

# Réduction de dimension des prix nodaux dans un réseau de distribution

Guénolé CHÉROT, Roman LE GOFF LATIMIER, Hamid BEN AHMED

SATIE, ENS Rennes, CNRS

**La généralisation des productions distribuées nécessite d'adapter les méthodes de gestion du réseau de distribution. La résolution du problème d'optimal power flow (OPF) indique que la gestion optimale d'un réseau ne peut être atteinte qu'en utilisant un nombre de variables minimales égales au nombre de bus le constituant. Or un déploiement réel devra vraisemblablement se limiter à l'envoi d'une ou de quelques variables aux agents présents sur le réseau, ainsi qu'en témoigne la riche littérature actuellement consacrée à ce sujet. Une question primordiale est alors de connaître l'écart d'optimalité entre la solution optimale et celle utilisant un nombre de variables réduit. En se basant sur les variables de contrôle issues de l'OPF et en utilisant l'analyse en composante principale (ACP), cette contribution propose de construire de nouvelles variables, dites latentes, décorrélées entre elles. Le contrôle du réseau par ces variables permet de montrer, sur le cas test considéré, que seulement 30% des variables sont nécessaires à l'obtention d'un contrôle proche de l'optimalité. Ce constat permet d'envisager l'apprentissage de ces variables pour des réseaux dans lesquels leur calcul opérationnel est impossible en temps réel.**

**Mots-clés – Réseau de distribution, prix nodaux, réduction de dimension, analyse en composante principale, Optimal power flow**

## 1. INTRODUCTION

La transition vers une production d'énergie moins carbonée ne pourra se faire sans la généralisation des énergies renouvelables (EnRs) [1]. Celles-ci sont distribuées et principalement connectées aux réseaux de distribution (RD) [2]. À l'heure actuelle, le faible taux de pénétration des EnRs, le dimensionnement des lignes et le peu de flexibilités disponibles conduisent à se passer de gestion. Cependant une évolution vers des réseaux où 100% des acteurs seraient équipés de moyen de production distribuée, et acheminant de surcroît les consommations liées à la mobilité électrique, la moitié des RDs pourraient avoir des problèmes de congestions ou de non-respect du plan de tension [3]. Le gestionnaire du réseau de distribution (GRD) doit donc trouver de nouveaux leviers afin de continuer à assurer la distribution d'une énergie de qualité.

De nombreuses stratégies de gestion s'appuyant sur l'existence de flexibilités distribuées –véhicules électriques, pompes à chaleur, etc.– et sur la démocratisation des compteurs communicants –Linky par exemple– sont actuellement étudiées activement. La résolution directe formulée comme un problème d'optimisation prenant en compte l'ensemble des contraintes physiques du réseau a été proposé la première fois par Carpentier [4] sous le nom de flux de puissance optimal (OPF pour optimal power flow). La solution de ce problème est usuellement un plan d'injection de puissances actives et réactives – variables primales du problème – à communiquer à chaque producteur – et éventuellement consommateurs si ceux-ci sont flexibles. La même situation optimale peut également être mise en application en communiquant à chaque nœuds les prix nodaux de l'électricité (PN) – variables duales du problème [5, 6]. Dans la perspective d'un déploiement opérationnel, ce second schéma de communication, non injonctif pour les utilisateurs, semble plus adapté à un contrôle distribué. Bien qu'optimale, cette solution n'est pas parfaitement adaptée au réseau de distribution

à cause du nombre d'agents impliqués [7] et des incertitudes du réseau [8]. La relaxation de certaines contraintes permet de rendre le problème convexe, ce qui accélère grandement l'optimisation. Dans certaines conditions, les relaxations ont un écart de sous-optimalité nul [9, 10] mais la résolution nécessite la connaissance parfaite des agents du réseau. Or ces derniers pourraient vouloir garder leurs préférences privées.

Ce problème a motivé l'essor des algorithmes pair à pair. Barroche [11] a par exemple proposé deux méthodes permettant l'intégration du GRD au sein d'un marché pair à pair. Celui-ci pouvait contraindre les échanges grâce à un prix d'utilisation du réseau, s'assurant ainsi du respect des contraintes de congestion. Faria [12] propose un autre type de pénalisation basée sur la distance entre agents, et le compare avec une approche basée sur un marché de flexibilité dans lequel le GRD proposerait des offres de flexibilités aux agents jusqu'à ce que l'ensemble des contraintes soient résolues. Cette approche, s'appuyant sur la présence d'aggrégateurs gérant des contraintes locales [13] semble être la plus prometteuse. Enfin, une dernière approche consiste à formuler la gestion des flexibilités sous la forme d'un problème d'apprentissage par renforcement dans lequel le GRD décide à chaque instant des puissances échangées par chaque agent et où le signal de récompense pénalise le non-respect des contraintes [14, 15].

La diversité de ces méthodes souligne le dynamisme des recherches actuellement consacrées à la question ouverte de la gestion des réseaux de distribution actifs. Elles ont pourtant un point commun : l'utilisation d'un nombre limité de signaux permettant de contrôler ou d'influencer les nombreux acteurs du réseau de distribution. Or ce choix implique nécessairement une sous-optimalité du contrôle qu'il est nécessaire de quantifier.

La présente contribution cherche donc à répondre aux questions suivantes : i) quelle est la sous-optimalité induite par le contrôle d'un réseau utilisant un nombre insuffisant de variables ? ii) pour un nombre de variables fixé, comment construire le jeu de variables optimales ?

Pour répondre à ces questions, nous nous appuierons sur la littérature des prix nodaux [5, 6]. Ces variables issues de l'OPF permettent le contrôle optimal du réseau si les agents qui le composent sont rationnels. En étudiant la structure de ces variables grâce à des méthodes d'extraction de caractéristiques [16] comme l'Analyse en Composante Principale ACP [17], nous montrerons qu'une majorité de l'information est contenue dans quelques variables.

Le reste de l'article sera organisé de la façon suivante. La section 2 présentera l'ACP et son utilisation comme outil de compression d'information. La section 3 présentera le cas test et les outils de simulation utilisés. Enfin la section 4 présentera les résultats.

## 2. MÉTHODE

### 2.1. Analyse en Composantes Principales

L'Analyse en Composante Principale (ACP) est une méthode d'analyse de donnée qui consiste à transformer des variables corrélées en nouvelles variables décorrélées les unes des autres. La transformation induite par l'ACP est linéaire et peut être vue

comme un changement de base [17, 18]. Les vecteurs unitaires formant la nouvelle base –appelé espace latent– sont orthogonaux entre eux. Ils sont ordonnés du plus au moins important au sens de la variance expliquée des données d’entrée.

Soit un vecteur de  $N$  variables aléatoires notées  $\mathbf{x} = (x_1, \dots, x_N)$  de moyenne nulle. On nomme  $x_i$  les composantes de  $\mathbf{x}$  dans la base initiale  $(e_1, \dots, e_N)$ . On note  $\mathbf{x}^{[m]}$  la  $m^{i\text{eme}}$  observation de  $\mathbf{x}$ . L’ensemble des  $M$  observations forme la matrice  $\mathbf{X} \in \mathcal{M}_{M,N}$  :

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{[1]} \\ \mathbf{x}^{[2]} \\ \dots \\ \mathbf{x}^{[M]} \end{pmatrix} = \begin{pmatrix} x_1^{[1]} & x_2^{[1]} & \dots & x_N^{[1]} \\ x_1^{[2]} & x_2^{[2]} & \dots & x_N^{[2]} \\ \dots & \dots & \dots & \dots \\ x_1^{[M]} & x_2^{[M]} & \dots & x_N^{[M]} \end{pmatrix} \quad (1)$$

La matrice  $\mathbf{T}$  permet de passer de l’espace initial à l’espace latent  $(e'_1, \dots, e'_N)$  de la manière suivante :  $x' = \mathbf{x}\mathbf{T}$ . Cette matrice est constituée des vecteurs propres de la matrice de covariance  $\mathbf{A} = \mathbf{X}\mathbf{X}^T/(M-1)$ . Les valeurs propres de  $\mathbf{A}$  indiquent un ordre d’importance des vecteurs propres. Le pourcentage d’information contenu dans la variable latente  $x'_i$  est :

$$\%_i = \frac{\lambda_i}{N} \quad (2)$$

où  $\lambda_i$  est la valeur propre associée au vecteur propre  $e'_i$ .

Il est courant que l’essentiel de l’information soit contenu dans seulement quelques variables latentes. Il est alors possible de définir une nouvelle observation contenant seulement une partie de l’information :

$$\tilde{\mathbf{x}}'^{(k)} = (x'_1, \dots, x'_k, 0, \dots, 0) \quad (3a)$$

$$= \mathbf{x}'\mathbf{K} \quad (3b)$$

où  $\tilde{\mathbf{x}}'^{(k)}$  est la nouvelle observation compressée et  $\mathbf{K} \in \mathcal{M}_{N,N}$  la matrice conservant uniquement les  $k$  premières composantes de  $\mathbf{x}'$ .

## 2.2. Application aux prix noraux

Dans le contexte de la présente étude, l’impact de la réduction de dimension sur la performance de la gestion d’un réseau de distribution est investigué. Pour ce faire, une méthode de réduction de dimension doit être sélectionnée. Le choix se porte ici sur l’ACP du fait de sa simplicité de mise en œuvre, de son faible coût calculatoire et de l’explicabilité des résultats obtenus. Il n’en demeure pas moins que cette méthode n’est pas la seule qui aurait pu être appliquée ici. À la lumière des résultats obtenus, la section 4.4 discutera la pertinence de ce choix et les différences pouvant être anticipées s’il venait à être modifié.

En utilisant l’ACP pour le présent problème, les corrélations entre les prix noraux seront mises en évidence. Il faut donc récolter un grand nombre d’observations représentatives du fonctionnement d’un réseau de distribution. Pour cela, nous disposons d’une base de données (BDD) contenant des séries temporelles de productions et de consommations décrivant la puissance demandée par chacun des agents sur le réseau. Combiné à un OPF, elles nous permettent de générer les données nécessaires à l’ACP. Pour le reste de cette contribution, cette étape sera nommée *phase d’entraînement*.

Une fois l’ACP effectuée, la performance de la réduction de dimension est testée sur une BDD de test. Le but de cette étape est de quantifier l’écart entre la solution optimale et la solution basée sur les prix latents compressés – écarts de tensions, de flux dans les lignes et de coût total. Cette étape constitue la *phase de test*.

La méthode de gestion à base de prix noraux réduits est décrite Fig. 1. Elle est détaillée ci-dessous.

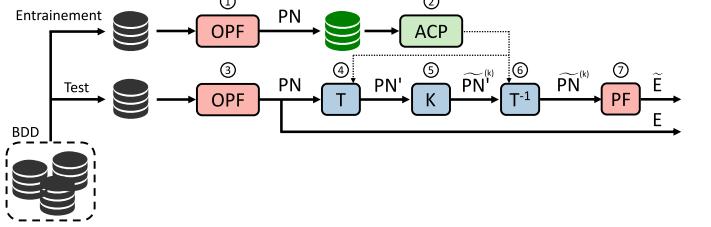


FIG. 1. Méthode de réduction de dimension des prix noraux. Durant la première phase – données d’entraînement – la matrice de transformation  $\mathbf{T}$  est identifiée grâce à une analyse en composante principale. Lors de la deuxième phase – données de test – un réseau de distribution est contrôlé à l’aide de prix noraux  $\tilde{\mathbf{PN}}$  dont l’information a été compressée.

**Entraînement :** ① Pour chaque point de fonctionnement issu de la BDD d’entraînement, nous calculons un OPF pour obtenir les prix noraux. On note  $\mathbf{pn} = (pn_1, \dots, pn_N)$  le vecteur des  $N$  prix noraux, correspondant à chaque nœud du réseau. La base de données de prix noraux est constituée de l’ensemble des  $\mathbf{pn}$  observés au cours de la période considérée pour cette phase d’entraînement. On a donc :  $\mathbf{PN} = (\mathbf{pn}^{[1]}, \dots, \mathbf{pn}^{[M]})$  avec  $M$  observations – issue des  $M$  points de fonctionnement de la BDD d’entraînement. ② Les prix noraux  $\mathbf{PN}$  sont normés selon la méthode décrite (4). Puis l’ACP est réalisée afin de calculer la matrice de transformation  $\mathbf{T}$ .

$$\mathbf{x}_{\text{normé}}^{[i]} = \frac{\mathbf{x}^{[i]} - \bar{\mathbf{X}}}{\sigma(\mathbf{X})} = \left( \frac{x_1^{[i]} - \bar{x}_1}{\sigma(x_1)}, \dots, \frac{x_N^{[i]} - \bar{x}_N}{\sigma(x_N)} \right) \quad (4)$$

**Test :** ③ Une fois l’entraînement réalisé, des configurations originales – c’est-à-dire non incluses dans la BDD d’entraînement – sont examinées. De même qu’en ①, pour chacune de ces configurations, les prix noraux optimaux sont calculés à l’aide d’un OPF. ④ Ils sont ensuite projetés dans l’espace latent grâce à la matrice de transformation  $\mathbf{T}$  obtenue lors de l’étape ②. ⑤ La matrice  $\mathbf{K}$  permet de conserver uniquement les  $k$  premières composantes, les autres sont remplacées par des zéros. Les composantes d’une ACP étant ordonnées, seule l’information la moins importante est supprimée. ⑥ Les vecteurs latents compressés sont projetés dans l’espace d’origine à l’aide de  $\mathbf{T}^{-1}$ . L’ensemble de l’opération peut se résumer de la manière suivante :  $\tilde{\mathbf{pn}}^{(k)} = \mathbf{pn}\mathbf{T}\mathbf{K}\mathbf{T}^{-1}$ . ⑦ Enfin, les  $\tilde{\mathbf{pn}}$  sont transmis à tous les acteurs du réseau pour qu’ils en déduisent leur injection de puissance. Un power flow est réalisé afin de comparer l’état du réseau respectivement sous gestion optimale  $E$  et sous gestion dégradée  $\tilde{E}$ .

## 3. CAS TEST ET OUTILS DE SIMULATION

Le *Low Voltage European Test Feeder* [19] a été sélectionné comme cas d’application, car il est représentatif des RD Européens. Une représentation graphique de ce réseau est donnée figure 5.

Les acteurs au sein du réseau sont tirés au sort à partir d’une base de données. Elle contient des séries temporelles au pas de temps minute, de puissances échangées par des foyers, des véhicules électriques (VE) et des panneaux photovoltaïques (PV). Le nombre de chaque type d’acteur est spécifié à la création du cas test. Un prix  $\theta$  est ensuite associé à chaque acteur. Il sera supposé fixe pour toute la simulation et tiré au sort suivant une loi normale  $\mathcal{N}(\mu, \sigma)$ . La puissance de chaque type d’acteur, et les valeurs de  $\mu$  et  $\sigma$  sont donnés tableau 1. La puissance échangée est volontairement plus importante que sur le cas test original de sorte à activer les contraintes du plan de tension sans quoi les prix noraux seraient toujours nuls.

TABLEAU 1. Valeurs utilisées pour la création du cas test.

	$\mu$	$\sigma$	Puissance totale pic
Foyer	0.25	0.10	1000 kW
VE	0.15	0.05	200 kW
PV	0.08	0.02	800 kW

L'OPF est un problème d'optimisation non convexe qui n'a donc pas de solution exacte dans le cas général. Nous utiliserons la formulation proposée par [20] qui combine une relaxation de la contrainte sur les angles des tensions et une relaxation de la contrainte d'égalité sur le flux dans les lignes [21]. Elles nous permettent de calculer la solution exacte de l'OPF dans le cas de réseaux non maillés et lorsque le flux de puissance inverse n'est qu'actif, que réactif, ou nul. L'optimalité de la solution est systématiquement vérifiée a posteriori. Le lecteur est invité à se référer à [10] pour une analyse complète.

Dans une démarche de science ouverte et reproductible, l'ensemble des données utilisées sont publiquement accessibles : [22] pour la consommation des foyers et les productions éoliennes, [23] pour les consommations de véhicules électriques et EPEX<sup>1</sup> pour le prix du marché de l'énergie. L'écoulement de puissance et l'OPF ont été réalisé grâce aux bibliothèques PandaPower [24] et PowerModels [25]. Nous utilisons l'implémentation de l'ACP proposée par Scikit-Learn [26]. Les codes sources de cette étude sont ouverts et accessibles sur GitLab<sup>2</sup>.

Les simulations ont été conduites en séparant les 365 jours de la base de données en deux groupes : 60 % des jours ont servi à l'entraînement, 40 % pour le test.

## 4. RÉSULTATS ET ANALYSES

Toutes les études présentées dans cette section se basent sur le cas présenté section 3. Les performances d'une gestion par OPF de cet exemple de réseau de distribution sont comparées avec les performances obtenues lorsque les prix nodaux sont obtenus après compression par une ACP.

### 4.1. Illustration d'une configuration particulière

La figure 2 montre deux exemples de plans de tension, obtenus à deux moments différents sur ce même cas d'étude. La première configuration – panneau supérieur – illustre un cas pour lequel la réduction conduit à une gestion peu efficace, voire dangereuse. La deuxième – panneau inférieur – est une situation où une très forte réduction peut être appliquée sans difficulté. Les tensions sont représentées en fonction de leurs distances au Point de Couplage Commun PCC afin de garder la structure du réseau et faciliter la lecture des résultats. Le PCC joue le rôle de référence, sa tension est fixée à 1 p.u. Le plan de tension obtenu en utilisant les PN optimaux issus d'un OPF est illustré en vert. Lorsque les PN sont compressés en utilisant seulement 1% – respectivement 2% – des dimensions de l'espace latent, le plan de tension obtenu est représenté en bleu – respectivement orange. Les limites de tension admissibles – fixées à 1.1 p.u. et 0.9 p.u. sont représentées en pointillés rouges.

La configuration n°1 illustre un cas pour lequel la réduction fonctionne très bien en utilisant 2 % des variables latentes, mais très mal en dessous. En effet, pour 1 % des variables latentes, 90 % de bus ne respecte pas les contraintes. L'écart maximal avec la solution optimale, de 0.22 p.u., est obtenu en bout de ligne. Dans cette configuration, le réseau optimalement géré aurait du être exportateur de puissance, ainsi qu'en témoigne les tensions qui sont supérieures à 1 p.u. Cependant une compression à 1% des prix nodaux engendre un écart qui rend le réseau importateur, d'où des écarts de tension significatifs.

La configuration n°2 correspond à un moment où le réseau

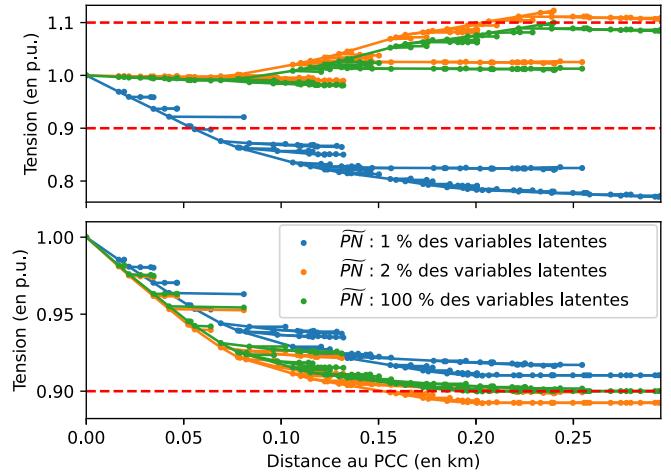


FIG. 2. Évolution de la tension de chaque bus en fonction de sa distance avec le PCC et selon la méthode de gestion utilisée : OPF en vert, en utilisant 1% des variables latentes en bleu et en utilisant 2% en orange. Les pointillés rouges représentent les contraintes du plan de tension  $\pm 10\%$ .

considéré est importateur de puissance. Les tensions sont donc décroissantes lorsqu'on s'éloigne du PCC. Elles présentent une chute de 0.05 p.u. en moyenne. On constate que plus le nombre de variables latentes augmente, plus les écarts entre le plan de tension optimal et les plans de tensions compressés se réduisent. En particulier, les écarts peuvent être positifs ou négatifs. Notons que des PN mal estimés ne conduisent pas obligatoirement à une violation du plan de tension. Il est remarquable que la stratégie conservant 2 % des variables latentes conduise à plus de violations que celle en utilisant uniquement 1%. En effet, une meilleure approximation des PN conduit les tensions à se rapprocher des tensions optimales issues de l'OPF. Or celui-ci les a souvent placées à leurs limites admissibles comme nous pouvons l'observer sur la courbe verte Fig. 2. Ainsi, une infime erreur de PN peut conduire à un non-respect des contraintes. Les phénomènes ici illustrés permettent de percevoir les difficultés et risques liés à une réduction de dimension. Bien qu'une amélioration globale de la gestion est attendue lorsque la compression est moins importante, beaucoup de phénomènes particuliers sont susceptibles d'apparaître ponctuellement. L'évolution des performances peut donc ne pas être monotone lors de l'étude d'un réseau en particulier.

### 4.2. Sous-optimalité induite par la réduction de dimension

Après avoir illustré l'impact de la réduction de dimension sur les configurations particulières de la figure 2, cette section systématisé l'étude des performances d'une gestion par des PN compressés. La figure 3 décrit les écarts entre la stratégie optimale (OPF) et la stratégie avec prix nodaux compressés, en termes d'information capturée, de différence des coûts, de différences sur les tensions observées et de fréquence des violations de tension. Elle est essentielle, car elle permet – pour un réseau fixé – d'estimer les performances maximales d'une stratégie de gestion utilisant un nombre de variables fixé. Cette figure agrège les résultats obtenus en simulant les stratégies optimales et compressées sur toute la durée du cas d'étude.

Le panneau 1 présente le pourcentage de la variance expliquée en fonction du nombre de composantes conservées dans l'espace latent – exprimé en % pour plus de généralité. Pour ce réseau, 89% de l'information contenue dans les PN est capturée dès la première composante et 99% dès deux composantes. Au sens de cette métrique, très peu de dimensions dans l'espace latent suffisent à expliquer l'ensemble des PN observés. Cette observation était prévisible : les PN sont tous identiques si aucune contrainte de tension ou de congestion n'est activée. Dans notre cas, et même en ayant surchargé le réseau (voir sec-

1. <https://ewoken.github.io/epex-spot-data/>

2. [https://gitlab.com/satие.sete/sge\\_dim](https://gitlab.com/satие.sete/sge_dim)

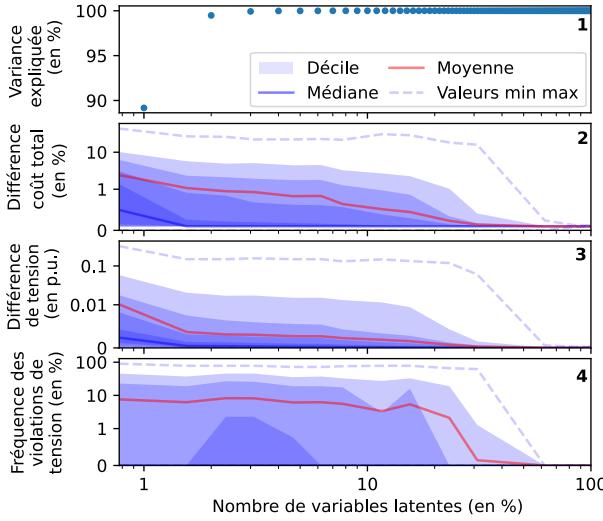


FIG. 3. Effet de la réduction de dimension sur la sous-optimalité du contrôle. Le panneau 1 représente l'information contenue dans chaque composante de la représentation latente. Les panneaux suivants représentent différents écarts  $|E - \tilde{E}|$  entre la stratégie OPF et la stratégie proposée en fonction du nombre de composantes conservées dans l'espace latent (étape ⑤). Les panneaux 2 et 3 représentent respectivement les écarts de coûts et du plan de tension. Le panneau 5 décrit la fréquence des violations de tension.

tion 3), cela ne se produit que 27% du temps. Cependant, à la lumière des évolutions décrites sur les panneaux suivants, cette perte d'information, semblant minime à ce stade, a cependant de très fortes répercussions sur les grandeurs électriques du réseau.

Les panneaux 2, 3 et 4 décrivent les écarts entre les deux stratégies. Le lecteur notera que pour l'ensemble des panneaux, l'axe des ordonnées est linéaire entre 0 et la première graduation puis logarithmique pour les valeurs supérieures. Les valeurs extrêmes sont représentées par des pointillés, et les déciles par des dégradés de couleurs. Ainsi, le premier trait plein en partant du haut correspond au décile à 90 %. Les valeurs médiane et moyenne sont également représentées. Ces statistiques ont été construites à partir des ensembles décrit équation 5 :

$$\text{Diff coût tot}(k) = \left\{ 100 \cdot \left| \sum_{i \in \mathcal{A}} \frac{\theta_i P_i^{[j]} - \theta_i \tilde{P}_i^{[j]}}{\theta_i P_i^{[j]}} \right|, \forall j \in \mathcal{O}_k \right\} \quad (5a)$$

$$\text{Diff de tension}(k) = \left\{ |V_i^{[j]} - \tilde{V}_i^{[j]}|, \forall (j, i) \in \mathcal{O}_k \times \mathcal{N} \right\} \quad (5b)$$

$$\text{Fréq violation tension}(k) = \left\{ \frac{100}{\#\mathcal{N}} \sum_{i \in \mathcal{N}} \mathbb{1}_{\tilde{V}_i^{[j]}}, \forall j \in \mathcal{O}_k \right\} \quad (5c)$$

où  $k$  est le nombre de dimensions conservées dans l'espace latent,  $\mathcal{O}_k$  est l'ensemble des observations conservant  $k$  dimensions,  $x^{[j]}$  dénote l'observation  $j$  d'une grandeur  $x$  et  $\tilde{x}$  d'une grandeur issue du contrôle par prix nodaux réduits,  $\theta_i$  et  $P_i$  sont le prix et la puissance échangée par l'agent  $i$ ,  $\mathcal{A}$  est l'ensemble des agents du réseau,  $\mathcal{N}$  est l'ensemble des nœuds du réseau et  $\#\mathcal{N}$  leur nombre,  $\mathbb{1}_{\tilde{V}_i^{[j]}}$  est la fonction indicatrice de la violation d'une contrainte de tension.

La panneau 2 illustre les écarts de coût total, c'est-à-dire la somme de coût payé par chaque agent du réseau. Ainsi qu'indiquée lors de la section précédente, la tendance générale est bien que les performances se rapprochent de la gestion optimale lorsque plus de dimensions sont prises en compte. Cette tendance n'est cependant pas tout à fait monotone ni de rythme uniforme. En particulier, les valeurs extrêmes sont les plus difficiles à estimer lors d'une simulation de type Monte-Carlo. En ne conservant qu'une seule variable lors de la compression – extrême gauche de la figure – des écarts de performances de l'ordre

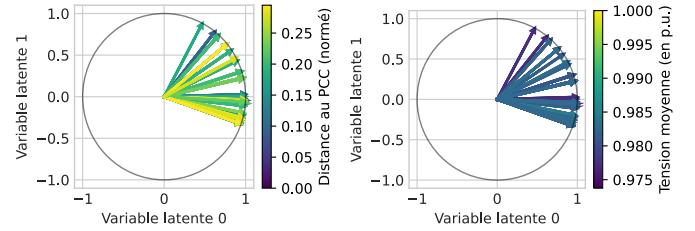


FIG. 4. Cercle de corrélation de l'ACP. Représente les corrélations entre les anciennes variables (prix nodaux) et les nouvelles variables décorrélées (vecteur de l'espace latent). Tous les prix nodaux ont une contribution équivalente. La distance au PCC et la tension moyenne ne permettent pas de dissocier les prix nodaux.

de quelques pourcents peuvent être constatés en moyenne. Cette valeur est tirée vers le haut par quelques configurations particulièrement mal gérées et coûteuses – voir extrême et dernier décile. Au maximum, on observe 60 % de différence. La médiane est significativement plus basse que la moyenne. Obtenir une gestion globalement performante peut donc être obtenu avec des taux de compression très élevés. En revanche, garantir que toutes les configurations seront bien gérées nécessite de conserver beaucoup plus de variables : un taux de compression de 25% est ainsi nécessaire pour que le dernier décile tombe en dessous de 1% de surcoût. On observe une différence de l'ordre de 0.1 % lorsque toutes les variables sont utilisées. Cet écart, qui devrait être nul, est dû à une erreur de précision numérique, pourrait être atténué en diminuant la tolérance de l'OPF.

Le panneau 3 décrit l'écart de tension en p.u. De manière similaire à l'interprétation du panneau 2, ces écarts décroissent de manière monotone jusqu'à atteindre une erreur nulle. En utilisant seulement 1 % des variables latentes, seules 10 % – 9<sup>eme</sup> décile – des tensions ont un écart supérieur à 0.09 p.u. Nous pouvons également noter que 90 % des écarts sont nuls avec l'utilisation de 30 % de variables latentes.

Le panneau 4 décrit la fréquence des violations de tensions. En ordonnée, 100 % signifie que l'ensemble des bus – à l'exception du PCC – viole le plan de tension. Cette situation se produit dans le pire des cas jusqu'à l'utilisation de 20 % des variables latentes. Cela signifie que certaines configurations sont très sensibles aux valeurs prises par les prix nodaux. Il s'agit de celles pour lesquelles les tensions du point de fonctionnement optimal sont très proches des contraintes. Ce phénomène explique également la non-monotonie du 7<sup>eme</sup> décile comme nous l'avons vu section 4.1.

#### 4.3. Analyse de la décomposition ACP et discussion

Le cercle de corrélations est un outil permettant l'analyse de la transformation construite par l'ACP. Pour l'obtenir, chaque vecteur de l'espace initial  $e_i$  correspondant au prix du nœud  $i$  est projeté sur les deux premières composantes de l'espace latent ( $e'_1, e'_2$ ) formant ainsi un vecteur dont la distance à l'origine décrit l'importance de ce prix vis-à-vis de l'ACP. Les prix positivement corrélés sont regroupés. Il est alors possible de colorier chaque vecteur en fonction d'un critère exogène pour voir s'il permet d'expliquer les groupes ainsi formés. À notre connaissance, six critères exogènes pourraient permettre d'expliquer la structure des données : la distance au PCC et les valeurs minimales, maximales et moyennes de la tension et de la puissance en chaque bus.

Deux cercles de corrélations sont représentés figure 4 pour lesquels les vecteurs représentant chaque prix nodal sont coloriés en fonction respectivement de la distance au point de couplage commun (PCC) et de la tension moyenne sur le bus. L'ensemble des prix nodaux sont très proches du cercle de rayon 1. Ils ont donc une importance similaire. De plus, les deux critères exogènes représentés ne permettent pas de dégager une tendance permettant d'expliquer la répartition des prix nodaux. Les cri-

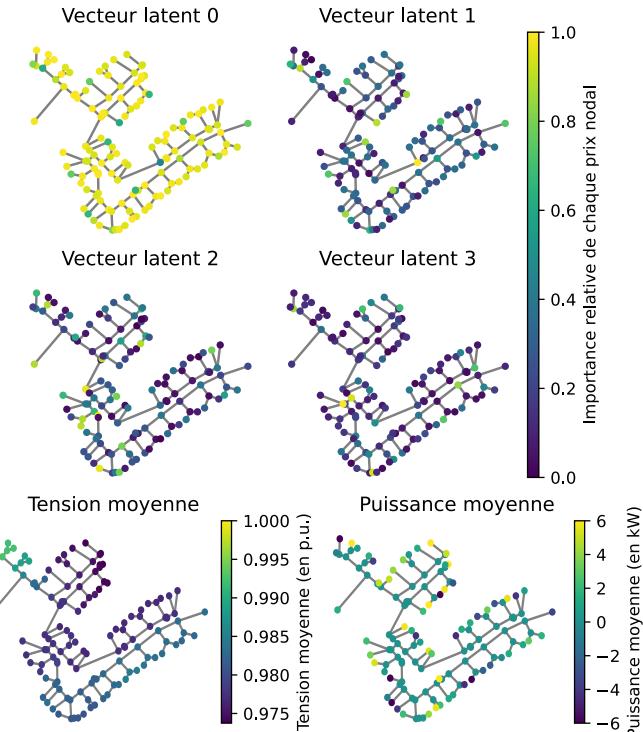


FIG. 5. Analyse des vecteurs propres de la matrice ACP via l'emplacement de chaque prix nodal sur le réseau. Les quatre premiers vecteurs latents sont représentés. Les contributions de chaque prix nodal sont normées de sorte que l'importance maximale pour chaque vecteur soit égale à 1. La figure représente également la tension moyenne et la puissance moyenne observée sur chaque bus lors de la simulation.

ères non représentés conduisent aux mêmes conclusions.

La figure 5 représente l'importance relative de chaque prix nodal dans l'ACP des quatre premiers vecteurs latents. La tension moyenne et la puissance moyenne en chaque bus sont également affichées. Une nouvelle fois, ces variables ne permettent pas d'expliquer intuitivement l'importance des prix nodaux dans l'ACP.

#### 4.4. Discussion de l'algorithme de réduction de dimension

Ainsi que mentionné section 2, l'ACP est une méthode de réduction de dimension parmi beaucoup d'autres. Ce choix a semblé pertinent pour cette première étude du fait de sa simplicité de mise en œuvre et de son explicabilité. Il convient cependant d'en discuter les limites à la lumière des résultats obtenus.

L'ACP est une transformation linéaire. Elle ne permet donc pas d'expliquer de façon certaine la structure des données d'un problème, en particulier si le motif sous-jacent est non linéaire, non orthogonal ou s'il existe des groupes dissimulés [27].

Une approche plus générale pourrait consister à utiliser un auto-encodeur [28, 29], un type de réseau de neurones particulier pour lesquels une des couches profondes possède volontairement un nombre de neurones réduit. En minimisant l'erreur de reconstruction, l'algorithme est contraint à apprendre une représentation compacte des données. La transposition de la présente étude en comparant les performances obtenues à l'aide d'une ACP et celles obtenues à l'aide d'autres outils de réduction de dimension serait donc un prolongement naturel de la démarche ici proposée.

## 5. CONCLUSIONS

Les réseaux de distribution étant un lieu incontournable de la transition énergétique, la gestion optimale des agents qui y sont connectés est un enjeu essentiel au sein de la littérature des réseaux de distribution actifs. La présente contribution s'in-

téresse au nombre de signaux nécessaires pour parvenir à une gestion satisfaisante, en se plaçant en amont de beaucoup des contributions de la littérature qui proposent des méthodes de gestion spécifiques. Les prix nodaux obtenus par un OPF ont ici été utilisés comme référence de performance optimale. Celle-ci est comparée avec les prix obtenus via une réduction de dimension par ACP. Il a été mis en évidence sur un cas d'étude que 99% de la variance peut être expliquée grâce à 2% des composantes de l'espace latent. Cependant, cette métrique n'est pas suffisante pour la gestion d'un RD, les prix nodaux étant structurellement fortement corrélés entre eux. De fortes violations du plan de tension peuvent donc être observées jusqu'à l'utilisation de 20% des composantes latentes. Afin de garantir d'éliminer ces configurations rares, seules des compressions conservant au moins la moitié des dimensions originales peuvent être envisagées. Les contributions de chaque composante de l'ACP ont été examinées. Aucune structure interprétable dans sa matrice de construction n'a pu être mise en évidence à ce stade.

À l'issue de ce travail, trois perspectives principales sont ouvertes. D'une part cet article constitue une contribution préliminaire qui doit être systématisée afin de gagner en générativité. En particulier une réduction de dimension par ACP constitue un choix justifié, mais nécessite une comparaison avec d'autres méthodes de réduction de dimension. En particulier répéter cette étude en remplaçant l'ACP par des algorithmes capables de capturer des comportements non linéaires – comme un réseau de neurones autoencodeur – est nécessaire pour confirmer les résultats obtenus. De même répéter l'étude sur d'autres configurations de réseau permettrait de généraliser les conclusions. Les performances de la réduction de dimension ne sont potentiellement pas les mêmes lorsque le réseau est importateur ou exportateur ainsi qu'illustre figure 2. Discuter les performances selon les caractéristiques du réseau serait donc prometteur.

D'autre part la présente étude permet de prédire une performance atteignable en fonction du nombre de grandeurs conservées. Cependant elle ne fournit pas un algorithme de gestion opérationnel permettant d'atteindre cette performance. Comparer des méthodes de gestion des réseaux de distribution actifs par rapport à la performance prédictive ici serait donc particulièrement intéressant. Cela permettrait d'évaluer les potentielles lacunes restantes dans les méthodes proposées par la littérature. Le potentiel de progrès restant serait ainsi quantifié.

Enfin, cette contribution ouvre la possibilité à de nouvelles méthodes de gérances opérationnelles du réseau de distribution basée uniquement sur les prix nodaux. Leur apprentissage ou leur prédiction, initialement trop complexe à cause du nombre de variables mises en jeu, est rendu possible par la réduction de dimension. L'objectif de ces algorithmes de gestion serait alors de prédire les variables latentes.

## 6. RÉFÉRENCES

- [1] L. Li, J. Lin, N. Wu, S. Xie, C. Meng, Y. Zheng, X. Wang, and Y. Zhao, "Review and outlook on the international renewable energy development," *Energy and Built Environment*, vol. 3, pp. 139–157, 4 2022.
- [2] S. Kerscher and P. Arbolea, "The key role of aggregators in the energy transition under the latest European regulatory framework," *International Journal of Electrical Power and Energy Systems*, vol. 134, p. 107361, 1 2022.
- [3] A. Navarro-Espinosa and L. F. Ochoa, "Probabilistic Impact Assessment of Low Carbon Technologies in LV Distribution Systems," *IEEE Transactions on Power Systems*, vol. 31, pp. 2192–2203, 5 2016.
- [4] J. Carpentier, "Optimal power flows," *International Journal of Electrical Power & Energy Systems*, vol. 1, pp. 3–15, 4 1979.
- [5] F. L. Alvarado, "Is system control entirely by price feasible?," *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*, 2003.
- [6] A. Papavasiliou, "Analysis of distribution locational marginal prices," *IEEE Transactions on Smart Grid*, vol. 9, pp. 4872–4882, 9 2018.
- [7] Z. Yang, H. Zhong, Q. Xia, and C. Kang, "Fundamental Review of the

OPF Problem : Challenges, Solutions, and State-of-the-Art Algorithms,” *Journal of Energy Engineering*, vol. 144, p. 04017075, 11 2017.

[8] G. Chérot, R. Le Goff Latimier, F. Sanchez, and H. Ben Ahmed, “Mi-  
-estimation of Impedance Values Within a Distribution Network Optimal  
Power Flow,” in *2023 IEEE Belgrad PowerTech, PowerTech 2023*, IEEE,  
2023.

[9] M. Farivar and S. H. Low, “Branch flow model : Relaxations and  
convexification-part II,” *IEEE Transactions on Power Systems*, vol. 28,  
no. 3, pp. 2565–2572, 2013.

[10] S. Huang, Q. Wu, J. Wang, and H. Zhao, “A Sufficient Condition on  
Convex Relaxation of AC Optimal Power Flow in Distribution Networks,”  
*IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1359–1368, 2017.

[11] T. Baroche, P. Pinson, R. Le Goff Latimier, and H. Ben Ahmed, “Exoge-  
-nous Cost Allocation in Peer-to-Peer Electricity Markets,” *IEEE Transac-  
-tions on Power Systems*, vol. 34, no. 4, pp. 2553–2564, 2019.

[12] A. S. Faria, T. Soares, T. Orlandini, C. Oliveira, T. Sousa, P. Pinson,  
and M. Matos, “P2P market coordination methodologies with distribution  
grid management,” *Sustainable Energy, Grids and Networks*, p. 101075, 5  
2023.

[13] T. Morstyn, A. Teytelboym, and M. D. McCulloch, “Bilateral contract net-  
-works for peer-to-peer energy trading,” *IEEE Transactions on Smart Grid*,  
vol. 10, pp. 2026–2035, 3 2019.

[14] A. Petrushev, M. A. Putratama, R. Rigo-Mariani, V. Debusschere, P. Rei-  
-gnier, and N. Hadjsaid, “Reinforcement learning for robust voltage control  
in distribution grids under uncertainties,” *Sustainable Energy, Grids and  
-Networks*, vol. 33, p. 100959, 3 2023.

[15] G. Chérot, R. Le Goff Latimier, B. Cajna, and H. Ben Ahmed, “Optimisa-  
-tion par apprentissage du prix d ’échange dans un marché hétérogène,” in  
*Symposium du génie électrique*, 2023.

[16] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu,  
“Feature selection : A data perspective,” *ACM Computing Surveys*, vol. 50,  
12 2017.

[17] J. Shlens, “PCA Tutorial,” *Measurement*, pp. 1–13, 2005.

[18] M. Rouaud, *Statistiques Et Analyses Multicritères*. 2017.

[19] K. P. Schneider, B. A. Mather, B. C. Pal, C. W. Ten, G. J. Shirek, H. Zhu,  
J. C. Fuller, J. L. Pereira, L. F. Ochoa, L. R. De Araujo, *et al.*, “Analytic  
Considerations and Design Basis for the IEEE Distribution Test Feeders,”  
*IEEE Transactions on Power Systems*, vol. 33, pp. 3181–3188, 5 2018.

[20] R. A. Jabr, “Radial distribution load flow using conic programming,” *IEEE  
Transactions on Power Systems*, vol. 21, pp. 1458–1459, 8 2006.

[21] L. Gan, N. Li, U. Topcu, and S. H. Low, “Exact Convex Relaxation of Opti-  
-mal Power Flow in Radial Networks,” *IEEE Transactions on Automatic  
Control*, vol. 60, pp. 72–87, 1 2014.

[22] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, and  
others, “Smart\* : An open data set and tools for enabling research in sus-  
-tainable homes,” *SustKDD, August*, vol. 111, no. 112, p. 108, 2012.

[23] A. L. Sørensen, K. B. Lindberg, I. Sartori, and I. Andresen, “Residential  
electric vehicle charging datasets from apartment buildings,” *Data in Brief*,  
vol. 36, p. 107105, 6 2021.

[24] L. Thurner, A. Scheidler, F. Schafer, J. H. Menke, J. Dollichon, F. Meier,  
S. Meinecke, and M. Braun, “Pandapower - An Open-Source Python Tool  
for Convenient Modeling, Analysis, and Optimization of Electric Power  
Systems,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–  
6521, 2018.

[25] C. Coffrin, R. Bent, K. Sundar, Y. Ng, and M. Lubin, “PowerModels.J1 :  
An Open-source framework for exploring power flow formulations,” *20th  
Power Systems Computation Conference, PSCC 2018*, 2018.

[26] C. M. Bishop, A. M. Saxe, J. L. McClelland, S. Ganguli, F. Pedregosa,  
G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*,  
“Scikit-learn : Machine Learning in {P}ython,” *Journal of Machine Lear-  
-ning Research*, vol. 12, pp. 2825–2830, 2011.

[27] J. Lever, M. Krzywinski, and N. Altman, “Principal Component Analysis,”  
*Nature Methods*, vol. 14, pp. 641–642, 6 2017.

[28] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mo-  
-hamed, and A. Lerchner, “beta-VAE : Learning Basic Visual Concepts  
with a Constrained Variational Framework,” in *International conference  
on learning representations.*, 2017.

[29] Q. Meng, D. Catchpoole, D. Skillicorn, and P. J. Kennedy, “Relational au-  
-toencoder for feature extraction,” in *Proceedings of the International Joint  
Conference on Neural Networks*, vol. 2017-May, pp. 364–371, Institute of  
Electrical and Electronics Engineers Inc., 6 2017.