{P}}^T$	$\tau^2$
$SPE$	$\tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \tilde{\mathbf{C}}$	$\delta^2$
$SWE$	$\tilde{\mathbf{P}}\tilde{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{P}}^T$	$\epsilon^2$
$\varphi$	$\delta^{-1}\tilde{\mathbf{C}} + \tau^{-1}\hat{\mathbf{P}}\hat{\Lambda}^{-\frac{1}{2}}\hat{\mathbf{P}}^T$	$\beta^2$
$D$	$\mathbf{P}\Lambda^{-\frac{1}{2}}\mathbf{P}^T = \Sigma^{-\frac{1}{2}}$	$\varrho^2$

TABLE 3.1 – Indices de détection

$$h_\gamma = \frac{(\text{tr}[\Sigma \mathcal{M}])^2}{\text{tr}[(\Sigma \mathcal{M})^2]} \quad (3.4)$$

où  $\Sigma$  est la matrice de covariance des données de  $\mathbf{X}$ . L'expression  $\text{tr}[\cdot]$  représente la trace d'une matrice carrée.

Nous rappelons que la présence d'un défaut multiple  $\mathcal{F}_{\mathcal{J}}$  dans les données a été considérée par le vecteur de mesures en défaut donné par l'équation (2.59). A partir de cette dernière et celle de (3.1), le vecteur duquel découle l'indice  $\gamma$  est exprimé par :

$$\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k) = \mathcal{M}^{\frac{1}{2}} (\mathbf{x}^*(k) + \Xi_{\mathcal{J}} \mathbf{f}(k)) = \mathcal{M}^{\frac{1}{2}} \mathbf{x}^*(k) + \mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k) \quad (3.5)$$

Deux conditions nécessaires doivent être considérées afin que le défaut  $\mathcal{F}_{\mathcal{J}}$  soit détectable par l'indice  $\gamma$  :

- i.* Sa projection dans le sous-espace engendré par les vecteurs colonnes de la matrice  $\mathcal{M}^{\frac{1}{2}}$  ne devrait pas être nulle, i.e.  $\|\mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| \neq 0$ ;
- ii.* Son amplitude devrait être suffisamment large afin que l'indice de détection dépasse sa limite de contrôle, i.e.  $\gamma(k) > \Gamma^2$ .

Pour déterminer la condition qui garantit la détection du défaut, nous avons besoin d'exprimer la norme euclidienne du vecteur  $\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k)$  comme suit :

$$\|\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k)\| = \|\mathcal{M}^{\frac{1}{2}} \mathbf{x}^*(k) + \mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| \geq \|\mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| - \|\mathcal{M}^{\frac{1}{2}} \mathbf{x}^*(k)\| \quad (3.6)$$

Puisque  $\mathbf{x}^*$  représente un vecteur de mesures prélevées lors du fonctionnement normal, alors :

$$\|\mathcal{M}^{\frac{1}{2}} \mathbf{x}^*(k)\| \leq \Gamma \quad (3.7)$$

La substitution de (3.7) dans (3.6) en considérant la positivité de la norme euclidienne mène à l'inégalité suivante :

$$\|\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k)\| \geq \|\mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| - \Gamma \geq 0 \quad (3.8)$$

Afin que le défaut soit suffisamment détectable, la contrainte  $\gamma(k) > \Gamma^2$  devrait être satisfaite. On doit alors imposer que

$$\|\mathcal{M}^{\frac{1}{2}} \mathbf{x}(k)\|^2 \geq (\|\mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| - \Gamma)^2 > \Gamma^2 \quad (3.9)$$

La résolution de l'inégalité (3.9) mène au résultat suivant :

$$\|\mathcal{M}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\| > 2\Gamma \quad (3.10)$$

Lorsque l'inégalité précédente est satisfaite, alors la détection du défaut  $\mathcal{F}_{\mathcal{J}}$  est garantie en utilisant l'indice de détection  $\gamma$ . Cette condition suffisante (sur l'amplitude du défaut) est valable pour tout indice de détection ayant une forme quadratique. En s'appuyant sur une telle inégalité, nous allons exprimer dans la suite la détectabilité relative à chaque indice donné dans le tableau 3.1.

### 3.2.1.1 Indice $T2$ de Hotelling

L'indice  $T2$  de Hotelling mesure les variations des projections dans le sous-espace principal pondéré par les inverses des variances des  $\ell$  premières CPs :

$$\begin{aligned} T2(k) &= \|\hat{\Lambda}^{-\frac{1}{2}} \hat{\mathbf{t}}(k)\|^2 = \hat{\mathbf{t}}^T(k) \hat{\Lambda}^{-1} \hat{\mathbf{t}}(k) \\ &= \mathbf{x}^T(k) \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \mathbf{x}(k) = \|\hat{\mathbf{P}} \hat{\Lambda}^{-\frac{1}{2}} \hat{\mathbf{P}}^T \mathbf{x}(k)\|^2 \end{aligned} \quad (3.11)$$

En considérant un tel indice, le processus est supposé sous contrôle à la  $k$ ème observation si :

$$T2(k) \leq \tau^2 = g_{T2} \chi_{(h_{T2}, \alpha)}^2 \quad (3.12)$$

où  $g_{T2}$  et  $h_{T2}$  sont calculés respectivement, à partir de (3.3) et (3.4), comme suit :

$$g_{T2} = \frac{\text{tr}[(\Sigma \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T)^2]}{\text{tr}[\Sigma \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T]} = \frac{\text{tr}[\mathbf{I}_{\ell}]}{\text{tr}[\mathbf{I}_{\ell}]} = 1 \quad (3.13)$$

$$h_{T2} = \frac{(\text{tr}[\Sigma \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T])^2}{\text{tr}[(\Sigma \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T)^2]} = \frac{(\text{tr}[\mathbf{I}_{\ell}])^2}{\text{tr}[\mathbf{I}_{\ell}]} = \text{tr}[\mathbf{I}_{\ell}] = \ell \quad (3.14)$$

Si le nombre d'observations  $N$  est faible, la limite de contrôle de la statistique  $T2$  de Hotelling peut être approximée par la relation suivante :

$$\tau^2 = \frac{\ell(N^2 - 1)}{N(N - \ell)} F_{(\ell, N - \ell, \alpha)} \quad (3.15)$$

où  $F_{(\ell, N - \ell, \alpha)}$  représente la distribution de Fisher avec  $\ell$  et  $(N - \ell)$  degrés de liberté ainsi qu'un seuil de signification  $\alpha$ .

Afin d'assurer la détection du défaut  $\mathcal{F}_{\mathcal{J}}$ , la condition suffisante de sa détectabilité par l'indice  $T2$  de Hotelling en se référant à (3.10) est donnée par :

$$\|\hat{\mathbf{P}} \hat{\Lambda}^{-\frac{1}{2}} \hat{\mathbf{P}}^T \Xi_{\mathcal{J}} \mathbf{f}(k)\| > 2\tau \quad (3.16)$$

Cette condition a été établie par Yue et Qin (2001). D'après (2.46) et (2.47), les mesures collectées sous des conditions de fonctionnement normal se projettent dans le sous-espace principal. Par conséquent, la performance de l'indice  $T2$  de Hotelling dans la détection de

défauts peut être limitée car les variations des projections des défauts dans le sous-espace principal peuvent être masquées par les variations normales. Nous avons montré que si le choix du modèle ACP est optimal, un tel indice est dédié à la détection de défauts des variables indépendantes et quasi-indépendantes (Mnassri et al., 2010a,b, 2011b,a). Sous la contrainte d'optimalité du modèle ACP, les défauts portés par les directions de ces variables se projettent totalement dans le sous-espace principal. Ainsi, leur détection n'est possible qu'avec des indices calculés dans ce sous-espace.

### 3.2.1.2 Indice *SPE*

L'indice *SPE* assure la détection de défauts dans le sous-espace résiduel. Son expression, à l'instant  $k$ , est donnée par :

$$\begin{aligned} SPE(k) &= \|\tilde{\mathbf{x}}(k)\|^2 = \|\tilde{\mathbf{C}}\mathbf{x}(k)\|^2 \\ &= \mathbf{x}^T(k)\tilde{\mathbf{C}}\mathbf{x}(k) = \|\tilde{\mathbf{t}}(k)\|^2 \end{aligned} \quad (3.17)$$

La distance *SPE* est un indicateur global qui somme les résidus sans tenir compte de leurs variances différentes. Toutefois, les résidus avec forte variance portent les erreurs de modélisation produites par l'ACP. Ainsi, ils ont plus d'effets sur la quantité *SPE* que les résidus ayant une faible variance et qui représentent réellement les relations de redondance linéaires ou quasi-linéaires. Par conséquent, l'indice *SPE* est très sensible aux erreurs de modélisation, ce qui peut entraîner de nombreuses fausses alarmes ou l'absence de la sensibilité à la détection de défauts en raison d'un seuil théorique élevé (Tharrault, 2008).

Avec un tel indice, le processus est considéré en fonctionnement normal à la  $k$ ème observation si :

$$SPE(k) \leq \delta^2 = g_{SPE}\chi_{(h_{SPE},\alpha)}^2 \quad (3.18)$$

où  $g_{SPE}$  et  $h_{SPE}$  sont respectivement données en se basant sur les expressions généralisées (3.3) et (3.4) par :

$$g_{SPE} = \frac{\text{tr}[(\Sigma\tilde{\mathbf{C}})^2]}{\text{tr}[\Sigma\tilde{\mathbf{C}}]} = \frac{\text{tr}[\tilde{\Lambda}^2]}{\text{tr}[\tilde{\Lambda}]} = \frac{\sum_{a=\ell+1}^m \lambda_a^2}{\sum_{a=\ell+1}^m \lambda_a} \quad (3.19)$$

$$h_{SPE} = \frac{(\text{tr}[\Sigma\tilde{\mathbf{C}}])^2}{\text{tr}[(\Sigma\tilde{\mathbf{C}})^2]} = \frac{(\text{tr}[\tilde{\Lambda}])^2}{\text{tr}[\tilde{\Lambda}^2]} = \frac{\left(\sum_{a=\ell+1}^m \lambda_a\right)^2}{\sum_{a=\ell+1}^m \lambda_a^2} \quad (3.20)$$

où  $\lambda_a$  représente la  $a$ ème valeur propre de la matrice  $\Sigma$ . En posant  $\theta_i = \sum_{a=\ell+1}^m \lambda_a^i$ , nous retrouvons ainsi la formule connue dans la littérature pour le seuil de contrôle de l'indice *SPE*. Il est important de mentionner qu'il existe également une autre expression proposée par Jackson et Mudholkar (1979) pour le calcul d'une telle limite de contrôle.

Ainsi, [Nomikos et MacGregor \(1995\)](#) ont montré que les deux expressions donnent des résultats identiques.

Le défaut  $\mathcal{F}_{\mathcal{J}}$  est garanti détectable par l'indice  $SPE$  si son amplitude calculée dans le sous-espace résiduel respecte d'après (3.10) l'inégalité suivante :

$$\|\tilde{\mathbf{C}}\Xi_{\mathcal{J}}\mathbf{f}(k)\| > 2\delta \quad (3.21)$$

Nous notons également qu'une telle condition a été proposée dans des travaux antérieurs ([Dunia et Qin, 1998b,c,a](#); [Yue et Qin, 2001](#); [Qin, 2003](#)). Dans la suite, nous allons proposer les conditions suffisantes de détectabilité pour le reste des indices en se basant sur notre étude généralisée développée dans la sous-section 3.2.1.

### 3.2.1.3 Indice $SWE$

L'indice  $T_H^2$  de Hawkins ou  $SWE$  est généralement plus sensible aux défauts ([Westerhuis et al., 2000b](#)). Il peut être défini comme l'indice  $SPE$  calculé avec des CPs résiduelles pondérées. Aussi, son expression est une implémentation symétrique de la statistique  $T^2$  de Hotelling dans le sous-espace résiduel :

$$\begin{aligned} SWE(k) &= \|\tilde{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{t}}(k)\|^2 = \tilde{\mathbf{t}}^T(k)\tilde{\Lambda}^{-1}\tilde{\mathbf{t}}(k) \\ &= \mathbf{x}^T(k)\tilde{\mathbf{P}}\tilde{\Lambda}^{-1}\tilde{\mathbf{P}}^T\mathbf{x}(k) = \|\tilde{\mathbf{P}}\tilde{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{P}}^T\mathbf{x}(k)\|^2 \end{aligned} \quad (3.22)$$

Un problème de conditionnement est considéré comme l'inconvénient d'un tel indice. En effet, le calcul numérique de cet indice peut rencontrer des erreurs lorsque les dernières valeurs propres de la matrice  $\Sigma$  sont très proches de zéro. Néanmoins, ce problème de conditionnement apparaît seulement en absence de bruit de mesures ([Tharrault, 2008](#)).

La présence d'erreurs de modélisation se traduit par une augmentation des variances des projections dans le sous-espace résiduel. Ainsi, les fortes variances peuvent limiter la performance de détection de défauts avec l'indice  $SPE$ . En effet, la pondération avec les valeurs propres de  $\Sigma$  en utilisant l'indice  $SWE$  peut être considérée comme une solution prometteuse. L'indice  $SWE$  est plus robuste pour la détection de défauts que celui du  $SPE$  ([Westerhuis et al., 2000b](#)).

Le comportement du processus à la  $k$ ème observation est considéré normal en se basant sur l'indicateur  $SWE$  si :

$$SWE(k) \leq \epsilon^2 = g_{SWE}\chi_{(h_{SWE},\alpha)}^2 \quad (3.23)$$

En se référant aux formules (3.3) et (3.4), les paramètres  $g_{SWE}$  et  $h_{SWE}$  sont respectivement exprimés comme suit :

$$g_{SWE} = \frac{\text{tr}[(\Sigma\tilde{\mathbf{P}}\tilde{\Lambda}^{-1}\tilde{\mathbf{P}}^T)^2]}{\text{tr}[\Sigma\tilde{\mathbf{P}}\tilde{\Lambda}^{-1}\tilde{\mathbf{P}}^T]} = \frac{\text{tr}[\mathbf{I}_{m-\ell}]}{\text{tr}[\mathbf{I}_{m-\ell}]} = 1 \quad (3.24)$$

$$h_{SWE} = \frac{(\text{tr}[\Sigma\tilde{\mathbf{P}}\tilde{\Lambda}^{-1}\tilde{\mathbf{P}}^T])^2}{\text{tr}[(\Sigma\tilde{\mathbf{P}}\tilde{\Lambda}^{-1}\tilde{\mathbf{P}}^T)^2]} = \frac{(\text{tr}[\mathbf{I}_{m-\ell}])^2}{\text{tr}[\mathbf{I}_{m-\ell}]} = \text{tr}[\mathbf{I}_{m-\ell}] = m - \ell \quad (3.25)$$



où  $\mathbf{I}_{m-\ell} \in \mathbb{R}^{(m-\ell) \times (m-\ell)}$  représente la matrice identité.

Lorsque le nombre d'observations  $N$  est faible, la limite de contrôle de l'indice  $SWE$  est approximée par une distribution de Fisher avec  $(m - \ell)$  et  $(N - m + \ell)$  degrés de liberté et un seuil de signification  $\alpha$  :

$$\epsilon^2 = \frac{(m - \ell)(N^2 - 1)}{N(N - m + \ell)} F_{(m-\ell, N-m+\ell, \alpha)} \quad (3.26)$$

D'après (3.10), le défaut  $\mathcal{F}_{\mathcal{J}}$  est garanti détectable par l'indice  $SWE$  si l'amplitude de sa projection dans le sous-espace résiduel pondéré satisfait l'inégalité suivante :

$$\|\tilde{\mathbf{P}}\tilde{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{P}}^T \Xi_{\mathcal{J}}\mathbf{f}(k)\| > 2\epsilon \quad (3.27)$$

### 3.2.1.4 Indice combiné

D'après leurs expressions, et selon l'étude pratique, les indices  $SPE$  et  $T2$  de Hotelling se comportent d'une manière complémentaire. Il est ainsi possible de combiner ces deux indices en un seul afin de simplifier la tâche de détection de défauts (Yue et Qin, 2001). En réalité, l'idée de la combinaison d'indices a été initialement proposée par Raich et Çinar (1996) :

$$\mathcal{R}(k) = c \frac{SPE(k)}{\delta^2} + (1 - c) \frac{T2(k)}{\tau^2} \quad (3.28)$$

où  $c$  est une constante comprise dans l'intervalle  $[0, 1]$ . Dans ce cadre, les mêmes auteurs suggèrent l'unité comme un seuil de contrôle pour cet indice ce qui a lourdement limité son efficacité. En effet, cet indice peut conduire à des résultats erronés en présence d'un défaut qui peut se manifester dans la possibilité d'avoir  $SPE > \delta^2$  et/ou  $T2 > \tau^2$ , alors que lui-même peut indiquer une situation normale. Pour cette raison, une forme quadratique connue sous le nom de la distance combinée a été proposée par Yue et Qin (2001). Une telle distance combine les statistiques  $SPE$  et  $T2$  de la manière suivante :

$$\begin{aligned} \varphi(k) &= \frac{SPE(k)}{\delta^2} + \frac{T2(k)}{\tau^2} = \mathbf{x}^T(k) \left( \frac{\tilde{\mathbf{C}}}{\delta^2} + \frac{\hat{\mathbf{P}}\hat{\Lambda}^{-1}\hat{\mathbf{P}}^T}{\tau^2} \right) \mathbf{x}(k) \\ &= \|(\delta^{-1}\tilde{\mathbf{C}} + \tau^{-1}\hat{\mathbf{P}}\hat{\Lambda}^{-\frac{1}{2}}\hat{\mathbf{P}}^T)\mathbf{x}(k)\|^2 \end{aligned} \quad (3.29)$$

La matrice caractéristique de cet indice est semi-définie positive. Afin d'utiliser l'indice combiné pour la détection de défauts, sa limite de contrôle est également établie en utilisant la distribution des formes quadratiques. Le processus est considéré en fonctionnement normal à l'instant  $k$  si :

$$\varphi(k) \leq \beta^2 = g_{\varphi} \chi_{(h_{\varphi}, \alpha)}^2 \quad (3.30)$$

D'après (3.3) et (3.4), les paramètres  $g_{\varphi}$  et  $h_{\varphi}$  ont respectivement les expressions suivantes :

$$g_{\varphi} = \frac{\text{tr}[(\Sigma(\delta^{-2}\tilde{\mathbf{C}} + \tau^{-2}\hat{\mathbf{P}}\hat{\Lambda}^{-1}\hat{\mathbf{P}}^T))^2]}{\text{tr}[\Sigma(\delta^{-2}\tilde{\mathbf{C}} + \tau^{-2}\hat{\mathbf{P}}\hat{\Lambda}^{-1}\hat{\mathbf{P}}^T)]} = \frac{\tau^{-4}\ell + \delta^{-4} \sum_{a=\ell+1}^m \lambda_a^2}{\tau^{-2}\ell + \delta^{-2} \sum_{a=\ell+1}^m \lambda_a} \quad (3.31)$$

$$h_\varphi = \frac{(\text{tr}[\Sigma(\delta^{-2}\tilde{\mathbf{C}} + \tau^{-2}\hat{\mathbf{P}}\hat{\Lambda}^{-1}\hat{\mathbf{P}}^T)])^2}{\text{tr}[(\Sigma(\delta^{-2}\tilde{\mathbf{C}} + \tau^{-2}\hat{\mathbf{P}}\hat{\Lambda}^{-1}\hat{\mathbf{P}}^T))^2]} = \frac{\left(\tau^{-2}\ell + \delta^{-2} \sum_{a=\ell+1}^m \lambda_a\right)^2}{\tau^{-4}\ell + \delta^{-4} \sum_{a=\ell+1}^m \lambda_a^2} \quad (3.32)$$

L'indice combiné permet de contrôler le processus dans l'ensemble de l'espace. D'après l'équation (3.10), le défaut  $\mathcal{F}_{\mathcal{J}}$  est garanti détectable si son amplitude exprimée dans l'espace engendré par les vecteurs colonnes de la matrice caractéristique de cet indice respecte l'inégalité suivante :

$$\|(\delta^{-1}\tilde{\mathbf{C}} + \tau^{-1}\hat{\mathbf{P}}\hat{\Lambda}^{-\frac{1}{2}}\hat{\mathbf{P}}^T)\Xi_{\mathcal{J}}\mathbf{f}(k)\| > 2\beta \quad (3.33)$$

### 3.2.1.5 Indice de Mahalanobis

La distance de Mahalanobis correspond à celle de la  $T^2$  de Hotelling étendue sur l'ensemble de l'espace :

$$\begin{aligned} D(k) &= \mathbf{x}^T(k)\mathbf{P}\Lambda^{-1}\mathbf{P}^T\mathbf{x}(k) = \mathbf{x}^T(k)\Sigma^{-1}\mathbf{x}(k) = \|\Sigma^{-\frac{1}{2}}\mathbf{x}(k)\|^2 \\ &= \|\mathbf{P}\Lambda^{-\frac{1}{2}}\mathbf{P}^T\mathbf{x}(k)\|^2 = \|\Lambda^{-\frac{1}{2}}\mathbf{t}(k)\|^2 = \mathbf{t}^T(k)\Lambda^{-1}\mathbf{t}(k) \end{aligned} \quad (3.34)$$

En présence de bruit de mesures, la matrice  $\Sigma$  est généralement de rang plein. Par conséquent, elle est inversible ce qui permet d'éviter le problème de conditionnement confronté également avec l'indice  $SWE$ .

Pour un nombre d'observations  $N$  important, le processus est considéré en fonctionnement normal à l'instant  $k$  en se basant sur la distance de Mahalanobis si :

$$D(k) \leq \varrho^2 = g_D \chi_{(h_D, \alpha)}^2 \quad (3.35)$$

Par analogie avec la limite de contrôle de l'indice  $T^2$  de Hotelling ou celle de  $SWE$ , les coefficients  $g_D$  et  $h_D$  auront respectivement les valeurs 1 et  $m$ . En se référant aux équations (3.3) et (3.4), on peut démontrer très aisément ces résultats :

$$g_D = \frac{\text{tr}[(\Sigma\Sigma^{-1})^2]}{\text{tr}[\Sigma\Sigma^{-1}]} = \frac{\text{tr}[\mathbf{I}_m]}{\text{tr}[\mathbf{I}_m]} = 1 \quad (3.36)$$

$$h_D = \frac{(\text{tr}[\Sigma\Sigma^{-1}])^2}{\text{tr}[(\Sigma\Sigma^{-1})^2]} = \frac{(\text{tr}[\mathbf{I}_m])^2}{\text{tr}[\mathbf{I}_m]} = \text{tr}[\mathbf{I}_m] = m \quad (3.37)$$

Lorsque le nombre d'observations  $N$  est faible, un tel seuil est approximé par une distribution de Fisher avec  $m$  et  $(N - m)$  degrés de liberté ainsi qu'un seuil de signification  $\alpha$  :

$$\varrho^2 = \frac{m(N^2 - 1)}{N(N - m)} F_{(m, N-m, \alpha)} \quad (3.38)$$

A partir de l'inégalité (3.10), la condition suffisante de détectabilité d'un défaut  $\mathcal{F}_{\mathcal{J}}$  basée sur l'utilisation de la distance de Mahalanobis est donnée par :

$$\|\Sigma^{-\frac{1}{2}}\Xi_{\mathcal{J}}\mathbf{f}(k)\| > 2\varrho \quad (3.39)$$

### 3.2.2 Influence de la modélisation sur la détectabilité de défauts

Bien que le thème du choix d'une structure optimale du modèle ACP ait reçu une attention considérable dans la littérature, l'objectif de l'optimalité est négligemment justifié théoriquement. Autrement dit, on n'est pas certain des conséquences d'un choix arbitraire du modèle ACP sur la démarche d'un diagnostic de défauts. Le point clé dans la construction d'un modèle ACP est de choisir un nombre adéquat de CPs afin de représenter le système de façon optimale. Si le nombre retenu est inférieur à celui requis, la représentation du processus sera incomplète. Par contre, si plus de CPs que nécessaires sont sélectionnées, le modèle ACP sera surestimé et contiendra le bruit de mesures (Valle et al., 1999). Néanmoins, ces arguments ainsi que d'autres sont difficiles à prouver mathématiquement.

Il se peut que les résultats obtenus par différents critères de choix d'un modèle ACP se diffèrent de plus ou moins d'une seule CP par rapport au cas optimal. Malgré cette faible différence, les effets peuvent avoir de lourdes conséquences, plus particulièrement dans la qualité de détection de défauts. Pour cela, nous allons montrer que ce choix doit être judicieux et peut être unique.

Puisque  $q$  représente le nombre optimal théorique des CPs, les vecteurs colonnes de la matrice  $\mathcal{M}_q^{\frac{1}{2}}$  engendrent le sous-espace optimal pour l'indice quadratique  $\gamma$ . Notre intérêt est de savoir l'effet d'un changement de la dimension du modèle ACP sur la détectabilité d'un défaut donné. Notamment, le plus intéressant est de se baser sur des hypothèses qui sont vérifiées dans le sous-espace optimal. Ensuite, on vérifie si ces hypothèses sont encore vérifiées ou non. Pour cela, nous supposons que le défaut considéré est garanti détectable par l'indice  $\gamma$  exprimé dans le sous-espace optimal :

$$\|\mathcal{M}_q^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k)\|^2 > 4\Gamma_q^2 \quad (3.40)$$

où  $\Gamma_q^2$  représente la limite de contrôle de l'indice  $\gamma$  dans le sous-espace optimal. En effet, la considération d'un critère quelconque pour le choix du nombre des CPs peut engendrer un modèle ACP qui se diffère par rapport à celui optimal suite à une réduction ou une augmentation du sous-espace optimal de  $\gamma$ . Par conséquent, nos démonstrations se déroulent en deux étapes.

#### 3.2.2.1 Effet d'une sous-estimation du modèle

Puisque les CPs sont orthogonales, une sous-estimation du modèle qui implique une réduction du sous-espace optimal nous permet d'établir la relation suivante :

$$\mathcal{M}_q^{\frac{1}{2}} = \mathcal{M}_{d^-}^{\frac{1}{2}} + \mathcal{M}_{red}^{\frac{1}{2}} \quad (3.41)$$

où  $\mathcal{M}_{d^-}^{\frac{1}{2}}$  représente la matrice dont les vecteurs colonnes engendrent le sous-espace retranché qui correspond aux CPs supprimées. Les vecteurs colonnes de la matrice  $\mathcal{M}_{red}^{\frac{1}{2}}$  engendrent le sous-espace réduit considéré optimal par un critère quelconque. Ainsi, nous montrons que :

$$\mathcal{M}_{d^-}^{\frac{1}{2}} \mathcal{M}_{red}^{\frac{1}{2}} = \mathcal{M}_{red}^{\frac{1}{2}} \mathcal{M}_{d^-}^{\frac{1}{2}} = \mathbf{0}_m \in \mathbb{R}^{m \times m} \quad (3.42)$$

D'après (3.40), nous pouvons déduire ce qui suit :

$$\|\mathcal{M}_q^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 = \|\mathcal{M}_{d^-}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 + \|\mathcal{M}_{red}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 > 4\Gamma_q^2 \quad (3.43)$$

Celle-ci peut nous permettre de définir le domaine de définition de l'amplitude du défaut dans le sous-espace réduit :

$$\|\mathcal{M}_{red}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 > 4\Gamma_q^2 - \|\mathcal{M}_{d^-}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 \quad (3.44)$$

Afin de conserver la qualité de détection du défaut en réduisant le sous-espace optimal, il faut que ce même défaut déjà garanti détectable dans un tel sous-espace demeure garanti détectable dans le sous-espace réduit, i.e.  $\|\mathcal{M}_{red}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\| > 2\Gamma_{red}$ . Pour cela, nous devons imposer ce qui suit :

$$4\Gamma_q^2 - \|\mathcal{M}_{d^-}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 \geq 4\Gamma_{red}^2 \quad (3.45)$$

où  $\Gamma_{red}^2$  représente le seuil de contrôle de l'indice de détection dans le sous-espace réduit. Ainsi, le défaut  $\mathcal{F}_{\mathcal{J}}$  reste garanti détectable suite à une réduction du sous-espace si son amplitude exprimée dans le sous-espace retranché satisfait l'inégalité suivante :

$$\|\mathcal{M}_{d^-}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\| \leq 2\sqrt{\Gamma_q^2 - \Gamma_{red}^2} \quad (3.46)$$

Cette inégalité n'est valable que si  $\Gamma_q^2 \geq \Gamma_{red}^2$  qui est généralement vérifiée. Elle représente une condition suffisante sur l'amplitude du défaut dans le sous-espace retranché afin qu'il reste garanti détectable même en réduisant le sous-espace optimal. Dans cette optique, nous pouvons conclure que le sous-espace retranché doit être insensible au défaut considéré. Ainsi, la majoration d'une telle amplitude prouve qu'une conservation de la qualité de détection de défauts par réduction du sous-espace optimal n'est pas garantie.

### 3.2.2.2 Effet d'une surestimation du modèle

Une surestimation du modèle implique une augmentation du sous-espace optimal par l'addition d'un ensemble de CPs. Cela nous permet d'écrire :

$$\mathcal{M}_{aug}^{\frac{1}{2}} = \mathcal{M}_{d^+}^{\frac{1}{2}} + \mathcal{M}_q^{\frac{1}{2}} \quad (3.47)$$

où  $\mathcal{M}_{d^+}^{\frac{1}{2}}$  et  $\mathcal{M}_{aug}^{\frac{1}{2}}$  représentent deux matrices dont les vecteurs colonnes engendrent respectivement le sous-espace ajouté qui correspond aux CPs insérées et le sous-espace augmenté considéré optimal par un critère donné. L'orthogonalité de ces deux sous-espaces, nous permet d'écrire ce qui suit :

$$\|\mathcal{M}_{aug}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 = \|\mathcal{M}_{d^+}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 + \|\mathcal{M}_q^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 \quad (3.48)$$

Puisque le défaut  $\mathcal{F}_{\mathcal{J}}$  est garanti détectable dans le sous-espace optimal (3.40), l'inégalité suivante est alors vraie :

$$\|\mathcal{M}_{aug}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 > \|\mathcal{M}_{d^+}^{\frac{1}{2}} \Xi_{\mathcal{F}} \mathbf{f}(k)\|^2 + 4\Gamma_q^2 \quad (3.49)$$

Afin de conserver la qualité de détection par addition de CPs supplémentaires, il faut que le même défaut demeure garanti détectable dans le sous-espace augmenté. Par conséquent, il faut s'assurer que :

$$\| \mathcal{M}_{d^+}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k) \|^2 + 4\Gamma_q^2 \geq 4\Gamma_{aug}^2 \quad (3.50)$$

ainsi,

$$\| \mathcal{M}_{d^+}^{\frac{1}{2}} \Xi_{\mathcal{J}} \mathbf{f}(k) \| \geq 2\sqrt{\Gamma_{aug}^2 - \Gamma_q^2} \quad (3.51)$$

où  $\Gamma_{aug}^2$  représente la limite de contrôle de l'indice de détection dans le sous-espace augmenté. Evidemment, cette inégalité n'est valable que si  $\Gamma_{aug}^2 \geq \Gamma_q^2$ . Une telle inégalité représente une condition de minoration sur l'amplitude du défaut dans le sous-espace ajouté. Cette minoration implique qu'un tel sous-espace doit être sensible au défaut considéré. Ainsi, la performance de détection prouvée dans le sous-espace optimal ne peut être garantie conservée par augmentation du sous-espace si les CPs ajoutées sont insensibles au défaut considéré.

Dans le cadre de l'utilisation de l'ACP pour une démarche de diagnostic, le nombre de CPs retenues pour la construction d'un modèle ACP aura certainement un impact sur les étapes d'une telle démarche notamment la détection et la localisation ou l'isolation de défauts. Si le nombre de CPs retenues est inférieur à celui optimal, cela risque de perdre des informations contenues dans les données initiales en projetant certaines variables dans le sous-espace résiduel. Par conséquent, on pourrait avoir des résidus entachés ce qui entraîne des fausses alarmes. Dans le cas contraire, le modèle est surestimé, et il peut ne pas permettre la détection de défauts.

### 3.3 Différentes variances non reconstruites

Le principe de l'approche de reconstruction des variables sera réalisé dans un contexte de défauts unidimensionnels en supposant qu'il ne peut y exister qu'une seule variable en défaut à chaque instant. Puisque les données collectées représentent le fonctionnement normal de processus, les défauts supposés sont réellement nuls. Par conséquent, l'objectif est de définir le modèle ACP permettant d'avoir une plus faible variance non reconstruite.

#### 3.3.1 Principe de la reconstruction unidimensionnelle

Le vecteur de données exprimant la présence d'un défaut simple  $\mathcal{F}_j$  affectant la  $j$ ème variable au  $k$ ème instant peut être déduit de l'équation (2.59) en remplaçant respectivement  $\Xi_{\mathcal{J}}$  et  $\mathbf{f}(k)$  par  $\xi_j$  et  $f(k)$ . L'élimination de l'effet d'un tel défaut revient à estimer un vecteur  $\mathbf{x}_j(k)$  insensible au défaut :

$$\begin{aligned} \mathbf{x}_j(k) &= \mathbf{x}(k) - \xi_j f(k) \\ &= \mathbf{x}(k) - \xi_j \hat{f}_j(k) \end{aligned} \quad (3.52)$$

où  $\hat{f}_j \in \mathbb{R}$  représente une estimation de  $f$ . Celle-ci permettra de calculer un indice de détection associé non influencé par le défaut :

$$\gamma_j(k) = \| \mathcal{M}^{\frac{1}{2}} \mathbf{x}_j(k) \|^2 \quad (3.53)$$

Ainsi, une estimation optimale de l'amplitude du défaut est celle qui minimise l'indice reconstruit  $\gamma_j(k)$  comme suit :

$$\hat{f}_j(k) = \arg \min_{f(k)} \{\gamma_j(k)\} \quad (3.54)$$

Cette minimisation est obtenue par l'application du principe des moindres carrés. En conséquence, l'amplitude estimée du défaut ainsi que le vecteur de données dont la  $j^{\text{ème}}$  variable est reconstruite sont respectivement exprimés par :

$$\hat{f}_j(k) = (\xi_j^T \mathcal{M} \xi_j)^{-1} \xi_j^T \mathcal{M} \mathbf{x}(k) \quad (3.55)$$

$$\mathbf{x}_j(k) = (\mathbf{I}_m - \xi_j (\xi_j^T \mathcal{M} \xi_j)^{-1} \xi_j^T \mathcal{M}) \mathbf{x}(k) \quad (3.56)$$

On peut déduire que la  $j^{\text{ème}}$  variable n'est reconstituée dans le sous-espace de  $\gamma$  que si le terme  $\xi_j^T \mathcal{M} \xi_j$  est non nul. Cela implique que le vecteur  $\mathcal{M}^{\frac{1}{2}} \xi_j$  ne soit pas nul.

Pour une meilleure mise en évidence du principe de reconstruction d'une variable, nous proposons de décomposer les vecteurs  $\mathbf{x}(k)$ ,  $\xi_j$  et la matrice  $\mathcal{M}$  comme suit :

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_{j-}(k) \\ x_j(k) \\ \mathbf{x}_{j+}(k) \end{bmatrix}, \quad \xi_j = \begin{bmatrix} \xi_{j-} \\ 1 \\ \xi_{j+} \end{bmatrix} \quad \text{et} \quad \mathcal{M} = \begin{bmatrix} \mathcal{M}_1 & \mathcal{M}_{j-} & \mathcal{M}_2 \\ \mathcal{M}_{j-}^T & \xi_j^T \mathcal{M} \xi_j & \mathcal{M}_{j+}^T \\ \mathcal{M}_2^T & \mathcal{M}_{j+} & \mathcal{M}_3 \end{bmatrix} \quad (3.57)$$

où  $\mathbf{x}_{j-}(k)$  et  $\mathbf{x}_{j+}(k)$  sont respectivement composés des  $(j-1)$  premières et  $(m-j)$  dernières mesures du vecteur de données  $\mathbf{x}(k)$ .  $x_j(k)$  représente la mesure de la  $j^{\text{ème}}$  variable au  $k^{\text{ème}}$  instant.  $\xi_{j-} \in \mathbb{R}^{j-1}$  et  $\xi_{j+} \in \mathbb{R}^{m-j}$  sont deux vecteurs nuls.  $\mathcal{M}_{j-}^T$  et  $\mathcal{M}_{j+}^T$  sont deux vecteurs composés respectivement des  $(j-1)$  premières et  $(m-j)$  dernières valeurs de la  $j^{\text{ème}}$  ligne de la matrice  $\mathcal{M}$ .  $\xi_j^T \mathcal{M} \xi_j$  est la  $j^{\text{ème}}$  valeur diagonale de la matrice  $\mathcal{M}$ .

A partir des équations (3.56) et (3.57), le vecteur de données dont la  $j^{\text{ème}}$  variable est reconstruite peut s'écrire de la façon suivante :

$$\mathbf{x}_j(k) = \begin{bmatrix} \mathbf{x}_{j-}(k) \\ -(\xi_j^T \mathcal{M} \xi_j)^{-1} (\mathcal{M}_{j-}^T \mathbf{x}_{j-}(k) + \mathcal{M}_{j+}^T \mathbf{x}_{j+}(k)) \\ \mathbf{x}_{j+}(k) \end{bmatrix} \quad (3.58)$$

D'après cette expression, seules les mesures des variables autres que celles de la variable en question sont utilisées pour sa reconstruction. En outre, la contribution des autres variables dans la reconstruction dépend de la dimension du modèle ACP considéré. Une telle contribution peut être illustrée par les coefficients des vecteurs  $\mathcal{M}_{j-}^T$  et  $\mathcal{M}_{j+}^T$ . Evidemment, les valeurs de ces coefficients changent en fonction du nombre des CPs utilisées dans le modèle ACP.

### 3.3.2 Variance non reconstruite généralisée

La reconstruction d'une variable dans un sous-espace donné est basée sur l'estimation de sa grandeur supposée être normale en utilisant les mesures des autres variables.

Néanmoins, la variance d'une telle variable ne peut être totalement reconstruite. Il est donc possible d'en extraire une variance non reconstruite qui dépend de la dimension du modèle ACP et du sous-espace dans lequel l'estimation est réalisée. Nous rappelons que seule la variance non reconstruite relative à une estimation dans le sous-espace résiduel a été étudiée dans la littérature (Dunia et Qin, 1998b,c,a; Qin et Dunia, 2000). Dans ce cadre, nous avons proposé une variance non reconstruite généralisée relative à un indice de détection quadratique quelconque (Mnassri et al., 2010a). Cette généralisation nous a permis d'étudier la variance de l'erreur de reconstruction relative à chaque indice de détection donné dans le tableau 3.1.

Considérons  $\mathbf{e}_j(k) \in \mathbb{R}^m$  le vecteur qui représente l'erreur de reconstruction de la  $j$ ème variable au  $k$ ème instant. Ainsi, l'erreur de reconstruction de cette variable est donnée par :

$$\xi_j^T \mathbf{e}_j(k) = \xi_j^T (\mathbf{x}(k) - \hat{\mathbf{x}}_j(k)) = \hat{f}_j(k) = (\xi_j^T \mathcal{M} \xi_j)^{-1} \xi_j^T \mathcal{M} \mathbf{x}(k) \quad (3.59)$$

Cette expression montre que l'estimation de la  $j$ ème variable dans n'importe quel sous-espace est non biaisée. Puisque les données sont centrées, la moyenne de l'erreur de reconstruction est nulle :

$$\mathbb{E} \{ \xi_j^T \mathbf{e}_j \} = \mathbb{E} \{ \hat{f}_j \} = (\xi_j^T \mathcal{M} \xi_j)^{-1} \xi_j^T \mathcal{M} \mathbb{E} \{ \mathbf{x} \} = 0 \quad (3.60)$$

La variance d'une telle erreur dans la direction de la variable en question dépend de la dimension du modèle ACP. Etant donné que les données utilisées dans l'objectif d'une modélisation par ACP sont censées être sans défauts, cela implique que le modèle ACP le plus adapté doit assurer une erreur de reconstruction ayant le moins possible de variance. Ainsi, on définit la variance de l'erreur de reconstruction de la  $j$ ème variable comme suit :

$$\begin{aligned} \sigma_{\gamma_j}^2(\ell) &= \text{Var} \{ \xi_j^T \mathbf{e}_j \} = \mathbb{E} \left\{ \left( \hat{f}_j - \mathbb{E} \{ \hat{f}_j \} \right)^2 \right\} = \mathbb{E} \{ \hat{f}_j^2 \} \\ &= \frac{\xi_j^T \mathcal{M} \mathbb{E} \{ \mathbf{x} \mathbf{x}^T \} \mathcal{M} \xi_j}{(\xi_j^T \mathcal{M} \xi_j)^2} = \frac{\xi_j^T \mathcal{M} \Sigma \mathcal{M} \xi_j}{(\xi_j^T \mathcal{M} \xi_j)^2} \end{aligned} \quad (3.61)$$

L'objectif étant alors de définir le nombre de CPs qui minimise l'expression précédente pour une meilleure reconstruction d'une variable donnée. En effet, le minimum de la variance non reconstruite d'une variable correspond à un nombre de CPs qui peut être différent de celui obtenu pour une reconstruction optimale d'une autre variable. Pour cette raison, ce compromis peut être résolu par la considération d'une variance non reconstruite globale. Ainsi, l'objectif sera plutôt de définir un nombre de CPs qui minimise la variance globale qui peut malheureusement ne pas assurer une variance non reconstruite minimale pour chacune des variables. Puisque le critère global représente la somme des variances non reconstruites individuelles des variables, il est préférable de les considérer dans la même échelle en pondérant chacune par la variance originelle de sa variable. Par conséquent, le critère de la variance non reconstruite globale relative à l'indice  $\gamma$  est donné par :

$$\mathbf{VNR}_\gamma(\ell) = \sum_{j=1}^m \frac{\sigma_{\gamma_j}^2(\ell)}{\xi_j^T \Sigma \xi_j} = \sum_{j=1}^m \frac{\xi_j^T \mathcal{M} \Sigma \mathcal{M} \xi_j}{(\xi_j^T \Sigma \xi_j)(\xi_j^T \mathcal{M} \xi_j)^2} \quad (3.62)$$

Dans le cadre d'une utilisation de la variance non reconstruite, nous considérons plus particulièrement que les données de  $\mathbf{X}$  sont centrées réduites. Par conséquent,  $\Sigma$  représente une matrice de corrélation, ce qui implique que  $\xi_j^T \Sigma \xi_j = 1$ .

### 3.3.3 Comportements des différents critères VNR

A partir de l'expression généralisée de la variance non reconstruite, nous développons et étudions théoriquement dans la suite les comportements des différents critères envisageables qui sont relatifs aux indices de détection utilisés dans le cadre de l'ACP.

#### 3.3.3.1 VNR utilisant l'indice *SPE*

En se référant au tableau 3.1, la matrice de l'indice *SPE* est  $\mathcal{M} = \tilde{\mathbf{C}}$ . La substitution de celle-ci dans la formule de la variance non reconstruite généralisée (3.61), nous permet de prouver en considérant l'équation (2.63) ce qui suit :

$$\sigma_{SPE_j}^2(\ell) = \frac{\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j}{(\xi_j^T \tilde{\mathbf{C}} \xi_j)^2} = \sigma_j^2(\ell) \quad (3.63)$$

Nous pouvons déduire que la variance non reconstruite d'une variable donnée en utilisant l'indice *SPE* est égale à celle étudiée dans le deuxième chapitre. Par conséquent, le critère de la variance globale de l'erreur de reconstruction relative à un tel indice n'est autre que le critère **VNR** qui a été l'objet d'une étude dans le chapitre précédent :

$$\mathbf{VNR}_{SPE}(\ell) = \sum_{j=1}^m \sigma_{SPE_j}^2(\ell) = \sum_{j=1}^m \sigma_j^2(\ell) = \mathbf{VNR}(\ell) \quad (3.64)$$

L'expression précédente représente l'unique critère proposé dans un contexte de choix d'une dimension optimale d'un modèle ACP en s'appuyant sur l'idée de la variance non reconstruite (Dunia et Qin, 1998b,c,a; Qin et Dunia, 2000).

#### 3.3.3.2 VNR utilisant l'indice *SWE*

En remplaçant la matrice  $\mathcal{M}$  exprimée dans l'équation (3.61) par  $\tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T$  qui représente la matrice de l'indice *SWE*, la variance de l'erreur de reconstruction de la *j*ème variable en utilisant un tel indice est donnée par :

$$\sigma_{SWE_j}^2(\ell) = \frac{\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \mathbf{P} \Lambda \mathbf{P}^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j}{(\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j)^2} = \frac{1}{\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j} \quad (3.65)$$

En effet, la matrice résiduelle des vecteurs propres ainsi que celle des valeurs propres peuvent être réécrites respectivement comme suit :

$$\tilde{\mathbf{P}} = \left[ \mathbf{p}_{\ell+1}, \tilde{\mathbf{P}}_r \right] \quad (3.66)$$



et

$$\tilde{\Lambda} = \begin{bmatrix} \lambda_{\ell+1} & 0 \\ 0 & \tilde{\Lambda}_r \end{bmatrix} \quad (3.67)$$

On peut alors déduire que :

$$\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j = \lambda_{\ell+1}^{-1} \xi_j^T \mathbf{P}_{\ell+1} \mathbf{P}_{\ell+1}^T \xi_j + \xi_j^T \tilde{\mathbf{P}}_r \tilde{\Lambda}_r^{-1} \tilde{\mathbf{P}}_r^T \xi_j \quad (3.68)$$

Puisque  $\lambda_{\ell+1}^{-1} \xi_j^T \mathbf{P}_{\ell+1} \mathbf{P}_{\ell+1}^T \xi_j \geq 0$ , alors

$$\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j \geq \xi_j^T \tilde{\mathbf{P}}_r \tilde{\Lambda}_r^{-1} \tilde{\mathbf{P}}_r^T \xi_j \quad (3.69)$$

donc

$$\frac{1}{\xi_j^T \tilde{\mathbf{P}} \tilde{\Lambda}^{-1} \tilde{\mathbf{P}}^T \xi_j} \leq \frac{1}{\xi_j^T \tilde{\mathbf{P}}_r \tilde{\Lambda}_r^{-1} \tilde{\mathbf{P}}_r^T \xi_j} \quad (3.70)$$

D'après l'inégalité précédente et l'équation (3.65), on peut établir ce qui suit :

$$\sigma_{SWE_j}^2(\ell) \leq \sigma_{SWE_j}^2(\ell + 1) \quad (3.71)$$

L'inégalité (3.71) prouve que la variance non reconstruite d'une variable donnée en utilisant l'indice  $SWE$  est monotone croissante en  $\ell$ . Ainsi, la variance globale de l'erreur de reconstruction est également monotone croissante en  $\ell$  :

$$\sum_{j=1}^m \sigma_{SWE_j}^2(\ell) \leq \sum_{j=1}^m \sigma_{SWE_j}^2(\ell + 1) \quad (3.72)$$

d'où

$$\mathbf{VNR}_{SWE}(\ell) \leq \mathbf{VNR}_{SWE}(\ell + 1) \quad (3.73)$$

Par conséquent, le minimum d'un tel critère correspond toujours à une seule CP qui est la première :

$$\min_{\ell} \mathbf{VNR}_{SWE}(\ell) = 1 \quad (3.74)$$

On peut conclure que la variance non reconstruite utilisant l'indice  $SWE$  ne peut pas servir dans le choix d'un nombre optimal de CPs.

### 3.3.3.3 VNR utilisant l'indice $T2$ de Hotelling

La reconstruction de la  $j^{\text{ème}}$  variable en se basant sur l'indice  $T2$  de Hotelling génère une variance non reconstruite dont l'expression est obtenue en substituant dans l'équation (3.61) et d'après le tableau 3.1 la matrice  $\mathcal{M}$  par  $\hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T$  comme suit :

$$\sigma_{T2_j}^2(\ell) = \frac{\xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \mathbf{P} \Lambda \mathbf{P}^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j}{(\xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j)^2} = \frac{1}{\xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j} \quad (3.75)$$

Notamment, l'augmentation de la dimension du sous-espace principal par l'addition d'une CP supplémentaire implique que la nouvelle matrice des vecteurs propres ainsi que celle des valeurs propres prennent respectivement les formes suivantes :

$$\hat{\mathbf{P}}_+ = \begin{bmatrix} \hat{\mathbf{P}} & \mathbf{p}_{\ell+1} \end{bmatrix} \quad (3.76)$$

$$\hat{\Lambda}_+ = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \lambda_{\ell+1} \end{bmatrix} \quad (3.77)$$

Ce qui nous permet de déduire que :

$$\sigma_{T_2}^2(\ell+1) = \frac{1}{\xi_j^T \hat{\mathbf{P}}_+ \hat{\Lambda}_+^{-1} \hat{\mathbf{P}}_+^T \xi_j} = \frac{1}{\lambda_{\ell+1}^{-1} \xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j + \xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j} \quad (3.78)$$

Puisque  $\lambda_{\ell+1}^{-1} \xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j \geq 0$ , alors

$$\frac{1}{\xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j} \geq \frac{1}{\xi_j^T \hat{\mathbf{P}}_+ \hat{\Lambda}_+^{-1} \hat{\mathbf{P}}_+^T \xi_j} \quad (3.79)$$

donc

$$\sigma_{T_2}^2(\ell) \geq \sigma_{T_2}^2(\ell+1) \quad (3.80)$$

et

$$\sum_{j=1}^m \sigma_{T_2}^2(\ell) \geq \sum_{j=1}^m \sigma_{T_2}^2(\ell+1) \quad (3.81)$$

Celle-ci implique que

$$\mathbf{VNR}_{T_2}(\ell) \geq \mathbf{VNR}_{T_2}(\ell+1) \quad (3.82)$$

Cette inégalité prouve que le critère de la variance non reconstruite utilisant à l'indice  $T_2$  de Hotelling est monotone décroissant en  $\ell$ . En se basant sur ce critère, la meilleure reconstruction est obtenue en considérant toutes les CPs puisque :

$$\min_{\ell} \mathbf{VNR}_{T_2}(\ell) = m \quad (3.83)$$

En effet, la variance de l'erreur de reconstruction basée sur l'indice  $T_2$  de Hotelling est incapable de déterminer un modèle optimal.

### 3.3.3.4 VNR utilisant l'indice de Mahalanobis

D'après le tableau 3.1, la matrice de la distance de Mahalanobis est indépendante du nombre des CPs. Par conséquent, la variance non reconstruite utilisant une telle distance ne dépend pas également d'une dimension donnée :

$$\sigma_{D_j}^2 = \frac{\xi_j^T \mathbf{P} \Lambda^{-1} \mathbf{P}^T \mathbf{P} \Lambda \mathbf{P}^T \mathbf{P} \Lambda^{-1} \mathbf{P}^T \xi_j}{(\xi_j^T \mathbf{P} \Lambda^{-1} \mathbf{P}^T \xi_j)^2} = \frac{1}{\xi_j^T \mathbf{P} \Lambda^{-1} \mathbf{P}^T \xi_j} = \frac{1}{\xi_j^T \Sigma^{-1} \xi_j} \quad (3.84)$$

ainsi,

$$\mathbf{VNR}_D = \sum_{j=1}^m \sigma_{D_j}^2 \quad (3.85)$$

Il est clair que ce critère est constant. Ainsi, sa minimisation par rapport à  $\ell$  n'est pas possible et ne peut apporter aucun avantage pour le choix d'un nombre optimal de CPs.

### 3.3.3.5 VNR utilisant un indice exprimé dans le sous-espace principal

Nous proposons ici d'étudier la variance non reconstruite en utilisant un indice exprimé dans le sous-espace principal autre que les indices de détection connus dans le cadre de l'ACP. Cet indice est complémentaire à celui du *SPE*.

En effet, nous remarquons dans la littérature l'inexistence d'intérêt à l'étude de la distance  $\|\hat{\mathbf{x}}\|^2$  malgré qu'elle représente une forme quadratique. La reconstruction d'une variable donnée en se basant sur une telle distance produit une variance non reconstruite. En substituant la matrice  $\mathcal{M}$  dans (3.61) par  $\hat{\mathbf{C}}$ , l'expression de la variance de l'erreur de reconstruction de la *j*ème variable est la suivante :

$$\sigma_{\|\hat{\mathbf{x}}_j\|^2}^2(\ell) = \frac{\xi_j^T \hat{\mathbf{C}} \Sigma \hat{\mathbf{C}} \xi_j}{(\xi_j^T \hat{\mathbf{C}} \xi_j)^2} = \left( \frac{\xi_j^T \hat{\mathbf{C}} \Sigma \hat{\mathbf{C}} \xi_j}{\xi_j^T \hat{\mathbf{C}} \xi_j} \right) \left( \frac{1}{\xi_j^T \hat{\mathbf{C}} \xi_j} \right) = \hat{u}_j(\ell) \hat{v}_j(\ell) \quad (3.86)$$

avec

$$\hat{u}_j(\ell) = \frac{\xi_j^T \hat{\mathbf{C}} \Sigma \hat{\mathbf{C}} \xi_j}{\xi_j^T \hat{\mathbf{C}} \xi_j} \quad (3.87)$$

et

$$\hat{v}_j(\ell) = \frac{1}{\xi_j^T \hat{\mathbf{C}} \xi_j} \quad (3.88)$$

L'ajout d'une CP supplémentaire dans le sous-espace principal nous permet d'écrire :

$$\hat{v}_j(\ell + 1) = \frac{1}{\xi_j^T \hat{\mathbf{C}}_+ \xi_j} = \frac{1}{\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j} \quad (3.89)$$

D'après l'équation (3.76), on peut déduire que :

$$\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j = \xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j + \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j \quad (3.90)$$

Celle-ci implique que

$$\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j \geq \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j \quad (3.91)$$

ainsi,

$$\hat{v}_j(\ell) \geq \hat{v}_j(\ell + 1) \quad (3.92)$$

L'inégalité précédente prouve que la fonction  $\hat{v}_j$  est monotone décroissante en  $\ell$ . En ce qui concerne la fonction  $\hat{u}_j$ , nous proposons de calculer la différence suivante :

$$\begin{aligned} \hat{u}_j(\ell + 1) - \hat{u}_j(\ell) &= \frac{\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \Sigma \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j}{\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j} - \frac{\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \Sigma \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j}{\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j} \\ &= \frac{\lambda_{\ell+1} \xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j + \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j}{\xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j + \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j} - \frac{\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j}{\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j} \\ &= \frac{\xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j \left( \lambda_{\ell+1} \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j - \xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j \right)}{(\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j) (\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j)} \\ &= \frac{\xi_j^T \mathbf{p}_{\ell+1} \mathbf{p}_{\ell+1}^T \xi_j}{(\xi_j^T \hat{\mathbf{P}}_+ \hat{\mathbf{P}}_+^T \xi_j) (\xi_j^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \xi_j)} \left( \xi_j^T \hat{\mathbf{P}} (\lambda_{\ell+1} \mathbf{I}_\ell - \hat{\Lambda}) \hat{\mathbf{P}}^T \xi_j \right) \quad (3.93) \end{aligned}$$

où  $\mathbf{I}_\ell \in \mathbb{R}^{\ell \times \ell}$  est une matrice identité. Notamment, toutes les valeurs propres de la matrice diagonale  $\hat{\Lambda}$  sont supérieures ou égales à  $\lambda_{\ell+1}$  ce qui implique que tous les éléments de la matrice diagonale  $(\lambda_{\ell+1}\mathbf{I}_\ell - \hat{\Lambda})$  sont alors négatifs ou nuls. Par conséquent, la différence exprimée par l'équation précédente est négative ou nulle, ainsi :

$$\hat{u}_j(\ell) \geq \hat{u}_j(\ell + 1) \quad (3.94)$$

A partir des inégalités (3.94) et (3.92), nous pouvons déduire que :

$$\hat{u}_j(\ell)\hat{v}_j(\ell) \geq \hat{u}_j(\ell+1)\hat{v}_j(\ell + 1) \quad (3.95)$$

ce qui nous permet également de déduire, d'après (3.86), que :

$$\sigma_{\|\hat{\mathbf{x}}_j\|^2}^2(\ell) \geq \sigma_{\|\hat{\mathbf{x}}_j\|^2}^2(\ell + 1) \quad (3.96)$$

En posant

$$\mathbf{VNR}_{\|\hat{\mathbf{x}}\|^2}(\ell) = \sum_{j=1}^m \sigma_{\|\hat{\mathbf{x}}_j\|^2}^2(\ell) \quad (3.97)$$

le critère qui représente la variance globale non reconstruite relative à l'indice considéré, on peut conclure de l'inégalité (3.96) que :

$$\mathbf{VNR}_{\|\hat{\mathbf{x}}\|^2}(\ell) \geq \mathbf{VNR}_{\|\hat{\mathbf{x}}\|^2}(\ell + 1) \quad (3.98)$$

Ce critère est alors monotone décroissant en  $\ell$ . Evidemment, la meilleure reconstruction est obtenue en considérant toutes les CPs dans le modèle ACP puisque :

$$\min_{\ell} \mathbf{VNR}_{\|\hat{\mathbf{x}}\|^2}(\ell) = m \quad (3.99)$$

Le comportement de cette variance non reconstruite ne peut pas servir pour le choix d'un nombre optimal de CPs.

### 3.3.3.6 VNR utilisant l'indice combiné

D'après le tableau 3.1, les vecteurs colonnes de la matrice  $\mathcal{M}$  de l'indice combiné engendrent deux sous-espaces orthogonaux. La substitution d'une telle matrice dans l'expression (3.61), nous a permis de montrer que la variance non reconstruite de la  $j$ ème variable en s'appuyant sur cet indice peut s'écrire comme suit :

$$\sigma_{\varphi_j}^2(\ell, \alpha) = \frac{\tau^4 \xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j + \delta^4 \xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j}{\left( \tau^2 \xi_j^T \tilde{\mathbf{C}} \xi_j + \delta^2 \xi_j^T \hat{\mathbf{P}} \hat{\Lambda}^{-1} \hat{\mathbf{P}}^T \xi_j \right)^2} \quad (3.100)$$

Cette expression dépend du nombre  $\ell$  des CPs retenus. En outre, nous remarquons qu'une telle variance dépend également des seuils de contrôle des indices  $SPE$  et  $T2$  de Hotelling. Ces limites sont en fonction de  $\ell$  et un seuil de signification  $\alpha$ . Par conséquent, la variance non reconstruite globale dépendra également des ces paramètres :

$$\mathbf{VNR}_{\varphi}(\ell, \alpha) = \sum_{j=1}^m \sigma_{\varphi_j}^2(\ell, \alpha) \quad (3.101)$$

Cette équation présente un nouveau critère basé sur la variance non reconstruite relative à l'indice combiné. Nous notons qu'il n'a pas été étudié auparavant dans la littérature.

Jusqu'à ce stade, nous avons montré théoriquement que tous les critères qui se basent sur la variance non reconstruite, à l'exception de celui de l'équation (3.101), ne peuvent pas définir la dimension adaptée d'un modèle ACP voire la plupart de ces critères sont inutiles pour effectuer une telle tâche. En revanche, l'unique critère dont nous ignorons son comportement est celui relatif à l'indice combiné. Pour cette raison, ce critère a été l'objet d'une étude sur un exemple simulé par Mnassri et al. (2010a). Nous avons remarqué qu'il peut nous renseigner sur la dimension optimale en s'appuyant sur à un choix très approprié du niveau de confiance  $(1 - \alpha)$ .

Afin d'assurer la détection de défauts, le seuil de signification  $\alpha$  doit généralement avoir des valeurs voisines de 5%. Malheureusement, ces valeurs sont inadéquates pour le nouveau critère. En effet, nous avons constaté que  $\alpha$  doit être proche de 80% pour que le minimum de  $\mathbf{VNR}_\varphi$  converge modérément vers le nombre désiré des CPs. Ce critère est conçu pour un objectif primordial qui est la minimisation de la variance de l'erreur de reconstruction. Toutefois, il est difficile d'expliquer théoriquement son comportement. Pour cela, nous pouvons le considérer comme un critère empirique puisque son efficacité s'appuie plus particulièrement sur l'expérience.

## 3.4 Nouveaux critères **VNR**

Dans l'objectif d'une sélection correcte de la dimension optimale d'un modèle ACP, nous contribuons dans ce chapitre par deux autres nouveaux critères qui se basent sur la théorie de la variance de l'erreur de reconstruction.

### 3.4.1 **VNR** utilisant un nouvel indice combiné

En se basant sur l'étude établie par Mnassri et al. (2010a) pour le critère  $\mathbf{VNR}_\varphi$ , nous avons essayé de remédier à l'inconvénient de celui-ci par la proposition d'un nouveau critère plus consistant en utilisant une nouvelle distance combinée (Mnassri et al., 2010b). Une caractéristique majeure constatée dans le premier critère réside dans son allure fluctuante face aux valeurs du niveau de confiance  $(1 - \alpha)$ . Notre objectif prioritaire est d'assurer par le nouveau critère une robustesse par rapport au choix de la valeur du seuil de signification au moins dans l'intervalle couramment utilisé comme  $[1\%, 5\%]$ .

D'après (3.100), nous constatons que le critère  $\mathbf{VNR}_\varphi$  est exprimé en fonction de l'inverse des valeurs propres. En effet, celles-ci ont une influence importante sur le comportement de ce critère. Dans ce cadre, Mnassri et al. (2010b) ont proposé un nouvel indice combiné qui combine la statistique  $SPE$  et une nouvelle distance  $\mathfrak{B}$  exprimée dans le sous-espace principal. La particularité de l'indice  $\mathfrak{B}$  se manifeste dans une matrice des valeurs propres modifiées :

$$\mathfrak{B}(k) = \|\hat{\mathbf{P}}\hat{\Lambda}_{\mathfrak{B}}^{-\frac{1}{2}}\hat{\mathbf{P}}^T \mathbf{x}(k)\|^2 = \mathbf{x}^T(k)\hat{\mathbf{P}}\hat{\Lambda}_{\mathfrak{B}}^{-1}\hat{\mathbf{P}}^T \mathbf{x}(k) \quad (3.102)$$

avec

$$\Lambda_{\mathfrak{B}} = \begin{bmatrix} \Lambda_v^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(m-v)} + \Lambda_{(m-v)} \end{bmatrix} \quad (3.103)$$

où  $\hat{\Lambda}_{\mathfrak{B}} \in \mathbb{R}^{\ell \times \ell}$  est une matrice diagonale constituée des  $\ell$  premiers éléments de la matrice diagonale  $\Lambda_{\mathfrak{B}}$ . Evidemment,  $\ell$  représente le nombre des CPs à retenir.  $\mathbf{I}_{(m-v)}$  est une matrice identité d'ordre  $(m-v)$ .  $v = \{1, \dots, m\}$  joue le rôle d'un deuxième paramètre pour le critère proposé en représentant également un nombre de CPs qui servira dans la minimisation par la suite.  $\Lambda_v \in \mathbb{R}^{v \times v}$  et  $\Lambda_{(m-v)} \in \mathbb{R}^{(m-v) \times (m-v)}$  sont deux matrices diagonales contenant respectivement les  $v$  premières et  $(m-v)$  dernières valeurs propres de la matrice  $\Lambda$ .

Nous constatons que cet indice ressemble dans sa forme générale à celui de la  $T^2$  de Hotelling. La modification est réalisée aux niveaux des valeurs propres de la matrice  $\Lambda$ . L'utilisation de l'indice  $\mathfrak{B}$  pour la détection de défauts impose qu'on lui propose une limite de contrôle. Puisque cet indice présente une forme quadratique, son seuil de détection peut être déterminé en se référant à la théorie de [Box \(1954\)](#). En considérant un tel indice, le processus est en fonctionnement normal au  $k$ ème instant si :

$$\mathfrak{B}(k) \leq \mathfrak{b}^2 = g_{\mathfrak{B}} \chi_{(h_{\mathfrak{B}}, \alpha)}^2 \quad (3.104)$$

avec

$$g_{\mathfrak{B}} = \frac{\text{tr}[(\Sigma \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T)^2]}{\text{tr}[\Sigma \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T]} \quad (3.105)$$

$$h_{\mathfrak{B}} = \frac{(\text{tr}[\Sigma \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T])^2}{\text{tr}[(\Sigma \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T)^2]} \quad (3.106)$$

Dans la pratique, on suggère généralement qu'on surveille le processus par un seul indice plutôt que deux statistiques de détection séparées. Pour cela, nous avons proposé de combiner l'indice  $\mathfrak{B}$  et celui du  $SPE$  en une seule statistique combinée comme suit :

$$\Psi(k) = \|\Upsilon^{\frac{1}{2}} \mathbf{x}(k)\|^2 = \frac{SPE(k)}{\delta^2} + \frac{\mathfrak{B}(k)}{\mathfrak{b}^2} \quad (3.107)$$

où

$$\Upsilon^{\frac{1}{2}} = \delta^{-1} \tilde{\mathbf{C}} + \mathfrak{b}^{-1} \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-\frac{1}{2}} \hat{\mathbf{P}}^T \quad (3.108)$$

$\Upsilon$  représente la matrice du nouvel indice combiné  $\Psi$ . Notamment, la reconstruction de la  $j$ ème variable en utilisant un tel indice engendre une variance non reconstruite dont son expression est déterminée en remplaçant dans l'équation (3.61)  $\mathcal{M}$  par  $\Upsilon$ . Ainsi, nous montrons qu'elle peut s'écrire comme suit :

$$\sigma_{\Psi_j}^2(\ell, v, \alpha) = \frac{\mathfrak{b}^4 \xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j + \delta^4 \xi_j^T \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T \xi_j}{\left( \mathfrak{b}^2 \xi_j^T \tilde{\mathbf{C}} \xi_j + \delta^2 \xi_j^T \hat{\mathbf{P}} \hat{\Lambda}_{\mathfrak{B}}^{-1} \hat{\mathbf{P}}^T \xi_j \right)^2} \quad (3.109)$$

Cette variance dépend de deux paramètres principaux qui sont  $\ell$  et  $v$  ainsi qu'un paramètre de lissage  $\alpha$  qui représente le seuil de signification. En revanche, le critère global dépend uniquement de  $\ell$  et  $\alpha$  :

$$\mathbf{VNR}_{\Psi}(\ell, \alpha) = \sum_{v=1}^m \sum_{j=1}^m \sigma_{\Psi_j}^2(\ell, v, \alpha) \quad (3.110)$$

Nous suggérons que le minimum de cette fonction peut correspondre au nombre optimal des CPs. En se basant sur un exemple simulé, nous avons observé que ce critère peut déterminer correctement le nombre désiré des CPs avec des valeurs pour  $\alpha$  qui sont comprises entre 1% et 5% (Mnassri et al., 2010b). Notamment, ce critère montre des résultats plus corrects que ceux du critère utilisant le classique indice combiné. Cependant, il demeure un critère empirique car seul l'expérience peut illustrer son efficacité.

### 3.4.2 Changement de représentation des données

La présence du seuil de signification dans les expressions des deux nouveaux critères proposés précédemment pose souvent des difficultés en rendant les résultats et les décisions incertains. Pour cette raison, nous avons étendu nos travaux dans ce cadre en proposant également un troisième nouveau critère que nous considérons le mieux performant et le plus consistant du point de vue théorique. Ce critère a été proposé initialement dans Mnassri et al. (2010c) puis il a été rectifié et adapté théoriquement dans Mnassri et al. (2011a).

#### 3.4.2.1 Motivation et théorie d'une nouvelle approche ACPVI

Le nouveau critère est fondé sur l'idée d'une analyse en composantes principales pondérées par leurs variances inversées (ACPVI, en anglais : *Inverse-Variance Weighting PCA* (IVWPCA)). En considérant des données normalisées, nous avons montré dans le deuxième chapitre que le critère **VNR** peut déterminer correctement le nombre optimal des CPs seulement en absence des variables indépendantes et quasi-indépendantes. En effet, un tel critère est insensible à la présence de ce type de variables. En outre, nous notons que ce critère peut également être plus limité en utilisant des données non normalisées. Une variable indépendante ayant une variance très importante peut occuper le premier ou le deuxième rang dans l'ordre d'importance des CPs en considérant la matrice de covariance. Dans ce cas et d'après la démonstration établie dans l'annexe A, le minimum du critère **VNR** correspondra à un nombre de CPs qui ne peut pas dépasser le rang d'une telle variable. Par conséquent, des CPs qui sont réellement significatives seront écartées. Ce critère peut garantir la sélection de ces CPs uniquement si les variables indépendantes et quasi-indépendantes disposent de variances moins inférieures aux valeurs propres de telles CPs.

Nous recommandons alors l'utilisation d'un tel critère avec des données normalisées car les CPs qui correspondent aux variables indépendantes sont hiérarchiquement placées après les CPs les plus significatives qui expriment de fortes corrélations. De cette manière, on est confronté uniquement au problème d'identification de telles variables. Notamment, les variables indépendantes se caractérisent par une invariabilité en termes de

leurs indépendances ainsi que leurs variances même en cas d'inversion de la matrice de corrélation.

Pour cette raison, l'idée principale s'appuie sur le principe des problèmes inverses par un changement de représentation des données. Puisque la normalisation des données influe souvent les relations linéaires entre les variables, nous supposons que les données normalisées de la matrice  $\mathbf{X}$  sont initialement issues d'une transformation linéaire d'autres données  $\mathbf{Y} \in \mathbb{R}^{N \times m}$  qui ne sont pas forcément normalisées. En notation vectorielle, cette hypothèse se traduit par :

$$\mathbf{x}(k) = \mathbf{A}\mathbf{y}(k) \quad (3.111)$$

$\mathbf{A} \in \mathbb{R}^{m \times m}$  est la matrice de transformation.  $\mathbf{y} \in \mathbb{R}^m$  représente le vecteur des données desquelles sont extraites celles du vecteur  $\mathbf{x}$ . A travers le changement de représentation, nous devons assurer que les données de  $\mathbf{X}$  sont normalisées. Dans ce contexte, on peut établir ce qui suit :

$$\begin{aligned} \Sigma &= \mathbb{E}\{\mathbf{x}\mathbf{x}^T\} = \mathbf{A} \mathbb{E}\{\mathbf{y}\mathbf{y}^T\} \mathbf{A}^T \\ &= \mathbf{A}\mathbf{S}\mathbf{A}^T \end{aligned} \quad (3.112)$$

où  $\Sigma$  et  $\mathbf{S}$  sont respectivement la matrice de corrélation de  $\mathbf{X}$  et la matrice de covariance de  $\mathbf{Y}$ .

Une contrainte que nous devons également l'assurer par une telle transformation linéaire concerne l'organisation des valeurs propres de  $\Sigma$  ainsi que celles de  $\mathbf{S}$ . En effet, nous souhaitons que leurs valeurs propres soient reliées en conservant un certain ordonnancement. Pour cela, la détermination des matrices  $\mathbf{A}$  et  $\mathbf{S}$  peut s'effectuer selon trois possibilités. Cependant, une unique possibilité peut satisfaire nos hypothèses :

- i.* D'après les équations (2.13) et (3.112), nous pouvons supposer que  $\mathbf{A} = \mathbf{P}$  et  $\mathbf{S} = \Lambda$ . Avec ces considérations, le problème à résoudre se ramène au cas classique ;
- ii.* Puisque  $\Sigma$  doit être nécessairement une matrice de corrélation, on peut considérer que  $\mathbf{S}$  est la matrice de covariance de  $\mathbf{X}$  et  $\mathbf{A}$  est une matrice diagonale constituée par les inverses des racines carrées des écarts types des variables originelles. Dans ce cas, on serait confronté à une étude sur les données de  $\mathbf{X}$  qui ne sont pas normalisées. Par conséquent, le critère **VNR** ne peut pas assurer une optimalité dans la sélection des CPs significatives, plus particulièrement en présence des variables indépendantes de variances importantes ;
- iii.* La dernière et l'unique possibilité envisageable est de supposer que :

$$\mathbf{A} = \Sigma \quad (3.113)$$

et

$$\mathbf{S} = \Sigma^{-1} \quad (3.114)$$

En présence de bruit de mesures de variances non nulles, la matrice de corrélation  $\Sigma$  est inversible. Ainsi,  $\Sigma^{-1}$  représente la matrice de covariance des données de  $\mathbf{Y}$ . Par conséquent, le vecteur de données  $\mathbf{y}(k)$  peut s'écrire comme suit :

$$\mathbf{y}(k) = \Sigma^{-1}\mathbf{x}(k) \quad (3.115)$$



Puisque  $\mathbf{P}$  est une matrice orthonormée, i.e.  $\mathbf{P}^{-1} = \mathbf{P}^T$ , la décomposition en valeurs singulières de la matrice de covariance  $\mathbf{S}$  des données de  $\mathbf{Y}$  mène à la forme suivante :

$$\mathbf{S} = \Sigma^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}^T \quad (3.116)$$

L'interprétation théorique de l'équation précédente prouve que les variables de  $\mathbf{Y}$  disposent des mêmes CPs que celles des variables de  $\mathbf{X}$ . La distinction se manifeste uniquement dans les variances de telles CPs. Posons  $\mathbf{h} \in \mathbb{R}^m$  le vecteur des CPs qui correspondent aux données de  $\mathbf{Y}$ , nous pouvons montrer que :

$$\begin{aligned} \mathbf{h}(k) &= \mathbf{P}^T \mathbf{y}(k) = \mathbf{P}^T \Sigma^{-1} \mathbf{x}(k) \\ &= \Lambda^{-1} \mathbf{P}^T \mathbf{x}(k) = \Lambda^{-1} \mathbf{t}(k) \end{aligned} \quad (3.117)$$

où, d'après (2.15),  $\mathbf{t}$  représente le vecteur des CPs de  $\mathbf{X}$ . Puisque ces CPs sont caractérisées par des variances qui sont égales aux valeurs propres de  $\Sigma$ , i.e.  $\mathbb{E}\{\mathbf{t}\mathbf{t}^T\} = \Lambda$ , on montre alors que :

$$\mathbb{E}\{\mathbf{h}\mathbf{h}^T\} = \Lambda^{-1} \mathbb{E}\{\mathbf{t}\mathbf{t}^T\} \Lambda^{-1} = \Lambda^{-1} \quad (3.118)$$

Par conséquent, nous déduisons que les CPs de  $\mathbf{Y}$  sont les mêmes que celles de  $\mathbf{X}$  mais de variances inverses. Pour cette raison, nous avons appelé cette approche analyse en composantes principales pondérées par leurs variances inversées (ACPVI).

D'après l'équation (3.116), nous remarquons qu'une telle décomposition engendre des valeurs propres dans l'ordre croissant. Cependant, le principe communément connu par une ACP est d'organiser les valeurs propres dans un ordre décroissant. Pour cela, l'équation (3.116) peut se réécrire comme suit :

$$\mathbf{S} = \mathbf{G}\mathbf{D}\mathbf{G}^T \quad (3.119)$$

avec

$$\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_{m-q}, \mathbf{g}_{m-q+1}, \dots, \mathbf{g}_m] = [\mathbf{p}_m, \dots, \mathbf{p}_{q+1}, \mathbf{p}_q, \dots, \mathbf{p}_1] \quad (3.120)$$

et

$$\mathbf{D} = \text{diag}\{d_1, \dots, d_{m-q}, d_{m-q+1}, \dots, d_m\} = \text{diag}\{\lambda_m^{-1}, \dots, \lambda_{q+1}^{-1}, \lambda_q^{-1}, \dots, \lambda_1^{-1}\} \quad (3.121)$$

Puisque  $\mathbf{S}$  est une matrice de covariance, cela semblerait contradictoire avec notre recommandation concernant la nature des données sur lesquelles on doit appliquer un critère basé sur le principe de la variance non reconstruite. Toutefois, l'organisation des valeurs propres de cette matrice de covariance qui est obtenue par un changement de représentation des données présentera un avantage majeur pour notre approche. Ainsi, ce type de matrice de covariance qui représente l'inverse d'une matrice de corrélation d'autres données sera le plus adapté pour une utilisation du critère **VNR** sur des données non normalisées.

Nous avons supposé que  $q$  représente le nombre optimal des CPs pour les données de  $\mathbf{X}$ . En présence des variables indépendantes et/ou quasi-indépendantes, la  $q$ ème CP représentera certainement l'une de ces variables. Ainsi, l'inversion de la matrice de corrélation

n'influe pas leurs natures. D'après (3.120) et (3.121), les  $q$  CPs significatives dans  $\mathbf{X}$  se transforment en des relations de redondance dans  $\mathbf{Y}$ . Inversement, les  $(m - q)$  relations de redondance dans  $\mathbf{X}$  se transforment en des CPs significatives dans  $\mathbf{Y}$ . Ces effets sont dus aux inverses des valeurs propres résiduelles de  $\Sigma$  qui produisent des valeurs propres très importantes dans  $\Sigma^{-1}$ . Evidemment s'il existe un ensemble de variables indépendantes dans  $\mathbf{X}$ , elles se mutent dans  $\mathbf{Y}$  sans modification considérable dans leurs indépendances. Ainsi, la dernière parmi elles sera associée au vecteur propre  $\mathbf{g}_{m-q+1}$  (3.120).

Afin de sélectionner les  $q$  CPs de la matrice  $\mathbf{X}$ , nous envisageons qu'il est plus facile d'identifier les  $(m - q)$  CPs de la matrice  $\mathbf{Y}$  en utilisant le principe de la variance non reconstruite.

### 3.4.2.2 VNR basée sur l'ACPVI

Posons  $\mathbf{y}^*(k)$  le vecteur des données obtenues lors du fonctionnement normal. En présence d'un défaut simple  $\mathcal{W}_i$  porté par la  $i$ ème direction  $\zeta_i$  et d'amplitude  $w(k)$  à l'instant  $k$ , le vecteur de données  $\mathbf{y}(k)$  peut s'exprimer comme suit :

$$\mathbf{y}(k) = \mathbf{y}^*(k) + \zeta_i w(k) \quad (3.122)$$

Le vecteur  $\zeta_i \in \mathbb{R}^m$  représente la  $i$ ème colonne de la matrice identité. Il correspond à la variable supposée en défaut. Ainsi, la reconstruction d'une telle variable engendre un vecteur reconstruit de données comme suit :

$$\mathbf{y}_i(k) = \mathbf{y}(k) - \zeta_i \hat{w}_i(k) \quad (3.123)$$

où  $\hat{w}_i$  est une estimation de  $w$ . Une telle estimation est optimale en minimisant l'erreur quadratique d'estimation  $\|\tilde{\mathbf{Z}}\mathbf{y}_i(k)\|^2$  de la façon suivante :

$$\begin{aligned} \hat{w}_i(k) &= \arg \min_{w(k)} \left\{ \|\tilde{\mathbf{Z}}\mathbf{y}_i(k)\|^2 \right\} \\ &= (\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^{-1} \zeta_i^T \tilde{\mathbf{Z}} \mathbf{y}(k) \end{aligned} \quad (3.124)$$

avec

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \quad (3.125)$$

Notamment,  $\tilde{\mathbf{Z}}$  est la matrice de projection dans le sous-espace résiduel de  $\mathbf{Y}$ .  $\tilde{\mathbf{G}} \in \mathbb{R}^{m \times (m-\kappa)}$  est composée des  $(m - \kappa)$  derniers vecteurs propres de la matrice  $\mathbf{G}$  et  $\kappa$  représente le nombre des CPs utilisées dans le modèle.

On note que la  $i$ ème variable est reconstructible si  $\zeta_i^T \tilde{\mathbf{Z}} \zeta_i \neq 0$ , i.e. le vecteur  $\tilde{\mathbf{Z}} \zeta_i$  est non nul.

Puisque les données de  $\mathbf{X}$  sont centrées alors celles de  $\mathbf{Y}$  le sont également. Par conséquent, la moyenne de  $\hat{w}_i$  est nulle. Ainsi, la variance de l'erreur de reconstruction de la  $i$ ème variable dans le sous-espace résiduel est définie par :

$$\begin{aligned} \sigma_i^2(\kappa) &= \text{Var} \left\{ \zeta_i^T (\mathbf{y} - \mathbf{y}_i) \right\} = \text{Var} \left\{ \hat{w}_i \right\} \\ &= \frac{\zeta_i^T \tilde{\mathbf{Z}} \mathbb{E} \{ \mathbf{y} \mathbf{y}^T \} \tilde{\mathbf{Z}} \zeta_i}{(\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^2} = \frac{\zeta_i^T \tilde{\mathbf{Z}} \Sigma^{-1} \tilde{\mathbf{Z}} \zeta_i}{(\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^2} = \frac{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{D}} \tilde{\mathbf{G}}^T \zeta_i}{(\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^2} \end{aligned} \quad (3.126)$$

On peut alors déduire le critère de la variance globale non reconstruite noté **VNRVI** et qui dépend de  $\kappa$  comme suit :

$$\mathbf{VNRVI}(\kappa) = \sum_{i=1}^m \frac{\sigma_i^2(\kappa)}{\zeta_i^T \Sigma^{-1} \zeta_i} = \sum_{i=1}^m \frac{\zeta_i^T \tilde{\mathbf{Z}} \Sigma^{-1} \tilde{\mathbf{Z}} \zeta_i}{(\zeta_i^T \Sigma^{-1} \zeta_i) (\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^2} = \sum_{i=1}^m \frac{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{D}} \tilde{\mathbf{G}}^T \zeta_i}{(\zeta_i^T \Sigma^{-1} \zeta_i) (\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)^2} \quad (3.127)$$

Notons par  $\kappa_{op}$  le nombre optimal des CPs dans  $\mathbf{Y}$ . En se basant sur ce critère, nous suggérons que son minimum correspond à  $\kappa_{op} = (m - q)$  CPs :

$$\kappa_{op} = \arg \min_{\kappa} \{\mathbf{VNRVI}(\kappa)\} = m - q \quad (3.128)$$

Par conséquent, le nombre optimal des CPs de  $\mathbf{X}$  est déduit de la manière suivante :

$$\ell_{op} = q = m - \kappa_{op} = m - \arg \min_{\kappa} \{\mathbf{VNRVI}(\kappa)\} \quad (3.129)$$

### 3.4.2.3 Consistance théorique du critère **VNRVI**

En posant

$$\tilde{u}_i^2(\kappa) = \frac{\zeta_i^T \tilde{\mathbf{Z}} \Sigma^{-1} \tilde{\mathbf{Z}} \zeta_i}{(\zeta_i^T \tilde{\mathbf{Z}} \zeta_i)} \quad (3.130)$$

et

$$\hat{u}_i^2(\kappa) = \tilde{u}_i^2(\kappa) \left( \frac{1 - \zeta_i^T \tilde{\mathbf{Z}} \zeta_i}{\zeta_i^T \tilde{\mathbf{Z}} \zeta_i} \right) \quad (3.131)$$

et d'après l'équation (3.126), la variance non reconstruite de la  $i$ ème variable pondérée par sa variance originelle peut s'exprimer comme suit :

$$\frac{\sigma_i^2(\kappa)}{\zeta_i^T \Sigma^{-1} \zeta_i} = \frac{\tilde{u}_i^2(\kappa) + \hat{u}_i^2(\kappa)}{\zeta_i^T \Sigma^{-1} \zeta_i} \quad (3.132)$$

Etant donné que la quantité  $\zeta_i^T \Sigma^{-1} \zeta_i$  est une constante qui ne dépend pas de  $\kappa$ , notre étude du comportement du nouveau critère peut se limiter aux deux termes  $\tilde{u}_i^2$  et  $\hat{u}_i^2$ . Dans ce cadre, nous avons montré théoriquement dans l'annexe C que la fonction  $\tilde{u}_i^2$  est monotone décroissante en  $\kappa$ . D'autre part, le terme  $\zeta_i^T \tilde{\mathbf{Z}} \zeta_i$  tend vers zéro quand  $\kappa$  tend vers  $m$ . Cela implique par conséquent que  $\hat{u}_i^2$  devient croissante en montant rapidement quand  $\kappa$  tend vers  $m$ . De cette manière, l'expression (3.132) doit avoir nécessairement un minimum unique qui correspond à un nombre de CPs dans l'intervalle  $[1, m]$ . Ainsi, nous pouvons déduire également que le critère **VNRVI** aura un seul minimum global pour un nombre de CPs dans le même intervalle.

En revanche, il est crucial d'établir les conditions qui garantissent d'avoir théoriquement ce minimum en  $(m - q)$  CPs. Dans ce cadre, il semblerait nécessaire de connaître tout d'abord la distribution du bruit dans les données de  $\mathbf{Y}$ . Selon l'équation (2.28), le vecteur de données  $\mathbf{y}(k)$  peut s'écrire comme suit :

$$\mathbf{y}(k) = \Sigma^{-1} \hat{\mathbf{x}}(k) + \Sigma^{-1} \mathbf{v}(k) = \hat{\mathbf{y}}(k) + \mathbf{w}(k) \quad (3.133)$$

où les vecteurs  $\hat{\mathbf{y}}(k)$  et  $\mathbf{w}(k)$  représentent respectivement les mesures sans bruit et le bruit au  $k$ ème instant. Ainsi, la matrice de covariance du bruit dans  $\mathbf{Y}$  est donnée par :

$$\mathbb{E}\{\mathbf{w}\mathbf{w}^T\} = \Sigma^{-1} \mathbb{E}\{\mathbf{v}\mathbf{v}^T\} \Sigma^{-1} \quad (3.134)$$

où  $\mathbb{E}\{\mathbf{v}\mathbf{v}^T\}$  représente la matrice de covariance du bruit dans les données de  $\mathbf{X}$ . D'après l'équation précédente, nous remarquons que le bruit dans  $\mathbf{Y}$  dépend fortement de celui dans  $\mathbf{X}$ . Néanmoins, la matrice de covariance exprimée par une telle équation est généralement non diagonale ce qui implique que le bruit  $\mathbf{w}$  est coloré.

D'après la démonstration que nous avons établie dans l'annexe D, le critère **VNRVI** garantit son minimum en  $\kappa = (m - q)$  CPs si :

$$\frac{d_{m-q+1}}{d_m} \leq \frac{\zeta_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \zeta_i}{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \zeta_i} \quad \text{pour } \kappa \geq m - q \quad (3.135)$$

et

$$d_{m-q} \geq \left( 1 + \frac{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \zeta_i}{\zeta_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \zeta_i} \right) d_{m-q+1} \quad \text{pour } \kappa < m - q \quad (3.136)$$

où  $\tilde{\mathbf{G}} \in \mathbb{R}^{m \times (m-\kappa)}$  et  $\tilde{\mathbf{G}}_q \in \mathbb{R}^{m \times q}$  sont deux matrices constituées respectivement par les  $(m - \kappa)$  et les  $q$  derniers vecteurs propres de la matrice  $\mathbf{G}$ . Notamment, nous pouvons réécrire les deux inégalités précédentes en fonction des valeurs propres de la matrice  $\Sigma$ . Selon l'équation (3.121), elles se transforment respectivement comme suit :

$$\frac{\lambda_1}{\lambda_q} \leq \frac{\zeta_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \zeta_i}{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \zeta_i} \quad \text{pour } \kappa \geq m - q \quad (3.137)$$

et

$$\lambda_q \geq \left( 1 + \frac{\zeta_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \zeta_i}{\zeta_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \zeta_i} \right) \lambda_{q+1} \quad \text{pour } \kappa < m - q \quad (3.138)$$

L'interprétation de l'inégalité (3.138) impose que la variance d'un signal quelconque doit être plus grande que celle du bruit, ce qui représente une condition très ordinaire. Cependant, la contrainte exprimée dans (3.137) signifie que les valeurs propres les plus significatives ( $\lambda_1, \dots, \lambda_q$ ) doivent avoir des valeurs très proches les unes des autres. Toutefois, la présence des variables indépendantes et/ou quasi-indépendantes peut alléger une telle contrainte en garantissant un minimum du critère **VNRVI** en  $(m - q)$  CPs pour  $\kappa \geq m - q$ . Cette déduction est affirmée grâce à la démonstration de l'annexe A qui est également valable pour ce nouveau critère. Une telle démonstration prouve que ces variables sont toujours considérées comme des CPs résiduelles. Effectivement, l'intérêt de notre approche par la proposition de la matrice  $\mathbf{Y}$  est d'avoir contrairement au cas classique, un sous-espace résiduel inversé dont les premières CPs que constituant un tel sous-espace représentent ce type de variables suivies par les CPs les plus significatives dans  $\mathbf{X}$ . Cette particularité du critère **VNRVI** peut garantir une sélection optimale des CPs tout en considérant les variables indépendantes et quasi-indépendantes.

### 3.5 Résultats de simulation

Afin d'illustrer les résultats obtenus par les critères proposés, nous avons considéré le même exemple du deuxième chapitre. Nous rappelons qu'un tel exemple représente quatre ensembles de variables  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  et  $\mathcal{D}$  dont chacun a été généré selon trois cas différents d'un bruit *i.i.d.* En s'appuyant sur l'analyse et les interprétations menées dans le chapitre précédent, nous avons conclu que ces ensembles de variables sont constitués respectivement de 5, 6, 7 et 8 CPs malgré le changement considérable de la variance du bruit.

Notamment, les critères proposés dans ce chapitre se basent sur des minimisations par rapport au nombre de CPs pour la détermination du modèle optimal. Exceptionnellement, les critères  $\mathbf{VNR}_\varphi$  et  $\mathbf{VNR}_\psi$  dépendent également d'un seuil de signification  $\alpha$ . Puisqu'on dispose d'une connaissance *a priori* sur les nombres optimaux des CPs constituant chaque ensemble des données, nous avons réalisé une étude de performance plus particulièrement pour ces deux critères par rapport à 100 réalisations en variant le niveau de confiance  $(1 - \alpha)$  dans l'intervalle  $[1, 99]\%$  avec un pas de 1%. Les figures 3.1, 3.2 et 3.3 affichent les pourcentages de sélections correctes des nombres appropriés des CPs dans les bases des données étudiées respectivement pour les trois différents cas de bruit *i.i.d.* Nous remarquons que le critère  $\mathbf{VNR}_\varphi$  sélectionne dans plus de 80% des réalisations les nombres désirés des CPs avec des valeurs de niveau de confiance généralement proches de 20%. En revanche, il montre parfois une allure fluctuante surtout lors d'un bruit de faibles variances. En comparant les courbes des deux critères, nous observons clairement que celui du  $\mathbf{VNR}_\psi$  est beaucoup plus performant. Sa minimisation par rapport à  $\ell$  converge vers les nombres corrects des CPs en considérant des valeurs habituelles pour le seuil de signification  $\alpha$ . Il montre ainsi une amélioration prometteuse par rapport au critère  $\mathbf{VNR}_\varphi$ .

En considérant des seuil de signification  $\alpha$  égaux à 80% pour  $\mathbf{VNR}_\varphi$  et 2% pour  $\mathbf{VNR}_\psi$ , les figures 3.4 et 3.5 illustrent les allures en  $\ell$  de ces deux critères appliqués sur les ensembles  $(\mathcal{A}, \mathcal{B})$  et  $(\mathcal{C}, \mathcal{D})$  respectivement pour un bruit *i.i.d.* de faible variance ( $\sigma^2 = 0.002$ ). Ces deux critères peuvent déterminer une dimension optimale d'un modèle ACP sous la contrainte d'un choix très approprié des valeurs de leurs seuils de signification.

Face à la difficulté d'une telle tâche, le troisième nouveau critère  $\mathbf{VNRVI}$  montre une efficacité très avantageuse en évitant les problèmes des quantiles par la considération d'un autre principe totalement différent. Selon les dernières courbes illustrées dans les premières et deuxièmes colonnes des figures 3.4 et 3.5, les minima d'un tel critère correspondent à  $\kappa_{op} = 7$  CPs dans toutes les bases des données transformées qui sont associées à celles de  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  et  $\mathcal{D}$ . Ainsi, les nombres des variables  $m$  dans ces ensembles sont respectivement 12, 13, 14 et 15. En se basant sur l'équation (3.129), nous pouvons déduire que les nombres optimaux des CPs dans les ensembles  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  et  $\mathcal{D}$  sont respectivement 5, 6, 7 et 8 CPs.

Puisque le bruit de mesures est représenté par des variables aléatoires, une simple réalisation ne peut pas valider ces résultats. De la même manière qu'au chapitre précédent, nous avons calculé un pourcentage par rapport à 1500 réalisations des nombres de CPs retenues à l'aide du critère  $\mathbf{VNRVI}$  en se référant à l'équation (3.129). En considérant un nombre d'observations  $N = 1500$  dans les trois différents cas de bruit *i.i.d.*, le tableau 3.2 prouve qu'un tel critère a déterminé dans 100% des réalisations les nombres désirés

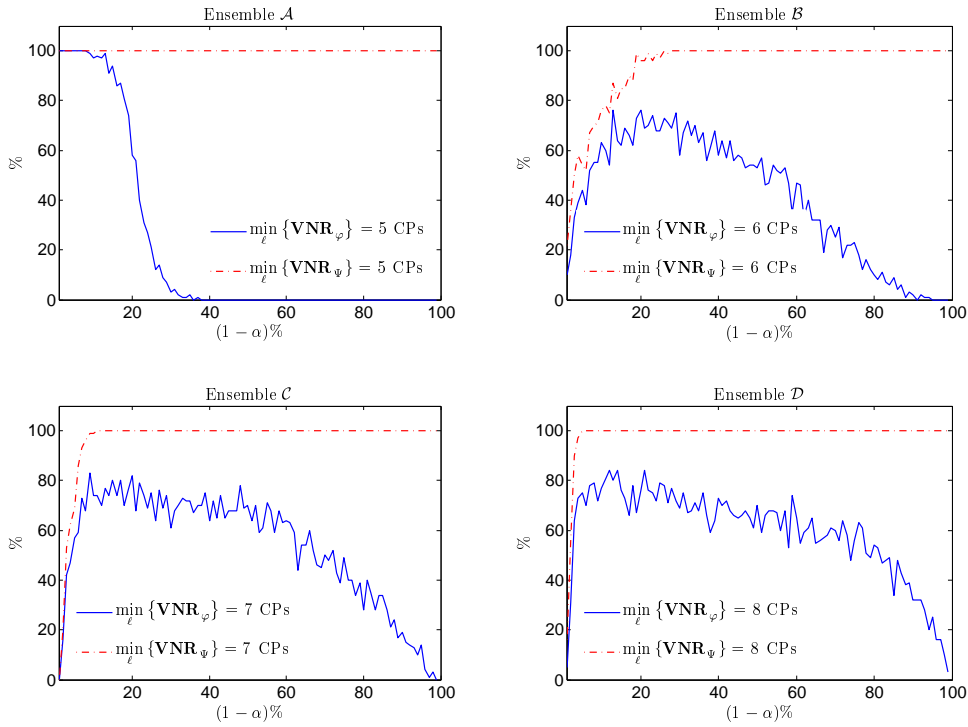


FIGURE 3.1 – Pourcentages de sélections correctes en fonction de  $(1 - \alpha)$  par  $\mathbf{VNR}_{\varphi}$  et  $\mathbf{VNR}_{\psi}$  sur tous les ensembles des variables avec un bruit *i.i.d.* de variance  $\sigma^2 = 0.002$

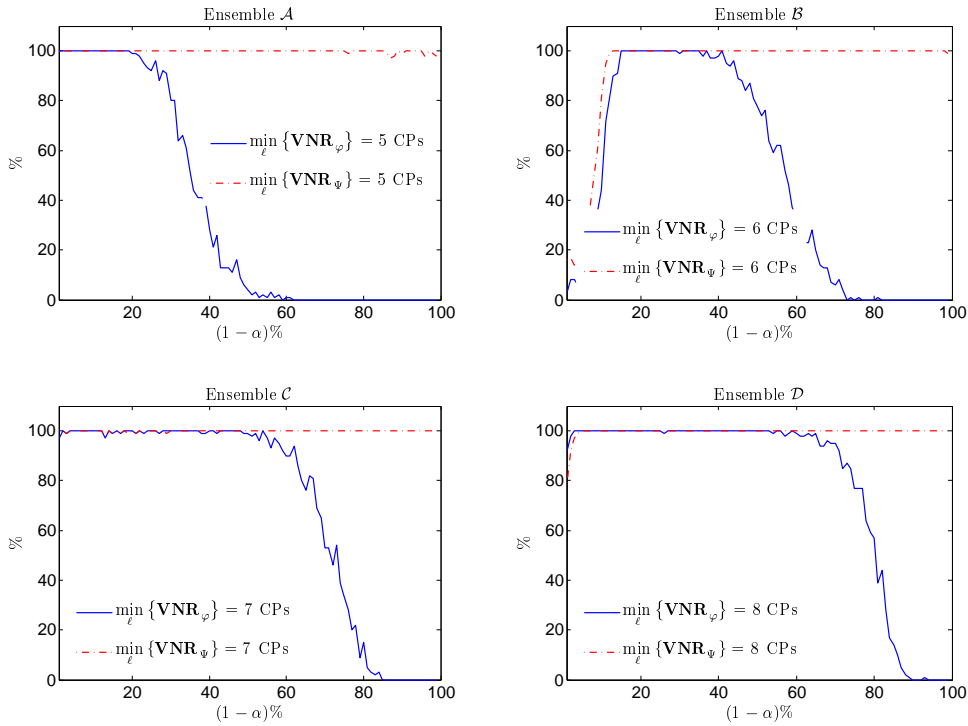


FIGURE 3.2 – Pourcentages de sélections correctes en fonction de  $(1 - \alpha)$  par  $\mathbf{VNR}_{\varphi}$  et  $\mathbf{VNR}_{\psi}$  sur tous les ensembles des variables avec un bruit *i.i.d.* de variance  $\sigma^2 = 0.2$

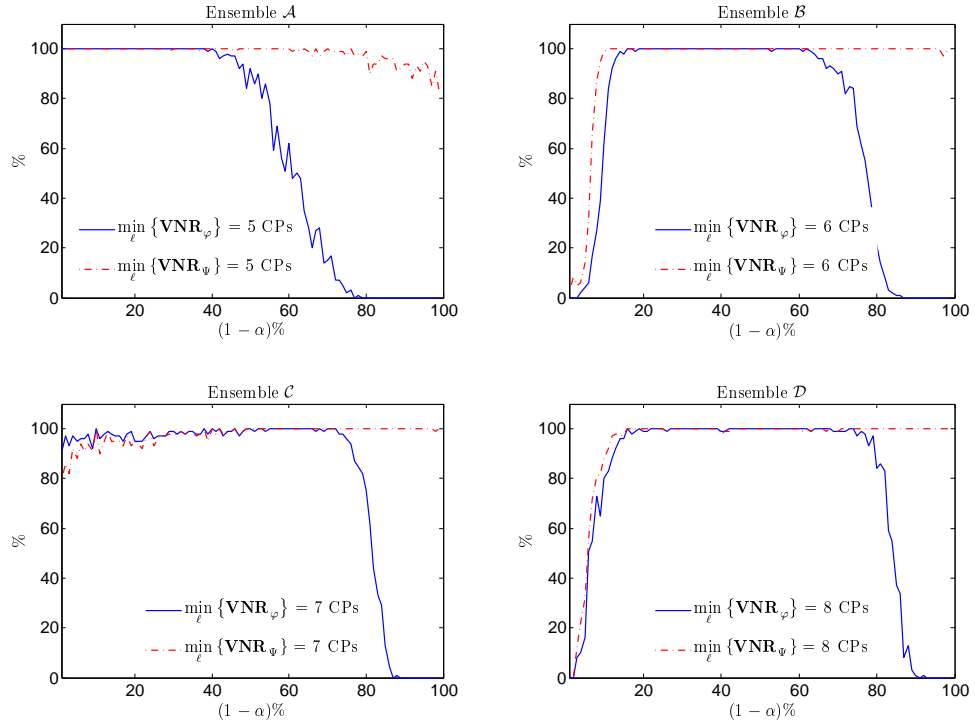


FIGURE 3.3 – Pourcentages de sélections correctes en fonction de  $(1 - \alpha)$  par  $\mathbf{VNR}_{\varphi}$  et  $\mathbf{VNR}_{\psi}$  sur tous les ensembles des variables avec un bruit *i.i.d.* de variance  $\sigma^2 = 0.5$

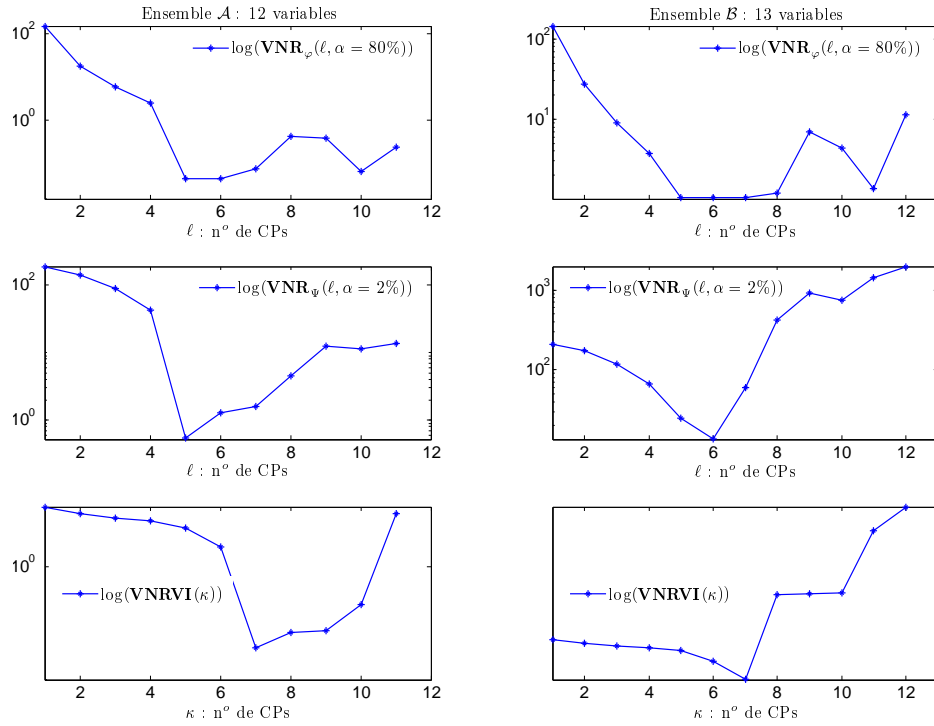


FIGURE 3.4 – Allures des nouveaux critères appliquées sur les ensembles  $\mathcal{A}$  et  $\mathcal{B}$ , respectivement en 1<sup>ère</sup> et 2<sup>ème</sup> colonne, en considérant un bruit *i.i.d.* de variance  $\sigma^2 = 0.002$

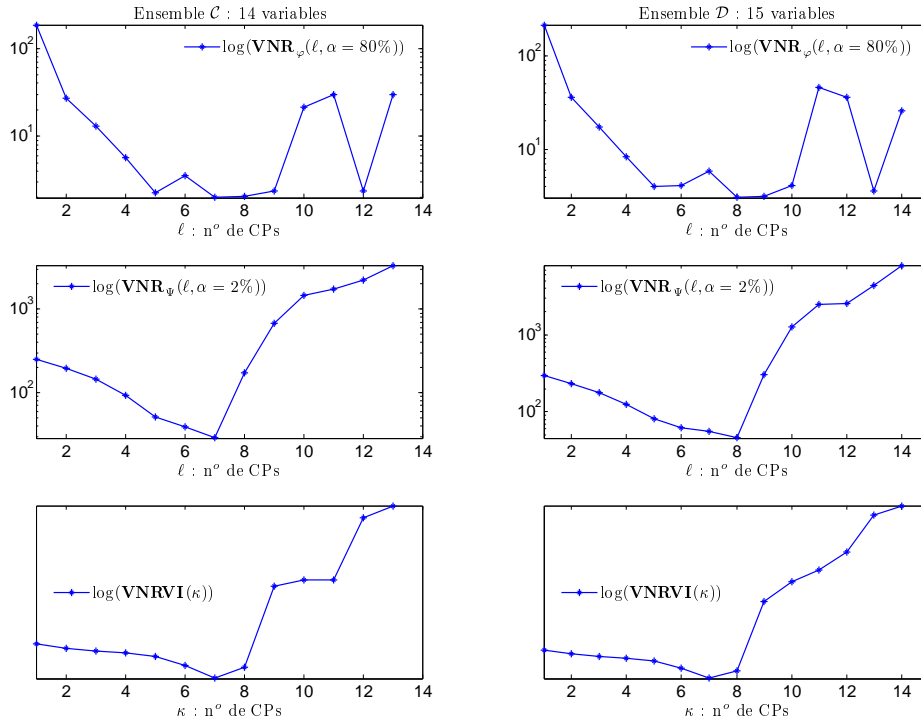


FIGURE 3.5 – Allures des nouveaux critères appliquées sur les ensembles  $\mathcal{C}$  et  $\mathcal{D}$ , respectivement en 1<sup>ère</sup> et 2<sup>ème</sup> colonne, en considérant un bruit *i.i.d.* de variance  $\sigma^2 = 0.002$

	Ensemble $\mathcal{A}$	Ensemble $\mathcal{B}$	Ensemble $\mathcal{C}$	Ensemble $\mathcal{D}$
Critère :	VNRVI	VNRVI	VNRVI	VNRVI
<b>Bruit <i>i.i.d.</i> : <math>\sigma^2 = 0.002</math></b>				
5 CPs (%)	100			
6 CPs (%)		100		
7 CPs (%)			100	
8 CPs (%)				100
<b>Bruit <i>i.i.d.</i> : <math>\sigma^2 = 0.2</math></b>				
5 CPs (%)	100			
6 CPs (%)		100		
7 CPs (%)			100	
8 CPs (%)				100
<b>Bruit <i>i.i.d.</i> : <math>\sigma^2 = 0.5</math></b>				
5 CPs (%)	100			
6 CPs (%)		100		
7 CPs (%)			100	
8 CPs (%)				100

TABLE 3.2 – Pourcentage, par rapport à 1500 réalisations, des nombres de CPs sélectionnées dans les ensembles des données en se basant sur le critère **VNRVI** ( $N = 1500$  observations générées selon trois différents cas de bruit *i.i.d.*)



des CPs. Son efficacité présente un avantage prometteur par rapport aux efficacités des autres critères de sélection présentés dans cette thèse.

Avec le critère **VNRVI**, l'optimalité de sélection est assurée par un nombre d'observations  $N$  assez inférieur à celui exigé pour le critère **MDL** présenté au deuxième chapitre. En comparant les résultats des tableaux 3.2 et 2.7, nous distinguons clairement la différence au niveau du nombre d'observations considérées. Dans ce contexte, le paramètre  $N$  est exprimé implicitement dans l'expression du **VNRVI** car il influe sur les valeurs et vecteurs propres considérés. Ainsi, cette influence sera négligeable voire nulle pour un nombre  $N$  suffisant et fini dans le sens où la matrice de covariance des données demeure constante.

## 3.6 Conclusion

Ce chapitre présente nos contributions dans le thème d'une modélisation optimale par ACP en utilisant le principe de la variance non reconstruite. Face à l'abondance des critères de sélection disponibles dans la littérature et qui ont montré effectivement une divergence remarquable dans les résultats obtenus, nous avons motivé théoriquement l'importance de la précision dans le choix d'une structure adaptée du modèle ACP. Dans la mesure où celui-ci n'est pas optimal, une conséquence évidente est prouvée sur la détectabilité de défauts et qui aura sans doute un effet également sur l'isolabilité de défauts.

En partant du principe de la variance de l'erreur de reconstruction, nous avons établi une expression généralisée pour une telle variance valable pour tout un indice de détection ayant une forme quadratique. Dans ce cadre, nous avons présenté les indices de détection proposés dans la littérature. Ainsi, nous avons établi pour chacun de ces indices sa variance non reconstruite en s'appuyant sur celle généralisée.

Cela nous a permis de conclure suite à une étude théorique des différentes variances obtenues que la plupart, exceptant quelques-unes, sont inutiles pour le choix du modèle. Plus particulièrement, nous avons montré que la variance relative à l'indice  $SPE$  n'est autre que celle exprimée par le critère **VNR** étudié dans le deuxième chapitre. Ainsi, notre contribution dans ce cadre est introduite par la variance non reconstruite ( $\mathbf{VNR}_\varphi$ ) associée à l'indice combiné. Celle-ci dépend du nombre des CPs considérées dans le modèle ainsi qu'un seuil de signification qui caractérise les limites de contrôle des indices de détection. Malheureusement, l'étude de ce critère n'a pas abouti à une consistance théorique de son comportement. En effet, seule l'expérience prouve qu'il est capable de déterminer la dimension du modèle en se référant à un choix très approprié du seuil de signification considéré. Dans ce contexte, nous avons proposé un deuxième nouveau critère ( $\mathbf{VNR}_\psi$ ) basé sur une nouvelle distance combinée. L'objectif étant de faciliter la tâche du choix de la valeur d'un tel seuil. Le nouveau critère montre des résultats plus efficaces et mieux stationnaires que ceux du  $\mathbf{VNR}_\varphi$ . Toutefois, ils restent dans leurs globalités incertains aussi faut-il disposer d'une connaissance *a priori* ou d'une expertise pour prendre la décision convenable. En effet, nous avons considéré que ces critères sont empiriques car ils s'appuient uniquement sur l'expérience.

La limitation qui a été prouvée théoriquement dans le chapitre précédent pour le critère **VNR**, a été un avantage d'une nouvelle proposition. Cette dernière est inspirée du

principe des problèmes inverses en proposant un changement de représentation des données. Nous avons supposé que les données normalisées de  $\mathbf{X}$  observées réellement peuvent être en réalité des dérivées linéaires d'autres données de  $\mathbf{Y}$ . Une ACP utilisant les nouvelles données de  $\mathbf{Y}$  engendre les mêmes CPs que celles de  $\mathbf{X}$  mais de variances inversées. Pour cela, la nouvelle approche est appelée ACPVI. En s'appuyant sur le principe de la variance de l'erreur de reconstruction, nous avons proposé un nouveau critère **VNRVI**. En effet, la détermination du nombre optimal des CPs dans  $\mathbf{Y}$  en utilisant ce critère sert à déduire celui de  $\mathbf{X}$  en considérant évidemment les variables indépendantes et quasi-indépendantes. En outre, nous avons établi théoriquement les conditions garantissant l'optimalité de sélection par un tel critère. Les résultats de simulation ont validé notre théorie en prouvant que le critère **VNRVI** apporte un avantage et une efficacité considérables par rapport à tous les critères présentés dans le deuxième chapitre pour le choix de la structure adaptée du modèle ACP.