



République du Sénégal

Un Peuple-Un But-Une Foi

MINISTERE DE L'ECONOMIE DU PLAN ET DE LA COOPERATION



Agence nationale de la Statistique et de la Démographie



Ecole nationale de la Statistique et de l'Analyse Economique

STATISTIQUE EXPLORATOIRE SPATIALE : Résumé des travaux pratiques

Rédigé par :

Mamadou Lamine DIABANG

Élève Ingénieur Statisticien Economiste

Sous la supervision de :

M. Aboubacar HEMA

Research Analyst – International Food Policy Research
Institute (IFPRI)

Février 2026

Table des matières

INTRODUCTION	3
1.1. Approche Pédagogique et Environnement Technique	3
1.2. Levier Technologique : Le TP0	3
1.3. Aperçu des Projets Réalisés	3
EXPERTISE EN ÉPIDÉMIOLOGIE SPATIALE : DYNAMIQUE DU PALUDISME AU TCHAD (2000-2024).....	5
ANALYSE DE L'ACCESSIBILITÉ AUX SERVICES SOCIAUX DE BASE ET RÉPARTITION DÉMOGRAPHIQUE AU TCHAD (TP2 & TP3).....	7
OPTIMISATION ET VIABILITÉ DES TERRES AGRICOLES (OUGANDA).....	9
DYNAMIQUES D'URBANISATION ET INDICATEURS DE DURABILITÉ (CAMEROUN)	11
COUPLAGE TÉLÉDÉTECTION ET ANALYSE SOCIO-ÉCONOMIQUE (BURKINA FASO)	13
CONCLUSION GÉNÉRALE ET PERSPECTIVES DÉCISIONNELLES	15

INTRODUCTION

Le cours de **Statistique Exploratoire Spatiale (SES)** a pour objectif fondamental d'initier à la statistique descriptive spatiale et à la cartographie thématique avancée. Dans un contexte où la donnée géolocalisée devient un levier stratégique pour les politiques publiques, ce cours a permis d'appréhender des concepts clés tels que les matrices de voisinage, l'autocorrélation spatiale et le traitement d'images satellitaires.

1.1. Approche Pédagogique et Environnement Technique

L'enseignement s'est déroulé selon une approche résolument pratique, favorisant le travail collaboratif en groupe et le développement de capacités de synthèse scientifique. Afin de répondre à la complexité des problématiques territoriales en Afrique, une pluralité d'environnements technologiques a été mobilisée :

- **R & Python:** Pour l'analyse robuste et la manipulation de données vectorielles et rasters en local.
- **Python (API GEE) :** Pour l'automatisation de scripts complexes et l'analyse de données massives.
- **JavaScript (Google Earth Engine) :** Pour le prototypage rapide et la visualisation interactive sur le Cloud.

1.2. Levier Technologique : Le TP0

En amont des analyses thématiques, un **TP0** a été consacré à la prise en main de l'environnement **Google Earth Engine**. Cette étape cruciale a permis de se familiariser avec l'interface du Code Editor, l'importation d'Assets (fichiers Shapefiles) et la manipulation des collections d'images (Sentinel, Landsat). Ce socle technique a rendu possible le traitement de données à l'échelle nationale sans les contraintes de puissance de calcul locale.

1.3. Aperçu des Projets Réalisés

Le présent document synthétise six travaux pratiques portant sur des enjeux majeurs de développement en Afrique subsaharienne. Chaque projet est structuré selon un format et une thématique particulière.

Travail	Thématique et Objectifs	Pays d'étude	Outils
TP 1 - 3	Accessibilité aux infrastructures de base et population	Tchad	Python
TP 4	Identification et filtrage des terres arables	Ouganda	Python (API)
TP 5	Analyse de l'étalement urbain (Ratio LCRPGR)	Cameroun	GEE
TP 6	Indices spectraux et fusion avec données d'enquête (EHCVM)	Burkina Faso	GEE

EXPERTISE EN ÉPIDÉMIOLOGIE SPATIALE : DYNAMIQUE DU PALUDISME AU TCHAD (2000-2024)

Le TP1 est consacré à l'analyse spatio-temporelle de l'incidence du paludisme à *Plasmodium falciparum* sur le territoire tchadien. L'enjeu est de transformer des données de surveillance sanitaire en indicateurs géographiques permettant d'identifier les zones prioritaires pour les interventions de lutte contre la maladie.

Consigne et Objectif Stratégique : L'objectif principal était de cartographier la répartition du taux d'incidence du paludisme au **Tchad** sur une période de 25 ans. Il s'agissait de passer d'une vision globale et nationale à une analyse granulaire, afin d'observer les tendances de recul ou de persistance de l'endémie et d'identifier les "hotspots" épidémiologiques dans un contexte sahélien et soudanien.

Inputs et Nature des Données : L'analyse repose sur un socle de données hétérogènes traitées via l'API Python de Google Earth Engine :

- **Séries Temporelles Raster :** Un ensemble de 25 fichiers **.tiff** (couches *Global Pf Incidence Rate*) couvrant chaque année de 2000 à 2024, importés comme actifs (*assets*) ou collections d'images.
- **Référentiel Administratif :** Les limites administratives de la base GADM pour le **Tchad**, déclinées du niveau national (Niveau 0) aux échelons des départements et sous-préfectures (Niveaux 1 à 3).

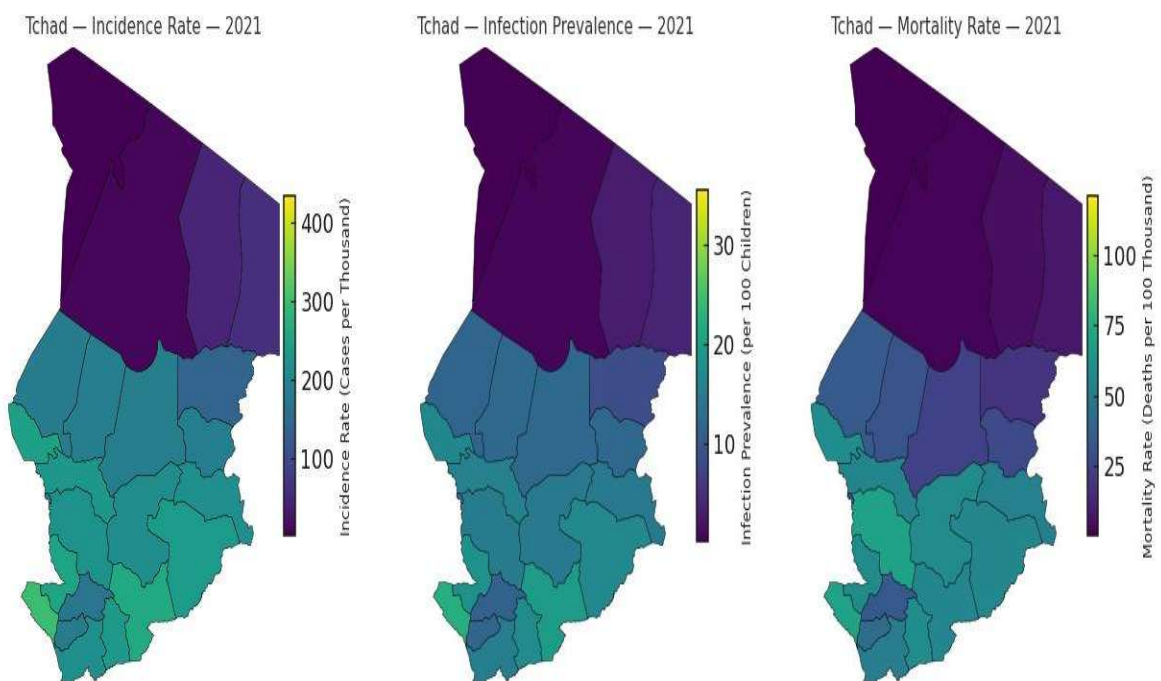
Méthodologie Opérationnelle (Python) : Le flux de traitement a été automatisé grâce aux bibliothèques **Pandas**, **Matplotlib** et **Geemap** :

1. **Ingestion et Prétraitement :** Utilisation de **ee.ImageCollection** pour charger la série temporelle et application d'un filtre spatial (**clip**) sur les frontières du Tchad.
2. **Statistiques Zonales :** Calcul des moyennes d'incidence par région administrative à l'aide de la fonction **reduceRegions** de GEE, puis conversion des résultats en **DataFrame Pandas** pour analyse.
3. **Visualisation Interactive :** Utilisation de **geemap** et **folium** pour générer des cartes interactives permettant de naviguer dans le temps (curseur temporel) et d'interroger les valeurs au clic sur chaque département.

4. **Analyse de Tendances** : Génération de graphiques linéaires avec `matplotlib` pour comparer l'évolution de l'incidence entre les régions du sud (plus humides) et celles du nord (arides).

Défis Techniques : La principale difficulté a résidé dans la gestion du volume de données sur une plateforme Cloud. L'automatisation en **Python** a nécessité une gestion rigoureuse des boucles pour traiter les 25 années sans dépasser les limites de calcul de l'API. De plus, l'harmonisation des palettes de couleurs sous `geemap` a été cruciale pour assurer que la réduction de l'incidence soit visuellement comparable d'une année à l'autre.

Outputs :



ANALYSE DE L'ACCESSIBILITÉ AUX SERVICES SOCIAUX DE BASE ET RÉPARTITION DÉMOGRAPHIQUE AU TCHAD (TP2 & TP3)

Plutôt que de simplement cartographier des points, cette section analyse la capacité de l'État tchadien à couvrir les besoins de sa population. L'enjeu est de transformer des données brutes (OSM et GADM) en indicateurs de rupture de service.

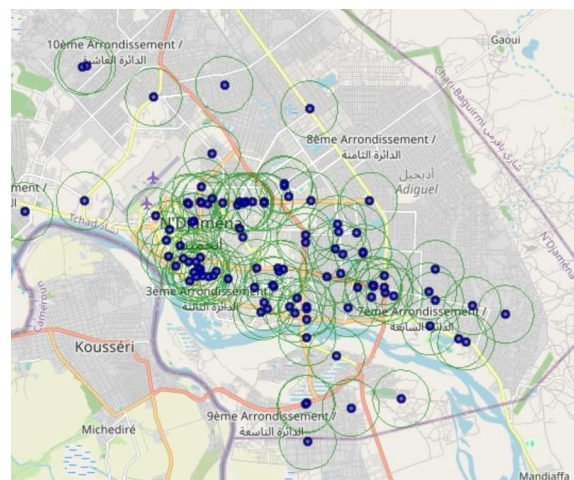
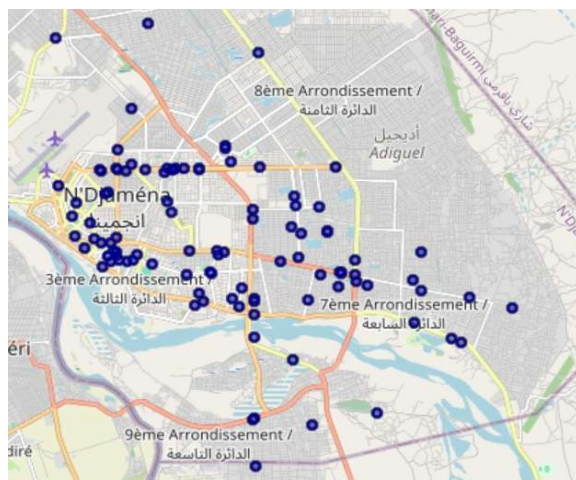
Consigne et Objectif Stratégique : L'objectif était de quantifier le déficit d'accès aux services de santé et d'éducation. Il ne s'agissait pas seulement de localiser les infrastructures, mais de mesurer l'efficacité de leur répartition spatiale face à une population en forte croissance.

Inputs et Architecture des Données : Le projet a nécessité le croisement de trois couches critiques :

- **Le référentiel administratif :** Niveaux 0 à 3 du Tchad (GADM).
- **L'offre de soins et d'éducation :** Points d'intérêt extraits d'OpenStreetMap (Hôpitaux, cliniques, écoles).
- **La réalité démographique :** Rasters WorldPop (100m et 1km) pour une précision granulaire de la distribution humaine.

Méthodologie Opérationnelle : Nous avons opté pour une approche par **zones de chalandise théoriques**. En générant des buffers multi-annulaires (1km, 5km et 10km) autour de chaque infrastructure, nous avons créé un "masque d'accessibilité". Ce masque a ensuite été superposé au raster de population sous Google Earth Engine. Grâce aux fonctions de réduction de régions (**reduceRegions**), nous avons pu extraire une donnée complexe : la "population hors-zone", soit le nombre réel d'individus situés au-delà des seuils de marche ou de transport acceptables.

outputs :



Leçon de Gouvernance : L'enseignement majeur de cette analyse est la mise en évidence des **zones d'ombre**. Nous avons constaté que la multiplication des infrastructures en zone urbaine masque souvent une exclusion quasi totale des populations rurales. La statistique spatiale permet ici de passer d'une gestion "au nombre d'écoles" à une gestion "par habitant desservi", changeant radicalement la vision de l'aménagement du territoire.

OPTIMISATION ET VIABILITÉ DES TERRES AGRICOLES (OUGANDA)

L'Ouganda possède un potentiel agricole immense, mais souvent contraint par des zones de conservation ou des topographies accidentées. L'objectif de cette section est de passer d'une vision brute du territoire à une cartographie précise du capital foncier exploitable.

Objectif Stratégique et Consigne : Le défi consistait à isoler les "terres arables" en éliminant systématiquement les zones impropres à la culture. Contrairement aux analyses classiques, ce travail repose sur une logique d'exclusion multi-critères pour garantir la durabilité environnementale des résultats.

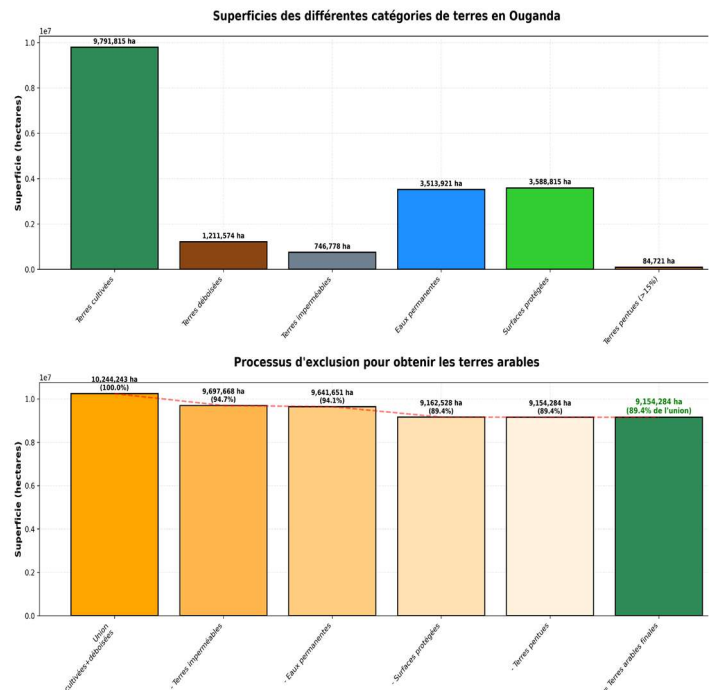
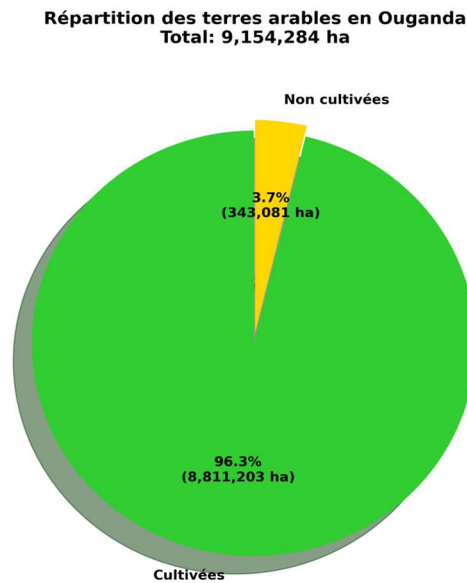
Inputs et Synergie des Données : Pour obtenir un diagnostic fiable, nous avons croisé des sources de données mondiales de haute résolution (GEE Data Catalog) :

- **Topographie :** SRTM (Pentes) pour limiter l'érosion.
- **Environnement :** Hansen Global Forest Change (Déforestation) et JRC (Eaux de surface).
- **Réglementaire :** WDPA pour le respect des aires protégées nationales.
- **Urbanisme :** GISA pour exclure les surfaces imperméabilisées (bâti).

Méthodologie par Filtrage Booléen : La démarche a été structurée autour d'un algorithme de "masquage" successif. Nous avons défini des seuils critiques : exclusion des pentes supérieures à 15%, des forêts denses, et des plans d'eau permanents. Un défi majeur a été l'intégration du raster de déforestation de Hansen ; sa densité de données étant trop élevée pour une importation directe, nous avons dû procéder à un **pré-découpage (clipping) via QGIS** sur l'emprise exacte de l'Ouganda avant son exploitation dans l'API Python de GEE.

Défis Techniques et Arbitrages : Le principal obstacle a été le compromis entre la précision spatiale et la puissance de calcul. L'utilisation de données à 30m de résolution (SRTM, Hansen) sur l'ensemble du territoire ougandais a nécessité une optimisation des fonctions de réduction (**reducer**) pour obtenir des statistiques de superficies par région GAUL sans erreur d'exécution (timeout).

Outputs :



Leçon de Planification Agricole : Ce projet démontre que la sécurité alimentaire ne dépend pas de la surface totale d'un pays, mais de sa **superficie utile**. La transformation des rasters thématiques en statistiques tabulaires (Pandas) a permis de mettre en évidence que certaines régions ougandaises, bien que vastes, disposent d'un potentiel arable très réduit une fois les aires protégées et les fortes pentes déduites. Cela souligne l'importance d'une planification agricole qui respecte les équilibres écologiques.

DYNAMIQUES D'URBANISATION ET INDICATEURS DE DURABILITÉ (CAMEROUN)

Le développement urbain en Afrique subsaharienne est l'un des plus rapides au monde.

L'enjeu de ce TP est de déterminer si l'expansion physique des villes camerounaises est en adéquation avec leur croissance démographique, ou si nous assistons à un phénomène d'étalement urbain (urban sprawl) non maîtrisé.

Le développement urbain en Afrique subsaharienne est l'un des plus rapides au monde.

L'enjeu de ce pôle est de déterminer si l'expansion physique des villes camerounaises est en adéquation avec leur croissance démographique, ou si nous assistons à un phénomène d'étalement urbain (urban sprawl) non maîtrisé.

Objectif Stratégique et Consigne : L'analyse s'est concentrée sur le calcul de l'indicateur **ODD 11.3.1**. L'objectif était de quantifier le ratio entre le taux de consommation des terres (LCR) et le taux de croissance de la population (PGR) au Cameroun sur la période 2015-2020. Il s'agissait de répondre à une question critique : les villes consomment-elles plus d'espace qu'elles ne gagnent d'habitants ?

Inputs et Synergie des Données : Pour cette analyse temporelle, nous avons croisé deux sources de données massives traitées sur le Cloud (GEE) :

- **Couche Bâti :** Le *Global Human Settlement Layer* (GHSL) pour extraire les surfaces construites en 2015 et 2020.
- **Couche Démographique :** Les données WorldPop de ces deux mêmes années pour obtenir la variation de la densité de population.
- **Référentiel Spatial :** Les limites GADM niveau 1 pour une désagrégation des résultats par région administrative camerounaise.

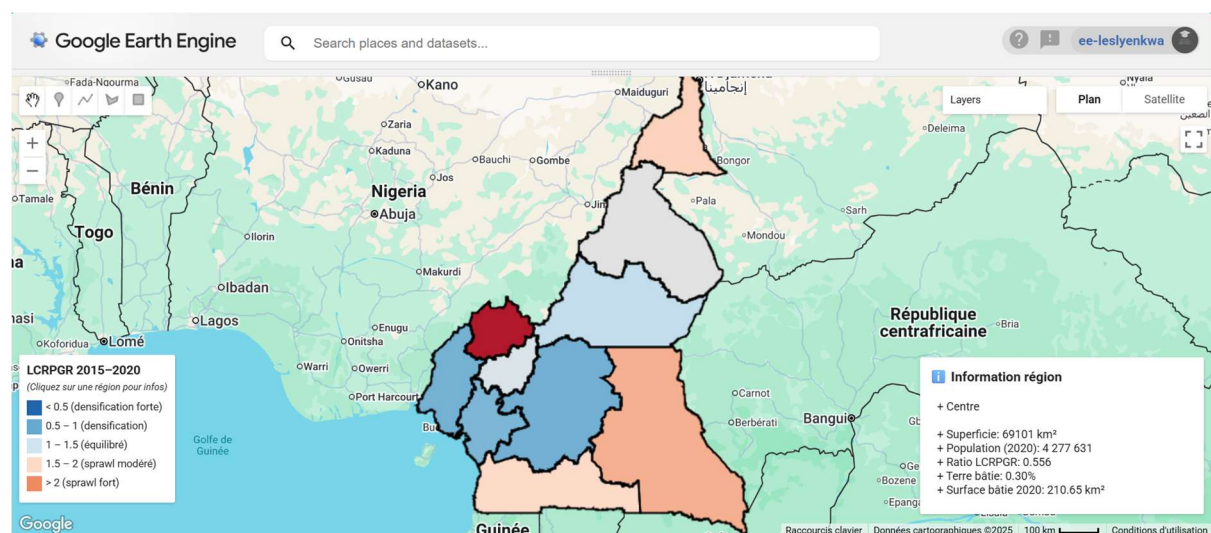
Méthodologie et Rigueur Mathématique : La méthodologie repose sur le calcul des taux de croissance annuels composés. Nous avons implémenté les formules logarithmiques préconisées par l'ONU :

1. $LCR = [\ln (\text{Surface_t1} / \text{Surface_t2}) / y] * 100$
2. $PGR = [\ln (\text{Pop_t1} / \text{Pop_t2}) / y] * 100$
3. $\text{Ratio LCRPGR} = LCR / PGR$

L'utilisation de la fonction `.evaluate()` dans l'environnement JavaScript de GEE a permis de générer des statistiques descriptives (moyenne, écart-type, min, max) directement exploitables pour le diagnostic régional.

Défis Techniques et Arbitrages : Le défi majeur a été l'harmonisation des échelles de résolution. Les données de bâti (GHSL) présentent une résolution de 100m tandis que WorldPop est à 1km. Nous avons dû opérer une **ré-échantillonnage spatial** pour assurer la cohérence des calculs au pixel près. De plus, la gestion des valeurs "zéro" ou négatives (dans le cas d'une dépopulation locale) a nécessité l'implémentation de filtres conditionnels pour éviter des erreurs de calcul dans les fonctions logarithmiques.

Outputs :



Leçon de Planification Urbaine : Ce travail met en lumière la dualité du développement urbain au Cameroun. Un ratio LCRPGR > 1 indique un étalement urbain qui pèse sur les finances publiques (coût des infrastructures), tandis qu'un ratio < 1 suggère une densification qui, bien qu'efficace spatialement, peut poser des problèmes de salubrité si elle n'est pas accompagnée de services adéquats. Cette analyse permet aux urbanistes d'identifier les régions où le foncier est "gaspillé" au détriment des zones agricoles périphériques.

COUPLAGE TÉLÉDÉTECTION ET ANALYSE SOCIO-ÉCONOMIQUE (BURKINA FASO)

L'innovation de ce dernier volet réside dans la capacité à faire dialoguer des données satellitaires (Sentinel-2) avec des données d'enquêtes de terrain (EHCVM). Cette approche permet de vérifier si les vulnérabilités environnementales perçues depuis l'espace se traduisent par une précarité économique pour les ménages au Burkina Faso.

Objectif Stratégique et Consigne : L'objectif était de générer une batterie d'indices spectraux par département pour la période 2021-2022, puis de fusionner ces indicateurs biophysiques avec la base de données de l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM). Il s'agissait de créer une base de données hybride capable d'expliquer les niveaux de vie par des facteurs environnementaux.

Inputs et Synergie des Données :

- **Imagerie Satellite :** Collection *Sentinel-2 Surface Reflectance* (MSI).
- **Données d'Enquêtes :** Micro-données EHCVM (Burkina Faso).
- **Référentiel Territorial :** Niveau 2 de la base FAO GAUL (Départements).

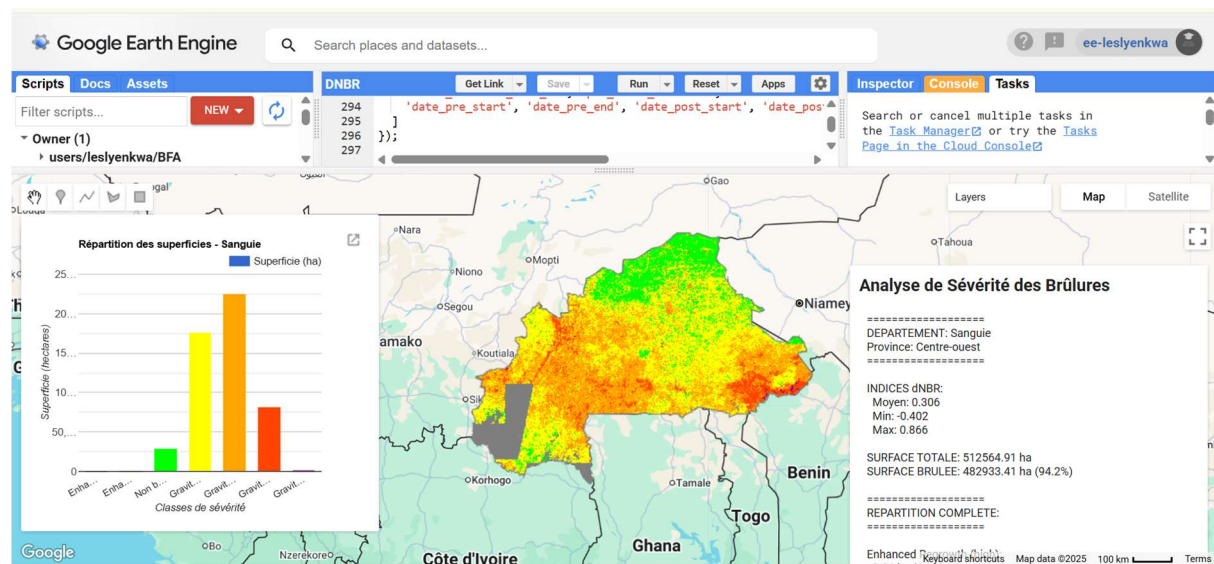
Méthodologie : La "Data Fusion" Spatiale : La démarche a été divisée en trois phases techniques distinctes :

1. **Calcul d'indices spectraux ciblés :** Nous avons implémenté le **NDVI** (Vigueur de la végétation), le **dNBR** (Sévérité des brûlis pour détecter les zones de feux de brousse), le **NDBI** (Zones bâties) et le **NDWI** (Ressources en eau).
2. **Statistiques Zonales et Nettoyage :** Un masquage rigoureux des nuages (via la couche SCL) a été appliqué avant d'agréger les valeurs moyennes de ces indices par département.
3. **Le "Merge" Statistique :** À l'aide du langage **R**, nous avons effectué une jointure entre les fichiers CSV exportés de GEE et la base EHCVM. Cette fusion a été réalisée sur la base des codes géographiques départementaux communs aux deux sources.

Défis Techniques et Arbitrages : Le défi majeur a été la gestion de la temporalité. Les données Sentinel-2 sont continues, tandis que l'enquête EHCVM est ponctuelle. Nous avons dû stabiliser les indices sur une fenêtre temporelle cohérente (juillet 2021 - juillet 2022) pour

correspondre au calendrier de collecte de l'enquête. Un autre obstacle a été l'harmonisation des noms de départements, qui différaient parfois entre le référentiel FAO et la base EHCVM, nécessitant un "fuzzy matching" ou une correction manuelle systématique.

Outputs :



Leçon d'Ingénierie des Données : Ce projet démontre que la télédétection est un complément indispensable aux statistiques classiques. En fusionnant ces données, nous pouvons désormais répondre à des questions complexes : *Les départements ayant subi les feux de brousse les plus sévères (dNBR élevé) sont-ils ceux où le score de consommation alimentaire est le plus faible ?* Cette approche permet de passer d'une observation passive à une **analyse causale** entre environnement et pauvreté.

CONCLUSION GÉNÉRALE ET PERSPECTIVES DÉCISIONNELLES

Ce cycle de travaux pratiques a permis de traverser les frontières techniques et géographiques de l'Afrique subsaharienne, du Tchad au Burkina Faso, en passant par l'Ouganda et le Cameroun. Au-delà de l'apprentissage des syntaxes R, Python ou JavaScript, cette expérience a révélé la puissance de la **Statistique Exploratoire Spatiale (SES)** comme outil de diagnostic et de prospective.

Un changement de paradigme technique

L'utilisation de **Google Earth Engine** a marqué un tournant dans notre approche de la donnée. En déplaçant le calcul du local vers le Cloud, nous avons pu manipuler des pétaoctets de données satellitaires (Sentinel-2, Landsat) et démographiques (WorldPop) à l'échelle de nations entières. Le passage d'un TP à l'autre a illustré une progression logique : de la simple mesure de distance (Accessibilité au Tchad) à la compréhension de phénomènes biophysiques complexes (Indices spectraux au Burkina Faso) et de dynamiques temporelles (ODD 11.3.1 au Cameroun).

La donnée spatiale au service du développement

Chaque pôle d'analyse a apporté une réponse concrète à des problématiques de développement :

- **Équité sociale** : En quantifiant la population hors-zone au Tchad, nous avons identifié les leviers d'une meilleure justice territoriale.
- **Souveraineté alimentaire** : En isolant les terres arables en Ouganda par filtrage multi-critères, nous avons fourni un outil de planification agricole durable.
- **Durabilité urbaine** : En calculant le ratio \$LCRPGR\$ au Cameroun, nous avons objectivé l'étalement urbain, permettant d'anticiper les besoins en infrastructures futures.
- **Analyse multidimensionnelle** : En fusionnant l'imagerie satellite avec les enquêtes **EHCVM**, nous avons ouvert la voie à une compréhension hybride de la pauvreté, mêlant conditions de vie et santé environnementale.

Défis globaux et leçons apprises

Le principal défi transverse a été l'**interopérabilité**. Faire dialoguer des formats disparates (Shapefiles, GeoTIFF, CSV), des résolutions différentes (100m vs 1km) et des nomenclatures administratives parfois divergentes a exigé une rigueur méthodologique constante. Nous en tirons une leçon fondamentale : la qualité de l'analyse spatiale dépend moins de la sophistication de l'algorithme que de la préparation et de l'harmonisation minutieuse des données d'entrée.

En conclusion, ces travaux démontrent que le statisticien moderne ne peut plus se contenter de tableaux de données “a-spatiaux”. L'intégration de la dimension géographique permet de donner un visage aux chiffres et une adresse aux problèmes. Ce rapport constitue désormais un socle de compétences prêt à être mobilisé pour accompagner les politiques publiques vers une gestion du territoire plus intelligente, résiliente et inclusive.

