



Dynamiques spatio-temporelles, typologie régionale et modélisation prédictive des effectifs étudiants à Madagascar

¹ZAFINIRINA Tsanganavelo, ²RAZAFIARISERA Ralay Tiana, ³RAKOTOSON Tolontsoa,
⁴TOVONIRINA Mamiharizo Jackie, ³Hery Zo ANDRIAMANOHISOA

1- ECOLE DOCTORALE EN SCIENCES ET TECHNIQUES DE L'INGENIERIE ET DE L'INNOVATION (EDSTII) UNIVERSITE D'ANTANANARIVO

2- ESPA - UNIVERSITE D'ANTANANARIVO

3- UNIVERSITE D'ANTANANARIVO

RESUME : Cette étude analyse les dynamiques spatio-temporelles des effectifs étudiants de l'enseignement supérieur à Madagascar entre 2017 et 2022, en combinant analyse territoriale, apprentissage automatique et méthodes non supervisées. En 2022, 178 665 étudiants sont recensés, avec une forte concentration à Analamanga (77 379 étudiants ; 43,3 %) et Haute Matsiatra (31 514 ; 17,6 %). À elles seules, ces deux régions regroupent 60,9 % des effectifs nationaux. L'analyse des trajectoires met en évidence une forte stabilité des principaux pôles universitaires, malgré certaines mobilités régionales, notamment Androy qui gagne sept rangs entre 2017 et 2022. La détection d'anomalies par Isolation Forest identifie Androy comme la trajectoire la plus atypique, avec un score d'environ 0,71, devant Analamanga (0,63). Le clustering non supervisé retient deux groupes régionaux, avec un score de silhouette proche de 0,65, traduisant une nette polarisation territoriale. Enfin, le modèle XGBoost présente une performance prédictive globale satisfaisante avec $R^2 = 0,833$, MAE = 1 684 et RMSE = 6 725, bien que les erreurs augmentent pour les régions à très forts effectifs.

Mots-clés : dynamique spatio-temporelle ; effectifs étudiants ; XGBoost ; clustering régional ; apprentissage automatique.

ABSTRACT: This study investigates the spatio-temporal dynamics of higher education student enrollment in Madagascar between 2017 and 2022 by combining territorial analysis, machine learning, and unsupervised learning methods. In 2022, a total of 178,665 students were recorded, with a strong concentration in the regions of Analamanga (77,379 students; 43.3%) and Haute Matsiatra (31,514; 17.6%). These two regions alone accounted for 60.9% of the national student population. The trajectory analysis highlights a high level of stability among the main university hubs, despite some regional shifts, particularly in Androy, which moved up seven ranks between 2017 and 2022. Anomaly detection using Isolation Forest identified Androy as the most atypical trajectory, with an anomaly score of approximately 0.71, followed by Analamanga (0.63). Unsupervised clustering revealed two regional groups, with a silhouette score close to 0.65, indicating a strong territorial polarization. Finally, the XGBoost model achieved satisfactory overall predictive performance, with $R^2 = 0.833$, MAE = 1,684, and RMSE = 6,725, although prediction errors increased for regions with very high enrollment levels.

Keywords: spatio-temporal dynamics; student enrollment; XGBoost; regional clustering; machine learning.

Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.23063741>

1 Introduction

L'enseignement supérieur constitue un levier majeur de formation du capital humain et de développement territorial. À Madagascar, son expansion s'accompagne toutefois d'importantes disparités régionales dans la répartition des étudiants et des établissements. Les statistiques officielles du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique montrent que les effectifs étudiants sont répartis de manière très inégale entre les régions, avec une concentration particulièrement importante autour des principaux pôles universitaires [1]. Ces disparités territoriales ne sont pas nouvelles et renvoient plus largement aux différences d'accès aux infrastructures éducatives et aux opportunités de formation observées entre les territoires malgaches [2]. Dans ce contexte, l'analyse des seules statistiques descriptives ne permet pas toujours d'identifier les trajectoires régionales, les comportements atypiques ou les structures sous-jacentes aux évolutions observées. Les méthodes de science des données et d'apprentissage automatique offrent des outils complémentaires pour modéliser ces dynamiques. En particulier, XGBoost permet de construire des modèles prédictifs performants à partir de relations complexes et non linéaires [3], tandis que Isolation Forest facilite la détection automatique de trajectoires atypiques [4]. Les approches de clustering permettent parallèlement d'identifier des groupes de régions présentant des profils temporels similaires, la qualité de cette classification pouvant être appréciée à l'aide du coefficient de silhouette [5]. Ainsi, cette étude analyse les dynamiques spatio-temporelles, la typologie régionale et la capacité prédictive des effectifs étudiants à Madagascar entre 2017 et 2022. Elle combine cartographie territoriale, analyse des trajectoires, détection d'anomalies, clustering non supervisé et modélisation XGBoost afin de caractériser les principales configurations régionales et d'évaluer leur évolution dans le temps.

2 Matériels et méthodes

2.1 Matériels

Les analyses ont été réalisées à partir de données statistiques sur les effectifs étudiants, organisées sous format Excel et extraites des annuaires statistiques du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique (MESUPRES) pour la période 2017–2022 [6]. Les traitements cartographiques ont mobilisé des données SIG vectorielles relatives aux limites administratives de Madagascar provenant du Foiben-Taosarintanin'i Madagasikara (FTM), organisme national de référence en information géospatiale [7]. L'ensemble des traitements statistiques, graphiques et de modélisation a été effectué sous Python [8], avec notamment Pandas et NumPy pour la préparation des données, GeoPandas pour leur intégration spatiale [9], Matplotlib pour les visualisations [10], ainsi que Scikit-learn et XGBoost pour la PCA, le clustering, la détection des trajectoires atypiques et la modélisation prédictive [11–12].

2.2 Méthodes

Les données ont d'abord été structurées sous forme de panel « région–année », puis contrôlées, harmonisées et enrichies par des variables dérivées décrivant l'évolution des effectifs, leur structure public–privé et leurs variations temporelles. L'analyse exploratoire a reposé sur des comparaisons interrégionales, des classements, des trajectoires temporelles et des représentations cartographiques. La réduction de dimension par analyse en composantes principales (ACP) a été utilisée pour synthétiser les profils régionaux [13], tandis que la classification des trajectoires a été réalisée par K-means, avec sélection du nombre de groupes à partir du coefficient de silhouette [14]. Les comportements régionaux atypiques ont été identifiés par Isolation Forest, méthode non supervisée adaptée à la détection d'anomalies multidimensionnelles [15]. La modélisation prédictive a mobilisé XGBoost, alimenté notamment par des variables temporelles retardées afin de prédire les effectifs des périodes ultérieures [16]. La robustesse prédictive a été évaluée selon une validation chronologique, en entraînant les modèles sur les années antérieures puis en testant leurs performances sur les années suivantes, à l'aide du coefficient de détermination, de l'erreur absolue moyenne et de la racine de l'erreur quadratique moyenne [17].

3 Résultats et discussions

3.1 Résultats

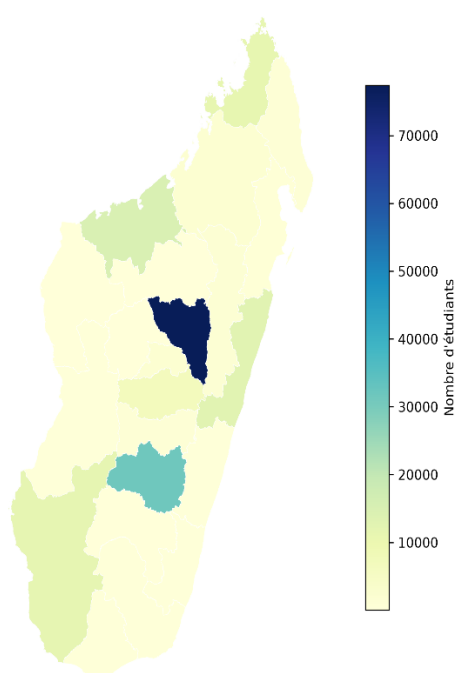
3.1.1 Forte concentration territoriale des effectifs étudiants dans l'enseignement supérieur à Madagascar en 2022

La figure 1 suivante présente la distribution spatiale des étudiants de l'enseignement supérieur à Madagascar en 2022 ainsi que le classement des régions selon leurs effectifs. La représentation cartographique met en évidence les disparités territoriales, tandis que le graphique de classement permet d'apprécier quantitativement l'ampleur des écarts entre les régions.

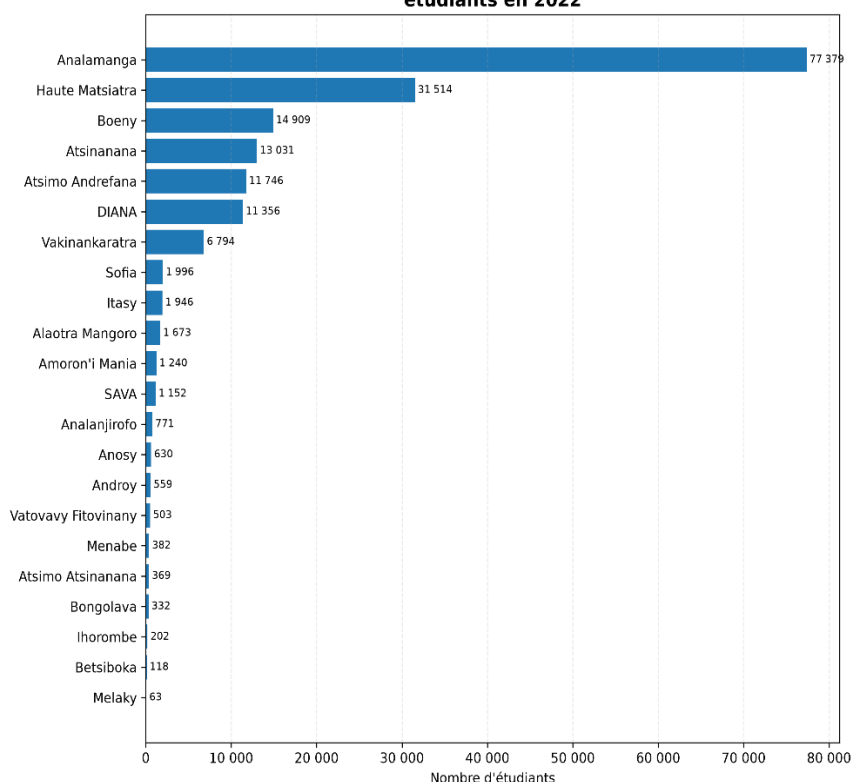
Figure 1: Forte concentration territoriale des effectifs étudiants dans l'enseignement supérieur à Madagascar en 2022

Enseignement supérieur à Madagascar : distribution territoriale des étudiants — 2022

A. Répartition spatiale des étudiants par région en 2022



B. Classement régional des effectifs étudiants en 2022



Source : MESUPRES, Annuaire Statistique - SSP / DSSIR, Traitement Python.

Source: Auteur 2026

Interprétation

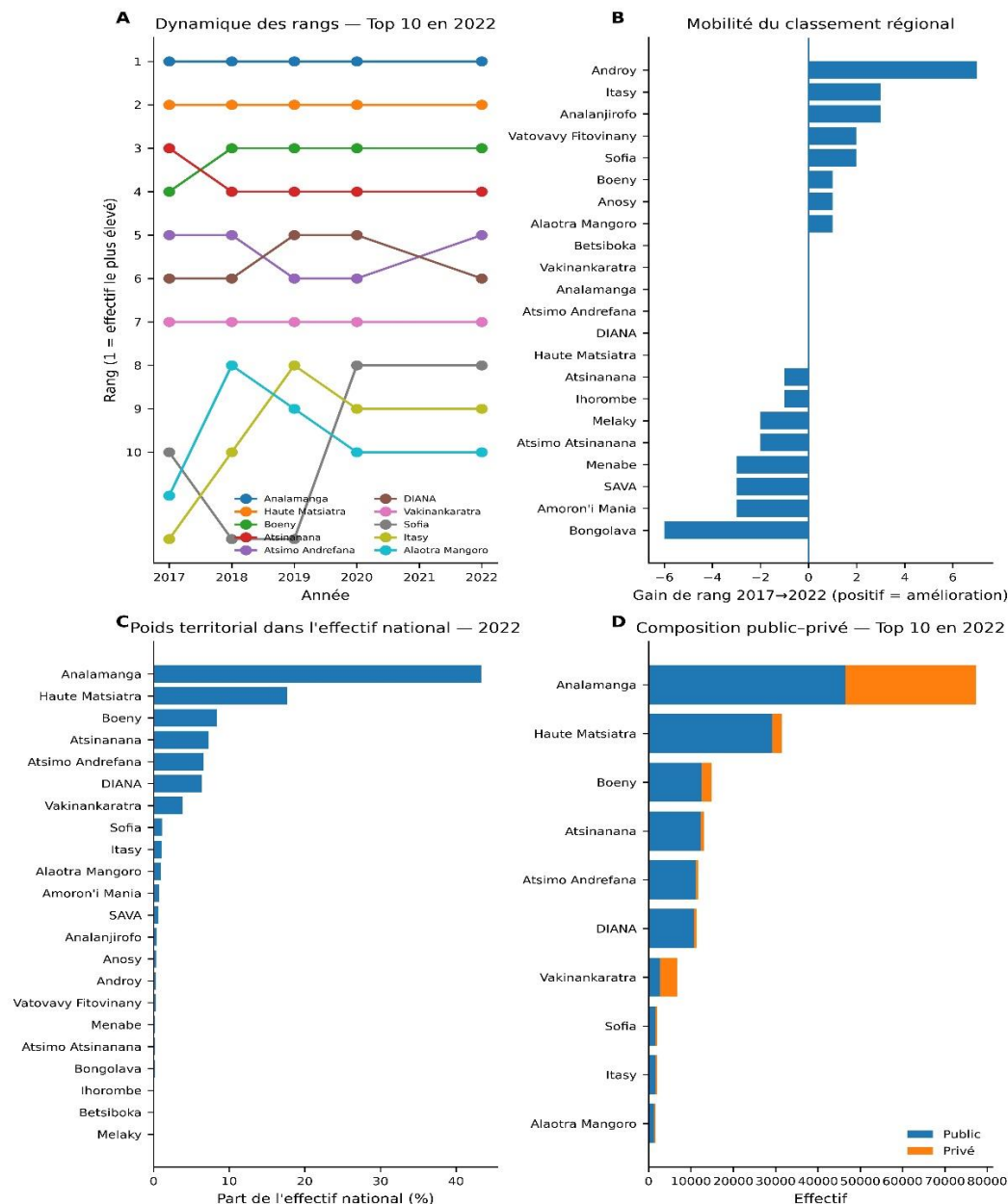
En 2022, l'enseignement supérieur malgache regroupe au total 178 665 étudiants, avec une concentration particulièrement marquée dans quelques régions. Analamanga domine très nettement la distribution avec 77 379 étudiants, soit environ 43,3 % de l'effectif national. Elle est suivie par Haute Matsiatra, qui compte 31 514 étudiants, représentant près de 17,6 % du total. À elles seules, ces deux régions regroupent ainsi environ 60,9 % des étudiants du pays, ce qui traduit une forte polarisation territoriale de l'offre et de la fréquentation de l'enseignement supérieur. Un deuxième groupe de régions présente des effectifs relativement importants, notamment Boeny avec 14 909 étudiants, Atsinanana avec 13 031, Atsimo Andrefana avec 11 746 et DIANA avec 11 356 étudiants. Ces six premières régions concentrent ensemble près de 89,5 % de l'ensemble des étudiants recensés en 2022. À l'inverse, plusieurs régions enregistrent des effectifs très faibles, notamment Melaky avec 63 étudiants, Betsiboka avec 118, Ihorombe avec 202 et Bongolava avec 332 étudiants. Ces résultats révèlent donc une forte asymétrie spatiale des effectifs universitaires à Madagascar, caractérisée par la prédominance d'un nombre limité de pôles régionaux, en particulier Analamanga et Haute Matsiatra. Cette configuration suggère que l'accès à l'enseignement supérieur demeure fortement structuré par la localisation territoriale des établissements

et des principaux centres universitaires. Elle justifie, dans la suite de l'analyse, l'examen des facteurs susceptibles d'expliquer ces disparités régionales à travers les approches de modélisation spatio-temporelle et d'apprentissage automatique.

3.1.2 Dynamique du classement régional, poids territorial et structure public-privé des effectifs étudiants à Madagascar entre 2017 et 2022

La figure 2 suivante met en perspective l'évolution du classement des régions selon leurs effectifs étudiants entre 2017 et 2022, les changements de rang observés sur la période, leur poids relatif dans l'effectif national ainsi que la composition public-privé des principales régions en 2022. Elle permet ainsi d'apprécier simultanément la stabilité de la hiérarchie territoriale, les mobilités régionales et les différences de structure de l'enseignement supérieur.

Figure 2: Dynamique du classement régional, poids territorial et structure public-privé des effectifs étudiants à Madagascar entre 2017 et 2022



Source : Auteur 2026

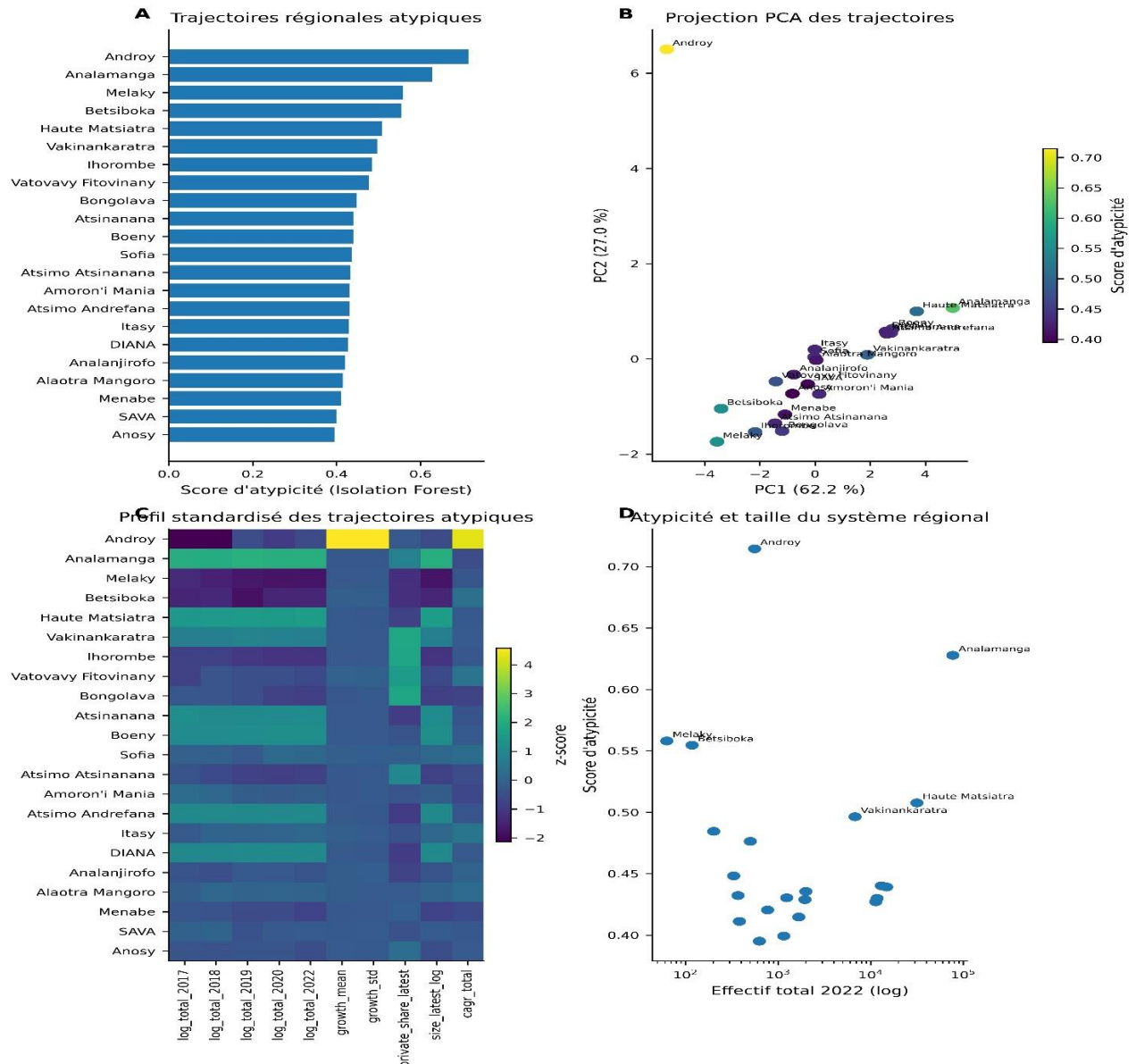
Interprétation

La hiérarchie des principales régions apparaît globalement stable sur la période étudiée. Analamanga conserve systématiquement la première position, tandis que Haute Matsiatra demeure deuxième de 2017 à 2022. Boeny progresse de la quatrième à la troisième place, alors qu'Atsinanana recule du troisième au quatrième rang. Atsimo Andrefana, DIANA et Vakinankaratra restent respectivement autour des cinquième, sixième et septième positions. Parmi les progressions les plus marquées, Androy gagne sept places entre 2017 et 2022, tandis qu'Itasy et Analanjirofo gagnent chacune trois rangs. À l'inverse, Bongolava perd six places, et Menabe, SAVA ainsi qu'Amoron'i Mania reculent chacune de trois rangs. Cette stabilité du sommet du classement s'accompagne d'une forte concentration des effectifs. En 2022, Analamanga représente à elle seule 43,3 % de l'effectif national, devant Haute Matsiatra avec 17,6 %, Boeny avec 8,3 %, Atsinanana avec 7,3 %, Atsimo Andrefana avec 6,6 % et DIANA avec 6,4 %. Les autres régions occupent des parts nettement plus réduites, Vakinankaratra ne représentant par exemple qu'environ 3,8 % du total national. La structure public-privé montre également des configurations territoriales contrastées. À Analamanga, sur 77 379 étudiants, 46 473 sont inscrits dans le public et 30 906 dans le privé, ce dernier représentant près de 39,9 % de l'effectif régional. Haute Matsiatra reste beaucoup plus largement dominée par le secteur public, avec 29 184 étudiants publics contre 2 330 privés. Une situation particulièrement distinctive apparaît à Vakinankaratra, où le privé compte 4 064 étudiants contre 2 730 dans le public, soit près de 59,8 % de l'effectif régional. Ces résultats montrent que, derrière une forte stabilité des principaux pôles universitaires, les trajectoires régionales et les configurations public-privé demeurent fortement différenciées, ce qui constitue un élément important à intégrer dans la modélisation des disparités territoriales de l'enseignement supérieur.

3.1.3 Détection des trajectoires régionales atypiques des effectifs étudiants par apprentissage automatique à Madagascar entre 2017 et 2022

La figure 3 ci-après présente l'identification des trajectoires régionales atypiques des effectifs étudiants à partir de l'algorithme non supervisé Isolation Forest, complétée par une analyse en composantes principales (PCA) et par l'examen des profils standardisés des régions. Cette approche permet de repérer les territoires dont l'évolution entre 2017 et 2022 s'écarte le plus du comportement général, sans présupposer au préalable la forme de ces trajectoires.

Figure 3: Détection des trajectoires régionales atypiques des effectifs étudiants par apprentissage automatique à Madagascar entre 2017 et 2022



Source : Auteur 2026

Interprétation

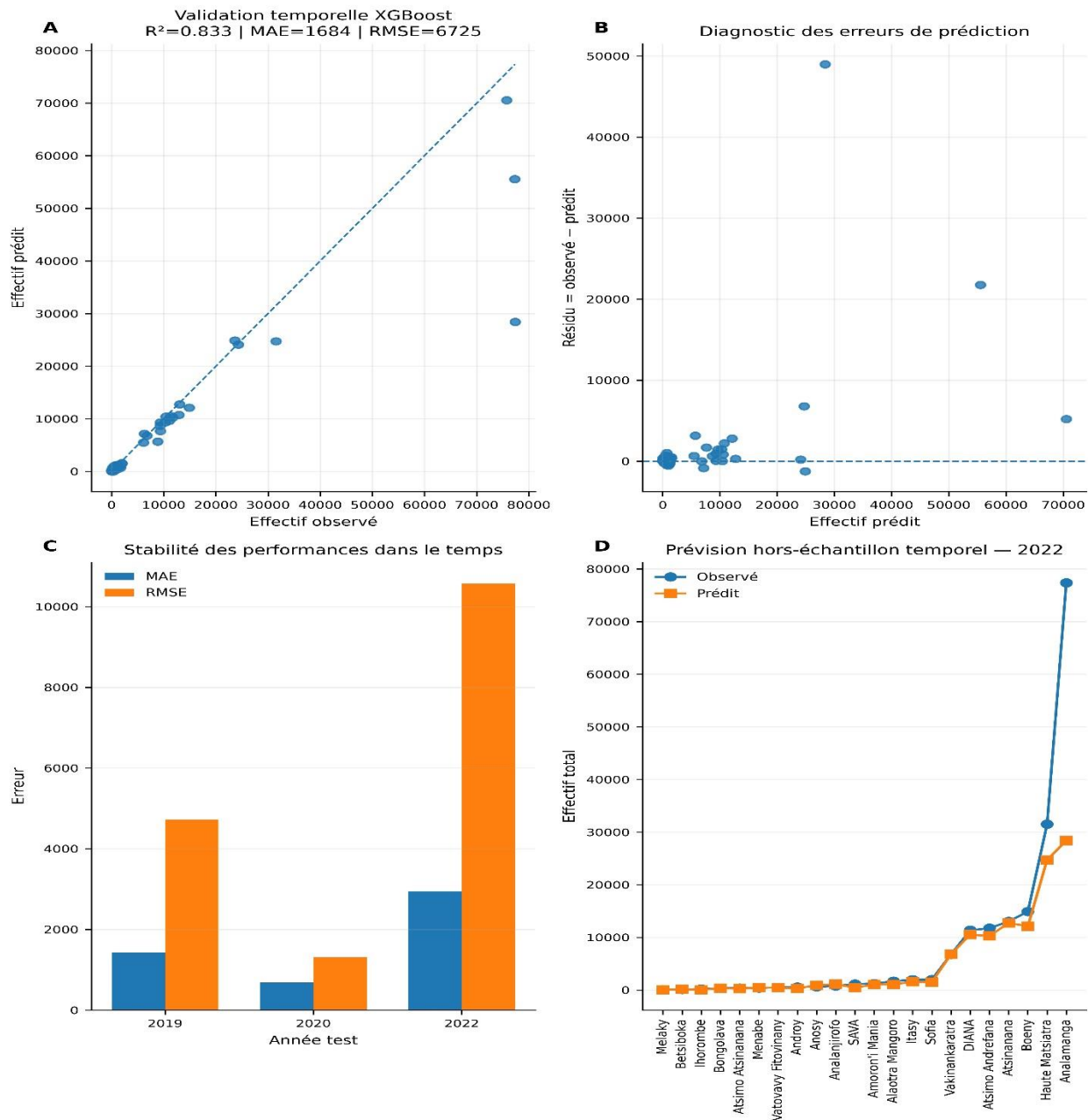
Les résultats montrent qu'Androy constitue la trajectoire la plus atypique, avec un score d'anomalie d'environ 0,71, nettement supérieur à celui des autres régions. Cette singularité est cohérente avec la forte transformation de ses effectifs, passés de seulement 7 étudiants en 2017 à 559 en 2022. Analamanga présente le deuxième score d'atypicité le plus élevé, autour de 0,63 ; contrairement à Androy, cette singularité résulte principalement de sa taille exceptionnelle, puisque la région concentre 77 379 étudiants en 2022, très au-dessus des autres territoires. Melaky et Betsiboka, avec des scores proches de 0,56, apparaissent également atypiques, mais à l'autre extrémité de la distribution, avec respectivement seulement 63 et 118 étudiants en 2022. Haute Matsiatra et Vakinankaratra présentent des niveaux intermédiaires d'atypicité, proches de 0,50, tandis que la majorité des autres régions se situent autour de 0,40 à 0,45. La projection PCA confirme cette différenciation. Les deux premières composantes expliquent ensemble 89,2 % de la variabilité totale des trajectoires, dont 62,2 % pour PC1 et 27,0 % pour PC2. Androy se détache très nettement du nuage principal sur la deuxième composante, tandis qu'Analamanga et Haute Matsiatra occupent des positions distinctes du fait de l'importance de leurs effectifs. À l'inverse, la majorité des régions restent relativement regroupées dans l'espace factoriel, ce qui traduit des profils d'évolution plus proches. L'analyse conjointe des profils standardisés montre ainsi que l'atypicité détectée ne correspond pas nécessairement à une contre-performance. Elle traduit plutôt une trajectoire statistiquement inhabituelle, pouvant provenir d'une croissance exceptionnelle, d'un niveau d'effectif extrêmement élevé ou, au contraire, d'un volume particulièrement faible. Ces résultats justifient l'intégration de méthodes non supervisées dans l'analyse spatio-

temporelle, car elles permettent d'identifier automatiquement les territoires nécessitant une investigation spécifique avant la construction des modèles prédictifs et explicatifs.

3.1.4 Performance prédictive et stabilité temporelle du modèle XGBoost appliqué aux effectifs étudiants régionaux à Madagascar

La figure 4 ci-dessous présente l'évaluation temporelle du modèle XGBoost utilisé pour prédire les effectifs étudiants régionaux à partir des trajectoires observées au cours de la période étudiée. Elle confronte les valeurs observées aux valeurs prédites, examine la distribution des résidus, compare les erreurs de prédiction selon les années de validation et détaille les résultats de la prévision hors échantillon pour 2022. Cette validation temporelle permet d'apprécier non seulement la qualité globale du modèle, mais également sa capacité à généraliser à des périodes non utilisées lors de l'apprentissage.

Figure 4: Performance prédictive et stabilité temporelle du modèle XGBoost appliqué aux effectifs étudiants régionaux à Madagascar



Source : Auteur 2026

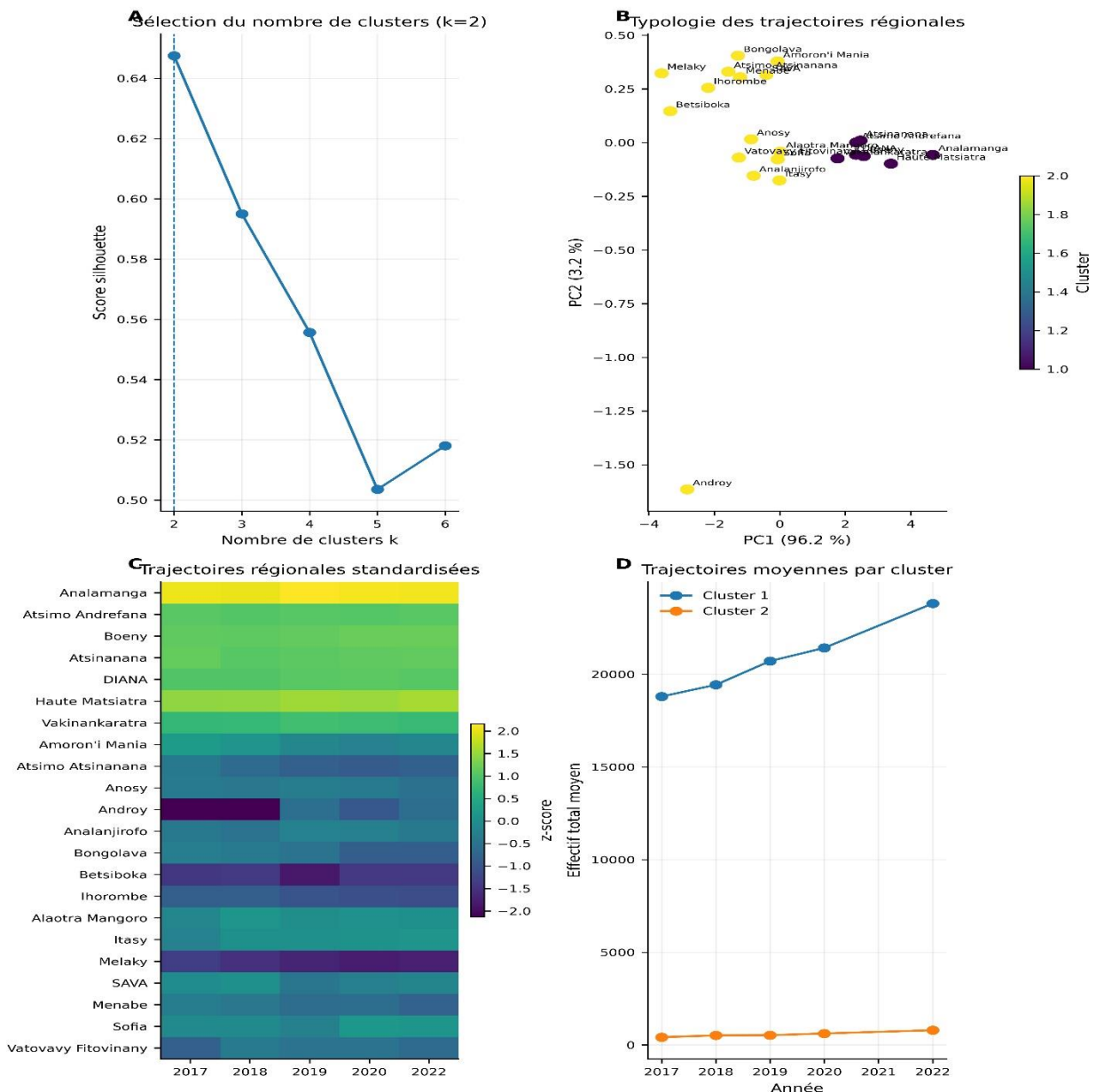
Interprétation

Globalement, le modèle présente une capacité prédictive satisfaisante, avec un coefficient de détermination de $R^2 = 0,833$, indiquant qu'environ 83,3 % de la variabilité des effectifs observés est expliquée par les prédictions du modèle. L'erreur absolue moyenne s'établit à 1 684 étudiants, tandis que la racine de l'erreur quadratique moyenne atteint 6 725 étudiants. Le nuage observé-prédit montre que la majorité des observations, notamment celles correspondant aux effectifs faibles et intermédiaires, se situent relativement près de la diagonale d'égalité. Les écarts deviennent cependant plus importants pour les régions présentant les effectifs les plus élevés, ce qui révèle une moindre capacité du modèle à reproduire certaines valeurs extrêmes. L'analyse temporelle met en évidence une variation sensible de la précision prédictive selon les années. Les meilleures performances sont observées pour 2020, avec une MAE d'environ 700 étudiants et une RMSE proche de 1 300 étudiants. En 2019, l'erreur augmente, avec une MAE d'environ 1 400 étudiants et une RMSE proche de 4 700 étudiants. La dégradation est nettement plus importante en 2022, pour laquelle la MAE approche 2 900 étudiants et la RMSE dépasse 10 000 étudiants. Cette hausse importante de la RMSE traduit l'existence de quelques erreurs de grande amplitude plutôt qu'une détérioration uniforme des prédictions pour toutes les régions. La prévision hors échantillon de 2022 confirme cette observation. Pour la majorité des régions à effectifs faibles ou moyens, les valeurs prédites restent relativement proches des effectifs effectivement observés. En revanche, le modèle sous-estime fortement certains grands pôles universitaires. Le cas le plus marqué concerne Analamanga, où l'effectif observé atteint 77 379 étudiants, alors que la prédiction se situe autour de 28 000 étudiants, soit un écart proche de 49 000 étudiants. Haute Matsiatra, avec 31 514 étudiants observés, est également sous-estimée, avec une prévision avoisinant 25 000 étudiants. Ces écarts expliquent largement l'augmentation de la RMSE en 2022. Ces résultats montrent ainsi que XGBoost restitue correctement la tendance générale et les niveaux d'effectifs de la majorité des régions, mais rencontre davantage de difficultés lorsque les trajectoires présentent des évolutions exceptionnelles ou des niveaux très éloignés de ceux observés durant l'apprentissage. La performance globale élevée du modèle doit donc être interprétée conjointement avec l'analyse des résidus et la validation temporelle. Cette hétérogénéité justifie, dans la suite de l'analyse, le recours aux méthodes d'intelligence artificielle explicable afin d'identifier les variables et les dynamiques régionales responsables des principales erreurs de prédiction.

3.1.5 Typologie des trajectoires régionales des effectifs étudiants par clustering non supervisé à Madagascar entre 2017 et 2022

La figure 5 suivante présente une classification non supervisée des régions malgaches selon leurs trajectoires d'effectifs étudiants entre 2017 et 2022. La détermination du nombre optimal de groupes repose sur le coefficient de silhouette, tandis que l'analyse en composantes principales permet de visualiser la structure des régions dans un espace réduit. Les trajectoires standardisées et les profils moyens des clusters complètent cette analyse en mettant en évidence les différences de niveau et d'évolution des effectifs au cours de la période.

Figure 5: Typologie des trajectoires régionales des effectifs étudiants par clustering non supervisé à Madagascar entre 2017 et 2022



Source : Auteur 2026

Interprétation

Le critère de silhouette identifie deux clusters comme la partition optimale, avec un score d'environ 0,65, supérieur aux valeurs obtenues pour trois à six groupes. Cette configuration traduit l'existence de deux profils régionaux suffisamment distincts dans les données. La projection factorielle confirme cette séparation : la première composante principale explique à elle seule 96,2 % de la variance totale, tandis que la deuxième en explique 3,2 %, soit environ 99,4 % de l'information représentée par les deux axes. La structuration des régions dépend ainsi principalement des différences de niveau et de trajectoire des effectifs étudiants. Le premier cluster regroupe essentiellement les principaux pôles de l'enseignement supérieur, notamment Analamanga, Haute Matsiatra, Boeny, Atsinanana, Atsimo Andrefana, DIANA et Vakinankaratra, caractérisés par des effectifs nettement supérieurs à la moyenne nationale. À l'intérieur de ce groupe, Analamanga se distingue particulièrement par des valeurs standardisées très élevées sur l'ensemble de la période. Le second cluster rassemble principalement les régions dont les effectifs demeurent beaucoup plus faibles, telles que Melaky, Betsiboka, Ihorombe, Bongolava, Menabe, SAVA ou encore Amoron'i Mania. Androy occupe cependant une position singulière dans cet ensemble, en raison de la forte modification de sa trajectoire au cours de la période, malgré un niveau absolu d'effectif encore relativement faible. Les trajectoires moyennes mettent clairement en évidence l'ampleur de cette différenciation. L'effectif moyen du premier groupe passe approximativement de 18 700 étudiants en 2017 à près de 23 800 en 2022, tandis que celui du second cluster demeure inférieur à 1 000 étudiants en moyenne, malgré une progression

au cours de la période. Cette classification révèle ainsi une structure fortement polarisée de l'enseignement supérieur malgache, opposant un nombre restreint de grands pôles universitaires à un ensemble beaucoup plus important de régions à faible densité étudiante. La mise en évidence de ces deux régimes territoriaux constitue un résultat important pour la modélisation spatio-temporelle, car elle suggère que les dynamiques régionales ne peuvent être considérées comme homogènes et qu'elles nécessitent des analyses différenciées selon les profils territoriaux.

3.2 Discussions

Les résultats mettent en évidence une forte concentration territoriale des effectifs étudiants à Madagascar, dominée par Analamanga et Haute Matsiatra, tandis que plusieurs régions restent faiblement représentées. Cette polarisation est confirmée par la stabilité du classement régional et par la structuration en deux grands profils territoriaux identifiés par le clustering. La détection d'anomalies montre par ailleurs que certaines régions, notamment Androy et Analamanga, présentent des trajectoires très différentes du comportement général. Le modèle XGBoost reproduit correctement une grande partie de la variabilité observée, avec une performance globale satisfaisante, mais ses erreurs augmentent pour les régions à effectifs très élevés ou connaissant des évolutions atypiques. Ces résultats suggèrent que les dynamiques étudiantes ne suivent pas un schéma homogène à l'échelle nationale. Ils confirment l'intérêt d'une approche combinant analyse spatiale, classification non supervisée, détection d'anomalies et apprentissage automatique pour mieux caractériser les disparités régionales de l'enseignement supérieur malgache.

4 Conclusion

Cette étude met en évidence de fortes disparités régionales dans la répartition et l'évolution des effectifs étudiants à Madagascar entre 2017 et 2022. Les analyses montrent une concentration marquée autour de quelques pôles universitaires, associée à des trajectoires territoriales différenciées et à l'existence de régions atypiques. Le clustering confirme cette structuration en profils distincts, tandis que le modèle XGBoost présente une capacité prédictive globalement satisfaisante malgré des limites pour certaines régions à effectifs très élevés ou à évolution exceptionnelle. L'approche combinant analyse spatio-temporelle, apprentissage non supervisé et modélisation prédictive apparaît ainsi pertinente pour mieux comprendre les dynamiques territoriales de l'enseignement supérieur et fournir des éléments d'aide à la planification et à la décision.

REFERENCES

- [1] MESUPRES. (2024/2025). Annuaire statistique de l'Enseignement Supérieur et de la Recherche Scientifique 2021–2022. Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Madagascar.
- [2] Carron, G., & Ta Ngoc Châu (Eds.). (1980). Regional Disparities in Educational Development: Diagnosis and Policies for Reduction. Paris: UNESCO, International Institute for Educational Planning.
- [3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [4] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. Proceedings of the IEEE International Conference on Data Mining, 413–422. <https://doi.org/10.1109/ICDM.2008.17>
- [5] 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](https://doi.org/10.1016/0377-0427(87)90125-7)
- [6] MESUPRES. Annuaire statistique de l'Enseignement Supérieur et de la Recherche Scientifique 2021–2022. Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Madagascar.

- [7] Foiben-Taosarintanin'i Madagasikara (FTM). Données géospatiales et limites administratives de Madagascar. Institut Géographique et Hydrographique National, Madagascar.
- [8] Python Software Foundation. Python Language Reference and Documentation. Python Software Foundation.
- [9] GeoPandas Developers. GeoPandas Documentation: Geographic Data Analysis in Python.
- [10] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- [11] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [12] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- [13] Jolliffe, I. T. (2002). *Principal Component Analysis*. 2nd ed. Springer, New York.
- [14] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297.
- [15] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. *Proceedings of the IEEE International Conference on Data Mining*, 413–422.
- [16] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- [17] Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450.