Revue Internationale de la Recherche Scientifique et de l'Innovation (Revue-IRSI)
ISSN: 2960-2823
Vol. 4, No. 5 , Octobre 2026

Profils multidimensionnels de vulnérabilité et priorisation des ménages à partir de données d'enquête en situation d'urgence : application d'une classification non supervisée aux données diem en République Démocratique du Congo

Assistant Kamanga Kamono Jean (Université de Mwene-Ditu)1 , Assistant Kalala Kayembe Barthelemy (Université de Mwene-Ditu)2, Assistant Kazadi Kabwe Landry (Université de Mwene-Ditu)3, Tshomba Sumaili Patty (Université de Mwene-Ditu)4, Attachée de recherche Kankolongo Mulumba Godelive (Université de Mwene-Ditu)5, Professeur Simon Ntumba Badibanga (Unikin)6,

Résumé:

En situation d'urgence, la profusion de données sur les ménages ne suffit pas à produire une lecture cohérente de leur vulnérabilité. La difficulté réside dans le rapprochement d'informations hétérogènes, l'identification de configurations pertinentes et la transformation des résultats en une information utile à la décision. Cet article présente une chaîne analytique complète, reproductible et traçable, destinée à dégager des profils multidimensionnels de ménages puis à établir une priorisation distincte de ces profils. L'étude mobilise le cinquième passage de l'enquête ménages du dispositif DIEM conduit par l'Organisation des Nations Unies pour l'alimentation et l'agriculture en République démocratique du Congo en 2023 (2 716 ménages, 310 variables). Quatorze indicateurs ont été sélectionnées et répartis en quatre dimensions : l'accès et la consommation alimentaires, les stratégies d'adaptation, l'exposition aux chocs ainsi que les capacités structurelles. Pour comparer les ménages, nous avons utilisé la dissimilation de Gower. Plusieurs regroupements ont ensuite été obtenus à l'aide d'un partitionnement autour des méloïdés et d'une classification hiérarchique agglomérative. Le choix de la partition finale repose sur l'analyse de la séparation entre les groupes, de leur stabilité et de leur structure. La solution finalement retenue distingue deux profils de ménage. Elle obtient une silhouette moyenne de 0,356 et présente une bonne stabilité lors du ré échantillonnage, avec un indice de Rand ajusté moyen de 0,884. L'examen des dimensions étudiées met en évidence des différences entre les deux profils. Le premier regroupe, dans l'ensemble, des ménages dont les situations sont relativement moins favorables.

La priorité a ensuite été réalisée indépendamment de la constitution de ces groupes. Pour chaque ménage, un score a été calculé afin de déterminer son rang. Les analyses de sensibilité montrent que les différents classements restent globalement proches. Les coefficients de corrélation de Spearman, compris entre 0,920 et 1,000, confirment cette concordance. Toutefois, certaines variations apparaissent davantage chez les ménages occupant les premières positions du classement. Enfin, ces résultats ont été intégrés dans un prototype logiciel. Celui-ci permet de consulter les profils obtenus, les scores, les rangs ainsi que différents scénarios de capacité.

L'intérêt principal de cette démarche tient à la mise en place d'une chaîne d'analyse traçable, depuis le traitement des données jusqu'à leur utilisation dans le processus de priorisation.

Mots-clés : vulnérabilité multidimensionnelle 1; classification non supervisée 2; partitionnement autour des médoïdes 3; dissimilarité de Gower 4; priorisation 5; aide à la décision 6; DIEM 7; République démocratique du Congo 8.

Digital Object Identifier (DOI): https://doi.org/10.5281/zenodo.**23857342**

0. Introduction

En République démocratique du Congo, les conflits armés, les déplacements de populations, les épidémies et d'autres chocs récurrents affectent durablement les moyens de subsistance et la sécurité alimentaire de nombreux ménages. Pour apprécier les besoins et orienter l'assistance, les acteurs humanitaires conduisent régulièrement des évaluations auprès des populations affectées. Ces évaluations produisent des informations de nature diverse : consommation alimentaire, expérience de la faim, stratégies d'adaptation, revenus, activités productives, exposition aux chocs et caractéristiques socioéconomiques. Cette richesse améliore la connaissance des situations, mais elle en complique également l'analyse. Un ménage ne peut être valablement caractérisé par une seule de ces variables, et leur simple juxtaposition ne garantit pas davantage une lecture cohérente de sa situation.

Dans cette étude, la vulnérabilité est abordée dans une perspective volontairement circonscrite. Elle peut, dans la littérature, renvoyer au risque de détérioration future de la situation d'un ménage. Dans cette étude, les données utilisées proviennent d'une enquête réalisée à un moment précis. Elles donnent donc une image de la suite des ménages au moment de la collecte des informations. Ce type de données ne permet pas de suivre l'évolution de chaque ménage dans le temps ni de mesurer avec précision sa vulnérabilité future. Cette lecture rejoint l'idée selon laquelle la vulnérabilité résulte de l'articulation de plusieurs facteurs plutôt que d'une privation considérée isolément (Ahmed & Gassmann, 2010 ; Chambers, 1989).

Cette délimitation conduit directement au problème analytique étudié. La base de données ne contient aucune variable pouvant être considérée, à elle seule, comme une classe de référence représentant l'ensemble de la vulnérabilité multidimensionnelle du ménage.

Le score de consommation alimentaire (FCS), l'échelle de faim du ménage (HHS) et l'échelle d'expérience de l'insécurité alimentaire (FIES) décrivent chacun une dimension particulière de la situation alimentaire, sans toutefois fournir, séparément, une mesure globale. L'étude privilégie ainsi une approche non supervisée permettrait d'analyser ces différentes dimensions et de regrouper les ménages présentant des caractéristiques similaires.la constitution de ces groupes ne permet pas, à elle seule, de finir l'ordre des interventions. Chaque groupe rassemble simplement des ménages qui présentent des caractéristiques proches selon les variables étudiées. Ainsi, le numéro attribué à un groupe ne correspond ni ç un degré de gravité, ni à une phase de l'IPC, ni à un critère d'éligibilité. La définition des priorités intervient donc dans une étape distincte, après l'identification des différents profils. Elle repose sur des indicateurs sélectionnés, normalisés puis combinés selon des critères préalablement établis. Cette démarche

permet de ne pas confondre le regroupement des ménages avec leur classement en vue de prise de décision. Les profils, les scores obtenus et les rangs sont ensuite présentés de manière à faciliter la lecture et l'utilisation des résultats. La classification constitue ainsi un appui à la décision et non une décision automatique : la responsabilité finale reste entre les mains de l'utilisateur (Sprague, 1980 ; Power, 202) Plusieurs travaux permettent de situer cette recherche. Ahmed et Gassmann (2010), dans une étude consacrée à l'Afghanistan, abordent la vulnérabilité en tenant compte de l'exposition aux risques et des ressources dont disposent les ménages ; cette approche montre l'intérêt d'une lecture qui ne réduit pas la situation du ménage à une seule privation, sans toutefois traiter la question de l'analyse conjointe de variables de différentes natures ni leur utilisation dans un dispositif de priorisation. En République démocratique du Congo, Marivoet et al. (2019) développent une typologie spatiale destinée à améliorer le ciblage des interventions de sécurité alimentaire et nutritionnelle ; l'analyse étant conduite à l'échelle territoriale, elle répond à une problématique différente de celle d'une typologie de ménages. D'autres travaux empiriques montrent l'intérêt d'analyses quantitatives rigoureuses pour améliorer le ciblage de l'assistance et l'évaluation de la sécurité alimentaire lorsque les ressources sont limitées (WFP, 2026). Ces démarches répondent à des problèmes pour lesquels un résultat à estimer est défini au préalable. Lorsque cette classe de référence n'existe pas, les méthodes non supervisées offrent une autre voie d'analyse : la mesure proposée par Gower (1971) permet notamment de traiter des observations décrites par des variables de types différents.

La question qui demeure est celle de l'articulation, au niveau du ménage, entre l'analyse de données hétérogènes, l'identification de profils sans classe prédéfinie, la validation des groupes obtenus et leur utilisation dans une procédure distincte de priorisation.

Le présent article s'intéresse à cette articulation et cherche à structurer les résultats afin qu'ils puissent être utilisés concrètement comme outil d'aide à la décision. La question centrale est donc la suivante : à partir de données hétérogènes recueillies auprès des ménages en situation d'urgence en République Démocratique du Congo, comment obtenir une classification non supervisée suffisamment stable et compréhensible pour dégager différents profils multidimensionnels et, sur cette base, définir des priorités claires pouvant être intégrées dans un système d'aide à la décision ?

Cette question peut être examinée à travers cinq interrogations spécifiques. La première concerne la préparation et la présentation des données, en tenant compte de la diversité des variables, des valeurs manquantes et des éventuelles redondances pouvant influencer la formation des groupes. La deuxième s'intéresse au choix de la méthode de classification ainsi qu'au nombre de groupes permettant d'obtenir une partition suffisamment stable, distance et interprétable. La troisième cherche à déterminer les caractéristiques alimentaires, socioéconomiques adaptables et contextuelles qui différencient les profils obtenus. La quatrième porte sur l'élaboration d'un score et d'un classement indépendants de l'appartenance aux groupes, tout en vérifiant leur stabilité lorsque les pondérations sont modifiées. Enfin, la cinquième examine la manière d'intégrer les profils, les scores et les rangs dans un dispositif

permettant de retracer les résultats, tout en laissant à l'utilisateur la responsabilité de la décision finale.

Les hypothèses de recherche découlent de cette problématique. L'hypothèse générale postule que, malgré l'absence d'une classe globale de vulnérabilité définie à l'avance, les informations disponibles dans l'enquête peuvent faire apparaître des profils différenciés lorsque plusieurs dimensions sont considérées conjointement, et qu'une priorisation peut être construite indépendamment de ces profils. Quatre hypothèses spécifiques en découlent : (i) au moins une méthode de classification adaptée à la représentation retenue produit des regroupements suffisamment stables et structurés pour être interprétés ; (ii) les profils identifiés se différencient sur plusieurs dimensions et non sur la base d'un seul indicateur alimentaire ; (iii) une priorisation construite indépendamment de l'appartenance aux profils conduit à un classement globalement stable, tout en faisant apparaître une sensibilité aux choix de pondération ; (iv) les profils, les scores et les rangs peuvent être intégrés dans un dispositif assurant la traçabilité sans transformer le classement en décision automatique.

L'objectif général de cette étude est de développer une démarche d'analyse reproductible permettant de mieux identifier et caractériser les différents profils multidimensionnels des ménages. A partir des données recueillies par le dispositif DIEM, cette démarche vise également à établir des priorités entre les ménages. L'intérêt de cet article se situe à deux niveaux. Sur le plan méthodologique, il combine la représentation des données mixtes, leur classification selon plusieurs critères de validation et l'analyse de sensibilité des résultats obtenus ? Sur le pal pratique, l'approche proposée facilite l'interprétation et l'utilisation des résultats dans un contexte humanitaire. La suite de l'article présente d'abord les données et la méthodologie adaptée, puis les résultats expérimentaux. Une discussion de ces résultats et de leurs implications est ensuite proposée avant la conclusion.

0.1. Données et dispositif méthodologique,

0.2. Nature de la recherche et unité d'analyse

Cette étude adopte une approche quantitative exploratoire fondée sur des données d'enquête.

Comme aucune classe de vulnérabilité n'a été définie au préalable, l'analyse cherche à identifier les groupes qui se dégagent des données, puis à les interpréter. Plusieurs configurations et deux méthodes de classification sont comparées afin de retenir les regroupements les plus cohérents et les plus stables.

Le ménage représente l'unité d'analyse, chaque observation correspondant à un ménage enquêté. Des variables comme la province, le sexe du chef de ménage, le type de résidence et la principale source de revenu servent surtout à décrire les groupes obtenus. Les données provenant d'un seul passage d'enquête, les résultats reflètent la situation au moment de la collecte et ne permettent ni de suivre son évolution dans le temps ni d'établir des relations causales.

1.1 Source et caractéristiques des données

Les données utilisées proviennent du cinquième passage de l'enquête ménages du Data in Emergencies Monitoring System (DIEM) de la FAO en République démocratique du Congo. Le jeu de données porte l'identifiant COD_2023_DIEM-HS-R5_v01_EN_M_v01_A_ESS et le

fichier analytique diffusé par le catalogue est intitulé data_anon_df_COD_R5 (FAO, 2023). L'enquête couvre 2 716 ménages agricoles et non agricoles et comprend 310 variables.

La collecte a été réalisée par entretiens téléphoniques assistés par ordinateur entre le 23 février et le 31 mars 2023. Elle concerne onze provinces : Ituri, Kasaï, Kasaï-Central, Kasaï-Oriental, Kwango, Nord-Kivu, Nord-Ubangi, Sud-Kivu, Sud-Ubangi, Tanganyika et Tshopo. Le catalogue indique que l'échantillon a été constitué suivant un plan aléatoire stratifié et qu'un système de pondération a été établi pour les agrégations au niveau administratif 1 (FAO, 2023). Les données retenues pour cette étude portent sur plusieurs aspects, notamment la consommation et l'accès aux aliments, les stratégies d'adaptation, les chocs subis, les revenus, les activités agricoles ainsi que les caractéristiques des ménages. Cependant, toutes les variables ne sont pas intégrées directement dans la classification. Leur sélection dépend de leur pertinence, de leur nature, de leur disponibilité et des liens observés avec les autres variables.

Table 1: Principales caractéristiques des données utilisées

Élément	Caractéristique
Source	FAO – Data in Emergencies Monitoring System
Passage	Round 5
Année	2023
Unité d'analyse	Ménage
Nombre de ménages	2 716
Nombre de variables	310
Mode de collecte	Entretien téléphonique assisté par ordinateur
Couverture	11 provinces
Nature des données	Transversales
Poids d'enquête	weight_final
Fichier analytique	data_anon_df_COD_R5

1.2. Préparation et contrôle de la qualité des données

Avant la classification, les variables sont organisées dans un dictionnaire analytique précisant leur signification, leur type, leur domaine, leur période de référence et leur rôle dans l'analyse. Cette étape permet de distinguer les variables qui participent à la formation des groupes de celles qui servent uniquement à les décrire ou qui doivent être écartées. Le contrôle porte ensuite sur les identifiants, les doublons, les variables constantes, les modalités inattendues et les valeurs manquantes.

Une attention particulière est portée aux questions conditionnelles. Dans ce type d'enquête, l'absence de réponse ne signifie pas toujours qu'une donnée manque : il se peut simplement que la question ne concerne pas le ménage interrogé. Il est donc nécessaire de faire la différence entre les données réellement manquantes et celles qui résultent du déroulement du questionnaire. La préparation des données permet aussi d'éviter les répétitions d'informations. Lorsqu'un score synthétique est utilisé, les variables ayant servi à son calcul ne sont pas reprises

simultanément dans la classification. Les catégories obtenues à partir de ce score servent plutôt à interpréter les profils. Cette précaution permet d'éviter qu'une même information influence plusieurs fois le calcul de la distance entre les ménages. Cette règle évite qu'une même information soit comptée plusieurs fois dans la distance entre ménages.

Table 2: Règles de préparation et de contrôle des données

Situation	Traitement retenu
Identifiant ou variable technique	Exclusion de la classification
Variable constante	Exclusion
Score et variables ayant servi à son calcul	Une seule représentation est retenue
Score et catégorie dérivée	Catégorie réservée à l'interprétation
Valeur réellement manquante	Pas d'imputation automatique
Question non applicable	Traitée comme absence structurelle
Variable géographique	Utilisée pour caractériser les profils
Assistance déjà reçue	Non utilisée pour former les groupes
Variable insuffisamment documentée	Non retenue dans la représentation principale

1.3. Opérationnalisation des dimensions et sélection des indicateurs

L'audit du fichier conduit à retenir quatorze indicateurs pour la représentation analytique. Leur sélection tient compte de leur signification, de leur disponibilité et du risque de redondance. La première dimension rassemble les informations relatives à l'accès et à la consommation alimentaires : expérience de l'insécurité alimentaire, consommation, diversité alimentaire et faim du ménage. Les stratégies d'adaptation forment une deuxième dimension à partir de l'indice réduit de stratégies d'adaptation (rCSI) et des stratégies touchant les moyens d'existence. Ces deux ensembles rendent compte, sous des angles différents, des difficultés alimentaires observées et des réponses mobilisées par les ménages.

Les deux autres dimensions portent sur le contexte de vulnérabilité. L'exposition est résumée par quatre familles de chocs : conflit ou insécurité, chocs économiques, chocs climatiques et événements affectant directement le ménage ou sa production. Après le recodage, les indicateurs sont harmonisés afin qu'une valeur élevée traduise une situation plus défavorable. Pour l'analyse, ils sont représentés par plusieurs variables dérivées, notamment fies_probability, food_consumption, dietary_diversity, household_hunger, reduced_coping et livelihood_coping. A celles-ci s'ajoutent les indicateurs relatifs aux différentes catégories des chocs et aux capacités structurelles. Les modalités de calcul et de recodage de ces variables sont précisées dans le dictionnaire analytique.

Table 3: Indicateurs retenus et construction des quatre dimensions

Dimension	Indicateurs retenus	Construction	Sens du score
Accès et consommation alimentaires	FIES ; consommation alimentaire ; diversité alimentaire ; faim du ménage	Moyenne de 4 indicateurs orientés	1 = situation plus défavorable
Stratégies d'adaptation	rCSI ; stratégies liées aux moyens d'existence	Moyenne de 2 indicateurs orientés	1 = recours plus important aux stratégies
Exposition aux chocs	Conflit/insécurité ; économie ; climat ; ménage/production	Moyenne de 4 indicateurs	1 = exposition plus forte
Capacités structurelles	Revenu ; éducation ; eau ; assainissement	Moyenne de 4 indicateurs orientés	1 = capacités plus faibles

Certaines variables sont volontairement maintenues hors de cette représentation. La principale source de revenu et le statut productif permettent de caractériser les profils obtenus après la formation. En revanche, les identifiants, les données constantes ou répétitives ainsi que les variables liées à l'assistance reçue ne sont pas prises en compte dans le regroupement. L'aide déjà reçue mérite une attention particulière, car un soutien alimentaire ou financier peut améliorer momentanément la situation d'un ménage sans nécessairement supprimer les causes de sa vulnérabilité. Par ailleurs, le fait de recevoir une aide peut simplement montrer que ce ménage avait déjà été considéré comme nécessitant un accompagnement. Elles ne peuvent donc être interprétées automatiquement comme un signe de moindre ou de plus grande vulnérabilité, ce qui justifie de l'exclure de la formation des groupes.

1.4. Construction des dimensions et mesure de dissimilarité

Les quatorze indicateurs ne sont pas introduits directement dans la matrice de distance. Ils sont d'abord transformés sur une échelle commune comprise entre 0 et 1 et orientés dans le même sens, afin de rapprocher des mesures exprimées à l'origine selon des échelles différentes sans confondre les valeurs favorables et défavorables. Pour chaque ménage, les indicateurs appartenant à une même dimension sont regroupés en calculant leur moyenne à partir des valeurs disponibles. Le score de la dimension de ces indicateurs est renseigné. Le ménage est ensuite retenu pour la classification si les données permettent de calculer au moins trois des quatre dimensions.

Le score du ménage i pour la dimension d s'écrit :

$$S_{id} = \frac{\sum_{p \in d} \delta_{ip} V_{ip}}{\sum_{p \in d} \delta_{ip}} \quad (1)$$

où V_ip désigne la valeur orientée de l'indicateur p et δ_ip vaut 1 lorsque cet indicateur est disponible, 0 sinon. Une fois les quatre dimensions constituées, la dissimilarité entre deux ménages est calculée selon le principe de Gower (1971). Dans cette implémentation, la dissimilarité est appliquée à des scores dimensionnels déjà bornés entre 0 et 1 ; la contribution d'une dimension correspond donc à l'écart absolu entre les deux ménages sur cette dimension :

$$D_{ij} = \frac{\sum_{d=1}^{4} W_d \, \delta_{ijd} \, |S_{id} - S_{jd}|}{\sum_{d=1}^{4} W_d \, \delta_{ijd}} \quad (2)$$

où W_d est le poids de la dimension d et δ_ijd indique si la dimension est disponible pour les deux ménages comparés. Les quatre dimensions ont la même importance dans l'analyse principale. La distance calculée varie entre 0 et 1 et tient compte uniquement des dimensions qui peuvent être comparées. Il faut toutefois préciser que les données DIEM regroupent plusieurs types de variables. La classification principale ne s'effectue pas directement à partir de ces données brutes, mais à partir des quatre scores dimensionnels obtenus après recodage et regroupement des données. Une analyse de sensibilité revient ensuite aux quatorze indicateurs afin d'examiner l'effet de cette représentation.

2. Méthodes de classification

La matrice de dissimilarité sert de base commune aux deux méthodes comparées. Le partitionnement autour des médoïdes (PAM) regroupe les ménages autour de médoïdes, c'est-à-dire d'observations réellement présentes dans les données. Contrairement aux approches fondées sur la moyenne, il ne repose pas sur le calcul d'un centre moyen et peut être appliqué directement à une matrice de dissimilarité, ce qui permet de l'associer à la dissimilarité de Gower (Kaufman & Rousseeuw, 1990). La classification hiérarchique agglomérative (CAH) fournit un second point de comparaison ; avec la liaison moyenne, elle rapproche progressivement les groupes en fonction de la dissimilarité moyenne entre leurs membres (Murtagh & Contreras, 2012). Les deux méthodes utilisent ainsi la même représentation des ménages, ce qui permet de comparer leurs partitions sans changer la base de calcul. Pour chacune d'elles, les solutions comprises entre k = 2 et k = 6 sont examinées : le nombre de profils n'est pas fixé à l'avance, mais choisi après comparaison des indicateurs de validation.

2.1. Validation et caractérisation des profils

Le choix d'une partition ne repose pas sur la seule valeur de la silhouette. Trois aspects sont examinés conjointement : la séparation des groupes, leur stabilité lorsque l'échantillon varie et la présence éventuelle de groupes très petits. La silhouette moyenne apprécie la cohésion d'un ménage avec son groupe et sa séparation par rapport aux groupes voisins (Rousseeuw, 1987) ; elle est définie par la relation suivante :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

où a(i) représente la dissimilarité moyenne entre l'observation i et les membres de son groupe, tandis que b(i) correspond à la plus faible dissimilarité moyenne entre cette observation et un autre groupe. Une valeur proche de 1 indique que l'observation est bien classée dans son groupe, tandis qu'une valeur négative révèle une affectation moins satisfaisante. L'indice de Dunn permet également d'évaluer la qualité des groupes en comparant leur séparation à leur dispersion interne (Dunn, 1074). Comme cet indice peut être influencé par les valeurs extrêmes, il est préférable de l'utiliser en complément d'autres critères plutôt que de le considérer seul. La stabilité des résultats est vérifiée à partir de plusieurs sous-échantillons. Pour cette étude, trente sous-échantillons sont constitués sans remise, chacun regroupant 80% des ménages. Les groupes obtenus sont ensuite comparés à la partition de référence à l'aide de l'indice de Rand

ajusté (ARI), calculé sur les observations communes. Cet indice permet de prendre en compte les concordances pouvant apparaître par hasard (Hubert & Arabie, 1985 ; Hennig, 2007). Une condition supplémentaire concerne la taille des groupes : lorsque le groupe le moins représenté contient moins de 5% des ménages, la solution correspondante est écartée. Le seuil de 5% a été retenu uniquement pour les besoins de cette étude ; il ne constitue donc pas une règle générale. Son utilisation permet surtout d'éviter de retenir une partition dont les résultats seraient fortement influencés par un groupe très peu représenté.

Table 4: Critères utilisés pour la sélection de la partition

Critère	Mesure	Contribution au score de sélection	Rôle
Séparation	Silhouette moyenne	40 % du score de sélection	Évaluer cohésion et séparation
Compacité / séparation	Indice de Dunn	20 % du score de sélection	Compléter l'examen de la structure
Stabilité	ARI moyen sur 30 rééchantillonnages	40 % du score de sélection	Vérifier la persistance de la partition
Taille des groupes	Part du plus petit groupe	Au moins 5 % pour l'admissibilité	Écarter les solutions dominées par un microgroupe
Parcimonie	Nombre de groupes	Plus petit k en cas d'égalité	Éviter une complexité sans gain mesurable

Cette pondération (40 %–20 %–40 %) appartient à la règle de sélection programmée dans cette recherche ; elle n'est pas présentée comme un standard méthodologique universel. L'interprétation des groupes intervient seulement après cette sélection, afin que le sens substantiel des profils ne serve pas à imposer a priori la partition attendue. Après la sélection, les groupes sont décrits à partir des quatre scores dimensionnels puis des quatorze indicateurs qui les composent ; les variables contextuelles complètent ensuite l'interprétation. Les groupes restent désignés par des numéros tant que leur caractérisation n'est pas achevée, ces numéros étant de simples identifiants sans signification de gravité.

2.2. Construction du score et du rang de priorité

Le profil décrit une configuration de caractéristiques ; il ne fournit pas, à lui seul, un ordre de priorité. Le classement est donc calculé séparément à partir des quatre scores dimensionnels, déjà orientés de sorte qu'une valeur élevée traduise une situation plus défavorable. Pour chaque ménage analysable, le score de priorisation correspond à la moyenne des dimensions disponibles :

$$P_i = \frac{\sum_{d=1}^{4} \delta_{id} S_{id}}{\sum_{d=1}^{4} \delta_{id}} \quad (4)$$

où P_i est le score de priorité du ménage i, S_id son score sur la dimension d et δ_id indique si cette dimension est calculable. Comme l'analyse nécessite au moins trois dimensions disponibles, le score est calculé à partir des informations effectivement renseignées. Lorsque les quatre dimensions sont disponibles, il correspond simplement à leur moyenne. Sa valeur varie de 0 à 1 : plus elle est élevée, plus la situation du ménage apparaît défavorable au regard des indicateurs considérés. Ce score ne doit toutefois pas être interprété comme une probabilité d'insécurité alimentaire future ni comme une décision automatique d'assistance. Dans le scénario de référence, les quatre dimensions ont le même poids, ce qui permet d'examiner leur contribution de manière équilibrée, conformément aux principes de construction des indicateurs composites (OECD & European Commission, Joint Research Centre, 2008). Les ménages sont ensuite classés du score le plus élevé au plus faible. En cas d'égalité, l'identifiant pseudonymisé permet uniquement de maintenir un classement reproductible, sans influencer l'interprétation des résultats. L'analyse principale ne fixe pas de capacité de prise en charge : les seuils de 10 %, 20 % et 30 % sont utilisés comme scénarios descriptifs de capacité de vérification et non comme seuils officiels de vulnérabilité.

2.3. Analyse de sensibilité

Le classement dépend en partie de la place accordée aux quatre dimensions. Pour vérifier que les conclusions ne reposent pas sur un seul choix de pondération, le scénario de référence est comparé à quatre variantes. Les variantes proposées modifient tour à tour l'importance accordée à l'accès-consommation, à l'adaptation et aux capacités structurelles. Elles ne créent donc pas de nouvelles catégories de ménages, mais permettent surtout d'examiner les changements de classement lorsque la pondération varie. Pour chaque variante, la corrélation de Spearman permet de comparer le classement obtenu à celui de référence. Cette analyse est complétée par une comparaison des ménages figurant dans les premiers 10%, 20% et 30% du classement. Cette seconde lecture est importante, car deux classements peuvent être fortement corrélés dans l'ensemble tout en différant près d'un seuil de capacité.

Table 5: Scénarios examinés dans l'analyse de sensibilité

Scénario	Modification examinée
Dimensions à poids égaux	Scénario de référence : contribution identique des quatre dimensions
Accès-consommation renforcé	Importance relative plus élevée de la dimension accès consommation
Adaptation renforcée	Importance relative plus élevée des stratégies d'adaptation
Sans capacités structurelles	Retrait de la dimension des capacités structurelles
Accès et adaptation	Pondération centrée conjointement sur l'accès consommation et l'adaptation

La sensibilité de la partition est aussi vérifiée de deux manières. Premièrement, le partitionnement autour des médoïdes est repris sur les quatorze indicateurs actifs, puis comparé

à celui obtenu à partir des quatre dimensions. Deuxièment, les quatre scores dimensionnels sont soumis à la méthode des l-moyennes afin d'effectuer une vérification supplémentaire. Pour chacune de ces analyses, l'indice de Rand ajusté permet d'évaluer la concordance avec la partition principale. Ces vérifications ne modifient pas la solution retenue, mais permettent de voir dans quelle mesure les regroupements obtenus dépendent de la représentation des données et de la méthode des données et de la méthode de partitionnement.

2.4. Reproductibilité, traçabilité et protection des données

L'analyse est réalisée sous Python, depuis la préparation des données jusqu'à la production des résultats. Afin de pouvoir reproduire les traitements, les principaux paramètres utilisés sont conservés, notamment la graine aléatoire, les versions des bibliothèques et l'empreinte du fichier source. Les transformations appliquées aux variables sont également documentées : les recodages, les regroupements et les pondérations peuvent ainsi être retracés, tandis que le fichier d'origine reste inchangé. La protection des données repose sur l'utilisation d'identifiants pseudonymisés et sur la présentation de résultats agrégés. Les observations ajoutées lors de la vérification humaine sont conservées séparément des résultats analytiques afin de ne pas modifier les profils, les scores ou les rangs calculés. Cette organisation permet de distinguer clairement les résultats produits par l'analyse des décisions prises ensuite par l'utilisateur.

2.5. Limites du protocole

Plusieurs limites doivent être prises en compte dans l'interprétation de cette démarche. La première concerne l'absence d'une classe de référence permettant d'évaluer extérieurement la justesse des profils : la validation reste donc principalement interne, une bonne silhouette ou une forte stabilité indiquant qu'une structure est cohérente dans les données sans démontrer l'existence d'une catégorie vraie de vulnérabilité dans la population. La deuxième tient aux informations disponibles dans l'enquête : les quatre dimensions utilisées représentent plusieurs aspects de la situation des ménages, mais elles n'épuisent pas toutes les composantes possibles de la vulnérabilité, de sorte que la qualité de la typologie dépend du périmètre et de la qualité des variables observées.

La troisième limite concerne les choix analytiques : l'équilibre accordé aux dimensions dans la dissimilarité et les poids utilisés dans le score de priorité restent des choix méthodologiques, dont la transparence et les analyses de sensibilité permettent d'apprécier les effets sans les transformer en pondérations universelles.

Nous n'avons pas intégré les poids d'enquête à la classification. Les profils décrivent donc la structure des ménages telle qu'observée dans l'échantillon ; les poids ne servent que pour les statistiques descriptives, lorsqu'il faut restituer des proportions représentatives du plan d'enquête. Les données étant transversales, aucune lecture prédictive ou causale n'est possible : le protocole repère des configurations existant au moment de l'enquête et les hiérarchise selon des règles explicites. Il ne dit rien d'une détérioration future d'un ménage, ni du rôle causal d'une caractéristique particulière.

3. Résultats

3.1. Base analytique effectivement obtenue

L'exécution finale a porté sur le fichier officiel DIEM Round 5 comprenant 2 716 ménages et 310 variables. Le contrôle préalable n'a relevé ni doublon d'identifiant, ni duplication complète de ligne. Les quatre dimensions prévues ont pu être calculées pour l'ensemble des ménages ; aucun cas n'a donc été écarté pour insuffisance d'information. La totalité des ménages a ainsi été conservée dans l'analyse. Cette complétude ne signifie pas que chaque variable du fichier soit entièrement renseignée : les valeurs manquantes et les variables peu couvertes restent documentés dans l'audit, conformément aux règles de préparation définies à la section 2. Le poids d'enquête a, pour sa part, été conservé pour les restitutions descriptives qui le nécessitent ; il n'intervient pas dans la formation des groupes.

Table 6: Résultats du contrôle de la base analytique

Élément contrôlé	Résultat
Ménages dans le fichier source	2 716
Variables dans le fichier source	310
Ménages retenus pour l'analyse	2 716
Ménages exclus pour données insuffisantes	0
Doublons d'identifiant détectés	0
Lignes entièrement dupliquées	0

3.2. Indicateurs analytiques et quatre dimensions

Quatorze variables ont été conservées au final. Recodées et orientées dans le même sens une valeur haute signifiant une situation plus défavorable, elles se répartissent en quatre dimensions : accès et consommation alimentaires, stratégies d'adaptation, exposition aux chocs, capacités structurelles. Leur agrégation fournit quatre scores compris entre 0 et 1. Ces scores, combinés à poids égaux, constituent la représentation retenue pour calculer la dissimilarité entre ménages. Les noms techniques des variables et les règles de recodage restent consignés dans le dictionnaire analytique, et les variables contextuelles servent uniquement à décrire la répartition des profils.

3.3. Choix du nombre de profils

Après la préparation de la base analytique, plusieurs solutions ont été testées, de deux à six groupes. Leur comparaison repose sur les critères définis à la section 2 : séparation, compacité, stabilité et taille des groupes. Une solution comportant un groupe représentant moins de 5 % des ménages n'a pas été retenue. À l'issue de cette comparaison, la solution à deux groupes a été retenue. Avec le partitionnement autour des médoïdes, elle présente une silhouette moyenne de 0,356, un indice de Dunn de 0,0126 et une stabilité exprimée par un indice de Rand ajusté moyen de 0,884, avec un écart-type de 0,094. Le plus petit groupe rassemble 1 156 ménages, soit 42,6 % de l'effectif analysé.

Table 7: Caractéristiques de la solution retenue

Nombre de groupes	Silhouette	Indice de Dunn	Stabilité (ARI)	Plus petit groupe
k = 2	0,356	0,0126	0,884 ± 0,094	1 156 (42,6 %)

Ces valeurs montrent que la partition retenue est stable au rééchantillonnage, même si la séparation entre les groupes reste modérée. La lecture des profils doit donc porter sur des configurations relatives de ménages plutôt que sur des catégories supposées naturelles.

3.4. Comparaison des méthodes de classification

La présence de deux groupes ne permet pas, à elle seule, de choisir la méthode la plus adaptée. Le partitionnement autour des médoïdes et la classification hiérarchique à liaison moyenne ont donc été comparés sur la même représentation des données. La classification hiérarchique présente une silhouette légèrement meilleure, avec 0,382 contre 0,356 pour le partitionnement autour des médoïdes. En revanche, sa stabilité est nettement plus faible, avec un indice de Rand ajusté de 0,223, et elle produit un groupe composé d'un seul ménage. Le partitionnement autour des médoïdes, au contraire, reste beaucoup plus stable, avec un indice de 0,884, et forme deux groupes de taille plus équilibrée.

Table 8: Comparaison des méthodes de classification pour k = 2

Méthode	Silhouette	Stabilité (ARI)	Plus petit groupe	Décision
Partitionnement autour des médoïdes (PAM)	0,356	0,884	1 156 (42,6 %)	Retenue
Classification hiérarchique (liaison moyenne)	0,382	0,223	1 (0,04 %)	Non retenue

Au regard de ces résultats, le partitionnement autour des médoïdes apparaît comme la solution la plus adaptée. Ce choix ne repose pas sur la silhouette seule, mais sur sa meilleure stabilité et sur la formation de groupes de taille plus cohérente. La comparaison confirme ainsi qu'une méthode ne doit pas être retenue à partir d'un seul indicateur.

3.5. Caractérisation des profils retenus

La partition retenue distingue deux profils, désignés par leur numéro afin d'éviter de leur attribuer d'emblée une qualification normative. Le profil 1 regroupe 1 560 ménages, soit 57,4 % de l'effectif, tandis que le profil 2 en regroupe 1 156, soit 42,6 %. La comparaison des scores moyens montre que le profil 1 présente des valeurs plus élevées dans les quatre dimensions. L'écart est particulièrement marqué pour l'accès et la consommation alimentaires ainsi que pour les stratégies d'adaptation. Le profil 2 présente des niveaux relativement plus faibles, sans que cela signifie une absence de vulnérabilité.

Table 9: Scores moyens des quatre dimensions selon le profil

Profil	Effectif	Accès consommation	Adaptation	Chocs	Capacités
Profil 1	1 560 (57,4 %)	0,664	0,625	0,224	0,448
Profil 2	1 156 (42,6 %)	0,349	0,248	0,094	0,309

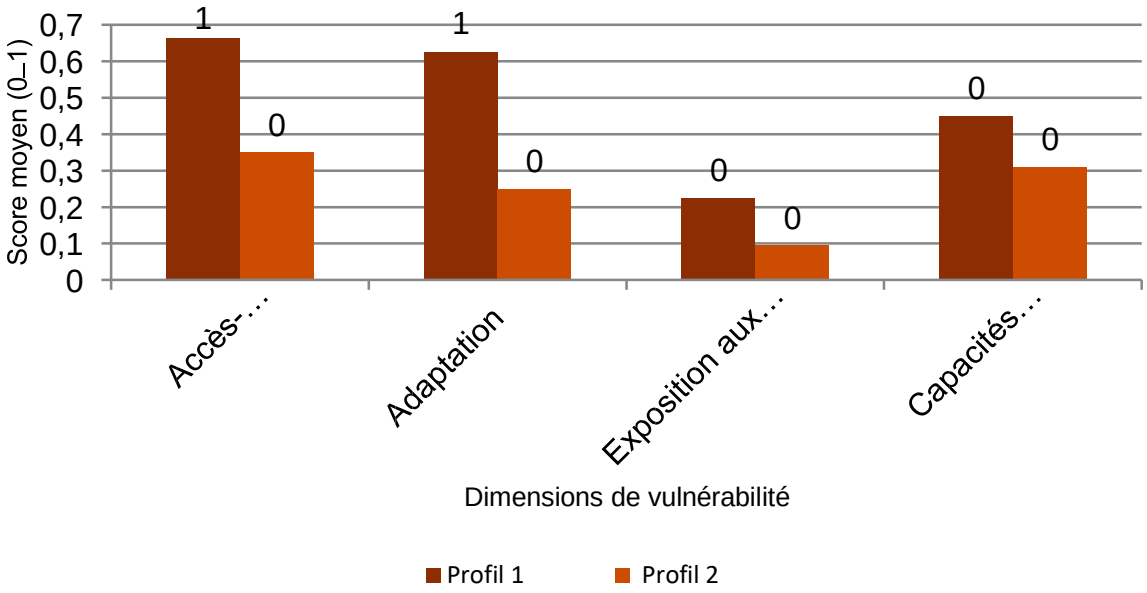


Figure 1: Scores moyens des quatre dimensions selon le profil de vulnérabilité

L'examen des variables analytiques montre que les écarts les plus nets entre les deux profils concernent l'échelle d'expérience de l'insécurité alimentaire, la faim du ménage et les stratégies liées aux moyens d'existence. Les écarts ne se limitent pas à un indicateur particulier : ils traversent les quatre dimensions, ce qui justifie précisément une lecture multidimensionnelle de la vulnérabilité. Ces profils proviennent des regroupements calculés sur les données ; rein ne les assimile à des phases IPC, à des catégories officielles de vulnérabilité ni à des critères d'éligibilité. La province n'est intervenue qu'n variable de contexte et n'a pas participé à la construction des groupes.

3.6. Priorisation des ménages

Après l'identification des profils, la priorisation est réalisée séparément. Pour chaque ménage, le score correspond à la moyenne des quatre dimensions orientées. Il varie de 0,0239 à 0,7928, avec une médiane de 0,3994. Les 2 716 ménages sont ensuite classés du score le plus élevé au plus faible, ce qui permet d'attribuer un rang allant de 1 à 2 716. En cas d'égalité, l'identifiant pseudonymisé sert uniquement à départager les ménages de manière reproductible, sans modifier leur score. L'exécution principale n'impose aucun statut prioritaire : les scénarios de 10 %, 20 % et 30 % servent uniquement à illustrer différentes capacités de vérification, et le

seuil qui en résulte dépend de la distribution des scores dans cette vague, sans constituer un seuil universel de vulnérabilité.

Table 10: Principaux résultats de la priorisation

Indicateur	Résultat
Score médian	0,3994
Étendue du score	0,0239 – 0,7928
Rang maximal	2 716
Capacité fixée dans l'exécution principale	Aucune
Scénario descriptif à 10 %	272 ménages
Scénario descriptif à 20 %	544 ménages
Scénario descriptif à 30 %	815 ménages

3.7. Sensibilité du classement

Afin de vérifier dans quelle mesure les choix de pondération influencent l'ordre de priorité des ménages, le classement obtenu avec les poids de référence a été comparé à ceux issus de quatre scénarios alternatifs. Dans l'ensemble, les corrélations de Spearman restent élevées, avec des valeurs comprises entre 0,920 et 1,000, ce qui traduit une forte concordance entre les classements. Cette stabilité d'ensemble doit toutefois être nuancée : elle n'implique pas que les mêmes ménages se retrouvent systématiquement parmi les premières positions.

Table 11: Sensibilité du classement aux scénarios de pondération

Scénario	ρ de Spearman	Top 10 %	Top 20 %	Top 30 %
Dimensions à poids égaux	1,000	100 %	100 %	100 %
Accès-consommation renforcé	0,966	79,0 %	86,4 %	89,0 %
Adaptation renforcée	0,964	77,6 %	84,2 %	86,4 %
Sans capacités structurelles	0,955	72,8 %	81,6 %	84,9 %
Accès et adaptation	0,920	64,0 %	73,7 %	80,6 %

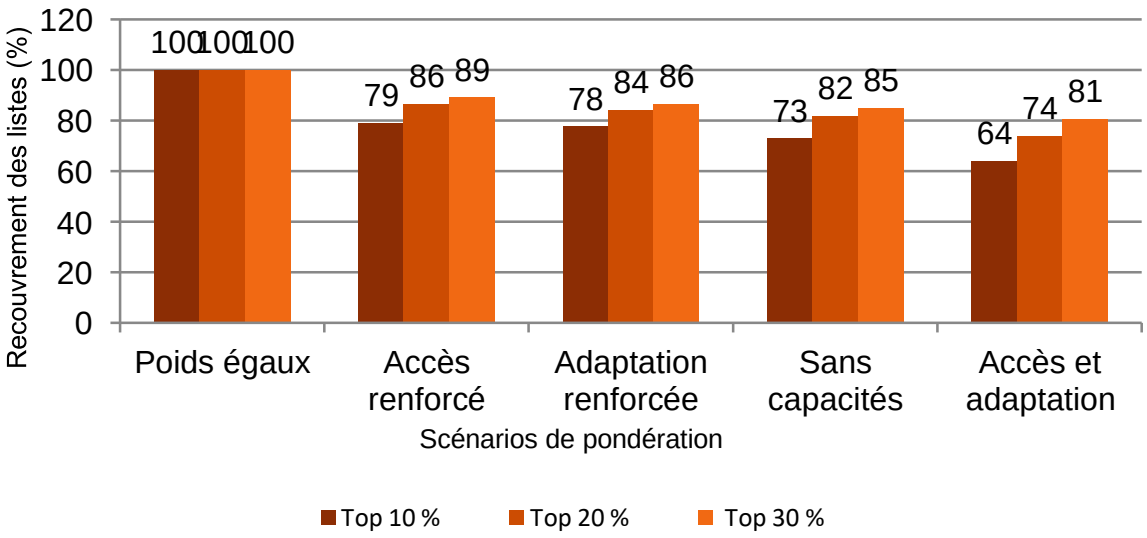


Figure 2: Recouvrement des listes prioritaires selon les scénarios de pondération

Le recouvrement des ménages prioritaires ne varie davantage que le classement global. Selon le scénario de pondération, il passe de 64,0 % à 100 % pour le top 10 %, de 73,7 % à 100 % pour le top 20 % et de 80,6 % à 100 % pour le top 30 %. Les pondérations influencent donc surtout les ménages placés en tête du classement. Deux analyses complémentaires montrent par ailleurs que la partition principale reste partiellement retrouvée avec d'autres méthodes : l'indice de Rand ajusté atteint 0,602 avec un partitionnement autour des médoïdes appliqué aux quatorze variables analytiques et 0,703 avec une méthode de partitionnement en k moyennes appliquée aux quatre dimensions. Cela traduit une certaine cohérence de la structure, sans obtenir exactement les mêmes groupes.

3.8. Vérification des hypothèses

Les hypothèses sont ici confrontées aux résultats obtenus, sans être traitées comme des tests statistiques au sens inférentiel. La première hypothèse reçoit un appui : la solution à deux profils présente une stabilité moyenne élevée et des groupes de taille substantielle, même si la silhouette de 0,356 invite à la prudence sur le degré de séparation.

L'hypothèse deux se trouve confirmée : les deux profils se démarquent sur les quatre dimensions à la fois, et non sur un seul indicateur alimentaire. L'hypothèse trois est confirmée elle aussi, avec une réserve. Le classement est bien construit indépendamment du numéro de profil et les rangs restent très corrélés d'une pondération à l'autre ; mais le haut du classement bouge davantage. Quand le score sert à répartir une capacité limitée de vérification, ce sont pourtant ces premières positions qui comptent : le choix des pondérations doit donc rester explicite. La quatrième hypothèse, relative à l'intégration des profils, des scores et des rangs dans un dispositif d'aide à la décision, est examinée dans la section suivante.

3.9. Intégration dans le dispositif d'aide à la décision

Filtrer, consulter un ménage ou appliquer un scénario de capacité dans l'interface ne relance pas la classification et laisse les résultats validés inchangés. Trois objets sont à distinguer : le profil,

qui situe le ménage dans une configuration multidimensionnelle ; le score, qui résume les quatre dimensions selon la pondération retenue ; le rang, qui donne sa position dans le classement général. Il s'agit d'éléments d'aide à la décision, non de décisions. Le statut affiché après un scénario de capacité signale simplement un ménage à examiner en priorité ; il ne vaut pas sélection automatique comme bénéficiaire. Les contrôles fonctionnels confirment que le prototype restitue la solution validée et que la consultation des résultats reste disjointe des calculs.

4. Principaux enseignements

Le profil 1 regroupe 57,4 % des ménages, avec des niveaux moyens plus défavorables, surtout pour l'accès et la consommation alimentaires et pour les stratégies d'adaptation ; le profil 2 se place en relative meilleure situation. La partition est stable au rééchantillonnage (indice de Rand ajusté moyen : 0,884), mais cette stabilité ne signifie pas une séparation nette des groupes : la silhouette moyenne (0,356) reste modérée. Les deux profils décrivent donc des configurations relatives, propres à cette vague d'enquête. Ils ne constituent ni des catégories fixes, ni des types naturels de ménages — précision indispensable si l'on veut éviter de leur prêter une valeur normative.

Les résultats montrent qu'il est possible, en l'absence d'une classe globale de vulnérabilité prédéfinie, de dégager des profils de ménages interprétables et suffisamment stables à partir de données d'enquête hétérogènes. La partition retenue comprend deux profils qui se différencient sur l'ensemble des quatre dimensions étudiées, ce qui conforte l'intérêt d'une lecture multidimensionnelle : les écarts ne se réduisent pas à un seul indicateur alimentaire, mais portent simultanément sur l'accès et la consommation alimentaires, les stratégies d'adaptation, l'exposition aux chocs et les capacités structurelles.

4.1. Mise en perspective avec les travaux antérieurs

Les résultats obtenus rejoignent plusieurs enseignements de la littérature. La lecture multidimensionnelle de la vulnérabilité, qui articule exposition aux risques et ressources des ménages, est cohérente avec le cadre proposé par Chambers (1989) et appliqué par Ahmed et Gassmann (2010). La démarche de typologie de ménages complète les approches territoriales, telles que celle développée pour la République démocratique du Congo par Marivoet et al. (2019), en déplaçant l'analyse de l'échelle du territoire à celle du ménage. Par ailleurs, l'intérêt d'une représentation adaptée aux données mixtes confirme l'apport de la mesure de Gower (1971), déjà mobilisée pour traiter conjointement des variables de différentes natures.

La contribution se distingue toutefois des démarches qui visent à reproduire des catégories institutionnelles de sécurité alimentaire à partir de règles et de seuils préétablis (WFP, 2025b ; IPC Global Partners, 2021). Ces démarches répondent à des problèmes pour lesquels la catégorie à retrouver est définie au préalable, alors que la présente étude se situe dans une situation où aucune classe de référence globale n'existe. La démarche adoptée ne cherche donc pas à reproduire une catégorie institutionnelle, comme le fait une approche harmonisée telle que le CARI (WFP, 2025b), mais à faire apparaître des configurations présentes dans les données avant de les interpréter. Enfin, la séparation explicite entre classification, priorisation et décision

s'inscrit dans la logique générale des systèmes d'aide à la décision (Sprague, 1980 ; Power, 2002), qui mettent des traitements analytiques au service de l'utilisateur sans se substituer à lui.

4.2. Portée méthodologique et pratique

Sur le plan méthodologique, l'apport principal réside dans l'articulation d'une chaîne complète, reproductible et traçable. La sélection raisonnée des variables, destinée à limiter les redondances, la représentation en quatre dimensions, la validation conjointe de la partition par la séparation, la stabilité et la taille des groupes, puis la construction d'une priorisation soumise à une analyse de sensibilité, forment un enchaînement dont chaque étape est documentée. Nos choix analytiques équilibre des dimensions dans la dissimilarité, poids retenus pour le score restent des conventions méthodologiques ; leur explicitation permet d'en mesurer les effets sans les présenter comme des règles universelles. Les données transversales forment une seconde limite : pas d'interprétation causale, pas de suivi des mêmes ménages dans le temps. Enfin, le prototype n'a pas été testé dans un cadre institutionnel réel : ce qui est établi ici relève de la faisabilité technique, non de l'efficacité opérationnelle. Ces limites bornent les conclusions sans les vider de leur intérêt ; elles appellent surtout de la prudence dès que les résultats servent à orienter une vérification sur le terrain.

4.3. Limites

Plusieurs limites doivent être soulignées. La validation des profils reste principalement interne : en l'absence d'une classe de référence, une bonne silhouette ou une forte stabilité attestent la cohérence de la structure dans les données, sans démontrer son adéquation à une catégorie externe. La qualité de la typologie dépend par ailleurs du périmètre et de la qualité des variables observées : les quatre dimensions retenues représentent plusieurs aspects de la situation des ménages, mais elles n'épuisent pas toutes les composantes possibles de la vulnérabilité.

4.4. Conditions d'utilisation opérationnelles du dispositif par les acteurs humanitaires

L'utilisation opérationnelle du dispositif proposé nécessite plusieurs précautions. Les profils et les scores doivent être considérés comme des instruments d'aide à la décision et non comme des critères automatiques d'attribution de l'assistance. Leur exploitation devait s'appuyer sur une vérification humaine des situations prioritaires, une actualisation régulière des données et une adaptation des indicateurs aux réalités locales. Une validation préalable auprès des organisations humanitaires permettrait également d'évaluer la pertinence des classements, leur interprétabilité et leur adéquation aux contraintes opérationnelles. La transparence des pondérations, la traçabilité des décisions et la protection des données des ménages constituent enfin des conditions essentielles ç une utilisation responsable du prototype.

Conclusion

L'étude s'est appuyée sur les données du cinquième passage DIEM réalisé en République démocratique du Congo en 2023. Les 2 716 ménages ont été analysés à partir de quatorze indicateurs regroupés en quatre dimensions, ce qui a permis de prendre en compte plusieurs aspects complémentaires de la vulnérabilité tout en limitant les redondances.

La comparaison des méthodes de classification a conduit à retenir une partition à deux profils obtenue par partitionnement autour des médoïdes. Cette solution présente une séparation modérée, avec une silhouette moyenne de 0,356, mais résiste bien au rééchantillonnage (indice de Rand ajusté moyen : 0,884). Les deux profils se distinguent sur toutes les dimensions, le premier cumulant les situations les plus défavorables. La priorisation, construite séparément de la classification, a doté chaque ménage d'un score et d'un rang ; elle reste globalement stable quand les pondérations changent, ses premières positions étant plus sensibles. La question posée trouve ainsi sa réponse : sans classe globale prédéfinie, on peut identifier des profils interprétables et assez stables, puis hiérarchiser à partir de critères explicites. Les résultats alimentent un prototype qui donne accès aux profils, scores et rangs, en gardant séparés le calcul et la décision. L'apport du travail tient moins à une méthode nouvelle qu'à l'enchaînement cohérent, de la préparation des données à leur restitution. Restent à venir : plusieurs passages d'enquête pour suivre l'évolution des profils, l'ouverture à d'autres sources de données et la validation du prototype par des utilisateurs institutionnels. Au bout du compte, la valeur d'une telle démarche dépend autant de la préparation et de la validation des données que des méthodes de regroupement, la décision finale relevant de l'utilisateur.

REFERENCES

1. Ahmed, M., & Gassmann, F. (2010). Measuring multidimensional vulnerability in Afghanistan (MGSoG Working Paper No. 2010/004). Maastricht Graduate School of Governance. https://doi.org/10.2139/ssrn.1829920
2. Ballard, T., Coates, J., Swindale, A., & Deitchler, M. (2011). Household Hunger Scale: Indicator definition and measurement guide. FANTA-2, FHI 360.
3. Chambers, R. (1989). Editorial introduction: Vulnerability, coping and policy. IDS Bulletin, 20(2), 1–7. https://doi.org/10.1111/j.1759-5436.1989.mp20002001.x
4. Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics, 4(1), 95–104. https://doi.org/10.1080/01969727408546059
5. Food and Agriculture Organization of the United Nations. (1998). Guidelines for national food insecurity and vulnerability information and mapping systems (FIVIMS): Background and principles. FAO.
6. Food and Agriculture Organization of the United Nations. (2008). An introduction to the basic concepts of food security. FAO.
7. Food and Agriculture Organization of the United Nations. (2017). The Food Insecurity Experience Scale: Measuring food insecurity through people's experiences. FAO.
8. Food and Agriculture Organization of the United Nations. (2023). Data in Emergencies (DIEM) Monitoring System – Household Survey – Round 5, Democratic Republic of the Congo, 2023 [Data set]. FAO Microdata Catalogue. https://microdata.fao.org/index.php/catalog/2726
9. Gower, J. C. (1971). A general coefficient of similarity and some of its properties. Biometrics, 27(4), 857–871. https://doi.org/10.2307/2528823
10. Hennig, C. (2007). Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis, 52(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025
11. Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of Classification, 2, 193–218. https://doi.org/10.1007/BF01908075
12. IPC Global Partners. (2021). Integrated Food Security Phase Classification technical manual version 3.1: Evidence and standards for better food security and nutrition decisions. IPC.
13. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
14. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computing Surveys, 31(3), 264–323. https://doi.org/10.1145/331499.331504
15. Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. John Wiley & Sons. https://doi.org/10.1002/9780470316801
16. Kennedy, G., Ballard, T., & Dop, M. C. (2011). Guidelines for measuring household and individual dietary diversity. FAO.

17. Marivoet, W., Ulimwengu, J. M., & Sedano, F. (2019). Spatial typology for targeted food and nutrition security interventions. World Development, 120, 62–75. https://doi.org/10.1016/j.worlddev.2019.04.003
18. Maxwell, D., & Caldwell, R. (2008). The Coping Strategies Index: Field methods manual (2nd ed.). CARE.
19. Maxwell, D., Vaitla, B., & Coates, J. (2014). How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. Food Policy, 47, 107–116. https://doi.org/10.1016/j.foodpol.2014.04.003
20. Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. WIREs Data Mining and Knowledge Discovery, 2(1), 86–97. https://doi.org/10.1002/widm.53
21. Negash, S. (2004). Business intelligence. Communications of the Association for Information Systems, 13, 177–195. https://doi.org/10.17705/1CAIS.01315
22. Organisation for Economic Co-operation and Development, & European Commission, Joint Research Centre. (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en
23. Power, D. J. (2002). Decision support systems: Concepts and resources for managers. Quorum Books.
24. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
25. Sprague, R. H., Jr. (1980). A framework for the development of decision support systems. MIS Quarterly, 4(4), 1–26. https://doi.org/10.2307/248957
26. Turner, B. L., II, Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., & Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. Proceedings of the National Academy of Sciences, 100(14), 8074–8079. https://doi.org/10.1073/pnas.1231335100
27. World Food Programme. (2008). Food consumption analysis: Calculation and use of the food consumption score in food security analysis (Version 1). WFP.
28. World Food Programme. (2025a). Livelihood Coping Strategies – Food Security. VAM Resource Centre. https://vamresources.manuals.wfp.org/docs/livelihood-copingstrategies-food-security
29. World Food Programme. (2025b). The Consolidated Approach for Reporting Indicators of Food Security (CARI). VAM Resource Centre. https://vamresources.manuals.wfp.org/docs/the-consolidated-approach-for-reportingindicators-of-food-security-cari
30. World Food Programme. (2026). Targeting and prioritization. VAM Resource Centre. https://vamresources.manuals.wfp.org/docs/targeting-and-prioritization