

République du Sénégal

Un Peuple – Un But – Une Foi



Ministère de l'Économie, du Plan et de la Coopération

Agence nationale de la Statistique et de la Démographie

École nationale de la Statistique et de l'Analyse économique Pierre Ndiaye



Statistique exploratoire spatiale

Synthèse des travaux pratiques collectifs

Rédigé par :

NKWA TSAMO Leslye Patricia
Elève Ingénieur Statisticien Économiste en
ISE 1 Cycle Long

Sous la supervision de :

M. Aboubacar HEMA
Analyste de recherche et data scientist

Février 2025

Sommaire

| | |
|--|----------|
| Introduction | 1 |
| TP 0 : Présentation de Google Earth Engine | 2 |
| TP 1 : Population et paludisme au Bénin | 2 |
| TP 2 : Accessibilité aux infrastructures au Bénin | 3 |
| TP 3 : Couverture sanitaire et aires protégées au Bénin | 4 |
| TP 4 : Identification des terres arables en Éthiopie | 5 |
| TP 5 : Étalement urbain et ratio LCRPGR au Cameroun | 6 |
| TP 6 : Indices spectraux au Burkina Faso | 7 |
| Conclusion générale | 8 |
| Appréciation du cours | 8 |

Liste des figures

| | | |
|---|--|---|
| 1 | Extrait de la présentation GEE. | 2 |
| 2 | Infrastructures de santé et d'éducation au Bénin. | 3 |
| 3 | Couverture sanitaire avec zones tampons différencierées. | 4 |
| 4 | Répartition des terres arables en Éthiopie. | 5 |
| 5 | Carte LCRPGR du Cameroun par région (2015 à 2020). | 6 |
| 6 | dNBR au Burkina Faso (Sentinel-2, 2021 à 2022). | 7 |

Introduction

L'analyse des données spatiales tient une place prépondérante dans la conception des dynamiques territoriales et l'aide à la décision. Que ce soit pour le territoire, le développement agricole, l'urbanisme ou encore la gestion de l'environnement, les données géospatiales et satellitaires deviennent incontournables. Dans ce cadre, notre cours de statistiques exploratoires spatiales donné à l'ENSAE par M. Aboubacar Hema a été très utile. L'approche adoptée, fondée sur le « learning by doing », nous a conduit à nous pencher sur des problématiques concrètes mobilisant des données et des outils directement applicables dans la vie professionnelle.

Au cours du semestre, sept travaux pratiques collec-

tifs avaient pour objectif de nous amener à maîtriser les méthodes d'analyse exploratoire spatiale. Les quatre premiers, hormis le TP0, ont été réalisés en Python, à l'aide notamment des bibliothèques geopandas, rasterio et folium. Dans les TP suivants, Google Earth Engine a été utilisé avec JavaScript. Ce document présente une synthèse des travaux menés au cours du semestre. Nous y abordons plusieurs thématiques, telles que l'accessibilité aux infrastructures au Bénin, le potentiel agricole en Éthiopie, l'étalement urbain au Cameroun et la caractérisation environnementale du Burkina Faso. L'objectif est de restituer les méthodes employées et les principaux enseignements tirés de cette expérience.

TP 0 : Présentation de Google Earth Engine

Le premier exercice du semestre consistait en une présentation collective de la plateforme Google Earth Engine (GEE) devant la classe. Cette activité préliminaire visait à nous familiariser avec l'environnement cloud de traitement de données géospatiales massives développé par Google.

Nous avons abordé l'architecture client-serveur de GEE, son catalogue de données (Sentinel, Landsat, MODIS, couverture terrestre), les objets fondamentaux (`ee.Image`, `ee.ImageCollection`, `ee.FeatureCollection`) ainsi que les opérations de base telles que le filtrage, le mosaïquage, le calcul d'indices spectraux et la réduction statistique. Les deux interfaces disponibles ont été présentées : l'éditeur JavaScript en ligne et l'API Python.



Figure 1: Extrait de la présentation GEE.

GEE a occupé une place importante dans le cours, notamment pour les TP 5 et 6. Cette présentation nous a permis de mieux cerner ce qui nous attendait en nous donnant les connaissances nécessaires aux travaux ultérieurs.

TP 1 : Population et paludisme au Bénin

Introduction

Ce premier travail pratique portait sur l'importation et la visualisation de données géospatiales relatives à la population et au paludisme au Bénin. Réalisé en Python à l'aide des bibliothèques geopandas, rasterio et folium, il visait à initier les étudiants à la manipulation conjointe de données raster (WorldPop, Malaria Atlas Project) et vectorielles (geoBoundaries) dans un contexte d'analyse territoriale appliquée à la santé publique.

1. Contexte

La répartition spatiale de la population et les indicateurs de santé constituent un préalable à la planification des politiques publiques. Au Bénin, population et prévalence du paludisme présentent des disparités régionales que ce TP cherchait à quantifier à l'aide d'outils géospatiaux.

2. Objectifs

L'objectif général était de maîtriser l'importation et la visualisation de données géospatiales multi-sources. Spécifiquement, il fallait :

1. Charger et harmoniser données vectorielles et raster (geoBoundaries, WorldPop, Malaria Atlas Project).
2. Calculer des statistiques zonales de population et de prévalence par unité administrative.
3. Produire des cartes statiques et interactives.

3. Méthodologie

Les limites administratives du Bénin ont été chargées depuis geoBoundaries et projetées en WGS 84. Les données WorldPop (100 m et 1 km, 2020) et les taux PfPR du Malaria Atlas Project ont été lus avec rasterio. Le croisement raster-vecteur a permis de calculer

population totale, densité et taux moyen de paludisme par département et par commune.

4. Résultats et discussion

Les statistiques zonales ont révélé des disparités notables entre les départements béninois. Les zones les plus densément peuplées se concentrent dans le sud du pays, autour de Cotonou et Porto-Novo, avec des densités pouvant dépasser 2 000 habitants par km². Les départements septentrionaux (Alibori, Atacora) présentent des densités nettement inférieures, souvent en dessous de 50 habitants par km². Concernant le paludisme, les taux de prévalence les plus élevés ont été observés précisément dans ces régions rurales du nord, où l'accès aux infrastructures de santé reste limité. Ce constat suggérait alors une corrélation spatiale entre faible densité, éloignement des services et prévalence élevée. Les cartes interactives folium ont permis de superposer ces couches et d'identifier visuellement les zones prioritaires pour l'intervention sanitaire.

5. Défis rencontrés

La gestion des systèmes de coordonnées (WGS 84 vs UTM) et le volume des rasters à 100 m ont constitué les principales difficultés. La compréhension de la différence entre projections géographiques et projetées a

représenté un apprentissage fondamental.

6. Apport du TP

Ce TP a posé les bases de la manipulation géospatiale en Python et illustré l'intérêt des statistiques spatiales pour l'aide à la décision sanitaire.

TP 2 : Accessibilité aux infrastructures au Bénin

Introduction

Ce deuxième travail pratique analysait l'accessibilité spatiale aux infrastructures essentielles au Bénin, couvrant les domaines de la santé, de l'éducation et du transport. Réalisé en Python avec geopandas, folium et matplotlib, il mobilisait des données OpenStreetMap et visait à évaluer la couverture territoriale des services publics, à calculer des distances d'accès par département et à identifier les zones sous-desservies.

1. Contexte

L'accès aux infrastructures conditionne le développement territorial. Au Bénin, la répartition des établissements de santé, des écoles et du réseau routier présente des inégalités marquées entre le sud urbanisé et le nord rural.

2. Objectifs

L'objectif général était d'évaluer l'accessibilité des populations aux services essentiels. Spécifiquement, il s'agissait de :

1. Cartographier la distribution des infrastructures de santé, d'éducation et de transport.
2. Calculer les distances moyennes aux infrastructures les plus proches, par département.
3. Identifier les zones sous-desservies par croisement avec les données de densité.

3. Méthodologie

Les données OpenStreetMap ont été extraites et projetées en UTM 31N. Des zones tampons et des calculs de distance au plus proche voisin ont été réalisés pour chaque catégorie d'infrastructure. Les résultats ont été agrégés par département sous forme de tableaux statistiques et de cartes choroplèthes.

4. Résultats et discussion

L'analyse a confirmé l'existence de disparités régionales profondes. Les départements du sud (Littoral, Atlantique, Ouémé) bénéficient d'une concentration élevée d'infrastructures, avec des distances moyennes d'accès inférieures à 5 km. À l'inverse, les départements du nord (Alibori, Atacora, Borgou) présentent des distances pouvant dépasser 20 km pour atteindre un hôpital. Le réseau de transport dessert essentiellement l'axe nord-sud, laissant de vastes portions du territoire à l'écart des corridors de circulation majeurs. La figure 2 met en évidence cette concentration des équipements dans le sud du pays. Ces résultats soulignent la nécessité d'investissements ciblés dans les zones rurales du nord pour réduire les

inégalités d'accès.

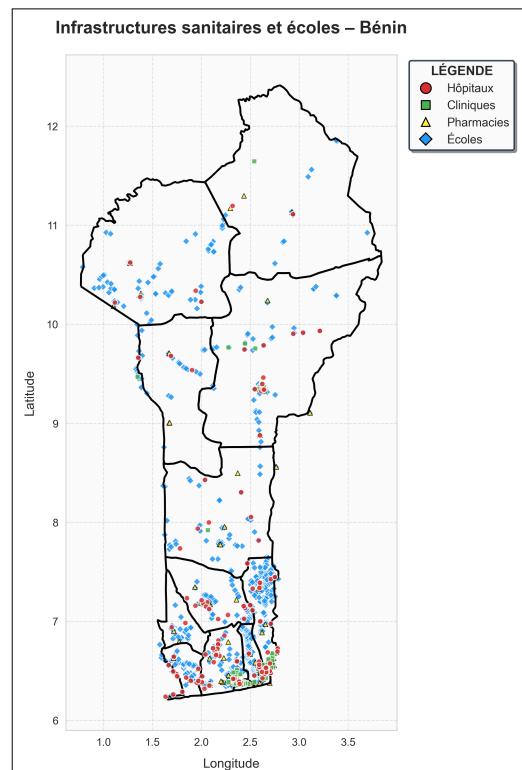


Figure 2: Infrastructures de santé et d'éducation au Bénin.

5. Défis rencontrés

La qualité inégale des données OSM a constitué un défi. Certaines catégories d'infrastructures étaient sous-recensées en zone rurale, entraînant une possible surestimation des distances. L'optimisation des calculs de distance sur de grands volumes ponctuels a également nécessité l'usage d'index spatiaux.

6. Apport du TP

Ce TP a permis de pratiquer les analyses de proximité et de mettre en évidence les déséquilibres territoriaux en matière d'accès aux services essentiels.

TP 3 : Couverture sanitaire et aires protégées au Bénin

Introduction

Ce troisième travail pratique approfondissait l'analyse géospatiale du Bénin en intégrant la couverture sanitaire, les aires protégées issues de la World Database on Protected Areas (WDPA) et les corridors de transport. Réalisé en Python, il mobilisait des zones tampons différenciées selon le contexte urbain ou rural et visait à produire une analyse territoriale intégrée croisant santé, environnement et mobilité.

1. Contexte

La planification territoriale nécessite une vision transversale. Au Bénin, santé publique, conservation et infrastructures de transport interagissent sur un même espace. Comprendre leur articulation à l'échelle départementale constitue un enjeu pour l'aménagement du territoire.

2. Objectifs

L'objectif général était de produire une analyse territoriale intégrée. Spécifiquement, il s'agissait de :

1. Évaluer la population desservie par les hôpitaux et pharmacies, en différenciant zones urbaines et rurales.
2. Quantifier la population dans les aires protégées et les corridors de transport.
3. Produire des cartes de synthèse identifiant les zones à enjeux multiples.

3. Méthodologie

Les localités ont été classées en urbain/rural selon la densité WorldPop dans un rayon de 2 km. Des buffers différenciés ont été appliqués : 5 km/15 km pour les hôpitaux, 2 km/7 km pour les pharmacies. Les corridors de transport ont été définis par un buffer de 2 km autour des axes principaux. Les aires protégées proviennent de la WDPA.

4. Résultats et discussion

Selon les analyses, la couverture hospitalière du sud a atteint plus de 80% de la population, mais reste inférieure à 40% dans les départements du nord. Les pharmacies se concentrent en zone urbaine et sont quasi absentes des zones rurales éloignées. Plusieurs départements septentrionaux abritent des populations significatives à l'intérieur ou en périphérie d'aires protégées, ce qui pose la question de la compatibilité entre conservation et accès aux services (figure 3). Par ailleurs, les localités situées hors des corridors de transport cumulent souvent éloignement sanitaire et enclavement économique.

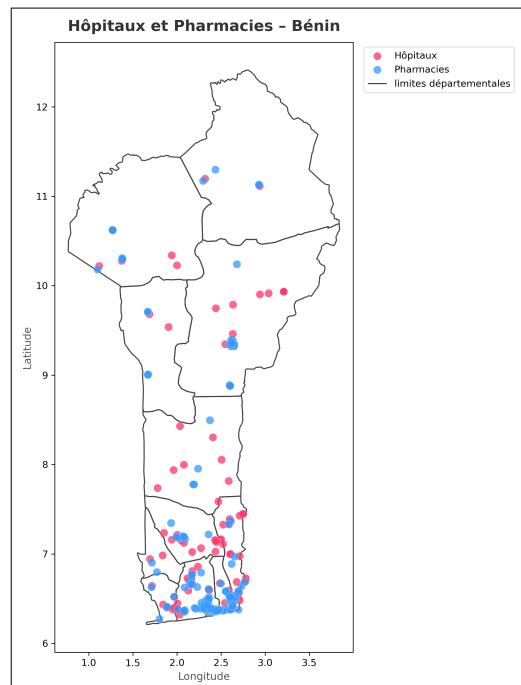


Figure 3: Couverture sanitaire avec zones tampons différencierées.

5. Défis rencontrés

La définition du seuil urbain/rural a influencé directement les rayons de couverture et les résultats. La superposition de multiples couches d'information a également complexifié la lisibilité des cartes. Par ailleurs, des différences ont été observées entre certains résultats obtenus et la réalité du terrain, notamment au niveau de certains chiffres nous a conduits à interroger la qualité et la fiabilité de certaines données open source utilisées dans l'analyse.

6. Appart du TP

Ce TP a montré l'intérêt de croiser plusieurs thématiques pour analyser un territoire et a permis de mieux comprendre l'utilisation des zones tampons différencierées.

TP 4 : Identification des terres arables en Éthiopie

Introduction

Ce quatrième travail pratique portait sur l'estimation des terres arables en Éthiopie à partir de données raster multi-sources : couverture terrestre (GFSAD30), déforestation (Hansen), surfaces imperméables (GMIS), occurrence d'eau (Pekel) et aires protégées (WDPA). Réalisé en Python, il visait à quantifier le potentiel agricole national par règles d'exclusion successives et à valider les résultats par confrontation aux données de la FAO.

1. Contexte

L'Éthiopie, deuxième pays le plus peuplé d'Afrique, fait face à des défis de sécurité alimentaire. L'estimation précise des terres cultivables constitue un enjeu stratégique pour la planification agricole.

2. Objectifs

L'objectif général était d'identifier et de quantifier les terres arables par combinaison de sources raster. Spécifiquement, il s'agissait de :

1. Déterminer un seuil optimal d'occurrence d'eau permanente (données Pekel).
2. Combiner cultures, forêts, surfaces imperméables et aires protégées pour estimer la superficie nette.
3. Désagrégérer les résultats par région administrative et produire des cartes.

3. Méthodologie

Trois scripts ont été développés. Le premier analysait l'occurrence d'eau pour fixer le seuil à 75%, détectant 0,65 million d'hectares d'eau permanente. Le deuxième combinait GFSAD30, Hansen, GMIS et WDPA par exclusions successives. Le troisième désagrégéait les résultats vers les unités ADM1 à ADM3 par facteurs d'allocation agro-écologiques.

4. Résultats et discussion

L'estimation finale s'établit à 15,05 millions d'hectares de terres arables, soit 13,3% du territoire éthiopien. Ce résultat présente un écart de seulement 0,3% par rapport aux données de référence de la FAO (15,0 millions d'hectares), validant la robustesse de la méthode. La désagrégation régionale a révélé que les régions Oromia et Amhara concentrent la majorité du potentiel cultivable du pays (figure 4), tandis que les régions périphériques (Afar, Somali) présentent des potentiels nettement plus limités en raison de conditions clima-

tiques arides. Le processus d'exclusion par étapes a permis de quantifier précisément la contribution de chaque facteur (forêts, eau, bâti, aires protégées) à la réduction de la superficie brute.

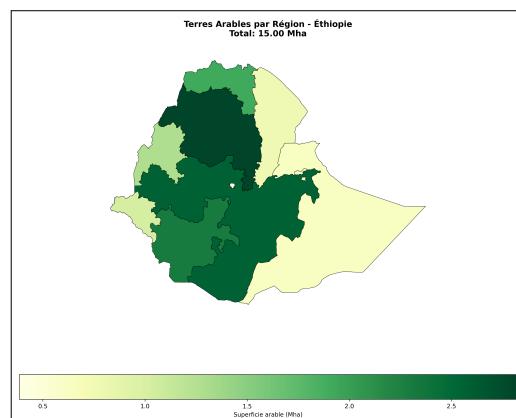


Figure 4: Répartition des terres arables en Éthiopie.

5. Défis rencontrés

La manipulation de rasters à 30 m sur l'ensemble du territoire Éthiopien a nécessité un travail par blocs et l'usage de masques binaires. Le choix du seuil d'eau a demandé une comparaison de plusieurs valeurs (75%, 85%, 90%, 95%) avant validation. Par ailleurs, des différences ont été observées entre certains résultats obtenus et la réalité du terrain, ce qui nous a conduits à interroger la qualité et la fiabilité de certaines données open source utilisées dans l'analyse.

6. Apport du TP

Ce TP a montré que les méthodes de télédétection peuvent fournir des estimations du potentiel agricole à une échelle fine. Toutefois, ces résultats restent dépendants des données utilisées et doivent être interprétés avec prudence, même lorsqu'ils sont comparés aux statistiques internationales.

TP 5 : Étalement urbain et ratio LCRPGR au Cameroun

Introduction

Ce cinquième travail pratique portait sur le calcul du ratio LCRPGR (Land Consumption Rate to Population Growth Rate), indicateur officiel de l'ODD 11.3.1 proposé par UN-Habitat. Réalisé en JavaScript sur Google Earth Engine, il mobilisait les données GHSL (Global Human Settlement Layer) pour les surfaces bâties, WorldPop pour la population et GADM pour les limites régionales du Cameroun, sur la période 2015 à 2020.

1. Contexte

L'urbanisation rapide constitue un phénomène structurant en Afrique. Le LCRPGR quantifie la relation entre consommation de terres et croissance démographique. Un ratio supérieur à 1 traduit un étalement urbain non maîtrisé, un ratio inférieur à 1 une densification.

2. Objectifs

L'objectif général était de spatialiser le ratio LCRPGR par région au Cameroun. Spécifiquement :

1. Extraire les surfaces bâties (GHSL) pour 2015 et 2020 et calculer le LCR par région.
2. Estimer le PGR à partir des grilles WorldPop par région administrative GADM1.
3. Produire cartes raster, tableaux régionaux et histogrammes du ratio LCRPGR.

3. Méthodologie

Le script GEE calcule, pour chaque région, la somme des surfaces bâties (GHSL, en m^2 par pixel) et de la population (WorldPop) aux deux dates. Le LCR correspond au taux de variation annuel des surfaces bâties, le PGR au logarithme du rapport des populations divisé par le nombre d'années. Le ratio $LCRPGR = LCR/PGR$ est calculé par région, rasterisé et exporté en CSV et GeoTIFF. Une interface interactive avec panneau d'information a été développée dans GEE.

4. Résultats et discussion

Les résultats révèlent des dynamiques contrastées entre les dix régions camerounaises. L'Extrême-Nord et l'Adamaoua présentent des ratios inférieurs à 1, suggérant une densification de l'habitat existant plutôt

qu'une extension spatiale. À l'inverse, le Centre et le Littoral affichent des ratios supérieurs à 2, témoignant d'une consommation de terres largement supérieure à ce que la croissance démographique justifierait (figure 5). La région du Centre, qui abrite Yaoundé, présente le ratio le plus élevé, signe d'un étalement périurbain rapide. Ces résultats offrent un diagnostic territorial utile pour orienter les politiques d'aménagement urbain.

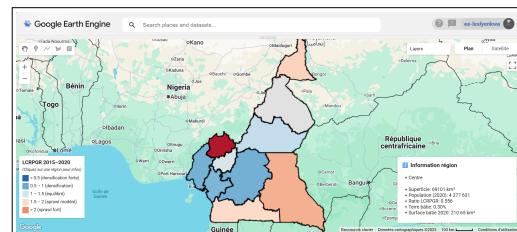


Figure 5: Carte LCRPGR du Cameroun par région (2015 à 2020).

5. Défis rencontrés

Le passage à GEE JavaScript a représenté un changement important. La syntaxe asynchrone côté serveur (`ee.Number`, `ee.Image`) et le paramétrage des échelles de réduction (`scale`, `tileScale`) ont nécessité un temps d'adaptation pour éviter les erreurs de calcul et les dépassements de mémoire.

6. Apport du TP

Ce TP a permis de maîtriser un indicateur internationalement reconnu pour le suivi de l'urbanisation durable et d'exploiter la puissance de calcul cloud de GEE à l'échelle nationale.

TP 6 : Indices spectraux au Burkina Faso

Introduction

Ce dernier travail pratique portait sur le calcul de six indices spectraux (dNBR, NDVI, SAVI, NDBI, NDWI, MNDWI) à partir d'images Sentinel-2 sur les 45 départements du Burkina Faso, en vue d'une fusion avec les données de l'enquête EHCVM. Réalisé en JavaScript sur Google Earth Engine pour le calcul des indices, puis en R pour la fusion des bases, il couvrait la période de juillet 2021 à juillet 2022 afin de capturer un cycle saisonnier complet.

1. Contexte

La caractérisation environnementale par télédétection représente un enjeu croissant pour les politiques de développement. L'EHCVM fournit des données socio-économiques détaillées mais sans variables environnementales spatialement référencées. Ce TP visait à combler cette lacune.

2. Objectifs

L'objectif général était de produire une base combinant indicateurs environnementaux et données socio-économiques à l'échelle départementale. Spécifiquement :

1. Calculer six indices spectraux à partir de composites Sentinel-2 dénuagés sur les 45 départements.
2. Agréger les valeurs moyennes par département et exporter les résultats en CSV et XLSX.
3. Fusionner les indices avec la base EHCVM en R pour constituer une base enrichie.

3. Méthodologie

Trois scripts GEE thématiques ont été développés : `Script_brulis_ndbr.js` pour le dNBR (sévérité des brûlis), `script_vegetation.js` pour le NDVI et le SAVI (végétation), et `BF_Indice_eau.js` pour le NDWI et le MNDWI (eau). Les images Sentinel-2 SR ont été filtrées et dénuagées par la couche SCL. Les indices ont été agrégés par département (FAO GAUL niveau 2). Le script R (`script_de_merge.R`) a fusionné les résultats avec l'EHCVM par code départemental pour produire une base finale au format Stata.

4. Résultats et discussion

Les résultats mettent en évidence des contrastes environnementaux marqués. Le NDVI présente un gradient nord-sud prononcé, avec des valeurs faibles dans la zone sahélienne et des valeurs plus élevées dans les régions soudanaises du sud-ouest, traduisant

la transition bioclimatique du pays. Le dNBR révèle l'intensité des feux de brousse, particulièrement fréquents dans les zones de savane (figure 6). Le NDBI confirme la concentration des surfaces bâties autour de Ouagadougou et Bobo-Dioulasso. Les indices d'eau (NDWI, MNDWI) permettent d'identifier les plans d'eau saisonniers, avec des différences notables entre saison humide et saison sèche. La base enrichie produite ouvre la voie à des analyses croisant conditions environnementales et indicateurs de bien-être des ménages.

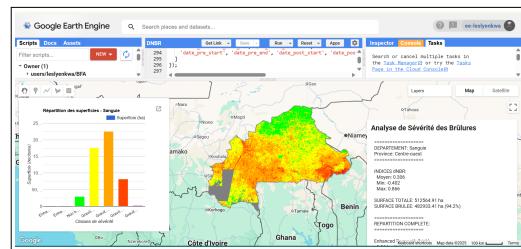


Figure 6: dNBR au Burkina Faso (Sentinel-2, 2021 à 2022).

5. Défis rencontrés

La couverture nuageuse variable a nécessité un filtrage rigoureux par la couche SCL. La compréhension physique de chaque indice spectral (bandes utilisées, phénomène mesuré) a demandé un effort d'apprentissage en télédétection optique. L'harmonisation des nomenclatures administratives entre GEE (FAO GAUL) et l'EHCVM a également posé des difficultés lors de la jointure.

6. Apport du TP

Ce dernier TP a permis de mobiliser plusieurs compétences travaillées au cours du semestre, en combinant les petites bases acquises en télédétection, agrégation statistique et fusion de bases hétérogènes. Il a ainsi mis en évidence l'intérêt d'articuler des outils comme GEE et R dans une même démarche d'analyse.

Conclusion générale

Les sept travaux pratiques collectifs ont constitué une progression bien structurée dans l'apprentissage de l'analyse spatiale. Les trois premiers ont introduit la manipulation de données géospatiales avec Python, en utilisant les bibliothèques geopandas, rasterio et folium. Ils ont couvert les reprojections, les statistiques zonales et la création de zones tampons. Le quatrième TP a porté sur le traitement de rasters multi-sources à haute résolution et la validation par données de référence. Les cinquième et sixième travaux ont présenté l'utilisation de Google

Earth Engine, permettant l'accès au calcul cloud et aux catalogues d'images satellitaires.

Aussi, les thématiques abordées concernaient des enjeux de développement territorial en Afrique : accessibilité des services au Bénin, potentiel agricole en Éthiopie, expansion urbaine au Cameroun et caractérisation environnementale au Burkina Faso. Cette diversité nous a montré l'utilité des statistiques spatiales pour l'appui aux politiques publiques et a renforcé la maîtrise des méthodes de télédétection et d'analyse spatiale.

Appréciation du cours

L'approche d'enseignement de M. Hema, qui repose sur le « learning by doing », nous a conduit à découvrir nous-mêmes les concepts clés de l'analyse spatiale grâce à notre formation. Chaque session nous confrontait à des enjeux, utilisant des informations authentiques et des instruments professionnels, ce qui a encouragé une participation dynamique et un degré croissant d'indépendance. Les erreurs de programmation, les décisions méthodologiques et les défis auxquels nous avons été confrontés ont toutes constitué des occasions d'améliorer nos compétences.

Toutefois, il convient de relever certains manquements de cette approche. L'évaluation des travaux pratiques ne présentait pas toujours un niveau de rigueur ou d'explicitation suffisant pour nous guider vers les pratiques optimales, ce qui a pu restreindre notre aptitude à déterminer avec précision la marche à suivre dans certaines circonstances. Par ailleurs, le focus des travaux pratiques sur l'aspect pratique a pu

affecter notre compréhension des principes théoriques essentiels, indispensables pour une maîtrise totale de l'exploration des données spatiales.

En dépit de ces contraintes, les contributions demeurent significatives. L'éventail d'outils (Python, Google Earth Engine, R) et la variété de contextes examinés ont favorisé le développement de notre polyvalence technique. La collaboration en équipe a amélioré nos aptitudes à travailler ensemble, particulièrement grâce à l'organisation de projets sur GitHub et la distribution des responsabilités. Ce cours a fortement contribué à améliorer notre approche des données géospatiales.

Nous tenons à exprimer notre gratitude envers M. Hema pour la qualité de son enseignement et sa pédagogie pratique, ainsi qu'envers le corps administratif de l'ENSAE, en particulier le directeur M. Idrissa Diagne, le directeur, et le chef de filière M. Ibrahima BARRY.