

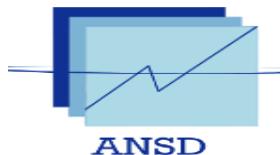
REPUBLIQUE DU SENEGLAL

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



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



STATISTIQUE EXPLORATOIRE SPATIALE

Résumé des travaux pratiques

Sous la supervision de

M. Aboubacar HEMA

Research Analyst - International Food Policy Research Institute (IFPRI)

Rédigé par

Sènou Michel-Marie Trésor TEVOEDJRE

Elève ingénieur statisticien économiste

Année académique 2025-2026

TABLE DES MATIERES

INTRODUCTION.....	I
1. TP 0.....	II
2. TP 1.....	III
3. TP 2.....	IV
4. TP 3.....	V
5. TP 4.....	VI
6. TP 5.....	VII
7. TP 6.....	VIII
8. TP Individuels.....	IX
CONCLUSION	X

INTRODUCTION

Le cours de Statistique exploratoire spatiale a pour objectif de présenter la statistique descriptive spatiale, la cartographie avec divers logiciels (R, Google Earth Engine, Python, Python API Google Earth Engine), les matrices de voisinage ainsi que le calcul des autocorrélations spatiales.

Le cours s'est déroulé suivant une approche très pratique afin de permettre aux étudiants de facilement appréhender les divers concepts enseignés et de s'exercer sur des projets concrets et utiles. Le présent document fournit un résumé des divers travaux pratiques effectués au cours des trente (30) heures de cours.

Chacun des travaux est présenté suivant le format :

- Consigne ;
- Inputs ;
- Etapes suivies et méthodes utilisées ;
- Défis ;
- Leçon principale tirée.

Cette structure permet d'offrir une vision transparente du processus d'apprentissage et de la progression technique réalisée durant ce semestre.

1. TP 0

Présentation de Google Earth Engine

Groupe 2

Logiciels : Canva & Powerpoint

Pays : Néant

1.1 Consigne

Il était question de réaliser une présentation sur la plateforme Google Earth Engine pour l'analyse de données géospatiales massives.

1.2 Inputs

La documentation provenait essentiellement des références fournies dans le repo github mis en place par le professeur dans le cadre du cours ainsi que de quelques sources extérieures.

1.3 Etapes suivies et méthodes utilisées

Le cheminement s'est structuré autour de la découverte et de l'appropriation de l'écosystème Google Earth Engine (GEE):

- Initialisation et accès : Création et configuration d'un compte Google spécifique pour obtenir l'approbation d'accès à la plateforme ;
- Exploration de l'interface : Différents assets, éditeur de scripts et console interactive ;
- Découverte du langage : Syntaxe JavaScript pour manipuler les objets spécifiques à GEE (Image, ImageCollection, Feature).

1.4 Défis

Cette expérience a présenté des obstacles techniques et conceptuels significatifs, particulièrement lors de la phase de démarrage :

- Complexité de l'inscription et de configuration initiale ;
- Surcharge informationnelle de la plateforme : Densité des fonctionnalités qui sont nombreuses ;
- Barrière linguistique (JavaScript) : Le passage à la programmation en JavaScript a constitué un défi majeur.

1.5 Leçon principale tirée

La leçon personnelle que nous tirons est que le Cloud Computing Géospatial (particulièrement GEE) révolutionne l'analyse spatiale en permettant l'accès à une puissance de calcul massive.

2. TP 1

Analyse Spatiale du Paludisme au Sénégal (2000-2024)

Groupe 2

Logiciels : R

Pays : Sénégal

2.1 Consigne

Il était question d'analyser et de cartographier l'évolution spatio-temporelle du taux d'incidence du paludisme au Sénégal sur une période de 25 ans (2000-2024) en utilisant le logiciel R.

2.2 Inputs

Les données utilisées comprenaient des couches raster (fichiers .tif) provenant du Malaria Atlas Project pour l'incidence annuelle, ainsi que des données administratives (Shapefiles) du Sénégal issues de la base GADM.

2.3 Etapes suivies et méthodes utilisées

Le travail s'est structuré comme suit :

- Inspection : Vérification des métadonnées des rasters (résolution, nombre de bandes, emprise) et des shapefiles ;
- Harmonisation : Alignement des systèmes de projection (CRS) pour garantir la superposition parfaite des couches ;
- Visualisation : Production de cartes thématiques statiques (tmap, ggplot2) et développement d'une application Shiny pour une exploration interactive.

2.4 Défis

Le principal défi de cette première manipulation était :

- Harmonisation des données : S'assurer que les données matricielles (rasters) et vectorielles (shapefiles) se superposent parfaitement (problèmes de CRS).

2.5 Leçon principale tirée

La leçon principale est la nécessité d'une vigilance absolue sur la superposition des données. L'inspection systématique des fichiers (projctions, résolution, nombre de pixels) doit être une étape obligatoire avant toute analyse, car la moindre incohérence technique fausse l'intégralité des résultats cartographiques.

3. TP 2

Analyse des Distances entre Infrastructures et Aires Protégées au Sénégal

Groupe 2

Logiciels : GEE JavaScript

Pays : Sénégal

3.1 Consigne

Nous avions ici effectué l'analyse spatiale de la proximité entre diverses infrastructures (écoles, hôpitaux, villes, etc.) et les aires protégées du Sénégal, notamment en calculant des statistiques de distance et en identifiant les points situés dans un rayon de 10 km.

3.2 Inputs

Les données ont été importées comme Assets dans le projet Cloud GEE :

- Limites administratives : Niveaux 0 à 4 issus de GADM.
- Aires protégées : Données vectorielles (points et polygones) de Protected Planet.
- Infrastructures : 9 catégories de points extraites d'OpenStreetMap (OSM).

3.3 Etapes suivies et méthodes utilisées

Le script JavaScript a été structuré pour automatiser :

- Prétraitement : Chargement des collections d'éléments (FeatureCollections) et filtrage géographique sur le Sénégal ;
- Analyse de distance : Création de rasters de distance euclidienne pour chaque type d'infrastructure par rapport aux aires protégées ;
- Géo-traitement : Génération d'une zone tampon (buffer) de 10 km autour des aires protégées et intersection avec les points d'infrastructures ;
- Statistiques et Export : Calcul des distances moyennes, minimales et maximales, puis programmation des tâches d'exportation des résultats vers Google Drive.

3.4 Défis

Les principaux défis étaient :

- Complexité de JavaScript ;
- Volume d'exportation : Gérer l'exportation simultanée de plus de 20 fichiers de résultats via l'onglet Tasks.

3.5 Leçon principale tirée

La leçon majeure de ce travail est la mise en évidence de la supériorité de Google Earth Engine pour le traitement de données massives par rapport à un logiciel local comme R.

4. TP 3

Accès des populations aux infrastructures.

Groupe 2

Logiciels : GEE JavaScript

Pays : Sénégal

4.1 Consigne

L'objectif était de concevoir une application interactive sous Google Earth Engine pour analyser l'accessibilité des populations aux services de base (santé, éducation, transport, eau) et quantifier la part de la population non desservie selon différents rayons d'action.

4.2 Inputs

Le projet a mobilisé une grande diversité de sources de données :

- Données de population : Rasters de population totale et de densité ;
- Santé : Points géolocalisés des hôpitaux, cliniques et pharmacies ;
- Éducation : Localisation des écoles, lycées, collèges et universités ;
- Transport et Ressources : Réseaux routiers (bitumés ou non), voies ferrées et cours d'eau ;
- Limites : Découpage administratif par régions.

4.3 Etapes suivies et méthodes utilisées

Le développement s'est concentré sur la création d'un outil d'aide à la décision :

- Calcul de couverture : Utilisation de la fonction buffer() pour créer des zones d'influence autour des infrastructures (ex: 5km, 10km, 20km) ;
- Analyse spatiale croisée : Utilisation de reduceRegion() pour croiser les zones tampons avec les rasters de population et calculer le nombre d'habitants hors zone de couverture ;
- Interface Utilisateur (UI) : Création de panneaux latéraux (ui.Panel), de boutons et de légendes interactives pour permettre à l'utilisateur de filtrer les données.

4.4 Défis

Le principal défi était :

- Gestion de l'interactivité : Synchroniser l'affichage des couches avec les éléments de l'interface utilisateur pour que la carte reste fluide.

4.5 Leçon principale tirée

JavaScript dans GEE permet de dépasser le rôle de simple cartographe pour devenir développeur d'applications géospatiales. Le professeur a d'ailleurs souligné l'aspect aide à la décision de ces cartes interactives.

5. TP 4

Identification des Terres Arables au Kenya

Groupe 2

Logiciels : R

Pays : Kenya

5.1 Consigne

L'objectif était d'identifier les terres arables disponibles au Kenya en croisant des données de zones cultivées avec plusieurs masques d'exclusion (pentes, zones urbaines, eaux permanentes et aires protégées).

5.2 Inputs

Le projet a nécessité une phase de collecte de données multi-sources :

- Zones cultivées : GFSAD30 (3 granules mosaïqués) ;
- Topographie : DEM Copernicus GLO-30 (pour le calcul des pentes) ;
- Occupation du sol : JRC Global Surface Water (eaux) et Global Man Impervious Surface (zones urbaines) ;
- Limites & Protection : GADM (Kenya) et WDPA (Aires protégées).

5.3 Etapes suivies et méthodes utilisées

Le travail a été séparé entre le Cloud et le local pour plus d'efficacité :

- Prétraitement (GEE) : Mosaïquage des granules GFSAD, calcul des pentes en pourcentage et extraction des données d'imperméabilité et d'eaux permanentes ;
- Traitement spatial (R) : Reprojection des surfaces imperméables (UTM vers WGS84). Application des masques d'exclusion (ex: pente > 15%, imperméabilité > 10%). Intersection avec les limites administratives.

5.4 Défis

Les défis majeurs de ce TP ont été :

- Lenteur de R : La manipulation des rasters à haute résolution sur l'ensemble du Kenya s'est avérée extrêmement lente sur une machine locale ;
- Hétérogénéité des données : Nécessité de gérer des formats et des systèmes de projection différents (re-projection de l'UTM vers le WGS84) ;
- Volume des données : La nécessité de mosaïquer plusieurs "granules" pour couvrir tout le pays.

5.5 Leçon principale tirée

La leçon principale tirée de ce TP est l'importance de l'approche hybride entre le Cloud (GEE) et le local (R).

6. TP 5

Analyse spatiale et développement urbain durable

Groupe 3

Logiciels : Python

Pays : Rwanda

6.1 Consigne

L'objectif était d'analyser le LCRPGR (Land Consumption Rate to Population Growth Rate). Il s'agissait de mesurer l'efficacité de l'utilisation des sols par rapport à la croissance démographique au Rwanda entre 2015 et 2025.

6.2 Inputs

Le projet a exploité les données suivantes :

- Données de Population : WorldPop (100m) pour 2015 et 2025 ;
- Surfaces Bâties : Rasters de l'occupation du sol (GHSL - Global Human Settlement Layer) ;
- Limites Administratives : Shapefiles complets du Rwanda (Provinces, Districts, Secteurs).

6.3 Etapes suivies et méthodes utilisées

Le travail a consisté à :

- Analyse Spatiale avec Python : Utilisation de Rasterio et GeoPandas pour extraire les statistiques de population et de bâti à travers 4 niveaux administratifs ;
- Calcul de l'indicateur : Programmation de la formule du LCRPGR pour identifier les zones d'étalement urbain ($LCRPGR > 1$) ou de densification ($LCRPGR < 1$) ;
- Développement Web : Création de l'interface avec le framework Dash ;
- Visualisations Avancées : Intégration de cartes de chaleur (Heatmaps), de graphiques de corrélation (Scatter plots) et de KPIs dynamiques.

6.4 Défis

La principale difficulté était :

- Réussir à rendre le dashboard fluide alors qu'il doit recalculer des statistiques pour des centaines de secteurs administratifs différents à chaque clic.

6.5 Leçon principale tirée

Nous retenons qu'un indicateur complexe comme le LCRPGR ne devient réellement utile pour un décideur que s'il peut être exploré visuellement à différentes échelles.

7. TP 6

Extraction et Fusion de Données Environnementales avec l'Enquête EHCVM

Groupe 3

Logiciels : Python

Pays : Mali

7.1 Consigne

L'objectif était d'enrichir la base de données socio-économique de l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM) en y intégrant des variables environnementales. Il s'agissait d'extraire des indicateurs précis (climat, végétation, pollution) pour chaque grappe de l'enquête et de fusionner le tout dans un fichier unique.

7.2 Inputs

Le projet a croisé deux types de sources hétérogènes :

- Données vectorielles : Les points GPS représentant les grappes de l'enquête EHCVM.
- Données Satellitaires (GEE) :
 - Indices spectraux : Sentinel-2 (NDVI, NDWI, NDBI, BSI, NDMI).
 - Pollution : Sentinel-5P (Aérosols, NO₂, CO, etc.).
 - Climat : Données de Température de Surface (LST).

7.3 Etapes suivies et méthodes utilisées

Le travail s'est concentré sur l'extraction de valeurs et la fusion de bases de données :

- Calcul des Indicateurs (GEE) : Développement de scripts pour calculer les moyennes (annuelles ou saisonnières) des indices biophysiques et des polluants sur la période de l'enquête ;
- Extraction Ponctuelle ("Reduce to Vectors") : Récupération des valeurs de pixels situées exactement sous les coordonnées GPS des grappes EHCVM ;
- Nettoyage et Fusion : Consolidation des différents fichiers de sortie (CSV) contenant la température, la pollution et les indices de végétation en un seul tableau final (EHCVM_indices_fusion).

7.4 Défis

Le principal défi de ce TP :

- **Gestion des "Null"** : Traiter les cas où les nuages ou l'absence de passage satellite empêchaient d'obtenir une valeur pour une grappe donnée.

7.5 Leçon principale tirée

Nous avons appris comment transformer des images brutes en données tabulaires.

8. TP Individuels

8.1 Analyse de l'incidence du paludisme

Logiciels : GEE

Pays : Togo

8.1.1 Consigne

Découvrir la manipulation de données géospatiales (rasters et shapefiles), effectuer des statistiques zonales et produire des visualisations cartographiques épidémiologiques.

8.1.2 Inputs

Les données utilisées :

- Données d'incidence : Provenant du Malaria Atlas Project, couvrant 25 années (2000-2024) sous forme de fichiers rasters (.tiff).
- Données administratives : Limites du Togo issues de la base GADM (niveaux 0, 1 et 2).

8.1.3 Etapes suivies et méthodes utilisées

Le script développé dans GEE permet de :

- Importer et préparer les données : Shapefiles et des couches d'incidence ;
- Calculer des statistiques : Génération de statistiques nationales.
- Visualiser les résultats ;
- Exporter les données.

8.2 Matrices de voisinage et Indices de Moran

Logiciels : R

Pays : USA

8.2.1 Consigne

Ce travail pratique, codé en R, se concentre sur la compréhension de la structure spatiale des données et la mesure de l'autocorrélation.

8.2.2 Inputs

Le script utilise des données spécialisées :

- Jeux de données classiques : baltimore (prix des maisons avec coordonnées X,Y) et columbus (données de criminalité au niveau des sections de recensement).
- Données géographiques : Shapefiles de la Caroline du Nord intégrés dans le package sf.

8.2.3 Etapes suivies et méthodes utilisées

Le travail est divisé en trois phases majeures :

- **Construction de matrices de voisinage complexes ;**
- **Mesure de l'autocorrélation (Indice de Moran) ;**
- **Visualisation et Diagnostics.**

CONCLUSION

La réalisation de ces travaux pratiques marque l'aboutissement d'un parcours d'apprentissage intensif en Statistique Exploratoire Spatiale (SES). Sous la supervision de M. Hema, ce cours s'est distingué par une approche résolument pratique, transformant des concepts théoriques complexes en solutions concrètes pour le développement de l'Afrique.

Ce cours a démontré que la donnée spatiale est un levier de décision stratégique. Grâce à l'application directe des concepts, nous sortons de ce module non seulement avec des compétences techniques, mais surtout avec la capacité d'apporter des réponses chiffrées et cartographiées aux défis majeurs de nos territoires.

La SES n'est plus une option pour l'Ingénieur Statisticien Économiste (ISE) moderne, c'est un langage essentiel pour décrypter le monde de demain.