
Caractériser les environnements pour comprendre les inégalités spatiales de santé : robustesse des méthodes de profilage territorial.

Tracy Lurant^{*1,2}, Martin Paumelle^{*1}, Florent Occelli^{*1}, and Annabelle Deram^{*1}

¹Laboratoire de Génie Civil et Géo-Environnement (LGCgE) - ULR 4515 – Université d’Artois, Université de Lille, Junia, IMT Lille Douai, 941 rue Charles Bourseul, CS 10838, 59508 Douai Cédex – France

²Agence de l’Environnement et de la Maîtrise de l’Énergie – ADEME – France

Résumé

Comprendre les associations géographiques entre qualité de l’environnement et santé suppose de disposer d’indicateurs spatialisés capables de décrire finement la diversité des contextes territoriaux d’exposition aux nuisances et aménités environnementales. Dans cette perspective, Paumelle et al. (2023) ont proposé un cadre méthodologique fondé sur la réutilisation de données environnementales ouvertes pour construire des profils environnementaux à l’échelle communale. Cette approche, reposant sur une Analyse en Composantes Principales (ACP) et une Classification Ascendante Hiérarchique (CAH), visait à représenter la complexité de l’environnement physique à travers 39 indicateurs décrivant les niveaux de contamination et d’émission, l’occupation des sols, les pratiques agricoles, la naturalité et le climat. Sept profils environnementaux ont ainsi été identifiés dans le nord de la France, offrant une lecture holistique des pressions et aménités territoriales susceptibles d’influencer la santé des populations.

Toutefois, l’ACP est sensible à la multicolinéarité entre variables, nécessite une présélection importante des indicateurs et tend à mieux discriminer les profils extrêmes (territoires sous pressions urbaines ou espaces naturels) que les contextes intermédiaires. De fait, afin d’optimiser la robustesse de la méthode de profilage, la stabilité des résultats selon (i) le périmètre géographique analysé et (ii) la méthode statistique employée ont été testées.

Le travail présenté ici vise à comparer la robustesse et la capacité de différentes méthodes multivariées et d’approches reposant sur les réseaux de neurones à produire des typologies environnementales cohérentes et transposables. L’analyse s’appuie sur la base nationale en libre accès développée par Paumelle et al. (2024), qui compile 112 indicateurs spatialisés issus de 19 jeux de données publiques harmonisées à l’échelle communale. Trois méthodes ont été testées : l’ACP, l’Analyse Factorielle Multiple (AFM) et le *Self-Organizing Maps* (SOM). Les analyses ont été menées à deux échelles, nationale et régionale (Nord de la France), afin d’évaluer l’influence du périmètre d’étude sur la détection et la caractérisation des profils.

Les premiers résultats montrent que la méthode SOM s’avère la plus performante, avec une variance expliquée de 79,9 % (contre 62,1 % pour l’ACP et 55,96 % pour l’AFM). Elle améliore la différenciation des contextes environnementaux mixtes, où aucune variable

*Intervenant

ne domine clairement. Toutefois, la comparaison des cartes obtenues à différentes échelles révèle des divergences notables pour une même zone d'étude : plus le territoire est vaste, plus la capture des spécificités locales devient difficile. Ces résultats soulignent la nécessité de développer des méthodes de profilage environnemental scalables, capables de traiter de larges ensembles de données tout en préservant la finesse territoriale.

À terme, ces profils environnementaux seront croisés avec des indicateurs socio-économiques et sanitaires, dans le but de mieux comprendre les interactions entre environnement, territoires et santé à l'échelle nationale. Cette démarche met ainsi en avant l'analyse des situations de multi-exposition grâce à la mobilisation de données ouvertes. Elle s'inscrit par ailleurs dans une approche intégrée environnement-santé pour documenter les inégalités territoriales.

Mots-Clés: Deep Learning, Typologies environnementales territorialisées, Modélisation à grande échelle, Analyse multi déterminants, Scalabilité