

Modélisation par analyse en composantes principales

Sommaire

2.1	Introduction	28
2.2	Analyse en composantes principales linéaire	29
2.3	Propriétés	33
2.3.1	Modélisation en absence de bruit	34
2.3.2	Modélisation en présence de bruit	36
2.4	Détermination d'une structure optimale du modèle ACP	36
2.4.1	Critères de la théorie de l'information	36
2.4.2	Critères heuristiques	37
2.4.2.1	Critère IE	37
2.4.2.2	Pourcentage cumulé de la variance	38
2.4.2.3	Scree Test	38
2.4.2.4	Critère de Guttman	39
2.4.2.5	Autocorrélation	40
2.4.2.6	Validation croisée	40
2.4.3	Minimisation de la variance de l'erreur de reconstruction	41
2.4.3.1	Critère VNR	42
2.4.3.2	Consistance théorique du critère VNR	43
	Cas de bruit <i>i.i.d.</i>	44
	Cas de bruit coloré	45
2.5	Etude comparative des différents critères	45
2.5.1	Présentation et interprétation de l'exemple simulé	46
2.5.2	Interprétations des critères basés sur des seuils	48
2.5.3	Interprétations des critères minimisés	51
2.6	Conclusion	58

2.1 Introduction

L'Analyse en Composantes Principales (ACP) est une méthode statistique multivariée qui permet d'extraire les corrélations d'un ensemble de données à travers un ensemble de fonctions empiriques orthogonales. Elle représente une méthode descriptive permettant d'étudier les relations linéaires entre les variables sans tenir compte *a priori* d'une quelconque structure (Jolliffe, 2002). Ses origines historiques peuvent être remontées jusqu'aux œuvres de E. Beltrami en Italie (1873) et C. Jordan en France (1874) puisque ce sont eux qui ont formulé la décomposition en valeurs singulières (SVD) d'une matrice carrée (Cinar et al., 2007). Cependant, la première application pratique de l'ACP peut être attribuée au travail de Pearson (1901) en biologie, puis elle a été de nouveau développée et formalisée par Hotelling (1933). Ensuite, elle est devenue une technique multivariée standard (Jackson, 1991; Jolliffe, 2002).

Depuis les années 1970, de nombreux travaux ont proposé d'utiliser l'ACP comme une technique de modélisation de processus à partir de laquelle un modèle ACP peut être obtenu (Kresta et al., 1991; MacGregor et Kourti, 1995; Jolliffe, 2002). Ce modèle est extrait en se basant sur un ensemble de données qui sont issues d'un fonctionnement normal. Il permet d'estimer les variables ou les paramètres du processus à surveiller. Ainsi, il peut être utilisé pour détecter les valeurs aberrantes dans les données, fournir la réconciliation de données et surveiller les écarts par rapport à un fonctionnement normal.

L'utilisation fréquente de l'ACP dans plusieurs domaines s'est justifiée par sa réduction des données caractérisant un espace de grande dimension en un ensemble de composantes principales (CPs) constituant un sous-espace de dimension réduite. Cependant, l'optimalité d'une telle réduction réside dans la détermination du nombre de CPs les plus significatives. Dans ce cadre, plusieurs critères et règles ont été proposés dans la littérature afin de définir la dimension optimale d'un modèle ACP (Jackson, 1991; Valle et al., 1999; Jolliffe, 2002). D'une façon non exhaustive, Jolliffe (2002) distingue selon son point de vue trois différentes catégories de critères.

La première famille constitue des critères empiriques ou heuristiques dont la justification de leur utilisation, malgré quelques tentatives pour les mettre sur une base plus formelle, reste subjective comme le pourcentage de la variance totale ou également le test du coude (*Scree Test*) proposé par Cattell (1966). Contrairement, la fonction d'*imbedded error* (IE) de Malinowski (1977) et beaucoup d'autres critères qui proviennent généralement de la communauté de la chimiométrie disposent des règles de décision plus objectives et simples. Ces règles se basent souvent sur la minimisation d'un critère par rapport au nombre des CPs. En revanche, la subjectivité s'exprime dans la théorie du critère lui-même. La deuxième catégorie représente des approches basées sur une série de tests d'hypothèses qui sont généralement très sophistiquées et surestiment souvent la dimension du modèle. A titre d'exemple, Bartlett (1954) et Lawley (1956) ont développé une méthode dont le problème confronté est associé au choix des niveaux des seuils pour les différents tests. La troisième catégorie constitue des critères basés sur des méthodes de calculs intensifs comme la validation croisée qui permet le calcul d'un critère dit **PRESS** (Wold, 1978; Eastment et Krzanowski, 1982). En effet, ce critère est fondé sur la capacité prédictive des différents modèles ACP.

Notamment, la littérature a offert une autre catégorie de critères qu'on ne peut pas

ignorer. Une telle catégorie intitulée la théorie de l'information est principalement issue du domaine du traitement du signal. Notamment, elle est constituée de deux critères qui sont communément connus sous les noms d'*Akaike Information Criterion* (**AIC**, Akaike (1973)) et *Minimum Description Length* (**MDL**, Schwarz (1978) et Rissanen (1978)).

Différemment aux principes des critères classiques, de nouvelles approches proposent de définir un modèle ACP afin d'assurer une meilleure détection et localisation des défauts plutôt que d'offrir une meilleure approximation des données. Pour obtenir le modèle ACP le plus sensible à un défaut, Wang et al. (2004) ont proposé d'utiliser un indice prenant en compte l'amplitude minimale du défaut nécessaire afin d'assurer sa détection. Notamment, cette approche nécessite une connaissance *a priori* sur les défauts. Plus récemment, Tamura et Tsujita (2007) ont proposé une procédure permettant de définir la dimension d'un modèle ACP en fonction des directions des défauts afin de leur offrir une meilleure sensibilité. Puisque ces directions sont généralement inconnues, les mêmes auteurs suggèrent de définir plusieurs modèles ACP. Par conséquent, une telle approche semble être plus pratique en considérant uniquement les défauts simples. Cependant, elle est difficilement utilisable dans le cas de défauts multiples où un grand nombre de modèles doivent être considérés.

Le critère qui représente un intérêt majeur dans nos travaux de recherche se base sur la minimisation de la variance de l'erreur de reconstruction également appelée la variance non reconstruite (**VNR**) (Dunia et Qin, 1998b,c,a; Qin et Dunia, 2000). Son expression représente la variance en fonction du nombre de CPs, de la différence entre une mesure observée et son estimée obtenue en utilisant l'ensemble des mesures des autres variables. En effet, nous montrerons que ce critère aide à identifier le nombre des axes principaux uniquement entre les variables qui sont linéairement corrélées.

Ce chapitre présentera un rappel du principe mathématique de l'ACP linéaire dans la deuxième section. Etant donné que cette méthode est considérée comme un outil de modélisation, la problématique souvent confrontée lors de son utilisation représente le choix de la dimension du modèle ACP. Pour cela, la troisième section définira quelques propriétés liées à la détermination d'un modèle ACP en absence puis en présence de bruit. Dans la pratique, le bruit de mesures ne peut pas être négligé, ce qui compliquera la détermination de la structure optimale du modèle. Dans ce contexte, la quatrième section présentera quelques critères de sélection parmi les plus connus dans la littérature. Ensuite, une étude comparative des critères choisis sera présentée dans l'avant dernière section. Ceci en considérant un exemple de synthèse. A travers ce dernier, nous contribuons par deux démonstrations montrant les limitations de deux critères. Finalement, nous concluons ce chapitre.

2.2 Analyse en composantes principales linéaire

L'ACP consiste à remplacer une famille de variables par de nouvelles variables appelées CPs. Ces dernières sont de variances maximales et non corrélées deux à deux. Elles sont des combinaisons linéaires des variables originelles.

Considérons un ensemble de données collectées lors d'un fonctionnement normal du

système étudié. Ces données peuvent être représentées par une matrice

$$\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(N)]^T \in \mathbb{R}^{N \times m} \quad (2.1)$$

où N représente le nombre des observations et m représente le nombre des variables mesurées. Chaque ligne de la matrice de données \mathbf{X} représente une observation sous forme d'un vecteur de mesures collectées à un instant k , généralement centrées

$$\mathbf{x}(k) = [x_1(k), \dots, x_m(k)]^T \in \mathbb{R}^m \quad (2.2)$$

où $x_j(k)$ avec $j = \{1, \dots, m\}$ représente la mesure de la variable j à l'instant k . Par définition, la matrice de covariance est donnée par :

$$\Sigma = \mathbb{E} \{ \mathbf{x}\mathbf{x}^T \} = \frac{1}{N} \mathbf{X}^T \mathbf{X} \in \mathbb{R}^{m \times m} \quad (2.3)$$

Selon le principe de l'ACP, on suppose qu'un vecteur de composantes $\hat{\mathbf{t}} \in \mathbb{R}^\ell$ est associé à chaque vecteur d'observation dont il optimise la représentation au sens de la minimisation de l'erreur d'estimation de \mathbf{x} ou la maximisation de la variance de $\hat{\mathbf{t}}$. A chaque instant k , les vecteurs $\hat{\mathbf{t}}$ et \mathbf{x} sont liés par une transformation linéaire de type $\hat{\mathbf{t}}(k) = \hat{P}^T \mathbf{x}(k)$ telle que la matrice de transformation $\hat{P} \in \mathbb{R}^{m \times \ell}$ vérifie la condition d'orthogonalité $\hat{P}^T \hat{P} = \mathbf{I}_\ell \in \mathbb{R}^{\ell \times \ell}$.

Les colonnes de la matrice \hat{P} sont les vecteurs d'une base orthonormée d'un sous-espace \mathbb{R}^ℓ de représentation réduite des données initiales. La transformation linéaire se traduit par la projection des données originelles exprimées dans un espace de dimension m vers un sous-espace orthogonal de dimension ℓ . Les composantes $t_j(k)$ avec $j = \{1, \dots, \ell\}$ du vecteur $\hat{\mathbf{t}}(k)$ sont les projections des éléments du vecteur de données $\mathbf{x}(k)$ dans le sous-espace \mathbb{R}^ℓ .

L'optimisation de la représentation en se basant sur la matrice de projection \hat{P} est obtenue par la minimisation de l'erreur quadratique d'estimation de \mathbf{x} . Notons par \hat{P} la matrice optimale de représentation, celle-ci peut être donnée par :

$$\hat{P} = \arg \min_{\hat{P}} \left\{ J_e(\hat{P}) \right\} \quad (2.4)$$

où J_e représente le critère de l'erreur d'estimation par ACP qui devrait être minimisé. Sous la contrainte d'orthogonalité de la matrice de projection \hat{P} , nous pouvons écrire :

$$\begin{aligned} J_e(\hat{P}) &= \mathbb{E} \left\{ \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \right\} = \mathbb{E} \left\{ \left\| \mathbf{x} - \hat{P} \hat{P}^T \mathbf{x} \right\|^2 \right\} \\ &= \mathbb{E} \left\{ \left(\mathbf{x} - \hat{P} \hat{\mathbf{t}} \right)^T \left(\mathbf{x} - \hat{P} \hat{\mathbf{t}} \right) \right\} = \mathbb{E} \left\{ \mathbf{x}^T \mathbf{x} - \hat{\mathbf{t}}^T \hat{\mathbf{t}} \right\} \\ &= \mathbb{E} \left\{ \text{tr} \left(\mathbf{x}\mathbf{x}^T \right) - \hat{\mathbf{t}}^T \hat{\mathbf{t}} \right\} = \text{tr} \{ \Sigma \} - \mathbb{E} \left\{ \hat{\mathbf{t}}^T \hat{\mathbf{t}} \right\} \\ &= \text{tr} \{ \Sigma \} - J_v(\hat{P}) \end{aligned} \quad (2.5)$$

où $\text{tr} \{ \cdot \}$ désigne la trace d'une matrice carrée. Etant donné que le terme $\text{tr} \{ \Sigma \}$ est une constante, la minimisation du critère J_e revient à maximiser celui du J_v donné par :

$$J_v(\hat{P}) = \mathbb{E} \left\{ \hat{\mathbf{t}}^T \hat{\mathbf{t}} \right\} = \mathbb{E} \left\{ \sum_{j=1}^{\ell} t_j^2 \right\} = \sum_{j=1}^{\ell} \mathbb{E} \left\{ t_j^2 \right\} = \sum_{j=1}^{\ell} \text{Var} \left\{ t_j \right\} \quad (2.6)$$

D'après l'équation précédente, la maximisation du critère J_v est équivalente à une maximisation de la variance de la composante t_j . Ainsi, le problème d'optimisation est reformulé comme suit :

$$\hat{\mathbf{P}} = \arg \min_{\hat{P}} \left\{ J_e(\hat{P}) \right\} = \arg \max_{\hat{P}} \left\{ J_v(\hat{P}) \right\} \quad (2.7)$$

Pour déterminer les vecteurs colonnes de la matrice $\hat{\mathbf{P}}$, on note par $t \in \mathbb{R}$ la projection du vecteur de données \mathbf{x} le long d'une direction représentée par un vecteur unitaire $p \in \mathbb{R}^m$. La composante t est obtenue par le produit scalaire $t = \mathbf{x}^T p = p^T \mathbf{x}$ sous la contrainte $\|p\|^2 = p^T p = 1$. Notamment, elle représente une nouvelle variable ayant une moyenne et une variance qui dépendent des propriétés statistiques de \mathbf{x} comme suit :

$$\mathbb{E} \{t\} = \mathbb{E} \{p^T \mathbf{x}\} = p^T \mathbb{E} \{\mathbf{x}\} = 0 \quad (2.8)$$

$$\begin{aligned} \text{Var} \{t\} &= \mathbb{E} \{(t - \mathbb{E} \{t\})^2\} = \mathbb{E} \{t^2\} \\ &= \mathbb{E} \{(p^T \mathbf{x}) (\mathbf{x}^T p)\} = p^T \mathbb{E} \{\mathbf{x} \mathbf{x}^T\} p \\ &= p^T \Sigma p \end{aligned} \quad (2.9)$$

La maximisation de la variance de projection, sous condition d'une norme unité du vecteur p , représente un problème d'optimisation sous contrainte égalité qui peut être formalisé par la fonction de Lagrange :

$$\mathcal{L}(p, \lambda) = J_v(p) - \lambda (p^T p - 1) = p^T \Sigma p - \lambda (p^T p - 1) \quad (2.10)$$

où $\lambda \in \mathbb{R}$ désigne le multiplicateur de Lagrange. En tenant compte de la symétrie de la matrice Σ , le vecteur p maximisant le critère d'optimisation J_v est solution du système d'équations suivant :

$$\begin{cases} \partial \mathcal{L}(p, \lambda) / \partial p = \Sigma p - \lambda p = 0 \\ \partial \mathcal{L}(p, \lambda) / \partial \lambda = p^T p - 1 = 0 \end{cases} \quad (2.11)$$

Par conséquent, la résolution de ce système d'équations est identifiée comme un problème d'estimation de valeurs et vecteurs propres normalisés de la matrice Σ . Un tel système d'équations admet des solutions réelles de la variables λ obtenues par résolution de l'équation caractéristique suivante :

$$\text{Det} \{\Sigma - \lambda \mathbf{I}_m\} = 0 \quad (2.12)$$

où $\text{Det} \{.\}$ représente le déterminant d'une matrice carrée. \mathbf{I}_m est la matrice identité d'ordre m . Les solutions de l'équation précédente représentent les valeurs propres de Σ . A chaque valeur propre λ est associé un vecteur propre p vérifiant $(\Sigma - \lambda \mathbf{I}_m)p = 0$.

Ceci permet d'avoir m vecteurs propres \mathbf{p}_i associés aux m valeurs propres λ_i de la matrice Σ vérifiant ainsi la relation $\Sigma \mathbf{p}_i = \lambda_i \mathbf{p}_i$ avec $i = \{1, \dots, m\}$. Sous forme matricielle, une telle relation mène à écrire ce qui suit :

$$\Sigma \mathbf{P} = \mathbf{P} \Lambda \quad (2.13)$$

$\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_m] \in \mathbb{R}^{m \times m}$ représente la matrice de projection de données. Elle est orthonormée puisque ses colonnes correspondent aux vecteurs propres de Σ :

$$\mathbf{P}^T \mathbf{P} = \mathbf{P} \mathbf{P}^T = \mathbf{I}_m \in \mathbb{R}^{m \times m} \quad (2.14)$$

$\Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_m \} \in \mathbb{R}^{m \times m}$ représente la matrice diagonale constituée en éléments diagonaux des valeurs propres de Σ .

D'après les équations (2.13) et (2.14), on peut déduire que $\mathbf{P}^T \Sigma \mathbf{P} = \Lambda$. Ceci nous permet de conclure que la première direction, ayant une variance maximale de projection de données \mathbf{x} , est portée par le vecteur propre \mathbf{p}_1 associé à la plus grande valeur propre λ_1 . Cette dernière représente la variance d'une telle direction. Le second axe factoriel aussi rend la variance maximale tout en étant orthogonal au premier. Sa variance λ_2 est moins importante que celle qui correspond à la première direction. Par conséquent, les éléments diagonaux de Λ sont arrangés dans l'ordre décroissant : $\lambda_1 \geq \dots \geq \lambda_m$.

En considérant la matrice \mathbf{P} , le vecteur de données $\mathbf{x}(k)$ peut se transformer sans aucune perte d'informations en un vecteur de composantes principales (CPs) :

$$\mathbf{t}(k) = [t_1(k), \dots, t_m(k)]^T = \mathbf{P}^T \mathbf{x}(k) \in \mathbb{R}^m \quad (2.15)$$

où les CPs t_j avec $j = \{1, \dots, m\}$ sont définies par :

$$t_j(k) = \mathbf{p}_j^T \mathbf{x}(k) = \mathbf{x}^T(k) \mathbf{p}_j \quad (2.16)$$

Celles-ci sont statistiquement non corrélées :

$$\mathbb{E} \{ t_i t_j \} = \mathbb{E} \{ \mathbf{p}_i^T \mathbf{x} \mathbf{x}^T \mathbf{p}_j \} = \mathbf{p}_i^T \Sigma \mathbf{p}_j = 0 \quad i \neq j \quad (2.17)$$

La notation sous forme matricielle nous permet de définir la matrice des CPs comme suit :

$$\mathbf{T} = [\mathbf{t}(1), \dots, \mathbf{t}(N)]^T = \mathbf{X} \mathbf{P} \in \mathbb{R}^{N \times m} \quad (2.18)$$

La détermination du vecteur de données $\mathbf{x}(k)$ à partir du vecteur associé des CPs $\mathbf{t}(k)$ est donnée par :

$$\mathbf{x}(k) = \mathbf{P} \mathbf{t}(k) = \sum_{j=1}^m \mathbf{p}_j t_j(k) \quad (2.19)$$

La réduction des données est réalisée à travers les ℓ premières CPs ayant les plus grandes variances. En conséquence, les ℓ premiers vecteurs propres forment le sous-espace vectoriel réduit pour les données initiales. L'estimation $\hat{\mathbf{x}}(k)$ du vecteur de données $\mathbf{x}(k)$ dans ce sous-espace réduit (souvent appelé sous-espace de représentation ou principal et noté $\hat{\mathcal{S}}$) est donnée par :

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{P}} \hat{\mathbf{t}}(k) = \hat{\mathbf{P}} \hat{\mathbf{P}}^T \mathbf{x}(k) = \hat{\mathbf{C}} \mathbf{x}(k) \quad (2.20)$$

où la matrice optimale de représentation exprimée dans l'équation (2.7) est définie comme suit :

$$\hat{\mathbf{P}} = [\mathbf{p}_1, \dots, \mathbf{p}_\ell] \in \mathbb{R}^{m \times \ell} \quad (2.21)$$

$\hat{\mathbf{t}}(k) = \hat{\mathbf{P}}^T \mathbf{x}(k) \in \mathbb{R}^\ell$ représente le vecteur des ℓ premières CPs. La matrice $\hat{\mathbf{C}} \in \mathbb{R}^{m \times m}$ caractérise ainsi le modèle ACP.

Toutefois, la réduction de dimension engendre généralement une perte d'informations qui sont récupérées dans un vecteur résiduel $\tilde{\mathbf{x}}(k)$. Ce dernier est exprimé dans un sous-espace résiduel $\tilde{\mathcal{S}}$ constitué par le reste des CPs associées aux $(m - \ell)$ derniers vecteurs propres :

$$\tilde{\mathbf{x}}(k) = \tilde{\mathbf{P}}\tilde{\mathbf{t}}(k) = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \mathbf{x}(k) = \tilde{\mathbf{C}}\mathbf{x}(k) \quad (2.22)$$

avec

$$\tilde{\mathbf{P}} = [\mathbf{p}_{\ell+1}, \dots, \mathbf{p}_m] \in \mathbb{R}^{m \times (m-\ell)} \quad (2.23)$$

et

$$\tilde{\mathbf{C}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I}_m - \hat{\mathbf{C}} \quad (2.24)$$

La matrice $\tilde{\mathbf{C}} \in \mathbb{R}^{m \times m}$ décrit le modèle résiduel. On entrevoit ici que l'ACP est une approche de modélisation permettant ainsi l'obtention d'un modèle ACP d'un système étudié.

L'interprétation du principe de la modélisation par ACP représente un partitionnement de l'espace \mathbb{R}^m des mesures $\mathbf{x}(k)$ en un sous-espace principal $\hat{\mathcal{S}}$ et un sous-espace résiduel $\tilde{\mathcal{S}}$. Par conséquent, le vecteur de mesures $\mathbf{x}(k)$ est décomposé comme suit :

$$\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \tilde{\mathbf{x}}(k) \quad (2.25)$$

Notamment, une propriété géométrique d'orthogonalité entre le vecteur estimé et celui résiduel est toujours vérifiée puisque :

$$\tilde{\mathbf{C}}\hat{\mathbf{C}} = \hat{\mathbf{C}}\tilde{\mathbf{C}} = \mathbf{0}_m \in \mathbb{R}^{m \times m} \quad (2.26)$$

Celle-ci implique que le sous-espace principal et le sous-espace résiduel sont orthogonaux pour toutes valeurs de ℓ . Ainsi,

$$\tilde{\mathbf{x}}^T(k)\hat{\mathbf{x}}(k) = 0 \quad (2.27)$$

Une analyse de la robustesse de l'ACP par rapport au bruit de mesures mène à la définition de quelques propriétés relatives à la détermination d'un modèle ACP.

2.3 Propriétés

Généralement, la présence de bruit de mesures est inévitable dans les données. Sous l'hypothèse de l'absence de perturbations et de défauts, il est possible de considérer que le vecteur \mathbf{x} est perturbé par un bruit $\mathbf{v} \in \mathbb{R}^m$ de moyenne nulle :

$$\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \mathbf{v}(k) \quad (2.28)$$

où $\hat{\mathbf{x}}(k)$ représente le vecteur de données en absence de bruit de mesures.

2.3.1 Modélisation en absence de bruit

Notons par $\overset{\circ}{\Sigma}$ la matrice de covariance des données $\overset{\circ}{\mathbf{x}}$. Sa décomposition en valeurs singulières peut s'exprimer comme suit :

$$\overset{\circ}{\Sigma} = \frac{1}{N} \overset{\circ}{\mathbf{X}}^T \overset{\circ}{\mathbf{X}} = \overset{\circ}{\mathbf{P}} \overset{\circ}{\Lambda} \overset{\circ}{\mathbf{P}}^T \quad (2.29)$$

où $\overset{\circ}{\mathbf{X}}$ représente la matrice des données supposées être sans bruit. Elle est exprimée en fonction des observations $\overset{\circ}{\mathbf{x}}(k)$ comme suit :

$$\overset{\circ}{\mathbf{X}} = [\overset{\circ}{\mathbf{x}}(1), \dots, \overset{\circ}{\mathbf{x}}(N)]^T \in \mathbb{R}^{N \times m} \quad (2.30)$$

En effet, le nombre optimal ℓ des CPs doit correspondre au rang q de la matrice $\overset{\circ}{\Sigma}$ car en présence des corrélations linéaires entre les variables, les $(m - q)$ dernières valeurs propres de $\overset{\circ}{\Sigma}$ sont nulles. Ainsi, les partitionnements appropriés des matrices $\overset{\circ}{\mathbf{P}}$ et $\overset{\circ}{\Lambda}$ conduisent à réécrire cette dernière de la façon suivante :

$$\overset{\circ}{\Lambda} = \begin{bmatrix} \hat{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(m-q)} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}}^T \\ \tilde{\mathbf{P}}^T \end{bmatrix} \overset{\circ}{\Sigma} \begin{bmatrix} \hat{\mathbf{P}} & \tilde{\mathbf{P}} \end{bmatrix} \quad (2.31)$$

où $\hat{\Lambda} \in \mathbb{R}^{q \times q}$ représente la matrice diagonale composée des q valeurs propres non nulles :

$$\hat{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_q\} \quad (2.32)$$

λ_ℓ , avec $\ell = \{1, \dots, q\}$, décrit la variance de la ℓ ème CP des données étudiées en l'absence de bruit de mesures. D'après l'équation (2.31), on peut déduire ce qui suit :

$$\tilde{\mathbf{P}}^T \overset{\circ}{\Sigma} \tilde{\mathbf{P}} = \mathbf{0}_{(m-q)} \in \mathbb{R}^{(m-q) \times (m-q)} \quad (2.33)$$

Ainsi, la substitution de la matrice $\overset{\circ}{\Sigma}$ par son expression donnée dans (2.29) mène à la relation suivante :

$$\tilde{\mathbf{P}}^T \overset{\circ}{\mathbf{X}}^T \overset{\circ}{\mathbf{X}} \tilde{\mathbf{P}} = \left(\overset{\circ}{\mathbf{X}} \tilde{\mathbf{P}} \right)^T \left(\overset{\circ}{\mathbf{X}} \tilde{\mathbf{P}} \right) = \mathbf{0}_{(m-q)} \quad (2.34)$$

Celle-ci implique que :

$$\overset{\circ}{\mathbf{X}} \tilde{\mathbf{P}} = \mathbf{0} \in \mathbb{R}^{N \times (m-q)} \quad (2.35)$$

En tenant compte des observations $\overset{\circ}{\mathbf{x}}(k)$ qui constituent la matrice $\overset{\circ}{\mathbf{X}}$, on peut également déduire que :

$$\tilde{\mathbf{P}}^T \overset{\circ}{\mathbf{x}}(k) = \mathbf{0} \in \mathbb{R}^{(m-q)} \quad (2.36)$$

La relation précédente forme un système composé de $(m - q)$ équations linéaires. En effet, les $(m - q)$ valeurs propres nulles de $\overset{\circ}{\Sigma}$ impliquent l'existence de $(m - q)$ relations linéaires entre les variables. Ainsi, ces équations représentent celles de redondance présentes entre les variables étudiées. En absence de bruit de mesures, les q valeurs propres non nulles indiquent l'existence de q variables linéairement indépendantes.

Exemple 2.1 Détermination d'un modèle ACP en absence de bruit

Afin d'illustrer un exemple, on considère le système suivant :

$$\begin{cases} y_1(k) = 2u_1(k) - u_2(k) \\ y_2(k) = u_2(k) + u_3(k) \end{cases} \quad (2.37)$$

y_1 et y_2 représentent les signaux de sorties d'un tel système. u_1 , u_2 et u_3 sont les signaux d'entrées dont les expressions en fonction du temps sont les suivantes :

$$\begin{cases} u_1(k) = 1 + \sin(k/3) \\ u_2(k) = 2 \cos(k/4) \exp(-k/N) \\ u_3(k) = \log(u_2(k)^2) \end{cases} \quad (2.38)$$

On a construit une matrice de données \mathbf{X} composée de $N = 2700$ observations. Chacune de ces dernières représente un vecteur des mesures associées aux variables étudiées : $\mathbf{x}(k) = [u_1(k) \ u_2(k) \ u_3(k) \ y_1(k) \ y_2(k)]^T$. La décomposition en valeurs singulières de la matrice de covariance $\hat{\Sigma}$ de \mathbf{X} mène aux valeurs propres données par le tableau 2.1.

13.026	5.276	1.315	0	0
--------	-------	-------	---	---

TABLE 2.1 – Valeurs propres de $\hat{\Sigma}$ de l'exemple 2.1

On remarque l'existence de deux valeurs propres nulles ce qui explique l'existence de deux équations de redondance dans le système. Les trois valeurs propres non nulles indiquent la présence de trois variables non corrélées qui représentent les entrées du système dans cet exemple.

Le sous-espace résiduel est engendré par les deux derniers vecteurs propres dont les transposées sont exprimées comme suit :

$$\tilde{\mathbf{P}}^T = \begin{bmatrix} 0.410 & 0.366 & 0.572 & -0.205 & -0.572 \\ 0.733 & -0.526 & -0.160 & -0.366 & 0.160 \end{bmatrix} \quad (2.39)$$

On peut alors déduire que les équations de redondance sont les suivantes :

$$\begin{cases} 0.410u_1(k) + 0.366u_2(k) + 0.572u_3(k) - 0.205y_1(k) - 0.572y_2(k) = 0 & \text{(a)} \\ 0.733u_1(k) - 0.526u_2(k) - 0.160u_3(k) - 0.366y_1(k) + 0.160y_2(k) = 0 & \text{(b)} \end{cases} \quad (2.40)$$

A partir de l'équation (2.40.a), la variable y_1 peut être exprimée en fonction des autres variables de la manière suivante :

$$y_1(k) = \frac{1}{0.205} (0.410u_1(k) + 0.366u_2(k) + 0.572u_3(k) - 0.572y_2(k)) \quad (2.41)$$

En remplaçant la variable y_1 de l'équation (2.40.b) par son expression donnée dans (2.41), on peut déduire que :

$$y_2(k) = u_2(k) + u_3(k) \quad (2.42)$$

Ainsi, la réinjection de celle-ci dans l'équation (2.41) mène à :

$$y_1(k) = 2u_1(k) - u_2(k) \quad (2.43)$$

On peut alors conclure qu'il est possible de retrouver à partir des équations de redondance le système initial donné par (2.37).

2.3.2 Modélisation en présence de bruit

D'après [Li et Qin \(2001\)](#), l'ACP fournit un modèle non biaisé uniquement dans le cas particulier où les mesures des variables du processus sont entachées par un bruit blanc (indépendant et identiquement distribué : *i.i.d.*).

Lorsque le bruit \mathbf{v} est *i.i.d.* de variance identique σ^2 , l'expression de la matrice de covariance des données bruitées est la suivante :

$$\Sigma = \mathbb{E}\{\mathbf{xx}^T\} = \mathbf{P}\Lambda\mathbf{P}^T = \mathring{\Sigma} + \sigma^2\mathbf{I}_m \quad (2.44)$$

Ainsi, [Anderson \(1963\)](#) a montré que les vecteurs propres des matrices Σ et $\mathring{\Sigma}$ sont identiques. Pour $\ell = q$ CPs, il est alors possible d'exprimer la matrice $\tilde{\mathbf{C}}$ comme suit :

$$\tilde{\mathbf{C}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \quad (2.45)$$

Par conséquent, le vecteur résiduel ainsi que celui estimé d'une observation bruitée $\mathbf{x}(k)$ sont respectivement exprimés de la manière suivante :

$$\begin{aligned} \tilde{\mathbf{x}}(k) &= \tilde{\mathbf{C}}(\mathring{\mathbf{x}}(k) + \mathbf{v}(k)) \\ &= \tilde{\mathbf{C}}\mathbf{v}(k) \end{aligned} \quad (2.46)$$

et

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{C}}(\mathring{\mathbf{x}}(k) + \mathbf{v}(k)) \quad (2.47)$$

En présence de bruit *i.i.d.*, le modèle ACP est constitué de q CPs. Dans ce cas et d'après l'équation (2.46), les données sans bruit ne sont pas projetées dans le sous-espace résiduel. En effet, ce dernier ne peut contenir que le bruit de mesures. Cependant, le sous-espace principal peut contenir les données non bruitées ainsi que le bruit.

2.4 Détermination d'une structure optimale du modèle ACP

En présence de bruit de mesures, les valeurs propres nulles de la matrice de covariance indiquent l'existence de relations linéaires et les plus petites valeurs propres indiquent l'existence de relations quasi-linéaires. Le choix du nombre de CPs joue donc un rôle crucial dans la détermination des relations de redondance entre les variables. Toutefois, un tel nombre révèle une difficulté pour sa détermination. Pour cela, de nombreux critères ont été proposés dans la littérature afin d'offrir des solutions aidant dans le choix de la structure adaptée du modèle ACP. D'une façon non exhaustive, nous avons choisi certains critères parmi les plus connus afin de réaliser une étude comparative sur leurs performances.

2.4.1 Critères de la théorie de l'information

Dans le domaine du traitement du signal, l'identification des sources indépendantes des signaux à partir d'un nombre fini d'observations bruitées est un problème crucial. Les solutions proposées dans ce cadre peuvent être explorées avec l'ACP pour déterminer le

nombre des signaux non corrélés. En effet, un vecteur d'observation peut être modélisé comme une superposition d'un nombre fini de signaux noyés dans un bruit additif. L'objectif est d'identifier ces signaux. Sous l'hypothèse que le bruit de mesures est *i.i.d.*, sa variance doit correspondre aux plus petites valeurs propres de la matrice de covariance. En se basant sur le principe de la vraisemblance, deux critères **AIC** (Akaike, 1973) et **MDL** (Rissanen, 1978) et (Schwarz, 1978) ont été proposés puis reformulés et adaptés par Wax et Kailath (1985) afin d'être utiles dans le choix du nombre des CPs significatives. En effet, le nombre des signaux non corrélés à identifier doit correspondre aux minima des critères **AIC** et **MDL** dont les expressions sont respectivement données par :

$$\mathbf{AIC}(\ell) = -2 \log f(\mathbf{X}|\hat{\theta}) + 2G(\hat{\theta}) \quad (2.48)$$

$$\mathbf{MDL}(\ell) = -2 \log f(\mathbf{X}|\hat{\theta}) + G(\hat{\theta}) \log N \quad (2.49)$$

avec

$$f(\mathbf{X}|\hat{\theta}) = N \sum_{a=\ell+1}^m \log \lambda_a - N(m - \ell) \log \left(\frac{1}{m - \ell} \sum_{a=\ell+1}^m \lambda_a \right) \quad (2.50)$$

et

$$G(\hat{\theta}) = \ell(2m - \ell) \quad (2.51)$$

où la fonction log représente l'opérateur du logarithme népérien. On remarque que les premiers termes (à droite) des expressions (2.48) et (2.49) sont identiques et décroissants en ℓ . En revanche, les seconds termes sont croissants en ℓ . Théoriquement, il existe un minimum qui correspond à un nombre de CPs pour chacun des deux critères. Notamment, Wax et Kailath (1985) ont montré que le critère **MDL** est capable d'estimer une structure adaptée du modèle ACP sous l'hypothèse que le nombre d'observations considérées soit assez important. Tandis que celui d'**AIC** présente une tendance asymptotique de surestimation du nombre des CPs retenues.

2.4.2 Critères heuristiques

La qualité des estimations par ACP dépend du choix du nombre des CPs retenues ou encore de la dimension du sous-espace de représentation. Dans la suite, nous présentons d'une façon non exhaustive les critères heuristiques les plus couramment utilisés.

2.4.2.1 Critère IE

L'analyse factorielle est une méthode conçue pour résoudre les problèmes multidimensionnels. Elle exprime un ensemble de données sous forme d'une somme linéaire des produits de fonctions. Ainsi, une réduction est réalisée afin de reproduire ces données à partir d'un sous-espace composé uniquement des variables latentes significatives qui ont été déterminées par l'analyse factorielle. Notamment, la première étape dans le processus de cette méthode fait appelle à l'ACP qui consiste à déterminer ce nombre de facteurs cachés. Puisque la reproduction des données engendre nécessairement des erreurs, Malinowski (1977) distingue trois types d'erreurs qui sont : *real error* (**RE**), *imbedded error*

(**IE**) et *extracted error* (**XE**). En investiguant de près le comportement de la fonction **IE**, cette dernière a été considérée par Malinowski (1977) comme un critère capable de déterminer la dimension d'un modèle ACP. Son expression dépend principalement des valeurs propres résiduelles :

$$\mathbf{IE}(\ell) = \left(\frac{\ell}{Nm(m-\ell)} \sum_{a=\ell+1}^m \lambda_a \right)^{\frac{1}{2}} \quad (2.52)$$

Chaque variable latente ou CP sélectionnée est composée des mesures observées et des erreurs. Lorsque le sous-espace constitué par les variables latentes retenues n'est pas encore optimal, le critère **IE** décroît en ℓ en représentant par conséquent une mixture des mesures observées et des erreurs. Au point où toute l'information normale est extraite, **IE** devient croissant en ℓ en exprimant uniquement les erreurs. Une représentation convenable des données est assurée par le nombre des CPs qui correspondent à la valeur minimale de ce critère.

Bien que les critères **AIC**, **MDL** et **IE** soient développés dans différents domaines, ils représentent deux points communs. En effet, ils se basent uniquement sur des valeurs propres issues de la matrice de covariance des données. Par conséquent, ces dernières ne doivent pas être normalisées. En outre, le bruit de mesures des variables est supposé être indépendant et identiquement distribué.

2.4.2.2 Pourcentage cumulé de la variance

Les valeurs propres de la matrice de corrélation représentent les variances des CPs. Par conséquent, elles ont été utilisées pour définir un critère **PCV** dans le but est la détermination de la structure optimale d'un modèle ACP. Ce critère représente le pourcentage de la variance expliquée par les CPs retenues comme suit :

$$\mathbf{PCV}(\ell) = 100 \left(\frac{\sum_{a=1}^{\ell} \lambda_a}{\sum_{a=1}^m \lambda_a} \right) \% \quad (2.53)$$

Généralement, on essaie de retenir avec ce critère un nombre de CPs ayant un **PCV** compris entre 90% et 95% ou 99% de la variance totale. La règle de décision basée sur un tel critère est très subjective, car il s'agit d'une réalisation d'un compromis entre une variance maximale et un nombre minimal de CPs retenues. Le choix de la dimension du modèle ACP est souvent difficile puisque ce critère est monotone croissant en ℓ . En outre, son efficacité à fournir le nombre optimal des CPs dépend fortement du rapport signal sur bruit puisque la variance du bruit est inconnue.

2.4.2.3 Scree Test

Le *Scree Test* ou test du coude est une approche empirique proposée principalement par Cattell (1966) pour la sélection du nombre optimal des CPs. Elle est basée sur un

test du coude de la courbe du pourcentage de la variance résiduelle dont l'expression est la suivante :

$$\mathbf{PVR}(\ell) = 100 \left(\frac{\sum_{a=\ell+1}^m \lambda_a}{\sum_{a=1}^m \lambda_a} \right) \% \quad (2.54)$$

Ce critère présente une allure décroissante en ℓ . Son principe est fondé sur l'idée que la variance résiduelle devrait atteindre un état stationnaire lorsque les CPs ignorées ressemblent à des erreurs aléatoires. Ainsi, le nombre des CPs à retenir est relatif au premier point d'inflexion détecté sur la courbe.

On observe le graphique du critère **PVR** et on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. Graphiquement, on part des composantes qui se trouvent à droite, apportant le moins d'informations. On relie par une droite les points presque alignés et on ne retient que les CPs qui sont au dessus de cette ligne. La mise en œuvre de cette méthode est relativement facile, cependant dans certains cas il est difficile de trouver un point d'inflexion ou le coude si la courbe décroît lentement.

2.4.2.4 Critère de Guttman

Le critère de Guttman pour la détermination du nombre approprié des variables latentes en analyse factorielle (ou le nombre des CPs significatives en ACP) a été initialement proposé par [Guttman \(1954\)](#) puis adapté et popularisé par [Kaiser \(1961\)](#). En s'appuyant sur un tel critère, on ne compte que les composantes ou facteurs avec des valeurs propres supérieures à l'unité qui représente la moyenne arithmétique de toutes les valeurs propres de la matrice de corrélation. L'expression de ce critère en ℓ est la suivante :

$$\mathbf{KG}(\ell) = \lambda_\ell \quad (2.55)$$

Chaque CP retenue dans le modèle ACP contribue par sa variance. Ainsi, sa contribution est considérée significative si elle dépasse la moyenne totale, sinon elle devrait être écartée. Cette idée est justifiée par [Guttman \(1954\)](#) afin de fournir une borne inférieure pour le nombre des variables latentes représentatives en considérant une matrice de corrélation. Plus intuitivement, l'argument a été avancé afin d'exprimer qu'aucune CP dont la variance est inférieure à celle d'une variable originelle ne peut être considérée comme représentative.

La popularité du critère **KG**, par rapport à d'autres plus opérationnels et mieux justifiables, apparaît plus particulièrement dans sa simplicité d'utilisation. Dans un cadre d'une étude par simulation, [Yeomans et Golder \(1982\)](#) ont examiné de plus près le comportement de ce critère afin de montrer l'ampleur probable des erreurs introduites par son utilisation sans précautions. Le seul cas où ce critère est efficace semble bien être lorsque le nombre des composantes représentatives est beaucoup moins inférieur que celui des variables originelles. Ainsi, la proportion de la variance de chaque variable, expliquée par les CPs retenues, doit être élevée.

2.4.2.5 Autocorrélation

L'autocorrélation est une méthode qualitative proposée pour le choix d'un modèle ACP. Généralement, les allures des CPs les plus significatives doivent être lisses tandis que les autres liées principalement aux bruits présentent des fluctuations rapides et aléatoires. Dans ce cadre, [Shrager et Hendler \(1982\)](#) ont proposé la fonction d'autocorrélation du premier ordre comme une mesure quantitative de ce comportement :

$$\mathbf{AC}(\ell) = \frac{\mathbb{E}\{t_\ell(k)t_\ell(k+1)\}}{\text{Var}\{t_\ell(k)\}} = \frac{1}{(N-1)\lambda_\ell} \sum_{k=1}^{N-1} t(k, \ell)t(k+1, \ell) \quad (2.56)$$

où $t(k, \ell)$ représente la k ème valeur de la ℓ ème colonne de la matrice des CPs notée \mathbf{T} dans l'équation (2.18).

Lorsqu'une CP est significative, son autocorrélation sera importante car il y aura certainement une corrélation entre ses éléments qui correspondent aux différentes observations. En revanche et si elle est principalement associée à un bruit de mesures, peu de corrélation est attendue entre ses éléments qui sont fortement aléatoires. Par conséquent, son autocorrélation sera relativement faible. D'après [Shrager et Hendler \(1982\)](#), une valeur d'autocorrélation supérieure à 0.5 indique que la CP correspondante est significative. Dans le cas contraire, la composante en question est constituée principalement de bruit. Dans ce cas, elle ne devrait pas être incluse dans le modèle. Le choix de la valeur d'un tel seuil est considérablement arbitraire, ce qui représente l'inconvénient de ce critère. En outre, une CP ayant une grande variance peut correspondre à une faible valeur d'autocorrélation, ce qui lui risque d'être exclue du modèle.

2.4.2.6 Validation croisée

L'idée de la validation croisée est basée principalement sur la prédiction à travers un modèle ACP d'une mesure $x_i(k)$ de la matrice de données \mathbf{X} . Selon [Wold \(1978\)](#) et [Eastment et Krzanowski \(1982\)](#), le nombre des CPs retenues est optimal si la moyenne de la prédiction globale d'une telle mesure n'est plus significativement améliorée par l'ajout de CPs supplémentaires. Ainsi, le modèle optimal est alors constitué du nombre minimum des CPs nécessaires pour une prédiction adéquate.

Dans ce cadre, le critère qui découle de l'approche de la validation croisée représente une minimisation de la somme des carrés des erreurs de prédiction. Il est connu sous le nom de **PRESS** :

$$\mathbf{PRESS}(\ell) = \frac{1}{Nm} \sum_{k=1}^N \sum_{i=1}^m (\hat{x}_i^{(\ell)}(k) - x_i(k))^2 \quad (2.57)$$

où $\hat{x}_i^{(\ell)}(k)$ représente la prédiction de $x_i(k)$, qui correspond à la k ème mesure de la i ème variable, en utilisant un modèle ACP constitué de ℓ CPs.

Toutefois, on distingue dans la littérature deux façons différentes pour le calcul de ce critère, car la manière de la prédiction proposée par [Wold \(1978\)](#) diffère de celle proposée par [Eastment et Krzanowski \(1982\)](#). Indépendamment de cette différence, il est important

de mentionner que ce critère présente une complexité dans son implémentation ainsi qu'un coût de calcul important.

Par ailleurs, Besse et Ferré (1993) ont montré théoriquement que l'usage du critère **PRESS** n'apporte pas une règle de décision plus objective que les critères heuristiques. Sous l'hypothèse que le nombre d'observations considérées est très important, un développement de Taylor a permis à ces auteurs de montrer que la quantité **PRESS** peut être approximée comme suit :

$$\mathbf{PRESS}(\ell) \approx \frac{1}{m} \sum_{a=\ell+1}^m \lambda_a \quad (2.58)$$

Ce critère est alors décroissant en ℓ . Par conséquent, l'idée d'identifier la CP qui correspond au minimum de **PRESS** ne peut servir dans le choix d'un modèle ACP. En outre, ce critère est équivalent à ceux prenant simplement la part de la variance résiduelle en particulier le critère **PVR**. Tandis que, Wold (1978) et Eastment et Krzanowski (1982) ont également proposé l'investigation d'autres critères issues de la quantité **PRESS** qui sont respectivement le ratio **R** et le critère **W**. L'utilisation des ces derniers pour le choix de la dimension d'un modèle ACP est basée sur une comparaison de leurs valeurs à des seuils jugés arbitraires dans la littérature limitant par conséquence de leurs efficacités. Malgré la célébrité de la validation croisée, cette dernière n'est plus considérée avantageuse par rapport aux restes des critères heuristiques. Pour cette raison, elle ne présentera pas l'objectif de notre étude dans ce chapitre.

2.4.3 Minimisation de la variance de l'erreur de reconstruction

Le critère qui représente un objectif principal d'une étude dans ce chapitre est lié à la variance de l'erreur de reconstruction ou la variance non reconstruite (**VNR**) (Dunia et Qin, 1998b,c,a; Qin et Dunia, 2000). En effet, la qualité de la reconstruction est liée à la capacité du modèle ACP à modéliser les relations de redondance entre les variables. Ainsi, ce critère aide à définir le nombre des CPs exprimant une meilleure reconstruction.

Le principe de reconstruction représente une élimination de l'effet d'un défaut. Autrement dit, cette approche estime le vecteur d'amplitudes d'un tel défaut.

Pour une raison de simplifications qui seront utiles pour le reste des chapitres, nous définissons ici d'une manière générale l'expression d'un vecteur de données en présence de défauts multidimensionnels. Nous notons par $\mathcal{S}_{\mathcal{J}}$ le sous-espace caractéristique d'un défaut réel $\mathcal{F}_{\mathcal{J}}$ affectant un ensemble \mathcal{J} constitué de r variables où $r = \dim(\mathcal{S}_{\mathcal{J}})$. Nous considérons également $\Xi_{\mathcal{J}} \in \mathbb{R}^{m \times r}$ une base orthonormée pour un tel sous-espace dont les colonnes sont celles de la matrice identité et qui correspondent aux variables en défaut. Posons $\mathbf{x}^*(k)$ qui est supposé inconnu le vecteur de mesures collectées lors du fonctionnement normal. En présence du défaut $\mathcal{F}_{\mathcal{J}}$, le vecteur $\mathbf{x}(k)$ peut s'écrire comme suit :

$$\mathbf{x}(k) = \mathbf{x}^*(k) + \Xi_{\mathcal{J}} \mathbf{f}(k) \quad (2.59)$$

où $\mathbf{f}(k) \in \mathbb{R}^r$ représente le vecteur d'amplitudes des composants de défaut à l'instant k .

On formule l'hypothèse que les données utilisées pour construire le modèle ACP sont "saines". Les données "saines" correspondent à des données obtenues lors du fonctionnement normal du système étudié.

2.4.3.1 Critère VNR

Dans l'objectif de choisir une structure adaptée du modèle ACP en s'appuyant sur l'approche de reconstruction, on suppose que les défauts sont unidimensionnels ou simples. Cela signifie que la matrice orthonormée des directions $\Xi_{\mathcal{J}}$ ainsi que le vecteur d'amplitudes des composants de défaut $\mathbf{f}(k)$ dans l'équation (2.59) sont respectivement remplacés par un vecteur ξ_j qui représente la $j^{\text{ème}}$ colonne de la matrice identité et $f(k)$ un scalaire qui désigne l'amplitude de défaut caractérisant la $j^{\text{ème}}$ variable. Dans ce cadre, la reconstruction d'une telle variable aboutit à un vecteur reconstruit d'observation ayant l'expression suivante :

$$\begin{aligned}\mathbf{x}_j(k) &= \mathbf{x}^*(k) = \mathbf{x}(k) - \xi_j f(k) \\ &= \mathbf{x}(k) - \xi_j \hat{f}_j(k)\end{aligned}\quad (2.60)$$

où $\hat{f}_j(k)$ est l'estimée de $f(k)$ dans la direction de la $j^{\text{ème}}$ variable. En effet, l'estimation de l'amplitude f est optimale par minimisation de l'erreur quadratique résiduelle $\|\tilde{\mathbf{C}}\mathbf{x}_j(k)\|^2$ comme suit :

$$\begin{aligned}\hat{f}_j(k) &= \arg \min_{f(k)} \left\{ \|\tilde{\mathbf{C}}\mathbf{x}_j(k)\|^2 \right\} \\ &= (\xi_j^T \tilde{\mathbf{C}}\xi_j)^{-1} \xi_j^T \tilde{\mathbf{C}}\mathbf{x}(k)\end{aligned}\quad (2.61)$$

Par conséquent, le vecteur reconstruit d'observation aura la forme suivante :

$$\mathbf{x}_j(k) = (\mathbf{I}_m - \xi_j (\xi_j^T \tilde{\mathbf{C}}\xi_j)^{-1} \xi_j^T \tilde{\mathbf{C}}) \mathbf{x}(k) \quad (2.62)$$

Certainement, la reconstruction n'est possible que sous conditions. En effet, la $j^{\text{ème}}$ variable est reconstituible si $\xi_j^T \tilde{\mathbf{C}}\xi_j \neq 0$. Cela implique que le vecteur $\tilde{\mathbf{C}}\xi_j$ n'est pas nul.

Après avoir présenté le principe de reconstruction, la variance non reconstruite de la $j^{\text{ème}}$ variable est définie par :

$$\begin{aligned}\sigma_j^2(\ell) &= \text{Var} \left\{ \xi_j^T (\mathbf{x} - \mathbf{x}_j) \right\} = \text{Var} \left\{ \hat{f}_j \right\} \\ &= \frac{\xi_j^T \tilde{\mathbf{C}} \mathbb{E} \{ \mathbf{x}\mathbf{x}^T \} \tilde{\mathbf{C}}\xi_j}{(\xi_j^T \tilde{\mathbf{C}}\xi_j)^2} = \frac{\xi_j^T \tilde{\mathbf{C}}\Sigma\tilde{\mathbf{C}}\xi_j}{(\xi_j^T \tilde{\mathbf{C}}\xi_j)^2}\end{aligned}\quad (2.63)$$

On déduit qu'une telle variance représente tout simplement la variance de l'amplitude estimée du défaut. Une telle variance justifie l'idée de son utilisation pour le choix d'un modèle ACP. En effet, les données utilisées pour la modélisation sont supposées être sans défauts. Ainsi, la reconstruction d'une variable donnée en utilisant un modèle judicieusement choisi doit fournir une variance non reconstruite minimale. Puisqu'on dispose de plusieurs variables qui doivent être étudiées simultanément, on doit alors identifier un modèle ACP qui minimise la variance globale non reconstruite (**VNR**) représentée par la somme suivante :

$$\mathbf{VNR}(\ell) = \sum_{j=1}^m \frac{\sigma_j^2(\ell)}{\xi_j^T \Sigma \xi_j} = \sum_{j=1}^m \frac{\xi_j^T \tilde{\mathbf{C}}\Sigma\tilde{\mathbf{C}}\xi_j}{(\xi_j^T \Sigma \xi_j)(\xi_j^T \tilde{\mathbf{C}}\xi_j)^2} \quad (2.64)$$

Dans le but d'éviter les problèmes d'échelles des variances non reconstruites, il est important de pondérer chaque σ_j^2 par $\xi_j^T \Sigma \xi_j$ qui représente la variance originelle de la j ème variable. Evidemment, si les variables étudiées sont réduites alors $\xi_j^T \Sigma \xi_j = 1$.

Le choix d'un modèle ACP en se basant sur ce critère s'est justifié par une sélection du nombre optimal (ℓ_{op}) des CPs offrant la meilleure reconstruction. Autrement dit, ℓ_{op} doit assurer une variance non reconstruite minimale :

$$\ell_{op} = \arg \min_{\ell} \{ \mathbf{VNR}(\ell) \} \quad (2.65)$$

2.4.3.2 Consistance théorique du critère VNR

D'après [Dunia et Qin \(1998b,c,a\)](#) et [Qin et Dunia \(2000\)](#), ce critère doit nécessairement admettre un minimum qui correspond à un modèle optimal offrant une meilleure reconstruction. Afin de montrer l'existence d'un tel minimum, [Dunia et Qin \(1998a\)](#) ont décomposé la variance non reconstruite de la j ème variable en deux variances en se basant sur la propriété exprimée dans l'équation (2.24) et qui donne :

$$\|\xi_j\|^2 = \|\tilde{\mathbf{C}}\xi_j\|^2 + \|\hat{\mathbf{C}}\xi_j\|^2 = 1 \quad (2.66)$$

Ainsi, la variance de l'erreur de reconstruction individuelle d'une variable donnée peut s'écrire comme suit :

$$\begin{aligned} \sigma_j^2(\ell) &= \frac{\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j}{(\xi_j^T \tilde{\mathbf{C}} \xi_j)^2} = \frac{\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j}{\|\tilde{\mathbf{C}}\xi_j\|^2 (1 - \|\hat{\mathbf{C}}\xi_j\|^2)} \\ &= \frac{\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j}{\|\tilde{\mathbf{C}}\xi_j\|^2} \left(1 + \frac{\|\hat{\mathbf{C}}\xi_j\|^2}{1 - \|\hat{\mathbf{C}}\xi_j\|^2} \right) \end{aligned} \quad (2.67)$$

En posant

$$\tilde{\sigma}_j^2(\ell) = \frac{\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j}{\|\tilde{\mathbf{C}}\xi_j\|^2} \quad (2.68)$$

et

$$\hat{\sigma}_j^2(\ell) = \tilde{\sigma}_j^2(\ell) \left(\frac{\|\hat{\mathbf{C}}\xi_j\|^2}{1 - \|\hat{\mathbf{C}}\xi_j\|^2} \right) \quad (2.69)$$

la variance non reconstruite exprimée dans l'équation (2.67) n'est autre que :

$$\sigma_j^2(\ell) = \tilde{\sigma}_j^2(\ell) + \hat{\sigma}_j^2(\ell) \quad (2.70)$$

[Dunia et Qin \(1998a\)](#) ont montré que la fonction $\tilde{\sigma}_j^2$ est monotone décroissante en ℓ . Par ailleurs, la fonction $\hat{\sigma}_j^2$ tend vers l'infini pour $\ell = m$. En conséquence, la variance non reconstruite σ_j^2 doit obligatoirement avoir un minimum qui correspond à un nombre de CPs $\ell \in [1, m]$. Ces déductions peuvent être également étendues au critère **VNR** puisqu'il représente la somme des variances non reconstruites de toutes les variables. Néanmoins, on s'interroge si le nombre des CPs qui correspond à un tel minimum représente le nombre théorique q défini auparavant. Dans ce cadre, [Valle et al. \(1999\)](#) ont étendu leurs travaux afin d'établir des théorèmes sur la consistance d'un tel critère en montrant qu'il est capable de définir correctement le nombre optimal des CPs sous des conditions qui dépendent de la distribution du bruit de mesures.

Cas de bruit *i.i.d.* D'après (2.44), la matrice des valeurs propres, en présence d'un bruit de mesures *i.i.d.* de variance σ^2 , peut être introduite de la façon suivante :

$$\Lambda = \hat{\Lambda} + \sigma^2 \mathbf{I}_m = \begin{bmatrix} \hat{\Lambda} + \sigma^2 \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{(m-q)} \end{bmatrix} \quad (2.71)$$

où $\hat{\Lambda}$ et $\hat{\Lambda}$ sont données respectivement par les équations (2.31) et (2.32).

Pour $\ell \geq q$, on a $\xi_j^T \tilde{\mathbf{C}} \Sigma \tilde{\mathbf{C}} \xi_j = \sigma^2 \xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j$ ce qui implique que la variance non reconstruite de la *j*ème variable peut s'exprimer comme suit :

$$\sigma_j^2(\ell) = \frac{\sigma^2}{\xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j} \quad \forall \ell \geq q \quad (2.72)$$

Dans le cas particulier où $\ell = q$, on a :

$$\sigma_j^2(q) = \frac{\sigma^2}{\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j} \quad (2.73)$$

avec la matrice des vecteurs propres $\tilde{\mathbf{P}}_q$ est donnée par :

$$\tilde{\mathbf{P}}_q = [\mathbf{p}_{q+1}, \dots, \mathbf{p}_\ell, \mathbf{p}_{\ell+1}, \dots, \mathbf{p}_m] = [\mathbf{p}_{q+1}, \dots, \mathbf{p}_\ell, \tilde{\mathbf{P}}] \quad (2.74)$$

Celle-ci implique que :

$$\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j \geq \xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j \quad (2.75)$$

ainsi,

$$\sigma_j^2(\ell) \geq \sigma_j^2(q) \quad \forall \ell \geq q \quad (2.76)$$

Indépendamment de la nature de la matrice Σ si elle représente une matrice de covariance ou une matrice de corrélation, on a :

$$\frac{\sigma_j^2(\ell)}{\xi_j^T \Sigma \xi_j} \geq \frac{\sigma_j^2(q)}{\xi_j^T \Sigma \xi_j} \quad \forall \ell \geq q \quad (2.77)$$

L'inégalité précédente nous permet de déduire que :

$$\mathbf{VNR}(\ell) \geq \mathbf{VNR}(q) \quad \forall \ell \geq q \quad (2.78)$$

d'où

$$\arg \min_{\ell} \{\mathbf{VNR}(\ell)\} = q \quad \forall \ell \geq q \quad (2.79)$$

Cette égalité prouve que le critère \mathbf{VNR} ne surestime plus la dimension d'un modèle ACP en présence de bruit *i.i.d.*

Par ailleurs et dans le cas où $\ell < q$, la matrice des vecteurs propres $\tilde{\mathbf{P}}_q$ sera englobée dans $\tilde{\mathbf{P}}$ comme suit :

$$\tilde{\mathbf{P}} = [\mathbf{p}_{\ell+1}, \dots, \mathbf{p}_q, \tilde{\mathbf{P}}_q] \quad \text{pour } \ell < q \quad (2.80)$$

Dans ce cas, [Valle et al. \(1999\)](#) ont montré que :

$$\sigma_j^2(\ell) \geq \sigma_j^2(q) \quad \text{si} \quad \hat{\lambda}_q \geq \frac{\xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j}{\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j} \sigma^2 \quad \forall \ell < q \quad (2.81)$$

Cette hypothèse mettra en évidence que la variance de la dernière CP théoriquement significative doit être supérieure à la variance du bruit de mesures. Par conséquent, le critère **VNR** définit correctement le nombre optimal des CPs sous une telle condition :

$$\arg \min_{\ell} \{\mathbf{VNR}(\ell)\} = q \quad \text{si} \quad \hat{\lambda}_q \geq \frac{\xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j}{\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j} \sigma^2 \quad \forall \begin{cases} \ell < q \\ 1 \leq j \leq m \end{cases} \quad (2.82)$$

Cas de bruit coloré Il semblerait que l'hypothèse qui consiste à considérer que le bruit est blanc ne soit pas toujours adaptée. Il se peut qu'on préfère modéliser le bruit différemment en le colorant i.e. soit en relâchant la contrainte que les variances sont identiques soit que le bruit est indépendant, soit les deux. Dans la pratique, les variances du bruit ne sont pas nécessairement identiques. Dans ce cas, la matrice des valeurs propres prend la forme suivante :

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_q, \sigma_{q+1}^2, \dots, \sigma_m^2\} \quad (2.83)$$

où les σ_i^2 sont les variances du bruit. En suivant le même principe que celui du cas de bruit *i.i.d.*, [Valle et al. \(1999\)](#) ont montré que le critère **VNR** atteint son minimum à $\ell = q$ CPs si :

$$\frac{\sigma_{q+1}^2}{\sigma_m^2} \leq \frac{\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j}{\xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j} \quad \text{pour} \quad \ell \geq q \quad (2.84)$$

et

$$\lambda_q \geq \left(1 + \frac{\xi_j^T \tilde{\mathbf{P}} \tilde{\mathbf{P}}^T \xi_j}{\xi_j^T \tilde{\mathbf{P}}_q \tilde{\mathbf{P}}_q^T \xi_j} \right) \sigma_{q+1}^2 \quad \text{pour} \quad \ell < q \quad (2.85)$$

L'interprétation de la première inégalité implique que l'étendue des variances du bruit doit être faible ce qui implique également que ces variances doivent avoir des valeurs très proches. Cependant, la deuxième condition indique que la q ème CP qui est la dernière supposée être théoriquement retenue doit avoir une variance au moins deux fois plus supérieure que celle de la $(q+1)$ ème CP.

2.5 Etude comparative des différents critères

Les différents critères présentés dans ce chapitre vont être comparés vis-à-vis de leurs efficacités et performances dans le choix de la dimension optimale du modèle ACP d'un ensemble de données simulées. Pour évaluer ces critères plusieurs enjeux sont considérés tels que le nombre de variables, le nombre d'observations et la variance du bruit de mesures.

$\hat{x}_1(k)$	$= 3 + \exp(\cos(\frac{k}{15\pi}))$
$\hat{x}_2(k)$	$= \cos(\frac{1}{2\pi k}) \sin(\frac{k}{2\pi})$
$\hat{x}_3(k)$	$= \arctan(50\pi k) \log(1 + \hat{x}_2(k)^2)$
$\hat{x}_4(k)$	$= \hat{x}_1(k) + 3\hat{x}_2(k)$
$\hat{x}_5(k)$	$= \hat{x}_1(k) - \hat{x}_2(k)$
$\hat{x}_6(k)$	$= \hat{x}_1(k) + \hat{x}_3(k)$
$\hat{x}_7(k)$	$= \hat{x}_2(k) + 3\hat{x}_3(k)$
$\hat{x}_8(k)$	$= \frac{10}{\pi} \arctan(\tan(\frac{\pi}{200}(k - 100)))$
$\hat{x}_9(k)$	$= 10 \cos(7\pi k)$
$\hat{x}_{10}(k)$	$= 2\hat{x}_8(k) + \hat{x}_9(k)$
$\hat{x}_{11}(k)$	$= -2\hat{x}_8(k) + 3\hat{x}_9(k)$
$\hat{x}_{12}(k)$	$= \hat{x}_2(k) + \hat{x}_8(k)$
$\hat{x}_{13}(k)$	$= (2 + \cos(0.2\pi k))^{-1}$
$\hat{x}_{14}(k)$	$= \frac{40}{\pi} \cos(7\pi k) \arctan(\tan(\frac{9\pi}{200}(k - 900)))$
$\hat{x}_{15}(k)$	$= \text{sgn}(\sin(0.007\pi k))$

TABLE 2.2 – Expressions analytiques des variables simulées

2.5.1 Présentation et interprétation de l'exemple simulé

L'exemple de simulation comporte différents ensembles de variables. Nous considérons un jeu de données qui fait apparaître des relations de redondance analytiques linéaires et non linéaires ainsi que des variables indépendantes. Les relations non linéaires entre quelques variables ne peuvent pas être exprimées dans un contexte d'une ACP linéaire. Par conséquent, ces variables sont linéairement indépendantes.

Notons par $\hat{\mathbf{x}}(k) = [\hat{x}_1(k), \dots, \hat{x}_{15}(k)]^T \in \mathbb{R}^{15}$ un vecteur de données composé des mesures non bruitées prélevées au k ème instant d'un ensemble de 15 variables. A titre d'exemple, $\hat{x}_j(k)$ représente la mesure non bruitée de la j ème variable au k ème instant. En absence de bruit, les mesures des variables sont générées par les expressions analytiques données dans le tableau 2.2.

L'interprétation théorique des équations du tableau 2.2 nous permet de déduire que les variables \hat{x}_1 , \hat{x}_2 , \hat{x}_3 , \hat{x}_8 et \hat{x}_9 ne sont pas corrélées. Par contre, elles sont corrélées avec d'autres variables. Cependant, les variables \hat{x}_{13} , \hat{x}_{14} et \hat{x}_{15} sont linéairement indépendantes et/ou quasi-indépendantes du reste des variables et entre elles. Afin de mieux interpréter ces relations de dépendance, nous avons proposé d'investiguer les valeurs propres des matrices de corrélation de quatre principales bases de données sans bruit de mesures. En considérant des mesures normalisées, les matrices de corrélation de ces quatre groupes de variables à étudier sont notées de la manière suivante :

$$\left(\overset{\circ}{\Sigma}\right)_{\mathcal{A}} = \mathbb{E} \{ \hat{\mathbf{x}} \hat{\mathbf{x}}^T \} \quad \text{où, à l'instant } k \quad \hat{\mathbf{x}}(k) = [\hat{x}_1(k), \dots, \hat{x}_{12}(k)]^T \in \mathbb{R}^{12} \quad (2.86)$$

$$\left(\overset{\circ}{\Sigma}\right)_{\mathcal{B}} = \mathbb{E} \{ \hat{\mathbf{x}} \hat{\mathbf{x}}^T \} \quad \text{où, à l'instant } k \quad \hat{\mathbf{x}}(k) = [\hat{x}_1(k), \dots, \hat{x}_{13}(k)]^T \in \mathbb{R}^{13} \quad (2.87)$$

$$\left(\overset{\circ}{\Sigma}\right)_{\mathcal{C}} = \mathbb{E} \{ \hat{\mathbf{x}} \hat{\mathbf{x}}^T \} \quad \text{où, à l'instant } k \quad \hat{\mathbf{x}}(k) = [\hat{x}_1(k), \dots, \hat{x}_{14}(k)]^T \in \mathbb{R}^{14} \quad (2.88)$$

$$\left(\overset{\circ}{\Sigma}\right)_{\mathcal{D}} = \mathbb{E} \left\{ \overset{\circ}{\mathbf{x}} \overset{\circ}{\mathbf{x}}^T \right\} \quad \text{où, à l'instant } k \quad \overset{\circ}{\mathbf{x}}(k) = [\overset{\circ}{x}_1(k), \dots, \overset{\circ}{x}_{15}(k)]^T \in \mathbb{R}^{15} \quad (2.89)$$

Le premier ensemble de données noté \mathcal{A} est composé des 12 premières variables du tableau 2.2. Théoriquement, ce groupe présente 5 variables linéairement non corrélées. Cela indique l'existence de 5 valeurs propres non nulles. Cette déduction est bien confirmée par les résultats donnés dans la deuxième colonne du tableau 2.3. L'addition de la treizième variable à celles de l'ensemble \mathcal{A} , permettra l'obtention d'un deuxième groupe de variables \mathcal{B} . La variable ajoutée est linéairement indépendante de toutes les variables. Théoriquement, elle représente toute seule une CP en gardant sa variance originelle qui vaut l'unité puisque les mesures sont normalisées. La troisième colonne du tableau 2.3 montre que la valeur propre $\overset{\circ}{\lambda}_6$ est celle qui est associée à la variable indépendante $\overset{\circ}{x}_{13}$. Ainsi, l'indépendance d'une telle variable se manifeste par l'insensibilité des autres valeurs propres à sa présence car elles sont inchangées. Cette variable indépendante constitue une CP ayant une variance moins faible qu'aux variances des autres CPs. Le groupe \mathcal{B} dispose alors de 6 valeurs propres non nulles.

Le troisième ensemble \mathcal{C} est composé des quatorze premières variables. Cet ensemble représente celui de \mathcal{B} auquel on a ajouté la variable $\overset{\circ}{x}_{14}$. En absence du bruit de mesures, l'ajout d'une telle variable à l'ensemble \mathcal{B} a légèrement modifié trois valeurs propres parmi les anciennes tout en produisant une nouvelle valeur propre $\overset{\circ}{\lambda}_7$ proche de l'unité. L'interprétation de la quatrième colonne du tableau 2.3 nous mène à conclure que la variable $\overset{\circ}{x}_{14}$ est très légèrement corrélée avec $\overset{\circ}{x}_{13}$ ainsi qu'avec d'autres variables difficilement identifiables. En effet, les seules valeurs propres qui ont été perturbées par la présence de $\overset{\circ}{x}_{14}$ sont $\overset{\circ}{\lambda}_6$ qui correspond à la variable $\overset{\circ}{x}_{13}$ ainsi que $\overset{\circ}{\lambda}_2$ et $\overset{\circ}{\lambda}_3$. Par ailleurs, cette perturbation peut être considérée très négligeable puisque $\overset{\circ}{\lambda}_6$ et $\overset{\circ}{\lambda}_7$ sont toutes les deux très proches de l'unité. En conséquence, $\overset{\circ}{x}_{13}$ et $\overset{\circ}{x}_{14}$ sont linéairement quasi-indépendantes voire indépendantes entre elles ainsi qu'aux autres variables. Notamment, l'ensemble \mathcal{C} présente 7 valeurs propres non nulles.

Le quatrième groupe de données, noté \mathcal{D} , est constitué de toutes les variables du tableau 2.2. D'après la cinquième colonne de ce tableau, on peut déduire que la variable ajoutée $\overset{\circ}{x}_{15}$ est indépendante puisque sa présence n'a aucun effet sur le reste des variables. En effet, elle a produit une valeur propre égale à l'unité, en même temps les autres anciennes valeurs propres n'ont subi aucun changement. Ce dernier groupe de variables dispose donc de 8 CPs associées aux valeurs propres non nulles.

Dans la pratique, les mesures sont entachées de bruit. Pour cette raison, des bruits *i.i.d.* de moyennes nulles et de variances identiques générées selon trois différents cas ont été superposés aux variables du tableau 2.2.

Selon le principe d'utilisation des critères étudiés dans ce chapitre, nous avons distingué deux principales familles. La première représente des critères qui se réfèrent à des seuils pour définir le nombre optimal des CPs. En revanche, la deuxième catégorie est constituée des critères dont la minimisation par rapport au nombre des CPs peut indiquer la dimension adéquate du modèle ACP.

Le premier cas correspond à la superposition d'un bruit *i.i.d.* de faible variance ($\sigma^2 = 0.002$) largement inférieure aux variances des variables originelles. Naturellement, les relations linéaires entre les variables ne sont pas trop perturbées. Ainsi, les nombres opti-

	Ensemble \mathcal{A}	Ensemble \mathcal{B}	Ensemble \mathcal{C}	Ensemble \mathcal{D}
	$(\hat{\Sigma})_{\mathcal{A}}$	$(\hat{\Sigma})_{\mathcal{B}}$	$(\hat{\Sigma})_{\mathcal{C}}$	$(\hat{\Sigma})_{\mathcal{D}}$
$\hat{\lambda}_1$	2.92	2.92	2.92	2.92
$\hat{\lambda}_2$	2.77	2.77	2.78	2.78
$\hat{\lambda}_3$	2.69	2.69	2.70	2.70
$\hat{\lambda}_4$	2.10	2.10	2.10	2.10
$\hat{\lambda}_5$	1.51	1.51	1.51	1.51
$\hat{\lambda}_6$	0	1	1.08	1.08
$\hat{\lambda}_7$	0	0	0.90	1
$\hat{\lambda}_8$	0	0	0	0.90
$\hat{\lambda}_9$	0	0	0	0
$\hat{\lambda}_{10}$	0	0	0	0
$\hat{\lambda}_{11}$	0	0	0	0
$\hat{\lambda}_{12}$	0	0	0	0
$\hat{\lambda}_{13}$	✓	0	0	0
$\hat{\lambda}_{14}$	✓	✓	0	0
$\hat{\lambda}_{15}$	✓	✓	✓	0

TABLE 2.3 – Valeurs propres des matrices de corrélation des quatre ensembles de variables en absence du bruit de mesures et pour $N = 1500$ observations

maux des CPs qui constituent les modèles ACP dans chaque ensemble étudié de variables, doivent correspondre aux nombres des valeurs propres non nulles du tableau 2.3.

Les courbes des critères étudiés pour la sélection du nombre des CPs dans les ensembles de variables \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} dont chacun est constitué de 1500 observations, sont illustrées respectivement par les figures 2.1, 2.2, 2.3 et 2.4. On note que les critères **AIC**, **MDL**, **IE** et **VNR** sont exprimées en coordonnées semi-logarithmique afin que leurs courbes soient mieux lisibles. Notamment, une majorité des critères étudiés est basée sur des données initialement normalisées donc en utilisant les valeurs propres des matrices de corrélation. Néanmoins, les critères **AIC**, **MDL** et **IE** utilisent des données non réduites puisqu'ils ont été définis valables uniquement avec les valeurs propres des matrices de covariance.

2.5.2 Interprétations des critères basés sur des seuils

Pour le premier cas de cette étude qui correspond à des variables entachées d'un bruit de faible variance, les nombres optimaux théoriques des CPs dans les ensembles \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} sont respectivement 5, 6, 7 et 8 CPs. On observe clairement dans toutes les figures (2.1, 2.2, 2.3 et 2.4) que les critères **AC**, **KG**, **PCV** et **PVR** ne peuvent en aucun cas assurer d'une façon précise et objective les modèles optimaux.

En effet, le critère **AC** montre que la deuxième CP possède la plus faible valeur d'autocorrélation. En outre, une telle valeur est très inférieure au seuil indiqué. Néanmoins, quelques auteurs suggèrent parfois un ajustement d'un tel seuil afin qu'il soit adaptable au critère. Il est bien évident qu'aucun ajustement ne peut apporter d'avantage puisque la CP en question présente la plus faible valeur d'autocorrélation. L'inconvénient de ce critère se résume dans le fait qu'une CP ayant une variance importante peut correspondre malheureusement à une faible valeur d'autocorrélation.

Le critère **KG** a correctement indiqué le nombre optimal des CPs uniquement dans

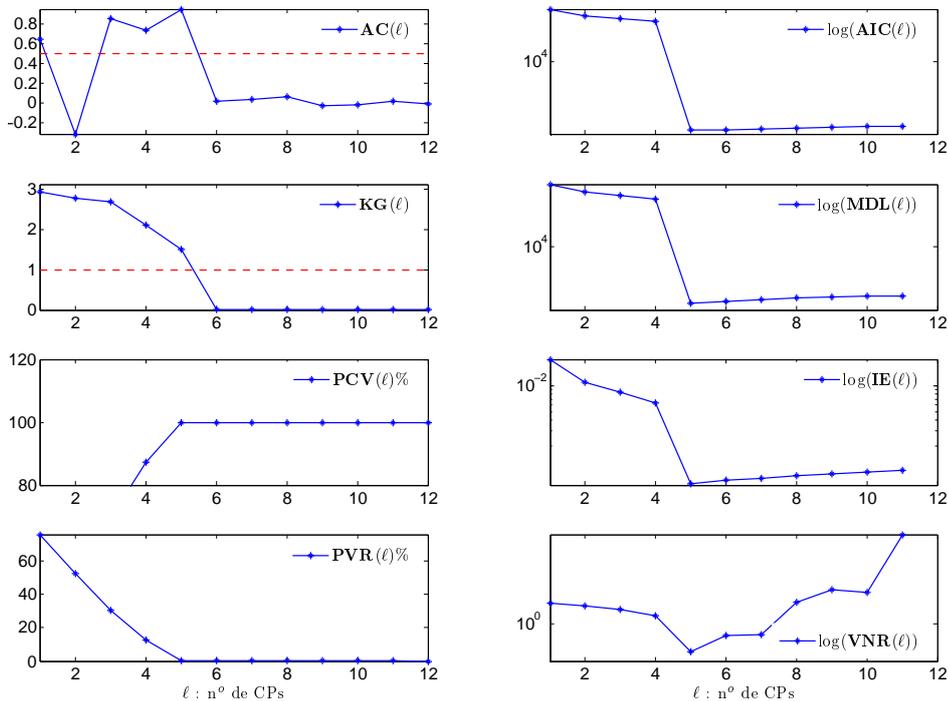


FIGURE 2.1 – Allures des critères de sélection pour l'ensemble \mathcal{A} dont les variables sont entachées par un bruit *i.i.d.* de variance $\sigma^2 = 0.002$

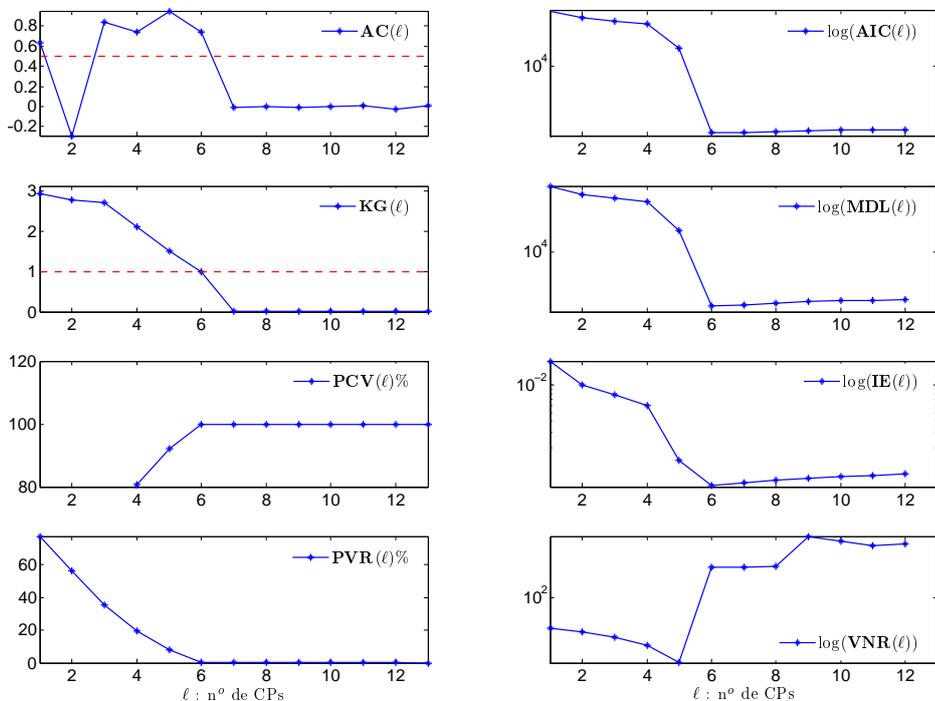


FIGURE 2.2 – Allures des critères de sélection pour l'ensemble \mathcal{B} dont les variables sont entachées par un bruit *i.i.d.* de variance $\sigma^2 = 0.002$

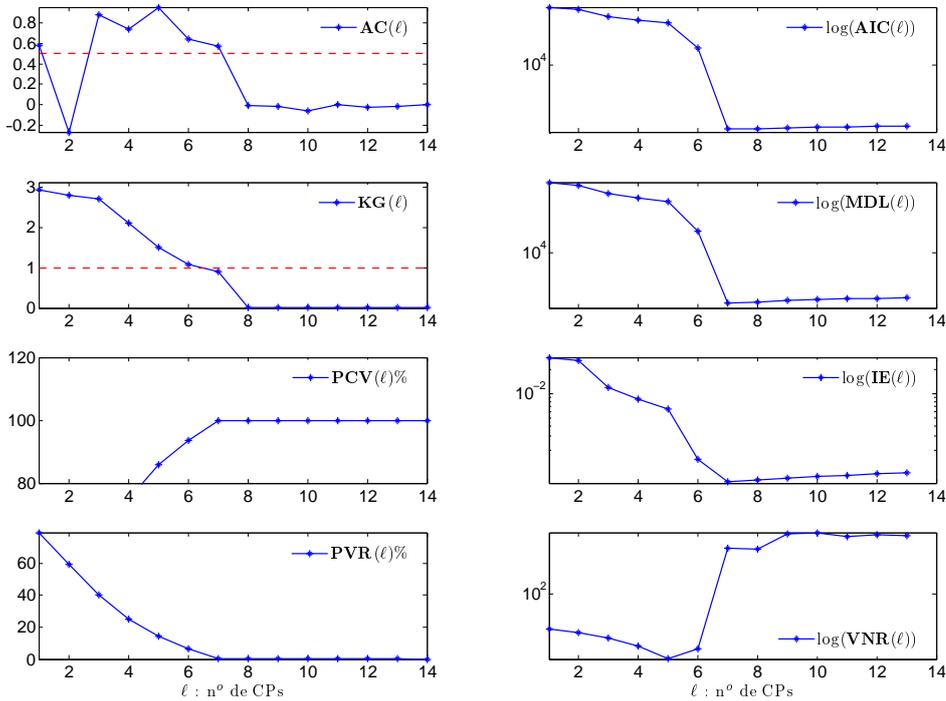


FIGURE 2.3 – Allures des critères de sélection pour l'ensemble \mathcal{C} dont les variables sont entachées par un bruit *i.i.d.* de variance $\sigma^2 = 0.002$

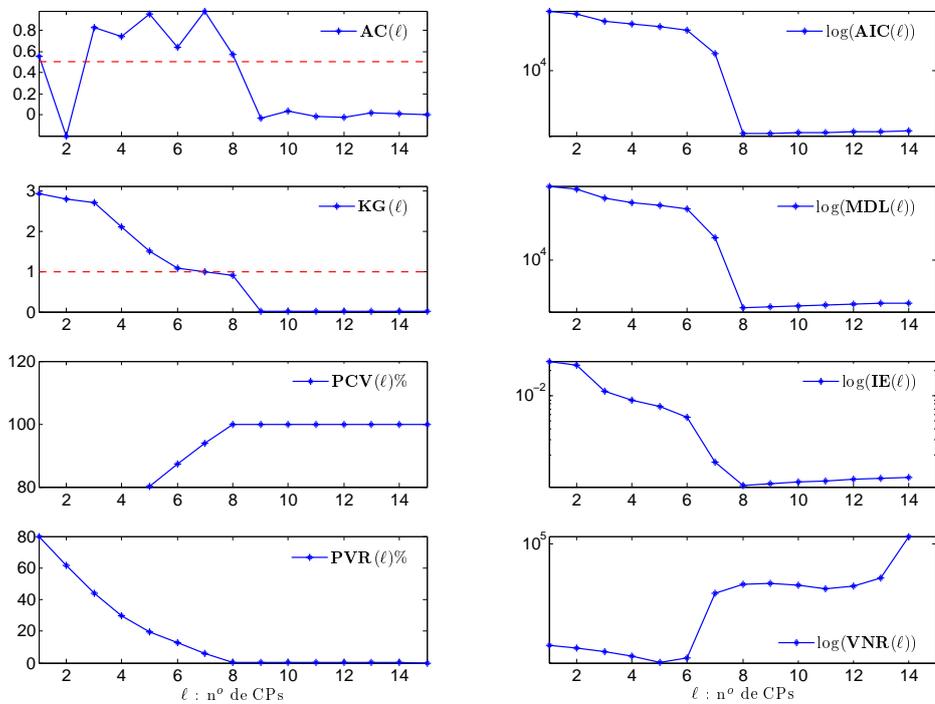


FIGURE 2.4 – Allures des critères de sélection pour l'ensemble \mathcal{D} dont les variables sont entachées par un bruit *i.i.d.* de variance $\sigma^2 = 0.002$

l'ensemble \mathcal{A} . Cependant, sa performance de sélection a été dégradée dans le reste des ensembles en manquant toujours une ou deux CPs par rapport aux nombres théoriques prévus. Cette limitation s'explique par l'existence des variables quasi-indépendantes \hat{x}_{13} et \hat{x}_{14} dans les ensembles \mathcal{C} et \mathcal{D} . Ces deux variables ont produit deux valeurs propres qui sont légèrement supérieure et inférieure à l'unité. En ajoutant un bruit, l'une des deux valeurs propres n'a pas franchi le seuil de l'unité. En effet, le critère **KG** ignore souvent les CPs qui correspondent aux variables indépendantes et plus particulièrement quasi-indépendantes dans le cas de données normalisées. En considérant des matrices de covariance, une variable indépendante ne peut être retenue par le critère **KG** que si sa variance est supérieure à la moyenne arithmétique de toutes les valeurs propres. Cela explique qu'une telle variable peut être retenue dans des cas et écartée dans d'autres.

En essayant d'expliquer une variabilité généralement supérieure à 80% de la variance totale, le critère **PCV** montre une diversité de choix qui se complique proportionnellement au nombre des variables considérées. En effet, il peut exister dans l'intervalle [80%, 99%] plus qu'une CP qui peut satisfaire la contrainte. En plus, cette situation est très confrontée lorsque les valeurs propres significatives sont trop proches les unes des autres. Par conséquent, la décision devient difficile voire même très subjective.

Le principe de l'utilisation du critère **PVR** se résume dans la recherche du coude ou du point d'inflexion. D'après les figures 2.1, 2.2, 2.3 et 2.4, on observe très nettement les points d'inflexion qui correspondent correctement aux nombres théoriques optimaux des CPs dans chaque ensemble de variables. Cependant, et selon le principe conventionné de l'utilisation de ce critère, on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. De cette manière, on remarque l'existence en permanence d'une CP manquante dans les modèles sélectionnés. Par conséquent, la meilleure solution semble d'en compter également les points d'inflexion. En revanche, cette situation qui montre clairement les points d'inflexion est très particulière pour deux raisons principales. Premièrement, la faiblesse de la variance du bruit considéré a rendu les valeurs propres significatives assez distinctes des autres. Deuxièmement, le nombre des variables étudiées n'est pas assez important. Si ces deux arguments ne sont pas vérifiés, la courbe du critère **PVR** sera caractérisée par une lente décroissance. Par conséquent, on peut se confronter à des situations où le point d'inflexion est quasiment inexistant.

2.5.3 Interprétations des critères minimisés

La deuxième catégorie est constituée des critères **AIC**, **MDL**, **IE** et **VNR**. Leurs courbes pour le cas des données entachées par un bruit *i.i.d.* de variance identique assez faible sont illustrées dans les deuxièmes colonnes des figures 2.1, 2.2, 2.3 et 2.4. L'avantage dans l'utilisation de ces critères se manifeste dans l'objectivité des choix puisqu'ils se basent sur leurs minima afin d'en définir un nombre optimal des CPs. En revanche, il reste à évaluer la consistance de leurs performances.

Afin d'approfondir l'étude comparative entre ces quatre critères, nous avons réalisé trois jeux de données pour chacun des ensembles \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} . Chaque jeu représente une matrice de données composé de N observations et correspond à une variance particulière du bruit *i.i.d.* Ainsi, les valeurs de ces variances sont respectivement $\sigma^2 = 0.002$, $\sigma^2 = 0.2$ puis $\sigma^2 = 0.5$. Puisque le bruit représente des variables aléatoires, une seule réalisation

peut ne pas exprimer continuellement le même minimum pour chaque critère. Pour cette raison, chaque jeu de données a été simulé 1500 fois afin d'exprimer en pourcentage le nombre des CPs retenues par chaque critère.

Le tableau 2.4 illustre les pourcentages des CPs retenues par les critères considérés pour des jeux de données dont chacun est composé de $N = 1500$ observations. En rejoignant les remarques données par Wax et Kailath (1985), ce tableau prouve que le critère **AIC** surestime souvent le nombre des CPs. Pour un bruit *i.i.d.* de faible variance ($\sigma^2 = 0.002$), le minimum d'un tel critère correspond dans 76.80%, 74.33%, 78.53% et 75.26% des réalisations à 5, 6, 7 et 8 CPs respectivement dans les ensembles \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} . Malgré la légère variance du bruit considéré, ce critère a retenu dans approximativement 25% des cas des nombres de CPs supérieures à ceux nécessaires. Par conséquent, son inconvénient se manifeste dans la surestimation de la structure du modèle. L'augmentation de la variance du bruit modifie les relations entre les variables en causant certainement l'apparition de nouvelles variables indépendantes. A ce stade, il est difficile de connaître les nombres optimaux de CPs dans les jeux de données qui correspondent aux deux autres variances du bruit. Cependant, et d'après les pourcentages exprimés dans le tableau 2.4, le comportement du critère **AIC** est toujours fluctuant.

D'après le même tableau, nous remarquons que le critère **VNR** présente avec 100% des réalisations un minimum pour 5 CPs dans les ensembles \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} qui correspondent aux deux premiers cas des variances du bruit qui sont respectivement $\sigma^2 = 0.002$ et $\sigma^2 = 0.2$. En revanche, ce critère indique avec plus de 97% un nombre de 4 CPs dans tous les ensembles des variables en considérant un bruit de variance plus forte $\sigma^2 = 0.5$.

Puisque le premier cas présente des données entachées par un bruit de faible variance, il peut être considéré comme une référence pour l'évaluation des critères étudiés car on sait *a priori* les nombres optimaux des CPs dans chaque ensemble. Dans ce cas, le critère **VNR** a défini correctement le nombre optimal des CPs uniquement dans l'ensemble \mathcal{A} . En investiguant la transition réalisée entre les ensembles étudiés, nous remarquons l'existence d'une variable indépendante ou quasi-indépendante qui s'ajoute à chaque transition d'un ensemble à l'autre dans le sens de \mathcal{A} vers \mathcal{D} . Notamment, les variables indépendantes conservent leurs indépendances quelle que soit la valeur de la variance du bruit *i.i.d.* En considérant cette propriété ainsi que les résultats exprimés par le tableau 2.4, on peut déduire que le critère **VNR** ne prend pas en compte ce type des variables. Cette déduction a été illustrée à travers un exemple simulé par Mnassri et al. (2010a). Néanmoins, elle n'a pas été prouvée ou montrée théoriquement dans la littérature. Dans le cadre de cette thèse, nous proposons dans l'annexe A une démonstration théorique prouvant la limitation du critère **VNR** dans la sélection des CPs en présence des variables indépendantes et quasi-indépendantes.

Une telle démonstration justifie les pourcentages donnés par ce critère dans le tableau 2.4 lors du premier cas de la variance du bruit. En augmentant cette variance ($\sigma^2 = 0.2$), nous remarquons que le critère **VNR** maintient les mêmes résultats. Toutefois, l'augmentation d'une telle variance peut causer l'apparition de variables indépendantes. Originellement, l'ensemble \mathcal{A} ne contient pas ce type de variables. Si les résultats de ce critère sont inchangés en faisant varier la variance du bruit ($\sigma^2 = 0.2$), cela ne peut pas nier la possibilité de leur apparition dans \mathcal{A} puisqu'un tel critère est prouvé insensible à leur présence.

Critère :	Ensemble \mathcal{A}				Ensemble \mathcal{B}				Ensemble \mathcal{C}				Ensemble \mathcal{D}			
	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.002$																
5 CPs (%)	76.80	100	100	100				100				100				100
6 CPs (%)	18.40				74.33	100	100									
7 CPs (%)	03.93				20.26				78.53	100	100					
8 CPs (%)	00.60				04.53				16.46				75.26	100	100	
9 CPs (%)	00.13				00.53				03.46				19.93			
10 CPs (%)	00.13				00.26				01.00				03.66			
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.2$																
5 CPs (%)	77.73	100	100	100	00.33	79.60	100	100				100				100
6 CPs (%)	17.80				76.73	20.40			00.33	79.06	100					
7 CPs (%)	03.06				17.80				76.00	20.93			00.20	79.53	100	
8 CPs (%)	00.80				03.60				19.53				78.06	20.46		
9 CPs (%)	00.46				01.40				03.20				17.40			
10 CPs (%)	00.13				00.13				00.73				03.40			
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.5$																
4 CPs (%)				97.20				97.66				98.20				98.20
5 CPs (%)	75.73	100	100	02.80	47.00	100	100	02.33				01.80				01.80
6 CPs (%)	19.40				42.13				45.80	100	100					
7 CPs (%)	03.53				08.53				41.86				44.93	100	100	
8 CPs (%)	00.80				01.86				10.00				42.46			
9 CPs (%)	00.33				00.46				01.80				10.40			
10 CPs (%)									00.53				01.86			

TABLE 2.4 – Pourcentage, par rapport à 1500 réalisations, des nombres de CPs retenues dans les ensembles des données par les critères minimisés ($N = 1500$ observations générées selon trois différents cas de bruit *i.i.d.*)

	Bruit <i>i.i.d.</i> : $\sigma^2 = 0.2$	Bruit <i>i.i.d.</i> : $\sigma^2 = 0.5$	
	$\ell = 5$ CPs	$\ell = 4$ CPs	$\ell = 5$ CPs
	$\xi_j^T \tilde{\mathbf{C}} \xi_j$	$\xi_j^T \tilde{\mathbf{C}} \xi_j$	$\xi_j^T \tilde{\mathbf{C}} \xi_j$
$j = 1$	0.593	0.653	0.604
$j = 2$	0.611	0.706	0.628
$j = 3$	0.398	0.965	0.362
$j = 4$	0.575	0.674	0.584
$j = 5$	0.629	0.622	0.622
$j = 6$	0.624	0.622	0.616
$j = 7$	0.579	0.760	0.593
$j = 8$	0.544	0.548	0.544
$j = 9$	0.628	0.629	0.628
$j = 10$	0.640	0.642	0.642
$j = 11$	0.611	0.611	0.610
$j = 12$	0.563	0.562	0.562

TABLE 2.5 – Coefficients diagonaux de la matrice $\tilde{\mathbf{C}}$ de l'ensemble \mathcal{A} et pour deux différentes variances de bruit *i.i.d.*

D'après l'expression du critère **VNR** ainsi que la démonstration de l'annexe A, les coefficients diagonaux ($\xi_j^T \tilde{\mathbf{C}} \xi_j$ avec $j = \{1, \dots, m\}$) de la matrice $\tilde{\mathbf{C}}$ peuvent nous aider dans l'identification des variables indépendantes et quasi-indépendantes. Par ailleurs, nous notons que cette technique ne garantit pas l'objectif désiré. Notamment, ces variables se transforment en des CPs de variances hiérarchiquement moins importantes que celles des premières CPs qui représentent les principales sources de corrélation. Ainsi, elles contribuent très faiblement dans les vecteurs propres associés aux premières CPs. Cela implique que les coefficients diagonaux qui leur correspondent dans $\tilde{\mathbf{C}}$ sont très proches de l'unité. En investiguant le tableau 2.5, la colonne qui correspond au cas d'un bruit *i.i.d.* de variance $\sigma^2 = 0.2$ n'exprime pas de particularité intéressante dans les coefficients diagonaux de $\tilde{\mathbf{C}}$. On peut ainsi déduire qu'avec une telle variance du bruit, l'ensemble \mathcal{A} ne présente pas de variables indépendantes.

Dans le troisième cas qui correspond à des données entachées par un bruit *i.i.d.* de variance plus élevée ($\sigma^2 = 0.5$), les pourcentages indiquent que le critère **VNR** a manqué une CP. Ce résultat implique certainement qu'une variable qui représente une source de corrélation s'est transformée avec l'effet du bruit en une variable indépendante ou quasi-indépendante. Par conséquent, l'étendue de cette transformation peut toucher d'autres variables en les rendant également indépendantes. Les variables touchées sont uniquement celles qui dépendent de la source disparue. Ainsi, la propriété du critère **VNR** garantit l'apparition d'au moins une variable indépendante dans l'ensemble \mathcal{A} avec un bruit *i.i.d.* de variance $\sigma^2 = 0.5$. D'après l'avant dernière colonne du tableau 2.5, les éléments diagonaux de la matrice $\tilde{\mathbf{C}}$ calculée à base des 4 CPs retenues par le critère **VNR** montre que la troisième variable dispose d'un coefficient égale à 0.965. Par conséquent, cette variable est devenue quasi-indépendante sous l'effet du bruit. Nous rappelons ainsi qu'elle a été une source de corrélation. Malgré sa transformation, les autres variables n'ont pas été influencées car leurs coefficients dans $\tilde{\mathbf{C}}$ calculée à base de 5 CPs ne présentent pas des valeurs proches de l'unité (dernière colonne du tableau 2.5).

Cette analyse nous permet de déduire les nombres optimaux théoriques des CPs sup-

posées être retenues dans les ensembles étudiés selon les différentes variances du bruit. Indépendamment des critères utilisés, les ensembles de variables \mathcal{A} , \mathcal{B} , \mathcal{C} et \mathcal{D} disposent respectivement de 5, 6, 7 et 8 CPs dans les trois différents cas du bruit *i.i.d.*

D'après le tableau 2.4, le critère **IE** retient avec une certitude de 100% un nombre approprié de CPs pour chaque ensemble de variables et dans les différents cas du bruit. En effet, il a correctement défini les nombres optimaux des CPs pour tous les ensembles de variables durant le premier cas qui correspond à un du bruit de faible variance. Exceptionnellement, ce critère a manqué une CP dans les ensembles \mathcal{B} , \mathcal{C} et \mathcal{D} pour les deux autres cas où la variance du bruit est plus élevée. Plus particulièrement, la CP manquée se déclare dans l'ensemble \mathcal{B} entaché d'un bruit de variance $\sigma^2 = 0.2$. En absence de ce dernier, nous rappelons que l'ensemble \mathcal{B} se distingue par rapport à \mathcal{A} par la variable indépendante \hat{x}_{13} (voir tableaux 2.2 et 2.3). Puisque ce critère a convenablement déterminé le nombre des CPs dans l'ensemble \mathcal{A} avec le même bruit, la CP manquée doit nécessairement correspondre à cette variable indépendante \hat{x}_{13} . En s'appuyant sur cette déduction, nous nous interrogeons alors sur la raison pour laquelle le critère **IE** a retenu les autres variables indépendantes dans le reste des ensembles.

Nous notons que Malinowski (1977) a montré qu'un tel critère est monotone croissant en ℓ en se limitant à l'intervalle $[q, m - 1]$ où q désigne le nombre théorique des CPs supposées être retenues. Cependant, cela n'implique pas nécessairement que le minimum de ce critère correspond à q CPs pour toutes valeurs de $\ell \in [1, m - 1]$. Pour cette raison, nous avons établi dans l'annexe B une condition nécessaire et suffisante garantissant le minimum de ce critère à q CPs. Théoriquement, le critère **IE** ne surestime pas la dimension du modèle. Par contre, il peut abandonner quelques CPs dont les valeurs propres en absence du bruit de mesures ne satisfont pas la condition établie (annexe B).

Avec des données non réduites, la variable indépendante \hat{x}_{13} abandonnée par ce critère dispose en l'absence du bruit d'une variance de 0.25. Ainsi, il a été supposé que $q = 6$ CPs dans l'ensemble \mathcal{B} entaché d'un bruit *i.i.d.* de variance $\sigma^2 = 0.2$. En considérant ces paramètres, nous pouvons prouver que la variance de la variable abandonnée ne satisfait pas la condition établie par l'inégalité (B.6) dans l'annexe B. De la même manière, nous pouvons vérifier également que cette variable n'est plus en mesure d'être retenue par le critère **IE** non seulement dans l'ensemble \mathcal{B} mais également dans \mathcal{C} et \mathcal{D} pour les deux cas des variances du bruit ($\sigma^2 = 0.2$ et $\sigma^2 = 0.5$).

Puisque le critère **IE** ne surestime pas le nombre des CPs, cela justifie notre raisonnement par l'apparition d'une variable indépendante qui n'a pas pu être retenue par le critère **VNR** dans l'ensemble \mathcal{A} contenant un bruit de variance $\sigma^2 = 0.5$.

Dans le premier cas qui correspond à un bruit *i.i.d.* de faible variance ($\sigma^2 = 0.002$), le critère **MDL** exprime dans 100% des réalisations les nombres corrects de CPs pour tous les ensembles de données (tableau 2.4). L'investigation du deuxième cas, caractérisé par un bruit de variance $\sigma^2 = 0.2$, montre que ce critère a défini convenablement le nombre théorique des CPs uniquement dans l'ensemble \mathcal{A} . Dans le reste des ensembles, il a exprimé avec 79% des nombres manquant une CP par rapport aux nombres théoriques prévus. Dans le troisième cas du bruit, le critère a totalement manqué une CP dans les ensembles \mathcal{B} , \mathcal{C} et \mathcal{D} . En comparant les résultats de ce critère avec ceux donnés par **IE**, nous remarquons que la CP non retenue par **MDL** est également liée à la variable indépendante \hat{x}_{13} .

Critère :	Ensemble \mathcal{A}				Ensemble \mathcal{B}				Ensemble \mathcal{C}				Ensemble \mathcal{D}			
	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR	AIC	MDL	IE	VNR
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.002$																
5 CPs (%)	76.66	100	100	100				100				100				100
6 CPs (%)	18.53				76.93	100	100									
7 CPs (%)	03.93				19.20				79.00	100	100					
8 CPs (%)	00.60				03.06				17.26				77.73	100	100	
9 CPs (%)	00.13				00.73				03.20				17.13			
10 CPs (%)	00.13								00.40				03.73			
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.2$																
5 CPs (%)	78.40	100	100	100			100	100				100				100
6 CPs (%)	18.06				77.46	100					100					
7 CPs (%)	02.93				18.33				79.00	100					100	
8 CPs (%)	00.40				03.06				16.93				74.00	100		
9 CPs (%)	00.13				00.86				03.06				21.06			
10 CPs (%)	00.06				00.13				00.80				03.86			
Bruit <i>i.i.d.</i> :																
$\sigma^2 = 0.5$																
4 CPs (%)				100				100				100				100
5 CPs (%)	74.13	100	100		01.93	99.40	100									
6 CPs (%)	20.46				75.86	00.60			01.60	99.40	100					
7 CPs (%)	04.13				17.80				77.60	00.60			02.06	99.06	100	
8 CPs (%)	01.13				03.66				15.93				76.53	00.93		
9 CPs (%)	00.13				00.60				03.80				17.40			
10 CPs (%)					00.13				00.86				02.86			

TABLE 2.6 – Pourcentage, par rapport à 1500 réalisations, des nombres de CPs retenues dans les ensembles des données par les critères minimisés ($N = 6000$ observations générées selon trois différents cas de bruit *i.i.d.*)

	Ensemble \mathcal{A}	Ensemble \mathcal{B}	Ensemble \mathcal{C}	Ensemble \mathcal{D}
Critère :	MDL	MDL	MDL	MDL
Bruit <i>i.i.d.</i> : $\sigma^2 = 0.002$				
5 CPs (%)	100			
6 CPs (%)		100		
7 CPs (%)			100	
8 CPs (%)				100
Bruit <i>i.i.d.</i> : $\sigma^2 = 0.2$				
5 CPs (%)	100			
6 CPs (%)		100		
7 CPs (%)			100	
8 CPs (%)				100
Bruit <i>i.i.d.</i> : $\sigma^2 = 0.5$				
5 CPs (%)	100	01.46		
6 CPs (%)		98.53	01.53	
7 CPs (%)			98.46	02.00
8 CPs (%)				98.00

TABLE 2.7 – Pourcentage, par rapport à 1500 réalisations, des nombres de CPs retenues dans les ensembles des données par le critère **MDL** ($N = 26000$ observations générées selon trois différents cas de bruit *i.i.d.*)

Néanmoins, [Wax et Kailath \(1985\)](#) ont montré que le nombre des CPs sélectionnées par le critère **MDL** converge vers le nombre optimal en considérant un nombre d'observations N assez important. Dans un premier temps, nous avons reconsidéré les mêmes ensembles des variables avec les mêmes variances du bruit mais pour un nombre d'observations $N = 6000$. Ainsi, nous avons refait le calcul des pourcentages des nombres de CPs retenues par les critères étudiés (tableau 2.6). Ce calcul a été réalisé en considérant 1500 réalisations.

La comparaison des résultats du tableau 2.6 à ceux du tableau 2.4 ne montre pas d'améliorations particulières aux niveaux des sélections par les critères **AIC**, **IE** et **VNR**. Toutefois, le critère **MDL** a été remarquablement influencé par l'augmentation du nombre d'observations en assurant avec une performance de 100% le nombre adéquat des CPs dans tous les ensembles de variables entachées par un bruit de variance $\sigma^2 = 0.2$. Pour le cas d'une variance $\sigma^2 = 0.5$, ce critère manque encore la CP posant le problème du départ dans les ensembles \mathcal{B} , \mathcal{C} et \mathcal{D} . Par contre, nous remarquons que les pourcentages de sélection ont quand même subi une très légère modification suite à l'augmentation du nombre d'observations. Malgré que ces observations soient largement suffisantes pour décrire correctement le comportement du système étudié, il semblerait encore insuffisantes pour que le critère **MDL** définisse correctement le nombre adéquat des CPs.

Dans ce cadre, nous avons augmenté le nombre d'observations à une valeur très importante $N = 26000$. Les nouveaux pourcentages obtenus pour le critère concerné sont affichés dans le tableau 2.7. D'après ce dernier, nous remarquons que le critère **MDL** converge dans 98% des réalisations vers les nombres souhaités des CPs dans le cas d'un bruit de variance $\sigma^2 = 0.5$. Pour les deux premiers cas du bruit, le critère atteint son optimal avec un nombre d'observations N moins inférieur.

Il est ainsi clair que la performance du critère **MDL** est proportionnelle au nombre d'observations. Dans la pratique, on dispose généralement d'un nombre fini d'observations. Par conséquent, l'utilisation d'un tel critère malgré son efficacité prometteuse ne garantit

pas la convergence vers le nombre convenable des CPs. Si les observations disponibles sont insuffisantes pour ce critère, ce dernier abandonne généralement les CPs qui correspondent aux variables indépendantes ou quasi-indépendantes ayant de faibles variances.

2.6 Conclusion

Le principe mathématique d'une ACP des données sans bruit de mesures nous a permis d'avoir une idée claire sur la définition d'une structure optimale d'un modèle ACP. Dans la pratique, la détermination d'une telle structure n'est pas assez simple qu'on l'imagine à cause de la présence du bruit dans les données. En se référant aux critères de choix de la structure adaptée du modèle ACP qui existent dans la littérature, nous avons choisi d'une manière non exhaustive quelques critères parmi les plus connus pour une évaluation de leur performance sur un exemple simulé. Selon le principe d'utilisation de ces critères, nous avons distingué deux principales familles.

La première présente des critères heuristiques qui se basent généralement sur des seuils pour la sélection du nombre optimal des CPs tels que l'autocorrélation (**AC**), le critère de Kaiser-Guttman (**KG**), le pourcentage cumulé de la variance (**PCV**) et le pourcentage de la variance résiduelle (**PVR**). Dans ce chapitre, ces critères ont été utilisés sur des données normalisées. Nous avons conclu à travers l'exemple de simulation qu'une décision basée sur les deux critères **PCV** et **PVR** se complique avec l'augmentation du nombre de variables étudiées ainsi que la variance du bruit. En effet, les allures des courbes représentant les valeurs propres peuvent devenir très lentes, ce qui favorise l'inexistence du point d'inflexion pour le critère **PVR**. Ceci implique également une multitude des nombres de CPs dans l'intervalle de sélection pour le critère **PCV**. Quant à la décision basée sur le critère **KG**, on constate qu'elle est discriminante car elle a éliminé les variables indépendantes et quasi-indépendantes. En ce qui concerne le critère **AC**, son inconvénient se résume principalement dans le fait qu'une CP significative peut avoir une faible valeur d'autocorrélation qui ne lui permis pas d'être retenue par un tel critère. Malgré leurs popularités, ces critères sont subjectifs et largement restreints en termes de décisions et efficacités respectivement.

La deuxième famille est constituée des critères qui se basent sur la minimisation pour déterminer la dimension du modèle ACP. Les trois premiers critères, notamment **AIC**, **MDL** et **IE** possèdent deux points communs. Premièrement, leur utilisation n'est valable qu'avec des données non réduites. Deuxièmement, le bruit de mesures est supposé être indépendant et identiquement distribué. Le quatrième critère de cette famille représente la variance non reconstruite (**VNR**). Ce dernier peut être utilisé aussi bien sur des données réduites ou non réduites. Dans le cadre d'une ACP, il est cependant préférable que les données soient exprimées dans la même échelle. Nous avons étudié le comportement du critère **VNR** en considérant des données normalisées car il est en relation directe avec les paramètres fournis par l'ACP.

Puisque les bruits de mesures représentent des variables aléatoires, une seule simulation ne permet pas de juger l'efficacité de ces critères. Pour cette raison, nous avons établi des pourcentages sur les nombres des CPs retenues par tous les critères en considérant 1500 réalisations. Ainsi, notre étude comparative a pris en compte plusieurs facteurs tels que

la nature des variables étudiées (corrélées ou indépendantes), la valeur de la variance du bruit considéré et le nombre d'observations.

Selon les résultats de simulation, le critère **AIC** surestime souvent le nombre des CPs nécessaires. L'avantage du critère **VNR** étant la considération de toutes les variables étudiées dans la même échelle. Malgré cette caractéristique, nous avons découvert que ce critère ne prend pas en compte les variables indépendantes et quasi-indépendantes même si en absence du bruit ces variables disposent des variances non nulles. Dans ce contexte, nous avons contribué par une démonstration théorique confirmant la limitation de ce critère. En effet, nous avons montré que le nombre des CPs qui correspond au minimum d'un tel critère ne change pas en ajoutant des variables indépendantes et quasi-indépendantes. Autrement dit, le critère **VNR** est insensible à la présence de ce type des variables. Ainsi, son minimum correspond au nombre des CPs uniquement pour les variables qui sont linéairement corrélées.

Dans cette étude comparative, nous avons remarqué que le critère **IE** abandonne souvent les CPs ayant de faibles variances bien qu'elles soient théoriquement supposées être retenues dans le modèle. Ce critère ne surestime pas le nombre des CPs. Dans ce contexte, nous avons étudié son expression afin de vérifier sa consistance théorique. Ainsi, nous avons établi une condition nécessaire et suffisante permettant au critère **IE** de garantir son minimum qui correspond au nombre optimal des CPs. En effet, ce critère ne retient plus la dernière CP si sa variance en l'absence du bruit est inférieure à une valeur établie par une telle condition. Alors qu'en théorie cette CP doit être retenue. En outre, cette valeur est proportionnelle au nombre des variables, ce qui présente un inconvénient majeur car cela implique que la CP en question peut être retenue dans des cas et écartées dans d'autres.

Le **MDL** semble être le critère le plus efficace. Cependant, sa performance est proportionnelle au nombre d'observations considérées. Sa sélection converge vers le nombre correct des CPs si le nombre d'observations utilisées est assez important. Dans la pratique, un système peut être suffisamment décrit par un nombre N d'observations. Ce nombre peut être insuffisant pour que le critère **MDL** exprime correctement le nombre adéquat des CPs. Dans ce cas, on ne peut pas savoir si un tel critère converge finalement vers le nombre correct des CPs. D'après les résultats de simulation, nous avons remarqué que si le nombre d'observations n'est pas suffisant, ce critère ne retient pas souvent les CPs associées aux variables indépendantes de faibles variances.

Cette étude comparative entre les critères choisis nous a permis d'illustrer leurs inconvénients et avantages. Dans ce cadre, seuls les critères **MDL** et **VNR** peuvent être considérés intéressants vis-à-vis de leur comportement et principe respectivement. Cependant, ils présentent des inconvénients rendant souvent les décisions incertaines. Pour cette raison, nous proposons dans le chapitre suivant d'autres critères utilisant le principe du celui de **VNR** et assurant les résultats obtenus par celui de **MDL** avec des conditions plus réalistes.